VARIABLE *QUE* IN THREE FRANCOPHONE REGIONS

by

JULIA STEELE JOSEPHS

(Under the Direction of Diana L. Ranson)

ABSTRACT

According to prescriptive grammar, the French complementizer *que* is obligatory. However, the absent form has been attested in more than one variety of French in informal spoken language. A combination of social, linguistic, and stylistic factors has been shown to condition *que* realization (Cedergren & Sankoff 1974; Connors 1975; Sankoff 1980; Martineau 1985; Warren 1994; Blondeau & Nagy 2008; Liang et al. 2021). Yet, there is disagreement among scholars as to which factors condition the absent variant and to what extent. This dissertation examines the absent complementizer in Europe, Quebec, and Africa in the early 21st century. The objectives are to determine how widespread absent *que* is in each region, the conditioning factors, and the trajectory of the phenomenon in Quebec since the first studies in the 1970s and 1980s.

Using two online corpora of informal spoken language, the Corpus of French spoken in Quebec (CFPQ) and Phonology of Contemporary French project (PFC), all examples where the complementizer could occur with 25 selected matrix verbal elements were extracted. Generalized linear mixed-effects models were fit using the *lmer* function in the *lme4* package (Bates et al. 2015) of R (R Core Team 2022). Absent *que* was found to exist in each region at a rate of at least 20%. None of the social factors was statistically significant. The lexical identity of the matrix

verbal element, the phonological and syntactic contexts following the complementizer, and intervening material in the matrix clause were significant for at least one model. It was determined that the matrix verb is the most important conditioning factor followed by the phonological context. The results indicate that the overall absent *que* rate has not changed since previous studies in Quebec. However, individual speaker use suggests that linguistic change may be occurring.

This study is the first to compare absent *que* rates in three geographically separated francophone regions. It is one of two studies to use mixed-effects models, which have several advantages over the methods used in previous studies. Additionally, this study proposes that syntactic analogy plays a role in *que* realization.

INDEX WORDS: Burkina Faso, Central African Republic, discourse marker, France, French complementizer, generalized linear mixed-effects models, Côte d'Ivoire, Quebec, sociolinguistics, Switzerland, variation

VARIABLE *QUE* IN THREE FRANCOPHONE REGIONS

by

JULIA STEELE JOSEPHS

B.A., Berry College, 2004

M.A., University of North Carolina, 2007

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

VARIABLE *QUE* IN THREE FRANCOPHONE REGIONS

by

JULIA STEELE JOSEPHS

Major Professor:	Diana L. Ranson
Committee:	Margaret E. L. Renwick
	L. Chad Howe

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

ACKNOWLEDGEMENTS

This dissertation represents the collective effort of several people without whose help it would not have been possible to complete. First, I would like to thank my major professor, Dr. Ranson, for all the hours she has invested in me and this work over the past few years. From my first meeting with her she has taken an interest not only in my academic pursuits but also in me as a person. I am grateful for her expertise, guidance, and mentorship throughout my doctoral journey. She has always gone above and beyond to help and support me, especially when unexpected circumstances arose. Her passion for French and linguistics is contagious, and I am thankful our paths crossed.

My committee members, Dr. Renwick and Dr. Howe, have gladly shared their knowledge to help this project come to fruition. The idea for this project came from Dr. Howe's class that I took my first year at UGA. I am grateful for his encouragement to take a term paper further to see where it led. His knowledge of sociolinguistics has been a tremendous help. Dr. Renwick made the statistical analysis possible by first showing me how to use R for linguistic studies and by helping me troubleshoot problems along the way.

I am grateful for the translation assistance provided by Dr. Ranson, Dr. Catherine Jones, and Dr. Julia Hernandez that helped add a new dimension to this study. I am also very appreciative of all the work that Jamie Kaiser did to help complete data collection. Her initiative and desire to learn outside the classroom are commendable. This project would not have been possible without the teams who built the PFC and CFPQ corpora. They spent countless hours conducting and transcribing interviews that were then made available to the public.

To my friends and peers who encouraged and supported me in various ways, I am so thankful. Tong, María, Betsy, Bénédicte, Kelly, Camila, and Jane, thank you for the pep talks, linguistic conversations, walks, picnic lunches, and words of encouragement. They meant more than you probably know.

From an early age my parents always supported me in academic endeavors. To Daddy, I wish I could show you this dissertation because you were always my biggest cheerleader. I know how proud you were that I was pursuing a PhD even though you did like to crack jokes about "fudds." Thank you for your curiosity and for showing an interest in what I was studying. Mama, thank you for the ways you have supported me during this process. To Stephanie, you have always believed in me and shown your support and encouragement. Thank you for listening to me talk about my struggles, for reminding me of what is true, for celebrating my successes, and for loving me. Jamal, I barely knew you when I began this endeavor, but you have always supported me even though it meant sacrifices on your part. Thank you for never complaining when I needed to work and for providing comic relief at just the right times. I am grateful for your prayers and your reminders of the bigger picture. I'm looking forward to being able to spend more time with you.

Last but most importantly, this dissertation would not have been possible without the love and grace of God who gave me the ability to learn and apply my skills. May the glory be His.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

Chapter 1

Variable *Que* from a Diachronic and Cross-linguistic Perspective

1.1. Introduction

The present study examines the complementizer clause, which is headed by the complementizer *que* 'that' and followed by a minimum of a subject and verb. The complementizer clause usually serves the function of a complement but can be a subject, although the complement function is the only one included in this study. The complementizer is required according to Standard French[1], but, as Bauche (1928:143) notes, in informal language the complementizer is sometimes absent.

Exactly which factors condition the absence of *que* is of primary interest for the present study. Two examples where the complementizer is absent can be seen in (1)[2] and (2) followed by two examples where the complementizer is present in (3) and (4). All four examples are from the Corpus of French spoken in Quebec (CFPQ, Corpus de français parlé au Québec, Dostie 2016), one of the corpora used for this study.

(1) moi j'ai trouvé ∅ c'était bon (QF74)

'me I thought ∅ it was good'

[1] Standard language refers to the set of prescriptive norms as established by the Académie Française, the council charged with deciding matters pertaining to the French language. Even though the majority of French users do not adhere to the rulings by the Académie Française, they have a remarkably solidified and unified understanding of the standard variety of the language (Lodge 1993:3).

[2] All examples are taken verbatim from the source cited. Glosses of Latin, Old French, and Modern French are mine unless otherwise indicated. Glosses of other languages are by the author cited unless otherwise indicated.

(2) je suis pas certain je pense ∅ je vas aller à Sherbrooke (QM80)

'I am not sure I think ∅ I am going to go to Sherbrooke'

(3) je trouve **que** c'est pas super grave de les laisser faire leur examen plus tard (QF24)

'I think **that** it's not a big deal to let them take their exam later'

(4) tu penses **qu'**il va l'accepter (QM27a)

'do you think **that** he is going to accept it'

After learning about *that*-deletion in English I became curious about the possible existence of *que*-deletion in French, a language with which I was already familiar. Even though I had never noticed this phenomenon in French, initial research indicated that it had indeed been documented, particularly among French speakers in Quebec and Ontario. As someone who is intrigued by regional variation, I set out to uncover the similarities and differences between France and Quebec. I also wondered about the trajectory of this phenomenon in Quebec since the first studies were published in the 1970s and 1980s. The more research I conducted, the more fascinated I became, and the more I desired to understand the intricacies of this phenomenon. The result of this quest is laid out in the pages that follow.

This phenomenon is referred to by several different names such as *que*-deletion, null complementizer, complementizer drop, complementizer omission, and complementizer absence. This study will refer to the phenomenon as the variable complementizer or variable *que*, which is either present or absent. Its absence will be indicated in examples and in shortened form by the symbol ∅.

Studies conducted on the variable French complementizer in the speech of the late 20th century focused primarily on North American varieties of French (Cedergren & Sankoff 1974;

Connors 1975; Sankoff 1980; Martineau 1985; Warren 1994; Blondeau & Nagy 2008[3]; Liang et al. 2021[4]). The first two studies combined quantitative and qualitative methods while the others focused on quantitative methods. The studies by Martineau (1985), Warren (1994), Blondeau and Nagy (2008), and Liang et al. (2021) took a variationist approach. Some of these studies examined only linguistic factors while others also looked at social factors. These studies revealed that the complementizer was sometimes absent in informal speech in Quebec and Ontario. More recent studies have taken a theoretical approach, sometimes incorporating a quantitative approach. They explored the variable complementizer through different lenses such as semantics and pragmatics. The present study analyzes the French complementizer within the variationist framework and includes both a quantitative and qualitative analysis.

Given the long-documented history of absent *que* in Quebec, the French of this region was selected for analysis along with the French spoken in Europe (France and Switzerland) and Africa (Burkina Faso, Central African Republic, and Côte d'Ivoire). Europe was chosen because the French language originated there and because the headquarters of the Académie Française, the governing body of Standard French, founded in 1634, is in Paris. French colonizers arrived in Africa in the 17th century where it is estimated that 115 million people speak French as a native or second language (Lafage 2002; Stein-Smith 2022). The variable complementizer has been documented in the spoken language of Côte d'Ivoire (Boutin 2007). The CFPQ and the Phonology of Contemporary French project or PFC (Phonologie du Français Contemporain, Durand et al. 2002) provide data that allow for quantitative and qualitative analyses of the informal spoken French in the three chosen regions.

[3] Interviews for this study were conducted in the mid-1990s.
[4] This study analyzed data from the *Montréal 84* corpus.

1.2. Research Questions

While the previous studies on the absence of the French complementizer provided insight into this phenomenon, some questions remain. Quantitative studies in the 21st century are lacking for all three regions being considered. There is also a lack of consensus on which factors condition the absent form of the complementizer. Most quantitative studies that have analyzed this phenomenon focused on Quebec. However, studies from the 21st century have not provided the information necessary to determine the trajectory of absent *que* since the extensive studies of the 1970s and 1980s. The following three research questions will guide the present study.

1. How widespread is use of absent *que* in Europe, Quebec, and Africa?

Previous research indicated that the overall rate of absent *que* in Quebec was 23% (Sankoff 1980), 32% (Martineau 1985), and 24.8% (Liang et al. 2021) in the 1970s and 1980s. Boutin (2007) found a rate of 8.8% in Côte d'Ivoire. Even though Bauche (1928) documents the absent complementizer in Europe, he gives no indication of the overall rate. Therefore, the present study seeks to determine the overall absent *que* rate in each of the three regions based on informal speech collected from 2001 to 2012. The overall absent *que* rate will provide an indication of language change in Quebec and will provide a baseline for Europe and Africa that will make it possible in the future to discern language change. It is hypothesized that examples of absent *que* will be found in each region and that Quebec will have the highest rate based on its long-documented history of the phenomenon. Whereas most French speakers in Quebec and Europe (France and western Switzerland specifically) acquired French as their first language, francophone Africa is comprised primarily of second language French speakers (Gadet 2011). It was noted by Gadet (2011) that French in Africa has gained new emergent processes. Also, French is spoken alongside several national and regional languages. This linguistic difference

and potential for influence from neighboring languages make it likely that Africa will display a higher rate of Non-Standard French. Since the absent complementizer is not part of Standard French, it is hypothesized that Africa will have a higher absent *que* rate than Europe, the birthplace of French and the seat of the governing body of Standard French grammar, the Académie Française.

2. What are the factors that condition absent *que* and how do they compare to those found in previous studies[5] (Cedergren & Sankoff 1974; Connors 1975; Sankoff 1980; Martineau 1985; Warren 1994; Blondeau & Nagy 2008; Liang et al. 2021)?

In their studies on the absent complementizer, Connors (1975) and Sankoff (1980) considered only linguistic factors, whereas Cedergren and Sankoff (1974), Martineau (1985), Warren (1994), Blondeau and Nagy (2008), and Liang et al. (2021) considered linguistic and social factors. There is some overlap in the results of these studies, but there is not a consensus on the role of each factor. The present study considers both social and linguistic factors. Based on a preliminary study I conducted and the results of the previous studies, it is anticipated that absent *que* will be linguistically conditioned, and that the primary conditioning factors will be the matrix verb and the phonological context following the complementizer.

3. What has been the trajectory of absent *que* in Quebec since the 20th century?

As previously stated, quantitative studies on absent *que* in Quebec date to the late 20th century. However, its status in the 21st century has not been studied. Quantitative data from a comparable time period in Europe and Africa are not available. The present study seeks to determine the overall absent *que* rate in the early 21st century in Quebec. This may in turn shed light on some

[5] Boutin (2007) is not included in this comparison since the purpose of her work is to document the linguistic variation in Côte d'Ivoire for the subordination markers *que* and *de* rather than to quantify the effect of social and linguistic factors.

of the results of previous studies. For example, do the age results of Martineau (1985) and Liang et al. (2021) indicate age grading or generational change? It is hypothesized that the rate of the absent complementizer in Quebec will have either remained stable or increased.

1.3. A Brief History of Absent *Que* in French: Parataxis or Hypotaxis?

1.3.1. Parataxis and Hypotaxis in Latin

The concepts of parataxis, or juxtaposition, and hypotaxis, or subordination, are crucial to the analysis of absent *que* since instances where there is no complementizer could be examples of parataxis or, on the other hand, hypotaxis with an absent *que*. Parataxis is defined by The *Oxford English Dictionary Online* as "the placing of propositions or clauses one after another, without indicating by connecting words the relation (of coordination or subordination) between them" (OED online 137669). Example (5) below from *La Chanson de Roland* offers an illustration of parataxis since no linguistic element occurs between *E or sai bien* and the following clause (Ritchie 1907:121).

(5) E or sai bien – n'avons guaires a vivre

'Et maintenant je sais bien – nous n'avons pas longtemps à vivre.'

'And now I surely know – we don't have long to live'

(Ritchie 1907:121, from *La Chanson de Roland*, unknown, 11th/12th century)

Subordination is defined as "the dependence of one clause or unit upon another; the construction of sentences using subordinate clauses" (OED online 192884). Example (6) provides an illustration of a subordinate clause following *j'ai prouvé* 'I proved' introduced by the complementizer *que*.

(6) Je vous ai prouvé **que** c'était dans le journal (Blanche-Benveniste 1989:62)

'I proved to you **that** it was in the newspaper'

Evidence of both parataxis and hypotaxis can be found in Latin, French's predecessor (Antoine 1899; Ritchie 1907). Parataxis was common in Colloquial Latin, while subordination, which came from the Greeks, appeared only in literary use and in the speech of the most cultivated members of society (Ritchie 1907:122). Even after the introduction of a present subordination marker, parataxis continued to be used and was even simultaneously used with hypotaxis in writings by the same author (Antoine 1899: 33-36). Example (7) below provides an illustration of parataxis in Latin, where two clauses are juxtaposed without a linguistic element between them. The following example, (8), shows hypotaxis in Latin, where *QUOD* overtly links the two clauses.

(7) Audivi. – Pater tuus aegrotat. (Antoine 1899: 28) [parataxis]

'J'ai entendu. – Ton père est malade.'

'I heard. – Your father is sick.'

(8) Audivi **QUOD** pater tuus aegrotat. (Antoine 1899: 28) [hypotaxis]

'J'ai entendu **que** ton père est malade.'

'I heard **that** your father is sick.'

1.3.2. Parataxis and Hypotaxis in Old French

Parataxis was very frequent in the last few centuries of the Roman Empire as well as in Old French (Ritchie 1907:122; Schneider and Glikman 2015:177). According to Arteaga (2009:31), examples of absent *que* in Old French indicate subordination, which is evidenced by the mood of the second clause being determined lexically by the verb in the first clause. Martineau (1993:80)

echoes this idea providing examples from Old French that she considers to be examples of subordination even though there is no *que* present, such as example (9).

(9) L'Empereres mout liez en fu; / Nouveles leur ha demandées / Comment les choses sunt alées, / Se li pelerin voir disoit Il dient ∅ de rien ne mentoit

'L'empereur en fut très content. Il leur a demandé des nouvelles (de) comment les choses sont allées. Si le pèlerin disait la vérité. Ils disent ∅ de rien ne mentait.'

'The emperor was very happy. He asked them for news (of) how things have gone. If the pilgrim was telling the truth. They say ∅ he wasn't lying about anything.'

(Martineau 1993:80, from *Roman du Saint Graal*, unknown, ca. 1450)

Environments where *que* is obligatory according to Modern Standard French but absent in Old French texts are considered by Ritchie (1907:121-122) to be parataxis. Example (10), which reproduces example (5) above, illustrates the grammatical independence of two expressions whose relationship is evident and reinforced by their side-by-side positioning. Instead of speaking of omitting *que* in Old French, he argues that instances where *que* is absent should simply be called coordination.

(10) E or sai bien – n'avons guaires a vivre

'Et maintenant je sais bien – nous n'avons pas longtemps à vivre.'

'And now I surely know – we don't have long to live'

(Ritchie 1907:121, from *La Chanson de Roland*, unknown, 11th/12th century)

Considering that parataxis was frequently used in Old French, it is possible that instances where there is no complementizer in Old French could be examples of parataxis. On the other hand, they might be examples of hypotaxis with an omitted *que*. As Moignet (1976:339) points

out, it is sometimes impossible to distinguish juxtaposition from subordination. It is also noteworthy that punctuation was sometimes used in place of *que* in Old French, such as the comma after *sachiez* in example (11). According to Foulet (1961:333), the punctuation "masks" the fact that a complementizer has been deleted. Thus, whether the complementizer has been deleted or was never there is uncertain.

(11) Mes, sachiez, je ne puis pas dire ce que volez que je vous die.
'Mais, sachez, ∅ je ne peux pas dire ce que vous voulez que je vous dise.'
'But, know, ∅ I can't say what you want me to say.'
(Foulet 1961:333, from *Chastelaine de Vergi*, unknown, 13th century)

1.3.3. Parataxis and Hypotaxis in Modern French

Texts from Latin and Old French highlight the use of parataxis and hypotaxis for many centuries and the difficulty in distinguishing the two. Not surprisingly, Modern French also attests to the use of these structures and the ongoing difficulty in distinguishing them. Examples (12) and (13), drawn from one of the corpora used for the present study, illustrate the difficulty in determining whether parataxis or hypotaxis with an omitted complementizer is the structure in use. In (12) *je trouve* could be used as a discourse marker in which case this would be an example of parataxis. On the other hand, it could be the matrix clause of the proposition that follows it, in which case it would represent hypotaxis with the absent complementizer. Similarly, in example (13), *tu vois* could be a discourse marker, in which case the structure would be parataxis. It could also be the matrix clause of the following proposition, which would make it hypotaxis with the absent complementizer. In instances such as these, a prosodic analysis is the only way to determine which it is. Such an analysis would measure the prosodic force of the verb and the last syllable of

the prosodic unit in which it occurs. The force of these two relative to one another would indicate whether it is used as a parenthetical verb or a discourse marker. This idea will be detailed in Section 1.7.

(12) c'est pas le principe de l'immersion je trouve ∅ c'est pas c'est vraiment pas réussi de ce côté-là (QF24)

'it's not the principle of immersion I think ∅ it's not it really isn't successful in that regard'

(13) tu vois ∅ elle est pas capable de décrocher de l'éducation là qu'elle a eu là (QM25a)

'you see ∅ she is not able to detach from the education that she had there'

The present study does not propose to make determinations between parataxis and hypotaxis with an omitted complementizer due to time constraints. Rather, its purpose is to determine how frequently all such structures occur in spoken language in different varieties of French. All instances where *que* is required according to Modern Standard French are included (Grévisse & Goosse 1993). In cases such as (12) and (13), the criteria outlined below in Section 2.3 were used to reduce the number of possible discourse markers included in the data. However, if an example did not meet the criteria for a discourse marker and it was possible for there to be a following complementizer, the example was included. Inevitably there may be some discourse markers included in the data since prosodic information was not examined. The terms present and absent complementizer/*que* are used, although it is understood that there is the possibility that a particular structure could be parataxis, in which case it was not possible to include *que* and so the term absent complementizer/*que* would be inaccurate.

1.4. The Absent Complementizer in Other Languages

Several languages historically related to French provide evidence of variation between the present and absent complementizer. An overview of this evidence situates the variable French complementizer in the broader linguistic context. Since these languages are all descendants of Latin, the absent complementizer could be linked to their shared origin. While the purpose of this work is not to conduct a diachronic study, some diachronic background information is helpful to provide the wider context in which this phenomenon occurs.

1.4.1. Gascon

Gascon, a Romance dialect spoken in Southwest France and in the Val d'Aran, Spain, has not two but three variants of the complementizer (Pusch 2003). In many sub-dialects of Gascon every conjugated verb form is preceded by an enunciative particle (Pusch 2003). The three forms of this particle are a preverbal *que*, a preverbal *e*, and a zero morpheme (Pusch 2003). In his study of the enunciative particle, Pusch (2003) points out that there is a regular distribution. The preverbal *que* is present in main clauses that are affirmative and that carry the primary message, as in example (14) where ENC indicates the placement of the enunciative particle. Preverbal *e* appears in subordinate clauses and interrogative phrases, two environments where there must be some type of reduction of the assertion, since they either express presupposed ideas or facts unknown to the speaker, as illustrated in example (15). Field (1985:81) states that "the entire series of particles can be viewed as a continuum expressing the degree of intensity with which the speaker is putting forward the proposition contained in the sentence." It follows that the zero morpheme is used when there is no speaker commitment to the propositional content. This morpheme is shown in example (16).

(14) cadun **que** pren sus eth lor temps (Pusch 2003:5)

'chacun **ENC** prend (un peu) de son temps'

'each one **ENC** takes (a little) of his time'[6]

(15) adara non podi pas díser [...] quant de centenats **e** n'i avè (Pusch 2003:5)

'maintenant je ne peux pas dire combien de centains il **ENC** y en avait'

'now I can't say how many hundreds there **ENC** were'

(16) cad/cada an ∅ pensi que i a mei e mei de monde (Pusch 2003:5)

'chaque année je ∅ pense il y a de plus en plus de monde'

'each year I ∅ think there are more and more people'

In his study of the enunciative particle in Gascon, Pusch (2003) used the Corpus Occitano-Gascon (COG) to search for examples of the Gascon equivalent of *penser*, *croire*, *dire*, *savoir* 'to know', *comprendre* 'to understand' and *voir* 'to see'. When *penser* and *croire* are used in the first person singular, they are preceded by the zero morpheme in just under half of the tokens and are preceded by *que* in just over half of the tokens. Thus, in roughly half the cases these verbs have an attenuated assertive force. The particle *e* appears just once. In the first person plural, *dire* is only used with the zero morpheme. Similarly, when *dire* is used to introduce indirect discourse in other persons, either the preverbal *e* or zero morpheme is used. For the verbs *savoir*, *comprendre*, and *voir*, Pusch (2003) searched for second-person singular and plural examples and found that they have an enunciative role that is parallel to the first-person form of verbs such as *penser* and *croire*, as asserted by Andersen (1997). In the COG all such forms are preceded by the zero morpheme, thereby confirming that they can have an attenuated assertive force in Gascon, just as *penser* and *croire* can.

[6] English glosses for (14) - (16) are mine.

1.4.2. Spanish and Italian

In Spanish, as in French, *que* is considered to be obligatory according to standard norms (Brovetto 2002). The extent to which it can be used and its frequency in modern spoken Spanish are unclear (Pountain 2015). After reading assertions that the absent complementizer in Spanish is both widespread in spoken language and also restricted to written language, Pountain (2015) found two examples of absent *que* and 434 examples of present *que* in a study he conducted using the Madrid *habla culta* corpus. Thus, he concludes that its status lies somewhere between widespread and restricted to written language.

Some instances of the absent complementizer are marked and associated with formal speech (Brovetto 2002; Pountain 2015). There are three classes of verbs that allow for the absent complementizer: verbs of prepositional attitude (such as *suponer* 'suppose', *dudar* 'doubt' and *parecer* 'seem'), 'lament' verbs (such as *lamentar* 'lament', *preocuparse* 'be worried', *alegrarse* 'be glad' and *sentir* 'be sorry') and volitional verbs/verbs of desire (such as *querer* 'want', *desear* 'desire' and *esperar* 'hope') (Brovetto 2002). Examples (17) - (19) illustrate the use of these classes of verbs.

(17) parece **que**/∅ no le importe nada de eso (Wanner 1981:54)
'it seems **that**/∅ she does not care at all about that'

(18) Lamento **que**/∅ no estés contenta con tu trabajo. (Torrego 1983:562)
'I lament **that**/∅ you are not happy with your job.'

(19) Espero **que**/∅ se solucionen pronto los problemas causados por el huracán. (Brovetto 2002:34)
'I hope **that**/∅ the problems caused by the hurricane will be solved soon.'

As noted by Torrego (1983), the absent complementizer is not allowed in every variety of Spanish. For example, it is allowed in Peninsular, Puerto Rican, and Dominican Spanish but not in Peruvian or Paraguayan Spanish. Additionally, Silva-Corvalán (2008) identifies Eastern Mexican Spanish as allowing this in informal spoken language.

The aforementioned classes of verbs use the subjunctive in the embedded clause, although Brovetto (2002) points out that this is not a condition. Rather, the restriction to these three classes relates to the class of the matrix verb and the semantic nature of the embedded clause. In Spanish, the complementizer can be absent when the embedded proposition relays uncertainty or an irrealis meaning (Brovetto 2002). Typically, the absent complementizer is not allowed with verbs of saying like *decir* 'say', *repetir* 'repeat' and factive verbs like *confesar* 'confess', *jurar* 'swear', and *admitir* 'admit', as illustrated in example (20) (Brovetto 2002). However, if these verbs convey an irrealis meaning, they can appear without *que*. The use of the conditional in (21) conveys an irrealis meaning thereby permitting the absent complementizer.

(20) Dijo **que**/*Ø llegó tarde a la reunión. (Brovetto 2002:34)

'He said **that**/Ø he arrived late to the meeting.'

(21) Me parece **que**/Ø podrían mejorarse algunos (Brovetto 2002:35)

'It seems to me **that**/Ø some (aspects) could be improved.'

Turning to another Romance language, Italian shows a very similar distribution to Spanish with regard to the absent complementizer (Brovetto 2002). It is primarily restricted to the subjunctive but can be absent with future and conditional. Examples (22) - (23b) show utterances in which *che* can be omitted.

(22) Credo **che**/Ø avesse appena comprato i gelati. (Scorretti 1981:35)

'I believe **that**/Ø he had just bought the ice creams.'

(23a) pare ∅ sia successo davvero (Wanner 1981:47)

'it seems ∅ it has really happened'[7]

(23b) pare **che** sia successo davvero (Wanner 1981:47)

'it seems **that** it really happened'

Thus, there is documentation of a variable complementizer in at least three languages related to French.

1.5. Overview of Research on Absent *Que* in French

1.5.1. Early Studies on Absent *Que*

Before the first studies on the absent complementizer were conducted in the 1970s, Bauche (1928:143) observed absent *que* in the spoken language of Parisians citing examples such as *tu veux ∅ je vienne* 'Do you want me to come' (lit, you want ∅ I come), *Faut ∅ je m'en alle*? 'Do I need to leave? (lit. it is necessary ∅ I leave?)' and *Il a dit ∅ i viendrait* 'He said ∅ he would come'. The first study conducted on this phenomenon was by Sankoff et al. (1971), cited by Connors (1975), whose primary results appear in Sankoff (1980). Early studies (Cedergren & Sankoff 1974; Connors 1975; Sankoff 1980; Martineau 1985) focused on speakers in Quebec but also looked at the phenomenon in speakers from Old Mines, Missouri (Connors 1975) and Ontario (Martineau 1985). A few studies published in subsequent decades examined speech in Montreal in the 80s and 90s. Warren (1994) compared her corpus collected in 1981 to that of Sankoff et al. (1971), whereas Liang et al. (2021) reexamined speech that was part of a 1984 corpus. Finally, Blondeau and Nagy (2008) studied Anglophones in Montreal who spoke French as a second language in the early 90s.

[7] English gloss mine.

While these studies analyzed informal spoken language, Martineau (1985) delved deeper into the role of formality in conditioning the absent complementizer. She found that an informal style favored absent *que* with a rate of 37% in informal speech compared to 29% for formal. The choice of matrix verb, which may be related to formality for some verbs, was also an important factor. For example, the verb *croire* 'to believe' is used more often in formal contexts than the verb *penser* 'to think', which has a very similar meaning. The verb itself was found to heavily influence the absent *que* rate. *Penser*, for example, one of the most frequently occurring verbs, had an absent *que* rate of 47%. While it is a very frequent verb, other frequent verbs such as *vouloir* 'to want' do not favor the absent complementizer (19% absent *que* rate). Compared to *penser*, *croire* had a very low absent *que* rate of 10%.

Regarding linguistic factors, the most commonly tested factor was the phonological context. This included both the phonological environment preceding the complementizer, the sound preceding *que*, such as the /s/ in *je pen<u>s</u>e que* 'I think that' as well as the environment following *que*, such as the /i/ in *je pense qu'<u>i</u>l* 'I think that he'. To test this factor, sounds were placed into groups, called sonority groups, based on the degree of constriction of air flow. The degree of constriction corresponds directly to the loudness of a sound "relative to that of other sounds with the same length, stress, and pitch" (Ladefoged 1975:219). Sonority groups are ordered according to the sonority hierarchy, which ranks speech sounds based on syllable well-formedness. While it is recognized that sonority constraints across languages have an almost identical nature, phonologists do not agree on the contents of one universal sonority hierarchy (Clements 1990:290-291). Clements (1990:291) points out that this is because the major classes of speech sounds have different properties depending on one's approach. For example, one could view them from an acoustic perspective or an articulatory one, both of which would have a

different classification than from an auditory perspective. While many versions of the sonority hierarchy have been proposed, Parker (2011:1161-1162) states that the most frequently cited hierarchy places vowels at the top as being highest in sonority followed by glides, liquids, nasals, and then obstruents. Previous studies on complementizer realization considered the sonority hierarchy when looking at the preceding and following phonological environments to determine the effects on *que* realization (Cedergren and Sankoff 1974; Connors 1975; Sankoff 1980; Martineau 1985; Blondeau & Nagy 2008; Liang et al. 2021). Warren (1994) considered the following context only. The primary groupings used for these studies were one of the following in order from least to greatest sonority:

- ◊ obstruents → sonorants → vowels
- ◊ sibilants → non-sibilant consonants → vowels

Cedergren and Sankoff (1974) and Sankoff (1980) found that sibilants favored the absent complementizer in the preceding and following environment. Martineau (1985) and Blondeau and Nagy (2008) found this to be true for the following context only. Both Warren (1994) and Liang et al. (2021) determined that obstruents favored the absent complementizer.

A somewhat later study conducted on Acadian French as spoken in New Brunswick in the late 90s found few instances of absent *que* (Wiesmath 2002). Of the examples found, it was determined that the syntactic context preceding *que* was important. Specifically, the expressions *je suis fière* 'I am proud', *je suis sûr* 'I am sure' and *il faut* 'it is necessary' favored absent *que*. An analysis of the matrix clause subject by Liang et al. (2021) revealed that there was no significant effect on complementizer realization. The syntactic environment following the absent complementizer was analyzed by Connors (1975), Martineau (1985), Blondeau and Nagy (2008), and Liang et al. (2021). The first two scholars classified the word following *que* as a pronoun,

such as *je trouve que tu* 'I think that you', or as 'other', such as *je pourrais dire que sur* 'I could say that on' and found that pronouns favored absent *que*. Blondeau and Nagy (2008) classified the following subject as a first-, second-, or third-person pronoun or a noun phrase and determined that first- and second-person pronouns favored absence of *que*. The salience of the following subject was studied by Liang et al. (2021) who looked at individual pronouns and then noun phrases as a whole. They concluded that the more salient pronouns favored absent *que*.

The number of dependent clauses preceding *que* within the same sentence also influenced the absent complementizer such that one clause favored absence while more than one favored presence. Examples (24) and (25) provide an illustration.

(24) Ça fait peut-être vingt ans **qu'**ils sont ici. (Martineau 1985:77)

'It's been maybe 20 years **that** they've been here.'

Number of clauses preceding *que*: 1

(25) Faut pas que tu te fies **que** si tu vas à la messe, tu es sauvé. (Martineau 1985:77)

'You must not trust **that** if you go to Mass, you are saved.'

Number of clauses preceding *que*: 2

Elements intervening between the verb and *que* can diminish comprehension of the sentence and therefore disfavor absent *que*. For example, in *je pense en fait* ***que*** 'I think in fact **that**', the expression *en fait* is considered intervening material since it comes between the matrix verb and *que*. For the embedded clause, anything the comes between *que* and the following subject is considered intervening material, such as *au Québec* in *je pense* ***qu'****au Québec il y a* 'I think **that** in Quebec there is'. Intervening material in both the matrix and embedded clauses was found by Martineau (1985) to disfavor absent *que*. Similarly, Warren (1994) considered the complexity of the preceding and following utterances and found that complex utterances disfavored absence,

whether preceding or following. A subject plus verb was considered a simple utterance while one with complements or prepositional phrases such as time expressions were considered complex. This view is roughly equivalent to the number of preceding clauses and intervening material.

One linguistic factor considered by Martineau (1985) that was not taken into consideration in other studies was the type of clause, which had the following rates of the absent complementizer: complementizer clause (32%), circumstantial clause (28%), and relative clause (22%). An example of each clause type is provided in (26) - (28). Only complementizer clauses are considered in the present study.

(26) Complementizer: Je pense **que** je vais y aller (Martineau 1985:2).
'I think **that** I am going to go there'

(27) Circumstantial: C'est parce **qu'**il veut (Martineau 1985:24).
'It's because he wants [to]'

(28) Relative: L'homme **que** j'ai vu est un voleur (Martineau 1985:1)
'The man **that** I saw is a thief'

Among social factors, social class is very commonly considered as a factor in sociolinguistic studies. However, the classification systems used can vary considerably. Guy (2011:162-163) identifies two primary views on class. The Marxist approach focuses on conflicting interests and differences in power, rather than on status. The other approach, used widely in sociolinguistics, focuses on social unity and status more than on conflict and power. According to this view, "class is seen as a relatively continuous scale on which individuals are ranked according to assorted personal characteristics such as level of education, income, occupation, etc., which collectively imply a certain degree of social esteem" (Guy 2011:163). Dodsworth (2011:192) echoes this by saying that social class originally and also largely today is

represented “as a set of discrete locations in a socioeconomic hierarchy”. Labov (1966) used a scalar index based on years of education, occupation of the family breadwinner, and family income. Most sociolinguistic studies since the 1990s have used some version of a scalar index to measure social class (Guy 2011:165). However, there are alternatives to this traditional approach that are used in current sociolinguistic studies. These include the linguistic market(s), life-modes, networks, and communities of practice (Dodsworth 2011:200-203).

Cedergren and Sankoff (1974) looked at social class based on whether a speaker belonged to the working-class or was a professional, whereas Warren (1994) divided speakers into three groups: blue collar, white collar, or professional. They found that the lowest social class favored absent *que*. Warren (1994) also found that those with an average level of education, which equated to the early years of university, favored absent *que*. Liang et al. (2021) considered education and occupation as an indicator of socioeconomic status and found that professionals, white collar workers, and university graduates disfavored absent *que*. Similarly, Martineau (1985) looked at socioeconomic status based on the neighborhood in which each speaker lived and education level. She concluded that residents of the neighborhood indicating the highest socioeconomic class disfavored absent *que*. Other social factors considered in previous quantitative studies were age (Martineau 1985; Warren 1994; Liang et al. 2021) and sex (Cedergren and Sankoff 1974; Martineau 1985; Warren 1994; Blondeau & Nagy 2008; Liang et al. 2021). Martineau (1985) was the only one who found an effect based on age. In her study, younger speakers favored absent *que*. None of the studies found sex to be a significant factor.

Results from these early studies inform the design of the present study, which is outlined in the following chapter. Overwhelmingly, the phonological context following *que* has been shown to significantly influence complementizer realization. Thus, this factor is tested based on

sonority groups. The sonority hierarchy used for the present study is similar to the most commonly cited one mentioned by Parker (2011). The only difference is that obstruents are divided into fricatives and stops[8]. Thus, the order from least to greatest sonority is stops, fricatives, nasals, liquids, and then vowels. This sonority scale will allow us to differentiate obstruents based on manner of articulation. Both Connors (1975) and Martineau (1985) found that a pronoun following the complementizer favored absent *que* while Blondeau and Nagy (2008) and Liang et al. (2021) found certain pronouns to favor the absence of *que*. The present study also considers the syntactic environment following *que* but takes a slightly different approach by classifying the word following the complementizer as a pronoun, lexical noun, or other word. Specific matrix verbs have been shown to favor absent *que* dating back to Old French (Martineau 1993:80-81; Arteaga 2009:24-27). Several scholars noted the lexical effect of specific verbs (Martineau 1985; Wiesmath 2002; Blondeau & Nagy 2008; Liang et al. 2021). The verbal elements chosen for the present study are those that allowed for the absent complementizer in Old French and/or are among the structures most frequently occurring with absent *que* in Modern spoken French. The final linguistic factor to be examined is material intervening between the verb and the complementizer and between the complementizer and the subject of the following clause, which has been shown to diminish comprehension of the utterance and disfavor absent *que* (Martineau 1985).

Regarding social factors, there is a long, documented history of absent *que* being used in popular spoken French in Quebec. It has also been documented in parts of francophone Africa. Given this information and the geographical distance from France, region is considered. Cedergren and Sankoff (1974) and Warren (1994) found an effect based on profession, and

[8] The data for this study do not include any glides.

Martineau (1985) found a slight effect based on neighborhood, which was indicative of socioeconomic class. Both Warren (1994) and Liang et al. (2021) ascertained that education level was a significant factor. Education level, available for European and Quebec speakers, is the only measure of social class used in the present study. Poplack (1989) asserted that education level is an indicator of generalized socioeconomic status. In her study, stratification by educational level confirmed the socioeconomic status indicated by other factors such as income, occupation, housing characteristics, and neighborhood (Poplack 1989). Since prescriptive grammar is necessarily part of formal education, and since the complementizer is obligatory according to prescriptive grammar, education is an important factor to consider on the assumption that a higher educational level will correlate with a lower rate of absent *que*. Finally, age is considered since Martineau (1985) found that age had a minor effect on complementizer realization whereby younger speakers had a higher rate of absent complementizers. An age analysis will allow us to see whether the patterns she and Liang et al. (2021) observed were indicative of age grading or generational change.

Among the linguistic factors tested in previous studies, the phonological context preceding *que* is not considered for the present study. In agreement with Martineau (1985), this environment seems to be lexically conditioned. There are specific verbs that have been shown to favor absent *que* dating back to Old French, and absent *que* is not attested for verbs outside of this group. Thus, it appears that the form of the complementizer is more closely linked to the specific verb preceding it than to the final sound of this verb, which determines the preceding phonological environment, except in the case of intervening material. The type of clause is not considered due to time limitations and the desire to focus solely on complementizer clauses. Style is not tested but may be reflected in the absent complementizer rates for *penser* and *croire*.

The number of preceding and following clauses is not included as a factor since the sentence domain cannot be determined solely from transcripts. Rather, intervening material is considered to provide insight into the larger category of items that can diminish comprehension of an utterance. Finally, the social factor of sex is not considered, since none of the five studies which considered it found it to be significant.

The present study is the first to compare complementizer rates in three francophone regions. Previous studies have yielded great insights into conditioning factors, particularly in Quebec. However, little is known about the phenomenon in the early 21st century, especially in Europe and Africa. All previous studies that used statistical analysis except for Liang et al. (2021) either did not use modern methods or used GoldVarb, which assumes a balanced distribution. This assumption leads to overestimating the significance of effects (Johnson 2009:363). However, mixed-effects models, which will be used for the present study, do not make this assumption. Further advantages of the statistical analysis for the present study are outlined in Section 2.6.

1.5.2. Later Studies on Absent *Que*

Many later studies on the absent complementizer took a more theoretical approach to understanding the phenomenon beginning with Andersen (1993), who focused on viewing the complementizer through a semantic and pragmatic lens. Using two corpora of speakers under the age of thirty recorded in the 1980s in Paris, she categorized each instance as belonging to one of three groups: absent complementizer, obligatory (present) complementizer, or constructions that avoid the complementizer. An example of each group is provided in (29) – (31). She found that *que* is avoided in spoken language, by using a construction where *que* is not allowed such as in

(31), in instances where there is no semantic or pragmatic subordination. In this example, the semantic role of *j'avoue* has been reduced such that it is not the primary idea of the utterance. Rather, *ça il est très précis* is the main idea. Andersen (1993) noted that this structure, which is often used in spoken language, indicates the main idea by the clause that is syntactically first. In this example, *j'avoue* does not indicate semantic subordination. If it did, the utterance would be, *ah oui mais j'avoue que ça il est très précis*. *J'avoue* also does not indicate pragmatic subordination since its role is not to get the attention of the interlocutor as is the case with *tu sais* in *tu sais c'est la mémoire d'écolier* 'you know it's the recollection of a schoolkid'.

(29) absent complementizer :

j'me rappelle bien ∅ quand j'étais à l'école par exemple euh on a bien appris comment écrire des lettres officielles (Andersen 1993:8)

'I well remember ∅ when I was at school for example uh we learned how to write official letters'

(30) obligatory (present) complementizer:

mais j'crois **qu'** c'est normal (Andersen 1993:10)

'but I believe **that** it's normal'

(31) construction that avoids the complementizer:

ah oui mais ça il est très précis j'avoue (Andersen 1993:11)

'ah yes but this it's very precise I admit'

Andersen (1993) also stated that the more abstract a verb is, the less certainly one can speak of subordination and the more often one finds absent *que*. For example, the verb *trouver* literally means 'to find'. However, it has another common use meaning 'to have the feeling that'. This latter meaning is considered more "abstract." Thus, Andersen (1993) observed that when

used with this abstract meaning, the absent complementizer is found more often than when used with the literal meaning.

Andersen (1993) argued that spoken language favors parataxis over morphologically marking a construction without semantic or pragmatic subordination. The complementizer was found to be obligatory with verbs that strongly modify the content of the complement, such as *croire, penser*, and *vouloir*. This group of verbs provides examples of subordination at the functional, semantic, and formal level. Enunciative verbs and verbs associated with indirect discourse were found to favor absent *que*. As examples of these groups, she cited *se souvenir* 'to remember', *se rappeler* 'to recall', *voir*, *dire*, *savoir*, and *trouver*.

When comparing results of the two selected corpora, Andersen (1993) noted a difference in complementizer rates. One corpus consisted of a discussion among nine people and included many examples of enunciative verbs with the present complementizer. The other corpus consisted primarily of interviews between two people and contained very few instances of the present complementizer. She suggested that, in a conversation between two people, abstract enunciative verbs occur very infrequently due to a reduced pragmatic need to state who is responding to a question or who holds the opinions stated. On the other hand, in a conversation among three or more people, there is a greater pragmatic need to state such information.

In an analysis of North American varieties of French, Roberge and Rosen (1999) noted that in Louisiana French and in other popular French varieties, *que* is the default relative pronoun. For example, in Louisiana French, *dont* 'of which/whose' and *lequel* 'which/who/whom' do not occur. Rather, *que* is used in their place as shown in example (32). Because of this, the relationship between the anteceding noun and the subordinate clause can no longer be recovered from the relative pronoun. Therefore, *que* is no longer the antecedent to the

noun phrase but has been reanalyzed as a complementizer. This reanalysis is said to be true of North American varieties of French. The additional function of *que* in turn makes it more difficult to recover its meaning. Since the relative pronoun function supersedes that of the complementizer, the latter is weakened. It is thus more likely to have the absent complementizer.

(32) Louisiana French: la femme...**que** le canal appartient (Roberge & Rosen 1999:162)

'the woman that the canal belongs [to] '

Standard French: la femme à **laquelle** le canal appartient

'the woman to whom the canal belongs'

The studies of Andersen (1993) and Roberge and Rosen (1999) provide theoretical explanations as to why absent *que* is used (more often) in certain contexts or locations. The ideas put forth by these scholars will be considered in the discussion portion of the present study in Sections 3.3.2 and 4.2.2. Other notable studies on the absent French complementizer will be detailed in Section 1.7, which are dedicated to specific questions raised in the literature.

1.6. Related Phenomena

In addition to drawing on previous research on absent *que*, the present study considers research findings on two related phenomena. The first of these phenomena is schwa [ə] deletion, which concerns the deletion of a single phoneme, and the second is *ne* deletion, which concerns the deletion of a word in certain syntactic contexts. Schwa deletion is understood to be obligatory when it is followed by a vowel at a word boundary (Green & Hintze 2021:125). This deletion is marked in written language by an apostrophe. For example, in *j'aime* 'I like/love', the deletion of schwa in *je* [ʒə] is marked by an apostrophe. Even though this type of schwa deletion is typically

viewed as obligatory, some variation in spoken language has been found as illustrated in the following two examples (Green & Hintze 2021:134-135).

(33) je crois qu'effectivement euh (Green & Hintze 2021:134)

'I believe that really uh'

(34) et quand je disais que actuellement en France (Green & Hintze 2021:135)

'and when I said that currently in France'

In their study on schwa realization, Durand et al. (1987:986) define a potential schwa as "the segment corresponding to any graphical 'e' not susceptible to an /e/ - /ɛ/ interpretation as in *fer*, *expliquer*, etc.". The following are examples of a schwa that is susceptible to deletion: *je* [ʒ(ə)], *demander* [d(ə)mãde], *tellement* [tɛl(ə)mã], or *pire* [piʁ(ə)] (Durand et al. 1987:986). The first two, where schwa is in a monosyllable and an initial syllable, are variably deleted in all francophone regions, while the last two, where schwa is word-medial and word-final, are only variable in the South of France. In all other regions they are categorically deleted after a single consonant. As Grevisse and Goosse (1993:36) point out, when schwa is word-interior or word-final it is deleted before a consonant as long as this deletion would not result in an infrequent consonant group. In the case of variable schwa deletion, henceforth referred to as schwa deletion, schwa is a vowel alternating with zero and is limited to unstressed syllables (Hutin et al. 2021:1).

Research on schwa realization in French is abundant, and as Malécot (1976:93) said, "No phenomenon of the French language has received as much attention as the mute-e". What follows is an overview of some of the studies conducted on this phenomenon that highlights the factors shown to influence schwa deletion. It should be noted that the French complementizer *que* consists of [k] plus schwa. If the schwa is deleted, *que* is pronounced [k]. As a point of reference, Table 1.1 summarizes the origin of the participants of each study as well as the

position of schwa studied and the overall deletion rate. Deletion rates are for casual speech in the event that a scholar looked at more than one register.

Table 1.1. Summary of previous studies on schwa deletion

Study	Origin of Speakers	Position of schwa						Overall percentage of schwa deletion
Malécot (1976)	Paris	UI	WI			UF		47%
Durand et al. (1987)	Southern France			WInt	WF		M	- -
Armstrong & Unsworth (1999)	Southern France			WInt	WF			- -
Hansen (2000)	Paris		WI	WInt			M	- -
Pustka (2009)	Southern France			WInt	WF		M	11%-45%**
Ranson & Passarello (2012)	Southern France				WF			65%
Eychenne (2019)	Southern France				WF			- -
Green & Hintze (2021)	Northern France				WF			75.7%*
Hutin et al. (2021)	Northern France				WF			92.32%

*The overall schwa deletion rate for *que*, including its compound forms *jusqu'à* and *qu'est-ce que*. **The overall range for individual participants.
Position of schwa: UI=Utterance-initial, WI=Word-initial, WInt=Word-internal, WF=Word-final, UF=Utterance-final, M=Monosyllables

It is evident from these studies that a wide range of linguistic, social, and stylistic factors affect schwa deletion. It can also be seen in Table 1.1 that schwa realization is variable in more than one variety of French. The factors shown to favor schwa deletion in these studies are summarized in Table 1.2. Phonological factors are by far the most prevalent among linguistic factors, and age is the most common for social factors. Both studies that tested register found casual speech to favor schwa deletion.

Table 1.2. Significant factors favoring schwa deletion in previous studies

Study	Position of schwa	Factors that favor schwa deletion
Linguistic factors		
		Syntactic factors
Green & Hintze (2021) speakers from the North of France	WF	a following personal pronoun or determiner beginning with a vowel
		Phonological factors
Malécot (1976) speakers from Paris	UI, WI, UF	soft enunciation, lax articulation, fast syllabic rate
Ranson & Passarello (2012) older men from the South of France	WF	a preceding /ʁ/, /l/, or /n/
Eychenne (2019) speakers from the South of France	WF	VC_#C, falling sonority, preceding voiceless fricative, preceding nasal vowel
Hutin et al. (2021) speakers from Central/Northern France	WF	simplex coda followed by pause, vowel, or single onset, falling sonority, preceding voiceless labial obstruent, a following voiceless obstruent
		Geographical factors
Eychenne (2019) speakers from the South of France	WF	Basque & Provençal (vs. Languedocian)
		Lexical factors
Durand et al. (1987) speakers from the South of France	WInt, WF, M	frequent words in initial position
Ranson & Passarello (2012) older men from the South of France	WF	frequent words
Eychenne (2019) speakers from the South of France	WF	frequent words
		Grammatical category
Eychenne (2019) speakers from the South of France	WF	function words
		Orthography
Hutin et al. (2021) speakers from Central/Northern France	WF	absence of graphical <e>

Social factors		
		Age
Malécot (1976) speakers from Paris	UI, WI, UF	30-49 years old
Durand et al. (1987) speakers from the South of France	WInt, WF, M	younger age group
Pustka (2009) speakers from the South of France	WInt, WF	younger age group
Eychenne (2019) speakers from the South of France	WF	males below 50
		Sex
Armstrong & Unsworth (1999) speakers from the South of France	WInt, WF	female
Hutin et al. (2021) speakers from Central/Northern France	WF	male
		Social class
Hansen (2000) speakers from Paris	WI, WInt, M	lower class adults (40-58 years old)
Stylistic factors		
		Register
Green & Hintze (2021) speakers from the North of France	WF	casual speech
Hutin et al. (2021) speakers from Central/Northern France	WF	casual speech

The study by Hansen (2000) looked specifically at the social effects of schwa deletion in monosyllables and in word-initial position among Parisians. All instances where there was a single preceding and following consonant were considered. In her 1989 corpus, she found a deletion rate in monosyllables of 69% in a casual conversation and 68% in an informal interview.

The younger group, aged 15 to 25, deleted schwa in monosyllables at a rate of 72% compared to adults, aged 40 to 55, who had a rate of 60%. There was no significant difference based on sex. For her 1992-1993 corpus, she found an overall deletion rate in monosyllables for the interview to be 70%. Younger speakers, aged 16 to 23, had a monosyllable deletion rate of 71.9% for the interview and adults, aged 42 to 58, had a rate of 68.6%. The difference between the two age groups was not found to be statistically significant. The only significant social distinction for this corpus was between lower-class and upper-class adults. The upper class had a deletion rate of 59% for monosyllables whereas the lower class had a rate of 69%. Based on this study, there is a high deletion rate for monosyllables, which include *que*.

We now turn to the second type of deletion, *ne* deletion, which is considered a syntactic, rather than phonological, phenomenon (Ashby 1981; Armstrong 2002). The preverbal negative particle *ne* is used in conjunction with a negative auxiliary that can be an adverb, pronoun, or determiner (Grevisse 1993:1447). The negative particle and auxiliary are separated by the verb, as in *je ne sais pas* 'I don't know', except when used with an infinitive as in *pour ne pas tomber* 'in order to not fall', or when the auxiliary is a subject pronoun as in *personne n'est venu* 'no one came'. Example (35) below shows *ne* used with the negative adverb *nulle part*, and (36) illustrates the use of *ne* with the determiner *aucun(e)*. To ascertain the rate of *ne* deletion, all examples where a productive *ne* can be used are considered (Sankoff & Vincent 1981:246). *Ne* is not considered productive in cases such as the *ne* explétif as in *avant qu'il ne soit venu* 'before he came' where the *ne* does not hold a negative meaning. Based on prescriptive grammar, *ne* is pronounced [nə] before a consonant, and prevocalically it is [n]. Frequently, it is pronounced [n] before a consonant in informal spoken speech (Coveney 1996:56). It is argued that this phenomenon is the result of the variable process of schwa deletion and thus should not be treated

as a distinct variant of *ne* (Coveney 1996:56). Rather, Coveney (1996) states that *ne* should be treated as having two variants: the realized variant, *ne*, which includes [nə] and [n], and the deleted variant ∅. This is the view traditionally taken when analyzing this variable.

(35) Je **ne** vais nulle part.

'I'm not going anywhere.'

(36) Il **n'**y a aucune boisson dans le frigo.

'There is no drink in the fridge.'

Examples of the deletion of the negative particle *ne* are found as far back as the 16th century, and its deletion had advanced considerably by the 18th century (Gougenheim 1951:217; Coveney 1996:58-59). This deletion sometimes happened in written language, such as in the works of Marguerite de Navarre, but primarily occurred in spoken language (Gougenheim 1951:217). Coveney (1996:60) explains that once the negative auxiliaries had taken on a negative meaning and had become obligatory, the particle *ne* became redundant. According to the Jespersen cycle, *ne* loses substance and is then joined by a postverbal reinforcement (Larrivée 2011:1). The postverbal negative then becomes the only marker of negation. This process combined with the particle not being phonetically heavy and being separated from the verb led to its absence in spoken language (Coveney 1996:60).

Previous studies on *ne* deletion found an overall deletion rate of 67.3% for speakers in Belgium and France (Pohl 1968:1355), 99.5% for speakers in Montreal (Sankoff & Vincent 1981:246), 63% for speakers in Tours, France (Ashby 1981:677), and 81.2% for speakers in Paris (Coveney 1996:64). Just as with schwa deletion, *ne* is variably deleted in more than one variety of French. Previous works identified a number of factors that condition the realization of the particle. A summary of factors shown to favor deletion is given in Table 1.3. In his study of

speakers from France and Belgium, Pohl (1968) speculated that articulatory economy plays a role in *ne* deletion. He noted that most sentences with a deleted *ne* resulted in a reduction of the number of phonemes. Those not resulting in fewer phonemes were sentences such as *Elle n'a même pas d'appartement* 'She doesn't even have an apartment' where the pronunciation with *ne* is [ɛna…], when [l] is deleted before a consonant, and without is [ɛla…] (Pohl 1968:1353).

Table 1.3. Significant factors favoring *ne* deletion in previous studies

Study	Factors that favor *ne* deletion
Linguistic factors	
	Phonological factors
Pohl (1968)* speakers from Belgium and France (primarily Paris)	lack of potential hiatus
Ashby (1981) speakers from Tours	preceding consonant, preceding oral or nasal vowel followed by a consonant, between two oral vowels, between two nasal vowels
	Syntactic factors
Pohl (1968) speakers from Belgium and France (primarily Paris)	subject: *il y a*, *ça* + consonant, *je*; before *pas* + infinitive
Ashby (1981) speakers from Tours	preformed sequences such as *je sais pas* & *c'est pas*, negative adverbs other than *que* & *plus*, declarative independent clauses, main verbs, aspectual auxiliary *aller*, clitic pronouns
Coveney (1996) speakers from the Somme department, France	preformed sequences such as *je sais pas* and *il y a pas*; subject: *ce*; negative adverb *pas*
Social factors	
	Age
Pohl (1968) speakers from Belgium and France (primarily Paris)	younger age group
Ashby (1981) speakers from Tours	younger age group
Coveney (1996)	younger age group

speakers from the Somme department, France	
	Sex
Ashby (1981) speakers from Tours	females
	Social Class
Ashby (1981) speakers from Tours	lower middle class
Stylistic factors	
Pohl (1968) speakers from Belgium and France (primarily Paris)	casual speech
Ashby (1981) speakers from Tours	second half of conversation, informal setting, informal topic
Sankoff & Vincent (1981) speakers from Montreal	topics & contexts that do not favor a careful speech style
Coveney (1996) speakers from the Somme department, France	informal setting
External factors	
Pohl (1968) speakers from Belgium and France (primarily Paris)	region: Paris (vs. rest of France), France (vs. Belgium), urban (vs. rural) (France vs Belgium), unilingual (vs. bilingual)

*Pohl (1968) did not conduct tests of statistical significance. Factors listed are based on deletion rates.

Regarding social characteristics, Sankoff and Vincent (1981) did not find a clear distinction based on age, social class, or sex among Montreal speakers. Coveney (1996) determined that age is the most important of these three social factors. He also noted that even though *ne* is seen as the more prestigious variant, there doesn't seem to be any stigma connected to its absence in informal language (Coveney 1996:58).

Indeed, there are many factors, both internal and external, that have quantitatively been shown to impact *ne* deletion. Regarding the importance of individual factors, Pohl (1968) stated that the place of domicile (urban versus rural), age, and formality of language most impact *ne*

deletion. Later, Sankoff and Vincent (1981) proposed that the metalinguistic context or themes favoring a careful speech style are the most important factors. They determined that among Montreal French speakers, *ne* remains a syntactic and stylistic resource, and that it is not in danger of completely disappearing from spoken language. While Ashby (1981) did not rank the importance of factors, he did point out that the fusion of the clitic pronouns and the verb, such as [ʃepa] for *je (ne) sais pas* 'I don't know' instead of [ʒəsepa] (*je sais pas*) or [ʒənəsepa] (*je ne sais pas*), was leading to an increase in *ne* deletion. Finally, Coveney (1996) pointed to the importance of age as a representation of stage of life as age grading. In this type of language change "each generation of speakers has virtually a zero rate of *ne* retention as children and adolescents, but then as they become older modify their speech, under pressure from, and in the direction of, the written language" (Coveney 1996:90). Coveney's (1996) conclusion about age was in contrast to Ashby's (1981, 1991) conclusion that the greater *ne* deletion rate by younger speakers represented a change in progress. He supported this view by showing an increase in *ne* deletion rates in both younger and especially older speakers in a corpus collected 19 years after the one for his first study (Ashby 2001).

Schwa and *ne* deletion have many of the same conditioning factors whether they be linguistic, social, or stylistic. Both *ne* and *que* are monosyllables that end in a variable schwa. Like *ne*, *que* is often shortened to [k] in informal language, even when followed by a consonant (Hansen 2000:45). Since variable schwa and *ne* deletion are related to absent *que*, the factors found to be relevant for these two types of deletion were considered for the present study alongside the factors found to be relevant in previous studies on absent *que*. For the absent complementizer, it is well-established that it is associated with informal spoken language just as is the case for variable schwa and *ne* deletion. Scholars seem to be in agreement regarding this

factor. Thus, formality was not tested. More than one previous study on the absent complementizer, schwa deletion, and *ne* deletion have indicated that phonological context, lexical identity, and age affect the variable. However, there is disagreement among scholars as to the role that each factor plays. Therefore, these factors are included in the present study. While two studies on schwa deletion found sex to be a significant factor, I decided not to include it. For absent *que* it was shown not to be a significant factor in five different studies. The syntactic context has been regarded less for schwa deletion but has been shown to be quite relevant to *ne* deletion and also to the absent complementizer. Regarding absent *que*, there has been considerable debate about this factor. It is therefore considered in the present study. Finally, social class has been considered in many studies on the absent complementizer but considered in fewer studies on schwa and *ne* deletion. The study by Hansen (2000) highlights the relevance of social class for variable schwa deletion in monosyllables such as *que*. It was decided to consider one aspect of social class based on her study and the studies on absent *que*.

The traditional treatment of *que* and *ne* in linguistic studies considers only two variants. Based on this precedent and the fact that data for the present study are extracted from written transcripts of spoken language, presence and absence of *que* are the only two variants that will be considered. Thus, presence of *que* will include both [kə] and [k], following the precedent set by studies on *ne* deletion (Coveney 1996).

1.7. Matrix Verbs or Discourse Markers?

1.7.1. Parenthetical and Non-parenthetical Verbs

It was noted by Benveniste (1966) and later echoed by other scholars (Recanati 1984; Blanche-Benveniste 1989; Martineau 1993; Andersen 1996; Apothéloz 2003; Avanzi 2012) that certain

French verbs among the group of performative verbs display a semantic asymmetry when used in the first person present indicative. These verbs are known as parenthetical, weak, epistemic, or incident verbs, and they indicate the degree of authenticity that a speaker attributes to the utterance (Andersen 1996:309). In order to illustrate this asymmetry, Benveniste (1966:263-264) provides the examples and arguments included in this paragraph. *Je mange* 'I eat', *tu manges* 'you eat', and *il mange* 'he eats' all describe an action. Similarly, *je souffre* 'I suffer', *tu souffres* 'you suffer', and *il souffre* 'he suffers' all describe a state of being. However, some verbs do not retain the same permanent meaning for all subjects. If one says, *Je sens que le temps va changer* 'I feel that the weather is going to change', the person is expressing a subjective impression. On the other hand, if one says, *Je crois que le temps va changer* 'I believe that the weather is going to change', the person is not expressing the act of believing but rather a mitigated assertion. A fact, *le temps va changer*, is changed into a subjective utterance by introducing it with *je crois*. Similarly, *je suppose* 'I suppose', *je présume* 'I presume', and *je conclus* 'I conclude' do not describe a mental operation. Rather, they indicate an opinion. *Raisonner* 'to reason' and *réfléchir* 'to think, reflect', closely related verbs, do not display this asymmetry between the objective and subjective meanings. When used in the first person indicative, they describe the act of reasoning and thinking.

The distinction between parenthetical and non-parenthetical verbs is important to the present study since parenthetical verbs favor absent *que*, while non-parenthetical verbs do not (Martineau 1993; Pusch 2003). Apothéloz (2003) explains that there has been morpho-phonological erosion of *que* where it is pronounced as [k] or left out altogether when used in a construction with a parenthetical verb. However, the present complementizer can appear in constructions involving parenthetical verbs (Pusch 2003). What follows is a description of both

types of verbs, tests for determining which type of verb is being used, and the function of parenthetical verbs. The discussion on parenthetical verbs helps us better understand the nuances of such verbs and brings to light the challenge of developing criteria to identify them definitively.

Verbs of the type *je crois* were termed *recteurs faibles* or 'weak verbs' by Blanche-Benveniste and defined as verbs allowing the following two types of constructions (Blanche-Benveniste 1989). First, the verb is the head of the construction and is followed by a *que*-phrase that has the appearance of a complement, as shown in (37). In the second construction, the verb is inserted after the phrase appearing to be a complement or within this phrase, as shown in (38).

(37) *je crois bien* **que** c'était signalé dans le journal (Blanche-Benveniste 1989:60)
'I believe **that** it was reported in the newspaper'

(38) c'était signalé dans le journal, *je crois bien* (Blanche-Benveniste 1989:60)
'It was reported in the newspaper, I believe'
c'était, *je crois bien*, signalé dans le journal
'It was, I believe, reported in the newspaper'

Verbs such as *croire* can function as a parenthetical or non-parenthetical verb, as will be shown later in this section. However, some verbs, such as *il paraît que* 'it appears that' and *on dirait que* 'one would say that', can only function as a parenthetical verb (Blanche-Benveniste 1989).

In contrast, non-parenthetical verbs are defined as verbs that allow double marking by a clitic pronoun and a following clause, as shown in (39) (Blanche-Benveniste 1989). The pronoun *l'* 'it' or *ça* 'this' can replace the utterance that follows. These verbs govern the following clause, which is considered a complement.

(39) il me l'a dit: « j'en suis donc à 14 leçons » (Blanche-Benveniste 1989:56)

'He said it to me: "I'm thus at 14 lessons"'

il m'a dit ça: « j'en suis donc à 14 leçons »

'He said this to me: "I'm thus at 14 lessons"'

Verbs requiring the subjunctive mood in the complement clause are considered by Blanche-Benveniste (1989) to be non-parenthetical verbs since they impose a certain mood. However, Martineau (1993) argues that verbs such as *vouloir* 'to want' and *falloir* 'to need to' that require the subjunctive should be considered parenthetical verbs even though there are some restrictions on their being inserted at the beginning of a sentence. This assertion is based on her conclusion that the restriction is dependent upon the use of the subjunctive rather than the verb. Example (40a) provides a sentence with *croire* as the matrix verb where *que* is absent. The clause in the complement position can be inserted at the beginning of the sentence using the conditional as shown in (40b). However, the clause cannot be inserted at the beginning of the sentence with the absent complementizer and a verb in the subjunctive as shown in (40c). Martineau (1993) argues that the grammaticality of the same sentence in (40c) with the present complementizer, as shown in (41), demonstrates that the restriction is tied to the use of the subjunctive and not the specific verb.

(40a) Je crois pas ∅ je serais / je sois capable de rester à rien faire (Martineau 1993:87)

'I don't believe ∅ I would be / I am (subjunctive) able to continue doing nothing'

(b) Je serais capable de rester à rien faire, je crois pas.

'I would be able to continue doing nothing, I don't believe [so].'

(c) *Je sois capable de rester à rien faire, je crois pas.

'I am (subjunctive) able to continue doing nothing, I don't believe [so].'

(41) **Que** je sois capable de rester à rien faire, je crois pas. (Martineau 1993:87)

'**That** I am able to continue doing nothing, I don't believe [so].'

If double marking is used as a criterion for non-parenthetical verbs, it can be applied to verbs such as *croire* that have both a parenthetical and non-parenthetical use to show the semantic difference between the two functions. When used as a non-parenthetical verb, it is possible to pronominalize the complement, as shown in (42). However, when *croire* is used as a parenthetical verb, the equivalence between the phrase that appears to be the complement and the pronoun is not very clear as illustrated in (43). In this case, *je crois bien* has the meaning of *à mon avis* 'in my opinion.' The double marking shown in (44) and (45) makes it clear that there is a semantic difference between the two uses. In (44) *je crois* has the literal meaning of 'I believe', whereas in (45) *je crois* expresses an opinion.

(42) je crois **qu'**il est innocent, je le crois (Blanche-Benveniste 1989:62)
'I believe **that** he is innocent, I believe it'

(43) je crois bien **qu'**il va pleuvoir (Blanche-Benveniste 1989:62)
'I believe **that** it's going to rain'
il va pleuvoir, je crois bien
'it's going to rain, I believe'
? je le crois bien
'I believe it'

(44) je le crois, **qu'**il est innocent (Blanche-Benveniste 1989:62)
'I believe it, **that** he is innocent'

(45) *je le crois bien, **qu'**il va pleuvoir (Blanche-Benveniste 1989:62)
'I believe it, **that** it's going to rain'

As for the function of the verbs *croire*, *supposer*, *présumer*, and *conclure* used in the first person present indicative, Benveniste (1966) asserts that they indicate subjectivity and communicate the attitude of the speaker regarding the utterance that follows. Recanati (1984) points out that the main clause is, in one regard, transparent since it does not add any additional information. Its purpose is to indicate the type of speech act. According to Urmson (1963:240), "they help the understanding and assessment of what is said rather than being a part of what is said." Benveniste (1966) specifies that this phenomenon only occurs with first person verbs. Expanding on this limitation, Blanche-Benveniste (1989) states that parenthetical verbs are limited to a modality, tense, or person. She states that there is no list of parenthetical verbs because their categorization is dependent upon their use. The same conclusion is reached by Martineau (1993). Andersen (1996:309) proposes that the following verbs are the only truly parenthetical verbs in French: *penser* 'to think', *croire*, *trouver* 'to find' (used in the sense of 'to have the feeling that'), *supposer*, *se souvenir* 'to remember', and *se rappeler* 'to recall'.

In a study using the Corpaix corpus, Willems and Blanche-Benveniste (2014) searched for and analyzed the three types of constructions in which parenthetical verbs can be used as illustrated in (46) - (48b). They found the construction in (46) to be far more frequent than the other two.

(46) Je trouve **que** c'est dommage. (Willems & Blanche-Benveniste 2014:117)
'I think **that** this is a pity.'[9]

(47) C'est dommage je trouve. (Willems & Blanche-Benveniste 2014:117)
'This is a pity I think.'

(48a) C'est dommage. (Willems & Blanche-Benveniste 2014:117)

[9] Glosses for (46) - (48b) are by Willems and Blanche-Benveniste (2014).

'It's a pity.'

(48b) Oui je trouve.

'Yes, I think so (too).'

Using the constructional framework, they view parenthetical verbs as pairing a syntactic representation with a semantic representation. These verbs are unique in that there is not one specific meaning associated with each syntactic structure, but rather, a cluster of three syntactically different structures are linked to a specific meaning. Willems and Blanche-Benveniste (2014) posit that the different syntactic realizations represent different degrees of autonomy and function at different discourse levels. The most frequent construction is also the least autonomous of the three. They note that the pragmatic function of these verbs is to put distance between the speaker and the statement as a way of attenuating the statement.

While many studies, such as Andersen (1993), analyze the presence or absence of subordination or the degree of subordination, Willems and Blanche-Benveniste (2014) state that *que* can be regarded as having an integrative function rather than a subordinative function. For example, they argue that in (49), *que* introduces a subject, and that it would not be accurate to say that *il fait beau aujourd'hui* is subordinate to *il est vrai*. In this case, *que* indicates the integration of the following clause. They continue by stating that the subordinative function is just one manifestation of a larger integrative function whereby a clause is considered to be integrated into a larger unit. As another example of the integrative function, in (50) there is not only a need to integrate what follows *que* but also what precedes *je trouve*.

(49) Il est vrai **qu'**il fait beau aujourd'hui. (Willems & Blanche-Benveniste 2014:129)

'It is true **that** the weather is nice today

(50) Il y a des jours où je trouve **que** les mots français me viennent plus facilement.
(Willems & Blanche-Benveniste 2014:129)
There are days when I think **that** French words come to me more easily.

1.7.2. Are Parenthetical Verbs Actually Verbs?

From a pragmatic point of view, Andersen (1996) likens parenthetical verbs to discourse markers. These verbs are engagement markers that indicate how the utterance should be interpreted. Barnes (1995:813) defines discourse markers as expressions of spoken language that are characterized by the loss of their typical lexical meaning and their function of structuring discourse. They can indicate emotions, epistemic value, the beginning or end of a turn, hesitation, a transition, new information, a correction, or reorientation, among others. Examples (51) and (52) below illustrate two uses of discourse markers in French. Parenthetical verbs can be used as discourse markers (Martineau 1993:85), but this use is distinct from one where the verb precedes a complement clause. With a discourse marker there is no complementizer that follows.

(51) Correction: J'y suis allé jeudi dernier, <u>enfin</u>, c'était vendredi.
'I went there last Thursday. <u>Actually,</u> it was Friday.'

(52) Hesitation: <u>Ben</u>, je sais pas, quand tu veux.
'<u>Er</u>, I don't know, whenever you want to.'

It was noted by Gachet (2015) that prosodic cues are sometimes the only means of determining whether a structure is a complementizer clause or a discourse marker when there is no present complementizer. Andersen (1999) places prosody on the same level as morpho-syntactic markers with regard to marking subordination in her study of adverbial clauses.

Focusing on prosodic units, Avanzi (2012) shows that categorization as a discourse marker or a complementizer clause is not always apparent when reading a transcript of spoken language. He conducted a study on six parenthetical verbs using data from the Phonology of Contemporary French (PFC) project and the Beeching Corpus. Analor was used to do an automated forced alignment to determine the force of the final syllable of prosodic groups. Analor detects major prosodic units based on global and local melodic variations (Avanzi et al. 2008:119). An utterance is segmented into prosodic periods within which prominent syllables are automatically detected (Avanzi et al. 2008:119). The results of Avanzi's (2012) study indicated that there is an encompassing prosody when absent *que* is used. In this case, the prosodic force of the parenthetical verb is inferior to that of the final syllable of the unit that follows it. Thus, the parenthetical verb forms a prosodic unit with what follows it. This is consistent with the prosodic pattern found with the present complementizer but not with a discourse marker. There was one instance where *que* was absent and the parenthetical verb seemed to refer to an earlier statement, not the one just before or after it. Analor showed that the prosodic force of the parenthetical verb was superior to that of the final syllable of the unit that followed. After listening to the utterance, Avanzi (2012) confirmed that it was a delayed parenthetical that indeed referred to something said earlier in the conversation. Prosodic analysis was used to determine that in (53) the parenthetical verb *pense* is part of the first sequence, thereby functioning as a discourse marker. However, in (54) it is part of the second sequence and functions as part of a complementizer clause.

(53) non ça a rien à voir je pense // c'est les villes qui sont un peu euh un peu plus grosses (Avanzi 2012:275)

‘No, this has nothing to do [with it] I think // It’s the cities that are a little uh a little bit larger’

(54) ils regardent la ferme aujourd’hui encore comme comme leur propriété // je pense ∅ les enfants sont encore aujourd’hui malheureux que le papa parce que... (Avanzi 2012:276)

‘They still look at the farm today as as their property // I think ∅ the children are still today unhappy that the dad because…’

While Andersen (1996) likens parenthetical verbs to discourse markers, she also compares them to adverbs. Due to the ambiguity of these verbs and the double construction possibility, they are sometimes likened to predicative adverbs such as *heureusement* ‘it’s just as well’, *peut-être* ‘maybe’, and *bien sûr* ‘of course’ (Recanati 1984; Blanche-Benveniste 1989; Gachet 2015) or sentence adverbs such as *enfin* ‘finally’ and *en effet* ‘indeed/in fact’ that serve as a complement to the whole phrase (Recanati 1984; Andersen 1996). To illustrate the similarity, Blanche-Benveniste (1989:68) provides example (55).

(55) oh, peut-être **qu’**ils sont comme ça (Blanche-Benveniste 1989:68)

‘Oh, maybe they’re like this’

ils sont comme ça, peut-être

‘They’re like this, maybe’

Adverbs can also appear sentence medially like parenthetical verbs, thus appearing in the same three syntactic constructions (Willems & Blanche-Benveniste 2014). Syntactically and typologically, parenthetical verbs serve the same function and have the same placement as sentence adverbs (Andersen 1996). However, Andersen (1996) points out that semantically they are more similar to predicative adverbs.

Further similarity is found between parenthetical verbs and adverbs in the phonetic realization of *que*. Following a non-parenthetical verb, *que* has the tendency to be realized as a full syllable, /kə/, even before a vowel (Blanche-Benveniste 1989). However, following a parenthetical verb, *que* is often realized as a single phoneme, /k/, that is affixed to the verb (Blanche-Benveniste 1989). Following a predicative adverb, *que* tends to have the same realization as after a parenthetical verb (Blanche-Benveniste 1989). For example, in *je crois qu'il est malade* 'I believe that he is sick' the speaker often treats *que* as the affix /k/ as in *je crois-k*.

For these reasons, Blanche-Benveniste (1989) considers parenthetical verbs not to be governing verbs but rather to be modality supports like adverbs. Due to the fact that parenthetical verbs express a greater degree of subjectivity than adverbs (Urmson 1963), and that this subjectivity expressed using the first person is not possible with adverbs, Andersen (1996) concludes that parenthetical verbs should be considered to be more like discourse markers. Willems and Blanche-Benveniste (2014) point out that adverbs, like *peut-être* 'maybe', can easily allow absent *que* in clause initial position, while parenthetical verbs, like *penser* 'to think', cannot. Morphologically, parenthetical verbs behave as verbs, and syntactically, they maintain certain argumentative properties and exercise selectional restrictions on their complements (Willems and Blanche-Benveniste 2014). Thus, Willems and Blanche-Benveniste (2014) conclude that parenthetical verbs should be considered verbs, not discourse markers or adverbs.

Regarding the status of parenthetical verbs, Gachet (2015) argues, in agreement with Recanati (1984) and Schneider (2007), that a parenthetical verb in initial position can function as a main clause governing a complement clause, even though it also fills a pragmatic mitigating function. When the parenthetical verb is in either of the other two possible positions, it is a peripheral element. When a parenthetical verb appears in initial position and is followed by the

absent complementizer, Gachet (2015) contends that the verb has a peripheral function, like in the other two possible positions. As evidence, he states that *paraît-il que* and *semble-t-il que*, both meaning 'Does it seem/appear/look that', should not allow inversion in initial position. However, inversion has been attested in informal spoken language, in literary imitations of spoken language, and in casual written language, such as in blogs and forums. He states that these structures have arisen based on analogy with adverbs such as *peut-être* and *heureusement*, which also play a mitigating role. Thus, what has traditionally been called *que*-deletion is the adoption of an initial adverbial position by a parenthetical verb via a process of analogy with adverbs, which can be viewed as a fourth use of parenthetical verbs, in addition to the three shown in Examples (22) - (24) above. Within this view there is no subordination and no complementizer to delete.

1.7.3. Implications for the Present Study

As previously stated, parenthetical verbs favor absent *que* (Martineau 1993; Pusch 2003). There is no definitive list of parenthetical verbs since their categorization as such is dependent upon their specific use (Blanche-Benveniste 1989; Martineau 1993). While Blanche-Benveniste (1989) proposed a test to determine whether a verb is used parenthetically, there is a lack of agreement among scholars as to which instances should be classified as parenthetical, as in the case of verbs requiring the subjunctive (Blanche-Benveniste 1989; Martineau 1993; Andersen 1996).

The aim of the present study is not to categorize each instance of a verb as parenthetical or non-parenthetical. Rather, the approach taken includes the verbs (parenthetical and non-parenthetical) that were used in what appeared syntactically to be a complementizer construction

with the absent complementizer in Old French or that frequently occur in such constructions in Modern Spoken French according to those identified in the following studies: Martineau 1985, Wiesmath 2002, Boutin 2007, Arteaga 2009, Glikman 2009, and Gachet 2015. Morphologically, parenthetical verbs are verbs. Their syntactic function is debated and has been proposed to be a governing verb, a modality support like adverbs, or a peripheral element. As stated in Vincent (1991:47), Martineau (1993:84), and Andersen (1996:307-309), the following verb forms can be used as discourse markers: *je crois* 'I believe', *je pense* 'I think', *je ne sais pas* 'I don't know', *je trouve* 'I believe/feel', *je veux dire* 'I mean', *je te dis* 'I tell you', *tu sais* 'you know', *tu vois* 'you see', and *voyons* 'let's see'. These can also be used as parenthetical verbs in a complementizer construction. For the present study, the syntactic function is considered to be a governing verb so as to allow for a comparison with early studies on this phenomenon and ascertain the trajectory of it in Quebec since the 1980s. Including discourse markers in a study on the complementizer would yield skewed results since these are not part of a complementizer construction. At the same time, it is not always possible to distinguish a discourse marker from a construction with an absent complementizer based on written transcripts. Since a prosodic analysis is outside the scope of this study and is not possible for all corpora, measures are taken to reduce the number of discourse markers that are included. Criteria for exclusion based on an increased possibility of a discourse marker will be detailed in Section 2.5.

1.8. Summary

The present variationist study on absent *que* in three francophone regions aims to determine how widespread the phenomenon is in the three regions, which factors condition it, and what the data tell us about the trajectory of absent *que* in Quebec since the late 20th century. It has been

suggested that the absent complementizer may not indicate that the complementizer was deleted but, rather, that it was never there. There are examples of parataxis and hypotaxis in Latin, Old French, and Modern French. The debate about whether the absent complementizer is an example of parataxis or hypotaxis with a deleted complementizer is ongoing. The two can be indistinguishable. The present study bears in mind that both are possibilities while considering all instances that could possibly represent hypotaxis.

The absent complementizer not only exists in French but also in other related languages such as Gascon, Spanish, and Italian. Early studies on the absence of the French complementizer were conducted in the last three decades of the 20th century and focused on Quebec speakers. A variety of social and linguistic factors were considered and shown to condition absent *que*. Later studies tended to focus on theoretical frameworks to better understand the phenomenon.

Variable schwa deletion has been shown to occur in monosyllables, such as *que*, even when followed by a consonant. Scholars have shown that linguistic, social, and stylistic factors condition this deletion. Another related phenomenon, *ne* deletion, also involves variable schwa deletion in a monosyllable. However, studies do not usually distinguish [nə] from [n]. Many of the same types of factors have been shown to affect the realization of this negative particle. The factors included in previous studies on both phenomena were considered for the present study. It was decided to treat [kə] and [k] as the same variant since the two cannot be distinguished in the transcripts used for the present study and since this is how such variants are usually treated in the literature.

Finally, a challenge related to the absent complementizer is that of determining which instances of certain verb forms are in fact part of a complementizer construction and which ones are discourse markers. This is an important distinction since discourse markers never appear with

a complementizer. Transcribed speech is not always sufficient for identifying discourse markers. In the absence of a fine-grained analysis of speaker audio to exclude all discourse markers, criteria are set to reduce the number that are included in the data. Also, potential discourse markers will be analyzed separately after the initial analysis of all data.

1.9. Dissertation Outline

This introductory chapter has served to provide the linguistic background necessary to place absent *que* in its larger context. The following chapter outlines the corpus and methods used for the present study. Chapter 3 then provides results for the generalized linear mixed-effect models, a discussion of the results for social factors, and a cross-regional comparison according to social factors. Chapter 4 continues with a discussion of the results for linguistic factors as well as a cross-regional comparison of these factors. Next, in Chapter 5, a qualitative analysis of individual speaker variation as it relates to the matrix verb and phonological groups following the complementizer is conducted. A summary of the results of the present study, limitations, and future research directions are presented in Chapter 6, the final chapter.

Chapter 2

Corpus and Analysis of Absent *Que* in Three Francophone Regions

2.1. Introduction

This chapter presents the corpora and methods for data collection and analysis used in the present study. Section 2.2 provides an overview of the corpora used, which includes criteria for speaker selection as well as demographic information for all participants. Methods used for data extraction and criteria for exclusions are outlined in Section 2.3. The following two sections, 2.4 and 2.5, present the social and linguistic factors tested in the present study. The tools used for quantitative analysis are presented in Section 2.6 followed by a summary of the chapter.

2.2. Corpus and speakers

Data for this study come from the Phonology of Contemporary French project or PFC (Phonologie du Français Contemporain, Durand et al. 2002), which is a 1.8-million-word corpus comprised of 50 subcorpora, and from the Corpus of French spoken in Quebec or CFPQ (Corpus de français parlé au Québec, Dostie 2016), a corpus comprised of 31 subcorpora and 112 speakers. Of the 50 subcorpora in the PFC, only 16 are currently available to the public. For each speaker, the PFC provides transcriptions and recordings of four tasks. Two are reading tasks designed to elicit a more formal speaking style, whereas the other two are conversations including guided conversations between an interviewer and an interviewee, and free conversations between a speaker and acquaintances. Both guided and free conversations are used

in this study since both elicit an informal speaking style (Mallet 2008:192). The recordings used for this study were made between 2001 and 2012. The CFPQ consists of transcriptions of spontaneous conversations among three to four speakers plus a student in charge of the recording. The speakers know each other well and typically know the student rather well. The conversations used for this study took place between 2006 and 2011.

Ten subcorpora were chosen from the PFC to represent European French. Each subcorpus is identified by the city or region from which participants were recruited, and these include different geographical areas within France as well as two cities in Switzerland. By including a variety of geographical areas within these two countries, there will be a more representative sample of the variation in European French. Three subcorpora, Brécey, Domfrontais, and Puteaux-Courbevoie, are from Northern France. Dijon and Ogéviller represent Eastern France, and Nantes represents Western France, while Aix-Marseille and Lacaune represent Southern France. Neuchâtel and Nyon are the cities representing Switzerland. Each of the six social groups, formed according to age and sex, consists of eight speakers as shown in Table 2.1. The younger age group consists of females and males ranging in age from 16 to 30, the middle group of speakers aged 42 to 55, and the older group of speakers from 65 to 86. The identifier for each speaker indicates the region (E=Europe), sex (M or F) and age followed in parentheses by the identifier assigned by the corpus. The education level of each speaker is indicated according to the following criteria:

P = primary school (some or completed)

S = secondary school (some or completed)

U = 1 year or more university / internship beyond secondary school / professional training

Table 2.1. Social Characteristics of European Speakers

Speaker	Age	Education	Subcorpus/City
Younger Females			
EF16 (61alh1)	16	S	Domfrontais
EF20 (50aev1)	20	U	Brécey
EF22 (44ams1)	22	U	Nantes
EF24 (21aml1)	24	U	Dijon
EF26 (81afv1)	26	U	Lacaune
EF27a (scajd1)	27	U	Neuchâtel
EF27b (scajc1)	27	U	Neuchâtel
EF30 (svacb1)	30	S	Nyon
Younger Males			
EM17 (61acm1)	17	S	Domfrontais
EM19 (50ajp1)	19	U	Brécey
EM21 (81aaa1)	21	U	Lacaune
EM23 (44ars1)	23	U	Nantes
EM24 (92acj1)	24	U	Puteaux-Courbevoie
EM26 (21ama1)	26	U	Dijon
EM29 (21abl1)	29	U	Dijon
EM30 (13bfa1)	30	U	Aix-Marseille
Middle-Aged Females			
EF42 (50atv1)	42	S	Brécey
EF43 (44aar2)	43	U	Nantes
EF46 (svanp1)	46	S	Nyon
EF48 (61ajh1)	48	S	Domfrontais
EF52 (svarb2)	52	U	Nyon
EF53 (13bma1)	53	S	Aix-Marseille
EF54 (scahd1)	54	S	Neuchâtel
EF55 (50alb1)	55	U	Brécey
Middle-Aged Males			
EM42 (scapm1)	42	U	Neuchâtel
EM43 (50ayp1)	43	U	Brécey
EM44 (scapy1)	44	S	Neuchâtel
EM45a (svaje1)	45	U	Nyon
EM45b (13brp2)	45	S	Aix-Marseille
EM51 (61abm1)	51	S	Domfrontais
EM53 (50ajm1)	53	S	Brécey
EM54 (81acc1)	54	U	Lacaune
Older Females			
EF65 (svaab1)	65	S	Nyon
EF68 (scamm1)	68	U	Neuchâtel
EF69 (50app1)	69	P	Brécey
EF78 (scajb1)	78	S	Neuchâtel
EF80 (61adp1)	80	P	Domfrontais

EF81 (44amr1)	81	S	Nantes
EF82 (13baa1)	82	S	Aix-Marseille
EF86 (21ash1)	86	U	Dijon
Older Males			
EM70 (svarv1)	70	U	Nyon
EM73 (81ajc1)	73	P	Lacaune
EM75 (scarp1)	75	U	Neuchâtel
EM78 (scaaf1)	78	S	Neuchâtel
EM79 (44ajs1)	79	U	Nantes
EM81a (13brp1)	81	S	Aix-Marseille
EM81b (50arm1)	81	P	Brécey
EM81c (54bpm1)	81	S	Ogéviller

Fourteen subcorpora were chosen from the CFPQ to represent Quebec French. Each subcorpus is identified by the number in parentheses in Table 2.2, which is part of the corpus-assigned speaker identifier. All participants are native speakers of Quebec French, but there is no further regional distinction provided in the corpus. Each of the six social groups consists of eight speakers with the exception of the older females and males, which include seven speakers each, since this is the maximum number available in the corpus. The younger age group consists of females and males ranging in age from 16 to 27, the middle group of speakers aged 42 to 55, and the older group of speakers from 65 to 95.

Table 2.2. Social Characteristics of Quebec Speakers

Speaker	Age	Education
Younger Females		
QF16a (17J)	16	S
QF16b (17C)	16	S
QF17 (17K)	17	S
QF22 (25C)	22	U
QF23 (25ME)	23	U
QF24 (10V)	24	U
QF25 (10M)	25	U
QF27 (14A)	27	U
Younger Males		

QM22 (25P)	22	U
QM24a (21O)	24	S
QM24b (10D)	24	U
QM25a (10J-M)	25	U
QM25b (21S)	25	U
QM26 (14H)	26	S
QM27a (14B)	27	U
QM27b (21Y)	27	S
Middle-Aged Females		
QF42 (6F)	42	U
QF47 (12J)	47	U
QF48a (23S)	48	S
QF48b (23T)	48	U
QF49 (15J)	49	S
QF50 (1C)	50	U
QF52 (1S)	52	U
QF53 (15S)	53	U
Middle-Aged Males		
QM42 (6M)	42	U
QM47 (6J)	47	U
QM48 (23J)	48	U
QM50 (23B)	50	U
QM51a (15R)	51	S
QM51b (1G)	51	S
QM53 (15B)	53	S
QM55 (1L)	55	S
Older Females		
QF65 (20AN)	65	S
QF71 (8E)	71	P
QF74 (8H)	74	P
QF79 (4M)	79	P
QF81 (11M)	81	S
QF82 (4H)	82	P
QF95 (11Mi)	95	P
Older Males		
QM65 (12S)	65	U
QM73a (8G)	73	P
QM73b (8N)	73	P
QM80 (4A)	80	S
QM81 (4R)	81	S
QM87 (11J)	87	S
QM95 (11H)	95	S

The PFC contains three subcorpora representing African French: Abidjan (Côte d'Ivoire), Bangui in the Central African Republic (or RCA, République centrafricaine), and Burkina Faso with no city mentioned in the PFC. All three subcorpora are used for this study. Due to the limited number of speakers, there are only two social groups with a sufficient number of speakers. There are eight females in the group of younger speakers who range in age from 19 to 29 as shown in Table 2.3 and fourteen females in the middle age group from 40 to 55 years old. In addition, there is one young male speaker who is 28 and two middle-aged males who are 40 and 50. The education level of speakers in this corpus is not provided.

Table 2.3. Social Characteristics of African Speakers

Speaker	Age	Subcorpus/City
Younger Females		
AF19 (ciaeo1)	19	Abidjan
AF23 (ciana1)	23	Abidjan
AF26 (ciaeo2)	26	Abidjan
AF27 (bfabo1)	27	Burkina Faso
AF28a (bfato1)	28	Burkina Faso
AF28b (rcadt1)	28	RCA Bangui
AF28c (rcayn1)	28	RCA Bangui
AF29 (bfanr1)	29	Burkina Faso
Younger Males		
AM28 (rcaat1)	28	RCA Bangui
Middle-Aged Females		
AF41a (rcarn1)	41	RCA Bangui
AF41b (bfadg1)	41	Burkina Faso
AF44a (ciapa1)	44	Abidjan
AF44b (rcamy1)	44	RCA Bangui
AF45 (rcamk1)	45	RCA Bangui
AF46a (bfakk1)	46	Burkina Faso
AF46b (bfasb1)	46	Burkina Faso
AF47 (bfann1)	47	Burkina Faso
AF48a (bfada1)	48	Burkina Faso
AF48b (bfasy1)	48	Burkina Faso
AF48c (ciasn1)	48	Abidjan
AF50 (bfati1)	50	Burkina Faso

AF54 (bfath1)	54	Burkina Faso
AF55 (ciavd1)	55	Abidjan
Middle-Aged Males		
AM40 (ciafk1)	40	Abidjan
AM50 (bfaki1)	50	Burkina Faso

2.3. Data Extraction and Exclusions

The data for this study were collected by manually searching subcorpora of both the PFC and CFPQ for all instances of the complementizer *que*, present and absent, for the following matrix clause verbs and verbal expressions: *avoir l'impression* 'to have the impression', *avoir raison* 'to be right', *c'est* 'it's', *croire* 'to believe', *dire* 'to say/tell', *falloir* 'must', *jurer* 'to swear', *paraître* 'to appear/seem', *penser* 'to think', *pouvoir dire* 'to be able to say', *promettre* 'to promise', *savoir* 'to know', *se dire* 'to think to oneself/say to oneself', *sembler* 'to seem', *trouver* 'to think/believe/feel', *voir* 'to see', *vouloir* 'to want', and *vouloir dire* 'to mean'. Additionally, the adjectival predicates *c'/il est vrai* 'it's true' and *être* 'to be' + adjective (also including the adverb *bien* since it is often used in this construction in place of *bon* 'good') were searched along with the modal adverbs *peut-être* 'maybe' and *heureusement* 'it's just as well' or 'it's fortunate' [lit. 'happily'] and the conjunctions *c'est-à-dire* 'in other words', *en cas* 'in case', and *être d'avis* 'to be of the opinion'. These elements have been shown to allow absent *que* in Old French and/or to be among the structures most frequently occurring with absent *que* in Modern Spoken French (Martineau 1985; Wiesmath 2002; Boutin 2007; Arteaga 2009; Glikman 2009; Gachet 2015).[10] Each matrix clause verb, verbal expression, adjectival predicate, modal adverb, and conjunction, heretofore referred to as a 'verbal element', was searched in all forms possible in the following

[10] There are 14 more verbs from Glikman (2009) that allowed absent *que* in Old French and still allow the complementizer in Modern French. These verbs do not appear in the other works cited in this paper and are not included in this study.

tenses of the indicative: present, perfect (*passé composé*), pluperfect, imperfect, conditional present, conditional past, simple future, and future perfect. Due to a sizable difference in the amount of data available in each subcorpus, each conjugation was limited to the first ten examples per speaker for the CFPQ. By so doing, the amount of data for Quebec speakers is closer to the amount for African and European speakers. Each instance was manually examined for its inclusion or exclusion based on the criteria below. In addition to a matrix clause containing one of the targeted elements, it is necessary for at least the subject and verb of an embedded clause to follow each instance. This was determined to be the minimum required to identify it as an embedded clause. Therefore, occurrences such as *Je pense que oui* 'I think so' were excluded. In addition, the following constructions were excluded:

◊ The phonological and/or syntactic context after *que* is unclear or cannot be determined, such as after a false start or correction where the word is not completed.

Il dit 'Oui c'est certain', ça m/ je veux dire, ∅ i/ je crois que d'un côté il est un petit peu jaloux. (EM17)

'He says, 'Yes, it's certain', this m- I mean, ∅ i- I believe that on the one hand he is a little bit jealous.'

◊ There is more than one possible placement of absent *que*.

je veux dire ∅ euh ∅ c'était comme elle mettait pas vraiment de sucre dans rien (QM25a)

'I mean ∅ uh ∅ it was like she wasn't really putting sugar in anything'

◊ The same utterance is repeated in succession in which case only one token is counted. For example, in the following utterance, only the last token, *je pense qu'ils l'ont fait aussi*, where the speaker completes the clause was counted:

je pense qu'ils l'ont mais je pense qu'ils l'ont je pense **qu'**ils l'ont fait aussi (QM25a)

'I think that they have (it) but I think that they have (it) I think **that** they have done it too'

◊ The use of the first person plural of *voir*. All instances, nine total, were from Quebec and required the figurative meanings *come on*, *really,* or *yeah right*. When used in this way, the grammatical construction does not include the complementizer.

ben non t'étais pas dans ça voyons t'aurais jamais voulu en tout cas (QF81)

'of course not you weren't in this come on you would never have wanted (to be) anyway'

Many examples of discourse markers were automatically excluded based on the absence of a following embedded clause. In instances where a verbal element could be part of a complementizer clause or used as a discourse marker according to the transcription, a prosodic analysis may be the only means to make this determination. For example, in (56) and (57) the underlined subject and verb could act as a discourse marker or could begin a complementizer clause. It is not possible to tell from the transcript alone. As outlined in Section 1.7, Avanzi (2012) showed that if the force of the final syllable of a prosodic group is superior to that of the verbal element, then it is part of a complementizer clause and is not used as a discourse marker. If the prosodic force of the last syllable in *attention* and *là* in the examples below were to be measured, it could be determined which structure was used. While it is not feasible to do such a prosodic analysis for this study, the criteria following (57) summarize exclusions that were made based on the increased likelihood of a discourse marker:

(56) Ouais toute façon je parle pas bien donc euh tu sais ils vont faire attention ils vont être plus cléments quoi. (EF24)

'Yeah anyway I don't speak well therefore uh you know they are going to pay attention they are going to be more lenient'

(57) oui entre les deux je pense c'est l'idéal là (QF24)

‘yes between the two I think it’s ideal there’

◊ If the verb is immediately followed by a pause, it was considered a discourse marker since this could indicate the end of an utterance. In the following example a pause was indicated in the transcription after *t’sais*.

- ♦ le deuil d'un d'une d'une vie t'sais je pensais que ça allait finir en beauté (QF27)

  ‘the loss of a of a of a life you know I thought that it was going to end on a high note’

◊ If the verbal element is used alongside a discourse marker it was considered a discourse marker. In the example below, *je sais pas* ‘I don’t know’ is preceded by the discourse marker *ben*.

- ♦ ben je sais pas c’est ce que tu m’as dit (QF27)
- ♦ ‘er I don’t know it’s what you told me’

Since verbal elements like those in (56) and (57) did not meet the criteria for exclusion, they were included in this study. Thus, it is possible that some discourse markers are included. For this reason, the verbs that can be discourse markers will first be analyzed with all verbal elements and then as a separate group.

Transcription conventions for each corpus were used to determine the occurrence of pauses. The PFC does not indicate pauses but does indicate hesitations by *euh* in the transcription. There were no instances of a matrix verbal element followed by *euh* and then *que*. Thus, all examples of a matrix verbal element immediately followed by *euh* were excluded. The CFPQ indicates pauses in the transcription by means of a period in parentheses.

While repetitions of an entire segment of speech were not included, tokens that contain the repetition of one or two words were included. An example of an inclusion of this type is shown in (58).

(58) certaines personnes pensaient **que** je j'étais de roman de langue romanche (EF78)

'certain people thought **that** I I was of Roman- of Romansh language [a Romansh speaker]'

The verbs *dire* and *se dire* can be used both in direct and indirect discourse. Each token was classified as belonging to one of these groups or possibly either one. Examples were classified as direct discourse if there were one or more words such as *bon* 'right/ok/well', *voilà* 'there you go', *tiens* 'hey/there you go/darn', *mais* 'but', or *voyons donc* 'come on' (used as a discourse marker) between *(se) dire* and the following subject. Also, if what followed *(se) dire* was a question, this was considered direct discourse and so was excluded from analysis. An example that meets the first two criteria is shown in (59). If a form of direct address was used (60) or there was a pronoun change from third to first person (61) this was considered direct discourse. Out of 877 tokens there were 13 that were clearly direct discourse introduced with *que* (62). All of these examples were from the African speakers except one, which was by a speaker from Quebec. Examples were classified as indirect discourse if the complementizer was present and it was not unequivocally an example of direct discourse as shown in (62).

(59) elle dit <u>mais</u> <u>qu'est-ce que vous dites</u> (EF27a)

'she says <u>but</u> <u>what are you saying</u>'

(60) Son papa lui a dit <u>mon fils</u> c'est trop loin. (AF47)

'his dad said to him <u>my son</u> it's too far'

(61) <u>il</u> dit faut que <u>je</u> te redonne un traitement (QF95)

'<u>he</u> says it's necessary for <u>me</u> to give you a treatment again'

(62) Alors il me dit **que** non <u>frère</u> il faudrait que <u>tu</u> me comprends (AM28)

'so he says to me **that** no <u>brother</u> it would be necessary for <u>you</u> to understand me'

If the subject in the matrix clause did not change in the embedded clause this was considered indirect discourse (63). If any indicator was present that would prohibit the utterance from being direct discourse, it was considered to be indirect. For example, in (64) *mon verre* 'my glass' would need to be *notre verre* 'our glass' or *nos verres* 'our glasses' for it to be direct discourse.

(63) si la petite fille elle dit ∅ à quatorze ou seize ans elle veut avoir une relation sexuelle (QM51b)

'if a young girl she says ∅ at 14 or 16 years old she wants to have a sexual relationship'

(64) Alors les positifs diront ∅ mon verre est à moitié plein. (EF65)

'Then the positives will say ∅ my glass is half full'

For some examples it was determined they could be interpreted either as direct discourse or indirect discourse with the absent complementizer, as in (65) and (66).

(65) il dit que/∅ elle en a un sur le sourcil (QF53)

'he says that/∅ she has one on her eyebrow'

(66) je lui dis que/∅ c'est pas ça (QF65)

'I say to him that/∅ it's not that'

These were classified in a separate group, referred to as the indeterminate group. An analysis of the rate of absent *que* by region for each of the 3 groups revealed that there are (near) categorical rates of absent *que* for each region for direct discourse and the indeterminate group. While the examples in the indeterminate group could possibly be indirect discourse, it appears that the speakers intended for them to be direct discourse. As expected, the rate of absent *que* is much lower for indirect discourse for each region. A summary of the rates for each group and region

are shown in Table 2.4. Given that the complementizer was almost always absent in European and Quebec French in direct discourse, only examples belonging to the indirect discourse group were included in the analysis of the present study.

Table 2.4. Absent *Que* Rates as percentages for *(Se) Dire* according to discourse type

Type of Discourse	% ∅
Europe	
Direct Discourse	100
Indirect Discourse	26.7
Either	100
Quebec	
Direct Discourse	99.5
Indirect Discourse	11.8
Either	100
Africa	
Direct Discourse	81.4
Indirect Discourse	18.0
Either	94.1

2.4. Social Factors

The social factors considered in this study are region, education level, and age. Region is available for the European and African speakers. In Europe, speakers were placed into one of the following groups: Northern France, Southern France, or Switzerland. African speakers were placed in groups according to country: Côte d'Ivoire, Central African Republic, or Burkina Faso. Education level is available only for speakers from Europe and Quebec. For each of these speakers, the education level indicated in the respective corpus was converted into one of the following three groups: primary school (some or completed), secondary school (some or completed), or university (one year or more university / internship beyond secondary school /

professional training). Each speaker was placed into one of three age groups: 16-30, 40-55, or 65 and older.

2.5. Linguistic Factors

Certain verbal elements have been shown to favor absent *que* as stated in Section 2.3. Among this group, particular verbs have been shown to favor absent *que* more than others. For example, Martineau (1993) found *croire*, *dire*, *penser*, *savoir*, *sembler*, and *trouver* to favor deletion to varying degrees. Therefore, the lexical role of the verbal element is examined to determine its possible effect on absent *que*. The phonological and syntactic contexts following the complementizer and intervening material in the matrix and embedded clauses are considered as independent factors. The sound immediately following the complementizer was placed into one of five sonority groups: stops, fricatives, nasals, liquids, or vowels. The word immediately following the complementizer was coded as a pronoun, lexical noun, or other word. Intervening material between the matrix verbal element and the complementizer and between the complementizer and the subject of the embedded clause was considered. These intervening elements include adverbs, negatives, pronouns, clauses, and prepositional phrases. An example of each is shown below.

◊ Adverb in the matrix clause:

elle pensait encore Ø j'étais un étudiant (QM25a)

'she still thought Ø I was a student'

◊ Negative in the matrix clause:

mais je pensais pas Ø c'était officiel (QM27a)

'but I didn't think Ø it was official'

◊ Clause within the embedded clause:

moi j'aurais pensé **que** si c'était juste dans dans le but de de parler de cours là elle aurait peut-être un petit peu exagéré (QF27)

'me I would have thought **that** if it was just in in order to to talk about classes then she might have exaggerated a little bit'

◊ Prepositional Phrase in the embedded clause:

Je veux dire ∅ de la pizza là tu peux faire de la pizza à n'importe quoi (QF25)

'I mean ∅ from the pizza there you can make any kind of pizza'

The following instances were not considered intervening material:

◊ Repetition of the complementizer:

Parce que si ils voulaient **que qu'**on comprenne ils auraient pu nous parler (AF26)

'Because if they wanted us to understand they could have talked to us'

◊ Repetition of the embedded subject, including a dislocation:

il faudrait **que** je euh j'aille directement à la maison (AM28)

'it's necessary for me uh for me to go directly to the house'

Je pense pas **que** grand-maman elle s'assagisse beaucoup (EF27a)

'I don't think **that** grandma, she is settling down much'

2.6. Analysis Tools

A generalized linear mixed-effects model (GLMM) was used to analyze the data from each region. The primary characteristic of such models is that they have both fixed effects and random effects. Fixed effects are explanatory variables whose effect is measured, while random effects are randomly sampled from the population and their effect is filtered out to give a more accurate

picture of the role of the explanatory variables. The variance measurement of each random effect indicates how much variation in the data is a result of the random effect. Thus, the effect is filtered out, but we are able to see how much variation it is responsible for.

Generalized linear mixed models show an outcome as the linear combination of fixed effects and conditional random effects. An example of fixed effects for the present study are education and phonological context. For any given speaker the education level will remain constant. For a given utterance, the phonological context will always belong to one of five groups based on sonority. There may be differences among individual speakers that could affect the model if they were not controlled for. Also, in corpus studies such as this one, there are usually a few speakers who contribute the majority of the data while the rest contribute very little. A mixed-effects model does not assume a balanced distribution. Assigning speaker as a random effect allows for a more accurate picture of the role played by different explanatory variables, such as education and phonological context. The parameters of the GLMM are fit to the data so that the resulting model optimally describes the data.

The ability to model random effects is a considerable advantage of the mixed model over generalized linear models. A GLMM has other notable advantages such as the following (Jaeger 2008:442-443):

- Does not make the assumption of homogeneity of variances
- Has greater power which makes it more likely to detect true effects
- Allows for the testing, rather than stipulation, of a hypothesized random effect to determine whether it should be included

Dummy coding is used to measure each level of each factor against the omitted reference level. The reference level is chosen by alphabetical order. Any manual change in the reference level is indicated and a justification is provided.

To build each GLMM I used the *lmer* function in the *lme4* package (Bates et al. 2015) of R (R Core Team 2022). One advantage of this function is that it maximizes penalized quasi-log-likelihood, thereby reducing the chance of overfitting the data to the sample (Jaeger 2008:443). The nonlinear optimizer bobyqa (Bates et al. 2014) was used to maximize the likelihood of convergence. A model was fit for each region with eight fixed effects and one random effect. The dependent variable has two levels, the absent or present complementizer. The eight fixed effects include the matrix verbal element, phonological context following *que*, syntactic context following *que*, intervening material in the matrix clause, intervening material in the embedded clause, region (for Europe and Africa), education level (for Europe and Quebec), and age. Speaker is considered a random effect. The statistical significance of model outputs is indicated by the p-values such that a p-value less than 0.001 indicates very strong significance. If it is less than 0.01 it indicates strong significance and less than 0.05 indicates significance. The degree of significance is also indicated in the model output by a corresponding number of asterisks as follows:

$p < 0.001$: *** $p < 0.01$: ** $p < 0.05$ * .

Any p-value that is less than 0.05 is considered significant in the present study.

2.7. Summary

The data for the present study were collected from two online corpora, the PFC and CFPQ, both of which allow for an analysis of informal spoken language in the early 21st century. Forty-eight

European speakers and 25 African speakers were selected from the PFC while 46 speakers from Quebec were chosen from the CFPQ. The verbal elements considered for the present study either allowed absent *que* in Old French or are frequently used with it in Modern Spoken French. The criteria used for data exclusions outline the way unclear elements, repetitions, possible discourse markers, and direct and indirect discourse were handled.

The social factors tested for an effect on complementizer realization are region, education, and age. The lexical role of the matrix verb, phonological and syntactic contexts after the complementizer, and intervening material are the linguistic factors considered. A generalized linear mixed-effects model using R (R Core Team 2022) was chosen as an optimal means of analyzing the data for the present study.

Chapter 3

Results of Social Factors in Three Francophone Regions

3.1. Introduction

As mentioned in Chapter 1, this study seeks to determine the extent of *que* absence and its conditioning factors. In order to answer these two questions, this chapter begins with the mixed-effects models for each region, which will be used as a point of reference in Chapter 4 as well. Then, in Section 3.3, an overview of variable *que* rates by region is provided according to social factors, followed by the results of the quantitative analysis and a discussion of each factor. These results will make it possible to ascertain the trajectory of this phenomenon in Quebec since the 20th century. A cross-regional comparison of these factors is also included in this section. Concluding remarks are presented in the final section.

3.2. Mixed-Methods Analysis

3.2.1. Matrix Verbal Element Frequency

The mixed-effects models that follow in Sections 3.2.2 - 3.2.4 do not include lexical frequency of the matrix verbal element. In order to determine whether there is a direct correlation between absent *que* rates and verbal element frequency, the log frequency of each verb form included in the data of the present study was obtained from the French Lexicon Project (Ferrand et al. 2010). This project provides frequency counts for 38,840 French words and 38,840 pseudowords. The log frequency scores, which come from French films, were recorded. Most verbal elements

contain more than one verb form in the data used for this study. Thus, the log frequency of each verb form was obtained, and then these frequencies were averaged for each verbal element. Table 3.1 provides this average along with the standard deviation in parentheses. Compound verb forms and expressions do not have an entry in the French Lexicon Project and thus do not have a frequency listed in Table 3.1. The minimum score for any word in the French Lexicon Project is -2, and the maximum is 4.4 with a mean of 0.0169. All verbal elements in Table 3.1 have a frequency well above this mean. While all verbs in the present study that have a frequency score are considered frequent, the frequency does not directly correlate with the absent *que* rates shown for each region. For example, *vouloir* has the highest log frequency but an absent *que* rate of zero in two regions and 1.8% in Quebec. *Sembler* has a relatively low frequency compared to the other verbal elements but an absent *que* rate of 43.2% in Quebec.

Table 3.1. Matrix Verbal Element Frequency and Absent *Que* Rates by Region

Matrix Verbal Element	Log Frequency	Absent *Que* Rate Europe	Absent *Que* Rate Quebec	Absent *Que* Rate Africa
vouloir	2.4 (σ=0.6)	0% (0/8)	1.8% (1/56)	0% (0/12)
voir	2.1 (σ=0.6)	**92.3%** (60/65)	24.2% (8/33)	35.0% (7/20)
falloir	2.1 (σ=0.6)	0% (0/46)	11.2% (36/321)	0% (0/42)
savoir	2.0 (σ=0.6)	52.1% (37/71)	**74.8%** (205/274)	42.9% (12/28)
dire	1.9 (σ=0.7)	28.4% (23/81)	11.7% (19/162)	18.6% (8/43)
promettre	1.9 (σ=0.1)	0% (0/1)	0% (0/1)	- -
trouver	1.8 (σ=0.5)	10.4% (5/48)	13.2% (9/68)	25.0% (1/4)
croire	1.7 (σ=0.7)	8.1% (60/74)	9.1% (1/11)	37.5% (6/16)
penser	1.7 (σ=0.8)	14.9% (15/101)	25.4% (59/232)	5.3% (6/119)
heureusement	1.6 (N/A)	0% (0/1)	- -	**100%** (1/1)
sembler	1.5 (σ=0.5)	16.7% (1/6)	43.2% (19/44)	- -
paraître	0.8 (σ=0.9)	0 % (0/6)	5.9% (1/17)	7.1% (1/14)
avoir l'impression	- -	8.3% (1/12)	0% (0/5)	0% (0/2)
c'est	- -	4.4% (1/23)	0% (0/25)	7.7% (1/13)
c'est-à-dire	- -	44.4% (4/9)	**83.3%** (5/6)	43.8% (7/16)
c'est vrai	- -	7.1% (9/126)	48.5% (32/66)	33.3% (1/3)

être + adj.	- -	4.8% (1/21)	14.1% (9/64)	- -
peut-être	- -	42.9% (9/21)	30.4% (17/56)	41.2% (7/17)
pouvoir dire	- -	0% (0/9)	0% (0/6)	20.0% (2/10)
se dire	- -	0% (0/5)	14.3% (1/7)	14.3% (1/7)
vouloir dire	- -	**76.9%** (20/26)	64.2% (43/67)	14.3% (1/7)

In the analysis that follows, potential discourse markers are separated from the non-discourse markers after the full data set from each region is analyzed. The verb forms that can be part of a complementizer construction or can be used as a discourse marker are known to be frequent (Blondeau & Nagy 2008:297). Table 3.2 provides the log frequency from the French Lexicon Project (Ferrand et al. 2010) for these verb forms where available. It should be noted that each of the verb forms listed includes the frequency for both the first and second person since these verb forms do not differ. For example, je *crois* can be a discourse marker. *Crois* is the conjugation not only for the first person singular but also for the second person singular, which cannot be a discourse marker. The log frequencies indicate that each of these forms is more frequent than the average for that particular verb or for any other verbal element shown in Table 3.1. *Sais* is the most frequent and has a very high absent *que* rate in each region. However, *crois* has very low rates even though it is the second most frequent form. Aside from *vois* in Europe and *crois*, the overall absent *que* rate for each of these forms decreases as frequency decreases. There may be a correlation between frequency and the absent *que* rate for the PDM forms, although there are a few exceptions.

Table 3.2. Matrix Verbal Element Frequency and Absent *Que* Rates by Region for PDM forms

Matrix Verbal Element	Log Frequency	Absent *Que* Rate Europe	Absent *Que* Rate Quebec	Absent *Que* Rate Africa
sais (*savoir*)	3.4	**74.4%** (29/39)	**91.6%** (196/214)	**71.4%** (5/7)
crois (*croire*)	3.0	10.3% (6/58)	0% (0/7)	40.0% (6/15)
vois (*voir*)	2.8	**98.1%** (51/52)	41.7% (5/12)	66.7% (2/3)
pense (*penser*)	2.7	15.6% (15/96)	32.5% (52/160)	7.7% (1/13)
trouve (*trouver*)	2.5	15.2% (5/33)	15.4% (6/39)	0% (0/3)

Based on the frequencies shown above, all verbs in the present study that have an entry in the French Lexicon Project (Ferrand et al. 2010) have a higher-than-average frequency, and the discourse marker forms have the highest frequencies. Overall, there does not appear to be a direct correlation between frequency and absent *que* rate. However, frequency may play a role in absent *que* rates for PDM forms.

3.2.2. Europe

European speakers used absent *que* in 192 out of 760 instances for an overall absence rate of 25.3%. A mixed-methods analysis of the full data set for European speakers was carried out along with two subsets of the data. These two subsets consist of all instances of verbal elements that cannot be used as discourse markers (non-discourse markers or NDMs) and all forms that can be used as discourse markers (potential discourse markers or PDMs). As mentioned in Section 1.7, these two subsets are analyzed to provide a more nuanced look at the data since it is possible that some discourse markers were inadvertently included in the data for this study. For each of the three data sets, a generalized linear mixed-effects model (GLMM) was fit as outlined in Section 2.6. Initially, I fit a GLMM with eight fixed effects consisting of the following social and linguistic factors: region, education level, age group, phonological context, syntactic context,

intervening material in the matrix clause, intervening material in the embedded clause, and matrix verbal element. Speaker was a random effect. The reference level remained at the default setting of alphabetical except for education and syntactic context. These were changed due to a much smaller number of tokens for the default level than the other levels. The reference level for education was university and for syntactic context it was pronoun.

Since a model with all these predictors did not converge for the full data set due to the large number of variables and considerable number of verbal elements, I reduced the levels of the matrix verbal elements by grouping them according to semantic properties. The groups were a combination of those put forth by Deulofeu and Véronis (2002:390) and my own categorization. The following were placed into the epistemic verbs category: *savoir*, *sembler*, *avoir l'impression*, *croire*, *c'est vrai*, *être* + adj., *paraître*, and *voir*. While *voir* 'to see' is an evidential marker, it was placed in the epistemic verbs category because of its likeness to *savoir* 'to know'. Examples (67) and (68) demonstrate this similarity.

(67) C'était très très dur, donc euh, quand j'ai vu qu'en quatre-vingt-six, on nous supprimait l'allocation aux, aux jeunes enfants... (EF48)
'It was very very hard, so uh, when I saw that in '86, they were getting rid of the allowance for young children on us…'

(68) vous voyez il y a deux ans j'aurais eu du mal à parler comme ça (EM17)
'You see two years ago I would have had a hard time talking like this'

Penser and *trouver* were classified as judgment verbs and *dire*, *se dire*, *pouvoir dire*, *c'est-à-dire*, *vouloir dire*, and *c'est* were grouped as verbs of speaking or restatement. *Falloir*, *promettre*, and *vouloir* were considered desired action verbs, while *heureusement* and *peut-être* were classified as adverbs.

For each of the three data sets, I first tried to fit a GLMM with all eight predictor variables. If it would not converge, I fit all eight predictor variables using verbal semantic groups instead of the individual verbal elements. If this was not successful, I fit a model with just the social predictor variables (region, education, and age) and verbal semantic groups. Finally, if none of these models converged, I fit a model with matrix verb as a random effect. This decision was made to reduce the degrees of freedom and make it possible to determine other factors that may significantly affect the complementizer form. Without fitting a GLMM in this manner it would not have been possible to obtain any useful results for one of the subsets. In the event that the variance of a random effect was very near zero (< 0.03), the effect was removed from the final model.

For each data set, I compared models based on AIC and also looked at the C index and Somers' Dxy for goodness of fit. To corroborate the p-values of each GLMM, I used 95% confidence intervals. If the confidence interval for a factor level contained zero, it was considered significant. Additionally, I checked for multicollinearity, the occurrence of high intercorrelations among two or more independent variables, by looking at the Variance Inflation Factors (VIF-scores) for each fixed effect. While there are different schools of thought on the upper VIF-score limit, I chose the conservative upper limit of 2.5 (Levshina 2015:272; Johnston et al. 2018:1958-1959). For each of the three final models, the confidence intervals confirmed the significance or lack thereof indicated by the GLMM p-values. Multicollinearity was not a concern for any of the models. The C index, Somers' Dxy, confidence intervals, and VIF-scores for the final models can be found in Appendix A. The summary statistics for the best model for each data set are shown below in Tables 3.1-3.3 as a point of reference for the discussion to follow in Sections 3.3 and 4.2.

Table 3.3 begins with the goodness of fit statistics followed by the variance and standard deviation of the random effects. The goodness of fit statistics were used to determine the best fit relative to the other models. The variance of a random effect indicates how much variation in the data is due to this factor. All fixed effects appear below the random effects. The intercept indicates the expected value of the response variable when all categorical variables are at their reference levels. Since the response variable represents the odds of the absence of *que*, a positive intercept indicates that there is a greater chance for the presence of *que*. A positive regression coefficient, found in the Estimate column, indicates an increased chance of present *que*. Conversely, a negative coefficient indicates an increased chance of *que* absence. The standard error represents the statistical accuracy of the estimate where zero indicates that the estimated value is exactly the true value. In the following column, the z-value indicates the number of standard deviations between a point and the mean. A z-value of zero indicates that the point is identical to the mean. A z-value of one is one standard deviation above the mean. The final column contains p-values, which indicate the probability of the absence of *que* for a certain factor level. For example, the p-value for the region of Southern France indicates the statistical probability of the absence of *que* when the region is Southern France compared to the reference level, Northern France. In this case, the probability is not significant. For the purpose of the results and discussion that follow in Sections 3.3 and 4.2, I will concentrate on the variance of the random effects, and the estimates and p-values for the fixed effects. All GLMMs have the same response variable, absence of *que*, so all tables of summary statistics should be interpreted in the manner described for Model 1E, shown in Table 3.3. Since it was not possible to fit a GLMM for each data set with the same random and fixed effects for the reasons outlined above, the following models are not identical with regard to included factors. The model for the full

European data set indicates that judgment verbs significantly increase the chance of present *que* compared to epistemic verbs.

Table 3.3. Summary of Generalized Linear Mixed-Effects Model 1E for the full European data set

	AIC	log likelihood		
Goodness of fit	770.7	-373.3		
Random effects	Variance	Standard Deviation		
Speaker	0.5908	0.7686		
Fixed effects	Estimate	Standard Error	z-value	p-value
Intercept	1.25145	0.453115	2.762	0.005751**
Region-Southern France	0.15159	0.51921	0.292	0.770316
Region-Switzerland	-0.34757	0.39366	-0.883	0.377291
Education-Primary	-0.08673	0.71806	-0.121	0.903858
Education-Secondary	0.15539	0.36741	0.423	0.672345
Age Group-Older	-0.07732	0.45121	-0.171	0.863932
Age Group-Younger	-0.31986	0.42773	-0.748	0.454577
Matrix Verb-Adverbs	-0.73476	0.48900	-1.503	0.132942
Matrix Verb-Desired Action	17.87303	109.15890	0.164	0.869941
Matrix Verb-Judgment	1.01488	0.29117	3.486	0.000491***
Matrix Verb-Speaking/Restatement	-0.14719	0.23146	-0.636	0.524818

Reference Levels: Region-Northern France, Education-University, Age Group-Middle, Matrix Verb-Epistemic

As detailed in Section 1.7.3, seven verbs chosen for this study can be used as discourse markers, and prosodic cues are sometimes the only way to determine the difference between a complementizer construction when *que* is absent and a discourse marker, which never appears before the complementizer. Vincent (1991:47), Martineau (1993:84), and Andersen (1996:307-309) identified the following verb forms as potential discourse markers: *je crois* 'I believe', *je*

pense 'I think', *je ne sais pas* 'I don't know', *je trouve* 'I believe/feel', *je veux dire* 'I mean', *je te dis* 'I tell you', *tu sais* 'you know', *tu vois* 'you see', and *voyons* 'let's see'. Even though criteria were developed to reduce the number of discourse markers included in this study, there remains the possibility that some are included in the data. Therefore, these forms were removed to create a data subset of NDMs. This subset includes 439 tokens, 47 speakers, and 21 matrix verbal elements. Thus, 321 tokens from the full data set were removed. The overall rate of absence of *que* for this data set is 12.1%, which is 13% lower than the full data set rate of 25.3%. The best model for this subset based on AIC is shown in Table 3.4. None of the social factors is statistically significant.

Table 3.4. Summary of Generalized Linear Mixed-Effects Model 2E for the European NDM subset

	AIC	log likelihood		
Goodness of fit	295.1	-139.5		
Random effects	Variance	Standard Deviation		
Matrix Verb	3.225	1.796		
Fixed effects	Estimate	Standard Error	z-value	p-value
Intercept	2.93467	0.75905	3.866	0.000111***
Region-Southern France	0.39699	0.67462	0.588	0.556219
Region-Switzerland	-0.24111	0.43482	-0.555	0.579229
Education-Primary	-0.62704	0.73491	-0.853	0.393543
Education-Secondary	0.04467	0.41158	0.109	0.913570
Age Group-Older	-0.11786	0.52816	-0.223	0.823411
Age Group-Younger	0.32192	0.44974	0.716	0.474118

Reference Levels: Region-Northern France, Education-University, Age Group-Middle

With the objective of better understanding how the PDMs might differ from the other verbal elements in the data set, I analyzed these dual function verb forms listed above. There are

321 tokens from 45 speakers and a total of 7 verbs. The overall absent *que* rate is 43.3%, which is 18% greater than the full data set rate of 25.3%. The best model for this subset is shown in Table 3.5. One observes that the verbs *dire, savoir, vouloir,* and *voir* strongly favor absence of *que* compared to *croire* and that the phonological context is significant.

Table 3.5. Summary of Generalized Linear Mixed-Effects Model 3E for the European PDM subset

	AIC	log likelihood		
Goodness of fit	269.1	-112.6		
Random effects	Variance	Standard Deviation		
Speaker	0.4025	0.6344		
Fixed effects	Estimate	Standard Error	z-value	p-value
Intercept	2.7410	0.8802	3.114	0.00185**
Region-Southern France	-0.3222	0.8920	-0.361	0.71791
Region-Switzerland	-0.5064	0.6023	-0.841	0.40054
Education-Primary	-1.2253	1.0371	-1.181	0.23745
Education-Secondary	0.1629	0.5541	0.294	0.76870
Age Group-Older	-0.3873	0.6666	-0.581	0.56123
Age Group-Younger	-0.8739	0.6766	-1.292	0.19652
Phonological-Liquid	1.4957	1.1342	1.319	0.18726
Phonological-Nasal	4.5987	1.9101	2.408	0.01606*
Phonological-Stop	1.1274	0.7334	1.537	0.12425
Phonological-Vowel	0.6347	0.4174	1.520	0.12839
Syntactic-Lexical Noun	0.0831	0.9775	0.085	0.93225
Syntactic-Other	2.8806	1.4908	1.932	0.05332
IM Matrix-Yes	0.8833	0.9427	0.937	0.34878
IM Embedded-Yes	-1.9657	1.2627	-1.557	0.11952
Matrix Verb-*Dire*	-3.1800	0.7574	-4.199	2.69e-05***
Matrix Verb-*Penser*	-0.7836	0.6044	-1.297	0.19480
Matrix Verb-*Savoir*	-3.8127	0.7191	-5.302	1.15e-07***
Matrix Verb-*Trouver*	-0.7709	0.7621	-1.011	0.31178
Matrix Verb-*Vouloir Dire*	-4.8236	1.0133	-4.760	1.93e-06***
Matrix Verb-*Voir*	-7.2575	1.3640	-5.321	1.03e-07***

Reference Levels: Region-Northern France, Age Group-Middle, Phonological-Fricative, Syntactic-Pronoun, IM Matrix-No, IM Embedded-No, Matrix Verb-Croire

3.2.3. Quebec

For a quantitative analysis of the Quebec data, the full data set plus the same two subsets as for the European data were analyzed. Quebec speakers used absent *que* in 470 out of 1,527 instances for an overall absence rate of 30.8%. A generalized linear mixed-effects model was fit for each of the three data sets according to the process outlined in Section 3.2.2 for Europe. The following variables were initially considered as fixed effects: education level, age group, phonological context, syntactic context, intervening material in the matrix clause, intervening material in the embedded clause, and matrix verbal element. These are the same fixed effects that were considered for the European data except region, which was not available for Quebec. The reference levels for the fixed effects were the same as for the European data except for the PDM forms. For these verb forms, the reference level was *savoir* since there were only seven tokens of *croire*. Speaker was considered a random effect. Due to a lack of convergence, the semantic groups for verbal elements mentioned in Section 3.2.2 were used in place of the individual verbal elements. The verbal element *avoir raison* was added to the judgment group and *jurer* was added to the desired action group because of its similarity in use to *promettre*.

The C index and Somers' Dxy, 95% confidence intervals, and VIF-scores for each of the three best models are shown in Appendix B. The confidence intervals confirmed the significance or non-significance of factors that were indicated by the p-values for each model. The VIF-scores revealed that multicollinearity was not a concern for any of the models. Below, Table 3.6 displays summary statistics for the best model for the full data set. Desired action, judgment, and speaking/restatement verbs significantly increase the chance of present *que* as do adverbs. Vowels strongly increase the chance of the present complementizer while liquids and nasals do to a lesser extent.

Table 3.6. Summary of Generalized Linear Mixed-Effects Model 1Q for the full Quebec data set

	AIC	log likelihood		
Goodness-of-fit	1635.7	-799.9		
Random effects	Variance	Standard Deviation		
Speaker	0.2189	0.4679		
Fixed effects	Estimate	Standard Error	z-value	p-value
Intercept	-0.68609	0.21391	-3.207	0.00134**
Education-Primary	-0.25789	0.42253	-0.610	0.54162
Education-Secondary	0.11065	0.22536	0.491	0.62344
Age Group-Older	0.12089	0.33497	0.361	0.71818
Age Group-Younger	0.22814	0.22823	1.000	0.31750
Phonological-Liquid	1.07654	0.42416	2.538	0.01115*
Phonological-Nasal	1.07377	0.49112	2.186	0.02879*
Phonological-Stop	0.01181	0.20114	0.059	0.95317
Phonological-Vowel	0.73330	0.14756	4.969	6.72e-07***
Syntactic-Lexical Noun	-0.19771	0.37492	-0.527	0.59795
Syntactic-Other	0.13801	0.47226	0.292	0.77012
IM Matrix-Yes	0.20495	0.23863	0.859	0.39043
IM Embedded-Yes	-0.29568	0.42532	-0.695	0.48694
Matrix Verb-Adverbs	0.92725	0.32069	2.891	0.00383**
Matrix Verb-Desired Action	2.54140	0.21066	12.064	< 2e-16***
Matrix Verb-Judgment	1.55937	0.17800	8.761	< 2e-16***
Matrix Verb-Speaking/Restatement	1.35340	0.17799	7.604	2.87e-14***

Reference Levels: Education-University, Age Group-Middle, Phonological-Fricative, Syntactic-Pronoun, IM Matrix-No, IM Embedded-No, Matrix Verb-Epistemic

Following the analysis of the full data set, the verbal elements that cannot be discourse markers were analyzed. This subset includes 1,023 tokens, 46 speakers, and 23 verbal elements. Therefore, there were 504 tokens from the full data set that were removed. The overall rate for absence of *que* is 16.0%, which is 14.8% lower than the full data set rate of 30.8%. The best model for this subset, Model 2Q, is shown in Table 3.7. In this model, matrix verb is a random effect since the model did not converge when it was a fixed effect. The speaker variance was

zero when included as a random effect. Therefore, it is not included in the final model. None of the factors included in this model is statistically significant.

Table 3.7. Summary of Generalized Linear Mixed-Effects Model 2Q for the Quebec NDM subset

	AIC	log likelihood		
Goodness-of-fit	914.1	-443.1		
Random effects	Variance	Standard Deviation		
Matrix Verb	0.1847	0.4298		
Fixed effects	Estimate	Standard Error	z-value	p-value
Intercept	1.67970	0.23373	7.187	6.64e-13***
Education-Primary	0.24417	0.39049	0.625	0.5318
Education-Secondary	-0.01818	0.19817	-0.092	0.9269
Age Group-Older	-0.37129	0.29613	-1.254	0.2099
Age Group-Younger	0.11461	0.20476	0.560	0.5757
Phonological-Liquid	-0.41019	0.60083	-0.683	0.4948
Phonological-Nasal	-0.73507	0.73771	-0.996	0.3190
Phonological-Stop	-0.13300	0.24776	-0.537	0.5914
Phonological-Vowel	0.35985	0.20905	1.721	0.0852
Syntactic-Lexical Noun	0.68583	0.58097	1.180	0.2378
Syntactic-Other	0.26661	0.81968	0.325	0.7450
IM Matrix-Yes	-0.18199	0.27401	-0.664	0.5066
IM Embedded-Yes	-0.51777	0.75455	-0.686	0.4926

Reference Levels: Education-University, Age Group-Middle, Phonogical-Fricative, Syntactic-Pronoun, IM Matrix-No, IM Embedded-No

The final model to be fit was the subset of PDM forms. This data set includes 495 tokens from 43 speakers, a total of 7 verbs, and an overall *que* absence rate of 61.8%, which is 31% higher than the full data set rate of 30.8%. Model 3Q, the best model for this subset, is summarized in Table 3.8. The matrix verb is a random effect along with speaker since a model with matrix verb as a fixed effect would not converge. Stops and nasals following the

complementizer as well as the presence of intervening material in the matrix clause increase the chance of present *que*.

Table 3.8. Summary of Generalized Linear Mixed-Effects Model 3Q for the Quebec PDM subset

	AIC	log likelihood		
Goodness-of-fit	428.9	-199.5		
Random effects	Variance	Standard Deviation		
Speaker	0.2739	0.5234		
Matrix Verb	6.2604	2.5021		
Fixed effects	Estimate	Standard Error	z-value	p-value
Intercept	-1.0514	1.0452	-1.006	0.314460
Education-Primary	-0.5589	0.7035	-0.794	0.426936
Education-Secondary	0.7330	0.3908	1.876	0.060685
Age Group-Older	0.5625	0.5528	1.018	0.308882
Age Group-Younger	0.2217	0.3800	0.583	0.559687
Phonological-Liquid	1.3856	0.9967	1.390	0.164468
Phonological-Nasal	0.6297	1.1440	0.550	0.582010
Phonological-Stop	1.3805	0.5650	2.443	0.014549*
Phonological-Vowel	1.2982	0.3433	3.782	0.000156***
Syntactic-Lexical Noun	0.3510	0.8450	0.415	0.677867
Syntactic-Other	0.9103	0.9627	0.946	0.344377
IM Matrix-Yes	1.1814	0.5617	2.103	0.035443*
IM Embedded-Yes	-0.6061	0.8214	-0.738	0.460586

Reference Levels: Education-University, Age Group-Middle, Phonological-Fricative, Syntactic-Pronoun, IM Matrix-No, IM Embedded-No

3.2.4. Africa

The African speakers used absent *que* in 57 out of 274 instances for an overall absence rate of 20.8%. For the mixed-methods analysis of the African data, the full data set as well as the same subsets as for the European and Quebec data were analyzed. For each of the three data sets, a generalized linear mixed-effects model was fit using the same process as outlined in Section

3.2.2 The following seven variables were initially considered as fixed effects: region, age, phonological context, syntactic context, intervening material in the matrix clause, intervening material in the embedded clause, and matrix verbal element. These are the same fixed effects that were considered for the European data except that education was not available for the African speakers. The reference levels for the fixed effects were the same as for the European speakers except for the matrix verb. Instead of *voir*, the reference level was *falloir* since there were relatively few tokens of *voir*. The reference level for region was Burkina Faso. There was initially one random effect, which was speaker. Matrix verb was added as a random effect in each model since the models would not converge when it was a fixed effect.

The C index and Somers' Dxy for each of the three best models are shown in Appendix C along with the confidence intervals and VIF-scores. The 95% confidence intervals for each model confirmed the statistical significance or non-significance indicated by the p-values for each model. The VIF-scores indicated that multicollinearity was not a concern for any of the models. A summary of the best model for the full data set, Model 1A, is shown in Table 3.9. For the full African data set, liquids increase the chance of the present complementizer as do words in the other syntactic category.

Table 3.9. Summary of Generalized Linear Mixed-Effects Model 1A for the full African data set

	AIC	log likelihood		
Goodness-of-fit	265.8	-118.9		
Random effects	Variance	Standard Deviation		
Speaker	0.1942	0.4406		
Matrix Verb	1.0254	1.0126		
Fixed effects	Estimate	Standard Error	z-value	p-value

Intercept	1.1565	0.4905	2.358	0.0184*
Region-Central African Republic	-0.2485	0.7069	-0.352	0.7252
Region-Côte d'Ivoire	-0.5899	0.4696	-1.256	0.2090
Age Group-Younger	0.6064	0.5126	1.183	0.2368
Phonological-Liquid	2.2344	1.0823	2.065	0.0390*
Phonological-Nasal	1.1830	1.1498	1.029	0.3035
Phonological-Stop	0.1646	0.6389	0.258	0.7966
Phonological-Vowel	0.8218	0.4400	1.868	0.0618
Syntactic-Lexical Noun	-1.7196	0.9545	-1.802	0.0716
Syntactic-Other	1.8825	0.9257	2.034	0.0420*
IM Matrix-Yes	2.0504	1.1799	1.738	0.0823
IM Embedded-Yes	-1.3363	0.8750	-1.527	0.1267

Reference Levels: Region-Burkina Faso, Age Group-Middle, Phonological-Fricative, Syntactic-Pronoun, IM Matrix-No, IM Embedded-No

The subset of matrix verbal forms that cannot be discourse markers was analyzed next. This data set includes 230 tokens, 25 speakers, and 18 verbal elements. Therefore, 44 tokens from the full data set were removed. The rate of absence of *que* is 17.4%, which is 3.4% lower than for the full data set rate of 20.8%. Summary statistics for the best model for this data set, Model 2A, are shown in Table 3.10. Liquids increase the chance of the present complementizer compared to fricatives.

Table 3.10. Summary of Generalized Linear Mixed-Effects Model 2A for the African NDM subset

	AIC	log likelihood		
Goodness-of-fit	208.6	-90.3		
Random effects	Variance	Standard Deviation		
Speaker	0.1994	0.4465		
Matrix Verb	0.9181	0.9582		
Fixed effects	Estimate	Standard Error	z-value	p-value
Intercept	1.288829	0.554418	2.325	0.0201*

Region-Central African Republic	0.482853	0.895465	0.539	0.5897
Region-Côte d'Ivoire	-0.582308	0.556836	-1.046	0.2957
Age Group-Younger	0.606260	0.576972	1.051	0.2934
Phonological-Liquid	2.537611	1.291518	1.965	0.0494*
Phonological-Nasal	1.605104	1.483416	1.082	0.2792
Phonological-Stop	0.006751	0.696006	0.010	0.9923
Phonological-Vowel	0.721039	0.519896	1.387	0.1655
Syntactic-Lexical Noun	-1.934436	1.156127	-1.673	0.0943
Syntactic-Other	2.064078	1.100814	1.875	0.0608
IM Matrix-Yes	1.819314	1.206876	1.507	0.1317
IM Embedded-Yes	-1.875603	1.007614	-1.861	0.0627

Reference Levels: Region-Burkina Faso, Age Group-Middle, Phonological-Fricative, Syntactic-Pronoun, IM Matrix-No, IM Embedded-No

The final subset that was analyzed included all matrix verb forms that can be discourse markers. This data set contains 44 tokens, 16 speakers, 6 verbs, and has an absent *que* rate of 38.6%, which is 17.8% higher than the full data set rate of 20.8%. There were no instances of *vouloir dire* for the African speakers. Table 3.11 provides a summary of Model 3A, which was determined to be the best fit for the data. None of the social factors is statistically significant.

Table 3.11. Summary of Generalized Linear Mixed-Effects Model 3A for the African PDM subset

	AIC	log likelihood		
Goodness-of-fit	63.3	-25.7		
Random effects	Variance	Standard Deviation		
Speaker	0.1845	0.4296		
Matrix Verb	3.2760	1.8100		
Fixed effects	Estimate	Standard Error	z-value	p-value
Intercept	0.3252	1.1737	0.277	0.782
Region-Central African Republic	-2.3522	1.8812	-1.250	0.211
Region-Côte d'Ivoire	-0.3176	1.0805	-0.294	0.769
Age Group-Younger	1.8764	1.9122	0.981	0.326

Reference Levels: Region-Burkina Faso, Age Group-Middle

3.3. Social Factors

3.3.1. Results and Discussion of Social Factors

3.3.1.1. Europe

The 48 European French speakers produced *que* in 568 of the 760 environments where it was possible, resulting in a *que* absence rate of 25.3% (n=192). Table 3.12 provides a summary of the *que* absence rates for each social factor. Speakers from Northern France have the highest rate, while those from the South have the lowest. These results align with the notion that speakers from the South pronounce more sounds than those from the North. For example, Myers and Ranson (2014) found that speakers from the North delete postconsonantal /ʁ/ and /l/, as in *autr'* and *tabl'*, more than those from the South. Comparing the results of Ranson and Passarello (2012), who looked at word-final schwa in speakers from Southern France to the results of Green and Hintze (2021) and Hutin et al. (2021), who looked at word-final schwa in Northern France, those from the North have a higher absence rate (75.7% and 92.32% respectively) than speakers from the South (65%). Furthermore, Josephs and Ranson (2020) found that Southern speakers have a higher rate of liaison (25.7%) compared to Northern speakers (16.3%). Since liaison is the pronunciation of a latent consonant before a vowel, such as the final /t/ of *est* in *c'est une langue*, the Southern speakers produce an additional sound more often than the Northern speakers. Speakers from Switzerland fall between the two groups from France with an absent *que* rate of 20.4%.

Table 3.12. Europe - Overall absence of *que* as percentages for social factors

Social Factor	% Ø	N=Ø/Total
Region		
Northern France	26.7	116/435
Southern France	20.4	11/54

Switzerland	24.0	65/271
Education Level		
primary	22.9	16/70
secondary	21.2	75/354
university	30.1	101/336
Age Group		
younger	29.5	106/359
middle	19.5	42/215
older	23.7	44/186

Results for education level show the highest absent *que* rate for those with an education at the university level followed by the primary level. Martineau (1985) also found that participants with a university degree had the highest absence rate while those who had finished secondary school had the lowest. However, this factor was not statistically significant in her study, just as it is not in the present study. Liang et al. (2021) found the opposite to be true, to a statistically significant level, so that those with the highest level of education had the lowest absent *que* rate. Warren (1994) found that of the three education levels she considered, those with an average level, which corresponded to the early stages of university, had the highest rate of absent *que*, regardless of occupation. She concluded that absence of *que* among Quebec speakers was not a misuse of the language due to a lower education level. Rather, it was one option that had developed within a systematic grammar distinctive to Quebec. In the present study, those considered to be average would be secondary-educated speakers. However, this is not the same level of education that Warren (1994) referred to. In the present study, those with one or two years of a university education are combined with those who have completed university or gone on for more advanced degrees. Nonetheless, since the most educated speakers have the highest absence rate, it is not a lack of formal education that contributes to the absent variant.

As for age, the youngest age group has the highest absent *que* rate followed by the older group and then the middle age group. Martineau (1985) and Liang et al. (2021) also found that the youngest speakers had the highest rate of absent *que* with rates of 36% and 32.9%, respectively. The middle and older age groups in Martineau's (1985) study had very similar rates to one another at 31% and 30% respectively. Liang et al. (2021) found a rate of 24.8% for the middle group and 17.6% for the older group. While age was not found to be significant for either of these studies, the ordering of age groups was the same. For the present study, the middle and older groups are ordered differently. It is possible that the middle age group has the lowest absent *que* rate due to professional needs to use the standard variant, since this group represents those who are established in a career.

All three social factors are included in the linear models of the three European data sets. However, none of these factors is significant regardless of the data set. For the full data set and the PDM data set, the matrix verb is statistically significant. In the absence of any statistically significant social factors, the rate of absent *que* for the NDM and PDM verbal elements according to social factors will allow us to determine social trends based on the two types of matrix verbs. Table 3.13 provides the rate of absent *que* for both of these groups.

Table 3.13. Europe - Absence of *que* as percentages for social factors according to type of verbal element

Non-Discourse Markers (NDM)			Potential Discourse Markers (PDM)		
Social Factor	% Ø	N=Ø/Total	Social Factor	% Ø	N=Ø/Total
Region			Region		
Northern France	11.8	32/272	Northern France	**51.5**	84/163
Southern France	11.8	4/34	Southern France	**35.0**	7/20
Switzerland	12.8	17/133	Switzerland	**34.8**	48/138
Education Level			Education Level		
primary	20.4	10/49	primary	**28.6**	6/21
secondary	10.0	23/229	secondary	**41.6**	52/125

university	12.4	20/161	university	**46.3**	81/175
Age Group			Age Group		
younger	9.4	18/192	younger	**52.7**	88/167
middle	11.6	15/129	middle	**31.4**	27/86
older	16.9	20/118	older	**35.3**	24/68

A regional comparison indicates that Switzerland has the highest absent *que* rate for the NDM forms and the lowest rate for the PDM forms. All three regions have very similar rates for the NDM forms, and both regions in France have identical rates for them. For the PDM forms, Southern France and Switzerland have almost identical rates but differ considerably from Northern France. Thus, the greatest regional divergence is seen with the PDM forms where Northern France has a much higher rate than the other two regions and where Southern France and Switzerland are nearly identical. The most striking result is the far higher rates of absent *que* with PDMs in all regions, especially Northern France.

Regarding education, a university level education is associated with a slightly higher absent *que* rate than a secondary level regardless of the verbal elements being considered. Speakers with a primary level of education have a rate 8% higher than the next group for the NDM forms and a rate 13% lower than the next group for the PDM verb forms. It should be noted that there are just four speakers with a primary level of education, which is considerably fewer than for the other education levels. The results for the NDMs are consistent with the pattern for a stable linguistic variable. This type of variable exhibits a monotonic (linear) distribution whereby the highest use of the non-standard variant is among the lowest class, followed by the middle class, and then the upper class (Dodsworth 2011:194). Stable variables may also pattern into two broad groups: working class and middle/upper class (Dodsworth 2011:194). Those with a primary level of education would represent the working class while

those with a secondary or university level education would represent the middle and upper class. Secondary- and university-educated speakers pattern very similarly for both types of verbs.

It should be noted that the distinction between a monotonic and curvilinear distribution is reliable only for adult speakers since adolescents are known to have extremely high rates of stigmatized variants (Dodsworth 2011:196). According to the World Health Organization, an adolescent ranges in age from 10 to 19 years (WHO 2023). Among the European speakers, there are three who are considered adolescents (EF16, EM17, EM19). Of these three, only one (EM17) has a relatively high rate of absent *que* (26.7%) compared to the group average of 15.1% for younger males. Therefore, it is unlikely that this one speaker invalidates the monotonic pattern.

One possible explanation for a secondary and university level of education having the highest rates for the PDM forms, but not the NDM forms, is due to the frequency of the verbs that can be discourse markers, as shown above in Table 3.2. These verbs are commonly used in complementizer clauses and as discourse markers. Since discourse markers never appear with a *que*, speakers would be accustomed to hearing and using these verbal forms without a following *que*. It is proposed that even when used as a matrix verb, these collocations frequently appear without *que* due to syntactic analogy with the discourse marker. Analogy in general refers to "a general cognitive process that transfers specific information or knowledge from one instance or domain to another" (Blevins & Blevins 2009:2). Syntactic analogy occurs when analogous portions of sentences replace each other in different sentence constructions (Blevins & Blevins 2009:8). Psycholinguists posit that a speaker's choice between alternative constructions is related to their experience with specific linguistic forms and meanings (Diessel & Hilpert 2016:1). Thus, the experience of French speakers with the discourse markers may make them more susceptible to use absent *que* with the same forms when used in a complementizer construction. Further

support comes from cognitive linguists who assert that the productivity of a grammatical schema is determined by frequency and similarity (Diessel & Hilpert 2016:1). Potential discourse markers serve a pragmatic role whether used as a discourse marker or as part of a complementizer construction. In a complementizer clause they indicate subjectivity and relay the attitude of the speaker with regard to what follows (Benveniste 1966:264). Discourse markers serve a variety of pragmatic functions, such as the introduction of new information, reformulation, hesitation, and beginning a turn in the discourse. Thus, these forms are both frequent and similar. Returning to the education results for European speakers, it may be more acceptable, even among the most educated, not to use *que* in a complementizer construction with the PDM forms. Those with a primary education have the lowest absent *que* rate for the more common, dual function PDM forms while they have the highest rate for the NDM forms. Since the NDM forms do not have a dual function, there may be more of a stigma attached to an absent *que* with these forms. In this case, it would be expected that those with a lower level of education would be more prone to use the nonstandard variant than those with a higher level of education.

As will be shown in Section 5.2, there are eight European speakers who never have an absent *que*. Of these eight, all except one have a university level education. The one exception has a secondary level education. Thus, 28% (7/25) of all university-educated speakers never have an absent *que*, or, in other words, always have a present complementizer. Of those who do use the absent variant, they use it at a much higher rate for the frequent, dual function PDM forms. Therefore, there does appear to be a distinction based on education level and type of matrix verbal element, although this distinction is not statistically significant.

Table 3.14 shows the number of speakers by region and education level. A university level education is highly represented by speakers from Northern France. Compared to speakers

from Southern France, there are twice as many secondary-educated Northern speakers and three times as many university-educated Northern speakers. The number of speakers for Northern France and Switzerland are more similar except for a primary education where Swiss speakers are not represented. The trends for a university education level that can be extracted from the European data primarily represent Northern speakers. The same is true for a primary education level.

Table 3.14. Number of European Speakers by Region and Education Level

	Primary	Secondary	University
Northern France	3	8	13
Southern France	1	4	4
Switzerland	0	7	8

Regarding age, the older group always has a higher absent *que* rate than the middle group. However, the younger group has the lowest rate for the NDM forms and the highest rate for the PDM forms. The younger group has a rate 17.4% higher than the next group for the PDMs. If one type of verb were more stigmatized than the other, the NDMs would be expected to be this group, as mentioned above with regard to education level. Also, the younger generation tends to use stigmatized variants more than middle-aged speakers (Labov 1994:73). However, this is not what these data indicate. Comparing the overall rates for all data shown in Table 3.12 with the PDM forms in Table 3.13, the same ordering is found with younger speakers having the highest rate followed by older speakers and then middle-aged speakers. Although the middle group has the second highest rate for NDM forms, it does not differ much from the lowest group (2.2% difference).

One final consideration for age is its interation with education level. Table 3.15 indicates that the younger group is made up primarily of university-educated speakers. The middle group is balanced between secondary and university-educated speakers. The older group is a little more heavily represented by secondary-educated speakers. Overall, the highest absent *que* rate corresponds to university-educated speakers and younger speakers. Table 3.15 indicates that it is the younger speakers who primarily determine the rates for university level, while the middle and older groups are the driving force behind the secondary level, and the older group for primary. The fact that the younger speakers are primarily from the North and university-educated has some bearing on the higher absent *que* rate for the PDM forms.

Table 3.15. Number of European Speakers by Age and Education Level

	Primary	Secondary	University
younger	0	3	13
middle	0	9	7
older	4	7	5

In summary, all three regions show similar absent *que* rates for the NDM forms, while Northern France diverges from the other two regions for only the PDM forms. Thus, these dual function forms are responsible for most of the variation by region. Speakers with a secondary or university level education have a lower absent *que* rate for NDMs than do those with a primary level of education. The opposite pattern is found for the verbal elements that can be discourse markers. Their dual function and frequency likely make an absent *que* less stigmatized. University-educated speakers are more likely to avoid the absent complementizer altogether, but those who do use it do so at a high rate. The overall pattern for age is the same as that for the PDM forms where the younger group has the highest absent *que* rate and the middle group has

the lowest. On the other hand, the NDM forms reveal that the older group has the highest rate, and the younger group has the lowest. The social trends shown in this section are in large part a reflection of younger, university-educated speakers from Northern France.

3.3.1.2. Quebec

The Quebec speakers produced a total of 1,527 sites where a complementizer was possible with a *que* absence rate of 30.8% (n=470). By comparison, Sankoff (1980) found an absence rate for *que* of 23% in Quebec, Liang et al. (2021) found a rate of 24.8% in the same province, and Martineau (1985) found it to be 32% in Quebec and Ontario. The present study's result is very similar to what Martineau (1985) found and between the ranges of the previous studies. There is no evidence that the overall absent complementizer rate has changed since these studies were conducted. Warren (1994) compared 16 speakers recorded in 1971 to the same speakers recorded in 1984 and found no change over real time. While the present study does not compare the same speakers as Warren (1994) did, it does provide a basis of comparison to the studies by Sankoff (1980), Martineau (1985), and Liang et al. (2021) by including speakers of various social groups recorded in an informal setting in the province of Quebec.

An overview of absent *que* rates by social factor is shown in Table 3.16. The highest absent *que* rate corresponds to the lowest level of education. However, those with a university education have a higher rate than those with only a secondary education, although the difference between these two groups is very small. There are a couple of comparisons with previous studies that indicate some similarity. Martineau (1985) found those with a university level of education to have an absence rate of 35% and those with a secondary level 30%. The overall rates for the present study indicate that a university and secondary level correspond to the rate she found for a

secondary level. Cedergren and Sankoff (1974) found that members of the working-class in Quebec used absent *que* more often than professionals, which would suggest that those with less education have a higher absence rate. This finding aligns with the overall education rates for the present study. Since speakers with a secondary level of education pattern almost the same as those with a university level, they can be grouped together. Based on this grouping, the overall rates indicate a stable linguistic variable where the least educated, or lower class, have a higher rate of the non-standard variant, the absent *que*, and the most educated, or middle and upper classes, have a lower rate of this variant.

Table 3.16. Quebec - Overall absence of *que* as percentages for social factors

Social Factor	% Ø	N=Ø/Total
Education Level		
primary	37.0	70/189
secondary	29.5	181/613
university	30.2	219/725
Age Group		
younger	28.0	172/614
middle	32.0	177/553
older	33.6	121/360

Regarding age, there is little difference among the groups. The older group has an absent *que* the most often and the younger group the least often. This pattern represents a monotonic distribution. Martineau (1985) and Liang et al. (2021) found the opposite monotonic pattern since their studies revealed that younger speakers had an absent *que* most often and older speakers the least. Figure 4.1 illustrates the distributions for absent *que* for these two previous studies and the present study. Two common types of age-related linguistic change are age grading and generational change. Both can appear as a monotonic distribution in apparent time,

which is the case for Martineau (1985) and Liang et al. (2021). While there are different definitions for age grading, I will use that of Labov (1994:84) which states that it is when individuals change their linguistic behavior throughout their lifetimes, but the community as a whole does not change. Generational change is when "individual speakers enter the community with a characteristic frequency for a particular variable, maintained throughout their lifetimes, but regular increases in the values adopted by individuals, often incremented by generations, lead to linguistic change for the community" (Labov 1994:84). If considering apparent time only, the two cannot be differentiated, but a real-time analysis can provide a means of disentangling the two (Labov 1994:84). For a real-time analysis, a trend study can be conducted whereby a sample of different but comparable speakers are recorded at each point in time (Wagner 2012:376). In the case of the present study, Martineau (1985) and Liang et al. (2021) are the most similar in that some or all of the speakers are from Quebec, represent different socioeconomic groups, include both sexes, and are in clearly defined age groups. There are certain differences such as number of speakers and precise domicile. However, all three studies provide social groups based on sex, age, and at least one socioeconomic indicator. The age groups in the present study are almost identical to those of Martineau (1985) and similar to Liang et al. (2021), and all three studies examine informal language. The participants in Martineau's (1985) study were from the provinces of Quebec and Ontario, those in Liang et al.'s (2021) study were from Montreal, and participants for the present study are from the province of Quebec. Thus, each study provides data for participants living in the province of Quebec. While a comparison of these two previous studies with the present is not a perfect scenario for a trend study, it can provide some insight into age-related change.

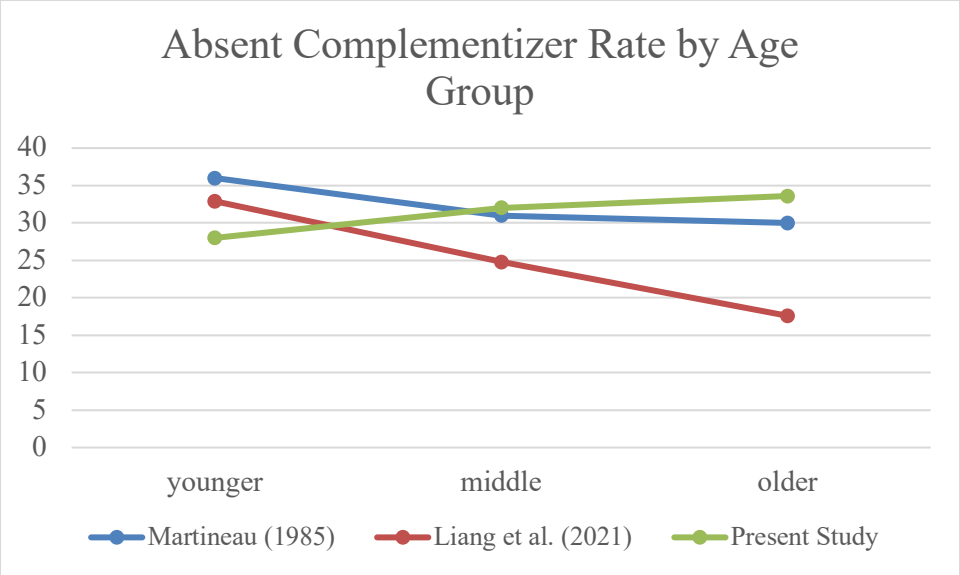

Figure 4.1. Absent complementizer rate by age group for two previous studies and the present study

Since the absent *que* rate trend by age group for the present study does not align with those of the two previous studies, Figure 4.1 makes it clear that the present study does not provide evidence of age grading. To investigate the possibility of generational change, the projected absent *que* rate is shown alongside the observed rates in Figure 4.2. The projected rate is based on the results of Martineau (1985). Most of the interviews for Martineau (1985) were carried out in 1982, whereas hose for the present study took place between 2006 and 2011. Thus, there is roughly a 25-year period between them. Assuming that the use of the absent complementizer was stable, the 1982 values are plotted, and then each value for 1982 is projected one unit to the right. These values are then compared to those observed in the present study, represented by the year 2010 in the figure. The projected rate and the observed rate for 2010 for the older group are extremely similar. However, the middle group is projected to have a higher rate than the older group, but the observed values show the opposite. Figure 4.3 provides the same information based on Liang et al.'s (2021) study. The projected and observed values for

the middle age group in 2010 are very similar. However, the older group has a higher rate rather than the lower one projected. Thus, there is no evidence of generational change based on these studies. The results of the present study do not indicate age grading or generational change compared to the results of Martineau (1985) or Liang et al. (2021). This could be due to differences in the study design and methodology or the age results for the present study could be due to a factor that is yet to be uncovered. It is important to keep in mind also that age is not a statistically significant factor in the present study.

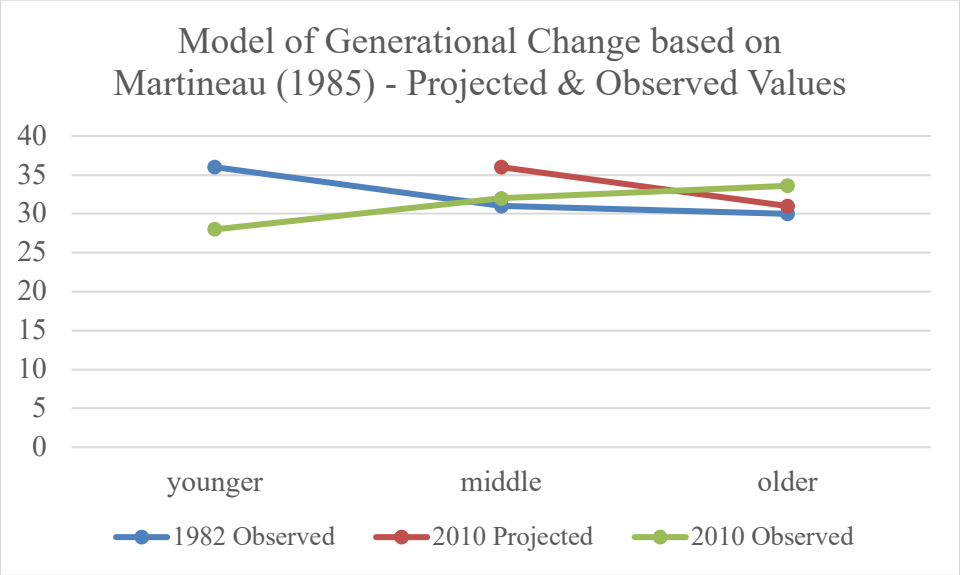

Figure 4.2. Model of generational change based on results of Martineau (1985)

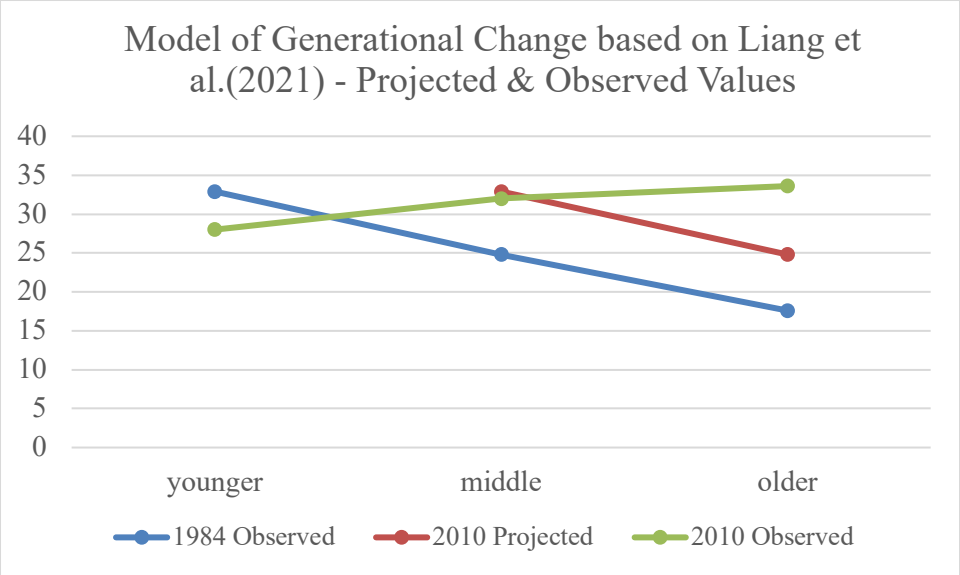

Figure 4.3. Model of generational change based on results of Liang et al. (2021)

Age and education, the two social factors considered for Quebec speakers, were included in all three of the mixed models shown in Section 3.2.3. Each model indicates that these factors are not significant. We now turn to a comparison of absent *que* rates for these social factors based on the two types of verbs, NDMs and PDMs.

Table 3.17 indicates that the ordering of education level based on absent *que* rates differs according to verb type. For the NDM forms, speakers with a secondary or primary level of education have almost identical rates, whereas the rates for speakers with a university education are roughly 1.5% higher. Even still, rates for each group are very similar. For the PDM forms, a primary level of education is associated with the highest absent *que* rate followed by university and then secondary. This ordering is in line with what Warren (1994) found, which indicated that a lack of formal education was not synonymous with a higher absent *que* rate. A university level education has a lower rate than a primary one for both verb types. Cedergren and Sankoff's (1974) finding that the working class had a higher absence rate for *que* than professionals holds

true for a primary versus university education level for both types of verbs but not a secondary-level education. Most of the complementizer variation is due to the PDM forms.

Table 3.17 Quebec - Absence of *que* as percentages for social factors according to type of verbal element

Non-Discourse Markers (NDM)			Potential Discourse Markers (PDM)		
Social Factor	% ∅	N=∅/Total	Social Factor	% ∅	N=∅/Total
Education Level			Education Level		
primary	16.7	19/114	primary	**70.7**	53/75
secondary	16.8	71/423	secondary	**57.9**	110/190
university	15.2	74/486	university	**62.2**	143/230
Age Group			Age Group		
younger	14.8	61/413	younger	**62.5**	120/192
middle	15.8	60/380	middle	**59.0**	102/173
older	18.7	43/230	older	**64.6**	84/130

There are three female Quebec speakers who qualify as adolescents (WHO 2023). These speakers have an absent *que* rate ranging from 20.0% to 27.9%. These rates are below the group average of 28.0% and below the average of every other social group except older males (27.4%). Excluding these three speakers, their social group has an average absent *que* rate of 30%. Thus, these three speakers do not have extremely high rates of the nonstandard variant. Therefore, the effect of these three speakers on rates by education level or age group should not lessen the validity of the results.

The numbers of Quebec speakers by age group and education level are shown in Table 3.18, which are used to investigate a possible interaction between the two factors. Primary-educated speakers are not represented among younger or middle-aged speakers, and university-educated speakers are under-represented among older speakers. Thus, age appears to be the underlying force that largely affects *que* absence or presence among primary- and university-

educated speakers. There is one speaker from Quebec (QM25b) who never produces an absent *que*, and this speaker has a university level education. This connection may purely be a coincidence. While a university level education results in a lower *que* absence rate than a primary level, this distinction appears to be largely due to age.

Table 3.18. Number of Quebec Speakers by Age and Education Level

	Primary	Secondary	University
Younger	0	6	10
Middle	0	6	10
Older	7	5	1

Turning now to age, the older age group has the highest absent *que* rate by 2-3% for both verb types. The younger age has the lowest for the NDM forms, while the middle group does for the PDM forms. The difference between the highest and lowest absence rates for each verb type is very small (~6%). However, there is a large difference in rates according to verb type. Thus, each age group does not differ much from the others, and all age groups have much higher absent *que* rates for PDMs.

In summary, the overall absent *que* rate of the present study falls between the ranges of the three previous studies that provide an overall rate for absent *que*. Therefore, the rate in Quebec has not changed since the 1970s and 1980s. Overall, speakers with a primary level of education use the non-standard absent *que* variant the most, while secondary- and university-educated speakers combined use it the least. According to matrix verb type, this holds true for the NDMs but not for the PDMs. Rather, these forms are used at very similar rates for all three education levels. Upon further investigation, it was found that a primary and university level of education are heavily influenced by age. For age, the older age group has the highest absent *que*

rate regardless of verb type. However, the middle group has the lowest rate for the PDM forms but the middle rate for the full data set and the NDM forms. There is little difference in absent *que* rates by age group for both types of matrix verbal element. Compared to Martineau (1985) and Liang et al. (2021), there is no evidence of age grading or generational change.

3.3.1.3. Africa

The African French speakers produced 274 tokens where a complementizer was possible with an absence rate of 20.8% (n=57). A summary of results by social factor is shown in Table 3.19. The country with the highest rate of absent *que* is Côte d'Ivoire followed by Burkina Faso and then the Central African Republic. Boutin (2007) found an absent *que* rate of 8.8% for Côte d'Ivoire, which is considerably lower than that of the present study. It should be noted that she considered all 14 speakers in the PFC from Côte d'Ivoire while the present study only considered those in selected age groups (n=7). Due to a smaller number of speakers in the subcopora of African speakers, there are only two age groups. The middle group has a higher absence rate than the younger group by 8.6%.

Table 3.19. Africa - Overall absence of *que* as percentages for social factors

Social Factor	% ∅	N=∅/Total
Region		
Burkina Faso	19.0	20/105
Central African Rep.	13.8	9/65
Côte d'Ivoire	26.9	28/104
Age Group		
younger	15.6	17/109
middle	24.2	40/165

Region and age, the two social factors considered for African speakers, are included in all three linear models shown in Section 3.2.4. All of the models indicate that the social factors are not statistically significant. In the absence of any significant social factors, I will look at complementizer rates for the two types of matrix verbal elements to determine social trends. While the matrix verbal element is not included as a fixed effect for any of the linear models, its variance as a random effect indicates that it accounts for some of the variation in the data, and that it is responsible for more of the variation than the speaker. Table 3.20 provides the absent *que* rates for each social factor.

Table 3.20 Africa - Absence of *que* as percentages for social factors according to type of verbal element

Non-Discourse Markers (NDM)			Potential Discourse Markers (PDM)		
Social Factor	% Ø	N=Ø/Total	Social Factor	% Ø	N=Ø/Total
Region			Region		
Burkina Faso	17.4	16/92	Burkina Faso	**30.8**	4/13
Central African Republic	7.0	4/57	Central African Republic	**62.5**	5/8
Côte d'Ivoire	24.7	20/81	Côte d'Ivoire	**34.8**	8/23
Age Group			Age Group		
younger	11.6	11/95	younger	**42.9**	6/14
middle	21.5	29/135	middle	**36.7**	11/30

A comparison of regions reveals that Côte d'Ivoire has the highest *que* absence rate for the NDM forms, whereas it has the second highest for the PDM forms. The Central African Republic has a rate difference of 55.5% based on the verbal elements chosen, indicating that this region varies the most by verb type. Burkina Faso has a similar but lower rate than Côte d'Ivoire for both verb types. Manessy (1992) notes that in Africa the French spoken in Côte d'Ivoire differs the most from Standard and Nonstandard European French. Francophones from other

African countries consider the French of Côte d'Ivoire to be very marked lexically and grammatically even to the point where it can impede mutual comprehension with speakers from other francophone countries (Manessy 1994). With regard to absent *que*, the full data set and the NDM forms show that of the three countries included in this study, Côte d'Ivoire differs the most from Standard European French since it has the highest *que* absence rate. However, this does not hold true for the PDM verb forms since the Central African Republic has a much higher rate. It is worth noting that there are extremely few tokens from the Central African Republic, and that Côte d'Ivoire has the second highest rate for this subset. It is possible that the Central African Republic shows such a difference in rates based on the verbal element due to a stigmatization for the NDMs, where one never encounters an absent *que* according to prescriptive grammar. On the other hand, when the PDM forms are used as discourse markers there is no complementizer. As described for European speakers, the frequency and dual function of the PDM forms could reduce stigmatization.

The second social factor is age with just two different groups, younger and middle. Looking at the absent *que* rates in Table 3.20, the NDM forms show the opposite trend from the PDM forms. It is the middle age group that has the higher rate for the NDM forms but the lower rate for the PDM forms. There is an approximate 10% difference between the rates of the two age groups for the NDM forms, and a slightly smaller gap between the age groups for the PDM forms. Additionally, this latter group aligns with the results of Martineau (1985) and Liang et al. (2021) who found that younger speakers had a higher absent *que* rate than the middle age group. An examination of representation by region and age group in Table 3.21 reveals that the middle age group is highly represented by speakers from Burkina Faso. This region has an absent *que*

rate between the other two regions regardless of verb type. The younger age group is evenly represented among the countries.

Table 3.21. Number of African Speakers by Region and Age

	Younger	Middle
Burkina Faso	3	9
Central African Rep.	3	3
Côte d'Ivoire	3	4

To summarize, the absent *que* rate by region varies the most for the Central African Republic. Côte d'Ivoire has a higher but similar rate to Burkina Faso for both verb types. Overall, Côte d'Ivoire displays the highest use of the non-standard variant. Speakers in the middle age group have a higher absent *que* rate for the NDM verbal elements, while they have a lower rate for the PDM forms.

3.3.2. Cross-regional Comparison and Discussion of Social Factors

This section compares informal spoken language in Europe, Quebec, and Africa to indicate in which ways absent *que* is similar cross-regionally and in which ways it diverges. I begin with a brief discussion of the overall absent *que* rates for the three regions followed by a comparison and discussion of each social factor.

3.3.2.1. Overall Absent *Que* Rates

Quebec is the region with the highest overall absent *que* rate at 30.8% followed by Europe at 25.3% and then Africa at 20.8%. Based on these rates, the absent complementizer is most

widespread in Quebec and the least widespread in Africa. Before analyzing the data, it was hypothesized that Quebec would have the highest rate based on its long-documented history of absent *que*. Following Quebec would be Africa due to its linguistic differences and its distance from the governing body of prescriptive French grammar, the Académie Française, headquartered in Paris. Europe was predicted to have the lowest absent *que* rate since it represents the birthplace of French and the seat of the Académie Française. As predicted, Quebec has the highest rate, although Europe has a higher rate than Africa.

Aside from a long history of absent *que* in Quebec dating back at least to the 1970s, there are two other reasons that support a high rate in this region. As mentioned in Section 1.5.2, *que* is the default relative pronoun in North American varieties of French (Roberge and Rosen 1999). Since it is no longer the antecedent to the noun phrase, it has undergone reanalysis as a complementizer. Due to its dual function and the fact that the relative pronoun function is primary, the complementizer has been weakened. As a result, absent *que* is more prevalent than it would have been otherwise. Example (32) from Section 1.5.2 is repeated below as Example (69) along with an additional example from Louisiana French. In the first, *que* has replaced the relative pronoun *laquelle*, and in the second, it has replaced *dont*.

(69) Louisiana French: la femme...**que** le canal appartient (Roberge & Rosen 1999:162)
'the woman that the canal belongs [to]'
Standard French: la femme à **laquelle** le canal appartient
'the woman to whom the canal belongs'

(70) Louisiana French: le tit enfant de cet homme-là qu'a brûlé...**que** sa maison a brûlé... (Roberge & Rosen 1999:162)

'the grandchild of that man...that burned...that his house burned'

Standard French: le petit enfant de cet homme...qui a brûlé...**dont** la maison a brûlé

'the grandchild of that man...who burned...whose house burned'

Section 1.6 provided a brief overview of schwa and *ne* deletion in French. Sankoff and Vincent (1981) found a *ne* deletion rate of 99.5% among Montreal speakers, higher than the rates in Pohl (1968) for France and Belgium (67.3%), in Ashby (1981) for Tours, France (63%), and in Coveney (1996) for Paris (81.2%). While not directly related to the complementizer, Côté and Milne (2009:slide 10) found that speakers from Quebec had a higher deletion rate of /ʁ/ and /l/ than speakers from Northern France. When looking at the pronunciation of consonant clusters by speakers from Quebec and Northern France, Green and Hintze (2021:132-134) found that speakers from Quebec use deletion to simplify consonant clusters more often than their French counterparts. These studies indicate that deletion is prevalent in Quebec and more so than in some francophone regions. It is thus not surprising that Quebec speakers use the absent complementizer, or "delete" *que*, more than in Europe and Africa.

It was mentioned in Section 1.6 that francophone Africa is mostly comprised of second language French speakers. While it has been shown that English influence does not affect presence or absence of *que*, the impact of other languages is unknown (Martineau 1985; Blondeau & Nagy 2008). In their study of complementizer use among Montreal Anglophones, Blondeau and Nagy (2008) found their greater contact with native French speakers significantly increased the chance of absent *que* when they spoke French. It should also be pointed out that in

Africa, the sociocultural context, status of French, and the context in which French is used can vary considerably by country (Manessy 1992). Factors that affect variability are location, French competency level of the speaker, and circumstances of French use (Manessy 1992). These linguistic and sociocultural differences most likely account for the lower absent *que* rate in Africa.

3.3.2.2. Comparison and Discussion of Social Factors

The social factors of region, education, and age were included in every linear model and none of them was shown to be statistically significant in any model. Therefore, we can conclude that these specific social factors do not condition the absence of *que*. While these social factors are not significant, cross-regional trends will now be explored to determine any similarities.

The absent *que* rates by region for Europe and Africa signal that there are regional differences, which can be seen in Table 3.22. The difference in rates between Switzerland and Southern France, the regions having the two lowest rates for Europe, is very similar to the difference between Burkina Faso and the Central African Republic, which hold the same positions for Africa. However, there is a somewhat more marked difference between Côte d'Ivoire and Burkina Faso, representing the highest and lowest rates, than there is between Northern France and Southern France. As pointed out by Manessy (1992), the French in Côte d'Ivoire is very marked lexically and grammatically and is noticeably different from that of other African Francophones. The fact that Northern France has a higher absent *que* rate than Southern France supports the idea that Southern speakers generally pronounce more sounds. Both Europe and Africa reveal different trends based on verb type. In Europe, it is Northern and Southern France that have the lowest rate for the NDM forms, and Northern France has the highest for the

PDM forms. The Central African Republic has the lowest rate for the NDM forms and the highest for the PDM forms. Regional differences in both Europe and Africa can clearly be seen, while they are slightly more pronounced in Africa. Additionally, the role of the matrix verb is highlighted by the differing trends according to which verb type is considered.

Table 3.22. Absence of *que* as percentages by region

Europe			Africa		
Region	% Ø	N=Ø/Total	Region	% Ø	N=Ø/Total
Northern France	26.7	116/435	Burkina Faso	19.0	20/105
Southern France	20.4	11/54	Central African Rep.	13.8	9/65
Switzerland	24.0	65/271	Côte d'Ivoire	26.9	28/104

The second social factor, education level, shows opposite patterns by region. The European data reveal that a university level education is associated with the highest absent *que* rate, whereas it is a primary level of education that holds this position in Quebec, as summarized in Table 3.23. A primary level of education has a higher absent *que* rate than secondary for both regions, although there is little difference between primary and secondary in Europe. Based on verb type, the rates for secondary and university vary the most in Europe. In Quebec, all three levels vary considerably by verb type. For this region, the rates for each education level are very similar to one another for the NDM forms but not for the PDM forms. For both primary- and university-educated speakers, Northern speakers largely affect the rates in Europe, and in Quebec, younger speakers affect the rates for university while older speakers affect the rates for primary.

Table 3.23. Absence of *que* as percentages by education level and region

Education Level	Europe		Quebec	
	% Ø	N=Ø/Total	% Ø	N=Ø/Total
primary	22.9	16/70	37.0	70/189
secondary	21.2	75/354	29.5	181/613
university	30.1	101/336	30.2	219/725

At first glance it appears that the three regions are dissimilar concerning the age patterns shown in Table 3.24. In Europe, the younger age group has the highest absent *que* rate while in Quebec it is the older group and the middle group in Africa. However, Quebec and Africa have the same pattern since there is no older group among African speakers. In both of these regions, the middle group has a higher rate than the younger group. With respect to the ordering of age groups for each region, it is the younger group that varies. Interestingly, the absent *que* rates for the younger group in Europe and in Quebec differ by only 1.5%. Each region indicates a difference in age group ordering based on verb type.

Table 3.24. Absence of *que* as percentages by age group and region

Age Group	Europe		Quebec		Africa	
	% Ø	N=Ø/Total	% Ø	N=Ø/Total	% Ø	N=Ø/Total
younger	29.5	106/359	28.0	172/614	15.6	17/109
middle	19.5	42/215	32.0	177/553	24.2	40/165
older	23.7	44/186	33.6	121/360	- -	- -

In summary, Quebec has the highest overall absent *que* rate most likely due to its longer history of the phenomenon and its relatively higher deletion rate with other linguistic variables. A comparison of absent *que* rates by social factor reveals that Europe and Africa show regional differences, but Africa indicates a more distinct difference for one region. Regarding education

level, Europe and Quebec display different patterns. For age, Europe stands alone, while Quebec and Africa pattern similarly.

3.4. Conclusion

The frequency of matrix verbal elements indicates that they are all highly frequent, and that the PDM forms are the most frequent. There is no obvious correlation between frequency and the absent *que* rates for PDM forms, although this does not mean definitively that frequency of the matrix verbal element has no effect on the realization of *que*. The generalized linear mixed-effects models for each region indicated that the social factors considered for the present study are not statistically significant. In the absence of any significant factors the absent *que* rates for each region and social factor were considered to determine trends. First, the overall absent *que* rates were discussed followed by the rates for the two types of matrix verbal elements. The first factor to be considered was region. It was determined that Northern France has a higher absent *que* rate than the other two European regions and that the greatest variation occurs in the absent *que* rates following potential discourse markers. In Africa, Côte d'Ivoire has the highest overall absent *que* rate, which is congruent with its status of having the variety of French that diverges most from Standard French. The variation in the African data comes from both NDMs and PDMs. The Central African Republic has the greatest variation based on verb type, possibly because the PDM forms are less stigmatized due to their frequency and dual function. Both Europe and Africa show regional differences with those in Africa being somewhat more pronounced.

In Europe, secondary and university levels of education have the highest rates for the PDMs but not the NDMs. This could be due to the frequency and dual function of the PDM

forms. The primary and university absent *que* rates are largely dependent on region since Northern French speakers are over-represented in both levels. Most of the variation for education levels in Quebec is due to the PDM forms. Also of note is that the primary-educated speakers are almost entirely comprised of those in the older age group. The university-educated speakers overwhelmingly belong to the younger and middle age groups. Europe and Quebec show opposite patterns with regard to education, which is likely due to a difference in social status of the non-standard variant.

The final social factor considered was age. The middle age group in Europe has the lowest overall absent *que* rate. This age group likely has a professional need to use the standard variant more often. Both types of matrix verbal element display variation for this region. Based on a comparison with Martineau (1985) and Liang et al. (2021), there is no evidence of age grading or generational change. These two studies and the present study show a monotonic distribution, but the one for the present study is in the opposite direction. The differences in study design and the participants may in part account for the age-related differences. In Quebec the older speakers show the most stability since they have the highest absent *que* rate regardless of verb type. The age variation seen in African speakers depends on the type of matrix verbal element. The middle age group is heavily represented by Burkina Faso, which may have an effect on rates. Quebec and Africa pattern the same both overall and by verb type.

It is worth bearing in mind that the GLMMs for all but one model in Europe and one model in Quebec included speaker as a random effect. In each of these models, the individual speaker accounted for some of the variation in the data. It accounted for the most in Europe followed by Quebec and then Africa. Individual speaker variation will be explored in Chapter 5.

For every social factor level, the *que* absence rate increased considerably for the PDM forms. The variation seen with the PDM forms may be due to the inclusion of some discourse markers. In some cases, however, syntactic analogy may contribute to the higher absent *que* rates for PDMs. In all three regions, the role of the matrix verb is highlighted by the variation based on verb type. Finally, the overall absent *que* rate for Quebec indicated that the rate has not changed since studies conducted in the 1970s and 1980s. The rate for the present study falls between the ranges of previous studies.

Chapter 4

Results of Linguistic Factors in Three Francophone Regions

4.1. Introduction

This chapter is devoted to the second part of the absent *que* analysis, which considers four linguistic factors. Section 4.2 discusses the results from the GLMMs presented in Section 3.2 as well as the overall absent *que* rates for each factor level. The results for Europe are presented first, followed by those for Quebec and then Africa. The final portion of the section consists of a cross-regional comparison and discussion of linguistic factors. Concluding remarks are presented in the final section.

4.2. Linguistic Factors

4.2.1. Results and Discussion of Linguistic Factors

4.2.1.1. Europe

A summary of the overall absent *que* rates for the linguistic factors is shown in Table 4.1. As presented in Section 1.5.1, the sonority hierarchy ranks sounds based on degree of constriction. The most sonorous group, vowels, are followed by liquids, then nasals, fricatives, and stops. This hierarchy, as opposed to the most common one as stated in Parker (2011:1161-1162), differentiates fricatives and stops rather than combining them into one group. This allows for a more finely tuned analysis. Based on this hierarchy, the least sonorous class, stops, would have the highest absent *que* rate. The [k] variant would be the least likely to be followed by a stop

since this environment would probably result in an undesirable consonant cluster. However, the [kə] variant should have more instances of present *que* than [k] when followed by a stop since there is a vowel immediately preceding the stop in [kə]. The most sonorous class, vowels, easily forms a syllable with [k]. As mentioned in Section 1.6, schwa deletion is obligatory when followed by a vowel at a word boundary (Green & Hintze 2021:125). Thus, vowels would have the lowest absent *que* rate. While the present study does not distinguish [kə] from [k], it is still expected that absent *que* is least likely when followed by a vowel. The analysis of the phonological context following *que* will provide an indication of the role phonological groups may play in the realization of *que*. However, in order to fully determine the phonological effect, it would be necessary to distinguish [kə] from [k].

The rates for the phonological context correspond to the sonority hierarchy with the exception of stops. The results of the present study for sonority groups vary little by group with only a 4% difference between the highest and lowest absent *que* rates. Turning now to syntactic context, we see that overwhelmingly the syntactic context after the complementizer is a pronoun. Not only are pronouns more frequent, but they also have the highest rate of absent *que* followed closely by lexical nouns. In both the matrix and embedded clauses, there is a higher absent *que* rate when there is no intervening material. Results for the verbal elements indicate three verbs with an absent *que* rate over 50%: *voir*, *vouloir dire*, and *savoir*. *C'est-à-dire* and *peut-être* also have a relatively high rate. Seven elements appear categorically with the present complementizer: *falloir, heureusement, paraître, pouvoir dire, promettre, se dire*, and *vouloir*. All of these except *falloir* have extremely few tokens. There are no tokens of *avoir raison*, *être d'avis*, *en cas*, or *jurer*.

Table 4.1. Europe – Overall absent *que* rates as percentages for linguistic factors

Factor	% ∅	N=∅/Total	Factor	% ∅	N=∅/Total
Phonological Context			Verbal Element		
fricative	27.0	80/296	***voir***	92.3	60/65
nasal	26.1	6/23	***vouloir dire***	76.9	20/26
liquid	25.0	11/44	***savoir***	52.1	37/71
vowel	24.2	74/306	*c'est-à-dire*	44.4	4/9
stop	23.1	21/91	*peut-être*	42.9	9/21
Syntactic Context			***dire***	28.4	23/81
pronoun	27.1	162/598	*sembler*	16.7	1/6
lexical noun	26.2	16/61	***penser***	14.9	15/101
other	13.9	14/101	***trouver***	10.4	5/48
Intervening Material Matrix Clause			*avoir l'impression*	8.3	10/12
no	26.3	182/693	***croire***	8.1	60/74
yes	14.9	10/67	*c'est vrai*	7.14	90/126
Intervening Material Embedded Clause			*être* + adj.	4.8	10/21
no	26.6	171/642	*c'est*	4.4	10/23
yes	17.8	21/118	*falloir*	0	0/46
			heureusement	0	0/1
			paraître	0	0/6
			pouvoir dire	0	0/9
			promettre	0	0/1
			se dire	0	0/5
			vouloir	0	0/8

Verbal elements having discourse marker forms are indicated in bold.

There were 154 exclusions for the European data with 116 of them having an absent *que*. The matrix verbal elements with the most exclusions were *savoir* (n=39), *dire* (n=28), and *voir* (n=21). Many instances were excluded because they could be discourse markers based on the criteria outlined in Chapter 2. There were also several that were excluded because the embedded clause did not contain the minimum required elements or because of ambiguous placement of absent *que*, as detailed in Section 2.3.

The phonological context following the complementizer was tested in several previous studies mentioned in Chapter 1. These studies stated that sibilants favor absent *que* the most and

that vowels favor it the least (Cedergren and Sankoff 1974; Sankoff 1980; Blondeau & Nagy 2008). To compare the present study with these previous studies, the sonority groups would need to be condensed into three. The results for these three are as follows: sibilants (27.0%), non-sibilants (24.1%), and vowels (24.2%). According to this grouping, sibilants have a higher absent *que* rate than the other groups, and non-sibilants and vowels have almost identical rates. Liang et al. (2021) found that the following phonological context was significant and that obstruents favored absent *que* the most followed by sonorants and then vowels. Similarly, Warren (1994) also found obstruents to favor absent *que* while sonorants and vowels had almost identical rates. Using these three sonority groups, the present study has the following results: obstruents (26.1%), sonorants (25.4%), and vowels (24.2%). Thus, the present study orders the groups the same as Liang et al. (2021) and similarly to Warren (1994), even though the rates are similar for the three sonority groups in the present study. Martineau (1985) used five sonority groups and found the following rates of absent *que*: sibilant fricatives (58%), non-sibilant fricatives (57%), stops (56%), liquids and nasals (17%), and vowels and pauses (6%). While she gave results for five sonority groups, she grouped the first three together and compared them to the last two groups. In this way, her groups are comparable to the ones used by Warren (1994) and Liang et al. (2021). Based on the sonority groupings used in these three studies, their results indicated that the phonological context following the complementizer follows the sonority hierarchy with the exception of Warren (1994). Using the obstruent/sonorant/vowel groupings, the present study also follows the sonority hierarchy.

Based on the overall absent *que* rates, stops do not follow the five-group sonority hierarchy, but the other sonority groups do. One reason that stops may not follow the hierarchy relative to fricatives is because the sonority of these two groups depends on the specific

phonemes considered. Voiced fricatives (/v/, /z/, /ʒ/) are more sonorous than voiced stops (/b/, /d/, /g/), which are more sonorous than voiceless fricatives (/f/, /s/, /ʃ/) (Parker 2011:1177). Voiceless stops (/p/, /t/, /k/) are the least sonorous of all. An analysis of the stops shows that voiceless stops (27.9% ∅) have a higher absent *que* rate than voiced stops (13.3% ∅), in agreement with the sonority hierarchy. Still, voiced stops have a rate considerably lower than the other phonological groups, which is not expected.

As was seen with the social variables, the absent *que* results usually vary considerably based on the type of matrix verbal element. In Table 4.2, the rates are shown for both types of verbs for the phonological and syntactic contexts as well as intervening material. Beginning with the phonological context following the complementizer, nasals show the greatest difference from the other sonority groups by verb type since they have the highest absent *que* rate for the NDM forms and the lowest for the PDM forms. Of the words beginning with a nasal, *moi* 'me' and the possessive adjective *mon/ma* 'my' are the only ones to appear with absent *que*. The former has absent *que* three out of five times, and the possessive adjective has it two out of five times. *Maintenant* 'now' also appears five times but always with the present complementizer. Other words, appearing two or fewer times, that never appear with the absent *que* are *nous* 'we', *maman* 'mom', New York, and *même* 'even'. All other sonority groups have a far higher rate for the PDM forms than for the NDM forms, except when the following sound is a nasal. Liquids are ordered the same relative to the other sonority groups regardless of the matrix verbal elements.

Table 4.2. Europe – *Que* absence as percentages for NDM and PDM forms for three linguistic factors

Non-Discourse Markers (NDM)			Potential Discourse Markers (PDM)		
Factor	% ∅	N=∅/Total	Factor	% ∅	N=∅/Total
Phonological Context			Phonological Context		
vowel	11.4	21/185	vowel	**43.8**	53/121
liquid	12.0	3/25	liquid	**42.1**	8/19
nasal	**28.6**	4/14	nasal	22.2	2/9
fricative	13.0	22/169	fricative	**45.7**	58/127
stop	6.5	3/46	stop	**40.0**	18/45
Syntactic Context			Syntactic Context		
pronoun	13.1	44/337	pronoun	**45.4**	118/260
lexical noun	15.4	6/39	lexical noun	**43.5**	10/23
other	4.8	3/63	other	**28.9**	11/38
Intervening Material Matrix Clause			Intervening Material Matrix Clause		
yes	12.8	6/47	yes	**20.0**	4/20
no	12.0	47/392	no	**44.9**	135/301
Intervening Material Embedded Clause			Intervening Material Embedded Clause		
yes	7.0	5/71	yes	**34.0**	16/47
no	13.0	48/368	no	**44.9**	123/274

Concerning statistical significance, Model 3E (PDM), the only GLMM to include linguistic factors for Europe, reveals that nasals differ significantly from fricatives and that they decrease the chance of absent *que*. This is reflected in the results by verb type in Table 4.2. These two sonority groups differ considerably for both verb types and especially for NDMs. Overall, the results indicate that the phonological context following the complementizer does not play a large role in conditioning *que* realization. This is evident by the similar overall absent *que* rates of each sonority group relative to others. The statistical significance of this factor is due to the difference between nasals and fricatives, which is evident for both types of matrix verbs. To further explore this factor, Chapter 5 will analyze individual speaker variation for the phonological context following the complementizer.

Regarding the syntactic context following the complementizer, Connors (1975) concluded that it was the primary conditioning factor followed by the phonological context. Martineau (1985) found the syntactic context to be important but not more so than the phonological context. Similarly, Liang et al. (2021) found the syntactic and phonological contexts to be independently significant. The syntactic context following the complementizer was only included in the GLMM for the PDM forms and is not statistically significant. However, in the same model, the phonological context is significant. Therefore, for the European speakers, the phonological context is more important than the syntactic one for the PDM forms.

The results of the present study align with the conclusion of Connors (1975) and Martineau (1985) that a pronoun after *que* favors the absent form when compared to all other word types. Thus, one is more likely to hear *c'est vrai Ø c'est un peu un choix qu'ils ont* 'it's true Ø it's kind of a choice they have' (EM26) where absent *que* is followed by the pronoun *ce* (*c'*). Connors (1975) found 44 out of the 56 pronouns following *que* to be forms of *ce* 'this/it' and *je* 'I'. She determined that the high proportion of tokens was due to the frequency of these pronouns rather than their phonological properties. For European speakers in the present study, there are 175 instances of *ce* (including *ça*), 113 of *il* 'he/it', 102 of *je*, 72 of *on* 'we/one', 53 of *ils* 'they', and 34 of *tu* 'you' out of 598 pronouns. Just as in Connors' (1975) study, *ce* is one of the two most common pronouns. However, *il* is slightly more common than *je*. It should be pointed out that both *ce* and *je* begin with a fricative, whereas the pronouns that Connors (1975) found to disfavor absent *que* began with a vowel. Thus, there may be a phonological component to what she found after all, especially since she did not use the modern statistical tools that are now available to test the phonological and syntactic contexts. Evidence that there is a phonological component to the syntactic context for the present study is found in the absent *que*

rates for pronouns according to sonority group. The rates listed in Table 4.3 indicate that pronouns beginning with a vowel have the lowest absent *que* rate compared to all other sonority groups, which represent an initial consonant.

Table 4.3. European absent *que* rates for pronouns according to sonority group

Sonority Group	% Ø	N=Ø/Total
vowel	25.5%	67/263
liquid	33.3%	1/3
nasal	37.5%	3/8
fricative	27.3%	77/282
stop	33.3%	14/42

The present study indicates that lexical nouns, while much less frequent than pronouns, have a very similar absent *que* rate to pronouns. The vast majority of lexical nouns consist of an article plus a noun such as *les années* '(the) years' or *son père* 'his/her father' and begin with a liquid. The remaining few are names or proper nouns, such as *grand-maman* 'grandma' or New York. The group of other types of words has a considerably lower absent *que* rate than pronouns and lexical nouns. This group consists primarily of prepositional phrases such as *pour le repas* 'for the meal' or adverbial phrases like *au depart* 'at first', and most begin with a stop or vowel. There are also a few instances of *quand* 'when' and *si* 'if'.

The syntactic context following the complementizer has a different pattern depending on the verbal elements chosen. Lexical nouns have the highest absent *que* rate for the NDM forms with a rate roughly 2% greater than the next group, pronouns. The PDM forms show the reverse with pronouns having the highest rate, which is roughly 2% higher than for lexical nouns. Other types of words have a much lower rate than the other two groups regardless of the verbal elements. Since pronouns refer to a noun that has already been identified, it is usually easy to

recover their meaning. Liang et al. (2021) found that pronouns that are easier to retrieve from memory (*je*, *tu*, impersonal *il*, *ce*, and *ça*) were found to favor absent *que* more than all other pronouns and noun phrases. If we apply this principle to pronouns as a whole, compared to lexical nouns and other phrases, then pronouns should appear with absent *que* more often. This is the case with the European data except for the NDM forms. As stated above, lexical nouns have the highest absent *que* rate for these forms. Two out of the six lexical nouns that appeared with absent *que* were used with the matrix verbal element *c'est-à-dire* 'that is to say/in other words', which is used to restate a previously expressed idea in somewhat different words. Thus, it would be easier to retrieve the subject appearing after this expression. It is therefore not surprising that both instances of *c'est-à-dire* that appeared with the NDM forms were used with absent *que*.

Linguistic material that intervenes between the matrix verb and the complementizer usually requires *que* to be present (Bolinger 1972:38). It is believed that this helps make the complementizer clause identifiable. Intervening material can be a complement or an insertion (Bolinger 1972:38). To test the idea asserted by Bolinger (1972) regarding English, Martineau (1985) considered intervening material in the matrix and embedded clauses and found that in the matrix clause intervening material disfavored absent *que*. Intervening material in the embedded clause was not statistically significant. She concluded that in a complex sentence, one that includes intervening material, the present complementizer allows the speaker to clearly mark the beginning of the complementizer clause. In the case of an intervening adverb, she illustrates that it can be impossible to determine which clause the adverb belongs to if *que* is absent, as shown in example (71) below (Martineau 1985:80). Intonation can help with understanding in the case of absent *que*, but this is not always marked in a transcription. Warren (1994) also considered intervening material in both the matrix and embedded clauses. She found that intervening

material in the embedded clause was slightly more significant than in the matrix clause. While she did not explore the difference between the two clauses, she concluded that the present complementizer helps the listener understand the message when there is intervening material.

(71) Je pense ∅? jamais ∅? je vais y aller. (Martineau 1985:81)

'I think ∅ I will never go there.'

'I never think ∅ I will go there.'

The absent *que* rates for intervening material in the present study indicate that overall, there is a higher absent *que* rate when there is no intervening material, regardless of the clause. Looking at this factor by matrix verb type, the NDM forms show that there is a 0.8% higher rate of absent *que* when there is intervening material in the matrix clause. However, the PDM forms have the same pattern as the overall results for the matrix clause. Thus, for the PDM forms, absent *que* would be less likely in an utterance such as *je crois pas* ***qu'****il y ait euh les sous-titres* 'I don't believe **that** there are uh subtitles' (EM42) since *pas* is considered intervening material.

Regarding the embedded clause, there is a higher rate of absent *que* when there is no intervening material regardless of matrix verb type. Therefore, absent *que* would be less likely in an utterance such as *et euh et c'est vrai* ***qu'****à aux Etats-Unis j'ai pas du tout vu ça quoi* 'and uh it's true **that** at in the United States I didn't see this at all' (EM29) since *à aux Etats-Unis* is intervening material. Intervening material was included as a factor in the GLMM for the PDM forms only. In this model it is not a statistically significant factor.

Based on the works of Bolinger (1972), Martineau (1985), and Warren (1994), we would expect there to be a lower absent *que* rate when there is intervening material, at least in the matrix clause. The only exception to this for either clause in the present study is the NDM forms. All of the examples of NDM forms with absent *que* when there is intervening material in the

matrix clause indicate that the intervening material is just one word rather than a phrase. For example, (72) below indicates that the intervening material is the word *après*. This indicates that the length of the intervening material likely impacts the complementizer variant.

(72) Parce que quand même on dira <u>après</u> ∅ c'est vache (EM65)

'because just the same we will say <u>afterward</u> ∅ it's harsh'

Comparing the rates of the seven verbs that have NDM and PDM forms in Table 4.4, we see that all seven of them have a higher absent *que* rate for the forms that can be discourse markers (PDMs). *Voir* has a rate of 69.2% for NDM forms and 98.1% for PDM forms. The corresponding rates for the other verbs are as follows: *vouloir dire* 0%/90.9%, *savoir* 25.0%/74.4%, *dire* 16.7%/61.9%, *penser* 0%/15.6%, *trouver* 0%/15.2%, and *croire* 0%/10.3%. *Vouloir dire* shows the most striking difference based on which forms are chosen followed by *savoir*. These results are most likely due to these forms favoring absent *que* because of their dual function and possibly due to the inclusion of some examples that are in fact used as discourse markers. The former is an idea echoed by Martineau (1985:65) when she says that the ability of *savoir* and *penser* to be used as discourse markers probably reinforced their favorableness to the absent complementizer.

Table 4.4. Europe – Absence of *que* as percentages for NDM and PDM forms for verbal element

Non-Discourse Markers			Potential Discourse Markers		
Factor	% ∅	N=∅/Total	Factor	% ∅	N=∅/Total
Verbal Element			Verbal Element		
croire	0	0/16	*croire*	**10.3**	6/58
dire	16.7	10/60	*dire*	**61.9**	13/21
penser	0	0/5	*penser*	**15.6**	15/96
savoir	25.0	8/32	*savoir*	**74.4**	29/39
trouver	0	0/15	*trouver*	**15.2**	5/33
voir	69.2	9/13	*voir*	**98.1**	51/52
vouloir dire	0	0/4	*vouloir dire*	**90.9**	20/22

avoir l'impression	8.3	1/12
c'est	4.4	1/23
c'est-à-dire	44.4	4/9
c'est vrai	7.1	9/126
être + adj.	4.8	1/21
falloir	0	0/46
heureusement	0	0/1
paraître	0	0/6
peut-être	42.9	9/21
pouvoir dire	0	0/9
promettre	0	0/1
se dire	0	0/5
sembler	16.7	1/6
vouloir	0	0/8

Wiesmath (2002) found that certain lexical items in the matrix clause favor absent *que* in Acadian French, and Martineau (1985) concluded that the matrix verb strongly influences the variable *que* rate in Quebec and Ontario French. Blondeau and Nagy (2008) found the lexical identity of the matrix verb to be the most significant factor among Anglophone French speakers in Montreal. Liang et al. (2021) tested the Uniform Information Density (UID) hypothesis developed by Levy and Jaeger (2007) and found that speakers are more likely to use absent *que* with a verb that is more likely to appear with a complementizer clause. Model 1E shows that judgment verbs differ significantly from epistemic verbs, and Model 3E indicates that *savoir*, *voir, dire*, and *vouloir dire* differ significantly from *croire*. Based on the estimates of the coefficients in Model 1E, judgment verbs decrease the chance of absent *que* compared to epistemic verbs. Since judgment verbs decrease the chance of the absent form, they usually appear with *que* as in *je pense **que** ça s'est fait comme ça* 'I think **that** it was done like this' (EF53). As in previous studies, the matrix verb is a statistically significant factor. However, we cannot rank its importance since the same factors are not included in every linear model for the European data.

Comparing the PDM forms to one another, it is clear that *croire*, *trouver*, and *penser* have relatively low absent *que* rates compared to *voir*, *vouloir dire*, *savoir*, and *dire*. In Model 3E, the difference in the last four verbs compared to the reference level *croire* is statistically significant. They increase the chance of absent *que* compared to *croire*. *Penser* has the most tokens but one of the lower rates. Its rate is higher than that of *croire*, which has a very similar meaning. Martineau (1985) proposed that *croire* favors formal speech, while *penser* favors informal speech. She suggested that this makes *croire* less favorable to absent *que* since it is associated with formal speech. There are roughly one-third more tokens of *penser* than *croire* and the rate 5.3% higher. This does show a difference between the two verbs but not a very sizeable one. Since Martineau (1985) observed the language of Canadian speakers from Ontario and Quebec, it may be that *croire* is used more frequently by Francophone from Europe than from Canada. Section 4.2.1.2 will provide results to allow a comparison of these two regions.

Looking at the specific verbs shown to favor absent *que* in previous studies, I focus first on *croire*, *dire*, *penser*, and *sembler*, which Martineau (1985) and Blondeau and Nagy (2008) found to favor absent *que*. *Savoir* and *trouver* are added to this list since Martineau (1993) found them to also favor the absent variant. Referring to the full data set in Table 4.2, *savoir* has a rate over 50% and *dire* has a rate above 25%, while the others are between 8% and 17%. The analyses presented in this chapter do not allow us to say with certainty that all these verbs favor absent *que*. However, we can see that they all allow absent *que* in European French and that *savoir* and *dire* occur with absent *que* more often than the others. Also, compared to *croire*, the coefficient estimates for the PDM forms show that the verbs *savoir* and *dire* increase the chance of absent *que* to a statistically significant level.

In her analysis of two corpora of Parisian French in the 1980s, Andersen (1993) found that *c'est* + adjective had an absent *que* rate of 35% and that *voir*, *dire*, *savoir*, *trouver*, *se souvenir*, and *se rappeler* had a rate of 30%. *Être* + adjective, which includes *c'est* + adjective, has a rate of 4.8% in the present study while the four other verbs included have the following rates: *voir* 92.3%, *dire* 28.4%, *savoir* 52.1%, and *trouver* 10.4%. Thus, the present study reveals a wide range of rates for these verbs that Andersen (1993) found to favor absent *que*, and at a much higher average rate (45.8%). The coefficient estimates for the PDM forms show that *voir*, in addition to *savoir* and *dire,* increases the chance of absent *que* compared to *croire*. In the same study, Andersen (1993) did not find any examples of absent *que* with *croire*, *penser*, *falloir*, or *vouloir*. The European data for the present study confirm this finding for the latter two verbs but not the first two. While *penser* and *croire* do not have absent *que* rates above 15%, they both appear with the absent complementizer.

To summarize the linguistic results, the phonological context following the complementizer is significant for the PDM forms indicating that nasals reduce the chance of absent *que* compared to fricatives. None of the data sets follows the five-group sonority hierarchy, but the overall data follow the obstruent/sonorant/vowel hierarchy. Nasals show the greatest difference according to matrix verb type. Regarding the syntactic context following the complementizer, pronouns and lexical nouns have very similar absent *que* rates, although lexical nouns are much less frequent in the corpus. The absent *que* rates for following pronouns and lexical nouns are much higher than those for other elements. The matrix verb type impacts whether intervening material in the matrix clause results in a higher absent *que* rate. However, the presence of intervening material in the embedded clause is always associated with a lower rate. The absent *que* rate is always higher for the PDM forms than the NDM forms of the same

verbs. When considering all the data, judgment verbs differ significantly from epistemic verbs. For the PDM forms, *dire*, *savoir*, *vouloir dire*, and *voir* differ significantly from *croire*.

4.2.1.2. Quebec

The overall absent *que* rates for each linguistic factor for Quebec speakers are shown in Table 4.5. Results for the phonological context following the complementizer do not adhere to the sonority hierarchy described in Section 1.5.1. Most of the words following *que* are pronouns. Their absent *que* rate is slightly higher than that of lexical nouns. Words classified as "other" have a considerably higher rate than the other two groups. The absent complementizer is used much more often when there is no intervening material in the matrix clause, but this does not hold true for the embedded clause. The verbal element *avoir raison* is categorically used with absent *que*, while *avoir l'impression*, *c'est*, *pouvoir dire*, and *promettre* are categorically used with *que* present. All of these have very few tokens except for *c'est*. Aside from the categorical elements, *c'est-à-dire*, *jurer*, *savoir*, and *vouloir dire* have an absent *que* rate greater than 50%. There are no instances of *être d'avis*, *en cas*, or *heureusement* in the Quebec subcorpora.

Table 4.5. Quebec – Overall absence of *que* as percentages for linguistic factors

Factor	% ∅	N=∅/Total	Factor	% ∅	N=∅/Total
Phonological Context			Verbal Element		
fricative	36.9	232/629	*avoir raison*	100	2/2
nasal	34.6	9/26	*c'est-à-dire*	83.3	5/6
stop	29.8	72/242	*jurer*	75.0	3/4
vowel	25.0	137/549	***savoir***	74.8	205/274
liquid	24.7	20/81	***vouloir dire***	64.2	43/67
Syntactic Context			*c'est vrai*	48.5	32/66
other	40.7	37/91	*sembler*	43.2	19/44
pronoun	30.2	403/1334	*peut-être*	30.4	17/56
lexical noun	29.4	30/102	***penser***	25.4	59/232
Intervening Material Matrix Clause			***voir***	24.2	8/33

no	31.9	440/1380	*se dire*	14.3	1/7
yes	20.4	30/147	*être* + adj.	14.1	9/64
Intervening Material Embedded Clause			***trouver***	13.2	9/68
yes	41.2	47/114	***dire***	11.7	19/162
no	29.9	423/1413	*falloir*	11.2	36/321
			croire	9.1	1/11
			paraître	5.9	1/17
			vouloir	1.8	1/56
			avoir l'impression	0	0/5
			c'est	0	0/25
			pouvoir dire	0	0/6
			promettre	0	0/1

Verbal elements having discourse marker forms are indicated in bold.

The Quebec data set had 440 exclusions, and 304 of these had absent *que*. Of the matrix verbal elements, *savoir* had the most exclusions (n=165) followed by *dire* (n=108). The majority of the exclusions were due to the embedded clause not having the minimum required components. Roughly 100 tokens were excluded because the grammatical structure of the utterance was unclear. The other exclusions were omitted primarily because they met the criteria for a discourse marker or because of ambiguous placement of absent *que*.

Based on the sonority groupings used by Cedergren and Sankoff (1974), Sankoff (1980), and Blondeau and Nagy (2008), the present study reveals that sibilants have the highest absent *que* rate at 36.9%, followed by non-sibilants with a rate of 28.9%, and vowels at 25.0%. Thus, the same order as in previous studies is found. The sonority groupings used by Martineau (1985), Warren (1994), and Liang et al. (2021) yield the following results for the present study: obstruents (34.9%), sonorants (27.1%), and vowels (25.0%). Just as for the European data, the Quebec speakers follow the same order. It would therefore be more likely to have absent *que* when it is followed by a fricative as in *je pense ∅ c'est universel aussi* 'I think ∅ it's universal

too' (QM24b) than when it is followed by a vowel as in *moi je pensais **qu'**elle s'appelait Tina* 'me, I thought **that** her name was Tina' (QM24b).

Absent *que* rates for the phonological and syntactic contexts as well as intervening material according to verb type are provided in Table 4.6. Based on the phonological context there are some similarities between the two verb types. Nasals have the highest absent *que* rate regardless of verb type, and fricatives and vowels have the two lowest rates. Of the words beginning with a nasal, there is an absent complementizer before *moi* 'me' five out of eleven times, *nous-autres* 'we' two out of four times, *mon/ma* 'my' one out of five times, and the name *Nate* the one time it is said. The words *maman* 'mom', *Madame* 'Mrs.', and the possessive adjective *nos* 'our' appear two or fewer times and never with absent *que*. All sonority groups have considerably higher rates for the PDM forms compared to the NDMs. While both verb types exhibit variation according to sonority group, there is more variation among the PDM forms.

Table 4.6. Quebec – Absence of *que* as percentages for NDM and PDM forms for three linguistic factors

Non-Discourse Markers			Potential Discourse Markers		
Factor	% ∅	N=∅/Total	Factor	% ∅	N=∅/Total
Phonological Context			Phonological Context		
vowel	13.1	50/382	vowel	**60.0**	99/165
liquid	17.9	10/56	liquid	**68.0**	17/25
nasal	23.5	4/17	nasal	**77.8**	7/9
fricative	16.9	65/384	fricative	**60.8**	146/240
stop	19.0	35/184	stop	**66.1**	37/56
Syntactic Context			Syntactic Context		
pronoun	15.8	144/909	pronoun	**60.8**	253/416
lexical noun	14.9	10/67	lexical noun	**65.7**	23/35
other	21.3	10/47	other	**68.2**	30/44
Intervening Material Matrix Clause			Intervening Material Matrix Clause		
yes	17.9	20/112	yes	**30.3**	10/33
no	15.8	144/911	no	**64.1**	296/462

Intervening Material Embedded Clause			Intervening Material Embedded Clause		
yes	22.4	13/58	yes	**67.9**	38/56
no	15.6	151/965	no	**61.0**	268/439

The linear model for the full data set indicates that vowels, liquids, and nasals differ significantly from fricatives and that all three decrease the chance of absent *que*. Vowels are the most statistically significant with a p-value less than 0.001. When considering the PDM forms only, vowels and stops are significantly different from fricatives, and both reduce the chance of absent *que*. Here again, vowels are the most statistically significant with a p-value less than 0.001. Since vowels are at the opposite end of the sonority hierarchy from fricatives, it is expected that they would differ from each other. However, it is not anticipated that stops would reduce the chance of absent *que* compared to fricatives. This could be due in part to the specific stops and fricatives involved (Parker 2011:1177). However, according to sonority, stops and fricatives pattern more similarly to each other than to other phonological groups. It should be noted that the linear model for the PDMs has matrix verb as a random effect, while the full data set does not. Rather, the full data set includes matrix verb semantic group as a fixed effect.

Concerning the importance of the phonological versus the syntactic context, the syntactic context is not statistically significant in any of the linear models, but the phonological context is significant for the full data set and the PDM forms. Thus, the present study supports the conclusion of Martineau (1985) and Liang et al. (2021) that the syntactic context is not more important than the phonological one.

The overall absent *que* rates for the syntactic context reveal that words other than lexical nouns and pronouns have the highest absent *que* rate. An example is *je veux dire Ø <u>si</u> c'est juste des femmes...* 'I mean Ø <u>if</u> it's just women...' (QM25a) where *si* 'if' belongs to the other words

category. Upon closer inspection, two-thirds of the occurrences of absent *que* with other words are for the verb *savoir*. The other verb with the most occurrences of absent *que* is *vouloir dire*, which has an absent *que* seven out of nine times. Both *savoir* and *vouloir dire* can be discourse markers. In Table 4.5 we see that *savoir* and *vouloir dire* are ranked fourth and fifth for the highest overall absent *que* rate. The three verbs with higher rates have six or fewer tokens. *Savoir* and *vouloir dire* have the highest rates of any PDM verbs. Therefore, the high absent *que* rate for other words is due to their high occurrence with *savoir* and *vouloir dire* and the favorableness of these two verbs to the absent complementizer.

Most words in the other category are prepositional or adverbial phrases such as *à un moment donné* 'at some point' or *dans le fond* 'basically'. Other words also have the highest absent *que* rate regardless of matrix verb type. The phonological context for other words reveals that stops are the most frequent and have the highest absent *que* rate followed by vowels and then fricatives. This does not provide evidence of there being a phonological component to the syntactic context.

For both the full data set and NDM forms, pronouns and lexical nouns have very similar rates with pronouns having a rate roughly 1% greater than lexical nouns. The PDM forms have a higher rate for lexical nouns than pronouns by 5%. The vast majority of the tokens with absent *que* for these forms are used with the verbs *savoir* and *vouloir dire*. Regarding frequency, pronouns are by far the most frequent for each data set. The most frequently occurring pronouns are *ce* (n=391), *il* (n=238), *je* (n=196), *tu* (n=188), *ils* (n=115), *elle* (n=81), and *on* (n=81) out of a total of 1333 tokens. Just as Connors (1975) found, forms of *ce* are one of the two most common types of pronouns after the complementizer. However, *il* is more common than *je* in the

present study. The lexical nouns primarily have the form of a definite article plus a noun such as *la ville* 'the city'.

The conclusion that pronouns favor absent *que* (Connors 1975; Martineau 1985; Liang et al. 2021) is not supported by the data for Quebec in the present study. This is largely due to the high occurrence of other words and lexical nouns with the verbs *savoir* and *vouloir dire*. These two verbs are relatively frequent and have very high absent *que* rates.

The overall absent *que* rates for intervening material indicate that there is a higher rate of absent *que* when there is no intervening material in the matrix clause. Based on matrix verb type, there is a higher rate when there is intervening material with NDM forms than when there isn't, but the opposite is true for the PDM forms. Thus, one is more likely to hear *je pense Ø il y a quand même du bon là* 'I think Ø there is still some good there' (QM25a) with a PDM and absent *que* without intervening material in the matrix clause. Looking at the specific examples of absent *que* with NDM forms when there is intervening material, all but three consist of just one word that intervenes. Two of the three exceptions were by the same speaker. The one intervening word is often the negative marker *pas* (n=11/20). Five out of ten of the verbal elements used with intervening material have low overall absent *que* rates. The five that do not are *c'est vrai*, *peut-être*, *penser*, *savoir*, and *voir*. For *c'est vrai* and *peut-être*, there was just one example of each. For *penser*, *savoir*, and *voir*, all three have discourse marker forms, and it is these forms that account for the high absent *que* rates. The NDM forms of these verbs have low rates. It appears that for the NDM forms, the verbs with low absent *que* rates allow the absent variant when intervening material in the matrix clause is limited to one word. Intervening material was included in all three GLMMs for Quebec speakers, and it is significant for the matrix clause for the PDMs. For these forms, the absent *que* rate is doubled when there is no intervening material.

Turning to the embedded clause, overall, there is a considerably higher absent *que* rate when there is intervening material. When the data are divided by matrix verb type, both types of verbs show this same pattern. It is more likely for the absent variant to occur when there is intervening material in the embedded clause as in *je pense Ø un mois avant les Olympiques j'ai dit…* 'I think Ø a month before the Olympics I said…' (QM48). An inspection of the examples revealed that many examples of intervening material consist of more than one word. Thus, the idea that the length of the intervening material affects the complementizer variant does not extend to the embedded clause. The linear models for Quebec speakers indicate that intervening material in the embedded clause is not a significant factor. This disagrees with what Warren (1994) found. However, the significance of intervening material in the matrix clause for the PDM forms agrees with the findings of Martineau (1985) and Warren (1994).

Turning to the matrix verb, those that have both PDM and NDM forms indicate a higher absent *que* rate for the PDM forms except for *trouver* and *croire*. The largest difference in rates by verb type can be seen with *vouloir dire* and *savoir*. For the NDM forms, *vouloir dire* has a rate of 4.0% but a rate of 97.6% for the PDM forms. Similarly, *savoir* has a rate of 9.8% for the first group but 91.6% for the second. On the other hand, *trouver* has a higher rate for the NDM forms (27.6%) compared to PDMs (15.4%). *Croire* has a rate of 25.0% for the first group and 0% for the second. In general, the verbs that have NDM and PDM forms may be more favorable to absent *que* because of their dual function of matrix verb or discourse marker.

Table 4.7. Quebec – Absence of *que* as percentages for NDM and PDM forms for verbal element

Non-Discourse Markers			Potential Discourse Markers		
Factor	% Ø	N=Ø/Total	Factor	% Ø	N=Ø/Total
Verbal Element			Verbal Element		
croire	**25.0**	1/4	*croire*	0	0/7
dire	24.1	34/141	*dire*	**28.6**	6/21
penser	12.5	9/72	*penser*	**32.5**	52/160
savoir	9.8	5/51	*savoir*	**91.6**	196/214
trouver	**27.6**	8/29	*trouver*	15.4	6/39
voir	19.0	4/21	*voir*	**41.7**	5/12
vouloir dire	4.0	1/25	*vouloir dire*	**97.6**	41/42
avoir l'impression	60.0	3/5			
c'est	24.0	6/25			
c'est-à-dire	16.7	1/6			
c'est vrai	10.6	7/66			
être + adj.	12.5	8/64			
falloir	16.2	52/321			
paraître	0	0/17			
peut-être	19.6	11/56			
pouvoir dire	0	0/6			
promettre	0	0/1			
se dire	14.3	1/7			
sembler	11.4	5/44			
vouloir	12.5	7/56			

The linear model of the full data set indicates that the matrix verb semantic groups are statistically significant. Desired action, judgment, and speaking/restatement verbs all differ considerably from epistemic verbs. Adverbs are also significantly different, and all four of these semantic groups decrease the chance of absent *que* compared to epistemic verbs. Therefore, one would be more likely to hear *il <u>veut</u> pas **que** la maison soit collée sur la sienne* 'he doesn't <u>want</u> the house to be right next to his [lit. 'stuck on his'] (QM27a) where the present *que* is used with a desired action verb such as *vouloir*. While it was not possible to test the matrix verb as a fixed effect for either of the data subsets, it was a random effect, and its variance indicated that it accounts for some of the variation in the data, particularly for the PDM forms.

Looking only at PDM forms, *vouloir dire* and *savoir* have absent *que* rates much higher than those of the other five verbs. *Savoir* and *penser* are by far the most frequent, and their absent *que* rates are ranked second and fourth, respectively. Martineau's (1985) finding that *croire* favors formal speech and disfavors absent *que* compared to *penser* is supported by the results of the present study for the PDM forms and the full data set. *Penser* is very frequent (n=160) and has an absent *que* rate of 32.5% for the PDMs, while *croire* is very infrequent (n=7) and has a 0% rate for the PDM forms. In comparing the NDM forms of these two verbs, *croire* has a higher absent *que* rate (25.0%) than *penser* (12.5%) but *croire* is still much less frequent (n=4) than *penser* (n=72). Overall, *croire* has a lower absent *que* rate and is much less frequent than *penser* in the informal speech of the Quebec corpus.

As mentioned in Section 4.2.1.1, Martineau (1993) and Blondeau and Nagy (2008) found *croire*, *dire*, *penser*, and *sembler* to favor the absent complementizer along with *savoir* and *trouver* (Martineau 1993). The overall absent *que* rates for the present study show that *savoir* has a rate over 50% and that *penser* and *sembler* have a rate greater than 25%. The other three verbs have a rate between 9.0% and 13.5%. Thus, we can see that all these verbs allow absent *que* and that it occurs more often after *savoir*.

In her Parisian French study, Andersen (1993) found *c'est* + adjective to have an absent *que* rate of 35% and the verbs *voir*, *dire*, *savoir*, *trouver*, *se souvenir*, and *se rappeler* to have a combined rate of 30%. The results from Quebec for the present study reveal a rate of 14.1% for *être* + adjective, 24.2% for *voir*, 11.7% for *dire*, 74.8% for *savoir*, and 13.2% for *trouver*. These results indicate a very wide range of rates with a combined rate (31.0%) that is comparable to what Andersen (1993) found. In the same study she found no examples of absent *que* with

croire, *penser*, *falloir*, and *vouloir*. In the present study all these verbs are used with absent *que*, although all except *penser* have a rate under 15%.

In summary, the phonological context following the complementizer is significant for the full data set and the PDM forms. For the full data set, vowels, liquids, and nasals significantly reduce the chance of absent *que* compared to fricatives, while vowels and stops do so for the PDM forms. None of the data sets follows the five-group sonority hierarchy. However, overall, Quebec speakers follow both of the three-group sonority hierarchies used in previous studies. Turning to the syntactic context, words other than pronouns and lexical nouns have a higher rate of absent *que*, which is due to their high occurrence with *savoir* and *vouloir dire*. It is clear from the statistical analyses that the phonological context is more important than the syntactic context. Intervening material in the matrix clause is significant for the PDM forms where its absence increases the chance of absent *que*. The opposite trend is found with the NDM forms. For the embedded clause, both verb types have a higher absent *que* rate when there is intervening material. Finally, with regard to the matrix verb, the absent *que* rate is higher for most of the PDM forms compared to the NDM forms of the same verbs. For the full data set, desired action, judgment, and speaking/restatement verbs as well as adverbs significantly decrease the chance of absent *que* compared to epistemic verbs.

4.2.1.3. Africa

The absent *que* rates for the linguistic factors shown in Table 4.8 indicate that the phonological context follows the sonority hierarchy with stops having the highest rate of absent *que* followed by fricatives, nasals, liquids, and then vowels. Lexical nouns and pronouns following *que* have almost identical rates, although the rate is slightly higher for lexical nouns. The absent

complementizer is used more often when there is no intervening material in the matrix clause. For the embedded clause, the rates are almost identical with presence of intervening material being associated with a slightly higher rate than absence of intervening material. The one token of the verbal element *heureusement* is used with absent *que*. On the other hand, *avoir l'impression*, *falloir*, and *vouloir* are categorically used with the present complementizer. *C'est-à-dire*, *savoir*, and *peut-être* all have an absent *que* rate in the 41%-44% range. There are no instances of *avoir raison*, *être* + adjectif, *être d'avis*, *en cas*, *jurer*, *promettre*, or *sembler* in the African data.

Table 4.8. Africa - Overall absence of *que* as percentages for linguistic factors

Factor	% ∅	N=∅/Total	Factor	% ∅	N=∅/Total
Phonological Context			Verbal Element		
stop	29.4	10/34	*heureusement*	100	1/1
fricative	28.1	25/89	*c'est-à-dire*	43.8	7/16
nasal	22.2	2/9	***savoir***	42.9	12/28
liquid	14.3	6/42	*peut-être*	41.2	7/17
vowel	14.0	14/100	***croire***	37.5	6/16
Syntactic Context			***voir***	35.0	7/20
lexical noun	21.7	10/46	*c'est vrai*	33.3	1/3
pronoun	21.5	37/172	***trouver***	25.0	1/4
other	17.9	10/56	*pouvoir dire*	20.0	2/10
Intervening Material Matrix Clause			***dire***	18.6	8/43
no	23.0	56/244	*se dire*	14.3	1/7
yes	3.3	1/30	***vouloir dire***	14.3	1/7
Intervening Material Embedded Clause			*c'est*	7.7	1/13
yes	20.9	14/67	*paraître*	7.1	1/14
no	20.8	43/207	***penser***	5.3	6/119
			avoir l'impression	0	0/2
			falloir	0	0/42
			vouloir	0	0/12

Verbal elements having discourse marker forms are indicated in bold.

There were 51 exclusions for the African data with 30 of them having absent *que*. Roughly 35% of the exclusions were for the matrix verb *dire* (n=18). For the other verbal elements, none of them had more than six tokens eliminated. Most tokens were omitted because the embedded clause did not contain the minimum required elements. Other reasons for exclusion were ambiguous placement of the absent complementizer, meeting the criteria for a discourse marker, having an unclear grammatical structure, and one case of a repetition.

The overall absent *que* rates for the present study indicate that the five-group sonority hierarchy as well as the two three-group hierarchies used in previous studies all indicate that the most sonorous groups have the lowest absent *que* rate and the least sonorous groups have the highest. If using the sibilant distinction, the results are as follows: sibilants (28.1%), non-sibilants (21.2%), and vowels (14.0%). Regarding the sonorant versus obstruent hierarchy the results are obstruents (28.5%), sonorants (15.7%), and vowels (14.0%).

The rates for the two types of matrix verbal elements are shown in Table 4.9. The NDM forms indicate that the sonority hierarchy is followed with the exception of liquids. Liquids would be expected to have a lower absent *que* rate than vowels. For the PDM forms, stops, liquids, and nasals have identical rates, and fricatives follow closely with a rate 2.6% lower. It should be noted that the number of tokens for the first three phonological groups is very low. Vowels have the lowest rate, and there is a larger gap between vowels and all other phonological groups than for the NDM forms. Both verb types account for variation in *que* realization, but there is more variation with the PDM forms.

Table 4.9 Africa – Absence of *que* as percentages for NDM and PDM forms for three linguistic factors

Non-Discourse Markers			Potential Discourse Markers		
Factor	% ∅	N=∅/Total	Factor	% ∅	N=∅/Total
Phonological Context			Phonological Context		
vowel	12.9	11/85	vowel	**20.0**	3/15
liquid	10.5	4/38	liquid	**50.0**	2/4
nasal	14.3	1/7	nasal	**50.0**	1/2
fricative	22.9	16/70	fricative	**47.4**	9/19
stop	26.7	8/30	stop	**50.0**	2/4
Syntactic Context			Syntactic Context		
pronoun	17.3	24/139	pronoun	**39.4**	13/33
lexical noun	15.4	6/39	lexical noun	**57.1**	4/7
other	**19.2**	10/52	other	0	0/4
Intervening Material Matrix Clause			Intervening Material Matrix Clause		
yes	**3.6**	1/28	yes	0	0/2
no	19.3	39/202	no	**40.5**	17/42
Intervening Material Embedded Clause			Intervening Material Embedded Clause		
yes	**21.3**	13/61	yes	16.7	1/6
no	16.0	27/169	no	**42.1**	16/38

Both Model 1A for the full data set and Model 2A for the NDM forms, the only models to include linguistic factors, indicate that liquids differ significantly from fricatives, and that they decrease the chance of absent *que*. Therefore, one is less likely to use absent *que* in an utterance such as *il faudrait **que** la paix revienne d'abord* 'first peace would have to return' (AF28b) where *que* is followed by a liquid than in *on peut dire ∅ c'est moi-même qui faisait tout le travail de la maison* 'one can say ∅ it's me who did all the housework' (AF48c) where absent *que* is followed by a fricative. Overall, the sonority hierarchy is followed by African speakers. There are some differences based on verb type, but it is likely that the small sample size for the PDM forms skews the results.

The syntactic context is a fixed effect for the GLMM for the full data set and the NDM forms. The group of other words is significantly different from pronouns for the full data set, and

the positive coefficient reveals that other words reduce the chance of absent *que*. Thus, it is more likely for a speaker to say *tu as vu **que** à l'école aussi c'est la même chose* 'you saw **that** at school it's also the same thing' (AF29) than the same utterance with absent *que*. Model 1A (full data set) indicates that both the syntactic and phonological contexts are significant and that the p-value for the phonological context is slightly lower than for the syntactic context. Connors (1975), Martineau (1985), and Liang et al. (2021) found both contexts to be important conditioning factors, although Connors believed the syntactic context to be more important. In the case of the full data set for the present study the phonological context is more important than the syntactic one. Model 2A (NDM forms) includes the same fixed and random effects as Model 1A. However, the syntactic context is not significant in Model 2A. Thus, the significance of the syntactic context lies with the PDM forms.

If one looks at pronouns versus all other words, as Connors (1975) and Martineau (1985) did, pronouns have a higher rate for the NDM forms but not for the PDM forms. The full data set shows that pronouns have a higher rate than other words, which is what Connors (1975) and Martineau (1985) found. Most of the pronouns following the complementizer were forms of *ce* and *je* for Connors' (1975) study. For African speakers, there are 51 instances of *ce*, 41 of *il*, 21 of *je*, 21 of *on*, and 13 for *ils* out of 172 tokens. Once again, *ce* is one of the two most common pronouns as in Connors' (1975) study. However, *il* is more common than *je*. Most pronouns begin with a vowel, which has the lowest overall absent *que* rate. Fricatives have just two fewer tokens and the highest overall absent *que* rate. The lexical nouns mostly take the form of a definite article plus a noun such as *les gens* "(the) people". The vast majority of the lexical nouns begin with a liquid. Overall liquids have one of the lowest absent *que* rates. Thus, it does not appear that there is a phonological component to the ranking of these two syntactic contexts. As

with the other regions, the other group consists primarily of prepositional phrases with adverbial phrases being the next largest group.

Comparing the absent *que* rates for the syntactic context following the complementizer by verb type, the NDM forms reveal that other words have the highest rate followed by pronouns and then lexical nouns. There is roughly a 2% difference between one group and the next. The PDM forms show a very different pattern with lexical nouns having the highest rate followed by pronouns and then other words. There is a much larger gap from one group to the next, and there are only four tokens for the other category. We now return to the result that the statistical significance of the syntactic context lies with the PDM forms. The tokens of absent *que* with lexical nouns for the PDM forms reveal that *penser*, *savoir*, and *croire* are the matrix verbs. For pronouns, *croire*, *dire*, *savoir*, and *voir* are the matrix verbs used. If we turn once again to the phonological context and look only at PDM forms, lexical nouns used with absent *que* always begin with a liquid, nasal, or stop. These three phonological groups have the same and highest absent *que* rates for PDM forms. On the other hand, pronouns that are used with absent *que* begin primarily with a fricative. There are a few that begin with a vowel and one with a stop. Fricatives and vowels both have lower absent *que* rates for PDM forms than liquids, nasals, and stops. It is thus possible that complementizer realization for PDMs has a phonological component, especially since the matrix verbs used with pronouns and lexical nouns do not indicate a motivation.

The presence of intervening material in the matrix clause is not commonly seen with absent *que* in Africa. There is just a 3.3% absent *que* rate when there is intervening material. The rate for NDM forms is almost identical to the overall rate, and PDMs never appear with the

absent variant when there is intervening material in the matrix clause. It should be noted that there are only two examples of intervening material in the matrix clause for PDMs.

In the embedded clause there are near identical absent *que* rates for presence and absence of intervening material. This is due to the NDM forms where there is roughly a 5% higher absent *que* rate when there is intervening material. An analysis of these forms reveals that the intervening material usually consists of more than one word. Once again, it seems that the length of the intervening material in the embedded clause does not correlate with variable *que*.

Intervening material was included in the linear models for the full data set and the NDM forms and is not significant for either. The matrix clause pattern follows what Martineau (1985) and Warren (1994) found for this clause, but the embedded clause pattern does not align with Warren's (1994) finding. In summary, it would be more likely for one to hear an African speaker say *tu sais Ø c'est toi qui, qui me suivais chaque fois* 'you know Ø it's you who, who followed me each time' (AF48a) with a PDM, absent *que*, and no intervening material in either clause than when there is intervening material as in *tu vois* ***que*** *moi j'ai des problèmes* 'you see **that** me I have problems' (AM28).

Regarding the matrix verbal elements, all forms that can be discourse markers have a higher absent *que* rate than the forms of the same verbs that cannot be discourse markers, with the exception of *trouver*, as shown in Table 4.10. *Dire* has a rate of 12.5% for the NDM forms and 100% for the PDM forms. The corresponding rates for the other verbs are as follows: *savoir* 33.3%/71.4%, *voir* 29.4%/66.7%, *croire* 0%/40.0%, *penser* 0%/7.7%, and *trouver* 100%/0%. It should be noted that there are very few tokens of most of these verbs. *Trouver* shows the largest difference in rates depending on which verb forms are chosen followed by *dire* and then *croire*.

While *trouver* has the highest rate for NDM forms and *dire* has the highest rate for the PDM forms, *savoir* and *voir* are ranked second and third, respectively, for both types of verbs.

Table 4.10. Africa – Absence of *que* as percentages for NDM and PDM forms for verbal element

Non-Discourse Markers			Potential Discourse Markers		
Factor	% Ø	N=Ø/Total	Factor	% Ø	N=Ø/Total
Verbal Element			Verbal Element		
croire	0	0/1	*croire*	**40.0**	6/15
dire	12.5	5/40	*dire*	**100**	3/3
penser	0	0/6	*penser*	**7.7**	1/13
savoir	33.3	7/21	*savoir*	**71.4**	5/7
trouver	**100**	1/1	*trouver*	0	0/3
voir	29.4	5/17	*voir*	**66.7**	2/3
vouloir dire	14.3	1/7	*vouloir dire*	- -	- -
avoir l'impression	0	0/2			
c'est	7.7	1/13			
c'est-à-dire	43.8	7/16			
c'est vrai	33.3	1/3			
falloir	0	0/42			
paraître	7.1	1/14			
peut-être	41.2	7/17			
pouvoir dire	20.0	2/10			
se dire	14.3	1/7			
vouloir	0	0/12			

Focusing on the PDM forms, we see that *penser* and *trouver* have low absent *que* rates compared to those of the other four verbs. Returning to the idea of *croire* favoring formal speech and being less favorable to absent *que* than *penser* as stated in Martineau (1985:66-67) for Canadian French speakers, both verbs occur with a similar frequency to one another, and *croire* has a higher rate than *penser*. Thus, African speakers do not use *que* with these two verbs in the same way as the Canadian speakers that Martineau observed.

Of the verbs that Martineau (1993) and Blondeau and Nagy (2008) found to favor absent *que*, all except *sembler* are present in the corpus of African speakers. *Savoir* and *croire* have

rates between 37% and 43% as shown in Table 4.8. *Trouver* and *dire* have rates between 18% and 25%, while *penser* has a rate of 5.3%. The analyses in this chapter do not allow us to say whether these verbs favor or disfavor absent *que*. However, we can see that they all allow this construction and that *savoir* and *croire* allow it more often than the others.

As for the verbs that Andersen (1993) found to favor absent *que*, *voir* has a rate of 35.0%, *dire* 18.6%, *savoir* 42.9%, and *trouver* 25.0%. This group, combined with *se souvenir* and *se rappeler*, had a rate of 30% in her study. The combined rate in the present study is 30.4%. *Être* + adjective, which she found to have a rate of 35%, did not appear in the subcorpora of African speakers. In the same study, she found that *falloir, vouloir, croire*, and *penser* were never used with absent *que*. The data from African speakers indicate that the same is true for *falloir* and *vouloir* but not for *croire* and *penser*.

In summary, the full data set follows the five-group sonority hierarchy, while the two subsets diverge slightly from this pattern. For both the full data set and the NDM forms, liquids differ significantly from fricatives and decrease the chance of absent *que*. All data sets follow the three-group sonority hierarchies used in previous studies. Lexical nouns have the highest rate for the full data set and PDM forms while pronouns have a very similar rate for the full data set. By contrast, lexical nouns have the lowest rate for the NDM forms. Pronouns are the most frequent of the syntactic groups. The syntactic context is significant for the full data set but not for the NDM forms. Compared to pronouns, words other than pronouns and lexical nouns decrease the chance of absent *que*. The variation according to syntactic context is attributed to the PDM forms, which indicate a possible phonological motivation. Intervening material in the matrix clause results in a lower absent *que* rate regardless of the verb type. However, for the embedded clause, there is a higher rate of absent *que* when there is intervening material with the NDM

forms but not for the PDM forms. Regarding the matrix verb, all verb forms that can be discourse markers have a higher absent *que* rate than the forms of the same verbs that cannot be discourse markers, except for *trouver*. The verb with the highest absent *que* rate depends on the type of verb, but *savoir* and *voir* are both ranked the same regardless of the verb type.

4.2.2. Cross-regional Comparison and Discussion of Linguistic Factors

Of the five linguistic factors tested, all except for intervening material in the embedded clause are significant for at least one linear model in at least one region. Across models and regions, the matrix verb and phonological context after the complementizer are the most common significant factors. Cross-regional trends based on absent *que* rates as well as the statistical significance of linguistic factors will now be presented.

4.2.2.1. Phonological Context

Based on overall absent *que* rates shown in Table 4.11, Europe and Quebec both indicate that fricatives have the highest rate. This finding is in accordance with previous studies (Cedergren & Sankoff 1974; Sankoff 1980; Martineau 1985; Blondeau & Nagy 2008) that found sibilants to favor absent *que* more than other sounds. Africa does not follow this pattern but rather follows the five-group sonority hierarchy. Stops have the highest rate followed by fricatives, nasals, liquids, and then vowels. Using the obstruent/sonorant/vowel hierarchy examined by Warren (1994) and Liang et al. (2021), all regions indicate a higher absent *que* rate for obstruents followed by sonorants and then vowels in agreement with the results of Liang et al. (2021). Three different sonority hierarchies have been considered in the present study. In the beginning, it was thought that the five-group hierarchy would be most suitable since it would allow for a

more detailed analysis. Upon further reflection, the obstruent/sonorant/vowel distinction is the most appropriate. The primary reason to prefer this hierarchy is because the sonority of stops and fricatives depends on the specific phonemes considered. Voiced fricatives (/v/, /z/, /ʒ/) are more sonorous than voiced stops (/b/, /d/, /g/), which are more sonorous than voiceless fricatives (/f/, /s/, /ʃ/) (Parker 2011:1177). Voiceless stops (/p/, /t/, /k/) are the least sonorous of all. Since the sonority of stops and fricatives is intermingled, it is most logical to group them together as obstruents.

In the linear models for the full data set of Quebec speakers and African speakers, the phonological context is statistically significant. In Quebec, liquids, nasals, and vowels differ significantly from fricatives with vowels differing the most. The three significant groups reduce the chance of absent *que*. For African speakers, liquids differ significantly from fricatives and reduce the chance of absent *que*. The most surprising trend is that stops have the lowest absent *que* rate in Europe. On a final note, the range of rates for each region reveals that Europe has a very small range of roughly 4% for the following phonological context, whereas Quebec and Africa have a range near 12% and 15%, respectively.

Table 4.11. Absent *que* rates as percentages by phonological context and region

	Europe		Quebec		Africa	
Phonological Context	% ∅	N=∅/Total	% ∅	N=∅/Total	% ∅	N=∅/Total
vowel	24.2	74/306	25.0	137/549	14.0	14/100
liquid	25.0	11/44	24.7	20/81	14.3	6/42
nasal	26.1	6/23	34.6	9/26	22.2	2/9
fricative	27.0	80/296	36.9	232/629	28.1	25/89
stop	23.1	21/91	29.8	72/242	29.4	10/34

Separating the data by verb type revealed that each region shows differences in the order of phonological groups based on the verb type as shown in Table 4.12. In Europe, the order of liquids does not change but the order of all other groups does. Nasals show the largest difference since they have the highest absent *que* rate for the NDM forms and the lowest rate for the PDM forms. Quebec speakers show the most consistency across verb types as liquids and stops are the only groups to change order, and the change is very slight. In Africa, the order of fricatives, liquids, and vowels changes depending on verb type. The rate from one phonological group to the next is more markedly different in Europe according to verb type versus the full data set. The linear models indicate that the phonological context is significant for the PDM forms in Europe and Quebec and for the NDM forms in Africa. In Europe, nasals differ significantly from fricatives, while stops and vowels do in Quebec. In Africa, it is liquids that differ from fricatives. Each phonological group that is statistically significant reduces the chance of absent *que* compared to fricatives.

Table 4.12. Absent *que* rates as percentages by phonological group and verb type

Non-Discourse Markers			Potential Discourse Markers		
Factor	% Ø	N=Ø/Total	Factor	% Ø	N=Ø/Total
Phonological Context			Phonological Context		
Europe					
vowel	11.4	21/185	vowel	**43.8**	53/121
liquid	12.0	3/25	liquid	**42.1**	8/19
nasal	**28.6**	4/14	nasal	22.2	2/9
fricative	13.0	22/169	fricative	**45.7**	58/127
stop	6.5	3/46	stop	**40.0**	18/45
Quebec					
vowel	13.1	50/382	vowel	**60.0**	99/165
liquid	17.9	10/56	liquid	**68.0**	17/25
nasal	23.5	4/17	nasal	**77.8**	7/9
fricative	16.9	65/384	fricative	**60.8**	146/240
stop	19.0	35/184	stop	**66.1**	37/56
Africa					

vowel	12.9	11/85	vowel	**20.0**	3/15
liquid	10.5	4/38	liquid	**50.0**	2/4
nasal	14.3	1/7	nasal	**50.0**	½
fricative	22.9	16/70	fricative	**47.4**	9/19
stop	26.7	8/30	stop	**50.0**	2/4

There are some similarities regarding the phonological context when looking at different regions. In Europe and Quebec the absent *que* rate is the highest before a fricative and the phonological context is significant for the PDM forms. In Quebec and Africa liquids are significantly different from fricatives. However, there is no clear cross-regional pattern. In more general terms, it can be said that the phonological context is statistically significant in each region and that the verb type affects the order of phonological groups.

4.2.2.2. Syntactic Context

The syntactic context following the complementizer shows a different pattern for each region based on overall absent *que* rates, as indicated in Table 4.13. Pronouns have the highest rate in Europe, while other words do in Quebec and lexical nouns in Africa. However, lexical nouns and pronouns have very similar rates to one another in each region. Pronouns have the second highest rate in Quebec and Africa, and other words have the lowest rate in Europe and Africa. The linear model for the full data set of African speakers indicates that the syntactic context is statistically significant. For these speakers, other words are significantly different from pronouns, and they reduce the chance of absent *que*. Regarding the most frequent pronouns following the complementizer, Connors (1975), in a study of Quebec French, found that most were forms of *ce* and *je*. The present study indicates that *ce* is the most common in each region followed by *il* and

then *je*. It is likely that *il* is so common since it can be used impersonally to mean 'it' or personally to mean 'he'.

Table 4.13. Absent *que* rates as percentages by syntactic context and region

	Europe		Quebec		Africa	
Syntactic Context	% Ø	N=Ø/Total	% Ø	N=Ø/Total	% Ø	N=Ø/Total
pronoun	27.1	162/598	30.2	403/1334	21.5	37/172
lexical noun	26.2	16/61	29.4	30/102	21.7	10/46
other	13.9	14/101	40.7	37/91	17.9	10/56

In Europe and Quebec, other words have the lowest and highest rates, respectively, regardless of verb type, but the other two groups change depending on the verb type. In Africa, pronouns retain the middle position, but the other two groups change based on the type of verb in the matrix clause. As for the full data set, lexical nouns and pronouns have very similar rates in Europe for both verb types, and they have similar rates in Quebec for the NDM forms. The syntactic context was included as a fixed effect for the NDM forms in Africa and is not significant. Since the same factors, social and linguistic, were included in the full data set and the NDM set, and since the syntactic context was significant for the full data set but not the NDMs, the significance of this factor stems from the PDM forms. While there is a difference in the overall pattern of the syntactic context according to region, each region has something in common with the other two regions.

Both the phonological and syntactic contexts were found to be important by Connors (1975), Martineau (1985), and Liang et al. (2012). Connors (1975) believed the syntactic context to be more important than the phonological one, whereas Martineau (1985) asserted the reverse, and Liang et al. (2021) found the two to be independently statistically significant. These two

factors were included as fixed effects in six out of the nine linear models. The phonological context is statistically significant in five, and the syntactic is significant in one. Thus, the results of the present study are consistent with those of Martineau (1985). For this latter model, the p-value of the phonological context is lower than that of the syntactic. Thus, we can conclude that the phonological context is more important in conditioning absent *que* than the syntactic context, and the significance of the syntactic context for African speakers is linked to the PDM forms.

4.2.2.3. Intervening Material

Regarding intervening material between the matrix verb and the complementizer, all three regions have a higher absent *que* rate when there is no intervening material as indicated in Table 4.14. This is the expected outcome since intervening material can diminish the comprehension of an utterance as in example (71), repeated below as (73).

(73) Je pense Ø? jamais Ø? je vais y aller. (Martineau 1985:81)

'I think Ø I will never go there.'

'I never think Ø I will go there.'

Table 4.14. Absent *que* rates as percentages by intervening material and region

	Europe		Quebec		Africa	
	% Ø	N=Ø/Total	% Ø	N=Ø/Total	% Ø	N=Ø/Total
Intervening Material Matrix Clause						
no	26.3	182/693	31.9	440/1380	23.0	56/244
yes	14.9	10/67	20.4	30/147	3.3	1/30
Intervening Material Embedded Clause						
no	26.6	171/642	29.9	423/1413	20.8	43/207
yes	17.8	21/118	41.2	47/114	20.9	14/67

When verb type is taken into account, Europe and Quebec display different patterns, while Africa remains the same, as seen in Table 4.15. The NDM forms have a higher rate when there is intervening material in Europe and Quebec. However, there is a very small difference between the rates for presence and absence of intervening material, 0.8% and 2.1% respectively. The linear model for the PDM forms in Quebec reveals that this factor is statistically significant. Intervening material in the matrix clause is the factor that has thus far indicated the most cross-regional similarity. All regions have the same overall pattern and verb type affects the results for Europe and Quebec.

Table 4.15. Absent *que* rates as percentages for intervening material in the matrix clause by region and verb type

Non-Discourse Markers			Potential Discourse Markers		
Intervening Material Matrix Clause	% Ø	N=Ø/Total	Intervening Material Matrix Clause	% Ø	N=Ø/Total
Europe					
yes	12.8	6/47	yes	**20.0**	4/20
no	12.0	47/392	no	**44.9**	135/301
Quebec					
yes	17.9	20/112	yes	**30.3**	10/33
no	15.8	144/911	no	**64.1**	296/462
Africa					
yes	3.6	1/28	yes	0	0/2
no	19.3	39/202	no	**40.5**	17/42

Less cross-regional similarity is found when considering intervening material between *que* and the subject of the embedded clause. Quebec and Africa have a higher absent *que* rate when there is intervening material, while Europe shows the opposite pattern. However, there is only a 0.1% difference between presence and absence of intervening material in Africa. The verb type is associated with a different outcome in Africa only, although there are only two tokens of

intervening material in the embedded clause for the PDMs in this region. Given these results, there are three different overall patterns, one for each region since Africa essentially has the same rate for both possibilities. The verb type appears to impact the absent *que* rates in Africa, although this is not a definitive claim since there are so few tokens for one category. Quebec and Africa have a higher absent *que* rate in the presence of intervening material for the embedded clause than for the matrix clause.

4.2.2.4. Matrix Verb

Beginning with the overall absent *que* rates and looking at the matrix verbs with the highest rates in each region, the one commonality across all regions is that *savoir* has a relatively high rate. It has the third highest rate in Europe and Africa and the fourth highest in Quebec. In Europe and Quebec *vouloir dire* has a rate above 60%. It is ranked second in Europe and fifth in Quebec. *C'est-à-dire* has the second highest rate in Quebec and Africa. Regarding the verbs that are never used with absent *que*, both *falloir* and *vouloir* fall into this category in Europe and Africa. *Avoir l'impression* is never used with absent *que* in Quebec and Africa. The verbs that have a low variable rate (<15%) are sometimes the same in two regions. *Trouver*, *croire*, and *être* + adj. belong to this category in Europe and Quebec. In Europe and Africa, *penser* and *c'est* have low variable rates. *Se dire* and *paraître* are in this category in Quebec and Africa. Thus, the use of some verbs is similar in two regions, but only one verb, *savoir*, displays consistency in its use with absent *que* across all regions.

Matrix verb semantic groups were included as a fixed effect in the linear models of the full data set in Europe and Quebec. In Europe, judgment verbs, *penser* and *trouver*, are statistically significant compared to epistemic verbs, and they reduce the chance of absent *que*. In

Quebec all semantic groups are statistically significant and differ from epistemic verbs. Compared to this latter group, they all decrease the chance of absent *que*. These results call to mind the finding of Thompson and Mulac (1991:243-249) that an epistemic matrix clause subject and verb favored *that*-deletion in English. While their study treated a language other than French, the present study indicates that epistemic verbs are associated with a higher absent *que* rate compared to at least one other group.

The division of matrix verb forms into NDMs and PDMs makes possible a more nuanced analysis. For each region, the PDM forms have an overall absent *que* rate higher than that of the NDM forms. In Europe the NDM forms have an overall rate of 12.1%, whereas the PDM forms have a rate of 43.3%. In Quebec the rates are 16.0% and 61.8% respectively, and in Africa they are 17.4% and 38.6%. Also, the distinction of verb type affects the ordering of factor levels for almost every factor in every region. These results are probably in some part due to the inadvertent inclusion of some discourse markers. However, I propose that the differences seen based on matrix verb type are primarily a case of syntactic analogy as described in Section 3.3.1.1.

Focusing on just the PDM forms, most verbs are ranked very similarly across regions according to absent *que* rates, which are provided in Table 4.16. The largest difference can be seen with *dire* and *croire*, which have a considerably higher ranking in Africa than in the other two regions. *Trouver* has a low absent *que* rate in all three regions. The matrix verb was a fixed effect in the linear model for the PDM forms in Europe. This model indicates that *dire*, *savoir*, *vouloir dire*, and *voir* differ significantly from *croire* and correlate with a higher absent *que* rate.

Table 4.16. Absent *que* rates as percentages by PDM verb and region

	Europe		Quebec		Africa	
Verbal Element	% Ø	N=Ø/Total	% Ø	N=Ø/Total	% Ø	N=Ø/Total
voir	98.1	51/52	41.7	5/12	66.7	2/3
vouloir dire	90.9	20/22	97.6	41/42	- -	- -
savoir	74.4	29/39	91.6	196/214	71.4	5/7
dire	61.9	13/21	28.6	6/21	100	3/3
penser	15.6	15/96	32.5	52/160	7.7	1/13
trouver	15.2	5/33	15.4	6/39	0	0/3
croire	10.3	6/58	0	0/7	40.0	6/15

The verbs *penser* and *croire* have very similar meanings and can be used in the same constructions. It was suggested by Martineau (1985) in her study of French in Quebec and Ontario that *croire* is used more often in formal speech and *penser* in informal speech. In turn, she proposed that this makes *croire* less favorable to absent *que*. The rates shown in Table 4.17 indicate that *penser* is used more frequently with absent *que* in Europe and Quebec but not in Africa. The difference between the absent *que* rates for *penser* and *croire* is markedly different in both Quebec and Africa, while there is only a small difference in Europe. *Penser* is used more frequently than *croire* except among PDMs in Africa, where there is little difference in the frequency of these two verbs. *Croire* is more frequent in Europe than it is in Quebec or Africa. Based on these results, I cannot say that African speakers associate *penser* with informal speech and *croire* with formal speech. The data indicate that this is possibly the case for Quebec and European speakers, although it is less clear for European speakers.

Table 4.17. Absent *que* rates as percentages for *penser* and *croire* by region

	Europe		Quebec		Africa	
	% Ø	N=Ø/Total	% Ø	N=Ø/Total	% Ø	N=Ø/Total
Overall						
penser	14.9	15/101	25.4	59/232	5.3	1/19
croire	8.1	6/74	9.1	1/11	37.5	6/16
Discourse Markers						
penser	15.6	15/96	32.5	52/160	7.7	1/13
croire	10.3	6/58	0	0/7	40.0	6/15

In their studies of spoken French in Quebec and Ontario, Martineau (1993) and Blondeau and Nagy (2008) collectively found the verbs *croire*, *dire*, *penser*, *savoir*, *sembler*, and *trouver* to favor absent *que*. The results of the present study, shown in Table 4.18, reveal that these six verbs allow absent *que* in each region except for *sembler*, which had no occurrences in Africa. Of these verbs, *savoir* has the highest rate in each region, and *croire* has the lowest rate in Europe and Quebec. Andersen's (1993) study of Parisian French in the 1980s revealed that *c'est* + adjective had an absent *que* rate of 35%. *Être* + adjective, which includes *c'est* + adjective, in the present study has a much lower rate in Europe and Quebec than what she found. There are no occurrences of this structure in the African data. The verbs *voir*, *dire*, *savoir*, *trouver*, *se souvenir*, and *se rappeler* had a combined rate of 30% in her study. The first four of these verbs are included in the present study and have a very similar combined rate in Quebec (31.0%) and Africa (30.4%). In Europe it is considerably higher (45.8%). The same study indicated that *croire*, *penser*, *falloir*, and *vouloir* did not have any instances of absent *que*. The present study confirms this for *falloir* and *vouloir* in Europe and Africa. However, in Quebec all of these verbs appear with absent *que*, even if at a very low rate. All three regions allow absent *que* with *croire* and *penser*.

Table 4.18. Absent *que* rates as percentages by matrix verb and region

	Europe		Quebec		Africa	
Matrix Verb	% Ø	N=Ø/Total	% Ø	N=Ø/Total	% Ø	N=Ø/Total
croire	8.1	6/74	9.1	1/11	37.5	6/16
dire	28.4	23/81	11.7	19/162	18.6	8/43
penser	14.9	15/101	25.4	59/232	5.3	1/19
savoir	52.1	37/71	74.8	205/274	42.9	12/28
sembler	16.7	1/6	43.2	19/44	- -	- -
trouver	10.4	5/48	13.2	9/68	25.0	1/4
être + adj.	4.8	1/21	14.1	9/64	- -	- -
voir	92.3	60/65	24.2	8/33	35.0	7/20
falloir	0	0/46	11.2	36/321	0	0/42
vouloir	0	0/8	1.8	1/56	0	0/12

Even though the matrix verb could not be included as a fixed effect in every linear model due to a lack of convergence, its effect can still be seen. When not included as a fixed effect, the matrix verb was considered a random effect. Each time, the variance indicates that it explains some of the variation in the data. When speaker was also a random effect, the variance of the matrix verb was the greater of the two. Additionally, each factor except for intervening material indicates a different pattern when comparing NDM forms to PDM forms. These trends are consistent across regions.

The result that NDM forms have lower absent *que* rates than PDM forms in almost every instance highlights the need to definitively identify any discourse markers in the data. Using the procedure outlined in Avanzi (2012), instances that are not clear could be tested prosodically and identified as a matrix verb or discourse marker. If a marked distinction persists between verb types after removing any discourse markers, this would indicate that syntactic analogy likely plays a role.

4.2.2.5. Summary

Phonological context has been shown to be statistically significant for the PDM forms in Europe and Quebec. These two regions also have the highest absent *que* rate for fricatives. Liquids differ significantly from fricatives in Quebec and Africa, although in Quebec this is for the PDMs, whereas in Africa it is for the NDMs. The syntactic context reveals different patterns by region, although each region has something in common with one other region. The three most common pronouns after the complementizer are the same in each region. The phonological context is more important than the syntactic context in all regions. Intervening material in the matrix clause is associated with a lower absent *que* rate in all three regions. On the other hand, the embedded clause has a higher absent *que* rate when intervening material is present than when it is absent, especially in Quebec and Africa.

Concerning the matrix verb, *savoir* has one of the highest rates in each region, and it has the highest rate of the verbs that Martineau (1993) found to favor absent *que*. In Europe and Quebec at least one semantic verb group differs significantly from epistemic verbs.

Overall, the PDM forms have a higher absent *que* rate than the NDM forms in each region. Also, the PDM verbs have similar rankings across regions with the exception of *dire* and *croire* in Africa. Of the verbs that Andersen (1993) found to never be used with absent *que*, *penser* and *croire* both have instances of absent *que* in all regions.

4.3 Conclusion

Based on the generalized linear mixed-effects models shown in Section 3.2 and absent *que* rates, the phonological context indicates that all regions follow the sonority hierarchy in which obstruents have the highest absent *que* rate followed by sonorants and then vowels. The

phonological context following the complementizer is statistically significant in each of the three regions, and the matrix verb type affects the ordering of the five sonority groups from lowest to highest absent *que* rate. It was determined that the phonological context is more important than the syntactic one in all regions and that the phonological context does not play a large role in *que* realization in Europe.

Regarding the syntactic context following the complementizer, pronouns and lexical nouns have similar rates to one another in each region. In Europe pronouns have the highest absent *que* rate, and the pronouns that start with a vowel have the lowest rate. In Quebec other types of words have the highest absent *que* rate due to their frequent occurrence with the matrix verbs *savoir* and *vouloir dire*. The syntactic context is statistically significant in Africa and lies with the PDM forms. There is potentially a phonological component to this context.

Intervening material in the matrix clause is statistically significant in Quebec. For all regions, there is a higher absent *que* rate when intervening material is absent. For Europe and Quebec, when intervening material is present in the matrix clause and absent *que* is used, the intervening material usually consists of one word. This is true for the NDM forms, which occur less often with absent *que*. If the intervening material is longer than one word, this almost always results in the present complementizer. Intervening material in the embedded clause does not condition *que* realization. There can be intervening material present and there be a higher absent *que* rate than when it is absent in Quebec and Africa. Also, the length of the intervening material in this clause appears to have no bearing on the complementizer form.

The matrix verb is statistically significant in Europe and Quebec, and its effect can be seen in all three regions by different results for almost every factor according to verb type. *Savoir* has a high absent *que* rate in all regions, and epistemic verbs are significantly different

from at least one other semantic group in Europe and Quebec. *Penser* has a higher absent *que* rate than *croire* in Europe and Quebec, which may be attributed to *croire* being viewed as more formal, especially in Quebec. Chapter 5 will explore individual speaker variation regarding the two factors shown to be significant in each region, matrix verbal element and phonological context.

Chapter 5

Individual Speaker Variation

5.1. Introduction

In Chapter 4 it was shown that when included as a fixed effect, the matrix verb and the phonological context following the complementizer are statistically significant in at least one model for each of the three regions and that these two factors are more frequently significant than other linguistic variables across regions and models. In the case of Africa, the matrix verb was not a fixed effect but rather a random effect. An analysis of the two types of matrix verbs strongly points to the importance of the verbal element for African speakers. In many models, speaker was a random effect and was shown to account for some of the variation in the data. The present chapter looks at the matrix verb and phonological context through the lens of individual speakers to gain an additional perspective. Section 5.2 provides the absent *que* rate for each speaker included in this study and a brief summary. Next, individual *que* variation according to the matrix verb is explored by region in Section 5.3 to arrive at a regional comparison. Section 5.4 examines the phonological context for each region. The final section presents concluding remarks.

5.2. Absent *Que* Rate by Speaker

The individual rates of absent *que* for all 48 speakers from the European corpus are shown in Table 5.1. There are eight speakers who never use absent *que*, and the speaker with the highest

rate is EM26 (70.0%) followed closely by EF24 (69.7%) and EM54 (66.7%). Half of the speakers who never use absent *que* are younger males. Middle-aged males are the only social group in which all speakers use absent *que*. Younger females as a group have the highest average rate (31.2%), while younger males have the lowest (15.1%). Younger speakers as a whole show the most variability in rates of absent *que*, while the older speakers show the least variability. Middle-aged females and males pattern similarly regarding the average rate and range as do older females with older males. The absent *que* rates indicate that a few of the younger speakers, such as EF24, EF20, EF22, EF30, and EM26, are leading the move toward absent *que*.

Table 5.1. Absent *que* rates as percentages for individual European speakers

Speaker	% ø	N=ø/Total	Speaker	% ø	N=ø/Total
Younger Females			Younger Males		
EF16	8.3	4/48	EM17	26.7	20/75
EF20	50.0	7/14	EM19	0	0/11
EF22	40.0	2/5	EM21	16.7	1/6
EF24	69.7	23/33	EM23	0	0/2
EF26	0	0/2	EM24	0	0/4
EF27a	28.0	7/25	EM26	70.0	28/40
EF27b	13.6	9/66	EM29	7.1	1/14
EF30	40.0	4/10	EM30	0	0/4
Group average	31.2		Group average	15.1	
Group range	69.7		Group range	70.0	
Middle-Aged Females			Middle-Aged Males		
EF42	11.1	2/18	EM42	18.2	2/11
EF43	14.3	1/7	EM43	33.3	1/3
EF46	45.5	5/11	EM44	13.3	2/15
EF48	15.4	4/26	EM45a	20.0	1/5
EF52	25.0	5/20	EM45b	14.3	1/7
EF53	10.0	1/10	EM51	15.6	5/32
EF54	36.0	9/25	EM53	14.3	1/7
EF55	0	0/15	EM54	66.7	2/3
Group average	19.7		Group average	24.5	
Group range	45.5		Group range	53.4	
Older Females			Older Males		

EF65	23.8	5/21	EM70	28.6	2/7
EF68	22.7	5/22	EM73	33.3	1/3
EF69	21.7	5/23	EM75	27.3	3/11
EF78	26.7	4/15	EM78	28.6	2/7
EF80	32.0	8/25	EM79	0	0/2
EF81	0	0/1	EM81a	37.5	3/8
EF82	18.2	2/11	EM81b	10.5	2/19
EF86	25.0	1/4	EM81c	14.3	1/7
Group average	21.3		Group average	22.5	
Group range	32.0		Group range	37.5	

The Quebec corpus consists of 46 speakers whose individual rates of absent *que* are shown in Table 5.2. One speaker, QM25b, a younger male, never uses absent *que*. QM42 has the highest rate (55.2%) followed by QF23 (53.8%). Older females have the highest average rate (35.2%), while older males have the lowest (27.4%). Younger speakers have the most variability in absent *que* rates, while older speakers, particularly females, have the least variability. With respect to social group averages and ranges, younger females and males pattern very similarly, and middle-aged females and males do as well, particularly regarding the group averages. Older females and males differ from each other more, but they both have lower ranges than the other groups.

Table 5.2. Absent *que* rates as percentages for individual Quebec speakers

Speaker	% Ø	N=Ø/Total	Speaker	% Ø	N=Ø/Total
Younger Females			Younger Males		
QF16a	26.4	14/53	QM22	25.0	11/44
QF16b	20.0	9/45	QM24a	35.7	10/28
QF17	27.9	12/43	QM24b	25.0	2/8
QF22	31.8	7/22	QM25a	38.9	21/54
QF23	53.8	7/13	QM25b	0	0/1
QF24	11.4	5/44	QM26	42.9	15/35
QF25	42.6	23/54	QM27a	20.8	5/24
QF27	10.3	10/97	QM27b	42.9	21/49

Group average	28.0		Group average	28.9	
Group range	43.5		Group range	42.9	
Middle-Aged Females			Middle-Aged Males		
QF42	42.1	16/38	QM42	55.2	16/29
QF47	44.8	13/29	QM47	40.5	15/37
QF48a	25.0	8/32	QM48	16.7	3/18
QF48b	31.2	15/48	QM50	16.7	1/6
QF49	19.3	11/57	QM51a	44.0	11/25
QF50	26.9	7/26	QM51b	17.8	8/45
QF52	35.8	24/67	QM53	32.3	10/31
QF53	28.8	15/52	QM55	30.8	4/13
Group average	31.7		Group average	31.8	
Group range	25.5		Group range	38.5	
Older Females			Older Males		
QF65	27.8	10/36	QM65	21.4	3/14
QF71	38.1	8/21	QM73a	21.7	5/23
QF74	31.2	6/16	QM73b	40.3	25/62
QF79	41.7	5/12	QM80	37.8	17/45
QF81	31.0	9/29	QM81	26.9	7/26
QF82	43.2	16/37	QM87	26.7	4/15
QF95	33.3	6/18	QM95	16.7	1/6
Group average	35.2		Group average	27.4	
Group range	15.4		Group range	23.6	

The absent *que* rates for the 25 speakers from the African corpus are shown in Table 5.3. There are four speakers who never use absent *que*. Two of these are younger females and the other two are middle-aged females. Two speakers, AF19 and AF48a, have the highest rate of absent *que* at 60.0%. Females show the most variability in absent *que* rates, and both younger and middle-aged females have the same range. Younger females use absent *que* less than middle-aged females, and females use it more than African males.

Table 5.3. Absent *que* rates as percentages for individual African speakers

Speaker	% Ø	N=Ø/Total	Speaker	% Ø	N=Ø/Total
Younger Females			Younger Males		
AF19	60.0	3/5	AM28	4.5	1/22
AF23	26.7	4/15			
AF26	6.7	1/15			
AF27	0	0/3			
AF28a	20.0	2/10			
AF28b	19.0	4/21			
AF28c	20.0	2/10			
AF29	0	0/8			
Group average	19.1		Group average	4.5	
Group range	60.0		Group range	N/A	
Middle-Aged Females			Middle-Aged Males		
AF41a	11.1	1/9	AM40	5.9	1/17
AF41b	12.5	1/8	AM50	27.3	3/11
AF44a	41.7	10/24			
AF44b	33.3	1/3			
AF45	11.1	2/18			
AF46a	16.7	1/6			
AF46b	42.9	3/7			
AF47	33.3	2/6			
AF48a	60.0	3/5			
AF48b	0	0/4			
AF48c	26.7	4/15			
AF50	23.1	3/13			
AF54	0	0/6			
AF55	38.5	5/13			
Group average	25.1		Group average	16.6	
Group range	60.0		Group range	21.4	

Comparing these three regions, Europe has the speaker with the highest individual absent *que* rate (70%), followed by Africa (60%), and then Quebec (55.2%). Europe also has the largest number of speakers who never use absent *que*, and Africa has half this number. In Quebec, almost all speakers use absent *que* to some extent. Of the speakers who never use absent *que*, half belong to the younger age group in Europe and Africa, and the one speaker in Quebec who

meets this criterion is a younger male. Thus, Quebec speakers show more use of absent *que* as a whole. Members of the younger generation vary more widely in their use of absent *que* in both Europe and Quebec, while the older generation shows more stability in these two regions. In Africa, the variability of individual speaker use patterns the same for females, regardless of age group. Figure 5.1 illustrates the average absent *que* rates by social group for each of the three regions. As a whole, absent *que* use by social group shows the most stability in Quebec, while Europe and Africa indicate a larger degree of variability.

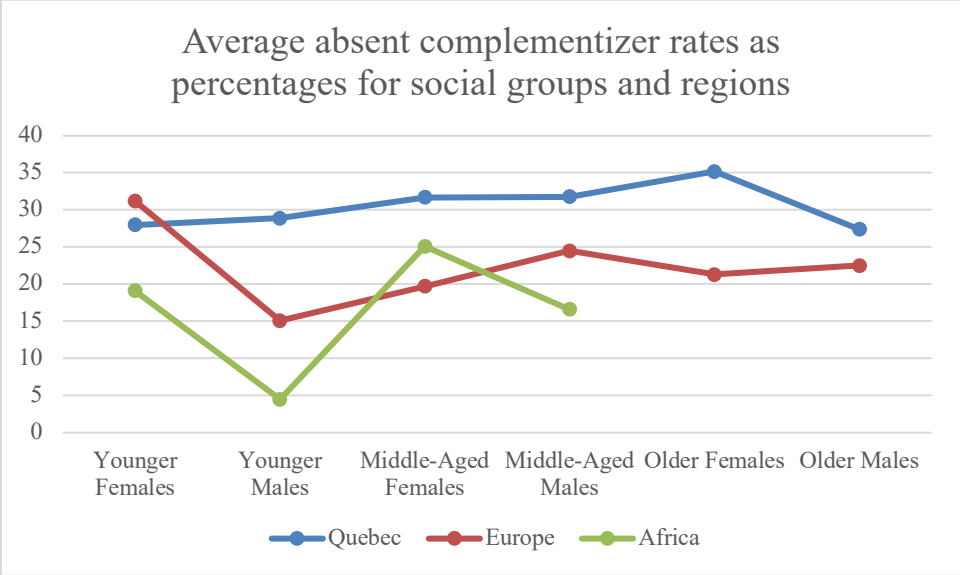

Figure 5.1. Average absent *que* rates as percentages for social groups and regions

5.3. Individual Variation by Preceding Verbal Element

5.3.1 Matrix Verb Trends by Individual Speaker

The statistical analysis of each region outlined in Section 4.2 indicated that the matrix verb is an important factor in conditioning absent *que*. In Europe and Quebec it is a statistically significant factor, while in Africa it shows that some variation in the data can be attributed to the matrix

verb. The full European data set indicated that judgment verbs differ significantly from epistemic verbs and decrease the chance of absent *que*. For the PDM forms, *dire*, *savoir*, *vouloir dire*, and *voir* differ significantly from *croire* and increase the chance of absent *que*. In Quebec all matrix verb semantic groups are statistically different from epistemic verbs, and they decrease the chance of absent *que* compared to the reference level. Also, NDMs and PDMs show different patterns for all but one factor, thereby underscoring the importance of the matrix verb. The following analysis adds another dimension to the one presented in Chapter 4 by looking at matrix verb trends by individual speaker.

In order to be included in the variability corpus a speaker must have at least two tokens for one verb, which is the minimum required to determine variability in an individual's speech. The European variability corpus is comprised of 43 speakers, 17 verbal elements, and 607 tokens. The verbal elements *c'est-à-dire*, *heureusement*, *paraître*, and *promettre* cannot be tested for variability because they do not have enough tokens for any given speaker. With the exception of *c'est-à-dire*, these verbal elements had an overall absent *que* rate of 0% in the European corpus, and all four of these elements had fewer than nine tokens each.

The Quebec variability corpus consists of 45 speakers, 19 verbs, and 1,376 tokens. The verbal elements *avoir raison*, *jurer*, and *promettre* can not be tested for variability because they do not have enough tokens for any given speaker. In the Quebec corpus, *avoir raison* had an absent *que* rate of 100%, while the rate of *jurer* was 75%, and *promettre* was 0%. All three of these elements had fewer than five tokens each.

There are 25 speakers, 12 verbs, and 150 tokens in the African variability corpus. The verbal elements *avoir l'impression*, *heureusement*, *pouvoir dire*, *voir*, *vouloir*, and *vouloir dire* can not be tested for variability because of the low number of tokens per individual speaker. The

number of tokens and absent *que* rates for these elements varied considerably in the African corpus. The individual absent *que* rates for all verbal elements containing two or more tokens in each region are shown in Appendix D.

The overall variability patterns for each region based on verbal element and speaker are summarized in Table 5.4. More than half of the speakers in each region show variability in at least one verb, but less than 5% show variability for all verbs. Most speakers in each region show variability for some verbs and always use the present complementizer for others. In all regions it is extremely rare or unheard of for a speaker to only use absent *que* for all verbs or never to demonstrate variable use by using exclusively the present complementizer for some verbs and the absent variant for the others.

Table 5.4. Overall *que* variability based on verbal element and individual speaker

Absent *que* use	Europe		Quebec		Africa	
Number of speakers showing *que* variation in at least one verb	**27**	**62.8%**	**44**	**97.8%**	**18**	**72%**
Variation in all verbs (V)	2*	4.7%	2	4.4%	1*	4.0%
Variation in some verbs Only absent *que* in some verbs (VA)	4	9.3%	1	2.2%	1	4.0%
Variation in some verbs Only present *que* in some verbs (VP)	16	37.2%	23	51.1%	10	40%
Variation in some verbs Only absent *que* in some verbs Only present *que* in some verbs (VAP)	5	11.6%	17	37.8%	0	0%
Number of speakers never showing *que* variation	**16**	**37.2%**	**1**	**2.2%**	**7**	**28%**
Only absent *que* in some verbs Only present *que* in some verbs (AP)	0	0%	1	2.2%	0	0%
Absent *que* with all verbs (A)	1*	2.3%	0	0%	0	0%
Present *que* with all verbs (P)	15	34.9%	1*	2.2%	7	28%
Total number of speakers	**43**		**45**		**25**	

V=variable, A=absent *que*, P=present *que*, *=based on one verb

While these three regions have some commonalities regarding these verbal elements, they show more differences than similarities. More speakers in Quebec show variability by individual verbal element than in the other two regions. Roughly 98% of Quebec speakers in the variability corpus show some type of variability by verbal element, which is at least 25% more than in both other regions. There are more than three times as many speakers in Quebec who show variability for some verbs, the present complementizer for some, and the absent variant for others. For speakers who show no variation for any verbs, Europe and Africa have similar percentages, while Quebec has a considerably lower percentage with only one speaker in this category. A very similar pattern is found among speakers who only use the present complementizer. There are more speakers in Europe who never use absent *que* followed by Africa and then Quebec. Since Quebec has the highest overall absent *que* rate, it is not surprising that there are fewer speakers from this region who never use absent *que* for a given verbal element. Interestingly, Europe has a higher overall rate of absent *que* than Africa but more individual speakers who never use it. Thus, fewer European speakers use absent *que*, but those who do, use it at a relatively high rate, whereas more African speakers use absent *que* but do so at a relatively low rate.

The variability categorization of each verbal element based on individual speakers in each region is displayed in Table 5.5. Verbs in bold represent those that have forms that can be discourse markers. For Europe, the individual speaker accounts for extremely little variation in the data for the NDM forms but accounts for a considerable amount of variation with the PDM forms, as shown in the linear models presented in Section 3.2. *Savoir* and *penser* show the most variability for this region since they display all possibilities: variable, absent *que*, and present *que* (VAP). Overall, *penser* has a low absent *que* rate (14.9%), while *savoir* has a high rate (52.1%). The same trend holds for the PDM forms where *penser* has a rate of 15.6% and *savoir* has a rate

of 74.4%. Thus, the variation shown in Table 5.5 for these two verbs is not correlated with the absent *que* rate. Most other verbal elements are either never used with absent *que* or used with it to varying degrees. *Sembler* is the only verb that is not used categorically, although there is only one speaker with two or more tokens of this verb.

Table 5.5. Variability of *que* according to individual verbal elements by region

Verbal Element	Europe	Quebec	Africa
avoir l'impression	VP	P*	- -
c'est	P*	P	V*P
c'est-à-dire	- -	VA	VP*
c'est vrai	VP	VAP	V*
croire	VP	P	V
dire	VP	VP	VAP
être + adj.	P*	VP	- -
falloir	AP	VP	P
paraître	- -	P	V*P
penser	VAP	VAP	V*P
peut-être	VP	VP	V*P
pouvoir dire	P*	P*	- -
savoir	VAP	VA	V
se dire	P*	VP	P
sembler	V*	VAP	- -
trouver	VP	VP	P*
voir	VA	VP	- -
vouloir	P*	VP	- -
vouloir dire	VA	VAP	- -
Total # of types	5P 6VP 1V 1AP 2VAP 2VA	5P 8VP 0V 0AP 4VAP 2VA	3P 5VP 3V 0AP 1VAP 0VA

V=variable, A=absent *que*, P=present *que*, *=based on one speaker, verbs in bold can be discourse markers

In Quebec, *c'est vrai*, *penser*, *sembler*, and *vouloir dire* show the most variation since they are classified as VAP. Of these four verbs, *vouloir dire* has the highest overall absent *que*

rate. Most verbs have variable use or are used with the present complementizer. Only two verbs, *c'est-à-dire* and *savoir*, are categorically used with absent *que* by some speakers and variably used with it for other speakers. Considering the verbs that can be discourse markers, *savoir* is the only one that is VA. In Chapter 4, it was shown that *savoir* has a very high absent *que* rate, and most forms of this verb included in the data are PDMs. *Penser* and *vouloir dire*, as previously stated, have all three classifications. There is no speaker who exclusively uses absent *que* with *dire*, *trouver*, *voir*, or *croire*. *Croire* is never used with absent *que*.

The final region, Africa, shows that *dire* exhibits the most variation since it belongs to all three groups. Its overall absent *que* rate is 18.6% with a 100% rate for the PDM forms. Most verbal elements are never used with absent *que* or are used with it at varying rates. Regarding the PDMs, *croire* and *savoir* are used with absent *que* by each speaker, although not categorically. *Dire* belongs to all three groups, while *trouver* is never used with absent *que*. *Penser* is either not used with absent *que* or used with it at varying rates.

There are a few similarities between and among regions with regard to verb classification. First, the verbal element *peut-être* has the same classification in all three regions. Second, European and Quebec speakers share the same classification for *penser*, *dire*, *trouver*, *c'est*, and *pouvoir dire*. Third, European and African speakers never use *se dire* with absent *que*. While all regions have at least one verb with all three possible categorizations, Quebec has four such verbs compared to Europe, which has two, and Africa, which has one.

As outlined in Section 1.7.3, at least one form of the following verbs can be used as a discourse marker: *penser*, *savoir*, *voir*, *vouloir dire*, *croire*, *dire*, and *trouver* (Vincent 1991:47; Martineau 1993:84; Andersen 1996:307-309). While measures were taken to reduce the number of discourse markers that were included in the data for this study, it is likely that some were still

inadvertently included. In this case, the discourse marker forms would have a high rate of absent *que* since the complementizer construction is not used with discourse markers. In Section 4.2.1.1 it was shown that for European speakers, many of the PDMs have high absent *que* rates, and all of them have higher rates than the NDM forms. *Vouloir dire* and *savoir* have the largest difference in absent *que* rates based on whether we consider the NDM forms or just the PDM forms as indicated in Section 4.2.1.1. Two of the PDM verbs, *voir* and *vouloir dire*, are sometimes or always used with absent *que* for every speaker in Europe. Additionally, it was indicated in Section 4.2.1.1 that these two verbs have the highest overall absent *que* rates and also the highest rates for the PDM forms. The other PDMs are used exclusively with the present complementizer by some speakers. This may indicate that *voir* and *vouloir dire* have more instances of discourse markers included in the data or this could be a case of syntactic analogy as discussed in Section 4.2.2.4. Most likely, it is a combination of the two.

Regarding Quebec, Section 4.2.1.2 showed that *vouloir dire* and *savoir* have absent *que* rates much higher than the other PDM verbs. Of the PDM verbs, *vouloir dire* and *penser* vary the most based on the individual speaker since they have all three variability categorizations in Table 5.5. *Vouloir dire* has the highest overall absent *que* rate of the four verbal elements that have this categorization. *Savoir* is the second most frequent verb of all 22 verbs in the Quebec corpus and the only potential discourse marker verb that is categorized as VA. These two verbs show the largest difference in absent *que* rates based on which type of verb is chosen.

The African data presented in Section 4.2.1.3 indicated that the verbs showing the largest difference in absent *que* rates based on which verb forms are considered are *trouver*, *dire*, and *croire*. Classification based on individual speaker variation indicates that *dire* varies the most. The two PDM verbs with which all individual speakers vary their use of the complementizer are

croire and *savoir*. Both of these verbs have one of the five highest overall absent *que* rates, and both have higher rates than the other PDM verbs.

In all regions, *savoir* is notable for at least one reason. It has the second largest difference in absent *que* rates based on which verb type is considered for the European data. In Quebec and Africa, it has the second highest absent *que* rate of the PDM forms. In Quebec it is the second most frequent verb and the only PDM to be classified as VA. *Vouloir dire* is notable in Europe and Quebec but not Africa. It has the second highest overall absent *que* rate and second highest of the PDM verbs in Europe. In Quebec it has the second highest rate of the PDM verbs. For both regions, this verb shows the largest difference in absent *que* rates based on whether we consider the NDM forms or just the PDM forms. However, in Africa, *vouloir dire* has a low overall absent *que* rate and no instances of this verb for the PDM forms.

While there is some similarity in verbal elements across regions or for two regions, most show different patterns by region. The lack of homogeneity indicates that most of these verbs are treated differently in these three regions regarding absent *que*. As previously stated, Quebec shows the most variability based on verbal element classification. This region's long history of absent *que* use may have over time removed any stigma attached to it, prompting more varied use. Africa and Europe do not have the same documented history of its use, and it may be used in fewer instances or at a lower rate in each region due to a stigma or a lack of time for it to be extended to and accepted by more speakers. While the present study does not include the components necessary to evaluate the potential role of a social stigma, it is a possibility worth including in future research. The social factors included in the present study are not statistically significant, but there are regional differences in the use of absent *que* that cannot fully be

explained by the linguistic factors shown to be significant in this study. Differences in absent *que* use by matrix verb is one such factor.

5.3.2. *Que* Variation by Matrix Verb and Individual Speaker

Some speakers in each region display variation in *que* realization for the same matrix verb. In Europe there are 23 speakers who have variable *que* realization for at least one matrix verb. An analysis of these instances indicates that most involve a potential discourse marker. In example (74) *tu sais* appears to be used as a discourse marker rather than asserting that someone knows something. In example (75) it is possible that *je pense* ends a sentence and *c'est* begins the next sentence. In this case, *je pense* would be a discourse marker. For the speaker who shows variable use for *vouloir dire*, all instances of absent *que* are used with the first person singular, whereas the present complementizer is used only with the third person singular. Since the first person singular can be a discourse marker, it is likely that the speaker is actually using it as a discourse marker instead of a complementizer construction. One example from this speaker is shown in (76). Other PDMs with individual speaker variation are *croire*, *trouver*, and *voir*.

(74) mais tu sais (Ø?) ils prennent euh ils sélectionnent des gens sur la télévision française (EF46)
'but you know (Ø?) they take uh they select people on French television'

(75) ouais je vous dis nonante pourcent je pense (Ø?) c'est c'est jamais (EM44)
'yeah I tell you ninety percent I think (Ø?) it's it's never'

(76) j'ai eu beaucoup de tu vois beaucoup d'aide quoi je veux dire (Ø?) il m'a rien j'ai pas j'ai très peu de carence quoi (EM26)

‘I had a lot of you see a lot of help that’s all I mean (Ø?) he didn’t [incomplete verb] anything to me I don’t have I have very little lack in short’

The matrix verbal elements *c’est vrai* and *sembler* are often used variably among European speakers. An analysis of these tokens suggests that when there is no present complementizer, the verbal element ends an utterance and therefore is not part of a complementizer construction. Examples (77) and (78) below illustrate how each of these verbal elements are used in this way.

(77) non mais c’est c’est vrai (Ø ?) c’est un peu un choix qu’ils ont (EM26)
‘no but it’s it’s true (Ø?) it’s kind of a choice that they have’

(78) euh y a deux ans je crois bien oui y a deux ans il me semble (Ø ?) ils ont encore livré y a deux ou y a trois ans (EF69)
‘uh two years ago I think yes two years ago it seems to me (Ø?) they delivered [it] again two or three years ago’

The aforementioned examples represent the majority of tokens of variation by verbal element. For those that are not used as stated above, some indicate a possible motivation. One speaker, EM26, uses present *que* when *je pense* ‘I think’ is followed by a vowel (n=2) as in *je pense qu’on* ‘I think that we’ but absent *que* when *je pense* is followed by *ça* ‘this/it’ or *c’est* ‘it’s’ (n=2). It is plausible that this is motivated by articulatory economy, which Pohl (1968) posited as a conditioning factor in *ne* deletion. Since *pense* ends in /s/ and both *ça* and *c’est* begin with /s/, using the absent form makes [k(ə)] and an additional [s] unnecessary. There is therefore less articulatory movement required even compared to other instances where *que* is absent. The two speakers who vary complementizer use for *peut-être* use the present

complementizer only when it is followed by a vowel. All other tokens are followed by a fricative. This is the case for all seven tokens containing *peut-être*.

For a few speakers there is only one token with absent *que* and one with present *que* for a given verb. In some of these cases there is a possible motivation for the variation, but this is just speculation with only two examples. For example, in (79), *j'ai l'impression* could end an utterance in which case there would be no complementizer construction. After looking at example (80) where the same speaker uses present *que*, another possible explanation is that the speaker uses absent *que* when it is followed by a fricative but present *que* when it is followed by a vowel. Thus, there could be a phonological explanation.

(79) oui j'ai l'impression (∅?) c'était le jour où on a pris les cours (EF42)

'yes I have the impression (∅?) it was the day when we took classes'

(80) elle a pas du tout envie de/ d'aller à la fac parce qu'elle a l'impression **qu'**elle va y perdre son temps (EF42)

'she really doesn't want to go to university because she has the impression **that** she is going to waste her time there'

Finally, there are a few instances where a motivation for variation is not visible. In examples (81) and (82) below, *je pense* appears in the same syntactic and phonological context, and there is no intervening material in the matrix or embedded clauses. There is a possibility that *je pense* acts as a discourse marker in (81), but the presence of *mais* before it and the repetition of *je pense* makes this less likely.

(81) Ouais normalement c'est o/ c'est plutôt le contraire mais <u>je pense</u> euh <u>mais je pense</u> ∅ c'est pas mal une question de prof (EF27b)

‘yeah normally it’s o/ it’s rather the contrary but I think uh but I think ∅ it’s not bad, a question for the teacher’

(82) et puis je pense **que** c'est s/ grâce à ce boulot-là que j'a/ j'ai pu aller ben où je suis maintenant (EF27b)

‘and then I think **that** it’s s/ thanks to this job here that I/ I was able to go er where I am now’

In Quebec there are 43 speakers who have variable *que* realization for at least one matrix verb. The majority of these are the PDMs *penser*, *savoir*, *trouver*, *voir*, and *vouloir dire*. As is the case with Europe, most of the tokens with absent *que* are probably discourse markers. There are, however, a few exceptions. For example, in (83) the speaker uses *que* with *je pense* but does not in (84). In the dialogue, example (84) immediately follows (83). In both examples the complementizer is followed by the same subject, verb, and verb tense. It is very unlikely that *je pense* is acting as a discourse marker in (84) based on the context and structure of the utterances. The verbal elements *être* + adjective, *c’est vrai*, and *vouloir* could end the utterances in which they are used with absent *que*. Example (85) below illustrates this potential use. The first English translation represents use of absent *que*. The second indicates how the sentence would be understood if *voulaient pas* ended an utterance. The period is added to show this.

(83) moi je pense que ça va être un beau projet (QF48b)

‘me I think that this is going to be a nice project’

(84) je pense ∅ ça va être un espace vert (QF48b)

‘I think ∅ this is going to be a green space’

(85) parce qu'ils voulaient pas (∅?) ça leur coûte cher t'sais (QF42)

‘because they didn’t want it to cost them a lot you know’ (absent complementizer)

‘because they didn’t want [to]. it costs them a lot you know’ (end of utterance)

A motivation for the variation of the verbal elements *sembler*, *c’est-à-dire, peut*-être, and *falloir* is less clear. For *sembler*, two out of the six speakers who show variation use absent *que* when a fricative follows and use present *que* when a vowel or liquid follows. There could be a phonological explanation for this since it has been shown that less sonorous sounds favor absent *que* (Cedergren and Sankoff 1974; Sankoff 1980; Martineau 1985; Blondeau & Nagy 2008; Liang et al. 2021). This is possible based on the absent *que* rates for NDM forms (to which the verbal elements in question belong) but unlikely since liquids have a slightly higher absent *que* rate than fricatives. Further doubt about a phonological motivation arises due to the fact that fewer than half of the Quebec speakers show this pattern.

There is only one speaker who varies the complementizer form for *c’est-à-dire*, and there are just two tokens. The only difference between the two is that the one with absent *que* is followed by a stop, while the other is followed by a fricative, which does not lend itself to a phonological motivation based on the sonority hierarchy since fricatives and stops have similar sonority ratings. The case of *peut-être* reveals an interesting pattern for some speakers. For five out of the eight speakers who variably use *peut-être* and also display variation in the following phonological context, absent *que* is used when a fricative follows, and present *que* is used when a vowel or liquid follows. Example (86) shows *peut-être* with a following fricative and (87) shows a following vowel. This could provide evidence of a phonological motivation for the variation attested, especially since vowels have a lower absent *que* rate than fricatives for NDM forms in Quebec.

(86) peut-être Ø c'est seulement comme euh le yogourt ou bien le sirop d'érable (QM24b)

'maybe it's only like uh yogurt or rather maple syrup'

(87) ben en tout cas peut-être **qu'**ils faisaient une activité bricolage (QM24b)

'well in any case maybe they were doing a DIY project'

While very frequent and variable among Quebec speakers in this corpus, *falloir* is the most perplexing. Examples such as (88) and (89) are very typical. There are 17 Quebec speakers who vary *que* for this verb, and yet, there is no apparent motivation for the variation.

(88) faut **que** tu cherches des euh: de l'inspiration (QF47)

'you need to look for some uh some inspiration'

(89) faut Ø tu laves ça (QF47)

'you need to wash this'

Among the African speakers there are 12 who show *que* variation for the same verbal element. As with the European and Quebec speakers, most instances of variation occur with PDMs that are probably acting as discourse markers. Among these verbs, *croire*, *penser*, *savoir*, and *voir* are variable for at least one speaker. One speaker who exhibits variability for *savoir* could do so for more than one reason. Comparing examples (90) and (91) below, the only two tokens of *savoir* by this speaker, we see that the difference could be phonological since absent *que* is used when the following segment begins with a stop but present *que* is used when the following segment begins with a vowel. Intervening material in the matrix clause could also play a role. As stated earlier, intervening material can diminish comprehension of an utterance (Bolinger 1972; Martineau 1985; Warren 1994). In Africa absent *que* is never used with a PDM

when there is intervening material present in the matrix clause. The present complementizer is used in (91) when there is intervening material, probably to aid in comprehension.

(90) Dans la cour tout le monde te voit on sait ∅ toi en tout cas tu as parlé f/ euh jula en classe (AF28a)

'in the courtyard everyone sees you we know ∅ you in any case you spoke f/ uh Dyula in class'

(91) Mais nous on a fait toute l'école jusqu'à finir on savait même pas **qu'**on donnait des livres là-bas aux gens à lire (AF28a)

but we we completed school we didn't even know **that** they gave books there to people to read'

The remaining tokens, which include the verbal expressions *c'est-à-dire*, *c'est*, *peut-être*, and *paraître*, do not indicate a motivation for the variation. For example, *peut-être* is followed by absent *que* in (92) whereas it is followed by present *que* in (93). The structure of the two sentences is the same, the syntactic and phonological contexts are the same, and there is no intervening material. Even the topic is the same.

(92) Peut-être ∅ il y a moins de naissances malgré le couvre-feu (AF44a)

'maybe there are fewer births despite the curfew

(93) Peut-être **qu'**il y a moins de naissances je sais pas. (AF44a)

'maybe there are fewer births I don't know'

For all three regions, the majority of matrix verb variation by speakers can be attributed to PDMs, many of which are likely being used as discourse markers. It is also possible that some of them result from syntactic analogy. While neither of these can be confirmed based on

transcriptions alone, they do present a plausible explanation for the results of the qualitative analysis.

It was mentioned above that most Quebec speakers who exhibit *que* variation with *peut-être* and also with the phonological context following the complementizer use the absent form when followed by a fricative and the present form when followed by a vowel or liquid. Both of the European speakers who meet these criteria follow the pattern of the Quebec speakers, but the one African speaker does not. With all regions combined, there are seven out of eleven speakers who display this pattern.

For Quebec speakers it was also noted that many of them show variability in *que* realization with *falloir*. However, speakers generally use the present and absent forms in the same contexts. There are no speakers in Europe or Africa who vary *que* use for this verb. Thus, it seems that there is a distinct use of this verb in Quebec.

After analyzing each region, one observation is that some speakers tend to use absent *que* exclusively when *je pense* is followed by *c'est* or *ça*. A review of the data reveals that two out of five European and seven out of nineteen Quebec speakers do this. However, the one African speaker who varies the complementizer for *je pense* uses present *que* with *c'est*. The speakers who do not categorically use absent *que* in this construction either exclusively use the present form or use both the present and absent forms. For those who use both, there are no indications as to what conditions the form. For the speakers who solely use the absent form, it could be that they do so for articulatory economy. Cedergren and Sankoff (1974) and Sankoff (1980) found that a preceding and following sibilant was the phonological environment that most favored absent *que*. *C'est* and *ça* are both frequent collocations that would be easily recognizable to other speakers. Therefore, using the absent form should not diminish comprehension.

5.4. Individual Variation by Phonological Context Following *Que*

The analysis presented in Section 4.2 indicated that the phonological context is statistically significant for at least one data set for each region. The sibilant/non-sibilant/vowel sonority hierarchy and the obstruent/sonorant/vowel hierarchy are both followed for the full data set in each region. It was decided in Section 4.2.2.1 that the obstruent/sonorant/vowel hierarchy is the most appropriate for the present study. Still, the five-group sonority hierarchy that was initially chosen for this study is only followed by African speakers. The analysis that follows examines individual speaker variation for each of the five phonological groups to provide more insight into the phonological context following the complementizer.

Before analyzing individual variation let us review the results of the GLMMs and the absent *que* rates for the phonological context in each region. In Europe nasals significantly decrease the chance of absent *que* compared to fricatives for the PDM forms. For the full data set, stops have the lowest absent *que* rate when they would be expected to have the highest. In Quebec liquids and nasals significantly decrease the chance of absent *que* compared to fricatives for the full data set. For the PDM forms, vowels and stops significantly decrease the chance of absent *que* compared to fricatives. For the full data set, stops and liquids diverge from the five-group sonority hierarchy. For African speakers, liquids significantly decrease the chance of absent *que* compared to fricatives for the full data set and the NDM forms. The full data set follows the five-group sonority hierarchy.

The variability corpus for the phonological context following the complementizer is comprised of all speakers who have at least one phonological group (stops, fricatives, nasals, liquids, or vowels) with two or more tokens. For example, speaker EF26 has two tokens of a vowel following *que*, so these are included in the variability corpus. If only one token of a

certain phonological group exists for a given speaker, it is not included in the corpus since variability cannot be determined. The European variability corpus for the phonological context consists of 46 speakers and 714 tokens. In Quebec there are also 46 speakers included in the corpus and 1,494 tokens. The African corpus comprises 25 speakers and 244 tokens. Absent *que* rates for each speaker and phonological group are shown in Appendix E.

Each of the five phonological groups is shown in Table 5.6 along with the number and percentage of speakers from each region who use the complementizer according to the following variability classifications: variable rate (V), absent *que* only (A), or present *que* only (P). In Europe only one speaker (EM26) categorically uses absent *que* for any group, which is stops. The most variability exists for fricatives followed by vowels, stops, liquids, and then nasals. Among PDM forms, it was shown in the linear model in Section 3.2.2 that nasals differ significantly from fricatives. The same pattern is shown below when considering all verbal elements in Europe. Fricatives and nasals show the largest difference in the three variability groups, and they have opposite patterns from each other. For all speakers combined, there are 43 instances of a phonological group with categorical use (A or P) and 68 instances with variable use (V). Thus, there is more variability than categorical use.

Table 5.6. *Que* variation by phonological group and region based on individual speakers

	Europe		Quebec		Africa	
	# of Speakers	% of Speakers	# of Speakers	% of Speakers	# of Speakers	% of Speakers
Stops						
V	9	52.9	25	71.4	5	62.5
A	1	5.9	2	5.7	0	0
P	7	41.2	8	22.9	3	37.5
Fricatives						
V	27	69.2	42	93.3	13	68.4
A	0	0	1	2.2	1	5.3

P	12	30.8	2	4.4	5	26.3
Nasals						
V	2	33.3	2	40.0	1	50.0
A	0	0	0	0	0	0
P	4	66.7	3	60.0	1	50.0
Liquids						
V	4	44.4	12	54.5	4	36.4
A	0	0	1	4.6	0	0
P	5	55.6	9	40.9	7	63.6
Vowels						
V	26	65.0	38	84.4	11	55.0
A	0	0	0	0	0	0
P	14	35.0	7	15.6	9	45.0
Total number of speakers	43		45		25	

V=variable, A=absent *que*, P=present *que*

There are four Quebec speakers who categorically use absent *que* for a phonological group, and these groups include stops, fricatives, and liquids. The largest amount of variability is seen with fricatives followed by vowels and then stops. Absent *que* is used the least with nasals and liquids. The linear model of the full data set shown in Section 3.2.3 indicated that liquids and nasals differ significantly from fricatives and reduce the chance of absent *que*. The variability rates for these groups indicate that fricatives have an extremely high percentage of speakers who use absent *que* at a variable rate, while nasals and liquids have a much higher percentage of speakers who never use absent *que*. For all speakers combined, there are 33 instances of a phonological group with categorical use and 119 instances with variable use. Therefore, there is much more variability than categorical use of present or absent *que* in Quebec.

In Africa there is only one speaker who categorically uses absent *que* for a phonological group, which is fricatives. The largest percentage of variable use is for fricatives followed by stops, then vowels, nasals, and liquids. The mixed-methods analysis for the African data indicated that liquids differ significantly from fricatives. Evidence of this can be seen in the

patterns for these two groups in Table 5.6. Fricatives have a high rate of variability, while liquids are most often used with present *que*. For all speakers combined, there are 26 instances of a phonological group with categorical use and 34 instances with variable use. Thus, there is slightly more variability than categorical use of absent or present *que*.

An analysis of individual speaker patterns for each phonological group reveals many similarities between European and Quebec speakers. In both regions, nasals have the largest percentage of categorical present *que* use and the least amount of variability. In Africa, liquids occupy this position, with nasals next in line. The phonological group with the greatest percentage of categorical absent *que* use is stops in Europe and Quebec but fricatives in Africa. Fricatives have the most variability in all three regions. Otherwise, Europe and Quebec pattern the same, while Africa diverges.

5.5. Conclusion

A look at the individual absent *que* rates of each speaker reveals that Europe has the largest number of speakers who never use absent *que* followed by Africa and then Quebec. Even though as a region Quebec has the highest absent *que* rate of the three, Europe and Africa both have individual speakers with higher rates than any of the Quebec speakers. In Europe and Quebec, the younger generation varies the most in their use of absent *que*, whereas the older generation shows more stability. Of the two age groups represented by African speakers, neither differs in its use of absent *que* based on this factor.

The matrix verb was shown to be statistically significant in at least one linear model for Europe and Quebec. While it was not included as a fixed effect for the African models, its influence on absent *que* can nevertheless be seen as indicated in Section 4.2. An examination of

individual absent *que* rates for each matrix verbal element yielded some insights. Most speakers in all regions do not use absent *que* with at least one verbal element. The majority of speakers in Quebec show variability by individual verbal element, and there is more variability based on verbal element classification in Quebec than in Europe or Africa. Europe has fewer speakers than Africa who use absent *que*, but the ones who do use it display a relatively high rate, whereas more African speakers use it but have a low rate of use. Some verbal elements share the same classification in two regions, but *peut-être* is the only one with the same classification in all three regions. It is variable for some speakers and only used with present *que* for others. *Savoir* is important in all three regions for having the second largest difference in absent *que* rates based on which verb type is considered. Similarly, *vouloir dire* is important in Europe and Quebec since it shows the largest difference in absent *que* rates based on which verb type is analyzed.

An examination of speakers who vary *que* realization for the same verb suggests that many tokens represent PDMs that speakers may be using as discourse markers. There are also many examples where the verbal expression could end an utterance and, therefore, would not be part of a complementizer construction. Among the remaining verbal elements are those that indicate a possible motivation, at least for some speakers, and those that provide no indication of what might condition the variation. Out of the 11 speakers in all three regions who have variable *que* use for *peut-être* and more than one following phonological context, more than half use absent *que* before a fricative and the present form before a liquid or vowel. *Je pense* + *c'est*/*ça* may be used with absent *que* for some speakers due to articulatory economy and the saliency of *c'est* and *ça*. This appears to be the case for 37.5% of the speakers. The most frequent verbal element without an apparent motivation for variation is *falloir* as it is used in Quebec.

As suggested in Section 4.2.2.4, syntactic analogy may play a role in *que* realization with regard to PDM forms. The examples given in this chapter illustrating PDMs with an absent *que* highlight the fact that we do not know for sure if they were being used as a discourse marker. An acoustic analysis that uses prosody to make such determinations is needed, such as the one proposed by Avanzi (2012). It is likely that some examples are indeed discourse markers while others are matrix verbs heading a complementizer clause but using absent *que* based on syntactic analogy. The percentage of each would yield great insight into *que* realization.

Aside from the matrix verb, the phonological context following the complementizer is the only factor to be statistically significant in at least two regions. In fact, this factor is significant in all three regions and for more than one model in Quebec and Africa. An analysis of overall patterns based on individual use of absent *que* for the five phonological groups indicates that more than half of the speakers in each region never use absent *que* for at least one group. In Europe and Quebec, this group is nasals, while it is liquids in Africa. At the other end of the spectrum, stops have the greatest percentage of categorical absent *que* use in Europe and Quebec, while fricatives occupy this position in Africa. One commonality is that fricatives have the highest variable rate in all three regions. Overall, Quebec speakers indicate more variation in use of absent *que* based on phonological group while Europe and Africa pattern more similarly with each other than with Quebec.

The analysis of individual speaker variation suggests that language change may be taking place in Quebec and Europe. In Quebec, almost all speakers use absent *que* to some extent, and individual speakers vary their use of *que* according to matrix verb more than speakers from Europe and Africa. Absent *que* is used in more contexts by more people in this region, suggesting that it is extending its scope. Europe is marked by some speakers with very high

absent *que* rates, and the most speakers of any region who never use absent *que*. Thus, it appears that some European speakers are leading a move toward absent *que*. Since Africa is characterized by more speakers who use absent *que* at a low rate compared to Europe, the absent variant seems to be more stable in this region.

Chapter 6

Conclusion

6.1. Introduction

The absent complementizer in Modern French has been the subject of studies dating back to the latter half of the 20th century with at least one mention of the phenomenon in the early part of the century (Bauche 1928). This construction has been analyzed from a theoretical and quantitative standpoint. It has been established that the absent complementizer exists in Gascon (Pusch 2003), Spanish (Wanner 1981; Torrego 1983; Brovetto 2002; Silva-Corvalán 2008; Pountain 2015), and Italian (Scoretti 1981; Wanner 1981; Brovetto 2002), all of which are historically related to French. Additionally, deletion of specific phonemes or words, notably schwa and *ne*, is widespread in the francophone world. The absent complementizer can be seen as a related phenomenon that belongs to the same system. The present study has situated the absent complementizer in its historical context and has examined its occurrence in informal Modern Spoken French in three francophone regions. This concluding chapter returns to the three research questions that guided this study and provides a summary of the findings in Section 6.2. Limitations of this study and future research are discussed in Section 6.3.

6.2. Research Questions

1. How widespread is use of absent *que* in Europe, Quebec, and Africa?

While there are documented uses of absent *que* in Quebec, Europe, and Africa, there is no quantitative study in the last twenty years that indicates the prevalence of absent *que* in Quebec and Europe apart from the present study. Boutin (2007) showed that in the early 2000s there was an absent *que* rate of 8.8% in Côte d'Ivoire. Using transcripts of informal spoken language from the Phonology of Contemporary French project (PFC) and the Corpus of French spoken in Quebec (CFPQ), this study provided results for a total of 119 speakers in Europe, Quebec, and Africa. The overall absent *que* rates were 30.8% in Quebec, 25.3% in Europe, and 20.8% in Africa. An analysis of individual speaker variation revealed that most speakers in Quebec show variability for at least one verbal element, and there is more variability based on the classification of verbal elements than in Europe and Africa. Between the other regions, more African speakers use absent *que* at a low rate while fewer European speakers use it, but those who do use it have a higher rate of use. This is suggestive of language change in Europe with a few speakers leading the change toward absent *que*. The results of this study confirm that absent *que* is found in each of the three regions and that it is most widespread in Quebec.

2. What are the factors that condition absent *que* and how do they compare to those found in previous studies (Cedergren & Sankoff 1974; Connors 1975; Sankoff 1980; Martineau 1985; Warren 1994; Blondeau & Nagy 2008; Liang et al. 2021)?

The generalized linear mixed-effects models for each region indicated that the social factors do not condition use of absent *que*. Four linguistic factors condition its use: matrix verbal element, phonological context, intervening material in the matrix clause, and syntactic context. The first

two are significant for at least one model from each region when included as a fixed effect while the latter two are significant for one model each.

The matrix verb is statistically significant in all models where it is a fixed effect. For the models where it could not be included as a fixed effect, it was a random effect. In these models, its effect on absent *que* can be seen in the different factor level outcomes based on whether one looks at NDM (non-discourse marker) forms or PDM (potential discourse marker) forms. In Europe and Quebec, matrix verb semantic groups were included as fixed effects. In Europe, judgment verbs differ from epistemic verbs and decrease the chance of absent *que*. In Quebec, judgment, desired action, and speaking/restatement verbs as well as adverbs differ from epistemic verbs and decrease the chance of absent *que*. Among the PDMs in Europe, *dire*, *savoir*, *vouloir dire*, and *voir* differ significantly from *croire* and increase the chance of absent *que*.

Martineau (1985) and Wiesmath (2002) concluded that the matrix verb strongly influences the absent *que* rate. It was suggested by Martineau (1985) that *croire* favors formal speech while *penser* favors informal speech and that this makes *croire* less favorable to absent *que*. In the informal speech analyzed for the present study, it was found that *penser* is used more frequently with absent *que* than *croire* in Europe and Quebec but not in Africa. Also, *croire* is used more frequently in Europe than it is in Quebec or Africa. Thus, it does not appear that African speakers associate *croire* with formal speech and *penser* with informal speech, although this appears to be the case in Europe and Quebec.

The verbs *croire*, *dire*, *penser*, *savoir*, *sembler*, and *trouver* were found to favor absent *que* in Ontario and Quebec (Martineau 1985). All of these verbs allow the absent complementizer in each region. It could not be determined whether *sembler* allows absent *que* in

Africa since there were no instances of this verb. Of these verbs, *savoir* has the highest absent *que* rate in each region. In a study of Parisian French, Andersen (1993) found that *c'est* + adjective had an absent *que* rate of 35%. The present study indicates a much lower rate of 4.8% in Europe and 14.1% in Quebec for *être* + adjective, which includes *c'est* + adjective. There were no instances of this construction in Africa. Andersen (1993) also found that the verbs *voir*, *dire*, *savoir*, *trouver*, *se souvenir*, and *se rappeler* had a combined rate of 30%. The first four verbs have a combined rate of approximately 30% in Quebec and Africa, and a much higher rate in Europe (45.8%). The latter two verbs were not included in the present study. Andersen (1993) also found that *croire*, *penser*, *falloir*, and *vouloir* did not have any instances of absent *que*. This is true for *falloir* and *vouloir* in Europe and Africa, but in Quebec, these four verbs are sometimes used with absent *que*.

Regarding the trends of matrix verbal elements in the three regions, *savoir* has one of the highest absent *que* rates in all three regions. The verbal element *peut-être* has the same variability classification based on individual speaker in all three regions. There are some similarities between two regions, such as the same variability classification for *penser*, *dire*, *trouver*, *c'est*, and *pouvoir dire* in Europe and Quebec. However, most verbs indicate different trends according to region. For example, of the verbs that are never used with absent *que* in one region, not one is found with a 0% rate in all three regions. The lack of homogeneity in the variability classification of these verbs indicates that they are treated differently according to region.

The phonological context following the complementizer is significant in five out of the six models where it is a fixed effect. In Europe, nasals differ significantly from fricatives for the PDM forms and reduce the chance of absent *que*. In Quebec, liquids, nasals, and vowels differ

significantly from fricatives and reduce the chance of absent *que*. For African speakers, only liquids differ significantly from fricatives and reduce the chance of absent *que*. An analysis of variability by individual speaker indicates that fricatives have the highest variability rate in all three regions. Cedergren and Sankoff (1974), Sankoff (1980), Martineau (1985), and Blondeau and Nagy (2008) all found sibilants to favor absent *que* the most of any following sound. In Europe and Quebec, fricatives have the highest overall rate, whereas they have the second highest rate in Africa. Thus, we see the same finding for Europe and Quebec as these three previous studies. The five-group sonority hierarchy used for the present study is followed only in Africa. After further consideration it was determined that the obstruent/sonorant/vowel hierarchy is the most appropriate for this study. All three regions follow this hierarchy in agreement with the findings of Martineau (1985) and Liang et al. (2021).

Comparing the phonological and syntactic contexts after the complementizer, Connors (1975) found the syntactic to be more important while Martineau (1985) found the phonological to be more important. Liang et al. (2021) found both to be independently significant. Based on statistical significance of individual factors, the phonological context is more important than the syntactic in each of the three regions considered for the present study.

Intervening material in the matrix clause is statistically significant for the PDMs in Quebec. Its presence between the matrix verbal element and the complementizer reduces the chance of absent *que*. This is in agreement with Martineau's (1985) claim that intervening material reduces comprehension and disfavors absent *que*. Warren (1994) found simple utterances (ones without intervening material) to favor absent *que*. While not shown to be statistically significant, it is interesting that in Europe and Quebec there is a higher absent *que* rate when intervening material is present in the matrix clause for the NDM forms. Almost all

examples of intervening material in this clause in both regions consisted of just one word. It appears that absent *que* can be used when there is intervening material in the matrix clause but that there are limitations on its length.

The final linguistic factor shown to be statistically significant is the syntactic context following the complementizer. In Africa, other words differ significantly from pronouns and decrease the chance of absent *que*. Overall, lexical nouns and pronouns have almost identical rates, which are higher than those for other words. It was determined that the statistical significance for this factor lies with the PDM forms and that there may be a phonological motivation underlying the syntactic forms. Connors (1975) and Martineau (1985) found pronouns to favor absent *que* compared to all other word types. If pronouns are compared to lexical nouns and other words together, the present study yields the same results as the two previous studies. Connors (1975) found forms of *ce* and *je* to be the most common pronouns following the complementizer. The present study indicates that *ce* is the most common in each region followed by *il* and then *je*.

The results of this study indicate that the matrix verb is the most influential factor in *que* realization. Not only is the matrix verb statistically significant in every model in which it is a fixed effect, but the type of matrix verb affects the outcome of almost every factor level in all three regions. The phonological context following the complementizer is the second most important factor. It is significant in the second greatest number of linear models and is also a possible underlying factor of absent *que* rates for the syntactic context. It may also be responsible for some variation that individual speakers show for the same matrix verb.

This study has made clear that there is a distinct difference in absent *que* rates for verbs that do not also have a discourse marker form and those that do. With the exception of Blondeau

and Nagy (2008), previous studies did not consider how the inclusion of some possible discourse markers might affect their results. However, Martineau (1985:65) suggested that the ability of *savoir* and *penser* to be used as discourse markers probably reinforced their favorableness to absent *que*. Blondeau and Nagy (2008) removed the most frequent forms that could be discourse markers, while the present study used criteria to reduce the number of these forms and analyzed the potential ones separately. While some examples in the PDM category are most likely discourse markers rather than matrix verbs, it is possible that the higher absent *que* rates for these forms is due, at least in part, to syntactic analogy. The dual function PDM forms occur frequently, and, when used as a discourse marker, are never followed by a complementizer. Thus, their frequent appearance without the complementizer when used as discourse markers probably makes it more likely for speakers to use their matrix verb forms without a complementizer. This may also decrease any stigma associated with absent *que* being used with these forms. In order to test this theory, it would be necessary to identify discourse markers definitively and remove them from the data. In some cases, this can only be done by an acoustic analysis that measures prosody like the one Avanzi (2012) conducted.

3. What has been the trajectory of the absent *que* in Quebec since the 20th century?

Studies in Quebec and Ontario in the 1980s found absent *que* rates of 23% (Sankoff 1980), 32% (Martineau 1985), and 24.8% (Liang et al. 2021). The present study indicates an absent *que* rate of 30.8% for Quebec speakers. Since this rate falls between those of previous studies, we can conclude that the overall rate of absent *que* use has not changed. However, individual speaker use suggests that absent *que* is extending its scope. When looking at age patterns, Martineau (1985) and Liang et al. (2021) both found the youngest speakers to have the highest absent *que* rate, and the oldest speakers to have the lowest. The present study indicates that, among Quebec

speakers, the youngest speakers have the lowest rate, and the oldest have the highest. Based on these data, there is no evidence of age grading or generational change.

6.3. Limitations and Future Research

The present study has provided background information that situates the absent complementizer in its linguistic context both historically and in the present. It has provided a tri-regional quantitative analysis of social and linguistic factors that could play a role in conditioning absent *que*. Also, it has provided a qualitative analysis of individual speaker variation. As with any study, there are limitations. This study was limited to the available transcriptions for informal spoken French and the metadata provided by the two selected corpora. There are many factors that have been shown to affect *que* realization in previous studies, and there are likely significant factors yet to be uncovered. One challenge with the statistical analysis of this study was the ratio of factors to sample size. Selected factors had to be limited in some models so that they would converge. Instead of testing every possible factor, this study presents an analysis of some of the social and linguistic factors previously shown to impact *que* realization.

The phonological context following the complementizer was shown to be statistically significant in this study as was the case in previous studies on variable *que* and on related phenomena such as schwa and *ne* deletion. However, a full analysis of this factor is needed. As a first step, three variants should be considered: [kə], [k], and ∅. The absence or presence of variable schwa should be considered since this has implications for syllable structure, stress, and consonant clusters. An analysis of the preceding and following phonological environments for each of these three variants would provide more complete results.

One potentially influential factor on *que* realization in the three regions considered is the sociocultural context in which French is used. In Africa, location, French competency level of the speaker, and the circumstances of French use affect variability (Manessy 1992). French is primarily spoken as an additional language in Africa, while it is the first language of most Quebec and European speakers in France and Western Switzerland. French is the primary language spoken in Quebec, France, and Western Switzerland, while it is not in many francophone African countries. The social status of French also varies widely by location. It is also possible that contact with other languages may account for some of the differences seen among these three regions. Future research on absent *que* would benefit from considering the sociocultural context and the linguistic makeup of each region. This would lend itself to exploring the idea of social stigma and any role that it may play in variable *que* use.

This work has been the first to compare *que* absence rates in three francophone regions. The statistical methods employed offer several advantages over methods used in all previous studies with the exception of Liang et al. (2021) who also used mixed-effects models in R (R Core Team 2022). It has been shown that absent *que* is used in all three regions at a rate over 20%, and that the rate in Quebec has not changed since the 1970s and 1980s. Social factors do not condition absent *que* in any of the regions. The lexical identity of the matrix verb is the primary factor followed by the phonological context following the complementizer. The findings of the present study shed light on similarities and differences in *que* realization in three francophone regions as well as trends in its use at the beginning of the 21st century. It also highlights the need for further research in key areas. More generally, it provides insight into language change, variation, and diffusion. Ultimately, this study brings us further along in our understanding of the intricacies and complexities of human communication.

References

Andersen, Hanne Leth. 1993. Les complétives non introduites en français parlé. In Claude Muller & Daniel Roulland (eds.), *Subordinations*, 5–14. Rennes, France: Presses universitaires de Rennes.

Andersen, Hanne Leth. 1996. Verbes parenthétiques comme marqueurs discursifs. In Claude Muller (ed.), *Dépendance et intégration syntaxique*, 307–315. Tübingen: Max Niemeyer Verlag.

Andersen, Hanne Leth. 1997. *Propositions parenthétiques et subordination en français parlé*. Copenhagen: University of Copenhagen Dissertation.

Andersen, Hanne Leth. 1999. Propositions adverbiales non introduites en français parlé. *Revue Romane* 34(2). 163–180.

Antoine, F. 1899. De la parataxe et de l'hypotaxe dans la langue latine. *Revue des études anciennes* 1(1). 27–46.

Apothéloz, Denis. 2003. La rection dite "faible": grammaticalisation ou différentiel de grammaticité? *Verbum* 25(3). 241–262.

Armstrong, Nigel. 2002. Variable deletion of French *ne*: a cross-stylistic perspective. *Language Sciences* 24. 153–173.

Armstrong, Nigel & Sharon Unsworth. 1999. Sociolinguistic variation in southern French schwa. *Linguistics* 37(1). 127–156.

Arteaga, Deborah. 2009. On the existence of null complementizers in Old French. In Pascual José Masullo, Erin O'Rourke & Chia-Hui Huang (eds.), *Romance Linguistics 2007: Selected Papers From the 37th Linguistic Symposium on Romance Languages (LSRL), Pittsburgh, 15-18 March 2007*, 19–36. Amsterdam: John Benjamins.

Ashby, William. 1981. The Loss of the Negative Particle *ne* in French: A Syntactic Change in Progress. *Language* 57(3). 674–687.

Ashby, William. 1991. When does variation indicate linguistic change in progress? *Journal of French Language Studies* 1(1). 1–19.

Ashby, William. 2001. Un nouveau regard sur la chute du ne en français parlé tourangeau: s'agit-il d'un changement en cours? *French Language Studies* 11. 1–22.

Avanzi, Mathieu. 2012. L'interface prosodie/syntaxe en français: Dislocations, incises et asyndètes (GRAMM-R. Études de Linguistique Française 13). Bruxelles: Peter Lang.

Avanzi, Mathieu, Anne Lacheret-Dujour & Bernard Victorri. 2008. ANALOR. A Tool for Semi-Automatic Annotation of French Prosodic Structure. 119-122. Campinas, Brazil. https://hal.science/hal-00334656.

Barnes, Betsy K. 1995. Discourse Particles in French Conversation: *(eh) ben, bon*, and *enfin*. *The French Review* 68(5). 813–821.

Bates, Douglas, Martin Mächler, Benjamin M. Bolker & Steven C Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Bates, Douglas, Katharine M. Mullen, John C. Nash & Ravi Varadhan. 2014. minqa: Derivative-Free Optimization Algorithms by Quadratic Approximation. *R package version 1.2.4*. https://CRAN.R-project.org/package=minqa.

Bauche, Henri. 1928. *Le Langage Populaire*. Paris: Payot.

Benveniste, Émile. 1966. *Problèmes de linguistique générale*. Vol. 1. 2 vols. Paris: Gallimard.

Blanche-Benveniste, Claire. 1989. Constructions verbales “en incise” et rection faible des verbes. *Recherches sur le français parlé* 9. 53–73.

Blevins, James P. & Juliette Blevins. 2009. Introduction: Analogy in grammar. In James P. Blevins & Juliette Blevins (eds.), *Analogy in Grammar: Form and Acquisition.* 1-12. Online. Oxford: Oxford Academic. (13 March, 2023).

Blondeau, Hélène & Naomi Nagy. 2008. Subordinate clause marking in Montreal Anglophone French and English. In Miriam Meyerhoff & Naomi Nagy (eds.), *Social Lives in Language - Sociolinguistics and multilingual speech communities* (IMPACT: Studies in Language and Society), vol. 24, 273–313. Amsterdam: John Benjamins.

Bolinger, Dwight. 1972. *That’s that*. The Hague: Mouton.

Boutin, Akissi Béatrice. 2007. *De* et *que* subordonnants, et variation en français. *Linx. Revue des linguistes de l’université Paris X Nanterre* 57. 57–67. https://doi.org/10.4000/linx.280.

Brovetto, Claudia. 2002. Spanish clauses without complementizer. In Teresa Satterfield, Christina Tortora & Diana Cresti (eds.), *Current Issues in Romance Languages: Selected Papers from the 29th Linguistic Symposium on Romance Languages (LSRL), Ann Arbor, 8-11 April 1999*, 33–46. Amsterdam: John Benjamins.

Cedergren, Henrietta J. & David Sankoff. 1974. Variable Rules: Performance as a Statistical Reflection of Competence. *Language: Journal of the Linguistic Society of America* 50(2). 333–355.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, 283–333. Cambridge: Cambridge University Press.

Connors, Kathleen. 1975. L'effacement du QUE - règle syntaxique. *Recherches Linguistiques à Montréal* 4. 17–33.

Côté, Marie-Hélène & Peter Milne. 2009. Les relations France-Québec: insertion de schwa vs réduction des groupes consonantiques. *Journées PFC-MSH*. Paris.

Coveney, Aidan. 1996. *Variability in Spoken French: A sociolinguistic study of interrogation and negation*. Exeter: Elm Bank Publications.

Deulofeu, José & Jean Véronis. 2002. L'utilité du recours au corpus pour rendre compte des différences entre les locuteurs du Sud et du Nord de la France dans l'emploi du morphème que en langue parlée. In Claus Pusch & Wolfgang Raible (eds.), *Romance Corpus Linguistics: Corpora and Spoken Language*, 377–392. Tübingen: Narr.

Diessel, Holger & Martin Hilpert. 2016. Frequency effects in grammar. *Oxford research encyclopedia of linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.120

Dodsworth, Robin. 2011. Social Class. In Ruth Wodak, Barbara Johnstone & Paul Kerswill (eds.), *The SAGE Handbook of Sociolinguistics*, 192–207. London: SAGE Publications.

Dostie, Gaétane. 2016. *Corpus du français parlé au Québec*. (CFPQ) https://applis.flsh.usherbrooke.ca/cfpq/index.php/site/index. [accessed April 2022 to August 2022].

Durand, Jacques, Catherine Slater & Hilary Wise. 1987. Observations on Schwa in Southern French. *Linguistics* 25(5). 983–1004.

Durand, Jacques, Bernard Laks & Chantal Lyche. 2002. La phonologie du français contemporain: usages, variétés et structure. In C Pusch & W Raible (eds.), *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics – Corpora and Spoken Language*, 93–106. Tübingen: Gunter Narr. http://www.projet-pfc.net. [accessed April 2022 to August 2022].

Eychenne, Julien. 2019. On the deletion of word-final schwa in Southern French. *Phonology* 36. 355–389.

Ferrand, Ludovic, Boris New, Marc Brysbaert, Emmanuel Keuleers, Patrick Bonin, Alain Méot, Maria Augustinova & Christophe Pallier. 2010. The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods* 42. 488–496.

Field, Thomas T. 1985. Speech act markers in modern Gascon. In Larry D. King & Catherine A. Maley (eds.), *Selected papers from the XIIIth Linguistic symposium on Romance languages, Chapel Hill NC, 24-26 March 1983*, 77–97. Amsterdam: John Benjamins.

Foulet, Lucien. 1961. *Petite syntaxe de l'ancien français* (Classiques Français Du Moyen Âge: 21). Paris: Champion.

Gachet, Frédéric. 2015. Syntactic hypotheses about so-called "*que*-deletion" in French. In Marlies Kluck, Dennis Ott & Mark de Vries (eds.), *Parenthesis and ellipsis: Cross-linguistic and theoretical perspectives* (Studies in Generative Grammar), vol. 121, 147–172. Berlin: Walter de Gruyter.

Gadet, Françoise. 2011. What can be learned about the grammar of French from corpora of French spoken outside of France. In M. Konopka, J. Kubczak, C. Mair, F. Sticha & U. Wassner (eds.), *Grammatik und Corpora 2009*, 87–120. Tubingen: Gunter Narr.

Glikman, Julie. 2009. *Parataxe et Subordination en Ancien Français: Système syntaxique, variantes et variation*. Potsdam, Germany: Universität Potsdam Dissertation.

Gougenheim, Georges. 1951. *Grammaire de la langue française du seizième siècle*. Lyon: IAC.

Green, John N. & Marie-Anne Hintze. 2021. L'élision en français: une catégorie qui n'est plus catégorique. *Studii de Lingvistica* 11. 123–146.

Grevisse, Maurice & André Goosse. 1993. *Le Bon Usage: Grammaire Française*. 13th edn. Paris: Duculot.

Guy, Gregory. 2011. Language, social class, and status. In Rajend Mesthrie (ed.), *The Cambridge Handbook of Sociolinguistics*, 159–185. Cambridge: Cambridge University Press.

Hansen, Anita Berit. 2000. Le E caduc interconsonantique en tant que variable sociolinguistique. *Revue des linguistes de l'université Paris X Nanterre* 42. 45–58.

Hutin, Mathilde, Adèle Jatteau, Yaru Wu, Ioana Vasilescu, Lori Lamel & Martine Adda-Decker. 2021. A corpus-based study of the distribution of word-final schwa in Standard French and what it teaches us about its phonological status. *Isogloss. Open Journal of Romance Linguistics* 7(21). 1–27.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* (Special Issue: Emerging Data Analysis) 59(4). 434–446. https://doi.org/10.1016/j.jml.2007.11.007.

Johnson, Daniel Ezra. 2009. Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass* 3(1). 359–383.

Johnston, Ron, Kelvyn Jones & David Manley. 2018. Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality and Quantity* 52(4). 1957–1976.

Josephs, Julia Steele & Diana L. Ranson. 2020. La liaison variable chez les locuteurs français septentrionaux et méridionaux. In F. Neveu, B. Harmegnies, L. Hriba, S. Prévost et A. Steuckardt, *Actes du CMLF 2020 – 7e Congrès Mondial de Linguistique Française,* 1-15. Paris: SHS Web of Conferences. https://doi.org/10.1051/shsconf/20207809012.

Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, William. 1994. *Principles of Linguistic Change*. Vol. 1. Cambridge, MA: Blackwell.

Ladefoged, Peter. 1975. *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Lafage, Suzanne. 2002. *Le lexique français de Côte-d'Ivoire. Appropriation et créativité*. Nice : Institut de Linguistique Française (ILF) et Centre National de la Recherche Scientifique (CNRS).

Larrivée, Pierre. 2011. Is there a Jespersen cycle? In Pierre Larrivée & Richard Ingham (eds.), *The Evolution of Negation: Beyond the Jespersen Cycle*, 1–22. Berlin: De Gruyter Mouton.

Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems* 19. 849–856.

Liang, Yiming, Pascal Amsili & Heather Burnett. 2021. New ways of analyzing complementizer drop in Montréal French: Exploration of cognitive factors. *Language Variation and Change* 33. 359–385.

Lodge, R. Anthony. 1993. *French, from Dialect to Standard*. London: Routledge.

Malécot, André. 1976. The Effect of Linguistic and Paralinguistic Variables on the Elision of the French Mute-e. *Phonetica* 33(2). 93–112.

Mallet, Géraldine-M. 2008. *La liaison en français : descriptions et analyses dans le corpus PFC*. Université Paris Ouest Nanterre La Défense Dissertation.

Manessy, Gabriel. 1992. Norme endogène et normes pédagogiques en Afrique Noire francophone. *Multilinguisme et développement dans l'espace francophone* 43–81.

Manessy, Gabriel. 1994. *Le français en Afrique noire: Mythe, stratégies, pratiques*. Paris: L'Harmattan.

Martineau, France. 1985. *Elision variable de (que) dans le parler d'Ottawa-Hull*. Ottawa, Canada: University of Ottawa Dissertation.

Martineau, France. 1993. Rection forte et rection faible des verbes : l'ellipse de *Que* en français du Québec et de l'Ontario. *Francophonies d'Amérique* (3). 79–90. https://doi.org/10.7202/1004445ar.

Moignet, Gérard. 1976. *Grammaire de l'ancien français : morphologie, syntaxe* (Problèmes et Méthodes: 2). Paris: Klincksieck.

Myers, Emily L. & Diana L. Ranson. 2014. L'élision variable des /R/ et /l/ postconsonantiques finals en français méridional et septentrional : L'effet de l'âge, du lieu d'origine et des facteurs linguistiques. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer & Sophie Prévost, *Actes du CMLF 2014 -- 4e Congrès*

Mondial de Linguistique Française, 1345-1364. Paris: Institut de la Linguistique Française & EDP sciences.

Oxford English Dictionary Online. 2022. Oxford University Press. http://www.oed.com/.

Parker, Steve. 2011. Sonority. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, vol. 2, 1160–1184. Malden, MA: Wiley-Blackwell.

Pohl, Jacques. 1968. *Ne* dans le français parlé contemporain: Les modalités de son abandon. In Antonio Quilis (ed.), Actas *XI Congreso Internacional de Lingüística y Filología Románicas*, 1343–1360. Madrid: Consejo Superior de Investigaciones Científicas.

Poplack, Shana. 1989. The care and handling of a mega-corpus: The Ottawa-Hull French project. In Ralph W. Fasold & Deborah Schiffrin (eds.), *Language Change and Variation*, 411–451. Amsterdam: John Benjamins.

Pountain, Christopher J. 2015. *Que*-deletion: the rise and fall of a syntactic fashion. In Francisco Dubert García, Gabriel Rei-Doval & Xulio Sousa (eds.), *En memoria de tanto miragre. Estudos dedicados ó profesor David Mackenzie*, 143–159. Santiago de Compostela: Universidade de Santiago de Compostela.

Pusch, Claus D. 2003. Verbes recteurs faibles en occitan. In Fernando Sánchez Miret (ed.), *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica*, vol. 2, 247–258. Tübingen: Max Niemeyer. https://doi.org/10.1515/9783110946253-022.

Pustka, Elissa. 2009. A prototype-theoretic model of Southern French. In Kate Beeching, Nigel Armstrong & Françoise Gadet (eds.), *Sociolinguistic Variation in Contemporary French*, 77–94. John Benjamins.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ranson, Diana L. & Matthew D. Passarello. 2012. L'élision variable du schwa en fin de mot chez des hommes méridionaux: L'effet des consonnes environnantes et de la fréquence de la lexie. In Franck Neveu, Valelia Muni Toke, Peter Blumenthal, Thomas Klingler, Pierluigi Ligas, Sophie Prévost et Sandra Teston-Bonnard, *Actes du CMLF 2012 -- 3e Congrès Mondial de Linguistique Française*, 1519-35. Paris: Institut de la Linguistique Française et EDP sciences. https://doi.org/10.1051/shsconf/20120100299.

Recanati, François. 1984. Remarques sur les verbes parenthétiques. In Pierre Attal & Claude Muller (eds.), *De la Syntaxe à la Pragmatique : Actes du Colloque de Rennes, Université de Haute-Bretagne*, 319–352. Amsterdam: John Benjamins.

Ritchie, Robert Lindsay Graeme. 1907. *Recherches sur la syntaxe de la conjonction "que" dans l'ancien français depuis les origines de la langue jusqu'au commencement du XIIIe siècle*. Paris: l'Université Paris 1 Panthéon-Sorbonne Dissertation.

Roberge, Yves & Nicole Rosen. 1999. Preposition Stranding and *que*-deletion in Varieties of North American French. *Linguistica Atlantica* 21. 153–168.

Sankoff, Gillian, Robert Sarrasin & Henrietta Cedergren. 1971. Quelques considérations sur la distribution sociolinguistique de la variable QUE dans le français de Montréal. Presented at *L'ACFAS, Université de Sherbrooke, Québec, 15-16 October 1971*.

Sankoff, Gillian. 1980. *The social life of language*. Philadelphia: University of Pennsylvania Press.

Sankoff, Gillian & Diane Vincent. 1981. L'emploi productif du *ne* dans le français parlé à Montréal. *Le Français moderne* 45. 243–256.

Schneider, Stefan. 2007. *Reduced Parenthetical Clauses as Mitigators: A corpus study of spoken French, Italian and Spanish.* Amsterdam: John Benjamins.

Schneider, Stefan & Julie Glikman. 2015. Origin and development of French parenthetical verbs. In Stefan Schneider, Julie Glikman & Mathieu Avanzi (eds.), *Parenthetical verbs* (Linguistische Arbeiten 557), 163–188. Berlin: Walter De Gruyter.

Scorretti, Mauro. 1981. Complementizer Ellipsis in 15th Century Italian. *Journal of Italian Linguistics* 6(1). 35–46.

Silva-Corvalán, Carmen. 2008. The Limits of Convergence in Language Contact. *Journal of Language Contact* 2(1). 213–224. https://doi.org/10.1163/000000008792525246.

Stein-Smith, Kathleen. 2022. Africa and the French language are growing together in global importance. *The Conversation*. https://theconversation.com/africa-and-the-french-language-are-growing-together-in-global-importance-179224.

Thompson, Sandra A. & Anthony Mulac. 1991. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics* 15(3). 237–251.

Torrego, Esther. 1983. More Effects of Successive Cyclic Movement. *Linguistic Inquiry* 14(3). 561–565.

Urmson, James O. 1963. Parenthetical Verbs. In Charles E. Caton (ed.), *Philosophy and ordinary language* (Illini Books, Ib-14), 220–240. Urbana: University of Illinois Press.

Vincent, Diane. 1991. Quelques études sociolinguistiques de particules du discours. *Revue québécoise de linguistique théorique et appliquée* 10(3). 41–60.

Wagner, Suzanne. 2012. Age Grading in Sociolinguistic Theory. *Language and Linguistics Compass* 6. 371–382.

Wanner, Dieter. 1981. Surface Complementizer Deletion: Italian *che*~∅*. *Journal of Italian Linguistics* 6. 47–82.

Warren, Jane. 1994. Plus ça change, plus c'est pareil: The case of "que" in Montreal French. *Culture* 14(2). 39–49.

Wiesmath, Raphaële. 2002. Présence et absence du relatif et conjonctif *que* dans le français acadien: tendances contradictoires? In *Romantische Korpuslinguistik: Korpora und gesprochene Sprache / Romance Corpus Linguistics: Corpora and Spoken Language*, 395–410. Tübingen: Gunter Narr.

Willems, Dominique & Claire Blanche-Benveniste. 2014. A constructional corpus-based approach to "weak" verbs in French. In Hans C. Boas & Francisco Gonzálvez-García (eds.), *Romance Perspectives on Construction Grammar*, 113–138. Amsterdam: John Benjamins.

World Health Organization. 2023. Adolescent Health. https://www.who.int/health-topics/adolescent-health#tab=tab_1. (5 March, 2023).

Appendix A

Generalized linear mixed-effects models for European data

Model 1E: Full data set

Model 1E Goodness of fit statistics

C index	Somers' Dxy
0.7912999	0.5825998
acceptable discrimination	moderate predictive ability

Model 1E Confidence intervals for fixed effects

Factor	2.5%	97.5%
Intercept	-0.3632794	2.1396132
Region-Southern France	-0.8660359	1.1692125
Region-Switzerland	-1.1191316	0.4240013
Education-Primary	-1.4941010	1.3206331
Education-Secondary	-0.5647225	0.8755023
Age Group-Older	-0.9616694	0.8070226
Age Group-Younger	-1.1581811	0.5184702
Matrix Verb-Adverbs	-1.6931785	0.2236489
Matrix Verb-Desired Action	-196.0744906	231.8205468
Matrix Verb-Judgment	0.4442002	1.5855533
Matrix Verb-Speaking/Restatement	-0.6008424	0.3064572

Model 1E Variance Inflation Factors corrected for degrees of freedom

Factor	VIF-score
Region	1.066304
Education	1.158400
Age Group	1.148742
Matrix Verb	1.003108

Model 2E: Non-discourse marker forms

Model 2E Goodness of fit statistics

C index	Somers' Dxy
0.859126	0.718252
excellent discrimination	good predictive ability

Model 2E Confidence intervals for fixed effects

Factor	2.5%	97.5%
Intercept	1.4469630	4.4223780
Region-Southern France	-0.9252389	1.7192205
Region-Switzerland	-1.0933377	0.6111153
Education-Primary	-2.0674409	0.8133695
Education-Secondary	-0.7620118	0.8513549
Age Group-Older	-1.1530347	0.9173068
Age Group-Younger	-0.5595563	1.2034038

Model 2E Variance Inflation Factors corrected for degrees of freedom

Factor	VIF-score
Region	1.095426
Education	1.248263
Age Group	1.194130

Model 3E: Potential discourse marker forms

Model 3E Goodness of fit statistics

C index	Somers' Dxy
0.9309234	0.8618468
outstanding discrimination	good predictive ability

Model 3E Confidence intervals for fixed effects

Factor	2.5%	97.5%
Intercept	1.01583221	4.4661812
Region-Southern France	-2.07051490	1.4260372
Region-Switzerland	-1.68691791	0.6742027
Education-Primary	-3.25803991	0.8074852
Education-Secondary	-0.92304315	1.2489329
Age Group-Older	-1.69388492	0.9192398
Age Group-Younger	-2.19996557	0.4522600
Phonological-Liquid	-0.72730339	3.7186899
Phonological-Nasal	0.85489376	8.3424478
Phonological-Stop	-0.31009495	2.5648890
Phonological-Vowel	-0.18343576	1.4527635
Syntactic-Lexical Noun	-1.83268703	1.9988801
Syntactic-Other	-0.04125039	5.8024945
IM Matrix-Yes	-0.96439632	2.7309364
IM Embedded-Yes	-4.44051768	0.5090576
Matrix Verb-Dire	-4.66446308	-1.6955068
Matrix Verb-Penser	-1.96831852	0.4010105
Matrix Verb-Savoir	-5.22220486	-2.4031951
Matrix Verb-Trouver	-2.16457525	0.7228421
Matrix Verb-Vouloir Dire	-6.80963662	-2.8374675
Matrix Verb-Voir	-9.93099173	-4.5840600

Model 3E Variance Inflation Factors corrected for degrees of freedom

Factor	VIF-score
Region	1.186353
Education	1.304315
Age Group	1.265798
Phonological Context	1.194262
Syntactic Context	1.654728
IM Matrix	1.027733
IM Embedded	2.051552
Matrix Verb	1.074878

Appendix B

Generalized linear mixed-effects models for Quebec data

Model 1Q: Full data set

Model 1Q Goodness of fit statistics

C index	Somers' Dxy
0.7848256	0.5696512
acceptable discrimination	moderate predictive ability

Model 1Q Confidence intervals for fixed effects

Factor	2.5%	97.5%
Intercept	-1.1053553	-0.2668245
Education-Primary	-1.0860281	0.5702417
Education-Secondary	-0.3310541	0.5523476
Age Group-Older	-0.5356473	0.7774300
Age Group-Younger	-0.2191861	0.6754705
Phonological-Liquid	0.2452081	1.9078711
Phonological-Nasal	0.1111870	2.0363610
Phonological-Stop	-0.3824053	0.4060303
Phonological-Vowel	0.4440820	1.0225165
Syntactic-Lexical Noun	-0.9325455	0.5371179
Syntactic-Other	-0.7876074	1.0636191
IM Matrix-Yes	-0.2627645	0.6726553
IM Embedded-Yes	-1.1292978	0.5379399
Matrix Verb-Adverbs	0.2987167	1.5557859
Matrix Verb-Desired Action	2.1285046	2.9542913
Matrix Verb-Judgment	1.2105005	1.9082455
Matrix Verb-Speaking/Restatement	1.0045461	1.7022458

Model 1Q Variance Inflation Factors corrected for degrees of freedom

Factor	VIF-score
Education	1.202433
Age Group	1.202613
Phonological	1.131987
Syntactic	1.673796
IM Matrix	1.011299
IM Embedded	1.930092
Matrix Verb	1.025296

Model 2Q: Non-discourse marker forms

Model 2Q Goodness of fit statistics

C index	Somers' Dxy
0.6496706	0.2993413
poor discrimination	poor predictive ability

Model 2Q Confidence intervals for fixed effects

Factor	2.5%	97.5%
Intercept	1.22159940	2.1377951
Education-Primary	-0.52117218	1.0095159
Education-Secondary	-0.40658596	0.3702307
Age Group-Older	-0.95168090	0.2091083
Age Group-Younger	-0.28670103	0.5159242
Phonological-Liquid	-1.58778970	0.7674020
Phonological-Nasal	-2.18095744	0.7108231
Phonological-Stop	-0.61861382	0.3526057
Phonological-Vowel	-0.04986956	0.7695736
Syntactic-Lexical Noun	-0.45284331	1.8245048
Syntactic-Other	-1.33992271	1.8731458
IM Matrix-Yes	-0.71904393	0.3550581
IM Embedded-Yes	-1.99666926	0.9611209

Model 2Q Variance Inflation Factors corrected for degrees of freedom

Factor	VIF-score
Education	1.173616
Age Group	1.843565
Phonological	1.177108
Syntactic	1.157752
IM Matrix	1.008583
IM Embedded	2.193074

Model 3Q: Potential discourse marker forms

Model 3Q Goodness of fit statistics

C index	Somers' Dxy
0.9191479	0.8382958
outstanding discrimination	good predictive ability

Model 3Q Confidence intervals for fixed effects

Factor	2.5%	97.5%
Intercept	-3.0998824	0.9971642
Education-Primary	-1.9375957	0.8198862
Education-Secondary	-0.0328943	1.4988614
Age Group-Older	-0.5209582	1.6459906
Age Group-Younger	-0.5231277	0.9664387
Phonological-Liquid	-0.5678677	3.3390246
Phonological-Nasal	-1.6125378	2.8720114
Phonological-Stop	0.2731455	2.4878851
Phonological-Vowel	0.6253842	1.9709732
Syntactic-Lexical Noun	-1.3051190	2.0070733
Syntactic-Other	-0.9765759	2.7971487
IM Matrix-Yes	0.0804924	2.2823995
IM Embedded-Yes	-2.2161066	1.0038608

Model 3Q Variance Inflation Factors corrected for degrees of freedom

Factor	VIF-score
Education	1.188501
Age Group	1.181100
Phonological	1.146031
Syntactic	1.650424
IM Matrix	1.013066
IM Embedded	1.782533

Appendix C

Generalized linear mixed-effects models for African data

Model 1A: Full data set

Model 1A Goodness of fit statistics

C index	Somers' Dxy
0.8587598	0.7175196
excellent discrimination	good predictive ability

Model 1A Confidence intervals for fixed effects

Factor	2.5%	97.5%
Intercept	0.19508986	2.1180034
Region-Central African Republic	-1.63391051	1.1369027
Region-Côte d'Ivoire	-1.51021216	0.3304298
Age Group-Younger	-0.39827016	1.6109800
Phonological-Liquid	0.11327429	4.3556173
Phonological-Nasal	-1.07057839	3.4365623
Phonological-Stop	-1.08759485	1.4168838
Phonological-Vowel	-0.04054621	1.6841672
Syntactic-Lexical Noun	-3.59029322	0.1511357
Syntactic-Other	0.06811088	3.6968636
IM Matrix-Yes	-0.26226709	4.3630023
IM Embedded-Yes	-3.05120681	0.3786552

Model 1A Variance Inflation Factors corrected for degrees of freedom

Factor	VIF-score
Region	1.095341
Age Group	1.169961
Phonological	1.228867
Syntactic	1.950296
Intervening Material-Matrix	1.054839
Intervening Material-Embedded	2.080162

Model 2A: Non-discourse marker forms

Model 2A Goodness of fit statistics

C index	Somers' Dxy
0.8676316	0.7352632
excellent discrimination	good predictive ability

Model 2A Confidence intervals for fixed effects

Factor	2.5%	97.5%
Intercept	0.202190185	2.3754687
Region-Central African Republic	-1.272226044	2.2379317
Region-Côte d'Ivoire	-1.673685444	0.5090700
Age Group-Younger	-0.524584107	1.7371034
Phonological-Liquid	0.006280983	5.0689400
Phonological-Nasal	-1.302337414	4.5125451
Phonological-Stop	-1.357395815	1.3708977
Phonological-Vowel	-0.297938436	1.7400161
Syntactic-Lexical Noun	-4.200402677	0.3315301
Syntactic-Other	-0.093477760	4.2216345
IM Matrix-Yes	-0.546119820	4.1847471
IM Embedded-Yes	-3.850489257	0.0992841

Model 2A Variance Inflation Factors corrected for degrees of freedom

Factor	VIF-score
Region	1.112580
Age Group	1.145285
Phonological	1.237008
Syntactic	2.034481
Intervening Material-Matrix	1.067568
Intervening Material-Embedded	2.197736

Model 3A: Potential discourse marker forms

Model 3A Goodness of fit statistics

C index	Somers' Dxy
0.8986928	0.7973856
excellent discrimination	good predictive ability

Model 3A Confidence intervals for fixed effects

Factor	2.5%	97.5%
Intercept	-1.975125	2.625568
Region-Central African Republic	-6.039209	1.334890
Region-Côte d'Ivoire	-2.435266	1.800132
Age Group-Younger	-1.871417	5.624142

Model 3A Variance Inflation Factors corrected for degrees of freedom

Factor	VIF-score
Region	1.652651
Age Group	1.285555

Appendix D

Absent *que* rate by speaker and matrix verbal element

Europe: Younger female speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Younger females			
EF16 (61alh1)		EF27a (scajd1)	
savoir	25.0 (n=4)	peut-être	75.0 (n=4)
croire	25.0 (n=4)	trouver	50.0 (n=2)
c'est vrai	8.3 (n=24)	penser	22.2 (n=9)
dire	0 (n=3)	avoir l'impression	0 (n=4)
penser	0 (n=4)	c'est vrai	0 (n=3)
c'est	0 (n=2)	EF27b (scajc1)	
EF20 (50aev1)		peut-être	66.7 (n=3)
savoir	100 (n=3)	dire	50.0 (n=2)
voir	100 (n=2)	penser	22.7 (n=22)
dire	50.0 (n=2)	avoir l'impression	0 (n=3)
penser	16.7 (n=6)	croire	0 (n=6)
EF22 (44ams1)		c'est vrai	0 (n=7)
savoir	100 (n=2)	être + adj.	0 (n=2)
EF24 (21aml1)		falloir	0 (n=2)
voir	100 (n=13)	se dire	0 (n=2)
falloir	100 (n=4)	savoir	0 (n=8)
penser	100 (n=3)	trouver	0 (n=6)
savoir	88.9 (n=9)	EF30 (svacb1)	
dire	50.0 (n=2)	voir	100 (n=2)
EF26 (81afv1)		dire	50.0 (n=4)
c'est vrai	0 (n=2)		

Europe: Younger male speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Younger Males			
EM17 (61acm1)		EM24 (92acj1)	
vouloir dire	100 (n=14)	vouloir	0 (n=2)
voir	100 (n=5)	EM26 (21ama1)	
c'est vrai	2.5 (n=40)	voir	100 (n=17)
croire	0 (n=6)	savoir	80.0 (n=5)
falloir	0 (n=4)	c'est vrai	50.0 (n=2)
trouver	0 (n=3)	penser	50.0 (n=4)
EM19 (50ajp1)		vouloir dire	42.9 (n=7)
croire	0 (n=4)	falloir	0 (n=3)
dire	0 (n=3)	EM29 (21abl1)	
savoir	0 (n=2)	croire	0 (n=2)
EM21 (81aaa1)		c'est vrai	0 (n=3)
dire	0 (n=2)	peut-être	0 (n=2)
EM23 (44ars1)		penser	0 (n=2)
- -	- -	EM30 (13bfa1)	
		croire	0 (n=2)

Europe: Middle-aged female speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Middle-Aged Females			
EF42 (50atv1)		EF52 (svarb2)	
avoir l'impression	50.0 (n=2)	voir	100 (n=3)
c'est vrai	0 (n=2)	trouver	50.0 (n=2)
être + adj.	0 (n=2)	croire	0 (n=3)
pouvoir dire	0 (n=2)	c'est vrai	0 (n=2)
penser	0 (n=4)	falloir	0 (n=2)
EF43 (44aar2)		pouvoir dire	0 (n=3)
c'est vrai	0 (n=2)	EF53 (13bma1)	
penser	0 (n=2)	c'est	0 (n=2)
EF46 (svanp1)		penser	0 (n=3)
voir	100 (n=2)	EF54 (scahd1)	
savoir	75.0 (n=4)	voir	100 (n=3)
c'est vrai	0 (n=2)	dire	60.0 (n=5)
EF48 (61ajh1)		penser	33.3 (n=3)
voir	33.3 (n=3)	c'est	0 (n=2)
dire	22.2 (n=9)	croire	0 (n=3)
croire	0 (n=2)	être + adj.	0 (n=4)

falloir	0 (n=3)	EF55 (50alb1)	
trouver	0 (n=2)	croire	0 (n=3)
		dire	0 (n=3)
		falloir	0 (n=4)

Europe: Middle-aged male speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Middle-Aged Males			
EM42 (scapm1)		EM51 (61abm1)	
savoir	50.0 (n=2)	c'est vrai	66.7 (n=3)
croire	25.0 (n=4)	dire	20.0 (n=5)
falloir	0 (n=3)	penser	12.5 (n=8)
EM43 (50ayp1)		c'est	0 (n=2)
falloir	0 (n=2)	être + adj.	0 (n=2)
EM44 (scapy1)		falloir	0 (n=2)
penser	12.5 (n=8)	trouver	0 (n=7)
croire	0 (n=2)	EM53 (50ajm1)	
dire	0 (n=2)	falloir	0 (n=3)
EM45a (svaje1)		penser	0 (n=2)
dire	0 (n=2)	EM54 (81acc1)	
EM45b (13brp2)		- -	- -
voir	50.0 (n=2)		
c'est vrai	0 (n=2)		

Europe: Older female speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Older Females			
EF65 (svaab1)		EF78 (scajb1)	
dire	66.7 (n=6)	c'est	0 (n=3)
voir	50.0 (n=2)	falloir	0 (n=2)
croire	0 (n=3)	trouver	0 (n=2)
c'est vrai	0 (n=4)	EF80 (61adp1)	
penser	0 (n=3)	dire	50.0 (n=2)
trouver	0 (n=2)	croire	33.3 (n=6)
EF68 (scamm1)		c'est vrai	33.3 (n=6)
savoir	60.0 (n=5)	falloir	0 (n=3)
dire	33.3 (n=3)	savoir	0 (n=3)
croire	0 (n=2)	EF81 (44amr1)	
c'est vrai	0 (n=2)	- -	- -
penser	0 (n=2)	EF82 (13baa1)	
trouver	0 (n=4)	savoir	40.0 (n=5)

EF69 (50app1)		c'est	0 (n=2)
sembler	50.0 (n=2)	croire	0 (n=2)
croire	20.0 (n=5)	dire	0 (n=2)
c'est vrai	33.3 (n=3)	EF86 (21ash1)	
dire	0 (n=2)	savoir	0 (n=2)
falloir	0 (n=2)		
trouver	0 (n=3)		

Europe: Older male speakers

Speaker/Verb	% Ø	Speaker/Verb	% Ø
Older Males			
EM70 (svarv1)		EM81a (13brp1)	
pouvoir dire	0 (n=2)	savoir	100 (n=2)
penser	0 (n=3)	dire	33.3 (n=3)
EM73 (81ajc1)		EM81b (50arm1)	
- -	- -	croire	50.0 (n=2)
EM75 (scarp1)		savoir	50.0 (n=2)
dire	66.7 (n=3)	c'est vrai	0 (n=2)
c'est	0 (n=3)	dire	0 (n=3)
trouver	0 (n=2)	penser	0 (n=3)
EM78 (scaaf1)		trouver	0 (n=2)
trouver	50.0 (n=4)	vouloir	0 (n=2)
EM79 (44ajs1)		EM81c (54bpm1)	
- -	- -	dire	25.0 (n=4)

Quebec: Younger female speakers

Speaker/Verb	% Ø	Speaker/Verb	% Ø
Younger Females			
QF16a (17J)		QF24 (10V)	
savoir	80.0 (n=10)	falloir	0 (n=4)
penser	16.7 (n=12)	dire	0 (n=7)
dire	0 (n=9)	c'est vrai	0 (n=2)
falloir	0 (n=7)	penser	25.0 (n=12)
vouloir	0 (n=7)	savoir	33.3 (n=3)
QF16b (17C)		peut-être	0 (n=3)
falloir	22.2 (n=9)	se dire	0 (n=2)
penser	14.3 (n=7)	trouver	20.0 (n=5)
trouver	0 (n=2)	vouloir	0 (n=2)
dire	0 (n=5)	QF25 (10M)	
savoir	71.4 (n=7)	savoir	42.9 (n=7)
sembler	0 (n=2)	trouver	20.0 (n=5)

c'est	0 (n=2)	dire	0 (n=5)
vouloir	0 (n=3)	falloir	10.0 (n=10)
voir	0 (n=4)	penser	40.0 (n=10)
QF17 (17K)		être + adj.	50.0 (n=2)
savoir	90.0 (n=10)	vouloir dire	100 (n=10)
penser	0 (n=6)	QF27 (14A)	
dire	0 (n=4)	c'est vrai	25.0 (n=4)
falloir	16.7 (n=12)	trouver	10.0 (n=10)
être + adj.	0 (n=2)	vouloir dire	0 (n=2)
vouloir	0 (n=5)	c'est	0 (n=5)
QF22 (25C)		dire	0 (n=13)
vouloir dire	100 (n=2)	croire	0 (n=5)
penser	20.0 (n=5)	peut-être	33.3 (n=6)
savoir	75.0 (n=4)	être + adj.	0 (n=9)
trouver	25.0 (n=4)	penser	0 (n=11)
QF23 (25ME)		falloir	0 (n=13)
sembler	100 (n=2)	savoir	80.0 (n=5)
savoir	100 (n=2)	sembler	0 (n=2)
falloir	0 (n=6)	vouloir	0 (n=7)
		voir	50.0 (n=2)

Quebec: Younger male speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Younger Males			
QM22 (25P)		QM26 (14H)	
dire	0 (n=4)	savoir	100 (n=10)
savoir	100 (n=10)	dire	0 (n=3)
c'est	0 (n=3)	être + adj.	0 (n=3)
falloir	0 (n=13)	trouver	0 (n=2)
être + adj.	0 (n=4)	penser	0 (n=5)
sembler	0 (n=2)	sembler	75.0 (n=4)
vouloir dire	33.3 (n=3)	vouloir	0 (n=3)
QM24a (21O)		voir	0 (n=2)
peut-être	50.0 (n=2)	QM27a (14B)	
penser	33.3 (n=3)	penser	33.3 (n=3)
falloir	0 (n=7)	savoir	100 (n=3)
savoir	85.7 (n=7)	peut-être	50.0 (n=2)
QM24b (10D)		dire	0 (n=3)
c'est	0 (n=2)	être + adj.	0 (n=3)
penser	33.3 (n=3)	falloir	0 (n=4)
peut-être	50.0 (n=2)	vouloir	0 (n=3)
QM25a (10J-M)		QM27b (21Y)	
penser	60.0 (n=10)	dire	40.0 (n=5)

savoir	53.3 (n=15)	être + adj.	33.3 (n=3)
falloir	7.14 (n=14)	falloir	14.3 (n=14)
peut-être	50.0 (n=2)	penser	50.0 (n=4)
être + adj.	25.0 (n=4)	c'est vrai	50.0 (n=2)
dire	0 (n=2)	peut-être	50.0 (n=4)
vouloir dire	75.0 (n=4)	se dire	50.0 (n=2)
QM25b (21S)		savoir	76.9 (n=13)
- -	- -		

Quebec: Middle-aged female speakers

Speaker/Verb	% Ø	Speaker/Verb	% Ø
Middle-Aged Females			
QF42 (6F)		QF49 (15J)	
vouloir dire	100 (n=2)	être + adj.	33.3 (n=3)
dire	0 (n=2)	penser	0 (n=6)
être + adj.	0 (n=2)	c'est vrai	0 (n=7)
penser	14.3 (n=7)	dire	0 (n=3)
falloir	14.3 (n=7)	falloir	0 (n=9)
savoir	90.0 (n=10)	peut-être	0 (n=3)
paraître	0 (n=2)	sembler	0 (n=2)
vouloir	50.0 (n=2)	savoir	58.3 (n=12)
QF47 (12J)		trouver	0 (n=3)
dire	0 (n=2)	vouloir	0 (n=4)
falloir	57.1 (n=7)	voir	66.7 (n=3)
penser	33.3 (n=6)	QF50 (1C)	
savoir	83.3 (n=6)	penser	0 (n=3)
trouver	33.3 (n=3)	falloir	0 (n=5)
vouloir dire	50.0 (n=2)	dire	0 (n=2)
QF48a (23S)		sembler	33.3 (n=2)
c'est vrai	83.3 (n=6)	savoir	60.0 (n=5)
dire	0 (n=2)	vouloir dire	75.0 (n=4)
falloir	11.1 (n=9)	vouloir	0 (n=2)
penser	20.0 (n=5)	QF52 (1S)	
savoir	25.0 (n=4)	dire	15.4 (n=13)
trouver	0 (n=2)	trouver	0 (n=3)
vouloir dire	0 (n=2)	être + adj.	50.0 (n=2)
QF48b (23T)		savoir	46.2 (n=13)
dire	25.0 (n=4)	c'est vrai	57.1 (n=7)
falloir	0 (n=7)	penser	22.2 (n=9)
peut-être	33.3 (n=3)	falloir	22.2 (n=9)
penser	38.5 (n=13)	vouloir dire	100 (n=6)
sembler	33.3 (n=3)	voir	0 (n=2)
savoir	75.0 (n=4)	QF53 (15S)	

trouver	16.7 (n=6)	savoir	80.0 (n=10)
vouloir	0 (n=4)	penser	33.3 (n=6)
		peut-être	0 (n=2)
		dire	42.9 (n=7)
		falloir	0 (=14)
		c'est vrai	66.7 (n=3)
		vouloir	0 (n=5)
		voir	0 (n=2)

Quebec: Middle-aged male speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Middle-Aged Males			
QM42 (6M)		QM51b (1G)	
falloir	57.1 (n=14)	penser	0 (n=5)
dire	0 (n=3)	falloir	0 (n=7)
penser	100 (n=2)	être + adj.	0 (n=2)
sembler	100 (n=2)	c'est	0 (n=2)
savoir	100 (n=3)	croire	0 (n=2)
QM47 (6J)		dire	66.7 (n=3)
falloir	0 (n=4)	peut-être	0 (n=3)
c'est vrai	100 (n=3)	sembler	50.0 (n=2)
dire	0 (n=2)	savoir	60.0 (n=5)
être + adj.	25.0 (n=4)	trouver	0 (n=3)
peut-être	33.3 (n=3)	vouloir dire	40.0 (n=5)
penser	0 (n=7)	vouloir	0 (n=2)
savoir	100 (n=9)	voir	0 (n=2)
QM48 (23J)		QM53 (15B)	
penser	37.5 (n=8)	penser	75.0 (n=4)
dire	0 (n=2)	être + adj.	0 (n=2)
falloir	0 (n=3)	c'est vrai	0 (n=3)
savoir	0 (n=2)	falloir	12.5 (n=16)
QM50 (23B)		savoir	100 (n=5)
falloir	33.3 (n=3)	QM55 (1L)	
QM51a (15R)		c'est vrai	100 (n=2)
falloir	40.0 (n=5)	penser	50.0 (n=2)
savoir	75.0 (n=4)	trouver	25.0 (n=4)
penser	40.0 (n=5)	vouloir	0 (n=2)
être + adj.	0 (n=2)		
c'est vrai	33.3 (n=3)		
c'est -à-dire	50.0 (n=2)		

Quebec: Older female speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Older Females			
QF65 (20AN)		QF81 (11M)	
dire	14.3 (n=7)	penser	0 (n=4)
être + adj.	50.0 (n=2)	falloir	0 (n=3)
savoir	100 (n=3)	savoir	87.5 (n=8)
c'est vrai	100 (n=3)	sembler	50.0 (n=2)
falloir	20.0 (n=10)	dire	0 (n=5)
penser	0 (n=6)	paraître	0 (n=2)
vouloir dire	0 (n=3)	trouver	0 (n=2)
QF71 (8E)		QF82 (4H)	
penser	25.0 (n=4)	dire	0 (n=2)
falloir	0 (n=3)	savoir	100 (n=8)
dire	0 (n=3)	paraître	0 (n=2)
savoir	60.0 (n=5)	falloir	20.0 (n=10)
peut-être	50.0 (n=2)	penser	40.0 (n=5)
QF74 (8H)		c'est vrai	50.0 (n=4)
falloir	0 (n=3)	peut-être	50.0 (n=2)
dire	0 (n=2)	voir	50.0 (n=2)
c'est vrai	0 (n=3)	QF95 (11Mi)	
savoir	100 (n=2)	penser	0 (n=3)
trouver	50.0 (n=2)	savoir	85.7 (n=7)
QF79 (4M)		falloir	0 (n=5)
penser	50.0 (n=2)	trouver	0 (n=2)
dire	0 (n=3)		
falloir	25.0 (n=4)		
savoir	100 (n=2)		

Quebec: Older male speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Older Males			
QM65 (12S)		QM80 (4A)	
dire	0 (n=2)	penser	71.4 (n=7)
falloir	0 (n=3)	dire	0 (n=2)
penser	50.0 (n=4)	falloir	11.8 (n=17)
vouloir dire	0 (n=2)	être + adj.	0 (n=2)
QM73a (8G)		savoir	77.8 (n=9)
savoir	100 (n=3)	avoir l'impression	0 (n=2)
sembler	33.3 (n=3)	c'est-à-dire	100 (n=3)
falloir	0 (n=8)	QM81 (4R)	
dire	0 (n=6)	penser	33.3 (n=3)

QM73b (8N)		peut-être	0 (n=3)
sembler	100 (n=3)	falloir	0 (n=5)
dire	50.0 (n=12)	savoir	66.7 (n=9)
falloir	0 (n=12)	trouver	0 (n=2)
peut-être	50.0 (n=2)	QM87 (11J)	
penser	9.1 (n=11)	dire	0 (n=2)
savoir	63.6 (n=11)	savoir	50.0 (n=8)
vouloir dire	87.5 (n=8)	pouvoir dire	0 (n=2)
voir	0 (n=2)	QM95 (11H)	
		falloir	0 (n=4)

Africa: Younger speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Younger Females			
AF19 (ciaeo1)		AF28a (bfato1)	
c'est	50.0 (n=2)	c'est	0 (n=2)
dire	100 (n=2)	dire	20.0 (n=5)
AF23 (ciana1)		AF28b (rcadt1)	
c'est	0 (n=3)	falloir	0 (n=13)
c'est-à-dire	75.0 (n=4)	AF28c (rcayn1)	
dire	20.0 (n=5)	dire	0 (n=4)
falloir	0 (n=2)	falloir	0 (n=2)
AF26 (ciaeo2)		AF29 (bfanr1)	
c'est vrai	50.0 (n=2)	paraître	0 (n=3)
penser	0 (n=3)	Younger Males	
dire	0 (n=2)	AM28 (rcaat1)	
AF27 (bfabo1)		dire	25.0 (n=4)
paraître	0 (n=2)	falloir	0 (n=5)

Africa: Middle-aged speakers

Speaker/Verb	% ∅	Speaker/Verb	% ∅
Middle-Aged Females			
AF41a (rcarn1)		AF48b (bfasy1)	
falloir	0 (n=4)	- -	- -
AF41b (bfadg1)		AF48c (ciasn1)	
se dire	0 (n=2)	falloir	0 (n=3)
trouver	0 (n=2)	penser	0 (n=2)
AF44a (ciapa1)		savoir	50.0 (n=2)
c'est	0 (n=2)	se dire	0 (n=2)
croire	66.7 (n=6)	AF50 (bfati1)	
dire	50.0 (n=2)	c'est-à-dire	50.0 (n=2)

penser	16.7 (n=6)	dire	0 (n=2)
peut-être	50.0 (n=6)	AF54 (bfath1)	
AF44b (rcamy1)		dire	0 (n=4)
- -	- -	AF55 (ciavd1)	
AF45 (rcamk1)		peut-être	0 (n=2)
dire	0 (n=4)	savoir	20.0 (n=5)
falloir	0 (n=4)	Middle-Aged Males	
paraître	20.0 (n=5)	AM40 (ciafk1)	
AF46a (bfakk1)		c'est-à-dire	0 (n=4)
croire	50.0 (n=2)	croire	33.3 (n=3)
AF46b (bfasb1)		falloir	0 (n=2)
- -	- -	penser	0 (n=2)
AF47 (bfann1)		peut-être	0 (n=2)
- -	- -	AM50 (bfaki1)	
AF48a (bfada1)		dire	33.3 (n=3)
- -	- -		

Appendix E

Absent *que* rate by speaker and phonological groups

Europe: Younger speakers

Phonological Context/ % Ø	Phonological Context/ % Ø	Phonological Context/ % Ø	Phonological Context/ % Ø
Younger Females		Younger Males	
EF16 (61alh1)	EF26 (81afv1)	EM17 (61acm1)	EM24 (92acj1)
S: 9.1 (n=11)	S: - -	S: 20.0 (n=5)	S: - -
F: 6.7 (n=15)	F: - -	F: 21.2 (n=33)	F: - -
N: - -	N: - -	N: 0 (n=3)	N: - -
L: 0 (n=5)	L: - -	L: - -	L: - -
V: 12.5 (n=16)	V: 0 (n=2)	V: 36.4 (n=33)	V: 0 (n=3)
EF20 (50aev1)	EF27a (scajd1)	EM19 (50ajp1)	EM26 (21ama1)
S: - -	S: 0 (n=2)	S: - -	S: 100 (n=5)
F: 42.9 (n=7)	F: 55.6 (n=9)	F: 0 (n=6)	F: 93.8 (n=16)
N: - -	N: - -	N: - -	N: - -
L: - -	L: 0 (n=3)	L: - -	L: 33.3 (n=6)
V: 50.0 (n=6)	V: 20.0 (n=10)	V: 0 (n=4)	V: 41.7 (n=12)
EF22 (44ams1)	EF27b (scajc1)	EM21 (81aaa1)	EM29 (21abl1)
S: - -	S: 6.7 (n=15)	S: - -	S: 50.0 (n=2)
F: 50.0 (n=2)	F: 16.0 (n=25)	F: 0 (n=2)	F: 0 (n=4)
N: - -	N: 0 (n=2)	N: 50.0 (n=2)	N: - -
L: - -	L: - -	L: - -	L: - -
V: 0 (n=2)	V: 16.7 (n=24)	V: - -	V: 0 (n=7)
EF24 (21aml1)	EF30 (svacb1)	EM23 (44ars1)	EM30 (13bfa1)
S: 71.4 (n=7)	S: - -	S: - -	S: - -
F: 76.9 (n=13)	F: 25.0 (n=4)	F: - -	F: 0 (n=2)
N: - -	N: - -	N: - -	N: - -
L: 66.7 (n=3)	L: - -	L: - -	L: - -
V: 60.0 (n=10)	V: 50.0 (n=4)	V: - -	V: - -

Europe: Middle-aged speakers

Phonological Context/ % ø	Phonological Context/ % ø	Phonological Context/ % ø	Phonological Context/ % ø
Middle-Aged Females		Middle-Aged Males	
EF42 (50atv1)	EF52 (svarb2)	EM42 (scapm1)	EM45b (13brp2)
S: 0 (n=6)	S: 50.0 (n=2)	S: - -	S: - -
F: 33.3 (n=6)	F: 25.0 (n=8)	F: 33.3 (n=3)	F: 20.0 (n=5)
N: - -	N: - -	N: - -	N: - -
L: - -	L: - -	L: - -	L: - -
V: 0 (n=6)	V: 22.2 (n=9)	V: 16.7 (n=6)	V: - -
EF43 (44aar2)	EF53 (13bma1)	EM43 (50ayp1)	EM51 (61abm1)
S: - -	S: 0 (n=2)	S: - -	S: 0 (n=2)
F: 0 (n=4)	F: 0 (n=5)	F: - -	F: 41.7 (n=12)
N: - -	N: - -	N: - -	N: - -
L: 0 (n=2)	L: - -	L: - -	L: 0 (n=2)
V: - -	V: 33.3 (n=3)	V: 33.3 (n=3)	V: 0 (n=15)
EF46 (svanp1)	EF54 (scahd1)	EM44 (scapy1)	EM53 (50ajm1)
S: - -	S: 0 (n=2)	S: 0 (n=2)	S: - -
F: 33.3 (n=3)	F: 26.7 (n=15)	F: 11.1 (n=9)	F: 25.0 (n=4)
N: 0 (n=2)	N: - -	N: - -	N: - -
L: - -	L: - -	L: - -	L: - -
V: 75.0 (n=4)	V: 57.1 (n=7)	V: 25.0 (n=4)	V: 0 (n=3)
EF48 (61ajh1)	EF55 (50alb1)	EM45a (svaje1)	EM54 (81acc1)
S: - -	S: - -	S: - -	S: - -
F: 20.0 (n=10)	F: 0 (n=10)	F: 33.3 (n=3)	F: - -
N: - -	N: - -	N: - -	N: - -
L: 33.3 (n=3)	L: - -	L: - -	L: - -
V: 8.3 (n=12)	V: 0 (n=4)	V: 0 (n=2)	V: 66.7 (n=3)

Europe: Older speakers

Phonological Context/ % ø	Phonological Context/ % ø	Phonological Context/ % ø	Phonological Context/ % ø
Older Females		Older Males	
EF65 (svaab1)	EF80 (61adp1)	EM70 (svarv1)	EM 79 (44ajs1)
S: - -	S: - -	S: - -	S: - -
F: 50.0 (n=4)	F: 42.9 (n=7)	F: 50.0 (n=4)	F: - -
N: 50.0 (n=2)	N: - -	N: - -	N: - -
L: - -	L: - -	L: - -	L: - -
V: 15.4 (n=13)	V: 25.0 (n=16)	V: 0 (n=2)	V: 0 (n=2)
EF68 (scamm1)	EF81 (44amr1)	EM73 (81ajc1)	EM81a (13brp1)
S: - -	S: - -	S: - -	S: 33.3 (n=3)
F: 12.5 (n=8)	F: - -	F: - -	F: - -
N: 0 (n=2)	N: - -	N: - -	N: - -
L: 0 (n=2)	L: - -	L: - -	L: - -
V: 33.3 (n=9)	V: - -	V: 0 (n=2)	V: 66.7 (n=3)
EF69 (50app1)	EF82 (13baa1)	EM75 (scarp1)	EM81b (50arm1)
S: - -	S: - -	S: 33.3 (n=3)	S: 0 (n=4)
F: 14.3 (n=7)	F: 0 (n=2)	F: 50.0 (n=2)	F: 0 (n=5)
N: - -	N: - -	N: - -	N: - -
L: 0 (n=4)	L: - -	L: 50.0 (n=2)	L: - -
V: 36.4 (n=11)	V: 25.0 (n=8)	V: 0 (n=4)	V: 25.0 (n=8)
EF78 (scajb1)	EF86 (21ash1)	EM78 (scaaf1)	EM81c (54bpm1)
S: 66.7 (n=3)	S: - -	S: - -	S: - -
F: 0 (n=7)	F: 0 (n=2)	F: 28.6 (n=7)	F: 0 (n=2)
N: - -	N: - -	N: - -	N: - -
L: - -	L: - -	L: - -	L: - -
V: 25.0 (n=4)	V: - -	V: - -	V: 20.0 (n=5)

Quebec: Younger speakers

Phonological Context/ % ø	Phonological Context/ % ø	Phonological Context/ % ø	Phonological Context/ % ø
Younger Females		Younger Males	
QF16a (17J)	QF23 (25ME)	QM22 (25P)	QM25b (21S)
S: 0 (n=6)	S: 50.0 (n=4)	S: 50.0 (n=2)	S: - -
F: 31.8 (n=22)	F: 66.7 (n=3)	F: 14.3 (n=21)	F: - -
N: 50.0 (n=2)	N: - -	N: - -	N: - -
L: 25.0 (n=4)	L: - -	L: 50.0 (n=2)	L: - -
V: 26.3 (n=19)	V: 40.0 (n=5)	V: 31.6 (n=19)	V: - -
QF16b (17C)	QF24 (10V)	QM24a (21O)	QM26 (14H)
S: 16.7 (n=6)	S: 0 (n=10)	S: 0 (n=2)	S: 100 (n=2)
F: 16.7 (n=18)	F: 22.2 (n=18)	F: 64.3 (n=14)	F: 35.3 (n=17)
N: 0 (n=3)	N: - -	N: - -	N: - -
L: 16.7 (n=6)	L: - -	L: 0 (n=2)	L: - -
V: 33.3 (n=12)	V: 6.7 (n=15)	V: 10.0 (n=10)	V: 43.8 (16)
QF17 (17K)	QF25 (10M)	QM24b (10D)	QM27a (14B)
S: 0 (n=11)	S: - -	S: - -	S: 50.0 (n=2)
F: 27.3 (n=22)	F: 37.9 (n=29)	F: 33.3 (n=6)	F: 25.0 (n=8)
N: - -	N: - -	N: - -	N: - -
L: 50.0 (n=2)	L: - -	L: - -	L: - -
V: 62.5 (n=8)	V: 45.5 (n=22)	V: 0 (n=2)	V: 15.4 (n=13)
QF22 (25C)	QF27 (14A)	QM25a (10J-M)	QM27b (21Y)
S: 66.7 (n=3)	S: 23.5 (n=17)	S: 18.2 (n=11)	S: 35.7 (n=14)
F: 50.0 (n=8)	F: 9.1 (n=44)	F: 46.4 (n=28)	F: 57.9 (n=19)
N: - -	N: - -	N: - -	N: - -
L: 50.0 (n=2)	L: 0 (n=6)	L: 50.0 (n=2)	L: 25.0 (n=4)
V: 0 (n=8)	V: 6.7 (n=30)	V: 38.5 (n=13)	V: 33.3 (n=12)

Quebec: Middle-aged speakers

Phonological Context/ % ø	Phonological Context/ % ø	Phonological Context/ % ø	Phonological Context/ % ø
Middle-Aged Females		Middle-Aged Males	
QF42 (6F)	QF49 (15J)	QM42 (6M)	QM51a (15R)
S: 66.7 (n=9)	S: 38.5 (n=13)	S: 58.3 (n=12)	S: 50.0 (n=4)
F: 41.7 (n=12)	F: 15.0 (n=20)	F: 71.4 (n=7)	F: 41.7 (n=12)
N: - -	N: - -	N: - -	N: - -
L: - -	L: 0 (n=4)	L: - -	L: - -
V: 29.4 (n=17)	V: 10.5 (n=19)	V: 44.4 (n=9)	V: 50.0 (n=8)
QF47 (12J)	QF50 (1C)	QM47 (6J)	QM51b (1G)
S: 66.7 (n=3)	S: 33.3 (n=3)	S: 40.0 (n=5)	S: 20.0 (n=5)
F: 71.4 (n=14)	F: 100 (n=3)	F: 55.0 (n=20)	F: 15.8 (n=19)
N: - -	N: - -	N: - -	N: - -
L: - -	L: 0 (n=3)	L: 0 (n=2)	L: 20.0 (n=5)
V: 0 (n=10)	V: 17.6 (n=17)	V: 11.1 (n=9)	V: 18.8 (n=16)
QF48a (23S)	QF52 (1S)	QM48 (23J)	QM53 (15B)
S: 14.3 (n=7)	S: 30.0 (n=10)	S: 0 (n=4)	S: 15.4 (n=13)
F: 25.0 (n=8)	F: 46.4 (n=28)	F: 25.0 (n=8)	F: 57.1 (n=7)
N: 0 (n=3)	N: 0 (n=4)	N: - -	N: - -
L: - -	L: 66.7 (n=3)	L: 0 (n=2)	L: - -
V: 30.8 (n=13)	V: 27.3 (n=22)	V: 25.0 (n=4)	V: 30.0 (n=10)
QF48b (23T)	QF53 (15S)	QM50 (23B)	QM55 (1L)
S: 0 (n=6)	S: 10.0 (n=10)	S: - -	S: - -
F: 55.0 (n=20)	F: 50.0 (n=22)	F: 0 (n=3)	F: 28.6 (n=7)
N: - -	N: - -	N: - -	N: - -
L: 100 (n=2)	L: 0 (n=2)	L: - -	L: - -
V: 10.5 (n=19)	V: 17.6 (n=17)	V: 0 (n=2)	V: 25.0 (n=4)

Quebec: Older speakers

Phonological Context/ % ∅	Phonological Context/ % ∅	Phonological Context/ % ∅	Phonological Context/ % ∅
Older Females		Older Males	
QF65 (20AN)	QF81 (11M)	QM65 (12S)	QM81 (4R)
S: 44.4 (n=9)	S: - -	S: 0 (n=3)	S: 33.3 (n=6)
F: 33.3 (n=12)	F: 15.4 (n=13)	F: 50.0 (n=6)	F: 18.2 (n=11)
N: - -	N: - -	N: - -	N: 50.0 (n=2)
L: - -	L: 0 (n=2)	L: - -	L: - -
V: 14.3 (n=14)	V: 53.8 (n=13)	V: 0 (n=5)	V: 33.3 (n=6)
QF71 (8E)	QF82 (4H)	QM73a (8G)	QM87 (11J)
S: - -	S: - -	S: 50.0 (n=4)	S: 33.3 (n=3)
F: 61.5 (n=13)	F: 58.3 (n=12)	F: 0 (n=6)	F: 28.6 (n=7)
N: - -	N: - -	N: - -	N: - -
L: 0 (n=5)	L: 50.0 (n=2)	L: 33.3 (n=3)	L: - -
V: 0 (n=2)	V: 33.3 (n=21)	V: 20.0 (n=10)	V: 25.0 (n=4)
QF74 (8H)	QF95 (11Mi)	QM73b (8N)	QM95 (11H)
S: - -	S: 100 (n=2)	S: 46.2 (n=13)	S: - -
F: 22.2 (n=9)	F: 33.3 (n=6)	F: 50.0 (n=20)	F: 33.3 (n=3)
N: - -	N: - -	N: - -	N: - -
L: - -	L: - -	L: - -	L: - -
V: 50.0 (n=6)	V: 20.0 (n=10)	V: 28.6 (n=28)	V: 0 (n=3)
QF79 (4M)		QM80 (4A)	
S: - -		S: 0 (n=2)	
F: 28.6 (n=7)		F: 62.5 (n=16)	
N: - -		N: - -	
L: - -		L: 50.0 (n=2)	
V: 66.7 (n=3)		V: 20.8 (n=24)	

Africa: Younger female speakers

Phonological Context/ % ∅			
Younger Females			
AF19 (ciaeo1)	AF26 (ciaeo2)	AF28a (bfato1)	AF28c (rcayn1)
S: - -	S: - -	S: 33.3 (n=3)	S: 50.0 (n=2)
F: - -	F: 20.0 (n=5)	F: 33.3 (n=3)	F: - -
N: - -	N: - -	N: - -	N: - -
L: - -	L: 0 (n=2)	L: 0 (n=2)	L: 33.3 (n=3)
V: 50.0 (n=4)	V: 0 (n=7)	V: 0 (n=2)	V: 0 (n=3)
AF23 (ciana1)	AF27 (bfabo1)	AF28b (rcadt1)	AF29 (bfanr1)
S: - -	S: - -	S: - -	S: - -
F: 16.7 (n=6)	F: 0 (n=2)	F: 20.0 (n=5)	F: 0 (n=2)
N: 50.0 (n=2)	N: - -	N: - -	N: - -
L: 33.3 (n=3)	L: - -	L: 14.3 (n=7)	L: 0 (n=2)
V: 33.3 (n=3)	V: - -	V: 25.0 (n=8)	V: 0 (n=4)

Africa: Younger male speaker

Phonological Context/ % ∅
Younger Males
AM28 (rcaat1)
S: 0 (n=3)
F: 11.1 (n=9)
N: - -
L: 0 (n=4)
V: 0 (n=5)

Africa: Middle-aged female speakers

Phonological Context/ % ∅			
Middle-Aged Females			
AF41a (rcarn1)	AF45 (rcamk1)	AF48a (bfada1)	AF54 (bfath1)
S: - -	S: - -	S: - -	S: 0 (n=2)
F: - -	F: 16.7 (n=6)	F: 100 (n=2)	F: 0 (n=4)
N: - -	N: - -	N: - -	N: - -
L: 0 (n=4)	L: 0 (n=2)	L: - -	L: - -
V: 0 (n=4)	V: 11.1 (n=9)	V: 50.0 (n=2)	V: - -
AF41b (bfadg1)	AF46a (bfakk1)	AF48b (bfasy1)	AF55 (ciavd1)
S: - -	S: - -	S: 0 (n=2)	S: 50.0 (n=6)
F: 16.7 (n=6)	F: - -	F: - -	F: 50.0 (n=4)
N: - -	N: - -	N: - -	N: - -
L: - -	L: - -	L: - -	L: - -
V: - -	V: 0 (n=3)	V: - -	V: 0 (n=2)
AF44a (ciapa1)	AF46b (bfasb1)	AF48c (ciasn1)	
S: - -	S: - -	S: 50.0 (n=2)	
F: 50.0 (n=12)	F: 0 (n=2)	F: 33.3 (n=6)	
N: - -	N: - -	N: 0 (n=3)	
L: 33.3 (n=3)	L: - -	L: - -	
V: 25.0 (n=8)	V: 50.0 (n=2)	V: 25.0 (n=4)	
AF44b (rcamy1)	AF47 (bfann1)	AF50 (bfati1)	
S: - -	S: - -	S: 33.3 (n=3)	
F: - -	F: 33.3 (n=3)	F: 66.7 (n=3)	
N: - -	N: - -	N: - -	
L: - -	L: - -	L: - -	
V: 50.0 (n=2)	V: - -	V: 0 (n=6)	

Africa: Middle-aged male speakers

Phonological Context/ % ∅	
Middle-Aged Males	
AM40 (ciafk1)	AM50 (bfaki1)
S: - -	S: - -
F: 0 (n=2)	F: 33.3 (n=3)
N: - -	N: - -
L: 0 (n=2)	L: - -
V: 8.3 (n=12)	V: 14.3 (n=7)