Information Structure, Syntax and Sociolinguistics through the History of French

Wyn Shaw

St. Hilda's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2024

Word Count: 69,985

Acknowledgements

I'd like to thank both the AHRC OOCDTP, Clarendon Scholarship and Bailie Gifford for providing the funding which made this work possible, without their continued support I would have been totally unable to carry out this research.

I'm deeply indebted to my DPhil supervisor: Sam Wolfe, for his continued support throughout my academic career, providing plentiful feedback and stimulating discussion in-person to develop my work. Sam has been relentlessly understanding and supportive through a variety of difficult life changes which have occurred over the course of my DPhil, and I do not think I would have been able to complete this degree without his flexibility and support.

I am grateful for the feedback provided by audiences at conferences and seminars as well as for my students' freshly critical eye on issues which relate to my thesis. Without this feedback innumerable errors would have survived into the submitted thesis. Likewise, the various theoretical discussions I've had with researchers of all kinds at conferences, in university rooms, pubs and elsewhere have been crucial to the development of the ideas in this thesis and my approach to academic work in general.

Finally, I'd like to express the deepest thanks to the emotional support of my communities both in Oxford and London throughout this process. I'm deeply appreciative for my former housemate Nico's endless emotional support as well as our frequent theoretical discussions over morning coffees. Similarly, my dearest friend Agnes has been ever-present over my DPhil, helping me through some difficult times with unrivalled patience and kindness. Finally, I'd like to express the deepest appreciation for the support which my partners Sioned, Josie and Faye have provided me in the last few months prior to submission.

Any errors which remain are my own.

Abstract

Information Structure, Syntax and Sociolinguistics through the History of French

Wyn Shaw

St. Hilda's College

A thesis submitted for the degree of *Doctor of Philosophy*

Hilary Term 2024

This thesis examines the impact of information-structural and social factors on syntactic change in declarative main clauses in the history of French. It provides a novel perspective on French historical syntax by synthesising methodologies from historical sociolinguistics and historical pragmatics with the cartographic approach to syntax across an extensive timespan, from the earliest texts to the contemporary spoken language. The thesis employs a wide variety of corpora and texts from varied registers and genres, including works of less-educated authors, alongside 17th-century metalinguistic texts, to provide a highly varied dataset through which examples of syntactic change are studied.

Through three case studies: post-verbal subjects, left and right dislocations and clefts, the thesis shows that information-structural and social factors have a crucial role to play in syntactic change. In each case study, we analyse the frequency of the construction across our dataset, assessing the impact of so-called external factors on syntactic variation and change. The thesis presents a qualitative cartographic analysis of the construction in question and its developments over the history of the language, with a particular focus on information-structural functions. Finally, the thesis concludes by contextualising these analyses with a discussion of the wider literature on information structure and syntactic change in French: it shows that the cartographic approach provides a powerful framework through which to analyse the role of information structure in syntactic change. However, in each case, external factors also play a considerable role, with social factors often mediating micro- and nano-parametric variation in a change in progress. Overall, the thesis shows that both external and internal factors must be considered to form an accurate description of a given syntactic change.

Table of Contents

List of Tables and Figures

List of Abbreviations

Adj(P)	Adjective (Phrase)
Adv(P)	Adverbial (Phrase)
Agr(P)	Agreement (Phrase)
ASP	Aspect(ual Adverb)
CL	Clitic
C(P)	Complementiser (Phrase)
Contr(P)	Contrastive Topic (Phrase)
D(P)	Determiner (Phrase)
EPP	Extended Projection Principle
FP	Unspecified Functional Phrase
Fam(P)	Familiar Topic (Phrase)
Fin(P)	Finiteness (Phrase)
Foc(P)	Focus (Phrase)
Frame(P)	Frame (Phrase)
Ground(P)	Ground (Phrase)
I(P)	Inflection(al Phrase)
LOC	Locative (clitic)
NEG	Negation
O	Object
Obj(P)	Object (Phrase)
P(P)	Preposition(al Phrase)
Pred(P)	Predicate (Phrase)
PTCL	Particle
Q(P)	Quantifier (Phrase)
REFL	Reflexive (clitic)
S	Subject

Shift(P)	Shifting Topic (Phrase)
Spec(-)	Specifier (of…)
Subj(P)	Subject (Phrase)
T(P)	Tense (Phrase)
Top(P)	Topic (Phrase)
v(P)	Little Verb (Phrase)
V(P)	Verb (Phrase)
V1	Verb First
V2	Verb Second
V3	Verb Third
XP	Phrasal Category

1. Introduction

The history of the French language and its structure have been much discussed over the last 200 years, and yet, as Lodge (1993:1) states, 'a subject as broad as this can never be exhausted. Even after all the work so far done in the history of French, a host of interesting problems remain unsolved, a great array of questions remain unanswered.' In particular, as Wolfe (2021) points out, 'a recurring characteristic of French linguistic studies in the last two centuries has been a frequently cursory treatment of syntax when discussing the history of the language and its present state' (Wolfe, 2021: 2).

While syntax has remained under-studied in the history of French, historical linguistics has continued to develop more broadly, with two major turns occurring in the last 40 years. Firstly, the development of historical sociolinguistics following Romaine's (1982) monograph and secondly, the development of historical pragmatics following Jucker's (1995) work. Both of these sub-fields can be seen to reflect a more general turn in linguistics towards social and functional elements of language use and change. Although these developments began in functionalist and sociolinguistic fields of research, we can see their influence in generative syntax too, namely in the development of the discourse-pragmatic left-periphery in cartographic syntax (Rizzi, 1997; Poletto, 2000, 2006; Benincà and Poletto, 2004; Benincà, 2001, 2006; Frascarelli and Hinterhölzl, 2007; Benincà and Munaro, 2011; Samek-Lodovici, 2015; Rizzi and Bocci, 2017; Samo, 2019; Wolfe, 2022), and in the development of syntactic theories of politeness markers and interactionally-oriented content (Speas and Tenny, 2003; Corr, 2022; Miyagawa, 2022). In this thesis, I endeavour to contribute to this socially and pragmatically oriented turn, all while utilising the formal rigour of cartographic syntax, to examine the ways in which social and pragmatic factors motivate syntactic change in the history of French.

In the study of French historical syntax, Wolfe's (2021) work provides by far the most extensive theoretically informed work to date, while the Grande Grammaire Historique du Français (2020) (henceforth: GGHF) provides the most extensive descriptive work. However, neither of these works applies historical pragmatic and sociolinguistic techniques to the history of French. While Wolfe's work makes use of information-structural concepts as they are relevant to cartographic

analysis, he does not make use of historical sociolinguistic approaches. Meanwhile, the GGHF (2020) is a purely descriptive reference work and does not make broader theoretical claims about the nature of language history and change, syntax, information structure or sociolinguistics.

Despite this lack of a comprehensive work on the interrelation of syntax, pragmatics and sociolinguistic pressures in language change, French historical linguistics has seen a swell in research in related areas in the last 20 years. Cartographic and information-structurally informed approaches to the left periphery have been used extensively to examine the Verb-Second (henceforth: V2) system of Old French and its collapse (Mathieu, 2012; Steiner, 2014; de Andrade, 2018; Labelle and Hirschbühler, 2018; Larrivée, 2021), while more recent work has examined the wider history of the syntax-information structure relationship (Wolfe, 2021). Several major studies have been conducted on the relationship between word order and information structure in the modern language as well, starting as early as the 1980s with Lambrecht's seminal work (Lambrecht, 1981; Lahousse, 2003, 2006; De Cat, 2007). With regard to historical sociolinguistics, French has seen some, albeit limited, study, particularly from the field of standardization studies (Lodge, 1993, 2004) and in discussions of periodisation (Smith, 2002). In the 2000s more fine-grained historical sociolinguistic work began to take place, particularly focusing on the Early Modern period (Martineau and Mougeon, 2003; Ayres-Bennett, 2004) and, in recent years, work on text-typological motivations for syntactic variation has been conducted on Old and Middle French (Simonenko, Crabbé and Prévost, 2018, 2019; Larrivée, 2022b).

Notably, however, there has been relatively little interaction between these different strands of historical linguistics. Additionally, the majority of studies on historical syntax and historical sociolinguistics take a single period in the history of French as their focus. This thesis, then, aims to synthesise the varying approaches of historical cartographic syntax, historical pragmatics and historical sociolinguistics to create a detailed description of syntactic change in declarative main clauses throughout the entire history of French.

The recent development of a wide variety of corpora is crucial to this work. While historical parsed corpora such as the MCVF (Martineau, Hirschbühler, Kroch and Morin, 2021) and Penn

Supplement (Kroch and Santorini, 2021)[1] have been used extensively in historical syntactic work, a number of more modern corpora have been under-utilised outside of sociolinguistic work. For instance, numerous spoken corpora are now available, as well as a wide variety of corpora of SMS and online written language. A guiding principle of this thesis is to use as varied a data set as possible to ensure that we gain an accurate understanding of syntactic variation and change. A particularly important element of this work is in our attempt to apply the same principles across distinct time periods. Thus, while we use letters to approximate spoken data in our early periods, our contemporary spoken data is compared against SMS data and data from internet fora, which we anticipate to be 'spoken-like' texts. I hope that this approach will lead to further discussion between synchronic and diachronic analyses, allowing us to see how synchronic variation may inform our understanding of diachronic change, as well as modern data deepening our understanding of the relation between historical written data and the spoken language of the day.

This thesis, then, aims to develop our understanding of syntactic change in French and, in particular, the roles that information-structural and sociolinguistic considerations play in this change. In order to explore this, I provide three case studies of structures which have been observed to have information-structural roles and analyse their distribution, including analysis of sociolinguistic variation. The case studies in this thesis are: post-verbal subjects, dislocations (both left and right) and *c'est* clefts. In each case, I focus only on matrix declarative clauses, since these are the primary locus of large-scale syntactic change. This approach does limit our analysis somewhat but seems methodologically necessary to cover such an ambitious set of data over such a timescale. I provide formal analysis of each construction throughout its history as well as discussing the relationship between the construction, information-structural constraints, and text-typological factors as well as the role of standardization, prescriptivism and sociohistorical pressures in the changes which occur.

Throughout, I make use of traditional periodisations. However, it should be noted that I do not understand these periodisations to represent ontologically distinct linguistic entities, rather, they are simply useful heuristic labels used to discuss the language of a given period in all its variation (on

[1] It should be noted that earlier pre-release versions of the corpus have been used in numerous studies.

the problematic nature of periodisation, see Smith, 2002). Thus, I use Old French to refer to the language of the earliest texts until the early 14th-century. I use Middle French to refer to the language of the 14th and 15th centuries. I use Early Modern French to refer to the language of the 16th-19th centuries. At times, I use the more specific Renaissance French for the language of the 16th century and Classical French for that of the 17th and 18th centuries. Finally, I use Modern French to refer to the language of the 20th century until the present day, and Contemporary French to refer to the colloquial language of the present day, with Contemporary Spoken French referring specifically to the spoken colloquial variety.

The thesis is structured as follows. In Chapter 2, I give an overview of the existing literature in the relevant sub-fields, namely information-structural studies, cartographic syntax and historical sociolinguistics, alongside a brief introduction to the case studies in this thesis. In Chapter 3, I present my own methodology, listing the corpora and texts used and describing the techniques and software used for analysis; the more particular methodologies relevant for each case study are reserved for their respective chapters.

In Chapter 4, I discuss post-verbal subjects. I conclude that cartographic syntax is an extremely powerful tool for the analysis of syntactic change. I show that a variety of post-verbal subject positions are available in French until the 17th century, at which point the stricter modern system develops. By utilising the fine-grained analytical approach of cartographic syntax, I demonstrate that a wider number of positions than previously acknowledged are present. With all this said, however, I also show that the relationship between syntax and information structure cannot be neatly mapped with a one-to-one mapping. Many syntactic positions are information-structurally under-specified and marking information-structural roles through syntactic movement appears to be optional.

In Chapter 5, I discuss dislocations. Again, I show that cartographic syntax is a powerful tool for mapping the relation between the dislocated constituent and its clause, although the situation does appear to be more complex than that of post-verbal subjects, with some inconsistencies. The strong relationship between syntactic position and information-structural role (near-categorical through the

entire history of French in dislocations) supports the importance of maintaining a cartographic approach to information structure. The most interesting finding of this case study, however, is the role which text-type and modality play in dislocations; almost all texts through the history of French show low rates of dislocations, while all Contemporary Spoken French data shows extremely high rates. By comparing Contemporary Spoken French data with written Contemporary French data of lower and more 'spoken-like' registers, I show how syntactic constructions can become modality-specific. Following interactional work on dislocations (Doehler, de Stefani and Hörlacher, 2015), I argue that, while cartographic syntax can model the position and information-structural roles of syntactic structures, it cannot currently analyse all the functions of dislocations. It seems that dislocations have taken on more interactionally-oriented and modality-specific roles in the spoken language. In this chapter, I also problematise the notion of 'spoken-like' text, showing that even the most spoken-like texts do not show rates of dislocations comparable to actual speech in Contemporary French.

Finally, in Chapter 6, I discuss clefts. Clefts represent one of the most distinctly challenging constructions for the cartographic approach; there has been extensive debate on the appropriate analysis of these constructions. In this chapter, I present a variety of descriptive data on clefts, arguing that the existing cartographic literature on the construction often ignores the variety of attested structures. I present a sketch of a potential analysis, while arguing that clefts may be difficult to analyse under the cartographic framework. Additionally, I suggest that the history of clefts' increasing frequency is linked to the loss of other focusing techniques during the Early Modern period (contra previous work by Dufter, 2008, 2009). I tentatively argue that clefts may be an example where alternative approaches to syntactic analysis may be more viable than a cartographic approach, particularly in light of arguments for the lexicalization of clefts in the contemporary language.

In Chapter 7, I present an overview of the foregoing chapters, discussing the many ways in which social, pragmatic and syntactic pressures intertwine and the extent to which cartographic frameworks can be used to describe this. I discuss the overall information-structural system of French syntax and its history in light of both my own case studies and the existing literature, particularly the

literature on focus fronting and topicalization without clitic resumption. I argue that the changes we see through the history of French typically begin at the micro- and nano-parametric level; larger parametric changes then follow from this low-level change. Although sweeping changes occur through the history of French, I show that few 'extinct' constructions have become totally extinct, often surviving in specific micro- or nano-parametrically determined environments. Additionally, I discuss some limits of a narrow cartographic approach. By focusing only on those elements of language which have been formalised in cartography, we can miss the other potential analyses available in interactional and historical pragmatic approaches. Under these approaches, we can see that these structures can develop new and particularised uses, particularly related to interactional functions such as turn-taking (dislocations) or attitudinal expression (clefts). I argue, then, that the cartographic approach allows us to provide a fine-grained analysis of the internal structure of the clause and the limitations on possible clauses which might exist, but it does not, currently at least, provide a full description of the function of syntax and syntactic structures, which may be repurposed as markers of specific, typically more socially and pragmatically oriented, purposes.

In Chapter 8, I turn to a more theoretical discussion, asking whether the cartographic approach is tenable in light of various critiques of generative assumptions. I argue that social, interactional and pragmatic factors impact language down to its most structural sub-systems. I also suggest, following the variation seen in our own data, that language change may be able to occur over the life of an individual and need not be driven by child language acquisition. However, I also argue that the cartographic approach has proven extremely useful throughout this thesis in providing fine-grained analysis of the internal structure of the clause. This analysis is not fundamentally tied to any generative assumptions, and it can be fruitfully combined with social, pragmatic and interactional data to give us a more accurate and in-depth description of language history. I follow recent work by Corr (2022) and Miyagawa (2022), arguing that the cartographic formalism can be adapted to include interactional use of language. Thus, I suggest that the integration of interactional and social analysis into the formal analysis of cartography is a necessary next step to further developing our analyses.

In Chapter 9, I present some concluding remarks and potential areas for further study, summarising the key findings of this thesis and noting areas where further work is required.

2. Background

Before proceeding to my own data, analysis and discussion, a variety of different concepts require introduction. In this chapter, I survey the existing literature in the fields of information structure, syntax, text typology and historical sociolinguistics, focusing on the schools of thought which I use throughout this thesis. I then turn to the three constructions I study in this thesis, giving a brief overview of their structure. It should be noted, however, that in-depth overviews of the relevant literature on each construction is provided in its own chapter.

2.1 Information Structure

I begin with a survey of the information-structural literature. A wide variety of approaches to information structure have been taken in the literature. In this thesis, I will focus on Lambrecht's (1994) system, but others will be discussed here.

Information structure can be seen as a sub-field of the wider study of discourse pragmatics. While discourse pragmatics at large is concerned with a wide range of ways in which meaning is expressed between clauses, information structure is more specific. Information structure developed out of the notion of information packaging in the works of Chafe (1976) and Prince (1981). Information packaging consists of strategies by which interlocutors affect and understand each other's assumptions and beliefs. Crucial to this is the notion of the common ground: those propositions which are assumed to be understood by both interlocutors. In Lambrecht's (1994) work, following Chafe (1976), information structure relates not just to the content of a discourse, but to the way it is transmitted. Thus, for Lambrecht (1994: 3) information structure is, by definition, those information packaging functions which are reflected in the grammar. Additionally, information structure, under this view, operates at the level of the clause, although it relates to features of the discourse. It should be noted that this approach, often explicitly (e.g. Lambrecht, 1994: 36), assumes the primacy of spoken language. It is not immediately obvious how one should apply these concepts to written language. For our purposes, however, we take written language to be a discourse between the author and an imagined audience.

Now that we have provided a broad definition of information structure, I turn to the particular practical concepts that will be used in our analyses. Throughout this thesis, three notions will be key to our approach to information structure: givenness, topics and foci.

2.1.1 Givenness

I begin with a discussion of givenness, before moving to the two notions of topic and focus. Givenness, also sometimes described as or combined with the notions of activation or accessibility, has played a central role in information-structural studies historically. In Lambrecht's (1994) terminology, the concept of activation is used. Activation refers to the degree to which a certain referent is accessible in the mind of discourse participants. If a referent is recently mentioned in the discourse, it can be seen as activated; if it is inferable from the extra-linguistic context or from assumptions about the semantic context, it can be seen as accessible; if it is not at all activated, it can be seen as new. In the terminology which we use, a referent may be given, accessible or new. In other approaches, particularly the work of Chafe (1976) and Lambrecht (1994) the notions of activation are used. Lambrecht (1994: 93) takes care to distinguish between this notion of activation (our givenness) and the notion of identifiability. Identifiability refers to the degree to which a referent can be identified by discourse participants, regardless of whether it is new to the discourse or not, if a referent is identifiable, it may then have differing degrees of activation: activated, semi-activated/accessible or new. In this approach, an unidentifiable referent is referred to as brand new. While these distinctions are highly necessary for a comprehensive account of information structure, they are excessive for our purposes. As such, throughout this thesis we focus primarily on activation and use the terminology of 'given'-ness, referring to referents as given, accessible or new, depending on their accessibility. Identifiability will not be discussed, since it is less relevant to the central concepts of topic and focus.

2.1.2 Focus

Having discussed givenness, I turn to foci. The concept of focus has numerous interconnected meanings. In the Prague School (Daneš, 1968; Firbas, 1971 and others), focus is taken as complementary to topic, with the topic the old information of a clause and the focus the new

information. A similar approach is taken up in the early generative tradition of Chomsky and Jackendoff; Jackendoff (1972) treats focus as the new information communicated in a clause, complementing the notion of presupposition, that which is already known or assumed. However, in more recent work, focus takes on a narrower meaning. Krifka and Musan (2012) discuss the history of the concept and identify two other key elements. Firstly, a focus can be seen to correspond to the *wh*-word in a corresponding question. For instance, consider the exchange below:

(1) 'What did John eat?'

'He ate the cake.'

In this example, 'the cake' can be said to be focused since it corresponds to the *wh*-word in the question. Secondly, following Rooth (1985), a focus suggests the possibility of alternatives, or rather, draws attention to them. For example, the focusing of 'the cake' in the above example suggests the possibility that John may have eaten something else. This function also relates to the notion that focus should be contrastive. For our purposes however, we focus on the *wh*-word test. In Lambrecht's (1994) work, the focus is understood as 'the element of information whereby the presupposition and the assertion DIFFER from each other' (Lambrecht, 1994: 207). Crucially, however, this does not mean that a focused referent is new. For instance, take example (2) below:

(2) 'John and Clare came to the party.'

'Who ate the cake?'

'John ate the cake.'

In this case, John is focused in 'John ate the cake'. However, John is also given, mentioned in the immediately previous discourse. John is nonetheless focused, since, in this section of discourse, the presupposition is that somebody ate the cake and specifying John as the person who did so can be seen as the difference between the presupposition and the assertion of the utterance.

In summary, the focus of a clause is typically new information which is not presupposed. It is the information which is being added to the discourse and which the speaker wishes to draw the listeners attention to. The focus often stresses the possibility of alternatives and should reflect the *wh*-

word in the equivalent *wh*-question. As a result of these elements of focus, while foci are typically informationally new, they may also be accessible or even given.

As we shall see, a wide variety of sub-types of focus have also been identified. A key distinction made in Lambrecht's (1994) work is that between argument, predicate and sentence focus. This distinction can also be referred to as broad vs. narrow focus. Argument or narrow focus is focus exclusively on the argument, rather than on the entire predicate, which can be referred to as broad or predicate focus. Sentence focus on the other hand represents the case where an entire clause is in focus, in such a case the clause is referred to as thetic.

However, further distinctions can be made beyond this distinction of the domain of focusing. The most commonly made distinction is that between contrastive and information focus. Contrastive foci must be explicitly contrasted against another option in the foregoing discourse, I provide an example below:

(3) 'John and Clare came to the party.'
'Did Clare eat the cake?'
'No, John ate the cake.'

In this case, John represents a contrastive focus, since he is being explicitly contrasted against the other option presented in the question: Clare. Information foci on the other hand are more akin to Jackendoff's (1972) notion of focus: the new information communicated in a clause, complementing the presupposition. These are of the type illustrated in (1) above. In addition to these two broad types of focus several other are relevant in this thesis: mirative focus, presentative focus and exhaustive focus. Mirative foci are those foci which contain an element of shock or surprise, showing some element of speaker-oriented attitudinal meaning. I illustrate mirative focus with a French-language example from Authier and Haegemann (2019) below:

(4)	Plus	de	cent	lettres	(que)	j'ai	trouvées	dans	notre
	More	than	one-hundred	letters	(that)	I-have	found	in	our
	boîte!								
	mailbox								

'I've found more than one hundred letters in our mailbox!' (Authier and Haegeman, 2019: 15)

In (4) we can see that the number of letters 'more than one hundred' is unexpected and elicits a surprise response, as such, this can be seen as an example of mirative focus. Presentative foci can be seen as a subset of new information foci, those which occur in presentative clauses or similar contexts in which no presupposition is found. Again, I provide a French-language example to illustrate:

(5)	Ouvert	24	h	sur	24	il	y	avait	écrit
	open	24	hours	on	24	it	there	had	written

"'Open 24/7" it was written.' (Larrivée, 2022a: 12; originally from Blanche-Benveniste, 1996)

Finally, exhaustive foci are those foci in which the focused referent is the only referent for which the predicate is true. As we can see then, while foci all share the features of adding information to the discourse (although their referent need not be new to the discourse) and drawing attention to the possibility of alternatives, they may perform a number of more specific functions, typically relating to the degree to which other alternatives are excluded (exhaustive focus) or marked as surprising (contrastive and mirative focus).

It should be noted that a variety of different sets of terminology have been used for different focus sub-types. For instance, while we noted one use of 'broad' and 'narrow' focus above, Kiss (1998) considers these terms to reference the same distinction as contrastive and presentational focus or, in her terminology, identificational and information focus (Kiss, 1998: 245). This diversity of terminologies and the frequent willingness of researchers to coin new topic and focus sub-types can make research in the area rather difficult. I hope to have made my own use of these terminologies clear in the discussion above.

2.1.3 Topic

Having provided a definition of focus and discussion of its sub-types, I turn to the concept of topic. Just as foci typically consist of new information, topics typically consist of old information. Indeed, in the Prague School, the two notions: topic and informational oldness, were conflated. Since Reinhart's (1982) seminal work, topics are typically considered to be the information-structural constituent which the clause is 'about'. Further formalising this notion, De Cat (2007: 65) argues that

the topic is that with regard to which predication is evaluated. In summary, topics are typically given information, which represent that which the rest of the clause comments on. For instance, in our example (1) above 'he' is the topic, old information with regard to which 'ate the cake' is evaluated. A good diagnostic for a topic is that it can take the place of X in a rephrasing of the clause starting with 'As for X...' or any other periphrasis which explicitly draws attention to the notion of aboutness.

Again, as with the notion of focus, several sub-types of topics have been described in the literature. The three most important for our purposes are: shifting topic, familiar topic and contrastive topic. The shifting topic must differ from the topic of the previous clause. If it does not differ from the topic of the previous clause, it is instead a familiar topic. Familiar topics are used in particular to analyse large chunks of discourse in which the topic remains continuous. Finally, contrastive topics are those topics which are explicitly contrasted against the previous topic in the discourse.

A final notion is relevant here: frame-setters. Frame-setters are those discourse constituents, typically adverbials at the beginning of a clause, which provide a frame in which the proposition should be evaluated (Krifka and Musan, 2012: 31). In many cases, frame-setters seem to simply indicate the type of information which is to be given in this clause and restrict the set of possible interpretations to a given frame (see Krifka and Musan, 2012: 31-2). An example of this can be seen below in (6):

(6) [As for his health situation]$_{\text{Frame}}$, he had a bypass operation recently

(Example from Krifka and Musan, 2012: 32; annotation mine)

In summary, in this thesis, following Lambrecht (1994) and work in the cartographic syntax tradition, I focus my analysis on givenness, topics and foci as the core elements of information structure. This approach is only one view on information structure and discourse pragmatics. Although alternative approaches will not be used in our analysis, it is worth acknowledging the wide variety of functional, interactional and dialogic approaches to discourse pragmatics which depart from those analytic tools laid out above. Interactional approaches will feature in our discussion in Chapter 5 and Chapter 8. That said, our focus here is on applying Lambrecht's (1994) system and its subsequent

expansions to the data we analyse, with an implicit assumption that it will bear some relevance to these data.

2.2 Syntax

A variety of approaches to syntactic analysis of information-structural features have been taken in the literature. Much early work on information structure and syntax emerged from the functionalist tradition, especially in the work of Givón (1983). This tradition can be seen reflected in the more recent information-structural approaches of Erteschik-Shir (2007). However, since the late 1990s, particularly following the work of Rizzi (1997), a variety of generative approaches to the information structure-syntax interface have been undertaken. I focus here primarily on the cartographic approach of Rizzi (1997) and others, although even within the generative approach, other approaches are available (see, for instance, the minimalist feature-driven approach of López, 2009)

In Rizzi's (1997) approach, the left periphery of the clause is treated as featuring numerous syntactic positions specialised for different, often information-structural, functions. Specifically, Rizzi (1997) argues for, at the outermost level of the C domain, a ForceP which is 'outward-looking' where we find elements which modify the clausal type and certain complementizers (Rizzi, 1997: 283). Below ForceP is TopP in which extraposed topics are to be found. Below TopP is FocP in which extraposed foci are found. Below FocP is another, optional TopP. Finally, the lowest point of the left periphery is FinP, an 'inward-looking' phrase which expresses the finiteness of the clause and certain lower complementizers (Rizzi, 1997: 284). While Rizzi's (1997) original description remains influential, several variants have been suggested. Here I will be following Wolfe's (2021: 37) version of the hierarchy, since this is the most recent version provided for the historical study of French. It should be noted that this version draws heavily on that of Ledgeway (2010).

(7) [$_{\text{Frame}}$ Topic_{HT}, $\text{Adverbial}_{\text{Scene Setting}}$/$\text{Clause}_{\text{Scene Setting}}$ [$_{\text{Force}}$ Complementizer, Verb_{Fin}
[$_{\text{Topic}}$ Topic_{LD}, $\text{Topic}_{\text{Moved}}$ [$_{\text{Focus}}$ $\text{Focus}_{\text{Contrastive}}$, $\text{Quantifier}_{\text{Indefinite}}$, $\text{Focus}_{\text{Information}}$
[$_{\text{Fin}}$ Complementizer, Verb_{Fin} [$_{\text{TP}}$ . . .]]]]]]

Two significant differences from Rizzi's (1997) work should be noted. While Rizzi (1997) suggests that each section in the hierarchy is specifically a phrase, much work since has utilised the notion of 'fields', which may feature further, finer grained phrasal-level distinctions. This can be used, for instance, to account for the recursive nature of the Topic field. Additionally, the illustration shown above adds the Frame field at the highest edge of the left-periphery to host hanging topics and scene setters (another term for frame-setters). The notion of hanging topics will be discussed further below in our review of previous analyses of left dislocation structures. Finally, in this version of the expanded left-periphery, the lower Topic field is removed, since the topicality of these lower elements is debated.[2] It should be noted that, here and throughout, we will refer to the tense/inflection head and phrase as the T(P), not, as in much of the literature, the I(P).[3] Nonetheless, when discussing literature which makes use of the IP, that term may still be used.

Further expansions to this analysis have been provided by Frascarelli and Hinterhölzl (2007). Frascarelli and Hinterhölzl (2007) have argued for numerous further distinctions in the left periphery, particularly focusing on sub-divisions in the Topic field. Most notably, Frascarelli and Hinterhölzl (2007) propose three sub-divisions of the Topic field, a phrase, ShiftP, for shifting topics, below this, a phrase, ContrP, for contrastive topics and, finally, a phrase, FamP, for familiar topics. Importantly, familiar topics are taken to fill the specifier of a phrase below the Focus field, while all other topics are above. Crucially, in this analysis, ShiftP is not recursive, while ContrP and FamP may be recursive. Debates as to which phrases in the cartographic spine are or are not recursible have been the focus of much of the cartographic literature in the last ten years.

Numerous further debates have emerged in the wake of Rizzi's (1997) work, in many cases, I do not settle on one approach here, but will discuss the various options throughout this thesis. One debate in particular is highly relevant to us: that surrounding the *v*P periphery. Belletti (1998, 2001, 2004) and Poletto (2006) have argued for a low left-periphery in the *v*P periphery in order to account for information-structurally marked post-verbal subjects. Indeed, Poletto (2006) expands this analysis,

[2] It should be noted that this debate has largely been resolved, particularly in the wake of Rizzi's (2022) work supporting the lack of a lower topic field, in contrast to his previous work.
[3] In some of the literature these are two distinct phrases and heads, but in much of the literature, the two are used interchangeably.

arguing for peripheries of this kind across the different levels of the clause, at least across DP (following work by Giusti (2005, 2006)), *v*P and CP. On the other hand, Cardinaletti (2018) argues that post-verbal subjects remain in their VP-internal thematic position, which is information-structurally under-specified. The differing predictions of these two accounts will be discussed in Chapter 4.

In addition to these elements of syntax directly tied to information structure, the cartographic approach has been used to discuss fine-grained word order variation at further levels within the clause. In particular, studies of adverb ordering in the work of Cinque (1999, 2004) and work on the multiple subject and inflectional positions within the TP (Cardinaletti and Roberts, 2002; Cardinaletti, 2004; Wolfe, 2021, 2022) are relevant here. However, these approaches are of more limited relevance to our work. As such, they are discussed in more depth in the relevant chapters, most prominently in Chapter 4.

As well as making use of these particular cartographic syntactic hierarchies, we will also follow standard assumptions of modern cartographic minimalism throughout this thesis. We will follow the copy-theory of movement (developed in its modern form by Chomsky, 1993, 1995) and assume that movement is triggered by feature agreement. We will also make use of the Extended Projection Principle (EPP) in our assumed analysis of V2. Following Chomsky (2001), Roberts and Roussou (2002) and Roberts (2010) amongst others, we assume that an EPP feature may be distributed on any of a number of functional heads in a given language. In languages with a V2 grammar, there is an EPP feature localised on a head in the C-domain, leading to the Specifier of said head being filled, alongside verb-movement to the head, producing a surface V2 order. This approach notably differs from the classical EPP as introduced by Chomsky (1982: 10), which simply stipulates that a clause must have a subject. Under our view then, the EPP feature may simply be viewed as 'a diacritic for triggering movement' (Roberts, 2010: 29).[4] In addition to this use of EPP-features, we will make use of phi-features to account for verb and subject raising. These mechanisms are discussed in our analysis of post-verbal subjects in Chapter 4, while through the rest of the thesis the particulars of syntactic movement and its causes will not generally be discussed in our analyses.

[4] The concept of movement diacritics has been further developed in more recent work by Biberauer, Holmberg and Roberts (2014) and Holmberg and Roberts (2018) amongst others.

In addition to these particular theoretical assumptions, we will make use of recent work on fine-grained parametric variation, particularly using the notions of micro- and nano-parameters (Roberts, 2019). Micro-parameters are parameters which operate on the level of a small subclass of functional heads, while nano-parameters operate on the level of individual lexical items. These will be crucial for much of our discussion of fine-grained variation and change, particularly in Chapter 7.

Overall, a wide variety of different approaches have emerged in the development of the cartographic approach. Even just within the study of information structure and syntax, many different approaches have emerged in the 26 years since the publication of Rizzi's (1997) seminal monograph. While we have illustrated some specific developments above, I do not dogmatically follow one approach here. Rather, I use the general heuristics of the cartographic research program: fine-grained descriptive analysis of word order variation and a focus on a highly developed functional spine of the clause, to analyse our data and the information structure-syntax interface.

2.3 Syntax and Information Structure in the History of French

The cartographic approach to syntax and information structure has been widely applied across the history of French. Here, I provide a brief overview of the existing literature, going through each period chronologically. Overviews of the literature on specific constructions are provided in the appropriate chapters, while a more in-depth discussion of the claims made in the literature can be found in Chapter 7.

Old French has seen the most study of any historical period of French with regard to syntax and information structure. Throughout the 2000s and early 2010s a wide variety of claims were made regarding the information-structural nature of Old French word order. Once it had become widely accepted that Old French was a V2 language (Adams, 1987; Roberts, 1993; Vance, 1995; 1997, De Andrade, 2018; Wolfe, 2021; pace Rinke and Meisel, 2009; Sitaridou, 2012), debate began around the information-structural nature of the pre-verbal constituent. Some researchers claim that it is topical (Rinke and Meisel, 2009; Larrivée, 2021) and others that it is a focus (Zimmermann, 2014). However, in most more recent work, it is now agreed that the immediately pre-verbal constituent is information-

structurally under-specified (Salvesen, 2013; de Andrade, 2018; Labelle and Hirschbühler, 2018; Wolfe, 2021), with information-structurally specialised left-peripheral positions available in V3+ clauses. In addition to the study of the left-periphery, some discourse particles have seen extensive study in this period, particularly the particle *si*. Despite this extensive study, Old French is often studied in comparative isolation, rather than in the larger diachronic context of the history of French. The work of Roberts (1993) and Vance (1995, 1997) on the collapse of V2 are an exception to this, as is Wolfe's (2021) recent book on the history of French syntax, which puts Old French in appropriate diachronic context. Nonetheless, taking the existing study of the period into a wider diachronic context is one of the major goals of this thesis.

Once we turn to Middle and Early Modern French, we can see how, despite a rich literature on Old French, much of the history of French is under-studied when it comes to syntax and information structure. As noted above, the collapse of V2 in this period has been studied quite extensively, but this research typically focuses only on the Middle French period, not going beyond to Renaissance and Classical French. Once we reach the Renaissance and Classical periods there is a particular void in the literature. Recently, the GGHF has provided some information on the period, but this is largely descriptive. Similarly, prior to the publication of the GGHF some other descriptive work was available, particularly the work of Gougenheim (1973) and Fournier (2001, 2009). As well as these works, Roberts (1993) undertakes a major diachronic study of post-verbal subjects but focuses only on their use in interrogatives. In addition, Dufter (2008, 2009) provides an in-depth analysis of clefts from Old and Middle French until the 18th century. These papers will be discussed in more depth in the background sections of their respective chapters. To my knowledge, the only work of diachronic theoretical syntax on the period is Wolfe's (2021) book. While this work provides a strong basis for further work, it is also broad in scope, leaving much space for more fine-grained analysis of individual constructions.

While the Early Modern period has seen an astonishing lack of syntactic research, Modern and Contemporary Spoken French have been heavily studied. In particular, numerous fine-grained studies of individual constructions have been undertaken, most notably De Cat's (2007) work on

dislocations and Lahousse's (2003, 2006, 2022) work on post-verbal subjects (both to be discussed in detail in their respective chapters). However, little work has been undertaken to develop an analysis of the general relationship between information structure and syntax in Modern French. Additionally, the overwhelming majority of work on the Modern and Contemporary language is synchronic in nature. It is my opinion that synchronic study of Modern languages will benefit from diachronic and historical context.

Thus, while historical work, particularly on Old French, is characterized by a focus on the role of information structure in declarative main clauses, work on the Modern and Contemporary language is characterized by fine-grained analysis of particular constructions. In this thesis, then, I intend to synthesize these two approaches, providing a fine-grained analysis of three different constructions in declarative main clauses which I then analyse within the broader context of the history of the language.

2.4 Text Typology and Historical Sociolinguistics

While a variety of works have been undertaken on the syntax-information structure interface in French, particularly Old French, a novel element of this thesis is our inclusion of text-typological data and historical sociolinguistic methodologies. Thus, I now provide a brief overview of text typology and its role in historical sociolinguistic analysis.

I begin with a discussion of text typology. Trosborg (1997) distinguishes three central concepts in text typology: register, genre and text-type. Register divides texts depending on the field which they are a part of, for instance the language of religion, legal documents, medical language etc. (Trosborg, 1997: 3). Genre on the other hand is defined by communicative purpose (Trosborg, 1997: 5-6) and forms sub-divisions of registers, for instance the scientific register could include diaristic field reports, formalised academic writing etc. It should, however, be noted that one genre can be used for multiple registers and vice versa. Finally, text-types are defined by the structuring of a text, for instance narrative, expository, argumentative etc. Importantly, one text can contain segments of different text-types. For our purposes, we will be focusing exclusively on register and genre, since text-

type variation is often intra-textual. Importantly, all of these words are routinely used imprecisely in the literature, for instance register is often used to refer to something approximately the same as formality, with 'high register' referencing more formal language use and 'low register' referencing more colloquial language use. Despite the technical use discussed by Trosborg (1997), it is this use of register, referring to formality, which I will use throughout the thesis, using genre to refer to both the register and genre elements of text typology.

Variation in register and genre have been shown to affect syntax in the history of French. Larrivée (2022b) shows that register (in the sense of formality) affects the frequency of surface-V2 clauses in late 13th-century French, with higher rates in higher register texts. Simonenko, Crabbé and Prévost (2018) show that the syntax of verse texts in Old French show a structured lag behind that of prose texts. Donaldson (2014) shows that representations of dialogue often show more advanced syntax than that of narrative text and similarly, Mazziotta and Glikman (2019) find that rates of different parts of speech vary between dialogue and narrative.

Text-typological work often underpins historical sociolinguistics work, in which certain texts are examined as being more "spoken-like" (see e.g. Elspaß, 2012). Historical sociolinguistics, then, can be seen as the attempt to access the spoken language of past periods by examining more varied data, particularly data from more colloquial registers and writing from less educated speakers. Elspaß (2007, 2012) has, in particular, pioneered the notion of language history from below, in which the focus is shifted away from hegemonic, standard-language texts and linguistic data, towards the available data on the language of working and lower-middle class people. In particular, historical sociolinguistic approaches often focus on so-called ego documents: auto-biographical texts including letters, diaries and memoires. Private diaries can be of particular value since they may have no intended reader, leading to a more candid and unmonitored linguistic style. By shifting the focus towards data closer to the spoken language, historical sociolinguistics gives us a fuller picture of the history of a language, allowing for more fine-grained analysis of language change.

Although much work in historical sociolinguistics has been conducted on the Germanic languages, in particular Elspaß's (2007) pioneering work on language history from below and recent

work on letters as historical linguistic data in Dutch (Rutten and van der Wal, 2014), historical sociolinguistic methodologies have also been fruitfully applied to the history of French. We have already seen the application of text-typological data to historical syntactic analysis of Old French; Ayres-Bennett (2004) has also conducted extensive historical sociolinguistic work on the Early Modern period, particularly focusing on the development of prescriptivism and the impact of standardisation in that period. Additionally, much work has been conducted on the historical sociolinguistic origins of Canadian French varieties (Martineau and Mougeon, 2003; Martineau, 2005). It should be noted, however, that both historical and modern sociolinguistics typically focus on phonetic and lexical variables. Indeed, it is only recently that the field of so-called socio-syntax has begun to emerge (Christensen and Jensen, 2022). Thus, I hope to break new ground by applying historical sociolinguistic methodologies to syntactic data at a broad timescale.

2.5 Towards a Cohesive History of Information Structure, Syntax and Sociolinguistics in French

It is clear that a wide variety of factors have intertwined in the history of French syntax. The goal of this thesis, then, is to synthesize the variety of approaches previously taken to the history of French syntax and information structure to achieve a more cohesive picture of this history.

It is the contention of this thesis that integrating the insights of formal syntactic approaches to linguistic history with historical sociolinguistic techniques will allow us to reach a more total understanding of the history of French syntax as well as providing insights for all work in historical linguistics. While historical sociolinguistics provides us with fine-grained variation in data and tools for describing this variation, cartographic syntax provides us with the formal machinery to analyse and explain this kind of fine-grained syntactic variation, making them two well-matched methodological approaches.

2.6 Case Studies

Having given an overview of the existing literature on the relevant theoretical topics, I turn to the subjects of my case studies. Full overviews of the literature on each structure are provided in the relevant chapters. In this section, I simply provide a brief descriptive overview of the constructions which form the content of this thesis.

2.6.1 Post-Verbal Subjects

I begin with a discussion of post-verbal subjects, the subject of Chapter 4. While these structures have traditionally been described as subject-verb inversions, this label is now widely viewed as being misleading and theoretically vacuous, as such I refer to them as post-verbal subjects, although specific constructions may still be referred to as the appropriate sub-type of inversion. In Modern French, three kinds are found:

(8) a. Quand est- elle partie? (pronominal inversion)
When is she left?

'When did she leave?'

b. Quand Claire est- elle partie? (complex inversion)
When Claire is she left?

'When did Claire leave?'

c. Quand est partie Claire? (stylistic inversion)
When is left Claire

'When did Claire leave?'

In the Modern language, these constructions are rare and limited to higher registers. Historically, however, both pronominal and stylistic inversion were considerably more frequent, while complex inversion only developed in the 15th century. Additionally, other structures existed, such as simple inversion, in which the DP subject follows the auxiliary but not the main verb, rather than following both as in Modern French stylistic inversion:[5]

[5] Note that, although non-standard diacritics have been removed, historical texts and those written by *peu-lettrés* have not been normalised and may show a variety of non-standard spellings.

(9) et sont **moult de gent** mis a mort et a pechié.
and are many of people put to death and to sin

'and many people turned to death and sin.'

Somme Royale, 13th century

In Old and Middle French, simple inversion and pronominal inversion are seen as evidence for V2 (Roberts, 1993; Vance, 1995; Vance, 1997). These structures have been analysed extensively in the Modern language as well as in Old and Middle French. However, little research has been undertaken on the full history of post-verbal subjects, limited to Roberts' (1993) work, which focuses exclusively on their use in interrogation.

Two central points of debate emerge from the literature. Firstly, there is debate as to the exact nature of the post-verbal pronominal subject in the Modern language, be it an inflectional morpheme (Roberts, 2009), a syntactic clitic (Rowlett, 2007), a phonological clitic (Rizzi and Roberts, 1989), or a full pronominal subject (Wolfe, 2021, 2022). As such, in our chapter, we discuss this debate from a diachronic perspective, examining the development of post-verbal pronominal subjects alongside the development of complex inversion as V2 is lost. I argue that the post-verbal pronominal subject becomes a syntactic clitic in the Early Modern period, allowing it to survive the loss of V2 while simple inversion is lost.

Secondly, the exact position of the post-verbal DP subject is a point of considerable debate in the existing literature (see Belletti, 2001; Lahousse, 2003, 2006; Cardinaletti, 2004; Ordóñez, 2007). As such, locating the position of the lower subject is another focal point of Chapter 4. I utilise a variety of cartographic techniques, including examining the relative position of the subject and object, subject and adverbs and the information-structural role of the subject to diagnose this position.

In summary, in Chapter 4, I apply a fine-grained cartographic approach to the full history of post-verbal subjects in declarative main clauses, investigating their variation and change across the history of French.

2.6.2 Dislocations

I next give a brief overview of dislocations, the focus of Chapter 5. Dislocations are those constructions in which a constituent surfaces at the far left (left dislocation) or right (right dislocation) of the clause, resumed within the clause itself. Both cases are exemplified below:

(10) Enfin	**l'**	**école**	**à**	**aires**	**ouvertes**	**c'**	est	une	école
In fact	the	school	with	areas	open	it	is	a	school

expérimentale. (left dislocation)
experimental

'Well, the open-air school, it's an experimental school'.

(11) On **le**	kiffait	bien	**le**	**troisième.** (right dislocation)	
We it	liked	well	the	*troisième*	

'We liked the *troisième*.'

While both left and right dislocation have been extensively studied in the Modern and Contemporary language, they have seen relatively little study in historical data, largely limited to the work of Mathieu (2012) and description in the GGHF. Given the extremely high frequency of dislocations in Contemporary Spoken French, it is surprising that little research has been devoted to analysing their history and origin. There is widespread agreement in the literature that dislocations are used to mark topic (Lambrecht, 1981; De Cat, 2007), however there is some debate as to the syntactic analysis of these constructions, whether the dislocated elements should be treated as moved (Angelopoulos and Sportiche, 2021) or base-generated in the periphery (De Cat, 2007). Additionally, some recent work has argued for further interactional functions for dislocations in Spoken Contemporary French, beyond the traditional topic analysis (Doehler, de Stefani and Hörlacher, 2015).

In Chapter 5, then, I provide an analysis of the history of dislocations, finding an unusual stability in their structure, despite recent increases in frequency. A considerable focal point of this chapter is the role of the spoken-written divide in determining the frequency of dislocations, with written Contemporary French, even the most spoken-like writing, showing much lower frequencies than Contemporary Spoken French.

2.6.3 *C'est* Clefts

Finally, I give a brief overview of *c'est* clefts, the focus of Chapter 6. *C'est* clefts are a complex construction, in which an introductory matrix clause, featuring an expletive pronominal and a copula, separates a clefted element, which may be a DP as well as several types of adverbials, from a subordinate or pseudo-subordinate relative clause. This can be seen illustrated below:

(12) C' est Jean qui est venu.
It is John who is came

'It's John who came.'

Our chapter focuses only on *c'est* clefts, although other cleft sub-types exist. There is some agreement in the literature that clefts typically mark focus, however some other more marginal functions are available, discussed at length in Chapter 6. Clefts have been studied thoroughly in French, including historically. They are frequent in the contemporary language, but rare historically. Traditionally, the increase in frequency in the Modern language was argued to be in response to the loss of other focusing constructions in the Early Modern period. However, Dufter (2008, 2009) has argued against this view, arguing instead that the increase in frequency of clefts is due to an expansion in function to include non-focusing clefts and adverbial clefts. Thus, it is a focal point of Chapter 6 to test Dufter's (2008, 2009) claims.

With regard to structure, clefts have proven extremely difficult to analyse, with a wide variety of analyses in existence (Meinunger, 1998; Belletti, 2008, 2012, 2015; Frascarelli and Ramaglia, 2009, 2013; Reeve, 2011, 2012; Wehr, 2011, 2015). Most notably, these analyses make wildly different assumptions about the nature of clefts, with some arguing that *c'est… que…* is a discontinuous morpheme (Wehr, 2011, 2015), others applying a monoclausal analysis (Meinunger, 1998) and still more applying a biclausal analysis of the construction (Belletti, 2008, 2012, 2015; Reeve, 2011, 2012). Although I present some tentative analyses in Chapter 6, I also argue that clefts may present a substantial challenge to cartographic approaches, since their internal structure can be seen as somewhat conventionalised and opaque, especially in cases where clefts have started to become lexicalised. Cartographic analysis is well-suited to analysing the internal structure and dependencies of

a clause, but once these structures are highly conventionalised, it is difficult to apply this type of analysis.

2.6.4 Summary

In summary, I approach all three case studies with a fine-grained cartographic approach, attempting to analyse the exact derivation of each construction through its history and how these constructions change through their histories. Throughout, I take into account both internal pressures, such as changes in the overall structure of the clause, and external pressures, such as the impact of prescriptivism and text-typological factors. I hope that by analysing an inclusive, wide-spanning dataset and utilising fine-grained analytic techniques, new light will be shone on the history of French syntax and the nature of syntactic change more broadly.

3. Methodology

I turn now to my own methodology, focused on addressing the issues raised in Chapter 2. The specific methodologies utilised in each case study are detailed in the relevant chapters. Here I present an overview of the data gathered and rationale behind it.

3.1 Frequency Data

In line with the historical sociolinguistic approach taken here, I have endeavoured to select data from as wide a set of sources as possible.

For Old and Middle French, I have used the MCVF (Martineau, Hirschbühler, Kroch and Morin, 2021) and Penn Supplement (Kroch and Santorini, 2021) parsed corpora,[6] which include a wide variety of Old and Middle French texts, both verse and prose in a variety of registers and genres. In each case, the full MCVF and Penn Supplement corpora have been searched utilising CorpusSearch[7] to generate frequency data for the relevant constructions. The particulars of each automated search are discussed in each individual case study chapter, and all queries used are provided in Appendix C.

In addition to the MCVF and Penn Supplement data, I have supplemented my Old and Middle French data with data from the *Correspondance de la famille Estouteville*: a collection of letters from a lower-middle class family in the 15th century, available through the EPELE project.[8] This data has been hand-searched, with the first 400 clauses examined for examples of the constructions in our case studies.

Turning to Early Modern French, no parsed corpora akin to the MCVF and Penn Supplement are available. As such, texts from the 17th-19th centuries were manually selected, a total of 16 texts. In order to ensure a variety of registers and mitigate against the effects of prescriptivism, several different genres were examined. Firstly, novels were examined as examples of standard, high-

[6] For more details, see the corpus documentation: https://github.com/beatrice57/mcvf-plus-ppchf

[7] See website for more details: http://corpussearch.sourceforge.net/CS.html

[8] For more information, see the EPELE website: https://www.unicaen.fr/epele/accueil

register language, likely to be broadly analogous to prose literary texts from earlier periods. Secondly, travel writing was examined. These texts were selected as an intermediary register. Written for publication, they were likely to adhere to the standard, but nonetheless as non-fictional texts, sometimes with diaristic elements, they would not have the same artistic flourishes found in novels. Finally, private *journaux intimes* (diaries) and published memoires were examined. These texts, particularly the private ones, were selected so as to capture language closer to speech which might evade the effects of prescriptivism. Texts from lower class writers and the so-called *peu lettrés* were prioritised. Printed editions were invaluable for private texts, particularly Ernst's (2018) edition of private texts from the 17th and 18th centuries. Novels and travel texts were accessed through Gallica, the Bibliothéque Nationale de France's portal for digitised texts. In addition, two 'spoken-like' texts were examine. Firstly, Héroard's record of the Dauphin's speech, dated from the first decade of the 1600s, in an edition provided by Ernst (1985). Secondly, letters from low-rank soldiers in the Guerre de la Vendée were used in an edition from Collet (2021). It should be noted that this is an edition intended for popular historical readers and is not an unedited linguistic edition. Nonetheless, it seems that the syntax of the letters is unedited in this text. I hoped that by examining texts which may be closer to speech, the issues of using textual data would be somewhat mitigated. As with the *Correspondance Estouteville* each of these texts had its first 400 clauses searched for examples of our case study constructions, which were then analysed further, as detailed in their respective chapters.

Finally, for my Modern and Contemporary French data both published texts and corpora were examined. From the 20th century, correspondence corpora from lower class authors are available. World War 1 soldiers' letters were taken from Steuckardt (2015). From the 1960s speech corpora become available. The ESLO corpus[9] was used for 20th century Modern French speech, while the CFPP2000 corpus[10] was used for Contemporary Spoken French. Finally, the 88MilSMS corpus[11] of text messages and YCCQA corpus (De Smet, 2009) of online forum interactions were used for written Contemporary French. In addition to these new media, unavailable in previous centuries,

[9] See website for more details: http://eslo.huma-num.fr/

[10] Branca-Rosoff et al. (2012), see website for more details: http://cfpp2000.univ-paris3.fr/index.html

[11] Panckhurst et al. (2014), see website for more details: http://88milsms.huma-num.fr/index en.htm

travel writing and diaristic writing were also examined for the 20th century, so as to maintain comparability with prior centuries. For a full list of all the texts and corpora used, see the appendices.

Throughout, examples taken from the MCVF and Penn Supplement will be given with the corresponding reference code from that corpus (rather than referencing line or page numbers of the original texts), examples from the ESLO and CFPP2000 corpora will be given with the appropriate timestamp and examples from the 88MILSMS and YCCQA corpus will be provided with their reference codes. Examples from non-corpus texts on the other hand will be provided with page, chapter, or section references, with a preference for referencing original sections rather than modern editions. For instance, examples from the texts in Ernst's (2018) edition will be provided with manuscript page numbers, not the relevant page number of Ernst's (2018) book. The texts themselves will be referred to with either a (sometimes shortened) form of the title or the author's surname, the appropriate shortenings are listed alongside the texts in Appendix B. For readability, this has not been followed in the introduction, where clearer references to text and approximate dating have been used instead.

In summary, I have endeavoured to gather a variety of texts from each period, covering both more and less colloquial contexts in a variety of genres, so as to maintain comparability. Most crucially, I will compare Contemporary Spoken French and written colloquial Contemporary French, rather than comparing the spoken contemporary language directly to written historical registers. I hope that this will avoid the frequent problem in historical linguistics of comparing text and spoken language, since these two modalities have such fundamentally different features. While technologies of text production change immensely over the history of French, such that direct comparison always comes with caveats, by maintaining a variety of types of data across each period, I hope that no variation will be erased, and we will achieve a full picture of the history of French syntax.

3.2 Metalinguistic Data

In addition to the examples of our case study constructions gathered directly, I have also analysed metalinguistic data. In our discussion above, we saw that a considerable issue develops in the Early

Modern period. The development of standardisation and prescriptivism leads to an increasing gap between written and spoken language. Those texts which we do have from the period are often higher register and likely to be imitating the prescriptive standard rather than representative of actual speech. These issues are exacerbated with the foundation of the Académie Française in 1635 and the concurrent development of more overt, state-sponsored prescriptivism. The influence of this prescriptivism was pernicious and pervasive. In his memoires, the peasant turned librarian Jamerey-Duval specifically notes that, in his childhood, he was initially educated in 'mauvais françois' 'bad French' (Goulemot, 1981: 112). This suggests that even lower-class speakers are aware of the sociolinguistic situation and will imitate the standard in their writing.

As outlined above, I have endeavoured to overcome this issue by gathering data from a wide variety of texts including those from the so-called *peu lettrés*. In addition to this, however, I have gathered metalinguistic data from a number of Early Modern grammarians. I have examined three 17th-century prescriptivist commentators (Maupas' Grammaire et syntaxe Françoise (1618), Vaugelas' Remarques sur la langue Françoise (1647) and Bouhours' Remarques nouvelles sur la langue Françoise (1692)). In each case, I have looked for any mention or example of our case study constructions so as to further understand the effects of nascent prescriptivism on word order. Additionally, I have examined the texts for general comments on the ordering of constituents, for example regarding 'clarity'.

I hope that, by supplementing my raw data with metalinguistic data of this period, a clearer image of the processes of syntactic change at play will emerge.

3.3 Summary

In summary, I have taken a wide-spanning, sociolinguistically informed approach to data gathering. My frequency data are drawn from a wide variety of texts, which I hope will display a wide range of linguistic variation in the history of the language. Meanwhile, I have supplemented this approach with metalinguistic data to mitigate against the effects of prescriptivism in the Early Modern period. This

approach will allow us to achieve an extremely detailed understanding of syntactic variation and change throughout the history of French.

4. Post-Verbal Subjects

4.1 Introduction

This chapter is concerned with the history of post-verbal subjects in French. Post-verbal subjects have been studied extensively over the last 30 years, in Modern French (Lahousse, 2003, 2006, 2022), Old and Middle French (Adams, 1987; Vance, 1995, 1997; Wolfe, 2021) and in intervening periods (Roberts, 1993; Wolfe, 2021, 2022). In the Old and Middle French periods, post-verbal subjects have also been heavily studied for their relationship to V2 (Adams, 1987; Roberts, 1993; Vance, 1995, 1997; Wolfe, 2021). Looking beyond French, an extensive literature has built up on post-verbal subjects in Romance more generally (Belletti, 1998, 2001; Cardinaletti, 2004; Ordóñez, 2007; Sheehan, 2010; Leonetti, 2014). However, very little large-scale work on the changes to post-verbal subjects over the entire history of French has been undertaken. Additionally, that which has been undertaken has typically focused only on either pronominal or DP subjects, and particularly on interrogative contexts.

In this chapter, I provide an account of the changes to post-verbal subjects in the history of French, focusing exclusively on declarative main clause contexts. We will see that this large-scale historical approach leads us to a detailed typology of subject positions which will be useful not just in the study of French but in the study of Romance more generally, particularly diachronic studies. I begin with an overview of previous studies on post-verbal subjects both in French specifically, and in Romance more broadly. Next, I provide details of the methodology specific to post-verbal subjects, before turning to data presentation. I provide descriptive data and formal analysis for each period in its own section. Regarding post-verbal pronominal subjects, I argue that, in Old and Middle French, they represent the same structure as so-called simple inversion. During the Early Modern period, however, post-verbal pronominal subjects cliticize, leading to the development of complex inversion. I provide an analysis of post-verbal pronominal subjects from Early Modern French onwards following Rizzi and Roberts' (1989) and Wolfe's (2022) analyses. Regarding post-verbal DP subjects, on the other hand, I argue that, while the traditional split between free and simple inversion has been fruitful for discussing the loss of V2 in French, a finer grained typology of subject positions exists. Following work from Ordóñez (2007) on Spanish, I elaborate on these positions and their historical

availability, particularly during the Early Modern period of the 16th-18th centuries. I provide an analysis of these various different subject positions in Old, Middle and Early Modern French, before discussing Lahousse's (2003, 2006, 2022) descriptive work on Modern French stylistic inversion against this historical backdrop.

4.2 Background

4.2.1 Post-Verbal Subjects in French

I begin with some background information on post-verbal subjects through the history of French. In Modern French, three types of post-verbal subject construction are found: pronominal inversion, complex inversion, and stylistic inversion, all shown below in (13).

(13) a. Quand est- elle partie? (pronominal inversion)
When is she left?

'When did she leave?'

b. Quand Claire est- elle partie? (complex inversion)
When Claire is she left?

'When did Claire leave ?'

c. Quand est partie Claire? (stylistic inversion)
When is left Claire?

'When did Claire leave?'

Stylistic inversion is marginal in main clause declaratives, either fulfilling an extremely narrow information-structural function of exhaustive focus, or triggered by a specific, limited set of pre-verbal adverbials, adjectivals and certain verbs (Lahousse, 2003, 2006). Pronominal and complex inversion are found somewhat more frequently in declaratives, triggered by a set of initial adverbs including *aussi* and *encore* (see Rowlett, 2007: 200; Smith, 2016: 310, for details). If we turn to Old French, the situation is considerably different. Three types of post-verbal subject construction are found: simple inversion, now extinct, free inversion, a precursor of Modern French stylistic inversion and pronominal inversion, which appears to function as in Modern French. These are illustrated below:

(14) a. Or leuet **om** l' espouse (pronominal inversion)
Or woke up one the husband

'We woke up the husband.'

(1190-SBERNAN-BFM-P,112.3724)

b. E si furent ocis **li** **sarrazin**. (free inversion)
And *si* was killed the Saracens

'And the Saracens were killed.'

(122X-PSEUDOTURPIN-MCVF-P,271.221)

c. Et sont **moult** **de** **gent** mis a mort et a pechié. (simple inversion)
And are many of people put to death and to sin

'And many people turned to death and sin.'

(1279-SOMME-ROYAL-PENN-P,1,44.1177)

While Adams' (1987) work was the first to develop the analysis of V-to-C movement in Old French, the differences between different post-verbal subject positions are not discussed. Vance (1995, 1997) observes that, in pronominal inversion, the subject always occurs between the two parts of the verb, leading her to analyse the subject as obligatorily occurring in Spec-TP with the verb raised higher to C. In Modern French, pronominal inversion and complex inversion are often analysed as either cases of subject pronominal cliticisation or affixation (e.g. Boeckx, 2001; Cardinaletti, 2004; Rowlett, 2007: 206; Roberts, 2009) to the verb, with a pre-verbal null subject. However, it should be noted that some Modern French analyses still suggest that pronominal inversion involves V-to-C movement (see Rizzi and Roberts, 1989; Roberts, 2009) or some other form of verb movement to an intermediary position above T but below the C-domain (Wolfe, 2022). These different accounts are discussed further below in section 4.2.2. As for DP inversion, Roberts (1993) distinguishes the free and simple DP inversion structures on the basis of the position of the subject DP. In free inversion the subject DP remains in a low position, thus following both parts of the verb. In simple inversion, on the other hand, the subject DP occurs between both parts of the verb. In Roberts' (1993) analysis, free inversion represents the case in which the subject DP remains in its initial position in Spec-VP. On the other hand, in simple inversion, the subject DP raises to Spec-TP while the verb is raised higher to C. This

distinction is illustrated in the bracketing in (15), adapted from Roberts' (1993: 173) illustration of post-verbal subjects in *wh*-questions to represent main clause declaratives:

(15) Free Inversion: [$_{CP}$ C [$_{TP}$ pro ... V+Agr [$_{VP}$ ~~V~~ ... DP]]]

Simple Inversion: [$_{CP}$ V [$_{TP}$ DP ....]]

Turning to the intervening periods, the Early Modern period, particularly the 15th and 16th centuries are key in the history of post-verbal subjects. Complex inversion is innovated during this period, first attested during the 15th century (Roberts, 1993: 189). Stylistic inversion on the other hand develops from free inversion, while simple inversion becomes extinct, at least in interrogatives, in the 16th century (Roberts, 1993: 190). At this same time, the V2 system of Old and Middle French is fully lost, as are null subjects (Wolfe, 2021). These historical details will be discussed in more detail in section 4.2.3.

4.2.2 Formal Analyses

I next turn to a discussion of formal analyses proposed for post-verbal subjects. The majority of the theories discussed in section 4.2.2.1 have been proposed for Romance languages other than French, thus they must be adapted to our own data, rather than taken uncritically.

4.2.2.1 Post-Verbal DP Subjects

In general, post-verbal subjects in Modern Romance languages follow both the main verb and the auxiliary, as in Old French free inversion. This is usually analysed as the subject occupying a low post-verbal position below T (Belletti, 1998, 2001; Cardinaletti, 2004; Ordóñez, 2007). While analyses of Old and Middle French free inversion often suggest that the subject remains in its base position in the VP (e.g. Roberts, 1993: 173), there is debate in the broader Romance linguistics literature as to where exactly the lower subject position is. In some accounts of Modern Romance languages, the subject simply remains in its base-generated position in Spec-VP (e.g. Cardinaletti, 2004). Additionally, in some early accounts, low subjects are treated as cases of right adjunction (e.g. Rizzi, 1982: 132) or a right Specifier of the VP (e.g. Giorgi and Longobardi, 1991). However, since the mid-1990s both right adjunction and rightward specifiers have been disallowed in most generative syntax. In order to

continue to fulfil the EPP, some accounts (e.g. Laenzlinger and Soare, 2005), argue that the VP must be fully evacuated in all clauses and thus that the subject is in a higher position but one still lower than TP. Indeed, this seems to be the consensus amongst most researchers (see Belletti, 1998, 2001; Laenzlinger and Soare 2005; Ordóñez, 2007), with debate as to the exact position which the post-verbal subject fills, and some suggestions that the subject may occupy a variety of positions along an expanded clausal spine (Laenzlinger and Soare, 2005: 17). For brevity, I do not here provide a full description of all described positions, focusing on the most relevant modern analyses.

I first turn to Belletti's (2001) analysis. In Belletti's (2001) analysis of Italian free inversion, the post-verbal subject is argued to be a focus, found in the specifier of a low focus projection directly above the VP. Indeed, Belletti (2004) further extends this analysis, arguing also for a low, clause-internal topic position. Poletto (2006) has also developed this analysis, arguing for both a focus and topic projection in the *v*P periphery (Poletto, 2006: 285-286). Belletti's (2001) analysis also disallows VSO orders with a focused subject, since the object could not be licensed due to Relativised Minimality, with the subject blocking object licensing. VOS orders on the other hand are analysed as examples of remnant movement, with the subject extracted to its focused position, as in other free inversion examples. The full remaining VP is then moved to the left-dislocated topic position, as supported by the fact that all of the objects in VOS orders are topical (Belletti, 2001: 70-71). These two analyses are illustrated in the schematised bracketings below, (16a) represents the VS case, while (16b) represents a VOS structure:

(16) a. [$_{FocP}$ [subject] [$_{Foc}$] [$_{VP}$ ~~subject~~ [$_{V'}$ [$_{V}$ verb] [$_{DP}$ object]]]]

b. [$_{TopP}$ [$_{VP}$ [~~subject~~ [$_{V'}$ [$_{V}$ verb] [$_{DP}$ object]]] [$_{Top'}$ … [$_{FocP}$ [subject] [$_{Foc'}$ [$_{Foc}$] [~~VP~~]]]]]]

An important part of Belletti's (2001) approach is her consideration of the role of the object in different word orders, while previous analyses of French data traditionally focus simply on the relationship of subject and verb (e.g. Roberts, 1993; Rowlett, 2007; Wolfe, 2021). By applying this more fine-grained approach to historical French data, I hope to be able to deepen our understanding of post-verbal subjects, the clausal syntax of French and its subject positions in general.

It is important to note that there is much variation in Modern Romance languages with regard to the position of the overt object relative to the verb. As we have seen, VOS orders are marginal but grammatical in Italian, assuming certain information-structural conditions are met (Belletti, 2001: 68). VSO orders on the other hand are disallowed (Belletti, 2001: 66-67). However, Spanish, Catalan, Portuguese and Romanian allow both VOS and VSO orders (Zubizaretta, 1998; Laenzlinger and Soare, 2005; Ordóñez, 2007). On the basis of this variation, as well as variation of the lower subject position with regard to adverbials, Ordóñez (2007) has suggested two separate lower subject positions in Spanish: the focus position identified by Belletti (2001), and a higher position, used in VSO orders, which does not display the same information-structural requirements of the lower position (Ordóñez, 2007: 264-5).

Laenzlinger and Soare (2005) take this approach to its extreme, suggesting a recursive structure of SVO chunks along the clausal spine between each of Cinque's (1999) adverbial projections. This would of course allow us to account for the variation between the Italian and Spanish data, however, it may also over-generate.

With regard to Modern French, Lahousse (2003, 2006, 2022) has provided in-depth, fine-grained descriptions of stylistic inversion. She argues that there are two sub-types of stylistic inversion. Firstly, genuine stylistic inversion, in which pre-verbal material triggers inversion. Secondly, focus inversion, in which the subject must be exhaustively focused. In genuine stylistic inversion the subject is argued to remain fully VP-internal while all other material moves to its relevant higher position. Evidence for this is adduced from the position of floating quantifiers, which cannot be found in genuine stylistic inversion, leading to the conclusion that the subject must be in a low position (Lahousse, 2006: 445). An example is given below (Lahousse, 2006: 445, translation and formatting mine):

(17) a. *Alors sont tous arrivés les enfants. . .
Then were all arrived the children. . .

(intended) 'Then all the children arrived.'

b. Alors sont arrivés tous les enfants.
Then were arrived all the children.

'Then all the children arrived.'

In (17a) we see that placing the floating quantifier in a higher position is ungrammatical. Floating quantifiers are typically analysed as evidence of a deleted copy. Thus, if the floating quantifier is not grammatical, it suggests that the subject cannot have moved through this position prior to remnant movement, as would be necessary for the remnant movement analysis of Kayne and Pollock (2001). As such, Lahousse (2006) argues for a base generation analysis. This is shown in a schematic bracketing below:

(18) [$_{TP}$ XP [$_{T}$ verb] [$_{VP}$ subject [$_{V}$ ~~verb~~]]]

In focus inversion, on the other hand, the subject is argued to raise to the Spec-FocP position in the clausal left periphery, while the rest of the clause then undergoes remnant movement to Spec-TopP. Focus inversion is exemplified, alongside a schematic bracketing, below.

(19) a.

Ont	mangé	une	pomme	**seules**	**les**	**filles**
Have	eaten	an	apple	only	the	girls

'Only the girls have eaten an apple' (Lahousse, 2003: 221)

b. [$_{TopP}$ [$_{TP}$ … verb + complements + adjuncts …] [$_{Top}$] [$_{FocP}$ subject[$_{Foc}$] ~~TP~~]]

Importantly, focus inversion may lead to V1 clauses and VOS orders, while genuine stylistic inversion may not. This analysis thus differs from Belletti's (2001) analysis in not utilising the low, VP-peripheral focus position. Following these analyses, Lahousse (2022) has further argued that some cases of genuine stylistic inversion, namely those triggered by [+anaphoric] or [+scalar] adjectives and adverbials, are both cases of the Romance phenomena of resumptive preposing and mirative focus fronting, suggesting that she believes the subject occupies a similar position to other Romance languages. However, she does not elaborate on the position of the subject in this analysis, rather providing an analysis of potential triggers. Importantly, Lahousse (2022) supports the view that Modern French post-verbal subjects maintain the same syntax as the other Romance languages but are simply more restricted in the contexts which license them.

Both Belletti's (2001) and Lahousse's (2003, 2006) analyses of post-verbal subjects provide a strong starting point for the analysis of post-verbal subjects. In examining our own data, the main distinguishing feature will be the distribution of floating quantifiers and the information-structural roles of the post-verbal subjects, which can help us distinguish between a remnant movement analysis and a VP-periphery analysis.

4.2.2.2 Post-Verbal Pronominal Subjects

Having described several analyses of post-verbal DP subjects, I turn to post-verbal pronominal subjects. Traditionally, post-verbal pronominal subjects in Old and Middle French are taken to be surface manifestations of the V2 constraint (see Adams, 1987, Vance, 1995, 1997), with the main verb raising past the pronominal subject in its canonical position, akin to simple inversion. In this analysis, pronominal subjects must always occupy the specifier of TP, while DP subjects may occur in lower positions, as in cases of free inversion. Thus, any case of a post-verbal pronominal subject in this period is a case of V-to-C movement, with the pronominal subject in Spec-TP (or Spec-AgrP in Vance's (1995, 1997) original formulation) surfacing post-verbally.

An analysis with V-to-C movement is further supported by Rizzi and Roberts (1989) for Modern French. However, here, pronominal subjects are analysed as cliticized to the verb, with the pronominal subject in Spec-TP left-incorporating to the verb in the C head. This cliticization is said to be necessary to achieve the appropriate feature-checking relations. The view that the post-verbal position of the subject must be caused by verb movement to C is supported by the ungrammaticality of post-verbal pronominal subjects in contexts where the C head is filled, as in Quebec French 'qui que tu as?' where '*qui que as-tu?' is ungrammatical. Complex inversion is then accounted for by arguing that CP has two specifiers, one occupied by the subject, the other by the *wh*-word in *wh*-questions. Another way of framing this is that the subject is in an adjunct position to the CP (Rizzi and Roberts, 1989: 302). This analysis was further developed in Cardinaletti and Roberts (2002). In this analysis, the outer layer of the TP, AgrP, is argued to have two sections, with the pronominal subject occupying a specifier of the lower AgrP, incorporating, as in the earlier analysis, and a DP subject occurring in the specifier of the higher AgrP. Rowlett (2007) has critiqued this analysis on the

basis that it is unclear why the verb could not check the DP's case feature prior to movement and then raise above it, if there is a higher position occupied by the verb (Rowlett, 2007: 203). Cardinaletti (2004) addresses some of these issues, developing an analysis in which the post-verbal subject occupies a head: Subj, a position above the rest of the TP. The verb moves to this head position and incorporates the clitic subject, while the DP subject occurs in Spec-SubjP (Cardinaletti, 2004: 140-141). Roberts (2009) also further develops this analysis, arguing that the enclitic subjects are in fact 'interrogative conjugation' which occupies the head of C, having no pronominal status. Roberts (2009) still argues for V-to-C movement, suggesting also that in the appropriate 'remnant V2' contexts, V moves to C to incorporate with the enclitic. There is then an EPP feature on C which may be fulfilled by either a null subject or an overt DP. If an overt DP is present it must move to Spec-CP, producing complex inversion.

I next turn to Rowlett's (2007) analysis of post-verbal pronominal subjects. In this analysis, similarly to Roberts' (2009) approach, the post-verbal subject is argued to be a fully grammaticalized syntactic clitic, occupying the head of T, which incorporates into the verb once it moves there. Several pieces of evidence are presented for this view. Firstly, while pre-verbal pronominal subjects can be omitted in a second conjunct, post-verbal pronominal subjects cannot. Secondly, not all pre-verbal pronominal forms can be used post-verbally. Pre-verbal *ça* systematically corresponds to post-verbal *il* for instance. This suggests that the post-verbal forms are meaningfully different from the pre-verbal forms. Additionally, Canadian French has evolved interrogative marker *-ti* from the original post-verbal subject form *-t-il*, suggesting some degree of grammaticalization of the subject (Rowlett, 2007: 207). This analysis has the advantage of allowing a straight-forward account of complex inversion structures, which are simply cases of an overt subject. It does, as such, require us to stipulate null subjects surviving in post-verbal pronominal subject environments. However, we can straight-forwardly argue that when the appropriate phi-features of the verb are checked by the clitic head in T, it needn't check them against an overt subject.

This analysis is extremely similar to that of Roberts (2009), except that it locates the enclitic inflection in T rather than C. Both Rowlett's (2007) analysis and that of Roberts (2009) leave us with

an open question of why complex inversion is impossible with a pronominal pre-verbal subject. Indeed, in varieties where interrogative inflection has truly developed, as in some colloquial Canadian varieties, we do see pre-verbal pronominal subjects alongside the interrogative inflection. I provide an example of this below, from Rowlett (2007: 207), glossing and translation original.

(20) Tu l'as-tu battu?
You him-have-tu beaten

'Did you beat him?'

Rowlett (2007) argues that such examples are not found in European varieties because it is 'uneconomical'. This is a problematic explanation, since redundancy is found frequently across language. Even though we would expect it to be significantly marked, we would not expect a pre-verbal pronominal subject to be fully ungrammatical, as it is.

Yet another alternative analysis is provided by Wolfe (2021, 2022), building on work by Cardinaletti (2021) on post-verbal subjects in interrogatives. In Wolfe's view, the pronominal subject is in fact a full subject when in the post-verbal position. However, pronominal and full DP subjects occupy distinct positions in the clausal spine. Full DP subjects occupy the specifier of a Subj head, as in Cardinaletti's (2004) analysis, higher than T but below the C-domain. Meanwhile, pronominal subjects obligatorily occupy Spec-TP. When pronominal subjects are found post-verbally, the verb has moved to the Subj head, a low-movement remnant of the V2 system. In this analysis, the loss of V2 was due to downwards reanalysis of the verb, gradually expanding the available pre-verbal positions until V2 was lost and the verb was in its current position in T. At one stage of this process, the verb was reanalysed to be in the Subj head, above T and thus preceding pronominal subjects, but below Spec-Subj and thus following DP subjects. This analysis has significant appeal in its account of post-verbal pronominal subjects as a remnant of the Old and Middle French V2 grammar. Although Wolfe's (2022) work does not directly address the parallels between pronominal inversion and complex inversion, we can adapt the analysis in Rizzi and Roberts (1989). If we adapt Wolfe's (2022) analysis to include the syntactic incorporation of the post-verbal pronominal subject, as in Rizzi and Roberts' (1989) analysis, we can then account for the acceptability of the pre-verbal DP subject

readily, while also accounting for why this position would be restricted to DP subjects, since they are the only permissible higher subject-type. This approach is strongly akin to that of Cardinaletti (2004). This analysis would also resolve Rowlett's (2007) critique of Cardinaletti and Roberts' (2002) analysis, since it provides us with a principled explanation of why DP subjects must be pre-verbal while pronominal subjects may be post-verbal: they simply inherently occupy different positions along the clausal spine.

In conclusion, an adapted version of Wolfe's (2022) analysis, including the use of incorporation, following Rizzi and Roberts (1989), provides a strong account of post-verbal pronominal subjects from which to begin our own analysis.

4.2.3 The History of Post-Verbal Subjects in French

Now that I have given the relevant descriptive background and discussed different theoretical approaches to post-verbal subjects in modern Romance languages, I turn to previous work on their history, particularly as they change from Old to Modern French.

I begin with a summary of descriptions of the state of play in Old and Middle French, starting with Vance's (1995, 1997) work. Vance (1995) argues that full DP subjects can occur post-verbally in non-V2 grammars, while pronominal subjects cannot. While Old French has a V2 grammar, even after the loss of V2 in Middle French, SVO word order does not become dominant for some time, with post-verbal subjects still found in over 50% of clauses in many 15th-century texts (Vance, 1995: 185). Indeed, post-verbal DP subjects increase in frequency during the Middle French period, despite the drop in the frequency of post-verbal pronominal subjects (Vance, 1995: 178). Vance (1995) argues that post-verbal pronominal subjects in Old and Middle French represent unambiguous cases of V2, since subject pronominals can only occur in Spec-TP and so, in order for the subject pronominals to follow the verb, the verb must raise higher, to C. On the other hand, post-verbal DP subjects, even in Old French, are taken to be ambiguous between a V2 analysis and the subject occurring in a low position. Kroch and Santorini (2009) further support the analysis of a low subject position in Old and Middle French, arguing that this position remains active, to varying degrees, until the 17th century. This description is further fleshed out by Roberts' (1993) distinction

between free and simple inversion, discussed above. In this view then, Old French shows simple inversion and pronominal inversion, both evidence of V2, as well as free inversion, which could be generated under a non-V2 grammar.

The authors of the GGHF suggest that earlier Old French texts allowed both simple and free inversion, but with a distinct preference for simple inversion (Marchello-Nizia and Prévost, 2020: 1085). This pattern seems to be even more marked in prose texts, leading to the suggestion that meter may influence the choice of free inversion. The GGHF also suggests that post-verbal subjects are typically focused, or at least that focused subjects are typically post-verbal alongside switch topic subjects (Marchello-Nizia and Prévost, 2020: 1101) while familiar topic subjects tend to be pre-verbal (Marchello-Nizia and Prévost, 2020: 1102). In contrast to DP subjects, pronominal subjects are typically pre-verbal from the earliest texts (Marchello-Nizia and Prévost, 2020: 1102). In the 16th and 17th centuries, pre-verbal subjects begin to predominate (Marchello-Nizia and Prévost, 2020: 1092), with post-verbal subjects restricted to increasingly specific contexts. The GGHF does, however, note that V1 contexts with overt post-verbal subjects are maintained well into the 17th century (Marchello-Nizia and Prévost, 2020: 1094). Indeed, this continues into Modern French, but in increasingly strictly specified contexts, particularly that of exhaustive focus (see Lahousse, 2003, 2006). At this time the pre-verbal elements which trigger the post-verbal position of the subject become limited to a closed set of specific, predominantly adverbial, elements (Marchello-Nizia and Prévost, 2020: 1094-6). Although some further variation develops in the 18th century, the GGHF states that 'En déclarative, les structures à sujet postverbal, diversifiées, n'ont pas évolué depuis le 18e s.' 'In declaratives, the diverse structures with a post-verbal subject have not evolved since the 18th century' (Marchello-Nizia and Prévost, 2020: 1098).

Beyond the GGHF, a handful of grammars of the Early Modern French period are available. Fournier (2009:45) argues that the subject is fixed pre-verbally by the beginning of the 17th century but that this change is still in progress during the 16th century. Gougenheim (1973), in his grammar of 16th century French, notes that post-verbal subjects can be triggered by adverbs and subordinate clauses (Gougenheim, 1973: 254) or even simply *et* 'and' (Gougenheim, 1973: 255). Gougenheim

(1973) even gives examples of V1 clauses with no apparent trigger (21) and SOV clauses (22). I exemplify these below:

(21)	Se	esveilloit	Gargantua	environ	quatre	heures	du	matin
	REFL	woke up	Gargantua	around	four	hours	of.the	morning

'Gargantua woke up around four o'clock in the morning.'

Rabelais, *Gargantua*, 23 (cited in Gougenheim, 1973: 254)

(22)	Pallas	sa	guide	estoit
	Pallas	their	guide	was

'Pallas was their guide.'

Du Bellay, *Regrets*, 40 (cited in Gougenheim, 1973: 255)

Roberts (1993) provides an in-depth, theoretical account of the development of post-verbal subjects from Old to Modern French, focusing on their development in interrogatives. In Roberts' (1993) account, the 16th century displays substantial change from the 15th century in the sorts of post-verbal subjects which are permissible. Simple inversion all but disappears (Roberts, 1993: 190), while complex inversion begins to develop, and free inversion structures are maintained. Importantly however, Roberts' (1993) work focuses only on interrogative contexts, without providing an analysis for post-verbal subjects in declaratives.

Wolfe (2021) provides one of the most in-depth accounts of French historical syntax and provides an entire chapter on the subject system of French. In Old French, Wolfe (2021) finds, like the GGHF, that simple inversion is consistently more common than free inversion. Additionally, although there are information-structural preferences for the free and simple inversion positions, neither shows a strict information-structural function and the distribution of information-structural roles varies across the Old French period. This situation continues into Middle French. In Wolfe's (2021: 175) view, this is suggestive of two different *v*P-peripheral positions, with Belletti's (2001) low focus position active in addition to a low topic position. This contrasts with Roberts' (1993) analysis in which Old French free inversion is interpreted as the subject remaining in its base-generated position. Wolfe's (2021: 190-2) discussion of the Early Modern French situation is brief, and no discussion is given to the status of post-verbal subjects in free inversion during that period, besides

noting their rarity. However, the central claim is simply that simple inversion of DPs is lost, and the Modern French system begins to establish itself.

Many of these works form useful descriptive background. Roberts (1993) and Wolfe (2021) provide a considerable theoretical background for this kind of diachronic work as well. However, Roberts' (1993) work focuses only on post-verbal subjects in interrogatives. On the other hand, Wolfe's (2021) work, although it does address post-verbal subjects, is focused on a full description of the broader subject system in the history of French and does not address the relationship of post-verbal subjects to overt objects or adverbials, although distinctions with regard to these are found to be extremely important in much modern comparative work (e.g. Cardinaletti, 2004; Ordóñez, 2007; Lahousse and Lamiroy, 2012). As such, there is still need for a more in-depth look at the history of post-verbal subjects in French.

4.3 Methodology

I now turn to my methodology. Text and corpus selection is covered in Chapter 3; here I provide further details on the methodology particular to the analysis of post-verbal subjects.

For the MCVF and Penn corpora, I have used a variety of search terms to search distinctly for main clause declarative pronominal inversion, simple inversion, free inversion and ambiguous DP inversion. I have then developed search terms to find instances of free inversion VSO and VOS orders as well as for examples of free inversion with a post-verbal adverb in order to further diagnose the position of the post-verbal subject.[12] The data from hand-analysed texts are analysed as with the other phenomena studied: the first 400 clauses are gathered, and any post-verbal subjects are noted to ascertain overall frequency data. Post-verbal subjects with *dire* and other verbs of speech were excluded from the frequency data, since they represent a particular idiomatised structure distinct from other post-verbal subjects. In all 400 clauses the subject type was annotated in order to provide relative frequency data, as well as raw frequency data. In all graphs in this chapter, where trendlines are present, they have been generated using a LOESS local regression in R, weighted for the number of

[12] See Appendix C for CorpusSearch queries.

clauses analysed in each text.[13] Information-structural work has then proceeded qualitatively, finding individual examples in order to demonstrate the variety of information-structural roles available to subjects in different positions.

4.4 Old and Middle French

Having provided an overview of the existing literature on post-verbal subjects in French, I turn to my data. I begin by discussing the quantitative corpus data, before turning to qualitative analysis.

4.4.1 Quantitative Analysis

I begin by presenting the quantitative data on all post-verbal subjects from the MCVF and Penn Supplement corpora, shown below in Figure 1.

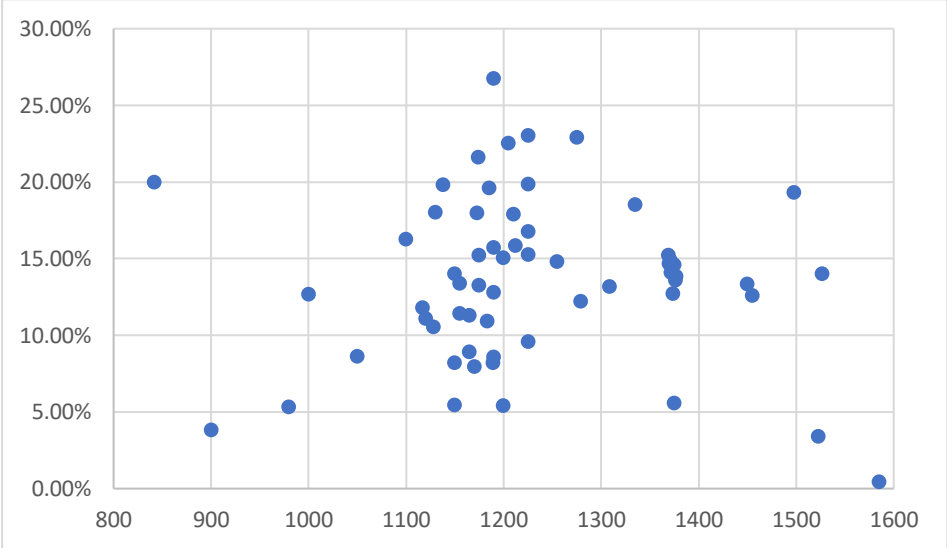

Figure 1 : Frequency of Post-Verbal Subjects in the MCVF and Penn Supplement

It is clear from this figure that post-verbal subjects are consistently common throughout the Old and Middle French periods. However, they are far from the dominant word order, occurring in less than one third of clauses in all texts. It is perhaps no coincidence that the lowest points are both in the 16th century, however, it seems like post-verbal subjects have not shown a considerable drop in frequency yet. There also appears to be a slight increase around 1200, and decrease thereafter, perhaps related to the shifting nature of the V2 constraint (see Wolfe, 2021: chapter 4).

[13] I am indebted to my examiners for the suggestion of the LOESS algorithm for generating useful trendlines.

4.4.1.1 Post-Verbal DP Subjects

I next turn to the relative rates of simple and free inversion in the MCVF and Penn Supplement. It should be remembered that the vast majority of post-verbal subjects are pronominal or ambiguous DPs (i.e. they occur in clauses with only a main verb, no auxiliary), so the rates of unambiguous simple and free inversion are low.

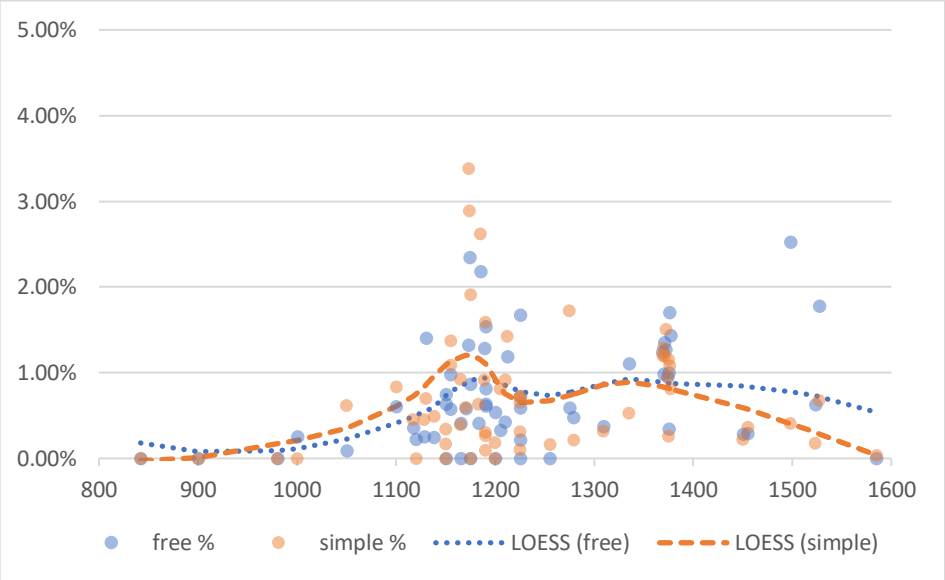

Figure 2 : Frequency of Simple and Free Inversion in the MCVF and Penn Supplement

Interestingly, according to LOESS trendlines, both free and simple inversion appear to increase in frequency from the earliest texts through to the 13th century. Although there is some decrease in frequency from the peak at 1200, the frequency of simple inversion only begins to drop considerably after approximately 1400, while free inversions seem to remain as frequent through to Early Modern French. During the 12th and 13th centuries, simple inversion becomes marginally more frequent than free inversion, as argued in the GGHF and Wolfe's (2021) work. It is important to note that even without the Early Modern French data, simple inversion appears to decline faster than free inversion during the Middle French period. However, for the majority of the period examined, free and simple inversion have comparable frequencies. I exemplify both simple and free inversion below.

(23) a.	E	si	furent	ocis	**li**	**sarrazin**.
	And	*si*	was	killed	the	Saracens

'And the Saracens were killed.'

(122X-PSEUDOTURPIN-MCVF-P,271.219)

b. Desuz un pin en est **li** **reis** alez
Beneath a pine CL is the king gone

'The king went beneath a pine.'

(1100-ROLAND-V,11.140)

We can see that, in (23a) the subject occurs after the verb and the auxiliary, thus, a case of free inversion. In (23b) on the other hand, the subject appears between the main verb and the auxiliary, thus a case of simple inversion.

4.4.1.2 Post-Verbal Pronominal Subjects

If we turn to post-verbal pronominal subjects, a decrease is visible over the course of this period. This finding is in line with Vance's (1995, 1997) work, in which post-verbal pronominal subjects are seen as an unambiguous indicator of V2.

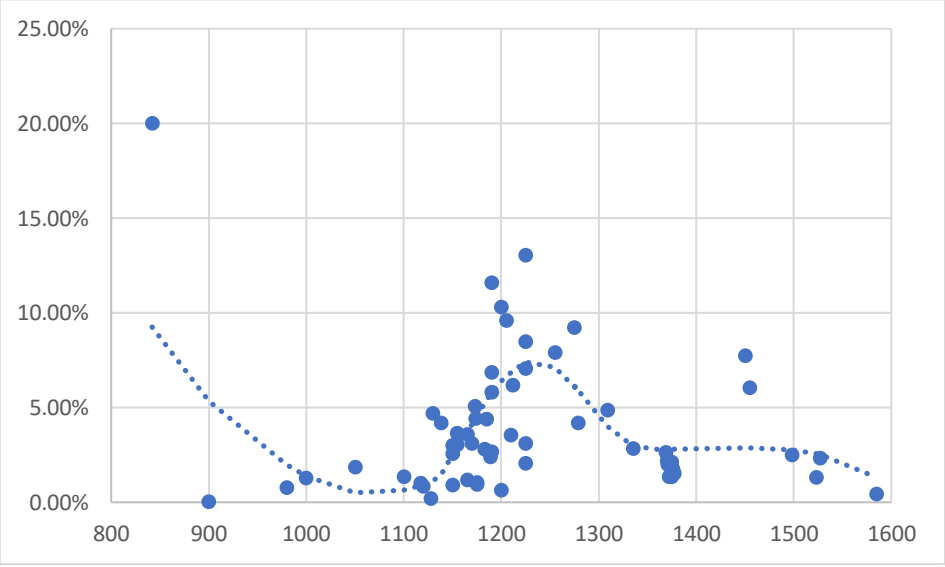

Figure 3 : Frequency of Post-Verbal Pronominal Subjects in the MCVF and Penn Supplement

Although the initial anomalous point from the earliest text distorts the LOESS trendline somewhat, we can still see that pronominal inversion decreases in frequency during the Middle French period, beginning in the latter half of the 13th century. As with post-verbal DP subjects, a peak is visible

around 1200, before the frequency begins to decrease. I illustrate with an example of a post-verbal pronominal subject from the 12th century.

(24) Tantost a **cele** demandees les noveles qu' ele queroit;
Immediately has this asked the news which she sought

'This person immediately asked for the news which she sought'

(1170-YVAIN-R,151.5249)

Notably, we see here that the pronominal subject occupies the same position as the DP subject in simple inversion, that is, it occurs between the auxiliary and main verb. It should be noted that complex inversion does not appear to be attested in these corpora, suggesting that it only develops later.

4.4.2 Formal Analysis

I now turn to the formal analysis of post-verbal subjects in this period. I argue here that, although many post-verbal subjects in Old and Middle French are the result of V2, free inversion is still widely attested and need not be the result of V2. By examining the relative position of the post-verbal subject to the auxiliary, adverbials and the object, we see that a variety of subject positions are attested, including at least two positions for subjects in free inversion. Before moving to my discussion of post-verbal DP subjects, however, I provide an analysis of post-verbal pronominal subjects, following previous analyses by Roberts (1993) and Vance (1995, 1997)

4.4.2.1 Post-Verbal Pronominal Subjects

I directly follow the analysis of post-verbal pronominal subjects in Old French proposed by Vance (1995, 1997). Under this analysis, the pronominal is seen as occupying its canonical position in Spec-TP, while the verb raises above it to C, or, in a cartographic analysis, either Fin or Force, depending on the time period (see Wolfe, 2021). Important evidence for this analysis comes from the parallels between post-verbal pronominal subjects and simple inversion, with the subject occurring between the auxiliary and main verb in both. To illustrate, I provide a bracketing of (24) below in (25):

(25) [FrameP [AdvP tantost] [… [FinP [Fin' [Fin a] [TP [DP cele] [T' [T ~~a~~] [… [FP [F' [F demandees]
[VP [DP ~~cele~~] [V' [V ~~a~~] [DP les noveles qu'ele queroit]]]]]]]]]]]

We see in this example that the pre-verbal position is filled by the adverbial 'tantost',[14] the verb raises to its left-peripheral position in Fin, while the subject remains in its canonical Spec-TP position, the past participle in a lower, unspecified F head, above the VP, and the object remains in a low position, here illustrated as VP-internal. Additionally, the lack of complex inversion at this time means that there are no further data to account for. We also do not see any notable changes in the distribution of post-verbal pronominal subjects in these data.

4.4.2.2 Post-Verbal DP Subjects

Having briefly discussed post-verbal pronominal subjects, I turn to post-verbal DP subjects. As we have already seen, Old and Middle French feature post-verbal DP subjects in both simple and free inversion constructions.

To re-illustrate the current predominant analyses, I give bracketings for the two examples in (23) below in (26), following Belletti's (2001) analysis of free inversion (26a) and Robert's (1993) analysis of simple inversion (26b).

(26) a. si furent ocis **li sarrazin**.

[$_{FinP}$ [$_{AdvP}$ si] [$_{Fin'}$ [$_{Fin}$ furent] [$_{TP}$ [$_{T'}$ [$_{T}$ ~~furent~~] [… [$_{FP}$ [$_{F'}$ [$_{F}$ ocis] [$_{FocP}$ [$_{DP}$ li sarrazin]

… [$_{VP}$ [$_{DP}$ ~~li sarrazin~~] [$_{V'}$ [$_{V}$ ~~furent~~]]]]]]]]]]

b. Desuz un pin en est **li reis** alez

[$_{FrameP}$ [$_{PP}$ Desuz un pin] [… [$_{FinP}$ [$_{Fin'}$ [$_{Fin}$ en est] [$_{TP}$ [$_{DP}$ li reis] [$_{T'}$ [$_{T}$ ~~en est~~] [… [$_{FP}$ [$_{F'}$ [$_{F}$ alez]

[$_{VP}$ [$_{DP}$ ~~li reis~~] [$_{V'}$ [$_{V}$ ~~en est~~]]]]]]]]]]][15]

As in our bracketing of post-verbal pronominal subjects, we have shown the non-finite verb position as a low FP, such that the focused free inversion position remains post-verbal. Importantly, the position occupied by the participle has been shown to vary between Modern French on the one hand, and Modern Italian and Spanish on the other. In Modern French, the quantifier object *tout* must

[14] I have shown the adverbial occupying Spec-FrameP. Some would claim that the immediately pre-verbal constituent always occupies Spec-FinP to fulfil the EPP feature there. I do not intend the illustrated position here to make any strong theoretical claims, as the focus of this example is the post-verbal subject position.
[15] For ease of reading, I have kept the clitic *en* in the deleted copies of the verb; this should not be read as a theoretical claim about the status of locative clitics.

precede the participle, while in Modern Italian and Spanish it may occur after it, suggesting that the participle must raise higher (Ordóñez, 2007: 263-4). In our Old French data, the same situation as Modern French obtains, with object quantifier *tout* only occurring before the participle:

(27) Et l' eit Joseph tout delivré.
And it had Joseph all delivered

'And Joseph had delivered it all.'

(1190-BORON-PENN-R,18.283)

It also seems to be the case that multiple low subject positions are available in free inversion constructions. Since, unlike Modern Italian and Catalan but like Modern Spanish, Portuguese and Romanian, VSO orders are attested in addition to VOS orders in our Old and Middle French data. We can see this illustrated below in (28):

(28) a. Lors a geté Helcanor ses yex
Then has thrown Helcanor his eyes

'Then Helcanor cast his eyes.'

(127X-CASSIDORUS-PENN-P,341.2764)

b. Et avoient ordonné une place li Hainnuier ou il
And had arranged a place the *Hainnuier* where they

se devoient tout retraire.
REFL had all to retreat.

'And the *Hainnuiers* had arranged a place where they all had to retreat.'

(1369-FROISSART-1-P,122.1843)

We can see that in (28a) the subject 'Helcanor' precedes the object 'ses yex', while in (28b) the subject 'li Hainnuier' follows the object 'une place'. We can further diagnose the position of the lower subject with adverbial ordering. Although the data from the MCVF and Penn Supplement are a little sparse to easily exemplify this, adverbials of the type found in Cinque's (1999) hierarchy seem to consistently precede the lower subject, although extraposed frame-setters and apposed adverbials can occur after the subject. These data are exemplified below:

(29) a. Malade ot geü longuemant la pucele,
Ill had lain for a long time the maiden

'The maiden had lain ill for a long time'

(1170-YVAIN-R,177.6224)

b. Pere, quant vus partistes del regne sudement,
Father when you left from.the kingdom suddenly

mult par en fu trublez li regnes erramment.
much by CL was troubled the kingdom wrongly.

'Father, when you suddenly left the kingdom, the kingdom was, wrongly, very troubled.'

(1173-BECKET-BFM-R,98.2552)

In (29a) we see that the adverb 'longuemant' precedes the subject. This is an aspectual adverb in $\text{ASP}_{\text{durative}}$, the middle area of Cinque's (1999) hierarchy. Although in (29b) the adverb 'erramment' does seem to follow the subject, it is clearly a modal adverb which should traditionally be analysed as higher in Cinque's hierarchy, in $\text{MOOD}_{\text{evaluative}}$. This example appears to involve some sort of parenthetical or focusing usage, as described in Cinque (1999: 30-32). However, it is difficult to show this clearly with textual data. Suffice to say that these examples are vanishingly rare in our data, compared to those in which the adverb precedes the subject, such as (29a).

Indeed, if we seek adverbs lower down in Cinque's (1999) adverbial spine, we find that the subject seems to even follow those in the lowest position: $\text{ASP}_{\text{frequentative(II)}}$, as illustrated by 'often'. We can see this below in (30):

(30) Si l' ad oï suvent mis sires acunter.
Si it has heard often my lord recount

'My lord has often heard it recounted'

(1173-BECKET-BFM-R,57.1547)

In this example we see that the low post-verbal subject follows both the auxiliary 'ad' and main verb 'oï', as well as the low adverb often 'suvent', in addition to preceding the infinitive 'acunter', suggesting that it must not be in its base-generated position.

The conclusion I draw from these data is that, as per Ordóñez' (2007) analysis of Modern Spanish free inversion, multiple post-verbal subject positions are available. However, extending Ordóñez's (2007) analysis to our data becomes problematic due to the positioning of the participle, which in Old and Middle French consistently follows object quantifier *tout*, while it precedes it in Spanish, where VSO subjects also precede object quantifiers. Thus, we must conclude that, in Old French, the participle has moved to a lower position than Spanish, while the higher subject in free inversion constructions occurs in a lower position than that of Spanish, below the adverbial spine. One might ask at this stage why we would not assume that the subject is in fact in the low focus position, and the object licensed by some other means in Old French, unavailable in Modern Italian. Importantly, however, in VSO examples in Old French, the subject need not be focused, thus suggesting that it is occupying a separate position. An example of a non-focused VSO subject is given below in (31). Due to the length of this example, I gloss only the relevant clause, while providing a full translation of the rest of the example.

(31) Quant Ydoine le vit, si fu moult courouciez, et dist : « Helcanor, comment fustes vous si hardiz que vous avez trespassé mon commandement? – Sire » dist Helcanor « dont ne me huchastes vous ? – Par Dieu, » dist il, « je non. – Ainsi, » dist Helcanor, « me fu il avis. »

Lors	**a**	**geté**	**Helcanor**	**ses**	**yex**
Then	has	cast	Helcanor	his	eyes

et vit le preudomme qu'il n'avoit nul dois en l'une de ses mains.

'When Ydoine saw him, he was very shocked, and said "Helcanor, how were you so bold that you broke my commandment?" "Sire," said Helcanor "did you not summon me?" "By god," he said, "not I" "It seemed to me that you did" said Helcanor. **Then Helcanor cast his eyes** and saw that the brave man had no fingers on one of his hands.'

(127X-CASSIDORUS-PENN-P,341.2764)

In (31) we can see that Helcanor has been one of the two topics of the ongoing discourse (alongside Ydoine, his interlocutor), as well as the subject of the previous matrix clause, thus it is clear that we could not consider Helcanor focused in this context. As such, I analyse these structures as involving a

dedicated low subject position with no information-structural requirements, above the object position but below the position of non-finite verbs and Cinque's (1999) adverbial spine. I tentatively suggest that this position may be Spec-*v*P, with a null subject fulfilling the EPP in T, the DP subject may then remain in a lower position. This position would also be expected to consistently follow both the finite and non-finite verb, assuming the non-finite verb occurs in an FP above *v*P (see Rowlett, 2007: 108-110, 226-229 for discussion of non-finite verb syntax).

In VOS orders, a variety of information-structural relationships are also available to both subjects and objects. It should be noted, however, that VOS orders are extremely scarce. I provide an example in (32). Again, only the relevant clause is glossed to conserve space.

(32) Sus cel estat et avis, fissent li chevalier de Hainnau pluisseurs bonnes ordenances pour euls mieuls garder et desfendre, par lesquelles ordenances il convenoit toutdis par nuit jesir en armes, et par jour euls tenir en lors hostels et lor harnas avoir apparilliet et les chevaus tous pres et ensellés. Et les convenoit toutdis, par nuit et par jour, gettier par connestablies les camps et les cemins de autour de la ville, et envoiier auqunes escoutes demi lieue en sus de la ville, pour savoir se chil archier faisioent nul agait ne asamblee**.**

Et	**avoient ordonné**	**une**	**place**	**li**	**Hainnuier**	**ou**	**il**
And	had decided	a	place	the	*Hainnuiers*	where	they

se	**devoient**	**tout**	**retraire.**
REFL	must	all	retreat

'Under this state and advice, the knights of Hainault made several good plans to better guard and defend themselves, by which plans it was agreed by all to lay in their armor by night and to stay in their hostels by day with the horses close and saddled. And they all agreed, by night and by day, to impose troops upon the fields and paths around the town and to send scouts half a league from the town, to know if this archer was hatching a trap or an army. **And the *Hainnuiers* [people from Hainault] had decided a place where they should retreat to**.'

(1369-FROISSART-1-P,122.1843)

In this example, the *Hainnuiers* have been the topic of the recent discourse, as those who are making the preparations, while the object, the place they decided on, appears to be focused, new information. We also see an example of an extraposed relative clause modifying the object here, further complicating the analysis. In Belletti's (2001) analysis, these VOS structures are analysed as remnant movement of the VP to Spec-TopP after the extraction of the subject to its canonical position. We can simply extend this analysis in Old and Middle French, with the subject moving out of the VP to Spec-*v*P before remnant movement of the VP to either Focus or Topic domains of the left periphery. We have now accounted for both VSO and VOS orders. It should also be noted that both (31) and (32) contradict the claims of the GGHF that post-verbal subjects tend to be informationally new or focused, while informationally old and topical subjects tend to be pre-verbal.

I argue that all post-verbal subjects, outside of cases of VSO or VOS, may occupy the informationally under-specified position, which is lower than that of Spanish, or a distinct focused position, potentially similar to that of Belletti (2001). Notably, these focused examples display a distribution noted for Modern French by Lahousse (2006). Floating quantifiers may occur above the subject's surface position, this is illustrated below in (33):

(33) Et furent **tout** mort chil qui avoecques le dit conte
and were all died those who with the said count

avoient pris terre.
had taken land

'And all who had taken land with the aforementioned count had died'

(1375-FROISSART-6-P,643.1760)

Here, the floating quantifier 'tout' occurs between the auxiliary 'furent' and main verb 'mort', above the position of the subject. Lahousse (2006: 445-446) argues that this distribution of floating quantifiers, found also in Modern French, is evidence that the subject must have moved through this position, and thus, moved to a left-peripheral position prior to remnant movement of the TP. In this case, we would analyse the subject as moving to a left-peripheral focus position, before the remnant TP 'furent tout mort' inverts.

Thus, we conclude that Belletti's (2001) focus position is not in fact active in Old French and that these information-structurally marked examples are cases of left-peripheral movement with remnant movement, while information-structurally under-specified low subjects occupy the low subject position described above.

In summary, free inversion may represent two different structures. Firstly, the subject may occupy a low subject position. This position is information-structurally under-specified and occurs below all adverbials in Cinque's (1999) adverbial spine, as demonstrated in (30). It is in this position which VSO orders are licensed. Given the information-structurally under-specified nature of this position, as well as its extremely low position compared to adverbials, I suggest here that it is Spec-*v*P. Secondly, the subject may move to a left-peripheral position before the remnant clause inverts. This accounts for information-structurally marked low subjects (following Lahousse, 2006), as well as VOS orders (following Belletti, 2001).

These findings may have further ramifications for the general study of subject positions cross-linguistically. Considering our analysis in comparison to Spanish data, it seems that there are not a small number of fixed subject positions, but rather a wide variety of positions along the clausal spine, perhaps even suggesting the recursive analysis of Laenzlinger and Soare (2005). In each language, multiple positions may be available, with null subjects licensing non-canonical positions. These positions may vary from language to language. These distinctions may be extremely fine-grained, as we see with the positioning of quantifier objects distinguishing Old and Middle French VSO from Ibero-Romance VSO. This wide variety of subject positions is noted in Cardinaletti's (2004: 118-120) discussion of the subject mid-field. However, even here, a simple distinction is made between Ibero-Romance and Italo-Romance data, supplemented by data from Hebrew and Icelandic, while further, even finer-grained distinctions are not discussed.

4.5 Early Modern French

I now turn to my Early Modern French data. I begin with quantitative data, before turning to formal analysis.

4.5.1 Quantitative Data

I begin with quantitative data. Firstly, I present the overall rates of post-verbal subjects, including both the MCVF and Penn Supplement data and my hand-collected Early Modern French data.

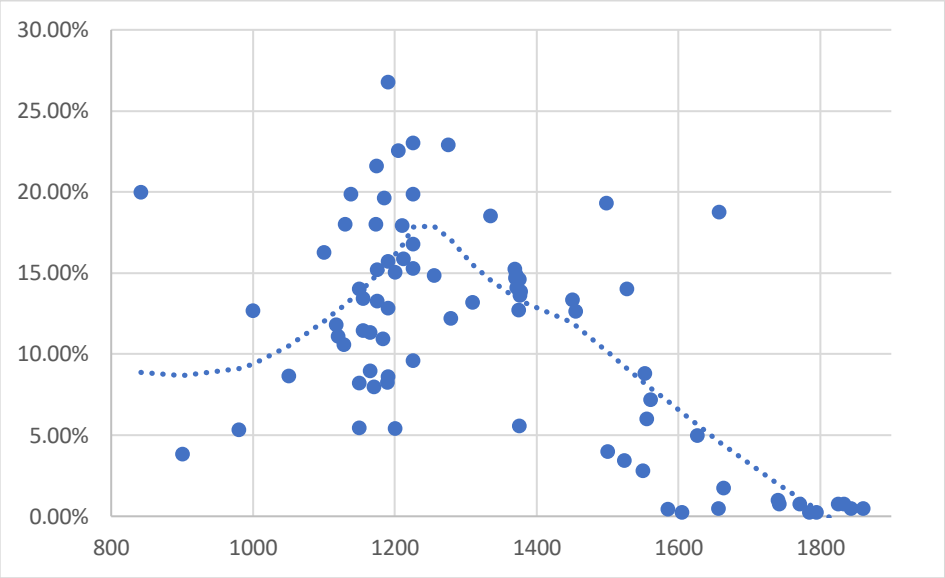

Figure 4 : Frequency of Post-Verbal Subjects in Old, Middle and Early Modern French

Interestingly, the trendline here seems to suggest that the decline in frequency of post-verbal subjects begins as early as the early 13th century and continues more or less uninterrupted. It should be noted that we have one anomalous high point in our Early Modern data: Chavatte, a sub-standard private chronical written by a *peu lettré*, with 18.25% of its clauses featuring a post-verbal subject. Although less marked with a 5% rate of post-verbal subjects, Valuche, a *journal intime*, also sub-standard, also shows unusually high rates. This seems to suggest that innovations in the syntax at this time were partially driven from above, with prescriptive standardisation in the wake of the Renaissance pushing towards an SVO word order. Indeed, grammarians of the time often commented on the 'clarity' of SVO order, in opposition to the variation found in earlier varieties of the language (Ayres-Bennett, 1996: 184-5). However, we have also seen that the decline in post-verbal subjects begins much earlier. As such, it seems more likely that grammarians and prescriptivists latched onto an existing change, rather than causing it. Additionally, it should be noted that the use of post-verbal subjects in Chavatte

appears to be somewhat formulaic, occurring with passive verbs and particularly with initial frame-setters. These are exemplified below.

(34) a. Au dite an fut deffaite la bonne maladri
In.the said year was demolished the good leper colony

de la porte des malades laquelle on mettois ceux qui estoient
of the gate of.the sick which one put those who were

lepreux ceste a dire l'ardre[16]
lepers that's to say leprosy

'In the said year, the good leper colony by the *porte des malades*, in which one put those who were lepers, was demolished'

Chavatte p. 167v

b. Le 11 jour de Juillet furent pendues en la ville
The 11th day of July was hanged in the city

de lille quatre soldat pour ce qui avoient pillees un chariot
of Lille four soldiers because had pillaged a wagon

'On the 11th day of July four soldiers were hanged in the city of Lille because they had pillaged a wagon'

Chavatte p. 166r

c. Le 2 jour d' avril mourut ce bon religieux prince Ferdinand
The 2nd day of April died this good religious prince Ferdinand

3e empereur tres auguste
third emperor very august

'On the 2nd day of April, the good, religious, very august emperor prince Ferdinand the third died'

Chavatte p. 166r

We can see in (34a) and (34b) that the combination of passive with an initial frame-setter seems to be a strong motivation for the post-verbal position of the subject. In (34c) only the frame-setter is present, but it seems clear that these structures are something of a formula found in this style of

[16] In Ernst's (2018) edition, footnotes clarify that *maladri* should be taken as *maladrerie* and *l'ardre* as *ladre*. However, as footnoted above, I maintain all spelling irregularities from the texts of *peu-lettrés*.

diaristic writing where the date is given at the head of the clause, triggering the post-verbal position of the subject.

I next turn to the relative rates of unambiguous simple and free inversion in these data.

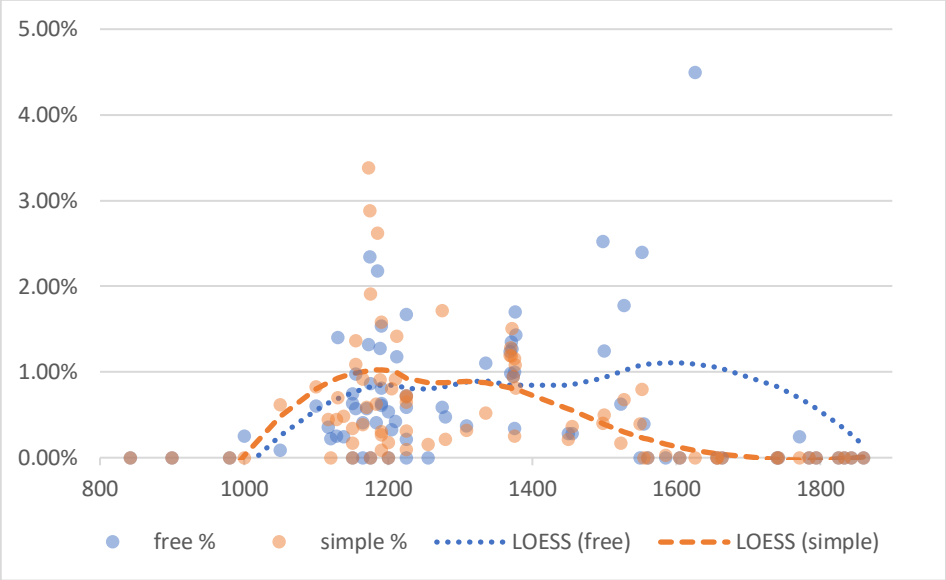

Figure 5 : Frequency of Free and Simple Inversion in Old, Middle and Early Modern French

We see here that, while free inversion remains readily available well into the 1700s, simple inversion is almost entirely extinct after 1600, with a considerable drop off beginning around 1400. Our anomalously high free inversion data point after 1600 is Valuche. The frequency of free inversions in Chavatte is not visible on this graph, since it is so anomalously high (17%) that it would render the other data unreadable if the axis was adjusted appropriately. Nonetheless, this data point has been taken into account in generating trendlines. In the 15th and 16th centuries some remnant cases of simple inversion remain, including in Pantagruel and the Correspondance Estouteville. Two examples are given below:

(35)	En	iceluy	estoit	Malicorne,	escuyer	tranchant	de	Gargantua	
	In	this	was	Malicorne,	squire	sharp	of	Gargantua	
	envoyé	expressément	de	par	luy	entendre	l'	estat	
	sent	expressly	of	by	him	hear	the	state	
	et	portement	de	son	fils	le	bon	Pantagruel	et
	and	behaviour	of	his	son	the	good	Pantagruel	and

luy porter letres de creance.
him carry letters of credence.

'In this [ship] Malicorne, Gargantua's sharp squire, was expressly sent by him to hear the state and behaviour of his son, the good Pantagruel, and bring him letters of credence'

Pantagruel, Book 4 Chapter 3

(36) depuis n' est aultre chose sourvenu que ne puissez scavoir
since NEG is other thing happened that NEG could know

'Since then, nothing else has happened that you could know'.

Correspondance Estouteville, Letter IX

It is clear that some remnant cases of simple inversion survive in texts which do not otherwise show a V2 grammar. However, due to the extremely rare nature of these remnant constructions, it is unclear if specific environments favour these remnant V2 structures. As such, it is hard to identify whether or not they represent cases of micro- or nano-parametric variation. This issue will be discussed further in Chapter 7 and Chapter 8. Whatever the case, the loss of V2, although relatively sudden, nonetheless leaves room for remnant constructions in texts which do not otherwise display a V2 grammar.

Finally, I turn to the frequency of post-verbal pronominal and DP subjects.

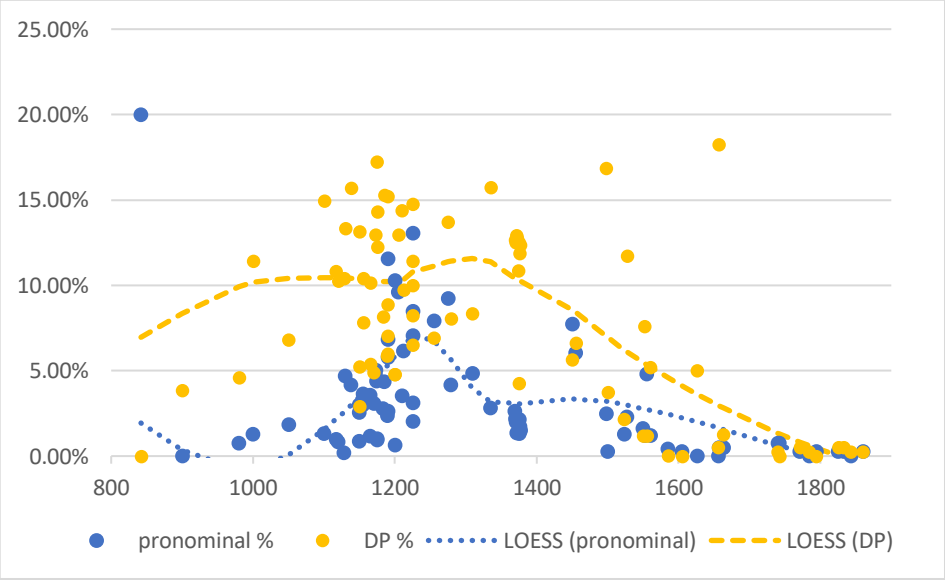

Figure 6 : Frequency of Post-Verbal DP and Pronominal Subjects in Old, Middle and Early Modern French

We see in this graph that the frequency of both post-verbal pronominal and DP subjects decreases considerably after 1600. In both cases, however, the decline begins much earlier, with post-verbal DP subjects declining in frequency from the 14th century and post-verbal pronominal subjects declining in frequency from a peak just after 1200. The earlier decline of post-verbal pronominal subjects can be attributed to their relationship to the V2 constraint, following Vance (1995, 1997), while DP inversions may be generated outside of a V2 structure, as in free inversion. Additionally, it is interesting to note that post-verbal DP subjects have historically been considerably more frequent than pronominal.

In summary, it's clear that post-verbal subjects become extremely marginal after 1600. However, the decline begins much earlier, during the 13th and 14th centuries. Additionally, the decline is somewhat delayed in texts from *peu lettrés*. Given the substantial changes in frequency which occur during the Early Modern period, it will be useful to analyse the formal and functional shifts which occur in this period.

4.5.2 Metalinguistic Data

Before proceeding to formal analysis, I provide a discussion of the mentions of post-verbal subjects in the metalinguistic texts examined, hoping to reveal more about the role of prescriptivism in the decline of post-verbal subjects.

Bouhours provides no discussion of post-verbal subjects at all, while Vaugelas mentions them in passing several times and Maupas provides extensive discussion.

In Vaugelas, the majority of the discussion of constituent order is either extremely vague and abstract, or so specific as to be of little use for our work. For instance, we find frequent discussion of the importance of clarity, but Vaugelas rarely specifies how one achieves clarity. He specifies that a great writer's work should have 'un grand jugement à composer la phrase claire & elegante, la douceur que demande l'oreille' 'a great judgement to compose a clear and elegant sentence, the gentleness that the ear demands' (Vaugelas, 1647: 123). Similarly, Vaugelas notes that 'c'est à faire aux paroles de faire entendre le sens' 'it is the role of speech to make the meaning understood' (Vaugelas, 1647: 590) but

without specifying exactly how this is achieved. When more specifics are given, we find something more akin to a style guide than grammatical comments. For instance, under the section entitled 'Certaine Reigle pour une plus grande netteté ou douceur de stile' 'Certain rules for greater clarity or gentleness of style' (Vaugelas, 1647: 528) discussion is primarily given to highly specific examples of genitive and adjective ordering in coordinated contexts.

When examples of post-verbal subjects are discussed, they are extremely specific cases, for example Vaugelas notes the grammaticality of 'n'ont-ils pas fait' 'they have not done' in passing in a discussion of negation. It is clear that Vaugelas understands post-verbal subjects to be both grammatical and unproblematic, with the notion of *netteté* representing a more abstract guiding principle of style, rather than suggesting that SVO order should dominate.

While Vaugelas rarely directly discusses post-verbal subjects, Maupas discusses them extensively. Maupas lists numerous potential triggers for a post-verbal subject, including specific verbs such as *dire* 'to say', *croir* 'to think' (Maupas, 1618: 122), a slew of adverbials such as *ainsi* 'thus', *aussi* 'also', *à peine* 'scarcely', *difficilement* 'difficultly' (Maupas, 1618: 123-125) and more specific contexts such as with emphatic subjunctives 'fusse-je aussi heureux que vous' 'would that I were as happy as you' (Maupas, 1618: 125). Despite Maupas describing SVO as 'l'ordre naturel de l'entendement' 'the natural order of the intellect' (Maupas, 1618: 252), it is clear that post-verbal subjects are not only accepted but encouraged. Indeed, Maupas goes on to say that 'il nous advient bien quelquefois de postposer le nominatif a son verbe, ce qui n'est pas trait d'apprenti, mais de bien versé en la langue, de la faire avec grace.' 'it is sometimes good to postpose the nominative from its verb, to do this with grace is not a trait of an apprentice, but of one well-versed in the language' (Maupas, 1618: 255). It is clear then, that neither a purely traditionalist nor innovative perspective is found in the grammarians. Rather, while SVO is understood to be the underlying word order of the language, post-verbal subjects are not considered aberrations, but artful flourishes used by the most proficient language users. These findings match the findings from our own quantitative data, where a complex and indirect relationship seemed to hold between text-typological factors and changes to subject positions, with the decline in frequency of post-verbal subjects beginning well before the Early Modern period.

Having discussed our quantitative data, as well as metalinguistic data on post-verbal subjects, I next turn to formal analysis of these structures in the Early Modern period.

4.5.3 Formal Analysis

The Early Modern French period is one of considerable shifts in post-verbal subjects. Thus, I will be focusing on how these changes occur here, taking my analysis of the Old and Middle French situation as my starting point. As with my Old and Middle French data, I begin with the analysis of post-verbal pronominal subjects, before turning to shifts in post-verbal DP subjects.

4.5.3.1 Post-Verbal Pronominal Subjects and Complex Inversion

I begin with post-verbal pronominal subjects. Although the analysis of post-verbal pronominal subjects in Old and Middle French proved straight-forward, the shifts which occur in Early Modern French leave us with several options for analysis, as described in section 4.2.2.2.

As we have seen in our quantitative data, post-verbal pronominal subjects are extremely infrequent after 1600. However, where they do survive, they are not motivated by discourse pragmatic considerations but typically by certain pre-verbal trigger words, something also noted by contemporaneous grammarians, as seen in section 4.5.2. Additionally, complex inversion develops during this period. Examples of both post-verbal pronominal subjects and complex inversion are given below:

(37) a. Aussi parlent ils toutes langues.
Also speak they all languages

'Also, they speak all languages.'

Belon du Mans Chapter 26

b. À peine le vieux soldat eut- il aperçu ma mère, qu'
Scarcely the old soldier had he noticed my mother that

il se crut délivré.
he REFL believed saved

'The old soldier had scarcely noticed my mother when he believed himself saved.'

La Russie en 1839 p. 29

It should be noted that not all examples of post-verbal pronominal subjects feature explicit triggering pre-verbal content. Two examples are given below.

(38) a. Et ne sceu- t- on jamais qui ce fut.
And NEG knew PTCL one ever who it was

'And nobody ever knew who it was.'

Chavatte p. 169r

b. Et le voiois t' on encore au fin de l' an
And it saw PTCL one still at.the end of the year

passee
passed

'And one could still see it once the end of the year had passed'

Chavatte p. 177v

These examples seem to typically feature an expletive or semantically weak subject, in this case the generic 3rd person 'on'.

Having discussed the surface realisations of post-verbal pronominal subjects found in this period, I turn to analysis. The key to our analysis must be in the striking fact that complex inversion develops from post-verbal pronominal subjects precisely when simple inversion has reduced in frequency and V2 features appear to be lost. It is my contention that this strongly suggests that the survival of post-verbal pronominal subjects is due to a reanalysis which allows their continued existence.

If an analysis like that of Rizzi and Roberts (1989) is correct in postulating V-to-C movement to account for post-verbal pronominal subjects, we would expect simple inversion to be maintained in the same contexts, at least initially. In Modern French Rizzi and Roberts (1989) account for this asymmetry by varying licensing restrictions on pronominal and nominal subjects. However, we would expect to see the changes to subject licensing play out diachronically and we do not. Thus, it seems that post-verbal pronominal subjects must be reanalysed in the 16th and 17th centuries and that this reanalysis also allows for the development of complex inversion. This then leaves us with two options

for our analysis. Firstly, Rowlett (2007) and Roberts' (2009) analyses, in which the post-verbal subject is argued to be inflectional, with the two analyses differing on the location of the verb (T in Rowlett (2007), but C in Roberts (2009)). Secondly, Wolfe's (2022) V-to-Subj analysis, in which the verb raises to the head of a phrase whose specifier hosts DP subjects, but above Spec-TP, where pronominal subjects are hosted.

In favour of Rowlett's (2007) and Roberts' (2009) analyses, it is clear that post-verbal pronominal subjects behave distinctly from their pre-verbal counterparts. We can see this clearly in the two examples from Chavatte given above in (38), in these cases the post-verbal form 't'on' is used. A search through the rest of the text shows that 't'on' is exclusively used in this post-verbal position, while 'on' is only used pre-verbally. However, as seen in section 4.2, Rowlett's (2007) and Roberts' (2009) accounts fail to account for the lack of subject doubling with a pronominal subject.

If we follow Rizzi and Roberts' (1989) analysis however, this is resolved. In this view, cliticization occurs by incorporation of the pronominal subject after the verb raises to a higher position. Thus, since the enclitic is truly a subject, a pre-verbal pronominal subject could not also be generated.[17] It should be noted that Rizzi and Roberts' (1989) account also requires further adaptation. Rizzi and Roberts' (1989) analysis involves V-to-C movement, which we have seen is problematic, especially in this historical context, since we would expect to see some surviving cases of simple inversion alongside post-verbal pronominal subjects, at least initially. Thus, I suggest that we take an element from Cardinaletti's (2004) and Wolfe's (2022) analysis: the two subject positions and V-to-Subj movement, rather than V-to-C movement. This allows us to elegantly explain why subject doubling may only occur with DP subjects and why simple inversion is no longer attested. The special behaviour of post-verbal pronominal subjects is then accounted for by their syntactic cliticization by incorporation, per Rizzi and Roberts' (1989) analysis. We might also suggest that Wolfe's (2022) account does occur in its original form at some stage, with the post-verbal pronominal subjects

[17] Roberts (2009) argues against this view, since strict cyclicity would result in the subject being proclitic to the verb, not enclitic. While I do not propose a solution to this issue, nor discuss the particulars of head movement and incorporation here, we have seen that this account is preferrable from a descriptive perspective. Roberts' (2009) account, meanwhile, struggles to account for the lack of reduplication with pronominal subjects, particularly in light of Canadian French data discussed in section 4.2.2.2.

uncliticized. This structure would then allow for the gradual development of post-verbal pronominals into their clitic form through standard processes of grammaticalization.

To summarize, we have seen that, during the Early Modern French period, post-verbal pronominal subjects become grammaticalized as syntactic clitics. In cases of pronominal inversion, the verb moves to the Subj head, a position above T the specifier of which hosts full DP subjects. The pronominal subject in Spec-TP then left-incorporates to the verb, as in Rizzi and Roberts (1989), accounting for the special behaviours of post-verbal pronominal subjects. Finally, the EPP on SubjP is fulfilled either by a null subject, as in pronominal inversion, or a DP subject, as in complex inversion. Since the pronominal subject is already present in Spec-TP no pronominal subject doubling may occur.

Under this analysis, cases of post-verbal pronominal subjects represent the case in which no overt DP subject fills Spec-SubjP, instead the EPP on SubjP is fulfilled by a null subject. This leaves us with a question regarding the licensing of a null subject in Spec-SubjP. I argue that the clitic is licensed by the specific contexts of pronominal and complex inversion. This clitic then checks the phi-features of the verb through incorporation into the Subj head, allowing for a null subject to surface in Spec-SubjP, rather than requiring an overt subject to check the verb's phi-features. If the null subject licensed the pronominal clitic, then we would not find cases of complex inversion, since no null subject would be present to license the clitic. Instead, since the clitic is licensed by the specific contexts, and in turn licenses the null subject, we can neatly explain why complex inversion shares the same distribution as pronominal inversion.

Additionally, we may treat the development of these syntactic clitics as a case of layering, a tell-tale sign of grammaticalization. In this case, we can see that the cliticized post-verbal subject grammaticalizes despite the survival of the independent pre-verbal pronominal subjects, which are not syntactically cliticized. Additionally, it seems to be the case that some instances of post-verbal pronominal subjects are analogous to cases of post-verbal DP subjects in Modern and Contemporary French. Lahousse (2022) shows several instances in which post-verbal pronominal subjects pattern

alongside a counterpart post-verbal DP subject, with specific adverbial or adjectival triggers appearing to cause verb movement.[18]

4.5.3.2 Post-Verbal DP Subjects

Turning to post-verbal DP subjects, the Early Modern period is a period of considerable shift. I will argue that, in certain 16th- and 17th-century texts we see a maintenance of the low subject positions found in Old French. However, by the 18th century, post-verbal DP subjects are highly restricted, both formally and information-structurally. I argue that the loss of these low subject positions is linked to the loss of null subjects in the 16th and 17th centuries.

Firstly, I exemplify the Old French low subject positions maintained in our 16th- and 17th-century texts, with examples of both VSO and VOS orders, as well as the variable position of the adverb given below:

(39) Ce mesme jour passa Pantagruel les deux isles de Thohu
This same day passed Pantagruel the two islands of Thohu

et Bohu
and Bohu

'On this same day, Pantagruel passed the two islands of Thohu and Bohu.'

Pantagruel Book 4 Chapter 17

(40) A ses paroles du roy de Portugal donnerent grant
at these speech of.the king of Portugal gave great

foy les seigneurs et dames
faith the lords and ladies

'The lords and ladies gave great faith to this speech of the king of Portugal.'

Roman de Jean de Paris p. 47 (BFM version)

We can see that this appears to be a straight-forward maintenance of the Old French situation, including information-structurally. In (39) Pantagruel is the topic of the entire text, and has been the topic of the previous clauses, a clear case of a topic, while the object is focused with a VSO order. In

[18] This seems to suggest that there may be some link between DP and pronominal inversions in Contemporary French, but I do not speculate on this further here.

the second case, an example of VOS, the 'lords and ladies' are directly addressed in the speech of the king of Portugal, evidently the focused information is the object: the faith which the speech garnered.

Some marginal examples of VSO and VOS orders also remain in the 17th century, as shown below.

(41)	jnvocquet	le	non	de	la	vierge	marye	les	uns	pour	avoir
	invoked	the	name	of	the	Virgin	Mary	the	ones	to	have
	alegeance	de	la	fiebvre	et	pour	aultres	maladye			
	protection	of	the	fever	and	for	other	illness			

'Some invoked the name of the Virgin Mary to have protection from fever and other illnesses'

Valuche p. 8v

These examples can only be found in texts of the *peu lettrés* and it seems that VSO and VOS orders are all but disallowed in more standardised and formal language from the 17th century. However, we will see in section 4.6 and discussion in Chapter 7 that some examples survive in Modern French. At this stage the available subject positions become more fixed. DP subjects now seem to typically occur post-verbally only when no object is present. Often, these examples are restricted to passives or presentative clauses. These examples are also almost always triggered by pre-posed frame-setting adverbials. Examples of this pattern were given above in (34).

In summary, it seems that post-verbal DP subjects initially continue the Old and Middle French patterns in the 16th and early 17th centuries, before the variety of positions reduces, with the loss of VSO orders. At this stage, specific factors related to pre-posed content, transitivity of the verb and information structure become most important. This diachronic shift is of particular interest given comparative work on post-verbal subjects. Lahousse and Lamiroy (2012) have specifically argued for an approach in which Spanish is less advanced than Italian which is less advanced than French with regard to changes in VS word orders. The fact that we can find further available subject positions in Old French is suggestive of the view that the Romance languages have undergone similar changes since the Medieval period, losing their complex and well-articulated set of subject positions, but to different extents. Thus, the study of change in French, which appears to be the most advanced in this

respect, allows us to diachronically examine the various layers which are synchronically visible in the Modern Romance languages. It would be extremely valuable to examine the different post-verbal subject positions available in Medieval Italo-Romance and Ibero-Romance varieties to further this work, particularly examining whether the extra low subject position visible in French is found in these varieties or not. It may also be the case that originally information-structurally unmarked orders in the medieval varieties of these languages developed their current information-structurally marked functions over time, just as the narrower functions of Modern French post-verbal subjects have developed since the Early Modern period.

It also seems that some of these post-verbal positions are linked to null subjects, at least until the 18th century. Those Early Modern texts with the highest rates of post-verbal DP subjects prior to the 18th century also have high rates of null subjects. For instance, Valuche features 17.5% null subjects, alongside 5% post-verbal DP subjects. Similarly, Chavatte features 9.25% null subjects, alongside 18.5% post-verbal DP subjects. All other texts show under 0.5% null subjects and under 1.5% post-verbal DP subjects. Notably, neither Valuche nor Chavatte show higher rates of post-verbal pronominal subjects, so this appears to be a distinct link between post-verbal DP subjects and null subjects. Examples of null subjects in Valuche (42) and Chavatte (43) are given below to illustrate:

(42) ont couche deulx nuictz a cande
have slept two nights at Candé

'They slept at Candé for two nights'

Valuche p. 2v

(43) et le 6e jour de decembre par le jour de st
and the sixth day of December by the day of Saint

nicolas environ les 7 heures du soir sorti hors de
Nicholas around the seven hours of.the evening went out outside of

la ville de lille
the city of Lille

'And on the sixth day of December on the day of St. Nicholas, around seven o'clock in the evening, he went out from the city of Lille'

Chavatte p. 167r

This apparent relationship between null subjects and post-verbal subjects is expected from a comparative perspective; those Modern Romance languages which maintain post-verbal subjects also maintain null subjects and it has often been suggested that some relationship exists between the two features (e.g. Rizzi, 1982; Cardinaletti, 2004; Laenzlinger and Soare, 2005: 39), while a link between V2 and null subjects has been noted in the history of French (Adams, 1987; Roberts, 1993; Wolfe, 2022 among others).

Overall, we have seen that Early Modern French exhibits a wide variety of subject positions similar to those of Old French. During the 16th century, many of these low positions are lost. Only a restricted set of post-verbal subject positions, as described in the GGHF and in the work of Lahousse (2003, 2006, 2022) remain.

4.6 Modern and Contemporary French

Finally, I turn to our Modern and Contemporary French data. These data are extremely sparse and so much of this discussion will rely on the existing literature on post-verbal subjects. It should be noted, however, that post-verbal subjects appear to have become a marker of written and formal language; the only texts of our 20th- and 21st-century data sources which featured any post-verbal subjects in their first 400 clauses were André Gide's diary and travel writing as well as our forum data. It is also interesting to note that post-verbal DP subjects were attested across all of these, while post-verbal pronominal subjects were only found in André Gide's writing. I would point out, however, that we are focusing exclusively on declarative main clauses here, while post-verbal pronominal subjects are typically discussed in their uses in interrogatives. An interesting text-typological stratification presents itself here, with post-verbal subjects all but extinct in speech and the most dialogic written genres (text messages and letters) but still found in more literary writing and even in the marginally more formal environment of the online forum.

In our data, it seems that the majority of post-verbal subjects appear to be cases of idioms or fossilised structures, this is particularly the case in more colloquial data, such as our forum data.

(44) Plus	vous	continuerez,	plus	dure	sera	la	chute
More	you	will.continue	more	hard	will.be	the	fall

'The longer you continue, the harder the fall will be.'

YCQQA corpus date-stamp: 2009-01-11 10:50:41

We see in this example that a somewhat formal style is used, with an idiomatic turn of phrase, one of the only times that post-verbal subjects are found in more colloquial data. On the other hand, post-verbal subjects seem to remain productive in the more literary and typically written data, such as André Gide's diary and travel texts.

(45) a. Mais peut-être vaut -il mieux qu' elles ne viennent
But maybe is.worth it better that they NEG come

pas trop en avant
NEG too in advance

'But maybe it's better if they don't come too far in advance'

André Gide, Diary, 3rd Oct 1924

b. Je me decide à quitter Paris (**ne fût -ce que pour**
I REFL decide to leave Paris NEG was it only for

quelques jours)
a few days

'I decide to quit Paris (although this was only for a few days)'

André Gide, Diary, 11th Nov 1924

c. Trop nombreux sont en France, aujourd'hui, ceux qui ne
Too numerous are in France today those who NEG

s' intéressent plus qu' aux vieilleries ou
REFL interest anymore just to.the old-fashioned things or

qu' aux sornettes
just to.the drivel

'In France, those who now only interest themselves in old-fashioned things and drivel are too numerous'

André Gide, Diary, 11th Nov 1924

In (45c) we see that post-verbal DP subjects are still productive in Gide's writing, particularly when triggered by certain pre-verbal adjectives. In (45a) and (45b) we see the productivity of post-verbal

pronominal subjects, where various factors, such as pre-verbal adverbs (45a), expletive subjects (45a,b) and use of negation (45b) can, in combination, trigger the post-verbal placement of the subject. This is in-line with the description given in the GGHF, where a combination of factors triggers post-verbal placement of the subject, rather than one particular trigger. Notably, this seems distinct from the triggers of post-verbal placement of a DP subject, where, as Lahousse (2006) describes, either a trigger word is found, in cases of genuine stylistic inversion (46), or the post-verbal subject is exhaustively focused, in cases of focus inversion (47). These two cases are exemplified below with examples from Lahousse (2006: 444) contrasting those in (45).

(46) Alors commença pour les miens la traversée de
Then began for the mine the crossing of

la douleur…
the pain…

'Then the painful years began for my loved ones…'

(47) Rendront un devoir les élèves qui ont raté l'
Will hand in an assignment the students who have failed the

examen de chimie.
exam of chemistry.

'The students who have failed the chemistry exam will hand in an assignment.'

It is also notable that VOS orders are attested here, as in (47), despite the considerable drop in frequency of such orders in our own data. We see here a distinction between productive use of post-verbal subjects, though rare, in higher register writing, and non-productive, fossilised examples in colloquial use. Indeed, this finding is in line with Lahousse's (2022: 19) argument that post-verbal subjects have remained productive in journalistic writing. Lahousse (2022: 19) specifically notes the productivity of *nombreux* as a trigger for post-verbal placement of the subject, shown here in our example (45c). It seems that in Modern French, post-verbal subjects in declarative main clauses are now particular to certain registers of writing, while elsewhere they only survive in idioms. Since their analyses are not reflected in detail in my own sparse data, I will provide further discussion of Lahousse's (2003, 2006, 2022) analyses in Chapter 7, in our discussion of the broader connections between syntax and information structure through the history of French.

4.7 Concluding Remarks

This chapter has provided a wide-ranging overview of the history of post-verbal subjects in declarative main clauses in French. From this discussion, it is clear that pronominal subject positions are highly limited, seemingly exclusive to Spec-TP. While the verb may move above this position and post-verbal pronominal subjects appear to become syntactically cliticised after the 16th century, no other positions are available. On the other hand, DP subjects show multiple different positions. Firstly, the subject may occupy the canonical position in the Spec-SubjP (a phrase just above TP). Secondly, the subject may occur in a mid-field position, higher than Belletti's (2001) VP-peripheral FocP, but lower than the mid-field subject position found in the Ibero-Romance languages according to Ordóñez (2007). I have tentatively argued that this position must be Spec-*v*P. This position has been diagnosed through the relative positioning of quantifier objects and adverbials, as well as the occurrence of VSO orders with minimal information-structural restrictions. This position appears to be available until approximately the 17th century. Next, a position akin to that of Belletti's (2001) VP-peripheral FocP appears to be available. However, following Lahousse's (2006) argumentation for Modern French, we have seen that the positions of floating quantifiers suggest a remnant movement analysis. Thus, this 'low' position in fact represents a case of movement to the left-periphery, with the remnant clause inverting. This analysis is also the one we provide for instances of VOS, which also survive into Modern French in Lahousse's (2006) data. During the 16th and 17th centuries, options for these lower positions become more limited, with licensing requirements developing as described by Lahousse (2022) alongside a loss of VSO structures. The motivation for this loss of VSO structures remains unclear, but it may relate to a lowering of the post-verbal subject, since Lahousse (2003, 2006) goes so far as to argue that it is VP-internal in Modern French. The situation described above continues until the Modern and Contemporary French period, with the potential pre-verbal triggers narrowing further in the 17th and 18th centuries. In the 20th and 21st centuries, new productivity develops in certain written registers, despite the loss of post-verbal subjects in declarative main clauses in speech (Lahousse, 2022).

To fully illustrate, I provide a bracketed schema below, illustrating the surface position of subjects, objects and the verb found at different periods in the history of French. I do not illustrate VOS orders, since, as described above, these are cases of a remnant movement construction, illustrated in section 4.2.2.1.

(48) a. [$_{SpecFinP}$ X (incl S) [$_{Fin}$ V$_{finite}$] … [$_{SpecSubjP}$ S$_{DP}$ … [$_{SpecTP}$ S$_{Pron}$ … [$_{FP}$ [$_{F}$ V$_{non\text{-}finite}$] …

[$_{Spec\textit{v}P}$ S$_{DP}$] … [$_{SpecObjP}$ O … [$_{VP}$]]]]]] (Old and Middle French)

b. [$_{SpecSubjP}$ S$_{DP}$ [$_{Subj}$ V$_{finite}$] [$_{SpecTP}$ S$_{Pron}$ [$_{T}$ V$_{finite}$] … [$_{FP}$ [$_{F}$ V$_{non\text{-}finite}$] …

[$_{Spec\textit{v}P}$ S$_{DP}$] … [$_{SpecObjP}$ O … [$_{VP}$]]]]] (Renaissance French – Contemporary French)

Example (48a) illustrates the subject positions of Old and Middle French. The pre-verbal position in Spec-Fin may be occupied by a variety of constituents, represented by X, including all types of subject; the finite verb occupies the head of Fin, due to the V2 constraint; the DP subject in cases of simple inversion occupies Spec-SubjP, while the post-verbal pronominal subject occupies Spec-TP; the non-finite verb is in the head of an FP, below TP and Cinque's (1999) adverbial spine but above *v*P; the low DP subject, found in free inversion, occupies Spec-*v*P; finally, the object occurs in a low specifier position just above the VP, but below *v*P, I use ObjP here, but do not make any strong theoretical claims about this position. Example (48b) illustrates the subject positions starting in Renaissance French. Here, the finite verb no longer moves to the left periphery, instead surfacing in either Subj (as in cases of pronominal and complex inversion) or T (as in SVO sentences or examples of free inversion). Subject movement to Fin is also lost due to the loss of V2, but all other subject and object positions remain the same. VSO orders are lost in the 17th century, potentially due to the development of an even lower subject position, although we have insufficient data to make a firm claim. The frequency of different positions varies considerably, but these positions are also those found in Contemporary French.

This diachronic account accounts for the majority of post-verbal subjects in declarative main clauses across the history of French. It should be noted that we have not discussed some more idiomatic examples, for instance post-verbal subjects with verbs of speech, focusing instead on the

overall structure of the clause. We have also seen that this diachronic account would be useful to supplement with comparative data, particularly following Lahousse and Lamiroy's (2012) comparative work on Romance subject positions. Further investigation of the histories of Spanish and Italian would be particularly useful to determine whether the information-structurally under-specified low subject position of Old French is attested elsewhere.

5. Dislocations

5.1 Introduction

I have now shown how the cartographic framework can provide a detailed and systematic description of post-verbal subject positions within the clause. Next, I turn to a left-peripheral construction: dislocations.

Dislocations have been extensively studied in Modern and Contemporary Spoken French (Ashby, 1982, 1988; Auger, 2003; Delais-Roussarie, Doetjes and Sleeman, 2004; De Cat, 2007; Donaldson, 2011). Similarly, a modest literature exists on dislocations in Old and Middle French (Troberg, 2004; Engel, 2009; Mathieu, 2012; Salvesen, 2013; Donaldson, 2016). Dislocations in Old and Middle French have sometimes been argued to be distinct constructions to their modern counterparts (Marchello-Nizia, 1998).[19] Left and right dislocations are also noted in grammars of the intervening periods (on 16th-century French see Gougenheim, 1973; on 17th-century French, Fournier, 2001). Before proceeding to an in-depth discussion of the structures and the literature regarding them, I provide some illustrative examples. Examples of both structures from Contemporary Spoken French are given below in (49) and (50), both taken from my Contemporary Spoken French data from the CFPP2000 corpus.[20]

(49)	Enfin	**l'**	**école**	**à**	**aires**	**ouvertes**	**c'**	est	une	école
	In fact	the	school	with	areas	open	it	is	a	school

expérimentale (left dislocation)
experimental

'Well, the open-air school, it's an experimental school.'

CFPP2000 OZGUR_KILIC_H_32_ALII_3E [49.72]

(50)	On	**le**	kiffait	bien	**le**	**troisième** (right dislocation)
	We	it	liked	well	the	*troisième*

[19] Marchello-Nizia (1998) makes this argument on the basis of rare examples in Old French in which the dislocated element is a focus, not a topic. See below for further discussion.

[20] It should be noted that I give idiomatic translations throughout, rather than directly translating the dislocated structures.

'We liked the *troisième*.'

CFPP2000 OZGUR_KILIC_H_32_ALII_3E [176.624]

In example (49), a left dislocation, the dislocated element 'l'école à aires ouvertes' is positioned to the left of the clause and resumed by the clitic 'c''. In example (50), a right dislocation, the dislocated element 'le troisième' is positioned to the far right of the clause, resumed by the clitic 'le'. It should be noted that several further types of XP may be dislocated, not just DPs, and different resumptives may be used, not just subject and object clitics (examples from Angelopoulos and Sportiche, 2021: 961).

(51)	**À**	**Paris**,	on	**y**	va	souvent
	to	Paris	we	LOC	go	often

'We go to Paris often.'

(52)	**Que**	**Marie**	**est**	**coupable**,	on	**le**	sait
	that	Mary	is	guilty	we	it	know

'We know that Mary is guilty.'

Despite the extensive existing literature on dislocations there is almost no diachronic literature on dislocations in French and no literature discussing their development from the Old and Middle French periods to Contemporary Spoken French. In this chapter, I examine the full history of dislocations in French. I show that the syntax of these constructions is extraordinarily stable across their history despite a recent increase in frequency and subtle shifts in function and distribution.

The chapter is structured as follows. In section 5.2, I provide an in-depth overview of the literature on dislocations throughout the history of French. In section 5.3 I provide details of the methodology for the present study, focusing on the specific methodological quirks of dislocations, while the general methodology of this thesis was discussed above in Chapter 3. In section 5.4 I discuss my Old and Middle French data, providing an analysis of the distribution, function, and syntax of dislocations in this period. In section 5.5, I discuss the data from Early Modern French texts from the 16th to 19th centuries, focusing primarily on function and distribution, since syntactic analysis is difficult with such sparse data. In section 5.6, I turn to Modern and Contemporary French

data, focusing on the written-spoken divide. Finally, in section 5.7, I provide concluding remarks and discuss the overarching history of dislocations.

I argue that, although dislocations are extremely rare throughout much of the attested history of French, they are also consistently found across texts of various genres and text-types. They appear to have had the same function throughout their history, with different distributional factors affecting their frequency. In the same way, they seem to exhibit the same syntactic behaviour throughout their history, in line with a base-generated left-peripheral analysis. Despite this stability in form and function, dislocations show a sharp increase in frequency in our Contemporary French data. I argue that these sorts of shifts in frequency despite stable form and function are an important phenomenon for historical linguists to describe.

5.2 Background

I begin with an overview of previous studies on dislocations. Firstly, I provide a brief overview of the descriptive literature on dislocations in Contemporary French before turning to previous work on the history of dislocations, particularly in Old and Middle French. Finally, I provide a more in-depth discussion of the pragmatic function and syntax of these constructions.

5.2.1 Contemporary French

Both left and right dislocations are found throughout Contemporary French, as illustrated from the CFPP2000 corpus in examples (53)-(54) (left dislocations) and (55) (right dislocation) below.

(53)	Speaker 1:	**Etienne**	**il**	était	dans	une	autre	école
		Etienne	he	was	in	an	other	school
	Speaker 2 (Etienne):	oui	**moi**	**j'**	étais	à	rue des Vertus	
		Yes	me	I	was	at	*rue des Vertus*	

Speaker 1: 'Etienne was at another school.'

Speaker 2 (Etienne): 'Yes, I was at *rue des Vertus*.'

CFPP2000 OZGUR_KILIC_H_32_ALII_3E [73.605] - [76.156]

(54)	**Ma mère**	**elle**	y	vit	toujours
	My mother	she	LOC	lives	still

'My mother still lives there.'

CFPP2000 OZGUR_KILIC_H_32_ALII_3E [163.719]

(55)	**Il**	est	tout	prêt	**le**	**jardin**	**des**	**Plantes**
	It	is	all	close	the	garden	of.the	plants

'The botanic garden is very close.'

CFPP2000 CHRISTOPHE_ANDRE_H_62_MARIE_ANNE_ANDRE_F_63_5E [676.706]

Dislocations are generally considered to be used to mark the topic of a clause (Lambrecht, 1981; De Cat, 2007). Left dislocations can also be used to introduce a new topic, as seen in (54), where the speaker's mother has not been mentioned in the previous discourse. Right dislocations on the other hand seem to require the referent to be given or at least accessible, as illustrated in (55), where the *Jardin des Plantes* was already the topic of the ongoing discourse. In Contemporary French, dislocations are extremely common, with Delais-Roussarie, Doetjes and Sleeman (2004) stating that '[dislocation] is even more used than the simple SVO order in spoken French' (Delais-Roussarie, Doetjes and Sleeman, 2004: 506). That said, outside of spoken French, dislocations are considerably less common, with Auger noting that they are 'banished from standard and literary French'[21] (Auger, 2003: 381). Despite this, sociolinguistic factors do not seem to be relevant within spoken French. Ashby (1988) finds that dislocations are not associated with age-grading or class but appear to be a feature of 'unplanned discourse' (Ashby, 1988: 226).

5.2.2 Old and Middle French

Both left and right dislocations can be found in Old French, with left dislocations occurring as early as the Chanson de Roland (56). Right dislocations are extremely rare and rarely discussed, although their presence is acknowledged in the GGHF (Marchello-Nizia et al. 2020: 316).

(56)	**Li**	**quens**	**Rollant,**	**il**	est	mult	irascut
	The	count	Roland,	he	is	very	angered.

'Count Roland is very angered'

(1100-ROLAND-V,62.737)

[21] My translation from the original French.

(57) Aprés **il** parlerent ensanle, **li** **pelerin et** **li**
Afterwards they spoke together, the pilgrims and the
Venicien
Venetians

'Afterwards the pilgrims and the Venetians spoke together'

(1205-CLARI-P,45.1119)

Traditionally dislocations are thought to be rare in Old French. Mathieu (2012) argues against this, providing numerous examples, although no quantitative data. Most studies of Old French simply assume the topicalisation function of left dislocations (see Mathieu, 2012; Salvesen, 2013; Donaldson, 2016). However, Marchello-Nizia (1998) challenges the assumption of the topic function with the following example:

(58) Orgoill oi e folage
Proud hear and madness

Ço set hom ben, n' ai cure de manace
This knows man well NEG have cure of threat

Mai **saives** **hom**, **il** deit faire message;
But wise man he must do message;

Si li reis vœlt, prez sui por vus le face!
If the king wishes ready am for you it do

'I hear proud words, and mad ones. It's well known, I have no cure for the threat, but a wise man must deliver the message. If the king wishes, I am ready to do it for you.'

Chanson de Roland ll. 292-295

In this section, the left-dislocated constituent 'saives hom' 'wise man' is clearly being used contrastively and emphatically, in the role of a focus. One could in no sense say that it is 'a wise man' that this clause is about, rather, delivering a message is distinctly the topic. Indeed, Marchello-Nizia (1998) shows this reflected in modern translations of the text, which often use clefts, a structure widely noted to mark focus, to translate the left dislocation. This example seems a strong case for the view that Old French left dislocations had different information-structural properties from Modern French. If this were the only example of an apparent focalisation function of dislocation in Old French, it may be best to treat it as anomalous. However, in the GGHF, several further examples are

provided. Although some seem to be disputable, several are quite clear in their focalisation function. For instance, another example from the GGHF (Marchello-Nizia et al. 2020: 1319) is given below (emphasis and translation mine):

(59) 'Dites, qui est ceste meson?
Tell who is this house?

Qui est li sires, conme a non?'
Who is the lord how for name?

Premiers parla Melyagés
Firstly spoke Meliagés

qui fu cousin Ethïoclés:
who was cousin Ethéocle

'Sire,' fet il, 'que voulez vos ?
Lord, said he what want you ?

De nostre seingneur et de nous
Of our lord and of us

vous respondrai em pes sanz ire ;
you will reply in peace without anger

Ethïoclés, **il** est mes sire'
Ethéocle he is my lord'

'Tell me, whose is this house? Who is its lord, what is he called?' Meliagés, who was Ethéocle's cousin spoke first: 'Lord,' he said, 'what do you want? Regarding our lord and us, I will reply to you in peace and without anger; Ethéocle is my lord.'

Roman de Thèbes, ll. 2969-2976

In this example, as the translation suggests, Ethéocle is being introduced not as a new topic for the continued discourse, but rather as a focus, this is emphasised new information. Indeed, the *wh*-question 'Qui est li sires, conme a non?' 'Who is its lord, what is he called?' clearly illustrates this, with Ethéocle reflecting the *wh*-word, a classic test for focus. It should be noted that both of these examples are from verse texts and thus may not be the most reliable sources for syntactic data. However, recent work by Samek-Lodovici (2015, 2022, Forthcoming) has shown that focused hanging topics can also be found in extremely marginal contexts in Modern Italian. Although the sparsity of

examples in Old French limits our discussion here, testing our own historical data for further examples of this focalisation function is important to exploring the history of dislocations.

5.2.3 Early Modern French

Turning to grammars of Early Modern French, we find that dislocations were attested throughout the history of French. Gougenheim (1973) notes, albeit in passing, the existence of dislocations in 16th-century French and likewise, Fournier (2001: 98) notes their presence in 17th-century French. It seems likely, then, that the structure has existed continuously with some narrowing of information-structural function overtime. The GGHF specifies the Middle French period as that in which the focalisation function was lost. Additionally, while Marchello-Nizia (1998) does cite an example of a focused left dislocation, topicalised left-dislocated content is by far the majority in her data.

Although minimal research has been conducted on the Early Modern period, the differences in the frequency of dislocations between Old and Modern French have led some to postulate a sudden change at this time. Wolfe (2021: 76), specifically suggests that the rise in the frequency of dislocations between Old and Modern French occurs in response to the drop in frequency of bare topicalization (topicalization without clitic resumption). The drop in bare topicalization seems to occur most sharply in the Early Modern French period, with Wolfe (2021) finding no cases of bare topicalization of objects in his corpus of 16th-18th-century texts (Wolfe, 2021: 76), despite their high frequency in Old French (Wolfe, 2021: 53). De Andrade (2018) notes the disappearance of aboutness topicalization by the 16th century. As such, we might expect to see a sudden rise in the frequency of dislocations during the Early Modern period.

5.2.4 Formal Studies

In the formal literature, the analysis of dislocations has been hotly debated. The difficulty in providing a satisfying analysis of left dislocations in large part derives from the mixture of movement and non-movement properties which they display, an issue known as Cinque's Paradox (Iatridou, 1995). Left dislocations do not yield weak cross-over effects, nor do they license parasitic gaps (see De Cat, 2007: 118-120 for a discussion). However, left dislocations are sensitive to strong islands in many Romance languages (Cruschina, 2021: 4-5), for instance standard Italian. While these features are known to be

features of left dislocations across a number of languages, De Cat (2007: 121-134) observes that French left dislocations are not in fact sensitive to strong islands, something which should be taken into account when evaluating theories of left dislocation developed from non-French data. Thus, there is no clear evidence for French left dislocations to be moved at all and Cinque's Paradox does not apply to Contemporary French data.

Turning to the existing analyses, numerous options are available. Even without Cinque's Paradox in action, some have still argued for a movement analysis (e.g. Angelopoulos and Sportiche, 2021), others a base-generation analysis making use of the expanded left periphery (Frascarelli, 2004; Frascarelli and Hinterhölzl, 2007) and yet others have rejected Rizzi's (1997) left periphery, opting for a stripped back base-generation account using adjunction (De Cat, 2007).

Various alternative ways of sub-dividing dislocated constructions should also be noted at this stage. While I focus here on only those which feature a clitic resumptive, numerous other formal distinctions have been made, including some which cut across the surface-level distinctions of resumed and unresumed dislocation. These distinctions are key to some of the analyses discussed above, for instance Angelopoulos and Sportiche's (2021) analysis. For Angelopoulos and Sportiche (2021), two types of dislocation are attested: Hanging Topic Left Dislocation (HTLD) and Clitic Left Dislocation (CLLD). For Angelopoulos and Sportiche (2021) HTLD is only found in root contexts, whereas CLLD is found in both root and non-root contexts. Reconstruction effects are found with CLLD but not HTLD. HTLD is limited to DPs while CLLD can use PPs, CPs or AdjPs. CLLD is recursive whereas HTLD can only occur once and CLLD must be resumed by a clitic pronominal, whereas HTLD can be resumed by a tonic pronoun or an epithet. I provide examples below to illustrate:

(60) a. **À** **Paris**, on **y** va souvent
to Paris we LOC go often

'We go to Paris often.' (Angelopoulos and Sportiche, 2021: 961)

b. **Plastic Bertrand**$_i$, j' ai tous les disques de **ce**
Plastic Bertrand, I have all the records of this

farfelu$_i$
weirdo

'I have all of Plastic Bertrand's records' (De Cat, 2007: 109)

In (60a) we see an unambiguous example of CLLD since the left-dislocated constituent is a PP. On the other hand, in (60b) we see an unambiguous HTLD, since the resumptive is an epithet, not a clitic pronoun. It should be noted that the diagnostics for CLLD and HTLD are not undisputed. Indeed, De Cat (2007: 106-108) discusses the various, often contradictory, proposed diagnostics at length, showing that many of the classic diagnostics do not seem to hold in French. Likewise, Helland, Nilsen and Lohndal (2020) have shown the extreme difficulties of sub-dividing topicalisation structures into HTLD and CLLD. Thus, in this chapter, we discuss all cases of clitic resumed dislocation, assuming that these represent the same construction, following De Cat (2007).

With this established, I give a description of several analyses of left dislocations. I begin with a movement analysis, from Angelopoulos and Sportiche (2021) and three cartographic analyses: Rizzi's (1997) original analysis, Frascarelli's (2004) base-generation analysis and Frascarelli and Hinterhölzl's (2007) elaboration on this. Finally, I discuss De Cat's (2007) base generation analysis, which rejects cartographic analysis of the left periphery. I note that other analyses, most notably Ott's (2015) biclausal analysis, are available, but these are not well-suited to French data. Ott's (2015) analysis only allows for one dislocated element, as we find in Modern German, while French allows an astonishing number of stacked dislocated elements. Thus, I do not provide further discussion of this analysis here.

As noted above, Angelopoulos and Sportiche's (2021) analysis distinguishes clearly between CLLD and HTLD. Thus, their analysis is only describing some of the phenomena we will be examining. Angelopoulos and Sportiche's (2021) analysis is entirely based on data from left-dislocated contrastive topics. This is potentially problematic given that the diagnostic of the contrastive topic as CLLD is highly disputed. Indeed, as De Cat (2007) points out, some have even made the inverse claim regarding information-structural roles, while others have claimed that HTLD and CLLD have the same pragmatic function (see De Cat, 2007: 106-108 for discussion). This sub-division appears to be considerably clearer in those languages where Cinque's Paradox is active in CLLD. However, in

French this is not the case, indeed, as De Cat (2007: 134-139) discusses, it seems that the same locality restrictions are found across both kinds of construction.

In Angelopoulos and Sportiche's (2021) analysis, left dislocation consists of two steps: firstly, an A-movement step from the base-position to the specifier of the appropriate CliticP in the mid-field, then two A'-movement steps, firstly to the T-domain and then to the left-periphery. To support this analysis, Angelopoulos and Sportiche (2021) present a variety of reconstruction effects which imply that some A'-movement steps must have occurred (Angelopoulos and Sportiche, 2021: 1001-1004). While these locality effects are indeed difficult to account for, I do not discuss them here since Angelopoulos and Sportiche's (2021) analysis suffers from more fundamental issues. Firstly, Angelopoulos and Sportiche's (2021) analysis is highly complex and yet accounts only for the dislocation of contrastive topics, since it is explicitly not applicable to HTLD. In addition, this analysis treats CLLD as directly analogous to clitic-doubling with added movement to the left-periphery (Angelopoulos and Sportiche, 2021: 985). When we consider that French subject and object clitics show such distinct levels of cliticization (see discussion in Chapter 4 and Chapter 7), we would expect significant differences between subject and object dislocations, which are not attested. Finally, although this analysis accounts for some difficult reconstruction effects, it is left with the issues of every movement analysis of dislocation: accounting for its ability to violate strong islands and other non-movement properties. As such, I do not adopt Angelopoulos and Sportiche's (2021) analysis.

Turning to some left-peripheral analyses, I describe Rizzi's (1997) classic analysis of left dislocations. Under this analysis, left-dislocated topics are base-generated in their standard clause-internal position before being moved to the left-periphery, to the specifier of a dedicated TopP. The lack of weak cross-over effects is then accounted for by suggesting that these effects are linked to quantificational A'-movement, rather than all A'-movement. Additionally, it is argued that this analysis is preferrable to adjunction due to a variety of locality effects, such as left dislocations' sensitivity to strong islands. However, as we have established, French left dislocations are not sensitive to strong islands.

As such, I turn to approaches which argue for base-generation of the left-dislocated topic in the left-periphery. Frascarelli (2004) and Benincà (2006) argue that the left-dislocated topic still occupies the specifier of TopP, as in Rizzi's (1997) analysis. However, rather than moving to the position, it is base-generated there. Against the movement analysis, Frascarelli (2004) argues that Rizzi's (1997) analysis fails to deal with some of the atypical features of left dislocations. Although the lack of weak cross-over effects and parasitic-gap licensing are accounted for by Rizzi's (1997) analysis of topics as non-quantificational A'-movement, other non-movement features are found. For instance, clitic-doubling requires us to provide an analysis which allows for multiple-realisation of the subject, although this is disallowed in all contexts other than left dislocations and right dislocations (Frascarelli, 2004: 102). In order to account for the strong island sensitivity found in Standard Italian, Frascarelli (2004) then suggests that resumed topicalization must happen locally, in order to license *pro*, while Italian clitics occupy a dedicated clitic position. Thus, left dislocation from subordinate clauses does involve movement, simply from the subordinate TopP to the matrix TopP (Frascarelli, 2004: 114-116). In order to derive right dislocations, Frascarelli (2004) proposes IP-inversion (in our terms, TP-inversion). In this analysis, right dislocations are not syntactically different from left dislocations but involve an extra step where the IP is raised to a higher position in the left-periphery, resulting in a right-dislocated order. Frascarelli's (2004) analysis was further developed by Frascarelli and Hinterhölzl (2007), who propose further subdivisions in the left periphery, arguing that the ordering of topics is in fact quite restricted, as described in section 2.3 above. This is then used to account for the limited information-structural function of right dislocations, which are said to obligatorily occupy Spec-FamP.

It should be noted that these analyses are based primarily on Italo-Romance data. To utilise Frascarelli's (2004) analysis on French data, some further stipulations are required, allowing dislocations to be base-generated in the non-local TopP. I now turn to the approach of De Cat (2007), an analysis focusing exclusively on dislocations in French.

In many ways, De Cat's (2007) analysis can be seen as analogous to that of Frascarelli (2004): the left-dislocated constituents are not analysed as moved, instead they are first-merged into their

clause-peripheral position. However, rather than utilising the cartographic analysis of the left periphery, De Cat (2007) argues that the dislocated constituent adjoins directly to the highest node of the clause. De Cat (2007) argues that this allows us to account for the lack of restrictions on the number and order of topics since they can simply be freely adjoined to the root node and this can be done recursively. As with Frascarelli's (2004) analysis, this approach also allows us to account for the presence of resumptives. The resumptive clitic can now be treated as a full syntactic component and the main clause as syntactically complete, with the adjunction of a topic as an essentially optional syntactic operation performed purely for information-structural purposes. Important to De Cat's analysis is the presentation of new grammaticality judgement data on French left dislocation examples, demonstrating that, unlike Italian (among others), French left dislocations are insensitive to strong islands (De Cat, 2007: 125-129). This new finding considerably strengthens the case for any base-generation or adjunction analysis of dislocations in French. One issue with De Cat's (2007) approach is that it struggles to capture the distribution of left dislocations and right dislocations with regard to subordinate clauses. De Cat (2007) argues that dislocations can be found in matrix clauses and 'dependent clauses with root properties' (De Cat, 2007: 152). However, at no point is a clear, independent definition of these 'root properties' given and, indeed, De Cat herself notes that the notion of 'root clause' used here is quite different from the traditional notion (De Cat, 2007: 158).[22] Thus, much of De Cat's argument becomes circular. A further issue with De Cat's (2007) account is her rather broad understanding of the notion of topic. In De Cat's (2007) view left dislocations are consistently clause-initial. However, numerous examples of frame-setting clauses preceding left dislocations can be found, as given below, from an Early Modern French text:

(61)	Le	31	de	juillet	la	grosse	cloche	de	l'	eglise	de	sainct
	The	31st	of	July	the	big	bell	of	the	church	of	saint
	pierre		elle	fut	menez		par	les	paroisses		de	la
	Peter		she	was	brought		through	the	parishes		of	the
	ville		de	lille								
	town		of	Lille								

[22] See Miyagawa (2022: 144-149) for an in-depth discussion of the concept of root and topicalization, arguing that non-root contexts are presupposed contexts.

'On the 31st of July, the big bell of the church of Saint Peter was brought through the parishes of the town of Lille'

Chavatte p. 166v

In De Cat's (2007: 71-74) view these frame-setting clauses should also be viewed as topics and thus their consistent position to the left of left dislocations is unaccounted for. In an account utilising the expanded left-periphery however, this is readily accounted for, since frame-setting clauses occupy a higher clausal position in Frame and will thus precede left dislocations. We have seen that De Cat's (2007) analysis captures a number of particular details of the Modern French data. However, on closer inspection, all of these same details can be captured by any base-generation approach. The addition of fine-grained distinctions between frame-setters and left dislocations provided by a cartographic approach marks this as the superior current theory.

5.3 **Methodology**

The broad methodology, including corpus and text-selection, has been described above in Chapter 3. Here, I provide further details of methodological particulars of this case study.

5.3.1 Frequency Data

5.3.1.1 Old and Middle French Data

For Old and Middle French data, the MCVF and Penn Supplement parsed corpora were searched using the CorpusSearch program. The MCVF and Penn Supplement's tagging system does not distinguish between resumed or unresumed dislocation and has no system of annotation for coreference. As such, this process required numerous rounds of searching for different sub-types of dislocations as well as manual checks of the results to confirm that all and only legitimate examples of resumed dislocations were counted.[23] One issue with the Old and Middle French period is the preponderance of null subjects. Following a base-generation analysis, these null subjects should be able to act as a resumptive, leaving no surface sign of clitic dislocation. The MCVF and Penn Supplement's annotation includes null subjects, however, many cases of dislocation are ambiguous

[23] For the set of CorpusSearch queries used, see Appendix C.

between dislocation with a null resumptive, bare topicalization or non-dislocation. Thus, we are left with two options: leaving this interpretive issue to the corpus annotators or focusing solely on those examples which feature an overt resumptive. Unfortunately, having performed test searches, some highly ambiguous examples are flagged as left dislocations, such as (62), where 'li empereres' appears more likely to be a case of bare topicalization, particularly given that it precedes frame-setting clauses. That said, more unambiguous examples are also found, marked by strong pronominals in the dislocated position (63).

(62) Li empereres, ce sachiez, quant l' oï, si en fu mout
The emperor this know, when it heard, *si* CL was much

liez
obliged

'Know this, the emperor, when he heard this, was much obliged'

(1190-BORON-PENN-R,57.881)

(63) Et luy, des courtois le plus honorable, la baisa
And him of.the courtiers the most honourable, her kissed

doulcement
softly

'And he, the most honourable of the courtiers, kissed her softly'

(145X-CNNA-P,431.9182)

As well as this issue with automated searching for left dislocations, right dislocations raise further issues, since several examples which may be cases of post-verbal subjects are flagged. Consider (64) below. In this case the weak pronominal 'je' is used in the potentially right-dislocated constituent 'je et ma mere aussi', but the constituent occurs between the main verb and a subordinate clause, thus seeming considerably more likely to be a case of a post-verbal subject.

(64) Au mains li prierai je et ma mere aussi que il
At hands him will pray I and my mother also that he

pense de vostre besoingne
think of your need

'My mother and I will pray at his hands that he might think of your needs'

(127X-CASSIDORUS-PENN-P,23.556)

As such, the corpus cannot currently be reliably interrogated for potential cases of dislocation with a null resumptive. As a result of these issues, an investigation of null subjects and dislocations is beyond the current scope of this work. Additionally, it should be noted that these null resumptive structures are extremely infrequent in the corpus, with 21 left dislocation examples and eight right dislocation examples across the entirety of the MCVF and Penn Supplement. Thus, the exclusion of these data is highly unlikely to skew our results.

5.3.1.2 Early Modern, Modern and Contemporary French Data

As noted in Chapter 3, a group of hand-selected texts and corpora were used for our Early Modern, Modern and Contemporary French data, representing a variety of different genres and registers. For every text or corpus, the first 400 clauses were gathered, and the frequency of dislocations noted. The word order of each clause was also noted, as well as the status of the dislocated constituent as an adverb, PP, CP, object or subject, as pronominal or full DP and as a light or heavy XP, as well as the person and number of any pronominal dislocations.

For a full list of all the texts and corpora used, see the appendices and discussion in Chapter 3.

5.4 Old and Middle French Data

5.4.1 Frequency Data

I now turn to data presentation, beginning with my Old and Middle French data. Before discussing information structural functions and the syntactic analysis of dislocations in Old and Middle French, I first discuss the frequency of dislocations in the language of these periods. I begin with the frequency of left dislocations throughout the Old and Middle French periods. These data are given in Figure 7:

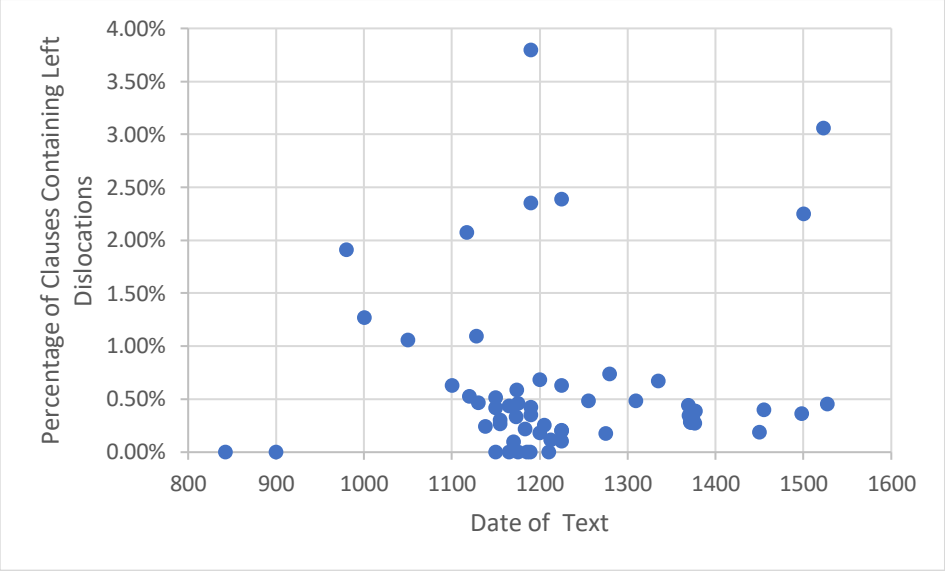

Figure 7 : Frequency of Left Dislocations in Old and Middle French Data

As can be immediately seen, left dislocations are consistently infrequent but also consistently present across the period. We can also see that, while the bulk of texts show fewer than 1% of clauses featuring dislocations, there are several outliers. These outliers do seem to group together in some way. All those texts with more than 3% of clauses featuring left dislocations are translations of religious texts, which likely leads to a less natural use of the language. Additionally, both the Lapidal, an encyclopedia of stones, and Bestiaire, a bestiary, show above 2% of clauses featuring left dislocations. If we examine the texts, we can see why. Formulaic structures, introducing each stone or animal to be discussed and then resuming them in the main clause, lead to an anomalously high rate of left dislocations. Two examples of this are given below:

(65) **Aptalon** **ço** est beste
Antelope it is beast

'The antelope is a beast'

(1128-BESTIAIRE-BFM-R,28.363)

(66) **Draconitides** **ço** est un nom de pere qui vient de
Draconitides it is a name of stone which comes from

dragon
dragon

'Draconitides is a name of a stone which comes from a dragon'

(1117-LAPIDAL-BFM-R,229.480)

These examples also show us that the left dislocations at the time fulfilled a similar function to Modern French: introducing a new topic. Each of these dislocations introduces the previously unmentioned topic of the entry, a new topic, before proceeding to discussion.

One outlier remains: the letters from the Correspondance Estouteville, these texts show a left dislocation rate of over 2%, shown on our graph as a point at 1500. While this rate is undoubtedly considerably higher than in literary texts and suggests that dislocations may have been more common in day-to-day language, it should be noted that this rate remains well below rates in Contemporary Spoken French, where over 15% of clauses contain a left dislocation.

Once these outliers have been accounted for, the remaining data show consistently low rates of left dislocations. These rates do not show any further patterning with any text-typological factors. It should also be noted that the rate of dislocations seems to be quite stable across this period.

Having discussed the frequency of left dislocations in this period, I turn to right dislocations. These data are given in Figure 8:

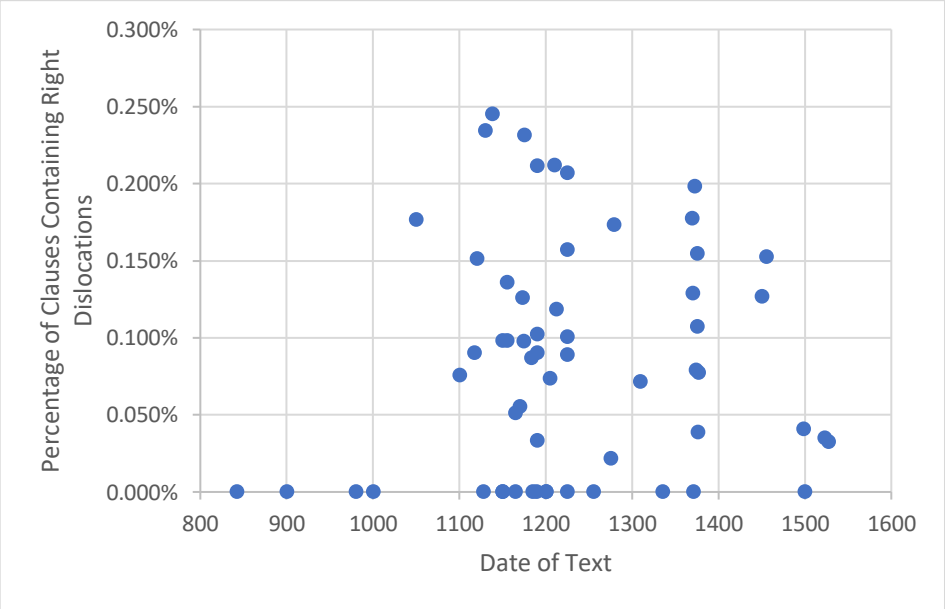

Figure 8 : Frequency of Right Dislocations in the MCVF and Penn Supplement

Right dislocations are even more vanishingly rare in Old and Middle French but, again, are consistently found throughout the corpus, including in the earliest text with over 1000 clauses: Alexis.[24] Once again, the variation in rates of right dislocations seems to be essentially random, with little or no influence of genre, register or other text-typological factors. This can be particularly well illustrated by the series of data points before 1400, all of which are sections of the first book of Froissart's Chroniques. In the MCVF corpus, this text has been split into 'eight roughly equal-sized sections'.[25] Thus, if any text-typological factors played a role, we would expect these sections to cluster, however, instead, we see that they display a range of frequencies of right dislocations, supporting the view that this variation is unpatterned.

In summary, it can be clearly seen that dislocations, both left and right, have been present since the earliest texts of the French language. These structures are not a new development, despite their much higher frequency in Contemporary Spoken French. Additionally, in the Old and Middle French data, they do not seem to be a particular stylistic marker but are instead found across texts from a wide variety of genres and registers. Having shown this, I turn to their function and syntax, to see how these structures have changed over the history of French.

5.4.2 Distribution and Function

5.4.2.1 Weight

One striking feature of left dislocations in all of the historical data is the importance that weight seems to play in their usage. Weight is a somewhat vague notion in linguistics, with two primary components. Firstly, a DP can be considered heavy if it contains a subordinate clause. Secondly, some studies have shown that long DPs act as heavy DPs (see Wasow, 1997 for an in-depth discussion of these two aspects of weight). The notion was initially utilised to account for so-called 'Heavy NP shift' in the Germanic languages, but it seems to also play a role here. I will also be extending the concept to non-DPs, with the same criteria of length and structural complexity. Of all DP left

[24] It should be noted that one anomalous outlier has been excluded from Figure 8, since it distorts the data. The text, *Quatre fragments de miracles de la Vierge* is only 98 clauses long, although only one of these clauses features a right dislocation, this produces a misleadingly high percentage, thus I have excluded it for the purposes of this discussion.

[25] For more information, see the corpus documentation: https://www.ling.upenn.edu/~beatrice/corpus-ling/french-corpora-sources/index.html

dislocations in the MCVF and Penn Supplement, 72% feature heavy DPs, either in the sense of featuring a subordinate clause, or in the sense of being four or more words long. To further illustrate, examples of heavy left dislocations are given below.

(67) Cel damoisel ki cele lance tient an sa main par
This squire who this lance holds in his hand as

conoissance, cil ira primes de ceüs al soverain air
emblem, this will go first of heavens to the highest air

la desus
there above

'This squire who holds this lance in his hand as emblem will go first to the highest air of the heavens'

(1155-ENEAS1-BFM-R,90.1948)

(68) Adam e Salemun e Davit e Samsun il furent deceü
Adam and Solomon and David and Samson they were deceived

e par femes vencu.
and by women defeated.

'Adam and Solomon and David and Samson were deceived and defeated by women'

(1128-BESTIAIRE-BFM-R,104.1380)

The importance of weight to Old and Middle French dislocations is of particular interest given the high frequency of extremely short, single word dislocations of the sort 'moi, je pense' 'me, I think' in Contemporary Spoken French. It seems that, during this time of low frequency, dislocations may have been associated with more particular functions related to syntactic weight.

5.4.2.2. Information Structure

I next turn to the information-structural function of dislocations in the texts from this period. As we saw in section 5.2, it has been claimed that a focalising function can be found for left dislocations in the Old and Middle French period, and some convincing examples have been provided. Looking through our data from the MCVF and Penn Supplement however, these examples were extremely rare. Only one further unambiguously focused example could be found, this is given below:

(69) Et coment ot ele a non? Abisac en droit non, et
And how has she as name Abisac in true name and

Sunamitis ce fut li sornons.[26]
Sunamitis this was the nickname

'And what's her name? Abisac as true name and Sunamitis was her nickname'

(122X-AGNES-PENN-P,252.406)

In this example, we can see that the left-dislocated element: Sunamitis corresponds directly to the *wh*-word in the previous clause. As discussed in Chapter 2, this is an extremely strong test for a focused constituent. The name itself: Sunamitis, is new, focused information, while the aboutness topic of these clauses is the notion of the name. Thus, focalised examples of left dislocations can be found in Old French data (as described by the GGHF, we do not find any examples in Middle French texts). However, the examples are extraordinarily scarce. Additionally, both examples given in section 5.2 are from verse texts, while the example given above is from a translation of a hagiography. This suggests that this extremely marginal function may only be used as an artistic flourish. The verse examples might be accounted for with meter, the resumptive used to fulfil the metric requirements of the form. As noted in our introduction, Samek-Lodovici (2015, 2022, Forthcoming) has shown that focused hanging topics can also be found in extremely marginal contexts in Modern Italian. Thus, I would argue that these Old French examples represent a similar phenomenon, where focusing can be applied to these constituents in extremely marginal contexts. Unfortunately, due to the extreme infrequency of these structures, further analysis is not possible here. However, it is interesting to note that these examples are restricted to Old French, disappearing in Middle French.

Turning to the topic-marking function of dislocations, we might expect, given the lower frequency of dislocations in Old and Middle French data, that their function would be more limited, for instance to one of the topic sub-types identified by Frascarelli and Hinterhölzl (2007). However, we in fact find that the dislocations in our data display the same functions as in Contemporary Spoken French. These are exemplified below:

[26] In the MCVF and Penn Supplement *sornons* is given as *somons* while, in the BFM, it is given as *sornons*. I have inferred that this is the result of a transcription error, since <rn> is routinely mistaken for <m>, as such, I have given it as *sornons* here.

(70) Jadis le poeit hum oïr
Once it could one hear

E sovent suleit avenir,
And often it was wont to happen

Hume plusur garval devindrent
Men several werewolf became

E es boscages meisun tindrent.
And in the woodland house kept

Garvalf, **ceo** est beste salvage;
Werewolf it is beast savage

Tant cum il est en cele rage,
As much as he is in this rage

Hummes devure, grant mal fait;
Men devours great bad does

Es granz forez converse e vait.
In the great forests lives and goes.

'Once, one could hear and it was often wont to happen, that several men would become werewolves and would keep a woodland house. A werewolf is a savage beast, as long as he is in this rage, he devours men, does great evils and travels among and lives in the great forests' (116X-MARIE-DE-FRANCE-R,61.1233)

(71) Vers Sarraguce les enchalcent ferant,
To Zaragoza them chased striking

A colps pleners les en vunt ocïant,
With strikes great them CL went killing

Tolent lur veies e les chemins plus granz.
took their paths and the roads most big

L' ewe **de** **Sebre,** **el** lur est dedevant:
The water of Ebro she them is before

Mult est parfunde, merveilluse e currant;
Much is deep, marvellous and fast-moving

Il n' en i ad barge ne drodmund ne caland
It NEG CL CL has barge nor lighter nor boat

'To Zaragoza they pursued them, with great strikes they went killing, they took their paths and greatest roads. The waters of the Ebro are in front of them. It is very deep, marvelous and fast-

moving, there is neither barge nor lighter nor boat.'

(1100-ROLAND-V,180.2445)

In example (70) the werewolf has been mentioned previously, and is the ongoing topic of the lais, thus we can see that the left dislocation is used to mark a continuing topic. In example (71), the Ebro has not been mentioned at all thus far in the poem but forms the topic of the subsequent lines, describing its depth and strong current. Thus, this shows a clear example of left dislocation being used to introduce a new topic. It seems, then, that despite a much lower frequency than in Contemporary Spoken French, dislocations in Old and Middle French do not display any considerable difference in function, excepting the marginal focused examples and the importance of weight.

5.4.3. Syntax

Having discussed the information-structural function of dislocations in Old and Middle French, I turn to the syntax of the constructions. For much of our data, the tests applied by De Cat (2007) and Angelopoulos and Sportiche (2021) regarding strong islands cannot be applied. However, the large scale of the MCVF and Penn Supplement corpora, yielding hundreds of examples, allows us to make some claims regarding the syntax of these constructions. One striking fact in our data is that examples of dislocations across strong islands are found from the earliest texts. Consider the following example from the Chanson de Roland:

(72) **Carles** **li** **velz** **a** **la** **barbe** **flurie**,
Charles the old with the beard flourishing

Jamais n' ert jurn qu' **il** n' en ait doel
Never NEG was day that he NEG CL has suffering

e ire
and anger.

'As for Charles the old, with the flourishing beard, there will never be a day when he doesn't have suffering and anger'

(1100-ROLAND-V,77.931)

Here we can see that the resumptive *il* is in the subordinate clause of a complex DP 'jurn qu'il n'en ait doel e ire', thus, were the left-dislocated constituent moved, it would have to be extracted from this

strong island. Likewise, we can find examples of dislocated constituents whose resumptive is found in one half of a coordinate structure:

(73)	Or	avint		ainsi	que	**mon**	**seigneur**		**Phelippe**,		que
	Now	happened		thus	that	my	lord		Philip		that
	toute	nostre	gent	estoient	pris	et	**il**	ne	le	fu	pas,
	all	our	people	were	taken	and	he	NEG	it	was	NEG
	pour ce que		il	estoit	message.						
	because		he	was	messenger						

'So it happened that all our people were taken but my lord Philip was not, because he was a messenger'

(1309-JOINVILLE-P,153.2192)

Here the resumptive *il* is found in the second half of the coordinate structure 'toute nostre gent estoient pris et il ne le fu pas', where A'-movement restrictions only allow movement out of both halves of a coordinate structure. It should be noted that coordinate structure constraint violations are also permitted in some languages where Cinque's Paradox is active, such as German (Fernández-Sánchez and Ott, 2020: 14). Nonetheless both these examples, particularly (72) are in line with De Cat's (2007) examples from Contemporary Spoken French and support the view that dislocations have had the same syntactic behaviour historically. One difference from Contemporary French can be seen in example (73), however. In (73), we see the reduplication of the complementizer *que* both before the dislocated constituent and before the clause from which it was extracted. Indeed, this seems to be the standard means of constructing left dislocations out of subordinate clauses in Old French, although the second complementizer does appear to be optional. An example with no second complementizer is shown below:

(74)	Et	dit	ainsi	que	**qui**	**vouloit**	**tuer**	**la**	**serpent**,	**il**
	And	said	thus	that	who	wanted	kill	the	snake,	he
	li	devoit	premier		esquacher	le	chief.			
	it	must	first		crush	the	head			

'And he said thus that whosoever wanted to kill the snake must first crush its head'

(1309-JOINVILLE-P,89.1229))

The topic of double complementizer constructions has been studied extensively in other Romance languages[27] and been discussed for Old French by Salvesen (2014), Salvesen and Walkden (2017) and Wolfe (2021) as well as being discussed for the Ternois dialect of Picard by Dagnac (2012) and Gascon by Ledgeway (2020). In these analyses, complementizer doubling is treated through the expanded left-periphery of Rizzi (1997). In this analysis, the higher complementizer is found in Force, while the lower complementizer is found in Fin, with any intervening material occurring in the intervening phrases. This analysis works perfectly for our data, with the left-dislocated elements occurring in the Topic field, between both complementizers. Indeed, this is the analysis proposed by Wolfe (2021: 68) for such structures. These data also create issues for De Cat's (2007) analysis, since the dislocated element does not appear to be attaching to the highest node of the clause, but instead, more complex ordering operations appear to be in place in the left-periphery, supporting the use of an expanded CP for the analysis of dislocated structures.

To summarise and fully illustrate the analysis proposed for our Old French data, I provide a bracketing of (69) below, showing a minimal example, as well as a bracketing of (72) to show how a more complex example would be analysed. Note that, for simplicity, I use a simplified TP to model the core of the sentence, rather than fully modelling the Verb-Second grammar of Old French.

(75) [ForceP … [TopP [DP Garvalf$_i$] … [TP [DP ceo$_i$] [T' [T est] [VP [DP ~~ceo$_i$~~] [V' [V ~~est~~] [DP beste saulvage]]]]]]]

(76) [FrameP [AdvP Or] … [TP [DP ø] [T' [T avint] … [AdvP ainsi] … [VP [V' [V ~~avint~~] [ForceP [Force que] [TopP [DP mon seigneur Phelippe]$_i$ … [FinP [Fin que] [ConjP [TP toute nostre gent estoient pris] [Conj' [Conj et] [TP il$_i$ ne le fu pas pour ce que il estoit message]]]]]]]]]]

With regard to right dislocations, even the data in the MCVF and Penn Supplement are too sparse to apply the tests utilised by De Cat (2007). As such, I do not discuss their syntax here.

In summary, we have found that, despite their considerably lower frequency, dislocations in Old and Middle French do predominantly behave as their Contemporary French equivalents. Although heavy constituents are considerably more frequent in left dislocations compared to

[27] On Italo-Romance see Ledgeway (2005), Paoli (2007) and Munaro (2018); On Spanish see Villa-García (2012); On Portuguese see Mascarenhas (2007)

Contemporary French, the constructions do not otherwise behave differently. They do not appear to be a product of syntactic movement and they are used to mark the topic of a clause, excepting marginal focalisation examples.

5.5 Early Modern Data

Having given my analysis of the function and syntax of dislocations in Old and Middle French, I turn to my Early Modern French data.

5.5.1 Frequency Data

I begin with a discussion of frequency data. Only left dislocation data is given since, outside of Héroard's record of the Dauphin's speech, no text contains more than one example of right dislocations in the first 400 clauses.

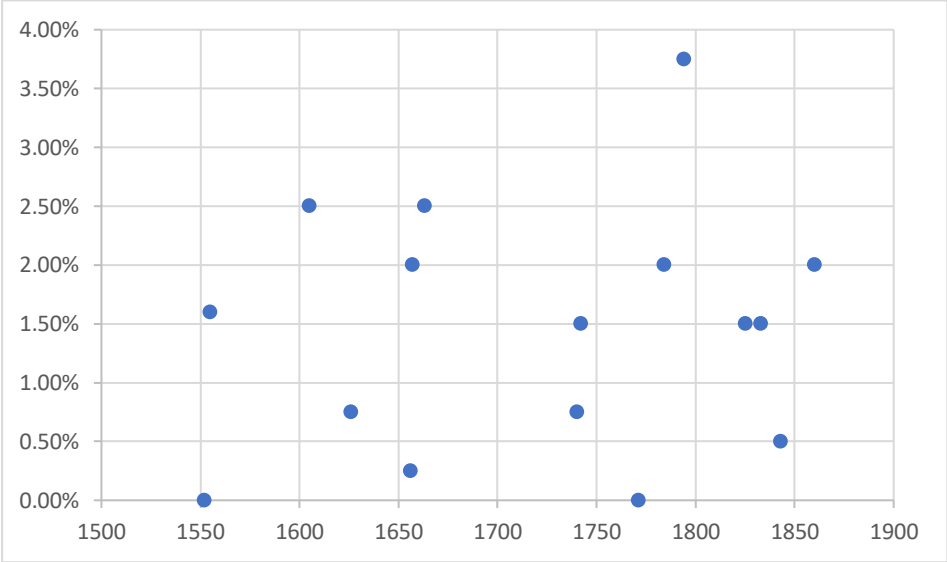

Figure 9 : Frequency of Left Dislocations in Early Modern French Data

In the Early Modern French period, we see a continuation of the Old and Middle French patterns. The frequency of dislocations of all kinds is consistently low throughout the entirety of the period examined, at most 3.75% of clauses in a given text. It should be noted that this does represent an increase from Old French but the frequency of left dislocations is still low compared to Contemporary Spoken French. While this small increase in frequency might be attributable to the loss of bare topicalization, as Wolfe (2021) suggests, we cannot attribute the higher frequency of

dislocations in the contemporary language to this change. It should be noted that Wolfe's (2021) argument is specific to the loss of object bare topicalization. Unfortunately, the rates of left dislocations in our data are too low to make quantitative claims regarding different types of dislocated constituents. However, it should be noted that, just as in Old French, both objects, subjects and non-DP constituents are found left-dislocated, suggesting, along with the stable frequency of dislocations, that no considerable change has occurred. Additionally, sociolinguistic variation in these data is minimal. Most notably, private texts of less educated authors do not show a higher rate of dislocations than published texts. Data comparing private with public texts is shown in Figure 10 to illustrate:

Figure 10 : Frequency of Left Dislocations in Public and Private Early Modern texts

There is one outlier text here: Héroard's record of the Dauphin's speech. This text is a pseudo-phonetic recording of the Dauphin's childhood speech by his physician Jean Héroard. As such, this seems to suggest that dislocations are associated specifically with spoken contexts, and that traditional notions that the private writing of lesser educated speakers is more 'spoken-like' (see, e.g. Elspaß, 2012) may be incorrect. Again, however, these texts still show considerably lower rates than Contemporary Spoken French. Notably, this pattern suggests that dislocations do not have any sociolinguistic association with class or register, but may be associated with spoken and interactional

language, regardless of register (note, for instance, how the Dauphin's speech, despite its extremely high register,[28] shows higher rates of dislocations than written texts).

5.5.2 Metalinguistic Data

Metalinguistic data further supports the view that these constructions are rare and do not vary across register. Two of the three metalinguistic commentators examined, Maupas and Vaugelas, make explicit reference to dislocations. Both these writers only discuss left dislocations with no mention of right dislocations. The discussion provided by Maupas is given below:

> 'on peut dire Moy, Je croy, Toy, Tu lis, comme distribuant a chacun son office et action'
> 'one can say me, I think, you, you read, as if distributing to each person their position and action' (Maupas, 1618: 121).

This example is of particular interest since it seems to show a striking awareness of the use of left dislocations to mark specifically contrastive topics. In this case, the left-dislocated elements *moy* and *toy* are clearly being used in contrast against each other, while still the topic of their respective clauses, reflecting an oft-observed function of left dislocations in Contemporary French.

Vaugelas provides some discussion of the appropriate contexts for repeated subjects (what we have been analysing as cases of left-dislocation). An example of this discussion is given below:

> 'si le nominatif […] est fort esloigné du verbe substantif, alors il est bien mieux de dire *ce*, que de ne le dire pas comme *en fin la cause de tant de malheurs et de miseres qui nous arrivent en ce monde les une sur les autres, c'est &c*' 'if the nominative […] is very distant from the verb, it is better to say *ce* [i.e. reduplicate the subject] than to not say it like "in the end, the cause of so many miseries which befall us in this world, one upon the other, it is etc."' (Vaugelas, 1647: 304).

Just as Maupas' discussion reflects the contrastive topic function of dislocations, the role of weight in dislocations can be seen here. Vaugelas' discussion suggests that dislocation was seen as necessary to remind the reader of the subject after extensive modifiers.

[28] Although the Dauphin's speech is extremely informal, being that of a child to their loved ones, I still consider it to be high register, due to his position at the apex of French society and his education, even at a young age.

Bouhours also gives an example of left dislocation as *bon usage*, although it is found in a separate discussion related to lists as subjects:

> 'Gloire, richesses, noblesse, puissance, ce ne sont que des noms imaginaires'
>
> 'Glory, riches, nobility, power, these are just imaginary names' (Bouhours, 1692: 15).

Again, we can see here that grammarians understand the role of weight in dislocations.

While some discussion of dislocations is visible in the metalinguistic texts examined, there is no extensive discussion and no evidence that dislocations are seen as constituting improper language use. This relative lack of commentary seems to suggest that the structures were indeed rare, as well as having little associated sociostylistic meaning, positive or negative.

5.5.3 Distribution and Function

I now turn to a discussion of the function and distribution of the dislocations in the Early Modern French period.

With regard to information structure, the marginal focalisation function found in Old French appears to have disappeared at this stage. As in Contemporary French, however, left dislocation can still be used to promote totally new information to the role of topic, this is exemplified below:

(77)	Ausie	**le**	**pere**	**jerome**	**quy**	**etoit**	**pecheur,**		je	**le**	faisoit
	Also	the	father	Jerome	who	was	fisherman		I	him	made
	bien	enrager	en	alant	sur	cest	bascule	et	prenant	son	
	well	angered	CL	going	on	this	plank	and	taking	his	
	bateaux	pour	me	promener		sur	la	riviere.	il	m'	aimoit
	boat	for	REFL	wander		on	the	river	He	me	liked
	et il	disoit	toujours	qu'	il	le	diray	a	mon	pere.	
	and he	said	always	that	he	it	say	to	my	father	
	Mais	comme	il	connoisoit		ses	promtitude		envert		moy
	But	since	he	knew		his	promptness		towards		me
	cela	faisoit	qu'	il	ne	disoit	rien.				
	This	meant	that	he	NEG	said	anything.				

'Also, I angered Father Jerome, who was a fisherman, by going on this plank and taking his boat to wander on the river. He liked me and always said that he'd tell my father. But, as he was aware of his promptness with me, it meant that he said nothing.'

Ménétra p. 16

In example (77), Father Jerome has not been mentioned at all in the prior discourse, nonetheless the focus here is that the author angered him. Indeed, Jerome goes on to be the subject and topic of the following clauses. This shows the use of left dislocation to promote a new discourse constituent to topic.

Turning to the distribution of dislocations. Weight seems to play a significant role in left dislocations, with 62.2% of dislocations at this period featuring a heavy XP, as in example (77) above, or example (61) reprinted below as example (78):

(78)	Le	31	de	juillet	la	grosse	cloche	de	l'	eglise
	The	31st	of	July	the	big	bell	of	the	church

de	sainct	pierre	elle	fut	menez	par	les	paroisses	de
of	saint	Peter	she	was	brought	through	the	parishes	of

la	ville	de	Lille
the	town	of	Lille

'On the 31st of July, the big bell of the church of Saint Peter was brought through the parishes of the town of Lille'

Chavatte p. 166v

Weight does not seem to play a role in right dislocations, although our data on them is so sparse that any meaningful commentary is not possible. It should also be noted that the effect of weight is considerably weaker in Héroard (where half of the left dislocations feature heavy XPs) and the Vendée letters (where a third of the left dislocations feature heavy XPs). We might take this to suggest that the use of dislocations in non-interactional writing differs somewhat from their use in interactional contexts.

In addition to information structure and weight, some formulaic structures seem to be used, for instance pseudo-clefts of the form 'ce que... c'est...':

(79)	**Ce**	**qui**	**me**	**consolait**	**c'**	est	que	je	n'	étais	pas
	it	which	me	consoled	it	is	that	I	NEG	was	NEG

maltraité
mistreated

'What consoled me was that I wasn't mistreated.'

Bédé p. 79

In summary, as in Old and Middle French, it seems that dislocations have a somewhat formulaic use and preference for heavy DPs which is linked to their lower frequency. Nonetheless, in terms of information-structural function and syntax, they do not seem to behave differently to Contemporary French.

5.6 Modern and Contemporary French Data

Having discussed our Old, Middle and Early Modern French data, I now turn to our data from Modern and Contemporary French.

5.6.1 Frequency Data

Firstly, I discuss the frequency data from those texts and corpora examined, beginning with left dislocations, shown in Figure 11:

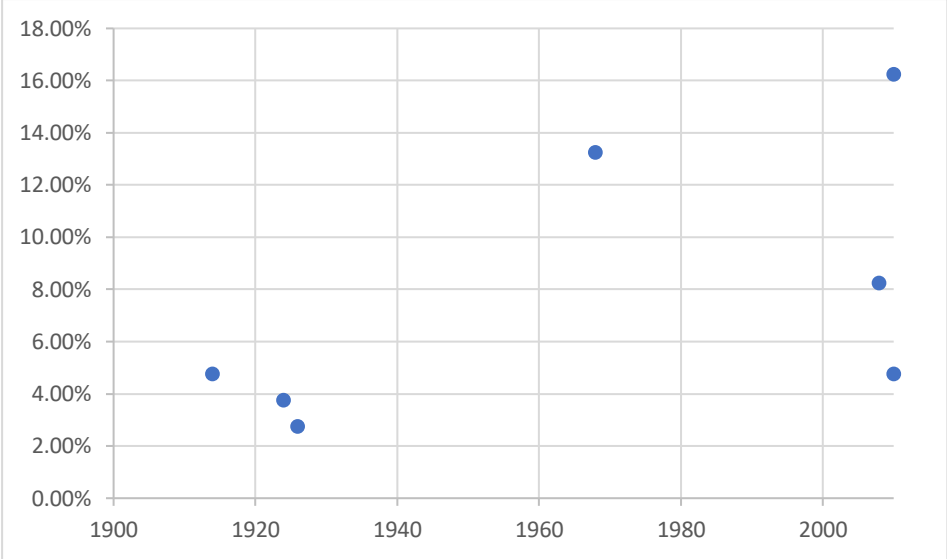

Figure 11 : Frequency of Left Dislocations in Modern and Contemporary French Data

A divide is immediately clear between the spoken and written data. The spoken corpora used are shown in the two highest points. Forum data, SMS messages and letters come next, being relatively closer to spoken language with an 8% rate of dislocations in the forum data and 4.75% in both the SMS and letter data. Finally, André Gide's diary and travel writing show the lowest rates. This stratification of text-types is striking for a number of reasons. Firstly, it confirms the oft-made observation that dislocations might be seen as a predominantly spoken phenomenon (Ashby, 1988; Auger, 2003; Delais-Roussarie, Doetjes and Sleeman, 2004). Secondly, it contrasts heavily with the historical data, in which text-typological factors appear to play little to no role. Particularly concerning is the large gap between the spoken data and even the spoken-like text sources: the letters, forum data and SMS data. This might suggest that some spoken phenomena are not fully represented in text, even the most spoken-like texts, and may cause an issue for any historical examination of a phenomenon like dislocation. Of course, the letters are a full 50 years prior to the earliest spoken data we have, so this may play some role. Additionally, SMS data is potentially unreliable, since SMS messages have been identified as exhibiting specific kinds of argument-drop and subject omission, even in languages where these phenomena are not licensed in any other context (see Stark and Robert-Tissot, 2017; Stark and Meier, 2017; Haegeman and Stark, 2021). However, our forum data,

although showing higher frequency than SMS data, still displays under half the frequency of left dislocations of contemporaneous spoken data.

I next turn to right dislocations, the frequency data for which are given in Figure 12:

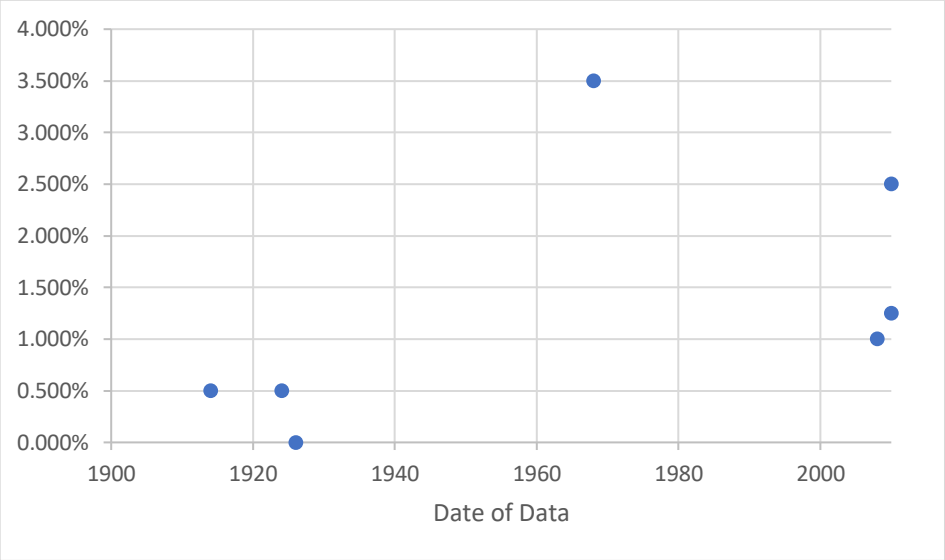

Figure 12 : Frequency of Right Dislocations in Modern and Contemporary French Data

Once again, we see that dislocations are by far most common in speech. However, we do not see the stratification so clearly seen in the left dislocations. Indeed, dating seems to be a factor here, with the early-20th-century texts showing by far the lowest rates. We should also consider whether right dislocations may be an even more characteristically spoken phenomenon than left dislocations, as suggested in the historical data by the higher frequency in Héroard's record of the Dauphin's speech. However, further data from this period are needed to make any stronger claims regarding the impact of text-type and the spoken-written divide.

5.6.2 Distribution and Function

Through the 20th century, the distributional factors identified for Old, Middle and Early Modern French appear to become less important. Firstly, weight seems to have become considerably less important, with a substantial increase in pronominal dislocations and single word dislocations. Dislocations of the type 'moi, je pense que...' are extremely frequent. Indeed, in our extract of the CFPP2000 corpus, only 15.4% of left dislocations feature heavy DPs, compare this to the

approximately 70% in both our Old, Middle and Early Modern French data. This seems to reflect a substantial shift in the use of dislocations, from a marginal structure with a particularised function for heavy topics, to a more generalised topicalization device. It seems that this distinction reflects both a historical trend and a distinction between spoken and written language, with 30.3% of left dislocations in our forum data featuring heavy DPs, much lower than in Old, Middle and Early Modern French, but still approximately twice that of our spoken data. Even in a relatively recent and non-literary text, such as André Gide's diary, 66.7% of left dislocations analysed feature a heavy DP dislocated, almost as high as the rates found in medieval texts. This strongly suggests that the distinction is primarily one between written and spoken language and secondarily a historical development, which has transferred back to written language in the most colloquial registers.

With regard to information-structural function, the claims of the existing literature seem to be correct in our data. Left dislocations can refer to any topic, including new and contrastive, while right dislocations tend to be familiar topics. All of these functions are illustrated below.

(80) Speaker 1 (interviewer) : et vous n' avez pas de [pause] cinéma
and you neg have NEG of [pause] cinema

de quartier [pause] rue de la Clef c' est peut être un peu
of *quartier* [pause] *rue de la Clef* it is maybe a bit

loin déjà ou des choses comme ça euh [pause]
far already or some things like that uh [pause]

Speaker 2 : on va de temps en temps [pause] au Ciné action en bas
we go sometimes [pause] to the *Ciné action* down

là qui passe des vieux films rue des Écoles [pause] **puis**
there which shows some old films *rue des Écoles* [pause] then

sinon moi je suis pas très cinéma et toi quand tu
otherwise me I am neg very cinema and you when you

y vas tu vas n'importe où
CL go you go wherever

Speaker 1: 'And you don't have a local cinema, *rue de la Clef* is already maybe a bit far or things like that uh'

Speaker 2: 'We occasionally go to the *Ciné action* down there which shows old films on *rue des Écoles* but otherwise, I'm not very into cinema and when you go, you go wherever.'

CFPP2000 CHRISTOPHE_ANDRE_H_62_MARIE_ANNE_ANDRE_F_63_5E [1616.561] - [1623.811]

(81) Speaker 1: le plus loin c' est [pause] la Cité de la musique
the most far it is [pause] the *Cité de la musique*

où on met [pause] quarante-cinq minu- quarante minutes
where we put [pause] forty-five minu- forty minutes

quoi [pause] en métro
what [pause] in metro

Speaker 2 : hein [pause] **et** **ça** **c'** **est** **le** **bout** **du** **monde**
uh [pause] and that it is the edge of the world

Speaker 1: 'The furthest is the *Cité de la musique*, which takes about 40-45 minutes by metro'

Speaker 2: 'and that's the edge of the world.'

CFPP2000 CHRISTOPHE_ANDRE_H_62_MARIE_ANNE_ANDRE_F_63_5E [1652.595] – [1660.178]

(82) Speaker 1: euh onzième [pause] onzième artisans plutôt euh [pause]
uh *onzième* [pause] *onzième* artisans more uh [pause]

onzième enfin
onzième really

Speaker 2 : autour de Saint Antoine
around *Saint Antoine*

Speaker 1 : euh [pause] autour de Saint Antoine [pause] euh [pause]
uh [pause] around *Saint Antoine* [pause] uh [pause]

Speaker 3 : **c'** **est** **le** **onzième** **ça**
it is the *onzième* that

Speaker 1: 'uh the *onzième*, the *onzième*'s more artisans, the *onzième* after all'

Speaker 2: 'around *Saint Antoine*'

Speaker 1: 'uh around *Saint Antoine* uh'

Speaker 3: 'That's the *onzième*'

CFPP2000 CHRISTOPHE_ANDRE_H_62_MARIE_ANNE_ANDRE_F_63_5E [2018.845] – [2024.71]

In the first of these examples, we see the use of left dislocation to mark contrastive topic. With *toi*, referencing a third interlocutor, the speaker's wife, left-dislocated in the second clause marking it as a topic contrasted against the previous clauses's topic *moi*, also marked by left dislocation. In the second example, we see left dislocation marking a continuing topic. Here *ça* refers to the area of the city the speakers were discussing prior to this point, thus continuing the existing topic of discussion. Finally, in the last example we see the use of a right dislocation to mark a continuing topic. Indeed, this use is directly comparable to that in the second example, with *ça* referring to the area of the city being discussed.

In summary, it seems that previous works on information structure correctly characterise the functions of these two structures in Contemporary Spoken French. However, looking beyond this approach, we do see a distinction between written uses of dislocations, typically resuming heavy elements and spoken uses, typically resuming short elements. As we have seen, this distinction maps to a sharp difference in frequency of dislocations between spoken and written language. This difference should not be dismissed. Indeed, some accounts of dislocations have described them in terms of interactional resources (Doehler et al., 2015), showing that they can play a role in turn-taking and interactional signalling beyond their simple information-structural roles. In this approach, left dislocations are seen to play a variety of roles in turn-taking, allowing speakers to initiate a turn without causing a total disruption to the flow of conversation. By simply uttering the left-dislocated constituent to signal their desire to begin a turn, pausing to allow other speakers to finish and then producing the full clause, chaos and confusion are avoided. For a full list of interactional functions of left dislocations, see Doehler et al. (2015: 129-131). Interestingly, Doehler et al. (2015: 130) note the list item function of left dislocations, something which we saw described above by prescriptive grammarians and which has been discussed on occasion in the generative literature as well (see Benincà and Poletto, 2004: 70). In our own, limited data, these turn-taking functions are not entirely obvious, with most dislocations featuring pronominals quickly followed by the main clause, rather than used to initiate a turn before a pause. However, this may simply reflect our relatively limited data.

Regardless, these functions are, intrinsically, specific to spoken language and this might explain the increased frequency of dislocations. It may be the case, then, that dislocations in speech and writing play somewhat different functional roles, despite the information-structural and syntactic similarities. It should also be noted that the interactional functions described by Doehler et al. (2015) relate quite directly to the information-structural roles played by these structures. It is as a result of the topical nature of left dislocations that they work so well for turn-initiation. By presenting a topical constituent, a speaker makes it clear what they are about to talk about, facilitating the smooth progress of conversation. This interaction between the syntactically encoded, information-structural functions of syntactic structures and interactional, external pressures seems central to small-scale language variation and change. Certain information-structural functions lead to a given construction being well-suited for achieving certain interactional goals, this then leads to the creation of increasingly routinised constructions, used in those particular interactional contexts, with knock-on effects in the rest of the syntax. The role of routinisation and syntactic opacity in variation and change is discussed further in Chapter 6.

5.6.3 Syntax

Turning to the syntax, the increased frequency of dislocations in our Contemporary French data allows us to apply the same analytical technique as we applied to the Old and Middle French data: qualitative analysis using specific examples.

As in our Old and Middle French data, several examples of dislocations occurring across strong islands can be found. In order to mark the multiple dislocation, one dislocate and its resumptive are marked in bold while the other pair is marked with italics:

(83)	Enfin	**moi**	*ça*	*c'*	est	pas	quelque chose	qui	**me**	perturbe
	Well	me	that	it	is	NEG	something	which	me	bothers

'Well, that's not something which bothers me.'

CFPP2000 CHRISTOPHE_ANDRE_H_62_MARIE_ANNE_ANDRE_F_63_5E [894.202]

(84)	**Moi**	*ce*	*qui*	***me***	*déplait*	*c'*	est	les	saletés	que
	Me	that	which	me	displeases	it	is	the	mess	which

laissent les animaux
leave the animals

'What upsets me is the mess that the animals leave.'

CFPP2000 CHRISTOPHE_ANDRE_H_62_MARIE_ANNE_ANDRE_F_63_5E [3315.409]

In example (83) we see that the left-dislocated pronominal *moi* is resumed in the subordinate clause of a complex DP 'qui me perturbe' a classic strong island. This suggests that the left-dislocated element cannot have been extracted via movement. In (84), the left-dislocated pronominal *moi* has been extracted from the subordinate clause in another left-dislocated DP: 'ce qui me déplait'. These examples are in line with De Cat's (2007) analysis, since they suggest the left-dislocated element is base-generated in this position rather than moved.

All of these examples can be accounted for through a modified version of Frascarelli's (2004) account. While Frascarelli (2004) suggests that dislocation out of subordinate clauses involves TopP-to-TopP movement, thus accounting for the impact of strong islands, we see here dislocation occurring out of strong islands. Thus, I argue that in French, as opposed to Italian, dislocated elements can be base-generated in either the subordinate or matrix clause, analogous to De Cat's (2007) argument that left dislocations can be adjoined to the root clause at various points in the derivation. Thus, in (84) we have a base sentence 'c'est les saletés que laissent les animaux' to which 'ce qui me déplait' is added as a left-dislocated topic, to which *moi* is then further added, also in the Topic field, since there are no locality restrictions on its relation to its resumptive.

However, De Cat (2007) also provides several counter-examples to a cartographic approach. Consider the following (De Cat, 2007: 156; bracketing original):

(85) Je pense sur son doigt, [ForceP qu' il montrait ou est
I think on his finger that he showed where is

ce que c' était sale]
it that it was dirty

'I think he was pointing to where it was dirty on his finger.'

(86) Je sais pas, les clients, [ForceP ce qu' ils veulent]
I know not the clients that which they want

'I don't know what the clients want.'

De Cat (2007) argues that these two examples show that left dislocations can attach to the highest point of the clause, including to Force, following Rizzi's (1997) analysis in which finite complementizers occupy Force. However, as we have already seen in Old and Middle French data, French finite complementizers were historically able to occur in both Force and Fin. Thus, for (85), we can simply suggest that the complementizer is in Fin, while the left-dislocated PP 'sur son doigt' is higher in the Topic field. Indeed, this is consistent with findings on French mirative focus fronting, which has been shown to be accompanied by a complementizer to the right, suggesting that a low left-peripheral complementizer survives in certain contexts in Modern French (see Authier and Haegemann, 2019: 5). Further examples of this structure can be found in our data, as below:

(87)	Par exemple		on	a	visité	un	appartement	moi	qui	me
	For example		we	have	visited	an	apartment	me	which	me
	plaisait	énorme	[pause]	énormément						
	pleased	enorm-	[pause]	enormously						

'For example, we visited an apartment which pleased me enorm-enormously'

CFPP2000 CHRISTOPHE_ANDRE_H_62_MARIE_ANNE_ANDRE_F_63_5E [560.206]

In this example we can see that, in the first clause, *moi* is left-dislocated from the subordinate clause 'qui me plaisait énorme [pause] énormément'.[29] This seems to be a clear parallel to De Cat's (2007: 156) example given in (85) above. Given the evidence for a low complementizer in some contexts in Modern French, and the regular occurrence of such examples, this seems like the clearest analysis.

Turning to (86), we can, in fact, still account for this example in a cartographic account. The key to solving this issue is the presence of null objects in Contemporary Spoken French (Cummins and Roberge, 2004, 2005; Grüter, 2006, 2009), such null objects can thus act as a clitic resumptive in left dislocation. Indeed, Cummins and Roberge's (2004, 2005) analysis of null objects argues specifically that these constructions involve a null clitic. Cummins and Roberge (2004) provide, in their own examples, cases of dislocation with a null resumptive, for instance 'Les steaks, moi, je

[29] One might suggest that this is a case of right dislocation from the matrix clause, however the pronominal mismatch between *moi* and the matrix subject *on* suggests to me a left dislocation interpretation.

préfère manger saignant.' 'I prefer to eat steaks bloody' (Cummins and Roberge, 2004: 5). Furthermore, De Cat (2007) acknowledges the possibilities of such constructions in a footnote, stating that 'the 'resumptive' can be omitted if and only if this is allowed by the grammar independent of the presence of a dislocated element.' (De Cat, 2007; 150) and going on to cite Cummins and Roberge's (2004, 2005) work. To summarize, an example such as (86) represents the exact same structure as (84) with the added step of TP-inversion. This is illustrated through bracketing below, with the landing site of TP-inversion left non-specific.

(88) [Force [XP [TP je sais pas] [TopP [DP les clients] … [TopP [CP ce qu'ils veulent] ... [FinP …

[TP ~~je sais pas~~]]]]]]

One remaining issue is how to account for those clauses in which both left and right dislocation occur simultaneously. Two examples of such clauses are given below:

(89) Je	trouve	la	Seine	à	Paris	ça	ça	c'	est	quelque chose
I	find	the	Seine	in	Paris	that	that	it	is	something
d'	important		ça							
of	important		that							

'I find the Seine in Paris is something important'

CFPP2000 CHRISTOPHE_ANDRE_H_62_MARIE_ANNE_ANDRE_F_63_5E [2178.895]

(90) Et	euh	moi	ça	ça	moi	ça	me	plaît	beaucoup	ça
and	uh	me	that	that	me	that	me	pleases	a lot	that

'And that pleases me a lot'

CFPP2000 OZGUR_KILIC_H_32_ALII_3E [430.289]

In both these examples the pronominal *ça* is both left and right dislocated simultaneously, indeed it is left-dislocated multiple times. In example (90) we also see multiple left-dislocation of *moi*. One potential solution to this is to utilise Frascarelli and Hinterhölzl's (2007) approach, in which right dislocations occur in FamP, the lowest topic position, and IP-inversion (or, in our terms, TP-inversion) occurs to an intermediary position between FamP and the higher topic positions. In many cases this would work. However, in the examples given above all the dislocated constituents, left and

right, are coreferent, additionally, in (89) the Seine was already the topic of discussion in the prior clause, thus all these constituents should occupy Spec-FamP. In truth, however, it is unclear if the pronominals in this example are dislocated at all, or whether they are simple fragments, interjected for interactional purposes. We might even suggest, then, that there is a continuum from those strictest syntactic dislocations, which may still be used for interactional purposes, as per Doehler et al (2015), and interjections with a surface-similar appearance, used purely as gap-fillers and pauses in speech. Regardless of the particular boundary between dislocations and other, less clearly syntactic constructions, we have already seen that a TP-inversion analysis appears to be the most powerful. It accounts for data which De Cat's (2007) analysis cannot and has much broader applicability than Angelopoulos and Sportiche's (2021) narrow analysis of contrastive topic dislocations.

To summarise, we have seen that Contemporary French shows a continuation of many of the patterns of older varieties with regard to dislocations; the information-structural functions remain largely unchanged. However, a huge surge in frequency occurs in the spoken language over the course of the 20th century. Additionally, previously important factors, such as the weight of left-dislocated constituents, become less important in this period. We have hypothesised that these shifts may be related. In spoken language, interactional uses of dislocations are found, leading to higher frequencies and shorter dislocated elements. In written language on the other hand, dislocations are used to resume heavy constituents where there may be confusion without resumption. Indeed, following Traugott (2010), a literature on dialogic and interactional causes of syntactic change has developed. This may be a fruitful path for further exploration of the history of dislocations.

Turning to syntactic analysis, we have seen that both Contemporary French and earlier varieties are best accounted for by a modified version of Frascarelli's (2004) analysis, in which dislocated topics are base-generated in the clausal left periphery either of the relevant subordinate clause or of the matrix clause. Right dislocations are then accounted for as cases of TP-inversion. We have seen that this analysis accounts for complementizer doubling data in Old French as well as several complex examples which De Cat's (2007) adjunction analysis fails to account for, such as (86). This account, however, does struggle to account for some cases of simultaneous left and right

dislocation and we have seen that there are other phenomena in spoken language which, while bearing surface similarities, may lie outside of the scope of such an account. I do not propose a solution to these issues here.

5.7 Concluding Remarks

Having examined each era individually, I provide an overview of the large-scale changes we have observed in dislocations across the history of French. The most notable finding of this study regarding long-term historical linguistics is the continued stability of both left and right dislocations through the history of French. We do not see considerable changes in the function or distribution of dislocations until the 20th century. Until this point their frequencies appear to be quite steady, even when considering varied text-types. This stability of frequency is particularly notable given the concurrent shifts in other types of information-structurally marked construction, for instance, bare topicalization.

Of further interest is the stability of the constructions' syntax as well as their distribution. We have seen that the Old French examples of the construction appear to obey the same syntactic rules as the Contemporary Spoken French examples, with examples of dislocations occurring across strong islands, as found by De Cat (2007) for Contemporary French.

This distinctly spoken, information-structurally marked structure remains extraordinarily stable, despite claims that interactional contexts and pragmatics drive syntactic change (Traugott, 2010; Paoli, 2020). It seems, rather than necessarily dying out or changing function, extremely marginal structures can exist in a stable form for centuries before suddenly becoming considerably more frequent. In the case of dislocations, the shift away from heavy DP topics to light topics and interactional uses of dislocation may explain the rise in frequency. However, further data from the 20th and 21st centuries are required to confirm this and a full explanation of this shift in distribution is still required. These results should be treated with considerable caution. The distinctions between spoken and written data in our 20th- and 21st-century data suggest a sociolinguistic roadblock to a historical analysis of dislocations. Even using lower register or more 'spoken-like' texts does not entirely

mitigate this divide. Thus, as argued in Ayres-Bennett (2004) and Rutten and van der Wal (2014:7-9) we should always question the potential 'orality' of a given text, and never take it as given that a certain text-type is necessarily more spoken-like.

Nonetheless, by examining the long-term history of dislocations, we have been able to provide novel syntactic analysis, deepening our understanding of the structures, as well as showing the ways in which function and frequency can vary independently of syntactic form.

6. Clefts

6.1 Introduction

This chapter is concerned with the history of clefts in French. While both post-verbal subjects and dislocations have been shown to be highly amenable to cartographic analysis, we will see that clefts present a much more complex case. Clefts in contemporary French have accrued an extensive literature (Prince, 1978; Clech-Darbon, Rebuschi and Rialland, 1999; Lambrecht, 2001; Titus-Brianti, 2014), as well as a sizeable theoretical syntactic literature cross-linguistically (Meinunger, 1998; Clech-Darbon, Rebuschi and Rialland, 1999; Belletti, 2008, 2012, 2015; Wehr, 2011, 2015; Reeve, 2011, 2012; Haegeman, Meinunger and Vercauteren, 2014; Bonan and Ledgeway, 2023). Unlike post-verbal subjects however, the historical literature is sparse, with two studies by Dufter (2008, 2009) and Rouquier's (2014) book on the topic, but little else regarding their history. Additionally, clefts have been given a variety of, often quite narrow, definitions, which typically exclude a number of surface-similar constructions which could shed light on the origins and development of clefts.

In this chapter, I discuss new data on clefts from my corpora, as well as comparing data on surface-similar constructions, with an aim to examining how clefts may have developed from or in response to those constructions. I begin by providing a descriptive overview of clefts and a summary of existing definitions of the construction. I then briefly describe several surface-similar constructions which may have intertwined histories. Next, I provide an overview of the existing literature on clefts, both the descriptive and the theoretical, before turning to my own data and discussion. I argue that, contra Dufter's (2008, 2009) work, both DP and adverbial clefts have been attested throughout the history of French as have the various diverse functions of clefts attested in Modern French. Following this, I provide a discussion of potential syntactic analyses of clefts, showing that they still elude a satisfying, unifying account.

I begin with an overview of the cleft structure, focusing here on *c'est* clefts. Although a wide variety of other constructions, including *il y a* clefts and *voici* clefts, exist, I focus here only on *c'est* clefts. A cleft is formed by an introductory matrix clause, featuring an expletive pronominal, copula

and clefted element, which may be a DP as well as several types of adverbials, followed by a relative or relative-like clause, this is illustrated in (91).

(91) C' est Jean qui est venu
It is John who is came

'It's John who came.'

Although this description provides an account of all the necessary conditions to be considered a cleft sentence, it does not exclude several surface-similar but distinct constructions. Thus, several narrower definitions have been provided, most notably that of Lambrecht (2001: 467):

(92) 'A CLEFT CONSTRUCTION (CC) is a complex sentence structure consisting of a matrix clause headed by a copula and a relative or relative-like clause whose relativized argument is coindexed with the predicative argument of the copula. Taken together, the matrix and the relative express a logically simple proposition, which can also be expressed in the form of a single clause without a change in truth conditions.'

This definition excludes surface-similar clauses with no similar function, such as the restrictive relative and the appositive, exemplified below in (93) and (94) respectively.

(93) C' est [une tarte que j' ai faite ce matin]
It is a tart which I have made this morning

'It's a tart which I made this morning.' (Rowlett, 2007: 184)

(94) C' est votre père, qui vous aime tant
It is your father who you love so much

'It's your father, who loves you very much.' (Rouquier, 2014: 21).

As well as simple, surface-similar clauses with no common function to clefts, Lambrecht's definition notably excludes clauses such as the following:

(95) C' est un grand mal que l' abus du temps
It is a great evil which the waste of time

'Wasting time is a great evil.' (Dufter, 2008: 6)

Here, the complementizer *que* does not introduce a relative-like clause, but instead seems to function as some sort of particle introducing a DP. Lambrecht's definition also excludes so-called inferential clefts, exemplified below:

(96)	C'	est	qu'	il	était	devenu vieux,	subitement
	It	is	that	he	was	became old	suddenly

'It's that he had become old, suddenly' (Dufter, 2008: 7)

Here, there is no clefted constituent. Both the types of clause in (95) and (96) are included as variants of clefts by Dufter (2008, 2009). For our purposes, the strict definition of cleft from Lambrecht (2001) will be used. However, these various other constructions will also be examined alongside clefts.

The function of clefts can typically also be fulfilled by a pseudo-cleft in which the structure is inverted, as in (97) below:

(97)	Celui	qui	est	venu,	c'	est	Jean
	The one	who	is	came	it	is	John

'The one who came was John.'

Notably, in French, pseudo-clefts require dislocation with clitic resumption (Clech-Darbon, Rebuschi and Rialland, 1999: 92), as seen in the example provided. For the purposes of our study, we will be examining both the narrowly defined cleft clauses, the variety of surface-similar clauses, and the information-structurally analogous clauses, in the hopes that this expansive analysis may provide further insights into the development of these structures.

A note on terminology is necessary before I precede. Several different terms have been used for the description of cleft clauses. For our purposes I will refer to the pronominal subject of the matrix clause (typically *ce*, although it may on occasion be another pronominal) as the expletive. The copula *est* will be referred to as the copula. The clefted element, distinguished from its relative clause, shall be referred to as the clefted constituent. Finally, the subordinate clause will be referred to as the relative clause (although it should be noted that I do not necessarily take it to be strictly speaking a relative clause).

6.2 Background

Having given a basic description of cleft constructions, I turn to the existing literature on clefts, giving an overview of both descriptive and theoretical work on the construction. I begin with a discussion of the existing literature on the function of cleft clauses, before discussing analyses of their syntax and then turning to the historical literature.

6.2.1 Information Structure

Traditionally, clefts have been seen to mark focus, even before this term was used in its modern, strict sense (e.g. Prince, 1978). However, in more recent studies, more specific functions have been identified. Usually, at least two functions are identified, with the corresponding types of clefts described by Lahousse and Borremans (2014: 803-4) as focus clefts and informative-presuppositional clefts. In focus clefts, the clefted constituent is focused. In informative-presuppositional clefts on the other hand, it is the relative clause which is focused, while the clefted constituent must be presupposed. In general, focus clefts are more common than informative-presuppositional clefts. In addition, focus clefts have been divided into two sub-types: stressed focus clefts and all focus clefts (Stark and de Cesare, 2014: 334). Stressed focus clefts are those in which only the clefted constituent is focused, while the relative clause is not. All focus clefts on the other hand, are thetic, i.e. both the clefted constituent and the relative clause are focused. We can also distinguish, amongst what Stark and de Cesare (2014) term stressed focus clefts, between those in which the clefted constituent is an informational focus, and those in which it is a contrastive focus. Indeed, informational focus clefts and contrastive focus clefts display some distinct syntactic behaviours too. These four types of clefts are illustrated below in (98) (examples a-c from Rowlett, 2007: 186, example d from Karssenberg and Lahousse, 2015: 3; translations my own):[30]

(98) a.	– Quelle		voiture	tu	as	conduit ?				
	- Which		car	you	have	driven				
	- C’	est	celle	de	Guy	que	j’	ai	conduit	
	- It	is	that	of	Guy	which	I	have	driven	

[30] For clearer exemplification I have translated all examples with clefts in English, even though these are not idiomatic in every case.

- Which car did you drive?

- It's Guy's car that I drove.

b. – C' est la voiture de Pierre que tu as conduit?
- it is the car of Pierre which you have driven

- Non, c' est celle de Guy (que j' ai conduit)
- No it is that of Guy which I have driven

- Did you drive Pierre's car?

- No it's Guy's car that I drove.

c. – Qu'est-ce qui se passe?
- What REFL happen?

- C' est Jean qui t' appelle
- it is John who you calls

- What's happening?

- It's John who's calling you.

d. Le Cécom est le distributeur de cette vidéo, c' est
Le Cécom is the distributor of this video it is

à lui qu' il faut s' addresser pour le procurer
to them that one must REFL address to it obtain

'Le Cécom is the distributor of this video, it's them who you must contact to obtain a copy.'

In (98a) we see that the specification of the car 'celle de Guy' is new to the discourse, focused information, but not contrastive against any existing suggestion, thus informational focus. In (98b) the focused constituent 'celle de Guy' is contrasted against the previously suggested 'de Pierre', thus an example of contrastive focus. In this case the relative clause may be omitted. According to De Cat (2002), the relative clause in (98b) is right dislocated, while that in (98a) is not. In (98c) the entire cleft, including the relative clause, consists of new, focused information, thus, an all-focus cleft. Finally, in (98d) the clefted constituent 'à lui' is given and topical, being the topic of the directly previous clause, while the information provided in the relative clause is new, focused information: an informative-presuppositional cleft. Notably, these different sub-types are also distributed differently across the

different kinds of clefted constituents, with adverbials strongly associated with informative-presuppositional clefts (see De Cesare, Garassino, García and Uth, 2018).

In addition to this typology of cleft functions, some unified analyses of the functions of clefts have been proposed. For instance, Lambrecht (2001) argues for a unified analysis in which the relative clause in a cleft is 'k-presupposed' where k-presuppositions are 'the set of propositions lexico-grammatically evoked in a sentence that the speaker assumes the hearer already knows or believes or is ready to take for granted at the time the sentence is uttered' (Lambrecht, 2001: 474). Cases such as informative-presuppositional clefts are then cases of pragmatic adaptation, where the structure is used to suggest that the subordinate element should be taken for granted or is beyond question. Similarly, Rialland, Doetjes and Rebuschi (2002) use prosodic evidence to suggest that the relative clause is consistently presupposed across all types of clefts. They argue, therefore, that the focused nature of the clefted constituent is generally epiphenomenal. If the relative clause is presupposed and, pragmatically, one must provide new information in an utterance, it stands to reason that the clefted constituent will generally match the function of focus.

As well as the functions of clefts proper, we should also discuss other functions of similar clauses. Two main similar clause-types are found, as noted above: restrictive relatives and appositives. Rowlett (2007) notes that clefts can be distinguished from restrictive relatives by the lack of past-participle agreement. This is exemplified below (example 88a from Rowlett, 2007: 186, 88b my own):

(99) a. C' est une tarte que j' ai faite ce matin.
it is a tart which I have made this morning

'It's a tart which I made this morning.'

b. – Qu'est-ce que tu as fait ce matin?
- What you have made this morning?

– C' est une tarte que j' ai fait(*e) ce matin.
- It is a tart which I have made this morning

- What did you make this morning?

- I made a tart this morning.

Here we see the contrast between the restrictive relative in (98a) displaying past participle agreement, and the relative clause of a cleft in (98b), not showing it. Additionally, in a restrictive relative, the relative clause is necessarily not given information and cannot be omitted, while it can be in a contrastive focus cleft. As well as these similar clause types which do not exhibit the formalised features of clefts, some specific clefts have become grammaticalized, for instance, the phrase 'c'est ainsi que…' has been argued to be lexicalised (Lahousse and Lamiroy, 2017).

In summary, clefts show a variety of information-structural functions, however, there is clear consensus in the literature on which functions are possible for clefts. Namely, informative-presuppositional, informational focus, contrastive focus and all focus functions are attested. In addition, it has been suggested that these various functions of clefts can be unified by the property of the relative clause being presupposed. As we shall see, consensus is not so well-established in the syntactic and historical literature.

6.2.2 Syntax

The syntax of clefts, both in French and cross-linguistically, is hotly debated, with a number of analyses available in the literature. These analyses can generally be divided into two main camps: the monoclausal (Meinunger, 1998; Frascarelli and Ramaglia, 2009, 2013) and the biclausal (Belletti, 2008, 2012, 2015; Reeve, 2011, 2012). In monoclausal analyses, the clefted constituent is taken to be in the left periphery of the matrix CP, with the relative clause tacitly a matrix clause, and the expletive and copula acting as some sort of discourse marker in the higher left periphery (the degree to which this is grammaticalized varies between languages and analyses). In biclausal analyses on the other hand, the relative clause is treated as a genuine relative clause, and the clefted constituent is extracted from the relative clause. It should also be noted that many of the analyses discussed here do not base themselves primarily on French data, so they should be applied to our data with caution.

A variety of both types of analysis are available. I begin with an overview of some monoclausal analyses. Firstly, the radical monoclausal analysis of Meinunger (1998), I then turn to that of Frascarelli and Ramaglia (2009, 2013). In Meinunger's (1998) rather stripped back analysis, clefting is treated as directly analogous to focus fronting. The relative clause is treated as the matrix clause,

with the clefted constituent raising to Spec-FocP, a complementiser materialising in Fin, and the expletive and copula base-generated in Spec-TopP and the Top head respectively, failing to project a full clausal projection (Meinunger, 1998: 5). It should be noted that this analysis is primarily driven by a further analysis of pseudo-clefts as derived from the simplex sentence. In this case, the remnant TP raises to above the expletive, copula and clefted constituent after they have been extracted to the left periphery. However, for our purposes I focus only on the analysis of clefts. To illustrate, I provide a bracketing below:

(100) [$_{TopP}$ c' [$_{Top'}$ [$_{Top}$ est] [$_{FocP}$ Jean … [$_{FinP}$ [$_{Fin'}$ [$_{Fin}$ qui] [$_{TP}$ ~~Jean~~ est venu]]]]]]

A similar analysis is supported by Wehr (2011, 2015), who argues that *c'est… que…* should be analysed as a discontinuous morpheme marking focus. It is unclear, under this analysis, whether clefts should be found in a complementary distribution to focus fronting, or the two should co-occur. In English, both are licit constructions. However, in Contemporary Spoken French, focus fronting is highly limited (Rowlett, 2007: 182-183), while clefting is common (Rowlett, 2007: 184). Bonan and Ledgeway (2023: 7) note a widely observed complementary distribution between clefting and focus fronting; testing this distribution through the history of French is one focus of this chapter's empirical work.

I next turn to Frascarelli and Ramaglia's (2009, 2013) analysis. This analysis is perhaps not strictly monoclausal, since there are two CPs involved, however the resultant analysis is nonetheless analogous to other monoclausal analyses since all the constituents are taken to be in the matrix left periphery. In this analysis, the base sentence consists of the expletive, copula and clefted constituent, while the relative clause is base-generated in Spec-FamP, a lower Topic position in the clausal left periphery. In order to derive the correct surface order, the clefted constituent is next raised to Spec-FocP, to mark its focused nature, before the rest of the TP inverts to the specifier of a head Ground, which hosts backgrounded information (Frascarelli and Ramaglia, 2013: 177). This is illustrated through a series of bracketings below, the full details of the moved remnant TP are not given in (101c) for brevity:

(101) a. [$_{FamP}$ [qui est venu] [$_{Fam'}$ … [$_{TP}$ c' [$_{T'}$ [$_{T}$ est] … [$_{VP}$ [$_{V'}$ [$_{V}$ ~~est~~] [$_{DP}$ Jean]]]]]]]

b. [$_{FocP}$ [Jean][$_{Foc'}$ … [$_{FamP}$ [qui est venu] [$_{Fam'}$ … [$_{TP}$ c' [$_{T'}$ [$_{T}$ est] [$_{VP}$ [$_{V'}$ [$_{V}$ ~~est~~] [$_{DP}$ ~~Jean~~]]]]]]]]

c. [$_{GroundP}$ [$_{TP}$ c'est] [$_{FocP}$ [Jean][$_{Foc'}$ … [$_{FamP}$ [qui est venu] [$_{Fam'}$ … [$_{TP}$ ~~c'est~~]]]]]]

As seen by (101a) this analysis has the appealing result of uniting analyses of clefts and pseudo-clefts. The derivation shown above is also analogous to that of right dislocations given by Frascarelli (2004) and, indeed, Frascarelli and Ramaglia (2013: 113-115) specifically argue that the relative clause is right dislocated in clefts. They support this with examples of adverbials intervening between the matrix clause and the relative clause as well as examples in which the relative clause may be omitted. It should be noted, however, that the relative clause may only be omitted in certain contexts, particularly contrastive focus clefts, as illustrated above in (98b). This analysis also predicts that *ce* is not an expletive but a referential subject co-referential with the right-dislocated relative clause, this is a notion also supported by Belletti (2008, 2012, 2015). As with Meinunger's (1998) analysis, however, it also suggests that clefting is fundamentally the same as focus fronting and the ramifications of this are unclear. The information-structural facets of both these analyses also mean that they cannot be readily extended to informative-presuppositional clefts or all-focus clefts, for which a different analysis would have to be suggested.

Turning to the biclausal analyses, I give an overview of analyses provided by Belletti (2008, 2012, 2015) and Reeve (2011, 2012). Belletti (2008) presents an initial analysis of clefts, in which they are treated as analogous to Italian free inversion. In this model, then, the relative clause is a small-clause CP, which does not project up to Force. There is an EPP feature on a left-peripheral phrase above Fin, causing the clefted subject to move to its specifier, making it visible to the low *v*P Focus of the matrix clause, allowing movement to the specifier of (*v*P) Focus of the matrix clause. There are then two options for the modelling of the small-clause, either the subject is first-merged in the left-peripheral EPP position, or it is moved there from the lower *v*P, in either case, Spec-TP is filled by a null subject *pro*. It strikes me that it should be equally possible for the subject to move from Spec-*v*P, through Spec-TP to the EPP position before raising into the matrix clause, thus removing the need for a null subject. It should be noted that this version of Belletti's analysis only applies to subject clefts

since an object would not be able to move through the EPP position. To further illustrate, I provide a bracketing below.

(102) [TP c' [T' [T est] … [FocP Jean [*v*P ~~est~~[CP … [EPP ~~Jean~~ [FinP [Fin qui] … [est venu]]]]]]]

As well as this analysis specific to subject clefts, Belletti (2008) provides an analysis of complement clefts. In this case, the clefted element is raised, as in focus fronting, to the left-peripheral Spec-Focus of the small clause. Belletti (2008) uses this to account for alleged distributional differences between contrastive and new information foci. In Italian, it is claimed, subject clefts can be both contrastive foci and new information foci, while complement clefts may only be contrastive foci, a restriction also placed on focus fronting. Belletti (2012) expands this analysis. In order to account for cleft-like clauses such as 'Gianni e che…' and the presence of intervening adverbs in some clefts, it is suggested that the entire small-clause from Fin downwards is extraposed (with the landing site left unspecified) after the clefted element has moved to the relevant higher position. Next, the entire remnant phrase, which in surface terms now only includes the clefted constituent, raises to the specifier of the low *v*P Focus position of the matrix clause. This analysis is further extended in Belletti (2015). In this new analysis, the expletive is not in fact a true expletive, but a special pronominal DP specific to clefts, which features a silent THING head. This subject is first-merged in Spec-PredP of the relative clause before raising to the subject position of the matrix clause. Then, the extraposition steps described above occur, with the clefted element moving to the Spec-FocP position of the relative clause before the rest of the relative clause extraposes and any other optional movement operations occur. This analysis then has the advantages of drawing together the formerly disparate analyses of subject and object clefts while maintaining the distinct predictions regarding their pragmatic limitations. Additionally, this analysis accounts for the cross-linguistically observed non-expletive behaviours of the subject of the cleft (Belletti, 2015, 48-50). However, it is not clear to what extent *ce* differs from a standard expletive in French. I provide a series of bracketings below to illustrate this analysis, several parts of the derivation have been simplified to draw out the key parts of the analysis more clearly:

(103) a. [CP [TP c' [T' [T est] … [PredP [Pred ~~ce~~] … [FinP [Fin que] … [TP j'ai mangé … [DP le gâteau]]]]]]

b. [$_{CP}$ [$_{TP}$ c' [$_{T'}$ [$_{T}$ est] … [$_{vP}$ … [$_{FocP}$ [$_{DP}$ le gâteau] … [$_{FinP}$ [$_{Fin}$ que] … [$_{TP}$ j'ai mangé …
[$_{DP}$ ~~le gâteau~~]]]]]]]

c. [$_{CP}$ [$_{TP}$ c' [$_{T'}$ [$_{T}$ est] … [$_{vP}$ … [$_{FocP}$ [$_{DP}$ le gâteau] … ~~que j'ai mangé~~]]]] [$_{FinP}$ que j'ai mangé]]

d. [$_{CP}$ [$_{TP}$ c' [$_{T'}$ [$_{T}$ est] … [$_{FocP}$ [$_{DP}$ le gâteau] [$_{vP}$ [$_{DP}$ ~~le gâteau~~]]]]] [$_{FinP}$ que j'ai mangé]]

One issue with Belletti's (2012, 2015) extraposition analyses is that they either require us to accept rightward movement of the small-clause, as illustrated in their bracketing in Belletti (2015: 55), or they require further movement operations to acquire the desired surface order. Additionally, it is unclear what motivates this extraposition movement in the first place. Thus, Belletti's (2012, 2015) analyses are, in some ways, incomplete. One important element of Belletti's (2015) analysis is the basis in subject-object asymmetries, a key part of this chapter, then, is to test these asymmetries in the history of French.

I now turn to Reeve's (2011) analysis. This analysis is much like that of Belletti (2008), although lacking some of the fine-grained cartographic descriptions. In Reeve's (2011) analysis the relative clause is taken to be, syntactically, a restrictive relative clause. Evidence for this includes the shared set of relative operators between both clause-types and the fact that the cleft clause behaves as a strong island. However, the clefted constituent is also moved out of this relative clause, into an adjoined position, which then merges in the object position of the matrix clause (Reeve, 2011: 161). This analysis is thus strongly analogous to Belletti's (2008) analysis, positing a dedicated left-peripheral position in the relative clause for the clefted constituent to occupy. An issue with both analyses, then, is this stipulated position, which does not clearly fit our general understanding of the left periphery.

Finally, I describe Clech-Darbon, Rebuschi and Rialland's (1999) unorthodox analysis. Clech-Darbon et al. (1999) reject typical descriptions of clefts, arguing that both traditional monoclausal and biclausal analyses are deficient. Importantly, Clech-Darbon et al. (1999) seem to take the term cleft to specifically refer to 'narrow focus clefts' i.e. our contrastive focus clefts. In their analysis the relative clause is simply directly adjoined to the right of the base clause (which constitutes the expletive, copula and clefted constituent) analogous to right dislocation. Clech-Darbon et al. (1999: 100) show that the intonational patterns of clefts specifically match right dislocation patterns. Thus, they 'reject

most of the analyses reviewed […] except possibly the one that relies on extraposition, simply because it is the only one that guarantees that the coda will BE TO THE RIGHT OF AN IP BOUNDARY'. In many ways this supports the analysis of Frascarelli and Ramaglia (2009, 2013).

It is clear from this overview of the literature that there is very little consensus regarding the syntax of clefts, and thus, a rich array of predictions can be tested on our data. It should be noted that almost all of these analyses, especially those which make use of information-structurally marked projections, treat only a sub-set of clefts, typically ignoring informative-presuppositional clefts and non-DP clefts. These analyses often leave little room for the analysis of all-focus clefts or for differentiation between contrastive and informational focus clefts. Thus, in this chapter, I hope to provide an analysis which covers the wide variety of cleft constructions available, as well as connecting them to the rest of the syntax of French, rather than providing an isolated analysis of one sub-type.

6.2.3 History

Having discussed the wide range of syntactic analyses, I turn to discussions of the history of clefts, summarising here the work of Dufter (2008, 2009) and Rouquier (2014).

Dufter's (2008, 2009) analysis is centred on the rejection of the traditional assumption that clefts' frequency increases in response to the loss of focus fronting (supported by, e.g., Wehr, 2005). Dufter (2008) instead show that, although they are rare, clefts are found throughout the history of French, including in Old French. Indeed, while we would expect the rate of object and subject clefts to increase once focus fronting is lost, we instead find only an increase in the rate of adverbial and inferential clefts during this period (Dufter, 2008: 18). Dufter (2008) thus argues that the rise in rates of clefts is related to shifts in their function, independent of their syntax, with the original clefts only being used for focusing before informative-presuppositional clefts develop alongside adverbial clefts in the Early Modern period, accounting for the rise found in the 16th century onwards. Importantly, as well as simply a rejection of orthodox assumptions regarding clefts, Dufter's (2008, 2009) work also rejects the assumption of coevolution of form and function central to generative approaches to language change. Thus, confirming or denying his claims is likely to have widespread ramifications for

our theory of language change. Rouquier (2014) examines the history of clefts in Old and Middle French, without examining the Early Modern period so central to Dufter's (2008, 2009) analysis. Examining texts from the 10th to 15th centuries, Rouquier finds clefts to be first attested in the 12th century, although limited to proper names or previously mentioned nouns (Rouquier, 2014: 11), perhaps suggestive of contrastive focus. So-called *constructions liées* of the kind 'c'est une chose singulière que la rumeur' 'rumour is a singular thing' are also found in this early period, although Rouquier (2014) argues that they should not be analysed as clefts. Additionally, regarding function, Rouquier (2014: 39) supports Dufter's findings, suggesting that clefts originally had a narrow focusing function, before expanding to the use of informative-presuppositional clefts as their frequency increased.

Regarding syntactic development, Rouquier (2014) notes that the copula agrees with the clefted constituent in Old French, while Dufter (2008: 4) specifies that the change to an invariantly 3rd person copula only occurs in the 16th century, although Rouquier (2014: 78) notes that Maupas still accepts *ce sui je* in 1625.

Although Dufter (2008) and Rouquier (2014) seem to be in agreement on much of the history of clefts, their claims have recently been challenged by Trips and Stein (2018). Trips and Stein's (2018) study focuses on Old and Middle French, comparing Middle English, with an analysis of the MCVF and Penn Supplement corpora. They show that, although rare, informative-presuppositional and adverbial clefts are both attested throughout the MCVF and Penn Supplement corpora, as well as finding quite different frequencies of clefts from those which Dufter (2008) finds.

It seems, then, that there is some debate on the history of clefts. Dufter (2008) has argued for an analysis in which clefts begin as a marginal construction in Old French, where they are only used for focusing, and where the copula agrees with the clefted constituent. This shifts in the 16th century, when the copula begins to become more invariant and informative-presuppositional clefts become acceptable. Following this shift, the rate of clefts gradually increases. However, the Old and Middle French portion of this analysis has been thrown into doubt by Trips and Stein's (2018) corpus work. It will be interesting to confirm or deny these two analyses and to further test Dufter's (2008) analysis

on later data, particularly from Early Modern French and with the added data from lower register texts. Additionally, adding fine-grained syntactic analysis to the descriptive analysis provided by Dufter (2008, 2009) and Rouquier (2014) will allow us to confirm or reject Dufter's (2009) claim that form and function do not co-vary in this case, and shine some further light on the variety of cleft analyses available for synchronic data.

In order to evaluate these claims further, some notes are also required on the history of focus fronting. As we noted above, focus fronting is no longer available in Contemporary Spoken French (Rowlett, 2007: 182-183), however, it was extremely common in early Old French (Wolfe, 2021: 51-52). It should also be noted that two types of focus fronting were attested in the earliest texts, contrastive focus fronting and informational focus fronting. This latter type was rare as early as late Old French despite being common in the 10th-12th centuries (Wolfe, 2021: 52). By the 17th century focus fronting is limited to QPs and certain adverbials and predicative adjectives (Wolfe, 2021: 75). Thus, if we were to analyse the increase in clefts as resulting from the decrease in focus fronting, we would expect an increase in frequency in the late Old French period with informational foci, contrary to Dufter's (2008) and Rouquier's (2014) findings.

Finally, a note on text-types and the spoken-written divide is needed. The precise way which clefts are divided across spoken and written language use is hotly debated, with claims made that clefts are more common in speech, writing and neither (Dufter, 2008: 7). Dufter (2008: 12) finds that they are more common in speech than literary writing in Contemporary French. With our set of corpora covering a wide-variety of text-types, it will be of particular interest to analyse this variance across medium, and we may gain further insights into the development of clefts, as we did with dislocations in Chapter 5.

In summary, the literature on clefts is wide-ranging and varied. With regard to function, there seems to be a general consensus on four types of clefts: informational focus, contrastive focus, all focus and informative-presuppositional. There has also been the suggestion that the distinctive feature of all clefts is that the relative clause is communicated to be presupposed (Lambrecht, 2001: 482-5;

Rebuschi, Doetjes and Rialland, 2002). In informative-presuppositional clefts, this has the effect of suggesting that the relative clause is beyond questioning.

Regarding syntax, however, the literature has come to no consensus on the analysis of clefts, and a wide variety of analyses are available. Importantly, these analyses primarily account for focus clefts, utilising the cartographic positions of Focus Phrases. Thus, in addition to testing these analyses, a central element of this chapter is to determine how and whether they can be extended to informative-presuppositional clefts. Finally, with regard to the historical literature, it is rather sparse, and one central claim is typically made: that clefts expand in function and frequency over the history of French, initially only marking focus on DPs, before expanding to the informative-presuppositional function alongside allowing clefted adverbials. Testing this claim will form the core of this chapter.

6.3 Methodology

I now turn to my methodology. Text and corpus selection is covered in Chapter 3 of this thesis; here I provide further details on the methodology particular to the analysis of clefts.

For the MCVF and Penn corpora, Trips and Stein (2018) provide search terms for the CorpusSearch program. These have been used here. Importantly, these corpora have a specialised annotation for clefts, however, Trips and Stein (2018) find that some clefts are incorrectly annotated, and thus use a secondary search to find all surface-similar forms, before hand-sorting all of the results. I have followed this methodology here. I give examples below of cases of non-clefts annotated as clefts (104) and clefts annotated as non-clefts (105):

(104)

Car	il	est	cil	qui	noz	pechiez	nos	pardonet.
since	he	is	this	who	our	sins	us	pardoned

'Since he is the one who pardoned our sins'

(1190-SBERNAN-BFM-P,170.5680)

(105)

Ço	fud	Maheu le Puigneur	sur	qui	vint	la	lance;
It	was	Maheu le Puigneur	on	who	came	the	lance

'Maheu le Puigneur was the one upon whom the lance came'

(1175-FANTOSME-BFM-R,8.68)

In (104) we see that the relative clause appears to be a restrictive relative, while the main clause has a presentative structure. This clause lacks the usual expletive *ce* subject of clefts, although this may be attributable to diachronic variation. (104) does not seem to have any of the information-structural properties of clefts, since *cil* is neither focused nor emphasised in any way. This is a clear case of a presentative with a complex DP. In (105) on the other hand, *Maheu le Puigneur* is being focused by the cleft, the preceding discourse features a discussion of those involved in the battle in question and so *Maheu le Puigneur* is being contrastively focused as the one who is defeated, while the notion that somebody would have been defeated, given in the relative clause, is presupposed. In addition to these examples, *il y a* clefts are annotated as clefts in these corpora, but should not be counted in this work on *c'est* clefts. These examples show the importance of hand-checking corpus data in cases such as this.

While the MCVF and Penn Supplement data were subject to this complex searching procedure, the data from hand-analysed texts are analysed as with the other phenomena studied: the first 400 clauses are gathered, and any clefts are noted to ascertain frequency data.

As well as these initial frequency data, all clefts have been annotated for cleft-type: contrastive focus, informative focus, all focus or informative-presuppositional. Additionally, clefts have been annotated for type of the clefted element: DP, adverbial or PP. In cases of adverbial clefts which have become conventionalised or, as per Lahousse and Lamiroy's (2017) analysis, lexicalised, such as 'c'est ainsi que…' and 'c'est vrai que…' I have erred on the side of including these in the data, although they will be discussed appropriately given their unusual behaviour.

It should be noted that throughout our texts, the rates of clefts are quite low. As such, our token numbers are low and much of what follows will focus on the qualitative rather than the quantitative analysis.

6.4 Data Presentation

I now turn to data presentation. I begin by describing the distribution and surface features of clefts throughout their history, before turning to an analysis of their syntax.

6.4.1 Old and Middle French

I begin with the data from the MCVF and Penn Supplement. These data are extremely sparse, with half of the texts having no clefts at all. Of those texts which do feature clefts, over half have them in under 0.1% of clauses. As such, I focus primarily on qualitative discussion of the data here.

From the existing literature, two main claims need to be evaluated. Firstly, that clefts are first found in the 12th century, limited to DPs, as per Rouquier (2014). Secondly, that informative-presuppositional clefts are not found until Middle French, as per Dufter (2008) and Rouquier (2014). In our data, neither of these claims appears to hold.

Firstly, clefts are attested as early as the 10th century, with examples in the Vie de saint Léger, one of which is given below:

(106)

cio	fud	Lusos ut	il	intrat.
It	was	Lusos where	he	entered

'It was Lusos where he entered'

(0980-LEGER-PENN-V,XVII.117)

Although there is a preference for DPs in early texts, temporal adverbials are also found in clefts as early as the 12th century. An early example is given below.

(107)

Ço	fud	juesdi	au	seir	ke	li	reis	ad	parlé,
It	was	Thursday	in.the	evening	that	the	king	has	talked

'It was Thursday in the evening that the king spoke.'

(1175-FANTOSME-BFM-R,126.1377)

Both information focus, contrastive focus and all focus clefts are found in this period, just as in Modern French.

(108)

C'	est	Dieu	qui	tout	set	et	tout	voit,	non	mie
It	is	God	who	all	knows	and	all	sees,	NEG	NEG

les	pures	creatures,	qui	ne	sont	fors	vanitez.
the	pure	creatures	who	NEG	are	only	vanities

'It's God who knows all and sees all, not the pure creatures, who are only vanities'

(1279-SOMME-ROYAL-PENN-P,1,60.1644)

(109)

Ço	fud	Waltier	le	fiz	Robert	ki	puinst	premièrement
It	was	Walter	the	son	Robert	who	attacked	firstly

'It was Walter, the son of Robert, who attacked first'

(1175-FANTOSME-BFM-R,74.789)

(110)

Quant	il	ço	sourent	qued	il	fud	si	alét,	Ço	fut
When	they	this	knew	that	he	was	*si*	gone	it	was

granz	dols	quet	il	unt	demenét	E	granz	deplainz
great	sadness	which	they	had	felt	and	great	grieving

par	tuta	la	citiét.
through	all	the	city

'When they knew this, that he had gone, there was a great sadness which they felt and great grieving through the whole city.'

(10XX-ALEXIS-PENN-V,21.198)

In (108) the clefted constituent 'Dieu' is explicitly contrasted against another DP 'les pures creatures' thus a case of a contrastive focus cleft. In (109) the clause occurs at the beginning of a new section of the text, rather than contrasting *Waltier* against other options, the clefted element simply adds new information to the discourse, an information focus cleft. Finally, in (110) neither the clefted constituent nor the relative clause have been previously mentioned or stand in contrast to one another, the entire element introduced by the copula is focused. Notably, (110) also represents an example of a clefted object outside of a contrastive focus cleft. Thus, it seems that the subject-object asymmetries observed by Belletti (2012, 2015) for Italian do not apply in Old and Middle French.

In addition to these focusing functions of clefts, however, informative-presuppositional clefts are also found as early as the 12th century. This early occurrence of informative-presuppositional clefts directly contradicts Dufter's (2008) claims on the evolution of clefts. This has been observed in previous work utilising the MCVF and Penn Supplement by Trips and Stein (2018: 295). I exemplify

an early informative-presuppositional cleft below. Due to the length of the example, I only gloss the cleft clause.

(111) En terre sen fil envoia qui aveques nous conversa. Nez fu de la virge Marie Sans pechié et san vilenie, Sanz semence d'omne engenrez, sanz pechié conceuz et nez;

Ce	**fu**	**cil**	**meïsmes**	**Jhesus**	**qui**	**o**	**nous**	**conversa**
It	was	this	same	Jesus	who	with	us	stayed

ça	**jus**	**et**	**qui**	**les**	**miracles**	**feisoit**;
here	down	and	who	the	miracles	did

'He sent his son, who lived with us, to Earth. He was born of the virgin Mary without sin and without evil, conceived without the seed of man, conceived and born without sin. It was this same Jesus who stayed with us down here and who performed the miracles'
(1190-BORON-PENN-R,76.1199)

In this example, the clefted constituent 'cil meïsmes Jhesus' has already been mentioned in the on-going discourse and, indeed, is the topic of the ongoing discourse, while the relative clause is providing new information. Interestingly, informative-presuppositional clefts seem to be restricted to a handful of texts, predominantly religious texts, where they are used to punctuate or re-emphasise the topic in an ongoing section with a continuing topic. Often, this section is describing the life and actions of Christ, with clefting used to re-emphasise his importance. Another example is given below to illustrate. Due to the length of this example, I only provide glosses of the cleft clauses.

(112) Lors fist li Pere çou que il avoit dit, que il envoieroit en terre son fil Jhesucrist, qui nasqui en a Bethleem de le virgne Marie.

Ce	**fu**	**cil**	qui	ala	par	terre	trente et trois	ans.
It	was	this	who	went	by	earth	thirty-three	years

Ce	**fu**	**cil**	qui	faisoit	les	beles	miracles	et
It	was	this	who	did	the	beautiful	miracles	and

les	buenes	ouevres	que	onques	nus	hom	ne	fist	se	il
the	good	works	which	ever	no	man	NEG	did	but	him

non	seulement.
NEG	alone

'Then the Father did that which he had said, that he would send to Earth his son Jesus Christ, who would be born of the Virgin Mary in Bethlehem. It was this man who went across the Earth for thirty-three years. It was this man who did the great miracles and good works which no man has ever done but him alone.'

(1210-BORON-PENN-P,50.924)

We see here that the cleft is used as a stylistic device to re-assert the central importance of Christ, as well as to assert the unquestionable, presupposed nature of his acts and experiences. It seems clear then, that from the earliest attestations of clefts, all of their functional sub-types are attested. As such, it does not seem that the later increase of frequency in clefts can be attributed to the development of new functions. These new data seem to refute Dufter's (2008) proposition that it is the development of informative-presuppositional and adverbial clefts which leads to the increase in the frequency of clefts from the 16th century onwards.

6.4.2 Early Modern French

I next turn to my Early Modern French data. In these data, the frequency increases considerably, allowing for some graphing and quantitative analysis. I present the frequency data on clefts, including data from the MCVF and Penn Supplement for comparison, in the graph below:

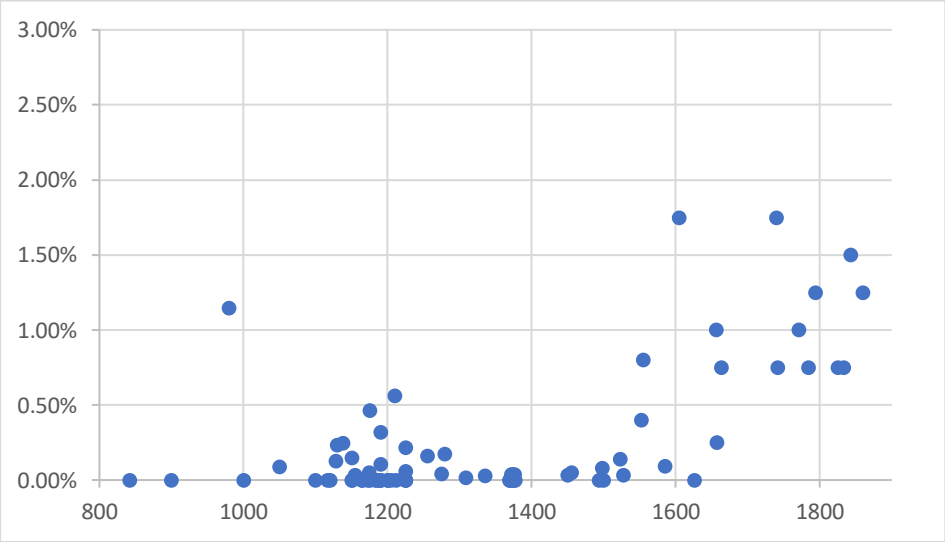

Figure 13 : Frequency of Clefts from the MCVF and Penn Supplement and Early Modern French Texts

We see that a sharp increase occurs just after the 16th century, with several texts showing clefts in over 1% of clauses, while the vast majority of Old and Middle French texts only had clefts in under 0.5% of clauses, often 0. It has been suggested that the rise in clefts in this period is linked to the loss of focus fronting (Wolfe, 2021: 92). Focus fronting drops in frequency considerably over the course of the 16th century, perhaps leading to this increase. These data seem to suggest that the increase in clefts may indeed be related to the loss of focus fronting. It may also be relevant that textual data is our only reference point for this period of French history. It has been suggested that, in Modern and Contemporary French, clefts are typical of writing, while in speech, focusing can be achieved through prosodic accentuation. It should be noted, however, that contrary claims have been made regarding the frequency of clefts in Modern French, with Dufter (2008: 12) suggesting that clefts are, in fact, more common in spoken than written French.

Although a link between the rise of clefting and fall of focus fronting seems intuitive, Dufter (2008) has argued against it, claiming instead that the increase in the frequency of clefts is in fact an increase in adverbial clefts, while DP clefts do not show a sharp increase. However, if we show only the frequency of DP clefts on a graph, an increase remains clear, although less marked.

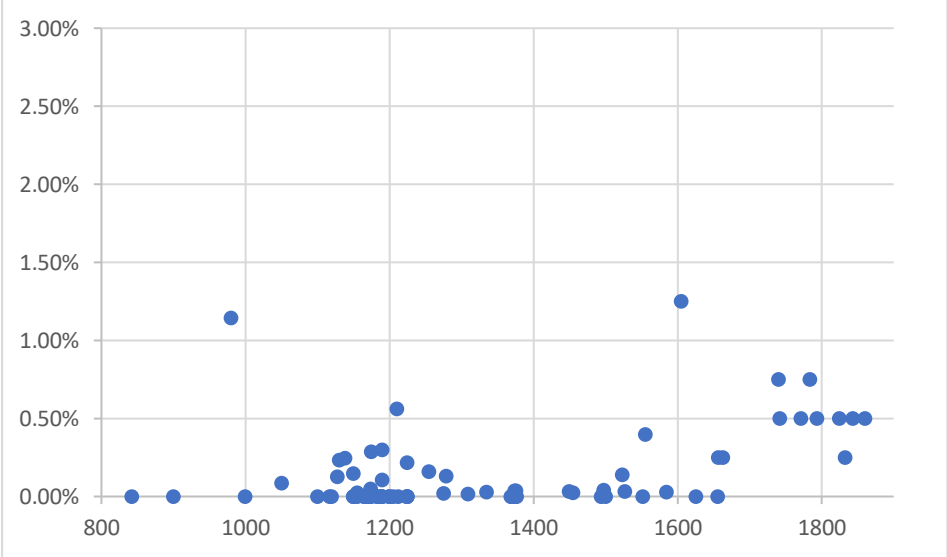

Figure 14 : Frequency of DP Clefts in the MCVF, Penn Supplement and Early Modern French Texts

This graph does show a striking contrast to the previous data. However, once we consider the high number of Old and Middle French texts in which no clefts are attested across thousands of clauses and that all of our Early Modern French texts have only had the first 400 clauses analysed, there remains a clear increase. Crucially, Dufter's (2008) argument rests on comparing the relative rates of adverbial and DP clefts rather than the absolute rates of said clefts in the corpora examined. While it is true that adverbial clefts and informative-presuppositional clefts increase in frequency after the 16th century, we cannot ignore the general increase in the frequency of all clefts, including DP clefts, which may well be a reaction to the loss of focus fronting.

It should be noted that our data is considerably sparser than Dufter's (2008) data, however, we also use a wider variety of texts, while Dufter's (2008) work is based entirely on the Frantext corpus of literary texts. Dufter's (2008) claims seem to overstate the role of adverbial and informative-presuppositional clefts in the history of French clefts. Although the rate of these clefts does raise faster than that of DP and focus clefts, DP and focus clefts do still increase considerably in frequency in the 16th and 17th centuries. Turning to the function of clefts, all functions are once again found, illustrated below. Example (116) is only partially glossed due to its length.

(113)

Nous	sommes	partis	pour	faire	face	à	l'	ennemi.	**C'**	**est**
We	are	left	to	make	face	to	the	enemy	it	is

notre	**battalion**	**qui**	**a**	**commencé**	**le**	**feu**
our	battalion	which	has	started	the	fire

'We left to face the enemy. It's our battalion which started firing'

Vendée Letters (Collet, 2021: 18)

(114)

Mr. Héroard	ce	sont	ces	femmes	qui	sont	cause	que
Mr. Héroard	it	are	these	women	who	are	reason	that

je	me	suis	levé	si	tard.
I	REFL	am	got up	so	late

'Mr. Héroard, it's these women who are the reason that I got up so late.'

Héroard 1609.6 (spelling normalised from Héroard's pseudo-phonetic spelling)

(115)

Ailleurs,	c'	étoit	des	amas	de	neiges,	aussy	anciens	que
Elsewhere	it	was	some	piles	of	snow	as	ancient	as

le	monde,	qui	entretenoient	la	rigueur	des	hyvers au
the	world	which	maintained	the	intensity	of.the	winters in.the

milieu	des	ardeurs	de	la	canicule.
middle	of.the	heat	of	the	heat wave

'Elsewhere, it was piles of snow, as ancient as the world, which maintained the intensity of the winters in the middle of the heat of the heat wave.'

Jamerey-Duval p. 126

(116) En général, il paroît que, dans cette partie, les courans varient, & portent quelquefois au Nord-Est, plus souvent au Sud-Ouest. Un coup d'œil sur le gissement de la côte suffit pour prouver qu'ils ne doivent suivre que l'une ou l'autre de ces deux directions, & il est toujours facile de distinguer laquelle regne, par les différences Nord ou Sud que donnent les observations de latitude.

C'	**est**	**à**	**ces**	**courans**	**qu'**	**il**	**faut**	**imputer**
It	is	to	these	currents	that	one	must	attribute

les	**erreurs**	**fréquentes**	**dont**	**les**	**navigateurs**	**se**
the	errors	frequent	of.which	the	navigators	REFL

plaignent.
complain.

'In general, it seems that, in this area, the currents vary, and move sometimes to the North-East, more often to the South-West. A glance at the position of the coast suffices to prove that one has to follow only one of these two directions and it's always easy to distinguish which one predominates by the differences to the North or South which observations of latitude provide. It's these currents to which we must attribute the frequent errors of which navigators complain.'

Bougainville p. 27

In (113) the clefted constituent 'notre battailon' is being contrasted against the enemy force, contrastively focusing it. In (114) the Dauphin is providing an explanation, the relative clause here is clearly presupposed (Héroard would know that the Dauphin had got up late) and the cleft is being used to put an informative focus on the fact that it was these women who were the cause. In (115)

none of the information in the clause has been mentioned yet, a clear example of a thetic clause and thus an all-focus cleft. Finally, in (116) the currents were already the topic of the ongoing discourse, while the relative clause is adding information, a case of an informative-presuppositional cleft.

In addition to these examples, we find some instances of clefts combined with other information-structurally marked constructions. For instance, we find examples of a clefted element left-dislocated from within the cleft, showing that it must be topical.

(117) Moyse auquel vous avez esperance: c' est celuy qui vous accuse.
Moses in.whom you have faith it is this who you accuses

'Moses, in whom you have faith, it is this man who accuses you'

(1523-NEW-TESTAMENT-PENN-P,y4V.1470)

In summary, the Early Modern period shows a considerable increase in the frequency of clefts, potentially related to shifts in other focus-marking constructions such as the drop in frequency of focus fronting and free inversion. Our data seems to suggest that, while adverbial clefts increase in frequency the most, all cleft-types increase in frequency in this period, contra Dufter's (2008) claim that the increased rate of clefts can be attributed to adverbial and informative-presuppositional clefts.

6.4.2.1 Sociohistorical Elements

Turning to text-type, we do not see any clear pattern. The relevant data are displayed below:

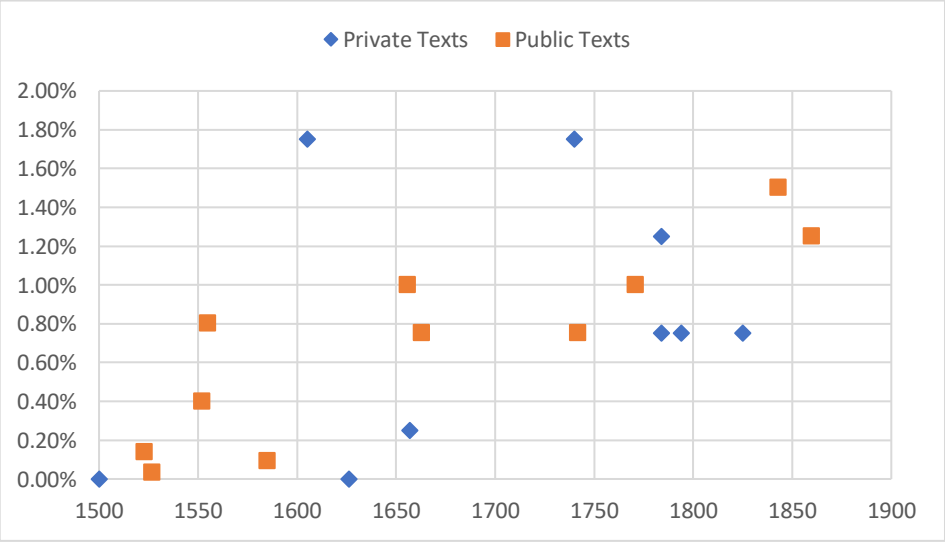

Figure 15 : Frequency of Clefts in Published and Private Early Modern French Texts

Although there are two anomalously high rates, one from Héroard's record of the Dauphin's speech, the other from Jamerey-Duval's Memoires, both less formal texts, the overall pattern seems to suggest that the more formal texts might have marginally higher rates. With the 19th century *journaux intimes* showing lower rates than any of the contemporaneous literary texts, and the 17th century *journaux intimes* showing some of the lowest rates of any texts from the period. It should again be noted, however, that our data are extremely sparse, so drawing clear conclusions is difficult. Notably, however, we do not see a clear divide between the spoken-like texts (letters and the Dauphin's speech) and the other texts, these data seem to be even less clear than our dislocation data in Chapter 5. This lack of clarity suggests to me a lack of relationship; however, larger quantities of data are certainly required to make any strong claims.

6.4.2.2 Metalinguistic Data

As in the case of dislocations, there is relatively little discussion of clefts in the metalinguistic texts examined. None of the three texts directly discuss clefts. However, Maupas does provide numerous examples of clefts as *bon usage* during a discussion of presentatives. One such example is provided below:

> 'C'estoit donc luy que vous prisiez tant' 'It was thus him who you prized so much' (Maupas, 1618: 142)

Indeed Maupas (1618: 142-143) goes on to discuss how, without the use of the expletive subject, the copula must agree with the clefted constituent and the subordinate clause must be incorporated into the main clause 'Car si la particule *ce*, n'y est mise, ledit verbe doit estre personnel avec les pronoms convenables & les articles *Le, La, Les,* seront en suite' 'Since if the particle *ce* is not placed there, the said verb must agree with the relevant pronoun and the articles *Le, La, Les,* should follow' (Maupas, 1618: 142), he provides the example '*Suis-je le respondant'* 'I am the respondent' (Maupas, 1618: 143). We can see, then, that grammarians both view clefts as licit constructions, and understand that they can be paraphrased by a non-complex clause. Again, however, as with dislocations, this relatively limited discussion shows that clefts do not have a strong sociostylistic association, supporting the

findings from our quantitative data. Rather, clefts constitute a relatively rare but fully acceptable construction, used across a range of sociostylistic registers.

6.4.3 Interim Summary

Taking stock of the history provided so far, from Old and Middle French through to 19th century, it seems that many of Dufter's (2008) claims regarding the history of clefts do not hold. In Old and Middle French, informative-presuppositional clefts and clefted adverbials can be found, contra Dufter's (2008: 20) claims. Indeed, Old and Middle French show remarkably well-developed clefts, with both object and subject DPs clefted, with no apparent asymmetry. Likewise, in Early Modern French, the increase in the frequency of clefts is not just an increase in the frequency of adverbial clefts, but an increase in the frequency of all clefts. All these developments point to the more traditional view that the frequency of clefts does in fact respond to other shifts in the information-structurally marked syntax of French, namely the loss of focus fronting (Wolfe, 2021: 92) and drops in the frequency of post-verbal subjects, seen in Chapter 4. Nonetheless, it should be noted that clefting and focus fronting are not in complementary distribution, since the two co-exist in Old and Middle French.

6.4.4 Modern and Contemporary French

Having given an overview of the historical situation, I turn to our Modern and Contemporary French data. These are shown below in Figure 16:

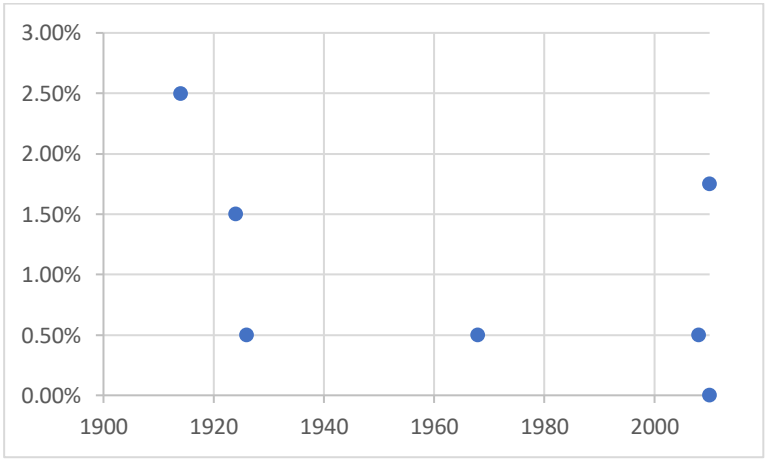

Figure 16 : Frequency of Clefts in Modern and Contemporary French Data

These data are rather difficult to interpret. Although we see the highest rates of clefts yet seen, there is not a clear increase, on average, from the historical data. Particularly of note is that there is no clear pattern distinguishing the spoken and written data, marking a clear difference from our analysis of dislocations. Indeed, our two spoken language corpora show wildly divergent frequencies of dislocations, with the CFPP2000 corpus showing 1.75% and the ESLO corpus showing 0.5%. It should also be noted that both SMS and forum data show some of the lowest rates, although letters show some of the highest, both here and in the Vendée Letters. These various inconsistencies suggest to me that register, genre and the spoken-written divide have a minimal impact on clefts in our data. Again, I caution against strong conclusions of this kind given the sparse nature of our data, but we can compare to our dislocation data, where an extremely clear pattern presents itself, and see that there is clearly a different situation at play here.

In terms of function and distribution of clefts, it is interesting to note that two different patterns appear to be at play. While the high rate in the soldiers' letters is due predominantly to a large number of clefted DPs (shown in (118)), the rate in our speech corpus is due to a large number of clefted adverbials and conventionalised clefts (shown in 118).

(118)

C'	est	la	23e	compagnie	qui	partira	demain
It	is	the	23rd	company	who	will leave	tomorrow

mercredi	21	octobre	pour	fort	de	Rembètant
Wednesday	21st	October	for	fort	of	Rembètant

'It's the 23rd company who'll leave tomorrow, Wednesday the 21st of October, for the fort of Rembètant'

WWI Letters, (Steuckardt, 2015: 382)

(119)

C'est	vrai	que	ça	compte
it is	true	that	it	counts

'It's true that it counts.'

CFPP2000 CHRISTOPHE_ANDRE_H_62_MARIE_ANNE_ANDRE_F_63_5E [1666.365]

Often these clefted adverbials appear to be conventionalised or perhaps even lexicalised, as has been argued for 'c'est ainsi que…' (Lahousse and Lamiroy, 2007). Examples of 'c'est vrai que…' and 'c'est

pour ça que…' constitute the majority of the examples in the CFPP2000 corpus, just as in (119) above.

If we follow a view in which these conventionalised examples are not to be considered true clefts, we might argue that spoken data seems to show an extremely low rate of clefts compared to writing. On the other hand, if we take these sorts of adverbial constructions to be clefts, we might argue that this represents a distinction between the spoken and written uses of the construction. We may also argue, as we have done for dislocations, that these sorts of clefts are playing a particularly spoken role, with some clefts such as 'c'est vrai que…' acting simply as gap fillers to signal a continuing conversational turn. An example of this role is given below:

(120) oui oui oui [pause] oui oui oui euh [pause] et donc c' est vrai que
yes yes yes [pause] yes yes yes uh [pause] and so it is true that

euh [pause] Marie-Anne a toujours été Parisienne
uh [pause] Marie-Anne has always been Parisian

'Yes yes yes [pause] yes yes yes uh [pause] and so it's true that uh [pause] Marie-Anne has always been a Parisian'

CFPP2000 CHRISTOPHE_ANDRE_H_62_MARIE_ANNE_ANDRE_F_63_5E [284.261]

In this example we clearly see that the speaker is trying to maintain their turn, initially filling the space repeating 'oui' before using 'c'est vrai que' before a pause to signal that they intend to continue speaking. Indeed, Lahousse and Lamiroy (2017) have shown that 'c'est ainsi que…' functions as a single unit as a connective. Further work on spoken French data would be needed to determine the possibility of such interactional functions, but I wish to be clear in showing that the information-structural account does not fully explain the functions of many of these conventionalised clefts. Clefts as they are used in Contemporary Spoken French seem to be playing some further roles.

6.5 Syntax

Having provided a history of clefts' functions and distribution, I next turn to their syntax. I begin with a descriptive history, before turning to the possible analyses which this history suggests.

6.5.1 Description

As described in section 6.2, the copula agrees with the clefted constituent, not the expletive, in Old and Middle French texts, exemplified below.

(121) Sire, Sire, çó sui jó que ai pecchied
Lord, Lord it am I which has sinned

'Lord, Lord, it is I who has sinned'

(1150-QUATRELIVRE-PENN-P,108.4044)

Interestingly, this behaviour is shared by some surface-similar examples of answers to *wh*-questions including those which feature an appositive relative. We can see this in the example below, from the 12th century:

(122) Ceo estes vus, chastes, net e religius, ki me servez
It are you, pure clean and religious, who me serves

mult bonement.
very well

'You are the one, pure, clean and religious, who serves me very well'

(1183-ADGAR-BFM-R,141.1297)

Regarding the types of DPs which may be clefted, Trips and Stein (2018) argue that objects are not found in clefts until the *Cent nouvelle nouvelles*, a 15th-century text. However, I have found examples as early as the 12th century.

(123) Ço fud Maheu le Puigneur sur qui vint la lance;
It was *Maheu le Puigneur* on who came the lance

'It was Maheu le Puigneur upon whom the lance came.'

(1175-FANTOSME-BFM-R,8.68)

We have also already seen that both informative-presuppositional clefts and adverbial clefts are attested during this period. Additionally, as early as the 12th century, we find instances of adverbials and appositive relative clauses intervening between the clefted element and the relative clause in both DP clefts and adverbial-clefts.

(124) Ço fud en mai enprés avril, quant l' erbe s' est
It was in May nearly April when the grass REFL is

verdie, ke David vint d' Escoce od fiere
growing green that David came from Scotland with proud

cumpaignie
company

'It was in May, near to April, when the grass was growing green, that David came from Scotland with proud company'

(1175-FANTOSME-BFM-R,82.868)

(125) Ço fud Waltier le fiz Robert, dunt vus oiez parlance,
It was Walter the son Robert of whom you heard tell

ki primes justad as Flamens e mist en grief
who first charged at.the Flemish and made CL grave

errance.
error.

'It was Walter, the son of Robert, whom you've heard talk of, who first charged at the Flemish and made a grave error'

(1175-FANTOSME-BFM-R,74.789)

Interestingly, we only find examples of frame-setting clauses and temporal adverbials in this intervening position, while the TP-internal adverbials of Cinque's (1999) hierarchy are not found here. Looking beyond the Old and Middle French periods, agreement between the copula and the clefted element disappears in the 16th century, although the first example in our Early Modern French data of a clefted 1st or 2nd person DP is from the 18th century.

(126) C' est moi qui le dis
It is me who it said

'It's me who said it'

Le Sopha p. 22

Importantly, this shift also involves a shift from the use of subject pronominals as the clefted element to object pronominals. This suggests that the change which has occurred is not just due to the

regularisation of the cleft formula but actually represents a change in the relationship between the copula, expletive and clefted constituent.

At this stage, clefts begin to behave essentially as in Modern and Contemporary French, with no other changes besides fluctuations in frequency.

6.5.2 Analysis

Having provided a descriptive overview of the history of clefts in our data, I turn to analysis. I begin with a discussion of the situation in Old and Middle French, before discussing how clefts evolve through their history. Due to the sparsity of our data as well as the simple difficulty of providing cartographic analyses of clefts, I do not provide an exhaustive analysis here. Instead, I show where faults can be found in existing analyses and make note of oft-ignored distributional facts which are crucial to an eventual successful analysis of clefts. I argue that clefts and other routinised constructions like them are extremely difficult to analyse under a cartographic approach for a number of reasons. Firstly, their internal structure is extremely opaque, with the surface structure potentially reflecting a wide variety of derivations. Secondly, while much of the cartographic literature on clefts treats them as a purely focusing construction, the descriptive literature notes a wider information-structural role, including some non-focusing functions. This lack of clear information-structural role, combined with a highly structurally ambiguous surface form makes the application of cartographic syntax, with its focus on fine-grained description and a close mapping of syntax and function, particularly difficult.

In Old and Middle French, the copula and clefted constituent agree. This seems to suggest an analysis in which the relative clause is extraposed from the expletive pronoun, which is actually the complement of the copula, while the clefted constituent is a post-verbal subject. This would also lead us to predict that the syntax of clefts would change at the time that the frequency of post-verbal subjects drops in the Early Modern period, since a new structure would be necessary to derive the surface form of the cleft. This shift is precisely what we observe.

Further evidence for this extraposition and inversion approach can be seen if we examine contemporaneous examples of extraposed relative clauses. The surface manifestation of these clauses is extremely similar to that of early DP clefts. Compare the examples below of a true cleft (127) and an extraposed relative (128):

(127)	Ço	fud	Waltier	le	fiz	Robert	ki	puinst	premièrement
	It	was	Walter	the	son	Robert	who	attacked	first

'It was Walter, the son of Robert, who attacked first'

(1175-FANTOSME-BFM-R,76.809)

(128)	Cil	est	mult	proz	ki	sunet	l'	olifant:
	This	is	very	brave	who	sounded	the	horn

'The one who sounded the horn is very brave'

(1100-ROLAND-V,230.3157)

Examples such as (128), in which the extraposed relative modifies the pronominal subject *cil*, seem strikingly similar to cleft constructions, where we can analyse the relative clause as modifying the expletive in the matrix clause. A similar example was seen above in (122). This account is also supported by our metalinguistic data, with Maupas' discussion of clefts, incorporated into a discussion of presentatives, suggesting that the two constructions are viewed as analogous by the grammarians. Thus, I suggest that it is these extraposed relative clause constructions which lead to the development of DP clefts.

Thus, we can argue that the original structure of DP clefts involves a post-verbal subject, with the expletive acting as the fronted object of the copula from which the relative clause has been extraposed. I exemplify this analysis of early DP clefts with a potential bracketing of (121) below in (129). (129a) shows the original clause, while (129b) demonstrates the derivation of an extraposition analysis. I use the unexploded CP and do not write out deleted copies that have already been illustrated for simplicity.

(129) a. çó sui jó que ai pecchied

b. [TP jó [T' [T sui] [VP [V' [V ~~sui~~] [DP çó [CP que ai pecchied]]]]]]

[CP [TP jó [T' [T sui] [VP … [DP çó [CP ~~que ai pecchied~~]]]]] [CP que ai pecchied]] (extraposition)

[CP [C' [C sui] [TP jó [T' [T ~~sui~~] [VP … [DP çó]]]] [CP que ai pecchied]]] (verb-raising)

[CP çó [C' [C sui] [TP jó … [VP … [DP ~~çó~~]]]] [CP que ai pecchied]] (object-fronting)

While this extraposition and inversion analysis holds for Old and Middle French, the loss of post-verbal subject positions in the Early Modern period and the shifts in agreement with the copula suggest that the analysis of DP clefts shifts in the Early Modern period. In Old and Middle French, the relative clause is analysed as extraposed from the expletive, in Early Modern French it is reanalysed as either extraposed from the clefted constituent or, potentially, as a restrictive relative modifying the clefted constituent. Meanwhile, the expletive is reanalysed as subject of the copula and maintains some semantic content, allowing for continued verbal number agreement. An example with bracketing is provided below to illustrate:

(130) a. C' est moi qui le dis

b. [TP C' [T' [T est] [VP [V' [V ~~est~~] [DP moi [CP qui le dis]]]]]]

In the bracketing provided in (130), we follow Reeve's (2011) analysis, treating the relative clause as a true restrictive relative. However, as we shall see, an extraposition analysis in which the clefted constituent raises out of the subordinate clause may more aptly capture the descriptive facts. Nonetheless, the crucial change in this period is that the subordinate clause is no longer syntactically connected to the expletive, only to the clefted constituent.

I now turn to adverbial clefts, since they present several complications for our analysis of DP clefts. Adverbial clefts in Old and Middle French do not follow from our extraposition and inversion analysis. If we took adverbial clefts to be derived from an extraposed relative with an added modifying adverbial, we would expect to find examples in which the modifying adverbial is omitted, what Dufter (2008) refers to as inferential clefts, structures of the type 'c'est que…'. However, these structures are not attested until Early Modern French. An adverbial is in fact required to make such a structure grammatical. Additionally, the simple restrictive relative analysis espoused by Reeve (2011) cannot

apply here, since adverbials cannot be modified by restrictive relatives. Nonetheless, adverbial clefts are attested, as shown below:

(131)	Ce	n'	ert	hui	que	saisiz	soie	de	s'	amor
	It	NEG	was	today	that	seized	was	by	his	love

'It was not today that I would be seized by his love'

(1155-ENEAS2-BFM-R,125.2844)

Thus, for these adverbial clefts, a distinct analysis is required. Initially, an analysis of clefts akin to that of Clech-Darbon, Rebuschi and Rialland (1999) and Rebuschi, Rialland and Doetjes (2002) seemed best-suited to these data. Under this analysis, we analyse the clefted clause as a simple presentative with a right dislocated clausal constituent (the relative clause). This allows us to account for the unattested nature of inferential clefts, since, as we saw in Chapter 5, right dislocation requires the clause to be grammatical without the right dislocated constituent. Additionally, under this analysis, there is no movement to the relevant cartographic focus positions, but rather the pragmatic focusing phenomena associated with clefting are epiphenomenal. This allows us to unify the syntax of DP and adverbial clefts, focus and non-focus clefts as well as capturing the intuition that clefts are linked to pseudo-clefts. I illustrate this with a bracketing of the derivation of (126) below in (132).

(132) a. [$_{FamP}$ [qui le dit] [$_{Fam'}$ … [$_{TP}$ c' [$_{T'}$ [$_{T}$ est] … [$_{VP}$ [$_{V'}$ [$_{V}$ ~~est~~] [$_{DP}$ moi]]]]]]]

b. [$_{GroundP}$ [$_{TP}$ c'est moi] … [$_{FamP}$ [qui le dit] [$_{Fam'}$ … [$_{TP}$ ~~c'est moi~~]]]]

This analysis directly follows that of Frascarelli and Ramaglia (2009, 2013), illustrated earlier in (101), but skips the focus-movement step in order to unify the analysis of focus and informative-presuppositional clefts. Essentially, on this analysis the subordinate clause is straight-forwardly a right-dislocated clause and the unification of focus and informative-presuppositional clefts is straightforward. The remnant TP raises to SpecGround and the cleft constituent may be of any information-structural status, with focused examples taken as cases of in-situ focusing. This analysis also allows us to derive cleft constructions from the equivalent pseudo-cleft.

However, it should be noted that pseudo-clefts do not show the same distribution as true clefts; as Clech-Darbon, Rebuschi and Rialland (1999: 105) point out 'the very idea of deriving clefts

from pseudo-clefts in French … also meets more empirical difficulties.' Most notably, adverbial clefts and clefted PPs cannot be generated under this analysis (Clech-Darbon, Rebuschi and Rialland, 1999: 105). The obligatory reduplication found in pseudo-clefts is also left unaccounted for in any approach which unifies clefts and pseudo-clefts. Additionally, a right-dislocation analysis fails to account for several crucial facts regarding clefts, most importantly, the obligatoriness of the relative clause in most cases. As we saw in Chapter 5, dislocations in French are always additions to a syntactically complete clause, thus, they can always be omitted, excepting semantico-pragmatic constraints. De Cat (2002) argues that it is only in contrastive focus clefts that the relative clause is right dislocated and may be omitted. In all other types, the relative clause is a necessary part of the expression. We can see this with an example of a typical informational focus cleft below:

(133)	a. C'	est	à	toi	que	j'	ai	donné	le	gâteau
	It	is	to	you	that	I	have	given	the	cake

'It's to you that I gave the cake.'

b. *C'	est	à	toi[31]
It	is	to	you

(intended) 'It's to you.'

In this example, we can see that the subordinate clause is necessary to achieve the correct interpretation. While analyses such as that of Belletti (2008, 2012, 2015) and Bonan (2017) manage to account for many of these issues, they also rely entirely on the focusing function of clefts to motivate movement, despite the existence of informative-presuppositional clefts which seem to display the same syntactic behaviours as focusing clefts. Indeed, Bonan and Ledgeway (2023: 5) appear to be unaware of the existence of informative-presuppositional clefts stating that 'clefts come in many shapes and sizes but, according to the existing literature, all of them constitute a form of focalisation'. Likewise, Belletti and Bocci (2023: 81) list only new information clefts and contrastive/corrective clefts as information-structural functions of the structure. This seems to be a major oversight in much of the cartographic work on clefts. If we observe our data, we can see that informative-presuppositional clefts behave, syntactically, like non-contrastive focus clefts, suggesting an analysis

[31] Example (133b) could also be interpreted as a conventionalised phrase meaning 'it's your turn' but this is not the meaning under discussion here.

which avoids the use of focus phrases. This can be seen if we consider the example below, reprinted from (98d) above:

(134) Le Cécom est le distributeur de cette vidéo, c' est
Le Cécom is the distributor of this video it is

à lui *(qu' il faut s' addresser pour le
to them that one must REFL address to it

procurer)
obtain

'*Le Cécom* is the distributor of this video, it's them who you must contact to obtain a copy.'

Just as with non-contrastive focus clefts, the subordinate clause of the cleft cannot be omitted here, suggesting that both informative-presuppositional and information focus clefts share an underlying structure despite information-structural differences. It should be noted that, in many of these examples, teasing apart ungrammaticality from semantico-pragmatic incoherence is rather difficult. Nonetheless, in cases like (133) and (134), it does seem that the truncated cleft is truly ungrammatical, since it does not seem possible to ameliorate the grammaticality by changing the semantico-pragmatic context. In this case, I suggest that this ungrammaticality is likely due to the lack of a licenser for the prepositional phrase. The prepositional phrase cannot be licensed as a simple complement of the copula, suggesting it must have moved out of the subordinate clause, where it was originally licensed, accounting for the obligatoriness of the subordinate clause in these examples. Clearly then, some sort of syntactic connection exists between the clefted constituent and the subordinate clause, in this case, it is this connection which allows the licensing of the prepositional phrase. Additionally, this connection is not dependent on specific information-structural functions.

A crucial issue here is that the structure of DP clefts is highly ambiguous on the surface. We have seen in the discussion above that a wide variety of analyses can be applied to the very same example. In recent comparative work on clefts (e.g. Tang, 2023) it has been suggested that clefts are not cross-linguistically a unified structure. It may be the case that this would also be true diachronically and across different cleft sub-types, with a number of analyses active for clefts

simultaneously, with their opacity as a structure limiting the ability to form one consistent analysis in the speaker community.

Given the structural parallels between informative-presuppositional and non-contrastive focus clefts, however, a singular analysis is desirable. Variation between adverbial and DP clefts as well as varying information-structural functions make such an analysis difficult. However, some combination of Reeve's (2011) simple restrictive relative analysis and Belletti's (2008, 2012, 2015) extraposition analysis seems closest to capturing the data, with a strong syntactic link between the clefted constituent and the relative clause. In order to illustrate, I provide a schematic bracketing below of a possible analysis. In (135a-c) I only show the relative clause, since it is yet to be integrated into the main clause. The lower portion of the relative clause is not explicitly mapped out since it is not directly relevant to the analysis.

(135) a. [$_{FinP}$ qui [$_{TP}$ Jean [[$_{T}$ mange] le gâteau]]]

b. [$_{FP}$ Jean [$_{FinP}$ qui [$_{TP}$ ~~Jean~~ [[$_{T}$ mange] le gâteau]]]]

c. [$_{CP}$ [$_{TP}$ c' [$_{T'}$ [$_{T}$ est] … [$_{v'P}$ [$_{v'}$ ~~est~~] [[$_{FP}$ Jean [$_{FinP}$ qui [$_{TP}$ ~~Jean~~ [[$_{T}$ mange] le gâteau]]]]]]]]]

d. [$_{CP}$ [$_{TP}$ c' [$_{T'}$ [$_{T}$ est] … [$_{v'P}$ [$_{v'}$ ~~est~~] [[$_{FP}$ Jean ~~[$_{FinP}$ qui [$_{TP}$ Jean [[$_{T}$ mange] le gâteau]]]]~~]]]]]] [$_{FinP}$ qui [$_{TP}$ ~~Jean~~ [[$_{T}$ mange] le gâteau]]]]

In (135a) we see the generation of the small relative clause. In (135b) the clefted constituent (here the DP 'Jean') moves to an unspecified left-peripheral position so as to be visible to the higher clause. In (135c) the entire lower clause structure is merged as the complement of the copula. Finally, in (135d) the relative clause extraposes. This analysis follows Belletti's in many areas, maintaining the extraposition step and the movement of the clefted constituent out of the lower clause. However, more in-keeping with Reeve's (2011) analysis, the analysis presented here suggests that the clefted constituent merges as the complement of the copula, not into an information-structurally marked low position. Additionally, movement out of the clause must be through some information-structurally under-specified FP in the left-periphery in order to account for informative-presuppositional clefts structural similarities to focus clefts. Although it leaves some questions unanswered, this analysis can

account for the strong relationship between the clefted constituent and the relative clause (through the movement of the constituent out of said clause), and the possibility of intervening material between the clefted constituent and the relative clause (through the extraposition operation). A further strength of this analysis is that it can be readily incorporated into a diachronic account in which extraposed relative constructions are a source for the original development of clefts, a possibility discussed above with examples (127-128).

In summary, I give a description of the diachronic analysis argued for here. In Old French, the relative clause in cleft constructions is extraposed from the expletive, which is in fact the complement of the copula, while the clefted constituent is a post-verbal subject. This construction parallels similar extraposed relative constructions, as shown above in examples (127-128). As post-verbal subjects are lost, the clefted constituent is reinterpreted as the complement of the copula and the relative clause is reinterpreted as linked to the clefted constituent, as per Reeve's (2011) analysis. It is difficult to account for adverbial and PP clefts in this analysis, but a derivation akin to Belletti's (2012, 2015), with movement of the clefted constituent out of the relative clause followed by extraposition of the relative clause seems to be a good starting point, as illustrated above in (135). This analysis leaves some questions unanswered. For instance, it is not clear how or why the complementiser is generated and selected for in the relative clause; the motivations of most of the movement operations and the extraposition operation also remain unclear. In addition, Frascarelli and Ramaglia (2013: 113) have argued that the existence of intervening adverbial PPs and frame-setters between the clefted constituent and the relative clause are counter-evidence to an extraposition analysis. In their view, these intervening elements can only be interpreted as right dislocates and thus the following clause must also be a right dislocate. While it may be the case that the relative clause in cleft constructions is not a traditional extraposed relative clause, it also does not display the behaviour of a right dislocated clause, as discussed above. Even if we must argue that the process at play in clefting is not the same as traditional relative clause extraposition, it does not seem to me that these data mean that clefting must represent a case of right dislocation.

While there are gaps in the analysis presented here, the outlined analysis above, taking into account all of the various functions and surface structures of clefts, can act as a starting point for further analyses. It should be noted that much of the recent cartographic literature on clefts relies on providing an analysis for a narrow sub-set of cleft constructions: focus clefts of DPs. There is very little discussion of PP and adverbial clefts and almost no discussion of informative-presuppositional clefts. This approach of excluding certain data as outside of the scope of investigation can lead to incomplete analyses.

As we have seen throughout this discussion, a number of contradictions appear to be at play in the case of clefts. The crucial point to be made here is that we must take account of all of the descriptive data before providing an analysis; analyses of clefts which describe only focus clefts or DP clefts are doomed to fail, since they exclude important data. We have provided a potential analysis here; however, we have left the issue of the extraposition landing site, as well as the motivations for the various movement operations at play for further study.

6.6 Concluding Remarks

Our investigation of clefts has several interesting ramifications. Firstly, as with dislocations, we have seen the remarkable stability of marginal, information-structurally marked constructions. Although clefts have seen some minor changes in their syntax over their history, their function has remained stable, and they have been consistently attested across the history of French despite being extremely rare at some stages.

We have seen the importance of fine-grained analysis of corpus data and hand-checking of said data to ensure correct results. Without this hand-checking several examples of clefts would have been missed as well as non-clefts included in our data. This fine-grained approach has also allowed us to provide counter-evidence to several of Dufter's (2008) claims regarding clefts, most importantly, the claims that adverbial and informative-presuppositional clefts are unattested in Old and Middle French. This same rigorous, data-driven approach has also led us to problematise much current work on clefts, showing how informative-presuppositional clefts and adverbial clefts in particular cause

considerable issues for many of the current strands of analysis applied to clefts. Additionally, as in Chapter 5, we've seen how contemporary spoken data can inform an expanded analysis which goes beyond the formal syntactic functions to further interactional functions.

It is only by developing a widely inclusive analysis which takes account of all of the potential functions and structures of clefts as well as incorporating a wide variety of approaches and analyses that we will be able to develop an analysis which truly captures the full function of clefts as well as providing a deeper understanding of the interaction between information structure, syntax and wider discourse pragmatics.

7. Information Structure and Syntax through the History of French

Having discussed the particulars of our three case-studies: post-verbal subjects, dislocations and clefts, I now turn to the broader history of information structure and syntax in French. Incorporating my own work into the existing literature on French historical syntax, I endeavour to present a cohesive picture of the changes which have taken place, as well as showing the limitations of the tools I have used.

This chapter is organised by period. As discussed in our introduction, periodisation should not be used to delineate strict boundaries between different historical varieties, but rather should be viewed as a useful heuristic for approaching historical varieties which cluster together. I begin with a discussion of Old and Middle French, focusing on the availability of movement to the left periphery, the role of the V2 parameter and the large set of subject positions. I then discuss Early Modern French, focusing on the loss of V2, loss of null subjects and reduction of movement options which characterise the period, as well as discussing the role which socio-political shifts played in these changes. Finally, I discuss Modern and Contemporary Spoken French, focusing on distinctions between the written and spoken language and the limitations of a purely cartographic analysis. I show that the cartographic framework provides a strong frame through which to view large-scale syntactic change of the kind found in the history of French, but that various other approaches, particularly historical sociolinguistic approaches to data and a functionally oriented understanding of syntax are useful to flesh out the finer details and truly understand the processes of change at play.

7.1 Old and Middle French

I begin with a discussion of Old and Middle French. Firstly, I discuss the clausal left periphery and its information-structural functions, before turning to the information structure of the middle-field and other approaches to marking information-structural relations.

7.1.1 The Clausal Left Periphery

To open our discussion of the left periphery, I summarize my own data from Chapter 5 on dislocations. Resumed dislocation is rare but consistently present in Old and Middle French. As discussed in Chapter 5, resumed dislocation is produced by base generation of the dislocated constituent in the left periphery. I provide a paradigm example from the period below:

(136) Li quens Rollant, il est mult irascut
The count Roland, he is very angered.

'Count Roland is very angered'

(1100-ROLAND-V,62.737)

While both left and right dislocation, produced by base generation in the left periphery, are found in Old and Middle French, it is notable that a wide number of left-peripheral movement options are also available.

Debate in the literature as to the nature of the immediately pre-verbal position makes analysis of the left periphery in this period somewhat difficult. In particular, Rinke and Meisel (2009) have claimed that Old French is not a V2 language but rather a topic-initial language, while Larrivée (2021) has argued that Old French is V2, but that the pre-verbal position is nonetheless a dedicated topic position and Zimmermann (2014) has argued that the pre-verbal position is a focus position. However, much work in the last 10 years has demonstrated that these approaches are inaccurate. Instead, we find a wide variety of different information-structural constituents in the left periphery (Salvesen, 2013; de Andrade, 2018; Labelle and Hirschbühler, 2018; Wolfe, 2021). Crucially, the immediately pre-verbal constituent is information-structurally under-specified throughout the period (Salvesen, 2013; de Andrade, 2018; Wolfe, 2021). To illustrate I provide examples of immediately pre-verbal topics and foci, taken from Wolfe (2021):

(137) *Un* *filz* lur dunet
A son them gave

'He gave them a son'

Alexis, l. 28 (11th century) (Wolfe, 2021: 52)

(138) Iceste espee porterai en Arabe
This sword will.bring in Arabia

'I will bring this sword to Arabia'

Chanson de Roland, l. 2282 (early 12th century) (Wolfe, 2021: 54)

In example (137), we can see that the indefinite 'un filz' is newly introduced information, a classic example of a new-information focus. In (138), 'this sword' is evidently old information, the topic of the ongoing discourse. Thus, we can see that the immediately pre-verbal position is information-structurally underspecified, allowing both topics and foci.

This immediately pre-verbal constituent is thus typically analysed as occupying the specifier of Fin, while some accounts argue that the locus of V2 changes to Force over the course of Old and Middle French French (Rouveret, 2004; Wolfe, 2018; Wolfe, 2021). Wolfe (2021) specifically suggests that we find a change from Fin-V2 in Early Old French to Force-V2 in the 13th century and back to Fin-V2 in the 14th century.

Besides the immediately pre-verbal position, however, both unresumed topicalization and focus fronting positions are available. To demonstrate this, I provide examples of V3 clauses with topics and foci in the initial position, both taken from Shaw (2022):

(139) Eurus si feri qui tost les remist a la voie
Eurus *si* was who all them put.back to the sail

'It was Eurus who put them all back to sail'

Histoire Ancienne Jusqu'à César, §606.7 (13th century) (Shaw, 2022: 14)

(140) De cele estoire si fu chevetaines Johans de Neele
Of this army *si* was leader Johans of Neele

'Johans of Neele was the leader of this army'

Villehardouin's *Conqueste de Constantinople* §48 (13th century) (Shaw, 2022: 13)

In both examples, the particle *si* fills the immediately pre-verbal position. Thus, we must analyse the focused and topicalized constituents as occupying higher left-peripheral positions. In example (139), Eurus, not previously mentioned in the text, the new information of the clause, is focused. In

example (140) on the other hand, the army has been the topic of the ongoing discourse, while *Johans de Neele* is the new information contributing to the discourse, the focus of the clause. Thus, this is an example of unresumed topicalization, with the army fronted to the left periphery to mark its topical role.

It should be noted that there is some debate as to how to distinguish hanging topics from dislocation, as discussed in Chapter 5. The crucial fact, however, is that both resumed and unresumed topicalization are available in Old and Middle French. While a variety of terms, including hanging topics and bare topicalization have been used for unresumed topicalization, I will refer to this construction as unresumed topicalization in this chapter, while dislocation will be used to refer to resumed topicalization constructions, as in Chapter 5. Unresumed topicalization is often claimed to be due to movement to the left periphery, while resumed dislocation is generated through base-generation. Likewise, focus fronting is typically viewed as a result of movement operations. As such, we can see that, throughout Old and Middle French, movement to the left periphery is widely available.

Throughout this period, as already noted, the nature of V2 has been argued to vary, with V2 localised in Fin during the early Old French period, shifting higher to Force during the 13th century (Rouveret, 2004; Wolfe, 2018; Wolfe, 2021). However, in our data, cases of left dislocation, typically analysed as occurring below Force, are found in prose texts during the 13th century, exemplified below:

(141)

Cil	qui	vivent	selonc	ypocrisie	il	sont	martir	au
Those	who	live	according.to	hypocrisy	they	are	martyrs	to.the

dyable
devil

'Those who live according to hypocrisy are martyrs to the devil'

(1279-SOMME-ROYAL-PENN-P,1,49.1340)

Wolfe (2022), following Benincà and Poletto (2004) argues that LDs can occur above Force. However, instances of initial foci in V3+ clauses are also available in 13th-century texts, as illustrated below (example from Shaw, 2022: 14):

(142)	Eurus	si	feri	qui	tost	les	remist	a	la	voie
	Eurus	*si*	was	who	all	them	put back	to	the	sail

'It was Eurus who put them all back to sail'

Histoire ancienne jusqu'à César §606.7

If, as these data suggest, the low left-left periphery remains active and pre-verbal, it seems that V2 in fact remained in a lower position. That said, Shaw (2022) finds that the *Histoire ancienne jusqu'à César* is anomalous, with other 13th century prose texts showing no evidence of Fin-V2, while verse texts maintain the Fin-V2 grammar. Thus, further research on the period is required to determine the exact nature of V2 and its shifts throughout the Old and Middle French periods.

Thus, we can see that Old and Middle French are characterised by widespread use of the clausal left periphery, both with move and first-merge, as well as a consistently enforced V2 constraint.

7.1.2 The Subject and the Middle-Field

Having discussed the roles of the left periphery in Old and Middle French, I turn to the position of the subject and the middle-field.

Before discussing the middle-field and the number of available subject positions, it is important to discuss null subjects in Old and Middle French. Null subjects are frequent throughout Old and Middle French. However, Old and Middle French are not consistent null subject languages. Null subjects appear to be limited to certain contexts. It was observed as early as Foulet (1935) that null subjects appear in contexts which would otherwise allow post-verbal subjects, with Adams (1987) associating the presence of null subjects with the presence of the V2 constraint. It has also been observed that null subjects are considerably more common in main clauses as compared to subordinate clauses. Finally, some have argued that null subjects show a preference for certain information-structural roles, namely, they are more frequent when they are topical. This can be most

clearly seen in cases of long sections of continuous discourse, where the subject is often left unrepeated. I provide an example of Old French null subject use below:

(143) Lors vint a son cheval.
Then came to his horse

'Then they came to their horse'

La queste del saint Graal, p. 145, (13th century), cited by Ingham (2018: 243)

Discussion of the exact licensing conditions on null subjects is beyond the scope of this thesis. It should simply be noted, however, that Old and Middle French are not consistent null subject languages. The nature of null subjects and their relation to other information-structural constructions is discussed further in our section on Early Modern French.

As we saw in Chapter 4, the Old and Middle French middle-field is characterised by information-structural under-specification and a variety of subject positions.

In simple inversion, the subject remains in its canonical spec-TP position (or Spec-SubjP in our more fine-grained analysis), with apparent 'inversion' the result of the verb moving to the left periphery due to the V2 constraint. Pronominal inversion is, in this period, a sub-type of simple inversion, since no degree of cliticization has begun with pronominal subjects; pronominal subjects obligatorily occupy Spec-TP (or higher positions) and only occur post-verbally when the subject raises higher. Complex inversion has not developed at this stage; I have argued that this is due to the non-cliticised nature of subject pronouns. Simple and pronominal inversion are illustrated below:

(144) Desuz un pin en est **li** **reis** alez,
Beneath a pine REFL is the king gone

'The king went beneath a pine'

(1100-ROLAND-V,11.140)

(145) En tanz lius les avum **nos** portees!
In so.many places them have we carried

'We have carried them in so many lands'

(1100-ROLAND-V,112.1454)

In addition to these post-verbal subject constructions, Old and Middle French display lower subject positions used in cases of free inversion. These positions behave differently with regard to adverb ordering and object ordering with both VSO and VOS orders attested. The full evidence for these positions is discussed in Chapter 4. We may see this variety of low subject positions as paralleling the optional movement of the subject to the pre-verbal position. Some examples of VSO and VOS orders, reprinted from Chapter 4, are provided below:

(146) Lors a geté Helcanor ses yex
Then has thrown Helcanor his eyes

'Then Helcanor cast his eyes'

(127X-CASSIDORUS-PENN-P,341.2764)

(147) Et avoient ordonné une place li Hainnuier ou il se
And had ordoned a place the *Hainnuier* where they REFL

devoient tout retraire.
had all to retreat.

'And the *Hainnuiers* had ordoned a place where they all had to retreat'

(1369-FROISSART-1-P,122.1843)

Crucially, these examples seem to be under-specified for information-structural role, with both topics and foci appearing in the low subject positions. This is in contrast to Modern Italian in which the low subject position is characterised by a focusing function (Belletti, 2001). That said, it should be noted that Belletti (2001) also argues for a low topic position, and her analysis could be extended to an information-structurally under-specified low subject. Nonetheless, unless clear ordering distinctions between low topics and foci can be demonstrated, a single information-structurally under-specified position accounts for the same data. Thus, we have argued that, cross-linguistically, a variety of information-structurally under-specified positions are available throughout the middle-field, in-line with Cinque's (2023) typological approach to the clausal spine. In Old and Middle French, one such position is available, while VOS orders are generated by movement of the subject to the clausal left periphery, followed by TP inversion. This account is contrary to Wolfe's (2021) analysis of inversion

in this period. See discussion in Chapter 4 for a full overview of the low subject system and full examples of information-structural variation.

It seems that the availability of lower subject positions may be related to the availability of null subjects in Old and Middle French, with the null subject filling Spec-TP and thus fulfilling the EPP condition on T, while the overt subject is free to occupy any of a variety of positions along the clausal spine. Although null subjects are more restricted in this period than in Modern Romance null subject languages such as Italian and Spanish (Ingham, 2018; Zimmermann, 2018), it has been observed that the availability of null subjects correlates with the availability of low subject positions (Rizzi, 1982; Sheehan, 2010) and with the V2 constraint (Adams, 1987; Roberts, 1993). The relation between lower subject positions and null subjects will be discussed further in section 7.2.

In summary, Old and Middle French are characterised by considerable freedom with respect to the placement of the subject, potentially linked to the availability of null subjects. Pronominal subjects, both pre- and post-verbal are not cliticised and must occupy the Spec-TP position (or higher positions in the left periphery), while low subject positions are reserved for DP subjects.

7.1.3 Other Strategies for Information Structure Marking

Having discussed the role of specific clausal positions in information-structural marking in Old and Middle French, I turn to other strategies for the marking of information-structural relationships, such as clefts. Old and Middle French are notable for the relative infrequency of routinised constructions to mark information-structural relationships. Additionally, these constructions seem to show many of the features of a standard affirmative matrix clause. Clefts are not yet routinised or invariant; the copula shows verbal agreement with the clefted constituent, rather than the expletive subject. I give an example of an early cleft as well as copula agreement below, both re-printed from Chapter 6.

(148)	cio	fud	Lusos ut	il	intrat.
	It	was	Lusos where	he	entered

'It was Lusos where he entered'

(0980-LEGER-PENN-V,XVII.117)

(149) Sire, Sire, ço sui jó que ai pecchied
Lord, Lord it am I which has sinned

'Lord, Lord, it is I who has sinned'

(1150-QUATRELIVRE-PENN-P,108.4044)

For our full discussion of clefts see Chapter 6. Interestingly, Rouquier (2014: 7) argues that pseudo-clefts are only attested from the 14th-century onwards, I would suggest that this may be due to the fact that clefts do not yet seem to be the invariant construction which they become.

Although clefts are not frequent in Old and Middle French, discourse particles and markers are frequently used to mark information-structural relations. Most notable amongst these is the oft-debated *si*, typically viewed as a marker of topic-continuity (Fleischman, 1991, 1992; van Reenen and Schøsler, 2000; Shaw, 2022). Van Reenen and Schøsler (2000) have also analysed a variety of other Old and Middle French particles through this lens, showing that many of them fulfil a role in marking topic (dis)continuity relations between clauses.

Thus, it should be noted that information-structural roles are not, at this stage, purely fulfilled by syntactic relations, but may also be marked by other means.

7.1.4 Text-Typological Factors

While the Old and Middle French periods represent an under-studied era in historical sociolinguistics, there is nonetheless a rich literature on the impact of text-type on the syntax of the language at this time. Thus, I provide a brief overview of both our own findings and those of the literature in general.

In this work, one key finding in the study of this period has been the frequent use of formulaic language, often specific to a specialised genre. For instance, in Chapter 5, we saw that the bestiary and lapidary texts made a particular use of dislocation structures to introduce the topic of each entry. Similarly, in Chapter 6, we saw that clefts are often used in religious texts to continually bring focus back to Christ in lengthy sections describing his acts. This highly particularised and formulaic use of language seems to be widespread in the period.

In the wider literature, much has been made of the prose-verse distinction in this period. Simonenko, Crabbé and Prévost (2019) find a structured lag in the syntax of verse texts compared to that of prose texts during the Old French period, while Shaw (2022) specifically argues that verse texts in the 13th century seem to demonstrate a Fin-V2, while prose texts shift to Force-V2. Meanwhile, Larrivée (2022b) shows that V2 is sensitive to register in late Middle French (15th century), to such an extent that Larrivée (2022b: 12) states 'what is clear is that V2 is not part of the ordinary competence of speakers by the end of the 15th century'. It seems clear then, that text-typological factors do play a major role in this period. However, we do not see, in our data, wide variation in the frequency of constructions between different genres or registers, excepting those cases of formulaic uses of structures. Nonetheless, it is clear that fine-grained analysis, taking into account small-scale genre variation is crucial to achieving accurate results from our data. By taking into account specific, formulaic functions in specific genres, we can be sure to achieve a more accurate understanding of historical linguistic variation and the causes of change.

7.1.5 Summary

In summary, four major features characterise the Old and Middle French declarative main clause: the V2 constraint, movement to the left periphery, null subjects and the availability of low subject positions. Despite the wide-ranging options for syntactic marking of information-structural relations, other options for marking information-structural relations are available, most notably the wide set of topic (dis)continuity markers explored by van Reenen and Schøsler (2000).

7.2 Early Modern French

Having given an overview of the information structure marking systems of Old and Middle French, I now turn to Early Modern French. I begin with a discussion of the shifts which occur in the left periphery, before discussing the middle-field and subject positions and finally, discussing other options for information structure marking. We shall see that a widespread reconfiguration of information structure marking begins during this period, due in large part to the loss of the V2 constraint and accompanying shifts in the accessibility of the clausal left periphery. I will also argue that many of these shifts are simultaneously enhanced and obscured by the development of a

prescriptive standard during the 16th and 17th centuries, making a fine-grained analysis of these changes particularly difficult.

7.2.1 The Clausal Left Periphery

In Early Modern French, the left periphery begins to shift considerably. This section will largely provide an overview of the literature on the Early Modern French left periphery, since our own work has focused exclusively on dislocations. As discussed above, dislocations are typically seen as the result of base-generation in the left periphery, while focus fronting and unresumed topicalization are seen as the result of movement. It is notable then, that while dislocations marginally increase in frequency during this period, unresumed topicalization and focus fronting, particularly of objects, decrease rapidly (de Andrade, 2018; Wolfe, 2021). Indeed, Wolfe (2021) shows that this decrease begins already in Middle French.

It is also widely agreed in the literature that it is during the late Middle and Early Modern French periods that the V2 constraint is lost in French (Roberts, 1993; Vance, 1995, 1997; Steiner, 2014; de Andrade, 2018; Larrivée, 2021; Wolfe, 2021). Indeed, we have seen this reflected in our own data on simple inversion, which decreases rapidly in frequency during this period. This seems to suggest then, that, in French, the loss of V2 maps onto a loss of accessibility of the left periphery, with only base-generated options remaining. The exact mechanism of this relationship is not clear, however.

Although Wolfe (2021) finds no examples of pre-posed focal objects in his textual samples from the 16th century onwards (Wolfe, 2021: 73), he does cite examples in the literature from Fournier (2007). Fournier (2007: 35) provides several examples of pre-posed objects from 16th-century texts:

(150)

Peu	de	prudence	eurent	les	pauvres	gens
Little	of	prudence	had	the	poor	people

'The poor people had little caution'

La Fontaine, Fables, 7.7 (17th century)

(151)

Telle	science	sceut	le	jeune	prince	Achille
Such	science	knew	the	young	prince	Achille

'The young prince Achille knew such science'

Ronsard (16th century)

This seems to suggest that there may be some gradience in the availability of the left periphery. Movement-based options do not disappear entirely during this period. For example, Wolfe (2021:76-77) argues that unresumed topicalization is strictly reserved for discourse-old topics from 1525 until the 19th century, when it is lost entirely. Indeed, many such options remain as marginal options even in Modern French, where movement to the left periphery is limited to extremely specific information-structural sub-types and specific registers. The fine-grained nature of the cartographic approach is well-suited to capture this. This fine-grained variation will be further discussed below in our discussion of Modern and Contemporary Spoken French.

Despite the survival of marginal left-peripheral movement, the fact remains that the Early Modern period is characterized by a considerable decline in the availability of left-peripheral movement.

7.2.2 The Subject and the Middle-Field

I now turn to the subject and the middle-field. As with the left periphery, the Early Modern period is characterised by rapid change.

The availability of null subjects declines considerably during this period. While in Old and Middle French null subjects were frequent but information-structurally contingent, in Early Modern French they become limited to a small set of constructions. Nonetheless, as with changes to the left periphery, this is a gradual process. Marchello-Nizia and Prévost (2020) provide percentage figures for subject expression in texts from the 9th to 20th centuries, showing that some Medieval French texts in fact have over 90% of their subjects expressed. Nonetheless, while the vast majority of texts before 1500 have fewer than 80% of subjects expressed, the vast majority after 1500 have over 90% of subjects expressed. After 1700, all texts overtly express 92% or more of their subjects. It is thus clear that the Early Modern period represents a turning point in the loss of null subjects in French.

However, Marchello-Nizia and Prévost (2020: 1074) point out that:

'la proximité des chiffres ne signifie pas que la situation moderne est acquise : jusqu'à la fin du 17e s., les sujets non exprimés apparaissent encore dans des contextes que le FMod n'accepte plus, et qui disparaîtront pour la plupart au siècle suivant.'

'The closeness of the figures does not mean that the modern situation obtains: until the end of the 17th century, null subjects still appear in contexts which Modern French no longer accepts and which will disappear for the most part in the following century.'

In particular, null subjects survive in the 16th and 17th centuries in impersonal constructions as well as in contexts where the subject is a recent topic. These two contexts are illustrated below:

(152) a. Et **faut** qu'à la longue la vanité de telle entreprise soit la fable du peuple
'And it must be that, in the long run, the vanity of such an enterprise would be the story of the people.'

Serres, *Le Théâtre d'agriculture et mesnage des champs*, 1603, p. 34, cited by Marchello-Nizia and Prévost (2020: 1074)

b. Au reste du moien, nostre Seigneur **vous** donnera la prudence de le disposer et **estes** sur le lieu pour mieulx pouvoir discerner ce que vos affaires portent.
'As for the rest of the way, our Lord will give you the prudence to deal with it and you are better able to understand what your affairs bring.'

Calvin, *Lettres à Monsieur et Madame de Falai*, 1549, p. 39, cited by Marchello-Nizia and Prévost (2020: 1075)

In (152a) we can see that an impersonal subject may still be omitted, while in (152b) we can see that a subject may be omitted when it is recently mentioned, even if its previous mention is not acting as a subject. In this case, *vous* is mentioned as the object of one clause, before being omitted as subject in the next. Thus, we can see that the contexts in which null subjects are licit gradually narrow over the Early Modern period.

Zimmermann (2020) has argued that the use of null subjects in formal texts during the Classical French period is in fact an archaising structure. Fournier (1998) also makes this assessment in their grammar of Classical French (Fournier, 1998: 22). It seems, then, that the survival of null subjects in specific constructions may in fact be the result of an archaising style in certain texts, rather than a representation of the true grammar.

Null subjects survive most tenaciously in coordinated contexts, where they continue to be used into Modern French. A historical example and a comparable 20th century example are given below:

(153) a. Mais il **eut** le courage tousjours constant, sans se perdre; et, d'un visage ferme, **alloit** au contraire ramentevant <rappelant> à haute voix l'honorable et glorieuse cause de sa mort

'But he always had constant courage, without ever losing himself and, with a firm face, went, to the contrary, retelling in a loud voice the honorable and glorious cause of their death.'

Montaigne, *Essais*, p. 9, (16th century),

cited by Marchello-Nizia and Prévost (2020: 1075)

b. Je reste à Sainte-Maxime jusqu'au 15 puis **remonte** doucement sur Paris par Marseille, Aigues-Mortes, Arles, les Baux, etc.

'I'm staying in Sainte-Maxime until the 15th and then will go back up to Paris via Marseille, Aigues-Mortes, Arles, les Baux, etc.'

Sartre, *Lettres au Castor et à quelques autres*, 1951, p. 353, cited by Marchello-Nizia and Prévost (2020: 1076)

It should be noted that this use of a dropped subject in coordination constructions may not represent a null subject, but a small clause which lacks a subject altogether. Thus, by the 17th and 18th centuries, many cases of omitted subjects are ambiguous between a genuine null subject and a small clause analysis. It is this ambiguity which allows these apparent null subjects to survive into Modern French,

with the relevant clause most likely reanalysed as a small clause. In summary, then, null subjects are lost rapidly in Early Modern French, surviving marginally in a narrowing set of contexts until the 18th century. At this point, the Modern French system is established, with some remaining omitted subjects in coordinated contexts surviving as these contexts are taken as small clauses with no subject position at all.

Turning to post-verbal subjects, I begin with a discussion of simple inversion and post-verbal pronominal subjects. Simple inversion is lost extremely quickly during this period, as the V2 constraint is lost. Only a few marginal examples survive in the 16th century and simple inversion is totally extinct by the 17th century. Remnant examples like this may be bound by micro- and nano-parametric variation, represent a case of a variable rule, or provide evidence for mixed grammars. Since the examples in our data are so scarce it is impossible to make any claims regarding patterned variation. However, these different possibilities are discussed further from a theoretical perspective in Chapter 8. Whatever, the case, simple inversion is totally unattested from the 17th century onwards.

It is surprising then, that post-verbal pronominal subjects, which I have argued to be structurally identical to simple inversion in Old and Middle French, survive (albeit at a decreased frequency). I argue therefore, in Chapter 4, that the nature of post-verbal pronominal subjects shifts, further evidenced by the appearance of complex inversion during this period. In our analysis, building on synchronic analyses on Modern French by Rizzi and Roberts (1989), Rowlett (2007) and Wolfe (2022), post-verbal pronominal subjects are syntactically cliticized, incorporating into the verb, while a null subject fulfils the EPP in Spec-SubjP. This then allows for the form of subject doubling seen with complex inversion, with an overt subject filling Spec-SubjP while the post-verbal pronominal is attached to the verb. The use of Spec-SubjP is adapted from Cardinaletti's (2004, 2021) and Wolfe's (2022) work. In this analysis, the verb moves from V, through the T head to the Subj head; DP subjects are hosted in Spec-SubjP, while pronominal subjects are hosted in Spec-TP, cliticizing onto the verb. The nature of post-verbal subject constructions as root phenomena can then be accounted for by limitations on the triggers of inversion. I do not repeat the full analysis here. For full details, see Chapter 4.

Regarding post-verbal DP subjects, the availability of middle-field subject positions changes rapidly. Although a post-verbal DP subject position is available still during the 16th century, this position becomes highly restricted during the 17th century. VSO orders become extinct and VOS orders become infrequent over the course of the 16th century, while verb-initial constructions decline considerably in frequency, although some marginal examples survive in 17th century diary writing, exemplified below:

(154) jnvocquet le non de la vierge marye les uns pour avoir alegeance
invoked the name of the virgin mary the ones to have protection

de la fiebvre et pour aultres maladye
of the fever and for other illness

'Some invoked the name of the Virgin Mary to have protection from fever and other illnesses'

Valuche p. 8v (17th century)

During this period, a variety of specific adverbs begin to act as triggers for post-verbal placement of the subject. Fournier (2007: 36) notes that these are typically spatio-temporal adverbs, while remaining examples of post-verbal subjects with other adverbials should be treated as archaisms, which become increasingly rare over the period. Examples of adverbially triggered post-verbal subjects are given below:

(155) là coulent mille divers ruisseaux d'une eau claire,
There flow a thousand various streams of.a water clear,

qui distribuent l' eau partout
which distribute the water everywhere

'There, a thousand varied streams of clear water, which distribute the water everywhere, flow'

Fenélon, Tél. 3, 164, (17th century) (Fournier, 2007 : 36)

As with the left periphery, however, it should be noted that verb-initial structures survive to Modern French in a highly limited form too, discussed below in section 7.3. Fournier (2007: 36) notes that 'La plupart des exemples classiques ne sont en rien différents de l'usage moderne' 'The majority of

classical examples are in no way different from Modern usage'. I give an example of such a case of inversion below:

(156)	Ainsi	fut	démontrée	physiquement,	pour	la	première	fois,
	Thus	was	demonstrated	physically	for	the	first	time
	la	sphéricité	&	l'	étendue	de	la	circonférence	de
	the	sphericity	and	the	size	of	the	circumference	of
	la	terre							
	the	earth							

'Thus, for the first time, the sphericity and size of the circumference of the Earth were demonstrated'

Bougainville p. 1 (18th century)

As with the left periphery, we see a considerable decrease in the availability of varied constructions but, crucially, not a total disappearance of these constructions. Notably, the set of adverbs which may trigger post-verbal placement of the subject narrows gradually over this period. Marchello-Nizia and Prévost (2020: 1095) note that 'jusqu'au milieu du 16e s., de nombreux adverbes sont encore suivis de VSnom, qu'ils aient une valeur spatiale, temporelle, de manière, et même épistémique ou discussive.' 'Until the middle of the 16th century, several adverbs are still followed by VS, whether they have a spatial, temporal, manner, or even epistemic or discursive meaning.' An example of a manner adverb triggering post-verbal placement of the DP subject is given below:

(157)	Ainsi	perdit	Nicias	l'	avantage	qu'	il	avoit	nettement
	Thus	lost	Nicias	the	advantage	which	he	had	clearly
	gaigné	sur	les	Corinthiens.					
	won	over	the	Corinthians.					

'Thus, Nicias lost the advantage which he had clearly gained over the Corinthians'

Montaigne Essais, p. 17 (16th century), cited by Marchello-Nizia and Prévost (2020: 1095)

By the end of the 17th century, however, a more-or-less Modern system has established itself. Thus, we can see the remaining contexts in which post-verbal subjects survive as the result of nano- and micro-parametric shifts in the licensing of post-verbal subject positions.

Notably, as part of this Modern system, certain verb-initial structures remain, exemplified below:

(158) Passe sur une civière un blessé ou un mort,
Passes on a stretcher a wounded or a dead

escorté par un peloton de mobiles.
escoted by a platoon of soldiers.

'A wounded or dead man passes on a stretcher, escorted by a platoon of soldiers'

Goncourt Journal, p. 616 (19th century), cited by Marchello-Nizia and Prévost (2020: 1099)

(159) Devront repasser l' examen tous les étudiants qui
Must resit the exam all the students who

ont raté le contrôle continu
have failed the coursework

'All the students who failed the coursework must resit the exam'

(Marchello-Nizia and Prévost, 2020 : 1099)

Since these examples also survive in high register written Modern French and have been analysed extensively in the work of Lahousse (2003, 2006, 2022), I leave discussion of them to section 7.3.

7.2.3 Other Strategies for Information Structure Marking

Having discussed the shifts in the left periphery and middle-field In Early Modern French, I turn to other strategies for information-structural marking. Clefts increase slightly in frequency during this period. It is not until Contemporary French, however, that these constructions increase considerably in frequency. Additionally, pseudo-clefts develop in this period, perhaps in response to the increasing routinisation of clefts. I have argued in Chapter 6 that the increase in frequency of clefts responds to the decrease of frequency of focus fronting, contra Dufter (2008, 2009). This shift also seems to correspond to increasing routinisation of clefts, which, during this period, begin to feature copula agreement with the expletive subject, not the clefted constituent, a step towards an increasingly invariant structure.

(160) c' est moi qui le dis
It is me who it said

'It's me who said it'

Le Sopha p. 22 (18th century)

In addition to these developments in the nature of clefts, changes to the distribution of information-structural particles occur during this period. Notably, many such particles decline in frequency or disappear. In particular, it is during this period when the discourse particle *si* disappears entirely (Fleischman, 1991, 1992).

In summary then, the loss of left-peripheral and middle-field positions seems to be reflected in a slight increase in the frequency of cleft constructions. However, this increase in frequency is neither sudden, nor overwhelming. Additionally, some other tools for information-structural marking, such as the particle *si* are lost during this period. Thus, it seems that information-structural marking is not driving syntactic change in this period.

7.2.4 The Role of Prescriptivism

We have seen, thus far, that the period from the 16th-19th centuries represents a period of radical shifts in French syntax, with a wide variety of left-peripheral positions, subject positions and null-subjects becoming considerably rarer and more restricted. However, this is also a period of radical social shift, with the establishment of the Académie Française and development of prescriptive ideology in the work of the grammarians. Indeed, these developments appear to obscure the shifts which are occurring in the spoken language. In our investigation of working people's written texts from this period, we found that some *journaux intimes* showed rates of null subjects and post-verbal subjects much more in-line with those of Medieval French texts (see discussion of Chavatte in Chapter 4). On the other hand, however, Zimmermann (2020), utilising similar data from working people's diaries, claims that the surviving cases of null subjects in formal texts are in fact archaizing. Examining metalinguistic texts suggested that a more complex situation was at play, with grammarians asserting that the basic word order of the language was SVO, but nonetheless that the "correct" use of inversion was an example of highly refined language use. Thus, it is hard to determine whether prescriptivism had any role in accelerating or slowing the change, since the views of the grammarians were not simplistic.

As such, it seems that, with currently available data, it is not yet possible to truly trace the full details of these shifts or to directly identify the role which prescriptivism played. Our data are also not sufficient to suggest that the prescriptive ideology drove these changes rather than accelerating them or making them apparent in the text-language earlier than they otherwise would be. It seems, however, that the opacity caused by the development of prescriptive ideology does pose a large problem for fine-grained historical linguistic research. When huge shifts occur in the technologies and social context of text production and linguistic structure, it is impossible to follow constructions continuously and track change at a fine-grained level. This problem is particularly acute for syntactic change, which is frequently under-discussed by prescriptive grammarians, who prefer to focus on lexical and phonological variation. This issue may be insurmountable but may also be addressed by the construction of more corpora and accessible datasets of writing from writers of varied backgrounds. The importance of this sociohistorical approach is further discussed in Chapter 8.

7.2.5 Summary

In summary, the Early Modern period is a period of considerable shift, both socially and linguistically. The left periphery becomes considerably less accessible to movement, while the various middle-field subject positions become less accessible and subject to more clear restrictions. Alongside this, null subjects become highly restricted to coordinated contexts, where they continue to survive in the modern language. Finally, clefts begin to increase in frequency as well as becoming more idiomatised, with the copula shifting from agreement with the clefted constituent to agreement with the expletive subject. Thus, the ground is laid at this stage for the development of the Modern French system.

7.3 Modern and Contemporary Spoken French

While Early Modern French represents a period of paradigmatic shift in the information structure-syntax relationship in the history of French, Modern and Contemporary Spoken French see much smaller shifts occur, alongside increasing specialisation of information-structural constructions.

7.3.1 The Clausal Left Periphery

While, in Early Modern French, we saw that focus fronting and other movement-based left-peripheral operations decline considerably in frequency, it is widely assumed in the literature that such operations are totally unavailable in Modern French (see Authier and Haegeman, 2019: 5), with Rowlett (2007: 182) stating that 'French doesn't allow focus fronting', although he notes that it is still available in high-register Modern French.

However, as Authier and Haegeman (2019) identify, mirative focus fronting remains available in French. Mirative foci are those foci which include an element of shocking or surprising content, an example is given below:

(161) Plus de cent lettres (que) j'ai trouvées dans notre
More than one-hundred letters (that) I-have found in our

boîte!
mailbox

'I've found more than one hundred letters in our mailbox!' (Authier and Haegeman, 2019: 15)

Although Authier and Haegeman (2019) identify only mirative focus fronting in Modern French, Larrivée (2022a) has argued that cases of contrastive focus and presentative focus fronting can be found too:

(162) Aux Forestiers, on a bien envoyé les soixantes bâtons
To.the Forestiers, we have well sent the sixty sticks

qu' ils avaient commandés, mais aux Gaulois cinquante bâtons
that they had ordered, but to.the Gaulois fifty sticks

(qu') on a envoyé (Contrastive Focus)
(that) we have sent

'To the Forestiers we did send the sixty sticks they ordered, but to the Gaulois we sent fifty sticks'

(Larrivé, 2022a: 192, originally from Authier and Haegeman, 2019: 20)

(163) Ouvert 24 h sur 24 il y avait écrit
open 24 hours on 24 it there had written

'Open 24/7" it was written.'

(Larrivée, 2022a: 193; originally from Blanche-Benveniste, 1996)

In (162), we can see that the fifty sticks in the second clause are not being presented as surprising or shocking, but simply contrastive against the sixty sticks in the previous clause. Meanwhile, in (163), there is certainly nothing surprising about the fronted quotative 'Ouvert 24 h sur 24', but it is nonetheless informationally new and seems a clear case of a presentative focus. Lahousse (2022: 14) on the other hand argues that only mirative and corrective focus fronting are available in Colloquial French, while other forms are available in standard and higher register varieties.

In addition to focus fronting, Authier and Haegeman (2019) suggest that some cases of unresumed topicalization can be identified with PPs, exemplified below:

(164)

Marie	a	réuni	ses	élèves.	Aux	filles,	elle	a	donné
Marie	has	gathered	her	pupils	to-the	girls	she	has	given

des	exercices	d'algèbre.	Aux	garçons,	elle	a	dicté
some	exercises	of-algebra	to-the	boys	she	has	read-out

un	problème	de	géométrie.
a	problem	of	geometry

'Mary has gathered her pupils, she has given some algebra exercises to the girls and she has read out a geometry problem to the boys'

Authier and Haegemann (2019: 5, originally from Delais-Roussarie, Doetjes and Sleeman, 2004: 512)

These examples seem to me to be limited to cases of contrastive topicalization. It seems then, that Modern French maintains some left-peripheral movement in a highly limited set of contexts, partially informed by socio-stylistic factors.

Having discussed the status of moved left-peripheral constructions in Modern and Contemporary Spoken French, I turn to resumed dislocation. As discussed above, this construction is derived by first-merge in the left periphery. As we saw in Chapter 5, both left and right dislocations increase rapidly in frequency during the 20th century. Additionally, both are particularly common in

spoken registers, although their frequency does increase considerably in written French too. During this period, dislocations also acquire interactional functions as turn-taking markers (Doehler, Stefani and Horlacher, 2015). It is perhaps this multi-functionality which leads to such a large increase in frequency.

With regard to their syntax, some have suggested that dislocations are cases of subject doubling, with subject pronominals functioning as agreement markers on the verb (Harris, 1978; Sportiche, 1996). However, this view has been largely discredited. Here, I follow Roberts' (2009) view that only post-verbal pronominal subjects are syntactically cliticised, while pre-verbal pronominals are full subjects, with some phonological cliticization occurring. This is particularly clear when we compare Gallo-Romance varieties, such as Picard, which have systematic subject doubling (Auger, 2003). For our full analysis, see discussion in Chapter 5.

In summary, the Contemporary Spoken French left periphery is characterised by severe limitations on left-peripheral movement, with some marginal constructions such as mirative focus fronting available, while resumed dislocation is overwhelmingly common, generated by first-merge in the left periphery.

7.3.2 The Subject and the Middle-Field

I now turn to the middle-field in Modern and Contemporary Spoken French, focusing on post-verbal subjects.

I begin with a discussion of post-verbal pronominal subjects. Post-verbal pronominal subjects continue to survive in a set of restricted contexts, with certain specific adverbial triggers such as *peut-être* and *toujours* (Lahousse, 2022: 6). Since these triggers are lexically specified, Lahousse (2022: 6-7) argues that they represent cases of nano-parametric variation. I exemplify below:

(165) Peut-être viendra-t-il.
maybe come.3SG.FUT-he

'Maybe he'll come'

(Lahousse, 2022: 6, originally Rowlett, 2007: 208)

At this stage, post-verbal pronominal subjects are fully incorporated clitics which license a null subject in Spec-SubjP. For a full discussion of our analysis of post-verbal pronominal subjects, see Chapter 4.

DP subject positions are likewise highly limited in Modern and Contemporary French. Main clause declarative post-verbal DP subjects are only found in the written data of our sample. However, the broader literature attests that such structures exist with extremely limited distribution restricted to certain specific triggers.

As discussed in Chapter 4, Lahousse (2003, 2006, 2022) provides a typology of the available post-verbal DP subject structures. In our analysis here, we can see this as the result of gradual micro-parametric change to the licensing conditions of post-verbal subjects. In Modern French two sub-types of post-verbal DP subject remain: genuine stylistic inversion, in which pre-verbal material triggers inversion, and focus inversion, in which the post-verbal subject must be exhaustively focused.

I begin with a discussion of genuine stylistic inversion. While the adverbs which trigger pronominal inversion are subject to nanoparametric variation, Lahousse (2022) demonstrates that the adverbs and adjectives which trigger genuine stylistic inversion have the semantic features [+anaphoric] or, more rarely, [+scalar].

(166) Il lui tenait (…) des discours aussi sentimentaux

'He would also give her sentimental speeches'

(sentimentaux	sont	aussi	ceux	que	tant	de	jeunes
sentimental	are	also	those	which	so.many	of	young

aristocrats	tiennent	à	quelque	ravissante	jeune fille...).
nobles	give	to	some	pretty	young girl…

'(sentimental also, are those which so many young aristocrats deliver to pretty young girls…)'

(Proust, cited in Lahousse, 2022: 9; originally cited in Le Bidois 1952: 181).

In (166) we see that the fronted adjective 'sentimentaux' is being used anaphorically, referring back to the already mentioned sentimentality of the character's speeches. In this case, the adjective is not lexically [+anaphoric] but its anaphoric nature is determined contextually. In contrast to the

nanoparametrically determined pronominal inversion triggers, this class is productive in registers such as journalistic writing where words which are both [+anaphoric] and [+scalar] such as *rare* 'rare' and *nombreux* 'numerous' have become triggers for post-verbal placement of the subject. We can also link this class of triggers to our discussion of the left periphery and Zimmermann's (2020) analysis of null subjects. Lahousse (2022) argues that these fronted adverbs and adjectives are cases of mirative focus fronting. In Lahousse's (2022) analysis this then licenses a V2 structure. However, the DP subjects consistently occur in a free inversion position, after both auxiliary and main verb. As such, I argue that these represent cases of free inversion, with the subject in a low subject position. Thus, mirative focus fronting can be said to license a null subject in Spec-TP, as Zimermann (2020) argues for both Classical and Modern French. This null subject then licences a low DP subject.

The position of the subject in genuine stylistic inversion is extremely low, contrasting against the variety of middle-field positions available in other periods. Indeed, Lahousse (2003, 2006) specifically argues that the subject remains VP-internal. However, I do not make a specific claim as to the position of the subject here, simply noting that it may not be in the same middle-field position as Old and Middle French, but an even lower position, accounting for the loss of VSO orders.

Notably, however, these post-verbal subject constructions are largely unattested in Contemporary Spoken French, with most attestations found in journalistic written French or literary writing.

Turning to focus inversion, Lahousse (2003, 2006) shows that exhaustively focused subjects can occur post-verbally, including in V1 structures. This is exemplified below:

(167)	Ont	mangé	une	pomme	**seules**	**les**	**filles**
	Have	eaten	an	apple	only	the	girls

'Only the girls have eaten an apple' (Lahousse, 2003: 221)

Lahousse (2003, 2006) argues that these cases should be modelled as cases of movement to the clausal left periphery followed by remnant movement of the TP to a topical position. Thus, Lahousse (2003, 2006) does not make use of Belletti's (2001) *v*P-periphery. For a full summary of our analysis of post-verbal subjects, see Chapter 4.

In summary, we see, as with left-peripheral movement, that post-verbal subjects only survive in highly restricted contexts, defined by a trigger adverbial or information-structural functions in combination with sociostylistic context. It seems that further research is required to formalise these fine-grained distinctions into a cartographic framework. While much work has been undertaken to elucidate fine-grained information-structural distinctions in the cartographic framework, sociostylistic contexts have not yet been explored through this lens.

7.3.3 Other Strategies for Information Structure Marking

Turning to other means of marking information structure, clefts are found more frequently in this period. We may claim that this is also the result of the pressures of spoken language leading to frequent use of idiomatised and lexicalised constructions. We have seen that work such as that of Lahousse and Lamiroy (2017) on *c'est ainsi que*… clefts suggest that these constructions have become lexicalised in some cases. Notably, however, clefts are still considerably less frequent than dislocations.

Much work has also been conducted on the role of phonological markers in French information structure, although it is typically argued that these do not map neatly to topic and focus distinctions. Full exploration of the phonological aspect of information structure is beyond the scope of this thesis.

7.3.4 Summary

In summary, Modern and Contemporary Spoken French continue the processes of paradigmatic change begun in the Early Modern period. Notably, however, while many constructions become highly limited in frequency, few become totally extinct, with focus fronting surviving in mirative focus fronting and post-verbal subjects surviving in a variety of limited environments. These structures also typically survive more in higher register speech and written varieties. It seems, then, that remnants with narrow functional and sociostylistic boundaries are typical of change in the syntax-information structure interface. We have also seen a distinct split in spoken and written use of information-structural constructions. This may present a substantial issue for sociohistorical linguistics, since even low register texts, such as text messages and forum writing, exhibit a different syntax to that of the spoken language. It is important, then, that sociohistorical work consistently engages with the

limitations of its data, rather than idealising low-register data as more spoken-like than high register texts. A truly socially engaged sociohistorical linguistics appreciates the value of both high and low register texts, while understanding the inherent limitations of all textual historical linguistics. These sociohistorical issues are discussed further in Chapter 8.

7.4 Summary and Limitations

Before summarising our findings with regard to information structure and syntactic change, I provide a schematization of the various information structure-syntax constructions found throughout the history of French. Given the frequent importance of micro- and nano-parametric changes, I use a five-point scale of frequency here: frequent, well-attested, marginal, extremely marginal and unattested. This schematisation is not based on specific frequency data but is intended to give an impressionistic, qualitative overview of the changes we find. From this tabulation we can clearly see a number of shifts in French syntax and information structure.

Period / **Construction**	Old French	Middle French	Renaissance French	Classical French	Modern French	Contemporary Spoken French
Focus Fronting	frequent	well-attested	marginal	marginal	marginal	extremely marginal
Unresumed Topicalization	frequent	frequent	marginal	extremely marginal	extremely marginal	unattested[32]
Left Dislocation	marginal	marginal	well-attested	well-attested	well-attested	frequent
Right Dislocation	extremely marginal	extremely marginal	marginal	marginal	marginal	well-attested
Simple Inversion	frequent	frequent	extremely marginal	unattested	unattested	unattested
Free Inversion	frequent	frequent	well-attested	well-attested	marginal	extremely marginal
Pronominal Inversion	frequent	frequent	well-attested	well-attested	marginal	marginal
Complex Inversion	unattested	unattested	extremely marginal	marginal	marginal	marginal
Null Subjects	frequent	frequent	well-attested	marginal	extremely marginal	extremely marginal
Clefting	marginal	marginal	well-attested	well-attested	well-attested	well-attested

Table 1: Frequency of Relevant Structures across the History of French

[32] Examples of contrastive topicalization were provided above in (164) but since it is unclear whether these are attested in more colloquial Contemporary Spoken French, I have only marked contrastive topicalization as extremely marginal in Modern French, and extended this to Classical French, since it seems implausible that such cases would become extinct before reappearing in Modern French.

First amongst these is the shift away from left-peripheral movement. While, in Old and Middle French, both topic and focus movement to the left periphery are well-attested phenomena, both decrease rapidly in frequency during the Renaissance and Classical French periods (de Andrade, 2018; Wolfe, 2021). By the Modern French period, both unresumed topicalization and focus fronting have become limited to highly specific information-structural roles, with some debate as to the exact distribution (see Authier and Haegeman, 2019; Larrivée, 2022a; Lahousse, 2022), representing an extreme narrowing of function. Meanwhile, non-movement left-peripheral strategies, such as dislocation survive and increase in frequency.

With regard to post-verbal subjeccts, shifts in the frequency of post-verbal pronominal subjects can be seen as resulting from the changing nature of post-verbal subject clitics. While simple inversion is lost as V2 is lost, post-verbal pronominal subjects, originally a reflex of V2, survive. I have argued that this survival is due to the cliticization of the post-verbal pronominal subject, with a null subject fulfilling the EPP in the canonical subject position. Complex inversion develops concurrently with this shift, representing the case in which a full DP subject occupies the canonical subject position, rather than a null subject. In our argumentation, the specific contexts of pronominal and complex inversion, i.e. specific adverbial triggers and interrogation, should be seen as licensing the post-verbal pronominal clitic. It is then this clitic which licenses the null subject in Spec-SubjP. For the full details of this analysis, see Chapter 4.

Alongside the shift away from left-peripheral movement, unmarked DP subject positions become more restricted, with the low subject positions available in Old, Middle and Renaissance French being lost over the course of Classical French, leaving us with highly limited low subject positions in the Modern and Contemporary Spoken language. It seems plausible that post-verbal subjects will disappear completely in the near future, given the total lack of such structures in some colloquial varieties (see Roberts, 2009; Lahousse, 2022). It should be noted that the low subject positions available in Medieval and Early Modern French and the features which distinguish them from low subject positions in other Romance languages has not been previously noted in the literature.

The remnant DP inversion contexts in Modern and Contemporary Spoken French seem to be linked to licensing conditions on null subjects, following Roberts' (2009) analysis of post-verbal subjects. These cases are then governed by micro- and nano-parametric shifts of certain pre-verbal licensers (namely adverbs such as *aussi* and *ainsi*) and sociostylistic factors, with journalistic writing making particularly productive use of post-verbal subjects. Formally accounting for the three-way interaction between pre-verbal triggers, null subjects and low subject positions, the latter two of which appear to be mutually dependent on each other, is an important next step in this research. Similarly, it is important for our formal analysis to begin work on formalising sociostylistic factors, since it seems that they are crucial to a proper understanding of syntactic change.

In summary, I follow Roberts' (2009) account of Modern French as a '(very) partial null-subject language' (Roberts, 2009: 311) in which the licensing of null subjects leads to the overt expression of the subject low in the clausal spine.

A notable finding of this work is that information-structural pressures do not play a central role in the syntactic changes which we see occurring. Although clefts increase in frequency in response to the loss of focus fronting, they do not increase significantly enough to fully replace its function. It seems that syntactic marking of information-structural roles is just one of many options for encoding them in the linguistic signal. As such, information structure cannot be said to drive syntactic change in the history of French. Each of the changes discussed so far is, instead, the result of gradual, probabilistic shifts in data, leading to gradual, often small reanalyses and leaving us with remnant structures which maintain aspects of earlier grammars due to micro- or nano-parametric change. It is, in fact, a confluence of strictly formal, information-structural, functional and social factors which drives change, with no single factor having priority over the others in all instances.

This discussion has thus far only focused on the interplay of the V2 constraint, null subjects, post-verbal subjects and left-peripheral movement. However, much of this thesis has focused on other information-structurally marked constructions: namely, dislocations and clefts.

With regard to dislocations, we have seen that these can be well-accounted for under cartography through an account in which they are first-merged in the clausal left periphery. Under this

account, their increase in frequency in the Modern language is readily accounted for as a response to the loss of other topicalization techniques, particularly hanging topics. However, as we have seen, this strict focus on information-structural functions leaves us with an incomplete picture. In textual examples, weight plays a significant role in the distribution of dislocations, with longer DPs requiring resumption in the main clause for simple, practical reasons. Likewise, in the contemporary spoken language, the high frequency of dislocations seems to reflect functions beyond the usual information-structural roles, with interactional functions as turn-taking markers playing some role in their high frequency. This expanded analysis, in which non-information-structural functions are taken into account, allows us to provide a clearer account of the variation and change in the history of dislocations, particularly of the wide differentiation found between the contemporary spoken and written language.

We have also seen that cartographic approaches have limits when applied to clefts. Clefts appear, in some cases, to lack the sort of internal structure which cartographic approaches are designed to analyse. As noted above, certain clefts have been argued to be lexicalised as discourse particles. This process of the lexicalisation of syntax is difficult to directly analyse under cartography. Even in their least lexical iterations, clefts demonstrate a perplexing type of internal structure which also seems difficult to account for. In these cases, it seems that an approach which goes beyond cartography's focus on internal syntactic structure of the clause may be useful. In Chapter 8, I discuss potential alternative approaches to these issues.

In addition to the specific limitations that we have found when applying a cartographic approach to these data, there have been practical limitations in this work. Firstly, throughout this chapter I have drawn on the broader literature, rather than my own data, to draw out a wider analysis of the relationship between information structure and syntactic change. It is fortunate that such a rich literature now exists on this topic, but to truly understand the gradual and fine-grained changes which occur, direct analysis of the data is necessary. Particularly, further work is required on the shifts to post-verbal subjects in Early Modern French, as well as the narrowing of function of focus fronting. Additionally, throughout this thesis we have discussed and compared only those cases in which

information-structural functions are overtly marked. For a full understanding of the interplay between information structure and syntax, we must also observe those cases in which no overt marking takes place. Annotating textual data for information-structural roles is notoriously hard and so conducting historical work on this topic is laborious. Additionally, while such analysis would allow us to distinguish between syntactically marked and unmarked information-structural roles, it would not allow us to examine the role of phonological marking. In analysis of spoken languages, phonological marking is often considered to play a role in the identification of foci, although the extent of this role in French is debated (see Lambrecht, 1994; German and D'Imperio, 2016; Vander Klok, Goad and Wagner, 2018). This may prove to be an absolute limitation on historical work on information structure and syntax. The spoken and written language are also fundamentally different in some ways, notably in the frequency of dislocations in the modern language. As such, there are hard limitations on what can be achieved with historical linguistics, even when using the most 'spoken-like' data with a well-considered, fine-grained sociohistorical approach. Thus, while we have uncovered much regarding the interplay of different factors in syntactic change, there is much left to discover and much which may be undiscoverable. In the following chapter, I further discuss these limitations and potential further directions, including ways in which formal analysis can incorporate sociostylistic factors.

8. Critical Directions

8.1 Introduction

As we have seen, the cartographic framework has its limitations in capturing the intricacies of our data. Although it can capture immense detail in elaborating the various subject positions available in French (through its elaborate clausal spine), as well as capturing the intuitive relation of the dislocated element to its clause (base-generated, not moved), I have struggled to adequately account for the structure and history of clefts with cartographic analysis and, at present, there is no clear means to account for the interactional roles of dislocations or sociostylistic variation in cartography. These issues may be simply considered outside of the scope of cartography, which has been described as 'the attempt to draw maps as precise and detailed as possible of syntactic configurations' (Cinque and Rizzi, 2008: 42) and need not provide a full elaboration of all possible functions of such configurations. Nonetheless, it is perhaps worth considering the underlying assumptions of the cartographic approach and how they may help or hinder us in our fundamental endeavour: describing the nature of human language and its processes of change.

In this chapter, then, I turn to a broader critique of the assumptions which underly the cartographic approach, namely, generative assumptions regarding the nature of language and their implications for historical linguistics in general and historical syntax in particular. I argue that many assumptions of the generative approach to historical syntax may be untenable in light of increasingly fine-grained sociohistorical data, including much of the data presented in this thesis. Nonetheless, I go on to argue that, despite these flawed assumptions, we can continue to utilise the descriptive formalisms of cartographic syntax and the generative research program in general, combined with a socially grounded understanding of language and language change.

I begin with an overview of generative assumptions in their strictest form: namely those utilised in so-called 'biolinguistics' (Hauser, Chomsky and Fitch, 2002; Boeckx, 2006) and the minimalist program (Chomsky, 1995). Although I will be drawing my summary of these assumptions from works from the 1990s and 2000s, these works typically discuss the history of the approach and

particularly the Principles and Parameters framework. As such, I will take it as given that similar assumptions can be said to underlie generative work from the mid-1970s onwards. I discuss the ramifications that such assumptions have for historical syntax, discussing a variety of generative approaches to language change, including early work on the notion of reanalysis and 'catastrophic' change from Lightfoot (1979), competing grammar approaches following Kroch's (1989) seminal work on the so-called Constant Rate Effect and the more recent parameter hierarchy approach of Roberts (2019).

I then turn to my own data's implications for these theories of change, suggesting that neither catastrophic change nor competing grammars appear to be at play in language change, rather, linguistic change seems to occur at a very fine-grained level, perhaps even involving mixed grammars.

From here, I critique the generative approach, touching on a variety of issues, largely from the sociohistorical literature. I show how a variable rule approach, following Labov (1969), allows us to incorporate intuitions about the socially grounded nature of language.

Crucially, I do not take my critique of generative assumptions to invalidate the results of generative research. Rather than arguing against this research, I simply argue that we must be open to the broader data available and the ways in which different levels of linguistic and social variation are highly interconnected, not independent. Thus, I argue that we do not need to abandon the formalism and insights of the generative research program, despite the problematic assumptions which have underpinned this research historically. Here, I follow recent work by Corr (2022) who fully abandons the approach to language at the core of generative work, while continuing to utilise the cartographic formalism and its insights, indeed, expanding on such a formalism. I also note, however, that varied formal approaches may be of use for achieving deeper insights into syntactic history, particularly in the case of interactional functions of syntax and opaque syntactic constructions such as clefts.

8.2 Generative Historical Syntax

I begin with a discussion of the history of generative approaches to historical syntax. Here, I hope to assess how well these approaches could be used to describe our fine-grained sociohistorical data. We

shall see that certain assumptions of the generative approach are problematic for our data, and some of our base assumptions need reconsidering or at least critiquing.

Before I turn to historical generative approaches, however, it's crucial to establish what exactly I am discussing when discussing the generative approach. The approach I will discuss and critique here is essentially the most recent generalised version of the generative assumptions provided by Chomsky and his associates, as found in Chomsky's (1995) work on the minimalist program, the biolinguistics approach outlined in Hauser, Chomsky and Fitch (2002) and works of other theorists in biolinguistics such as Boeckx (2006). Under the view presented in Boeckx (2006) the language faculty should be viewed as a discrete, innate and universal 'mental organ' (Boeckx, 2006: 27). Chomsky (2015[1995]: 12) lays out the concerns of the generative research program most clearly:

> 'We are concerned, then, with states of the language faculty, which we understand to be some array of cognitive traits and capacities, a particular component of the human mind/brain. The language faculty has an initial state, genetically determined; in the normal course of development it passes through a series of states in early childhood, reaching a relatively stable steady state that undergoes little subsequent change, apart from the lexicon. To a good first approximation, the initial state appears to be uniform for the species. Adapting traditional terms to a special usage, we call the theory of the state attained its grammar and the theory of the initial state Universal Grammar (UG).'

Two key assumptions can be seen in this summary. Firstly, language and grammar are innate and essentially biological objects of study (on this biocentric view of language, see particularly Boeckx (2006)). Secondly, a language's syntax is fixed in child language acquisition. This view has several important ramifications for historical and diachronic research. Firstly, since the grammar of a speaker cannot change after child language acquisition, change must occur through child acquisition, this view is widely espoused in the generative historical literature (Lightfoot, 1979; Battye and Roberts, 1995: 7; van Gelderen, 2011: Chapter 1; Ledgeway, 2012: 188). Secondly, the biocentric and internal view of language encourages us to take all linguistic data as being representative of the same underlying object

of study, an abstracted internal grammar of the language. Thus, this theoretical approach indirectly leads us towards a methodology disengaged from social factors in variation and change.

Having established the generative assumptions that will form the basis for this chapter's discussion, I turn to the generative historical linguistic literature. I begin with the one of the earliest works of generative historical syntax: Lightfoot's (1979) seminal work. Here, it is argued that syntactic change occurs through 'catastrophic reanalyses.' In this view, syntactic change begins with small changes to the surface form without change to the underlying structure. Once sufficient 'piecemeal changes' have built up, a catastrophic reanalysis occurs, and the entire structural syntactic system shifts, with a large number of changes occurring concurrently as one grammar is replaced by the next. Central to this theory is the view that grammars are fixed during child language acquisition and cannot be further changed thereafter (Lightfoot, 1979: 137). As such, syntactic change must happen in catastrophic leaps as parametric change occurs when children reanalyse the linguistic data in the process of acquisition. While we will critique this theoretical approach, it is important to acknowledge those cases in which it is useful. We saw, in our discussion in Chapter 7, that the Early Modern French period is one of great change: the V2 constraint and null subjects are lost concurrently, while movement to the left periphery is highly reduced. While Lightfoot's (1979) original claim that such shifts happen 'catastrophically' does not seem to hold, it does seem clear that large parametric shifts can occur quickly, over the course of two centuries or so (compare the language of 15th- and 17th-century French and you will find wildly differing syntactic systems).

While Lightfoot's (1979) work is well worth critiquing, the tradition of generative diachronic syntax has moved beyond this 'catastrophic' approach to change. Most crucial to developing beyond this catastrophic change model was the work of Kroch (1989), who developed the Constant Rate Hypothesis (CRH). The CRH states that 'although the rate of use of grammatical options in competition will generally differ across contexts at each period in time, the rate of change will be the same across contexts' (Kroch, 1989: 206). Kroch (1989) argues that this constant rate of change is attributable to change in the underlying grammar's parameters, with the differing start points across different contexts attributable to sociostylistic variation. As such, this approach allows the generative

historical linguist to dismiss sociostylistic variation as outside the scope of study. The CRH has been argued to be 'a useful tool in the armoury of diachronic syntacticians who wish to argue for "the controlling effect of abstract grammatical analyses on patterns in usage data" (Kroch 1989:239)' (Kauhanen and Walkden, 2018: 484). The CRH allows us to maintain the view that linguistic change happens during child language acquisition, with each speaker having a single grammar.

These more nuanced approaches to syntactic change have been further developed in recent years by Roberts' (2019) work on micro- and nano-parameters, which show how linguistic change can take place piecemeal through child language acquisition, affecting only individual lexical items or semantic classes of lexical items during each generational shift. We have seen, in Chapter 4 of this thesis, that variability frequently occurs in single texts, with marginal examples of simple inversion surviving in texts which do not, in general, exhibit a V2 grammar. Unfortunately, we have insufficient data on simple inversion to determine whether these surviving examples can be seen to be micro- or nano-parametrically delineated or instead represent something closer to a mixed grammar. However, in the case of free inversion and pronominal inversion, we saw that changes to licensing conditions from Early Modern French to Contemporary Spoken French seem to be clearly governed by nano- and micro-parametric changes.

If we were to find evidence of mixed grammars in language change, it would be highly problematic for the current biolinguistics assumptions of the generative framework. However, these assumptions are rarely directly relevant to our analysis. The bulk of our theories of syntax, particularly in more descriptive models such as the cartographic framework, do not directly rest on these assumptions. Regardless of whether these assumptions hold, we can continue to make use of these descriptively powerful formalisms.

8.3 Critical Directions

We have seen that, although cartographic approaches are limited and some of their assumptions problematic, they remain a useful tool for fine-grained analysis of syntactic structures. Nonetheless,

the issues of generative assumptions should be taken seriously. As such, I now turn to alternative approaches to modelling linguistic variation and change alongside critiques of generativism.

The formal-functional divide is widely known within linguistics. Generativists have long been criticised by those in the functional tradition for focusing on a supposed biologically inherent formalised structure unique to language. For functionalists, language is a result of domain-general cognitive processes, where the function of language (typically taken to be communication) is primary, with the form shaped by this function (see van Valin, 2003 for an overview). Much has been made of the functionalist-formalist debates in linguistics (most notably, see Newmeyer, 2000). However, I will not provide a full discussion of functionalist approaches here. Many of the original critiques of functionalist authors have been increasingly dealt with by cartographic approaches, most notably with the inclusion of information-structural and discourse-oriented analyses in syntax following Rizzi's (1997) work. The issues we have faced in this thesis are largely related to sociostylistic factors in syntactic change. Thus, socially oriented critiques of the generative approach are of particular interest to us.[33] Before I move to these however, it seems important to discuss existing approaches which are based in the formalism of cartography but make divergent assumptions.

8.3.1 Corr's Un-Cartesian Hypothesis

In recent years, some researchers in the cartographic tradition have critiqued the limits of the current cartographic framework. In this section, I will discuss Corr's (2022) work on the syntax of the utterance in Ibero-Romance and the un-Cartesian hypothesis she presents here.

In Corr's (2022) work, the cartographic framework is taken as a jumping off point for a wider discussion of utterance syntax in Ibero-Romance. The crucial findings of Corr's (2022) work are that syntactic structuring extends beyond even the left-periphery of Rizzi (1997) to a structural framework for the entire utterance and discourse. Crucially, however, Corr (2022) argues that, in this approach,

[33] It should be noted that, while much of the functionalist literature discusses sociostylistic variation as little as the generative literature, two approaches do merit further discussion: interactional and dialogic linguistics. The interactional approach (e.g. Fox, Thompson, Ford & Couper-Kuhlen, 2013) and the dialogic approach (e.g. Linell, 2007) provide alternatives to the highly formalised approaches found in both generative and functionalist linguistics. The interactional approach proposes a refocusing of linguistic work onto language-in-use in conversational and interactional contexts, sometimes including laughter and gesture in their analysis. The dialogic approach, on the other hand, presents a more radical philosophical shift, arguing that 'knowledge is largely communicatively constructed in and through its sociohistorical genesis' (Linell, 2007:8).

we must include a much wider variety of elements into our conception of grammar. She shows that, while these utterance-oriented items do not take part in the phasal derivation and are, in some sense, 'hierarchically superior to the matrix clause' (Corr, 2022: 300) they nonetheless participate in grammatical structure and are 'formally connected to the 'host' sentence' (Corr, 2022: 300). We can see that such an analysis would be useful at points where we have struggled in this thesis, for instance in bridging the gap between dislocations and interjected particles illustrated in examples such as (89) and (90) in Chapter 5.

Corr (2022) goes further, however, even going so far as to argue for the inclusion of gestures (Corr, 2022: 5). This last inclusion resonates well with functionalist approaches in the interactional and dialogic tradition, where language-in-use is viewed as one part of embodied communication which includes elements of non-linguistic semiotics, such as laughter (see Ford and Fox, 2010). Indeed, other researchers, such as Colasanti (2023a,b) have begun to analyse gestural communication using a generative approach. In Corr's (2022) view then, the assumptions of traditional generative work, in which grammar operates as an independent 'mental organ' must be abandoned, replaced with a view that grammar structures all of our thinking and cognitive processes. Rather than the functionalist view that our general cognitive processes shape our linguistic processes, Corr (2022) argues that all our cognitive processes are fundamentally of the same type as linguistic processes. Corr (2022) defines this viewpoint as the 'un-Cartesian' hypothesis. Crucially, however, Corr (2022) does not find it necessary to abandon cartographic formalism or methodology in order to apply this approach, making it clear that we can expand the scope of our research as cartographic researchers and radically change our assumptions regarding language, without needing to radically change our methodological approach or our formalism.

Corr's (2022) work presents a fascinating departure from generative assumptions coupled with a maintenance of the formalism. In Corr's (2022) book, the focus remains on a cognitive and internalist approach to language, with social variation still largely undiscussed. Nonetheless, Corr's (2022) un-Cartesian approach does make room for a radically socially oriented approach. As she states, 'taken to its logical conclusion, the [un-Cartesian approach] predicts that the distinction

between language, the mind, and the world should be collapsed.' (Corr: 46).[34] If we go so far as to collapse not just language and the mind, but also the world, we have a natural explanation of the simultaneously external and internal nature of language, and it becomes clear that the social and biological sides of language cannot be meaningfully separated. Having established that there is precedent for utilising formalist tools alongside radically reframed assumptions, I proceed to a discussion of socially-oriented approaches to language and how they might inform our work.

8.3.2 Sociolinguistic Approaches

I begin with a discussion of non-historical work in sociolinguistics, before showing how these principles can be applied to historical data, including diachronic data.

8.3.2.1 Labov and Variable Rules

The work of William Labov is foundational to any discussion of sociolinguistics. His works from the 1960s and 70s laid the groundwork for variationist sociolinguistics. Fine-grained quantitative work was used to show the role which social categories play in language variation and changes in progress. In this approach, linguistic variables are shown to vary down to the finest-grained elements of social categorisation. This variationist approach is somewhat limited, however, as it often fails to engage with the reasons behind the social structures at play. As Carter (2013: 581) points out 'variationist sociolinguists interested in using language to illuminate questions about the social have from time to time lamented the lack of adequate social theory in which to ground their work.'

One particular concept is key to Labov's work: variable rules (e.g. Labov, 1969). This approach considers linguistic rules in themselves to be capable of variability in their application. Not only do rules vary but they are themselves variable. Interestingly, this approach was initially embraced by some generative linguists (e.g. Cedergren and Sankoff, 1974). My conjecture here, then, is that language consists of both strict, inflexible rules as well as rules which may be breached in some cases, producing gradient grammaticality or, in some cases, stylistic effects. Indeed, Fraser (1972: 4-5) even suggests 'a continuum of optionality, determined by the function of the rule in the language.' Indeed,

[34] The original quote is describing the Topological Mapping Theory, which Corr (2022: 43-44) states 'presents a framework for implementing the conceptual insights of un-Cartesian linguistics in our technical analyses of natural language grammar.'

we can even see that these variable rules could be mapped onto Roberts' (2019) parameter hierarchies, with nano-parameters, which represent variation at the level of individual lexical items, showing a potential means of formalising some variable rules. As we saw in Chapters 4 and 7, these nano-parameters, in the case of post-verbal subjects, also seem sensitive to sociostylistic variation, with post-verbal subjects surviving particularly in journalistic writing but not in colloquial French.

I begin here with Labov to show that, while strict approaches to rules are dominant in much of current generative linguistics, a more variable approach is possible. Sankoff and Labov (1979) show that these variable systems can still be formalised, thus, they do not necessarily represent a foundational issue for generative approaches, at least at the level of description. This sort of model can make concrete predictions and, indeed, we might expect there to be a marked difference between the areas in which variable and invariable rules apply. Indeed, this approach seems well-suited to modelling the variation we have found throughout this thesis, where change happens gradually, even in the language of single speakers or authors.

Labov's theories have been further developed for application to historical data and data on language change, before we discuss this, however, I turn to more radical sociological approaches in non-historical sociolinguistics.

8.3.2.2 Sociocultural Linguistics

In the years since Labov's seminal work sociolinguistics has developed in a number of directions beyond the variationist approach. As we have already seen, the variationist approach can be somewhat limited, as Carter notes:

> 'Eckert and McConnell-Ginet (1992: 484) describe sociolinguistics' indifference to theories of social construction, writing that 'With only a few exceptions, linguists have ignored recent work in social theory that might eventually deepen our understanding of the social dimensions of cognition. Even less attention has been paid to the social (including the linguistic) construction of gender categories: The notions of 'women' and 'men' are typically taken for granted in sociolinguistics.' (Carter, 2013: 581).

These issues have begun to be addressed in various works starting in the 1990s, perhaps most notably the collection *Queerly Phrased* (Livia and Hall, 1997). Scholars have begun to take on theoretical perspectives from sociology, anthropology and queer and feminist theory. These approaches have been variously labelled critical sociolinguistics, sociocultural linguistics and critical ethnographic sociolinguistics among other terms. This approach, in which language is understood as part of the social and cultural context in which it occurs, rather than in relation to it, not only shines further light on sociocultural questions but also develops our understanding of linguistic structure. Indeed, in recent years, these critical socioculturally oriented approaches have been incorporated into variationist data-driven linguistics in the work of sociophoneticians. Calder (2019, 2023) and Steele's (2019) works on the sociophonetics of /s/ in queer communities in the US make extensive use of queer, post-colonial and critical race theory in engaging the underlying sociocultural structures at play. Not only does this allow a more politically engaged analysis, but it also shines further light on the surface linguistic data and structures under scrutiny by allowing a more developed understanding of the types of identity-formation at play. In the view of these critical approaches then, we should always keep in mind the sociopolitical and cultural ramifications of our analyses.

One might question how relevant these sorts of approaches are to the historical linguist. However, it is precisely this engagement with complex sociocultural questions which allows us to make the best use of 'bad data' (Labov, 1972). By taking seriously the view that all linguistic data is valid, and truly engaging with the sociocultural forces underpinning it, we can glean far more information regarding linguistic structure itself than if we focus on certain types of data.[35]

[35] I also wish to acknowledge the wide variety of critiques of linguistics found outside of the field, several of which have fed into the critical sociolinguistics approach. In the continental philosophical tradition, linguistics has routinely been critiqued for its insistence on structural dichotomies such as *langue/parole*, competence/performance etc. To thinkers in this tradition, such as Deleuze and Guattari (1980) and Lecercle (1990, 2006), we should understand language as a socio-political construct. The primary function of language then, is embodied socio-political action (Deleuze & Guattari, 1980: 80, 105-106). This approach has some similarities to the linguistic thought discussed here, with Deleuze and Guattari providing extensive discussion of Labovian sociolinguistic theory (Deleuze & Guattari, 1980: 117-119). Linguistics has also long been criticized in the Natural Language Processing literature, where it is sometimes entirely dismissed as a field. Edelman (2019: 11) cites a supposedly 'famous engineer's quip' from Jelinek (1988) 'Whenever I fire a linguist, our system performance improves'. Edelman (2019) goes so far as to propose a conception of language as 'a set of tools for dynamically and occasionally strategically influencing the behaviour of others' (Edelman, 2019: 13). It seems, then, that thinkers in both the empirically oriented and critical theoretical literatures have argued for a radical

8.3.3 Interim Summary

In summary, then, generative assumptions and their ramifications for historical syntax may prove problematic in light of fine-grained sociohistorical data. However, we have seen that approaches which move beyond these assumptions (e.g. Corr, 2022) need not abandon the methodological formalism used nor the insights of these approaches. At the fundamental level, I argue that language appears to be inherently variable to a much higher degree than the generative tradition would admit, with Labov's stable variable rules showing this most clearly.

We can view the cartographic approach as a formalist attempt to understand and model this complexity and variation. While cartography still uses discontinuous categories in the form of heads, phrases and fields along the clausal spine, the number of such categories is extremely high, such that it effectively becomes a model for continuous variation, mapping out the limits of such variation in typological terms in recent work such as that of Cinque (2023). This is precisely the reason for cartography's success: its focus on fine-grained description and a fundamental understanding that a vast number of categories appear to exist cross-linguistically. As cartographers have pointed out, cartography is not a model per se, but an object of study: the mapping out of linear order (Cinque and Rizzi, 2008: 42). We can thus continue to study this in much the same way, while rejecting the problematic assumptions of strict generativists.

8.4 Historicising the Social Approach

While our critiques of generativism have taken into account its diachronic predictions, we have thus far only discussed synchronic alternatives, focusing on the cognitive nature of grammar in Corr's (2022) work and the social dimension of language use in Labov's (1969, 1972) work. In this section, then, I discuss how to apply this socially grounded approach to historical linguistic work and why one should do so. Firstly, however, I discuss existing approaches to historical linguistics which move beyond generative assumptions. I begin with a discussion of historical pragmatics and functionalist approaches, critiquing these, before turning to historical sociolinguistics.

---

reframing of the linguistic object of study. Although I do not believe such a radical reorientation is necessary, I believe linguists should be aware of these critiques.

8.4.1 Historical Pragmatics

I begin by discussing historical pragmatics, which has folded functional, interactional and dialogic approaches to language into approaches to language change. Since the 1990s, pragmatics has become an increasing focus of historical linguists, particularly in the wake of Andreas Jucker's (1995) work and developments in functional historical linguistics with grammaticalization theory (Heine, Claudi and Hünnemeyer, 1991; Hopper and Traugott, 1993). These approaches have been applied extremely fruitfully in French historical linguistics, particularly in the work of Hansen (2005, 2014) on pragmatic markers and Hansen and Visconti (2014) on negation. We have been using some elements of the historical pragmatics approach throughout this thesis. For instance, our use of letters, SMS messages and spoken data is in-line with, particularly, the more recent historical pragmatic turn towards interactional and dialogic historical approaches (see Traugott, 2010).

A constructional approach to historical pragmatics has recently been developed (Traugott and Trousdale, 2013). Börjars, Vincent and Walkden (2015) provide an in-depth critique of Traugott and Trousdale's (2013) work. In their review they note a number of issues with the approach. One repeated criticism of Traugott and Trousdale's (2013) work can be seen in the variety of new terminology introduced but imprecisely used, for instance 'neoanalyzed', rather than traditional 'renalayzed', 'analogization' rather than traditional 'analogy' etc. Börjars et al. (2015: 365-6) provide a critique of one of Traugott and Trousdale's (2013) specific case studies, showing how their imprecise use of syntactic terms such as head and phrase leads to an imprecise analysis. Indeed, this case study shows the ways in which functionalists can tend to restate descriptive facts in unnecessary formalism. Börjars et al. point out:

> 'It is important to stress in this context that we are not saying that the only acceptable theories are formal ones, but simply that formalism aids precision just as argued more than half a century ago by Chomsky in the preface to Syntactic Structures. We agree with Pollard and Sag (1994: 7) when they write: 'this does not mean that the empirical hypotheses must be rendered in a formal logic as long as their content can be made clear and unambiguous in

> natural language'. However, what we miss in this book are those 'clear and unambiguous' definitions of terms used.' (Börjars et al. 2015: 380).

The key issue is that these approaches often deploy novel terminology in a way which is not conducive to rigorous and accurate description or analysis. Indeed, historical pragmatics, taken on its own, does not help us move beyond the criticisms laid out above, but simply represents a sub-field of historical linguistics with a particular object of study. Historical pragmatics could even be seen as one of the dominant paradigms in recent generative historical syntax, with considerable focus in recent years on the role of the left-periphery and discourse pragmatic functions in driving syntactic change. Thus, while historical pragmatics' contributions to recent historical linguistics should not be understated, we should note that it does not represent a meaningful alternative to generative historical syntax, except in acknowledging the possibility of grammatical change in adult language users. Indeed, work such as that of Traugott and Trousdale (2013) may function as a cautionary tale; it is crucial that any new theoretical concepts justify themselves and, just as formalism should not be used for its own sake, we should not abandon formalism for the sake of abandoning formalism.

I now go on to discuss the work of Labov and historical sociolinguists to show how we may expand our historical analysis.

8.4.2 Labov's Theories of Change

As discussed above, variable rules are a central element of Labov's work. Variable rules are key to both his synchronic description of language use and his views on language change. As Labov (1994: 579-583) shows, variable rules are readily acquired by children entering into the speaker community: they are not reinterpreted as invariable rules. As such, Labov links his synchronic and diachronic accounts of language arguing that, with the stability of variable rules over multiple generations 'it is then easier to understand that much of synchronic variation is a residue of historical processes, rather than the immediate product of linguistic or physiological principles.' (Labov, 1994: 583). As Labov points out, however, the stability of such variable systems means that 'the problem of accounting for linguistic change becomes more difficult' (Labov, 1994: 584). For Labov, linguistic change, particularly phonological linguistic change, is driven by processes of misunderstanding which lead to gradual

systemic shifts. However, crucially 'to the extent that general linguistic principles apply, they form the favorable undercurrent, or perhaps prevailing wind, for changes now in progress. Given enough social motivation or contrary linguistic pressures, retrograde movements can be set in motion, just as a boat may tack into the wind' (Labov, 2001: 499).

Regarding the initiation of change, Labov argues that misunderstandings drive phonological change. With a sufficiently fine-grained syntactic description, it seems clear that similar misunderstandings, for instance with the position of the post-verbal subject, would be readily possible, since fully unambiguous clauses are rare. Indeed, in Chapter 4, we saw that Middle and Early Modern French low subjects and Ibero-Romance low subjects were distinguished by data as fine-grained as the position of object quantifiers.

In summary, for Labov, 'the evolutionary and historical perspective […] demands an understanding of human behavior in its social context' (Labov, 1994: 598-9). This sort of approach emerges immediately from our philosophical assumptions if we reject generative assumptions and take on the view, as argued above, that language can be seen as both social and biological, with neither factor having primacy. This is precisely where the functionalist approach often falls down, since, rather than taking into account all factors in linguistic variation and change, the functional factors are argued to have primacy. I hope that our approach has revealed itself to be highly powerful throughout this thesis. Without the use of varied genres, registers and spoken data, we would not have revealed the lag in the loss of post-verbal subjects in Early Modern French texts, we would not have identified the spoken-written divide in dislocations and we would not have been able to describe the interactional functions at play in the distribution of dislocations and clefts. It is crucial to take into account a highly varied dataset and a wide variety of potential functions in order to truly describe linguistic variation and change.

8.4.3 Historical Sociolinguistics

In addition to Labov's seminal work on sociohistorical approaches to linguistic change, the historical sociolinguistic approach has been widely applied in the last 30 years. Pioneered by Romaine (1982), the field has since developed considerably. In French historical linguistics, considerable work has been

done, most notably, Ayres-Bennett's (2004) study of 17th century French and recent text-typological work on Old French (Simonenko, Crabbé and Prévost, 2018; Larrivée, 2022b). Outside of French, the field has seen particularly large developments since the publication of new corpora of letters and texts from lesser educated authors, such as the Letters as Loot corpus (Rutten and van der Wal, 2014).

However, much work in historical sociolinguistics has neglected to adopt any of the approaches found in sociocultural linguistics and critical sociolinguistics, fields discussed above in 8.3.2. Typically, variationist sociolinguistics dominates, especially focused on text-typological factors. We will not discuss these approaches in-depth here. Instead, I turn to those occasional moves which have been made to go beyond variationism in historical sociolinguistics. Most notably, Elspaß (2007) develops the notion of language history from below. This approach emphasises a focus on the under-studied language of lesser-educated writers and writers from under-represented backgrounds. Often, the use of hand-written texts is emphasised, texts 'below the surface of the printed language' (Elspaß, 2007: 4). Crucially, for Elspaß, this focus on under-studied works is not simply an extension of existing historical linguistics work, nor an interesting corollary, but is essential to forming a full image of a language's history. For Elspaß, traditional language histories from above are inherently flawed and 'the inevitable consequence of these practices for language historiography are incomplete language histories full of "blank areas"' (Elspaß, 2007: 4).

Indeed, we can find resonances between Elspaß's work and that of radical critics of linguistics such as Lecercle (2006). Both Elspaß and Lecercle provide a strong critique of the way standardised language norms are reflected in linguistic approaches. As Lecercle states:

> 'It is not enough to decide that the 'language' (a given language) is a set of heterogeneous phenomena in a state of continuous variation. We also need to explain why linguists fetishise this heterogeneity into a homogeneous, or, rather, homogenised, system. The answer is simple: there is a homology between the scientific model of the language system and the political model which makes the centralised national language a vector of power.' (2006: 135),

Compare Elspaß's 'language change appeared to be of interest only when it contributed to the linguistic architecture of a modern standard language' (2007: 4) and Elspaß's wider critiques of

'standard language ideology'. To take this sociolinguistic approach further then, we must more meaningfully adopt Labov's notion of variable rules, fully accepting the variability that is found in our data.

Standard language ideology has also been the target of significant criticism in French linguistics. Zribi-Hertz (2006) and Massot (2008) have argued that some sort of diglossic system obtains in contemporary France, masked by the dominance of standard language ideology and prescriptive norms in the country. However, this very argument itself is born out of a desire to construe variation as occurring between discrete grammars. Indeed, Coveney argues that the examples of register variation are 'for the most part, not categorical constraints, applying in 100% of possible cases, but are rather variable (probabilistic) linguistic constraints – or in some cases simply co-occurrence restrictions' (Coveney, 2011: 74, quoted in Hornsby, 2019: 579-580). In more recent works by Rowlett (2013) and Zribi-Hertz (2013), it has been argued that we should view the situation as best-described by a single grammar, either with a 'grammatical bolt-on' (Rowlett, 2013) or with a learnt academic norm (Zribi-Hertz, 2013) which has no internalised grammar, what we might call a 'pseudo-grammar'. This view is in-line with our findings in this thesis. It has been consistently clear that variability between text-types is rarely a matter of clearly distinct grammars. For instance, while dislocations are far more frequent in informal speech than in more formal written French, they are still attested in both (See Chapter 5 for discussion). Likewise, although post-verbal subjects are associated with more formal contexts, they are nonetheless still found in the informal language of internet fora (See Chapters 4 and 7 for discussion). It is important to expand our understanding of the sociolinguistic situation in the contemporary language by exploring the history of this situation, and engaging critically with the textual data which is at the core of historical linguistics.

As discussed above, the precise nature of the linguistic variation we have found, be it micro- or nano-parametrically mediated, or a case of variable rules in the Labovian sense, is unclear. Nonetheless, standard language ideology restricts our thinking about linguistic variation, enforcing a view in which strictly delineated forms of language are presupposed to be the norm, with little room

in our analysis for variable rules or the possibility of learnt, sociostylistically delineated pseudo-grammars.

8.4.3.1 Sociohistorical Issues in Our Case Studies

We have seen then, that Labov's approach to language change insists on a complementary understanding of both external and internal factors. As I have argued, in-line with Corr's (2022) un-Cartesian hypothesis, these two elements of language change are not only complementary, but in fact inseparable. In our own data, discussed in this thesis, we have repetitively seen the role of so-called external factors in language change. Here, I briefly summarise the roles played by sociopolitical factors in our three case studies, focusing primarily on post-verbal subjects and dislocations.

With regard to post-verbal subjects, we found that the establishment of the SVO word order was a natural progression from the Old and Middle French word order to that of Modern French. That said, we also saw in our 17th-century texts, that texts of less highly educated writers utilised considerably more post-verbal subjects than printed texts of highly educated authors. As discussed in Chapter 7, we do not have sufficient evidence from prescriptive texts to suggest a specific role for prescriptivism in this process, but it does seem that some sociopolitically informed variation was occurring at this time. Further investigation is required to clarify the exact role of prescriptive ideology in this situation, however. In Modern and Contemporary French, post-verbal subjects are highly sensitive to sociostylistic distinction, with Lahousse (2022) discussing the survival of post-verbal subjects in journalistic writing, despite their disappearance in spoken French.

With regard to dislocations, we found that they are minimally discussed by grammarians, with almost no mentions in the prescriptive works. This highlights an issue in applying traditional historical sociolinguistic methodologies to our work. While phonological and lexical variation are highly discussed by prescriptive grammarians, and vary widely depending on sociopolitical variables, syntactic variation appears to be subject to less explicit manipulation by speakers and writers. Nonetheless, this variation is still highly affected by social, political and interactional factors. In particular, modality has a huge effect on the frequency of dislocations. We have argued that the much higher frequency of dislocations in speech is due to the number of further functions available to dislocations, especially in

turn-taking. Thus, the physical nature of each modality has an effect in these cases, with dislocations able to play a role in turn-taking in speech due to the temporal nature of speech, while in writing they may only play their role in information-structural marking. We saw in Chapter 5 and Chapter 7 that this may present a considerable issue for historical syntax in particular as well as historical linguistics more broadly, with spoken language more inaccessible than previously thought, differing considerably from even the most spoken-like texts.

In the example of clefts, we see that text-typological, social and political factors seem to play quite a minimal role. Some functions of highly routinised and idiomatised clefts appear to relate to turn-taking and other interactional pressures. However, this does not seem to have any bearing on the overall distribution of clefts.

As we can see then, it has proved more difficult than anticipated to analyse and detect sociopolitical roles in syntactic change. Nonetheless, these factors are, in some way, at play in all changes, and although they may not be the driving force, totally ignoring them or excluding them from analysis at best puts us at risk of an over-narrow analysis and at worst could lead to totally incorrect analyses of the data. The example of dislocations is particularly pertinent, since it shows not just a relation between frequency of dislocations and modality in the strictly variationist sense, but instead shows how the very nature of modality defines the potential set of functions a structure can take on, showing how these features cannot be ignored.

8.4.4 Towards a Theory of Change

Here then, I wish to argue for an approach to linguistic change synthesising the fine-grained work of Labov, the theoretical radicalism of Elspaß (2007), critical sociolinguists and Corr (2022) and the formal and descriptive rigor of generative and particularly cartographic syntax.

Throughout this chapter we have identified the merits of the Labovian view of variability in language, i.e. that language is by its nature variable, and that variable rules govern this variation. Therefore, the contention of this thesis is that these variable rules apply within the minds of individual speakers rather than being epiphenomena from data drawn from across speaker

communities. This is demonstrated by the variability found within individual texts discussed in this thesis (see, particularly, variable rates of post-verbal subjects and the available subject positions in the Middle French period, discussed in Chapter 4 and section 8.2). Nonetheless, the generative view of parametric change is useful, in particular Roberts' (2019) parameter hierarchies may be well-suited to formalising Labov's work. It seems likely that some degree of parameter setting, particularly at the macro-end of the parameter scale, occurs permanently in child-language acquisition while micro- and nano- parameters may vary throughout a speaker's life and exhibit sociostylistic sensitivity.[36]

This latter point relates to notions of obligatoriness and optionality. The emphasis on obligatoriness in generative work, while often advantageous as a research heuristic, can create more issues than it solves. In recent work, Cinque (2023: 102-103) has suggested that base orders should be modelled through some form of meaningless movement. We can then model optional variation in base orders as openness to a variety of types of meaningless movement in a given language.

Turning to the distinction between written and spoken language, we should be open to the varied ways in which different modalities shape language use, not just in the variationist sense but in the sense of the communicative pressures at play in different modalities. The physical realities of speech allow for extra functions such as turn-taking markers; the way in which external pressures shape our language use is central to the approach I am proposing here. Crucially, however, this does not represent an abandonment of the formalism I have been using but rather, an understanding that structures within this formalism may be appropriated for use in new and interesting ways, as with dislocations being used as turn-taking markers or with clefts becoming routinised and lexicalised structures. Recent cartographic works have begun to attempt to formalise these utterance-oriented structures (see Speas and Tenny 2003, Corr, 2022; Miyagawa, 2022).

Thus far, I have simply argued for a combination of formalist approaches with sociolinguistic and variationist work. However, it is at this stage where I begin to take a more radical turn. Both the formalist and functionalist traditions typically understand language as being primarily a cognitive

[36] Biberauer and Roberts (2016: 261) note a study from Tsimpli (2014) in which macro-parameters are shown to be stable even in bilingual and language contact contexts. Given the stability of parameters at the macro-end of the scale, it seems likely that micro- and nano-parameters would be less stable, even so unstable as to be variable after child acquisition.

object with some biological basis (be it as an innate 'organ' as in the case of Universal Grammar, or as a result of general human neurological and psychological capabilities). While I do not deny that language has a biological and neurological basis, I argue here that centring the biological basis of language as more fundamental than its social roles has been primarily harmful to linguistic research. As discussed throughout this chapter, sociostylistic factors cannot be seen as mere externalities. Although Labov still supports the notion of an internal-external divide, his research program broadly supports this view, focusing on the social processes of language use which lead to language variation and change.

Crucially, however, biological, neurological and other 'internal' factors are also still at play. The issue then, from both approaches, is prioritising the internal or external to the total exclusion of the other.[37]

Importantly, this approach is meaningfully useful for research. In my work on dislocations, examining modality variation with an open mind to the different functions we may find has led us to understand the distinction between written and spoken frequencies of dislocations and a better understanding of the historical written record. In Nevalainen and Raumolin-Brunberg's (2003) work on 3rd person singular endings in English, taking sociopolitical and historical data into account allows us to account for the seemingly inexplicable changes in the frequency of different endings. Several works in generative syntax have fruitfully taken into account social and political variation in the syntactic sphere (see Emonds, 1986), and we are now, in the last 10 years, seeing the birth of a new field of sociosyntax (see Christensen and Jensen, 2022). Work in critical sociophonetics such as that of Calder (2019, 2023) and Steele (2019), discussed above, shows us how extremely fine-grained distinctions in linguistic data can belie complex sociopolitical issues. Crucially, though, it is not just socially oriented work which can inform narrow linguistic work, but narrow linguistic work can also inform more traditionally socially oriented work. For instance, Loporcaro (2003) shows how historical phonological data from the early Romance languages allows us to provide assistance in the dating of

[37] Lecercle (1990) has argued for the concept of the 'remainder': however we cut up the object of study leaves out a 'remainder' which could be crucial to future analysis. Instead, then, we must take an expansive and inclusive approach.

various sociohistorical shifts in Western Europe in the wake of the Roman Empire. Recent work by Bach and Esher (2023) applies a fine-grained morphological lens to the development of new systems of gender reference, arguing that typological fit is important in the acceptance of these systems in the wider speaker community.

As we saw above, Corr's (2022) work provides a basis for further exploration beyond the theoretical assumptions of generativism. I argue then, that we should follow this approach and extend it, applying the formal rigor of cartographic work to a wide variety of historical and social data, incorporating elements of language typically excluded from generative work. If we follow Corr's (2022) un-Cartesian hypothesis and the topological mapping theory, collapsing the distinctions between language, mind and world, it becomes clear that social factors cannot and should not be distinguished from psychological and formal factors in analysing language use and language change. In practice, this approach consists in analysing varied data sources, taking into account the social and functional pressures at play in those sources and examining metalinguistic sources to gain a deeper understanding of social, stylistic and political pressures at play. In future, following this approach to its logical conclusion would also entail formalising interactional and social levels of variation within the cartographic framework. It would have been unthinkable in the 1980s to suggest that information structure would one day be the purview of generative syntacticians, and yet we now have a highly developed system of information-structural analysis. Likewise, while incorporating sociological analysis into generative formalism may seem an outlandish suggestion, I argue that it is crucial for further developing our analyses.[38]

8.5 Summary

This chapter represents a variety of critical approaches to generative syntax. I hope to not have fallen where many previous critics of generativism have: throwing out the baby with the bathwater. As we saw in our discussion of historical pragmatics, we often find researchers reinventing the wheel,

[38] I am indebted to an in-person conversation with my supervisor for this comparison between information-structural and social factors.

labelling already known processes with novel names in order to distinguish their work from that of their theoretical opponents.

The generative syntactic research program has proven immensely successful in developing fine-grained analyses of the internal syntactic structure of the clause. Even within this rather critical thesis we have had great success in applying its mechanical formalism to our objects of study, particularly post-verbal subjects. As such, before concluding, I would like to note what this critical turn is not. This is not an argument for an abandonment of the generative project, nor is this an argument for fully abandoning our assumptions, at least as research tools. Although I have argued that the assumptions laid out in section 8.2 are not materially, ontologically true, untrue assumptions can act as an extremely useful research heuristic. The Neo-Grammarian hypothesis exemplifies this well: although it has long been known that phonological change does, at least on occasion, spread through the lexicon and speaker community gradually (see discussion in Labov 1994: chapter 16), it is only by maintaining the assumption of the Neo-Grammarian hypothesis that the astonishing results of the Comparative Method were achieved. Incorrect assumptions need not be harmful as long as we are aware that they constitute heuristic tools, not ontological truths.

This chapter does not argue to abandon all biological and neurological factors, simply to be more realistic about their roles. My contention here is that social and political factors, alongside a huge variety of other pressures (including the biological and psychological) all intermingle to produce a given linguistic utterance. Formal analyses can still be brought to bear on this, as shown by Labov's formalisation of variable rules. I have argued that Corr's (2022) recent work moving beyond the assumptions of generativism while maintaining its formalism represents fertile ground for further work, incorporating even more factors into our formalism. Although I have not developed a specific formalisation here, the literature in the fields of sociosyntax and historical sociolinguistics seem good places to start, using Corr's (2022) approach to formalism as a guide. If we actively incorporate a much wider variety of data, we will produce more complete linguistic analyses. Likewise, taking varied approaches to linguistic analysis, including incorporating sociopolitical factors, beyond treating them as rarely relevant 'external' factors, will also allow us to deepen our analysis. I end with a quote from

Weinreich, Labov and Herzog: 'Linguistic and social factors are closely interrelated in the development of language change. Explanations which are confined to one or the other aspect, no matter how well constructed, will fail to account for the rich body of regularities that can be observed in empirical studies of language behavior.' (1968: 188).

9. Conclusion

9.1 Summary

This thesis aimed to determine the role of social and information-structural factors in syntactic change in French, focusing on three case studies through the history of the language. Throughout, we have found that the role of these factors is less direct and less clear than anticipated. Rather than social and information-structural factors directly driving change in transparent ways, these factors seem to play smaller roles in the development of the complex system of language. Nonetheless, by incorporating these under-discussed factors in linguistic change, we have uncovered a wide variety of perspectives previously undiscussed in the literature and, as a result, achieved a more detailed description of French syntax both historically and in contemporary varieties.

Throughout this thesis we have also tested the power of the cartographic framework for analysing syntax, syntactic change and incorporating information-structural and social factors. We have found that the cartographic approach represents an extraordinarily strong descriptive framework for analysing syntactic variation and change. Nonetheless, its focus on the internal structure of the clause limits its capacity to analyse the role of syntax in language-in-use, as in the interactional functions of dislocations and its capacity to analyse syntactically opaque, routinised structures, such as clefts. We have seen that an expansion of cartography to include a wider set of interactional and social factors, as argued for by Corr (2022), may resolve some of these issues. Additionally, further explorations of sociosyntax through the cartographic framework are crucial to developing a fully-fledged understanding of syntactic change.

In Chapter 4, we discussed the history of post-verbal subjects in French. By analysing a wide variety of data, including texts from less highly educated writers and unedited texts, we found that a previously undiscussed low subject position was available in Old, Middle and Early Modern French. We found that the cartographic framework was extremely well-suited to analysing these positions, both in providing the requisite formalism to describe them formally and in encouraging the methodology of fine-grained analysis, considering relative positions of objects and adverbs, which led

us to these findings. Low subject positions become highly restricted in Early Modern French, but we have left the mechanism by which VSO orders were lost open to future research. The changes in low subject positions appear to be partially conditioned by the processes of prescriptivism at play at the time. However, the evidence in this thesis is too sparse to make specific claims about the role of prescriptivism. In addition, in Modern and Contemporary French, we found, in-line with Lahousse (2022), that post-verbal subjects are sensitive to some sociostylistic factors.

In Chapter 5, we discussed the history of dislocations in French. We found dislocation structures to be astonishingly stable, with no apparent structural changes through the history of French. As such, the cartographic framework provides a clear formal analysis of the construction which seems to capture the relation of the dislocated constituent to the clause. The most crucial finding of this chapter, however, was the wide gap in the frequency of dislocations between Contemporary Spoken French and written Contemporary French, suggesting that the linguistic modality directly affects the frequency and function of dislocations. Dislocations display distinct functions related to speech as a modality which cannot be exhibited in writing. If, as this suggests, historical texts cannot reliably be used to access information regarding the spoken language, historical linguists need to be more cautious regarding the reliability of their findings. Since we have only used a narrow set of so-called 'spoken-like' texts, however, further work is required to explore the exact relation between written and spoken language in historical linguistics. In particular, theatrical texts and direct representations of speech merit further exploration, while we have shown that letters and particularly ego-documents do not seem to be as 'spoken-like' as typically thought. Nonetheless, the findings of this chapter demonstrate the usefulness of including Contemporary sociostylistic data into our analyses of historical data.

In Chapter 6, we discussed the history of clefts in French. We found, contra Dufter (2008, 2009), that clefts increase in frequency as focus fronting decreases in frequency during the Early Modern French period. Regarding their syntax, we found that clefts were astonishingly difficult to analyse under cartography, with several pieces of complex and seemingly contradictory evidence causing issues for many existing analyses. Additionally, much of the current cartographic work on

clefts assumes the primacy of their focusing functions (Belletti, 2012, 2015; Bonan and Ledgeway, 2023), while we show that informative-presuppositional clefts are frequently found not only in Contemporary French but also throughout the history of French. The wide variety of surface-similar but functionally divergent structures is a central difficulty to the analysis of clefts, particularly when attempting to account for the similarities and differences between adverbial and DP clefts. We saw that Contemporary French clefts have, in some cases, lexicalised, or at least become highly routinised, developing particular interactional functions. I presented a potential outline for future analyses under cartography, while also suggesting that work under different theoretical frameworks may be required to truly understand clefts as a construction.

In Chapter 7, I provide an overview of the foregoing three case studies, contextualising them into a summary of the existing literature on the roles of information structure and sociolinguistics in French historical syntax. Two key findings emerged from this wide-spanning overview. Firstly, linguistic change and variation is a chaotic process, with remnants surviving for centuries after a construction becomes extinct as a generalised part of the syntax. Crucially, during periods of change, we find variability in the works of single authors, rather than clear distinctions between different grammars. It remains unclear whether this variability can be explained through micro- and nano-parametric variation or if a variable rule or mixed grammar approach would be needed. Secondly, we found that social and information-structural factors did not play a dominant role in syntactic change. Rather, they appear to be two of many probabilistic factors which determine which changes do or do not occur. A clear instance of this can be found in the history of clefts; contra Dufter (2008, 2009) clefts appear to increase in frequency when other focusing strategies are lost, however, they do not increase overwhelmingly in frequency, 'nor are they in complementary distribution with these constructions. As such, it seems that information-structural pressures provide some explanation for some changes in historical syntax, but they are not a driving force. Indeed, this second finding relates directly to the first, to quote Labov yet again, as quoted in Chapter 8 'to the extent that general linguistic principles apply, they form the favorable undercurrent, or perhaps prevailing wind, for

changes now in progress. Given enough social motivation or contrary linguistic pressures, retrograde movements can be set in motion, just as a boat may tack into the wind' (Labov, 2001: 499).

Finally, in Chapter 8, I considered the limitations of cartography and generative syntax more broadly, particularly as applied to historical linguistics. I discussed traditional generative approaches to historical syntax, showing that strict generative assumptions limit our inclusion of interactional and sociolinguistic data. I argued that a sociolinguistically informed approach, driven by Labov's notion of variable rules and formalised through Roberts' (2019) parametric hierarchy and Corr's (2022) radical new set of theoretical assumptions would allow us to incorporate sociological intuitions into a formal analysis of language. This then allows us to maintain the formal rigor of the generative and cartographic approaches whether or not the biolinguistic assumptions which traditionally underpin generative work hold true. We have also seen that several pieces of recent work, such as that of Corr (2022), have also been moving beyond the traditional assumptions of generative syntax. Although I have not provided the formalism for this new direction of cartographic work, I hope to contribute here to this emerging tradition.

9.2 Further Directions and Flaws

Throughout the writing of this thesis, many questions have been left unanswered. Firstly, the core data and interpretation of our three central case studies have raised a number of unanswered questions.

In our study of post-verbal subjects, several finer points of the analysis were left unclear. Firstly, the loss of VSO orders in the Early Modern period was left unaccounted for, although we have tentatively suggested that it may be linked to a change in the post-verbal DP subject position. Secondly, I have left the process of clitic incorporation as beyond the scope of our analysis of post-verbal pronominal subjects. If, as our analysis proposed, the pronominal moves to incorporate with the verb, we would expect to find a proclitic, not an enclitic. Nonetheless, we have argued that our account is necessary to describe all the relevant data. Due to the scale of this thesis, these gaps in our analysis cannot be further explored, but they do merit further research.

Through the study of dislocations, it became apparent that an exhaustive history of the structures would have to account for the development of interactional functions, particular to the spoken language. Given the difficulties we found with the written-spoken divide, it is unclear how to further develop historical work with the inclusion of interactional approaches. Since we have focused here on applying cartographic analyses, we have left this as beyond the scope of the thesis. However, recent work in interactional and dialogic historical linguistics has taken an approach focusing on analysing represented speech and represented dialogue in theatrical and prose texts (e.g. Paoli, 2020). We have not analysed these genres in our approach to spoken-like texts, and testing the differences between written speech and spoken language may be a further area for research on this topic. Regardless, it would considerably advance historical syntax to develop a methodological approach which incorporates interactional analyses. As discussed in Chapter 8, some cartographic work has begun to incorporate interactional functions, as in the work of Speas and Tenny (2003), Corr (2022) and Miyagawa (2022). However, none of these works apply this approach to historical data. Meanwhile, in historical syntax, many of the current approaches fail to fully account for the extent of the spoken-written divide we have found in our work. It is crucial that we take seriously the gap between written and spoken language and develop methodologies to overcome it.

While our work on dislocations raised questions regarding the inclusion of interactional analysis in historical work, our work on clefts left us without a clear syntactic analysis of the construction. A construction-grammar oriented approach may be useful to analyse clefts, particularly given the routinisation and lexicalisation processes which appear to be at play in Contemporary Spoken French. Unfortunately, such approaches are outside the scope of this thesis. However, it is clear from both dislocations and clefts that an approach to historical syntax which is theory-neutral and allows for the inclusion of varied theoretical approaches may be useful in making progress on these hard problems.

Finally, throughout the writing of this thesis, it has been clear to me that further cross-linguistic data would be useful to expand our analysis. This is particularly clear in the analysis of subject positions in Chapter 4, where data from low subject positions in Ibero-Romance helped to

develop the analysis of such positions in Early Modern French. Further incorporation of this cross-linguistic work, both within and beyond Romance, would also assist in the analysis of dislocations and particularly clefts. Recent work on clefts has been highly focused on cross-linguistic comparison and variation (Bonan and Ledgeway, 2023), incorporating these data directly into the studies presented here would likely help to develop our understanding of the situation in French. We can see how effective this can be in the work of Wolfe (2018) where the various V2 systems of Medieval Romance languages are compared to great effect.

In addition to these core questions regarding the methodologies and case studies of this thesis, further methodological questions have presented themselves through the process of writing. These were discussed to some degree in Chapter 8, where we critiqued the assumptions of generative historical linguistics and argued for a historical sociolinguistic approach. However, through this thesis we have seen that there are considerable difficulties in applying such an approach, particularly in the case of historical syntax. While variation between genres and registers is found, the structure of this variation is often chaotic and opaque and, as we saw in Chapter 5, the spoken-written divide exists independent of the notion of 'spoken-like' texts. Through analysing the works of prescriptive grammarians of the 16th and 17th centuries we gained some further insights. However, this approach is considerably more effective with lexical and phonological variation. The grammarians rarely spoke directly on issues of word order and syntactic variation and, in the cases where they did, they typically discuss specific word order patterns found in highly idiomatised structures. Thus, general historical sociolinguistic techniques are not as easily applicable to historical sociosyntax work. Unfortunately, the scale of the study undertaken here has limited our findings on particular historical periods. A more fine-grained examination of genre variation in the Early Modern period may yield clearer results on socially conditioned variation and allow us to begin overcoming the issues of the spoken-written divide. Developing a clear methodology for historical sociosyntax work is an important area for further research in this direction. Likewise, applying the methodologies here at a smaller scale with more rigor is likely to achieve better results. The ambitious scale of this thesis has meant that some details which may otherwise be identified have been left out.

In addition to the methodological question of how to conduct historical sociosyntax work, a theoretical question has been raised with regard to how to approach such work. In Chapter 8 we argued for an approach which abandons the strict generative assumptions common in much of the historical syntactic literature, instead refocusing on the inherently social nature of language, all the while maintaining the highly descriptive formalism of the cartographic framework. We specifically argued to take Corr's (2022) un-Cartesian hypothesis as a starting point. Corr (2022) reaches similar conclusions to ours regarding the generative approach to syntax, continuing to use its formalism while abandoning its core assumptions regarding the independence of the language faculty. Corr's (2022) work certainly presents a hopeful new direction for syntactic research, moving beyond the limitations of internal clause-structure and generative assumptions, including the ambitious suggestion that gestural analysis could be included in linguistic description. Nonetheless Corr's (2022) work is still focused on the internal, psychological elements of language. Throughout this thesis, however, we have seen that social and external factors play major roles in language change. Thus, as argued in Chapter 8, we should take Corr's (2022) approach further, including not just interactional and gestural features of language but also sociostylistic elements of variation. We can see that sociolinguistic data is already included in some syntax work in the emerging field of sociosyntax, in work like that of Adger and Smith (2020).

Finally, we have also seen the importance of challenging prescriptivism and standard language ideology, with new findings emerging from the varied data we have examined. In my view, we must challenge standard language ideology and prescriptivism at much deeper levels however, going beyond simply tinkering our methodology to include under-examined data. As discussed in Chapter 8, both Elspaß (2007) and Lecercle (2006) have suggested that prescriptive ideologies and standard language ideologies do not just affect language use but have been central to the development of modern linguistics. While linguists now of course reject overtly prescriptive analyses, a variety of assumptions allow us to exclude certain natural language data as being inadequate or inaccurate. This occurs both on the micro-level, where some examples of natural language use are rejected as poor evidence due to representing 'performance' not 'competence' as well as on a more overtly political level, with the

concept of the native speaker[39] allowing us to dismiss natural language data from speakers who do not acquire a language in childhood, a concept which emerges both from the older norms of standard language ideology as well as the assumptions of generative linguistics. To truly move forward, we must be even more radically inclusive in our use of data, and truly understand all linguistic data as valid natural language data. Just as using varied historical data has allowed us to overcome some of the issues of historical linguistics, 'researchers have begun to make a virtue of so-called 'bad data' by showing the wealth of insights that can be gleaned from data that are often discarded as 'inauthentic" (Bucholtz and Hall, 2008: 411).

I hope that, as the field of historical sociolinguistics develops, it develops more rigorous methodologies while also taking into consideration more critical approaches to social conceptions of language. We have seen throughout this thesis the importance of taking into account fine-grained data, both in formal and social terms. If we are to continue progressing in our understanding of language change, we must continue to embrace new data, new methodologies and new theoretical assumptions. This thesis has made modest attempts to move beyond the traditional limitations of historical syntax work, including using a variety of texts outside of the usual datasets of historical syntacticians, while also embracing atypical methodologies and analytical frames from historical sociolinguistics and interactional linguistics. Simultaneously, we have applied the cartographic framework as rigorously as we can with our comparatively limited data, examining variation at a very fine-grained level and discovering a variety of new syntactic positions as a result, particularly with regard to low subject positions. This combination of fine-grained empirical rigor with an expansive methodological and theoretical approach seems to be crucial to truly moving forward in historical linguistics.

[39] For a broad collection of critical discussion of the notion of native speaker, see Slavkov, Melo-Pfeifer and Kerschhofer-Puhalo (2022)

Bibliography

Adams, M. (1987). From Old French to the theory of pro-drop. *Natural Language and Linguistic Theory* 4. pp. 1–32.

Adger, D. and Smith, J. (2020) Explaining variability in negative concord: A sociosyntactic analysis. In: Beaman, K. V., Buchstaller, I., Fox, S. and Walker, J. A. (eds.) *Advancing Socio-grammatical Variation and Change: In Honour of Jenny Cheshire*. London: Routledge

Angelopoulos, N., & Sportiche, D. (2021). Clitic dislocations and clitics in French and Greek. *Natural Language & Linguistic Theory*, 39(4), pp. 959–1022. https://doi.org/10.1007/s11049-020-09500-z

Ashby, W. J. (1982). The drift of french syntax. *Lingua*, 57(1), pp. 29–46. https://doi.org/10.1016/0024-3841(82)90069-9x

Ashby, W. J. (1988). The syntax, pragmatics, and sociolinguistics of left- and right-dislocations in French. *Lingua*, 75(2), pp. 203–229. https://doi.org/10.1016/0024-3841(88)90032-0

Ashby, W. J. (1994). An acoustic profile of right-dislocations in French. *Journal of French Language Studies*, 4(2), pp. 127–145. https://doi.org/10.1017/S0959269500002192

Auger, J. (2003). Le redoublement des sujets en picard. *Journal of French Language Studies*, 13(3), pp. 381–404. https://doi.org/10.1017/S0959269503001200

Authier, J. M., & Haegeman, L. (2019). The syntax of mirative focus fronting: Evidence from French. In Arteaga, D. L. (ed.) *Contributions of Romance languages to current linguistic theory*. Berlin: Springer. pp. 39-63

Ayres-Bennett, W. (1996) *A History of the French Language through Texts*. London: Routledge

Ayres-Bennett, W. (2004). *Sociolinguistic Variation in Seventeenth-Century France: Methodology and Case Studies*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486630

Bach, X. & Esher, L. (2023). Slotting non-binary determiners into the French inflectional system. Presentation given at *Challenging the Binary Non-binary, genderqueer, and gender-neutral language* 13-15 June 2023, UCL, London.

Battye, A., & Roberts, I. G. (Eds.). (1995). *Clause structure and language change*. Oxford: Oxford University Press.

Bédé, J. E., & Gossez, R. (1984). *Un ouvrier en 1820: Manuscrit inédit de Jacques Étienne Bédé*. Paris: FeniXX.

Belletti, A. (1998). *Inversion as focalization*. Ms. University of Siena.

Belletti, A. (2001). Inversion as focalization. In Hulk, Aafke & Pollock, Jean-Yves (eds.) *Subject inversion in Romance and the theory of Universal Grammar*. Oxford: Oxford University Press pp. 60-90

Belletti, A. (2004). Aspects of the low IP area. In Rizzi, L. (ed.) *The structure of CP and IP: The cartography of syntactic structures Vol. 2*. Oxford: Oxford University Press pp. 16-51.

Belletti, A. (2008). The CP of clefts. *Rivista di Grammatica Generativa* 33, pp. 191–204.

Belletti, A. (2012). Revisiting the CP of Clefts. In Grewendorf, G. & Zimmermann, T. (eds.) *Discourse and grammar. From sentence types to lexical categories*, Berlin: De Gruyter. pp. 91-114.

Belletti, A. (2015). The focus map of clefts: Extraposition and Predication. In Shlonsky, U. (ed.) *Beyond functional sequence: the cartography of syntactic structures*, Oxford: Oxford University Press. pp. 42-60

Belletti, A. & Bocci, G. (2023). Subject versus object clefts: A fresh perspective on a robust asymmetry. Insights from French answering strategies. In Bonan, C. & Ledgeway, A. (eds.) *It-Clefts, Empirical and Theoretical Surveys and Advances*. Berlin: De Gruyter. pp. 81-104

Benincà, P. (2001). The position of topic and focus in the left periphery. In Cinque, G. and Salvi, G. (eds.) *Current studies in Italian syntax*. Leiden: Brill. pp. 39-64

Benincà, Paola (2006). A detailed map of the left periphery of medieval romance. In Sanuttini, Rafaella (ed.) *Crosslinguistic research in syntax and semantics: Negation, tense, and clausal architecture*. Washington, DC: Georgetown University Press pp. 53–86

Benincà, Paola and Cecilia Poletto. (2004). Topic, Focus and V2: defining the CP sublayers. In Rizzi, L. (ed.), *The Structure of CP and IP. The Cartography of Syntactic Structures vol. 2*, Oxford: Oxford University Press. pp. 52-76

Benincà, P., & Munaro, N. (Eds.). (2011). *Mapping the Left Periphery: The Cartography of Syntactic Structures, Volume 5*. Oxford University Press.

Biberauer, T., & Roberts, I. (2016). Parameter typology from a diachronic perspective. In Bidese, E., Cognola, F. & Moroni, M. C. (eds.) *Theoretical approaches to linguistic variation*, Amsterdam: John Benjamins. pp. 259-291.

Boeckx, C. (2001). French Complex Inversion. In Camps, Joaquin and Caroline Wiltshire, (eds.) *Romance Syntax, Semantics and L2 Acquisition: Selected Papers from the 30th Linguistic Symposium on Romance Languages, Gainesville, Florida, February 2000*. Amsterdam: John Benjamins. pp. 43-56

Boeckx, C. (2006). *Linguistic minimalism: Origins, concepts, methods, and aims*. Oxford: Oxford University Press.

Bonan, C. & Ledgeway, A. (2023). It-clefts: State-of-the-art, and some empirical challenges. In Bonan, C. & Ledgeway, A. (eds.) *It-Clefts, Empirical and Theoretical Surveys and Advances*. Berlin: De Gruyter. pp. 1-10

Börjars, K., Vincent, N., & Walkden, G. (2015). On constructing a theory of grammatical change. *Transactions of the Philological Society*, 113(3), pp. 363-382.

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

Bucholtz, M., & Hall, K. (2008). All of the above: New coalitions in sociocultural linguistics. *Journal of sociolinguistics*, 12(4), pp. 401-431.

Calder, J. (2019). The fierceness of fronted/s: Linguistic rhematization through visual transformation. *Language in Society*, 48(1), pp. 31-64.

Calder, J. (2023). Towards a queer and trans sociophonetics. In Kibbey, T. (ed.) *Linguistics Out of the Closet: The Interdisciplinarity of Gender and Sexuality in Language Science*, Berlin: De Gruyter. pp. 117-138.

Cardinaletti, A. (2004) Towards a cartography of subject positions. In Rizzi, Luigi (ed.) *The structure of CP and IP*. Oxford: Oxford University Press. pp. 115-165

Cardinaletti, A. (2018). On different types of postverbal subjects in Italian. *Italian Journal of Linguistics*, 30(2), pp. 79-106.

Cardinaletti, A. (2021). The position of subjects in Germanic and Romance questions. In Wolfe, S. and Nilsen, C. M. (eds.) *Continuity and Variation in Germanic and Romance*. Oxford: Oxford University Press. pp. 44-69

Cardinaletti, A., & Roberts, I. (2002). Clause Structure and X-second. In Cinque, G. (ed.) *Functional Structure in DP and IP: The Cartography of Syntactic Structures*. Oxford: Oxford University Press. pp. 123–166

Carter, P. M. (2013). Poststructuralist theory and sociolinguistics: Mapping the linguistic turn in social theory. *Language and Linguistics compass*, 7(11), pp. 580-596.

Cat, C. D. (2005). French subject clitics are not agreement markers. *Lingua*, 115(9), pp. 1195–1219. https://doi.org/10.1016/j.lingua.2004.02.002

Cedergren, H. J., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, 50(2), pp. 333-355.

Chafe, Wallace L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In: Charles N. Li (ed.): *Subject and topic*, New York: Academic Press. pp. 27–55

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Boston MA: MIT Press.

Chomsky, N. (1993). A minimalist program for linguistic theory. In Hale, K. & Keyser, S. J. (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*. Cambridge, MA: MIT Press. pp. 1–52

Chomsky, N. (1995). *The minimalist program*. Cambridge MA: MIT Press.

Chomsky, N. (2001). Derivation by Phase. In Kenstowicz, M. (ed.). *Ken Hale: A Life in Language*, Cambridge, MA: MIT Press. pp. 1-52.

Chomsky, N., & Fitch, W. T., Hauser, M. D. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598), pp. 1569-1579.

Christensen, T. K., & Jensen, T. J. (Eds.). (2022). *Explanations in Sociosyntactic Variation*. Cambridge: Cambridge University Press.

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

Cinque, G. (2004). Restructuring and functional structure. In Belletti, A. (ed.) *Structures and beyond: The cartography of syntactic structures, vol. 3*, Oxford: Oxford University Press. pp. 132-191.

Cinque, G. (2023). *On linearization: toward a restrictive theory*. Cambridge, MA: The MIT Press.

Cinque, G., & Rizzi, L. (2008). The cartography of syntactic structures. *CISCL Working Papers* vol. 2 pp. 42-58.

Clech-Darbon, A., Rebuschi, G., & Rialland, A. (1999). Are there cleft sentences in French. In Tuller, Laurice & Rebuschi, Georges (eds.) *The grammar of focus*, Amsterdam: John Benjamins. pp. 83-118.

Colasanti, V. (2023a). Functional gestures as morphemes: Some evidence from the languages of Southern Italy, *Glossa: a journal of general linguistics* 8(1), pp. 1–45.

Colasanti, V. (2023b). Gestural focus marking in Italo-Romance, Isogloss. *Open Journal of Romance Linguistics,* 9/(4)/5, pp. 1-39.

Collet, J. (2021). *Soldats de la liberté!: Lettres de soldats républicains pendant la guerre de Vendée: 1793-1795*. Le Lys et le Lin.

Corr, A. (2022) *The grammar of the utterance: How to do things with Ibero-Romance.* Oxford: Oxford University Press.

Coveney, A. (2011). A Language Divided against Itself ? Diglossia, Code-Switching and Variation in French, in Martineau, F. and Nadasdi, T. (eds.) *Le Français en contact: hommages à Raymond Mougeon*, Quebec City: Presses de l'Université Laval. pp. 51–85.

Cruschina, S. (2021). Topicalization in the Romance Languages. *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.650

Cummins, S., & Roberge, Y. (2004). Null objects in French and English. In Auger, J., Clements, J. C. & Vance, B. (eds.) *Contemporary Approaches to Romance Linguistics: Selected Papers from the 33rd Linguistic Symposium on Romance Languages (LSRL), Bloomington, Indiana, April 2003*, Amsterdam: John Benjamins. pp. 121-138.

Cummins, S., & Roberge, Y. (2005). A modular account of null objects in French. *Syntax*, 8(1), pp. 44-64.

Dagnac, A. (2012). How do you double your C? Evidence from an Oïl dialect. *Empirical Issues in Syntax and Semantics*, 9, pp. 77–94.

Daneš, F. (1968 [1957]). Some thoughts on the semantic structure of the sentence. *Lingua* 21, pp. 55-69.

De Andrade, A. (2018). Aboutness Topics in Old and Middle French: A corpus-based study on the fate of V2. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 63(2), pp. 194–220.

De Cat, C. (2002). *French dislocation*. Doctoral dissertation. University of York.

De Cat, C. (2007). *French Dislocation. Interpretation, syntax, acquisition*. Oxford: Oxford University Press. https://eprints.whiterose.ac.uk/10747/

De Cesare, A. M., Garassino, D., García, M. G., & Uth, M. (2018). Adverbial cleft sentences in Italian, French and English. In García García, M. & Uth, M. (eds.) *Focus realization in Romance and beyond*, Amsterdam: John Benjamins. pp. 201-255.

De Smet, H. (2009). Yahoo-Based Contrastive Corpus of Questions and Answers.

Delais-Roussarie, E., Doetjes, J., & Sleeman, P. (2004). Dislocation. In Corblin, F. and de Swart, H. (eds.) *Handbook of French Semantics*, Stanford CA: CSLI Publications. pp. 501–528.

Deleuze, G. & Guattari, F. (1980). *Mille Plateaux*, Paris: Les Éditions de Minuit.

Doehler, S. P., De Stefani, E., & Horlacher, A.-S. (2015). *Time and emergence in grammar: Dislocation, topicalization and hanging topic in French talk-in-interaction*. Amsterdam: John Benjamins.

Donaldson, B. (2011a). Left dislocation in near-native French. *Studies in Second Language Acquisition*, 33(3), pp. 399–432. https://doi.org/10.1017/S0272263111000039

Donaldson, B. (2011b). Nativelike right-dislocation in near-native French. *Second Language Research*, 27(3), pp. 361–390. https://doi.org/10.1177/0267658310395866

Donaldson, B. (2014). Socio-stylistic reflexes of syntactic change in Old French. *Journal of French Language Studies*, 24(3), pp. 319–345. https://doi.org/10.1017/S0959269513000203

Donaldson, B. (2016). Preverbal subjects, information structure, and object clitic position in Old Occitan. *Journal of Linguistics*, 52(1), pp. 37–69.

Dufter, A. (2008). On explaining the rise of c'est-clefts in French. In Detges, U. & Waltereit, R. (eds.) *The Paradox of Grammatical Change. Perspectives from Romance*, Amsterdam: John Benjamins. pp. 31-56

Dufter, A. (2009). Beyond focus marking: Fine-tuning the evolution of cleft types from Latin to Modern French. Paper presented at the DGfS-Jahrestagung 31 (AG 9: Focus marking strategies and focus interpretation). 4–6 October 2009, Osnabrück.

Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual review of anthropology*, 21(1), pp. 461-488.

Edelman, S. (2019). *Verbal behavior without syntactic structures: beyond Skinner and Chomsky*. arXiv preprint arXiv:4783474.

Elspaß, S. (2007). A twofold view 'from below': New perspectives on language histories and language historiographies. In Elspaß, S., Langer, N., Scharloth, J. & Vandenbussche, W. (eds.) *Germanic language histories 'from below' (1700-2000)*, Berlin: De Gruyter. pp. 3-9.

Elspaß, S. (2012). The use of private letters and diaries in sociolinguistic investigation. In Hernández-Campoy, J. M. & Conde-Silvestre, J. C. (eds.) *The Handbook of Historical Sociolinguistics*, New York: Wiley. pp. 156–169.

Emonds, J. (2004). Unspecified categories as the key to root constructions. In Adger, D., De Cat, C. & Tsoulas, G. (eds.) *Peripheries: Syntactic Edges and their Effects*, Berlin: Springer. pp. 75–120.

Engel, H. (2009). La dislocation dans le Miroir historial de Jean de Noyal. In Ferm, O. & Förnegård, P. (eds.) *Regards sur la France du Moyen Âge: Mélanges offerts à Gunnel Engwall* Stockholm: Sällskapet Runica et Mediaevalia. pp. 246-260

Ernst, G. (1985). *Gesprochenes Französisch zu Beginn des 17. Jahrhunderts: Direkte Rede in Jean Héroards' Histoire particulière de Louis XIII' (1605-1610)*. Tübingen: Niemeyer.

Ernst, G. (2018). *Textes français privés des XVIIe et XVIIIe siècles*. Berlin: De Gruyter. https://doi.org/10.1515/9783110482003

Erteschik-Shir, N. (2007). *Information Structure: The Syntax-Discourse Interface*. Oxford: Oxford University Press.

Fernández-Sánchez, J., & Ott, D. (2020). Dislocations. *Language and Linguistics Compass*, 14(9), e12391. https://doi.org/10.1111/lnc3.12391

Firbas, J. (1971). On the concept of communicative dynamism in the theory of Functional Sentence Perspective. *Brno Studies in English* 7, pp. 12-47.

Fleischman, S. (1991). Discourse Pragmatics and the Grammar of Old French: A Functional Reinterpretation of 'si' and the Personal Pronouns. *Romance Philology* 44.2, pp. 251-283.

Fleischman, S. (1992). Discourse and diachrony: The rise and fall of Old French SI. In M. Gerritsen and D. Stein (eds.*) Internal and external factors in syntactic change*. Berlin: De Gruyter. pp. 433-474 https://doi.org/10.1515/9783110886047.433

Ford, C. E., & Fox, B. A. (2010). Multiple practices for constructing laughables. In Barth-Weingarten, D., Reber, E. & Selting, M. (eds.) *Prosody in interaction*, Amsterdam: John Benjamin. pp. 339-368.

Foulet, L. (1935/6). L'extension de la forme oblique du pronom personnel en ancien français. *Romania* 61(243): pp. 257–315; pp. 401–463; *Romania* 62(245), pp. 27–91.

Fournier (1998). *Grammaire du français classique*. Paris: Belin SUP Lettres

Fournier, N. (2001). Expression et place des constituants dans l'énoncé en français classique: la relation sujet-verbe et la relation verbe-objet. *Langue française*, 130 pp. 89–107.

Fournier, N. (2009). Les Insertions entre le Sujet et le Verbe dans la Prose Classique. In Fontvielle, A. and Thonnerieux, S (eds.) *L'ordre des mots: à la lecture du text*. Lyon: Presses Universitaires de Lyon. pp. 45-67

Fox, B. A., Thompson, S. A., Ford, C. E., & Couper-Kuhlen, E. (2013). Conversation Analysis and Linguistics. In Sidnell, J. & Stivers, T. (eds.) *The handbook of conversation analysis*, New York: Wiley. pp. 726-740.

Francis, E. J. (2021). *Gradient Acceptability and Linguistic Theory*. Oxford: Oxford University Press.

Frascarelli, M. (2004). Dislocation, Clitic Resumption and Minimality: A comparative analysis of left and right Topic constructions in Italian. In R. Bok-Bennema, B. Hollebrandse, B. Kampers-Manhe and P. Sleeman (eds) *Romance Languages and Linguistic Theory 2002 Selected papers from 'Going Romance', Groningen, 28–30 November 2002*, Amsterdam: John Benjamins. pp. 98-118.

Frascarelli, M., & Hinterholzl, R. (2007). Types of topics in German and Italian. In Schwabe, K. & Winkler, S. (eds.) *On information structure, meaning and form*. Amsterdam: John Benjamins. pp. 87-116.

Frascarelli, M., & Ramaglia, F. (2009). Cleft Constructions at the Interfaces. *Lingbuzz*, < http://ling. auf. net/lingBuzz/000841>.

Frascarelli, M., & Ramaglia, F. (2013). (Pseudo) clefts at the syntax-prosody-discourse interface. In Hartmann, K. & Veenstra, T. (eds.) *Cleft structures*, Amsterdam: John Benjamins. pp. 97-138.

Fraser, B. (1972). Optional rules in grammar. *Monograph series on languages and linguistics*, *Georgetown University* 25, pp. 1-16.

Van Gelderen, E. (2011). *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.

German, J. S., & D'Imperio, M. (2016). The status of the initial rise as a marker of focus in French. *Language and speech*, 59(2), pp. 165-195.

Giorgi, A., and Longobardi, G. (1991). *The Syntax of Noun Phrases*, Cambridge: Cambridge University Press.

Giusti, G. (2005). At the left periphery of the Romanian noun phrase. In Coene, M. & Tasmowski, L. (eds.) *On space and time in language*, Cluj: Clusium, pp. 23-49.

Giusti, G. (2006). Parallels in clausal and nominal periphery. In Frascarelli, M. (ed.) *Phases of interpretation*, Berlin: De Gruyter. pp. 163-184.

Givón, T. (1983). *Topic Continuity in Discourse: A quantitative cross-language study*. Amsterdam: John Benjamins.

Gougenheim, G. (1973). *Grammaire de la langue française du seizième siècle*. Paris: Éditions A&J Picard.

Goulemot, J. M. (1981). *Valentin Jamerey Duval. Mémoires: enfance et éducation d'un paysan au XVIIIe siècle.* Paris: Le Sycomore

Grüter, T. (2006). *Object clitics and null objects in the acquisition of French*. Doctoral Dissertation, McGill University

Grüter, T. (2009). A unified account of object clitics and referential null objects in French. *Syntax*, 12(3), pp. 215-241.

Haegeman, L. (2004). Topicalization, CLLD and the left periphery. In Maienborn, C., Frey, W. & Shaer, B. (eds.), *ZAS Papers in Linguistics 35: Proceedings of the Dislocated Elements Workshop*, Berlin: Leibniz-Zentrum Allgemeine Sprachwissenschaft. pp. 157–192. https://doi.org/10.21248/zaspil.35.2004.226

Haegeman, L. (2012). *Adverbial Clauses, Main Clause Phenomena, and Composition of the Left Periphery: The Cartography of Syntactic Structures, Volume 8*. Oxford: Oxford University Press.

Haegeman, L., Meinunger, A., & Vercauteren, A. (2014). The architecture of it-clefts. *Journal of Linguistics*, 50(2), pp. 269-296.

Haegeman, L., & Stark, E. (2021). Register-specific subject omission in English and French and the syntax of coordination. In Wolfe, S. & Nilsen, C. M. *Continuity and variation in Germanic and Romance* (pp. 15-43). Oxford: Oxford University Press.

Hansen, M. B. M. (2005). From prepositional phrase to hesitation marker: The semantic and pragmatic evolution of French *enfin*. *Journal of historical pragmatics*, 6(1), pp. 37-68.

Hansen, M. B. M., & Visconti, J. (Eds.). (2014). *The diachrony of negation*. Amsterdam: John Benjamins.

Harris, M. (1978). *The Evolution of French Syntax: A Comparative Approach*. London: Longman

Heine B, Claudi U, Hünnemeyer F. (1991). *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago Press

Helland, H. P., Meklenborg Nilsen, C., & Lohndal, T. (2020). La topicalisation en français et en norvégien. Scolia. *Revue de linguistique*, 34, pp. 51-69.

Hopper, P. J. & Traugott, E. C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press

Hornsby, D. (2019). Variable liaison, diglossia, and the style dimension in spoken French. *French Studies*, 73(4), pp. 578-597.

Iatridou, S. (1995). Clitics and island effects. *University of Pennsylvania Working Papers in Linguistics*, 2(1), 11-31.

Jacob, D., & Garassino, D. (2018). Polarity Focus and Non-Canonical Syntax in Italian, French and Spanish. Clitic Left Dislocation and sì che / sí que-constructions. In Dimroth, C. & Sudhoff, S. (eds.) *The Grammatical Realization of Polarity Contrast: Theoretical, empirical, and typological approaches*. Amsterdam: John Benjamins. pp. 227-254

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

Jelinek, F. (1988). *Applying information theoretic methods: Evaluation of grammar quality*. Workshop on Evaluation of Natural Language Processing Systems, Wayne, PA.

Jucker, A. H. (ed.) (1995). *Historical pragmatics: Pragmatic developments in the history of English*. Amsterdam: John Benjamins.

Karssenberg, L., & Lahousse, K. (2015). Two types of clefts in French: information structure, existential presupposition and syntax. Paper presented at Incontro di Grammatica Generativa. 26-28 February 2015, Perugia, Italy.

Kauhanen, H., & Walkden, G. (2018). Deriving the constant rate effect. *Natural Language & Linguistic Theory*, 36, pp. 483-521.

Kayne, R., & Pollock, J. Y. (2001). New thoughts on stylistic inversion. In Hulk, A. & Pollock, J. (eds.) *Subject inversion in Romance and the theory of Universal Grammar*, Oxford: Oxford University Press. pp. 107-162.

King, R., & Nadasdi, T. (1997). Left dislocation, number marking and (non-)standard French. *Probus* 9(3), pp. 267–284. https://doi.org/10.1515/prbs.1997.9.3.267

Krifka M., & Musan R. (2012). Information structure: Overview and linguistic issues. In Krifka, M. & Musan, R. (eds.) *The Expression of Information Structure*. Berlin: De Gruyter. pp. 1-44.

Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language variation and change*, 1(3), pp. 199-244.

Kroch, A., & Santorini, B. (2009). *Revisiting the loss of V2 in French*. Paper presented at Digs11. 22-24 July 2009, University of Campinas.

Kroch, A. and Santorini, B. (eds.) (2021). *Penn-BFM Parsed Corpus of Historical French, version 1.0*.

Labelle, M., & Hirschbühler, P. (2018). Topic and focus in Old French V1 and V2 structures. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 63(2), pp. 264–287. https://doi.org/10.1017/cnj.2017.52

Labelle, M., & Valois, D. (1996). The status of post-verbal subjects in French child language. *Probus* 8(1), pp. 53–80. https://doi.org/10.1515/prbs.1996.8.1.53

Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, 45, pp. 715-762.

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, 1(1), pp. 97–120.

Labov, W. (1994). *Principles of Linguistic Change, Volume 1: Internal Factors.* New York: Wiley

Labov, W. (2001). *Principles of Linguistic Change, Volume 2: Social Factors* New York: Wiley

Labov, W. (2010). *Principles of Linguistic Change, Volume 3: Cognitive and Cultural Factors* New York: Wiley

Laenzlinger, C., & Soare, F. (2005). A cartographic approach to the Romance mittelfeld. *Rivista di Grammatica Generativa*, 30, pp. 17-69.

Lahousse, K. (2003). *The distribution of postverbal nominal subjects in French: A syntactic, semantic and pragmatic analysis* (Doctoral dissertation, Paris 8).

Lahousse, K. (2006). NP subject inversion in French: two types, two configurations. *Lingua*, 116(4), pp. 424-461.

Lahousse, K. (2022). Parametric variation in Romance synchronic V2 through a French lens. *Isogloss. Open Journal of Romance Linguistics*, 8(3), pp. 1-30.

Lahousse, K., & Borremans, M. (2014). The distribution of functional-pragmatic types of clefts in adverbial clauses. *Linguistics*, 52(3), pp. 793-836.

Lahousse, K., & Lamiroy, B. (2012). Word order in French, Spanish and Italian: A grammaticalization account. *Folia linguistica*, 46(2), pp. 387-416.

Lahousse, K., & Lamiroy, B. (2017). C'est ainsi que: grammaticalisation ou lexicalisation ou les deux à la fois?. *Journal of French Language Studies*, 27(2), pp. 161-185.

Lambrecht, K. (1981). *Topic, antitopic and verb agreement in non-standard French*. Amsterdam: John Benjamins.

Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.

Lambrecht, K. (2001). A framework for the analysis of cleft constructions. *Linguistics* 39, pp. 463-516.

Larrivée, P. (2021). An information structure scenario for V2 loss in Medieval French. *Diachronica*, 38(2), pp. 189-209.

Larrivée, P. (2022a). The curious case of the rare focus movement in French. In Garassino, D. and Jacob, D. (eds.). *When data challenges theory. Unexpected and paradoxical evidence in Information Structure*. Amsterdam: John Benjamins. pp. 183-202.

Larrivée, P. (2022b). Is Medieval French diglossic? New evidence on remnant V2 and register. *Isogloss*, 8(2).

Le Bidois, R. (1952). *L'Inversion du sujet dans la prose contemporaine (1900-1950)*. Paris: Artrey

Lecercle, J. J. (1990). *The violence of language*. London: Routledge.

Lecercle, J. J. (2006). *A Marxist philosophy of language*. Amsterdam: Brill.

Ledgeway, A. (2005). Moving through the left periphery: The dual complementiser system in the dialects of Southern Italy 1. *Transactions of the philological society*, 103(3), 339-396.

Ledgeway, A. (2010). Introduction: The clausal domain: CP structure and the left periphery. In D'Alessandro, R. Ledgeway, A. & Roberts, I. (eds.), *Syntactic Variation: The Dialects of Italy*, Cambridge: Cambridge University Press. pp. 38–51.

Ledgeway, A. (2012*). From Latin to Romance: Morphosyntactic typology and change*. Oxford: Oxford University Press.

Ledgeway, A. (2020). Variation in the Gallo-Romance left periphery: V2, complementizers, and the Gascon enunciative system. In Wolfe, S. & Maiden, M. (eds.) *Variation and Change in Gallo-Romance Grammar*. Oxford, Oxford University Press. pp. 71-99.

Leonetti, M. (2014). Spanish VSX. In *Romance Languages and Linguistic Theory 2012 Selected papers from 'Going Romance' Leuven 2012*. Amsterdam: John Benjamins. pp. 37-64.

Lightfoot, D. W. (1979). *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.

Linell, P. (2007). *Essentials of dialogism. Aspects and elements of a dialogical approach to language, communication and cognition*. Unpublished manuscript. Retrieved from 'https://edisciplinas.usp.br/pluginfile.php/5669304/mod_resource/content/1/2%20-%20Linell-%20Essentials%20of%20Dialogism.pdf'

Livia, A., & Hall, K. (Eds.). (1997). *Queerly phrased: Language, gender, and sexuality*. Oxford: Oxford University Press.

Lodge, R. A. (1993). *French: From Dialect to Standard*. London: Routledge.

Lodge, R. A. (2004). *A sociolinguistic history of Parisian French*. Cambridge: Cambridge University Press.

López, L. (2014). *A Derivational Syntax for Information Structure*. Oxford: Oxford University Press.

Loporcaro, M. (2003). Muhammad, Charlemagne, and Apocope, or the Cultural Relevance of Romance Historical Linguistics. *La corónica: A Journal of Medieval Hispanic Languages, Literatures, and Cultures*, 31(2), pp. 57-65.

Marchello-Nizia, C. (1998). Dislocations en ancien français: Thématisation ou rhématisation ? *Cahiers de praxématique*, 30, pp. 161-178 https://doi.org/10.4000/praxematique.2762

Marchello-Nizia, C. and Prévost, S. (2020). Expression et position des constituants majeurs dans les divers types de propositions. In Marchello-Nizia, C., Combettes, B., Prévost, S. and Scheer T. (eds.) *Grande Grammaire Historique du Français*. Berlin: De Gruyter. pp. 1055-1219.

Martineau, F. (2004). Un corpus de textes français pour l'analyse de la variation diachronique et dialectale. *Lexicometrica*, 5, pp. 1–4.

Martineau, F. (2005). Perspectives sur le changement linguistique: aux sources du français canadien. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 50(1-4), pp. 173-213.

Martineau, F., Hirschbühler, P., Kroch, A., and Morin, Y. C. (eds.) (2021). *MCVF corpus, parsed, version 2.0*

Martineau, F. & Mougeon, R. (2003). A sociolinguistic study of the origins of ne deletion in European and Quebec French. *Language*, 79(1), pp. 118-152

Mascarenhas, S. (2007). Complementizer doubling in European Portuguese. Ms., Amsterdam/New York: ILLC/NYU.

Massot, B. (2008). *Français et diglossie. Décrire la situation linguistique française contemporaine comme une diglossie: arguments morphosyntaxiques*. PhD Thesis, Université Paris VIII.

Mathieu, É. (2013). The left-periphery in Old French. In Arteaga, D. L. (ed.) *Research on Old French: The state of the art*. Berlin: Springer. pp. 327–350

Mazziotta, N. & Glikman, J. (2019). Oral représenté et narration en ancien français. Spécificités syntaxiques dans trois textes de genres distincts. *Linx*, 78. https://doi.org/10.4000/linx.3151

Meinunger, A. (1998). A monoclausal structure for (pseudo-) cleft sentences. *Proceedings of NELS*, 28, pp. 283-298.

Miyagawa, S. (2022). *Syntax in the treetops*. Cambridge MA: The MIT Press

Munaro, N. (2018). Complementizer doubling and subject extraction in Italo-Romance. *Linguistic Variation*, 18(2), pp. 315-335.

Nadasdi, T. (1995). Subject NP doubling, matching, and minority French. *Language Variation and Change,* 7(1), pp. 1–14. https://doi.org/10.1017/S0954394500000879

Nevalainen, T. & Raumolin-Brunberg, H. (2003). *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. London: Routledge.

Newmeyer, F. J. (2000). *Language form and language function*. Cambridge, MA: MIT press.

Ordóñez, Francisco. (2007). Cartography of postverbal subjects in Spanish and Catalan. In Baauw, S., Drijkoningen, F. & Pinto, M. (eds.) *Romance Languages and Linguistic Theory 2005 Selected papers from 'Going Romance', Utrecht, 8–10 December 2005*, Amsterdam: John Benjamins. pp. 259-280.

Ott, D. (2015). Connectivity in left-dislocation and the composition of the left periphery. *Linguistic Variation*, 15(2), pp. 225–290.

Panckhurst, R., Détrie, C., Lopez, C., Moïse, C., Roche, M., & Verine, B. (2014). *88milSMS. A Corpus of Authentic Text Messages in French*.

Paoli, Sandra (2007). The fine structure of the left periphery: COMPs and subjects; evidence from Romance. *Lingua*, 117, pp. 1057–1079.

Paoli, S. (2020). A step forward in understanding pas: The post-verbal negator in Old Occitan from the perspective of communication and rhetorical strategizing. *Zeitschrift Für Romanische Philologie*, 136(4), pp. 1018–1048. https://doi.org/10.1515/zrp-2020-0056

Poletto, C. (2000). *The Higher Functional Field: Evidence from Northern Italian Dialects*. Oxford: Oxford University Press.

Poletto, C. (2006). Parallel Phases: a study on the high and low left periphery of Old Italian. In Frascarelli, M. (ed.) *Phases of interpretation*, Berlin: De Gruyter. pp. 261-294.

Poletto, C. (2014). *Word Order in Old Italian*. Oxford: Oxford University Press.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago IL: University of Chicago Press.

Priestley, L. (1955). Reprise constructions in French. *Archivum Linguisticum*, 7(1), pp. 1–28.

Prince, E. F. (1978). A comparison of wh-clefts and it-clefts in discourse. *Language*, 54(4), pp. 883-906.

Prince, Ellen (1981). Towards a taxonomy of given-new information. In Peter Cole (ed.): *Radical pragmatics*, New York: Academic Press. pp. 223–56.

Reeve, M. (2011). The syntactic structure of English clefts. *Lingua*, 121(2), pp. 142-171.

Reeve, M. (2012). *Clefts and their relatives*. Amsterdam: John Benjamins.

Reinhart, T. (1981). Pragmatics and Linguistics: An Analysis of Sentence Topics in Pragmatics and Philosophy. *Philosophica*, 27(1), pp. 53–94.

Rialland, A, Doetjes, J & Rebuschi, G. (2002). What is Focused in C'est XP qui/que Cleft Sentences in French? In Bel, B. & Marlien, I. (eds.), *Proceedings of the First International Conference on Prosody, 11–13 April, Aix-en-Provence*, Aix-en-Provence: Laboratoire Parole et Langage. pp. 595–598.

Rinke, E. & Meisel, J. M. (2009). Subject-inversion in Old French: Syntax and Information Structure. In Kaiser, G. A. & Remberger, E.-M. (eds.) *Proceedings of the Workshop "Null-subjects, expletives and locatives in Romance"*. Konstanz: Fachbereich Sprachwissenschaft, Universität Konstanz. pp. 93-130.

Rizzi, L. (1982). *Issues in Italian Syntax*. Dordrecht: Foris Publications.

Rizzi, L. (1997). The fine structure of the left periphery. In Haegeman, L. (ed.) *Elements of grammar: Handbook in generative syntax*. Berlin: Springer. pp. 281-337.

Rizzi, L. (2013). Notes on cartography and further explanation. *Probus*, 25(1), pp. 197–226. https://doi.org/10.1515/probus-2013-0010

Rizzi, L. (2022). *Description and Explanation in the Left Periphery: some results, problems and prospects.* Paper presented at Mapping Syntax - Universals, Variation, Acquisition, and Change workshop. 16-18 November 2022, Oxford.

Rizzi, L., & Bocci, G. (2017). Left Periphery of the Clause: Primarily Illustrated for Italian. In M. Everaert & H. C. van Riemsdijk (Eds.), *The Wiley Blackwell Companion to Syntax, Second Edition*. New York: Wiley. pp. 1–30. https://doi.org/10.1002/9781118358733.wbsyncom104

Rizzi, L. & Roberts, I. (1989). Complex inversion in French. *Probus* 1, pp. 1–30.

Roberts, I. (1993). *Verbs and diachronic syntax: A comparative history of English and French*. Berlin: Springer.

Roberts, I. (2009). Varieties of French and the Null Subject Parameter. In Biberauer, T., Holmberg, A., Roberts, I. & Sheehan, M. (eds.) *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge: Cambridge University Press. pp. 303-327.

Roberts, I. (2010). *Agreement and Head Movement: Clitics, Incorporation, and Defective Goals*. Boston MA: MIT Press.

Roberts, I. (2019). *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press.

Roberts, I. & Roussou, A. (2002) The Extended Projection Principle as a condition on the tense dependency. In Svenonius, P. (ed.) *Subjects, Expletives, and the EPP*, Oxford: Oxford University Press. pp. 125-155.

Romaine, S. (1982). *Socio-historical linguistics: Its status and methodology*. Cambridge: Cambridge University Press

Rooth, M. E. (1985). *Association with focus (montague grammar, semantics, only, even)*. Doctoral Dissertation, University of Massachusetts Amherst.

Rouquier, M. (2014). *L'émergence des constructions clivées, pseudo-clivées et liées en français*. Paris: Classiques Garnier.

Rouveret, A. (2004). Les clitiques pronominaux et la périphérie gauche en ancien français. *Bulletin de la Société de Linguistique de Paris* 99.1, pp. 181-237.

Rowlett, P. (2007). *The syntax of French*. Cambridge: Cambridge University Press.

Rowlett, P. (2013). Do French speakers really have two grammars? *Journal of French Language Studies* 23(1), pp. 37-57.

Rutten, G., & Van der Wal, M. J. (2014). *Letters as loot: A sociolinguistic approach to seventeenth-and eighteenth-century Dutch*, Amsterdam: John Benjamins.

Salvesen, C. M. (2011). Stylistic Fronting and Remnant movement in Old French. In Berns, J., Jacobs, H. & Scheer, T. (Eds.), *Romance Languages and Linguistic Theory 2009 Selected papers from 'Going Romance' Nice 2009*. Amsterdam: John Benjamins. pp. 323–342. https://doi.org/10.1075/rllt.3.19sal

Salvesen, C. M. (2013). Topics and the left periphery: A comparison of Old French and Modern Germanic. In Lohndal, T. (ed.), *In Search of Universal Grammar: From Old Norse to Zoque*, Amsterdam: John Benjamins. pp. 131–172. https://doi.org/10.1075/la.202.10sal

Salvesen, C. (2014). Le complémenteur que et la péripherie gauche : Analyse diachronique. *Syntaxe et Sémantique* 15, pp. 47–80.

Salvesen, C. M., & Walkden, G. (2017). Diagnosing embedded V2 in Old English and Old French. In Mathieu, E. & Truswell, R. (eds.) *Micro-change and Macro-change in Diachronic Syntax*, Oxford: Oxford University Press. pp. 168-181.

Samek-Lodovici, V. (2015). *The interaction of focus, givenness, and prosody: A study of Italian clause structure*. Oxford: Oxford University Press.

Samek-Lodovici, V. (2022). Focused Hanging Topics. Paper presented at Mapping Syntax - Universals, Variation, Acquisition, and Change workshop. 16-18 November 2022, Oxford.

Samek-Lodovici, V. (Forthcoming). Focalization in-situ vs Focus Projection Focused topics, focused questions, focused heads, and other challenges. In Gallego, A. J. & Ott, D. (eds.), *Cartography and Explanatory Adequacy*. Oxford: Oxford University Press.

Samo, G. (2019). *A Criterial Approach to the Cartography of V2*. Amsterdam: John Benjamins.

Sankoff, D., & Labov, W. (1979). On the uses of variable rules. *Language in society*, 8(2-3), pp. 189-222.

Shaw, W. (2022). Old French SI, Syntax and Function in Diachrony. *Isogloss*, 8(2)/9 1-18.

Sheehan, Michelle. (2010). 'Free' Inversion in Romance and the Null Subject Parameter. In Biberauer, T, Holmberg, A. & Roberts, I. (eds.) *Parametric Variation: Null Subjects in Minimalist Theory*, Cambridge: Cambridge University Press. pp. 231-262.

Simonenko, A., Crabbé, B. & Prévost. S. (2018). Text form and grammatical changes in Medieval French: a treebank-based diachronic study. *Diachronica* 35.3: pp. 393-428.

Simonenko, A., Crabbé, B., & Prévost, S. (2019). Agreement syncretization and the loss of null subjects: quantificational models for Medieval French. *Language Variation and Change*, 31(3), pp. 275-301.

Sitaridou, I. (2012). A comparative study of word order in Old Romance. *Folia linguistica*, 46(2), pp. 553-604.

Slavkov, N., Melo-Pfeifer, S., & Kerschhofer-Puhalo, N. (eds.). (2021). *The changing face of the "native speaker": Perspectives from multilingualism and globalization*. Berlin: De Gruyter.

Smith, J. C. (2002). Middle French: When? What? Why?. *Language Sciences*, 24(3-4), pp. 423-445.

Smith, J. C. (2016). French and northern gallo-romance. In Ledgeway, A. and Maiden, M., eds. *The Oxford Guide to the Romance Languages*. Oxford: Oxford University Press. pp. 292–318

Speas, P., & Tenny, C. (2003). Configurational properties of point of view roles. In Di Sciullo, A. M. (ed.) *Asymmetry in grammar*, pp. 315-345. Amsterdam: John Benjamins Publishing Company

Sportiche, D. (1996). Clitic Constructions. In Rooryck, J. & Zaring, L. (eds.) *Phrase structure and the lexicon*. Dordrecht: Kluwer. pp. 213–277

Stark, E., & de Cesare, A. M. (2014). Frequency, form and function of cleft constructions in the Swiss SMS corpus. In De Cesare, A.-M. (ed.), *Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic*. Berlin: De Gruyter. pp. 325-344.

Stark, E., & Meier, P. (2017). Argument Drop in Swiss WhatsApp Messages: A Pilot Study on French and (Swiss) German. *Zeitschrift für französische Sprache und Literatur*, 127(3), pp. 224-252.

Stark, E., & Robert-Tissot, A. (2017). Subject drop in Swiss French text messages. *Linguistic variation*, 17(2), pp. 251-271.

Steele, A. J. (2019). *Non-binary speech, race, and non-normative gender: Sociolinguistic style beyond the binary*. Doctoral dissertation, The Ohio State University.

Steiner, B. D. J. (2014). *The evolution of information structure and verb second in the history of French*. Doctoral dissertation, Indiana University.

Steuckardt, A. (2015). *Entre village et tranches : l'écriture de poilus ordinaires*. Parthenay: Éditions Inclinaison.

Tang, W. (2023). A cartographic approach to Chinese V de O clefts. In Bonan, C. & Ledgeway, A. (eds.) *It-Clefts, Empirical and Theoretical Surveys and Advances*. Berlin: De Gruyter. pp. 235-256.

Titus-Brianti, G. (2014). Cleft sentences. A translation perspective on Italian and French [Chapitre de livre]. In *Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic : Contrastive, corpus-based studies*. De Gruyter. pp. 277-321

Traugott, E. C. (2010). Dialogic contexts as motivations for syntactic change. In Cloutier, R. A., Hamilton-Brehm, A. M. & Kretzschmar Jr., W. A. (eds.), *Studies in the History of the English Language V: Variation and Change in English Grammar and Lexicon: Contemporary Approaches*, Berlin: De Gruyter. pp. 11–27.

Traugott, E. C., & Trousdale, G. (2013). *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Trips, C., & Stein, A. (2018). Cleft sentences in the history of French and English. In García García, M. and Uth, M. (eds.) *Focus Realization in Romance and Beyond*. Amsterdam: John Benjamins. pp. 287-310

Troberg, M. (2004). Topic-comment resumptive pronouns in Modern French and Old and Middle French. *Toronto Working Papers in Linguistics* 23, pp. 133–156.

Trosborg, A. (1997). Text typology: Register, genre and text type. In Trosborg, A. (ed.) *Text Typology and Translation*, Amsterdam: John Benjamins. pp. 3-24.

Tsimpli, I. M. (2014). Early, late or very late?: Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism*, 4(3), pp. 283-313.

van Reenen, P. and Schøsler, L. (2000). The pragmatic functions of the Old French particles ainz, apres, donc, lors, or, puis, and si. In Herring, S., van Reenen, P., and Schøsler, L., (eds.), *Textual Parameters in Older Languages*, Amsterda; John Benjamins. pp. 59–105.

van Valin, R. D. (2003). Functional Linguistics. In Aranoff, M. & Ress-Miller, J. (eds.) *The Handbook of Linguistics*. Oxford: Blackwell. pp. 319-336.

Vance, B. S. (1995). On the Decline of Verb Movement to Comp in Old and Middle French. In Battye, A. and Roberts, I. (eds.) *Clause Structure and Language Change*. Oxford: Oxford University Press. pp. 173-199.

Vance, B. S. (1997). *Syntactic change in medieval French: Verb-second and null subjects*. Berlin: Springer. https://doi.org/10.1075/dia.16.2.15sch

Vander Klok, J. M., Goad, H., & Wagner, M. (2018). Prosodic focus in English vs. French: A scope account. *Glossa: a journal of general linguistics*, 3(1): 71, pp. 1-47.

Villa-García, J. (2012). Recomplementation and locality of movement in Spanish. *International Journal of Latin and Romance Linguistics*, 24(2), pp. 257-314.

Wasow, T. (1997). Remarks on grammatical weight. *Language variation and change*, 9(1), pp. 81-105.

Wehr, B. (2011). La phrase clivée en français: problèmes de description. In Dufter, Andreas & Jacob, Daniel (eds.) *Syntaxe, structure informationnelle et organisation du discours dans les langues romanes*, Lausanne: Peter Lang. pp. 189-214.

Wehr, B. (2015). Cleft constructions from a typological perspective. In Diao-Klaeger, S. & Thörle, B. (eds.) *Linguistique interactionnelle contrastive. Grammaire et interaction dans les langues romanes*. Tübingen: Stauffenburg.

Weinreich, U., William, L. & Marvin, H. (1968). Empirical foundations for a theory of language change. In: Lehmann, W. P. & Malkiel, Y. (eds), *Directions for Historical Linguistics: A Symposium*. Austin: University of Texas Press. pp. 95-195.

Wolfe, S. (2018). *Verb Second in Medieval Romance*. Oxford: Oxford University Press.

Wolfe, S. (2021). *Syntactic change in French*. Oxford: Oxford University Press.

Wolfe, S. (2022). Residual Verb Second in French and Romance. *Isogloss. Open Journal of Romance Linguistics*, 8(3), pp. 1-34.

Zimmermann, M. (2014). *Expletive and referential subject pronouns in Medieval French*. Berlin: De Gruyter.

Zimmermann, M. (2020). Investigating the setting of the null-subject parameter in Early Classical French. In Hinzelin, M., Pomino, N. & Remberger, E. (eds.) *Formal Approaches to Romance Morphosyntax*, Berlin: De Gruyter. pp. 117-143.

Zribi-Hertz, A. (2006). *Français standard et français dialectal: pour une reconnaissance du phénomène diglossique dans les descriptions syntaxiques du français*. Unpublished Master's thesis. Université Paris, 8.

Zribi-Hertz, A. (2013). De la notion de grammaire standard dans une optique diglossique du français. *Journal of French Language Studies*, 23(1), pp. 59-85.

Zubizarreta, M. L. (1998). *Prosody, Focus and Word Order*. Cambridge MA: MIT Press.

Appendices

A. List of Corpora and Collected Texts

Modéliser le changement: Les voies de français (MCVF) Martineau. Hirschbühler, Kroch & Morin (2021) https://github.com/beatrice57/mcvf-plus-ppchf

Penn Supplement. Kroch and Santorini (2021) https://github.com/beatrice57/mcvf-plus-ppchf

L'écriture des peu lettrés : Français vernaculaire dans la Normandie Médiévale (EPELE) https://www.unicaen.fr/epele/accueil

Soldats de la liberté!: Lettres de soldats républicains pendant la guerre de Vendée: 1793-1795. Collet (2021) (Vendée Letters)

Entre village et tranchées: l'écriture de poilus ordinaires. Steuckardt (2015) (WWI Letters)

Portrait sonore d'Orléans (ESLO). http://eslo.huma-num.fr/

88MilSMS. Panckhurst, Détrie, Lopez, Moïse, Roche & Verine (2014) http://88milsms.huma-num.fr/index_en.html

Yahoo-based Contrastive Corpus of Questions and Answers (YCCQA) de Smet (2009) https://varieng.helsinki.fi/CoRD/corpora/YCCQA/

Corpus de Français Parlé Parisien (CFPP2000), Branca-Rosoff, Fleury, Lefeuvre & Pires (2012) http://cfpp2000.univ-paris3.fr/index.html

B. List of Primary Sources

Linguistic Sources

Shortened Title	Full Title
Pantagruel (Book 4)	François Rabelais, *Le Quart Livre de Pantagruel*, 1552, Accessed through Gallica.
Belon du Mans	Pierre Belon du Mans *Les Observations de plusieurs singularitez et choses memorables de divers pays estranges*, 1555, Accessed through Gallica.
Héroard	Héroard's Record of the Dauphin's speech, 1605-1610. From Ernst (1985)
Valuche	Jacques Valuche, *Journal*, 1626-. From Ernst (2018)
Segrais	Jean Regnault de Segrais, *Les Nouvelles Françaises, Ou, Les Divertissements De La Princesse Aurélie*, 1656. Accessed through Gallica
Boucher	Pierre Boucher, *Histoire véritable et naturelle des moeurs et productions du pays de la Nouvelle-France*, 1663. Accessed through Gallica
Chavatte	Pierre Ignace Chavatte, *Chronique Memorial*, c. 1657-1693. From Ernst (2018)
Le Sopha	Claude-Prosper de Crebillon-Fils, *Le Sopha, Conte Moral*, 1742. Accessed through Gallica.
Jamerey-Duval	Valentin Jamerey-Duval, Memoires, *Enfance et éducation d'un paysan au XVIIIe siècle,* c. 1733-1747. From Goulemot (1981).
Bougainville	Louis Antoine de Bougainville, *Voyage autour du monde par la frégate du roi "la Boudeuse" et la flûte "l'Étoile" en 1766, 1767, 1768 & 1769, 1771*. 1766-1771. Accessed through Gallica.
Ménétra	Jacques-Louis Ménétra, *Journal de ma Vie*, 1764-1803. From Ernst (2018).
Bédé	Jacques Étienne Bédé, *Mémoires*, 1821-1830. From Gossez (1984).
Guérin	Maurice de Guérin, *Journal*, 1832-1835. Accessed through Gallica.
La Russie en 1839	Astolphe de Custine, *La Russie en 1839*, 1843. Accessed through Gallica.

Lui	Louise Colet, *Lui, Roman Contemporain*, 1860. Accessed through Gallica.
André Gide's diary	André Gide, *Journal*, 1924. Accessed in physical copy.
André Gide's travel writing	André Gide, *Voyage au Congo*, 1926. Accessed in physical copy.

Metalinguistic Sources

Shortened Title	**Full Title**
Maupas	Charles Maupas, *Grammaire et syntaxe Françoise* (1618) Accessed through Gallica
Vaugelas	Claude Favre de Vaugelas, *Remarques sur la langue Françoise* (1647) Accessed through Gallica
Bouhours	Dominique Bouhours, *Remarques nouvelles sur la langue Françoise* (1692) Accessed through Gallica

C. CorpusSearch Queries

Post-Verbal Subjects

Free inversion:

node: IP-MAT*

query: (IP-MAT* iDoms EJ|AJ)

AND (IP-MAT* iDoms VPP*)

AND (IP-MAT* iDoms NP-SBJ*)

AND (NP-SBJ* iDoms !*pro*|*con*)

AND (EJ|AJ Precedes NP-SBJ*)

AND (VPP* Precedes NP-SBJ*)

AND (EJ|AJ Precedes VPP*)

AND (VPP* iDoms !*ICH*)

Simple inversion:

node: IP-MAT*

query: (IP-MAT* iDoms EJ|AJ)

AND (IP-MAT* iDoms VPP*)

AND (IP-MAT* iDoms NP-SBJ*)

AND (NP-SBJ* iDoms !*pro*|*con*|PRO*)

AND (EJ|AJ Precedes NP-SBJ*)

AND (NP-SBJ* Precedes VPP*)

AND (VPP* iDoms !*ICH*)

Pronominal inversion:

node: IP-MAT*

query: (NP-SBJ* iDoms PRO)

AND (IP-MAT* iDoms VJ|EJ|AJ)

AND (IP-MAT* iDoms NP-SBJ*)

AND (VJ|EJ|AJ Precedes NP-SBJ*)

Ambiguous DP inversion:

node: IP-MAT*

query: (IP-MAT* iDoms VJ)

AND (IP-MAT* iDoms NP-SBJ*)

AND (NP-SBJ* iDoms !*pro*|*con*|PRO*)

AND (VJ Precedes NP-SBJ*)

Free inversion VOS:

node: IP-MAT*

query: (IP-MAT* iDoms EJ|AJ)

AND (IP-MAT* iDoms VPP*)

AND (IP-MAT* iDoms NP-SBJ*)

AND (IP-MAT* iDoms NP-ACC*|NP-DTV*)

AND (NP-SBJ* iDoms !*pro*|*con*)

AND (EJ|AJ Precedes NP-SBJ*)

AND (VPP* Precedes NP-SBJ*)

AND (EJ|AJ Precedes VPP*)

AND (VPP* iDoms !*ICH*)

AND (NP-ACC*|NP-DTV* Precedes NP-SBJ*)

AND (VPP* Precedes NP-ACC*|NP-DTV*)

Free inversion VSO:

node: IP-MAT*

query: (IP-MAT* iDoms EJ|AJ)

AND (IP-MAT* iDoms VPP*)

AND (IP-MAT* iDoms NP-SBJ*)

AND (IP-MAT* iDoms NP-ACC*|NP-DTV*)

AND (NP-SBJ* iDoms !*pro*|*con*)

AND (EJ|AJ Precedes NP-SBJ*)

AND (VPP* Precedes NP-SBJ*)

AND (EJ|AJ Precedes VPP*)

AND (VPP* iDoms !*ICH*)

AND (NP-SBJ* Precedes NP-ACC*|NP-DTV*)

Free inversion with adverbial:

node: IP-MAT*

query: (IP-MAT* iDoms EJ|AJ)

AND (IP-MAT* iDoms VPP*)

AND (IP-MAT* iDoms NP-SBJ*)

AND (IP-MAT* iDoms ADVP*)

AND (NP-SBJ* iDoms !*pro*|*con*)

AND (EJ|AJ Precedes NP-SBJ*)

AND (VPP* Precedes NP-SBJ*)

AND (EJ|AJ Precedes VPP*)

AND (VPP* iDoms !*ICH*)

AND (VPP* Precedes ADVP*)

Dislocations

Subject left dislocation:

node: IP-MAT*

query: (IP-MAT* iDoms NP-SBJ-RSP)

AND (NP-SBJ-RSP Doms PRO)

AND (IP-MAT* iDoms *-LFD)

AND (*-LFD Precedes NP-SBJ-RSP)

Object left dislocation:

node: IP-MAT*

query: (IP-MAT* iDoms CL-NP-ACC-RSP|CL-NP-RSP-2)

AND (IP-MAT* iDoms *-LFD)

AND (*-LFD Precedes CL-NP-ACC-RSP|CL-NP-RSP-2)

Non-nominal left dislocation:

node: IP-MAT*

query: (IP-MAT* iDoms CL-*-RSP*|ADVP-LOC-RSP|ADVP-DIR-RSP|PP-*-RSP)

AND (IP-MAT* iDoms *-LFD)

AND (IP-MAT* iDoms !NP-LFD)

AND (*-LFD Doms !CONJS)

AND (*-LFD Precedes CL-*-RSP*|ADVP-LOC-RSP|ADVP-DIR-RSP|PP-*-RSP)

Subject right dislocation:

node: IP-MAT*

query: (IP-MAT* iDoms NP-SBJ-RSP)

AND (NP-SBJ-RSP Doms PRO)

AND (IP-MAT* iDoms *-LFD)

AND (NP-SBJ-RSP Precedes *-LFD)

Object right dislocation:

node: IP-MAT*

query: (IP-MAT* iDoms CL-NP-ACC-RSP|CL-NP-RSP-2)

AND (IP-MAT* iDoms *-LFD)

AND (CL-NP-ACC-RSP|CL-NP-RSP-2 Precedes *-LFD)

Non-nominal right dislocation:

node: IP-MAT*

query: (IP-MAT* Doms CL|ADVP-LOC*)

AND (IP-MAT* Doms *-LFD)

AND (IP-MAT* Doms !NP-LFD)

AND (*-LFD Doms !CONJS)

AND (CL|ADVP-LOC* Precedes *-LFD)

To search for heavy (long) dislocations, add this to the relevant query:

AND (*-LFD domsWords> 4)

To search for heavy (complex) dislocations add this to the relevant query:

AND (*-LFD Doms IP-SUB*|IP-INF*)

Clefts

Clefts (tagged as such):

node: IP-MAT*

query: (CP-CLF* exists)

Surface-form search: (taken from Trips and Stein (2018))

node: IP-MAT*

query: (IP-MAT* idoms NP-SBJ) AND (NP-SBJ idoms PRO)

AND (PRO idoms [CcÇçZz][eio]|[CcÇçZz][eio].)

AND (IP-MAT* idoms NP-PRD*) AND (IP-MAT* idoms !CP-CLF*)

AND (NP-PRD* idoms CP-REL*)