



A corpus-based study of the prosody and information structure of English *it*-clefts and French *c'est*-clefts

Thesis submitted for the degree of Doctor of Philosophy

by

Charlotte Bourgoin

Supervisor: Prof. dr. Kristin Davidse

Co-supervisors: Prof. dr. Karen Lahousse

Prof. dr. Gerard O'Grady

Leuven/Cardiff, October 2022

“All we have to decide is what to do with the time that is given to us.” – LOTR

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors Kristin Davidse, Gerard O’Grady and Karen Lahousse for giving me the opportunity to conduct this PhD research. Their constant guidance, encouragement and support over the last four years have been invaluable. It has been a real privilege working on this project with such an incredibly kind and inspiring team of mentors. I would especially like to thank Kristin Davidse for whom I have nothing but complete admiration both on a professional and personal level. Kristin, I will forever be in awe of your passion, strength, generosity and immense knowledge.

I gracefully acknowledge that this joint doctoral project was made possible by the research grant ‘Beyond the clause: encoding and inference in clause combining’ (C14/18/034) (promoter: Jean-Christophe Verstraete, co-promoters: Bert Cornillie, Kristin Davidse and Elwys De Stefani), awarded by the research council of the University of Leuven.

Many thanks further go out to Lise Fontaine, Shirley Carter-Thomas, Andreas Dufter and Stefania Marzo for accepting to act as members of my doctoral examination board and for reading my work. My gratitude likewise goes to Michèle Goyens for accepting to oversee my doctoral defence. I further thank Hendrik De Smet and Anne-Catherine Simon for having been part of my advisory committee. I would also like to express my gratitude to Sylvia Lescauwaet without whom the organisation of my doctoral defence would have been a lot more challenging.

I would like to thank my colleagues from the project *Beyond the Clause* for useful data sessions and great conversations. Thanks also to my colleagues from the KU Leuven Linguistics department. Though the last two years have mainly been spent working from home, I will leave with very fond memories of our coffee room lunches. A special thanks goes to the 2018 cohort of PhD students, Virginia, Giulia, Marie-Anne, Marlieke, Catho, Afra, Ellison, Alexandra and Jorina, alongside whom I embarked on this PhD journey. Being able to relate to each other’s experiences has brought some much-needed relief in the toughest of times. I feel very lucky to have met such brilliant colleagues, all of whom I have no doubt will all go on to achieve great things. I would also like to give a shout out to Manon, Tom and Thomas with whom I have shared great conversations and venting sessions.

I also thank my colleagues from the English Language and Communication department at Cardiff University, most of whom I unfortunately only had the chance to meet online.

I wish to warmly thank Laure Lansari for believing in me and for so selflessly pushing me to pursue greater goals. I would not have had the opportunity to advance my academic career if it had not been for her support.

My appreciation also goes to Jozefien, Simone, Olivier, Clémence, Anna and Louise whose friendship I hold close to my heart and whose attentive ears have been more than welcome on many occasions.

Enfin, je tiens à remercier les deux personnes les plus importantes à mes yeux, et celles sans qui tout ceci n'aurait jamais été possible, mes parents. Papa, Maman, votre soutien sans faille m'a permis de surmonter bon nombre de moments difficiles lors de ces quatre dernières années et c'est en très grande partie grâce à vous que j'ai pu trouver la force d'arriver à bout de ce doctorat. Je ne peux que vous remercier encore et encore pour tous les sacrifices que vous avez faits pour moi et pour tous les précieux conseils que vous m'avez donnés. For everything you have done for me, this work is as much your accomplishment as mine.

Table of content

Acknowledgements	v
Table of content	vii
Introduction	1
Chapter 1: State of the art and research questions	6
1.1 Theoretical premises	7
1.1.1 Theoretical framework	7
1.1.2 Information structure	10
1.1.3 The notion of focus	15
1.1.3.1 The functional approach	16
1.1.3.2 The formal-pragmatic approach	20
1.1.3.3 Focus marking in French	21
1.2 The cleft construction: an overview	29
1.2.1 General overview of <i>it</i> - and <i>c'est</i> -clefts	29
1.2.1.1 The constituents of the cleft	30
1.2.1.2 The rationale of the cleft	33
1.2.1.3 The representational meaning of the cleft	36
1.2.2 The information structure of <i>it</i> - and <i>c'est</i> -clefts	41
1.2.2.1 Typologies of the discourse-familiarity of <i>it</i> - and <i>c'est</i> -clefts	42
1.2.2.2 Typologies of the prosody of <i>it</i> - and <i>c'est</i> -clefts	48
1.2.2.3 Shortcomings of the existing typologies	59
1.2.3 Functional studies of <i>it</i> - and <i>c'est</i> -clefts	62
1.3 Aims and research questions	68
Chapter 2: Data and methodology	72
2.1 Corpora	72
2.1.1 Comparability of corpora	73
2.1.2 London-Lund Corpus	74
2.1.3 Corpus de Référence du Français Parlé	75
2.2 Delineation and extraction of datasets	76
2.2.1 Recognition criteria for <i>it</i> -clefts	77
2.2.2 Recognition criteria for <i>c'est</i> -clefts	80
2.2.3 Recognition criteria of reduced <i>it</i> - and <i>c'est</i> -clefts	82
2.2.3 Extraction of data and building of the datasets	89

2.3 Prosodic analysis	90
2.3.1 Review of prosodic annotations and analysis of English data	91
2.3.2 Prosodic annotation and analysis of French data	96
2.4 Analytical model of discourse-familiarity	100
2.4.1 Nominal referent in the value	103
2.4.1.1 Evoked	103
2.4.1.2 Inferable	104
2.4.1.3 New-anchored	105
2.4.1.4 Brand-new	106
2.4.2 Open proposition in the variable	106
2.4.2.1 Evoked	107
2.4.2.2 Inferable	107
2.4.2.3 New-anchored	108
2.4.2.4 Brand-new	109
2.4.3 Prospective vs. retrospective analysis	110
2.5 Concluding remarks	111
Chapter 3: The English <i>it</i>-cleft	113
3.1 Morphosyntactic properties	113
3.1.1 Syntactic category of the clefted constituent	114
3.1.2 Grammatical function of the clefted constituent	117
3.1.3 Paradigm of relative markers	119
3.1.4 Modality, negation and exhaustivity particles in <i>it</i> -clefts	120
3.2 Prosodically coded information structure of <i>it</i> -clefts	122
3.2.1 Location of prosodic focus	123
3.2.2 Hierarchy of prosodic foci	126
3.2.3 Contrastive vs. presenting focus	131
3.3 Discourse-familiarity in <i>it</i> -clefts	135
3.3.1. Distribution of discourse-given and discourse-new information	136
3.3.2 Towards a new typology of the referential information structure of <i>it</i> -clefts	142
3.3.3 Functions of <i>it</i> -clefts	145
3.4 The interplay between the syntax and relational and referential information structure of <i>it</i> -clefts	149
3.4.1 The relation between syntax and prosodic foci	150
3.4.2 The relation between discourse-familiarity and focus assignment	153
3.4.3 Selective focus	157
3.4.4 Onsets and information flow management	161

3.5 Concluding remarks	165
Chapter 4: The French <i>c'est</i>-cleft	168
4.1 Morphosyntactic properties	168
4.1.1 Morphology of the copula and the relative marker	169
4.1.2 Paradigm of relative markers	173
4.1.3 Syntactic category of the clefted constituent	176
4.1.4 Grammatical function of the clefted constituent	179
4.1.5 Modality, negation and exhaustivity markers in <i>c'est</i> -clefts	181
4.2 The relational information structure of <i>c'est</i> -clefts	186
4.2.1 Prosodic patterns	187
4.2.2 The influence of interactional aspects on the prosody of <i>c'est</i> -clefts	200
4.2.3 The role of the initial accent	203
4.3 The referential information structure of <i>c'est</i> -clefts	209
4.3.1 Discourse-familiarity patterns	209
4.3.2 A revised typology of the discourse-familiarity of <i>c'est</i> -clefts	217
4.4 The interplay between the relational and referential information structure of <i>c'est</i> -clefts	221
4.4.1 The interaction between syntax and prosody	222
4.4.2 The interaction between discourse-familiarity and prosody	224
4.4.3 <i>C'est là que, c'est ainsi que, c'est comme ça que</i> : evidence for grammaticalisation of information structure?	229
4.5 Functional description of <i>c'est</i> -clefts	239
4.6 Concluding remarks	245
Chapter 5: Comparison of <i>it</i>-clefts and <i>c'est</i>-clefts	248
5.1. Morphosyntactic properties	248
5.1.1 Frequency of occurrence	248
5.1.2 Morphology of the copula and paradigm of relative markers	250
5.1.3 Form and function of the clefted constituent	252
5.2. The operationalisation of information structure	256
5.2.1 The relational information structure of <i>it</i> - and <i>c'est</i> -clefts	256
5.2.2 The referential information structure of <i>it</i> - and <i>c'est</i> -clefts	259
5.3. Pragmatic features of <i>it</i> - and <i>c'est</i> -clefts	262
5.3.1 Exhaustivity	262
5.3.2. Contrast	271
5.3.3 Discourse functions of <i>it</i> - and <i>c'est</i> -clefts	281
5.4 Concluding remarks	287

Chapter 6: The case of reduced clefts	289
6.1 State of the art	289
6.2 General characteristics of reduced clefts	294
6.2.1 Frequency of use of reduced clefts	294
6.2.2 Morphosyntactic properties	297
6.2.3 French PRICs: reduced clefts or pseudo-clefts?	300
6.3. Information structure of reduced clefts	304
6.3.1 The referential information structure of reduced clefts	304
6.3.2 The relational information structure of reduced clefts	310
6.3.3 A functional perspective	313
6.4 Concluding remarks	317
Conclusion	320
Glossary	327
References	329

Introduction

The present thesis investigates the use of cleft constructions in spoken data. The term ‘cleft’ sentence, first coined by Jespersen (1927), refers to constructions in which the information of a simple clause, e.g. (1a) and (2a), is repackaged into a matrix clause and a cleft-relative clause, e.g. (1b) and (2b) (Huddleston and Pullum 2002). Clefts are productive in a number of languages which include, but are not limited to, English and French (Dufter 2009a), the two languages this study compares. The clefts under study are *it*-clefts for English and *c’est*-clefts for French.

- (1) (a) The grammar is interesting.
(b) It’s the grammar which is interesting. (London Lund Corpus 1, LLC-1)
- (2) (a) Ton œil voit.
‘Your eye sees.’
(b) C’est ton œil qui voit. (Corpus de Référence du Français Parlé, CRFP)
‘It’s your eye that sees.’

It is generally accepted that the bi-clausal syntax of clefts foregrounds the postcopular element in the matrix clause (henceforth referred to as clefted constituent) vis-à-vis the information conveyed in the cleft-relative clause which is consequently backgrounded (Quirk et al. 1985; Huddleston and Pullum 2002). Semantically, clefts are specificational-identifying constructions, whereby the copula *be/être* expresses a relation of identification between the clefted constituent, e.g. *the grammar* in (1b) and *ton œil* in (2b), and the presupposed open proposition introduced in the cleft-relative clause, e.g. *x is interesting* and *x voit* (Bolinger 1972; Davidse 2000). Hence, the clefted constituent is construed as the value filling the semantic gap in the open proposition making up the variable.

Because they modify the linear order of syntactic constituents, clefts have been categorised as ‘information-packaging’ constructions (Huddleston and Pullum 2002)

which speakers use to manipulate information structure (Chafe 1974). As such, clefts are a good example of the type of choices speakers make in real time to accommodate to the speaker's ever-changing states of mind (Chafe 1976). These choices are operationalised with different information structural phenomena including focus, i.e. focal/non-focal, and discourse-familiarity, i.e. new/given which, as I show in this study, are distinct, but interacting layers of information structure.

The literature on the information structure of clefts has been strongly influenced by both the formal-pragmatic approach inspired largely by Lambrecht (1994, 2001) and the functional approach instigated by Halliday (1967a, 1967b, and Greaves 2008), which this study adheres to. Theoretically, both of these approaches make claims on the way the information structural phenomena of discourse-familiarity and focus are operationalised. Lambrecht (2001: 474) conceives of information structure as the pragmatic structuring of propositions, whereby the focus is the new and typically accented information that makes the pragmatic assertion differ from the presupposition. On Lambrecht's (2001) account, the focus of clefts is conveyed by the clefted constituent. In the Hallidayan tradition, by contrast, information units and information foci are intrinsically coded by prosody in English spoken discourse (Halliday 1967a: 200). Speakers are viewed as having considerable freedom to mark off information units whose internal structure always features focal information, typically related to non-focal information. Within the Hallidayan approach, clefts are predicted to afford several information structures in which the clefted constituent is not necessarily focal.

In descriptive practice, there exist a number of typologies (among others Prince 1978; Declerck 1984; Doetjes et al. 2004; Mertens 2012), which this study reacts to, accounting for the ways in which *it*- and *c'est*-cleft partition information. These typologies seek to identify the different informational statuses, e.g. new/given or focal/non-focal, the complement of the matrix clause and the cleft-relative clause may have. In doing so, most of the typologies, or at least those describing *it*-clefts, tend to conflate the two layers of information structure, i.e. focus assignment and discourse-familiarity. This conflation characterises the two types posited in Prince's (1978) influential typology. The *stressed-focus* cleft is defined as having a new clefted element with strong stress and a cleft-relative clause with given and weakly

stressed information. The *informative-presupposition* cleft is said to have an unstressed, given clefted element and a cleft-relative clause with new and normally stressed information. Corpus-based studies in both English (Delin 1990; Collins 2006; Kimps 2016) and French (Rialland et al. 2002; Doetjes et al. 2004; Avanzi 2011; Mertens 2012) have revealed greater prosodic variation which is not in fact bound to Prince's (1978) two types of clefts, but they continue to work with a modified form of her binary typology. In this study, I argue that a model of the information structure of clefts has to make a principled distinction between discourse-familiarity (Kaltenböck 2005) and focal versus non-focal status as marked by prosody (Halliday and Greaves 2008).

Along with the full variant exemplified in (1a) and (2a), clefts also have a reduced form, e.g. (3) and (4), in which only the matrix clause is expressed while the cleft-relative clause is omitted.

- (3) and how she heard repeated bangs on the ceiling thinking **it was her son** [~~who was banging on the ceiling~~] (LLC-1)
- (4) dans la journée c'est moi qui vais livrer ou alors **c'est ma patronne** [~~qui va livrer~~] (CRFP)
'the during the day it's me who do the deliveries_or it's my boss [~~who does the deliveries~~']

Reduced clefts, whose existence has only been acknowledged in a handful of studies (e.g. Declerck 1988; Hedberg 2000; Belletti 2005; Mikkelsen 2007), are generally viewed as the monoclausal variant of full clefts whose variable is not realised, but still presupposed, because it has already been mentioned in the prior context. Outside of this high salience analysis, the overall description of reduced clefts has so far remained limited, and their use does not appear to be validated as a fairly common alternative choice. Hence, this study aims at filling the gap in the treatment of reduced clefts by providing an empirically based account of their use.

From a contrastive viewpoint, *it*- and *c'est*-clefts have been recognised as cross-linguistic equivalents which share the same basic syntactic structure (Bourns 2014; Lambrecht 2001; Carter-Thomas 2009; Dufter 2009a). This syntactic resemblance does not, however, translate into a similar frequency of occurrence.

Clefts are generally more common in French (Carter-Thomas 2009; Bourgoin 2017) which is partly explained by their use as a compensating device to circumvent the more rigid SVO order of French (Carter-Thomas 2009). The difference in use has also been justified with the underlying pragmatic differences in the two languages (Bourns 2014), but yet no in-depth comparative account of their information structure is available.

This study contributes to the literature on clefts in several ways. First, by combining quantitative and qualitative corpus-based analyses, the present research not only refines but also empirically substantiates the main claims bearing on the morphosyntactic, semantic, pragmatic, and information structural properties of *it*- and *c'est*-clefts. Furthermore, by extensively investigating the use of *it*- and *c'est*-clefts individually first, and then in comparison with one another, the present study simultaneously adds to the respective literatures and the contrastive state-of-the-art. This research also holds methodological value in that it offers a fine-grained analytical model for the study of discourse-familiarity and precise reference points for the investigation of prosodically coded focus. Finally, treating the reduced variant of clefts as a category in its own right allows me to shed light on its main properties and establish its role as a full-fledged type of cleft.

This thesis is structured as follows. In Chapter 1, I introduce the two general categories under study, namely information structure and cleft constructions. In view of the numerous approaches to both notions developed in the literature, which I detail and assess in relation with one another, the overview also includes clarification on the framework within which this research operates. In particular, I address the ways the information structural phenomena of focus and discourse-familiarity are conceived. Taken all together, this enables me to establish the goals and research questions by which this study is driven.

In Chapter 2, I present the data used to test the different working hypotheses, describing the two comparable corpora, i.e. London-Lund Corpus (LLC-1) and Corpus de Référence du Français Parlé (CRFP), from which the English and French datasets are built. I also explain the methods adopted for the prosodic analysis of data and the investigation of discourse-familiarity. This includes the proposal of a refined analytical model the different types of given and new which builds on Kaltenböck's (2005) quinary taxonomy and an instrumental methodology for the

study of information focus, both of which allow me to better capture the way information structure is operationalised.

Chapter 3 is devoted to the main findings on the English *it*-cleft which include results relating to its morphosyntax, semantics, pragmatics, and information structure, the latter of which inform the typologies of prosodic and discourse-familiarity patterns I propose. More generally, these allow me to characterise the relation between syntax and prosody on the one hand, and the different layers of information structure on the other.

Chapter 4 follows the same model as Chapter 3 but deals with the French *c'est*-cleft. In addition to what is already spelled out for Chapter 3, this chapter also addresses the question of grammaticalisation of clefts and their information structure, which is arguably more advanced in French.

Building on Chapters 3 and 4, Chapter 5 offers a contrastive perspective. For that reason, this chapter is once again built on the same model in which the syntactic, semantic, pragmatic and information structural systems of *it*-clefts are this time contrasted with that of *c'est-clefts* to determine the extent of their comparability. The comparison of the two languages allows me to go into the questions of exhaustivity and contrast in depth, which had otherwise only been concisely examined in the previous chapters.

The last chapter, Chapter 6, is devoted to the reduced variant of *it*- and *c'est*-clefts. I establish their syntactic, semantic, pragmatic, and information structural profile, which reveals itself to differ from that of full clefts in more than one way. I conclude that reduced clefts are a construction in their own right.

Finally, I conclude this thesis by summarising the key findings of my quantitative-qualitative study, which reset base-line thinking about clefts. Rather than constructions dedicated to focus marking, I have shown that they are constructions whose specificational meaning interacts with multiple, versatile information structural patterns, whose overall profiles in English and French differ considerably. As such, these findings can generate new descriptive, contrastive, and diachronic research questions to be examined in future studies.

Chapter 1: State of the art and research questions

It- and *c'est*-clefts, which this study focuses on, have received a lot of attention in both functional and formal approaches. As 'information-packaging' constructions (Huddleston and Pullum 2002), they are generally regarded as one of the many types of choices speakers continually make to accommodate to the hearer's knowledge at the time of utterance (Chafe 1976) and to manage information flow and interactional needs in real time (O'Grady 2010, and Bartlett 2019). More specifically, they exemplify how prosody and alterations to the linear order of syntactic elements can combine to produce certain information structural patterns (Lehmann 2008). Along with this information packaging function, cleft constructions also convey identifying-specificational meaning whereby the postcopular constituent in the matrix clause is construed as the value for the variable introduced in the cleft-relative clause. These functions are coded by different aspects of the structural organisation of the cleft. While specificational meaning is coded by syntax, information packaging is generally linked to focus marking for which different accounts exist. The present thesis develops a functional-cognitive analytical model aiming to capture how the discourse embedding of clefts, i.e. the discourse familiarity of their constituents, interacts with focus assignment in naturally occurring data. As such, this research is strongly grounded in empirical study of spoken contextualised data. No such study contrasting English and French has been carried out yet.

The goal of this chapter is to set the scene for the present study by offering an overview of the treatment of information structure and clefts. Because of the extensive nature of the literature on both components, I do not provide an inventory of works which is meant to be exhaustive by any means but instead which introduces the background necessary to motivate the theoretical approach adopted here and fully grasp the very rationale of this research. I start by establishing some necessary theoretical premises in Section 1.1. These include the theoretical framework I adopt, the main underlying principles of information structure and the different approaches to the notion of focus. I then move on the state of the art of *it*- and *c'est*-clefts in Section 1.2. I account for the structure of the cleft, its specificational

semantics, its information structural properties and its discourse functions. Finally, I introduce the main goals and research questions this study examines.

1.1 Theoretical premises

After defining the theoretical framework this study subscribes to in Section 1.1.1, I introduce the main concepts of information structure in Section 1.1.2 which I relate to the multi-layered description offered by Gundel (2003). I then provide an in-depth account of focus-marking in English and French in Section 1.1.3¹, which is encoded by different prosodic systems in the two languages.

1.1.1 Theoretical framework

To account for the use of *it*- and *c'est*-clefts, the present study heavily relies on the functional theory started by Halliday (1967a, 1967b, 1994, and Greaves 2008, and Matthiessen 2004, 2014), and developed further within the British school of intonation (Tench 1996; Brazil 1997; O'Grady 2013, 2014, and Bartlett 2019).² What follows is a description of the main principles of the Hallidayan tradition.³ The primordial tenet of Halliday's (1994) functional linguistic theory is that the grammatical meaning of linguistic signs is naturally symbolised by the grammar and prosody of a language. The function and form of a grammar cannot be separated from each other. They constitute the two sides of the same symbolic relation. Grammatical meaning is language-specific, not universal. From language-specific usage, the semantics of grammar can be derived in basically two ways: firstly, by cracking the coding relation between form and meaning, and, secondly, by

¹ Substantial parts of Section 1.1 consist of revised material from Bourgoïn and Davidse (in press).

² In addressing how the management of interactional needs and expectations is dealt with incrementally by speakers, the work carried out by the British school of intonation (e.g. O'Grady 2013, 2014, and Bartlett 2019) also relates to that developed in the interactional linguistics field (see Couper-Kuhlen 2014).

³ I am grateful to Kristin Davidse for enlightening me and guiding me through the many intricacies of Halliday's functional approach to language.

generalising from specific usage tokens to increasingly schematic structural representations.⁴

The second crucial tenet in Halliday's (1967, 1994) theory is that utterances convey three types of highly generalised and interacting functions, which are all naturally coded by grammatical and phonological form, and which, very roughly, can be summarised as follows. The *representational*⁵ organisation of the clause is concerned with representing experienced processes in the world as well as processes within our consciousness in terms of process-participant structures. This organisation is coded by what are often referred to as verb-argument structures (Halliday 1967a, 1994). The *interpersonal* organisation of the clause moulds these representations into *speech acts* coded by mood, modality and prosody. The *textual* organisation signals relations internal to the messages being communicated as well as relations with both the co-text and situational context, through word order and prosodic choices.

It has sometimes erroneously been thought that Halliday's three-level organisation of the clause is equivalent to the traditional distinction between semantics, syntax and pragmatics, as is indeed assumed by Lambrecht (1994: 6-7).⁶

⁴ This does not mean that Halliday rejects, for instance, a logical approach to language. He merely states that logic does not capture language-specific meaning, and is, of its nature, not concerned with the symbolic relation defining linguistic signs. He does recognise a type of meaning, referred to by Hjelmslev ([1947] 1961) as the 'purport' of utterances, which is "intertranslatable" between different languages and "could, for example, be analyzed from one or another logical, or ... psychological, point of view" (Hjelmslev [1947] 1961: 51) or within (social) anthropology (Hjelmslev [1947] 1961: 78).

⁵ Halliday (1967a, 1994) in fact uses the term "ideational" for this layer of organisation, rather than "representational" (Hengeveld 1989). I use the latter term as it has wider currency and seems more transparent. There are certain but by no means complete overlaps between how the representational/ideational and interpersonal layers of functional organisation of the clause are conceived of in the Hallidayan traditional and Functional Discourse Grammar (Hengeveld and Mackenzie 2018).

⁶ As Bateman (2017: 17) notes, while many schools of linguistics take the three-way system of syntax-semantics-pragmatics as the basis for their description of language characteristics, it is not the case of Systemic Functional Linguistics which instead adopts an 'in contexts' perspective and examines the "properties of [the] situations, generally social, in which language is occurring" (Bateman 2017: 11).

This is emphatically not the case. The traditional distinction assumes that syntax is concerned with purely structural combinatory relations, semantics with logical or conceptual representations of meaning, and pragmatics with how context contributes to meaning. At first sight, the non-adherence to this traditional distinction might appear surprising, but the alternative view of grammatical signs as symbolic form-meaning pairings, also advocated by Bolinger (1977), has in fact been gaining ground in such theories like Langacker's (1987, 1991, 2021) Cognitive Grammar and a number of Construction Grammars. The specific Hallidayan take on grammatical signs has been most significantly developed in McGregor's (1997, 2021) Semiotic Grammar, while it has arguably gained the greatest common currency in Traugott's (1982, 1989, 2010) work on language change. The first formulation of her hypothesis about subjectification as a process of change referred directly to the three types of functional structure distinguished by Halliday: change was predicted to start from representational "meanings situated in the described external or internal (evaluative/perceptual /cognitive) situation" over "meanings situated in the textual situation" to "interpersonal [meanings] ... such as expressions of speech function, exchange structures, and attitude" (Traugott 2010: 31). Importantly, Traugott (2010: 46-49) stresses that one can only speak of change if the new meaning has become conventionally associated with a specific form, yielding a new form-meaning pairing. In other words, in this school of linguistics, it is also assumed that representational, interpersonal and textual meaning all involve conventionalised form-meaning pairings. In the approach assumed here, pragmatic meaning is not directly coded by grammatical or prosodic form.⁷ Instead, it is inferred from coded meanings at two levels, firstly, the "systematicity of inference deeply interconnected to linguistic structure and meaning" (Levinson 1995: 93-94) and secondly highly context-dependent "nonce-speaker meaning" (Levinson 1995: 93). Note that, on the view that 'focus' is coded by prosody, which I adopt in this study, it is in Levinson's (1995) terms not 'pragmatic' meaning. Following the main

⁷ The SFL tradition differs from other schools of linguistics in its treatment of 'pragmatics' in that it does not view it as a separate system. As Bateman (2017) puts it, because language is taken to be a 'social semiotic' system (Halliday 1978), and because SFL is primarily concerned with the study of language 'in contexts', then "a separate designation of an area of 'pragmatics' within this theoretical orientation is considered redundant" (Bateman 2017: 20).

Hallidayan principles laid out in this section, I analyse the use of *it*- and *c'est*-clefts in context and characterise their grammatical and prosodic properties, and coded and pragmatic meanings.

1.1.2 Information structure

Having set out the theoretical framework, I now introduce the main concepts encompassed under the general term information structure. When producing discourse, speakers exchange information that is organised in a sequence of temporally ordered units. They structure the information within and across these units in such a way that the common ground shared between speaker and hearer is constantly updated. In doing so, they make a number of choices in how they organise the information which may be motivated by certain considerations of the hearer's temporary state of mind or by specific communicative goals. This process is referred to as information structure (Halliday 1967; Lambrecht 1994; Gundel 2003) or information packaging (Chafe 1974, 1976). Tied to this general notion of information structure are phenomena such as that of topic, givenness, and focus, which all relate to how referents are introduced in discourse (Chafe 1976; Krifka 2008) but which are coded by different components.

Information structure can be first characterised by the dichotomy between *referential* and *relational* information structure deriving from Gundel's (1988, 1999, 2003) influential distinction between referential and relational givenness. Gundel (2003) defines *referential* givenness as the "relation between a linguistic expression and a corresponding non-linguistic (conceptual) entity in (a model of) the speaker/hearer's mind, the discourse, or some real or possible world" (Gundel 2003: 125). Put simply, in the referential domain, the given/new status of an entity primarily relates to the presence or absence of a corresponding entity either in the discursive context, the common ground or in the world. Beyond its mere presence in the context, it is also the degree of salience of the corresponding entity that determines the extent of the entity's referential givenness.

Other definitions of referential givenness, or simply 'givenness' (e.g. Chafe 1976; Prince 1981, 1992), can be found in the literature. For example, Chafe (1976) defines given information as information which the speaker is expecting to already

be present in the hearer's consciousness at the time of the utterance and new information as information the speaker is expecting to be introducing for the first time in the hearer's consciousness. In associating givenness with consciousness, Chafe (1976) departs from the definition offered by Clark and Haviland (1977) according to which given information is differentiated from new information on the basis of it being already known to hearer, and not simply consciously activated. For Chafe (1976) and Prince (1981), this difference is significant in so far as a piece of information can be familiar to the hearer but not be in his/her mind as the conversation is happening and vice versa. Krifka (2008) makes a similar claim by treating referents as given if, and only if, their denotation is at least somewhat present in the common ground. While Krifka (2008) does not elaborate on which factors determine the degree of givenness of a referent, Chafe (1976) designates prior mention of a referent as the main linguistic indicator of its salience in the hearer's consciousness. Because referents gradually lose their referential status as speech progresses, this raises the question of the measure of the degree of referentiality of a given referent. On this, Givón (1983: 7, 2001: 123) proposes to rely on 'anaphoric distance' which is used to measure referentiality according to the number of clauses between the current and last reference of a given referent. Chafe (1994: 183) relates this to 'activation cost', which he suggests to measure recoverability in terms of clauses or tone units. In Halliday's (1967a, and Hasan 1976) textual approach, items are classified as given if they are recoverable from the preceding context. In languages such as English and French, the determiner system is generally taken to be the main recoverability marker. Thus, the definite-indefinite contrast of NPs system generally marks the distinction between referent presumed retrievable and referent not presumed retrievable. According to Du Bois (1980) and Martin (1992), indefinite determiners typically signal that the referent of NP is mentioned for the first time, e.g. *a* boy in (1), while definite NPs generally code their referent as retrievable from the co-text or context, e.g. *the* boy in (1).

- (1) then **a boy** [1st mention] comes by... on a bicycle; the man is in the tree,... and **the boy** [2nd mention] gets off the bicycle (Du Bois 1980: 206, my emphasis)

As another layer of information structure, *relational* givenness is defined by Gundel (2003) as the relation between “two elements of the same level of representation” (Gundel 2003: 125). The propositional content is here partitioned into two parts, the subject and what is predicated about the subject, whose givenness/newness is dependent on that of one another. Unlike referential givenness, relational givenness is not concerned with the speaker’s assumptions but with the way the propositional content is packaged within the sentence boundaries. Although Gundel (2003) assimilates relational givenness/newness with phenomena such as subject/predicate or topic/comment, she also appears to draw a fairly straightforward link between relational newness and prosodically coded focus. This is instantiated in example (2), which Gundel (2003) uses to argue that referential and relational givenness/newness need not coincide. If the pronoun *she* is taken to co-refer with the referent *Pat* mentioned in the same proposition, it can be analysed as referentially given. But from a relational viewpoint, *Pat* is the topic of the proposition and *she* the prosodically coded focus which makes the pronominal referent relationally new.

(2) A: Who called?

B: Pat said **SHE** called. (Gundel 1980, my coding)

Hence, when extending the definition of relational givenness to the function of focus coded by prosody, elements of a message unit can be analysed as either focal or non-focal.

The prosodically-based distinction between focal and non-focal, which is also developed in other accounts (e.g. Halliday 1967a), is to be distinguished from that established between focus and background by some authors (e.g. Karszenberg and Lahousse 2015; Rochemont 2016; Hedberg 2021) in which the focal status of a constituent is not wholly dependent on prosody. A difference should also be made with descriptions of focus as a semantic proposition (Akmajian 1970: 220; Rooth 1985) or a “pragmatic” role (Lambrecht 2001: 474), i.e. pertaining to the information structural role assigned to referents. Within the functional approach advocated by Halliday (1967a), a constituent is considered to be focal if it bears prosodic focus, which, in English, not only marks the information as focal, but also

presents it to the hearer as if it were new, regardless of its actual discursive availability. This was inspired by the Prague school view that intonation and word order, in interaction with retrievability, convey information structural meanings (Firbas 1971, 1992; Daneš 1987). Discourse in English is segmented into tone units, a process referred to as tonality (Halliday 1967a: 211). Each tone unit realises an information unit (Halliday 1967b). Focus is marked through tonicity, that is, the placement of the nuclear accent, i.e. the main pitch accent, within the tone unit. (3) is for instance uttered as two separate units, separated by //, each containing a nuclear accent, on *yesterday* on *grandfather* respectively. Because they bear a nucleus, these constituents are introduced to the hearer as focal new information worthy of attention.

(3) I sold my watch **YESTERDAY**. // It belonged to my **GRANDFATHER**. //

Interestingly, an item like *yesterday*, which is “interpretable only by reference to today” (Halliday 1967: 206) can take on a focal status and hence be coded as new, as is the case in (3), despite being deictic in nature. Halliday (1967a) explains this in terms of contrastiveness (see Section 1.1.3.1). What this reveals is that the elements of the information unit presented as focal show some, but by no means complete or straightforward, correlation with what is actually given or new in the discourse (Halliday 1967a: 208, 1994: 368). Thus, as far as English is concerned, referential and relational information structure present some overlaps which shows at least a partial dependency between one another.

In French, the prosodically coded distinction between focal and non-focal is said by some to be encoded through accentuation (Di Cristo 1998, 1999; Beyssade et al. 2009; Mertens 1993, 2006) and others through intonational phrasing (Féry 2001). Beyssade et al. (2009), who argue for an accentual description of focus, demonstrate that focused constituents typically bear a nuclear pitch accent on their right edge and display information highlighting through a combination of initial accent and pitch contour. On a pragmatic analysis such as Lambrecht’s (1994), the assignment of a nuclear pitch accent allows the speaker to indicate which information is new with regard to the assertion. An example of this is given in (4). As the answer to the partial question in (4a), the object *des pommes* in (4b) is the

logical focus carrier of the sentence. On Beyssade et al.'s (2009) analysis, *pommes* would be expected to carry a pitch accent at its rightward boundary. If it does, then the PP occurring immediately after will be deaccented. By contrast, if focus is only marked through intonational phrasing and the nuclear accent is located at the end of the unit, the post-focal material will not be deaccented.

- (4) (a) Qu'as-tu acheté au marché?
'What did you buy at the market?'
(b) J'ai acheté [**des pommes**]_{NPA} avec Mathilde.
'I bought apples.'

Given the central position of the notion of focus in this study and given the plurality of the descriptions offered in the literature, I come back to it in greater detail in Section 1.1.3. There, I explain the different approaches adopted to describe its marking in English in Sections 1.1.3.1 and 1.1.3.2 and in French in Section 1.1.3.3.

Along with the binary distinction between referential and relational information structure, information structure also encompasses the dichotomy between focal/non-focal material and topic/comment which, within the pragmatic approach, both relate to the "relative predictability vs. unpredictability of the relations between propositions and their elements in given discourse situations" (Lambrecht 1994: 6). In other words, the notions of focus and topic are here viewed as pertaining to the relation between the elements of a given utterance and the propositional content. The topic is broadly defined as what the sentence is about and the comment as what is predicated about the topic (Reinhart 1981; Lambrecht 2004; Gundel 2003). Although less disputed than the given/new and focal/non-focal contrasts, the functions of topic and comment are rejected by authors like Halliday (1967a) who considers that they tend to be wrongly conflated with newness and givenness. To avoid this, he labels the roles as theme and rheme. In terms of sequential organisation, the theme "is assigned initial position in the clause, and all that follows is the rheme" (Halliday 1967a: 212). This is illustrated with (5), repeated below.

- (5) [I]_{TOPIC/THEME} [sold my watch yesterday]_{COMMENT/RHEME}. [It]_{TOPIC/THEME} [belonged to my grandfather]_{COMMENT/RHEME}.

Functionally, the theme is viewed as the point of departure of the message, which is different from the traditional notion of topic. The theme may coincide with given information but the two need not concur.

Thus, information structure subsumes different distinct, yet interrelated, phenomena for which multiple approaches exist in the literature. This particular study is concerned with the interaction between referential, i.e. given/new, and relational, i.e. focal/non-focal information structure. I take referential information structure to be operationalised by discourse-familiarity (Kaltenböck 2005) and coded by the cohesive relations which are continuously constructed in the text. I assess its coding in the data with a strictly textual approach which departs from Gundel's (2003) and Prince's (1981) cognitively-oriented accounts (see Section 2.4). For relational information, I adopt Halliday's (1967a, and Matthiessen 2014) functional approach in which it is viewed as the manipulation of prosodic features of the spoken linear string (O'Grady & Bartlett 2019) and as relating to focus marking. I investigate focus assignment by carrying out a combination of instrumental and auditory analysis (see Section 2.3). Finally, topic/comment, which, like prosodically-coded focus, is part of the functional organisation, is coded by the different grammatical constituents of the clause. Because clefts are generally viewed as focus-marking constructions, its operationalisation in discourse will not be explored in the present study.

1.1.3 The notion of focus

In the following two sections, I present and clarify my position towards the different approaches to the notion of focus in English and French. Section 1.1.3.1 focuses on the functional approach in the Hallidayan tradition, which I adopt for the analysis of English clefts, Section 1.1.3.2 on the formal-pragmatic approach defended by Lambrecht (2001), and Section 1.1.3.3 on the treatment of focus in French.

1.1.3.1 The functional approach

In the Hallidayan (1967a,b, 1994, and Greaves 2008, and Matthiessen 2004, 2014) functional tradition, intonation is viewed as the main coding means by which English speakers structure information into focal and non-focal information (Tench 1996; O’Grady 2013). English speech progresses as a succession of melodic units, which, according to Halliday (1967a: 202) “represents the speaker’s blocking out of the message into quanta of information, or message blocks”. Each tone unit realises an information unit (Halliday 1967b), or what Cruttenden (1997) calls a presentation unit. While information units may correspond to a clause, they frequently do not do so (O’Grady 2014b). The segmentation is done by marking off units by melodic contours (Halliday 1963), which are falling or rising tones (transcribed by \ and /), or combinations of these. Information focus is coded by the placement of the nuclear accent on a specific syllable of the tone unit, which “carries the main pitch movement” (Halliday 1994: 296), i.e. the tonic syllable. The domain of the information focus is typically not just the tonic syllable as such but the larger constituent it is part of (Halliday 1967a: 204).

The most typical, or unmarked, information structure starts off with recoverable information and continues with non-recoverable information, which partly or wholly coincides with an information focus marking the most salient new information, whereby ‘new’ means information ‘freshly’ introduced in the discourse (Halliday 1994: 295-299). An unmarked information focus occurs unit-finally, that is to say, it is marked by a tonic on the last lexical element of the information unit. The actual domain of the focus (underlined in the examples) is the constituent containing the tonic syllable, like the whole NP with its NP-internal restrictive relative clause in // I’m looking for the caretaker who looks after this block // (Halliday 1967a: 207). The unmarked information structure tends to have a left to right form of organisation with given information, if present, preceding new information (Halliday 1967a: 205). If the lexical element is followed by an anaphoric grammatical constituent, then it is not part of the focus because it is inherently recoverable, e.g. // I saw her //. What precedes the focal domain may be entirely recoverable, or it may be recoverable shifting into non-recoverable information without being focal, as is the case in the caretaker example, where / is situationally

recoverable, but *'m looking for* is not. This creates a certain indeterminacy. This is, in fact, why unmarked focus “may be ambiguous” (Halliday 1967a: 208), and why a second subtype needs to be distinguished, where the whole unit contains non-recoverable information, as is often the case at the beginning of a story. In this subtype, the focal domain is “the whole of the information unit” (Halliday 1967a: 208).

The marked type of information structure, then, contains a focus that is “informationally contrastive [...] within a closed system or lexically” (Halliday 1967a: 207). The notion of ‘contrastive focus’ subsumes both contrast with, and addition to, another option from a finite set (Halliday 1967a: 226). This marked focus is, through its shared set membership with another element, semantically cohesive (Halliday 1994: 295-299). A marked focus always relates to a block of presupposed information (Halliday 1967a: 206). This entails that marked information structures imply a question pertaining to one specific constituent like a *wh*-question or an echoic polar question (Halliday 1967a: 207-211). By contrast, unmarked information structures correspond simply to general questions like *what is happening?* (Halliday 1967a: 208). A marked focus need not, and often does not, form the final element of the tone unit. The domain of a marked focus relation may be a clause constituent, as in // John painted the shed yesterday //, which implies the *wh*-question *What did John paint yesterday?*, which presupposes that John painted something yesterday (Halliday 1967a: 207-208). Contrastive focus may also be a constituent of a phrase, as in // I’ve seen better plays //, which presupposes that the speaker has seen at least some plays, or a constituent of a word, as in // the damage was only external//, which presupposes that there was some sort of damage (Halliday 1967a: 208). Finally, marked focus may also involve contrast with options of grammatical systems, as in // he d\id take it //, which implies the echoic polar question // d\id he take it // (Halliday 1967a: 211).

From a functional perspective, the broad distinction between unmarked and marked information structure is conceived of by Halliday (1967a) in terms of presenting vs. contrastive focus, both of which can be summarised as follows. A presenting focus does not mark any information as presupposed. It always falls on the last lexical constituent of the information unit, which it marks as the most salient new, without specifying the information status of the remainder, which, at

the beginning of discourse may be wholly new. A contrastive focus relates to presupposed information, which may precede and/or follow the focus. This focus is informationally contrastive and relates to information presented as presupposed from the preceding discourse, i.e. as recoverable “anaphorically, by reference, substitution or ellipsis” (Halliday 1967a: 206). The possibility of a contrastive focus relating to an elided but recoverable presupposition is important for research on clefts as it allows speakers to anaphorically presuppose the variable in reduced clefts.

The two types of information focus⁸ are exemplified in (6), in which the saleslady of a silver department is prepping a new job student, Anne. The different pitch movements on the nuclei are indicated as falling (\) or rising (/) or a combination of both, and the tone unit boundaries are marked by double slashes. In the first information unit, which consists of a prepositional phrase only, *this* has a marked, non-final, information focus. Contrastive *this* points exophorically to the context of situation and “signal[s] the taken-for-grantedness that Anne is there to do a job” (Halliday 1994: 369). The second information unit presents information as corresponding simply to an implied question like ‘what is happening?’ (Halliday 1967a: 208) and introduces *silver* as an unmarked final focus.

(6) // ^in thV\is job Anne we're // working with s\ilver // (Halliday 1994: 368)

As illustrated in (6), what the elements of the information unit present as focal, presupposed, etc., shows some, but by no means complete or straightforward, correlation with what is actually given or new in the discourse. Halliday (1994: 301)

⁸ The basic contrast just outlined between unmarked and marked focus is referred to in É. Kiss (1998: 249) as a “neglected distinction”, which yet has “been present in the literature for a long time (see, for example, Halliday 1967; Rochemont 1986), although the interpretations attributed to the two focus notions (variously called CONTRASTIVE FOCUS versus PRESENTATIONAL FOCUS, NARROW FOCUS versus WIDE FOCUS, or, in [É. Kiss’s] article, IDENTIFICATIONAL FOCUS versus INFORMATION FOCUS) have not always been exactly the same”. In the remainder of this study, I will use the terms ‘contrastive focus’ versus ‘presenting focus’.

stresses that speakers, within given discursive conditions, may “play with the system (...) to produce an astonishing variety of rhetorical effects”.

My approach to information structure takes the Hallidayan model further in two ways. Firstly, the rather static model according to which information structure is ‘mapped onto’ transitivity and mood structure is replaced by the more dynamic model of speakers balancing grammatical and prosodic choices moment by moment in real time (O’Grady 2010, and Bartlett 2019). The choices are shaped both by the nature and extent of information shared between speaker and hearer and by their awareness of communicative purposes. Utterances are produced in real time as part of a textual chain which moves the discourse forward while simultaneously grounding the discourse in the shared context.

Secondly, the investigation of information management is extended beyond the sentence to larger discursive (dialogic) contexts. Initial and final positions of intonational units are especially important as they signal relations with the previous unit and anticipate hearer response to the upcoming update of the common ground (O’Grady 2010, 2014a). Onsets in particular have been shown to serve as an interactional device to signal how the upcoming material relates to the previously generated expectations (O’Grady 2013, 2014a). The onset is the “first prominent syllable in a tone unit” (O’Grady 2014a: 691), which indicates the ‘key’, i.e. the pitch level at which the current utterance starts. Brazil (1997) classified key choices as high, mid, or low relative to the height of the previous onset, which may be part of discourse by the same or by the previous speaker. A high onset indicates a change or disruption in the discourse such as the introduction of a new topic or disagreement, while a low onset signals that the content of the following unit is equivalent to the previously generated expectations. A mid onset is the unmarked option and projects that the upcoming unit is not contrary to expectations. Studying a construction like *it*-clefts with this broader view of information management allows us to link their internal information structure to their larger discursive-interactional context.

1.1.3.2 The formal-pragmatic approach

For Lambrecht (1994: 5), information structure is concerned with how propositions as conceptual representations of states of affairs are paired with lexicogrammatical structures that interlocutors can interpret as pragmatic units. Focus is defined as the “component of a pragmatically structured proposition whereby the assertion differs from the presupposition” (Lambrecht 2001: 474). The presuppositions are the set of propositions lexicogrammatically evoked by a sentence which the speaker assumes the hearer already knows or is ready to take for granted, “which is more or less equivalent to the notion “hearer-old” in the system of Prince (1992)” (Lambrecht 2001: 474). As noted by Lambrecht (1994: 52), the grantedness of the presupposition can be brought out by Ertheshik-Shir and Lappin’s (1979, 1983) “lie test”, whereby challenging a proposition with a response like ‘That’s not true’ only entails challenging the focus while the presupposition still remains granted. The assertion is the proposition the speaker expects the hearer to know or take for granted as a result of hearing the utterance, i.e. the ‘new information’. The focus is the denotatum that makes the utterance into an assertion and is “by definition an unpredictable part of the proposition” (Lambrecht 2001: 474), which explains why, with a few motivated exceptions, the focus constituent “necessarily requires an accent” (Lambrecht 2001: 479).

Verstraete (2007: 81) has pointed out the *prima facie* similarity between the three types of focus distinguished by Halliday (1967a) and Lambrecht’s (1994, 2001: 485) ternary typology of predicate-focus, sentence-focus and argument-focus. In cases where the information unit as defined by Halliday coincides with a clause, his and Lambrecht’s three types of focus may in practice roughly coincide. Halliday’s unmarked information structure whose focus presents the most salient new resembles Lambrecht’s predicate focus, and the structure whose focus domain is formed by the whole information unit resembles sentence-focus. Halliday’s marked information structure with a contrastive/additive focus relating to a presupposed message block resembles Lambrecht’s argument focus. However, there are fundamental differences between the pragmatic and functional approaches to information structure which I detail in what follows.

First, the pragmatic approach takes the pragmatically structured proposition, rather than the prosodically coded information unit, as the basic unit, which entails the importance of the lie test previously mentioned for recasting new information as a pragmatic assertion. In the case of constructions like clefts, this creates the expectation that only one constituent is focal while the rest of the construction is merely presupposed material. As will be further discussed in Section 1.2.2, this does not align with empirical evidence brought out by a number of studies. Another difference between the formal-pragmatic approach and the functional one lies in the treatment of prosodic accents. On Lambrecht's (2001) analysis, accents are treated as a possible corollary and not the intrinsic realisation of focus. Finally, the formal-pragmatic approach also correlates components, like presuppositions, more or less straightforwardly with discourse-old and discourse-new information in the cognitively oriented sense of Prince (1992) while the Hallidayan tradition does not make such claims. In placing itself within the functional philosophy, the present study treats focus as being marked in English by prosody and as being situated within the boundaries of the information unit.

1.1.3.3 Focus marking in French

There exist competing views on the encoding of focus in French, which this section aims at clarifying. But before delving into the description of the different approaches, I first start by covering the basic notions of the French intonation system.

With regard to speech segmentation, French follows a three-tiered model in which utterances consist of one (or more) Intonation Units (marked by []), which are themselves divided into Prosodic Words (marked by | |), which are further split into Tonal Units (marked by |) (Di Cristo 1998).⁹ An illustration of the phonological hierarchy is given in (7).

⁹ The terminology used to describe the different prosodic units of French varies significantly in the literature. Other names for the intonation unit include Intonational Phrase (Jun and Fougeron 2000), Groupe Intonatif (Mertens 1987, 1993, et al. 2001) and Breath Group (Vaissière 1991). The prosodic word is also labelled Groupe Clitique (Di Cristo 2016), Accentual Phrase (Jun and Fougeron 2000) or

- (7) [Mon FILS | | et son voiSIN] [se sont DIS | puTÉS]. (Di Cristo 1998: 199)
'My son and his neighbour had an argument.'

Intonation units are the highest phonological constituent, followed by prosodic words and tonal units. As shown with (7), the hierarchy is directly reflected in the relative length of each segment as well as in the syntactic constituents each accepts. The list of syntactic segments which can form intonation units include any major syntactic constituent, with the exception of clitic subjects (Di Cristo 1998: 214). The intonation units of (7) are the NP *mon fils et son voisin* and the VP *se sont disputés*. Di Cristo (*ibid.*) notes that the segmentation into intonation units is not language-dependent, i.e. not determined by the inherent principles of a language system, but rather tied to the speaker's choices and/or the situational context. Conversely, the division of utterances into prosodic words and tonal units derives from phonological and phonotactic constraints specific to French. Prosodic words, which represent the second highest phonological category, correspond to the domain of the primary stress and as such include the accented syllable and the preceding syllables up to the previous primary accent, e.g. *mon fils* and *et son voisin* in (7) (Di Cristo 2016). Tonal units, which are the minimal unit, are made of any accented syllable - with a primary or secondary accent - combined with the preceding syllables up to the previous accent, e.g. *se sont dis* and *putés*. In the remainder of this study, only intonation units will be considered and annotated for in in-text examples.

The description of prosodic phrasing in French highlights the existence of two types of basic accents: primary accents and secondary ones. For Mertens (1993) there only exists one primary accent which is the lexical accent. The lexical accent, also called the final accent, is the rhythmic stress located in the last full syllable of a prosodic word, e.g. *FILS*, *voiSIN*, *dispuTÉS*. In contrast with English, French does not have inherent word stress and therefore does not have the same rigid constraints on accent placement within the word (Quirk et al. 1985: 1590). Given

Groupe Accentuel (Mertens 1987, 1993, et al. 2001). Finally, the Tonal Unit is comparable to Di Cristo's Pied (Di Cristo 2016).

its mandatory nature, some authors have described it as a fixed rhythmic stress (Garde 1968: 137). The final accent is realised by means of pitch and temporal cues. It typically triggers a significant lengthening of the accented syllable as well as the insertion of a pause after it. In the case of phrase-final tonal units, it generally coincides with the nuclear accent and is therefore accompanied by the main pitch movement (Mertens 1987: 82, 1993: 4). Di Cristo (1998, 2016) describes the nuclear accent as the highest prominence of the intonation unit comparable to a ‘sentence accent’ (Di Cristo 2016: 273). Unlike the lexical accent, the nucleus can be moved within the unit to create a contrast or highlight an ‘apport informatif’ (Di Cristo 2016: 274). Besides the final accent, French also offers the possibility to encode an optional secondary stress typically carried by the initial syllable of a content word, e.g. DISputés, (Hirst and Di Cristo 1984: 561; Di Cristo 1998: 197). The secondary accent is weaker than the primary one, with regard to both the duration and pitch of the accented syllable (Jun and Fougeron 2000). Unlike the final accent, the initial accent can only be realised with a static tone void of any illocutionary pitch movement (Mertens 1993: 4). Despite being the markers of leftward edge for the initial accent and rightward edge for the final accent, the two types of prominences do not function wholly independently of one another. The initial accent may, in some cases, either reinforce or modify the boundaries set by the final accent (Simon 2004: 295) or further mark the focal status of a constituent (Di Cristo 2016: 271).

Focus-marking in French has been accounted for in several ways. In a first approach, Jun and Fougeron (2000) link focus-marking to tone movement. The encoding of focus is linked to the insertion of a rising pitch movement characterised by a large peak located either on the initial syllable (Hi) of the focused word or on its final syllable (H*). In the declarative sentence in (8), the focal status of the word *mangera* is, for instance, marked by the noticeable F0 peak on its last syllable.

- (8) Marion **MANGERA** au petit-déjeuner des bananes. (Jun and Fougeron 2000: 227)
‘Marion will eat bananas at breakfast.’

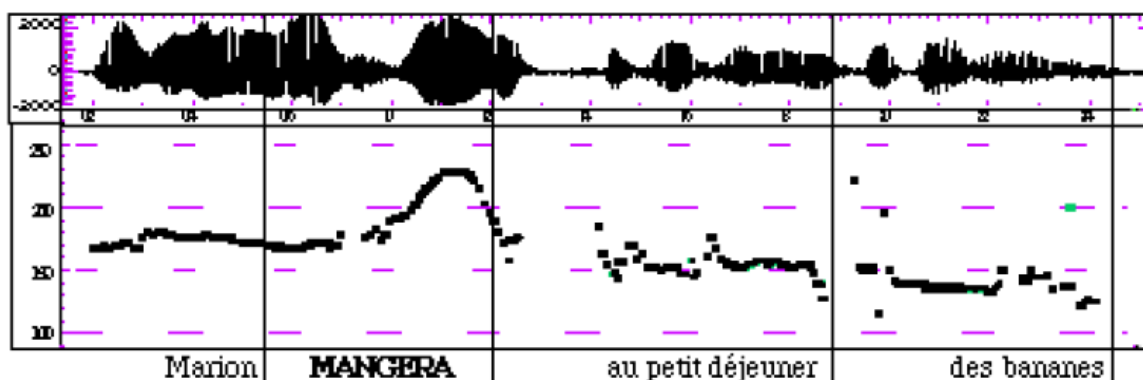


Figure 1. Prosodic realisation of (8) (Jun and Fougeron 2000: 227)

The pitch curve of (8), displayed in Figure 1, also shows that the syllable found immediately after the focused item, which is typically from the next word but sometimes within the focused word itself, displays a fall (L) ensuing from the focus peak. As far as the post-focal sequence is concerned, Jun and Fougeron's (2000) argue that, when part of the same intonation unit as that of the focused constituent, it is systematically deaccented and displays a plateauing of the pitch curve. It is the case of the segment *au petit-déjeuner des bananes* in (8) which does not exhibit any tonal movement and is produced with a low register. The post-focal sequence, however, typically maintains the same durational cues, i.e. relative duration and lengthening, as non-focused sentences, thus showing that the post-focal material is deaccented but not dephrased. Interrogative sentences follow the same pattern but with a steep rise on the final syllable of the focused constituent followed by a high plateau instead of a low one. Jun and Fougeron (2000) link this flat plateau to the phrasal tone of an intermediate phrase, whose phonological position lies between that of the intonation unit and the accentual phrase. The tone is not bound to a specific syllable but instead spreads over the whole post-focal sequence whose rightward boundary is accompanied by a boundary tone marking the end of the Intonation Unit. Thus, Jun and Fougeron (2000) treat focus as a primarily tonal characteristic which also interacts with boundary formation.

Di Cristo (1998, 2016) attributes focus-marking in French to accentuation. For him, focus is marked by the combination of the nuclear pitch accent, which is by default located at the end of the intonation unit, and the illocutionary contour. When the accent occurs in its default position, the focus can be described as broad. When it is moved to an earlier position in the unit, the focus becomes narrow and

it may induce a contrastive reading, though Di Cristo (2016: 269) argues that contrast mainly stems from contextual cues and not prosodic marking. The focused constituent may at times be further highlighted with the use of a double marking ('double balisage') (Di Cristo 2016: 271) whereby the constituent displays prosodic marking in both its initial and final syllable. It is especially useful in the case of focus operators such as exclusive particles which readily take on an initial accent. In addition to the movement of the nuclear accent, Di Cristo (1998, 2016) also relates focalisation to the use of a specific emphatic accent¹⁰ which can be of 'insistance' or 'contraste' depending on its pragmatic function. The intensification accent is typically located on the initial syllable of the word the speaker wishes to lay narrow emphasis on and triggers an extra pitch prominence. In spoken discourse, it is frequently found in adverbs and clitics (Di Cristo 1998: 209). The contrastive accent, as its name suggests, serves to mark a given item as contrastive through a global rising-falling tonal movement. When the contrast is objective, the rising movement may occur on any of the syllables of the focused item, except for the last. In the case of an expressive contrast, the rise is located on the last syllable. In both types of contrast, the fall always occurs on the last syllable. As far as segmentation is concerned, Di Cristo (1998: 210) claims that the focused constituent can be analysed as "an autonomous Intonation Unit embedded within a higher-level intonation unit constituting the utterance as a whole". The post-focal sequence, on the other hand, lies outside of the unit boundaries and is realised with a compressed register and no accentuation or significant tonal movement. Di Cristo's (1998, 2016) description agrees with that of Jun and Fougeron (2000) on all but one point. In both accounts, focus is characterised by a rising-falling movement borne by the focused item. The focused syllable is immediately followed by a fall and then by the deaccented post-focal tail. The two approaches thus both treat focus-marking as being intonational in nature. Jun and Fougeron's (2000) and Di Cristo's (1998, 1999) accounts however appear to diverge with regard to the question of phonological phrasing. While Jun and Fougeron (2000) accept that the post-focal

¹⁰ Other terms describing the focus-inducing accent in French include 'accent d'insistance' (Blanche-Benveniste (1990) and 'accent de focalisation' (Rossi 1999).

sequence may be uttered in its own Intonation Unit or as part of that of the focused constituent, Di Cristo (1998) systematically analyses it as a separate intonation unit.

In contrast with Sun and Fougeron (2000) and Di Cristo (1998, 1999), Féry (2001) rejects the postulate that focus-marking in French is primarily tied to accentuation and instead credits it to phrasing. Put simply, Féry (2001) argues that any focused constituent, in the case of narrow focus, is realised in a separate intonation unit while post-focal material is typically dephrased¹¹, though some speaker variation was found in the results of her experimental study. The prototypical phrasing pattern is instantiated in (9) below. Here, (9b) comes in answer to the question (9a) which is meant to induce a narrow focus in Féry's (2001) experimental setting. As corroborated by the F0 in Figure 2, the focused subject *les deux enfants* is phrased in a separate unit while the remainder of the sentence is uttered with a low and flat pitch.

- (9) (a) Qui va à l'école en vélo?
'Who goes to school by bicycle?'
 H L
- (b) [Les deux enfants] vont à l'école en vélo
'The two kids go to school by bicycle.'
 H L H L
- (c) Ce sont [les deux enfants] qui vont à l'école en vélo (Féry 2001:
 21)
'It is the two kids that go to school by bicycle.'

¹¹ Féry (2001) uses the term dephrased to mean constituents which show an absence of pitch variation and phrasing correlates. As she notes, "in an analysis requiring exhaustive prosodic parsing, [dephrased material] should be integrated into a PhP [Intonation Unit] to their left or right" (Féry 2011: 201).

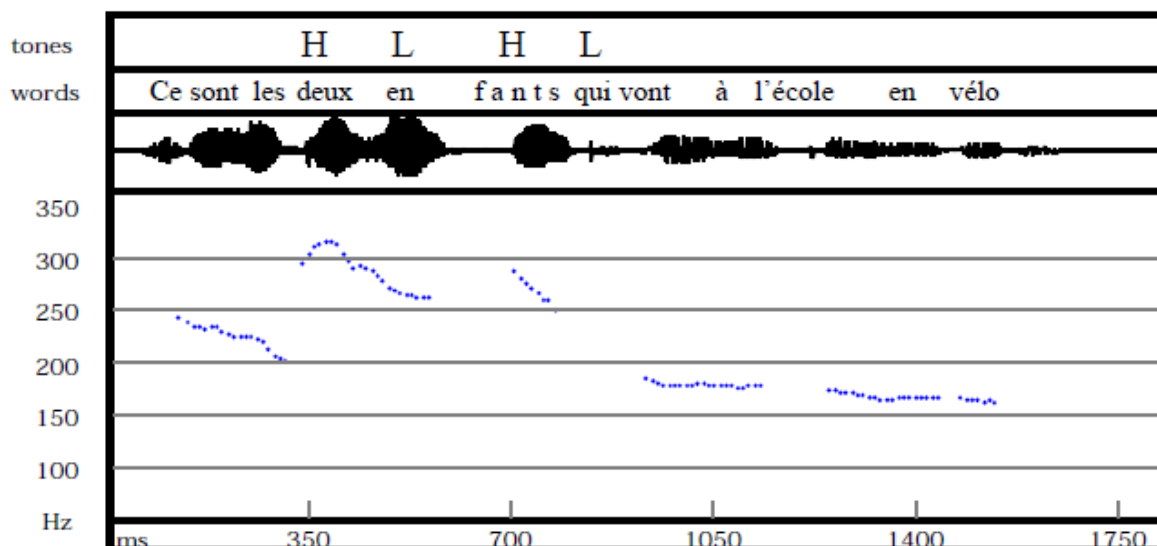


Figure 2. Prosodic realisation of (9c) (Féry 2001: 21)

Interestingly, the question in (9a) also led to the use of a cleft such as (9c) for some participants. There again, the constituent in focus forms its own intonation unit while the *qui*-clause is dephrased. The comparison of (9b) and (9c) reveals the existence of variation in tonal realisation occurring at the edges of the focused constituents. Thus, unlike Jun and Fougeron (2000) and Di Cristo (1998, 1999), Féry (2001) argues that focus phrasing is not associated to a single tone movement. Féry (2001) further notes that the pre-focal constituent is generally forced into its own intonation unit and may as a result display a boundary tone marking the transition from one unit to the other. In view of the phrasing and tonal behaviour of French, Féry (2001) argues that the pitch variations and temporal cues observed at the right edge of focused constituents are simply the corollary of phrasing and not the other way round. More generally, she rejects the general assumption that French, like English, has stress. This idea has led other authors like Vaissière (1991: 118) to argue that French, in comparison with English, is better described as a ‘boundary language’ than a ‘stress language’.

From a general perspective, the apparent mapping of prosodic units onto syntactic ones within the French intonation system¹² has raised the question of the

¹² The brief overview of the French intonation system provided here is not meant to exhaustively characterise its complex functioning. For a more complete account see Mertens (1987), Rossi (1999), Simon (2004) or Di Cristo (2016, 2019).

congruence between the two systems. On this, Rossi et al. (1981) claim that although interacting with one another, syntax and prosody are not fully congruent. For them (Rossi et al. 1981: 249), a perfect congruence between the syntactic and intonational hierarchies would cancel the *raison d'être* of the latter which would instead be marked syntactically. Incongruence can, for example, be observed in the case of non-coplanar utterances (Martin 1978), i.e. utterances with constituents of different syntactic ranks, or multiple embedded clauses. Mertens (1993) further corroborates this idea with a number of examples in which intonation units spill over more than one syntactic constituent, and cases of deaccentuation of syntactically salient constituents. For him, syntax and prosody cannot entirely concur in so far as some of the limits of the latter, e.g. limited number of tones, do not allow to fully transcribe the complexity of the former (Mertens 1993: 44). Féry (2001: 9) also appears to defend this argument by assuming that the two systems are “processed in parallel” rather than built on top of each other. Against this, Delais-Roussarie et al. (2011) argue that apart from minor incongruencies caused by phonological constraints, the boundaries of syntactic and prosodic units tend to overlap, thus supporting the idea that there is prosodic alignment stemming from syntax. For this study, I follow Di Cristo (1981) and Mertens (1993) and treat the syntax and prosody of French as interacting, but not fully congruent, layers. I reflect on the issue of the syntax-prosody interface in French with empirical data in Section 4.4.1.

From a cross-linguistic viewpoint, very few studies have delved into how French encodes focus in comparison with English. The only in-depth comparative study available, which was carried out by Vander Klok et al. (2018), point to a number of interesting findings. First, Vander Klok et al.’s (2018) experimental results show that French, like English, makes use of focus shifting, i.e. moving focus from its unmarked position, as had been suggested by Di Cristo (1998, 2016) and Jun and Fougeron (2000). However, French differs from English in that focus shifting is limited to cases of corrective focus only while other contexts, such as parallelism or contrastive focus, do not induce prominence shifting. Vander Klok et al. (2018) attribute this discrepancy to the syntactic scope of focus in French which appears to be restricted to speech-act-sized constituents while that of English shows a lot more variation. With regard to prosodic cues associated with focus-marking, the

initial accent was not found to be of significant use either in English or French, thus contradicting Di Cristo's (1998) claim for French.¹³ Another finding of Vander Klok et al.'s (2018), which is especially relevant for the present study, suggests that French and English differ with regard to the selection between prosodic strategies and syntactic ones such as clefts to mark focus, but that the choice between one or the other might be not fully inter-dependent as had been previously claimed (Lambrecht 2001). In other words, when it comes to focalisation, neither French nor English speakers appear to make 'either or' types of choices between syntactic and prosodic means. Thus, focus-marking in English and French is only comparable to a certain extent.

1.2 The cleft construction: an overview

This section provides an overview of the literature on *it*- and *c'est*-clefts relating to their formal and semantic properties (Section 1.2.1), information structure (Section 1.2.2), and discourse functions (Section 1.2.3).

1.2.1 General overview of *it*- and *c'est*-clefts

Ever since Jespersen's (1927) defining work on the syntax of modern English, cleft constructions have remained at the forefront of linguistic studies across languages. Studies on English *it*-clefts have ranged from formal and functional accounts (see among others Declerck 1984; Geluykens 1988; Delin 1990; Hedberg 1988, 1990, 2000, 2013, and Fadden 2007; Collins 1991, 2006; Weinert and Miller 1996; Lambrecht 1994, 2001; Nelson 1997; Reeve 2011; Büring and Križ 2013; Destruel and Velleman 2014) to prosodic accounts (Herment and Leornarduzzi 2012), and have also included diachronic descriptions (Filppula 2009; Patten 2012). The same diversity of research is available for French *c'est*-clefts with studies focusing on their formal and discourse properties (e.g. Katz 2000b; Doetjes et al. 2004; Carter-

¹³ This does not mean that onsets are not functionally significant either in English or French, but rather that their role is not tied to the encoding of the information structural phenomenon of focus.

Thomas 2009; Dufter 2009a; Destruel 2012; Scappini 2013; Roubaud and Sabio 2015, 2018; Lahousse and Lamiroy 2017; Destruel et al. 2019), their prosody (Rialland et al. 2002; Doetjes et al. 2004; Hamlaoui 2010; Avanzi 2011; Mertens 2012), and their diachronic evolution (Muller 2003; Bouchard et al. 2007; Rouquier 2007; Dufter 2008, 2009a, 2009b). Because of their seemingly similar syntax and use, there also exists a number of cross-linguistic studies between English and French clefts - or more generally Germanic and Romance languages - aiming at uncovering the extent of their comparability (e.g. Miller 1996; Carter-Thomas 2002; Lahousse and Borremans 2014; Dufter 2009a; Bourns 2014; Bourgoin 2017).

Before delving into the key aspects of *it*- and *c'est*-clefts, I start by detailing in the next two sections the competing views on the syntax of clefts as well as the prevailing rationale for their existence.

1.2.1.1 The constituents of the cleft

Cleft constructions are bi-clausal constructions consisting of a matrix clause introduced by the sequence *it/ce* + copula *be/être* followed by a cleft-relative clause. Both *it*- and *c'est*-clefts, in their prototypical declarative form¹⁴, follow the same syntactic order, as exemplified with (10) and (11) from the LLC-1 and CRFP below. In both languages, the cleft may also be reduced, as instantiated in (12) and (13).¹⁵ In this case, the cleft-relative clause is omitted primarily due to its easy recoverability. The omission of the cleft-relative clause is symbolised by strikethrough text formatting.

(10) it was Badger himself that took it (LLC-1)

(11) c'est la grappe qui se détache (CRFP)

'it's the grape that is coming off'

¹⁴ By contrast, the syntactic structure of interrogative clefts differ in both languages, as shown with (10') and (11').

(10') Is it your sister or your brother who's older?

(11') Est-ce que c'est ta sœur ou ton frère qui est plus vieux ?

'Is it [the case] that it is your sister or your brother who is older?'

¹⁵ I delve into the literature on reduced clefts in greater detail in Chapter 6.

- (12) even afterwards I don't mind but it's beforehand [~~that I mind~~] (LLC-1)
- (13) elle elle prend ses fleurs en Hollande mais nous c'est que des fleurs de France [~~qu'on prend~~] (CRFP)
- 'she she buys her flowers in Holland but for us it's only flowers from France [~~that we buy~~']

Because of their syntactic structure, clefts resemble other types of copular clauses introduced by *it/c'est* including extraposed subjects (Calude 2008), *it/c'est*-clauses with a restrictive relative clause and *it/c'est*-clauses with an appositive relative clause (Rouget and Salze 1986; Huddleston and Pullum 2002; Rouquier 2007; Laurent 2018). On the 'non-derivational' structural analysis argued for in Davidse (2000) and Huddleston and Pullum (2002), which I adhere to in this study, clefts are differentiated from other types of *it/c'est*-clauses based on the nature of their cleft-relative clause. Unlike *it/c'est*-clauses with a restrictive or appositive clause, the relative clause of clefts is not syntactically part of the subject *it/c'e* and does not form a constituent with its antecedent (Lambrecht 2001; Huddleston and Pullum 2002: 1416; Muller 2003; Rouquier 2007). Importantly, the antecedent of the cleft-relative clause is the full complement NP, e.g. *the boy* in (14), in contrast with restrictive relative clauses, whose antecedent is the nominal head, e.g. *boy* in (15). As a consequence, the antecedent of the cleft-relative clause refers to a fully determined instance, viz. *the boy* whom the hearer is presumed to be able to identify in (14) (Davidse 2000: 1107-1113). By contrast, the antecedent of a restrictive relative clause designates a mere entity-type. Its combination with the restrictive relative clause creates a contextually unique description, *boy who caused all the trouble*, which can, therefore, receive definite determination.

- (14) Who caused all the trouble? – It was [the boy] [who caused all the trouble]. (Huddleston 1984: 460)
- (15) Who was that on the phone? – It was [the [boy who caused all the trouble]]. (Huddleston 1984: 460)

The relative clause of *it*- and *c'est*-clefts thus displays features which are distinctive to clefts, which is why it will be referred to as the cleft-relative clause in the remainder of this study.

As far as the subject of clefts is concerned, *it* and *ce* are either analysed as semantically empty expletive pronouns (Chomsky 1971; Akmaijan [1973] 1979; Collins 1991; É. Kiss 1998; Lambrecht 2001; Huddleston and Pullum 2002) or as pronouns carrying some sort of semantic content (Bolinger 1972; Gundel 1974; Davidse 2000; Hedberg 2000; Reeve 2011). Bolinger (1972: 108) argues, for instance, that the pronoun functions as a full-fledged lexical item which carries low information but encodes identity. Davidse (2000) likewise argues that it construes the postcopular complement as 'identifier' (Halliday 1967a: 224) but that it does not refer to a specific constituent. For Hedberg (2000) and Reeve (2011), *it* is referential and forms a pragmatically discontinuous definite relation with the cleft-relative clause. Their account is a modified version of the syntactic extraposition one, which claims that the relative clause modifies subject *it*, with the relative clause extraposed from its antecedent (Jespersen 1927; Akmaijan [1973] 1979; Gundel 1974; Patten 2012), in spite of the fact that neither diachronically nor synchronically a non-extraposed variant has ever been attested, as recognised by Patten (2012) (see Section 1.2.2).¹⁶ In this study, I subscribe to the analysis developed by Davidse (2000) and consider *it/ce* to be non-referential pronouns which assign the role of identifier to the postcopular complement. To refer to the different constituents of the cleft, I will use the following terms:

- (16) pronoun *it/ce* + copula *be/être* + clefted constituent + cleft-relative clause
- (17) pronoun *it/ce* + copula *be/être* + clefted constituent + [~~omitted cleft-relative clause~~]

Since reduced clefts do not have an overt cleft-relative clause, its omission will be marked by the crossing out of the reconstructed cleft-relative clause.

¹⁶ Authors who ascribe meaning to *it* but who reject the extraposition analysis include Davidse (2000).

1.2.1.2 The rationale of the cleft

For some authors, clefts are primarily focus-marking constructions (Lambrecht 1994, 2001; Krifka 2008; Lehmann 2008) which derive from a corresponding monoclausal counterpart whose information is repackaged into the bipartite construction (Huddleston and Pullum 2002; Collins 2006). In this view, an example like (18a) is said to convey the same propositional meaning as the canonical sentence in (18b) whose information is simply reorganised in the matching cleft.

- (18) (a) It was the boy who caused all the trouble.
(b) The boy caused all the trouble.

Through this repackaging, the clefted constituent, which can select a wide range of syntactic categories, e.g. NP, PP, AdvP, VP, etc. (Delahunty 1982; Collins 1991; Weinert and Miller 1996; Huddleston and Pullum 2002), is argued to be given a certain emphasis (Jespersen 1927; Quirk et al. 1985) vis-à-vis the presupposed open proposition in the cleft-relative clause.

Jespersen, however, later revises his earlier account and notes that “emphasis is better given by stress and or intonation” (Jespersen [1937] 1984: 75) which may work in combination with the syntax of the cleft, but can also be encoded without the use of a cleft (see Section 1.2.2 for an account of the prosody of clefts). On Lambrecht’s (2001) analysis, *it*-clefts such as (18a) assert a simple proposition similar to that in (18b), which their bi-clausal syntax constructs into pragmatic components. The postverbal NP in the matrix is the focus, i.e. the element by which the assertion differs from the presupposition expressed by the cleft-relative clause. The distinction between the focus and the presupposition is, therefore, directly reflected by the bi-clausal syntax (Lehmann 2008: 212). According to Lambrecht (2001), the syntax of clefts is dedicated to coding these pragmatic functions. *It + be* in the matrix is a focus marker which unambiguously codes the “predicative argument of the copula” (Lambrecht 2001: 467) as an argument-focus, e.g. *the boy* in (18a), whose predication is coded by the cleft-relative clause, *x caused all the trouble*. By contrast, the non-cleft counterpart (18b) is pragmatically ambiguous between having argument focus, which answers an implied *wh*-question, or

sentence focus, which answers a question like ‘what is the case?’ (Lambrecht 1994: 221-223).

A number of accounts of *it*-clefts project the same description onto *c’est*-clefts and describe them as focalising devices whose primary purpose is to construe the clefted constituent as the focal item of an argument focus structure (Lambrecht 1994, 2001; Lehmann 2008). In the case of *c’est*-clefts, in contrast with *it*-clefts, this argument is heavily based on the widespread assumption that clefts allow speakers to circumvent the SVO order of French which is said to be more rigid than that of other Romance languages such as Spanish or Italian and which is argued to be stricter in written French than spoken French (Sornicola 2011). Because of this strict SVO order, focal objects in French cannot be moved in preverbal position other than by using specific focalisation mechanisms such as left dislocation or clefting. Thus, the object of a proposition like (19a) cannot be placed in initial position to mark its focal status as illustrated in (19b). Instead, the argument should either be clefted as in (19c) or fronted as in (19d). These strategies may be accompanied by prosodic prominence on the constituent. For a subject in initial position to be focal, e.g. *Marie* in (19e), the speaker also needs to resort to a cleft in which the same SVO order is kept but in which the clefted constituent *Marie* is now emphasised through the clefting. Here too, prosody may be used in combination with syntactic reordering to further highlight the clefted constituent. (19f) shows, however, that prosodic means (marked by capital letters) alone cannot readily be used in French to encode focalisation in preverbal subject position. Conversely, because of the freer positioning of prosodic accents in English, the same proposition translated into English (20a) can make use of either of the two strategies, i.e. syntactic or prosodic, to convey the same meaning (Carter-Thomas 2009).

- (19) (a) Marie a mangé le gâteau qui était sur la table.
(b) *Le gâteau Marie a mangé.
(c) C’est le **GÂTEAU** qui était sur la table que Marie a mangé.
(d) Le **GÂTEAU**, Marie l’a mangé.
(e) C’est **MARIE** qui a mangé le gâteau qui était sur la table.
(f) ?? **MARIE** a mangé le gâteau qui était sur la table.

- (20) (a) Mary ate the cookie that was on the table.
(b) Mary ate the **COOKIE** which was on the table.
(c) It is the **COOKIE** which was on the table that Mary ate.
(d) It is **MARY** who ate the cookie which was on the table.
(e) **MARY** ate the cookie which was on the table.

Lambrecht (2001) and Carter-Thomas (2009) argue that it is precisely the availability of more varied strategies to mark focus in English, as opposed to French, which explains the higher frequency of use of *c'est*-clefts over *it*-clefts. Though relying on written material, a number of comparative studies (e.g. Carter-Thomas 2002, 2009; Bourgoin 2017) have corroborated the greater use of *c'est*-clefts.

Against this description of *c'est*-clefts as grammaticalised focus constructions, Dufter (2008, 2009a, b) shows that their use extends beyond focalisation. From a diachronic viewpoint, he justifies this by showing that the increasing restrictions in Old French on focus marking in situ, i.e. through accentual means, and ex situ, i.e. with syntactic rearrangement within the construction boundaries, only partly contributed to the rise of *c'est*-clefts. Using corpus data, he demonstrates that this rise progressed even after French reached its current syntactic rigidity with regard to its SVO order. Dufter (2008) further challenges the description of *c'est*-clefts as compensatory focalising structures by showing that the clefting of subjects and objects remained stable throughout the period from Old French to Modern French. As a way to better account for the diachronic development of *c'est*-clefts, Dufter (2008) invokes two arguments relating to their presuppositional and discursive properties. Firstly, while the combination of the left dislocation and accentuation of *moi* in (22) would suffice to construe the pronoun as focal, only the cleft sentence in (21) introduces the presupposition that *someone is going to pay*.

- (21) C'est pas moi qui vais payer
'It's not me who is going to pay.'
(22) **MOI**, je vais pas payer
'I am not going to pay.' (Dufter 2008: 19, my emphasis)

Secondly, from a discursive viewpoint, Dufter (2008: 19) argues that the cleft allows to avoid any ambiguity with regard to which constituent is in focus and hence is a way to “optimise the argument-focus structure”. More importantly, Dufter (2009a) shows that the increase in the frequency of clefts was accompanied by a diversification of the information-structural patterns available for *c’est*-clefts. Because clefts with a lesser level of presuppositionality and hence a higher need for pragmatic accommodation (Lambrecht 2001: 485) emerged, *c’est*-clefts as presuppositional structures used for rhetorical purposes became more common. In other words, whereas clefts started out as mainly focus-background structures, the possibility for *c’est*-clefts to introduce material in their cleft-relative clause which has not been given in the prior discourse became more acceptable over time. Related to this is the “demarking” of the information-structural profile argued to be typical of *c’est*-clefts with objects as clefted constituents (Sornicola 2011). When an object is construed as the clefted constituent, the construction goes through a process of “demarking” whereby the clefted constituent is no longer pragmatically coded as the focus of the sentence and the construction becomes a presentational/all-focus structure functioning at a higher textual level (Sornicola 2011: 408). This is especially true for clefts used in spoken interaction. Lehmann (2008) analyses this in terms of the grammaticalisation of information structure. According to him, information structural devices such as clefts can only become grammaticalised through the grammaticalisation of specific information structure articulations. Hence, in clefts whose cleft-relative clause carries the biggest informative load, the focality aspect is gradually levelled out for the clefted constituent. As a result, against the traditional description, the cleft does not serve its focalising purpose anymore (Lehmann 2008: 213). I will return to the information structure and motivations behind the use of *c’est*-clefts in Section 1.2.3.

1.2.1.3 The representational meaning of the cleft

In this section, I present an account of the representational meaning coded by the grammatical structure of *it* and *c’est*-clefts, which, in accordance with the theoretical tenets set out in Section 1.1.1, I analyse compositionally (Davidse 2000, Davidse and Kimps 2016, Davidse and Njende 2019). The representational meaning

coded by the cleft syntax is to be distinguished from the various information structures mapped onto the constituents, which I present in Section 1.2.2.

From a semantic viewpoint, the grammatical structure of clefts codes specification whereby the clefted constituent is construed as the value satisfying the variable in the cleft-relative clause. More specifically, the cleft-relative clause formulates an open proposition, i.e. *x who caused all the trouble* in (23a), whose semantic gap is represented by a relative marker.¹⁷ In pseudoclefts, e.g. (23b), the relative clause likewise presupposes an open proposition, but its semantic gap is represented more explicitly by the general head noun like *person* or *one* heading the relative clause, resulting in the NP *the one ('x') who caused all the trouble*. This NP, *the one/person who caused all the trouble*, construes a 'superscriptional' entity (Higgins 1979), or, as Declerck (1988) puts it, a definite description used attributively in the sense of Donnellan (1966), i.e. used essentially, specifying the necessary properties of "whatever or whoever fits that description" (Donnellan 1966: 46). The concrete person fitting the definite description in (23b) is referred to by the postverbal NP, *the boy*. These semantics of pseudoclefts are accounted for in the literature by analysing them as specificational-identifying constructions, in which the main verb *be* expresses a relation of identification between the value of a variable (Higgins 1979; Declerck 1984, 1991; Huddleston and Pullum 2002). That is, the two NPs in pseudoclefts are said to fulfil the semantic roles of value and variable. The variable is the superscriptional entity, the *x* with the role marked by the relative pronoun in the state-of-affairs of the relative clause, and the value is the concrete entity that fills the open proposition, i.e. *the boy*.

- (23) (a) It was the boy who caused all the trouble.
(b) The one/person who caused all the trouble was the boy.
(c) value: the boy
(d) variable: *x* caused all the trouble

¹⁷ As noted by Destruel et al. (2019: 6), the presupposition, as an existential type of inference, may be conveyed with different degrees of likelihood.

It-clefts likewise express a specificational-identifying relation (Bolinger 1972; Collins 1991: 37; Davidse 2000; Huddleston and Pullum 2002: 1416-1417; Van Praet and Davidse 2015) and a relation between the open proposition conveyed by the relative clause and the NP filling the semantic gap. But the structural assembly of these two relations is different. In *it*-clefts, e.g. (23a), the value *the boy* is related to the variable *x who caused all the trouble* by being inserted directly into the antecedent position of the cleft-relative clause. The value can be related to the variable in this way because the cleft-relative and its antecedent do not form a single constituent (Huddleston and Pullum 2002) but are instead structurally ‘binarised’. It is the matrix that relates its postverbal NP as antecedent to the relative anaphor, thus specifying the value for the semantic gap in the cleft-relative clause’s open proposition. The matrix is an identifying clause (Huddleston and Pullum 2002: 1416-1417) with definite subject *it*, or occasionally *that* (Hedberg 2000), which construes the value NP as an identifier (Halliday 1967a: 224; Davidse 2000: 1120). The same description can be made of French *c’est*-clefts and pseudoclefts, whose specificational meaning is structured in different ways too. The fact that the argument of the presupposed proposition, i.e. *x a causé tous ces problèmes*, is underspecified in the pseudocleft (24b) is made explicit with the initial demonstrating pronoun *celui*. Its identity is then revealed in the postcopular position as being *le garçon*. By contrast, the value of the cleft in (24a) is introduced in the postverbal position of the matrix clause and is specified for the variable retrievable from the cleft-relative clause.

- (24) (a) C’est le garçon qui a causé tous ces problèmes.
(b) Celui qui a causé tous ces problèmes est le garçon.
(c) value: le garçon
(d) variable: *x a causé tous ces problèmes*

The identifying syntax of the matrix and the definite pronominal subject *it* or *that* combine to trigger an implicature of exhaustive specification to the effect that the individuals referred to by the value NP coincide with the set of all instances corresponding to the variable. If the value NP is singular, then it is taken to be the only one filling the semantic gap, and if it is plural, then it is taken to exhaustively

state the entities that correspond to the variable. This is a conversational implicature as it can be cancelled without causing any incongruence (Horn 1981; Declerck 1988; De Cesare and Garassino 2015), as illustrated by (25).

- (25) A. He-he saved my life ...
B. ... **It was the Italian that saved your life**
A. Oh, but he [the American] did too. (from de Mille, J. 1871. *The American Baron*, quoted in Schwenter and Waltereit 2010: 90)

The matrix of a cleft construction can also be an existential clause (Davidse 2000; Lambrecht 2001), as in (26). As in *it*-clefts, the entities filling the gap in the open proposition conveyed by the relative clause are inserted directly in the antecedent position of the cleft-relative clause, and it is the matrix which relates the structurally binarised antecedent and cleft-relative clause. As we have an existential matrix here, it merely states that instances corresponding to the variable exist. The existential matrix in (26) has an enumerative meaning (Davidse 1999, 2000), which triggers a conversational implicature of non-exhaustive specification (Lambrecht 2001: 205). It is implied that the individuals referred to in the postverbal complement, *Paul and his wife*, do not exhaust the set of instances that fill the open proposition *x were there*.

- (26) A: And were they ...were they well-received these visits? Erm were there a lot of people interested in them?
B: Well I suppose there were about I've got photographs actually if you'd be interested to see. There must have been about nine of us altogether. **There's Paul and his wife were there.** (Wordbanks)

The proposed compositional analysis sheds light on so-called reduced clefts (Büring 1998; Declerck and Seki 1990) or truncated clefts (Hedberg 2000; Huddleston and Pullum 2002; Collins 2006; Mikkelsen 2007). Importantly, as argued by Davidse (2000: 1121) and Davidse and Njende (2019), both *it*-clefts and *there*-clefts can be 'reduced' in the sense that they can presuppose the variable from the preceding discourse without realising it overtly as a cleft-relative clause, as in (27),

where the hearer is instructed to infer the variable from the preceding discourse, i.e. *x for supper*. Examples like (27a) and (27b) can naturally be accounted for as monoclausal constructions that are, in their own right, specificational constructions. The construction in (27a), with subject *it*, verb *be* and postverbal complement expresses as such identifying-specificational semantics with exhaustivity implicature, while that with existential *there* as subject (27b) expresses enumerating-specificational semantics with non-exhaustive implicature.

- (27) (a) What's for supper? It's spaghetti! (*ibid.*)
(b) What's for supper? There's spaghetti! (*ibid.*)

While it is generally accepted that *it*- and *c'est*-clefts express specificational-identifying meaning (Akmajian 1970), the use of predicational clefts, though infrequent, is attested in a number of studies (Declerck 1983, 1988; Hedberg 1990, 2000; Hartmann 2011, 2016). When a cleft is predicational, it does not serve to identify a given item but instead functions to add information to it (Declerck 1988). In (28a), what is conveyed is that the ceremony that the speaker watched from his/her living room is a touching one, as supported by the rephrasing in (28b).

- (28) (a) It was an odd televised ceremony that I watched from my living room, and a touching one.
(b) The televised ceremony that I watched from my living room was an odd one. (Hedberg 2000: 916)

As illustrated in (28a), the predicational element of predicational clefts can only be a modifier, e.g. *odd*, as opposed to the head noun of the NP, e.g. *ceremony*, and its associated NP cannot have specific reference. The variable of the cleft is here not 'superscriptional', i.e. the heading of a list, but is referential. Because the predicational reading is encoded in the cleft-relative clause, predicational clefts cannot be reduced. As Declerck (1988) shows, the predicationality of clefts can be revealed through a number of morphosyntactic tests which include, but are not restricted to, insertion of degree or comparison marker, conjunction with another predicational cleft (but not a specificational one), deletion of copula when

preceded by *seem*, replacement of the copula by *become*, unrestricted possibility for adverbial addition in pre-copula position, impossibility to neutralise tense of the copula. Predicational clefts can further be distinguished from specificational ones on the basis of their pragmatic and information structural properties. Unlike specificational clefts, predicational ones do not trigger any exhaustivity implicature nor code contrast (Declerck 1988), or at least they only do weaker contrast (Hartmann 2016). It should be noted that Declerck (1988) concludes his analysis of predicational clefts by showing that although they are predicational in nature, sentences like (28a) still retain some sort of specificational behaviour, thus showing the overbearing significance of the specificational-identifying reading of clefts. Purely predicational clefts, on the other hand, remain marginal. Because this study is interested in looking at the interplay between the grammatical meaning of clefts, i.e. specificational meaning, and their information structural properties, only specificational clefts will be investigated.

The view that the cleft syntax codes specificational meaning contrasts with accounts developed by authors like Prince (1978), Lambrecht (2001) or Hedberg (1990, 2000) who take the grammatical structure of clefts to simply code a focus-presupposition articulation. The same position is also advocated by Collins (1991, 2006) who, despite following Crystal's (1969) and Halliday's (1967a) approach to prosodically coded focus, still uses focus and presupposition to describe the grammatical constituents of the cleft. In the approach assumed in the present study, the cleft syntax and the prosodic patterns are viewed as coding different semantics. Grammatical structure codes specification, as argued in this section, while prosodic patterns code a variety of focus assignment patterns, as I describe in greater detail in the following section.

1.2.2 The information structure of *it*- and *c'est*-clefts

After having looked at the syntactic and semantic features of clefts, I now examine the treatment of their information structure. To do so, I first present what has been said about their referential information structure, i.e. the distribution of new and given information, in Section 1.2.2.1, after which I turn to their relational information structure, i.e. prosodically coded focus assignment, in Section 1.2.2.2.

I then discuss the different shortcomings of the different typologies in Section 1.2.2.3.

1.2.2.1 Typologies of the discourse-familiarity of *it*- and *c'est*-clefts

As information-packaging constructions for some (Huddleston and Pullum 2002) and focalising structures for others (Lambrecht 2001; Lehmann 2008), cleft sentences naturally fall within the scope of studies on information structure aiming at uncovering the properties of aspects such as focus, topic and givenness. Many authors have delved into the description of the relational and referential layers of the information structure of clefts, as well as the interaction between the two (see for instance Prince 1978; Geluykens 1988; Hedberg 1990; Clech-Darbon et al. 1999; Lambrecht 2001; Collins 1991, 2006; Huber 2006; Herment and Leornarduzzi 2012, 2015; Garassino 2014; Doetjes et al. 2004; Mertens 2012; Avanzi 2011). For the sake of clarity, I will first summarise the main approaches which have been adopted to study referential information structure, i.e. the way new and given information is signalled in clefts. Aspects relating to their relational information structure, i.e. the prosodic coding of focus, will be dealt with in the next section.

Within the formal-pragmatic approach to information structure advocated for by Lambrecht (2001) (see Section 1.1.3.2), clefts are analysed as argument-focus structures in which the newness of the focal item, i.e. the unpredictable component responsible for the difference between the pragmatic assertion and the pragmatic presupposition, is established relationally between said item and the rest of the proposition. In the case of clefts, Lambrecht (2001: 474) takes the clefted constituent to be the new information which “makes the utterance into an assertion” while he views the cleft-relative clause as carrying the pragmatically presupposed, and hence assumed-to-be-known, information. For Lambrecht (2001), the cleft-relative clause typically triggers a knowledge presupposition (K-presupposition), a consciousness presupposition (C-presupposition) and a topicality presupposition (T-presupposition). What this means is that, in Lambrecht’s (2001) view, the information introduced in the cleft-relative clause is most frequently hearer-old, mentally activated and of current interest to the hearer. When one or more of the presuppositional aspects is missing, then the cleft exemplifies a marked

use of the construction. Thus, while Lambrecht (2001) admits to the non-systematicity of discourse-givenness for the cleft-relative clause, he still argues in favour of a prototypical and prevalent cleft displaying a new-given structure.

In considering that clefts departing from the new-given pattern are instances of clefts whose use is not significant enough to form a category of their own, Lambrecht (2001) goes against the conclusions drawn in a number of studies investigating the information structure of *it*- and *c'est*-clefts. Delin (1990), for instance, finds that new referents may be introduced in the postcopular position but also in the cleft-relative clause. Her study not only shows that patterns such as new-new or given-new are attested in clefts, but also that their use is far from being scarce. Out of 150 *it*-clefts analysed by Delin (1990), 90 introduced some at least some new information in both the matrix clause and the cleft-relative clause, which amounts to 60% of all clefts. Hedberg and Fadden (2007) likewise conclude, with a smaller dataset, that *it*-clefts display a certain freedom in the distribution of discourse-new and discourse-given information in either of their components.

Given the different patterns of discourse-new and discourse-given information attested in data, some authors within the functional approach have attempted to devise typologies based on the relational and referential layers of the information structure of *it*- and *c'est*-clefts. These range from binary (Prince 1978; Geluykens 1988; Doetjes et al. 2004; Collins 2006) to ternary classifications (Declerck 1984; Huber 2006; Scappini 2013; Karssenbergh and Lahousse 2015).¹⁸ In addition to these information structure-based taxonomies, there also exist pragmatically based classifications such as Hedberg's (2000: 895) which do not rely on the binary new/given distinction but on eight levels of activation of referents based on the speaker's assumptions on the hearer's mental state. To avoid any misguided comparison of typologies which investigate relating, but yet distinct, aspects of the information structure of clefts, only those dealing with the distribution of discourse-new and discourse-givenness are included in the summary in Table 1 below.

¹⁸ Not included in the overview but worth mentioning are studies focusing on the functional viability of the typologies aforementioned (e.g. Dufter 2009a; Garassino 2014). These will be discussed in Section 1.2.3.

Type of <i>it/c'est</i> -clefts	Clefted constituent	Cleft-relative clause
Stressed-focus (Prince 1978)		
Old-presupposition (Collins 2006)		
Contrastive (Declerck 1984)		
Contrastive (Huber 2006)		
Filler-focus (Geluykens 1988)		
Focus-background (Karssenbergh and Lahousse 2015)	new	given
Focus-ground (Doetjes et al. 2004)		
Focus-post-focus (Rialland et al. 2002)		
Narrow focus (Scappini 2013)		
Type I (Mertens 2012)		
Informative-presupposition (Prince 1978)		
Unstressed-anaphoric-focus (Declerck 1984)		
Continuous topic (Huber 2006)		
Clause-focus (Geluykens 1988)		
Topic-comment (Karssenbergh and Lahousse 2015)	given	new
<i>Broad focus (Doetjes et al. 2004)</i>		
<i>Broad focus (Rialland et al. 2002)</i>		
<i>Broad focus (Scappini 2013)</i>		
Type II (Mertens 2012)		
New-presupposition (Collins 2006)		
Discontinuous (Declerck 1984)		
All new (Huber 2006)		
All focus (Karssenbergh and Lahousse 2015)		
<i>Broad focus (Doetjes et al. 2004)</i>	new	new
<i>Broad focus (Rialland et al. 2002)</i>		
<i>Broad focus (Scappini 2013)</i>		
Type II (Mertens 2012)		

Table 1. Summary of information structure-based typologies of *it*- and *c'est*-clefts

From the eight typologies included in Table 1, one can derive the simplified ternary taxonomy in Table 2 below. The remainder of this section will be devoted to the detailed description of the overlaps existing between the three discourse-familiarity patterns.

Type of <i>it/c'est</i> -clefts	Clefted constituent	Cleft-relative clause
Type 1	new	given
Type 2	given	new
Type 3	new	new

Table 2. Simplified information structure-based typology of *it*- and *c'est*-clefts

Type 1 encompasses all clefts in which the clefted constituent introduces discourse-new information while the cleft-relative clause conveys discourse-given information. These correspond to *stressed-focus*, *old-presupposition*, *contrastive*¹⁹, *filler-focus*, *focus-background*, *focus-ground*, *focus-post-focus*, *narrow focus*, and *type 1* clefts. The new-given pattern is illustrated in example (29) taken from Prince (1978). Here, the clefted constituent *the covers* constitutes new information and the presupposed open proposition *x that are rotten* given information retrievable from the mention of having to mend books in the prior discourse.

- (29) So I learned to sew books. They're really good books. **It's just [the covers]_{NEW} [that are rotten]_{GIVEN}.** (Prince 1978: 896)

Clefts exemplifying type 2, i.e. *informative-presupposition*, *unstressed-anaphoric-focus*, *continuous topic*, *clause-focus*, *topic-comment*, and *broad focus*

¹⁹ Declerck (1984) makes a distinction between the occurrence of information and the way it is presented. In the case of contrastive clefts, he explains that regardless of whether it occurs or not in the preceding context, it is still presented as new information as it has not yet been specified as value for the variable. With this view, it is not in fact the value that is new but the specification relation.

clefts, are clefts with a discourse-given clefted constituent and discourse-new material in the cleft-relative clause. In (30), also borrowed from Prince (1978), the fact that there was a vast internal migration from the south northwards is logico-semantically presupposed but is new at the level of discourse. The clefted constituent, on the other hand, conveys discourse-given information as shown by the anaphoric determiner *these*.

- (30) **It was also [during these centuries]_{GIVEN} [that a vast internal migration (mostly by the Galla) from the south northwards took place]_{NEW}, a process no less momentous than the Amhara expansion southwards during the last part of the nineteenth century and the beginning of the twentieth century. (Prince 1978: 898)**

In the third type of clefts, only accounted for by Declerck (1984), Collins (2006), Huber (2006), and Karssenbergh and Lahousse (2015), *new-presupposition*, *discontinuous*, *all-new* and *all-focus* clefts have discourse-new information in both the clefted constituent and the cleft-relative clause, as is the case in (31).

- (31) **It is [through the writings of Basil Bernstein]_{NEW} [that many social scientists have become aware of the scientific potential of sociolinguistics]_{NEW}. (Declerck 1984: 267)**

As far as reduced clefts are concerned, their existence is only acknowledged in Prince's (1978) and Collins's (2006) typologies and is mainly justified by the high degree of predictability of their variable.²⁰ As such, they are categorised as *stressed-focus/old-presupposition* clefts. The treatment of reduced clefts overall remains limited, and their use does not appear to be validated as a fairly common alternative choice. I will return to this in Chapter 6.

A few things regarding the debate on the different typologies should be noted. First, in defending the view that clefts have a prototypical new-given pattern,

²⁰ While Declerck (1984) does not mention reduced clefts in his classification, he recognises their existence and analyses their use in later work (Declerck 1988, and Seki 1990).

Lambrecht (2001) rejects the argument made by Prince (1978) that *informative-presupposition* clefts, i.e. given–new clefts, should form a full-fledged category. Instead, Lambrecht (2001) analyses them as strongly marked instances of clefts whose cleft-relative clause is K-presupposed, i.e. pragmatically presupposed, but not T-presupposed, i.e. which is encoded as a topic of interest for the hearer.²¹ He further supports this claim by arguing that the missing T-presupposition of these clefts simply requires accommodation on the hearer’s part, which is not enough to seclude them into a specific category. On this, Dufter (2009a) notes that type 2 clefts, or *informative-presupposition* clefts, arose in French in 14th-15th century and achieved an established status with a significant increase in the 16th century. Before then, the same information structure articulation could be achieved through the fronting of the non-subject anaphoric constituents of various types (Muller 2003; Bouchard et al. 2007). Dufter (2009a) identifies the same pattern for type 3 clefts, i.e. *all-focus* clefts, which are neither recognised by Prince (1978) or Lambrecht (2001) as a unique informationally-motivated class of clefts, but whose emergence is traced back to the 16th century and is justified by the growing restrictions on VS ordering in French. This suggests that a three-way typology better accounts for the variety of informational statuses available for the clefted constituent and cleft-relative clause of *it*- and *c’est*-clefts.

A similar prototypicality claim associated with type 1 new–given clefts is made for *c’est*-clefts by Scappini (2013: 83) who argues that *narrow focus* clefts are the most frequently found in spoken French. But rather than attributing this greater use to the focalisation function of clefts, Scappini (2013) explains the prevalence of type 1 clefts in terms of the contrast that is established between the clefted constituent, and one or more alternatives of a given paradigm. Thus, similarly to Declerck’s (1984) *contrastive* category, Scappini’s (2013) *narrow focus* clefts have a strongly contrastive item as their value. Unlike Lambrecht (2001), Scappini (2013) acknowledges the existence of other types of information distribution in *c’est*-clefts

²¹ Prince (1978) and Doetjes et al. (2004) note that the content of the cleft-relative clause of *it*- and *c’est*-clefts respectively is always logico-semantically presuppositional regardless of its information status. This contrasts with Declerck’s (1984) view on clefts of the new–new type which he argues do not convey any presupposed material in their variable and are therefore hardly specificational.

and treats her *broad focus* category as a full-fledged alternative choice with specific information structural properties.

Interestingly, Scappini's (2013) and Doetjes et al.'s (2004) *broad focus* categories do not perfectly match that of other typologies and instead span over both type 2 and type 3 clefts due to the informational freedom of the clefted constituent.²² In either case, however, whether the same distribution of the different types is corroborated by the data examined this study will be discussed in Section 5.2.

1.2.2.2 Typologies of the prosody of *it*- and *c'est*-clefts

Most of the studies investigating the referential information structure of clefts also make a number of predictions on the way their relational information structure is encoded. In the Hallidayan (1967a, and Matthiessen 2004, 2014) tradition, focus marking in English is done prosodically through the placement of nuclear accents. The source of focus in French, on the other hand, has been disputed and is said to rely either on accentuation (Di Cristo 2016) or intonational phrasing (Féry 2001). Because of the differences in the intonation systems of the two languages, I will first examine the treatment of *it*-clefts and then move on to *c'est*-clefts.

Within Lambrecht's (2001) formal-pragmatic approach, *it*-clefts typically indicate a new-given division in which the new clefted constituent is the focus denotatum and the given open proposition in the cleft-relative clause the presupposition. While Lambrecht concludes that the focus-presupposition articulation is coded by the syntax of clefts, he nevertheless makes descriptive claims about the pitch accents carried by them which are not rooted in corpus-based evidence (Lambrecht 2001: 478-493). For him, the focus phrase is

²² Although Doetjes et al. (2004) and Scappini (2013) both take Prince's (1978) binary taxonomy as a starting point, the status of the clefted constituent in *broad focus c'est*-clefts - which are compared to Prince's (1978) *informative-presupposition* clefts - remains rather unclear and is not unequivocally described as conveying either discourse-new or discourse-given. Looking at the examples presented as instances of broad focus in both studies nonetheless suggests that the clefted constituent of broad focus *c'est*-clefts tends to contain anaphoric references and hence discourse-given material. These would make *broad focus c'est*-clefts fall within the type 2 category.

“necessarily accented” (2001: 493) and the presupposition typically “unaccented” (2001: 479), unless it adds something to the presupposed current concern that was not ratified or “sufficiently salient in the discourse” (2001: 480), e.g. *I crave* in (32c). In such a case, there may be a pitch accent on an element of the cleft-relative clause, which, because of that unit’s inherent presupposed status, is analysed as a topic accent and not a focus accent (2001: 480).

- (32) (a) It is **CHAMPAGNE** that I like.
(b) I like **CHAMPAGNE**.
(c) It is **CHAMPAGNE** that I **CRAVE**. (Lambrecht 2001: 469)

Thus, the pattern he posits as being the most prototypical and frequent one for *it*-clefts is the one illustrated in (32a) in which the focus denotatum representing the new information is accented and the presupposed and hence given information is in the post-nuclear tail.

This alleged prototypicality is challenged by the description found in the main typologies of *it*-clefts (Prince 1978; Declerck 1984; Geluykens 1988; Collins 2006; Huber 2006) and is also contradicted by the results of different corpus-based studies (Delin 1990; Hedberg 1990; Collins 1991; Herment 2008; Herment and Leonarduzzi 2012; Kimps 2016). Let us first examine the extent to which the different information-structure based taxonomies of *it*-clefts compared in Table 3 below cast doubt on pragmatic accounts.

Type of <i>it</i> -clefts	Clefted constituent		Cleft-relative clause	
	RF	RL	RF	RL
Stressed-focus (Prince 1978)				
Old-presupposition (Collins 2006)				
Contrastive (Declerck 1984)	new	strong	given	weak/no
Contrastive (Huber 2006)		stress		
Filler-focus (Geluykens 1988)				
Informative-presupposition (Prince 1978)				
Unstressed-anaphoric-focus (Declerck 1984)	given	weak/no	new	normal
Continuous topic (Huber 2006)		stress		
Clause-focus (Geluykens 1988)				
New-presupposition (Collins 2006)				
Discontinuous (Declerck 1984)	new	stress	new	stress
All new (Huber 2006)				

Table 3. Overview of typologies of the referential (RF) and relational (RL) information structure of *it*-clefts

Each type of *it*-clefts recognised in the typologies, viz. new–given, given–new and new–new, is coupled with at least one prosodic pattern based on the presence of absence of prosodic stress and, in cases of multiple foci, their degree of prominence. The different combinations can be simplified as in Table 4 below.

Type of <i>it</i> -clefts	Clefted constituent		Cleft-relative clause	
	RF	RL	RF	RL
Type 1	new	strong stress	given	weak/no stress
Type 2	given	weak/no stress	new	normal stress
Type 3	new	stress	new	stress

Table 4. Simplified information structure-based typology of *it*-clefts

Starting with type 1 *it*-clefts, *stressed-focus*, *old-presupposition*, *contrastive*, and *filler-focus it*-clefts are said to carry a discourse-new stressed clefted constituent that receives strong stress and a discourse-given cleft-relative clause that bears either weak or no stress at all. It is the case of (33) retrieved from the LLC-1 and analysed by Collins (2006). In this example, the clefted constituent and the cleft-relative clause are realised in separate tone units both of which contain one nuclear accent, on *regions* and *matter* respectively. Despite the presence of multiple information foci, only the first accent is analysed by Collins (2006: 1709) as indicative of newsworthiness.

- (33) we repudiate Mr Heath and we repudiate any such undertaking for him the junketing for us# it's [jobs in the R\EGIONS#]_{NEW} [that M\ATTER#]_{GIVEN} (Collins 2006: 1709)

Type 2 *it*-clefts, i.e. *informative-presupposition*, *unstressed-anaphoric-focus*, *continuous topic*, and *clause-focus* clefts, display the opposite order whereby the clefted constituent introduces discourse-given information that is either weakly accented or deaccented altogether and the cleft-relative clause discourse-new information marked by normal stress. In (34) below, the cleft is uttered as a single information unit with a unique nuclear accent on *invited* in the cleft-relative clause. The clefted anaphoric pronoun, on the other hand, is located in the prenuclear segment.

- (34) A: did you meet Fuller
B: yes it was [he]_{GIVEN} [who INVITED me#]_{NEW} (Geluykens 1988: 828)

Finally, the third type of *it*-clefts, whose information is all new, carries prosodic stress in both of its constituents as instantiated in (35). Here, the cleft stretches over ten tone units, two of which make up the clefted constituent and eight of which the cleft-relative clause. As a result, both components bear multiple nuclear accents which mark most of the information throughout the cleft focal.

- (35) on the whole the whole of the village was happy but it was rather a hand-to-mouth existence it was a time of gentry I use the term specially gentry because I was brought up in this era and who on the whole were respected they influenced the village ran the Parish Council the Church and they banded together to build the village hall **it was only [\AFTER# W\ORLD War Two#]_{NEW} [that \WE# R\EALLY# the P\EOPLE# in our V\ILLAGE# H\AD# a little to S\AY# of how TH\INGS# should be R\UN#]_{NEW}** (Collins 2006: 1710)

Overall, what the comparison of the five typologies highlights is an apparent existence of a systematic match between the referential and relational layers of the information structure of *it*-clefts. In other words, Prince (1978), Declerck (1984), Geluykens (1988), Collins (2006) and Huber (2006) appear to conflate, to a certain extent, discourse-newness with the presence of (strong) prosodic stress and discourse-givenness with the absence of prosodic prominence or presence of a weaker stress. As a result, each pattern of discourse-familiarity is associated with a unique prosodic articulation. Upon closer inspection, Collins (2006) and Huber (2006) somewhat nuance this uniqueness by acknowledging a wider variety of focus placement than that seemingly projected by the categories of clefts they establish. Thus, Collins (2006) further divides his *new-presupposition* category into two subclasses, one of which introduces a focal clefted constituent and the other which introduces a deaccented one. This distinction is, however, not reflected in Collins's (2006) typology as he primarily bases it on the informational status of the cleft-relative clause. As for Huber (2006), he discusses a fourth type of clefts, whose status is semantically motivated rather than information structurally, in which both components of the cleft carry one of more prosodic foci, which makes the

information in both the clefted constituent and the cleft-relative clause contrastive in nature.

The comparison of the typologies aforementioned with the findings of the different empirical studies bearing on focus placement in *it*-clefts leads to a number of observations. Like the typologies under study, the accounts of Delin (1990), Hedberg (1990), Herment and Leonarduzzi (2012), and Kimps (2016) uncover a significant variation in the number of prosodic patterns available for *it*-clefts, which brings into question the ‘focalisation’ process traditionally invoked to explain their use. Herment and Leonarduzzi’s (2012) findings show, for instance, that clefts with multiple foci constitute the most frequent prosodic articulation in the speaking component of ICE-GB corpus, followed by clefts with a single nucleus in the cleft-relative clause and clefts with a single nuclear accent in the clefted constituent. This is consistent with Delin’s (1990) results which demonstrate that focus is not restricted to the clefted constituent and which, consequently, further exposes the mismatch being syntax and prosody in coding information as salient in *it*-clefts. The conclusion both studies arrive at contrasts with Lambrecht’s (2001) description of *it*-clefts in which the clefts with a unique nucleus in the postcopular complement are posited as being the prototypical realisation. They accord, however, with Halliday’s earlier observations on the possibility of multiple information structural patterns in *it*-clefts (Halliday 1967a: 236-237, 1968: 179). For Halliday (1967a), the value-constituent can be focal or non-focal, and the cleft-relative clause can feature no, one or multiple foci.

Though less extensive, the literature on the prosody of *c’est*-clefts also provides insights on where and how focus is encoded in French clefts. Table 5 below presents the mapping of discourse-familiarity patterns and prosodic articulations.

Type of <i>c'est</i> -clefts	Clefted constituent		Cleft-relative clause	
	RF	RL	RF	RL
Focus-ground (Doetjes et al. 2004)				
Focus-post-focus (Rialland et al. 2002)	new	focal	given	post-focal
Type I (Mertens 2012)				
Broad focus (Doetjes et al. 2004)				
Broad focus (Rialland et al. 2002)	given/new	not focal	new	focal/non-backgrounded
Type II (Mertens 2012)				
Explicative (Rialland et al. 2002)				
All focus (Rialland et al. 2002)	new	focal	—	—

Table 5. Overview of typologies of the referential (RF) and relational (RL) information structure of *c'est*-clefts

Given that Scappini's (2013) taxonomy does not make predictions on the prosody of *c'est*-clefts, it is excluded from Table 5. Note, furthermore, that the third subclass corresponding to *explicative* and *all-focus* clefts is not presented as a full-fledged category by Rialland et al. (2002) and Doetjes et al. (2004) but rather as a minor type of reduced clefts. It remains nonetheless included in the comparative table so as to draw a picture which is as exhaustive as possible.

Like *it*-clefts, the different subtypes of *c'est*-clefts, viz. new-given, given-new and new-new, are associated with a corresponding prosodic pattern based on the focal/non-focal status of the component. The taxonomies of *c'est*-clefts, however, break with *it*-clefts in that given-new and new-new are merged within the same class. Another difference relates to the existence of an additional category encompassing marked cases of reduced clefts. Table 3 below summarises the information structural properties attributed to the three types.

Type of <i>it</i> - clefts	Clefted constituent		Cleft-relative clause	
	RF	RL	RF	RL
Type 1	new	focal (narrow focus)	given	non-focal
Type 2	given/new	focal (broad focus)	new	focal (broad focus)
Type 3	new	focal	–	–

Table 6. Simplified information structure-based typology of *c'est*-clefts

In type 1 *c'est*-clefts, i.e. *focus-ground*, *focus-post-focus* and *type I* clefts, the clefted constituent is directly followed by a major prosodic boundary which marks the end of an intonational phrase and assigns focus to the postcopular complement. As such, the clefted constituent is presented as the new information to the hearer, irrespective to its discourse-newness or givenness. The cleft-relative clause, on the other hand, is realised in a separate intonational phrase whose illocutionary tone, e.g. assertive or interrogative, replicates that of the previous phrase. Put simply, if the clefted constituent ends with an assertive terminal boundary tone, marked L% in Rialland et al.'s (2002) and Doetjes et al.'s (2004) annotation system and L-L in Mertens's (2012), the cleft-relative clause will also display a final assertive appendix L% tone or l-l-. As noted by Rialland et al. (2002: 1) and Doetjes et al. (2004: 544-545), this replication is typically accompanied by some sort of register reduction illustrated in the pitch curve by a downstep. Intonational agreement and downstepping combine to bring out the post-focal nature of the intonational phrase and the backgrounding of the information it conveys. The prosodic profile of type 1 *c'est*-clefts is illustrated in (36) below.

(36) [C'est pour Tournier] [qu'elle va voter] (Rialland et al. 2002: 2)

L% L%

'It's for Tournier that she's going to vote'

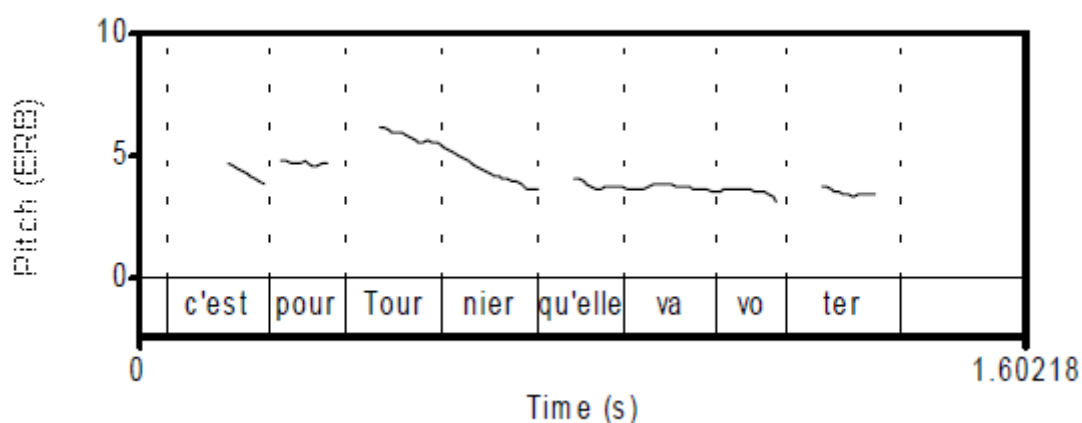


Figure 3. F0 curve of (36) (Rialland et al. 2002: 2)

As corroborated by the pitch curve in Figure 3, the matrix clause and cleft-relative clause of (36) each form an intonational phrase (marked by square brackets) whose boundary tone is the terminal assertive falling tone L% (l-l-). A clear downstepping occurs in the second intonational phrase as shown by the compression of the tone movement. As a result, the information conveyed in the cleft-relative clause, i.e. that she is going to vote for someone, is backgrounded while that in the clefted constituent coincides with the focus of the construction. In comparison with Jun and Fougeron's (2000) and Di Cristo's (1998, 2016) descriptions of focus, (36) illustrates the movement of the pitch nuclear accent on *Tournier* as well as the deaccentuation of post-focal material argued for in both accounts.

Type 2 *c'est*-clefts, i.e. *broad focus* and *type II* clefts, differ from type 1 in a number of ways. First, despite following a similar intonational segmentation, the clefted constituent is this time not marked by a focus-inducing terminal boundary tone but by a major continuation rising tone such as H%[cont] (HH). The information introduced in the postcopular position may be new or given and is part of broad focus scoping over the whole cleft. Unlike in type 1, the cleft-relative clause of type 2 clefts does not exhibit any sort of intonational agreement with the previous intonational phrase and is therefore not encoded as a post-focal backgrounded segment but as new information. The boundary tone occurring after the cleft-relative clause may be terminal or continuative. In the case of (37), the cleft is made up of three intonational phrases, all of which have a continuative rise at their rightward edge as illustrated in Figure 4.

- (37) [Ce sont les visiteurs][qui les déposent][le plus simplement du monde]
(Rialland et al. 2002)

H%[cont] H%[cont] H%[cont]

‘It’s the visitors that drop them off as simply as that.’

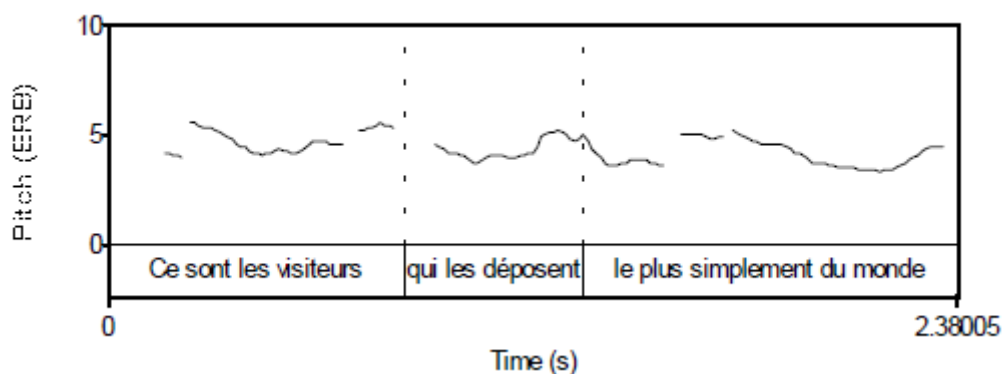


Figure 4. Prosodic realisation of (37) (Rialland et al. 2002: 3)

Through this prosodic articulation, the information conveyed in the cleft-relative clause is not marked as backgrounded information. In contrast with type 1 clefts, the clefted constituent of type 2 clefts does not receive a narrow focal status. The focus instead extends over the whole construction and there is no deaccented portion. Mertens (2012) argues that attaching focus specifically to the material in the cleft-relative clause of broad-focus sentences can be achieved with the insertion of a rise-fall continuative tone at the end of the cleft. In Di Cristo’s (1998, 2016) terms, this would entail a ‘double balisage’ around the focused constituent. For him, type 2 clefts constitute the most striking example of the mismatch between the syntax, information structure and prosody of clefts as well as the more richly exploited information structure articulation in spoken data.

In the last category, which is only accounted for by Rialland et al. (2002) and Doetjes et al. (2004), the cleft is reduced and is generally used as an answer to a broad question such as ‘What happened?’. Only the clefted constituent is expressed while the cleft-relative clause is omitted. The cleft is typically realised as a single intonational phrase bearing a terminal boundary tone, thus making the proposition an all-focus structure and the clefted constituent focal. It is the case of (38) which seemingly resembles a full cleft but in which only the matrix clause, which contains a relative clause, is expressed. Along with the boundary tone L% seen in Figure 5,

prosodic realisation of the cleft. In addition to the clefts deriving from the type 1/type 2 distinction encountered by Avanzi (2011: 120) are clefts whose clefted constituent bears prosodic focus but whose cleft-relative clause does not go through intonational agreement and instead shows pitch variations which are more significant than that of the preceding intonational phrase. Avanzi (2011) demonstrates that this may be due to the nature of the immediate co-text, e.g. a causal relative clause whose information is linked to that of the cleft, or to the speaker wanting to indicate to the hearer that s/he wishes to hold the conversational floor.

For both *it*- or *c'est*-clefts, the survey of the different studies focusing on the prosody of clefts bring to light the availability of a wider range of articulations than that predicted by earlier accounts. Sections 3.2 and 4.2 will address the question of whether this variation is corroborated by the data under study.

1.2.2.3 Shortcomings of the existing typologies

Although the ten information structure-based typologies discussed in the previous sections overlap to some degree, a number of issues emerge from their comparison. From a terminological viewpoint, different definitions are attributed to the notion of focus depending on whether it is treated as a syntactic, pragmatic or prosodic feature. In Prince's (1978) and Declerck's (1984) typologies, focus is tied to the syntactic position of postcopular complement in the matrix clause and hence systematically refers to the value of clefts.²³ By contrast, Geluykens (1988), Huber (2006), Scappini (2013) and Karszenberg and Lahousse (2015) define focus pragmatically as what is marked as salient by the speaker within the focus/background paradigm. Rialland et al. (2002), Doetjes et al. (2004) and Mertens (2012) adopt a dual approach to focus and distinguish between pragmatic focus and prosodically coded focus. Only Collins (2006) does not make use of the notion of focus and instead refers to syntactic focus as highlighting. To prevent any

²³ Declerck (1988: 13) revises this terminology in later work and redefines focus as the new information introduced by the cleft. He notes that focus is not necessarily restricted to the clefted constituent but can also extend beyond the matrix clause boundary.

unintentional conflation between the different aspects relating to the use of clefts, I will follow in Rialland et al. (2002), Doetjes et al. (2004) and Mertens's (2012) footsteps and treat syntactic focus and prosodic focus as two distinct, but yet interacting, notions. I will refer to the emphasis deriving from the bi-clausal syntax as syntactic *highlighting*, and will, following the functional tradition set out by Halliday (1967a), designate the prosodic coding of emphasis as *information focus*.²⁴

In the same vein, the comparability of the different typologies also suffers from the competing approaches to discourse-familiarity.²⁵ Prince (1978), Declerck (1984) and Huber (2006) all adopt a cognitive approach in which the speaker's assumptions are taken into account. On this view, items are considered given when the speaker expects them to already be familiar to the hearer while new items "cannot be assumed to be in the hearer's consciousness" (Declerck 1984: 256). By contrast, the description of discourse-familiarity developed by Geluykens (1988), Collins (2006), Rialland et al. (2002), Doetjes et al. (2004), Mertens (2012) and Karssenbergh and Lahousse (2015) primarily relies on textual givenness/newness. As such, a constituent is only considered discourse-given if, and only if, it is retrievable, whether explicitly or implicitly, from the prior context. Given that the cognitive approach is difficult, if not impossible, to apply to corpus data, I adopt the latter approach and treat the discourse-givenness/newness of a referent as being strictly dependent on its presence or absence in the preceding discourse.

A third shortcoming revealed by the comparative review of typologies of *it*-clefts specifically lies in the lack of technical description of the prosodic patterns associated with the different types of clefts. None of the five typologies of *it*-clefts, viz. Prince's (1978), Declerck's (1984), Geluykens's (1988), Collins's (2006) and Huber's (2006), present precise acoustic measures or reference points accounting for the variation between weak, normal, and strong stress. Most importantly, the pattern in which both clefted constituent and cleft-relative clause contain discourse-given information remains largely overlooked and excluded from existing

²⁴ See Section 1.1.3 for a detailed description of the notion of information focus.

²⁵ The differences between the cognitive and textual approaches to discourse-familiarity are explored in greater detail in Section 2.4.

typologies, despite Huber (2006) briefly acknowledging its existence as a subtype of *contrastive* clefts.

More generally, most of the typologies somewhat exemplify the conflation of referential and relational information structures of clefts. This is particularly true for taxonomies of *it*-clefts which predict that discourse-new constituents typically receive prosodic focus while discourse-given items generally remain unaccented. This one-to-one matching goes against Halliday's (1967a) argument that prosodically coded focus need not be borne by discourse-new information. For him, information foci allow speakers to mark the newness of a given item but "not in the sense that it cannot have been previously mentioned (...) but in the sense that the speaker presents it as not being recoverable from the preceding discourse" (Halliday 1967: 204). Collins (2006: 1709) illustrates this with example (39) from the LLC-1.

(39) we REPUDIATE Mr H/EATH# - and we repudiate any such
UNDERTAKING# - for him the J\UNKETING# - for \US# **it's [jobs in the
REGIONS# that MATTER#]**GIVEN (LLC-1)

Here, both the clefted constituent and the presupposed open proposition carry a nucleus on *regions* and *matter* respectively. Despite the fact that both components contain an information focus, i.e. prosodically coded focus, Collins (2006) argues that the information conveyed in the cleft as a whole, i.e. that the issue of unemployment is of concern, is given as it can be inferred from the preceding with the speaker warning of the effect of pound fluctuations. Collins's (2006) typology therefore partly reflects the distinction between discourse-familiarity and prosody, and hence also between referential and relational information structure. By contrast, the accounts of French *c'est*-clefts do not admit such a straightforward link between the two layers. Doetjes et al. (2004: 534) for instance note that "a focused constituent may have a discourse referent which is present in the preceding discourse". Instead, the taxonomies mainly rely on the distinction between narrow and broad focus, neither of which are associated with newness of givenness in a systematic manner. With this thesis, I aim at further developing the distinction between discourse-familiarity and prosody in order to provide a more fine-grained

description of the referential and relational information structure of *it*- and *c'est*-clefts.

1.2.3 Functional studies of *it*- and *c'est*-clefts

From a functional viewpoint, clefts have been associated with a wide range of discourse uses which are motivated either locally or at a higher textual level and which I detail in this section. In addition to the focalisation function, whose predominant role has been refuted by some authors, as discussed in Section 1.2.1.2, clefts have also been described as encoding ‘contrastive focus’ (É. Kiss 1998). As such, the lexical item introduced as the value satisfying the specificational relation is contrasted with other alternatives of a set that can be open or closed. Sornicola (2011: 403), for instance, notes for French that “any movement of a basic constituent which is not linked to the topicalization process gives rise to a contrastive structure”. Destruel et al. (2019) confirm this by showing that clefts are particularly suited to express contrast and contradiction with prior discourse, a feature which is not as salient in canonical sentences due to non-at-issueness of the antecedent of the contrast, i.e. the variable of the cleft (Destruel et Velleman 2014). This is true for both French and English, though Lambrecht (2001: 506) underlines that the semantic distinction between closed-set and open-set clefts tends to be more consistently marked in French than in English. In (40a), the question uttered by speaker M calls for an open set of answers which allows for answers like the simple pronoun ‘*Moi*’. By contrast, M’s question in (40b) implies a closed set of alternatives which more easily yields a cleft in French than in English.

(40) (a) M: Qui veut encore de la viande?

‘Who wants some more meat?’

A: *Moi*. / #*C’est moi*.

‘Me.’ ‘It’s me’

B: *Moi aussi*. / #*C’est moi aussi*.

‘Me too.’ ‘It’s me too.’

C: *Pas moi*. / #*C’est pas moi*.

‘Not me’ ‘It’s not me’

- (b) M: Qui c'est qui a pris le journal?
'Who is it that took the newspaper?'
A: C'est pas moi. / #Pas moi.
'It's not me.' 'Not me.'
B: C'est moi. / #Moi.
'(It's) me.' 'Me.' (Lambrecht 2001: 506)

For Hartmann (2011), whether the set of alternatives is given in the prior context or not impacts the strength of the contrast at stake. When it is expressed, as in (40b) through the negation, the contrast is more strongly marked than when alternatives remain implicit, which inevitably raises the question of whether all clefts are inherently contrastive. A partial answer to this is provided by Destruel (2012) whose experimental study shows that clefts are not systematically employed in contrast-prone environments. On the anaphoric analysis of clefts, contrast is generated by the comparison of the clefted item with one or more alternatives but also by the presence of an antecedent giving rise to the situation in which the contrast is relevant. In clefts, this antecedent corresponds to the presupposition conveyed in the cleft-relative clause which “pull[s] out a proposition in order to establish it as the basis for achieving a contrastive operation” (Delin 1992: 9). The presupposition conveyed in the cleft-relative clause is thus viewed as a type of propositional anaphora which “remind[s] rather than inform[s]” (1992: 8). As an anaphora, the presupposition allows to establish contrast with elements from the preceding discourse. The most obvious illustration of this antecedent-referent relation are corrective clefts in which the paradigmatic set and antecedent are both made explicit. An example of this is provided in (41) in which speaker B provides a correction on which is the appropriate value, between *John* and *Alex*, for the antecedent variable *x* *cooked beans*.

- (41) A: I wonder why Alex cooked so much beans.
B: Actually, **it was John who cooked the beans**. (Destruel et al. 2019: 198)

For Katz (2000a, b), only *corrective* clefts of the kind, which have a strongly presupposed variable, express true contrast while *value-specifying/variable fulfilment* clefts aim at identifying the missing argument fulfilling the assertion without contrasting it with any other lexical items. The distinction argued for by Katz (2000a, b), along with the anaphoric view of clefts, create the expectation that the presupposition of a cleft needs to be somewhat given in the prior discourse for contrast to have sufficient ground. This prediction seems to be corroborated by the accounts developed by Geluykens (1988) and Scappini (2013) who link contrastiveness with type 1 clefts, i.e. clefts selecting a new postcopular complement and conveying given material in the cleft-relative clause. It is however contradicted by Delin (1992: 10) who argues that the presupposition need not actually be explicitly given but may only be inferentially related or referred to attributively. In (42) below, the value *the new textures and fittings* is contrasted with *the doubling [of] the selling space* mentioned in the previous sentence. The contrast is established with regard to the antecedent that something was costly which can be inferred from the mention of the greatest expense. The variable is therefore not explicitly expressed in the context, but is yet sufficiently salient to add a contrastive reading to the cleft.

- (42) Doubling the selling space to 700 square feet was not to be the greatest expense. **It was the new textures and fittings to fill this space that would be costly.** (Delin 1992 : 10)

On the nature of the contrast marked in clefts, Destruel and Velleman (2014) suggest that it is not just linguistic contrast which is at play but also metalinguistic contrast. As such, clefts may, in some cases, be sourced by the will to indicate a conflict with the expectations about the world and/or the discourse. As Destruel et al. (2019) conclude, the stronger contrariness between speaker A's assertion and speaker B's rejection is, the more likely it will be that speaker B will resort to a cleft construction over the canonical order to express said contrariness.

Another function operating within the cleft identified by Prince (1978) for *informative-presupposition*, i.e. type 2, clefts, specifically relates to hedging and stance taking, or, in this case, lack thereof. By introducing information that is

discursively new but logico-semantically presupposed in the cleft-relative clause, the speaker presents it as an indisputable fact which is only unknown to the hearer. As a result, Prince (1978) compares type 2 clefts to hedges or impersonal constructions in which the speaker's stance towards the truth of the information is reduced to a minimum. She lists historical accounts, e.g. (43), and persuasive discourse, e.g. (44), as the most common settings in which type 2 clefts are productive.

- (43) It is through these conquests that the peasantry became absorbed into a single form of dependent Lord-tenant relationship. (Prince 1978: 900)
- (44) It is for this reason that Halle's argument against autonomous phonemics (Halle, 1959) is of such importance: it demonstrates that... (Chomsky 1972:127, as cited by Prince 1978: 900)

Delin (1992) explains this effect in terms of non-negotiability of the information introduced in the presupposed open proposition. Taking (44) as an example, the idea that Halle's argument is of significant importance is introduced by Chomsky as an undisputable fact that the hearer can only accept. With type 2 clefts, hearers are, therefore, not offered the possibility to reject the information at the time of utterance, which makes them a particularly useful device for argumentation.

At the discourse level, clefts have at times been equated with discourse markers working as cohesive, scene setting and circumstantial structures. It is mainly the case for type 2 clefts, which are said to hold less communicative dynamism (Collins 2006: 1710), in both English (Collins 2006) and French (Scappini 2013; Roubaud and Sabio 2015, 2018; Lahousse and Lamiroy 2017). In these clefts, the clefted constituent is typically anaphoric and belongs to a restricted lexical set, e.g. *là* 'here', *ainsi* 'how', *comme ça* 'how', etc. As a result, the clefted constituent, when used as a discourse marker, is non-paradigmatic and cannot contrast with another lexical item. The cleft as a whole therefore does not truly function as a bipartite structure pragmatically speaking but rather as a connective structure which builds on a large portion of the preceding context (Roubaud and Sabio 2018). This is illustrated in (45) with the *c'est comme ça*-cleft coming as an

answer to speaker A's question and as a conclusion to the three previous clauses detailing the life of the speaker B's grandparents.

(45) A: d'accord et euh v- vous vous savez pourquoi ils sont venus à Saint-Ouen pourquoi ils sont installés

B: eh ben [ils se sont mariés] et puis [mon grand-p- mon père ses parents étaient grainetiers à Pantin] et [il a voulu repren- prendre un acompte grainetier] et **c'est comme ça qu'ils sont ils ont é- exploré un peu toute la région** puis ça c'est trouvé que à Saint-Ouen il y avait un coin à vendre

A: d'accord d'accord OK ils se sont installés là d'accord (Roubaud and Sabio 2018: 12-13)

'A: okay and uh y- do you know why they came to Saint-Ouen why they settled down

B: well [they got married] and then [my grand f- my father his parents were seed sellers in Patin] and [he wanted to ma- make a down payment for a seed seller] and that's how they are they kind of e- explored the whole region and then it turned out that there was a land to sell

A: alright alright okay they settled down there alright'

In detail, cohesive clefts may take on a number of functions at the discourse level which include, but are not limited to, that of illustrating an argument (Lahousse and Lamiroy 2017), introducing a consequence (Roubaud and Sabio 2018), marking the speaker's stance (Roubaud and Sabio 2015), summarising (Dufter 2009a), etc.²⁶ Though of lesser interest for this study, it is worth noting that these functions are mirrored in written data in which clefts have been shown to act as stylistic devices (Carter Thomas 2002), cohesive structures (Garassino 2014), or as markers of authorial identity (Bourgoin 2017) aiming at guiding the reader and clarifying the speaker's desired interpretation (Carter-Thomas 2009).

²⁶ Cohesive clefts, as a subclass of *c'est*-clefts will be discussed in section 5.3.3.

Related to the cohesive use are the roles clefts may play in interactional settings. When used in conversations, clefts may serve structuring and rhetorical purposes by appearing as discourse openers (Prince 1978; Declerck 1984; Hedberg 1990; De Stefani 2008; Dufter 2009a; Garassino 2014; Hasselgård 2014), especially when introducing all new information, i.e. type 3 clefts. For Hasselgård (2014: 301), when used in written texts, these clefts allow the speaker to avoid taking responsibility with regard to what is shared in the presupposition. Garassino (2014: 121) relates the use of all-new clefts to the notion of ‘topic launching’ (Hasselgård 2004) whereby clefts introduce new referents which then become topics in the immediate co-text. In specific discourse contexts such as that of sport commentaries, clefts are particularly useful to mark the ever-changing identity of the topic under discussion (Nelson 1997). In the same vein, Garassino (2014) shows that type 2 clefts may introduce ‘topic relaunching’ when the given clefted constituent re-introduces a topic which had previously lost its topical status and on which additional new information is provided. For Hedberg (1990), clefts of the kind give rise to relevance implicatures which directly participate in the organisation of discourse. Besides heading a discourse segment, clefts may also appear in segment-final position (Hedberg 1990) and may, as a result, signal the end of a given conversational episode after which the hearer will be offered the conversational floor (De Stefani 2008). On this, Delin (1990) adds that clefts may also serve the opposite function of being a floor-holding strategy.

When put against the division of discourse-familiarity patterns, the different functions of *it*- and *c’est*-clefts can be summarised with the Table 7 below.

Type of <i>it/c'est</i> -clefts	Clefted constituent	Cleft-relative clause	Function
Type 1	new	given	focalisation contrastive focus
Type 2	given	new	hedging device non-negotiability cohesive marker topic relaunching
Type 3	new	new	discourse opener topic launching
Other			floor holding

Table 7. Overview of discourse functions of *it*- and *c'est*-clefts

As made evident by the summary of discourse uses, *it*- and *c'est*-clefts are multifunctional constructions. While the majority of discourse uses can be mapped onto a specific discourse-familiarity type, others, such as floor holding, do not appear to be linked to any specific cleft subclass. Overall, what this overview of the functions of *it*- and *c'est*-clefts shows highlights in the diversity of functional motivations, which further challenges the description of clefts as focus-marking devices.

1.3 Aims and research questions

Now that I have introduced the theoretical framework and the state of the art my research draws on, let me introduce the main goals and research questions the present study sets out to investigate. The aims of this study can be summarised as follows:

- (i) to provide a refined comparative account of the information structure of *it*- and *c'est*-clefts

Among the numerous studies focusing on the information structure of clefts, several authors have proposed binary or ternary typologies based, in both languages, on the distribution of given and new information and focus assignment patterns within the construction. While these have proven especially useful in accounting for the use of clefts in discourse, very few have substantiated their conclusions with quantifiable results (with the exception of Collins 2006 for *it*-clefts and Rialland et al. 2012 for *c'est*-clefts). In view of this, the primary goal of this study is to not only typify the different focus assignment and discourse-familiarity patterns observed in clefts in spontaneous conversations but also to quantify their frequencies. I present the results for English in Chapter 3 and for French in Chapter 4. By combining a qualitative and quantitative study of the information structure of clefts in relation to their semantic and pragmatic characteristics, I will be in a better position to address a number of recurrent questions in the literature such as *Are clefts true focalising constructions?*, *What discursive purposes do clefts serve?* or *Are clefts inherently contrastive?*. The systematic analysis is carried out on extensive datasets. With regard to the contrastive aspect of this study which I develop in Chapter 5, the qualitative-quantitative approach also allows me to better sketch the cross-linguistic points of convergence and divergence whose description has so far mainly been limited to the question of frequency of use in the two languages (Carter-Thomas 2009). This translates into the broad question *Do English clefts and French clefts display the same information structural behaviour with regard to focus-marking and distribution of discourse-new and discourse-given information?*

- (ii) to characterise the interplay between referential and relational information structure and the interaction between syntax and prosody

This study also proposes to closely examine the interplay between the different components under study, viz. between the referential and relational layers of information structure and between syntax and prosody. To achieve this goal, I distinguish the semantics of specification coded by the grammatical structure of clefts from the information structural patterns speakers map onto the

specificational relation. Given this, I ask the questions *Do the syntactic constituents of value and variable only code an argument focus-presupposition articulation as has been argued in pragmatic accounts? How do contrastive focus and presenting focus relate to the information shared in it- and c'est-clefts?* In the same vein, the present study aspires to determine the nature and extent of the moment-by-moment interaction between the syntax of the cleft, which highlights the value, and the prosodic choices speakers make to manage the information flow. On this, Portes and Reyles (2022) note that *c'est-clefts* are a good illustration of how the syntactic and prosodic systems independently participate but also combine to generate information packaging through specific overlapping strategies. By the same token, Vander Klok et al.'s (2018) experimental findings suggest that, when it comes to focus-marking, syntactic and prosodic strategies should not be taken as working either fully jointly or fully independently of one another. Instead, both studies support the view that the interaction between the two systems is gradient and context-bound. This raises the question *Does syntactic highlighting and prosodically coded focus assignment systematically co-occur in it- and c'est-clefts?* Whether the data from the London-Lund Corpus and Corpus de Référence du Français Parlé also warrant the same conclusion as that of Portes and Reyles (2022) will be addressed sporadically throughout Chapter 2 to 4.

(iii) to shed light on the hitherto overlooked reduced clefts

While the literature on clefts has acknowledged the existence of a reduced variant, only a handful of studies have explored the range of its formal and functional motivations. Its very existence has mostly been attributed to the fact that its variable is activated and in focus in the context hence making it deletable (Declerck 1988; Hedberg 1990, 2000). One of the goals of this research is to go beyond the 'high salience' analysis and determine whether reduced clefts serve specific purposes and if so, which ones. For this, in Chapter 6 I address the questions *Do reduced clefts constitute a full-fledged category of clefts? Which discourse contexts do reduced clefts typically appear in?*

- (iv) to offer a comprehensive methodology for the study of multi-faceted notions such as givenness/newness and prosodically coded focus

Finally, given the variety of the existing approaches to the notions of discourse-familiarity, e.g. textual vs. cognitive, and focus, e.g. functional vs. formal-pragmatic vs. semantic, the present study also aims at providing a methodological framework applicable to constructions such as clefts when investigated in corpus data. Hence, in Chapter 2, I build on Kaltenböck's (2005) analytical model of given and new information, which he originally developed for *it*-extraposition, and propose a refined version relevant for the study of discourse-familiarity in both written and spoken discourse. In order to circumvent the well-known limitations of corpus data with regard to speaker's assumptions, the model I propose relies on a strictly textual view of givenness and newness which takes the available linguistic discourse as its sole basis. In order to provide a complete methodology for the prosodic analysis of constructions like clefts, I also offer an framework for the instrumental comparison of prominences in English based on Esser's (1988) theoretical account of hierarchy of foci. As such, I offer concrete references points for the description of 'weak' and 'strong' stress which had remained vague in the typologies of clefts focusing on focus assignment patterns.

Beyond the mere characterisation of the use of cleft constructions, the goals of the present study are thus manifold and pertain to both theoretical and methodological gaps observed in the literature.

Chapter 2: Data and methodology

My approach to *it*- and *c'est*-clefts differs from the mainstream accounts in that it distinguishes the semantics of specification coded by grammatical structure from the many information structural patterns that can be mapped onto the constituents construed as value and variable. This involves examining the discourse-familiarity, i.e. given/new status, of the different constituents and the way prosodically coded focus, i.e. focal/non-focal status, is assigned within the construction. As such, the aim of this study is to investigate the information structural and prosodic behaviour of English *it*-clefts and French *c'est*-clefts using a corpus-based quantitative-qualitative approach. The choice of corpora of spoken interaction and methods was motivated by the principles established in Chapter 1, i.e. Halliday's functional approach to focus. This chapter details the data and methodology used for the investigation. In Section 2.1, I introduce the two corpora from which I compiled my datasets of spoken sound files. In Section 2.2, I describe the extraction process and the criteria used to distinguish clefts from cleft lookalikes. In Section 2.3, I focus on the prosodic analysis of the data. I explain how I re-analysed the pre-existing annotations of the English data and annotated the French data from scratch. In Section 2.4, I discuss how discourse-familiarity was assessed and I propose a refined analytical model for the distinction between different types of given and new. Finally, I conclude this chapter by summarising the main points of the methodological framework I adopt in this study.

2.1 Corpora

In this section, I introduce the two corpora of spoken data I used to investigate the information structural properties of *it*-clefts (Section 2.1.2) and *c'est*-clefts (Section 2.1.3). Given that cross-linguistic comparison of English and French is one of the aims, I first start by addressing the question of comparability between the two corpora in Section 2.1.1.

2.1.1 Comparability of corpora

Cross-linguistic studies of linguistic phenomena can be of two types and this affects the type of corpora used. Studies in contrastive linguistics tend to use comparable corpora containing monolingual or multilingual original texts while translation studies primarily make use of multilingual translation corpora (Granger 2010). This study falls within the scope of contrastive linguistics. This, together with the fact that spoken data in the form of sound files were needed, has a number of methodological consequences. Firstly, while translation corpora of English and French spoken data are in principle possible, e.g. in contexts of simultaneous interpreting, no such corpora of the required size are available. Moreover, Carter-Thomas's (2002) study on the use of French *c'est*-clefts and English *it*-clefts in translated texts has shown that the frequency of the former is significantly higher than that of the latter, which called into question the possibility to gather sufficient data for spoken English data. Finally, this study primarily aims at presenting a comprehensive study of information structural patterns in *it*-clefts and *c'est*-clefts in their own environment while the comparative aspect is a secondary goal. These methodological considerations led me to use two monolingual corpora consisting of original and non-translated texts with available sound files, namely the first London-Lund Corpus (LLC-1) and the Corpus de Référence du Français Parlé (CRFP). To ensure maximal comparability, I opted for two corpora of similar sizes containing spoken data of the same nature recorded in the same types of settings. This choice was further reinforced by the availability of a written transcription with the accompanying sound files for both corpora. A comparative overview of the LLC-1 and the CRFP can be found in Table 1 below.

	LLC-1	CRFP
Number of words	500,000	440,000
Date of compilation	1957-1983	2000s
Types of conversations	Dialogues and monologues	Dialogues and monologues
Types of settings	Private and public settings	Private and public settings
Audio files	Yes	Yes
Written transcription	Yes	Yes
Prosodic annotations	Yes	No

Table 1. Overview of the characteristics of the LLC-1 and the CRFP

2.1.2 London-Lund Corpus

The first London-Lund Corpus (LLC-1)²⁷ is a corpus compiled between 1957 and 1983 as part of two distinct projects, the Survey of English Usage (SEU) and the Survey of Spoken English (SSE) (Svartvik and Quirk 1980; Svartvik 1990). The goal of the LLC-1 is to provide an overview of the use of spoken and written British English. The corpus comprises 200 texts which amount to 1,000,000 words. Because the prosodic features of cleft constructions are an essential element in the research questions of this study, only the spoken part of the corpus was used. The spoken data represent 100 texts amounting to 500,000 words. As shown in Figure 1 below, the spoken texts from the LLC-1 include monologues and dialogues which range from spontaneous or prepared monologues to face-to face conversations and telephone calls.

²⁷ The second LLC-2 (see Pöldvere, Johansson and Paradis forthcoming) compiled in the 2010s had not been made available at the time this study was carried out. This is why this thesis uses the older LLC-1. I was supported in this by O’Grady’s (p.c.) experience with corpus-based analyses of the intonation of English, which found no major changes in the prosodic coding of information structure in the period that elapsed since the first LLC-1.

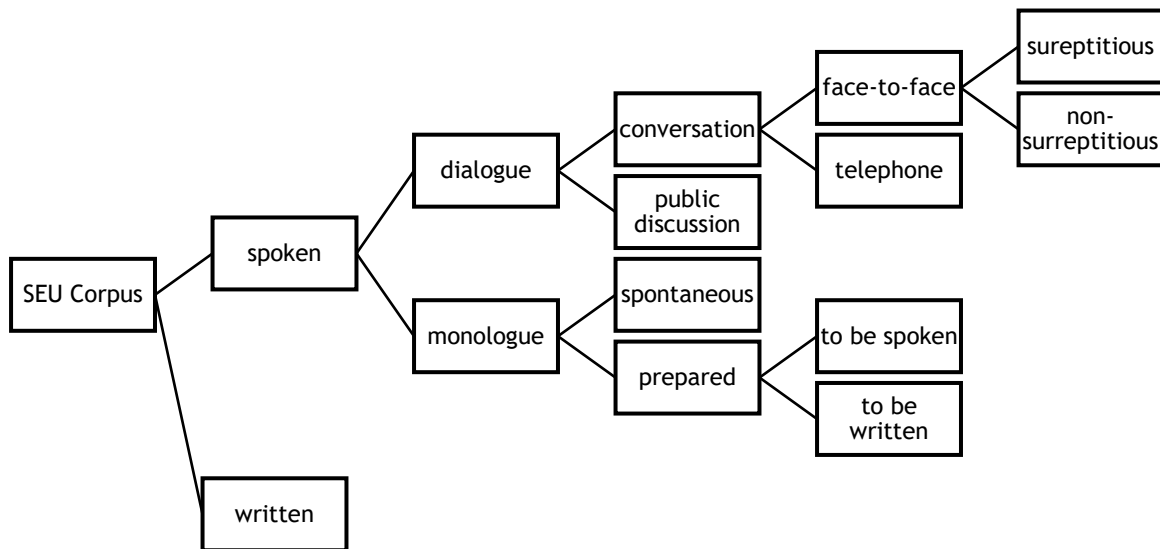


Figure 1. Overview of the texts of the Survey of English Usage (Svartvik 1990)

The LLC-1 was originally annotated according to Crystal’s (1969) model of prosodic and paralinguistic systems. Two written transcriptions and a grammatical analysis are available for the LLC-1. The full transcription includes phonological transcriptions, e.g. tone unit boundaries, nucleus location, tone movements, etc., and paralinguistic annotations, e.g. whisper and creaky voice, while the reduced one only contains the former. This study took the reduced transcription of the LLC-1 as starting point, changing and correcting some elements and omitting a number of features that were not relevant. A detailed description of the revision of the prosodic annotations can be found in Section 2.3.1.

2.1.3 Corpus de Référence du Français Parlé

The Corpus de Référence du Français Parlé (CRFP) is a corpus of spoken Metropolitan French compiled by DELIC (2004). It consists of 134 texts adding up to 36 hours of recordings and 440,000 words. In the same way the LLC-1 is meant to be representative of British English usage, the CRFP aims at providing an overview of the “general every-day usage” of French (*Autour du Corpus de Référence du*

Français Parlé 2004). Recordings were completed in the Parisian area and 37 cities of the provinces. The corpus includes monologues and dialogues recorded in private, professional and public settings as summarised in Figure 2 below.

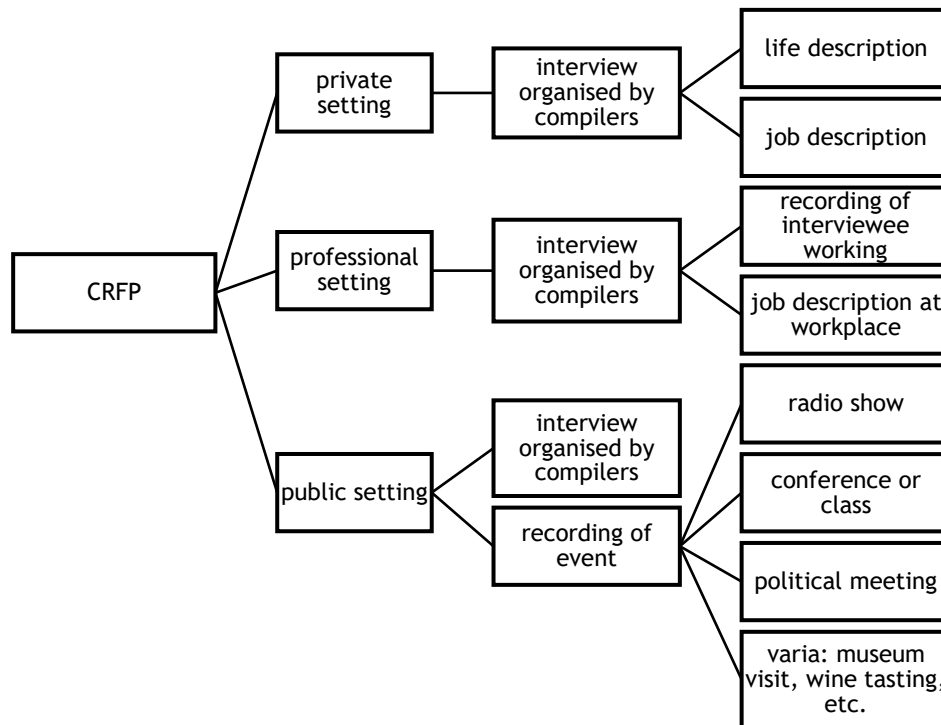


Figure 2. Overview of CRFP recordings

Unlike the LLC-1, the written transcription of the CRFP is solely orthographic and was completed according to the principles set out by the Groupe Aixois de Recherche en Syntaxe (GARS) (Blanche-Benveniste 1990: 228-229). The paralinguistic mark ups are limited to pauses and false starts, and no punctuation or prosodic annotations are included. I added prosodic annotations using a protocol I describe in Section 2.3.2.

2.2 Delineation and extraction of datasets

This section provides an overview of the methodology followed to delineate the classes of full *it-* (Section 2.2.1) and *c'est*-clefts (Section 2.2.2) and their reduced variant (Section 2.2.3). I then discuss the extraction of occurrences from the corpora (Section 2.2.4).

2.2.1 Recognition criteria for *it*-clefts

There exist in the literature a number of recognition criteria and tests one can rely on to delineate the class of *it*-clefts (see Delin and Oberlander 1995; Lambrecht 2001; Huddleston and Pullum 2002). These are mostly based on the inherent syntactic and semantic characteristics of *it*-clefts and can be used to distinguish clefts from sentences consisting of seemingly similar syntagms. The first and most frequently used test in the literature on *it*-clefts is known as the ‘decleftability’ (Lambrecht 2001) or ‘declefting’ (Delin and Oberlander 1995) test and consists in retrieving the canonical ‘declefted’ counterpart of the cleft. This criterion is based on the widespread assumption that the ‘propositional’ meaning of a cleft is fully equivalent to a simple proposition.²⁸ A clause like (1a) can, for instance be ‘declefted’ into the simple clause in (1b). Both the sequence *it is* introducing the matrix clause and the relative marker *that* linking the matrix clause and the cleft-relative clause are dropped and the positions of the syntactic constituents are shifted. Despite this reorganisation of the linguistic content, its meaning in relation to representation remains unchanged.

- (1) (a) it’s the present and our responsibility for the future that matters
(LLC-1)
(b) the present and our responsibility for the future matter

By contrast, the ‘declefting’ of the clause in (2a) leads to the simple clause in (2b) which is not grammatical. The sentence becomes grammatical again in (2c) by retaining the complementiser *that*, which shows that (2a) is not in fact a cleft but an extraposed subject.

²⁸ As detailed in Chapter 1, I depart from this position and instead view the grammar of clefts as compositionally coding its ‘representational’ semantics, whereby the identifying matrix asserts the state (see Delin 1992) of the identifying-specificational relation between the value and the variable.

- (2) (a) yes it is perfectly true that it includes those marginal cases that you've referred to (LLC-1)
(b) * yes it includes those marginal cases that you've referred to is perfectly true
(c) yes **that** it includes those marginal cases that you've referred to is perfectly true

As illustrated with (1) and (2), the decleftability test is particularly useful to separate clefts from extraposed subjects. It has been noted, however, that a declefted counterpart is not always available (Karszenberg 2018) and that the decleftability test is a test which can often, but not always, help to recognise clefts (Davidse et al. 2022). This is illustrated in (3) in which the *it*-cleft in (3a) can hardly yield the declefted counterpart (3b).

- (3) (a) it's not just port that goes into wine (LLC-1)
(b) ?? not just port goes into wine

That the decleftability test cannot always be used can be explained by the fact that the matrix clause and the antecedent - relative anaphor relation, in the compositional analysis of clefts which I ascribe to, involve meaning. It is particularly the case when clefts involve negation or quantified values. Apart from these cases, the fact that the decleftability test often works ultimately lies in the NP status of the antecedent which can, as such, fill the gap in the proposition in the cleft-relative clause (see Davidse 2000).

A second criterion for the identification of *it*-clefts relates to the grammatical status of the relative clause. Studies have indeed shown that clefts differ from constructions containing an embedded relative clause in that the cleft-relative clause does not form a constituent with its antecedent (Declerck 1988: 52; Davidse 2000, 2014, et al. 2022; Huddleston and Pullum 2002: 1416). On this basis, all constructions with an integrated relative clause can be discarded as non-clefts. This is the case of (4a), in which the relative clause *which entirely dissociates this movement from the actions of Mr Heath and his government* modifies its NP antecedent *decision* to which it is linked through the relativiser *which*. The relative

clause in (4a) is therefore of the restrictive kind while cleft constructions introduce a relative clause which is neither restrictive nor appositive. As such, the paradigm of relative markers available for *it*-clefts, e.g. *that*, *which*, *who*, \emptyset , *where*, is wider than that of appositive relative clauses but similar to that of restrictive relative clauses. The argument that (4a) is not a cleft is further reinforced by the impossibility to decleft it as shown in (4b).

- (4) (a) what I'm asking of conference it is this that conference stands firm by the decision we took so overwhelmingly at Brighton last year **it is a decision which entirely dissociates this movement from the actions of Mr Heath and his government** and puts us into full conflict with it (LLC-1)
- (b) ?? a decision entirely dissociates this movement from the actions of Mr Heath

In view of example (4), it is therefore crucial to note that the two criteria are best used when assessing cleft-like structures in their immediate co-text. Looking at the construction with subject *it* in (4a) in isolation could lead to a cleft diagnosis in which *a decision* acts as the value satisfying the variable *x dissociates this movement from the actions of Mr Heath and his government*. This reading is however revoked when we take the co-text into account.

Finally, *it*-clefts can also be distinguished from other types of copular clauses with *it* based on the non-referentiality of the introductory pronoun. As detailed in Section 1.2.1.1, *it* has been argued to be either an empty expletive pronoun (Chomsky 1971; Akmaijan [1973] 1979; Collins 1991; É. Kiss 1998; Lambrecht 2001; Huddleston and Pullum 2002) or a pronoun carrying some sort of semantic content (Bolinger 1972; Gundel 1974; Davidse 2000; Hedberg 2000; Reeve 2011). One common argument to both views is that *it* is never co-referential with a specific item from the preceding context which differentiates it from the introductory pronoun of ascriptive-identifying *it*-clauses.

2.2.2 Recognition criteria for *c'est*-clefts

Since *it*-clefts and *c'est*-clefts share the same syntactic properties, the criteria mentioned in the previous sub-section can also be applied to French *c'est*-clefts. This is illustrated in (5a) in which the *c'est*-clause can easily be converted from a cleft to a mono-clausal alternate, (5b). In (6), however, no declifted counterpart can be obtained from the *c'est*-clause in (6a) without altering the meaning conveyed.

- (5) (a) je vais reprendre mes activités mais très vite au début de l'année scolaire **c'est toi qui vas me remplacer** (CRFP)
 'I'm going to take up my occupation but very soon at the beginning of the academic year **it is you who's going to replace me**'
- (b) je vais reprendre mes activités mais très vite au début de l'année scolaire **tu vas me remplacer**
 'I'm going to take up my occupation but very soon at the beginning of the academic year **you're going to replace me**'
- (6) (a) c'est un travail euh qui est assez euh moi je le trouve très intéressant évidemment **c'est le travail que j'ai choisi** (CRFP)
 'it's a job that's uh that's rather uh I find it very interesting obviously **it's the job I chose**'
- (b) ?? c'est un travail euh qui est assez euh moi je le trouve très intéressant évidemment **j'ai choisi le travail**
 'it's a job that's uh that's rather uh I find it very interesting obviously **I chose the job**'

Similarly to *it*-clefts, *c'est*-clefts are also described as consisting of a matrix clause followed by a non-integrated relative clause (Rouget and Salze 1986). *C'est*-clefts should therefore be distinguished from other types of *c'est*-clauses with a restrictive relative clause such as (7) in which *une version* is the nominal antecedent of the relative clause *qui est assez contestée*.

- (7) et ces mecs-là en ont profité pour prendre du pouvoir et asseoi--
asseoir en fait leur pouvoir sur en prenant des terres etcetera **c'est
une version bon qu qui apparemment est assez contestée (CRFP)**
'and those guys used this opportunity to take power and impto- actually
impose their power on by seizing lands etcetera **it's a version well th
that's apparently quite disputed'**

Both integrated and non-integrated relative clauses rely on the same range of relative markers which include *qui*, *que*, *où*, *dont*, etc. (Muller 2003). In comparison with *it*-clefts, *c'est*-clefts cannot take \emptyset . *C'est*-clefts should also be differentiated from cases of extraposed subjects as in (8).

- (8) nous sommes revenus sur euh sur la France moi euh avec beaucoup de
beaucoup de regrets parce que **c'est vrai que j'aurais aimé rester plus
longtemps (CRFP)**
'we came back to uh to France moi uh with a lot a lot of regret because
it's true that I would have liked staying longer'

Here, the *c'est*-construction seemingly resembles a cleft with its structure consisting of the pronoun *ce* followed by the copula *être* and a *que*-clause. The deletion of *ce* and the shifting of the sentence-final clause to the front bring about the extraposition reading *que j'aurais aimé rester plus longtemps est vrai*.

An additional recognition criterion shared between *it*- and *c'est*-clefts recognised by authors such as Kunstmann (1990), Muller (2003) and Rouquier (2007) is the non-referentiability of *ce*. While the pronoun *ce* in clefts has been argued to be either an expletive pronoun or a pronoun holding some meaning, one common argument is its absence of co-referentiality between *ce* and any other linguistic constituents outside of the construction. When *ce* is referential and identifies a specific element, Kunstmann (1990) argues that it relates to the upcoming relative clause, which is therefore not a cleft-relative clause but a regular relative clause. This is illustrated with (9) in which *ce* is used anaphorically to refer to the NP *MIDI* introduced in the preceding discourse before the leftward boundary of the *c'est*-cleft.

- (9) à savoir qu'on a un poste euh informatique MIDI oui MIDI c'est tout ce qui est c'est tous les **c'est tous les systèmes de de de de codes ou de communication qui fonctionnent avec les euh hum avec les synthétiseurs** euh (CRFP)

'worth noting that we have a station uh computer MIDI yes MIDI it's everything that is it's all the **it's all the systems of of of of coding or communication that function with the uh with the synthesisers uh'**

As a structure with an introductory anaphoric pronoun *ce*, (9) can also be discarded as a non-cleft on account of the relative clause being a restrictive relative clause modifying the co-referent of *ce*, that is the head noun *les systèmes de codes ou de communication*. (9) illustrates the tendency for French to have left dislocation or double subjects.

Whether they are used to identify *it*- or *c'est*-clefts, the different recognition tests listed above, i.e. decleftability, status of the relative clause and non-referentiality of *it/ce* should be used in combination so as to keep the margin of error to a minimum. In other words, *it*- and *c'est*-clefts should be decleftable, have a relative clause that is non-restrictive and whose introductory pronoun *it/ce* should not display any co-referentiality beyond the construction boundaries.

2.3.3 Recognition criteria of reduced *it*- and *c'est*-clefts

While the different criteria described in the previous two sections effectively allow to extract full *it*- and *c'est*-clefts from corpora, they may be insufficient in some cases for reduced clefts, i.e. clefts with an omitted cleft-relative clause. Like for full clefts, the postcopular complement of reduced clefts identifies a value for a variable. However, while this variable is explicitly expressed in the cleft-relative clause of full clefts, that of reduced clefts has to be retrieved from the preceding text, which may prove challenging, e.g. when it is only inferable and not textually evoked, and thus makes the decleftability test inadequate. Another issue lies in the perceived ambiguity of the co-referentiality of *it/ce* which may result in unclarity in the meaning ascribed to the sentence. More specifically, there exist four

different lexicogrammatical structures realised by the string *it/ce + be/être + NP* but which all code different meanings.

In the first reading, the sentence is a reduced cleft which has a specificational meaning. As explained in Section 1.2.2, the introductory pronoun *it/ce* is non-referential and cannot be replaced by *that/cela* as it does not identify a specific referent. The postcopular complement is construed as the value filling the semantic gap of a variable. In other words, the postcopular complement is not just interpreted in terms of its own referent but as the qualifying value within the set of alternatives established by the variable. In the cleft reading of (10), for instance, the variable can be reconstructed as *qu'on est en entreprise* or *qu'on est en alternance*, thus referring back to the description of *alternance* provided by the speaker at the beginning of the excerpt. The postcopular NP *qu'une fois par mois* is thus not just an identifier but also the value *x* of the variable aforementioned. As such, it represents the one value that qualifies while the other alternatives within the paradigm, e.g. *deux fois par mois*, *trois fois par mois*, etc., are implicitly discarded as non-qualifying options.

- (10) c'était euh du travail en entreprise donc c'est là qu'on apprenait ou qu'on apprend le plus en fait donc c'est oui alternance est travail et école en même temps mais c'est pas euh **c'est qu'une fois par mois** c'est il y a pas d'autres d'autres moments (CRFP)
 'it was uh company work so that's where we learned or we learn the most actually so it's yes work/study training programme is work and school at the same time but it's not uh **it's only once a month** it's there are no other other moments'
- (11) A: as you probably know we teach a set author course at Beaton which is very very rigid especially in for instance in the summer term which is a a terrible time (...)
 B: what is the disadvantage of the set author course (...)
 A: the advantages are quite good for for two years or so in so far as you learn an awful lot about authors you didn't perhaps know too much about to start with but the second time the cycle comes round well you're beginning to petrify a bit vis-a-vis the material I

B: you mean **it's always the same set authors** is that what you're saying is that what you mean by rigid (LLC-1)

In the same vein, the NP *the same set authors* in (11) can also be analysed as the value intrinsically related to the reconstructed variable *you teach x*.

In the second reading, *it/ce + be/être + NP* corresponds to a specificational-identifying clause with referential subject *it/ce*. This time, *it/ce* anaphorically points back to an antecedent which is typically found in the preceding discourse. Both (10) and (11) are ambiguous instances of copular clauses in that they can be analysed as reduced clefts, as shown previously, but also as specificational-identifying clauses introduced by referential *it/ce*. In (10), *ce* can be taken to anaphorically refer to the NP *alternance* given in the preceding clause. This analysis would be corroborated by the false start *c'est pas euh* where the speaker adds information on the referent just mentioned. The postcopular complement *qu'une fois par mois* would then solely be an identifier. In the case of (11), *it* can be analysed as being co-referential with *a set author course*. The specificational-identifying clause then is *a set author course is always the same set authors*.²⁹

The string *it/ce + be/être + NP* can also exemplify ascriptive-identifying copular clauses, or what Declerck (1988) calls 'descriptively-identifying' clauses, with referential subject *it/ce*. Like specificational-identifying clauses, *it/ce* refers to an antecedent which can be retrieved from prior discourse and can be replaced by *s/he* in the case of proper names. In Halliday's (1967a: 224) terms, the subject referent is the identified, i.e. the entity to be identified, while the postcopular complement is the identifier, i.e. the entity bringing about identification. The identifier is therefore processed in terms of its descriptive information *as such* and

²⁹ The same referentiality ambiguity is identified by Ward et al. (2007) for English reduced *that*-clefts with variable-referring *would* such as (10') below where *that* can be construed as referring to the open proposition *x is the best singing* but also as a co-referent of *the best known Zip Code in the nation*.

(10') Targeted at 6- to 12-year-olds, the "Zoom" revival, like the original, pits seven telegenic youngsters...against a variety of activities and challenges that viewers send in.... If nothing else, it should once again make that of Boston PBS affiliate WGBH, where this is produced, the best known Zip Code in the nation. **That would be [singing] 0-2-1-3-4.** (Ward et al. 2007: 85)

does not display any sort of relation to an inferred variable. The central semantic relation is between the subject/identified and the postcopular complement/identifier, or in Van Praet's (2019) terms, the describee and the description. This type of clause is illustrated in (12) in which the postcopular complement links the new and defining information of the woman's name, *Mrs Billie Pavane* to the subject *she*. Here, *It's Mrs Billie Pavane*, with *it* referring to *she* (the woman physically identifiable in the context), is also possible.

- (12) 'So she's not - ' Clements looked at his notebook; he still carried it like an old family heirloom. 'Not Mrs Belinda Paterson?' 'No. **She's Mrs Billie Pavane**. She's the wife of the American Ambassador.' (Wordbanks)

Finally, the last type of lexicogrammatical structure realised by the sequence *it/ce + be/être + NP* is that of reduced extraposition. *It/ce* is in this case non-referential and the extraposed subject is inferable from the preceding context. (13) can, for example, be interpreted as *it's all guesswork [what Shakespeare's private life was]*.

- (13) no I just think we've got to admit we haven't got anybody to put against Burns in England you see we haven't got a self-made man who was a great gay lover and a fine poet and all we can do is rake up somebody like Piers Plowman who was a literary oddity otherwise we can't do it Shakespeare is too long ago we don't know about his private life **it's all guesswork** (LLC-1)

In view of the double reading accepted for a number of copular *it*-clauses, e.g. (10) and (11), which the prototypical criteria for the delineation of full *it*- and *c'est*-clefts do not allow to unequivocally and exhaustively disambiguate, criteria tailored for reduced clefts specifically are needed. In the remainder of this section, I attempt to develop new criteria based on the semantic and informational features of reduced clefts. These include high degree of predictability of the variable, presence of the state of affairs in the preceding context, and presence of a (non-)qualifying set.

The omission of the cleft-relative clause of reduced clefts is traditionally linked to its high degree of recoverability from the preceding discourse (Declerck 1988; Hedberg 2000). As such, the cleft-relative clause of any reduced cleft should be, to a certain extent, easily reconstructable and retrievable based on the information shared by speakers beforehand. This is the case with (14) and (15) where the variables *that he suggested me* and *qui mangent* are given verbatim in the preceding context.

- (14) A: why did Mr Power suggest me - because I was the first name that came into his head
 B: absolutely he said he there's no question about it (...)
 A: **it's not not because he thought that I knew the Ford Foundation or anything like that [~~that he suggested me~~]** (LLC-1)
- (15) A: et qui est-ce qui mange dans ces restaurants à part vous euh est-ce que vous avez pu
 B: euh ben **c'est des c'est des locaux en fait [~~qui mangent dans ces restaurants~~]** des gens de là-haut (CRFP)
 'A: and who is it that eat in those restaurants except you uh could you
 B: uh well it's **it's locals actually [~~that eat in those restaurants]~~'**

That a variable can be inferred from the prior discourse shows that the postcopular complement acts as a value and not just as a complement with the role of identifier. The high predictability of the variable therefore allows, to a certain extent, to disambiguate examples which allow both readings, i.e. identifying copular clause with anaphoric *ce/it* vs. reduced cleft.

In some cases, the variable might be more than just inferable and might be explicitly given in the preceding context, in which case the cleft reading is more strongly supported. The state of affairs can, for instance, be introduced in an introductory subclause (underlined in the following examples) similar in form to that of pre-modified reduced *it*-clefts as described by Declerck and Seki (1990). I analyse these examples and attempt to define their class in more detail in Section 6.2.3.

- (16) I always feel that if I fancy something it's because I need it [~~that I fancy it~~] (LLC-1)
- (17) si je m'intéresse au tarot c'est pas parce que je te vais tirer les cartes [~~que je m'intéresse au tarot~~] j'en ai rien à foutre de tirer les cartes mais c'est pour la l'initiation symbolique il y a derrière le tarot (CRFP)
 'if I'm interested in tarot it's not because I'm going to draw cards [~~that I'm interested in tarot~~] I don't give a fuck about drawing cards but it's for the symbolic initiation that's behind tarot'

In both (16) and (17), the *if*- and *si*-clause before the reduced cleft introduces the content of the respective variables *I fancy something because x* and *je m'intéresse au tarot parce que x* thus rendering them wholly predictable. When the variable cannot be retrieved as such, lexical items may nevertheless allow the analyst to infer the open proposition carried by the cleft-relative clause. This is the case in (18) below, where A justifies his dislike of comic book drawing with it being too closed ('*trop fermé*'), after which B proposes a value, i.e. *le cadre de l'image*, for the inferred variable *x est fermé*.

- (18) A: ben je suis un peu spécialiste de la bande dessinée ici à la bibliothèque déjà et puis j'en ai fait un peu oui parce que ça me plaisait et puis j'ai essayé puis en fait je me suis rendu compte que ça ne me correspondait pas c'était trop carré trop fermé pour moi
 B: trop carré trop fermé
 A: oui
 B: c'est-à-dire c'est le cadre de l'image [~~qui est fermé~~]
 A: eh ben oui il y a dé-- il y a déjà ça on est limité à une planche avec tel ou tel cadre (CFRP)
 'A; well I'm a bit of a specialist of comic books here at the library first of all and then I also did a little bit of it yes because I liked it and then I tried then actually I realised that it wasn't for me it was too strict too narrow for me
 B: too strict too narrow
 A: yes

B: what do you mean **it's the frame** [~~that's too narrow~~]

A: well yes there's- there's that already we're limited to a plate with one frame or the other'

Whether the information carried by the cleft-relative clause is directly recoverable or inferable, I consider its high predictability and its presence in the context as a determining clue for the recognition of reduced clefts. When no variable can be reconstructed, I assume that we are dealing with monoclausal *it*- and *c'est*-constructions and discard them from the dataset.

The third recognition criterion I propose relates to the identifying-specificational meaning coded by the grammatical structure of *it*- and *c'est*-clefts. Through their syntax, clefts establish a relation between one (or more) value(s) introduced in the postcopular position in the matrix clause and the associated variable expressed as an open proposition cleft-relative clause, which, in the case of reduced clefts, is elided. The set of qualifying value(s) (often singleton) are identified in a context where some potential, but ultimately non-qualifying values, may be more or less explicitly available. Unless stated otherwise, the non-qualifying values are always discarded on the basis of the implicature of exhaustivity triggered by the cleft, i.e. *it is X and only X*. The comparison between the qualifying and non-qualifying values introduced with the exhaustivity implicature gives rise to a notion of 'contrast' which is established on the representational meaning. When one (or more) values from the non-qualifying set are expressed, then the likelihood of the *it*-construction being a reduced cleft is very high. However, given the fact that the implicature exhaustivity may be weaker in some cases (Destruel et al. 2019), I do not treat this criterion as being systematically excluding but rather as being complementary to the first two criteria. The non-qualifying alternatives can be expressed in a paired full cleft as a realised contrast (Scappini 2013), whereby contrast between two alternatives is explicitly expressed, as in (19), or as lexically-suggested contrast with an exhaustive particle as in (20).

- (19) A: is this so with all tapes or is it just your particular machine
B: no no **it's just this particular set up here** [~~for which it is so~~] (LLC-1)

- (20) A: vous passez deux ans à regarder
B: ouais regarder puisque la plupart en apprentissage **c'est que shampooings couleurs [~~que l'on fait~~]** après petit à petit ben on fait d'autres choses (CRFP)
'A: you spend two years watching
B: yeah watching since most of the work/study programme **it's only shampooing dyeing [~~that we do~~]** then gradually well we do other things'

Using clearly established criteria such as high degree of predictability of the variable, presence of the state of affairs in the preceding context, and presence of a non-qualifying set of values, allows one to generate coherence and uniformity in the treatment of reduced clefts.

2.2.3 Extraction of data and building of the datasets

In order to build datasets from the LLC-1 for English and the CRFP for French, occurrences of full and reduced *it*-clefts were retrieved by looking for all sequences of *it* + form of *be* while full and reduced *c'est*-clefts were retrieved by looking for all sequences of *ce* + form of *être*. More specifically, the LLC-1 was queried for *it*-clauses with a copula in the present and past tenses and with and without a negation marker. The same basic queries were used for *c'est*-clauses in the CRFP, but an additional query was added to account for occurrences in which the copula agrees with a plural subject. These very general queries allowed me to retrieve full and reduced clefts at the same time and full *it*-clefts with zero relative marker [∅], e.g. *I thought perhaps it was the the soprano had got drowned* (LLC-1), which would have otherwise been overlooked. The strings used for the queries of both corpora are listed in (21) below.

- (21) LLC-1: *it + is / it + 's / it + is not / it + isn't / it + was / is it / was it*
- CRFP: *c' + est / ce + sont / ce + n'est / ce + ne sont / c' + était / c' + étaient / ce + n'était / ce + n'étaient*

Since the prosodic annotations of the LLC-1 are directly integrated in the text file of the written transcription, and hence make exhaustive automatic searches difficult to complete, the corpus was queried manually by individually reviewing each hit of the sequences in (21). By contrast, since the written transcription of the CRFP is void of any annotation, the extraction of *c'est*-clauses was done automatically with AntConc (Anthony 2019). Using the different recognition tests described in the previous sections, all instances of *it*- and *c'est*-clauses were manually sorted so as to retain clefts only. Among the discarded constructions were for instance copular sentences with referential *it*, copular sentences with a restrictive relative clause, and extraposition constructions. An overview of the datasets obtained is given in Table 2 below. It includes the raw and normalised frequencies per 10,000 words of the occurrences initially extracted and the tokens of full and reduced *it*- and *c'est*-clefts making up to the LLC-1 and CRFP datasets.

	Occurrences extracted		Full clefts		Reduced clefts		Total of clefts	
	RawF	NormF	RawF	NormF	RawF	NormF	RawF	NormF
LLC-1	4695	93.9	143	2.9	95	1.9	238	4.8
CRFP	9341	212.3	392	8.9	88	2	480	10.9

Table 2. Raw frequencies (RawF) and normalised frequencies (NormF) per 10,000 words) of extracted and retained occurrences

2.3 Prosodic analysis

The following two sections detail the procedure followed to carry out the prosodic analysis of *it*-clefts (Section 2.3.1) and *c'est*-clefts (Section 2.3.2).

2.3.1 Review of prosodic annotations and analysis of English data

Homing in on the question of how speakers combine the prosodic marking of focal vs. non-focal and the distribution of new vs. given information to control the progression of discourse, this study focuses on specific prosodic features of cleft constructions.³⁰ To achieve an exhaustive overview of the various information structural patterns available for clefts, a number of prosodic features relating to the coding of information focus are analysed. These include tone units, nuclear accents, i.e. the main pitch accents, tones, i.e. types of pitch movement, and onsets, i.e. the first prominence within the tone unit.³¹

For the LLC-1, I verified and adapted the original prosodic annotations, which were based on auditory analysis only. Among the different prosodic features coded, only a handful relate to my research questions. More specifically, the present study focuses on the prosodic features which code the different informational structural meanings taken on by clefts (Halliday 1963; O’Grady and Bartlett 2019). These include the different prosodic choices made by speakers which are subsumed under the three systems of the intonation system of English, namely tonality, tonicity and tones (Halliday 1967b). Tonality corresponds to the segmentation of discourse into tone units whose boundaries (marked by #) are typically accompanied by a number of pitch discontinuities such as pauses and final lengthening (Cruttenden 1997; Dehé and Braun 2013). Tonicity is the placement of the nuclear accent within the tone unit. The syllable bearing the nuclear accent is referred to as the tonic syllable, or simply tonic, and displays the main pitch variation. The segment preceding the tonic is the pre-tonic or pre-nuclear segment and the one following the tonic the post-tonic segment or post-nuclear tail. Finally, tones are the different pitch movements occurring on the tonic. These are Fall \, Rise /, Rise-Fall /\, Fall-Rise \/ and Level

³⁰ Section 2.3.1 is heavily inspired from the methodological description offered in Bourgoin et al. (2021).

³¹ A glossary summarising the definitions of the key terms pertaining to the English and French intonation systems is provided at the end of this thesis.

tone =. The exhaustive list of annotations and corresponding prosodic features included in the LLC-1's transcription are shown in Table 3 below.

Prosodic feature in the LLC-1 annotations	Code	Included in revised annotation (Y/N)
Fall	\	Yes
Rise	/	Yes
Rise-fall	/\	Yes
Fall-rise	\/	Yes
Onset	^	Yes
Level tone	=	Yes
Brief pause	.	Yes
Unit pause of one stress unit	-	Yes
Normal stress	'	No
Heavy stress	"	No
Higher pitch than preceding syllable	:	No
Booster higher than preceding prominent syllable	!	No
Tone unit boundary	#	Yes
Subordinate tone unit	}	No

Table 3. Overview of prosodic annotations in the LLC-1

Along with tone unit boundaries and tones, I also retained pauses and onsets. The remaining annotations, e.g. type of stress, booster, pitch height, which were studied in a more systematic manner with a combination of auditory and instrumental analysis than the original annotations, were removed from the coding. To improve the readability of examples, only annotations pertaining to the topic of the section they are found in are included.

In addition to the simplification of the prosodic coding, I tested the accuracy of the annotations by carrying out an inter-rater auditory and instrumental

analysis³², visualising the sound wave of each cleft with the assistance of Praat (Boersma 2001). Upon comparing my analysis with the original prosodic transcriptions of the LLC-1, I implemented changes of three types.³³ I first corrected the analysis of tones in a number of cases based on the pitch curve shown in Praat and on the auditory analysis of the corresponding segment. I also implemented two systematic changes regarding compound tones and subordinate tone units, which are part of the original LLC-1 transcription. Compound tones were defined by Halliday (1967b) as the fusion of two tones, yielding two tonics within a single tone unit. The existence of compound tones was rejected by Tench (1996) and O’Grady (2017) on two criteria. Firstly, Tench (1990) shows contra Halliday (1967b) that a pre-tonic segment can in fact be inserted before the second tonic, which justifies adding a tone unit boundary breaking the compound tone into two. Secondly, Tench (1990: 51) and O’Grady (2017) point out the incompatibility between Halliday’s (1967b) postulate that one tone unit always codes one information unit and the very existence of compound tones. Example (22b) illustrates the changes to the original transcription (22a) for the two issues just discussed.³⁴ The final tone is falling rather than rising, as shown by the Praat image in Figure 3, and the compound tone is reanalysed into a sequence of two distinct tones.

- (22) (a) he said you’re ^sure it’s Marks and Sp\arks you’re going to w\ork
for# (LLC-1)
(b) he said you’re ^sure it’s Marks and Sp\arks# you’re going to w\ork
for#

³² I am grateful to Gerard O’Grady for his assistance in the inter-rater review of the prosodic annotations.

³³ Such a comparison is possible because LLC-1 was transcribed following the annotation system of Crystal (1969), which like Halliday’s (1967b) system, is rooted in the British School of intonation.

³⁴ To improve readability, I only include prosodic annotations in the clefts while the preceding and following co-text are stripped of all coding.

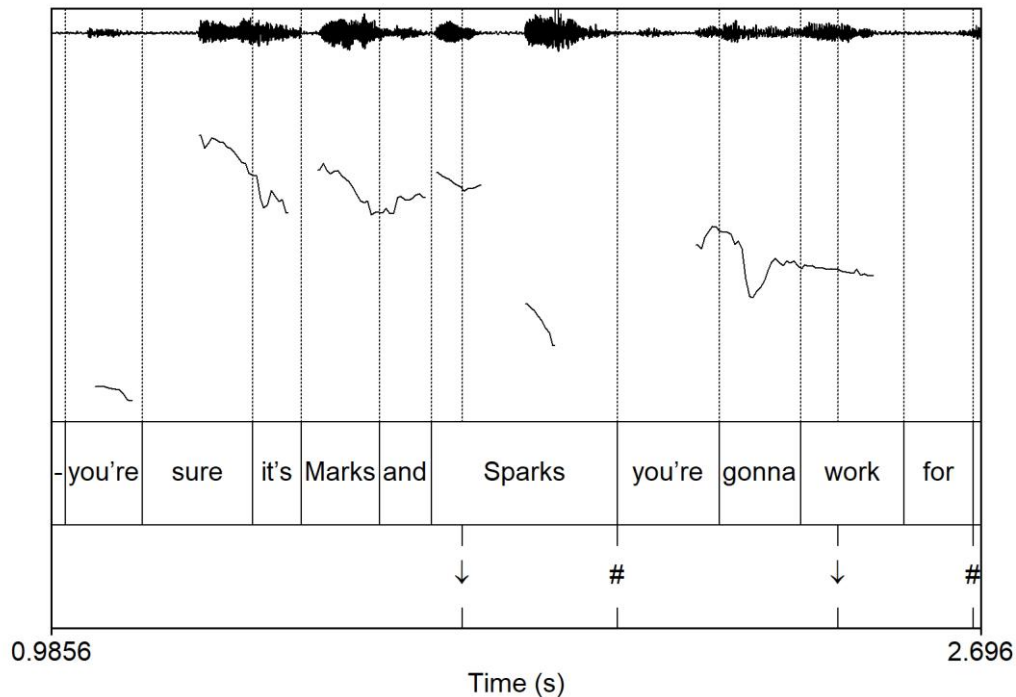


Figure 3. Prosodic realisation of (22)

Subordinate tone units, indicated by braces in the LLC-1 transcription as in (23a), were also eliminated in view of the flat nature of phonological patterning which prohibits any kind of recursion (O’Grady 2013). They were either replaced by full-fledged tone units, as shown in (23b), or fused with the preceding or following tone unit resulting in a single tone unit with one tonic as in (24b).

- (23) (a) n\o# it’s the \Union# - - ^as dist\ /inct from SRC {that runs th/at#}#
(LLC-1)
(b) n\o# it’s the \Union# - - L^as dist^M\ /inct from SRC# that runs
th^M/at#
- (24) (a) ^it’s the {r\olling} \ /out really# that . g\ives# the ^flaky p/astry#
its ^very individual ch\aracter# (LLC-1)
(b) ^it’s the r\olling out really# that . g/ives# the fl/aky pastry# its
very individual ch\aracter# (LLC-1)

Finally, I analysed the onsets, i.e. the first accented syllable of the intonation unit, which may be a pitch accent preceding the nuclear accent, e.g. as in (23b), whose

onset status is symbolised by \wedge , or the nuclear accent itself if it comes first or is the only accent in the tone unit (O'Grady 2010, 2014a). Onsets are analysed in terms of three degrees of relative pitch height, high, mid and low, indicated by small capitals H, M and L immediately prior to the onset. Relying on an instrumental analysis of the pitch curve, I assessed the pitch height of each onset relative to that of the previous onset in the preceding tone unit. The typical threshold I assumed for significant step ups, i.e. movements from one plateau to another, was 0.05 logHertz or more.

On the basis of this analysis, I typified and quantified the prosodic patterns realising information management in the *it*-clefts. The first question I addressed is one that has been central to the various typologies of clefts as information packaging constructions proposed in the literature, i.e. which constituents of clefts carry prosodically marked information foci? I categorised each pattern in terms of the focal/non-focal status of the value and of the cleft-relative clause if expressed, and the copula *be*. The status of the copula was only included in the list of patterns if it was focal.

The second issue I investigated was whether the foci differ in terms of degrees of prominence if a cleft has more than one focus. This operationalises the hypothesis that the focus on the value is realised by a 'stronger' accent than foci on other parts of the cleft (e.g. Prince 1978; Declerck 1984). Relying on instrumental analysis of the pitch curve, I examined the question of different degrees of prominence of nuclei with reference to Esser's (1988) hierarchy of foci. I posited a hierarchy based on the combination of two parameters: pitch height of the nucleus and tone movement. To determine the pitch height of nuclei, I queried its articulatory correlate, the fundamental frequency F0. This allowed me to distinguish three levels, high, mid, and low, for each type of tone movement. Tone movements were then ranked in the following order: High fall > mid fall > low fall > high rise > mid rise > low rise (Esser 1988; Van Praet 2019). No level tones were found in any of the cleft tokens and were therefore not included in the hierarchy. Rise-fall and fall-rise are included under the fall and rise categories respectively. For values and/or variables realised with multiple tone units, I only considered the highest-ranked tone movement. Thus, when both value and variable have a final falling tone, the value is considered higher if the pitch height of the information focus and the fall are higher and vice-versa. The value is also labelled as higher in

the hierarchy when the value has a final falling tone and the variable a rising tone. By contrast, if the value has a final rising tone and the variable a falling tone, the value is categorised as being lower in the hierarchy. The value and variable are treated as being at the same level in the hierarchy when they have the same final tones and the difference in pitch height is not auditorily perceptible. If step ups were noted in the auditory analysis, they were typically no less than 0.05 logHertz.

2.3.2 Prosodic annotation and analysis of French data

Unlike the LLC-1, the CRFP does not offer any pre-existing prosodic annotations in the written transcription of the sound files. Because this study aims at sketching out a comprehensive survey of the different prosodically coded patterns of focus assignment of clefts, the data were manually annotated for a number of parameters. This was done for the 480 clefts and, when needed, the immediate context. The annotation process was carried out auditorily and instrumentally with Praat. To stay in line with the more recent studies describing the prosodic coding in *c'est*-clefts (e.g. Avanzi 2011; Mertens 2012), and to make the comparison of results easier, I relied on the annotation system developed by Mertens (2006, 2008, 2012, 2019) for the representations of prosodic forms. Like Halliday (1967a), Mertens (2006, 2008, 2012, 2019) considers prosody to have a full-fledged functional role that is not merely derived from syntax. His descriptive framework subsumes a number of aspects relating to French intonation system which are speech segmentation, accentuation and tone contours and which are similar to Halliday's (1967a) tonality, tonicity and tones systems.

Starting with speech segmentation, Mertens (2006) defines the basic prosodic unit of his system as the *intonation unit* ('groupe intonatif') which is built around prosodic stress. More specifically, Mertens (2006) treats the *final accent* ('accent final'), i.e. the mandatory stress located on the last full syllable of the unit, as the marking strategy for the internal organisation of the intonation unit. The intonation unit may also contain an *initial accent* ('accent initial'), which, unlike the final accent, is optional and which typically appears on the first syllable of a word. There may be optional unstressed syllables appearing before the initial accent and in

between the initial and final accent. Intonation units in French are typically realised as follows:

- (25) Intonation Unit = ((unstressed) (AI)) (unstressed) AF (appendix)
(Mertens 2006: 8)

The parentheses mark the optionality of the element. Thus, an intonation unit, in its minimal realisation, consists of at least one accented syllable such as ‘*oui*’. By contrast, a maximal intonation unit is made up of five parts. These include the unstressed syllables leading up to the initial accent, the initial accent, a series of unstressed syllables, the final accent, and an appendix. The appendix corresponds to the final part of the utterance which occurs after the final accent, and which is realised with an extra-low flat contour and a compressed register. The optional syllables forming the appendix are unstressed.³⁵

With regard to accentuation, Mertens (2006) distinguishes between the final accent (AI) and initial accent (AF) on the basis of their distribution, function and realisation. Only the final accent is obligatory in the intonation unit whereas the initial accent is optional. Functionally, the final accent indicates where the rightward prosodic boundary of the intonation unit is, i.e. at the right edge of the item bearing the accent or of the last item of the appendix, while the initial accent serves emphatic or expressive purposes. The final and initial accent also differ in their articulatory properties. The final accent is characterised by a lengthening of the syllable by which it is carried and by pitch movement, e.g. rise, fall, etc. It is also prototypically accompanied by prominence in the intensity curve (Mertens 1993: 4) and may, in some cases, be followed by a pause (Mertens 2019: 67). By contrast, the initial accent generally triggers a lengthening of the initial consonant or the insertion of a glottal stop in case of an initial vowel (*ibid.*). The pitch contour on the initial accent shows no movement and is instead static, e.g. high or low

³⁵ Because the appendix consists of unstressed material, Mertens (2006: 9) remarks that it could alternatively be considered as a special type of intonation unit in which no stress occurs. However, because the appendix necessarily follows certain final tones, e.g. L-L- and H/H, it cannot appear on its own and is therefore best analysed as an optional part of the intonation unit. I follow Mertens’s (2006) reasoning and also treat it as being an optional post-focal tail.

(Mertens 1993: 4). The initial accent can be borne out by all kinds of lexical items, including clitics, and may be preceded by a pause (Mertens 2019: 67).

As far as pitch contours are concerned, Mertens (2006) offers an overview of the different relative pitch heights, which he refers to as ‘tones’³⁶, available for each position of the intonation unit. These are summarised in Figure 4 below.

unstr		AI	unstr		AF	appendix
l	l	H	l	l	L-L-	l...l-
h	h	L	h	h	H+H+	
					HL-	
					H/H	h...h
					/HH	
					\HH	
					HL	
					LH	
					HH	
					/LL	
					LL	
					\LL	

Figure 4. Distribution of tones in the maximal intonation unit (Mertens 2006: 10)

Mertens’s (2006) annotation system makes a distinction between four pitch levels, namely low (L), high (H), extra low (L-) and extra high (H+). These are determined relative to the local pitch levels and the speaker’s pitch range. The L and H pitch levels are separated by a major melodic interval, which is typically of ≥ 5 semitones. The L- and H+ pitch levels correspond to the pitch floor and ceiling, i.e. the lowest and highest level of the pitch range respectively. Minor intervals, which are typically of ≤ 5 semitones, are indicated by / and \ depending on whether the pitch level is raised, e.g. /L, /H, or lowered \L and \H. Unstressed syllables (unstr) are coded with a lower-case h when they are realised with a high pitch level and with a lower-case l when they are realised with a low pitch level. Likewise, the appendix

³⁶ Mertens’s (2006) ‘tone’, which corresponds to relative pitch level such as high or low, is to be distinguished from Halliday’s (1967a) ‘tone’, which is the pitch movement as a whole, e.g. fall, rise, fall-rise, rise-fall, level.

is marked as l-l- when the flat contour is low and hh when it is high. The initial accent can also be high H or low L. By contrast, the contour tones associated with the final accent show a lot more diversity. Terminal tones ending in L-, e.g. L-L-, HL-, and LL- constitute a major prosodic boundary indicative of the completion of a maximal prosodic unit and hence of an information object. Continuative tones such as /HH, \HH, HH, H/H, LH, HL also mark a major prosodic boundary but indicate that the utterance is not yet complete and that more information is to come. Continuative tones such as LL, /LL or \LL occur within the intonation unit and mark a minor prosodic boundary. As Mertens (2006: 10) explains, the horizontal lines in Table 4 indicate the existing constraints on the co-occurrence of certain tones. For example, the appendix tone l-l- can only appear after a terminal tone as instantiated in (26). The appendix is then uttered with a compressed register and no pitch variation.

- (26) [c'est un solvant]_{L-L-} [qu'il y a dedans]_{l-l-} (CRFP)
 'it's a solvent that's in it'

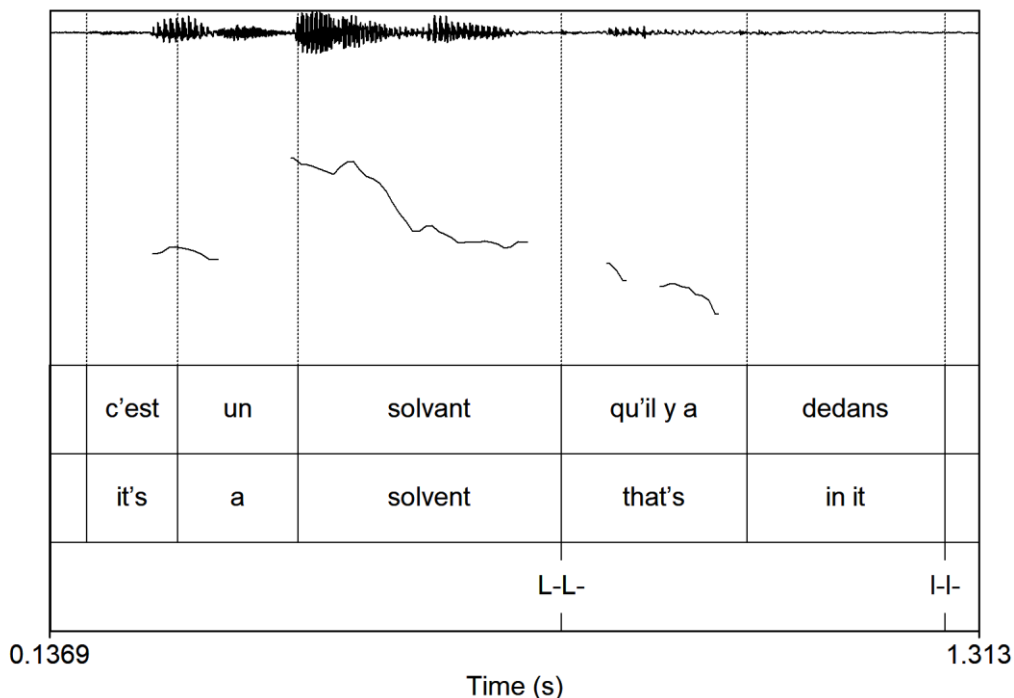


Figure 5. Praat realisation of (26)

Using the acoustic measures provided by Mertens (2006), I carried out the prosodic annotation of clefts strictly from what could be observed in F0 - and especially in the pitch range - and then formulated generalisations based on patterns observed in the data themselves. While I applied Mertens's (2006, 2012) analysis of the intonation unit in French, I do not adhere to the functional typology he associates with it, which failed to cover the diversity observed in my data. Among the 480 cleft tokens, 14 could not be annotated for prosodic features as the corresponding excerpt were either of poor quality or there was a speaker's overlap so prevalent that it made it impossible to provide even an auditory-based analysis.

Following the binary taxonomy established by Rialland et al. (2002), Doetjes et al. (2004) and Mertens (2012), summarised by Avanzi (2011) as (27) below, I labelled clefts with a terminal falling tone followed by a low appendix tone as Type 1 and clefts with a continuative tone followed by a terminal falling tone as Type 2. Any cleft with a prosodic articulation other than Type 1 or 2 was placed in a third unlabelled category which was then re-analysed to bring out any other recurrent patterns.

- (27) Type 1: [c' est X]L-L- [qui V]l-l-
Type 2: [c' est X]HH [qui V]L-L- (Avanzi 2011: 115)

Thus, unlike *it*-clefts, the prosodic patterns available for *c'est*-clefts were primarily differentiated on segmentation patterning and contour tones rather than with regard to focus placement. For this reason, no hierarchical investigation was carried out.

2.4 Analytical model of discourse-familiarity

In treating the prosodically coded focus assignment of clefts as being distinct from the distribution of new and given information within the construction, I depart from the view adopted in most of the existing typologies whereby the two aspects are conflated. As such, I do not see focus as tied up with what is not shared knowledge between speaker and hearer. Instead, I follow McGregor (1997) and Verstraete (2007) and treat the referential and relational layers of information structure

(Gundel 1988) to be separate, but interacting, aspects.³⁷ In this view, and in the Hallidayan tradition, what is discursively given can still be represented as focal for rhetorical reasons (Halliday 1967a, 1985, 1994). The referential layer of information structure relates to the discourse familiarity of linguistic expressions denoting discourse referents (Chafe 1976; Clark and Haviland 1977; Prince 1981) while relational information structure is concerned with the degree of salience and focal/non-focal status of information within a given information unit (see Section 1.1). Referential information structure relates to 'presenting' vs. 'presuming' reference (Martin 1992) which pertains to textual properties of NPs and reference by clauses to situations and state-of-affairs (Kaltenböck 2005). By contrast, relational information structure involves "contrasts between prominence and non-prominence in meaning as an aid in the processing of text" (Matthiessen 1992: 42). To account for the different referential information structural patterns of *it*-clefts, the discourse familiarity of both the discourse referents designated by the value, i.e. the entity referred to by the clefted NP, and the variable, i.e. the open proposition in the cleft-relative clause, have to be assessed. While this is recognised in existing typologies, the criteria of the analysis have been left largely implicit.

To study the discourse-familiarity of both entities and propositions in both *it*- and *c'est*-clefts, I develop an analytical model with more explicit recognition criteria. This model takes as a starting point Kaltenböck's (2005) classification of discourse-familiarity. His model, initially developed for the study of *it*-extraposition, is particularly suited for the analysis of constructions in corpus data as it favours a textual analysis of discourse-familiarity over a cognitive one. As such, Kaltenböck's (2005) model excludes any categories based on hearer-familiarity found in previous models such as Prince's (1981). With this text-oriented approach, the speaker's assumptions and background knowledge are not taken into account, the main advantage being that assumptions are difficult, if not impossible, to establish in corpus data. Because linguistic entities gradually lose their

³⁷ As underlined by an anonymous reviewer of Bourgoin (submitted), the systems of given/new and focal/non-focal as investigated in this study can be assimilated to emic categories (Pike 1954), i.e. as coded categories, rather than as etic or context-dependent features, with focus being investigated as a prosodically coded category.

recoverability as speech progresses (Givón 1975; Chafe 1987, 1994), only a portion of the preceding context is included in the analysis. For this study, I adjusted Givón's (1983: 13) recommendation of taking 20 clauses in the leftward context and considered a maximum of 10 previous turns in the case of dialogues and 20 clauses for monologues.³⁸ Given the prospective approach adopted for the analysis of discourse-familiarity, I also took into account the immediate rightward context up to 5 turns or clauses. I revised Kaltenböck's (2005) model from a quinary to a quaternary one. I retained the inferable, brand new and new-anchored categories and merged the situationally- and textually-evoked types under the single label evoked. I am thus left with two main categories of discourse-given vs. discourse-new information, which are themselves divided into two subtypes, i.e. evoked vs. inferable and brand new vs. new-anchored. The resulting model is visualised in Figure 6 below. I present recognition criteria for the discourse familiarity of entities and propositions in the following sections.

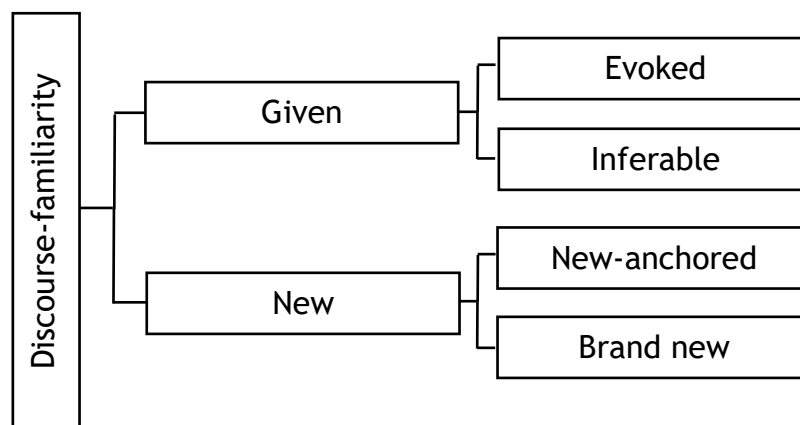


Figure 6. Model of discourse-familiarity based on Gentens (2016), itself based on Kaltenböck (2005)

³⁸ Given the lack of agreement concerning the quantification of lexical item recoverability, Givón's (1983) recommendation of 20 turns and/or clauses seems the most reasonable limit logistically to set for the amount of leftward context taken into account.

2.4.1 Nominal referent in the value

Kaltenböck's (2005) description of discourse-familiarity, which serves as basis for the analytical model proposed in this thesis, is directly inspired by Prince's (1981) taxonomy of givenness/newness. However, while Kaltenböck's taxonomy (2005) was developed to assess the discourse-familiarity of propositions as a whole, Prince's (1978) is primarily used for nominal referents. Despite Prince's (1981) model seemingly being more appropriate for the analysis of clefted constituents and value NPs in particular, the model proposed in this thesis departs from her classification on one major point which is the consideration of the definite/indefinite article system (Fontaine and Schönthal 2020). Prince's (1981) model does not take into account the determiner system as main recoverability marker, despite it being dedicated to marking in English the distinction between referents presumed retrievable and referents presumed irretrievable (Du Bois 1980; Martin 1992: 120). This claim is also true for French but with a few notable differences in the use of the different determiners. One of these differences relates to the use of the zero marker in English to refer to abstract concepts (\emptyset philosophy), generic descriptions (\emptyset men), materials (\emptyset copper), titles (\emptyset President Obama), for which French makes use of the definite article (*la* philosophie, *les* hommes, *le* cuivre, *le* Président Macron). Unlike in English, the zero article in French is used when preceded by the preposition *de* itself preceded by certain verbs (avoir besoin de \emptyset farine), adjectives (être dépourvu de \emptyset bon sens) or adverbs or nouns indicating quantity (beaucoup de \emptyset choses, un litre de \emptyset vin). When the following noun is specified (il a besoin de *la* bougie qui est sur la cheminée), the definite article replaces the zero marker. Despite those minor differences, the approach adopted in the present study considers the definite-indefinite system to be the primary marker of recoverability in both languages. The four discourse-familiarity categories are described and exemplified below.

2.4.1.1 Evoked

In the proposed model, evoked NPs are linguistic entities that are directly retrievable from the discursive or situational context. This category includes

referents involving anaphoric retrieval from discourse (Martin 1992) as well as exophoric and homophoric retrievals from outside discourse. Exophoric retrieval (from the context of situation) is typically realised by demonstratives as in (28), while homophoric retrieval (from the context of culture) is primarily realised by NPs with definite article such as (29) or proper nouns as in (30).

- (28) A: that perhaps something of Gertrude's character is thrown up in that scene
B: how many times is Hamlet alone with his mother as he is in the closet can you remember
A: I think it's just that one scene (LLC-1)
- (29) donc tu as tu as juste à brancher ta machine et puis après tu introduis euh ou les draps ou les nappages (...) et c'est la machine qui fait tout le travail (CRFP)
'so you just you just have to plug in your machine and then you put in uh or linen or table linen (...) and it's **the machine that does all the work**'
- (30) that is Hemsley Hemsley playing the ball into the centre on the arc up go the heads and it's Steve James that plays it forward back into **Sheffield United territory** (LLC-1)

2.4.1.2 Inferable

Inferable NPs are entities which are only indirectly retrievable and which involve inferential bridges based on different types of conceptual relations. These comprise hyponymy (general-specific), meronymy (part-whole), antonymy, contiguity, or entailment. (31) exemplifies a metonymic inferential relation where *Rufford people* are a subset of *Rufford colleges* and (32) a hyponymic relation between *un travail manuel* (a manual job) and *le manuel* (manual labour). Due to the prospective approach adopted for the study of discourse-familiarity (see Section 2.4.3), the inferential bridges considered are only those which are predictable and not merely retrievable.

- (31) A: a university committee was formed secretly to discuss the future of the English department
 B: a university committee or a Rufford colleges committee cos this was before separation
 A: a university committee I think Patrick Muir was on it too
 B: really really
 A: but it was mainly Rufford people Simon Crawley was on it so I heard (LLC-1)
- (32) donc lui a aussi un travail manuel lui c'est le manuel qui l'intéresse aussi oui (CRFP)
 'so he also has a manual job for him it's manual labour that he's interested in too yes'

2.4.1.3 New-anchored

NPs are considered new-anchored when they present new instances but still retain a link with the previous discourse which prevents them from being interpretable “outside of context” (Gentens 2016: 21). New-anchored NPs contrast with inferable ones in that at least some part represents wholly new information while inferable NPs are fully given, albeit not explicitly. As such, new-anchored NPs typically consist of more than just a noun head. In detail, they are often NPs introducing new notions, like *application* and *criticism* in (33), but which are anchored to given referents, like *the very techniques themselves*, allowing the overall referential status to be marked as definite. In (34), the value combines given information with *cette journée* which refers to the date announced beforehand, and new information in the coordinated relative clause.

- (33) well yes there are techniques to learn but it's very much the application or the criticism of the very techniques themselves which is the important thing (LLC-1)
- (34) je me suis mariée le le quatre septembre de cette année dix-neuf cent quatre-vingt-dix-neuf et euh c'est euh cette journée et ce qui l'a

précédée le voyage de noce qui qui a suivi dont je dont je vais parler aujourd'hui (CRFP)

'I got married on on September fourth that year of ninety ninety-nine and uh it's **that day and what preceded it the honeymoon that followed** that I'm going to talk about today'

2.4.1.4 Brand-new

The brand-new category of NPs includes all NPs introducing new instances of a new type designated by the head noun as in (35) and (36).

(35) A: well you give them the lot you see that's the point and make sure that there's something fairly closely related to what they've studied

B: it's just **one question** that they have to do isn't it (LLC-1)

(36) A: dans une grande surface il y a énormément de gens qui défilent pour venir chercher des fleurs après leurs courses

B: hum hum

A: et autrement alors que là c'est **des gens de passage qui passent dans la rue** qui ont besoin de fleurs (CRFP)

'in a supermarket there are a lot of people that pass by to pick up flowers after doing their food shopping

B: hm hm

A: and otherwise there by contrast it's **people passing through that pass by in the street** that need flowers'

2.4.2 Open proposition in the variable

The same basic distinctions can also be applied to the referents of the cleft-relative clauses, i.e. open propositions. Given that the variable takes the form of a proposition rather than a nominal complement, the role of the determiner system is logically lessened.

2.4.2.1 Evoked

Evoked open propositions have already been explicitly mentioned such as *x said no* in (37) or *x oriente* in (38) and are thus directly retrievable from the preceding discourse context.

- (37) A: there was one firm which said no too
B: it was **Gulbenkian** who said no (LLC-1)
- (38) A: ah donc allez voir cette personne c'est toi qui orientes
B: voilà et c'est moi qui voilà qui oriente (CRFP)
'A: oh so you go up to that person it's you who guide
B: exactly and it's me exactly who guide'

2.4.2.2 Inferable

Open propositions are considered inferable when they are linked to other linguistic entities from which they are derivable. Similarly to nominal referents, the link between two entities can take the form of inferential bridges, such as hyponymy, meronymy, antonymy, contiguity, or entailment. In (39), a link can for instance be inferred between the general abstract idea of *dealing with recognition* at the academic level and the specific concrete action of *putting people forward*. Open propositions are also considered to be inferable when they are linked to another proposition itself creating an inference. In (40), the fact that someone took money from the association is heavily implied in the preceding sentence and this inference is then taken up in the cleft-relative clause.

- (39) A: no boards of studies don't don't deal with recognition this is a bloody complicated university it's **it's the the faculty in the school**
B: oh no they don't of course do they no no they don't know
A: that that puts you forward you see (LLC-1)
- (40) malheureusement euh but lucratif certaines ont pensé que euh ils on pouvait mettre dans ses poches donc évidemment **c'est la directrice**

qui était un peu comme ma deuxième mère d'ailleurs qui s'est un peu servie (CRFP)

'unfortunately uh for-profit some of them thought that uh they one could take money so obviously it's the manager who was a bit like my second mum by the way who took some money'

2.4.2.3 New-anchored

The new-anchored category encompasses all open propositions that present a new state of affairs which cannot be interpreted without taking into consideration at least one link with the preceding context (Gentens 2016). It also includes propositions consisting of given information, generally inferable, which contains one wholly new component. Like nominal referents, new-anchored propositions differ from inferable ones on the basis that they must contain at least one new item. In (41), the open proposition in the variable introduces new information on the causes of a bad world which is only interpretable by taking into account the change of polarity in the previously shared information about what makes a good world. While the open proposition builds on the implication that some things in the world must be somewhat bad for it to have room for improvement, the very notions of unhappiness and violence are not predictable enough to make them inferable. In (42), the fact that someone has developed the specialised circuit in a proper manner is new information which is shared about *cet appareil* which refers anaphorically to the recorder mentioned by the speaker in their previous turn.

(41) A: if the the individual's understanding is improved then somehow if enough individuals are better understanding then you're going to produce a better world now let me rephrase that and try and get it a bit clearer

B: it sounds all right

A: it is prejudice that causes unhappinesses that causes violence and all sorts of things (LLC-1)

(42) A: si on prend genre l'exemple de l'appareil que vous avez à côté là l'enregistreur (...) euh ben là c'est miniaturisé au maximum donc les

circuits proprement dit c'est des circuits spécialisés qui font des fonctions je sais pas euh lecture enregistrement euh éjection peut-être du truc

B: hum hum

A: euh ce qui s- disons que **c'est le fabriquant qui a proprement développé ce circuit-là pour cet appareil** quoi (CRFP)

'A: if we take let's say the example of the device you put next to there the recording machine (...) uh well there it's miniaturised as much as possible so the circuits themselves it's specialised circuits that have functions I don't know uh play record uh eject maybe from the thing

B: hm hm

A: uh what s- let's say that **it's the manufacturer who properly developed that circuit for this device**'

2.4.2.4 Brand-new

Finally, brand new information corresponds to propositions that do not include any anaphoric links tying the information to previous referents as in (43) or (44). Personal pronominal reference to speech participants is not considered to create anaphoric links and brand-new information may therefore contain one or more of them.

- (43) there's something that makes us feel savage about these rock and roll singers and I hate it in myself and I see it in a lot of other people now **it's only about a year ago that on this programme we were asked about I think it was Tommy Steele being mobbed** (LLC-1)
- (44) c'est comme ça **que le champagne par exemple est fait avec du Pinot Noir** (CRFP)
'it's like that that champagne for instance is made with some Pinot Noir'

2.4.3 Prospective vs. retrospective analysis

The recognition criteria for the different categories of discourse-familiarity may be applied either retrospectively, i.e. with a backward-looking approach, or prospectively, i.e. in a forward-looking manner. While a retrospective analysis of discourse-familiarity entails establishing whether a given referent can be linked to an antecedent in the prior context, a prospective investigation involves establishing whether a given antecedent evokes, or predicts a referent from the following discourse. This distinction is particularly important for the dividing line between the categories of new-anchored and inferable whose scope is typically broader in retrospective analysis due to the possibility of reconstructing more inferential bridges. For instance, in (45), a retrospective analysis would probably lead one to treat the open proposition *x has revolted against the conception of the eleven plus* as inferable from the mention that grammar schools were abolished. However, looked at prospectively, the abolition of the grammar school does not as such imply that it was motivated by a revolt of any kind. I therefore analyse the proposition as new-anchored.

- (45) finally there is something which I ought to allude to [...] and that is the effect of changes in the curriculum the ways of teaching in the schools this is not anything to do necessarily with comprehensive schools or the abolition of the grammar school it is notable that in this country **it is the middle classes themselves who have revolted against the conception of the eleven plus** (LLC-1)

For this particular study on the discourse-familiarity of cleft referents, I rely on a prospective analysis only. This choice is primarily motivated by the very existence of reduced clefts in both English and French, which do not have an overt variable precisely due to their high degree of predictability. It also stems from the nature of the data this study deals with. Spontaneous speech tends, in contrast with writing, to have a forward-looking directionality (Sinclair 1992; Emmott 1992, 1997; p.c. Martin) with spoken utterances setting the scene for what follows (Sinclair 1992: 11).

2.5 Concluding remarks

In this chapter, I have laid out the methodology and analytical framework I adopt to study the use of *it*- and *c'est*-clefts in spoken data. The contrastive analysis is carried out using comparable corpora of recorded monologues and dialogues between one or more participants. Full and reduced clefts are extracted and sorted out manually from both corpora. The resulting dataset is annotated for a number of parameters pertaining to the syntactic, semantic, pragmatic, prosodic and information structural profile of clefts.

Clefts are analysed as constructions whose grammatical structure codes identifying-specificational meaning whereby the clefted constituent in the matrix clause is construed as the value satisfying the variable conveyed as an open proposition in the cleft-relative clause. Onto this value-variable relation are mapped different information structural patterns which encompass the distribution of focal and non-focal material and discourse-new and discourse-given information.

To study how focus is assigned in *it*- and *c'est*-clefts, I adopt Halliday's (1967a, 1985, 1994) functional approach in which focus is treated as a prosodically coded function marked by the placement of the nuclear accent, i.e. main pitch change, within the tone unit, i.e. the basic prosodic unit of speech segmentation. The focal/non-focal status of the cleft constituents is assessed through the combination of instrumental and auditory analysis which aims at determining the number and location of information foci in English clefts and the number of intonation units and nature of pitch contours in French clefts. In the case where *it*-clefts contain more than one information focus, the prominence level of each nuclear accent is examined relative to one another using Esser's (1988) hierarchy of foci. That this analysis is carried out on extensive datasets and with an in-depth systematic approach, which had not been done before, constitutes one of the novel aspects of the present study.

To establish how new and given information is packaged in clefts, I have proposed a comprehensive quaternary analytical model building on Kaltenböck's (2005) quinary model of discourse-familiarity. The new and given categories are further divided into two sub-categories, brand new and new-anchored, and evoked and inferable respectively. These categories are applied to the data using a

prospective approach in accordance with the forward directionality of spontaneous speech (Sinclair 1992; Emmott 1992, 1997).

Overall, the methodology presented in this chapter not only allows me to test a number of claims and hypotheses formulated in the literature but it also allows me to substantiate them with corpus-based empirical results.

Chapter 3: The English *it*-cleft

In this chapter³⁹, I set out my findings about the functional-structural patterns of *it*-clefts in spoken English, based on qualitative and quantitative corpus analysis carried out using the methodology described in Chapter 2. As discussed in the previous chapters, I distinguish the semantics of specification coded by grammatical structure from the information structural patterns speakers map onto the specificational relation. The investigation of those information structural patterns involves examining the relation of focus to non-focal information coded by prosody, the given/new discourse status of the elements of the cleft, and the interaction between the two. In Section 3.1, I summarise the findings on the morphosyntactic properties of *it*-clefts relating to the form and function of the clefted constituent, the paradigm of relative markers introducing the cleft-relative clause, and the expression of modality. Section 3.2 focuses on the relational layer of the information structure of *it*-clefts coded by prosody. In Section 3.3, I delve into the layer of discourse-familiarity which relates to the referential information structure of *it*-clefts. The relation between the two layers of information structure is addressed in Section 3.4. Finally, I wrap up this chapter in Section 3.5 by providing interim conclusions on the English *it*-cleft.

3.1 Morphosyntactic properties

I start by providing an overview of the morphosyntactic properties of *it*-clefts addressing the form (Section 3.1.1) and function (Section 3.1.2) of the clefted constituent. I then discuss the paradigm of relative markers used in *it*-clefts (Section 3.1.3), after which I examine the expression of modality and negation (Section 3.1.4).

³⁹ Substantial parts of this chapter are updated versions of sections found in Bourgoin et al. (2021), Bourgoin and Davidse (in press) and Bourgoin (submitted).

3.1.1 Syntactic category of the clefted constituent

As argued in Chapter 1, *it*-clefts express a specificational-identifying relation (Bolinger 1972; Collins 1991: 37; Davidse 2000; Huddleston and Pullum 2002: 1416-7; Van Praet and Davidse 2015). Studies on *it*-clefts (e.g. Delahunty 1982; Quirk et al. 1985; Collins 1991; Weinert and Miller 1996; Huddleston and Pullum 2002) have shown that the clefted constituent, i.e. the identifier, may be of various syntactic categories including NPs, PPs, AdjPs, AdvPs, finite clauses and in some cases *wh*-words. This wide range of syntactic categories is reflected in the results of the analysis of the LLC-1 dataset summarised in Table 1 below.

	Full clefts	Reduced clefts	Total
Noun phrase	109 (76.2%)	68 (71.6%)	177 (74.4%)
Prepositional phrase	12 (8.4%)	14 (14.7%)	26 (10.9%)
Adverbial phrase	5 (3.5%)	1 (1.1%)	6 (2.5%)
Finite clause	8 (5.6%)	6 (6.3%)	14 (5.9%)
<i>WH</i> -word	9 (6.3%)	4 (4.2%)	13 (5.5%)
Adjectival phrase	—	2 (2.1%)	2 (0.8%)
Non-finite clause	—	—	—
Total	143 (100%)	95 (100%)	238 (100%)

Table 1. Distribution of syntactic categories of the clefted constituent in the LLC-1

Among the different categories, NPs are the most frequent ones with 74.4% of all clefted constituents introducing nominal values. The prevalence of NPs as clefted constituents is typical of *it*-clefts whereas pseudo-clefts generally select phrasal and clausal complements (Weinert and Miller 1996). In more detail, value NPs can take the form of NPs with a common noun head, e.g. (1), proper nouns, e.g. (2), or pronouns, e.g. (3).

- (1) it's **people like that** that are looking after the the four and five year olds (LLC-1)
- (2) it was **Serena** who discovered that she'd got married (LLC-1)
- (3) it was I who said I wanted to sell out (LLC-1)

Within the category of nominals, common noun heads (76.3%) represent the majority of occurrences while proper nouns (18.1%) come second and pronouns (5.6%) last. The second most frequent syntactic class is that of prepositional phrases, as in (4), which amount to 10.9% of all clefted constituents. The distribution of classes in the remaining clefted constituents include, in order of frequency, finite clauses, (5), wh-word, (6), adverbial phrases, (7), and adjectival phrases, (8).

- (4) it was **through her inspiration** that possibly the Women's Institute it and thing the Women's Institute and things like that really developed (LLC-1)
- (5) it is only **as Christ condescends to come in** that we can be made by his grace what he would have us be (LLC-1)
- (6) when I was ten ten years ago I knew **what** it was that got people pregnant (LLC-1)
- (7) even afterwards I don't mind but it's **beforehand** [~~that I mind~~] (LLC-1)
- (8) he's got an Iraqi mother and an Indian or or I think it's **Pakistani** [~~that his father is~~] is it (LLC-1)

These results are for the most part in line with those of Collins (1991), whose study uses a combination of data from the LLC-1 and the Lancaster-Bergen/Oslo Corpus (LOB). The distribution of syntactic classes is similar for all but two categories. Unlike Collins (1991), no cleft with a non-finite clause as their clefted constituent was found in the LLC-1. It can be assumed that the six occurrences of clefts with a non-finite clause retrieved by Collins (1991), e.g. (9), were extracted from the LOB clefts, which would explain their absence in my dataset.

- (9) It is not **so very long ago** that Brahms met with bored incomprehension in Latin countries, that Bruckner and Mahler were regarded as exclusively Teutonic, Fauré exclusively French, and Nielsen exclusively Scandinavian, while Anglo-Saxons and Scandinavians marvelled at their own particular appreciation of Sibelius, Delius or Vaughan Williams. (LOB, Collins 1991: 55-56)

Another difference relates to ‘inferential’ constructions (Delahunty 1995), e.g. (10), which have sometimes been analysed as clefts (Collins 1991; Declerck 1992). In inferential constructions such as (10) below, the matrix clause only consists of *it* + copula *be* and is directly followed by the cleft-relative clause. According to Collins (1991: 57), these clefts serve to “highlight non-ideational items relating to tense, modality, aspect and polarity”.

- (10) If so, it must be \emptyset that their God was more powerful than the Kikuyu’s Ngai, (LOB, Collins 1991: 55)

Constructions of the kind can also be analysed as cases of extraposition in the sense of Davidse and Van Linden (2020), whereby *it*-extraposition is analysed compositionally rather than transformationally. On their analysis, (10) exemplifies complementation with an impersonal matrix and a dependent complement clause introduced by complementiser *that*. Because of the ambiguous nature of inferential constructions like (10), all instances were discarded from the dataset.

With regard to mood, it should be noted that instances of clefts with a *wh*-word as their clefted constituent, e.g. (6), constitute a subclass of interrogative clefts. The interrogative category subsumes dependent *wh*-interrogatives like (6) but also alternative questions such as (11). In both cases, the illocutionary force given to the cleft allows the speaker to inquire into the value that fills the semantic gap in the open proposition, with the identifier remaining underspecified. Thus, in (11), the speaker proposes two values and requests the hearer to indicate which is the correct one.

- (11) is it she who won’t do the allowing or he (LLC-1)

What the distribution of syntactic categories shows is that *it*-clefts in spoken interaction display a certain freedom in the selection of their clefted constituents, but that they retain a strong preference for NPs. The same order of frequency is found in both full and reduced clefts. A more detailed comparison of the two forms is provided in Section 6.2.

3.1.2 Grammatical function of the clefted constituent

As far as the function taken on by the clefted constituent in the cleft-relative clause is concerned, six are attested in the dataset: subject, direct object, indirect object, adjunct, and complement of a preposition. The distribution of the different functions is shown in Table 2 below.

	Full clefts	Reduced clefts	Total
Subject	87 (60.8%)	33 (34.7%)	120 (50.4%)
Direct object	13 (9.1%)	15 (15.8%)	28 (11.8%)
Indirect Object	—	1 (1.1%)	1 (0.4%)
Adjunct	32 (22.4%)	27 (28.4%)	59 (24.8%)
Complement of a preposition	11 (7.7%)	18 (18.9%)	29 (12.2%)
Subject complement	—	1 (1.1%)	1 (0.4%)
Total	143 (100%)	95 (100%)	238 (100%)

Table 2. Distribution of syntactic function of the antecedent clefted constituent in the LLC-1

Clefted constituents are pre-dominantly subjects and adjuncts while direct objects, complements of prepositions are less frequent. Indirect objects and subject complements only occur once each. The different functions are illustrated in examples (12)-(17) in the order of frequency of occurrence. It should be noted that the syntactic function assigned to the clefted constituent of reduced clefts, e.g. (16) and (17) below, are attributed according to the most salient and logical reconstructed cleft-relative clause.

- (12) it's **these individual copies** which are really the core of the system (LLC-1)
- (13) it's not **until you're there** that you realize that at university you can go your own way (LLC-1)
- (14) it was **this one** I was offering Barry (LLC-1)
- (15) it's **the academic structure of the university** that that we're concerned about (LLC-1)

- (16) he gives some of them in this room his undergraduate ones he gives in this room because it's **his own group** [~~that he teaches to~~] during the day (LLC-1)
- (17) he's got an Iraqi mother and an Indian or or I think it's **Pakistani** [~~that his father is~~] is it (LLC-1)

Halliday (1967a, 1985) motivates the fact that subjects are the most frequently clefted function as follows. Through their syntactic composition, clefts can make the subject of the corresponding canonical, or 'declefted', clause, like *these individual copies* in (18a), into a theme with unmarked non-initial focus.⁴⁰ By contrast, in the corresponding simple clause (18b), a focal subject/theme in initial position is very marked in English.

- (18) (a) it's **these individual copies** which are really the core of the system (LLC-1)
- (b) these individual copies are really the core of the system

Not observed in the LLC-1 data but still available for clefting in some cases are conjuncts such as *thus* in (19).

- (19) I thought it was needless to repeat what could be found there; and **thus it is that** so few manuscripts have descended to us which are marked in this way. ([the weird 'thus it is that' | The Grammar Exchange \(infopop.cc\)](#))

Here, the conjunct does not have a function in the cleft relative clause but acts as a linker between the proposition expressed by the cleft relative clause and the preceding discourse. Interestingly, in these cases there is no open proposition with a semantic gap in the cleft relative clause.

In providing an overview of the different syntactic categories and functions of the clefted constituent and of the different relative markers used in clefts, the last

⁴⁰ Marked theme can be all the other functions listed in Table 2.

two sections show that *it*-clefts typically select subject NPs as their syntactically-emphasised constituent, but that a wide array of less frequent forms and functions may also be found. Whether this syntactic highlighting is reinforced by way of prosodic means, i.e. whether focus is encoded syntactically and prosodically on the same constituent, will be discussed in Section 3.2.

3.1.3 Paradigm of relative markers

It is generally agreed upon that cleft-relative clauses in *it*-clefts, like restrictive relative clauses, take relative markers from either the *wh*-paradigm or the *th*-paradigm, i.e. *that* or \emptyset . According to Quirk et al. (1985) and Huddleston and Pullum (2002), however, cleft-relative clauses more frequently use the *th*-paradigm than restrictive relative clauses. The distribution of the different types of relative markers observed in the LLC-1 is summarised in Table 3 below.

Relative marker	n (%)
that	85 (59.4%)
which	10 (7%)
who	29 (20.3%)
where	1 (0.7%)
\emptyset	17 (11.9%)
on which	1 (0.7%)
Total	143 (100%)

Table 3. Distribution of types of relative markers

While both the *wh*-paradigm and *th*-paradigm are represented, the latter is more frequently used than the former. Importantly, 11.9% of occurrences display the zero relative marker \emptyset . Moreover, whereas restrictive relative clauses mostly allow the zero relative marker with object complements, cleft-relative clauses extend it to the subject function, as in (20)-(22). In all three clefts, the subject NPs are linked to the cleft-relative clause, also referred to as “contact clauses” by Jespersen ([1937] 1984), with the zero relative marker.

- (20) I think it was Mrs Corley \emptyset told your Lordship (LLC-1)
- (21) well it's Bill Gravy \emptyset wants to speak to him from William Martin (LLC-1)
- (22) it's much more recently than that may have been Ivor Bond \emptyset told me (LLC-1)

In the LLC-1, \emptyset is used after clefted constituents which are either subjects, direct objects, adjuncts or complements of prepositions, thus showing that its use is not restricted to one grammatical function. That the zero relative marker is productive in LLC-1 is in line with Greenbaum's (1996: 176) observation that it is generally restricted to spoken informal English. Hence, zero subject relatives can be, to a certain extent, viewed as a recognition criterion of *it*-clefts (e.g. Huddleston 1971: 324, Quirk et al 1985: 1387, Collins 1991: 52). As far as other relative markers are concerned, Huddleston and Pullum (2002) point out, contra Quirk et al. (1985), that clefts with adverbials as clefted constituent can take either *that* or a *wh*-adverbial as relative marker. No instance of *wh*-adverbial was found in the LLC-1 dataset.

3.1.4 Modality, negation and exhaustivity particles in *it*-clefts

Regardless of the syntactic class and function of their clefted constituents, *it*-clefts are, for the vast majority, specificational constructions. As such, they establish a relation between one (or more) value(s) referred to in the clefted constituent and the associated variable expressed in the cleft-relative clause. This specificational meaning always involves identifying the qualifying values for the variable in a context where some potential, but ultimately non-qualifying values, are more or less explicitly available. It is to this set of qualifying values that the implicature of exhaustivity triggered by the cleft, i.e. *it is X and only X*, applies. The comparison between the qualifying and non-qualifying gives rise to a notion of 'contrast' which is established on the representational meaning, or what Halliday (1967a) refers to as 'cognitive content'. The specificational relation, and hence the notion of contrast tied to it, may in some cases be impacted by various elements such as modality, negation or exhaustivity particles. This section uncovers the extent to

which these three aspects engage with the specificational meaning associated with the clefts of the LLC-1 dataset.

The role of modality is to indicate the stance of the speaker on the likelihood of the information conveyed being true through the use of modal verbs or adverbials (Quirk et al. 1985). Although cases of modal markers are rare in the LLC-1 data, they occur. Example (23a) below instantiates the use of the modal *may* to express epistemic possibility.

- (23) (a) **it's much more recently than that *may have been Ivor Bond* \emptyset told me** (LLC-1)
 (b) value: *Ivor Bond*
 (c) variable: *x told me*

Because the modal is found in the matrix clause, its scope spans the identification of the value for the variable designated by the cleft-relative clause. As such, it qualifies the relevance of the tentatively proposed value *Ivor Bond* for the given variable *x told me*. While the adequacy of the value is put into question, that of the variable is not, thus keeping the open proposition carried by the cleft true. The specification relation established between (23b) and (23c) is therefore preserved, but its accuracy is neither fully asserted nor rejected.

Similarly to modal verbs, the negation marker *not*, when used in *it*-clefts also provides the hearer with information on how to treat the specificational meaning. In (24a), the specification relation is set up between the value NP (24b) and the variable (24c) in the cleft-relative clause.

- (24) (a) **it wasn't Kilmarnock who who came to see you** actually that was Hawkins his partner [~~who came to see you~~] (LLC-1)
 (b) value: *Kilmarnock*
 (c) variable: *x came to see you*

The negation in the matrix clause does not cancel this relation but rather marks the value *Kilmarnock* as being incorrect. The appropriate value is then specified in the following clause as being *Hawkins*. The open proposition in the variable of the cleft

thus still holds true regardless of the inclusion of the negation marker. This corroborates the argument that presupposition in *it*-clefts functions - much like presupposition in any type of embedded clause - remains unscathed in face of negation operators (Delin 1992).

In the same vein as modality and negation, the specification relation at play in *it*-clefts also remains unchanged in face of exhaustive particles such as *only* or *just*, though the pragmatic mechanism of the conversational implicature triggered is modified (Velleman et al. 2012; De Cesare and Garassino 2015). For instance, the use of the particle *only* in (25) below asserts the exhaustivity of the value *medicine* which is specified for the variable *it's more or less general for x*. Without the particle, the exhaustive reading would only be implied and therefore cancellable, which is not the case in (25). It is worth noting that *only* receives the onset of the first tone unit, thus benefiting from additional prominence through prosodic means.

- (25) A: I think you've got to be either Scots or something else before apply
in the social sciences or something like that
B: oh I see yes yes
A: but it's ^only m\medicine# that it's more or less g\eneral for#

Whether stemming from modality, negation or the use of exhaustive particles, changes brought to the specification relation only apply when such grammatical items are added to the matrix clause. These may be emphasised through specific prosodic means, which will be discussed in the following section.

3.2 Prosodically coded information structure of *it*-clefts

I now turn to the results of the prosodic analysis of *it*-clefts, which relates to the relational layer of their information structure. The goal of this section is to examine a number of hypotheses and claims found in the literature, which I verify empirically and systematically. I first test the claim in Section 3.2.1 that the value constituent of clefts carries the only information focus coded by a nuclear accent (Clark and Haviland 1977; Givón 2001; Lambrecht 2001). I do so by assessing the different locations of prosodic focus. In the case of multiple-foci *it*-clefts, I investigate

whether the main information focus marked by the relatively most prominent accent is typically borne by the value as argued by Collins (2006) or whether it varies in its location, i.e. either on the value or variable, as claimed by Prince (1978), Declerck (1984) and Huber (2006). I do so in Section 3.2.2 by quantifying the different hierarchical orders established in multiple-foci *it*-clefts. Finally, I examine the claim that the clefts involving contrastive focus code argument focus (Lambrecht 2001) in Section 3.2.3.

3.2.1 Location of prosodic focus

In this thesis, I adhere to the view that information focus is coded by the placement of the tonic accent on a specific syllable of the tone unit (see Section 1.1.3.1), which marks the constituent it is part of as the information focus, i.e. the most salient new information in relation to the non-focal information within the tone unit (Halliday 1994: 296-9). In what follows, I present my findings on the location of prosodically coded foci within the LLC-1 dataset. Table 4 below summarises the different focal/non-focal patterns observed in *it*-clefts. A principled distinction is made between the patterns attested in full and reduced clefts, as they offer a different syntactic environment for prosodic choices. For reduced clefts, the absence of a cleft-relative clause is marked by 0. A more detailed comparison of full and reduced forms is provided in Section 6.3.2.

	Full clefts	Reduced clefts
Focal + non focal	12 (8.4%)	—
Focal + focal	91 (63.6%)	—
Non-focal + focal	34 (23.8%)	—
Non-focal + focal + focal copula	6 (4.2%)	—
Focal + 0	—	89 (93.7%)
Focal + 0 + focal copula	—	1 (1%)
Non-focal + 0 + focal copula	—	5 (5.3%)
Total	143 (100%)	95 (100%)

Table 4. Distribution of prosodic patterns of *it*-clefts

In total, four prosodic patterns can be identified for full clefts and three for reduced clefts. To classify a value or variable as focal, I made no distinction between realisation by single or by multiple tone units. The results show that full *it*-clefts typically exhibit a focal + focal pattern, as in (26), in which both the value and the variable are realised in separate tone units and each contain one or more nuclei. The second most frequent pattern, i.e. non-focal value and focal variable both realised in a single tone unit, e.g. (27), is found in 23.8% of occurrences. The same pattern with added focus on the copula is found in 4.2% only. This pattern is mainly exhibited in dependent interrogatives such as (28) or *it*-clefts with a fronted clefted constituent such as (29). The focal + non-focal pattern, illustrated in (30), which Lambrecht (2001) posits as the prototypical one, is only found in 8.4% of tokens, showing that it can hardly be described as the typical prosody of *it*-clefts. This finding is in line with Herment and Leonarduzzi's (2012) own conclusion on the scarce nature of the focal-non-focal pattern in spontaneous spoken data.

- (26) it's [^]how much they m\love it# that c\ounts# (LLC-1)
- (27) [^]for it is the terms that really m\atter# (LLC-1)
- (28) I'm [^]trying to remember where it w\as# that I h\ear'd# that you were [^]likely to get s\upport# (LLC-1)

- (29) ^he it w\as# who ^built Saint Paul's Ch\urch# ^in Stoke R/oad#
hims/elf# at his ^own exp\ense# (LLC-1)
- (30) it's the gr\ammar which is interesting# (LLC-1)

By contrast, reduced clefts show much less variety, mainly due to the absence of the cleft-relative clause, with only three patterns available. The majority of reduced *it*-clefts have one (or more) information focus on the value only, e.g. (31). In some cases, like (32), only the copula bears prosodic focus. Like full *it*-clefts, reduced *it*-clefts of this type are mainly indirect questions in declarative form. Finally, one occurrence, (33), displays a matrix clause uttered with two tone units and thus two information foci.

- (31) ^when does she ret\ire# it's ^not this y\ear# [~~that she retires~~] /is it#
(LLC-1)
- (32) I ^forget who it w/as [~~who asked me about what I thought~~]# (LLC-1)
- (33) and ^all that he asked me at the /interview# ^think it w\as# C P Sn/ow
[~~which he asked me about~~]# (LLC-1)

The results of the location of information foci in *it*-clefts warrant the conclusion that the prosodic profile of *it*-clefts is a diverse one which is not limited to a single prototypical realisation. Clefts may be realised in one or more tone units and information focus is not restricted to either part of the cleft. More specifically, the prosodically coded focus may be carried by the value, but it may also be borne by a constituent of the variable, e.g. in interrogative clefts and in one of the informational declarative subtypes, or by both value and variable. That the most frequent pattern in full clefts is focal + focal contrasts with the claim made by Clark and Haviland (1977), Givón (2001) and Lambrecht (2001) according to whom clefts carry a unique information focus on their value. By contrast, my findings not only corroborate the diversity of focus placement underlined by Prince (1978), Geluykens (1988) and Huber (2006), but they also provide a clear quantification of the different patterns. With regard to the broad vs. narrow focus distinction, which Halliday (1967a) conceptualises in terms of unmarked vs. marked, what the findings suggest is that the majority of clefts display broad focus whereby the information

in the whole cleft is marked informationally salient through the presence of multiple foci. Whether the focus on the value is prosodically, and hence informationally, more prominent or not will be examined in the next section.

3.2.2 Hierarchy of prosodic foci

Some authors, e.g. Prince (1978), Declerck (1984), Geluykens (1988) and Huber (2006), have argued that the prosodic hierarchical order of multiple-foci clefts differs according to their information structural subtype. According to them, the main information focus characterised by greater prominence may be located either in the value or the variable depending on where the new information is. Declerck (1984) recognises a third pattern in which both value and variable are equally prominent. In his corpus-based study, Collins (2006) shows that multiple-foci clefts generally have a stronger nuclear accent on the value but does not describe the remaining 8.8% of occurrences diverging from that pattern. The claims brought forward by the different authors, all of whom acknowledge at least some variation in the hierarchical order of information foci, are not contradictory by any means, but they do not rely on a clear and exhaustive inventory of the different patterns. Hence, the aim of this section is to verify and substantiate their claims by presenting a quantitative description of the hierarchies of foci observed in the LLC-1 dataset. The relative degrees of prominence of foci are analysed using Esser's (1988) analytical framework (see Section 2.3.1) and are assessed instrumentally using two parameters, namely pitch height of the nuclei and tone movement. This allows me to better describe what has been referred to as 'strong', 'normal' and 'weak' stress (Prince 1978; Declerck 1984) and 'primary' and 'secondary' stress (Huber 2006). Table 5 below visualises all the attested prosodic patterns.⁴¹

⁴¹ Unclear cases correspond to sound files that could not be analysed either due to technical issues, e.g. corrupted files, or to overlaps in the conversation.

Hierarchy observed	Tone movement in value	Tone movement in variable	Full clefts	Total
Value is higher	fall	fall	38 (41.7%)	56 (61.5%)
	rise	rise	1 (1.1%)	
	fall	rise	17 (18.7%)	
Variable is higher	fall	fall	4 (4.4%)	20 (22%)
	rise	fall	16 (17.6%)	
Value and variable are similar	fall	fall	11 (12.1%)	12 (13.2%)
	rise	rise	1 (1.1%)	
Unclear	—	—	3 (3.3%)	3 (3.3%)
Total			91 (100%)	

Table 5. Hierarchy of nuclei in full multiple-tone unit *it*-clefts

In the most frequent pattern, the value is hierarchically higher than the variable, either because the nucleus of the value has a higher pitch peak than the nucleus of the variable, i.e. with a step up of more than 0.05 logHertz as in (34), or because the value contains a fall while the variable contains a rise.

- (34) Hemsley chipping the ball into the centre onto the head of Scullion#
 from the ^Mhead of Scullion it's J^Mames# that g^Lets it# but only as
 far as Hockey (LLC-1)

As shown in the Praat image in Figure 1, the accent on *James* has a larger pitch excursion size than the accent on *gets*. Relative to other accents, the nucleus on *James* corresponds to a mid single-tone pitch accent while the nucleus on *gets* corresponds to a low single-tone pitch accent.

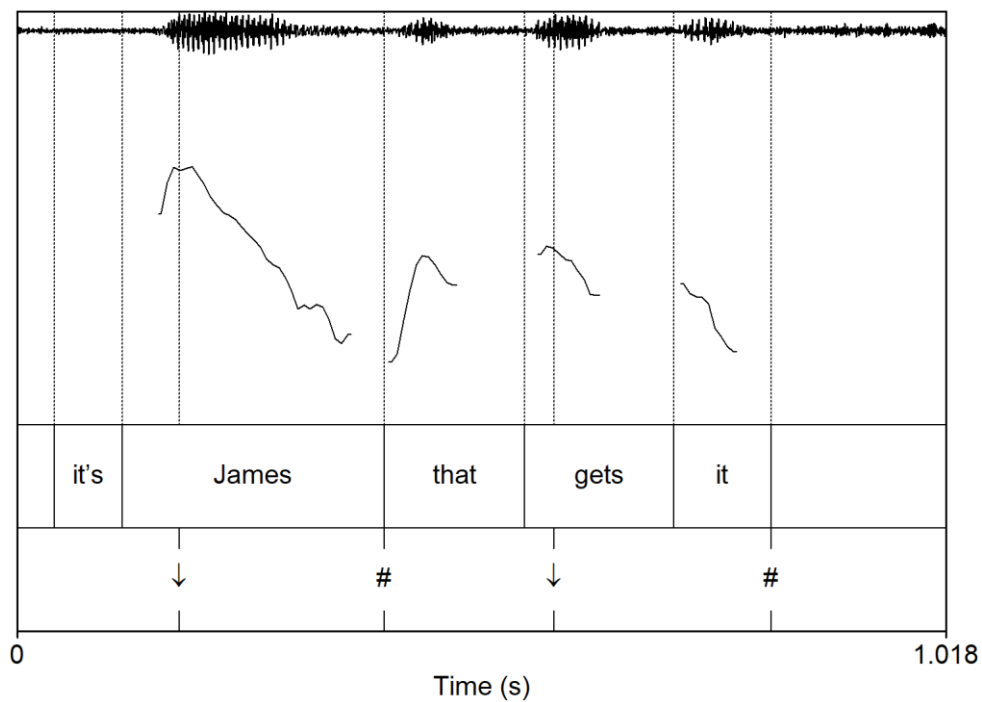


Figure 1. Praat image of (34)

The second most frequent pattern, in which the variable carries the highest-pitched nucleus, as seen in example (35) and its F0 in Figure 2 below, or in which the final tone of the value is a rise and that of the cleft-relative clause a fall, is found in 22% of cases.

(35) it's s\alads# that d\o it# (LLC-1)

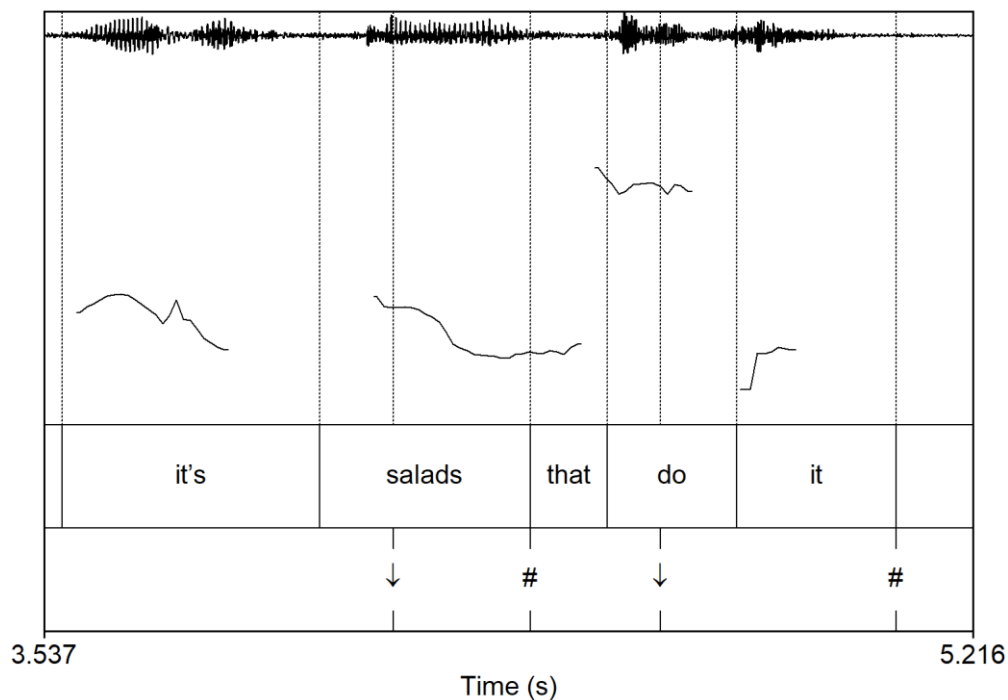


Figure 2. Praat image of (35)

The results also show that clefts do not always display a hierarchy of information foci. This is illustrated in (36), in which the accents on *beginning* and *difficult* are both mid-falling tone accents with a similar degree of prominence. What the Praat image in Figure 3 reveals is that there is no auditory step up or step down in pitch between the two elements bearing information focus.

- (36) A: [...] we've only got about thirteen hundred pounds in capital
 B: hm hm
 A: so although I could imagine that we could hm on our joint salary
 get perhaps quite a a a high mortgage# **it's the paying it back at the
 beg^M\inning# that's ^Mgoing to be d^M\ifficult# (LLC-1)**

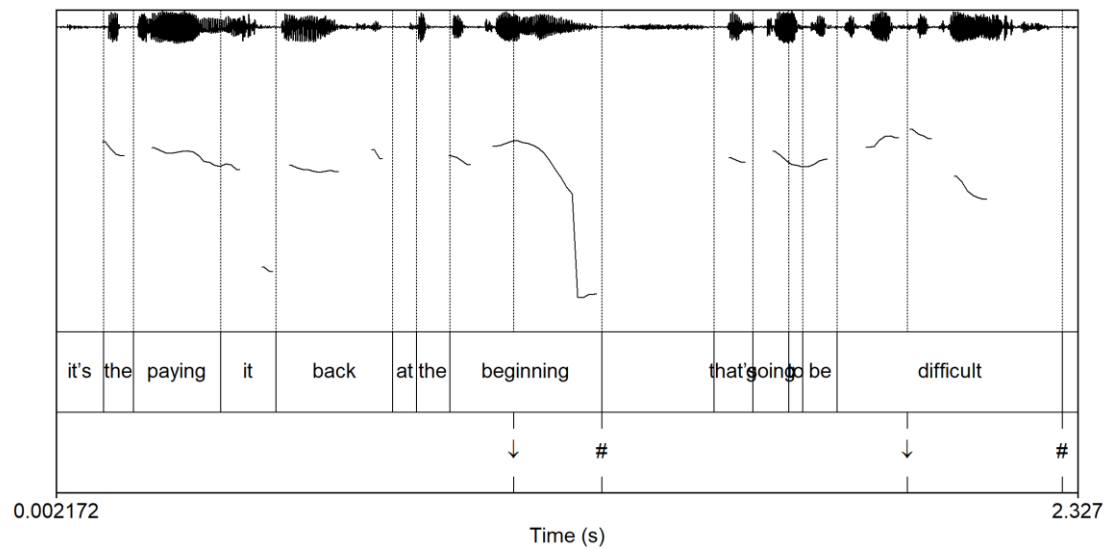


Figure 3. Praat image of (36)

What the instrumentally based operationalisation of the hierarchy claims shows is that full *it*-clefts most frequently exhibit a stronger degree of prominence on the value, which in Halliday's terms represents the unmarked prosodic realisation. This unmarked pattern is complemented by two marked choices in which there is either greater prominence on the variable or equal prominence on value and variable. Hence, the claims made by Prince (1978), Declerck (1984), Huber (2006), on the one hand, and that made by Collins (2006), on the other, are all verified. The conclusion I arrived at on the prosodically coded information structure of *it*-clefts in the previous section can be completed as follows. Clefts may be realised in one or more information units, but despite a preference for a stronger degree of prominence on the clefted constituent, information focus is not systematically restricted to the matrix clause. On the assumption that information focus is coded by prosody, descriptions of *it*-clefts that view the construction as a focusing device and link focus to prosodic accent therefore paint an incomplete picture of their use which overlooks cases in which the clefted constituent is not accented relative to the rest of the tone unit.

3.2.3 Contrastive vs. presenting focus

In this section, I interpret the results I have presented so far in terms of Halliday's (1967a) distinction between presenting and contrastive focus (see Section 1.1.3.1). A presenting focus, i.e. wide unmarked focus, does not mark any information as presupposed. It always falls on the last lexical constituent of the information unit, which is marked as the most salient new information while the status of the remainder is underspecified. When appearing at the beginning of discourse, the information unit may be wholly new. A contrastive focus, on the other hand, relates to presupposed information, which may precede and/or follow the focus, and which is presented as recoverable "anaphorically, by reference, substitution or ellipsis" (Halliday 1967a: 206). This focus is "informationally contrastive [...] either within a closed system or lexically" (Halliday 1967a: 207). Contrastive focus may, in some cases, relate to an elided but recoverable presupposition which is especially important for research on clefts as it allows speakers to anaphorically presuppose the variable in reduced clefts. On the nature of focus in *it*-clefts, Lambrecht (2001: 486-487) argues that clefts are argument-focus constructions in which the value necessarily receives the unique narrow and contrastive focus (Lambrecht 1994: 17). I put this claim to the test by examining how contrastive and presenting focus are encoded in the clefts of the LLC-1 dataset.

The most frequent information structure, if we add up full and reduced clefts, is the marked information structure with a single contrastive information focus relating to a presupposed message block as in (37). Here, the value NP *one question* bears contrasting focus (contrasting with two or more questions) which relates to the presupposed information *that they have to do*. The same analysis can be made for the reduced cleft in (38) in which the PP *through the British Council* receives narrow contrastive focus which relates to the omitted presupposition *he came through x*.

- (37) B: well you give them the lot you see, that's the point, and make sure that there's something fairly closely related to what they've studied
A: it's just **one question** that they have to do# /isn't it# (LLC-1)
- (38) A: he came through the British Council did he

B: ^well n\o# it's ^not through the British Council# [~~that he came~~]
(LLC-1)

Despite there being no overt presupposition, reduced clefts still express the same identifying-specificational semantics as full clefts and instruct the hearer to infer the variable from the preceding discourse. Thus, I consider the open proposition *he came through x* in (38) to be anaphorically recoverable by ellipsis (Halliday 1967a: 206) and consequently to be part of the information structure.

However, looking at full clefts only, the focal + non-focal realisation exemplified in (37) reveals itself as representing only a minority of occurrences of full clefts. Lambrecht (2001) claims that this prosodic pattern is the prototypical one for clefts, which is a corollary of his claim that cleft syntax codes the information structural relation between argument focus and presupposition. The low relative frequency of this type in my data does not support this prototypicality claim, and, as I will explain in Section 3.4.3, the possibility of selective focus goes against viewing contrastive focus as inherently argument focus.

The other type that is a clear instantiation of a specific information structure is the minority of *it*-clefts that manifest what is in general the unmarked information structure, i.e. non-focal + focal, in which information focus is located at the end of the construction. In *it*-clefts, this is typically operationalised with recoverable information preceding a presenting focus, i.e. non-contrastive, associated with the final lexical item. In the tradition that views cleft syntax as coding the focus-presupposition relation, this type has been said to have an ‘informative presupposition’, which “mark[s] a piece of information as fact, known to some people although not yet known to the intended hearer” (Prince 1978: 899). Prince (1978) explains the *informative-presupposition* cleft with example (39) below, which, on her account, can be used to mark the proposition in the relative clause as an uncontroversial fact that the hearer just happens not to know.

- (39) It was just about fifty years ago that Henry Ford gave us the week-end.
(Prince 1978: 898)

While Prince's (1978) *informative-presupposition* category allows one to make sense of examples like (39), the mapping between syntax, focus and discourse-givenness she proposes hardly applies to examples like (40), in which the material introduced in the cleft-relative clause represents irrecoverable information.

- (40) A: did you meet Fuller
B: yes# **it was ^he who invlited me#** (LLC-1)

Here, the matrix identifies the value, *he*, as filling the gap in the variable, triggering an exhaustiveness implicature that only *he* invited me. Onto this specification relation, an unmarked information structure is mapped, which starts with the recoverable information of the value and has presenting focus on the last lexical item *invited* of the variable. In this context it is very clear that there is no way in which there is even a hint of presupposition about the information in the cleft-relative clause, in which the very *unpredictability* of the fact that Fuller actually *invited* the speaker motivates the focus on the variable. Discursively, these *it*-clefts are chosen when the speaker wants to combine the exhaustive specification relation conveyed by the grammatical constituents with presenting focus on the open proposition. Whereas presenting focus is the unmarked option in spoken English at large, it is the marked information structure choice within a cleft environment.

How can we then analyse the full clefts with information foci on both value and variable, which form the largest portion of the whole dataset and the prototypical information structure of full clefts? In the literature, some authors, e.g. Declerck (1984), have suggested that two subtypes have to be distinguished which are prosodically and information structurally distinct. The first subtype is claimed to feature a stronger accent on the value and a weaker one on the variable to code a contrastive value relating to an 'old presupposition' in the cleft-relative clause. The second subtype is claimed to feature a normal (vs. weakly) accent on the variable, conveying a contrastive value and a new proposition in the cleft-relative clause. The results of the hierarchy of foci confirm the prosodic side of the hypothesis: clefts may have a more prominent focus on the value, a more prominent focus on the variable or a value and variable which are equally prominent. However, this prosodic difference does not convey the different information structures predicted

by these authors. In all of these subtypes, subject *it + be* is pre-tonic and there is a narrow, contrastive focus on the value NP, which conveys that the speaker chooses (contrasts or adds) an entity or person from a contextually given set of options. This contrastive focus on the value is combined with a typically unmarked presenting focus on the open proposition in the variable which may be used to inject ‘communicative dynamism’ (Firbas 1971, 1992).⁴² This is instantiated in (41) in which the matrix clause is realised as a single tone unit and whose cleft-relative clause is split into two tone units. The highest-pitched nucleus, and thus narrow contrastive focus, is carried by the value NP *Johnny* (contrasting with other members of the alternative set) while *money* and *France* both receive presenting focus.

- (41) it was J\ohnny# that stole her m\oney# ^while we were away in Fr\ance
(LLC-1)

Marked contrastive focus on an element of the variable may also occur but remains limited in use. This is the case in (35) which is repeated as (42) below.

- (42) A: why does one’s sort of regular habits alter in summer
B: no no but you just
A: and
B: wear less clothing so that’s it’s easier to take the nappies off
A: maybe
B: and stick them onto a potty (laughs)
A: maybe it’s s\alads# that d\o it# (LLC-1)

The fall on *salads* in (42) has a smaller peak amplitude than the fall on *do*. A and B are engaged in a humorous conversation about potty-training infants in summer.

⁴² Firbas (1971: 136) defines Communicative Dynamism as “a property of communication, displayed in the course of the development of the information to be conveyed and consisting in advancing this development”. In other words, the more a given constituent contributes to the progress of discourse, the more Communicative Dynamism it holds.

The cleft is A's answer to the speaker's own earlier probe *why* 'being regular' changes in summer, with contrastively focal *salads* providing the reason. The cleft-relative clause uses the substitute verb *do*, which relates anaphorically to 'alter one's sort of regular habits in summer' (Halliday 1994: 316-323) and hence provides material that would lend itself well to being wrapped as the presupposition part of an information unit. Yet, it is an information unit in its own right and one with marked contrastive focus on a closed system choice, i.e. the positive polarity form *do*.

The analysis of the hierarchy of foci thus allows me to refine the hypothesis formulated in Section 3.2.1. It can be concluded that, when full and reduced *it*-clefts are taken together, the unmarked choice is to have contrastive focus on the value either with or without presenting focus in the cleft-relative clause. That the variable carries, in some cases, presenting focus on one or more of its elements is an important pattern that has to be taken into account when discourse functions of clefts are studied. These findings reflect, to a certain extent, the different claims which the previous sections reacted to in that Prince (1978), Declerck (1984), Geluykens (1988) and Huber (2006) captured the nature of the unmarked option for focus marking in *it*-clefts, i.e. contrastive focus on the value. My analysis has, however, underlined the significance of the marked the option of presenting focus which is quantitatively more important than what had been predicted by these non-technical accounts. All in all, what the distribution shows is that the unmarked-marked distribution of information structures posited by Halliday (1967a, 1994) for spoken English at large is in fact reversed in clefts, with contrastive focus being more prevalent than presenting focus. Having now characterised the relational information structure of *it*-clefts, I now turn to their referential information structure, i.e. discourse-familiarity, which I investigate in the following section.

3.3 Discourse-familiarity in *it*-clefts

For some authors, e.g. Lambrecht (2001), the cleft syntax codes information-structural meaning, which entails that clefts convey one basic information structure, e.g. argument focus-presupposition. As I showed in Section 3.2, this is not tenable. For others, e.g. Declerck (1984) and Huddleston and Pullum (2002),

the syntax of clefts codes specificational meaning, onto which a number of different informational patterns may be mapped. Some authors recognising different information structural patterns have proposed typologies, e.g. Prince (1978), Declerck (1984), Geluykens (1988), Collins (2006), Huber (2006), in which each pattern is typified in terms of the informational givenness/newness of the value and the variable and is associated with a typical prosodic pattern. This section focuses on the referential information structure of *it*-clefts, i.e. the partition of information into new and given information. After addressing the distribution of discourse-new and discourse-given information in their constituents (Section 3.3.1), I propose a fine-grained typology of the discourse-familiarity of *it*-clefts (Section 3.3.2) which I relate to their functional profile (Section 3.3.3).

3.3.1. Distribution of discourse-given and discourse-new information

Table 6, below, summarises the findings of my analysis of the discourse-familiarity of value and variable. Applying Kaltenböck's (2005) approach and the analytic model described in Section 2.4, I analysed the discourse-familiarity of the referents in the value and the state-of-affairs in the variable, the latter of which is depicted by the cleft-relative clause in full clefts but derivable from the preceding discourse in reduced clefts. I classified evoked and inferable referents as given, and anchored and brand-new referents as new.

	Full clefts	Reduced clefts	Total
New-given	36 (25.1%)	54 (56.8%)	90 (37.8%)
New-new	35 (24.5%)	—	34 (14.3%)
Given-new	23 (16.1%)	—	24 (10.1%)
Given-given	39 (27.3%)	36 (37.9%)	73 (30.6%)
Unclear ⁴³	10 (7.0%)	5 (5.3%)	17 (7.1%)
Total	143 (100%)	95 (100%)	238 (100%)

Table 6. Distribution of discourse-familiarity patterns in *it*-clefts

The resulting taxonomy consists of four patterns of information distribution: (i) discourse-new value followed by a discourse-given variable, e.g. (43), (ii) discourse-new value followed by a discourse-new variable, e.g. (44), (iii) discourse-given value followed by a discourse-new variable, e.g. (45), and (iv) discourse-given value followed by a discourse-given variable, e.g. (46).

- (43) well Mallet is is is hopping mad about all this because Mallet sees a hundred and fifty thousand pounds for a building and various other things going down the drain what I what I think he doesn't realize# is that it's ^very largely bec/\ause# he's ^been b\uilding# ^this kind of peripheral \thing# in Appleby# that it h/\as gone down# (LLC-1)
- (44) only one comment I'm very glad to know that it's not only the m\issiles# and the \H bombs# that are ^going to be set upright by this g\overnment# (LLC-1)
- (45) I want here chairman to ask if you could tell us what notion is about college at the moment about what we might call interdisciplinary studies Scottish could you see be seen as a service industry in some respects# it's ^not that aspect I'm \after# (LLC-1)
- (46) A: I think I'm going to write I was thinking about this on the way down to the station this morning of writing every week to mom (...) and saying what I'd done in the week because she'd be quite interested and

⁴³ The 'unclear' category contains the cases for which we could not unequivocally determine the information status of either value or variable.

intrigued and eventually there'd be so much you'd find great difficulty in saying what it was (laughs) d'you know what I mean

B: how d'you mean

A: well sort of union account of what's happened

B: yes but there's a tendency then for her to go and repeat everything annoyance of (...)

A: it was for ^that r\reason# that I ^thought of d\oing this# (LLC-1)

The given-given pattern was not pointed out explicitly in existing taxonomies of *it*-clefts (e.g. Prince 1978; Declerck 1984; Geluykens 1988; Collins 2006) but was nonetheless acknowledged for full clefts in a handful of studies (see Johansson 2001, 2002; Huber 2006; Hasselgård 2014). While all four patterns occur in full clefts, only two of them are found in reduced clefts. The implied variable of reduced clefts always contains given information, either textually-evoked or inferable. This comes as no surprise as the cleft-relative clause can be omitted precisely because the variable is recoverable from the context (Declerck and Seki 1990). Overall, we observe a balance of all four types in full clefts, with the given-new type being slightly less frequent than the other three. This does not tally with Lambrecht's (2001) descriptive claim that the cleft-relative clause codes the 'presupposition', which "is more or less equivalent to the notion 'hearer-old'" (*ibid*: 474). Instead, the choice between given and new for both values and variables comes close to a balanced distribution. With reduced clefts, the new-given pattern is more frequent than the given-given one, the latter of which is recognised for the first time.

A point of special interest is the discourse familiarity of values carrying information focus, which is shown in Table 7 below.

	Full clefts	Reduced clefts	Total
New	68 (66%)	55 (61.1%)	123 (63.7%)
Given	33 (32%)	35 (38.9%)	68 (35.2%)
Unclear	2 (2%)	—	2 (1.1%)
Total	103 (100%)	90 (100%)	193 (100%)

Table 7. Discourse familiarity of focal values in *it*-clefts

While two-thirds of focal values in full clefts are discourse-new referents such as *the felt* in (47), one third designate referents already given in the preceding discourse as illustrated with *these individual copies* in (48). For reduced clefts, the proportion of discourse-given focal values is slightly higher at almost 39%.

- (47) A: roofing isn't the sort of thing you can get done by just a local builder in fact the specification that I most favour only licensed contractors can do anyway but they'd come with
 B: what this
 A: ten year guarantee and
 B: this tapered stuff no the insulation is is part of it# but **it's the felt that ^is the imp\ortant thing#** (LLC-1)
- (48) and you can see how this information we're putting in relates just to this this copy and that's how it's held in a in a separate entry and the a little separate record for each copy which will have circulation data (...) and in some ways this is the circulation system in fact# **it's ^these individual copies# which are ^really the core of the system#** (LLC-1)

That clefted constituents bearing a nuclear accent may be discourse-new but also discourse-given shows that there is no systematic mapping of prosodic prominence and discourse-newness, as has sometimes been suggested (e.g. Lambrecht 2001).

To further characterise the articulation of discourse-familiarity in *it*-clefts, let me now turn to the analysis of referents of all values also using the analytical model described in Section 2.4. The frequencies of the different types of discourse-familiarity are displayed in Table 8.

		Full clefts	Reduced clefts	All clefts
Given	Evoked	47 (32.9%)	29 (30.5%)	76 (31.9%)
	Inferable	12 (8.4%)	7 (7.4%)	19 (8%)
New	New-anchored	23 (16.1%)	15 (15.8%)	38 (16%)
	Brand new	61 (42.6%)	44 (46.3%)	105 (44.1%)
Total		143	95	238

Table 8. Distribution of discourse-familiarity types in values in *it*-clefts

The results of the discourse-familiarity of values are overall similar for the two forms of clefts. The values of full clefts typically introduce either evoked or brand new information, as in (47) and (48) above, while inferable and new-anchored information as shown in (49) and (50) below is less frequent.

- (49) oh yes oh yes I am yes because I'd like to do general medicine I find it very very interesting it contains a lot of the things I'd like it contains a lot of quite a lot of science you see quite a lot of people who are ill you know who are really ill rather than just and you can actually use diagnostic skills
 A: general science general medicine now I thought was the non-elitist
 B: oh no it's very elitist
 A: I I ^thought it was the sp\ecialists# who are e\litists#
 B: no but general medicine is a speciality I mean (LLC-1)
- (50) A: thought of driving that clapped out old van for seventy-five miles
 B: hundred and fifty miles
 A: yes# but it's the ^first seventy-five# ^which will be w\orst# (LLC-1)

By the same token, reduced clefts also most frequently introduce either evoked or brand new information in their value NPs while inferable and new-anchored values are not as frequent. That the two forms of clefts share a similar ratio for all types of values suggests that the choice between full and reduced clefts is not primarily attributable to the referential information structure of the values. This hypothesis will be examined in greater detail in Section 6.3.1. With regard to the general

distribution of discourse-new and discourse-given, discourse-given and discourse-new values of all values of full and reduced clefts are roughly equally distributed. The LLC-1 data thus does not uncover any significant preference for either discourse-new or discourse-given referents in the value of either form of clefts. The fact that the value may introduce either recoverable or unrecoverable information is in line with the description offered by the four typologies under review (Prince 1978; Declerck 1984; Geluykens 1988; Collins 2006).

As far as the analysis of open propositions in the variable is concerned, the categorisation proved to be more challenging than for NPs, despite the proposed criteria, which resulted in a small number of unclear cases. The distribution of the different types is summarised in Table 9.

		Full clefts	Reduced clefts	All clefts
Given	Evoked	36 (25.2%)	70 (73.7%)	106 (44.5%)
	Inferable	42 (29.3%)	25 (26.3%)	67 (28.1%)
New	New-anchored	22 (15.4%)	—	22 (9.2%)
	Brand new	41 (28.7%)	—	41 (17.2%)
	Unclear	2 (1.4%)	—	2 (0.8%)
	Total	143 (100%)	95 (100%)	238 (100%)

Table 9. Distribution of discourse-familiarity types in variables in *it*-clefts

The distribution of discourse-familiarity types in variables varies more than that of values. While the variable of full clefts exhibits the whole range of discourse-familiarity statuses, that of reduced clefts, as expected, conveys only discourse-given information. Indeed, the givenness of the information carried by the cleft-relative clause of reduced clefts has consistently been invoked to account for their very existence. This is exemplified in (51), where the variable *I need the car on x* can be inferred from speaker A's request to have B lend her car on Wednesday.

- (51) A: well if you can really very handsomely lend me your car on Wednesday (...)
 A: oh do you want the car for Wednesday - yes
 B: on on on Thursday morning

A: yes you take the car for that of course

B: or

A: no it's only Wednesday [~~that I need it~~] I'd be really very grateful

(LLC-1)

As far as full clefts are concerned, the results on the distribution of new and given information show that their variables can be either discourse-given or discourse-new. Thus, although reduced clefts always have a given variable, not all given variables necessarily yield reduced clefts and full clefts may restate already given material. In (52), for instance, the open presupposed proposition *x that runs the existing coffee bar* is introduced in the preceding clefted question and is restated in speaker B's answer.

(52) A: is it the Union Society is it the S'C that runs the existing coffee bar

B: no# it's the \Union# ^as dist\inct from SRC# that ^runs th/at#

(LLC-1)

These findings raise the question as to why full clefts would have a retrievable variable that is stated despite being given. This thwarts the expectations created not only by extant typologies but also by the descriptions of reduced clefts as the variant used when the variable is retrievable. I will attempt to answer this question by delving into the functional aspect of *it*-clefts in Section 5.3.

3.3.2 Towards a new typology of the referential information structure of *it*-clefts

As shown in the previous section, the analysis of the distribution of discourse-new and discourse-given reveals the existence of four possible patterns of referential information structure for *it*-clefts. These are summarised in the typology in Table 10 below.

Type of <i>it</i> -cleft	Value	Variable
new-given	new	given
new-new	new	new
given-new	given	new
given-given	given	given

Table 10. Typology of discourse-familiarity of *it*-clefts

This typology differs from those discussed in Section 1.2.2 in several ways. The first difference lies in the strict dissociation between referential and relational information structure. Instead of predicting prosodic patterns for each pattern of discourse familiarity, this classification focuses solely on the distribution of discourse-new and discourse-given information. The typology I propose also differs in the number of patterns identified. In contrast with existing typologies, which include either two (Prince 1978; Geluykens 1988; Collins 2006) or three categories (Declerck 1984) of *it*-clefts, my empirical study shows that four logical possibilities are attested: new-given, new-new, given-new and given-given. The new category in my typology is (iv) in which both the value and variable referents carry discourse-given information. Though this category is not represented in the extant typologies this thesis reacts to, its existence has however been acknowledged in studies by Johansson (2001) and Huber (2006). Its frequency of occurrence has also been documented in Johansson (2002) and Hasselgård (2014) who found that it is in fact the most frequent pattern in *it*-clefts.⁴⁴ Where the typology in Table 10 differs from Johansson (2002) and Hasselgård (2014) is that it includes reduced clefts and not just full clefts. Importantly, examining the given-given type more closely reveals that it includes examples in which the specification relation between the given constituents itself may be either new, as in (46), repeated as (53) below, or may be given, as in (54).

⁴⁴ It should be noted that the assessment of the discourse-familiarity of value and variable in *it*-clefts was carried out by Johansson (2002) and Hasselgård (2014) using a different classification than that used in the present study. One striking difference is that the category of ‘anchored’ is subsumed under givenness by Johansson (2002) and Hasselgård (2014) while it is treated as a subtype of newness in the analytical model I propose in Section 2.4.

- (53) A: I think I'm going to write I was thinking about this on the way down to the station this morning of writing every week to mom (...) and saying what I'd done in the week because she'd be quite interested and intrigued and eventually there'd be so much you'd find great difficulty in saying what it was (laughs) d'you know what I mean
 B: how d'you mean
 A: well sort of union account of what's happened
 B: yes but there's a tendency then for her to go and repeat everything annoyance of (...)
 A: **it was for "that reason# that I thought of doing 'this# (LLC-1)**
- (54) A: from Marlborough she has hit Reading at half
 B: splendid
 A: past eight in the morning
 B: splendid agreed but you're not
 A: that is that is the other side of Reading going into Reading I will be the other side of Reading going into Reading that's where she hit the traffic traffic going in to Reading from either side
 B: no you've missed the point the traffic you are worried about is the traffic going towards London
 A: no Petey at half past eight in the morning
 B: there is not a mysterious line which divides traffic going to London immediately at Reading
 A: no but there is traffic there is a traffic rush hour at Reading when traffic piles into Reading and it is **a'bout eight th'irty# that my ^mother has gotten st'luck# in ^traffic trying to get into R\reading (LLC-1)**

Taking example (53) again, both the clefted constituent and the presupposed open proposition in the cleft-relative include anaphoric references to elements stated in the prior context. Speaker A invokes *that* reason, i.e. x = the fact that his/her mum repeats everything s/he shares with her, to explain why s/he thought of doing *this*,

i.e. y = writing a shared account to his/her mum every week. Although both notions are already given, the relation of specification, *I did y because of x*, is only made explicit with the use of the cleft. In (54), the two speakers are arguing about the traffic between London and Reading in which A's mother got stuck. The value of the cleft is given verbatim in one of the previous turns, as underlined, and the variable is easily derivable from *that's where she hit the traffic [...] going in to Reading*. Unlike (53), the relation between the value and variable of (54) has already been established by speaker A in the first utterance of this excerpt. The speaker appears to be reiterating with a cleft the point that s/he had made earlier in the conversation for convincing purposes. The givenness/newness of the specification relation thus creates different motivations for clefts of the given-given type.

The distribution of the different patterns presented in Table 9 above goes against earlier accounts of *it*-clefts such as Lambrecht's (2001), who argued that the value prototypically introduces new information and the variable given information. This pattern comes as - and admittedly close - second to the given-given pattern. Strikingly, given-given, which had only been acknowledged in some studies (e.g. Johansson 2001, 2002; Huber 2006; Hasselgård 2014) but which had not been included in the extant typologies, is also the most common type for full and reduced clefts taken together. What the overall results show is that full clefts display a near-balanced distribution between given and new for both value and variable which illustrates the kind of 'equi-probable' choices described by Halliday and James (1993). In this respect, the discourse familiarity results differ from the focus assignment ones presented in Section 3.2, where one option was unmarked and more frequent, i.e. typically most prominent and contrastive focus on value. This shows that the referential and relational layers of information structure are manipulated as distinct systems by users of clefts.

3.3.3 Functions of *it*-clefts

The motivations behind the use of *it*-clefts are manifold as well as multi-layered. Semantically, *it*-clefts act as specifying devices while their bi-clausal syntax and information structure allow to lay emphasis on the clefted constituent (Huddleston

and Pullum 2002). As far as their discourse functions are concerned, clefts have been described as devices used to mark contrast (Givón 2001; Huddleston and Pullum 2002), organise text (Collins 1991), reinforce text cohesion (Dufter 2009a), set tone as discourse openers (Dufter 2009a), and establish question-answer relations (Delin and Oberlander 1995), especially in the case of reduced clefts (Declerck and Seki 1990). This leads to the question: can functionally coherent sub-categories of *it*-clefts be identified among the discourse-familiarity typology set out in the previous section?

Let me start with the new-given type. This type, in which the value introduces new information and the variable old information, occurs both in full and reduced clefts, as illustrated in (55).

- (55) A: you see I wondered who ultimately has more say (...) whether it is this daughter who's taken on the role of
 B: hm
 A: mother or# ^whether it is father# (LLC-1)

The open proposition *x who has taken on the role of mother* is mentioned verbatim in the full cleft, which creates an opposition between the two proposed values *this daughter* in the full cleft and *father* in the reduced cleft. Pairings of this kind are the most obvious use of contrastive *it*-clefts. However, example (56) illustrates that contrastive sets may also occur in full clefts in which the variable is restated despite being discourse-given.

- (56) A: Steven Peel supported you
 B: yes most curious
 A: now where did I hear that from
 B: probably me on the phone was it it was the day after on when I rang and we and we
 A: may have been may have been
 B: fixed up to meet in our house that
 A: yes yes
 B: was very curious Mallet is as I think possibly

- A: may have been this was from Chirk people that told me
 B: perhaps it was
 A: it's much more recently than that# ^**may have been Ivor B\ond told me#** (LLC-1)

Here, the variable *x told me* is already introduced in the preceding turn by the same speaker, although with a different value associated with it. I argue, however, that the full *it*-cleft at the end of the exchange can be interpreted as the conclusion of speaker A's thought process regarding the source of his knowledge that Steven Peel supported speaker B. Full clefts of the new-given type may therefore be used for textual organisation purposes, a use already noted by Collins (1991). This shows that although full and reduced clefts may function with the same information structural pattern, their discourse function differs. Reduced clefts exhibiting a new-given pattern are also used as mere answers to questions laying out the variable as in (57). This use was already highlighted by Declerck and Seki (1990) and Delin and Oberlander (1995).

- (57) A: and what is the pardoner being driven by
 B: well not not something outside himself really# **it's his ^own desire for m/oney I supp/ose# ^and a s\ort of power#** (LLC-1)

It-clefts of the new-new type, which are logically only full clefts in my dataset, introduce new information in both parts of the cleft. Because of the absence of a straightforward link with the preceding discourse, they typically introduce new discourse topics as in (58).

- (58) A: now what was the other thing I wanted to ask you is# ^**is it this year# that N\ightingale goes#**
 B: no next year
 A: hm sixty-four sixty-five
 B: sixty-five yeah
 A: I thought it was bef\ore sixty-five# so it's ^not until next year that the job will be \advertised# (LLC-1)

In this excerpt, three clefts are used. Speaker A first uses an interrogative pseudo-cleft announcing a topic shift, which is followed by the interrogative cleft in boldface, which features the new-new pattern and makes explicit the fact that speaker A is shifting to a new conversation topic. As the conversation continues, speakers A and B attempt to match the correct value with the variable introduced in the first cleft by exchanging their assumptions about the time Nightingale will retire. This type of use of new-new clefts is similar to Declerck's (1984) discontinuous category.

The third type, corresponding to given-new clefts, is a category of clefts for which two main discourse functions can be identified. In the first discourse-function exemplified in (59), a contrast is set up between two variables, e.g. *x where the fun is* and *x which is interesting*, linked to the same value, e.g. *the grammar*, which ultimately becomes evoked material for the second cleft. This type of cleft occurs specifically when two turns of two different speakers are involved.

- (59) A: it's the grammar where the fun is
 B: yes# **it's the grammar which is interesting#** one finds one | | | |
 find myself adopting phrases having this kind of chameleon approach
 to (LLC-1)

The second function associated with the given-new category is that of comment as in (60). In this case, the value is a continuous topic with mostly evoked NPs and a new presupposed open proposition presenting a comment on the value.

- (60) I've got some dough that's ready chilled here now and it's ready to roll
 out and# **^it's the r\olling out really# that ^g/ives# ^fl/aky pastry#**
its ^individual ch\aracter# (LLC-1)

Finally, given-given *it*-clefts also exhibit two functions. Firstly, a small number of cases in the LLC-1 of the given-given articulation correspond to mere repetitions as in the third cleft of excerpt (61).

- (61) A: at Gothenburg I've got all sorts of support from the Germans Turner and Pickering and so on you know and those English philologists that I have met are in that I've talked to are most enthusiastic
 B: yes *it's not philologists you want to convince*
 A: hm
 B: it's the people with money
 A: ^it's the people with the money# (LLC-1)
- (62) A: no but there is traffic there is a traffic rush hour at Reading when traffic piles into Reading and it is a^bout eight th\irty# that my ^mother has gotten st\uck# in ^traffic trying to get into R\reading (LLC-1)

Similarly to the given-new type, cases of given-given *it*-clefts may also allow a speaker to mark a contrast like the first cleft in (61). However, unlike in (59), the first cleft in (61) does not make use of contrastive listing of values, but rather cancels the specification relation between the negated value *the philologists* and the variable. Speaker B then puts forward the correct value *the people with money* with the following reduced cleft. It is therefore neither the value nor the variable that are new but the specification relation between the two referents. In instances of the given-given type in which the specification is given, e.g. (62), the speaker may use the cleft to strengthen his/her argument. This section was an attempt at determining the different discourse functions associated use of full and reduced *it*-clefts as found in the LLC-1.

3.4 The interplay between the syntax and relational and referential information structure of *it*-clefts

In some approaches, the syntactic constituents of value and variable are said to code argument focus-presupposition and to correlate more or less straightforwardly with discourse-new and discourse-given information in the cognitively-oriented sense of Prince (1992) (Lambrecht 2001). The results I have thus far presented have

disproved this one-to-one match. Hence, it is important to establish the interplay between the relational, i.e. focal/non-focal, and referential, i.e. relating to discourse-familiarity, information structure of *it*-clefts by treating them as separate aspects. This section examines the interplay between the two layers (Section 3.4.1) as well as the relation between the bi-clausal syntax and prosody of *it*-clefts (Section 3.4.2). I also address cases of selective focus (Section 3.4.3) and the role of onsets (Section 3.4.4).

3.4.1 The relation between syntax and prosodic foci

One of the shortcomings of the existing informational taxonomies of *it*-clefts (e.g. Prince 1978; Declerck 1984; Geluykens 1988; Collins 2006; Huber 2006) is that they all collapse discourse familiarity and focus assignment, the former of which they use to predict a predominant prosodic pattern. I have shown in Section 3.2 that speakers actually afford a great variety of information focus choices. In Section 3.3, I have provided evidence that the discourse-familiarity of value and variable likewise displays variation. In what follows, I examine the moment-by-moment interplay between the syntax of clefts, which highlights the value, and the prosodic choices speakers make to manage the information flow. I illustrate this first for clefts with a single information focus, as is the case in (63)–(65), and then for clefts with foci on both value and variable.

- (63) A: was it his first novel the first one he actually wrote
 B: no no he was writing A Passage to India at the same time hm and he he stopped writing Passage to India and hm stopped off to write Maurice anyway we
 A: oh it was much later than I thought I always got the impression it was
 B: I think# **it was nineteen fift\een that he wrote it#** (LLC-1)
- (64) because there's not the same pressure on the material# **it's the the p\op material that counts#** (LLC-1)
- (65) A: did you meet Fuller
 B: yes# **it was ^he who inv\ited me#** (LLC-1)

When speakers choose to use a cleft, it always syntactically foregrounds the whole postverbal NP as the value being specified for the variable, but it offers great possibility of choice for the assignment of information focus via nucleus placement. The selection of information focus is motivated by ‘communicative purpose and the extent of presumed shared information’ (O’Grady and Bartlett 2019: 192). In (63), the whole value NP *nineteen fifteen* is focal. It is what Halliday (1967a: 207) characterises as a marked, i.e. contrastive, focus, i.e. one that is informationally contrastive (in this context, with the other dates considered for Forster’s writing of *Maurice*) and packages the rest of the information in the unit as a presupposition. Example (64) also has a marked focus, which moreover singles out the premodifier *pop* of the value NP only, evoking contrast with the other types of music ‘material’ (classic, etc.) in the library, mentioned earlier in the discourse. This selective focus on *pop* is related to *material that counts*, which is discursively anchored-new, but is packaged as a presupposition within the information unit. Interestingly, the informational presupposition and focus do not coincide with the value and variable of the specification relation. *The pop material* is the value being identified for *x that counts*, the narrow focus is placed on *pop*. All of this contributes to the communicative foregrounding of *pop*. Example (64) illustrates how my functional approach brings out the moment-by-moment interplay between the choice of elements of the prosodically coded information structure (focal/non-focal) and of constituents of the syntactic structure (value-variable) and their discourse-familiarity. In (40), repeated above as (65), the anaphoric pronoun value, *he* referring back to *Fuller*, carries an onset accent⁴⁵ but the information focus of the unit is put on the final lexical element of the variable, *invited*. The cleft-relative clause conveys discourse-new, non-predictable information, put forth by B as an indirect response to A’s question on whether B had met Fuller. The emphasis produced by the prosodic coding of the cleft is thus shifted towards the end of the tone unit which, as a result, creates a discrepancy between the syntactic highlighting and the prosodic one. Given that examples like (65) represent 23.8% of

⁴⁵ Which indicates a ‘disruption’ with the expectations of the previous turn, in that Fuller, not the speaker, initiated the encounter (see section 3.4.4 on the discourse function of onsets).

all full clefts, it further reinforces the argument that the notion of focus, in the general sense of highlighting or emphasising, in *it*-clefts should be treated with caution and should be clearly defined as either coded by syntax, prosody, or both.

I now turn to clefts with information focus on both value and variable, which account for 63.6% of our data (see Table 3). Within that portion, I found that foci on the value are relatively more prominent than foci on the variable in about 65%, but not as a rule (Table 4), as had been suggested by e.g. Prince (1978) and Declerck (1984). I will reconsider examples (34) and (36), reproduced as (66) and (67), to bring out the interplay between the syntactic constituents and the hierarchically ordered foci as they unfold in real time.

- (66) Hemsley chipping the ball into the centre onto the head of Scullion#
from the ^{M^}head of Scullion it's J^Mames# that g^Lets it# but only as
far as Hockey (LLC-1)

In (66), by construing *James* as the value, the speaker foregrounds the point that James, not any of the other players, got the ball. In the prosodically coded information structure, *James* is marked by the higher-pitched nucleus as the most salient, but yet already given, information that the speaker wishes the hearer to attend to, while *gets* has a secondary information focus.⁴⁶ This is in line with Nelson's (1997: 346) explanation⁴⁷ of the frequent use of clefts in live sports commentaries, in which the speaker has to describe and react to a series of fast-paced actions.⁴⁸ Clefts allow the speaker to give prominence to the ever-changing identity of the player in possession of the ball, while also putting a secondary focus on the specific actions being described.

⁴⁶ An alternative analysis of (66) suggests that the low nucleus on *gets* exemplifies the low termination which is typical of a point of completion in discourse.

⁴⁷ I thank one of the anonymous reviewers who reviewed Bourgoïn (submitted) for bringing this reference to my attention.

⁴⁸ The sports commentary considered here contained 7 out of the 238 clefts in the whole corpus.

Example (67) is taken from a conversation between a salesperson from a building society (B) and a prospective customer (A), discussing problems associated with taking out a mortgage loan.

- (67) A: so that that's the kind of problem the other problem is that we haven't got an awful lot of capital we've only got about thirteen hundred pounds in capital
 B: hm hm
 A: so although I could imagine that we could hm on our joint salary get perhaps quite a a high mortgage# **it's the paying it back at the beg^M\inning# that's ^Hgoing to be d^M\ifficult#** (LLC-1)

The value consists of unpredictable information, as signalled by the high onset, that has not been mentioned in the preceding discourse, while the open proposition *that is going to be difficult* is inferable from speaker A's admission of the problems she faces taking out a big loan. With the cleft, speaker A identifies the initial payments - rather than the capital - as the main difficulty. In the information structure, the speaker first focuses on *paying it back at the beGINning* and then on *DIFFicult* in the cleft-relative clause, presenting the two prosodically equal foci as also equally prominent informationally.

From the discussion of the aforementioned examples, I conclude that clefts allow speakers to highlight elements by means of two strategies, syntactic and prosodic, which may reinforce each other or create their own different types of prominence in sequence motivated by specific discourse functions. The de-coupling and cross-coupling of prosody and lexicogrammar allows to do justice to the many different information structures with a great variety of focus assignments actually attested in usage.

3.4.2 The relation between discourse-familiarity and focus assignment

The analytical part of this chapter has so far focused on the referential and relational information structure of *it*-clefts separately. In this section, I present an

enhanced typology in which focal/non-focal patterns are mapped onto the proposed discourse familiarity typology. In the four main typologies of *it*-clefts, cleft types are each matched with a single prosodic realisation in which new information is claimed to be coded with (strong) stress and old information with weak stress (Prince 1978; Declerck 1984; Collins 2006) or absence of tonic prominence altogether (Geluykens 1988; Huber 2006). To establish the relation at play between discourse-familiarity and focus assignment in the LLC-1 dataset, I adopt Halliday's (1967a, 1994, and Greaves 2008) functional approach and treat information focus as revealing the speaker's choices on what to present as the most salient piece of information in the linear chain of successive intonation units. Unlike Prince (1978), Declerck (1984), Geluykens (1988), Collins (2006) and Huber (2006), I do not find any systematic mapping between focus placement and distribution of given vs. new information. Instead, each discourse-familiarity pattern has a main unmarked, i.e. most frequent, focus assignment pattern and at least one secondary marked, i.e. less frequent, one. These patterns are summarised in Table 11 below.

	Value	Variable
	Discourse-new	Discourse-given
Unmarked	foc	foc
Marked		foc
	foc	
	Discourse-given	Discourse-new
Unmarked		foc
Marked	foc	foc
	Discourse-new	Discourse-new
Unmarked	foc	foc
		foc
	Discourse-given	Discourse-given
Unmarked	foc	foc
Marked		foc

Table 11. Overview of focus assignment patterns of each discourse-familiarity pattern

The diversity of focus assignment patterns existing in *it*-clefts corroborates the untenability of the conflation between the referential and relational dimensions of the information structure of clefts. Among the different discourse-familiarity categories, the new-given one can for instance receive three different focus patterns. In the most frequent unmarked pattern, both value and variable are in focus and in which the value receives contrastive focus and the variable presenting focus. This is instantiated in (68) where both items *themselves* and *question* bear a nuclear accent. In the second most frequent pattern, only the variable bears focus as in (69). In the third pattern, (70), the value bears the unique, contrastive focus. While the pattern exemplified in (70) would typically be analysed as the ‘prototypical’ focus-presupposition pattern by some authors (e.g. Lambrecht 2001), it represents the least frequent pattern for new-given *it*-clefts in the LLC-1.

(68) it is the professional bodies **themselves**# which have got to ask this **question**# (LLC-1)

foc

foc

(69) it was Norma Harley that did most of the **talking**# (LLC-1)

foc

(70) it’s just one **question** that they have to do# (LLC-1)

foc

Clefts of the given-new type are associated with one unmarked and one marked focus assignment pattern. They can either have a unique presenting or contrastive focus in the variable, as in (71), or one (or more) focus in each part of the cleft, as is the case in (72), with contrastive focus generally located in the value and presenting focus in the variable.

(71) **yes**# it was he who **invited** me# (LLC-1) (LLC-1)

foc

- (72) it's the academic **str\ucture#** of the **univ\ersity#** that that we're
 foc foc
concerned about# (LLC-1)
 foc

The focus assignment patterns available for new-new clefts are the same as for given-new clefts, but their frequency of use is the opposite. Thus, in the unmarked pattern in (73), both value and variable are prominent, with contrastive focus on the value and presenting focus in the variable, while the marked cases such as (74) display one (or more) information focus in the variable.

- (73) it was **p\ort#** r\eally# that kept them **w\arm#** in the eighteenth
 foc foc
c\entury# (LLC-1)
 foc
- (74) it is prejudice that causes **unh\appinesses#** that causes **v\iolence#** and
 foc foc
 all sorts of **th\ings#** (LLC-1)
 foc

Finally, clefts which carry given material in both their value and variable may be prosodically coded in two manners. In the most frequent pattern, illustrated in (75), both value and variable carry one (or more) nuclear accents, once again with contrastive focus in the value and presenting focus in the variable for the most part. In the marked prosodic pattern, only (part of) the variable is accented as in (76).

- (75) so that it's the faculty of **/arts#** ^or the faculty of **econ/omics#** ^or
 foc foc
b/oth# with^in the **NF\O#** that'll be ^putting him **f\orward#** (LLC-1)
 foc foc foc
- (76) A: but if you're going to provide anything at all useful about the
 language it's can't be limited by the time factor it doesn't seem to
 make sense actually

B: well# it's it's ^money you know that \limits you# (LLC-1)

foc

Because the possibility for the copula *be* to be prosodically prominent relates to the syntax of the cleft - the few cases from LLC-1 being dependent interrogative clefts for the majority - and not to its discourse-familiarity, clefts with an accented copula are not included in the proposed taxonomy. As far as reduced clefts are concerned, they typically carry a unique nuclear accent on the value. Only five tokens out of all reduced clefts have a different pattern in which focus was borne by the copula. These results will be addressed at greater length in Section 6.3.2.

The typology proposed in Table 11 further backs up the argument that a distinction should be drawn between presenting and contrastive focus in *it*-clefts. In the majority of tokens, contrastive focus is coded in the value while presenting focus relates to the information conveyed in the variable. As such, the content of the cleft-relative clause does not relate to a mere presupposition but rather introduces new information that pushes communication forward.

3.4.3 Selective focus

I have so far presented a general overview of the relation between coded forms and coded meanings in *it*-clefts and have argued that these layers, although distinct, interact in a number of ways. Given that *it*-clefts are specificational constructions, a further distinction should be made between the syntactically coded specification relation from prosodically coded information structures. This is particularly significant for examples with contrastive focus on only a part of the value NP, which I refer to as selective focus.

So far, little substantial study has been devoted to cases of selective focus in *it*-clefts with the exception of Velleman et al. (2012) and Jespersen ([1937] 1984: 75) who had already noted that the “emphasis (and stress) is even frequently laid on another word than the one singled out by being made the predicative of *it is*”. Collins (2006: 1708) observed with regard to example (77) that the focal status of *garden* makes the following noun *space* fall within the post-nuclear tail which is, as a result, presented by the speaker as given information, regardless of its actual

informational status.⁴⁹ This exemplifies marked tonicity in Halliday's (1967a, 1994) terms, i.e. a contrastive focus, which in (77), is on the classifying modifier *garden space*, as opposed to *house space*. In (78), the marked contrastive focus is also borne by a lexical item other than the last one of the tone unit. However, the matrix clause in (78) instantiating unmarked focus is realised as a separate tone unit whose nuclear accent is on the last lexical item of the qualifying value *the paying it back at the beginning*.

- (77) no# and it's g\arden space# that is s\o# pr\ecious# for k\ids# not
h\ouse space# (LLC-1)
- (78) I could imagine that we could on our joint salary get perhaps quite a a
a high mortgage# ^it's the paying it back at the beg\inning# that's
^going to be d\ifficult# (LLC-1)

In my dataset, I found 22 tokens of selective focus on the value NP, i.e. 9.2 %, of which 59% are full clefts and 41% reduced clefts, as shown in Table 12 below.

	Full clefts	Reduced clefts	Total
Raw count (%)	13 (59%)	9 (41%)	22

Table 12. Distribution of selective focus

Besides defining modifiers of the head noun as in (77), selective focus can also be borne by quantifiers and determiners, e.g. (79) and (80) below. Prosodically, selective focus can be realised with the full set of tones available for English, i.e. fall, rise, fall-rise, rise-fall, level. The prosodically-prominent constituent is contrasted with alternatives which can be explicitly or implicitly stated in the context while the rest of the information unit is deaccented and therefore presented as presupposed. Importantly, the extent of the open proposition introduced in the cleft differs depending on whether the variable forms an

⁴⁹ An alternative analysis of (77) would consist in treating *garden space* as a compound item which would then receive unmarked information focus and not selective focus.

information unit in its own right, as in (80), or is represented as presupposed information, as is the case in (81).

- (79) A: something of Gertrude's character is thrown up in that scene
 B: how many times is Hamlet alone with his mother as he is in the closet
 can you remember
 A: I think it's just that *\one scene# [in which Hamlet is alone with his mother]* (LLC-1)
- (80) but I understand that we've been honoured by the visit of several distinguished pro-Market editors for this debate with those editors here# *it is th\eir credibility# that's in qu/estion#* (LLC-1)
- (81) but it it's the... *p\op* is the area# where adult and junior mainly overl\ap# [...] there's not the same pressure on the material# *it's the p\op material that counts#* (LLC)

If the variable is represented as presupposed information - whether overt in a cleft-relative clause or anaphorically inferable - then it also includes the presupposed information in the value, and this whole extended presupposition is represented as given in, or inferable from, the preceding discourse. Thus, in (81), the presupposition is not just coded by the cleft-relative clause but also includes the deaccented head noun *material*, yielding *x material that counts*. It is to this presupposition that the prosodically coded selective focus *p\op* is related. The speaker is concerned with the tagging of books in libraries and singles out the particular pressure on 'popular' books and CDs (which are in demand with both adults and juniors). As pointed out by Velleman et al. (2012), the selective focus on *p\op (material)* impacts on the exhaustivity effect of the cleft. The identifying matrix triggers a conversational implicature of exclusiveness to the effect that all other alternatives, i.e. other types of material, do not count. But because it is a conversational implicature, the exhaustive/exclusive effect can be cancelled (cf. Velleman et al. 2012: 443), for instance by adding *there's the cl\assic material that counts too*. In (79), the presupposition includes the anaphorically retrievable variable plus the deaccented head noun *scene*, yielding *x scene(s) that Hamlet is alone with his mother*. The selective focus on *that \one scene* also contextualises

as exclusive focus, excluding the existence of ‘more’ scenes in which Hamlet is alone with his mother. But in this example, the implicature triggered by the identifying main clause is also asserted by the exclusive focus particle *just* (Velleman et al. 2012: 443). As pointed out by De Cesare and Garassino (2015), the use of exclusive focus particles like *just* and *only* modifies the pragmatic mechanism of the conversational implicature of clefts.

In full clefts with selective focus, the cleft-relative clause is more frequently uttered in information units of its own, as in (77) or (80). This type of cleft has not only a different internal information structure but it also relates differently to the preceding discourse. Firstly, the selective focus defines *only* the deaccented head noun as presupposed. In (77), the notion of *space* is indeed given in the preceding discourse, but in (80), *credibility* is arguably less strongly predictable from the preceding discourse so that the information structure chosen by the speaker makes the hearer ‘accommodate’ the presupposition that credibility is at issue (Schwenter and Waltereit 2010). Secondly, the contrastive selective focus itself activates an anaphoric link with the preceding text, signalling “contrast with what has been said before or what might be expected” (Halliday 1967a: 206). In (77) the selective focus on *g\arden space* positions that *type* of space as the element from the preceding discourse to be attended to while in (80) the selective focus on *th\eir credibility* turns the tables against the Pro-Market editors (to whom *their* refers).

In their internal information structure, these clefts first feature a selective focus-presupposition structure within the field of the matrix, typically followed by an unmarked information structure with initial recoverable information (the relative anaphor) followed by presenting foci. In (77), the cleft-relative clause features three presenting foci with emphatic rhetorical effect. The speaker presents her view that *garden space* is *s\o# pr\ecious# for k\ids#* as fresh information the hearer has to attend to. In this sense, this example clearly illustrates why ascribing the information structural notion of ‘presupposition’ to the cleft-relative clause-constituent does not work. As I have argued throughout this chapter, it is the identifying matrix that triggers the exclusiveness implicature applying to the specificational relation between the value and the variable. As a conversational implicature it can in principle be cancelled, unless the exclusiveness is asserted by another element, like *just* in (79). This is also the case in (77), where

the selective focus associates the exclusiveness implicature with *g\arden space* being so precious for kids, but then adds explicitly that the only (binary) contrasting element in this context, *h\ouse space*, is not precious for kids. Thus it is contextually not possible to cancel the exclusiveness implicature by adding another option. In (80) the selective focus triggers the exclusiveness implicature for *their*, and not anyone else's, credibility being called into question.

Example (82) illustrates the infrequent pattern with a selective focus on an element of the value and a contrastive focus on an element of the variable. The selective focus on *this* is motivated by the exophoric pinpointing of *which chair*, and not just *what*, was offered to Barry by the speaker. In the cleft-relative clause, there is a non-final marked focus on *\offering*, which, as becomes clear in the following information unit, entails potential retraction of the offer to Barry, as the speaker asks another person whether he wants the chair.

- (82) A: now you see the disadvantage of these damned ch\airs Barry#
B: yeah holding both his sides
A&B: (2 to 3 seconds untranscribable)
A: it was th\is one# I was \offering Barry# do you w\ant it# (LLC-1)

I conclude that the information structural patterns of clefts proposed in Section 3.4.2 can also account for selective focus on an element of the value. In case the information pattern is of the focus-presupposition type, then the selective focus relates to a presupposition that includes both the head of the value NP and the variable.

3.4.4 Onsets and information flow management

In this section I consider the interplay between actual textual givenness/newness and the speaker's selection of information foci and onsets. As shown in Section 3.3, values and variables can both be either textually new or given. Importantly, full clefts can have a given-given pattern, as in (83), which has not received the attention it deserves in the literature.

- (83) A: from Marlborough she has hit Reading at half
B: splendid
A: past eight in the morning
B: splendid agreed but you're not
A: that is that is the other side of Reading going into Reading I will be the other side of Reading going into Reading that's where she hit the traffic traffic going in to Reading from either side
B: no you've missed the point the traffic you are worried about is the traffic going towards London
A: no Petey at half past eight in the morning
B: there is not a mysterious line which divides traffic going to London immediately at Reading
A: no but there is traffic there is a traffic rush hour at Reading when traffic piles into Reading# and it is a ^H^bout eight th\irty# that my ^mother has gotten st\uck# in ^H^traffic trying to get into R\reading#
(LLC-1)

In this excerpt, the two speakers are arguing about the traffic between London and Reading in which A's mother got stuck. The value of the cleft is given verbatim in one of the previous turns, as underlined, and the variable is easily derivable from *that's where she hit the traffic [...] going in to Reading*. Moreover, the relation between these elements has already been established by speaker A in the first utterance of this excerpt.⁵⁰ The cleft comes as a closing statement meant to resolve the disagreement between the two speakers. Speaker A reasserts all the points she had already made about her mother's traffic conundrum, assigning information foci to the crucial elements, *about eight th\irty, has gotten st\uck, into R\reading*, all carrying assertive falling tones. Moreover, each of these three tone units contains a high onset (indicated by small capital H before ^), by which speaker A signals that she projects a position contrary to that assumed by speaker B in their discussion.

⁵⁰ I thank one of the anonymous referees for Bourgoïn et al. (2021) for pointing this out.

In (84), the speaker uses a cleft to move from a general current concern, i.e. having negative feelings towards rock and roll singers, to the recalling of a specific instance of this. Both value and variable contain discourse-new information. Collins (2006: 1713) observes that in examples like these the proposition is held ‘in store’ in anticipation of its relevance to the unfolding discourse.

- (84) there’s something that makes us feel savage about these rock and roll singers and I hate it in myself and I see it in a lot of other people# **now it’s** ^{H^}**only about a year ago**# - that ^{H^}**on this programme**# ^{H^}**we were asked about I think it was Tommy StVeele**# **being mobbed**# and I remember making some perfectly horrible remarks (LLC-1)

Here, a new temporal setting, carrying information focus, is introduced in the syntactically highlighted value-position, *it’s only about a year ago*. This new temporal setting is added to the common ground shared between speaker and hearer and provides the knowledge necessary to engage with and react to the upcoming information in the cleft-relative clause. The new information in the cleft-relative clause is put across in three information units. Three of the four information units of the cleft have high onsets including at the beginning of the cleft construction, which signal the reset involved in shifting from general current concern to a specific temporally located instance of it.

High onsets are an important feature of the prosody of clefts, whose informational and interactional meanings warrant study. In the *it*-clefts in my data, the onsets at the beginning of the whole construction are all high. This means that clefts are always used in our data to signal a reset vis-à-vis what preceded - be it another speaker’s turn, as in (83), or the speaker’s own previous utterance, as in (84). Onsets in the tone unit of the value NP serve specific rhetorical and interactional effects, which may link up with the typical discursive functions of specificational clefts.

In the full cleft in (85), the high onset on the negator *not* reinforces the contrastive focus on the value *philologists*, which is rejected as the value of *x you want to convince* by means the negation marker *not* which bears a high onset. The following reduced cleft then provides the accurate value *people who have the*

money. (85) illustrates how the speaker projects that the assumption held by the hearer is not correct and that the information added is contrary to the hearer's expectations.

- (85) A: [...] those English philologists that I have met are in that I've talked to are most enthusiastic
 B: uh it's ^{H^}not phil\ologists# you want to c\onvince#
 A: hm
 B: it's the people with m\oney# (LLC-1)

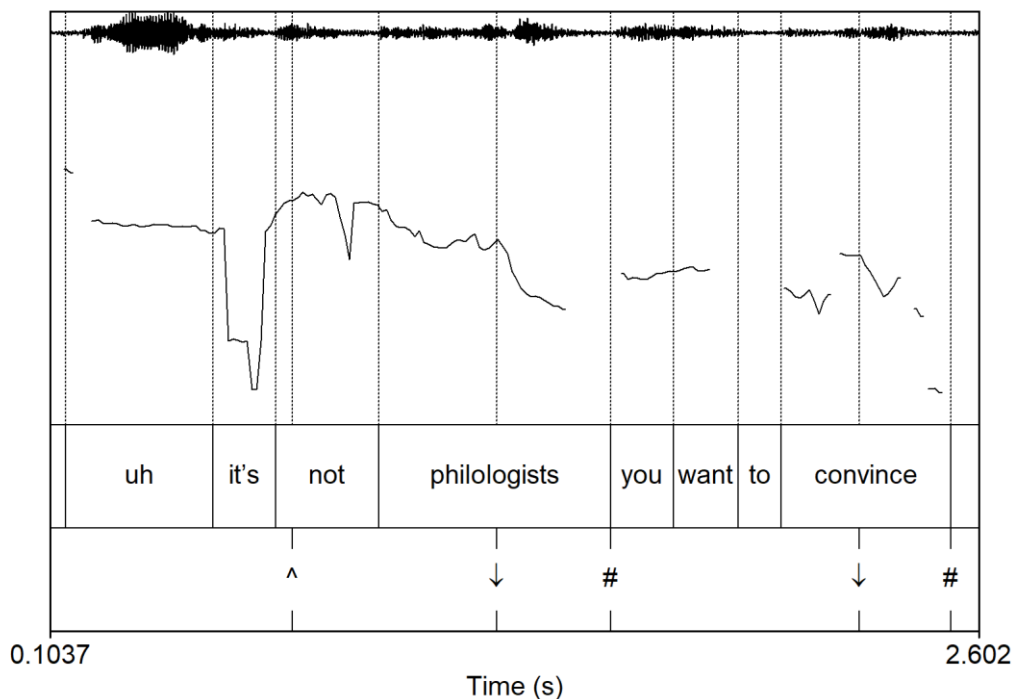


Figure 4. Praat image of (85)

In (86), the high onset on the adverb *only* makes explicit the implicature of exhaustivity triggered by the identifying matrix: *when they turn facing us* specifies the only condition under which *you get the underside full on*.

- (86) A: yes two of them sitting there the thing that that catches the eye if anything does is the white underside when they're sitting upright but uh perhaps the the most obvious thing is how well camouflaged they are in fact how inconspicuous they are hm

B: yeah# - it's ^{H^}only when they turn facing us# that you ^{M^}get the

A: yes

B: the \underside# - ^{H^}full \on as it were# (LLC-1)

In sum, clefts attest each possible distribution of discourse given or new information over value and variable, including given-given and new-new. Speakers can impose multiple foci on the points they want the hearer to sequentially attend to, irrespective of whether all the propositional material is already present in the context, as in (85), or is textually new, as in (86), or is partly given and new. But when it comes to the initial onset of the cleft, the speakers in the LLC-1 dataset generally realise a high onset. In my data, *it*-clefts hence always convey that there is some reset (change of topic, disruption, contradiction, etc.) vis-à-vis the preceding utterances.

3.5 Concluding remarks

The goal of this chapter was to present a fine-grained account of the use of the *it*-cleft in spoken English by examining the distinct layers of their functional-structural organisation in naturally-occurring data. In what follows, I summarise the main conclusions I arrived at with regard to the syntactically coded specification relation and the relational and referential information structure of *it*-clefts before commenting on the interplay between these three layers of organisation.

First, I have shown that the syntax of clefts involves a matrix whose post-verbal complement, the clefted constituent, is the antecedent of the relative anaphor of the cleft relative clause. The matrix is an identifying clause, which directly inserts the value(s) qualifying for the variable expressed by the cleft relative clause into the antecedent position. That is, the morphosyntax codes identifying-specificational meaning. The clefted constituent, which is assigned the role of identifier, may be of a wide range of syntactic categories and functions, but displays a preference for subject NPs. Aspects related to modality, negation and exhaustivity are also found to be at play in some occurrences of *it*-clefts, all of which may interact with the specification relation.

Secondly, my qualitative and quantitative analysis of the prosodically coded relational information structure of *it*-clefts has shown that there exists a higher number of information structural patterns than previously described. Nuclear accents may be borne by the clefted constituent, the presupposed open proposition in the cleft-relative clause, or both. Full clefts predominantly carry at least one information focus in both value and variable, in which case the accent on the value is typically more prominent. As far as reduced clefts are concerned, the vast majority of them bear a unique information focus on the value. In terms of contrastive vs. presenting focus, both full and reduced *it*-clefts most frequently have contrastive focus on the value and presenting focus on the variable, regardless of whether it is explicitly stated or not. With this, I have shown that informational highlighting is achieved through the interplay between the cleft's bi-clausal syntax and focus-marking through accent placement, which may reinforce each other or create their own different types of prominence in sequence. I have also highlighted the role of onsets, i.e. the first pitch accent of the cleft, in projecting how the upcoming propositional material relates to the expectations generated by the preceding context. Strikingly, in my data clefts generally have a high onset, signalling some disruption of expectations, which can be linked to typical discursive functions of clefts such as establishing an overt contrast or expressing exclusive focus.

Relying on the comparison between four typologies of *it*-clefts (Prince 1978; Declerck 1984; Geluykens 1988; Collins 2006) and the analysis of the discourse-familiarity of *it*-clefts, I have proposed a quaternary typology of the referential information structure of *it*-clefts. In order to assess the discourse-familiarity of both nominal referents in their value and open propositions in their variable, I relied on the newly-proposed analytical model distinguishing between two types of discourse-given, i.e. evoked and inferable, and two types of discourse-new, i.e. new-anchored and brand new, all of which were assessed with a textual approach. The results of the analysis of discourse-familiarity showed that given and new information are partitioned into four logical patterns, namely new-given, new-new, given-new and given-given. The given-given pattern, which is not included in any of the typologies and is only documented in a handful of studies (Johansson 2001, 2002; Huber 2006; Hasselgård 2014), turned out to be the most frequent pattern for full clefts. It was

found to be productive in reduced clefts too, which had hitherto not be acknowledged in the literature. Interestingly, the specification relation in clefts of the given-given type was observed to be either new or given.

From a functional perspective, different pragmatic motivations for the use of each type were established, among which a new type of contrast, between variables, was identified.

Finally, the analysis of the interplay between cleft syntax, specification meaning, focus assignment and discourse-familiarity led to a number of conclusions. First, the de-coupling and cross-coupling of prosodically coded focus and syntactic highlighting shows that the two strategies do not systematically combine to mark the clefted constituent focal. This constitutes one of the choices made by speakers which may be motivated by different discourse functions and rhetorical purposes. Secondly, the need to treat the referential and relational layers of information structure as non-conflatable, but yet interacting, aspects of *it*-clefts is highlighted by the fact that multiple focus assignment patterns, i.e. typically one unmarked and one or more marked patterns, may be mapped onto each of the four discourse-familiarity subtypes of clefts. Finally, distinguishing between syntactically coded specification relation in *it*-clefts from their prosodically coded information structures allows to make sense of cases of selective focus in which only part of the value is in focus. These clefts typically feature a selective focus-presupposition structure in the matrix clause typically followed by an unmarked information structure in the cleft-relative clause consisting of initial recoverable information and a presenting focus.

Thus, what this chapter has shown is that English *it*-clefts emerge as a construction whose different functional-structural layers can be rhetorically manipulated by speakers in multiple ways to achieve varying rhetorical and communicative goals.

Chapter 4: The French *c'est*-cleft

In the previous chapter, I examined how the information structural patterns of the English *it*-cleft map onto its syntactic structure, which codes its specificational meaning. In the present chapter I focus on its French equivalent, the *c'est*-cleft. In order to lay an adequate groundwork for the cross-linguistic analysis developed in Chapter 5, I follow the same analytical course as that adopted for *it*-clefts. As such, I develop an account of *c'est*-clefts which studies how patterns of prosodically coded focus assignment and discourse-familiarity map onto the constituents of the specificational relation, i.e. the value and the variable. I start by summarising the main morphosyntactic properties associated with *c'est*-clefts as specificational constructions in Section 4.1. In Section 4.2, I describe their relational information structure by outlining their numerous prosodic profiles. Section 4.3 focuses on the referential layer of the information structure of *c'est*-clefts, which involves the discourse-familiarity of their referents. I then investigate the nature and extent of the interplay between the two layers in Section 4.4. The discourse functions of *c'est*-clefts as observed in empirical data are addressed in Section 4.5. Finally, I conclude this chapter by summarising the key findings on French *c'est*-clefts.

4.1 Morphosyntactic properties

Like the syntax of the English *it*-cleft, that of the French *c'est*-cleft codes identifying-specificational semantics. The biclausal syntax consists of an identifying matrix and a relative clause whose antecedent, the clefted element, is a nominal constituent of the matrix. The antecedent and cleft-relative clause code the value and the variable of a specificational relation. The identifying matrix directly inserts the value(s) qualifying for the variable expressed by the cleft relative clause into the antecedent position. In this section, I investigate the general morphosyntactic properties of the identifying matrix clause of *c'est*-clefts which include the morphology of the copula *être* (Section 4.1.1), the paradigm of relative markers (Section 4.1.2), the syntactic category and function of the clefted constituent (Sections 4.1.3 and 4.1.4) and the expression of modality and negation (Section 4.1.5).

4.1.1 Morphology of the copula and the relative marker

To start with, I investigate the morphological flexibility of the copula *être*. Although the syntax of the *c'est*-cleft as a whole is relatively fixed - with a matrix clause introduced by non-referential *ce* and copula *être* followed by the clefted constituent and then the cleft-relative clause - the morphosyntactic profile of the matrix clause itself exhibits variation with regard to person agreement and tense of the copula *être*. This variation can be traced back to the first uses of *c'est*-clefts in Old French and later on in Middle French in which the agreement between the person of the copula and that of the referential NP coding the value was more common than in present-day French (Hatcher 1948; Dufter 2008). Along with the loss of person agreement, Modern French also shows more restriction in number agreement between the copula and the value NP which is not the case in all Romance languages (Dufter 2008). Both tendencies are shown in the cleft structure in (1) from Middle French taken from Dufter (2008: 5). Here, the clefting of the item *vous* imposes the use of second person plural form *estes* of the copula.

- (1) Ce **estes** vous que je doy remercier
 it be.2PL you.PL that I must.1SG thank
 ‘It is you (PL / SG.POLITE) whom I have to thank.’ (Dufter 2008: 5, my emphasis)
- (2) *c'est* vous qui les mettez ou quand vous dites *c'est* en prenant le parti du lecteur (CRFP)
 ‘it’s you who put them on or when you way it’s by siding with the reader’

The combination of the loss of person agreement in all types of copular clauses with subject *ce* from the Middle French period onwards and the spread of non-number agreement has led to the generalised use of third-person *est* (Dufter 2008).⁵¹ This

⁵¹ Hatcher (1948: 1056) attributes this ‘figement’ to the equative nature of the cleft. He argues that it “must constitute a (potential) challenge to the rules of normal syntax; and it is not surprising if, occasionally, the fundamental law of the agreement of subject and verb is broken.”

is exemplified in (2) retrieved from the CRFP. Like (1), the clefted constituent in (2) is the pronominal form *vous*, but the copula is this time conjugated in the third-person singular. The same person mismatch is found in clefts with other types of clefted pronouns such as *eux* in (3) and nominal clefted constituents such as *les réactions de l'enfant* in (4). In both cases, the copula is in singular form despite its complement being plural. By contrast, the copula in (5) agrees with the third plural anaphoric clefted constituent. *Ce sont* represents the only person agreement other than *c'est* which is still productive.

- (3) *c'est eux* qui donnaient la preuve hein le le l'enregistrement où ils nous cassaient là la fameuse cassette (CRFP)⁵²
 'it's them who were providing evidence wasn't it the the recording in which they were talking us down there the infamous cassette'
- (4) moi je trouve que ça joue énormément et où ça joue énormément *c'est les réactions de l'enfant* (CRFP)
 'I for one think that it plays a big role and where it plays a big role it's the child's reactions'
- (5) et ce *sont eux* qui nous donnent le carnet de bord pratiquement (CRFP)
 'and it's them who give us the log book almost'

Despite this tendency to move towards a generalised use of third-person singular (Dufter 2008, 2009a, b, 2015), *c'est*-clefts with a plural form of *être* are still productive in Modern French. In the CRFP dataset, 13 occurrences of plural *sont*, as

⁵² The translations provided for examples of French *c'est*-clefts are meant to closely reflect the syntactic structure of clefts in the original language in order to facilitate the understanding of the claims made in this study. The resulting *it*-clefts may, at times, appear ill-formed due to pragmatic differences between the two languages (see Bourns 2014). (3a') would typically be translated into English as (3b') but is translated as (3c') in this study for the sake of argument clarity.

- (3') (a) *c'est comme ça que le champagne par exemple est fait avec du Pinot Noir* (CRFP)
 (b) 'that's how champagne for instance is made with some Pinot Noir'
 (c) 'it's like that that champagne for instance is made with some Pinot Noir'

shown in Table 1 below, were identified in 12 conversations, which amounts to 2.7% of all clefts.

	Full clefts	Reduced clefts	Total
Singular <i>est</i>	382 (97.4%)	85 (96.6%)	467 (97.3%)
Plural <i>sont</i>	10 (2.6%)	3 (3.4%)	13 (2.7%)
Total	392	88	480

Table 1. Distribution of copulas in the singular and plural forms

In (5) above, the third-person plural form *sont* shows agreement with the third-person plural pronominal form in the clefted constituent. For Blanche-Benveniste (2002), the possibility for the copula to take the plural form is purely mechanic and is mainly observed in clefts with clefted subject, e.g. *eux* in (5), or non-prepositional complement, or in cases of hypercorrection. The imbalanced distribution between *est* and *sont* in the CRFP supports Lehmann's (2008) argument that clefts are showing a loss of agreement which is a sign that the modern *c'est*-cleft is undergoing grammaticalisation where grammaticalisation pertains to the reduction of members of a paradigm and of syntactic choices in a specific syntactic environment (De Mulder and Lamiroy 2012).

Along with the fixation of non-number agreement, *c'est*-clefts in present-day French are also characterised by the virtual absence of variation in the tense of the copula. In Blanche-Benveniste's (2002: 91) view, the tense of the copula in *c'est*-clefts can always be 'neutralised' regardless of the tense of the verb in the cleft-relative clause. This is also reinforced by the existing restrictions on tense and mood in French (Dufter 2008). Thus, whether the verb in the cleft-relative clause is in present tense, as in (6), in 'passé composé', as in (7), 'imparfait', as in (8), or even future, as in (9), the copula in the matrix clause is most frequently in the present tense.

- (6) *c'est* à Pau qu'ils ont le grand dépôt pour les boissons (CRFP)
'it's in Pau that they have the big warehouse for drinks'
- (7) *c'est* ça qui qui a fait que bon maintenant on entend partout parler de ça quoi (CRFP)

'it's that that that made it so that well now we hear about it everywhere'

- (8) *c'est* pour ça qu'il y avait pas de production (CRFP)
'it's for that that there was no production'
- (9) *c'est* pas demain la veille qu'il y aura une culture biologique comme *c'est* à la mode maintenant de la vigne (CRFP)
'the day hasn't come where there will biological crop growing like it's on-trend now for grapevine'

This invariance is verified for all but 22 occurrences in the CRFP, as Table 2 below shows.

	Full clefts	Reduced clefts	Total
Present <i>est/sont</i>	377 (96.2%)	81 (92%)	458 (95.4%)
Past <i>était</i>	15 (3.8%)	7 (8%)	22 (4.6%)
Total	392	88	480

Table 2. Distribution of copulas in the present and past tenses

Among the 22 occurrences, 12 of them introduce a habitual action in the past with the 'imparfait' tense, e.g. (10), 6 an action that took place in the past with the 'passé composé', e.g. (11), 3 an action that occurred before another one in the past with the 'plus-que-parfait', e.g. (12), and 1 wish expressed with the subjunctive, e.g. (13). In all of these instances, the specification relation established between value and variable is set in the past which, beyond the temporal location of the situation referred to in the cleft-relative clause, may be the main motivator for the past tense of the copula.

- (10) le premier endroit où j'ai travaillé *c'était* des poubelles entières qui partaient à la poubelle de fleurs (CRFP)
'the first place where I worked it was whole bins that would go into the flower bin'
- (11) *c'était* en fait un sénégalais qui nous l'avait pris (CRFP)
'it was actually a Senegalese that had taken it from us'

- (12) il y a deux trois mois *c'était* avec beaucoup de plaisir qu' elle est revenue (CRFP)
'two months ago it was with great pleasure that she came back'
- (13) *c'était* là tout près que ça m'intéresserait (CRFP)
'it was there really close that I would be interested'

While removing the past tense in (10)–(13) leads to the removal of the past location in which the speaker situates the specification relation, it does not make the utterance ungrammatical or unnatural. It appears, then, that the use of the past tense for the copula is not strictly required for *c'est*-clefts with past tense in the verb of their cleft-relative clause, which explains its low number of instances in the CRFP. The two options, present and non-present, can be viewed as meaningful options which signal the temporal location of the specification (Davidse et al. 2022). In the unmarked choice, i.e. present, the relation between value and variable is established in the here and now. This is particularly frequent in spontaneous dialogue where clefts are used to express *hic et nunc* construed specificational relations. However, a speaker can also construe him or herself as involved in a specificational relation established in the past like in (10)–(13).

Thus, as far as the morphology of the copula is concerned, my results are consistent with the reports on the growing invariance noted by some authors (Hatcher 1948; Dufter 2008). The reduction of agreement choices supports the view that *c'est*-clefts are progressively grammaticalising over time with the reduction of tense choices also manifesting a certain fixation.

4.1.2 Paradigm of relative markers

Along with person and number agreement in the copula *être*, the third factor illustrating the gradual 'figement' of the French *c'est*-cleft relates to the rigidity in the types of relative marker accepted as connectors between the matrix clause and cleft-relative clause (Muller 2003; Dufter 2015), which constitutes another reduction of syntactic choices. On this, Dufter (2015) notes that Modern French clefts tend to only allow for *que* or *qui* relativisers, but with a few exceptions

brought forth by his corpus data. This is largely corroborated by the CRFP data as shown in Table 3 below.

Relative marker	n (%)
que	180 (45.9%)
qui	202 (51.5%)
où	8 (2%)
dont	1 (0.3%)
∅	1 (0.3%)
Total	392

Table 3. Distribution of types of relative markers

Although *c'est-clefts* predominantly use *qui* - for clefted subjects - and *que* - for clefted adjuncts, conjuncts, direct objects and indirect objects - a small number of occurrences (2.6%) rely on the use of either *où*, *dont* or zero relativiser \emptyset . Just like the invariance in person and tense, the strong preference for *que/qui* hints at an acute rigidity in the morphosyntax of *c'est-clefts*.

From a diachronic perspective, Muller (2002; 2003) discusses the variation in the relativiser by distinguishing between three types of clefts with clefted PPs according to the location of the prepositional marking. In the 'old' type, which Muller (2002; 2003) argues is not productive in Modern French anymore, prepositional marking is coded in the relative marker as in (14a). (14b) illustrates the second type, the 'redundant' type, in which the functional marking is coded in both the clefted constituent and the relative marker. Finally, in the 'modern type', which Muller (2002; 2003) describes as the most commonly used type in Contemporary French, the oblique relation between the clefted constituent and the cleft-relative clause is directly coded in the clefted constituent, as shown in (14c).

- (14) (a) C'est ma mère **à qui** que tu as parlé.
 (b) C'est à ma mère **à qui** tu as parlé.
 (c) C'est à ma mère **que** tu as parlé. (Muller 2003 : 140, my coding)
 'It's my mum whom you talked to'

Upon applying Muller's (2002; 2003) taxonomy to the CRFP dataset, only two of the three types reveal themselves to be productive. Out of 392 full clefts, 384 are of the modern type and 8 of the redundant type. Occurrences (15a) and (16a) illustrating the 'modern' type are therefore more frequently used than their 'old' counterparts (15b) and (16b).

- (15) (a) *c'est chez lui **que** je suis partie en Mauritanie cet été (CRFP)*
 (b) *c'est lui **chez** **qui** je suis partie en Mauritanie cet été*
 'it's at his place that I went to Mauritania this summer'
- (16) (a) *c'est sur sa réussite **qu'**on juge s'il est capable de franchir un pas supérieur (CRFP)*
 (b) *c'est sa réussite **sur laquelle** on juge s'il est capable de franchir un pas supérieur (CRFP)*
 'it's on his success that we judge whether he can move up'

Within the redundant type, all 8 instances correspond to *c'est là que*-clefts, a subtype of clefts with a prepositional clefted constituent singled out by Roubaud and Sabio (2015) and which will be investigated in more detail in Section 4.4.3. (17a) exemplifies the redundant clefts of the CRFP whose relative marker is always *où*.

- (17) (a) *c'est là **où** on se rend compte euh si le travail était bien fait avant ou pas (CRFP)*
 (b) *c'est là **qu'**on se rend compte euh si le travail était bien fait avant ou pas*
 'it's then that we realised uh whether work was well done before or not'
- (18) *c'est là **qu'**on peut s'apercevoir qu'on peut faire des économies (CRFP)*
 'it's then that we can see that we can save money'

A non-redundant equivalent of (17a) would be (17b) where the cleft-relative clause is introduced by *que* instead of *où*. Strikingly, in addition to these 8 *c'est là que*-clefts with *où* relativiser, 28 additional *c'est là que*-clefts with *que* relativiser, e.g.

(18), are found in the CRFP, thus showing a preference for *que* and hence a tendency to avoid the doubling of prepositions. This co-existence of both relative markers was already highlighted by Roubaud and Sabio (2015) who note that the use of *c'est là où* is typically excluded from formal and normative registers.

I have so far discussed three morphosyntactic aspects of the *c'est*-cleft, i.e. person and tense agreement in the copula and type of relative marker. While the three aspects have been argued to illustrate the increasing 'figement' of the *c'est*-cleft, the reduction in person agreement has further been claimed to signal the ongoing grammaticalisation of the construction. Despite corroborating these two arguments for the most part, my analysis also shows that disappearing variants, though restricted in their use, are still found in present-day French and for some, like tense, still represent meaningful grammatical options.

4.1.3 Syntactic category of the clefted constituent

Let us now turn to the form and function of the clefted constituent of *c'est*-clefts. First, with regard to their syntactic form, *c'est*-clefts have been described as allowing a high categorial freedom with the most frequent categories being NPs, PPs, AdvPs and the most marginal ones being AdjP and VPs (Katz 1997, 2000b; Doetjes et al. 2004; Dufter 2009a; Rouquier 2018). The same distribution is found in the CRFP dataset as shown in Table 4.

	Full clefts	Reduced clefts	Total
Noun phrase	221 (56.4%)	40 (45.4%)	259 (54%)
Prepositional phrase	100 (25.5%)	22 (25%)	121 (25.2%)
Adverbial phrase	66 (16.8%)	—	68 (14.4%)
Finite clause	3 (0.8%)	22 (25%)	25 (5.2%)
Adjectival phrase	—	—	—
Non-finite clause	2 (0.5%)	4 (4.6%)	6 (1.2%)
Total	392 (100%)	88 (100%)	480 (100%)

Table 4. Distribution of syntactic categories of the clefted constituent in the CRFP

The range of syntactic categories accepted by the clefts in the CRFP include, in order of frequency, NPs, PPs, AdvPs, and finite and non-finite clauses. (19)–(23) exemplify each of the five categories productive in the CRFP. No AdjP were found in the dataset. Moreover, because inferential clefts with \emptyset clefted constituent, e.g. (24), were not included in the query of the corpus, their use is not investigated in this particular study.⁵³

- (19) *c'est toujours la la commission qui statue jamais nous* (CRFP)
 'it's always **the the commission** that rule never us'
- (20) *c'est à cet âge-là qu'il faut refaire les mentalités* (CRFP)
 'it's **at this age** that we need to change ways of thinking'
- (21) *c'est là où justement internet peut nous sauver* (CRFP)
 'it's **then** that internet can actually save us'
- (22) *c'est pas parce qu'il y a une histoire de drogue qu'on doit la laisser mourir sur un trottoir* (CRFP)
 'it's not **because there's some drugs involved** that we should leave her to die on a the sidewalk'
- (23) *ce n'est pas en la rendant obligatoire que nous allons réussir à la favoriser* (CRFP)
 'it's not **by making it compulsory** that we'll be able to favour it'
- (24) *c'est pas que je voudrais pas mais là je suis sûre de je suis sûre de bien souvent de me faire avoir* (CRFP)
 'it's not that I wouldn't want to but I'm sure that I'm sure that I often get played'

In an effort to make this account as exhaustive as possible, a few notes should be made. First, clefted NPs in *c'est*-clefts can be of three types: common noun (53%), proper noun (4%) or pronoun (43%). The fact that pronouns make up 43% of all clefted NPs, that is, almost half of them, seems to be in line with Dufter's

⁵³ For an overview of the use of inferential *c'est*-clefts, see Lambrecht (2001), De Stefani (2008) and Dufter (2008). It is worth noting that some authors (e.g. Davidse and Van Linden 2020) do not view these constructions as clefts but as extraposition constructions.

(2009a, 2015) findings on the diachronic increase use of clefts introducing anaphoric material in their clefted constituent. A large proportion of pronouns such as clefted pronouns can be situationally deictic and introduce references to speech participants as in (25) where *nous* refers to the speaker and their partner, or textually anaphoric or cataphoric as in (26) where the clefted NP refers to *les/la langue* given in the preceding and following co-text.

- (25) *c'est nous* qui faisons les enfants elle les garde (CRFP)
 'it's us that make the children she babysits them'
- (26) A: et ça alors justement le problème ça va être les langues
 B: ouais les langues ouais ouais moi c'est ça moi qui me dérange c'est la langue (CRFP)
 'A: and this well precisely the problem is going to be languages
 B: yeah languages yeah yeah it's **that** that bothers me it's language'

When added together, all types of NPs represent the most frequent grammatical category, which contrasts with De Stefani's (2008) corpus study relying on spoken data which yielded only a handful of nominal clefted constituents.

Second, adverbial phrases, which are only found in full clefts, can be divided into four discrete categories according to the adverb in the postcopular position. These include *c'est là que*-clefts (n= 35), *c'est ainsi que*-clefts (n= 5), *c'est comme ça que*-clefts (n= 18) and clefts with other sporadically-used adverbs. The first three constructions are particularly of interest when investigating the information structure of *c'est*-clefts and have, as a result, received attention in the literature (see Blanche-Benveniste 2006; Sabio and Benzitoun 2013; Roubaud and Sabio 2015, 2018; Lahousse and Lamiroy 2017). I will discuss these three subtypes of *c'est*-clefts in Section 4.4.3.

Finally, the comparison of full and reduced clefts shows that while the distribution of NPs and PPs is similar for the two forms, reduced clefts in the CRFP do not allow for AdvP in the postcopular position. Reduced clefts, however, more readily accept finite clauses which amount to 25% of all clefted constituents in reduced clefts but only 0.8% of that of full clefts. The higher proportion of finite

clauses can be explained by the use of ‘pre-modified *it/c'est*-clefts’ constructions (Declerck and Seki 1990) with a causal value as in (27).

- (27) *si je vous raconte cette histoire c'est parce que ça m'a touché* [~~que je vous raconte cette histoire~~] (CRFP)
'if I'm telling you this story it's **because it touched me** [~~that I'm telling you this story~~']

These constructions, whose grammatical and semantic status I discuss in Section 6.2.3, contain a pre-form, e.g. *si je vous raconte cette histoire*, introducing the variable, e.g. *je vous raconte cette histoire parce que x*, whose missing argument, e.g. *x = parce que ça m'a touché*, is shared in the reduced cleft that follows.

To conclude, the corpus findings on the syntactic category of the clefted constituent strongly corroborate the categorial freedom generally attributed to *c'est*-clefts. Although a tendency to favour NPs can be observed, *c'est*-clefts may select a wide array of syntactic classes.

4.1.4 Grammatical function of the clefted constituent

To further characterise the morphosyntax of *c'est*-clefts, let me now turn to the grammatical function taken on by the clefted constituent as antecedent of the cleft-relative clause. Table 5 below presents the frequency of each grammatical function found in the CRFP.

	Full clefts	Reduced clefts	Total
Subject	202 (51.5%)	27 (30.7%)	228 (47.5%)
Direct object	17 (4.4%)	9 (10.2%)	27 (5.6%)
Indirect Object	4 (1%)	—	4 (0.8%)
Adjunct	108 (27.5%)	50 (56.8%)	158 (33%)
Conjunct	61 (15.6%)	—	61 (12.7%)
Complement of a preposition	—	2 (2.3%)	2 (0.4%)
Subject complement	—	—	—
Total	392 (100%)	88 (100%)	480 (100%)

Table 5. Distribution of the grammatical functions of the clefted constituent

In total, six functions are attested: subject, direct object, indirect object, adjunct, conjunct and complement of a preposition. Subjects and adjuncts are the most frequent ones with 47.5% and 33% of all occurrences respectively, followed by conjuncts (12.7%), direct objects (5.6%), indirect objects (0.8%) and complements of a preposition (0.4%). The six functions are illustrated in order frequency in (28)–(33) below.

- (28) c'est **eux** qui nous font parvenir le fromage (CRFP)
 'it's **them** who send us the cheese'
- (29) c'est **pour ça** que le week-end je suis très rarement sur Troyes (CRFP)
 'it's **for that reason** that I'm rarely in Troyes during the weekend'
- (30) c'est **comme ça** que je suis venu à Amiens en dix-neuf cent vingt-huit après mon service militaire (CRFP)
 'it's **how** I came to Amiens in nineteen twenty-eight after my military service'
- (31) c'est tout le temps **le même système de nœuds** qu'on fait (CRFP)
 'it's always **the same knotting system** that we do'
- (32) c'est euh **cette journée et ce qui l'a précédée le voyage de noce qui a suivi** dont je dont je vais parler aujourd'hui (CRFP)
 'it's uh that **day and what preceded it the honeymoon that followed** that I'm going to talk about today'
- (33) et où ça joue énormément c'est **les réactions de l' enfant** (CRFP)

‘and where it plays a big role it’s **the child’s reactions**’

The prevalence of subjects and adjuncts observed in my data is consistent with that observed by Doetjes et al. (2004) and Carter Thomas (2009) in Contemporary French. It also appears to illustrate the diachronic evolution underlined by Dufter (2008) whereby subjects and adjuncts have asserted their dominance over the period from Old French to 20th-century French at the expense of objects. These results also contribute to further cast doubt on the assumption that the primary motivation behind the use of *c'est*-clefts relates to the constraints imposed on the focalisation of pre-verbal subjects (Dufter 2008, 2009a).

4.1.5 Modality, negation and exhaustivity markers in *c'est*-clefts

C'est-clefts, much like *it*-clefts, establish a specification relation between the clefted constituent, i.e. the value *x*, and the presupposed open proposition in the cleft-relative clause, i.e. the variable. Thus, (34) presupposes that the Roman linguistic atlas was born *somewhere*. The matrix clause fills the semantic gap in the presupposition by specifying the corresponding value as being *au centre de dialectologie*.

- (34) (a) l'atlas linguistique roman euh c'est euh c'est à partir du centre de dialectologie **c'est au centre de dialectologie qu'est né ce projet** (CRFP)
 ‘the Roman linguistic atlas uh it’s uh it’s from the dialectology centre **it’s at the dialectology centre that this project was born**’
 (b) value: au centre de dialectologie
 (c) variable: ce projet est né à *x*

The value in the matrix clause may be accompanied by a wide range of modality markers expressing negation, e.g. *pas*, *même pas*, restriction, e.g. *ne...que*, or epistemic indications, e.g. *presque*, *vraiment* (Blanche-Benveniste 1990: 60), which, in some cases, may interact with the specification meaning coded by the cleft syntax. Let us compare, for instance, (35) and (36).

- (35) (a) *c'est peut-être ça qui manque dans le sport maintenant* (CRFP)
'it may be this that's missing in sports now'
(b) value: *ça*
(c) variable: *x qui manque dans le sport maintenant*
- (36) (a) *c'est moi qui ai peut-être plus précipité les choses* (CRFP)
'it's me who may have rushed things more'
(b) value: *moi*
(c) variable: *x qui a peut-être précipité les choses*

In (35), the scope of the modal adverb *peut-être* bears on the content of the matrix clause it is part of and hence also on the postcopular complement *ça*. What is conveyed here is the extent to which the value introduced in the clefted constituent is the accurate value for the variable carried by the cleft-relative clause. In other words, through the use of modality, the speaker is expressing his/her stance on the relevance of the specification relation. By contrast, the same modal expression used in the cleft-relative clause of (36) bears on the content of the cleft-relative clause and is therefore part of the open proposition designated by the cleft-relative clause. As a result, the specificational meaning and the speaker's positioning towards it remain unchanged. Unlike in (35), the relation between the value and the variable is still asserted while the information in the open proposition, i.e. that *someone* rushed things, is what is nuanced by the modal adverb. This difference is in fact easily detected when formulating the variable of each cleft, (35c) and (36c). The modal adverb *peut-être* only appears in the variable of (36a).

The same contrast can be drawn for clefts with a negation marker, which are frequently used in informal spoken French (De Stefani 2008). I illustrate this with (37) and (38) below.

- (37) (a) *un soixante-quatre mal garé un soixante-cinq à côté c'est pas le soixante-cinq qui aura la contravention* (CRFP)
'a sixty-four badly parked a sixty-five next to it it's not the sixty-five that's going to be fined'
(b) value: *le soixante-cinq*

- (c) variable: x qui aura la contravention
- (38) (a) *c' est les mêmes problèmes pour l'informaticien c'est pour ça que on sait **jamais trop** (CRFP)*
 'it's the same issues for the IT engineer it's for that reason that we **never** really know'
- (b) value: ça
- (c) variable: on ne sait jamais trop parce que x

In (37a), the negation is marked before the clefted constituent with the marker *pas*. Similarly to modality, when negation bears on the content of the matrix clause, it also scopes over the semantic relation established between the value and the variable. In this case, the veracity of the relation is not simply put into question but fully rejected. The meaning conveyed can be paraphrased as 'there will be a car who will receive a fine but it will *not* be the one from the 65th (Haute-Pyrénées) department'. Thus, the insertion of a negation marker in the matrix clause of a cleft does not affect the encoding of specificational meaning. Instead, what is encoded by the insertion of the negation marker is that the correct value is not the one introduced in the cleft but is another element of the set of potential values. For De Stefani (2008), this is one of the most obvious manifestations of the discourse function of contrast associated with *c'est-clefts*. In Scappini's (2013) taxonomy, an example such as (37a) instantiates a 'virtual' contrast in which the contrastive reading is made explicit through the negation but the paradigm itself is not explicitly expressed. By contrast, when the negation is encoded in the cleft-relative clause, as in (38a), it does not add any contrastive reading to the clefted constituent. Instead, it is included in the presupposed open proposition and hence the variable for which the value is specified, as shown in (38c). The specification relation is therefore not impacted. On the encoding of negation and contrast in *c'est-clefts*, Rouquier (2007) underlines the tendency for *c'est-clefts* to first use negative polarity and then positive polarity by means of a corrective cleft. In other words, when a value is discarded in a negative cleft, the appropriate value is typically suggested in the immediate rightward co-text with a second cleft. In (39), for instance, the missing argument explaining why the speaker is interested in tarot is shown not to be because s/he wants to draw but because of the symbolic initiation

behind tarot. (39) thus follows the *not x but x* typical sequence identified by Rouquier (2007).

- (39) si je m'intéresse au tarot **c'est pas parce que je te vais tirer les cartes** [~~que je m'intéresse au tarot~~] j'en ai rien à foutre de tirer les cartes mais c'est pour la l'initiation symbolique il y a derrière le tarot (CRFP)
 'if I'm interested in tarot **it's not because I'm going to draw cards** [~~that I'm interested in tarot~~] I don't give a fuck about drawing cards but it's for the symbolic initiation that's behind tarot'

Associated with the contrastive aspect of *c'est-clefts* is the conversational implicature of exhaustivity triggered by the cleft syntax. Like contrast, the exhaustive reading can be enhanced with the use of markers which include adverbs such as *ne...que*, *uniquement*, *seulement* or *juste*. In the CRFP, only instances of the first two are found which number 4. With both adverbs, *ne...que* in (40), *uniquement* in (41), the otherwise implicit readings 'it is shampoos and colouring, and *only* shampoos and colouring, that we do' and 'it is to have a more modern palace and *only* to have a more modern palace that we replaced it' are brought forward and made unequivocal.

- (40) ouais regarder puisque la plupart en apprentissage c'est **que** shampoos couleurs [~~qu'on fait~~] après petit à petit ben on fait d'autres choses coupes euh des permanentes (CRFP)
 'yeah watching since most of the work/study programme it's **only** shampooing dyeing [~~that we do~~] then gradually well we do other things haircuts uh perms'
- (41) A: et on l'a remplacé au dix-huitième siècle par celui-ci
 B: et c'était **uniquement** dans le but d'avoir un palais plus moderne [~~que vous l'avez remplacé~~]
 A: alors oui oui (CRFP)
 'A: and we replaced it in the eighteenth century with this one
 B: and it was **only** with the goal of having a more modern palace [~~that you replaced it~~]

A: well yes yes'

Exhaustivity can be reinforced with exhaustive particles, but it can also be cancelled (Horn 1981; Declerck 1988; De Cesare and Garassino 2015). (42) demonstrates this with the value *par contact* being introduced as only one of the appropriate values for the variable *ça traite par x*.

- (42) *c'est systémique ça pénètre dans la sève et ça ça traite de l'intérieur si vous voulez c'est pas uniquement par contact [que ça traite] avant nous quand on traitait par contact avec essentiellement le cuivre (CRFP)*
 'it's systemic it penetrates into the sap and it it treats from the inside if you will it's **not just** by contact [~~that it is treated~~] before when we treated by contact with essentially copper'

While modality, negation and exhaustivity can be encoded and/or emphasised in *c'est-clefts* through lexical means, certain subtypes of *c'est-clefts* bar the insertion of paradigmatic adverbs in the matrix clause (Roubaud and Sabio 2015; Lahousse and Lamiroy 2017; Roubaud and Sabio 2018). This is the case of subclasses of *c'est là que-clefts*, *c'est comme ça que-clefts* and *c'est ainsi que-clefts* whose clefted adverbs have lost their original spatio-temporal and manner meanings and function instead as conjunctive adverbials. Because the clefted adverbs are not governed by the verb, the insertion of items such as *uniquement* or *seulement* is not possible, as shown in (43b).

- (43) (a) *j'avais perdu mon père à douze ans et je ne connaissais pas tellement la fabrication ma mère a fait tout ce qu'elle a pu mais eh eh elle était pas du métier bon alors j'ai cherché à ayant deux frè- deux frères et une sœur à leur laisser la place pour t avoir une profession et moi me perfectionner dans mon métier et **c'est comme ça que je suis venu à Amiens en dix-neuf cent vingt-huit après mon service militaire (CRFP)***

'I had lost my father at twelve years old and I didn't really know manufacturing my mother did everything she could but eh eh she wasn't in the field so well then I tried to having two bro~ two brothers and a sister give them the opportunity to t have a job and for me to get better in my job and **it's how I came to Amiens in nineteen twenty-eight after my military service**'

- (b) ?? (...) *c'est uniquement* comme ça que je suis venu à Amiens en dix-neuf cent vingt-huit après mon service militaire
'it's **only** how I came to Amiens in nineteen twenty-eight after my military service'

For Lahousse and Lamiory (2017), this is due to the very nature of paradigmatic adverbs whose purpose is to oppose one alternative to the rest of the other alternatives within the paradigm. As connectives, *là*, *comme ça* and *ainsi* do not allow for a contrastive reading.

I have shown in this section that *c'est*-clefts accept, for the most part, the insertion of modality and negation markers as well as exhaustive particles. When found in the matrix clause, these may interact with the general semantic and pragmatic properties of *c'est*-clefts by enhancing or cancelling their contrastive and/or exhaustive meanings. When located in the cleft-relative clause, their effect bears on the content of the cleft-relative clause alone.

4.2 The relational information structure of *c'est*-clefts

I now turn to the prosodically coded relational information structure of *c'est*-clefts. I first detail the prosodic patterns observed in the CRFP data and propose a prosody-based taxonomy of information foci in Section 4.2.1. I then discuss the effects that the interactional context may have on the prosodic realisations of *c'est*-clefts in Section 4.2.2. Finally I characterise the role of the initial accent within the intonation unit in Section 4.2.3.

4.2.1 Prosodic patterns

Before moving on to the discussion of focus-marking in *c'est*-clefts, I will first detail the prosodic patterns attested in the CRFP dataset. To start with, I assess the productivity of the types previously described in the literature on *c'est*-clefts. As noted in Section 1.2.2.2 of Chapter 1 on focus assignment in clefts, Rialland et al. (2002), Doetjes et al. (2004) and Mertens (2012) all distinguish two main prosodic articulations for clefts. In the first one, which can be summarised as (44), the clefted constituent bears a major terminal boundary such as L-L- or HL- at its right edge, making it prosodically focal. The cleft-relative clause, on the other hand, is realised as a string of deaccented syllables with an appendix contour l-l-. The two parts of the cleft are realised with the same movement but the cleft-relative clause displays a clear downstepping in pitch height and hence a compression of the pitch register. In the second type, (45), the clefted constituent is delineated by a major continuative tone HH at its right boundary, but the cleft-relative clause is not deaccented, as shown by the fact that it takes on a major prosodic boundary L-L- or HH. Rialland et al. (2002) and Doetjes et al. (2004) also recognise a third pattern for explicative reduced clefts shown in (46).⁵⁴ Because this pattern only concerns a very restricted subclass of reduced clefts, I do not treat it as a full-fledged type and therefore do not include it in the preliminary analysis of the CRFP data.

(44) Type 1: [c' est X]_{L-L-/HL-} [qui V]_{l-l-}

(45) Type 2: [c' est X]_{HH} [qui V]_{L-L-/HH} (Avanzi 2011: 115)

(46) Type 3: [(cest X)_h (quoi)]

Upon applying the binary taxonomy to the CRFP dataset, here is what can be observed. The distribution of the two types is shown in Table 6 below.

⁵⁴ Doetjes et al (2004: 540) illustrate their category of explicative reduced clefts with the following example (see Section 1.2.2.2):

(46') (Why are you worried?)
C'est le petit qui est tombé dans l'escalier.
'It's because the little one fell down the stairs.'

	Full clefts	Reduced clefts	Total
Type 1	22 (5.6%)	–	22 (4.6%)
Type 2	208 (53.1%)	–	208 (43.3%)
Other	148 (37.7%)	88 (100%)	236 (49.2%)
Unclear	14 (3.6%)	–	14 (2.9%)
Total	392 (100%)	88 (100%)	480 (100%)

Table 6. Distribution of prosodic patterns in *c'est*-clefts

The overall results show that only half of all clefts fall within one of the two types established by Rialland et al. (2002), Doetjes et al. (2004) and Mertens (2012). In detail, type 1 clefts represent 4.6% of all tokens while type 2 clefts amount to 43.3% of all *c'est*-clefts. The prevalence of type 2 clefts found in the CRFP dataset is consistent with Mertens's (2012: 8) results. This leaves 49.2% all occurrences which illustrate neither type 1 or 2. 2.9% of clefts are labelled 'unclear' as they could not be analysed either instrumentally or auditorily due to technical issues or overlaps. The fact that roughly half of all occurrences display a prosodic pattern other than the two accounted for in the literature is in line with Avanzi's (2011) results on a sample of 60 clefts. Besides the quantitative comparability, my findings also corroborate a number of observations made by Avanzi (2011) on the nature of prosodic contours and phrasing in *c'est*-clefts. First, type 2 clefts were found not to be strictly limited to the HH/L-L- realisation put forth by Mertens (2012). Instead, the cleft-relative clause was found to accept a wider variety of boundary tones including the terminal contour HL- and other continuative tones such as HL and LH. The second prosodic property found to be instantiated in the CRFP relates to prosodic segmentation, whereby the clefted constituent and cleft-relative clause need not necessarily be separated by a boundary tone. Thus, *c'est*-clefts may also be realised as a single intonation unit, which goes against Rialland et al.'s (2002), Doetjes et al.'s (2004) and Mertens's (2012) predictions. In order to account more accurately for the diverse prosodic patterns exhibited by *c'est*-clefts, I propose the following typology, in which I associate the different prosodic patterns with the types of focus they code. Following Di Cristo (1998, 2016), focus is analysed as broad when the nuclear accent occurs in its default position, i.e. at the end of the

intonation unit, and narrow when it is moved to an earlier position in the unit (see Section 1.1.3.3).

Type	Subtype	Typical realisation	Clefted constituent & CRC within same unit	Type of focus	Full or reduced
Type 1	–	[c' est X] _{L-L-/HL-} [qui V] _{L-L-}	yes	narrow	full
Type 2	2a	[c' est X] _{HH} [qui V] _{L-L-/HL-}	no	broad	full
	2b	[c' est X] _{HH} [qui V] _{HH}	no	broad	full
Type 3	3a	[c'est X qui V] _{L-L-/HL-}	yes	broad	full
	3b	[c'est X qui V] _{HH}	yes	broad	full
Type 4 interrogatives	4a	[c'est X qui V] _{H/H}	yes	broad	full
	4b	[c'est X] _{HH} [qui V] _{H/H}	no	broad	full
Type 5 reduced clefts	5a	[c'est X] _{L-L-/LH/HL}	–	narrow	reduced
	5b	[c'est X] _{HH}	–	broad	reduced
Rest	–		yes/no	narrow / broad	full

Table 7. Overview of prosodic patterns available for full and reduced clefts

The typology summarised in Table 7 consists of five types, four of which are further divided into two subtypes. The different patterns are differentiated according to three criteria: prosodic segmentation, type of focus exemplified, and form, e.g. full or reduced. A detailed description is provided for each type below.

In type 1 clefts, which correspond to Rialland et al.'s (2002) and Doetjes et al.'s (2004) focus-ground and Mertens's (2012) type I, the cleft forms a single intonation unit. The matrix clause ends with a terminal boundary tone such as L-L- or HL- while the cleft-relative clause represents the post-focal tail realised as an appendix. This is shown by the fact that it replicates the low tone with a compressed register and no major pitch variation. This type is illustrated in (47) in which the end of the clefted constituent *un solvant* is accompanied by a terminal boundary and a L-L- contour. The cleft-relative clause *qu'il y a dedans* takes the appendix contour l-l-.

- (47) [c'est un solvant]_{L-L-} [qu'il y a dedans]_{I-I-} (CRFP)
 'it's a solvent that's in it'

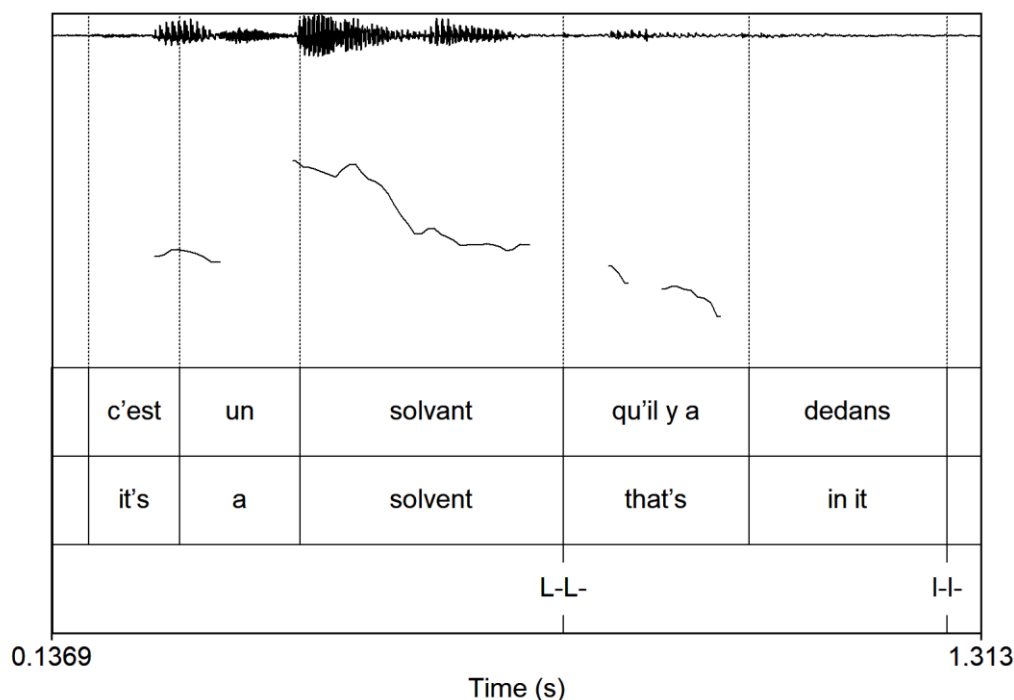


Figure 1. Prosodic realisation of (47)

Informationally, this translates into a focus/post-focus sequence in which the focus is borne by the clefted constituent and the post-focus tail subsumes the information carried in the whole cleft-relative clause.

The type 2 category encompasses all clefts which are realised with at least two intonation units and in which the clefted constituent and cleft-relative clause are separated by a non-terminal major prosodic boundary. Each of the components may be realised as more than one intonation unit, though this is more frequently the case for the cleft-relative clause. Type 2a and 2b are differentiated in terms of the nature of the construction-final boundary tone. The cleft-relative clause of type 2a clefts displays a terminal tone such as L-L- or, in some rarer cases HL-, while that of type 2b clefts is bounded by a major continuative tone HH. Both types have been recognised by Rialland et al. (2002) and Doetjes et al. (2004), but have remained merged into one category. The subdivision proposed in my typology finds its basis in the role of the L-L- and HL- tones, as opposed to the HH and LH tones. The L-L- and HL- tones are terminal tones which mark the end of the informational object

(Mertens 2008: 98). In addition to its demarcating role, the HL- tone also marks the constituent which carries it, when in non-final position, or the whole proposition as contrastive and informationally salient. By contrast, HH and LH are continuative tones which leave the information unit open and which indicate to the addressee that more information is to come. With regard to focus, both types of tones typically assign broad focus to the construction. Unlike in type 1 clefts, the clefted constituent is therefore not focal and the cleft-relative clause not backgrounded. In (48), for instance, there is a continuation rise HH occurring after the clefted constituent *pour ça* which is followed by a gradual fall in the cleft-relative clause reaching an infra-low frequency and the tone L-L-. (48) illustrates type 2a.

- (48) donc [c'est pour ça]_{HH} [que j'espère pouvoir refaire euh du cirque]_{L-L-}
 'so it's for that reason that I hope to be able to be a circus performer again'

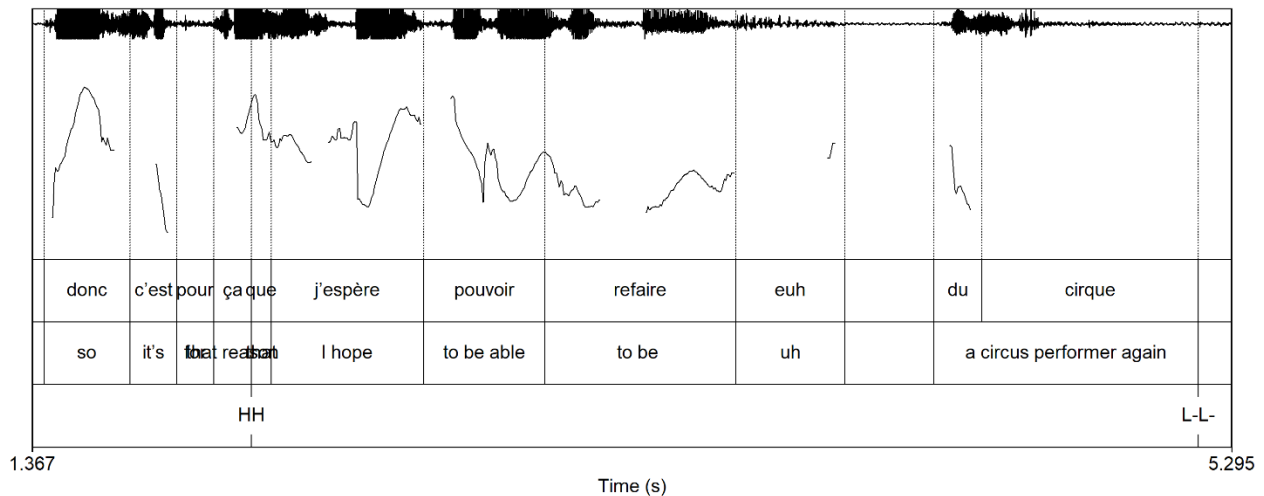


Figure 2. Prosodic realisation of (48)

An example of type 2b is given in (49) below. Here, the cleft is once again made up of two intonation units whose boundaries meet at the right edge of the clefted constituent *comme ça*. Unlike (48), both parts of the cleft are marked by a continuative boundary tone at their right edge. Because HH is not a tone triggering an illocutionary force, e.g. assertion, question, the fact that the cleft-relative clause exhibits the same rising pitch movement as that of the matrix clause does not entail intonational agreement. In declaratives, intonational agreement can only

take the form of a L-L-...l-l- sequence where L-L- marks the utterance as being assertive (Doetjes et al. 2004).

- (49) [c'est comme ça]_{HH} [que j'ai trouvé le moyen de venir]_{HH}
 'it's in that way that I found the way to come'

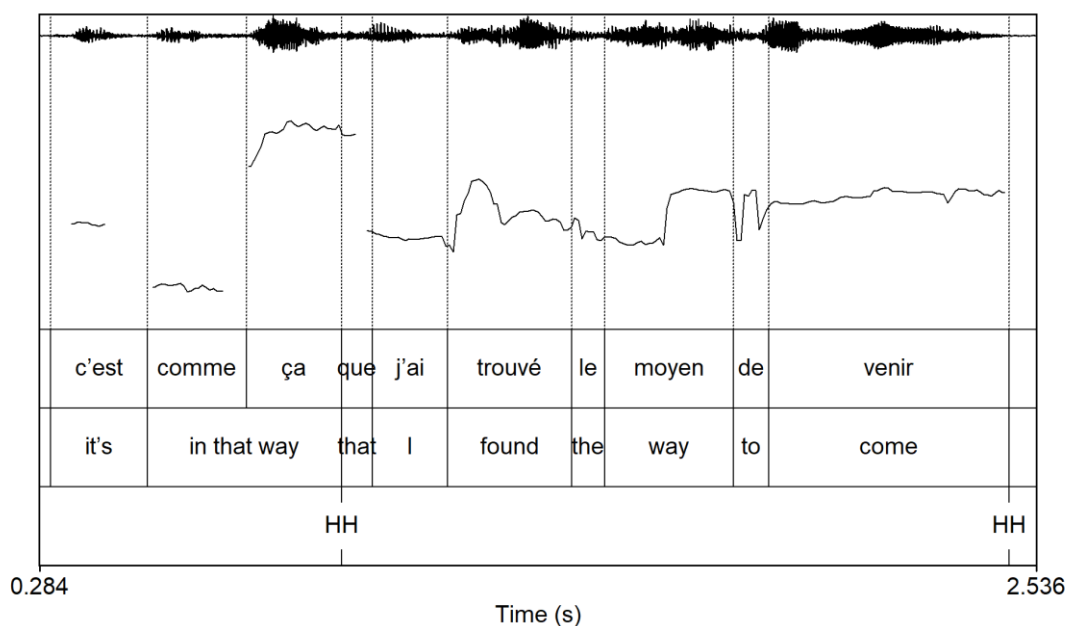


Figure 3. Prosodic realisation of (49)

Clefts of type 3 are uttered as a single intonation unit with no major boundary between the clefted constituent and the cleft-relative clause. This pattern is not predicted by Rialland et al.'s (2002), Doetjes et al.'s (2004) and Mertens's (2012) taxonomies but is acknowledged by Avanzi (2011). Like type 2, type 3 is further split into two categories. Clefts with a final terminal tone are subsumed under 3a and those with a continuative tone under 3b. An example of type 3a is given in (50). Here, the cleft only exhibits one boundary tone at the end of the cleft-relative clause with the tone HL-. For Mertens (2019), the HL- tone occurring in unit-final position encodes focalisation of the whole proposition and hence of the information conveyed in the entire unit. In the case of (50), the whole specification relation is taken to be focal.

- (50) [c'est pour ça que ça m'a beaucoup émue]_{HL-} (CRFP)
 'it's for that reason that it touched me a lot'

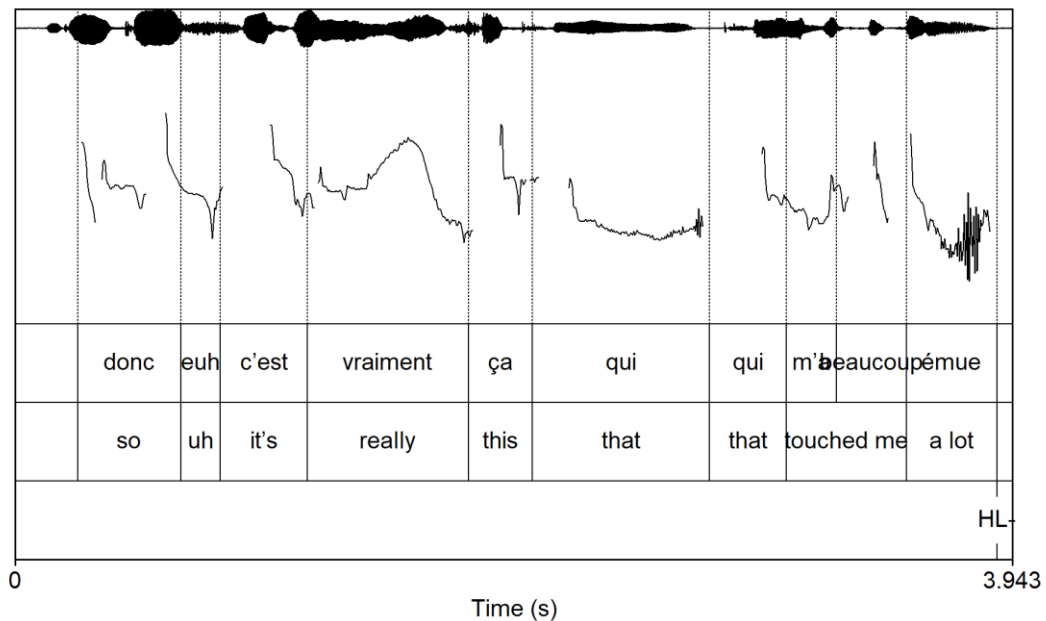


Figure 4. Prosodic realisation of (50)

Type 3b is illustrated in (51) in which the boundary of the intonation unit extends beyond the clefted constituent and is bounded by a continuative tone LH. The clefted constituent represents unstressed material leading up to the final tone. There is, in this case, a complete lack of focalisation of the clefted constituent, which is instead presented to the speaker as non-salient information.

- (51) [la première demi-heure]_{HH} [c'est vous qui mettez l'ambiance]_{LH} [et tout]_{h+h+}
 'the first half hour it's you who liven things up and all'

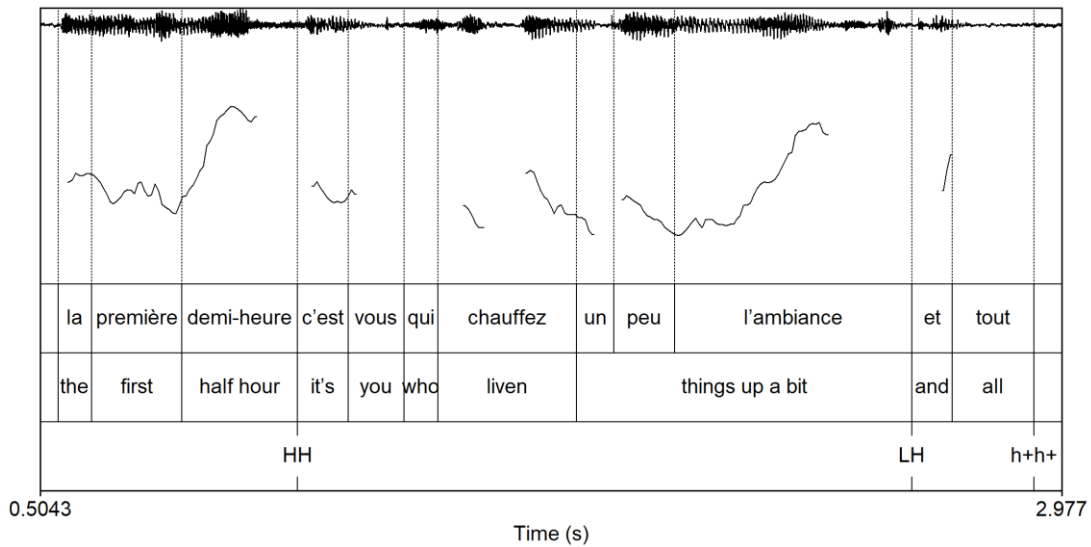


Figure 5. Prosodic realisation of (51)

Type 4 corresponds to interrogative clefts, which are set aside due to the difference in tonal encoding. Interrogatives are characterised in French by the H/H tone which ends in the infra-high section of F₀. Like declaratives, they may be realised as a single intonation unit (type 4a) or as multiple units (type 4b). The latter is instantiated in (52) in which the illocutionary force is encoded by the H/H tone on *mettez*. The interrogative H/H tone differs from the continuative HH one in the starting point of the rise as well as in the range of the rise which is typically greater for the former (Rossi et al. 1981). In the case of H/H, the penultimate syllable, e.g. **MET**tez, is realised with a high pitch and is then followed by a rise in pitch in the final syllable, e.g. met**TEZ**. By contrast, the rise occurring in an HH tone is not necessarily bound to the final syllable only and typically starts from a lower pitch height.

- (52) [c'est vous]_{HH} [qui les mettez]_{H/H}
'it's you who put them on'

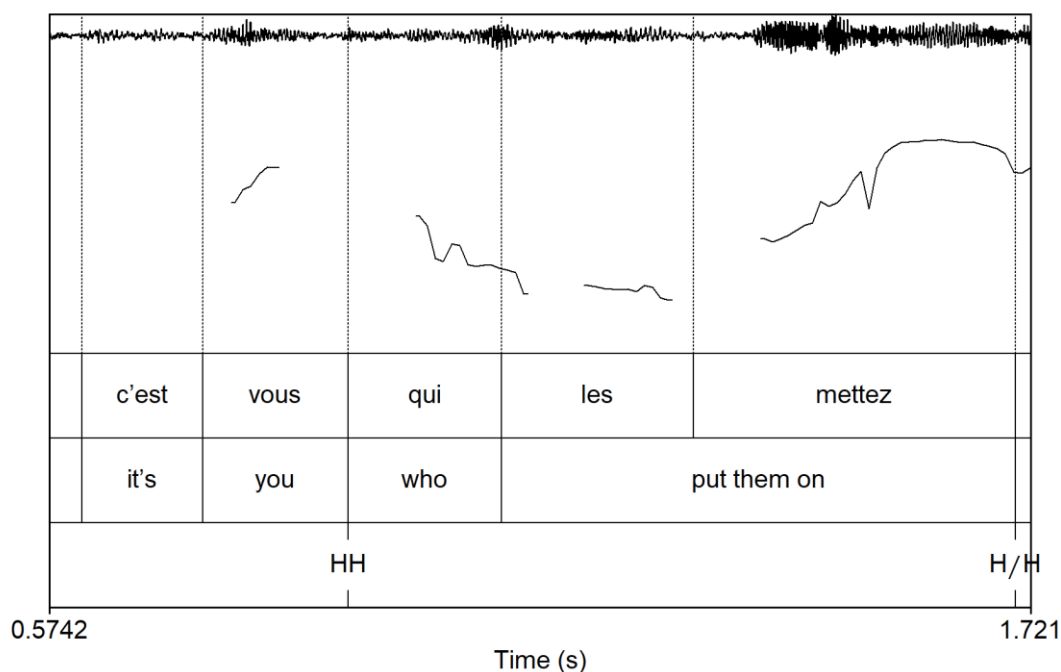


Figure 6. Prosodic realisation of (52)

Finally, type 5 includes all cases of reduced clefts. Like interrogatives, reduced clefts are placed within their own category as their prosodic articulation is significantly different due to the omission of the cleft-relative clause. The prosodic phrasing of reduced clefts is relatively fixed, with the majority forming a single intonation unit, but the nature of the unit-final tonal realisation varies. Once again, the clefted element may display either a terminal tone, i.e. type 5a, or a continuative tone, i.e. 5b, at its right edge. The first subtype, 5a, is found in (53), in which the string *que des fleurs de France* is focal as shown by the L-L- tone.

- (53) par exemple on a on a une c'est une concu~ une concurrente elle elle prend ses fleurs en Hollande [mais nous]_{HH} [c'est que des fleurs de France]_{L-L-}

'for instance we have we have a it's a comp~ a competitor she she buys her flowers in Holland but for us it's only flowers from France'

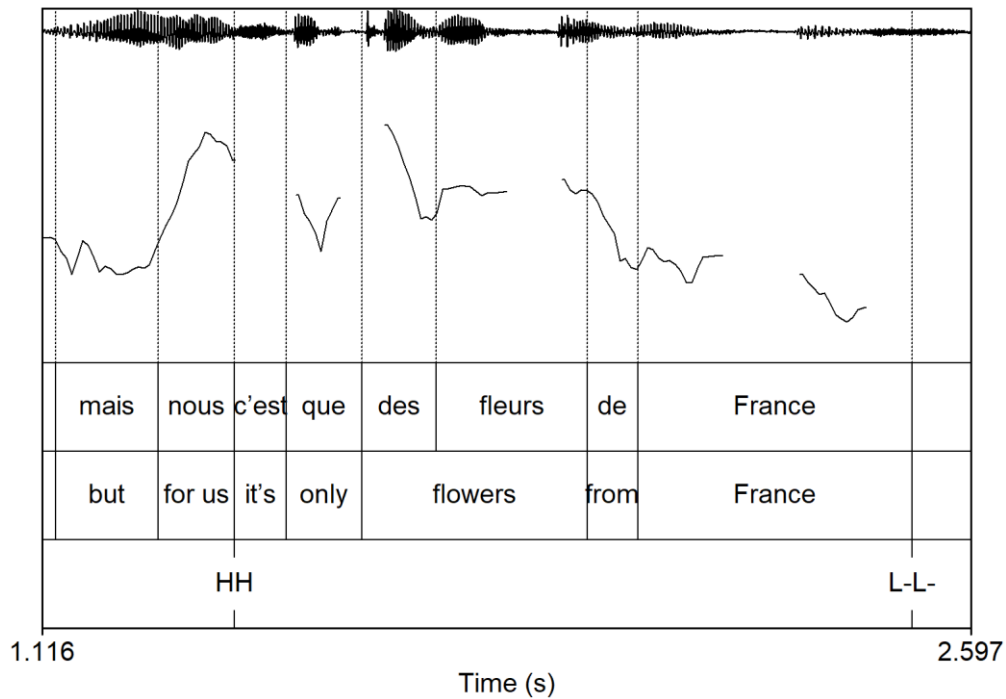


Figure 7. Prosodic realisation of (53)

In (54), which illustrates type 5b, the cleft is uttered as a two-unit structure. The continuative LH tone occurring on *Américain* indicates a request for confirmation on the hearer's part. It is then followed by another continuative tone HH at the right edge of *donc*.

- (54) [c'est par euh un Américain]_{LH} [donc]_{HH} (CRFP)
 'so it's by uh an American'

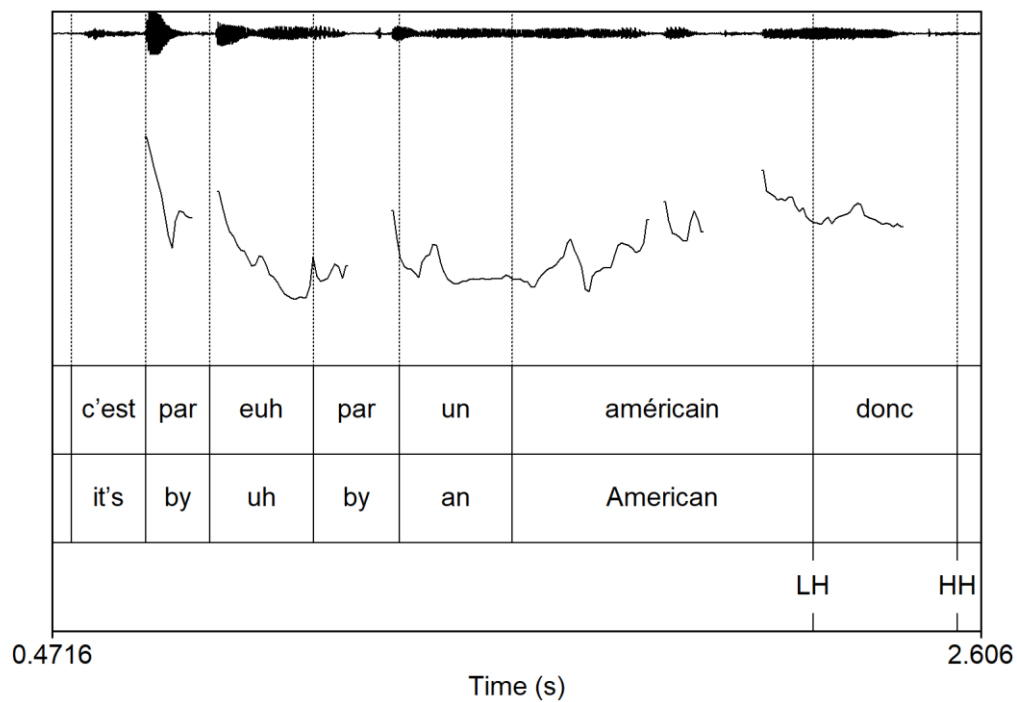


Figure 8. Prosodic realisation of (54)

By including interrogative and reduced clefts, the typology presented in Table 7 allows me to categorise a higher proportion of all the occurrences. The distribution of the different types is shown in Table 8 below.

Prosodic type	Subtype	n (%)	Total
Type 1	—	22 (4.6%)	22 (4.6%)
Type 2	2a	133 (27.7%)	211 (43.9%)
	2b	78 (16.2%)	
Type 3	3a	35 (7.3%)	114 (23.8%)
	3b	79 (16.5%)	
Type 4	4a	6 (1.2%)	16 (3.3%)
	4b	10 (2.1%)	
Type 5	5a	41 (8.5%)	88 (18.3%)
	5b	47 (9.8%)	
Uncategorised	—	15 (3.1%)	15 (3.1%)
Unclear	—	14 (3%)	14 (3%)
Total	—	480	480

Table 8. Distribution of prosodic types in the CRFP

Out of the 480 tokens making up the CRFP dataset, only 15 (3.1%) do not fit into any of the aforementioned categories and are therefore labelled ‘uncategorised’. Among them are full clefts whose clefted constituent is bounded by an L-L- or HL tone, which marks it as salient, but in which the cleft-relative clause is not deaccented, thus preventing the clefted constituent from being simultaneously informationally salient and in focus. This is the case in (55) whose clefted constituent displays a HL tone on its last word but whose cleft-relative clause then includes tonal variation with a continuation rise and a HL- tone in final position which marks the “speaker’s involvement” (Mertens 2019: 90).

- (55) [enfin c'est pas parce qu'il y a une histoire de drogue]_{HL} [qu'on doit la laisser mourir]_{HH} [sur un trottoir]_{HL-} (CRFP)

'well it's not because there are drugs involved that we should leave her to die on a sidewalk'

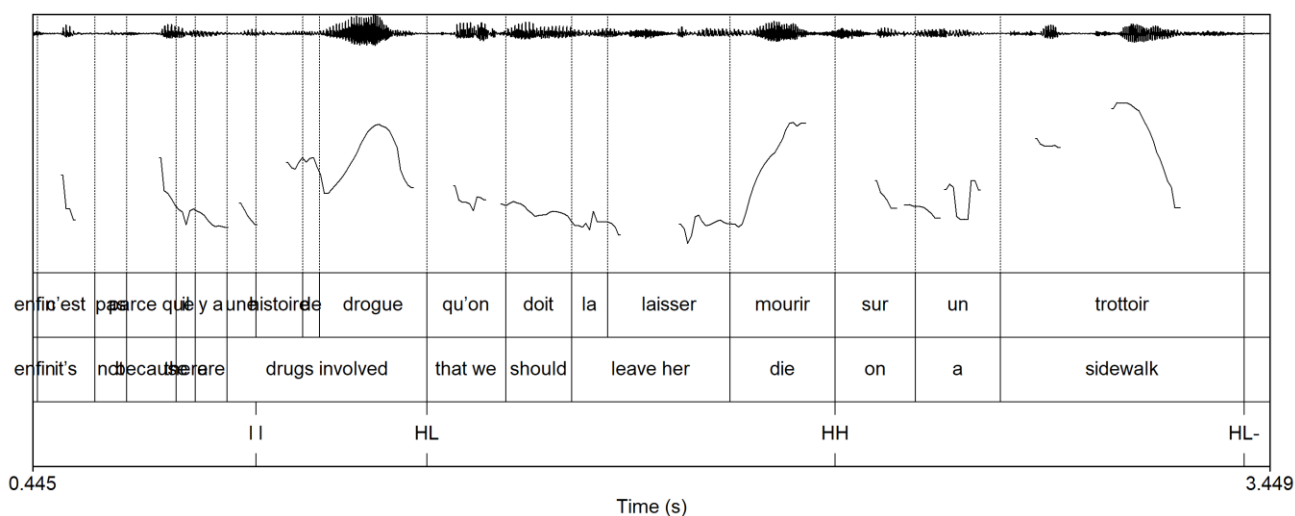


Figure 9. Prosodic realisation of (55)

Uncategorised clefts also include a few which exhibit distinctive prosodic coding mainly due to the interactional setting in which they are used. I detail these in the following section. That the uncategorised occurrences represents only 3.1% of all data supports the viability and applicability of the proposed taxonomy. The overall distribution of the different prosodic patterns points to a preference for types 2 and 3 for full clefts, which make up 43.9% and 23.7% of the data respectively. *C'est*-clefts are therefore most frequently realised as broad focus structures in which the clefted constituent and cleft-relative clause may belong to the same intonation unit or be separated by a major prosodic boundary that is continuative and not final. This prevalence of broad focus realisations had already been observed by Mertens (2012). By contrast, type 1 clefts, which exemplify the pattern posited as the prototypical realisation of clefts by Lambrecht (2001), is only found in 4.6% of occurrences in the CRFP. Interrogatives, with their prosodic range, likewise represent a minority of cases numbering 3.3%. In contrast with Rilland et al.'s (2002), Doetjes et al.'s (2004) and Mertens's (2012) typologies, the taxonomy presented in Table 7 along with its quantification in the CRFP in Table 8 reveal a wider variation than previously assumed. In doing so, it also counters the expectation created by accounts such as Lambrecht's (2001) according to which *c'est*-clefts are said to typically encode focus on the clefted constituent.

4.2.2 The influence of interactional aspects on the prosody of *c'est*-clefts

As De Stefani (2008: 711) notes, *c'est*-clefts used in interactional contexts not only contribute to the organisation of information within the proposition but also typically participate in the structuring of discourse.⁵⁵ This is directly reflected in the prosody of *c'est*-clefts in a number of ways. The first prosodic manifestation of the speaker attending to interactional needs relates to the relative prevalence of the HH tone in the final position of the cleft-relative clause of clefts. This is especially true for types 2b, 3b and 5b whose very rationale as subcategories is the presence of the continuation rise. In Mertens's (2006) typology of contours, the HH tone belongs to the class of major prosodic boundaries and is assigned the function of intermediate continuation. Information-wise, a continuation rise HH signals that the information unit is not yet complete and that more information is going to be added by the speaker. From an interactional perspective, this can be regarded as a strategy for the speaker to let the addressee know that s/he intends to keep the floor beyond the information that has just been shared, for instance in cases in which the informational object is complete but not the speaker's turn. This idea is especially prevalent in Conversation Analysis (CA), in which speech is viewed as being built incrementally, in collaboration, and as relying on interlocutors' knowledge and understanding of linguistic structures (Auer 2005; Couper-Kuhlen and Ono 2007). Within this approach, continuation rises can be taken to indicate the absence of a transition relevance place, which means that the addressee is *not* invited to take over (Couper-Kuhlen and Ono 2007). This was already noted by Di Cristo (2016: 227-228) who shows that continuative tones typically appear in moments where a transition relevance place would be expected and where negotiation between speaker and hearer may take place. In (56), for instance, the cleft forms two units both delineated by a continuation rise. In both cases, the boundary tone indicates that the information unit is left open. In the case of the clefted constituent, additional information is immediately added by the speaker. The cleft-relative clause, on the other hand, is followed by the conjunction *et* and

⁵⁵ See Section 4.5 for the detailed functional interactional description of *c'est*-clefts.

a pause. Despite this, the combination of the rising tone HH and the additive conjunction act as a warning that the conversational turn is not complete yet and the conversational floor is hence not yet released. The speaker then goes on to further explain why still life paintings are the ones she tends to focus on.

- (56) A: et c'est les mêmes genres de de motifs sur les toiles ou
 B: c'est euh pour pas lasser mon mon éventuel public quand je fais quelques quelques petites expos euh je je je je n'aimerais pas entendre oh la la que que des natures mortes c'est lassant donc je je j'attaque d'autres sujets mais [mais disons]_{HH} [que **ce sont les natures mortes**]_{HH} [**qui me demandent le moins de souci de construction**]_{HH} [et]_{hh} et c'est ce qui naît le plus naturellement chez moi (CRFP)
 'A: it's the same kind of of patterns on the canvas or
 B: it's uh not to bore my my potential audience when I do some some small exhibitions uh I I I wouldn't like to hear omg only only still life paintings it's boring so I I tackle other topics but let's say that **it's still life paintings that require the least construction trouble** and it's what comes most naturally to me'

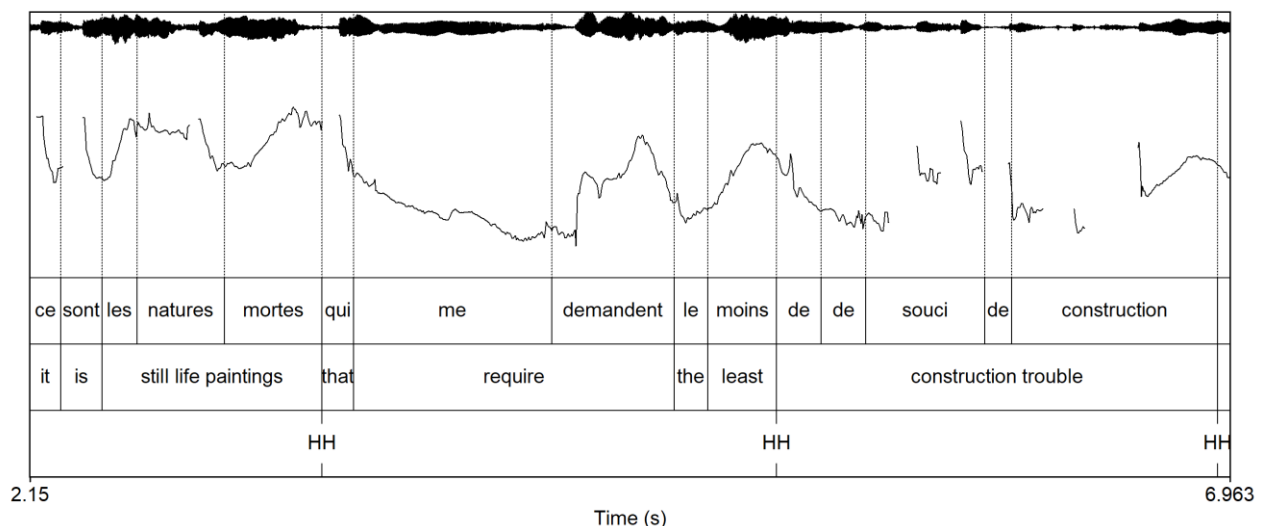


Figure 10. Prosodic realisation of (56)

Interactional effects are also found to be at play in some of the prosodic articulations subsumed under the uncategorised subclass. For example, a number

of these clefts appear to start as a focus-post-focus pattern whose clefted constituent takes on an L-L- tone at the right edge but whose cleft-relative clause ultimately ends in a continuation rise HH. According to Avanzi (2011) such cases are particularly striking instances of the influence of interactional needs. For him, in the case of a L-L-/HH pattern, the speaker 'recycles' the cleft-relative clause into information that is prosodically presented as incomplete in order to make room for the additional information coming afterwards (Avanzi: 120). This is illustrated in (57) below.

- (57) on obtient résultat cent pour cent avec eux parce que [ce n'est pas H
uniquement la partie sportive]_{L-L-} [qui joue/LL pour un joueur]_{HH} [C'
est vrai vous recevez un élément]_{HH} vous dites c'est un très bon joueur
(CRFP)

'we get a hundred percent result with them because **it's not just the sports aspect that plays a role for a player** it's true if you receive an individual you say it's a very good player'

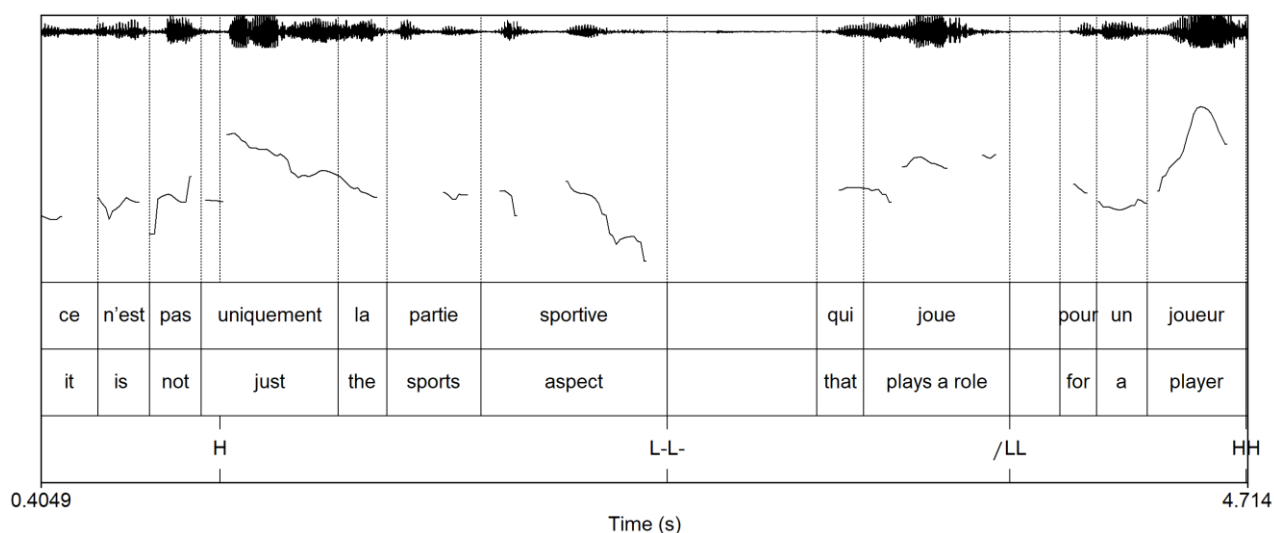


Figure 11. Prosodic realisation of (57)

Here, the speaker starts the construction as a type 1 cleft. The clefted constituent bears a unit-final L-L- tone and is followed by a pause seemingly indicating that his conversational unit is complete. Yet, the reduced cleft is here expanded through the addition of the cleft-relative clause. The fact that the cleft-relative clause is realised with a continuation rise HH at its rightward boundary allows the speaker to

continue on with the same topic by proposing an hypothetical situation to the addressee to illustrate his argument. In CA terms, the cleft-relative clause exemplifies the concept of increment, i.e. the extension of conversational units, whereby the speaker extends his turn beyond what is presented as a point of completion (Couper-Kuhlen and Ono 2007). Thus, the taxonomy proposed in the present study not only better allows one to account for the variety of prosodic patterns associated with *c'est*-clefts in spontaneous data, it also reflects the impact of contextual factors which otherwise tend to be overlooked by non-corpus-based studies.

4.2.3 The role of the initial accent

There exist different types of accent within the French accentuation system, some of which are compulsory, e.g. the final accent, and some optional, e.g. the initial accent or the emphatic accent ('accent d'insistance' Di Cristo (2018)). Both the initial and emphatic accents occur in initial position and may therefore be mapped onto one another. While the initial accent is a secondary rhythmic stress participating in the segmentation of discourse by combining with the final accent (Simon 2004), the emphatic accent serves to highlight a given lexical item and, in some cases, to signal the speaker's stance on the information conveyed (Di Cristo 2019: 42). The initial accent attaches to the first syllable of words whose leftward boundary it delineates. The emphatic accent also typically occurs on the first syllable of the word it lays emphasis on, but its use is more strongly dependent on specific lexical choices the speaker makes (*ibid*). With regard to its realisation, the initial accent is characterised by a rise in pitch accompanied by a slight lengthening of the onset of the word bearing the accent. By contrast, the emphatic accent triggers a rise in both pitch and intensity and is typically preceded by a short pause (Di Cristo 2019: 47). To illustrate the contrast between the two accents, let us look at (58) and its corresponding prosodic realisation in Figure 12.

- (58) A: et d'où vient cette passion pour la montagne
B: oh la passion pour la montagne elle est un peu elle a pris naissance
d'une manière un peu bizarre comme celle du vol à voile par euh ma

tentative de d'envol avec un parapluie euh mon fils qui est l'aîné de mes six enfants a eu euh la rougeole et j'avais entendu dire que euh pour se guérir de la rougeole euh il fallait changer d'air et changer d'air brusquement j'ai donc eu l'idée d'aller en montagne et puis en montagne j'ai d'abord fait de la luge j'ai emmené mon fils sur un sommet et sur une pente neigeuse assez rapide je suis descendu avec lui sur la luge je ne sais pas si c'est ça qui l'a guéri mais en tous les cas sa rougeole s'est passée mais ensuite je me suis dit mais pourquoi je ne ferais pas du ski j'ai acheté une paire de skis et [c'est ^Hdonc cet aspect]_{L-L} [de la maladie de NNAAMMEE]_{HH} [qui m' a amené]_{HH} [à pratiquer le ski]_{L-L} (CRFP)

'A: where does this passion for mountains come from

B: of the passion for mountains it's a little it stemmed in kind of weird manner like that of gliding with an umbrella uh my son who's the eldest of my six kids had uh measles and I had heard that uh in order to cure measles uh one needed a change the scenery and have a sudden change of scenery so I had the idea to go to the mountains and in the mountains I first went sledging I took my son to the summit and on a rather fast snowy slope I went down with him on the sledge I don't if it's that that cured him but in any case measles went away but then I told myself why wouldn't go skiing I bought skis and **it's thus that aspect of NNAAMMEE's sickness that led me to skiing'**

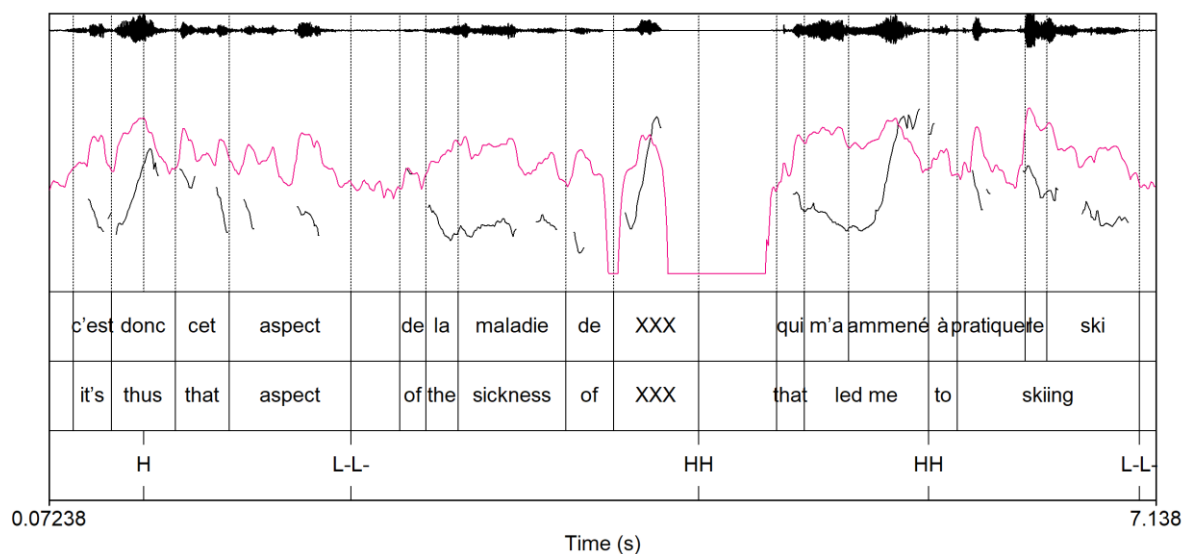


Figure 12. Prosodic realisation of (58)

The pitch curve is shown in black and the intensity curve in magenta. The cleft consists of four intonation units, two of which form the matrix clause and the other two the cleft-relative clause. The lexical items *donc*, *aspect*, *maladie* and *pratiquer* all display a slight rise in pitch which is not mirrored in F0 except for *donc* and *pratiquer*. As such, all four words instantiate the non-emphatic initial accent but only *donc* and *pratiquer* appear to illustrate the mapping of an emphatic accent onto the non-emphatic one. Within the first intonation unit running from *c'est* until *aspect*, both the pitch and intensity curves display a peak on the conjunctive adverb *donc*, which is also lengthened in comparison with the duration of other syllables in the unit. All three prosodic observations point to the encoding of an emphatic initial accent on *donc*. The same cannot be said for *pratiquer*, which does not display any significant lengthening. Informationally, this entails an increased emphasis on the concluding nature of the cleft which wraps up speaker's B lengthy explanation of his passion for mountains and skiing. This is further corroborated by the anaphoric nature of the clefted constituent *cet aspect de la maladie* which is understood in reference to measles being curable through a sudden change of scenery. The information conveyed in the cleft-relative clause, i.e. that speaker B's passion for skiing comes from *somewhere*, is likewise already mentioned in the preceding discourse and in fact serves as a point of departure for the whole conversational episode. That the cleft ends with the assertive tone L-L- reinforces the terminating

nature of the cleft with regard to the overall interactional structure, a use which had already been identified by De Stefani (2008).

Other contexts in which the emphatic initial accent was found to play a significant role include those in which the matrix clause contains a modal element placed before the clefted constituent. This is the case of (59) in which the adverb *toujours* indicates the moment, in this case *still up to this day*, in which the specification relation holds true. Prosodically, *toujours* is uttered with a high peak in F0 relative to the pitch height within the rest of the intonation unit. Though the peak is less perceptible in the intensity curve, the very short pause before the adverb along with the duration of its initial consonant are enough to mark it as bearing emphatic stress. Through this prosodic encoding, the expression of modality is therefore given a salient status which allows the speaker to add and underline the circumstantial information to the specification relation.

- (59) [ce sont ^Htoujours les mêmes personnes]_{HH} [qui taillent chez nous depuis trente ans]_{L-L} (CRFP)
'it's still the same people that have been doing the cutting for us for thirty years'

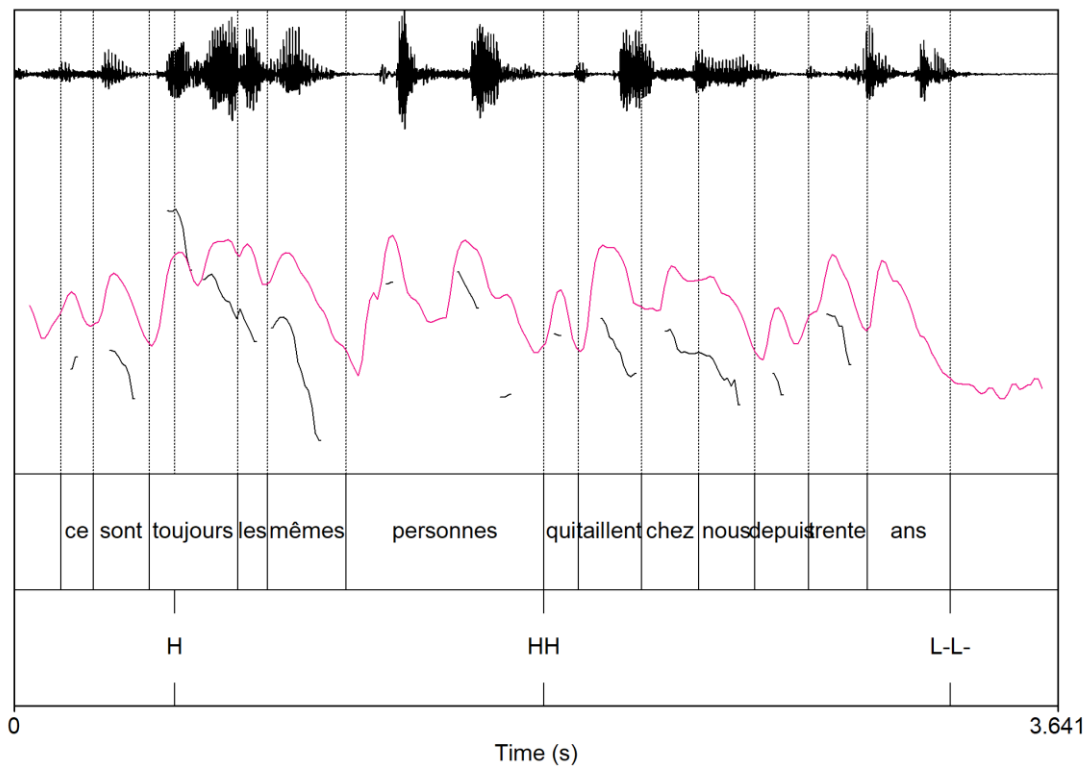


Figure 13. Prosodic realisation of (59)

Similarly, the emphatic initial accent may also attach to a negation operator such as *pas* as shown in (60). Here, the negation scopes over the following adverb *forcément* and combines with it to add a sense of tentativeness to the establishment of the specification relation. Thus, what the speaker is expressing is that it *may* be that work will be better if one spends more time on it, but it may also be the case that it will not. This epistemic modality is emphasised by means of the emphatic initial accent.

- (60) [c'est ^Hpas forcément quand on étale son travail dans le temps_{LL} qu'il est tellement mieux]_{L-L}. (CRFP)
 'it's not necessarily when we spread out our work in time that it's much better'

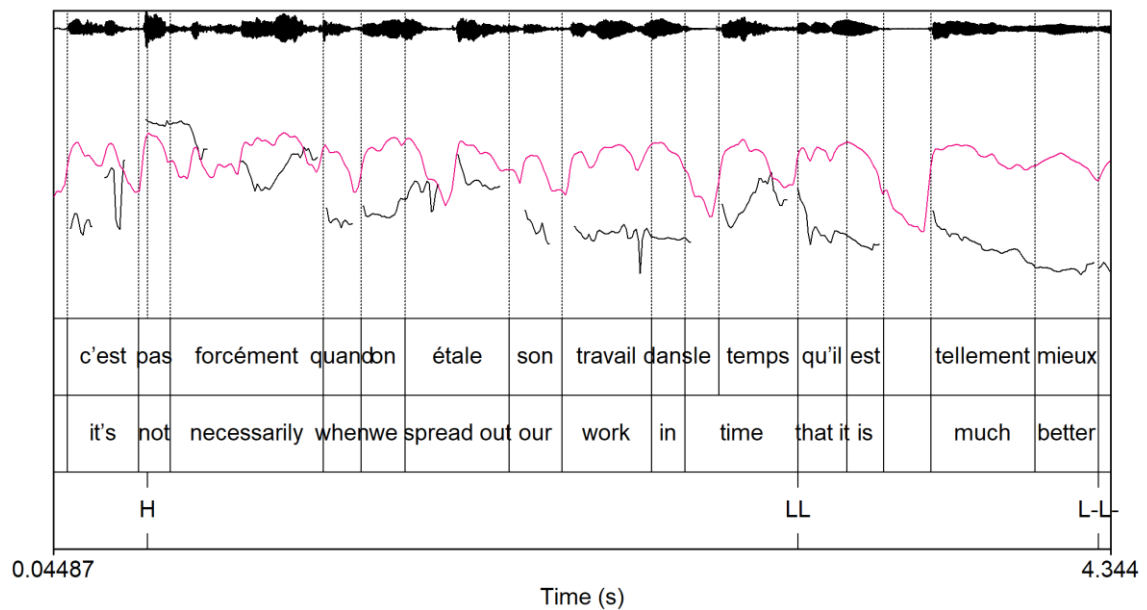


Figure 14. Prosodic realisation of (60)

It is worth noting that operators such as negation or epistemic markers do not always bear an emphatic accent, in which case the operators do not display any ‘double balisage’ and are thus not prosodically emphasised. In (61), for example, the pitch height of the negation operator *pas* is not relatively more prominent than that of other items within the intonation unit it is part of. In fact, it appears to be part of the unstressed onset leading up to the continuation rise HH occurring on the anaphoric clefted constituent *là*.

- (61) [c'est pas là]_{HH} [qu'on va réfléchir]_{HL}- [euh en fait]_{L-L}- [je dis bien oui]_{HL}-
 (CRFP)
 ‘it’s not at that point that we’re going to think things through uh
 actually I say so yes’

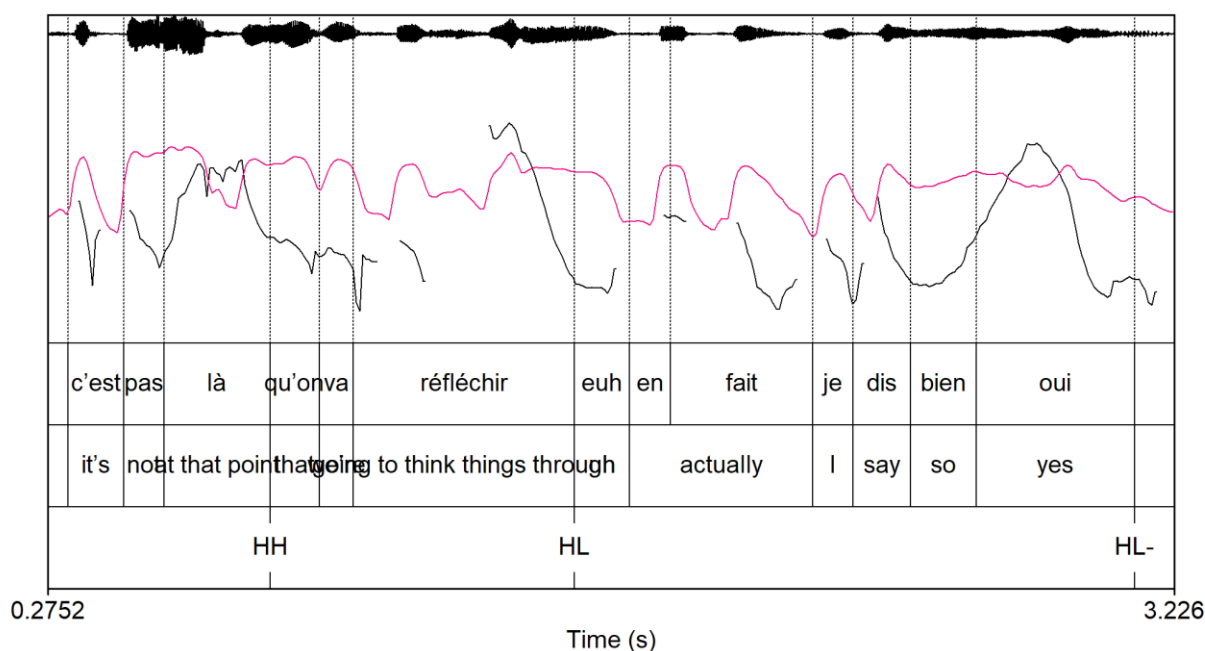


Figure 15. Prosodic realisation of (61)

Thus, the different types of prosodic accents afforded by French give the opportunity to speakers to disseminate prominences within intonation units. While the final accent is relatively fixed in its location, the emphatic initial accent can be borne by a variety of items such as modal operators, in the case of clefts, whose meaning is hence emphasised.

4.3 The referential information structure of *c'est-clefts*

I now turn to the referential layer of the information structure of *c'est-clefts* which is concerned with the distribution of discourse-new and discourse-given information within the cleft. After reviewing the different discourse-familiarity patterns attested in the CRFP dataset in Section 4.3.1, I propose a revised taxonomy of the relational information structure of *c'est-clefts* in Section 4.3.2.

4.3.1 Discourse-familiarity patterns

The application of the categories of discourse-familiarity included in the analytical model proposed in Section 2.4 allows me to classify tokens of *c'est-clefts* according

to the givenness/newness of their clefted constituent and cleft-relative clause. In total, the CRFP data display four patterns, viz. new-given, new-new, given-new and given-given, whose respective frequencies of occurrence are shown in Table 9 below.

	Full clefts	Reduced clefts	Total
New-given	46 (11.7%)	58 (65.9%)	104 (21.7%)
New-new	38 (9.7%)	—	38 (7.9%)
Given-new	178 (45.4%)	—	178 (37.1%)
Given-given	128 (32.7%)	30 (34.1%)	158 (32.9%)
Unclear	2 (0.5%)	—	2 (0.4%)
Total	392 (100%)	88 (100%)	480 (100%)

Table 9. Patterns of discourse-familiarity of *c'est*-clefts

The different patterns are exemplified in (62)–(65) in order of appearance in Table 9. Underlined segments correspond to material giving way to inferable or evoked information in the clefted constituent and/or cleft-relative clause.

(62) *new-given*

on a fait le dessin hein [*c'est ma fille*]_{HH} [*qui av-- qui a fait le dessin*]_{HH} elle me dit oh mais ça ça doit être un cheval qu'elle me fait (CRFP)

'we did the drawing huh **it's my daughter who ha-- did the drawing** she says to me oh but it it must be a horse that she's doing for me'

(63) *new-new*

A: elle m' a vraiment euh marquée cette dame

B: et comment vous avez été amenée à la connaître

A: ben parce que [*c'est la la secrétaire de mairie*]_{HH} [*qui m' a dit*]_{LL} euh écoute faudrait que tu ailles voir euh madame untel]_{HH} (CRFP)

'A: she really stood out this lady

B: and what led you to knowing her

A: well because **it's the townhall secretary who told me uh listen you should go see Ms. whatshername'**

(64) *given-new*

A: je voyais les photos euh enfin il y a des manifestations quand même à Caen je vois que ah c'est vrai oh totalement c'est pas dans le cadre d'un festival incroyable

B: non mais ça [c'est nous]_{HH} [qui les organisons]_{L-L-} (CRFP)

'A: I was seeing the pictures uh finally there are still demonstrations in Caen I see that ah it's oh totally it's not as part of an amazing festival

B: no but this **it's always us that organise them**'

(65) *given-given*

ben disc-jockey euh à l'origine c'était un passeur de disques mais maintenant disc-jockey est bien plus qu'un passeur de disques à mon goût c'est également euh ça ils font eux-mêmes leurs propres morceaux de techno tout ça et puis c'est vrai que la la dance tout ça c'est parti des DJ et c'est ce qui a fait leur popularisation quoi [c'est ça]_{HH} [qui qui a fait que bon]_{LL} maintenant on entend partout parler de ça quoi]_{L-L-} (CRFP)

'well disc-jockey uh originally it was someone who plays discs but now disc-jockey is much more than someone who plays discs to me it's also uh they make their own techno pieces and all and also it's true that dance music and all of that it started from DJs and it's what made them popular it's that that made it so that now we hear about it everywhere huh'

In revealing the existence of four patterns of discourse-familiarity, the taxonomy presented in Table 9 only partially corroborates the description offered by Riiland et al. (2002), Doetjes et al. (2004), Mertens (2012), Scappini (2013) and Karssenber and Lahousse (2015)⁵⁶ who only recognise the first three patterns, i.e. new-given, new-new, given-new. In addition to these, the CRFP data reveal the existence of a given-given distribution whereby the information conveyed in both the value and variable of the cleft is either explicitly evoked in the prior discourse, or at least

⁵⁶ Karssenber and Lahousse (2015) did not consider the referential status of the clefted constituent and cleft-relative clause separately but in relation to one another.

somewhat inferable. In (64), for instance, the anaphoric *ça* in the postcopular position refers back to *la dance tout ça*, which makes it evoked, and the presupposed open proposition that *something* caused dance music to be talked about everywhere is also clearly explained by the speaker beforehand with the mention that it becoming mainstream is incumbent upon DJs.

While all four patterns are instantiated in full clefts, as shown with (62)–(65), the information of reduced clefts is only organised into new-given or given-given patterns. That material in the cleft-relative clause of reduced clefts can only be given is consistent with the fact that it is its salience that makes it easy to omit without causing any incongruence.

With regard to the distribution of the different patterns, full *c'est-clefts* typically exhibit a given-new or given-given partition of information while new-given and new-new clefts are less frequent. As for reduced clefts, the new-given pattern is twice as frequent as the given-given one, which still numbers 30 occurrences out of 88. Given its proportion in both full and reduced clefts, the added given-given pattern reveals itself to be a significantly productive distribution of information, thus marking the need to revise the ternary typologies proposed in the literature on *c'est-clefts*.⁵⁷ Overall, the distribution of the different patterns thus does not uncover any particularly striking preference for one over the others.

Let us now look more closely at the informational status of the referents in the value and variable of *c'est-clefts*. Table 10 below summarises the frequencies of the different subcategories of given and new, in accordance with the model described in Section 2.4.1, and of the general categories given and new.

⁵⁷ As pointed out in Section 3.3 on the discourse-familiarity patterns of *it-clefts*, the given-given pattern has been identified for English clefts by Johansson (2001, 2002), Huber (2006) and Hasselgård (2014). It has not however been acknowledged for French.

		Full clefts	Reduced clefts	All clefts
Given	Evoked	283 (72.2%)	26 (29.6%)	309 (64.4%)
	Inferable	25 (6.4%)	4 (4.5%)	29 (6%)
New	New-anchored	15 (3.8%)	19 (21.6%)	34 (7.1%)
	Brand new	69 (17.6%)	39 (44.3%)	108 (22.5%)
	Total	392 (100%)	88 (100%)	238 (100%)

Table 10. Distribution of discourse-familiarity types in values of *c'est*-clefts

The results presented in Table 10 reveal a number of things. Within the full cleft category, values carry evoked information significantly more than brand new, inferable and new-anchored information. As a result, they more frequently convey given than new information. This overwhelming prevalence of givenness in the value can be related to the great use of clefts with a pro-form as their value, as the syntactic analysis provided in Section 4.1.2 had already shown, and as is illustrated in (66) and (67).⁵⁸ The pro-form may be a pronoun referring to one of the speech participants as in (66) or an anaphoric adjunct as in (67). In both cases, the referent is understood in reference to the discourse or situational context.

- (66) là l'exposition qu'il y a dans le hall le concept général de l'exposition
[c'est moi qui l'ai mis en place]_{HH} (CRFP)
 'there the exhibition that's in the hall the general concept of the exhibition **it's me who set it up**'
- (67) je suis passé au dériveur du dériveur je suis passé au bateau ponté et
[c'est ainsi]_{HH} **[que avec plusieurs de mes enfants]**_{HH} **[j'ai navigué euh à Carnac]**_{HH} (CRFP)
 'I switched to sailing dinghy's from sailing dinghy's I switched to bay liners and **it's then that with several of my children I sailed uh in Carnac**'

⁵⁸ It is worth noting that references to speech participants like *moi* would be considered new information in typologies like Karssenbergh and Lahousse's (2015) in which discourse-familiarity is assessed relationally.

By contrast, reduced clefts show more variation in the type of givenness/newness they allow in the value. All four categories are instantiated, with brand new values being the most frequent followed in order by new-anchored, evoked and inferable ones. In terms of the givenness/newness contrast, reduced clefts tend to select non-recoverable information over recoverable one. The distribution is therefore opposite to that observed for full clefts. This suggests that reduced *c'est*-clefts tend to be more readily used by speakers when the value specified in the matrix clause is discursively new while full clefts appear to be more appropriate for the reiteration of already-mentioned lexical items. To account for this asymmetry, I delve into the functional aspect of full and reduced *c'est*-clefts in Section 6.3.3.

Turning to the discourse-familiarity of the referents introduced in the cleft-relative clause, the frequencies of the discourse-familiarity subtypes are shown in Table 11.

		Full clefts	Reduced clefts	All clefts
Given	Evoked	49 (12.5%)	77 (87.5%)	126 (26.3%)
	Inferable	126 (32.1%)	11 (12.5%)	137 (28.5%)
New	New-anchored	39 (10%)	–	39 (8.1%)
	Brand new	178 (45.4%)	–	178 (37.1%)
Total		392 (100%)	88 (100%)	238 (100%)

Table 11. Distribution of discourse-familiarity types in the variable of *c'est*-clefts

In comparison with the results on the discourse-familiarity types of values, those of variables show an overall wider homogeneity. In full and reduced clefts combined, the new-anchored category is the least represented category behind the evoked, inferable and brand new ones. Given that reduced clefts only convey discourse-given information in their implicit cleft-relative clause, these results are however more relevant when the two variants of *c'est*-clefts are taken separately. Thus, reduced clefts typically convey evoked information in their cleft-relative clause, but may also introduce information that is recoverable through inferential bridges. That evoked variables are more frequent than inferable ones is in line with the fact that the stronger the salience of the variable is, the more likely it will be omitted. This is exemplified in (68) and (69). In (68), the presupposed open proposition at

play in the specification relation can be reconstructed as *je vous dis ça parce que x* 'I tell you that because of x' from the speaker's own underlining that s/he is telling the addressee that the reduction of working hours would help fight wealth discrepancies. As such, (68) illustrates the subcategory of pre-modified reduced clefts in which the variable is introduced in a pre-modifying clause and is therefore not restated (Declerck and Seki 1990).⁵⁹ In (69), the variable for which the value *c'est parce que je suis pas là depuis longtemps* is arguably less salient and is only indirectly retrievable through the implication attached to having hit it off more quickly elsewhere.

- (68) *c'est bien qu'il y a un problème ça veut bien dire que la richesse délogée va pas dans la bonne poche d'où l'idée pour lutter contre ça c'est de dire ben réduction du temps de travail d'accord donc si je vous si je vous dis ça [c'est parce que il faut pas croire que toutes les entreprises]_{HH} [sont sur le fil du rasoir]_{HH} [~~que je vous dis ça~~] (CRFP)*
 'it goes to show that there's an issue it means that the wealth that is cleared doesn't go into the right pocket which is where stems the idea to fight against that it's to say well reduction of working time alright so if I'm telling you that **it's because you shouldn't believe that every company is on a razor's edge** [~~that I'm telling you this~~']
- (69) *les non oui c'est sûr que bon j'y suis là il y a pas très longtemps ici mais à la fac quoi mais c'est vrai qu'à l'IUT c' était plus sympathique on a j'ai sympathisé plus vite qu' ici les étudiants ils étaient plus euh plus sympa euh bon [c'est peut-être parce que je suis pas là depuis longtemps]_{HH} [~~que j'ai sympathisé moins vite ici~~] ça va peut-être (CRFP)*
 'the no yes of course I'm here not too long ago ici but at university but it's true that at UIT it was more pleasant we I got on well more quickly than here students they were more uh nicer well **it may be because I haven't been here for too long** [~~that I got on more slowly here~~']

⁵⁹ Declerck and Seki (1990) only discuss the use of pre-modified clefts in English which they refer to as pre-modified *it*-clefts (PRICs). I investigate the existence of a French equivalent in Section 6.2.3.

Whether directly or indirectly recoverable, the cleft-relative clause remains characterised by its high degree of identifiability and hence of reconstructability. Full clefts, on the other hand, generally have brand new or inferable referents in the variable, or, to a lesser extent, evoked or new-anchored ones. This shows that an evoked or inferable variable does not necessarily entail that it will be omitted by the speaker. In (70), for instance, the fact that somebody is running for the local elections is given in speaker A's question in the preceding turn. After uttering the same reduced cleft three times indicating that it's older people who typically are candidates, speaker B goes on to further specify the type of people who are running in these elections, in this case always the same ones. As she does so, she reiterates the cleft-relative clause *qui se présentent* which had been left implied for the three previous reduced clefts and the one in speaker A's turn.

- (70) A: hum hum hum par exemple est-ce qu'il y a des jeunes qui se présentent aux élections ici ou c'est tous des vieux eh ben soyons brutaux brutaux
 B: c'est euh des vieux ici c'est des vieux c'est des vieux hé [c'est toujours les mêmes]_{HH} [déjà]_{h+h+} [qui se présentent]_{L-L} euh mais il y a jamais eu des jeunes qui ont osé se présenter (CRFP)
 'A: hm hm hm for instance are there young people who run for election here or is it only old people ah well let's be brutal brutal
 B: it's uh old people here it's old people it's old people eh **it's always the same people first of all who run for election** uh but there have never been young people who dare running for election'

As with reduced *it*-clefts (see Section 3.3.1), the findings on reduced *c'est*-clefts raise the question as to what motivates the speaker's choice between omitting or restating a discourse-given variable. From (70), I hypothesise that this binary choice is mainly tied to the communicative needs specific to the discourse context in which the cleft occurs, which would justify the use of reduced clefts beyond the mere discourse-givenness of their cleft-relative clause. In (70), for instance, speaker B restates the variable which had already been introduced in speaker A's question but which had remained implicit in the first three clefts in speaker B's answer. While

those three reduced clefts introduce *des vieux* as the appropriate value for the variable *x se présente(nt)*, the full cleft in boldface introduces a new value, i.e. *toujours les mêmes*, which does not cancel the previous value but rather subsumes it. The speaker's effort to further characterise the type of people who run for election is thus made evident by the repetition of the cleft-relative clause. In view of this, the tension between the two variants observed in the CRFP dataset, which has not been addressed in any of the existing studies on reduced *c'est*-clefts, is best accounted for in terms of discourse contexts and interactional needs which are both addressed in the functional account offered in Section 4.5.

4.3.2 A revised typology of the discourse-familiarity of *c'est*-clefts

The empirical findings on discourse-familiarity presented in the preceding section highlight the need to rethink the typologies found in the literature. In order to better account for all patterns attested in the CRFP dataset, a revised classification is proposed in Table 12.

Type of <i>c'est</i> -cleft	Value	Variable
new-given	new	given
new-new	new	new
given-new	given	new
given-given	given	given

Table 12. Typology of discourse-familiarity patterns of *c'est*-clefts

Taking the general given vs. new distinction as a starting point, the taxonomy establishes the different patterns afforded by *c'est*-clefts based on the givenness of their value and variable. In total, four patterns are attested, which correspond to the four logical possibilities deriving from the two-tiered categorical parameter 'discourse-familiarity'. As exemplified in the previous section, *c'est*-clefts may partition information as: (i) new-given with a discourse-new value followed by a discourse-given variable, (ii) new-new whereby both value and variable are discourse-new, (iii) given-new with a discourse-given value followed by a discourse-

new variable and (iv) given-given whereby both value and variable are discourse-given.

In contrast with the information structural-based typologies of *c'est*-clefts described in Section 1.2.2, the typology in Table 12 does not imply a one-to-one match between the referential and relational layers of the information structure of *c'est*-clefts. As such, it solely accounts for discourse-familiarity independently of prosodically coded information structure. The interaction between the two is addressed in Section 4.4.2. The typology I propose also differs from the existing ones in the number of patterns recognised. Contra Rialland et al. (2002), Doetjes et al. (2004), Mertens (2012), Scappini (2013) and Karszenberg and Lahousse (2015), I propose a quaternary typology to which the given-given pattern, whose existence has been acknowledged for English (Johansson 2001, 2002; Huber 2006; Hasselgård 2014) but not for French, is added. In this pattern, all referents are recoverable either from explicit mentions in the prior context, or through inferential bridges. In (71), for instance, the anaphoric clefted constituent and value *ça* is understood as referring to the fact that the speaker is not interested in biology stated in the preceding clause. The information conveyed in the cleft-relative clause, i.e. that something is difficult, is also recoverable from the context with the speaker's repeated mentions, as underlined, that he has trouble with the biology curriculum. Taken separately, the information in both parts of the cleft is given. Yet, the specification relation between the value and variable is only made explicit by the use of the cleft. In other words, without the cleft, the fact that it is the speaker's lack of interest in biology above everything else that made him perceive the subject as difficult might not necessarily be obvious to the addressee. The speaker brings it out with the cleft. Thus, a cleft of the given-given type may still hold some newness in the semantic relation that is established between the different referents of the cleft.

- (71) la biologie je bloque sur la biologie j'ai une formation littéraire donc euh la biologie t-- enfin quoique cette année ils ont changé un petit peu le programme c'est un peu plus léger c'est un peu plus accessible quoi mais l'année dernière c'était vraiment enfin lourd moi j'avais beaucoup de mal à tout tout ce qui est anatomie du cerveau tout ça j'a

j'avais vachement du mal bon ça ça y est encore quoi mais ils le ils le présentent un petit peu sous une autre forme et euh c'est plus basé euh psychologie quand même que que l'année dernière là c'est plus euh sur l'étude la dépression et tout ça quoi alors que l'année dernière c'était vraiment lourd on faisait la motricité tout ça je enfin je sais pas si euh si ça peut-être peut-être que ça sert en en psycho ou dans certains domaines quoi mais euh moi j'avais moi j'avais du vraiment du mal à visualiser en fait euh le fonctionnement de de des muscles par rapport au cerveau tout ça j' a- j' arrivais pas trop à suivre le circuit quoi et puis ça m' intéresse pas en fait trop trop donc [**c'est surtout ça**]_{L-L} [**qui est difficile**]_{L-L} (CRFP)

'biology I go blank on biology I have a literary training so uh biology t- I mean this year they changed the programme a little it's lighter it's more accessible do you know what I mean but last it was very well heavy I had a lot of trouble with everything everything related to the brain's anatomy all of that I ha- I had very much trouble well this this is still the case but they present it in a different way and uh it's more psychology-based after all than than last year now it's more uh on the study of depression and all of that whereas last year it was really heavy we were doing motricity and all that I mean I don't know if uh if it may be may be useful in psychology or in certain fields but uh I had I had a lot of trouble visualising well uh how muscles function compared to the brain and everything I cou- I couldn't really followed the circuit and wasn't really interested in it actually not really not really so **it's especially that that's difficult**'

By contrast, in (72), both the content of the cleft and the specification relation are given verbatim in the prior context. In this excerpt, the speaker is making the case that social workers assigned to specific areas typically work independently on cases and are therefore directly exposed to disabled people's problems. She contrasts this with cases in which multiple workers are in charge of a given file which is forwarded from one to another which, according to the speaker, causes the involved workers to have a more limited first-hand experience of issues encountered by

disabled people. The cleft occurs after the speaker has already described the two situations she opposes and after she tentatively attempts to clarify her argument by stating that even independent workers may at some point collaborate with colleagues. In stressing the fact that even in those cases it is still *them*, i.e. independent workers, that ultimately know about the problems, the cleft re-introduces and emphasises discourse-given information with a clarification, and arguably persuasive, purpose.

- (72) pour moi la réinsertion a été très bien j'ai repris du travail à la DASS où je me j'étais auprès des handicapés vous voyez je recherche toujours un petit peu ce qui pas ce qui est facile enfin c'est pas c'est pas que c'est pas facile mais où on sent que les gens ont plus besoin de vous les familles ont besoin de vous hein parce que c'est vrai sur le secteur aussi on a besoin mais j'avais l'impression que sur le secteur d'abord il faut vous dire que ce sont des filières on on a on quand on fait une enquête des fois d'abord solliciter les une autre assistante sociale qui retransmet à une autre assistante sociale et bien souvent à la fin de la chaîne vous savez pas trop ce qui s'est passé tandis que dans un tout à fait déterminé dans un endroit très déterminé euh là on sait on s'occupe des handicapés c'est handicapé point final et on fait tout bien sûr on a des relations avec les autres secteurs sociaux mais euh [**c'est quand même nous**]_{HH} [**qui sommes au courant**]_{HH} on a fait les démarches du départ on assiste à la souffrance des gens la souffrance des enfants (CRFP)

'for me the process of reintegrating was very good I took up work at the DASS (note: former Department of Health and Social Security) where I was around disabled people you see I always look a little for what isn't what isn't easy I mean it's not it's not that it's not easy but it's where you feel like people need you the most families need you right because it's true that in the area too we need but I had the feeling that in the area you need to tell yourself that it's sectors we we have we when do an investigation sometimes ask the another social worker who forwards to another social worker and most of the time at the end

of the line you don't really know what happened whereas in a very specific in a very specific place uh there we know we take care of disabled people it's disabled period and we do everything of course we have relations with other social sectors but uh **it's still us who know about it** we take the necessary steps at the beginning we experience people's suffering children's suffering'

With examples (71) and (72), I show that clefts of the given-given type may still introduce new information by exposing the otherwise implicit specification relation at play between the different referents, but it may also be entirely given when both the content and the specification relation are given. From a functional viewpoint, the two examples also offer a glimpse of the range of uses associated with *c'est-clefts*, which I will come back to in Section 4.5. Interestingly, the prosodic realisations of (71) and (72) exemplify two different prosodic patterns, viz. type 1 and 2b respectively, despite presenting the same partition of information in its referents, i.e. given-given. (71) displays the focus-post-focus articulation whereby the anaphoric referent *ça* is bounded by an L-L- tone at its right edge and is therefore in focus while the cleft-relative clause is realised with the appendix contour l-l-. In (72), both the matrix clause and cleft-relative clause take on a continuation rise HH indicating that focus is broad and scopes over the whole construction. This suggests that given information in French may still be focal and that different patterns of discourse-familiarity are not always prosodically realised. Whether these hypotheses are verified in other patterns is examined in the following section. From a contrastive viewpoint, the fact that the four possibilities for the partition of information are the same ones exploited by English and French is dealt with in Section 5.2.2.

4.4 The interplay between the relational and referential information structure of *c'est-clefts*

Having characterised the relational and referential layers of the information structure of *c'est-clefts* independently of each other, I now investigate the ways and extent in which they interact. For this, I start by examining the interplay

between syntax and prosody in Section 4.4.1, then between discourse-familiarity and prosody in Section 4.4.2, which leads me to devise a fine-grained information structural-based typology which relates referential and relation information structure to each other. Finally, I address the question of the grammaticalisation of the information structure of *c'est*-clefts in Section 4.4.3 by focusing on anaphoric clefts with a cohesive function.

4.4.1 The interaction between syntax and prosody

As noted by Lehmann (2008: 211), relational information structure is encoded by means of “immaterial features” such as word order and prosody, both of which are at play in clefts. This interaction has been recognised by a number of studies (see Vander Klok et al. 2018 and Portes and Reyle 2022) which have argued that prosody and syntax generally combine to code focalisation of the clefted constituent. Using the CRFP data, I examine the extent of the interplay between the two strategies which I argue plays out in an online and moment-by-moment fashion. For this, I rely on (73)–(75) which exemplify prosodic types 1, 2a and 2b, and in which the scope of prosodic focus differs.

- (73) Paul NNAAMMEE est-ce que est-ce que v~ vous remettez en cause euh l'implantation par exemple d'une usine d'incinération à Portes-lès-Valence est-à-dire à proximité d'une grosse agglomération ou bien **[c'est le système même de l'incinération]**_{L-L} **[que que vous rejetez]**_{L-L}. (CRFP)

‘Paul NNAAMMEE do you do you~ call into question uh the creation of an incineration plant in Portes-lès-Valence that is to say close to a large urban area or **it is the incineration system itself that that you reject**’

- (74) bon bon alors il y a des gens qui ont des u~ des u~ des u~ des usines qui décolorent la le la cire c' est pourquoi **[c'est comme ça]**_{HH} **[qu'on a des cierges blancs]**_{HL}. (CRFP)

‘well well okay there are people who have f~ f~ f~ factories which bleach was that's why **it's in that way we have white church candles**’

- (75) A: donc vous ne cherchez pas les on vous amène à demeure les si on peut dire les fleurs pour faire les compositions et tout vous n'allez pas les chercher aux halles ou
 B: non ça c'est les [**c'est les grossistes**]_{HH} [**qui s'en occupent**]_{L-L-} (CRFP)
 'A: so you don't pick up the someone brings to you the so to speak the flowers to make the arrangements and all you don't go pick them up at the markets or
 B: no this it's the **it's the wholesalers that take care of it'**

For Jespersen and Haislund ([1954] 2007), the bi-clausal syntax of the cleft systematically allows to 'single out' the element placed in postcopular position in the matrix clause whose newsworthiness may be further highlighted prosodically. This is the case in (73) in which the value NP *le système même de l'incinération* is not syntactically highlighted by the cleft syntax also prosodically construed as focal while the rest of the cleft is backgrounded. As such, it illustrates narrow focus which Di Cristo (2019: 278) deems highly contrastive and revealing of some sort of contradiction with the prior context. In this case, the value is proposed as one of two alternatives, the other one being *l'implantation d'une usine d'incinération à proximité d'une agglomération*, potentially fulfilling the variable *vous rejetez x*. The focus-post-focus-sequence attributed to the cleft thus spells out the communicative intention of the speaker to lay out paradigmatic alternatives from which the hearer is expected to pick one. In (74), by contrast, the combination of the continuation tone HH after the clefted constituent and the HL- tone located at the right edge of the cleft-relative clause indicate that focus is to be interpreted as bearing over the whole cleft. This time, what the speaker is presenting as being informationally salient is the fact that there exist white church candles and that they are made by means of a bleaching process, as marked by the 'involvement' tone (Mertens 2006: 11). This creates an asymmetry in the syntactic highlighting triggered by the bi-clausal syntax and the prosodically-marked function of focus. The same discrepancy is also observed in (75) in which the assertive boundary tone L-L- is located at the right edge of the cleft-relative clause thus including the information in the whole cleft in the focus. The value *les grossistes* represents irrecoverable information freshly introduced into the discourse while the

presupposed open proposition *x s'en occupe* is arguably predictable from speaker A's request for confirmation on whether speaker B is involved in the picking up of flowers from the markets. With focus scoping over the whole cleft, both value and variable are presented as equally prominent pieces of information despite the latter being somewhat discourse-given. Unlike in (73), the speaker is not directing the hearer's attention to a specific lexical item but to the whole cleft.

Thus, the emphasis placed on the clefted constituent through the information re-packaging is not always reflected in the prosodic articulation of the clefts. The two strategies, syntactic and prosodic, may co-occur to construe the clefted constituent as particularly salient informationally but they may also be de-coupled to give rise to different patterns of highlighting motivated by communicative needs which the hearer attends to as speech progresses.

4.4.2 The interaction between discourse-familiarity and prosody

Unlike accounts of *it*-clefts, the different studies of the discourse-familiarity of *c'est*-clefts this study reacts to do not establish as straightforward a link between newness to prosodic focus and givenness to deaccentuation. Instead, Rialland et al. (2002) and Doetjes et al. (2004) distinguish between different types of *c'est*-clefts based on the type of focus displayed, viz. narrow or broad focus, but they do not systematically associate them with specific discourse-familiarity patterns. In fact, Doetjes et al. (2004: 534) claim that focus is not restricted to discourse-new referents and that discourse-given items may also be focused. This claim is largely corroborated by the findings for the CRFP data. To illustrate this, let me focus on the informational status of value referents bearing narrow focus. First, Table 13 summarises the discourse-familiarity of clefted constituents in focus.

	Full clefts	Reduced clefts	Total
New	12 (55%)	26 (63%)	38 (60%)
Given	10 (45%)	15 (37%)	25 (40%)
Total	22 (100%)	41 (100%)	63 (100%)

Table 13. Information status of values carrying narrow focus

The overall results do not uncover any particular preference for discourse-new or discourse-given information in focused values. Rather, both full and reduced clefts may have either a new or given focused referent in postcopular position. (76) illustrates the former with the anaphoric pronoun *eux* referring to *capos*. The latter is instantiated in (77) in which the PP *à Pau* represents irrecoverable information which is introduced in discourse for the first time.

- (76) après tu as les capos les capos est eux qui gèrent tous les lieutenants ou~ c'est eux qui gèrent une armée [c'est eux]_{L-L} [qui gèrent l'armée]_L. (CRFP)
 'then you have corporals corporals are them who manage all the lieutenants or~ **it's them** who manage an army it's them who manage the army'
- (77) non le camion est à l'entreprise aussi ils ont trois trois camions mais euh le grand le grand dépôt vraiment [c'est à Pau]_{L-L} [qu'ils ont le grand dépôt pour les boissons]_{L-L}. (CRFP)
 'no the truck belongs to the company too they have three three trucks but uh the big the big warehouse really **it's in Pau** that they have the big warehouse for drinks'

When focus is broad, the referents in the matrix clause and the cleft-relative clause still likewise show variation in their given/new status. In (78), all referents are new and their informational salience is marked by means of an HL- boundary tone. By contrast, the referent *malle* in (79), which also carries an HL- tone encoding broad focus, is co-referential with the underlined NP *une malle* and hence discourse-given as shown by the demonstrative determiner that precedes it.

- (78) bon j'habite en ville hein j'avais pas de voiture [c'est le taxi qui m'a déposé_{BB} euh devant la maison avec mes bagages]_{HL}. (CRFP)
 'well I live in the city eh I didn't have a car **it's the taxi that dropped me off uh in front of the house with my luggage**'
- (79) euh on a créé une malle euh à à soixante-dix pour cent euh qui s'appelle Roule ta boule voilà et roule ma poule et en fait non elle est pas elle

est là-bas et en fait donc [**c'est Cardère**]_{HH} [**qui est à l' initiative de de cette malle**]_{HL} (CRFP)

'uh we created a trunk uh up up to seventy percent uh which is called Roule ta boule (note: pun literally translated as Rowl your ball) there you go and off we go and actually no it's not it's over there and so **it's Cardère who's behind the initiative for for this trunk'**

From this, I conclude that the givenness of a referent does not obligatorily entail its deaccentuation within the unit and similarly newness does not always involve prosodically coded focus. The empirical findings of the interaction between focus and givenness/newness are thus in line with Doetjes et al.'s (2004) claim that the two phenomena are not strictly correlated.

How then can we typify the information structure of *c'est*-clefts? Mapping the prosody-based taxonomy of *c'est*-clefts proposed in Section 4.2.1 onto that of discourse-familiarity patterns proposed in Section 4.3.2 leads to the typology in Table 14. Due to the differences in their tonal coding, interrogatives are not included in the table.

	Prosodic type	Value	Variable
		Discourse-new	Discourse-given
Unmarked	2a/2b		broad focus
Marked	1	narrow focus	deaccented
	3b		broad focus
		Discourse-new	Discourse-new
Unmarked	2a/2b		broad focus
Marked	3a/3b		broad focus
	1	narrow focus	deaccented
		Discourse-given	Discourse-new
Unmarked	2a/2b		broad focus
Marked	3a/3b		broad focus
	1	narrow focus	deaccented
		Discourse-given	Discourse-given
Unmarked	2a/2b		broad focus
Marked	3a/3b		broad focus
	1	narrow focus	deaccented

Table 14. Overview of focus patterns of each discourse-familiarity patterns

As revealed by the taxonomy in Table 14, none of the patterns of discourse-familiarity of *c'est-clefts* are associated with a unique prosodic profile. Instead, each of the four patterns of discourse-familiarity is associated with more than one prosodic articulation. Among these, the most frequent one is described as the unmarked and hence expected choice while the other ones are marked choices. All four patterns of givenness/newness rely on the same unmarked prosody corresponding to type 2a/2b whereby the cleft consists of two or more units, the clefted constituent is followed by a continuation rise, and the cleft-relative clause either an assertive tone L-L- or a continuation tone like the clefted constituent. In both cases, focus is broad and scopes over the whole cleft, as is the case in (80). Here, the clefted constituent *le nul* is identified as the appropriate value for the presupposed open proposition *x a besoin de concentration* which is already

introduced in the immediate co-text. Prosodically, focus is not tied to a specific lexical item, thus making the content of the whole cleft informationally salient.

- (80) en règle générale déjà primo *c'* est très bruyant primo donc le gamin qui a besoin de concentration lui il arrive plus du tout à bosser donc en règle générale [*c'est le nul*]_{HH} [*qui a besoin de concentration*]_{HH} (CRFP)

'as a general rule first of all it's very noisy firstly so the kid that needs to focus he can't work at all anymore so generally it's the slow kid that needs to focus'

Interestingly, the symmetry observed for the unmarked prosodic realisation of the different informational types of *c'est-clefts* is largely mirrored in the number and frequencies of the marked patterns. The marked realisations of new-new, given-new and given-given, i.e. the less frequent realisations, fall either within the 3a/3b or 2a/2b categories, which are types of broad focus exemplified in (81) and (82), or category 1 with a narrow focus on a specific lexical item such as *calandre* in (83).

- (81) à partir du mercredi donc là [*c'est là qu'on a notre plus grosse livraison*]_{HH} (CRFP)

'from Wednesday onwards so then it's then that we have our biggest delivery'

- (82) donc [*c'est des choses comme ça*]_{HH} [*que je recherche*]_{HL-} (CRFP)

'so it's things like that that I'm looking for'

- (83) *c'*est nous qu'on le qu'on le fait donc [*c'est la calandre*]_{L-L-} [*qui travaille aussi*]_{I-I-} (CRFP)

'it's us that it that do it so it's the radiator that works too'

Only new-given clefts show more restrictions in their prosodic encoding, with only a handful of type 3b tokens and no 3a ones.

The typology proposed in Table 15 not only backs up Doetjes et al.'s (2004) observation on the varied informational statuses of focused elements, but it also confirms the lack of direct correlation between the relational and referential layers

of the information structure of *c'est*-clefts. The concurrence of broad and narrow focus, neither of which are tied to a specific partition of information, in *c'est*-clefts appears to be indicative of the discursive versatility of the construction and its role in the moment-by-moment management of information. I further delve into this by providing a functional overview of *c'est*-clefts in Section 4.5.

4.4.3 *C'est là que, c'est ainsi que, c'est comme ça que*: evidence for grammaticalisation of information structure?

As shown in Section 4.1.3, *c'est*-clefts exhibit a high number of anaphoric clefted constituents. Besides pronominal values, anaphoric clefts also include those which contain an adverbial pro-form such as *là* 'there/then', *ainsi* 'how/that way' or *comme ça* 'like that/in that manner' which are most richly exploited in spoken French (Roubaud and Sabio 2015). All three types have been singled out by a number of studies (see Blanche-Benveniste 2006; Sabio and Benzitoun 2013; Roubaud and Sabio 2015 for *c'est là que* clefts; Lahousse and Lamiroy 2017 for *c'est ainsi que* clefts; Roubaud and Sabio 2018 for *c'est comme ça que* clefts) as clefts whose contrastive and focalising functions are, in some specific discursive contexts, weakened. Hence, this makes them an ideal subtype of clefts to study the interaction between syntax, focus marking and information structure which I will do in this section.

58 occurrences of the three abovementioned types could be retrieved from the CRFP. Among those 58 occurrences, only full clefts were found. Their distribution is shown in Table 15 below.

	Full clefts	Reduced clefts	Total
<i>c'est là que</i>	35	—	35 (60.3%)
<i>c'est ainsi que</i>	5	—	5 (8.6%)
<i>c'est comme ça que</i>	18	—	18 (31.1%)
Total	58	—	58 (100%)

Table 15. Distribution of *c'est là que*-, *c'est ainsi que*- and *c'est comme ça que*-clefts

Among the three subtypes, *c'est là que*-clefts are the most frequent with 35 tokens, followed by *c'est comme ça que*-clefts with 18 instances and *c'est ainsi que*-clefts with 5 tokens. This distribution is in line with that found by Roubaud and Sabio (2015) in the spoken component of their corpus. That adverbial anaphoric clefts with *là*, *ainsi* and *comme ça* are only of the full type is also consistent with Roubaud and Sabio's (2015) conclusion that clefts with non-contrastive clefted *là* cannot be reduced because of the informative aspect of their cleft-relative clause.

More specifically, the three types of anaphoric clefts display two uses: contrastive or discursive. In the contrastive use, *là*, *ainsi* and *comme ça* express referential spatio-temporal meanings or manner, and code contrast, which can be lexicalised or implicit. These options are illustrated in (84)–(86).

- (84) A: *c'est-à-dire que bon euh dans ma famille euh mon père est déjà un manuel*
 B: *hum hum dans quelle branche*
 A: *ben lui il est dessinateur euh industriel professionnel mais euh il a toujours euh dessiné fait de un peu de peinture fait du fer forgé euh de la gravure donc avec lui je me suis initié un peu à tout ça aussi*
 B: *à la gravure*
 A: *à gravure sur bois euh pyrogravure enfin vous voyez un peu tous ces domaines et il fais~ il tenait un une section dans un Cen-- Centre Culturel dans un une petite ville du territoire ici à Delle*
 B: *hum hum*
 A: *et pou-- dans le Centre Culturel et bon [**c'est là que j'ai un peu commencé à faire certaines choses**]_{HL} bon je dessinais beaucoup quand j'étais gamin (CRFP)*
 'A: I mean well uh in my family uh my dad is already a manual worker
 B: hm hm in which field
 A: well he's a draughtsman uh professional in the industry but uh he's always uh drawn done a bit of painting done wrought iron uh done engraving so with him I also got a little bit into all of that
 B: into engraving

A: wood engraving uh pyroengraving well you see a little bit of all these domains and he di- he had a pitch in a cultural centre in a small city of the region here in Delle

B: hm hm

A: and fo~ in the cultural centre and well **it's there that I slowly started doing some things** well I drew a lot when I was a kid'

- (85) ce budget se situe dans la continuité de l'action engagée par l'équipe municipale depuis quatre-vingt-quinze il est marqué tout comme les exercices précédents par la volonté de notre équipe de poursuivre le développement de Nanterre au profit de l'ensemble des Nanterriens et avec la population [c'est ainsi qu'il privilégie une série de grandes priorités]_{HH} [définies avec les citoyens au long de ces dernières années]_{LL} dont découlent d'importantes dépenses nouvelles]_{L-L} (CRFP)

'this budget is in line with the action that has been started by the local team since ninety-five it's marked like every preceding exercise by our team's will to go on with the development of Nanterre in favour of all the Nanterre inhabitants and with the population it's in that way that it favours a series of big priorities defined with the citizens throughout the last years from which stem important new expenses'

- (86) s'il réussit arrive à avoir euh des convictions des connaissances suffisantes pour que il puisse équiper entre guillemets parce que [c'est comme ça qu'on appelle]_{LL} la vente]_{HH} [chez nous]_{HH} (CRFP)

'if he manages to have uh convictions sufficient knowledge to equip quote unquote because **it's like that that we call sales here'**

In (84), *là* is taken to refer to the place where the speaker started experimenting with hobbies which is identified as *une section dans un Centre Culturel* in the preceding turn. Although no explicit contrast is established with another member of the paradigm, the cleft naturally gives rise to the reading *it is there and nowhere else*. In the same vein, the clefted *ainsi* in (85) expresses the manner in which the budget favours a number of priorities which have been previously agreed upon with citizens. The contrastive reading can be verified by inserting an additional

modifying adverb before *ainsi* such as *c'est justement ainsi qu'il privilégie de grandes priorités* without rendering the sentence agrammatical. This is only possible because contrastive *ainsi* is governed by the verb⁶⁰ as an adverb of manner (Lahousse and Lamiroy 2017). In (86), the co-referent of *comme ça* can be linked back to *équiper* and the paradigm relating to ways to refer to sales from which it is extracted. Syntactically, all three examples given below accept negation and adverb insertion in their matrix clause and their copula shows a certain morphological flexibility with regard to tense which is not found in discursive uses of the same clefts. Interestingly, all three clefted adverbs in (84)–(87) bear no rightward major prosodic boundary, which makes them type 3a/3b clefts and which, if the opposite is true for the same adverbs in discursive clefts, would constitute a defining property for contrastive ones. When coding contrast, *c'est ainsi que-*, *c'est là que-* and *c'est comme ça que-* clefts can therefore be considered to be 'canonical' clefts with the semantic and pragmatic properties which are generally associated with *c'est*-clefts.

By contrast, in their second broader sense, the use of *là*, *ainsi* and *comme ça* is non-spatial, non-contrastive and non-paradigmatic. Instead, the three adverbs have the function of conjuncts which act as discourse markers serving contextual purposes at a higher textual level. These clefts are referred to as discursive or cohesive *c'est*-clefts (Roubaud and Sabio 2015, 2018; Lahousse and Lamiroy 2018). As anaphoric clefts with a contextual use, they tend to net in a much larger portion of the preceding discourse as their antecedent which is shared as some sort of discourse elaboration leading to the information shared in the cleft. From a

⁶⁰ When characterising the properties of the clefted adverbs in relation to the verb, Sabio & Benzitoun (2013), Roubaud & Sabio (2015, 2018) and Lahousse & Lamiroy (2017) all adopt the pronominal approach developed by Blanche-Benveniste et al. (1984) and Van den Eynde (1995). In this approach, clefts are treated as constructions in which the different constituents governed by the verb are arranged in such a way that one of them, i.e. the clefted constituent, is treated as more prominent than the others. When *ainsi*, *là* and *comme* express spatio-temporal or manner meaning, the corresponding verb is said to have a normal rection and therefore exhibits a number of properties, e.g. possibility of a co-referential relation between constituent and pro-form, possibility of negation and adverb insertion in the matrix clause, etc., which are not observed in clefts with a discursive use of adverbs *ainsi*, *là* and *comme ça*. In the latter, neither paradigmatic adverbs nor negation can be introduced in the matrix clause.

conversational viewpoint, it makes them a particularly useful device for the structuring of long conversation turns or for the ending of ‘conversational episodes’ (De Stefani 2008: 714). Unlike the first subtype, the discursive class of anaphoric clefts only accepts negation or adverb insertion in their cleft-relative clause. While they share the same morphosyntactic properties, the three types of clefts differ in their discourse functions.

C'est là que-clefts are used to code information in the cleft-relative clause as particularly salient either as part of a storytelling segment or when the speaker’s stance is introduced (Roubaud and Sabio 2015). The former is shown in (87) in which the speaker is recounting his exploration of Africa while the latter is illustrated in (88) in which the speaker is signalling the moment after which s/he understood why a certain piece of art the Louvre was of interest to him/her. In both cases, there is a major prosodic boundary (HH) after the clefted adverb thus making (87) and (88) type 2b clefts and thus signalling that more information is to come.

- (87) euh ceci étant terminé nous sommes partis pour Mopti qui est un un port un sur le fleuve euh dont je me souviens plus du nom et euh qui est assez original euh là nous avons euh pas eu de problèmes particuliers euh nous avons visé-- visité ce que nous avons à faire nous étions dans un bon euh campement nous nous entendions bien avec les deux les deux anglaises et [*c'est là*]_{HH} [*qu'on a décidé de se séparer de du Dogon*]_{HH} parce que en fait ça faisait quand même quelques jours qu'on le trimbalait (CRFP)

‘uh with that being over we left for Mopti which is a a port a on the river uh of which I don’t remember the name and uh which is quite original uh there we uh didn’t have any particular issues uh we visited what we had to do we were at a good uh campsite we got along well with the two the two English girls and **it’s then that we decided to part ways with the Dogon** because actually we had been carting him around for several days’

- (88) alors bon euh il me téléphone dans l’après-midi en fin d’après-midi et il me dit monsieur NNAAMMEE euh euh votre pièce elle est prête vous pouvez venir la chercher et j’ai regardé ma montre écoutez dix-sept

heures précises dix-sept boum c'était la Vierge et euh qui se manifestait à sa façon et alors [c'est là]_{HH} [où j'ai j'ai compris]_{HH} [pourquoi]_{h+h+} parce que ce ce bas-relief contient ma date de naissance (CRFP)

'well so uh he calls me in the afternoon in the late afternoon and he tells me Sir NNAAMMEE uh uh your ID it's ready you can come pick it up and I look at my watch listen five o'clock on the dot bam it was the Virgin Mary and uh who manifested herself in her own manner and so **it's then that I understood why because this this bas-relief contains my date of birth**'

C'est comme ça que-sentences are comparable to recapitulative devices which introduce a consequence or an illustration. Similarly to *c'est là que*-clefts, *c'est comme ça que*-clefts may, in some cases, mark the information shared in the cleft-relative clause as particularly salient (Roubaud and Sabio 2018). To illustrate this, I use two examples from the CRFP which are also analysed in Roubaud and Sabio (2018). In (89), the speaker is narrating his/her life which consists of a series of events leading up to him/her moving to Amiens. This is presented as being the most significant climactic step of the process. (90) exemplifies the illustrative meaning of *comme ça* with the Pinot Noir being used as an example, as shown by the locution *par exemple*, of what product the wine-making process described at length in the preceding context is used for. Note here that the clefted adverb in (89) bears a major continuative prosodic boundary to its right, but that the one in (90) is lumped together with part of the cleft-relative clause. (89) corresponds to type 2a in my prosody-based typology and (90) to 2b. It therefore appears that cohesive *c'est comme ça que*-clefts are not necessarily restricted to one unique prosodic pattern.

- (89) j'avais perdu mon père à douze ans et je ne connaissais pas tellement la fabrication ma mère a fait tout ce qu'elle a pu mais eh eh elle était pas du métier bon alors j'ai cherché à ayant deux frè- deux frères et une sœur à leur laisser la place pour t avoir une profession et moi me perfectionner dans mon métier et [c'est comme ça]_{HH} [que je suis

venu à Amiens]_{L-L}- en dix-neuf cent vingt-huit]_{HH} [après mon service militaire]_{L-L}. (CRFP)

'I had lost my father at twelve years old and I didn't really know manufacturing my mother did everything she could but eh eh she wasn't in the field so well then I tried to having two bro- two brothers and a sister give them the opportunity to t have a job and for me to get better in my job and **it's how I came to Amiens in nineteen twenty-eight after my military service'**

- (90) on avait aussi dans notre région l'abbaye de Cîteaux les moines de Cîteaux qui se sont beaucoup beaucoup a-- attachés là aussi au vin pourquoi parce que le vin rouge c'est une symbolique très forte dans l'église chrétienne le vin rouge c'est le sang du Christ or notre Pinot Noir notre cépage qui donne de bons arômes c' est un raisin rouge à jus blanc voyez quand on presse comme ça on prend une grappe on la serre tout de suite on la presse le jus qui coule c'est un vin blanc **[c'est comme ça que le champagne]_{HH} [par exemple]_{h+h+} [est fait avec du Pinot Noir]_{HH}** (CRFP)

'we also had on our region the Cîteaux monks' abbey the Cîteaux monks who grew very very a-- attached there too to wine why because red wine has a strong symbolism in the Christian Church red wine is Christ's blood or our Pinot Noir our grape variety that gives good flavour it's red grapes with white juice you see when we press like that we take one grape we squeeze it straight away we squeeze it the juice that flows out it's white wine **it's how champagne for instance is made with Pinot Noir'**

Finally, *c'est ainsi que*-clefts express logical continuity by taking on either a consequential or illustrative meaning and by rendering the clefted adverb topical rather than focal (Lahousse and Lamiroy 2017). The first meaning is exemplified in (91) in which the fact that the speaker navigated towards Carnac with his children is introduced as the logical consequence of his/her many learning steps in sailing which are enumerated beforehand.

- (91) un autre sport qui touche également euh à l'air et au vent euh c'est la voile sur l'eau j'ai commencé par équiper une petite barque d'une voile que j'ai coupée dans un drap et j'ai pratiqué sur des plans d'eau euh mes premières navigations à la voile et tout ça a évolué du petit bateau équipé d'un drap je suis passé au dériveur du dériveur je suis passé au bateau ponté et [*c'est ainsi*]_{HH} [*que avec plusieurs de mes enfants*]_{HH} [*j' ai navigué euh à Carnac*]_{HH} (CRFP)

'another sport that is also linked to uh the air and wind uh it's windsurfing I started by equipping a small boat with a sail that I cut from a sheet and I practiced in bodies of water uh my first sailing experiences and all that it's evolved from the small with a sail I switched to sailing dinghy's from sailing dinghy's I switched to bay liners and it's then that with several of my children I sailed uh in Carnac'

Prosodically, the adverb in (91) and the matrix clause it is part of are realised as part of the broad focus extending over the whole cleft. The illustrative use of *ainsi* was not found in either of 4 occurrences of *c'est ainsi que*-clefts. Among the 5 tokens, 1 is of the contrastive type while the remaining 4 function as discourse markers.

The analysis of the frequency of use of the different subtypes of clefts summarised in Table 16 below reveals that both uses, i.e. contrastive and discursive, are attested in the CRFP dataset, with a preference for the discursive one for all three types clefts. This is consistent with Lahousse and Lamiroy's (2017) results on written and spoken French in which contrastive clefts amount to 40% of all tokens and discursive to 52%. The remaining 8% encompass all cases displaying a double reading. My findings, however, contrast with that of Roubaud and Sabio (2015) who uncovered a perfect balance of the two uses in the spoken category of their dataset.

	Use	Tokens <i>n</i> (%)	Total
<i>c'est là que</i>	contrastive	13 (37.1%)	35
	discursive	22 (62.9%)	
<i>c'est ainsi que</i>	contrastive	1 (20%)	5
	discursive	4 (80%)	
<i>c'est comme ça que</i>	contrastive	8 (44.4%)	18
	discursive	10 (55.6%)	
Total		58 (100%)	58

Table 16. Frequency of use of contrastive and discursive *là*, *ainsi* and *comme ça*

With regard to prosody, 15 tokens of contrastive clefts are realised as type 3a/3b clefts with a single intonation unit and no prosodic boundary between the clefted constituent and the cleft-relative clause while the remaining 7 bear a major prosodic boundary after the adverb and hence illustrate types 2a/2b. With regard to the narrow vs. broad focus distinction, all contrastive clefts are associated with broad focus. Discursive clefts also mainly exemplify types 3a/3b and 2a/2b but with the opposite distribution. Thus, 27 of them have a major boundary after the clefted constituent and 9 of them none between the two parts of the cleft. Once again, no cases of narrow focus are found. Overall, neither contrastive or discursive clefts have a focused clefted constituent, thus showing that it is typically the information conveyed in the cleft as a whole that is presented as informationally salient. All of this shows that although the two subtypes of anaphoric clefts tend to favour one realisation over the other, there is no clear-cut nor systematic prosodically-based distinction to be made between the two. Instead, it is more appropriate to assign one unmarked, i.e. most frequent, and one marked pattern, i.e. less frequent, to each of the subtype. This can be done as follows:

- (92) contrastive: *unmarked* [clefted constituent + CRC]_{MAJOR BOUNDARY}
marked [clefted constituent]_{MAJOR BOUNDARY} [CRC]_{MAJOR BOUNDARY}

- (93) discursive: *unmarked* [clefted constituent]_{MAJOR BOUNDARY} [CRC]_{MAJOR BOUNDARY}
marked [clefted constituent + CRC]_{MAJOR BOUNDARY}

For Lehmann (2008) and Lahousse and Lamiroy (2017), when used as a cohesive structure, *c'est*-clefts like the ones investigated in this section do not function as bipartite constructions⁶¹ but rather as simple sentences which are introduced by a connective marker such as *c'est là que*, *c'est ainsi que*, or *c'est comme ça que* (Lahousse and Lamiroy 2017: 180). The information that is introduced after the adverb, i.e. in the cleft-relative clause, therefore carries the biggest informative load while the adverb is merely used as a connective which links two segments of discourse.

As a result, the adverbs *là* and *ainsi* in cohesive clefts have been argued to display signs of lexicalisation (De Stefani 2008; Lahousse and Lamiroy 2017), and also of grammaticalisation (Lahousse and Lamiroy 2017). The grammaticalisation hypothesis for *ainsi* is supported by the co-existence in the CRFP dataset of the different meanings assigned to it, which shows that the process is still ongoing (Lahousse and Lamiroy 2017).

The same argument of grammaticalisation has been made for the cleft structure itself based on the weakening of its focalising function, and hence of its focus-background articulation (De Stefani 2008; Lehmann 2008; Dufter 2009b; Lahousse and Lamiroy 2017). This is corroborated by the prevalence of broad focus which attaches to the whole cleft rather than the clefted constituent. It is also supported by the relative discourse-newness of the information shared in the cleft-relative clause of such clefts - only 11 tokens out of 58 carry given information in their cleft-relative clause, out of which only 1 introduced explicitly evoked information - thus making the majority of them informative-presupposition clefts in Prince's (1978) terms. These results are in line with Lehmann's (2008: 212-213) argument against the 'incipient grammaticalisation' of cleft constructions, that is, grammaticalisation of the construction through the individual grammaticalisation of its constituents themselves. In other words, because information structure is

⁶¹ With this, Lehmann (2008) and Lahousse and Lamiroy (2017) mean that the bi-clausal structure of clefts, which is said to be brought about by the repackaging of information into an argument focus structure, appears to lose its very functional motivation. What is left then is "the simple topic-comment clause, where the topic is highly activated, i.e. represented by a clitic pronoun" (Lehmann 2008: 227).

relational, i.e. it always functions in relation to the components involved such as assertion vs. presupposition, only specific information structural articulations, and not information structural components, can become grammaticalised. In the case of given/new-new clefts, i.e. Prince's (1978) informative-presupposition clefts, in which the presupposition is not shared in the prior discourse, the hearer is instructed to pragmatically accommodate said presupposition. When this process becomes conventionalised in the construction, then the construction becomes more grammaticalised. As a result, the very function of the *c'est*-cleft, i.e. the coding of contrastive focus, is gradually lost. This is what Lehmann (2008: 213) refers to as levelling of focality. He also uses the example of the loss of agreement between copula *être* and clefted constituent to show a process of grammaticalisation but as has been shown in Section 4.1.1, this is to be nuanced.

4.5 Functional description of *c'est*-clefts

From a functional perspective, *c'est*-clefts assume a number of functions which range from focus and contrastive marking (É. Kiss 1998; Sornicola 2011; Scappini 2013) at the syntactic and semantic levels to hedging (Prince 1978), cohesion (Roubaud and Sabio 2015, 2018; Lahousse and Lamiroy 2017), floor holding (De Stefani 2008), and topic (re)launching at the textual one (Garassino 2014). Building on the information structure-based typology proposed in Section 4.4.2, I detail the discourse motivations of new-given, new-new, given-new and given-given *c'est*-clefts.

To start with, new-given clefts are particularly suitable for the expression of contrastive focus, as defined by É. Kiss (1998), which opposes the value of the cleft to other paradigmatic alternatives. In (94), for instance, the speaker is detailing her family history. In mentioning that her father is of Spanish origins, she creates the expectation that the Spanish roots come either from her father's own father, his mother, or from both, which renders the variable of the cleft *x a les origines* easily inferable. The value satisfying the variable is identified as *sa mère* which is then contrasted with *son père* in the immediate rightward co-text. Through the cleft, the mother is specified as the one and only value which is confirmed with the

father being dismissed immediately after. (94) not only highlights the contrastive reading induced by clefting but also the exhaustivity implicature it triggers.

- (94) non euh mon père est pied-noir donc il est il est venu en France en m-mille neuf cent soixante-deux je crois donc lui est d'origine espagnole c'est de c'est donc **[c'est sa mère]_{L-L} [qui a les origines]_{I-I}** son père est un pied-noir français (CRFP)

'no uh ma father is Pied-Noir so he he came to France in ni~ nineteen sixty-two I believe so he is of Spanish origins it's from **it's therefore his mother that has the origins** his father is a French Pied-Noir'

- (95) A: donc c' est hum le comptage du le ni- comment dire ça le léchage euh le hum quand le l'animal vient euh pff boire donc de compter le nombre de fois où il vient boire (...)

B: hum hum c'est s~ ouais c'est pa-- **[c'est par cellule photo-électrique]_{H/H} [ou euh par contact]_{HL} [que ça se produit]_{I-I}**

A: ben euh là c'est plus par contact la cellule photo-électrique je pense pas vraiment que c'est euh comme les les biberons

'A: so it's hm the counting of the ~ how can I put this the licking uh the hm when the the animal comes uh pff to drink so to count the number of times that it comes to drink (...)

B: hm hm it's s~ yeah it's b~ **it's with a photoelectric sensor or by contact that it happens**

A: well uh there it's more by contact the photoelectric sensor I don't really think that it's uh like bottles'

Given the way contrast is discursively built in (94), it is worth noting that it need not necessarily be established with a constituent from the preceding context. In some cases, like in (94), it may also rely on semantic connections created with referents appearing in the following discourse. From a methodological viewpoint, clefts should therefore always be analysed in relation to context on both sides, and even more so with a predictive approach like the one adopted in this study. As Scappini (2013) argues, contrast in *c'est*-clefts may take different forms and may be more or less explicit. In the case of (94), the two alternatives are introduced, of

which only one is validated by the speaker as fulfilling the given variable. In (95), however, speaker B presents speaker A with two potential values and is here requested to clarify which one is accurate, as shown by the interrogative boundary tone H/H. The presupposed open preposition is understood to refer to the fact that researchers count the number of times animals drink from the apparatus provided to them. The semantic gap is therefore to be filled by a description of the mechanism involved, either a photoelectric sensor or a contact mechanism. The value is identified by speaker A in the following turn by means of a reduced cleft selecting *par contact* as the correct value. This shows that combinations of full and reduced clefts are particularly useful for question-answer pairs.

Clefts of the new-new type typically introduce a new topic in discourse as in (96), which means that they can be used as discourse openers or as introductory sequences.

- (96) il y a euh beaucoup d'événements qui peuvent ponctuer la vie d'une d'une personne d'une femme en-- entre autres en ce qui me concerne euh je me suis mariée le le quatre septembre de cette année dix-neuf cent quatre-vingt-dix-neuf et euh [**c'est euh cette journée_{LL} et ce qui l'a précédée_{LL} le voyage de noce qui qui a suivi**]_{HH} [**dont je dont je vais parler aujourd'hui**]_{HH} alors pourquoi le pourquoi le mariage souvent euh on entend dire que le mariage est le plus beau jour de la vie euh souvent d'une femme c'est souvent les femmes qui tiennent ce genre de discours (CRFP)

'there are many events that can intersperse the life of a a person of a woman a~ for instance as far as I'm concerned uh I got married on on September fourth that year of ninety ninety-nine and uh **it's that day and what preceded it the honeymoon that followed that I'm going to talk about today** so why weddings we often hear people say that a wedding is the most beautiful day in life uh usually of a woman it's usually women who say this kind of thing'

The excerpt in (96) corresponds to the beginning of a private recording - PRI in the taxonomy of CRFP texts - in which the speaker is asked to talk about her life without

any specific prompt. The cleft does not constitute the very opening of the excerpt but functions, along with the preceding clause, as the introductory sequence of the conversation. This is shown by the information conveyed in the cleft-relative clause which is the speaker is going to talk about the day which she has just mentioned, i.e. her wedding day. While some of the referents of the clefted constituent, e.g. *cette journée*, are co-referential with items previously given, e.g. *le quatre septembre de cette année dix-neuf cent quatre-vingt-dix-neuf*, new information interpretable wholly out of context is still added, which makes it new anchored in my model of discourse-familiarity. The discourse topic is thus refined by the cleft with the addition of material that had not yet been introduced in discourse. The broad focus assigned to the cleft marks the information conveyed in both of its clauses as equally informationally salient. As a whole, the cleft in (96) exemplifies Hasselgård's (2004) and Garassino's (2014) idea of 'topic launching' whereby the topics of weddings and honeymoons are explicitly made available as topics for the speaker for the continuation of discourse.

Given-new clefts in the CRFP data are mainly characterised by the selection of an anaphoric item as their clefted constituent and exhibit two main discourse functions as shown in the previous section. Like new-given clefts, given-new clefts may encode contrast in the postcopular position but also, in some cases, in the cleft-relative clause. This is the case in (97) in which a double contrast is set up between the speaker and her sister and their mother, on the one hand, and between producing children and babysitting children on the other. As such, what is opposed is the speaker and her sister making children and their mother babysitting them. The cleft once again sets the scene for the contrast that is established with elements present in the following co-text.

- (97) A: et ma soeur elle a deux enfants aussi un de huit ans et l'autre de quatre ans
B: hum hum donc votre mère est bien occupée
A: oui elle est bien occupée avec les petits-enfants [**c'est nous qui faisons les enfants**]_{HH} elle les garde non mais ça se passe bien on a on a une maman adorable (CRFP)

'A: and my sister has two children too one who's eight years old and the other one four years old

B: hm hm so your mother is very busy

A: yes she's very busy with the grandchildren **it's us who make children** she babysits them no but it's going well we have we have an adorable mum'

The second use available for given-new is that of a cohesive marker. As shown in Section 4.4.3, cohesive clefts such as (98) typically occur within a storytelling segment. In this case, the speaker is recounting a story of him being bothered by a bunch of people wanting to steal his money. The speaker elaborates on what happened by means of an enumeration of actions. The cleft, as a cohesive marker, allows the speaker to signal that the information it conveys is particularly salient marking it a turning point in his story.

- (98) donc ils sont venus ils ont commencé à me brancher à chercher la bagarre il y en a un qui a essayé de me tirer un coup de poing que j' ai esquivé heureusement de justesse et euh donc ils voulaient me tirer du fric tout ça enfin bon je leur ai expliqué que de toute façon en ce moment c'était pas le moment de me faire chier parce que j'étais au chômage arrivais pas joindre les les les deux bouts alors euh il y en a un qui m'a demandé mais comment tu fais pour vivre alors j'ai dit ben comme tous comme tous les chômeurs hein les ASSEDIC bon alors que c'est pas tout à fait vrai parce que même les ASSEDIC euh ça fait un moment que je les avais pas eues et je vis justement euh par l'intérim sinon euh donc voilà donc après bon ils m'ont lâché quand ils ont quand ils ont compris que que ça irait pas plus loin après donc j'ai pris le train et alors là [**c'est là que j'ai commencé à flipper ma race**]_{HH} parce que donc je me suis installé dans le train tranquille (CRFP)
- 'so they came they started messing with my looking for a fight there's one that tried to punch which I dodged narrowly thankfully and uh they wanted to steal some of my money and all well I explained to them that anyway these days were not the best time to mess with me

because I unemployed could not make ends meet so uh one of them asked me but how do you go by then I said well like every unemployed person uh the ASSEDIC (note: type of government aid) I mean that's not even really true because even the ASSEDIC I hadn't received in a while and I actually live on uh temping otherwise uh so there you go so after well they gave up on me when they when they understood that that it wouldn't any further afterwards so I took the train and well then **it's then that I started freaking out** because so I got on the train'

Finally, clefts in which both value and variable are discourse-given may serve as a concluding device marking the end of a discourse segment. Like cohesive clefts, their use is therefore less localised and instead pertains to a bigger chunk of discourse. This is illustrated in (71), repeated as (99) below, in which the speaker elaborates on the fact that he finds biology a difficult subject for a number of reasons. The speaker ends the segment by putting forth the reason that makes it such a hard subject and marks it as the most salient through the insertion of the modal adverb *surtout*.

- (99) A: la biologie je bloque sur la biologie j'ai une formation littéraire donc euh la biologie t~ enfin quoique cette année ils ont changé un petit peu le programme c'est un peu plus léger c'est un peu plus accessible quoi mais l'année dernière c'était vraiment enfin lourd moi j'avais beaucoup de mal à tout tout ce qui est anatomie du cerveau tout ça j'a j'avais vachement du mal (...) mais euh moi j'avais moi j'avais du vraiment du mal à visualiser en fait euh le fonctionnement de de des muscles par rapport au cerveau tout ça j' a- j' arrivais pas trop à suivre le circuit quoi et puis ça m'intéresse pas en fait trop trop donc **[c'est surtout ça]_{L-L} [qui est difficile]_{I-I}**

B: ouais ouais hum hum et quel sport euh vous pratiquez (CRFP)

'A: biology I go blank on biology I have a literary training so uh biology t~ I mean this year they changed the programme a little it's lighter it's more accessible do you know what I mean but last it was very well heavy I had a lot of trouble with everything everything related to the

brain's anatomy all of that (...) uh I had I had a lot of trouble visualising well uh how muscles function compared to the brain and everything I cou~ I couldn't really followed the circuit and I wasn't really interested in it actually not really not really so **it's especially that that's difficult**
B: yeah yeah hm hm and which sport uh were you doing'

Within the interactional context, the prosodic sequence of assertive tone L-L- and appendix contour l-l- signals what Couper-Kuhlen and Ono (2007) call a transition relevance place whereby the speaker offers the conversational floor to the hearer. Here, speaker B uses this opportunity to move on to the next question of the interview.

The qualitative analysis of the CRFP dataset corroborates the diversity of discourse functions sourcing the use of *c'est*-clefts identified in the literature. It also further illustrates the extent of the interplay between the different layers of their information structure.

4.6 Concluding remarks

In this chapter, I aimed at refining the description of French *c'est*-clefts by providing a quantitative-qualitative account of their morphosyntactic, semantic, prosodic and information structural characteristics as observed in the Corpus de Référence du Français Parlé.

Starting with their morphosyntax, the findings have uncovered a limited flexibility in the tense and person agreement in the copula of *c'est*-clefts which is consistent with the claims on the gradual morphological fixation of the French cleft. By contrast, the form and function of the clefted constituent display a lot more variation, with a preference for NPs and PPs, and subjects and adjuncts. *C'est*-clefts also typically accept the insertion of a wide range of modal markers which may reinforce or clash with the specificational semantics and pragmatic features.

The analysis of the information structure of *c'est*-clefts was carried out by treating the relational, i.e. focal vs. non-focal, and referential, i.e. new vs. given, aspects separately. Prosodically, *c'est*-clefts can be divided into five main categories. To better account for the different syntactic and illocutionary forms of

c'est-clefts, reduced clefts and interrogatives were placed within their own categories. Overall, *c'est*-clefts were found to mainly instantiate broad focus whereby the whole cleft is in focus and hence informationally salient. Occurrences in which the focus is narrow and bears on the clefted constituent turned out to represent a minority of tokens. As such, my findings have uncovered an asymmetry between the informational highlighting triggered by the bi-clausal syntax and the prosodically coded function of focus. I have also shown that the prosodic profiles of *c'est*-clefts may be influenced by interactional aspects such as the management of the conversational floor. Related to this is the use of the initial accent which was found to serve pragmatic and discursive purposes.

As far as the distribution of discourse-given and discourse-new information is concerned, the empirical findings led to the proposal of a new quaternary typology which adds a fourth layer to the three already discussed in the existing typologies (e.g. Rialland et al. 2002; Doetjes et al. 2004; Mertens 2012). Thus, *c'est*-clefts may be of the new-given, new-new, given-new and given-given types. The additional given-given pattern was found to be further subdivided into two classes of clefts depending on whether the specification relation itself was new or given.

On the question of the interaction between the different layers of information structure, my results provided evidence in favour of the separate treatment of the phenomena of givenness/newness and prosodic focus. More precisely, given information is not systematically deaccented and new information is likewise not always in focus, which is in line with most descriptions of *c'est*-clefts. As a result, each discourse-familiarity type of *c'est*-clefts exhibits an array of prosodic articulations. The relational and referential information structures of *c'est*-clefts are thus best described as distinct but interacting layers.

C'est-clefts were moreover found to illustrate an array of discourse functions which are not limited to that of focalisation despite what is predicted by a number of accounts. The discursive functions of *c'est*-clefts include hedging, cohesion, floor holding and topic (re)launching. The analysis sheds particular light on a subclass of anaphoric *c'est*-clefts serving structural and cohesive purposes in which focality is 'levelled out' (Lehmann 2008) and contrast absent. As such, the existence of these clefts supports the claim of grammaticalisation of information put forth by a number

of authors (see Lehmann 2008; Lahousse and Lamiroy 2017) and illustrate a process of specialisation for the subcategory.

Chapter 5: Comparison of *it*-clefts and *c'est*-clefts

The goal of this chapter is to provide a cross-linguistic comparison of the English *it*-cleft and the French *c'est*-cleft by building on the findings presented in Chapters 3 and 4. Starting from the postulate that *c'est*-clefts resemble *it*-clefts syntactically but less so pragmatically (Bourns 2014), I reflect on the morphosyntactic, prosodic and information structural similarities and differences observed between clefts in the two languages. In Section 5.1, I compare the morphosyntactic properties of *it*- and *c'est*-clefts and discuss their respective frequency of use. I then focus on the operationalisation of their relational and referential information structures in Section 5.2 before moving to the description of their pragmatic features in Section 5.3. I conclude this contrastive chapter by summarising the main cross-linguistic findings in Section 5.4.

5.1. Morphosyntactic properties

I start by examining a number of aspects relating to the morphosyntax of *it*-clefts in comparison with that of *c'est*-clefts. These include their respective frequencies of use in Section 5.1.1, their structural characteristics in Section 5.1.2, and the form and function of their clefted constituent in Section 5.1.3.

5.1.1 Frequency of occurrence

As a first step to the characterisation of the use of *it*- vs. *c'est*-clefts, I start by comparing their frequencies of use in the LLC-1 and CRFP. A number of studies (e.g. Carter-Thomas 2002, 2009; Bourns 2014; Bourgoïn 2017) have shown that *c'est*-clefts are typically more frequent than their English counterpart *it*-clefts. In her study investigating the use of clefts in research articles, Bourgoïn (2017) for instance found the frequency of *c'est*-clefts to be twice as high as that of *it*-clefts. Similarly, Carter-Thomas's (2002) study revealed that only half of the *c'est*-clefts under review were translated into clefts in the English translated texts. For Carter-Thomas (2002, 2009) and Bourns (2014), this statistical difference is primarily due to the underlying pragmatic and prosodic properties of both languages. The raw and

normalised frequencies of *it*- and *c'est*-clefts in the LLC-1 and CRFP are presented in Table 1 below. As a reminder, the LLC-1 corpus consists of 500,000 words and the CRFP of 440,000 words. Despite the smaller size of the latter, I consider the two corpora to present acceptable comparability as discussed in Section 2.1.3.

	Full clefts		Reduced clefts		Total of clefts	
	RawF	NormF	RawF	NormF	RawF	NormF
LLC-1	143	2.86	95	1.9	238	4.76
CRFP	392	8.9	88	2	480	10.9

Table 1. Raw and normalised (per 10,000 words) frequencies of *it*- and *c'est*-clefts

In the CRFP, *c'est*-clefts have a normalised frequency of 10.9 per 10,000 words while *it*-clefts number 4.76 per 10,000 words in the LLC-1. When it comes to spontaneous spoken interaction, *c'est*-clefts are therefore twice as frequent as *it*-clefts, which is consistent with results found for written data. Both *it*- and *c'est*-clefts have two variants, full and reduced, illustrated in (1)–(4).

- (1) it's Derby that take up the count (LLC-1)
- (2) you know the Abbey National's not to blame in the least it's the surveyor and the person who did the structural [~~who's to blame~~] (LLC-1)
- (3) on dira que c'est le grand-père qui a fourni l'occasion de s'établir (CRFP)
'let's say that it's the grandfather that gave the opportunity to settle down'
- (4) alors troisième chose sur laquelle on travaille mais cela c'est pas moi [~~qui travaille dessus~~] (CRFP)
'so the third thing we are working on but this it's not me who's working on it'

In both languages, full clefts like (1) and (3) are more frequent than reduced clefts like (2) and (4). However, the cross-linguistic comparison shows that the higher frequency of the full form is more strongly marked in French than in English. The empirical investigation thus highlights two tendencies: a higher frequency of French

c'est-clefts and a higher proportion of reduced clefts in English. The comparable nature of my data does not allow me to make claims about the type of contexts in which clefts are more likely to occur. In order to further characterise the differences and similarities between *it*- and *c'est*-clefts, I move on to their morphosyntactic properties in the following section.

5.1.2 Morphology of the copula and paradigm of relative markers

It- and *c'est*-clefts are generally taken to be equivalent constructions (Jespersen [1937] 1984; Lambrecht 2001; Carter-Thomas 2002; Dufter 2009a; Bourns 2014) built with the same syntactical structure in their declarative form consisting of subject *it/ce*, copula *be/être*, postcopular complement and cleft-relative clause. However, while the morphology of the copula *be* is fixed with regard to person, that of *être* displays some number variation. For example, a plural NP value such as *philologists* in (5) does not trigger any number agreement with the copula *be*. By contrast, the copula form *sont* in (6) shows third-person plural agreement with the plural value *les labos pharmaceutiques*.

- (5) *it's* not philologists you want to convince (LLC-1)
- (6) *ce sont* les labos pharmaceutiques qui financent (CRFP)
'it's the pharmaceutical labs that provide funding'
- (7) *c'est* eux qui gèrent l'armée (CRFP)
'it's them who manage the army'

While it may be tempting to assume that the category of French clefts as a whole is characterised by a greater variation in the morphology of the copula, it should be noted that person agreement has been found to gradually decline starting from Middle French onwards (Dufter 2008: 4). Moreover, plural copulas are not systematically required, as illustrated in (7). In fact, the plural form *sont*, which is the only variant still accepted in Modern French, only occurs in 2.7% of all tokens of *c'est*-clefts (see Section 4.1.1). Instances such as (6) can therefore be considered to be remnants of historical structural variance (Dufter 2008) which merely constitute marginal cases of *c'est*-clefts in Contemporary French.

With regard to tense, both *it*- and *c'est*-cleft may have a past-tense copula, as is the case in (8) and (9).

- (8) it **wasn't** Kilmarnock who who came to see you actually (LLC-1)
- (9) *c'était pas moi qui l'avais fait ce projet* (CRFP)
'it wasn't me who did this project'

Past-tense copulas are found in 32% of all of tokens of *it*-clefts and in 5% of all occurrences of *c'est*-clefts. There is therefore a stronger tendency for past tense in English clefts than in French clefts.

Another difference between the two languages relates to the type of relative markers introducing the cleft-relative clause. On this cross-linguistic aspect, Dufter (2008) notes that the matrix clause and cleft-relative clause of the *c'est*-clefts are typically linked by means of the *que/qui* relativiser while English selects a wider range of relative markers. This claim is largely corroborated by the LLC-1 and CRFP datasets. English clefts may select *that*, *which*, *who*, *where* and \emptyset (with both subject and complement function - see Section 3.1.3), all of which are illustrated in (10)–(13) below.

- (10) it's Fire Raiser **who's** coming up to join Carbon (LLC-1)
- (11) it's the grammar **which** is interesting (LLC-1)
- (12) it's the grammar **where** the fun is (LLC-1)
- (13) may have been Ivor Bond \emptyset told me (LLC-1)

French, on the other hand, heavily relies on the use of *que/qui*, with only 10 occurrences diverging from this tendency and selecting other relativisers such as *où*, *dont* or \emptyset (see Section 4.1.2).

- (14) *c'est au-- à ce moment-là que ça craint le plus* (CRFP)
'it's a~ at that moment that it sucks the worst'
- (15) *c'est ça moi qui me dérange* (CRFP)
'it's that that bothers me'
- (16) *c'est là où le débat politique actuellement est passionnant* (CRFP)

- 'it's there that the political debate is currently fascinating'
- (17) *c'est euh cette journée et ce qui l'a précédée le voyage de noce qui qui a suivi dont je dont je vais parler aujourd'hui* (CRFP)
 'it's that day and what preceded it the honeymoon that followed that I'm going to talk about today'
- (18) *c'est pour ça ø je te disais c'est vraiment euh ma grosse heure* (CRFP)
 'it's for that reason I was telling you it's really my busiest hour'

For Dufter (2008), the variety of relativisers in English, on the one hand, and the omnipresence of *que/qui*, or at least its overwhelming use, in French, on the other, can be treated as a sign of a higher degree grammaticalisation of *c'est*-clefts (Dufter 2008: 5). That *c'est*-clefts are becoming structurally invariable is mostly supported by the findings on both relativisers and the copula, but it should nonetheless be stressed that this study deals with synchronic data and therefore only provides a still picture of the use of *it*- and *c'est*-clefts.

5.1.3 Form and function of the clefted constituent

I have shown in Sections 3.1 and 4.1 that the range of syntactic categories and grammatical functions available for the clefted constituent was varied in both English and French. Starting with the grammatical category of the postcopular complement, *it*- and *c'est*-clefts may select constituents which include NPs, PPs, AdvPs, and so forth. The distribution of each category for *it*- and *c'est*-clefts is summarised in Table 2 below. Full and reduced clefts are merged together.

	<i>It</i> -clefts	<i>C'est</i> -clefts
Noun phrase	177 (74.4%)	259 (54%)
Prepositional phrase	26 (10.9%)	121 (25.2%)
Adverbial phrase	6 (2.5%)	68 (14.4%)
Finite clause	14 (5.9%)	25 (5.2%)
<i>WH</i> -word	13 (5.5%)	—
Adjectival phrase	2 (0.8%)	—
Non-finite clause	—	6 (1.2%)
Total	238 (100%)	480 (100%)

Table 2. Distribution of syntactic categories of the clefted constituent in *it*- and *c'est*-clefts

At first glance, the overall order of the different categories appears to be similar in both languages. NPs are the most frequent syntactic class before PPs, AdvPs and finite clauses. In addition to these four types, *it*-clefts also select *wh*-words and AdjPs, which do not occur in *c'est*-clefts. Likewise, non-finite clauses are only found in *c'est*-clefts. However, looking at the frequencies of each category independently brings out a number of cross-linguistic differences. First, the preference for NPs is more marked in English than French - with 74.4% of them in the former and 54% in the latter.⁶² This gives rise to a higher frequency for both PPs and AdvPs in French, which remain more marginal in English. That NPs, PPs and AdvPs occur more often in French can be elucidated by examining the use of anaphoric clefts, which are particularly frequent in French and which can be divided into two subtypes.

In the first subtype, the clefted constituent is a personal pronoun referring either to one of the speech participants, as in (19), or to a third-party, as in (20). These clefts number 112 in the CRFP, which represents 23% of all tokens. Only 8 clefts of the kind are found in English, e.g. (21) and (22), which only amounts to 3% of all *it*-clefts. The prominent ability for pronouns to be clefted in French is explained by Bourns (2014) in terms of prosody. For her, pronouns in subject positions, such as *moi* in (19) and *elle* in (20), cannot be accented the same way

⁶² A Chi-Square test was run to test the significance of the difference in distribution of NPs, which yielded a *p*-value of *p*=.027558.

they would be in English, which makes clefting “vital for emphasizing them” (Bourns 2004: 205).

- (19) *c'est vrai que c'est moi qui m'en suis chargée de d'une façon générale* (CRFP)
 ‘it’s true that it’s **me** who took care of it generally speaking’
- (20) *après j'ai ma collègue qui arrive à deux heures donc c'est elle qui s'occupe de la blanchisserie* (CRFP)
 ‘then I have my colleague who gets here at two o’clock so it’s **her** who takes care of laundry’
- (21) *you English call us Irish stupid it's you that are stupid you should send them all home* (LLC-1)
- (22) *you didn't tell me to sell out it was I who said I wanted to sell out* (LLC-1)

Following Bourns’s (2014) argument, the *c'est*-clefts in (19) and (20) illustrate the unmarked strategy used in French to highlight subject pronouns. By contrast, the *it*-clefts in (21) and (22) exemplify a marked type of emphasis for English pronouns, which, in this case, appears to be primarily sourced by the encoding of contrastive focus. I describe this in greater detail in Section 5.3.2.

In the second subtype of anaphoric *c'est*-clefts, the clefted constituent is either a PP or an AdvP containing anaphoric reference. These include, but are not limited to, phrases such as *pour ça*, *comme ça*, *là*, *ainsi*, etc., for which examples are provided in (23)–(25).

- (23) *c'est pour ça qu'on l'appelle en reconnaissance le Corton-Charlemagne* (CRFP)
 ‘it’s **for that reason** that we call it to pay homage Corton-Charlemagne’
- (24) *et c'est là qu'elle a arrêté l'offensive allemande de soixante-dix* (CRFP)
 ‘and it’s **there** that she stopped the German attack of ninety seventy’
- (25) *c'est comme ça que ça doit être* (CRFP)
 ‘it’s **like that** that it should be’

Like clefted pronouns, anaphoric clefted PPs and AdvPs represent a majority of the occurrences within the general subclass they are part of. Anaphoric AdvPs, for instance, subsume 92% of all AdvPs. The fact that clefts of this kind do not occur in the LLC-1 dataset allows one to account for the overall lower frequency of PP and AdvP clefted constituents in *it*-clefts.

Let me now turn to the grammatical function fulfilled by the clefted constituent in the cleft-relative clause. The distribution of the different functions is shown in Table 3 below.

	<i>It</i> -clefts	<i>C'est</i> -clefts
Subject	120 (50.4%)	228 (47.5%)
Direct object	28 (11.8%)	27 (5.6%)
Indirect Object	1 (0.4%)	4 (0.8%)
Adjunct	59 (24.8%)	158 (33%)
Conjunct	—	61 (12.7%)
Complement of a preposition	29 (12.2%)	2 (0.4%)
Subject complement	1 (0.4%)	—
Total	238 (100%)	480 (100%)

Table 3. Distribution of syntactic functions of the clefted constituent in *it*- and *c'est*-clefts

The results of the grammatical analysis shows that *it*- and *c'est*-clefts assign the same types of functions to the clefted constituent, with the exception of conjuncts, which are only attested in *c'est*-clefts, and subject complement which is only found in one *it*-cleft, but not with the same relative frequencies. Subjects constitute the most instantiated function with a similar frequency in both languages. Adjuncts which represent the second most frequent function, are almost as common as subjects in *c'est*-clefts but only half as frequent in *it*-clefts. Differences are also observed in direct objects and complements of prepositions, whose frequency of occurrence is higher in *it*-clefts, and in indirect objects which are slightly more common in *c'est*-clefts. The most striking difference therefore relates to the prevalence of adjuncts and conjuncts in French, which can directly be linked to the

high proportion of anaphoric PPs and AdvPs previously described. Tackling (23)–(25) again, adjuncts may be causal, (23), temporal, (24), or of manner (25). While different types of adjuncts also occur in English, their use remains limited. Their higher frequency in French is in line with observations made by a number of authors (see Nølke 1983; Doetjes et al. 2004; Roubaud and Sabio 2015, 2018; Lahousse and Lamiroy 2017).

Overall, the cross-linguistic comparison of the LLC-1 and CRFP datasets suggests that *it*- and *c'est*-clefts, though similar on the surface, each have specific functional-structural profiles which present some dissimilarities. Whether this is tied to the information structural properties of the two constructions is discussed in the following section.

5.2. The operationalisation of information structure

In both English and French, clefts are a productive resource serving information structural purposes. The following sections compare the operationalisation of the relational and referential information structures in *it*- and *c'est*-clefts. Relational information structure, i.e. focal/non-focal, is examined in Section 5.2.1 and referential information structure, i.e. new/given, in Section 5.2.2.

5.2.1 The relational information structure of *it*- and *c'est*-clefts

In the approach developed in this study, relational information structure pertains to the encoding of prosodic focus. For English, this takes the form of the nuclear accent while French relies on tonal variation.⁶³ As shown in Sections 3.2 and 4.2, typifying the prosodic patterns of *it*- and *c'est*-clefts can be done by assessing the number of nuclei in the former and of intonation units and tone contours in the latter. Doing this has led me to devise a septenary classification for the former and a quinary one for the latter. Though there exist some overlaps between the two taxonomies, it should be noted that the typology of *it*-clefts makes a distinction between full and reduced clefts but merges assertive and interrogative clefts

⁶³ See Section 1.1.3 for a detailed description of focus-marking in English and French.

together. Because of the differences in the encoding of focus, the taxonomy of *c'est*-clefts treats both full and reduced clefts, on the one hand, and assertive and interrogative clefts, on the other, as separate categories. In what follows, interrogative clefts and reduced clefts will not be considered. The cross-linguistic analysis of prosodic articulations is carried out by conceiving of the different types in terms of the broad vs. narrow distinction. A cleft is analysed as a broad-focus construction when all of the information is presented as informationally salient. This is the case in (26) and (27) below.

- (26) it's the gu\ests# who say ah well can we help you with the ^washing-up n\ow# (LLC-1)
- (27) [c'est d'ailleurs par cet effondrement]_{HH} [que la Saône est venue s'installer comme ça dans notre plaine]_{L-L-} (CRFP)
 'it's incidentally with this collapse that the Saône river settled like this in our plains'

In (26), information focus is encoded by means of the two nuclei on *guests* in the clefted constituent and on *now* in the cleft-relative clause, both of which are realised with a similar pitch height. This not only shows an absence of prosodic hierarchy, but the hearer also does not draw the hearer's attention to a specific constituent. In other words, there are no narrow foci. Instead of choosing a specific constituent to draw the hearer's attention to, the speaker presents the whole cleft as conveying newsworthy information. By the same token, the *c'est*-cleft in (27) is realised as a single intonation unit dividing into two smaller units illustrating the HH/L-L- (type 2a in my typology) tone sequence. The clefted PP displays a continuation rise at its right edge, which marks the incompleteness of the turn and of the informational object, and the cleft-relative clause by an assertive tone, which marks the end of the unit. In this case, the scope of the focus bears over the whole cleft, thus marking the information conveyed in both of its parts equally as salient. Both cases exemplify the already identified de-coupling of lexicogrammar and prosody. In other words, the presence of narrow focus on the clefted constituent predicted by approaches such as Lambrecht's (2001) is not systematically found either in English or French. To further back up the cross-

linguistic observations made about (26) and (27), I quantify narrow and broad focus in the LLC-1 and CRFP datasets. Narrow focus is here understood as focus which is moved from final position to another element in the unit. It subsumes Halliday's (1967a) 'contrastive' information focus on the value (see Section 5.3.2), whether as unique information focus or hierarchically primary focus, for English and the paradigm of termination tones for French, e.g. L-L- or HL-. Broad focus, on the other hand, encompasses all patterns in which the clefted constituent is not in focus, which translates into absence of contrastive or hierarchically primary information focus in English and absence of terminal prosodic boundary in French. The results are presented in Table 4.

	<i>It</i> -clefts	<i>C'est</i> -clefts	Total
Narrow focus	127 (89.4%)	22 (6.1%)	149 (29.6%)
Broad focus	12 (8.5%)	325 (89.8%)	382 (75.8%)
Unclear	3 (2.1%)	15 (4.1%)	18 (3.6%)

Table 4. Distribution of broad and narrow focus in *it*- and *c'est*-clefts

The results show an unequivocal asymmetry between the two languages. *It*-clefts exhibit a clear preference for narrow focus while *c'est*-clefts significantly favour broad focus. More specifically, narrow focus represents 89.4% of all *it*-clefts and only 6.1% of *c'est*-clefts. In the same vein, broad focus subsumes 89.8% of all *c'est*-clefts but only 8.5% of *it*-clefts. What this suggests is that English speakers primarily rely on the use of *it*-clefts to emphasise a specific lexical item, which may be (part of) the clefted constituent or (part of) the cleft-relative clause, while French speakers typically make use of *c'est*-clefts to highlight a bigger chunk of information. This discrepancy not only thwarts the expectations created by some of the accounts of *it*- and *c'est*-clefts⁶⁴ (e.g. Lambrecht 2001; Bourns 2014) but it also further backs up the argument that, though syntactically similar, the two constructions do not function the same way. Thus, as far as their relational information structure is concerned, *it*- and *c'est*-clefts cannot be taken as true

⁶⁴ The tendency for *c'est*-clefts to exemplify broad focus in spoken French was already noted by Mertens (2012) but not substantiated with quantifiable results.

equivalents. Whether this difference is also reflected in their referential information structure is examined in the following section.

5.2.2 The referential information structure of *it*- and *c'est*-clefts

Having compared the relational information structure of *it*- and *c'est*-clefts, I now delve into their referential information structure, i.e. the distribution of discourse-new and discourse-given information. As shown in Sections 3.3 and 4.3, four patterns were identified in both languages, which correspond to the four logical possibilities offered by the binary division of information. The frequencies of each pattern are summarised in Table 5.

	<i>It</i> -clefts	<i>C'est</i> -clefts
New-given	90 (37.8%)	104 (21.7%)
New-new	34 (14.3%)	38 (7.9%)
Given-new	24 (10.1%)	178 (37.1%)
Given-given	73 (30.7%)	158 (32.9%)
Unclear	17 (7.1%)	2 (0.4%)
Total	238 (100%)	480 (100%)

Table 5. Distribution of patterns of discourse-familiarity in *it*- and *c'est*-clefts

Both *it*- and *c'est*-clefts may be of the new-given, new-new, given-new or given-given type. However, the distribution differs cross-linguistically. The most common pattern found in *it*-clefts is new-given whereas *c'est*-clefts are most frequently given-new. *It*-clefts and *c'est*-clefts share the same second most frequent pattern which is that of given-given. The third most frequent pattern is new-new for *it*-clefts and new-given for *c'est*-clefts. Finally, a small number of *it*-clefts are of the given-new type and *c'est*-clefts of the new-new type. As far as the comparison of each type in English and French is concerned, only the given-new type is significantly more frequent ($p=.002372$) in the latter than in the former. The contrastive analysis warrants two main conclusions.

First, both *it*- and *c'est*-clefts display wider informational variance than previously described. This supports Delin's (1990) findings on *it*-clefts, which show that the pattern consisting of new information in the clefted constituent and given information in the cleft-relative clause is not in fact prototypical. Clefts may introduce new information in the cleft-relative clause and vice versa. The comparative summary provided in Table 5 reveals that this is not only true for *it*-clefts but also for *c'est*-clefts. Delin's (1990) corpus study also reveals that clefts do not display a strict 'segregated' distribution between new and given information tied to either of the clauses in the cleft (Delin 1990: 84). In other words, discourse-new information may be introduced in both the matrix clause and the cleft-relative clause. This is once again largely corroborated in both languages with the existence of the new-new and given-given types.

Secondly, the fact that the category of given-new is represented more in *c'est*-clefts is consistent with the findings of the syntactic analysis covered in Section 5.1.3. These uncovered a significantly higher number of clefts with an anaphoric, and hence discourse-given, clefted constituent in French. These may be coupled with new, e.g. (28), or given information, e.g. (29) in the cleft-relative clause.

- (28) on a essayé de contacter les gens et on a essayé de leur dire qu'on pouvait faire ce qu'on pouvait leur proposer avec au moins aussi bien que ce qu'ils avaient après [c'est eux]_{L-L} [qui choisissent]_{I-I}. (CRFP)
 'we tried to contact these people and we tried to tell them that we could do what we could offer them with at least as good as they had but **it's them who choose**'
- (29) ça été quelque chose de très très difficile [c'est pour ça]_{HH} [que à côté je vais voir]_{/LL} euh certaines petites personnes âgées]_{HH} (CRFP)
 'it's something that's been really really hard **it's for that reason that besides that I go visit uh some little old people**'

For Bourns (2014), anaphoric given-new clefts with a clefted subject like (28) are particularly productive in French but less so in English because of the possibility to use prosodic shift in the canonical counterpart. This appears to be confirmed by the higher frequencies of given-new *c'est*-clefts. Interestingly, Bourns (2014) rejects

the cleft analysis of *c'est pour ça*-sentences like (29) because, she argues, they do not have a cleft alternative in English and do not illustrate the prototypical focal + non-focal model posited by Lambrecht (2001). On her account, (29) is a mere cleft lookalike that serves a pragmatic purpose that differs from that taken on by what she views as typical clefts. Relying on the same argument, she also treats new-new *it*- and *c'est*-clefts with a non-presupposed variable as non-clefts whose function is to share factual information. This type is illustrated in (30) and (31), both of which require presupposition accommodation⁶⁵ on the hearer's part (Lambrecht 1994; Lehmann 2008). Besides the 'fact-stating' pragmatics specific to this category, Bourns (2014) views the possibility of past tense in the copula as an additional criterion against the cleft reading.

- (30) he attended the Coronation when he was five years old today his twenty-fifth birthday in the uniform of a lieutenant and it's interesting to remember that# **it was \as a lieutenant# that his f\ather# the ^Duke of \Edinburgh# ^entered the /Abbey# for h\is wedding# in ^nineteen forty-s\even#** (LLC-1)
- (31) quand on peut partager je pense que on est d'autant plus heureux [***c'est un petit peu l'école de vie que j'ai appris en pénitencier***]_{HH} ça m'a é~ permis de de comprendre certaines choses (CRFP)
 'when one can share I think one is all the more happy **it's a little bit the school of life that I learned in prison** it's a~ allowed me to better understand some things'

Against Bourns (2014), I do not take either *it*- or *c'est*-clefts to fulfil a unique focalisation function and therefore admit new-new clefts such as (30) and (31) as subtypes of information distribution in their own right. To better account for the information structural difference between *it*- and *c'est*-clefts, I focus on their pragmatic features in the following section.

⁶⁵ Unlike Bourns (2014), Muller (2003) does not view the lack of an overt presupposition as an excluding factor for *c'est*-clefts.

5.3. Pragmatic features of *it*- and *c'est*-clefts

I take the cross-linguistic comparison of cleft constructions further by focusing on their pragmatic characteristics. I first discuss the nature of the exhaustive and contrast meanings associated with clefts in Sections 5.3.1 and 5.3.2 respectively. I then further refine the account of the functions of *it*- and *c'est*-clefts by comparing localised and textual uses in Section 5.3.3.

5.3.1 Exhaustivity

It- and *c'est*-clefts have traditionally been described as triggering exhaustivity⁶⁶, whereby it is Johnny and *only* Johnny who stole money in (32). Semantically, this means that Johnny is identified as the unique value which exhaustively specifies the variable *x stole her money*.

- (32) it was Johnny# that stole her money# ^while we were away in France# (LLC-1)

While the exhaustive effect is generally accepted for both *it*- and *c'est*-clefts, its description varies. For É. Kiss (1998) and Krifka (2008), exhaustivity stems from the type of focus laid on the postcopular complement which is identificational⁶⁷ and which “performs exhaustive identification on a set of entities given in the context or situation” (É. Kiss 1998: 248). This type of focus, É. Kiss (1998) argues, precludes a number of lexical items such as universal quantifiers or *something/somebody* from appearing in postcopular position, and also severely restricts the use of particles such as *also* or *even*. These observations are confirmed by the analysis of the LLC-1 and CRFP data. The few cases in which additive meaning is attached to the clefted constituent correspond to the restrictive contexts which É. Kiss (1998: 252) accepts

⁶⁶ Other terms used to refer to the exhaustive meaning of clefts include ‘maximality’ (Weinert and Miller 1996), ‘exclusiveness’ (Collins 1991) and ‘uniqueness’ (Delin and Oberlander 1995; Clech-Darbon 1999).

⁶⁷ Identificational focus can be understood as contrastive focus.

as allowing multiplicity of values.⁶⁸ This is illustrated in (33) in which *wine* is identified by speaker A as *only* one of the appropriate values fulfilling the variable *x goes into port*. Speaker B then humorously adds *toenails* as another alternative of the paradigmatic set filling the semantic gap in the presupposed open proposition by completing the second reduced cleft started by speaker B. Two values are therefore identified as instantiations of the variable, but the rest of the set is still excluded from the specification relation, which is what allows the additive meaning in the first place. The interrupted cleft in (34) also exemplifies this idea. Here, speaker A is explaining how the complex process behind the solving of an IT issue typically involves multiple teams working together. Speaker A uses the cleft to add another reason as to why it is important to precisely define the type of issue at the very beginning in order for it to be treated efficiently. It should be noted that although the additive particle *aussi* is added in the matrix clause to mark the clefted constituent as only one of the correct alternatives, no second value is explicitly provided. Even without this, (34) falls within the restricted class of contexts which É. Kiss (1998) considers as allowing for additive specification as the specification relation still only applies to a closed set of alternatives.

- (33) A: cos all sorts of things go into port don't they I mean **it's not just wine [~~that go into port~~] it's**
 B: like t/oenails [~~that go into port~~] (LLC-1)
- (34) A: là ils tombent au niveau d'un d'un SVP la personne là elle va pouvoir qualifier le problème et derrière euh router l'appel euh au niveau d'une personne qui serait plus qualifiée donc là après ils ont deux catégories de population ils ont les techniciens euh dit CISI mais ça c'est leurs

⁶⁸ Contra É. Kiss's (1998), Davidse (2000) and Davidse et al. (2022) show with (32') that universal quantifiers can in fact be inserted in postcopular position. What is entailed by the universal quantifier *all* in (32') is that it the entire of set of houses which require attention that are being done.

- (32') A: They 've done all those up and erm but ... I don't know whether it's just council houses ...
 B: Mm as far as I 'm aware I think **it's [all the [houses that require attention]] [Ø are being done]**. (Wordbanks)

normes à eux quoi c'est ce que eux ils utilisent comme comme comme
 nom qui eux vont être purement euh techniques informatique x réseau
 etc et ils ont tout ce qui est D R A qui eux vont être métier c'est-à-dire
 que ils ont des logiciels où ils se-- où ils vont euh ils vont aller avec les
 autres organismes mais euh je les plus tous en tête euh si le etc mais
 enfin bon enfin le problème quoi c'est que pour comprendre leur
 spécificité faut comprendre leurs leurs besoins et leur métier quoi
**[c'est aussi ça]_{HH} [qui fait que euh ben la phase de spécification
 chez]_{L-L} [au départ]_{LH}**

B: hum hum

A: on a besoin d'eux comme eux ils ont besoin de nous quoi c' est un
 travail en commun qui permet de d'arriver sur une solution qui euh qui
 corresponde (CRFP)

'then they're directed towards a a SVP the person there she'll be able
 to describe the issue and after that uh to forward the call uh to
 someone who'd be more qualified so then there are two categories of
 people they have the technicians uh called CISI but that's their own
 norms that's what they use as as as a name who are purely computer
 network experts etc and they have everything surrounding DRA who are
 professionals that is to say that they have software where they s~ they
 uh they go hand in hand with other bodies but uh I don't recall all of
 them uh if the uh but well so well the issue is that in order to
 understand their specificity one must understand their their needs and
 their job **it's also that that makes the specification phase at at the
 beginning**

B: hm hm

A: we need them like they need us after all it's a common job that
 allows us to arrive to a solution which uh which fits'

The low frequency of *also*-clefts in both English and French data provides evidence for É. Kiss's (1998) description of identificational focus in *it*- and *c'est*-clefts as being largely exhaustive. It is also consistent with the claims that exhaustivity takes the form of a conversational implicature which I discuss below.

Although most descriptions of clefts accept the exhaustivity reading, not all of them agree on its pragmatic source. More specifically, two competing approaches are developed based on whether exhaustivity is treated as a presupposition or an implicature. For Delin and Oberlander (1995) and Büring and Križ (2013), the exhaustive effect is presuppositional in nature. While Delin and Oberlander (1995) do not develop their argument further, Büring and Križ (2013) claim that exhaustivity is introduced in the cleft as a conditional presupposition. For them, what differentiates it from an implicature is that it must be “met in the local context in which the sentence triggering [it] is evaluated” (Büring and Križ 2013: 9-10). As such, exhaustivity cannot be attributed to the speaker if embedded in negation or in a belief statement concerning another participant, which is not the case when analysed as an implicature (Horn 1981). In the second approach, proponents of the implicature analysis (e.g. Horn 1981; Declerck 1988; Collins 1991; Byram Washburn et al. 2014; De Cesare and Garassino 2015) view exhaustivity as a conversational implicature which derives from the specificational semantics of the cleft and which is at least somewhat conventionalised. As an implicature, the exhaustive effect can be cancelled, though Declerck (1988) notes that this is only the case in specific contexts in which the cleft makes use of restrictive adverbs to convey non-exhaustivity. This claim is corroborated by the analysis of my data. In (35) and (36), for instance, the particles *not only* and *surtout* ‘mainly’ imply the existence of more than one acceptable value and therefore exclude any exhaustive reading.

- (35) I’m very glad to know that **it’s not only the missiles# and the \H bombs# that are ^going to be set upright by this g\overnment#** (LLC-1)
- (36) en règle générale c’est pas toujours ceux qui ont pas beaucoup d’argent qui sont le le plus sale contrairement à ce qu’on pourrait croire parce que en règle générale c’est surtout l’inverse et [**c’est surtout**]_{HH} [**en règle générale**]_{LL} [**ces gens-là**]_{HH} [**qui se plaignent le plus**]_{L-L} (CRFP)
 ‘in general it’s not always those who don’t have a lot of money who are the the the dirtiest contrary to what one may think because in

general it's mostly the opposite **it's mostly in general those people who complain the most'**

Further evidence for the acceptability of non-exhaustive clefts is brought by Byram Washburn et al.'s (2014) experimental findings for *it*-clefts and Destruel and De Veugh-Geiss's (2019) for *c'est*-clefts. In by Byram Washburn et al.'s (2014), participants did not find non-exhaustive clefts significantly more unacceptable than exhaustive ones, thus showing that exhaustivity is not obligatory in clefts and that it arises pragmatically rather than semantically or syntactically. In the same vein, Destruel and De Veugh-Geiss (2019) found that exhaustivity violation was treated as less problematic by participants in *c'est*-clefts than in German *es*-clefts. Declerck (1988) further argues that, besides being cancellable, the exhaustive effect may also be absent altogether when negation markers are inserted in the matrix clause. This claim is also supported by the results of my analysis in both English and French. Tackling (37) first, the negation located before the clefted constituent cancels the specification relation between the value *through the British Council* and the variable *he came through x*. Because no specification is established, no exhaustive listing is either. It is, however, brought back with the following reduced cleft identifying on a Goodman Fellowship as the actual value satisfying the variable. Thus, as argued by Declerck (1988), this is not a case of exhaustivity cancellation, but rather of a lack of specification causing a logical absence of exhaustivity.

- (37) A: he came through the British Council did he#
 B: ^well n\o# **it's ^not through the British Council [~~that he came~~]**#
 it's ^on a Goodman Fellowship [~~that he came~~]
- (38) j'avais en-tissé de tellement bons rapports que ses enfants sont en France à Paris [**c'est pas moi**]_{HH} [**qui les fait venir**]_{HH} mais enfin si j'avais pu je l'aurais fait (CRFP)
 'I had established such a good relationship that her children are in France in Paris it's not me who had them come but well if I had had the chance to I would have done so'

The same conclusion can be drawn for the *c'est*-cleft in (38). What is presupposed is that someone had the children come to France, which, as the cleft indicates, is not the speaker but someone else. Here too, the negation blocks the triggering of the exhaustivity effect. Unlike in (37), the correct value is not provided in the following co-text but the presupposition still remains active. Doetjes et al. (2004) also recognise the possibility of non-exhaustivity in a subtype of broad-focus clefts, i.e. new/given-new, whose copula does not take on an equative function and which does not select a referential expression as their postcopular constituent. Doetjes et al. (2004) illustrate this with (39), which they argue can be completed with *et par ailleurs aussi avec fierté* 'and besides with pride too', which would further cancel the exhaustive reading. Only one example of similar non-referential syntagms, shown in (40), was found in the CRFP. However, unlike (39), (40) displays a new-given pattern whereby the fact that she came back is mentioned verbatim in the preceding discourse. This arguably confers a stronger contrastive, and hence exhaustive, reading on the cleft.

- (39) C'est avec plaisir que je vous invite à ce séminaire.
 'It is with pleasure that I am inviting to this seminar.'
- (40) c'est vrai que lorsqu'elle est revenue là il y a deux trois mois [**c'était avec beaucoup de plaisir**]_{L-L} [**qu'elle est revenue**]_{I-I} (CRFP)
 'it's true that when she came back here two months ago it was with a lot of pleasure that she came back'

In fact, no instances of non-exhaustive broad-focus clefts similar to the one described by Doetjes et al. (2004) were attested in my English or French data. Non-exhaustivity in my data is thus primarily expressed lexically with the use of non-restrictive particles such as *not only/pas seulement*, *mainly/surtout*, etc.

The exhaustive effect of *it*- and *c'est*-clefts may be cancelled but it may also be reinforced (Clech-Darbon 1999; Velleman et al. 2012; De Cesare and Garassino 2015). De Cesare and Garassino's (2015) cross-linguistic study of English *it*- and Italian *c'e*-clefts, for instance, highlights the use of exhaustive particle *only/solo* to intensify exhaustivity. The same observation was made for the data under study in Section 3.4.4, in which I additionally showed that lexical means such as

exhaustive markers may be combined with prosodic means like a high onset for English. This is the case in (41), for which I have argued that the high onset on *only* fulfils a dual function. At the sentence level, it allows the speaker to emphasise and spell out the implicature of exhaustivity triggered by the identifying matrix. Within the interactional context, the high pitch level encodes some sort of disruption and unexpectedness with has been shared in the prior discourse (O'Grady 2013). In this case, the speaker first makes a general argument about the wild behaviour people generally develop towards rock and roll singers, after which the cleft allows him to recall a specific event in which he happened to discuss this specific topic.

- (41) there's something that makes us feel savage about these rock and roll singers and I hate it in myself and I see it in a lot of other people# **now it's** ^{H^}**only about a year ago**# - **that** ^{H^}**on this programme**# ^{H^}**we were asked about I think it was Tommy StVeele**# **being mobbed**# and I remember making some perfectly horrible remarks (LLC-1)
- (42) A: et au bout de combien de temps on vous laisse essayer avec des ciseaux
 B: et pff deux ouais deux ans
 A: vous passez deux ans à regarder
 B: ouais regarder puisque la plupart en apprentissage [**c'est que shampooings**]_{HH} [**couleurs**]_{HH} [**qu'on fait**] après petit à petit ben on fait d' autres choses coupes euh des permanentes (CRFP)
 'A: and how long does it take until you are allowed to try cissors
 B: and pfff two yeah two years
 A: you spend two years watching
 B: yeah watching since most of the work/study programme **it's only shampooing dyeing** [~~that we do~~] then gradually well we do other things'

In total, ten occurrences of *only*-clefts were found in the LLC-1 and seven of *uniquement/que*-clefts in the CRFP, as illustrated in (42). Similarly to (41), the exhaustive particle *que* in (42) takes the exhaustivity meaning from being simply

implied to being asserted. The information shared and explicitly stated by the speaker is that it is shampooing and dyeing and *only* that that the speaker and her colleagues get to do during the two years of the study-programme. The prosodic realisation of (42) differs from that of (41) in that no particular prominence is borne by the particle. Though not occurring in this specific example, the initial accent in French is known to frequently attach to focus operators, e.g. *seulement*, *surtout*, *même*, etc., and to emphasise their meaning (Féry 2001; Di Cristo 2016: 272). Even without prosodic marking, what the discussion of the previous examples shows is that exhaustivity may be reinforced through lexical means, and, in some cases, may display a three-way encoding involving the conversational implicature, the insertion of an exhaustive particle and prosodic prominence.

On the relation between focus and exhaustivity, Velleman et al. (2012) note that the exhaustivity expressed in *it*-clefts is focus sensitive. As such, it is directly tied to the location of the main pitch change within the clefted constituent. In cases of selective focus, i.e. focus carried by a lexical item of the clefted constituent other than the last one, the exhaustivity effect does not apply to the clefted constituent as a whole. Instead, only the focused item within the clefted constituent exhaustively satisfies the presupposition conveyed by the rest of the clefted constituent and the cleft-relative clause. Thus, in (43a), it is not *garden space* but *garden* alone that exhaustively satisfies the presupposition *x space is so precious for kids*, giving rise to the implicature that only one type of space is precious for kids. Because *garden* only, and not *garden space*, is in focus, the cleft may be completed with (43b), which specifies another value, *playgrounds*, for the open proposition *x that is so precious for kids*. *Playgrounds* are not part of the *house* versus *garden* (space) contrast, which the exhaustivity implicature was triggered for. By contrast, a proposition like (43c) added immediately after (43a) would be infelicitous with regard to the presupposition that *x space is precious for kids*.⁶⁹

⁶⁹ Declerck (1988) notes that while the insertion of adverbs cancelling exhaustivity such as *not only/pas seulement* or *mainly/surtout* is perfectly acceptable, as illustrated in (35) and (36), the addition of material beyond the boundaries of the cleft to achieve the same effect, as in (43c), is much less acceptable.

- (43) (a) A: no and it's garden space# that is so# precious# for kids#
and not house space# (LLC-1)
(b) ... And they also like playgrounds.
(c) # ... And house space is also precious to them.

In an effort to further describe exhaustivity in clefts, Velleman et al. (2012) analyse the exhaustivity and the use of exhaustive particles like *only/que* within the Question Under Discussion (QUD) approach. According to them, the two types of exhaustivity differ in the nature of the meaning component that is at issue. Both types convey a minimal (MIN) component, which requires there be a true answer to the QUD above a certain lower bound, and a maximal (MAX) component, which requires there be no true answer to the QUD above a certain upper bound. The cleft in (44a), for instance, introduces the presupposed open proposition *x laughed* answering to the *wh*-question *Who laughed?*

- (44) (a) It was MARY who laughed.
(b) Requires a question of the form “Who laughed?”
(c) Asserts that at least Mary laughed.
(d) Presupposes that there is no true answer strictly stronger than “Mary laughed.” (Velleman et al. 2012: 443)

The MIN term asserts that there is at least one true answer, e.g. Mary, to the question of who laughed. The MAX term, on the other hand, presupposes that there is no other true answer stronger than Mary to the question who laughed. In the corresponding exclusive construction in (45a), the MIN and MAX are reversed with the former being presupposed and the latter asserted.

- (45) (a) Only MARY laughed.
(b) Requires a question of the form “Who laughed?”
(c) Presupposes that at least Mary laughed.
(d) Asserts that there is no true answer strictly stronger than “Mary laughed.” (Velleman et al. 2012: 443)

Thus, while clefts make the MIN component at-issue, exclusive constructions with *only* make the MAX component at-issue. When applied to cases of selective focus such as (46a), this formal description makes sense of the change in their semantics. The underlying question is not what counts, but which type of material counts. The resulting assertion and presupposition do not, however, change.

- (46) (a) because there's not the same pressure on the material# **it's the pop material that counts#** (LLC-1)
(b) Requires a question of the form "Which kind of material counts?"
(c) Asserts that at least pop material counts.
(d) Presupposes that there is no true answer strictly stronger than "Pop material counts".

When exhaustivity is reinforced lexically, the exclusive particle takes over the implicature of exhaustivity which means that the MAX component is at issue. On Velleman et al.'s (2012) account, *only/que* are not a mere reinforcement of the exhaustivity effect of the cleft but instead serve to change the pragmatic mechanisms at play in the cleft. As a result, clefts with exclusive particle *only/que* function as a type of *it*- or *c'est*-clefts whose pragmatic mechanisms differ from that of bare clefts (Declerck 1988; De Cesare and Garassino 2015).

What I have shown in this section is that both *it*- and *c'est*-clefts generally trigger an implicature of exhaustivity whereby the clefted constituent is presented as the unique value exhaustively satisfying the variable. This exhaustive reading may be cancelled, thus excluding analyses of exhaustivity as a presupposition, or emphasised either through lexical means, prosodic means, or both. While cancellation occurs in both languages, reinforced exhaustivity was found to be slightly more frequent in English than in French.

5.3.2. Contrast

Along with exhaustivity, clefts have also traditionally been associated with the notion of contrastivity which is said to typically apply to the clefted constituent (Rouget and Salze 1986; Declerck 1988; É. Kiss 1998; Huber 2006; Rouquier 2007;

Scappini 2013; Destruel and Velleman 2014). The cleft not only allows to exclude all irrelevant set members from the specification relation but it also contrasts the accurate value with said members. Taking (32) again, repeated as (47) below, what is conveyed is that it is Johnny, as opposed to other people, who stole her money.

- (47) it was J\ohnny# that stole her m\oney# ^while we were away in Fr\ance# (LLC-1)

While this is true for some *it*- and *c'est*-clefts, it has been argued that not all clefts are contrastive. In the typologies of *it*-clefts described in Section 1.2.2, only the first type of cleft, i.e. Prince's (1978) *stressed-focus*, Declerck's (1984) and Huber's (2006) *contrastive*, and Geluykens's (1988) *filler-focus* clefts, is said to introduce a contrastive complement in postcopular position. As such, contrast is associated with discourse-new clefted constituents only while anaphoric ones are analysed as non-contrastive. This strict binary distinction is not borne out by my findings in the LLC-1 dataset.

Unlike what is argued in the existing typologies, a discourse-given value, even when followed by a discourse-new variable, may be contrastive. It is the case of (48) in which the speaker is explaining how the circulation system works. The clefted NP *these individual copies*, which is arguably derivable from the prior context, is re-introduced in the cleft as the component at the core of the circulation system. The combination of the high onset on the demonstrative *these* and the nuclear accent on *copies* gives the whole NP a prosodic and informationally prominent role. The cleft-relative clause likewise bears a nuclear accent on *core*, but which is lower-pitched than that on *copies*. What is to be understood is that it is the very individual copies, and not general reports, on which the whole system relies.

- (48) you can see how this information we're putting in relates just to this this copy and that's how it's held in a in a separate entry and the a little separate record for each copy which will have circulation data (...) and in some ways this is the circulation system in fact# it's^H^these

individual copies# which are really the core of the system# this is why it's the copy number which is indexed on line (LLC-1)

Thus, contrastiveness in *it*-clefts does not appear to be dependent on the distribution of given/new information. Instead, as discussed in Section 3.2, contrast is better conceived in terms of contrastive focus vs. presenting focus. Within the Hallidayan approach, contrastive focus represents the marked type of information structure which contains a focus that is “informationally contrastive [...] within a closed system or lexically” (Halliday 1967a: 207). The notion of contrastive focus subsumes both contrast with, and addition to, another option from a finite set (Halliday 1967a: 226). Presenting focus, on the other hand, is the unmarked focus occurring unit-finally. This distinction allows one to better make sense of the relation between contrast and the different prosodic types of *it*-clefts. As a reminder, these include clefts with a unique focus on the clefted constituent and clefts with multiple foci (see Section 3.2 for discussion of this point). With the proposed analytic model of contrast, the higher-pitched nucleus in the value of (48) is taken to illustrate marked contrastive focus while the open proposition in the variable exhibits an unmarked presenting focus, regardless of its discourse-newness or discourse-givenness. By contrast, the unique nucleus carried by the clefted constituent of (49) illustrates contrastive information focus on the value NP *one question* which relates to the deaccented presupposition *that they have to do*.

(49) A: well you give them the lot you see, that's the point, and make sure that there's something fairly closely related to what they've studied

B: **it's just one question that they have to do# /isn't it# (LLC-1)**

Clefts which have a single nuclear accent in the variable instantiate another type of specific information structure in which recoverable information precedes a presenting focus associated occurring towards the cleft-relative clause. This is the case in (50).

- (50) A: did you meet Fuller
 B: yes# **it was ^he who invlited me#** (LLC-1)

In the LLC-1, *it*-clefts are used most frequently with contrastive focus on the value and presenting focus on the open proposition, as illustrated in (48). This type of cleft thus combines the information structural affordances of both the marked and unmarked information structure. Their open proposition is typically not presupposed, which I can capture by describing them as discursively “discontinuous” (Declerck 1984). *It*-clefts may also display a marked information structure with a contrastive focus on the value, e.g. (49), which defines the variable as presupposed information. Finally, in a minority of cases, e.g. (50), *it*-clefts manifest the unmarked information structure with a presenting focus occurring on the final lexical element of the open proposition. This pattern has received little attention in the literature so far.

As far as *c'est*-clefts are concerned, Scappini (2013) develops a similar binary account in which she opposes *contrastive* clefts ‘clivées à contraste’ to *broad-focus* clefts ‘clivées à enchaînement’. Contrastive clefts display narrow focus on the clefted constituent, which is contrasted, either explicitly or implicitly, with another member of a paradigmatic set. Scappini (2013) further differentiates between a number of contrastive types according to the salience of the paradigm in the context. Contrast is realised when the alternative(s) set against the clefted constituent is expressed and virtual when it is not. The contrastive effect may derive from instantiations of a lexical field or from the use of paradigmatic adverbs. Broad-focus clefts, in comparison, have a non-contrastive and non-paradigmatic clefted constituent which typically belongs to closed lexical category, e.g. *ainsi*, *là*, *pour ça*. Scappini’s (2013) categories are all observed in the CRFP dataset.

I start by discussing contrastive clefts, whose types are illustrated in (51)–(54) below.

- (51) si je m'intéresse au tarot [**c'est pas parce que je te vais tirer les cartes**]_{HH} [~~que je m'intéresse au tarot~~] j'en ai rien à foutre de tirer les cartes mais c'est pour la l'initiation symbolique il y a derrière le tarot [~~que je m'intéresse au tarot~~] (CRPF)

'if I'm interested in tarot **it's not because I'm going to draw cards** [~~that I'm interested in tarot~~] I don't give a fuck about drawing cards but it's for the symbolic initiation that's behind tarot [~~that I'm interested in it~~']

- (52) tout ce qui est autour du papier du corps du texte c'est tot- [c'est pas du tout le journaliste]_{HH} [~~qui le fait~~] (CRFP)

'everything surrounding the paper the body it's t- it's not the journalist at all [~~who makes it~~']

- (53) A: et la boulangerie avait une bonne réputation elle a une grosse
B: ouais ouais ça ça avait une bonne réputation et en plus comme ma tante qui était avant il y avait pas mal de personnes enfin qui étaient âgées euh à l'époque de mes parents vu qui il y avait un petit décalage qui étaient habituées à venir et puis sachant que c'était le neveu je pense que ça y a fait au au début quand ils sont arrivés quoi que ce soit quelqu'un de la même famille

A: donc en fait d- [c'est ton père]_{HH} [qui était boulanger]_{H/H} ta mère s'est retrouvée embarquée dans la ah d'accord (CRFP)

'so actually t- it's your father who was a baker you mother found herself involved in the ah okay'

- (54) ouais bah ben tiens question question que je me posais euh c' est vrai que en général dans dans les concerts hardcore [c'est ^Hsurtout les filles]_{HH} [qui prennent des photos]_{L-L} (CRFP)

'yeah well okay here's a question question that I have uh it's true that in general at at hardcore concerts **it's mostly girls who take pictures**'

Realised contrast is exemplified in (51) in which the two alternatives *parce que je vais tirer les cartes* and *pour l'initiation symbolique* are contrasted by means of two reduced clefts. Here, only the latter is accepted as accurately filling the semantic gap in the presupposed open proposition. The temporal sequence 'rejected value-accepted value' found in (51) and in a number of other tokens exemplifies Rouquier's (2007) observation that French *c'est*-clefts with an explicit contrast typically present a negative polarity followed by a positive one. (52) illustrates virtual contrast in which the paradigm set is not expressed but the clefted

constituent still retains a contrastive undertone through the negation implying that the correct value is another set member. In this case, it is not the journalist but someone else, who remains underspecified, that creates the text. Lexically-suggested contrast is instantiated in (53) where the paradigmatic set is already established as being that of kinship in the preceding context. Because the opposition between *mother* and *father* is made explicit, I argue that lexically-suggested contrast in fact constitutes a subtype of realised contrast. Finally, contrast may also be implied with the use of paradigmatic adverbs such as *surtout* in (54), which Scappini (2013) analyses as a privilege marker. As such, it marks the clefted NP *les filles* as being the most appropriate value out of all those exhaustively satisfying the variable. The contrast established therefore differs from that developed in the three previous examples. The clefted constituent is not contrasted with rejected values but with accepted values whose relevance is lesser. In other words, the information conveyed in (54) is that it is mostly girls, but sometimes boys too, who take pictures.

Broad-focus clefts with a non-paradigmatic and anaphoric clefted constituent are also frequent in the CRFP dataset, as illustrated with (55) and (56). In (55) the anaphoric adverb *ainsi* is not in focus and does not exhibit any paradigmatic or contrastive behaviour. As shown in Section 4.4.4, the cleft has weakened focality and mostly serves cohesive and textual purposes. (56) can be analysed in a similar way with the clefted adverb *là* introducing the result of the different facts enumerated by the speaker in the prior discourse. Through the given-new partition of information and through the encoding of broad focus, the speaker draws the hearer's attention to the content of the cleft-relative clause rather than that of the matrix clause. Attempting to bring out the contrastive aspect typically found in clefts does not prove effective in the case of (56) which instead exemplifies a discursive, and not contrastive, use of the cleft. In both (55) and (56), the clefts do not function as bipartite constructions but as simple clauses introduced by a connective marker (Lehmann 2008; Lahousse and Lamiroy 2017). Clefts of the kind number 61 in the CRFP.

- (55) euh ceci étant terminé nous sommes partis pour Mopti qui est un un port un sur le fleuve euh dont je me souviens plus du nom et euh qui

est assez original euh là nous avons euh pas eu de problèmes particuliers euh nous avons visé-- visité ce que nous avons à faire nous étions dans un bon euh campement nous nous entendions bien avec les deux les deux anglaises et **[c'est là]_{HH} [qu'on a décidé de se séparer de du Dogon]_{HH}** parce que en fait ça faisait quand même quelques jours qu'on le trimbalait (CRFP)

'uh with that being over we left for Mopti which is a a port a on the river uh of which I don't remember the name and uh which is quite original uh there we uh didn't have any particular issues uh we visi-visited what we had to do we were at a good uh campsite we got along well with the two the two English girls and **it's then that we decided to part ways with the Dogon** because actually we had been carting him around for several days'

- (56) donc bon une fois qu'il a eu défini ça donc j'étais pas là pour prendre tous les enfants s difficulté de l'école que j'étais pas non plus animatrice hein mais que donc euh je devais euh avoir à l'esprit le l'axe du projet d'école qui est euh la maîtrise la favoriser la maîtrise de la langue donc euh après et bien je suis allée voir chacun en disant bon qu'est-ce que qu' est-ce qu'on pourrait mettre en place tu vois en fonction du du profil de ta classe pour euh justement euh ben faire en sorte bon ben de travailler quoi cette cette maîtrise de de la langue qu' elle soit euh orale ou écrite donc **[c'est comme ça qu'est né le projet de journal]_{HH}** (CRFP)

'so well once he had settled that so I wasn't here to take all the kids school struggle that I wasn't a team leader either right but that so uh I had uh to bear in mind the school project axis which is uh language command so uh after well I went to see each of them saying so what could what could we put into place you see based on your your class's profile to uh actually uh well make it so well that they work well you know this this command of language whether it's speaker or writing so **it's how this journal project was born'**

Although (55) and (56) support Scappini's (2013) claim about the non-contrastive nature of *c'est*-clefts with an anaphoric adverbial clefted constituent, other examples go against this analysis. As discussed in Section 4.4.4, *c'est ainsi que*-, *c'est comme ça*- and *c'est là que*-clefts are not always cohesive and may very well display paradigmatic behaviour in which contrast is established informationally with other alternatives. This is the case in (57a) in which contrastivity can be revealed by the addition of *et pas autrement* as shown in (57b). This is because the adverbial *comme ça* is in this case an adjunct of manner, modifying the verb of the cleft relative clause. By contrast, in an example like (56), the adverbial *comme ça* functions as a causal conjunct linking the proposition about the birth of the journal project in the cleft relative clause to the preceding discourse describing all the events that led to it.

- (57) (a) s'il réussit arrive à avoir euh des convictions des connaissances
suffisantes pour que il puisse équiper entre guillemets parce que
[c'est comme ça qu'on appelle_{LL} la vente]_{HH} [chez nous]_{HH}
(CRFP)
'if he manages to have uh convictions sufficient knowledge to
equip quote unquote because **it's how we call sales here**'
- (b) ... c'est comme ça et pas autrement qu'on appelle la vente chez
nous
'... that's how we call sales here not anything else'

What my findings show so far is that *c'est*-clefts may be contrastive or non-contrastive and that contrast may be more or less explicit depending on the salience of the paradigmatic set. These results support all but one argument developed by Scappini (2013) which is that *c'est*-clefts with an anaphoric AdvP in postcopular position systematically exhibit a lack of contrast.

While Scappini (2013) does not explicitly formulate hypotheses concerning the prosodic realisations of the different types of *c'est*-clefts, the contrast she draws between narrow and broad-focus clefts nonetheless creates certain expectations. Thus, narrow contrastive clefts are predicted to display a tonal realisation which focalises the clefted constituent. This can take the form of a focus/post-focus

sequence L-L-/l-l- or a contrastive accent HL- occurring on the clefted constituent. This is not the case for (51)–(54). The reduced clefts in (51) and (52) are bounded by a continuation rise at their right edge and the full clefts in (53) and (54) display broad focus through a sequence of continuation rises. This suggests that, in French, contrast developed semantically might not always be tied to a specific tonal realisation. On this, Di Cristo (2016: 270) notes that broad focus does not necessarily preclude semantic highlighting (in Bolinger's sense) from occurring within the intonation unit.⁷⁰ In (54), for instance, the high initial accent on *surtout*, shown in Figure 1, not only interacts with the pragmatic implicature of exhaustivity, or in this case the lack thereof, but also with the expression of contrast. As such, it emphasises the fact that girls, but boys too at times, take pictures at hardcore concerts.

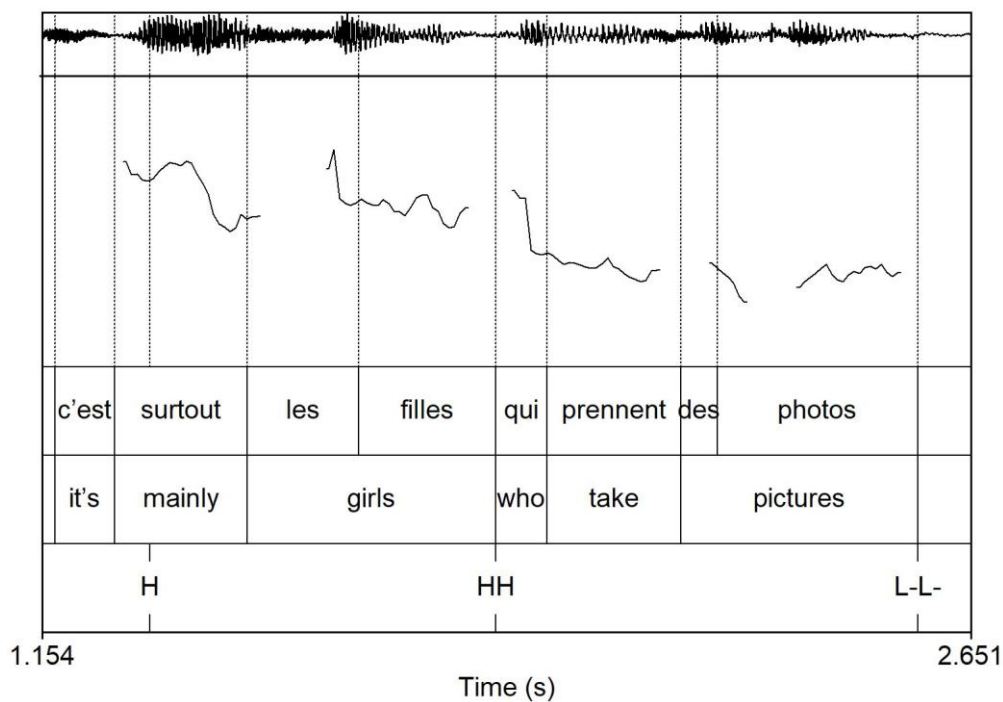


Figure 1. Prosodic realisation of (54)

⁷⁰ Citing authors like Halliday, Rochemont, Rooth and Jacobs, Di Cristo (2016: 270) also underlines the role played by topicalisation in the encoding of contrast marking whereby a contrastive effect may also arise from a constituent being placed in topic position. As such, contrast is not solely linked to focus but is instead subsumed under several information structural phenomena relating to functional organisation.

Interestingly, the set member receiving prosodic prominence in (53) is not the clefted constituent itself but the alternative provided in the following co-text. Like in (54), the contrastive reading is marked by a pitch peak which, this time, occurs on the set member *mère*, as shown in Figure 2.

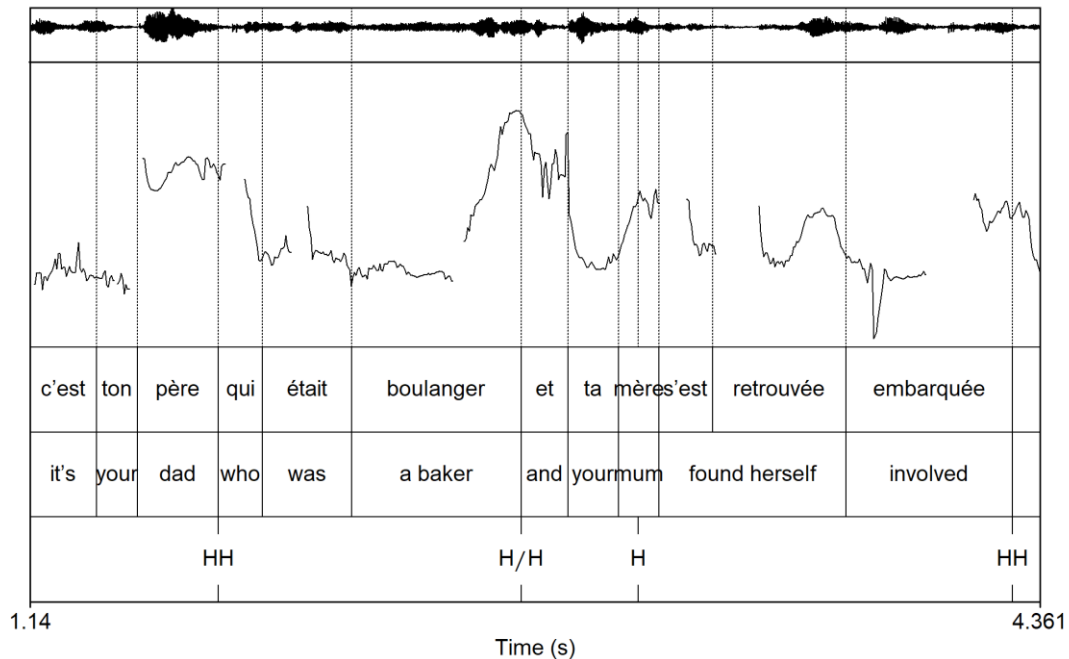


Figure 2. Prosodic realisation of (53)

This brief prosodic analysis allows me to complete the interim conclusion provided earlier. Like *it*-clefts, *c'est*-clefts generally establish a contrast between the value introduced in the matrix clause and one or more alternatives. The alternatives may be explicitly expressed in the context or suggested by the insertion of a negation marker or paradigmatic adverb. Contrast may be reinforced prosodically through the use of a focus-post-focus sequence or by means of the initial accent. Non-contrastive clefts, which represent a minority of cases in the CRFP, typically correspond to clefts which select an anaphoric AvdP such as *là*, *ainsi*, *pour ça*, *comme ça*, etc., but only when their use is cohesive. Thus, the contrastive effect associated with *c'est*-clefts is primarily generated by the 'processus d'extraction' (Scappini 2007, 2013; Blanche-Benvéniste 1990), but is lost when the cleft functions as a cohesive marker.

To conclude, both *it*- and *c'est*-clefts are typically contrastive. In *it*-clefts, contrastivity is best conceptualised in terms of contrastive focus, as opposed to presenting focus. In most occurrences, contrastive focus is borne by the value and presenting focus by the open proposition, thus marking the clefted constituent as contrastive. In *c'est*-clefts, contrast arises from the extraction process itself and is only lost in the minority of cases in which the cleft is used as a discourse marker. Unlike in *it*-clefts, the role of prosody in *c'est*-clefts seemingly remains limited and mostly allows to encode contrast in focal + non-focal clefts or to reinforce it in broad-focus clefts. This cross-linguistic difference allows me to account for the more general asymmetry between the distribution of broad and narrow focus summarised in Table 4. *It*-clefts more frequently display narrow focus which typically attaches to the clefted constituent while *c'est*-clefts favour broad focus whereby the whole cleft is in focus.

5.3.3 Discourse functions of *it*- and *c'est*-clefts

It has been argued on multiple occasions in this study that the functional profile of *it*- and *c'est*-clefts is a varied one that is far from restricted to the focalisation function generally attributed to clefts. As argued in Sections 3.3.3 and 4.5, this variety of discourse functions is directly tied to the different information structures afforded by clefts in both languages. In order to account for the different functions cross-linguistically, I propose to make a distinction between localised and general uses.

In their localised use, the functional scope of *it*- and *c'est*-clefts bears on a specific element of the preceding context or on a small portion of the immediate co-text. The most common instantiation is that of explicit contrast which is available in both languages as discussed in the previous section. In (58), the cleft reacts to the paradigm *staff*/*student* set out in the same turn and allows the speaker to reject the value *staff*, as emphasised by the high onset on the negation marker. In (59), the cleft introduces a first negated value *pas parce qu'il était un artiste* which is contrasted with a second asserted one *parce qu'il n'y avait que la lithographie qui permettait de reproduire une affiche* appearing in the rightward

co-text. Here too, prosody and grammar combine to highlight the negative marker and hence the contrastive reading.

(58) you mean in other words in the the business of the staff student relations# it's it's ^H^not the st\aff who are# ^who are making a very poor b\usiness# (LLC-1)

(59) mais lorsque par exemple Toulouse-Lautrec faisait une affiche à en lithographie [c'est ^Hpas parce qu'il était un artiste]_{L-L}. c'est parce qu'il n'y avait que la lithographie qui permettait de reproduire une affiche (CRFP)

'but when for instance Toulouse-Lautrec would make a poster in as a lithography **it's not because he was an artist** it's because there was only lithography that allowed one to reproduce a poster'

With regard to the types of lexical items typically receiving narrow contrastive focus, French clefts much more readily select unstressed pronominal subjects.

While contrast is typically associated with the postcopular position, both *it*- and *c'est*-clefts also allow, in some cases, to create a contrast between two variables. This is the case in (60) in which the value *the grammar* is introduced as the value filling the gap in two open propositions, *the fun is x where the fun is* and *x is interesting*. Here, the two clefts are spread over two turns from two different speakers. In (61), the contrast is established by the same speaker who includes contrasts to both the value and the variable of the cleft; *nous - elle, qui faisons les enfants - les garde*.

(60) A: it's the grammar where the fun is

B: yes# **it's the gr\ammar which is interesting**# one finds one | | | |
find myself adopting phrases having this kind of chameleon approach
to (LLC-1)

(61) A: et ma soeur elle a deux enfants aussi un de huit ans et l'autre de quatre ans

B: hum hum donc votre mère est bien occupée

A: oui elle est bien occupée avec les petits-enfants [**c'est nous qui faisons les enfants**]_{HH} elle les garde non mais ça se passe bien on a on a une maman adorable (CRFP)

'A: and my sister has two children too one who's eight years old and the other one four years old

B: hm hm so your mother is very busy

A: yes she's very busy with the grandchildren **it's us who produce children** she babysits them no but it's going well we have we have an adorable mum'

Another function observed in both languages and occurring at the sentence level relates to the use of question-answer pairs (Delin and Oberlander 1995). This function is mainly embodied by reduced clefts which are particularly suitable to provide an answer to a question laying out the variable, as is the case in (62) and (63).

(62) A: and what is the pardoner being driven by

B: well not not something outside himself really# **it's his ^own desire for m/oney#**

A: hm

B: I supp/ose# ^**and a s/vort of power#** [~~that drives him~~] (LLC-1)

(63) A: mais alors comment vous trouvez vos élèves c'est des rouennais aussi qui viennent

B: oh bah vous savez j'en j'en refuse tout le temps euh s [**c'est par bouche à oreille**]_{HH} [~~que je trouve mes élèves~~] je ne fais aucune publicité (CRFP)

'A : but so how do you find your students it is people from Rennes who come too

B: oh well you know I turn some of them down all the time uh s **it's by word of mouth** [~~that I find my students~~] I don't do any publicity'

In the second category of discourse functions displayed by clefts, which I call general uses, the construction plays a broader role pertaining to the discourse level.

First, clefts may act as topic-(re)launching devices by freshly introducing new topics or reintroducing topics which had previously lost their topical status (see Declerck 1984; Hasselgård 2004; Garassino 2014). This discourse function is instantiated in a small number of tokens of both the LLC-1, e.g. (64), and the CRFP, e.g. (65). Clefts of the kind may appear within the text at the start of a new conversational episode to mark a topic shift as in (64) or towards the beginning of the text itself to introduce a topic as in (65).

- (64) A: now what was the other thing I wanted to ask you is# ^is it this year# that N\lightingale _goes#
B: no next year
A: hm sixty-four sixty-five
B: sixty-five yeah
A: I thought it was before sixty-five so it's not until next year that the job will be advertised (LLC-1)
- (65) je me suis mariée le le quatre septembre de cette année dix-neuf cent quatre-vingt-dix-neuf et euh c'est euh cette journée et ce qui l'a précédée le voyage de noce qui qui a suivi dont je dont je vais parler aujourd'hui (CRFP)
'I got married on on September fourth that year of ninety ninety-nine and uh it's that day and what preceded it the honeymoon that followed that I'm going to talk about today'

In the same vein, clefts may also be used at the end of discourse segments as concluding or persuasive devices. In this case, clefts only introduce recoverable information which is repeated to either wrap up the conversational episode or to present the information as non-negotiable. In (66), both aspects combine to allow the speaker to end the discussion and assert her point. Similarly, the speaker in (67) is merely repeating a point that he has already made in the prior context so as to assert its relevance.

- (66) A: (...) the other side of Reading going into Reading that's where she hit the traffic traffic going in to Reading from either side

B: no you've missed the point the traffic you are worried about is the traffic going towards London

A: no Petey at half past eight in the morning

B: there is not a mysterious line which divides traffic going to London immediately at Reading

A: no but there is traffic there is a traffic rush hour at Reading when traffic piles into Reading# and it is a ^{H^}bout eight th\irty# that my ^mother has gotten st\uck# in ^{H^}traffic trying to get into R\reading# (LLC-1)

- (67) on s'aperçoit que les gens euh sont terriblement attachés à leur premier maître et à leurs maîtres d'une manière générale ah absolument ça c'est ça ça a certainement été une des des non pas des révélations s'en doutait bien un peu mais vraiment ils sont très très attachés et le le et ils nous disent euh bon bé c'est quand même grâce à vous que j'ai fait ça grâce à vous bon le grâce à vous je ne vous dis pas euh on a fait son boulot on fait son boulot quoi on fait son on fait son travail bon on fait ce qu'on peut c'est vrai pour les pour les aider mais euh pour eux c'est c'est c'est énorme c'est et on [**c'est presque nous qui**]_{HH} [**que qui les qui les avons fait ce qu' ils sont**]_{L-L} (CRFP)
 'we realise that people uh are terribly attached to their first teacher and first teachers in general ah absolutely this it's this this has probably been one of the the no note of the revelations I kind of knew it but really they're really really attached and the the and they tell us uh well it's thanks to you after all that I did that thanks to you well the thanks to you I'm not telling you uh we did our job we do our job we do our we do our job well actually we do what we can it's true to to help them but uh for them it's it's it's huge it's and we **it's almost us who who made them what they are'**

Finally, clefts may also participate in the structuring of discourse by acting as cohesive markers. This function, unlike the other ones I have detailed so far, is primarily found in French and is exhibited by clefts with an anaphoric AdvP such as *là, ainsi, comme ça, pour ça*, etc. (Blanche-Benveniste 2006; Sabio and Benzitoun

2013; Roubaud and Sabio 2015, 2018; Lahousse and Lamiroy 2017). In (68), for instance, the *c'est*-cleft occurs after the speaker lists a number of issues stemming from being in a different time zone than the people he needs to get in touch with. The clefted constituent acts as a discourse-marker which is coupled with discourse-new and informationally salient information in the cleft-relative clause.

- (68) A: tu supportes bien le décalage horaire
 B: là ça s'est bien fait euh il y a cinq heures de décalage euh quatre heures et demie exactement alors ça par contre ça pose énormément de problèmes au niveau du boulot parce que pour quand moi je veux joindre euh par exemple un technicien en France bon c'est une vraie galère euh il y a quatre heures et demie de moins euh il est dix-sept heures chez nous mince il sont en train de manger euh après ah ben zut moi je vais partir oui eux ils seront en train de bosser euh c'est euh invraisemblable c'est c'est des gros gros problèmes et **[c'est là]_{HH} [où justement internet]_{HH} [peut nous sauver]_{L-L}** parce qu'on peut s'envoyer des courriers électroniques (CRFP)
 'A: do you handle the time difference well
 B: this time it went well uh there's a five-hour time difference uh four fours and a half to be exact so this on this other hand causes a lot of issues when it comes to work because for when I want to contact uh for instance a technician in France well it's a real pain uh I'm four and a half hours behind uh it's five pm here crap they're eating uh then well shoot I'm going to go they'll be working uh it's uh unbelievable it's it's big big issues and **it's then that the Internet can actually save us** because we can send electronic emails'

Overall, *it*- and *c'est*-clefts appear in various discourse contexts and as such take on a range of discourse functions. The cross-linguistic analysis shows that the functions available for both languages are similar, with the exception of cohesive clefts which are much more frequently found in French. The main differences are therefore quantitative and are due to the different other focusing devices available in French and English. That *it*- and *c'est*-clefts share comparable functional profiles

is consistent with accounts treating the two constructions as cross-linguistic equivalents.

5.4 Concluding remarks

Building on Chapters 3 and 4 of the present study, the goal of this chapter was to compare the syntactic, semantic, pragmatic and information structural properties of *it*- and *c'est*-clefts. In doing so, this chapter examined the question of whether *it*- and *c'est*-clefts are true cross-linguistic equivalent constructions beyond their similar basic syntactic structure.

Starting with the frequency of occurrence of *it*- and *c'est*-clefts, my findings have uncovered a greater use of the French construction which, in spoken data, is twice as frequent as its English counterpart. Although the two constructions appear to share the same syntactic structure at first glance, the cross-linguistic comparison of their morphosyntax has revealed a number of differences. First, the morphology of the copula was found to differ on two aspects. While the copula of *it*-clefts typically displays wider variation in tense, that of *c'est*-clefts shows greater number agreement variance. The second cross-linguistic difference brought out by the analysis of the LLC-1 and CRFP datasets relates to the form and function of the clefted constituent. *C'est*-clefts exhibit a predominance of anaphoric PPs and AdvPs which results in a higher frequency of clefted adjuncts and conjuncts in French.

Following the accounts developed in the previous chapters, I investigated the information structure of *it*- and *c'est*-clefts by examining the relational and referential layers separately. Prosodically, the two constructions turned out to have an asymmetric encoding of focus. While *it*-clefts exhibit a clear preference for narrow focus, *c'est*-clefts significantly favour broad focus. This not only confirms that syntactic highlighting and prosodic focus are not systematically coupled in clefts, but it also shows that English and French clefts greatly differ in their prosodic focus profile. At the same time, the comparison of the typologies of discourse-familiarity revealed that both constructions afford the same four patterns, new-given, new-new, given-new and given-given. However, *c'est*-clefts were found to instantiate the given-new and given-given types more frequently than *it*-clefts which can be explained by the more numerous anaphoric clefted constituents.

Overall, the analysis of information structure further confirmed the different ways in which *it*-clefts and *c'est*-clefts operationalise prosodic focus assignment and information distribution.

In the last part of this chapter, I reflected on a number of pragmatic features associated with clefts including exhaustivity, contrast and their typical discourse functions. Both *it*- and *c'est*-clefts typically trigger an implicature of exhaustivity which may be cancelled or emphasised either through lexical means, prosodic means, or through a combination of both. Double exhaustivity, i.e. exhaustivity reinforced through lexical and/or prosodic means, was found to be slightly more frequent in English than in French, which constitutes the only asymmetry between the two languages. As far as contrast is concerned, I have shown that *it*- and *c'est*-clefts differ in the way it is encoded. Thus, contrast in English can be linked to contrastive focus which Halliday (1967a) views as a marked type of prosodically coded information structure. French does not make the same use of narrow focus to code contrast. Unlike *it*-clefts, the vast majority of *c'est*-clefts illustrate broad focus but may, in some cases, rely on the initial accent to underline the contrastive reading. Along with the general results on relational information structure, this suggests that English and French do not function in the same way when it comes to the prosodic realisations of clefts. Despite this, the functional perspective showed that *it*- and *c'est*-clefts display a similar array of functions whose scope is either local, i.e. at the sentence or turn level, or more general, i.e. at the textual level. Out of the several uses taken on by clefts, only the cohesive one was found to be more typical of *c'est*-clefts while the rest are more or less equally instantiated.

The main conclusion to draw from this chapter is that while *it*- and *c'est*-clefts share similar characteristics pertaining to their general properties, they also display several major asymmetries in their relational information structure and the way contrast are encoded and some minor differences in the distribution of information. As such, *it*- and *c'est*-clefts should not be taken as true cross-linguistic equivalents but rather as counterparts with specific information structural behaviour.

Chapter 6: The case of reduced clefts

The present chapter focuses on the reduced variant of *it*- and *c'est*-clefts which has received only limited attention in comparison with its full counterparts. The few studies addressing its existence primarily account for it in terms of givenness of its cleft-relative clause, which makes it deletable, but do not delve into its individual properties. The goal of this chapter is to fill this gap by providing a detailed account of the use reduced clefts in spontaneous data. The structure of this chapter is as follows. I start by providing a brief overview of the state of the art in Section 6.1. Then, following the same analytical path as for full *it*- and *c'est*-clefts, I examine the morphosyntactic, pragmatic and information structural characteristics of reduced clefts in both English and French in Section 6.2. I reflect on the functional profile of reduced *it*- and *c'est*-clefts in Section 6.3. Finally, I conclude this chapter with a summary of the main findings in Section 6.4.

6.1 State of the art

In order to better ground the analysis of reduced *it*- and *c'est*-clefts developed in this chapter, this section provides a general overview of the treatment of reduced clefts up to date.

Most descriptions of *it*- and *c'est*-clefts focus on their full variant which is generally treated as the default form of the construction. A few studies have also recognised the existence of a reduced variant whose basic form and function are said to resemble those of full clefts (e.g. Jespersen 1927; Declerck 1988, and Seki 1990; Hedberg 1990, 2000; Büring 1998; Rialland et al. 2002; Doetjes et al. 2004; Belletti 2005; Mikkelsen 2007; Birner et al. 2007; Ward et al. 2007).⁷¹ As has been made clear by now, full clefts always consist of a matrix clause and a cleft-relative

⁷¹ Though they also acknowledge the existence of reduced *it*-clefts, it should be noted that Birner et al.'s (2007) and Ward et al.'s (2007) accounts mainly bear on reduced *that*-clefts, e.g. (1'), which they argue are analogous to reduced *it*-clefts. Because these lie outside the scope and the corpus query of the present study, reduced *that*-clefts will not be addressed.

(1') A: Me? I never wallow. I suffer in silence. B: No, **that's Christine**.

clause. In reduced clefts⁷², by contrast, only the matrix clause is expressed while the cleft-relative clause is omitted due to its high salience in discourse. In this case, the hearer is instructed to infer the variable from the preceding discourse. Full and reduced clefts are productive in both English and French, as exemplified in (1)–(4).

- (1) I think **it is the professional bodies themselves which have got to ask this question** (LLC-1)
- (2) and how she heard repeated bangs on the ceiling thinking **it was her son** [~~who was banging on the ceiling~~] she finally dashed upstairs to to confront him with it (LLC-1)
- (3) comme on est deux on se relaie c'est jamais pareil donc **c'est jamais la même qui va qui va faire le plus d'heures** (CRFP)
 'because there's two of use we take turns it's never the same so it's never the same person who's doing the most hours'
- (4) bon c'est souvent moi mais autrement le soir non le soir il y a on a un livreur qui vient exprès mais le dans la journée c'est moi qui vais livrer ou alors **c'est ma patronne** [~~qui va livrer~~] tout dépend où se situent les livraisons (CRFP)
 'so it's usually me but otherwise in the evening no in the evening there's a we have a delivery man who comes on purpose but the during the day it's me who do the deliveries or **it's my boss** [~~who does the deliveries~~] it all depends on where the delivery is'

The deleted presupposed open proposition may be inferentially recoverable as in (2), in which the fact that someone is banging on the ceiling is inferred from the mention of repeated bangs. It may also be given verbatim in the prior context as shown in (4). In this case, the variable of the reduced cleft is understood as being *x qui va livrer* which is already given in the preceding full cleft. For Hedberg (2000), the presupposed open proposition must not only be given but it must also be in the addressee's focus of attention for it to be deletable. When it is merely activated,

⁷² Reduced clefts are also sometimes referred to as 'truncated' clefts (Hedberg 1990, 2000; Mikkelsen 2007; Birner et al. 2007; Ward et al. 2007).

reduction becomes infelicitous. Hedberg (2000: 900) illustrates this idea with the cleft in (5a) whose variable, i.e. that someone said that, is not salient enough to produce the reduced *it*-cleft in (5b). In comparison, a reduced *this/that*-cleft such as (5c) would be acceptable as “pronoun *this* in general only requires activation” (Hedberg 2000: 900).

- (5) (a) NF: And then, one morning, about three or four or five mornings before I was due to get out, I was lying in bed and someone, one of, one of my fellow soldiers came by and shook my bed and said, 'Come on Fredzo, get up' ... and the Sergeant himself said, 'Leave him alone, he's too short'.
 KF: Hmm.
 NF: I mean, the, **that was the platoon sergeant that said that**. I call that a pretty good guy.
- (b) ?? it was the platoon sergeant
 (c) *this/that* was the platoon sergeant
 (d) *this/it* was the platoon sergeant that said that (Hedberg 2000: 900)

In relation to this, Declerck (1988) argues that a cleft can only be reduced when its clefted NP has specific reference. If the specificity is brought about by the variable, then it cannot be omitted. Declerck (1988: 205-206) illustrates this with the following examples.

- (6) A: One of the pupils ran away.
 B: Was it a friend of yours (who ran away)?
- (7) A: I'd like an ice cream, please.
 B: Is it a big one that you'd like?
 * Is it a big one? (Declerck 1988: 205)

In (6), the specificity of the value is already established through the definite head *the pupils* which makes the variable *x ran away* omissible without rendering the proposition either infelicitous or incongruent. The same cannot be said for (7) in which the co-referent of the value corresponds to the indefinite NP *an ice cream* in

the previous turn and in which the presupposed open proposition must be restated. Thus, the fact that a variable is discourse-given does not necessarily entail its possible omission.

Because of the givenness of their cleft-relative clause, reduced clefts have been argued to illustrate Prince's (1978) *stressed-focus*, Doetjes et al.'s (2004) *focus-ground* and Collins's (2006) *old-presupposition* categories. In other words, reduced clefts are taken to systematically exhibit a new-given information distribution and prosodic prominence on the clefted constituent. From a functional perspective, the speaker's choice to use a reduced cleft has only been illustrated with question/answer pairs in which the reduced cleft takes up the variable (Belletti 2005; Mikkelsen 2007). Strikingly, the same argument is made by Delin and Oberlander (1995) for full focus-background clefts in which the clefted constituent is new and the cleft-relative clause given. The use of reduced clefts therefore does not appear to be validated as an alternative option in its own right.

Within the category of reduced clefts, two further subtypes are identified, but these differ in English and French. For *it*-clefts, Declerck and Seki (1990) distinguish between regular reduced *it*-clefts, e.g. (8) and pre-modified reduced *it*-clefts (PRICs), e.g. (9) and (10).

(8) Who said that?

It was Bill [~~who said that~~].

(9) If anyone can help us, it's John.

(10) When we went somewhere, it was always to some small village or other. (Declerck and Seki 1990: 19)

Declerck and Seki (1990) define PRICs as reduced clefts whose matrix clause is preceded by a pre-modifying clause relating to the omitted presupposed open proposition. In detail, they classify PRICs as relevance conditionals in which the pre-modifying clause, usually introduced by *if* or *when*, asserts the situation under which the matrix clause of the cleft is relevant. This introductory subclause may sometimes contain a pre-form (underlined in (9) and (10)) referring to the type of value conveyed in the cleft. In the case of (9), the *if*-clause presupposes the open proposition *x can help us*. With the pronoun *anyone*, the pre-modifying clause also

narrows down the value to a person, which is then identified as *John* in the cleft. What is presupposed in (10) is that the speaker went *somewhere* which indicates that the value appearing afterwards is likely going to be a spatio-temporal one. Thus, the pre-modifying clause not only spells out the variable but it may also, in some cases, hint at the nature of the value with the insertion of a preform. When the pre-modifying clause does not contain a preform, Declerck and Seki (1990) argue that more than one cleft-relative clause can be retrieved. While Meier (1988) classifies *if*-PRICs as a third type of cleft in addition to *it*- and pseudo-clefts, Declerck and Seki (1990) do not consider the *if*-clause to be part of the cleft and therefore treat the cleft itself as a subtype of reduced cleft. The formal properties of PRICs are also investigated by Lambrecht (2001: 498) who equates *if*-PRICs to a special type of pseudo-cleft motivated by the “lack of a semantically appropriate *wh*-marker”. On his account, (9) cannot be expressed as the pseudo-cleft ‘*Who can help us is John*’ and therefore requires the *wh*-clause to be reformulated into an *if*-clause. Contrary to Lambrecht’s analysis (2001), Declerck and Seki (1990) argue that the *if*-clause of PRICs and the variable of pseudo-clefts do not in fact share the same function. Whereas the latter introduces the presupposed open proposition serving as the variable in the specification relation, the former spells out the variable and involves the speaker’s stance on the likelihood of the matrix clause. In what follows, I adhere to Declerck and Seki’s (1990) description and treat instances such as (9) and (10) as PRICs. Whether the same subclass of reduced clefts is attested in French is examined in Section 6.2.3.

Similarly to Declerck and Seki (1990), Rialland et al. (2002) and Doetjes et al. (2004) also identify a subclass of reduced *c’est*-clefts with specific properties which they refer to as ‘explicative’ cleft sentences. They define explicative clefts as all-focus clefts which provide an answer to a general question such as ‘*What is going on?*’ and whose syntactic structure resembles that of a full cleft. According to them, an answer like a *c’est XP qui* sentence could be interpreted as corresponding to the matrix clause of a reduced cleft whose cleft-relative clause is not the relative *qui*-clause but an unexpressed clause. In (11a), for instance, the reduced cleft provides the hearer with the information that the little one fell down the stairs in answer to the interrogative cleft ‘*Qu’est-ce qui te tracasse ?*’. Though similar structurally, the *qui*-clause is not the cleft-relative clause of the cleft but a restrictive relative

clause which is part of the clefted constituent. The full variant of (11a) is reconstructed in (11b) with the addition of *qui me tracasse*.

- (11) (a) Qu'est-ce qui te tracasse ?
C'est le petit qui est tombé dans l'escalier.
'Why are you so worried?
It's the little one who fell down the stairs.'
- (b) C'est le petit qui est tombé dans l'escalier qui me tracasse.
'It's the little one who fell down the stairs that bothers me.'
(Doetjes et al. 2004: 540)

To summarise the account provided in this section, reduced clefts can be broadly described as full clefts whose variable is so salient in the context that it is omitted. They are productive in both English and French in which they display the same unique information structural behaviour. Relying on data from the LLC-1 and CRFP, the remainder of this chapter will put these assumptions to the test.

6.2 General characteristics of reduced clefts

Before delving into the information structure of reduced clefts, I first focus on their basic properties in both languages. I first comment on their frequency of occurrence, which I compare to that of full clefts in each language and also cross-linguistically in Section 6.2.1. I then report my findings on the morphosyntactic and semantic properties of reduced *it*- and *c'est*-clefts in Section 6.2.2. Building on Declerck and Seki's (1990) study, I end this section by investigating the use of PRICs in French in Section 6.2.3.

6.2.1 Frequency of use of reduced clefts

Applying the different recognition tests tailored specifically for reduced clefts I proposed in Section 2.3.3 allowed me to build a dataset of reduced *it*- and *c'est*-clefts which I then annotated for the same parameters as full clefts pertaining to their general properties. However, despite establishing clear criteria, the analysis

of a number of tokens still proved challenging. To illustrate this, let me take (21) and (22) as examples.

- (12) (a) A: Stalker was involved in a rather long tension about that thing that came up about the sessional exams the other day
 B: sorry what was this
 A: the point is that
 B: can you ink ink me in on this
 A: well I talked to you about it it it's quite
 B: I if **it's the Renaissance paper** they were too difficult (LLC-1)
- (b) ...if it's the Renaissance paper [~~you're talking about~~] they were difficult
- (c) ...if the thing that came up is the Renaissance paper they were too difficult
- (13) (a) A: et les fleurs elles viennent comment elles vous allez les chercher ou
 B: non c'est des fournisseurs qui passent on a des fournisseurs qui passent tous les jours on monte dans le camion choisit nos fleurs donc suivant s'il y a des commandes si on a une commande de roses par exemple on va prendre notre botte de roses si **c'est une commande un bouquet varié** bon ben on va prendre les couleurs les couleurs que le client déjà demandées (CRFP)
 'A: and the flowers how do they come they do you pick them up or
 B: no it's the suppliers who pass by we have suppliers passing by every day we get in the truck pick our flowers so depending on whether there are orders if we have an order for roses for instance we take our bundle of roses if **it's an order for a varied bouquet** well we take the colours the colours that the customer already asked for'
- (b) ...if it's an order for a varied bouquet [~~that we have~~]
- (c) ...if the order that we have is an order for a varied bouquet

In (21a), two readings are available for the *it*-sentence in boldface. In the cleft reading, one could assume that speaker B is using the *c'est*-sentence as a way to get confirmation on the nature of the thing that is being talked about. The corresponding full cleft could then be reconstructed as the cleft in (21b) whose variable *you're talking about x* arguably lacks salience in the context. The *c'est*-sentence could also be analysed as a specificational-identifying clause with referential subject *it*. This would be consistent with the multiple anaphoric references to *that thing* (underlined in the excerpt) in the turns leading up to the *it*-sentence. In this case, it can be rephrased as (21c) where *it* is replaced by the coreferential NP *the thing that came up*. In the same vein, the *c'est*-sentence in (22a) can also be construed as the cleft in (22b) or the monoclausal copular construction in (22c). Here, the salience of the proposed cleft-relative clause in (22b) appears to be stronger than that of (21b), as shown by the hypothetical scenario given in the prior context. In cases in which no reading revealed itself to be significantly more felicitous or acceptable than the other, as in (21) and (22), the token was dismissed from the dataset. The distribution of full, reduced and ambiguous clefts attested in the LLC-1 and CRFP is shown in Table 1.

	Full clefts		Reduced clefts		Ambiguous		Total of clefts	
	RawF	NormF	RawF	NormF	RawF	NormF	RawF	NormF
LLC-1	143	2.86	95	1.9	93	1.86	238	4.76
CRFP	392	8.9	88	2	46	1	480	10.9

Table 1. Raw (RawF) and normalised (NormF - per 10,000 words) frequencies of full, reduced, and ambiguous *it*- and *c'est*-clefts

The distribution reveals a higher frequency of occurrence for full clefts in both languages. However, the ratio between the two variants is higher for *c'est*-clefts than for *it*-clefts. The same is true for ambiguous cases of reduced clefts which are proportionally more numerous in English than French. This suggests that the choice to omit a discourse-given variable is more typical of English speakers.

6.2.2 Morphosyntactic properties

To account for the properties specific to reduced clefts, I first comment on the form and function of their clefted constituent, which I compare to that of full clefts. Table 2 below summarises the distribution of syntactic categories.

	<i>It</i> -clefts		<i>C'est</i> -clefts	
	Full	Reduced	Full clefts	Reduced clefts
Noun phrase	109 (76.2%)	68 (71.6%)	221 (56.4%)	40 (45.4%)
Prepositional phrase	12 (8.4%)	14 (14.7%)	100 (25.5%)	22 (25%)
Adverbial phrase	5 (3.5%)	1 (1.1%)	66 (16.8%)	—
Finite clause	8 (5.6%)	6 (6.3%)	3 (0.8%)	22 (25%)
<i>WH</i> -word	9 (6.3%)	4 (4.2%)	—	—
Adjectival phrase	—	2 (2.1%)	—	—
Non-finite clause	—	—	2 (0.5%)	4 (4.6%)
Total	143 (100%)	95 (100%)	392 (100%)	88 (100%)

Table 2. Distribution of syntactic categories of the clefted constituent in full and reduced *it*- and *c'est*-clefts

Starting with *it*-clefts, the distribution of the different syntactic categories is similar for both full and reduced forms. Only one exception is to be noted for PPs whose frequency is higher in reduced *it*-clefts. By contrast, full and reduced *c'est*-clefts show more variation. Among the differences observed, AdvPs, which are productive in full *c'est*-clefts, are non-existent in reduced *c'est*-clefts. The opposite distribution is found for finite clauses which are significantly more frequent in reduced *c'est*-clefts and scarce in full ones. These differences can be further characterised by carrying out a qualitative analysis of the tokens. What doing so shows is that full *c'est*-clefts with an adverbial clefted constituent mainly select three specific adverbs which are *ainsi*, *là* and *comme ça*. That reduced clefts do not allow for the selection of any of these anaphoric AdvPs hints at a disfavour of discourse-given clefted constituents. Whether this is confirmed by the analysis of discourse-familiarity will be dealt with in Section 6.3.1. With regard to the higher frequency of finite clauses in the postcopular position, the qualitative review of

occurrences reveals that all but two reduced *c'est*-clefts with a finite clause as their value select a causal *parce que*-clause as their clefted constituent. It is the case of (23) in which the reduced cleft is a PRIC-like construction. In comparison, only two full clefts have a causal value of the kind, which suggests that reduced clefts are particularly suited to syntactically highlight causal clauses.

- (14) ben ouais parce que moi enfin si je connais le mot Boche **c'est bien pour parce qu'on me l'a dit hein** et puis on me l'a répété hein mes grands-parents euh c'est les allemands non c'est pas les allemands c'est des Boches (CRFP)
 'yeah I mean because I well if I know the word Boche (note: pejorative term used to refer to German soldiers during World War I) **it's well for because someone said it to me it right** and well it was repeated to me right my grand parents uh it's the Germans no it's not the Germans it's the Boches'

In addition to differences between full and reduced variants in each language, the findings also highlight a number of cross-linguistic asymmetries. The most striking dissimilarity is that reduced *c'est*-clefts instantiate fewer syntactic categories than *it*-clefts, though it should be noted that the categories exclusively found only in *it*-clefts are not greatly productive. As for the categories shared by the two languages, NPs are more common in reduced *it*-clefts while reduced *c'est*-clefts select PPs and finite clauses more frequently.

Turning to the grammatical function of clefted constituents, Table 3 below presents the quantitative findings for full and reduced *it*- and *c'est*-clefts.

	<i>It</i> -clefts		<i>C'est</i> -clefts	
	Full	Reduced	Full	Reduced
Subject	87 (60.8%)	33 (34.7%)	202 (51.5%)	27 (30.7%)
Direct object	13 (9.1%)	15 (15.8%)	17 (4.4%)	9 (10.2%)
Indirect Object	—	1 (1.1%)	4 (1%)	—
Adjunct	32 (22.4%)	27 (28.4%)	108 (27.5%)	50 (56.8%)
Conjunct	—	—	61 (15.6%)	—
Complement of a preposition	11 (7.7%)	18 (18.9%)	—	2 (2.3%)
Subject complement	—	1 (1.1%)	—	—
Total	143 (100%)	95 (100%)	392 (100%)	88 (100%)

Table 3. Distribution of grammatical functions of the clefted constituent in full and reduced *it*- and *c'est*-clefts

The overall distribution of functions in *it*-clefts shows that full *it*-clefts typically select subjects and, to a lesser extent, adjuncts, direct objects, and complements of a preposition. By contrast, reduced *it*-clefts do not display as clear of a preference for one category over the others and instead select a more balanced mix of subjects, adjuncts, complements of a preposition, and direct objects. Subjects and adjuncts are still the most frequently exemplified categories but less so than in full *it*-clefts. When using full *c'est*-clefts, French speakers tend to select either subjects or adjuncts as their clefted constituent. For reduced *c'est*-clefts, adjuncts are the most common function before subjects, direct objects and complements of a preposition. The preference for adjuncts does not come as a surprise given the higher frequency of causal values identified earlier. Conjuncts, however, are only attested in full *c'est*-clefts. The contrastive analysis of reduced clefts reveals a significantly higher proportion of adjuncts in *c'est*-clefts while the complement of a preposition and direct object categories are more common in reduced *it*-clefts. As such, reduced *c'est*-clefts display the same preference for adjuncts observed in full *c'est*-clefts. Finally, subjects appear with a similar frequency of occurrence in both languages.

Whether for English or French, the reduced variant in the LLC-1 and CRFP generally differs from its full counterpart in the number and proportion of syntactic categories and grammatical functions instantiated in the clefted constituent. This

difference, which had not been addressed in the literature, shows that full and reduced clefts should be distinguished as distinct and fully-fledged alternatives displaying specific morphosyntactic properties.

6.2.3 French PRICs: reduced clefts or pseudo-clefts?

Pre-modified reduced *it*-clefts, or PRICs, are a subtype of clefts which have only been formally recognised in English but not in French. While Declerck and Seki (1990) analyse them as a subtype of *it*-clefts, Lambrecht (2001: 498) categorises them as pseudo-clefts. In this section, I examine whether a similar construction is also productive in French, and if it is the case, whether their characteristics are that of reduced *c'est*-clefts or pseudo-clefts. The starting dataset comprises 15 occurrences of reduced clefts whose basic structure seemingly resembles that of PRICs. Like PRICs, the instances of reduced *c'est*-clefts have different types of pre-modifying clauses which include *si*-, *quand*- and *où*-clauses, all of which are also found in English. The first two types are instantiated in (24) and (25) below.

- (15) (a) ben de toute façon si je donne **c'est parce que euh j'espère beaucoup recevoir** (CRFP)
 'well in any case if I give **it's because uh I hope to receive a lot'**
 (b) If $x = I$, if x gives, and if x does so for reason y , then $y =$ because x hopes to receive a lot
- (16) euh ouais j'ai pas précisé que quand on mettait en œuvre les dispositifs électro-acoustiques **c'est toujours sur des bases d'improvisation d'improvisation de l'interaction** (CRFP)
 'uh yeah I didn't mention that when we would use electro-acoustic devices **it's always on an improvisation basis improvisation of interaction'**

When the subclause is introduced by *if/si*, *when/quand* or *where/où*, as in (24a) and (25a), it serves to define the cases in which the specificational relation set up between the value and the variable is true (Declerck and Seki 1990: 39). In (24a), the relation established between the value *parce que j'espère beaucoup* and the

variable *je donne parce que x* is only relevant in the case where the speaker *does* give. Following Declerck and Seki's (1990) semantic analysis, the logical structure of (24a) is then (24b). This reading allows to bring out the relevance conditional nature of the *si*-clause. While *if*-clauses bind causal variables, *when*- and *where*-clauses bind temporal and spatial variables respectively. They differ from *if*-clauses in so far as they do not imply uncertainty and instead present the state-of-affairs conveyed in the pre-modifying clause as a fact. It is the case of (24), in which the fact the speaker used electro-acoustic devices at some point is asserted rather than presented as a condition. Of the 15 instances of PRIC-like constructions extracted from the CRFP, 10 have a conditional *si*-clause and 5 a factual *quand/où/dès que*-clause, none of which contain a pre-form lexicalising of the variable. So far, my findings suggest that PRICs also exist in French and that the two cross-linguistic equivalents share the same syntactic and semantic properties.

Let me now consider the question of whether French PRICs are reduced *c'est*-clefts, as projected by Declerck and Seki's (1990) account, or a subclass of pseudo-clefts, as posited by Lambrecht (2001: 498). In contrast with PRICs, pseudo-clefts are traditionally treated as specificational copular clauses forming a full-fledged type of clefts along with *it*-clefts and reverse pseudo-clefts (Huddleston and Pullum 2002).⁷³ In French, their pre-subclause, can either be either a syntactically complete clause, e.g. *il y a une chose qui m'étonne* in (26), or a NP-like incomplete clause, e.g. *une chose qu'on peut espérer* in (27).

- (17) *il y a une chose qui m'étonne c'est qu'un éditeur n'ait pas pensé à vous*
(Apothéloz and Roubaud 2015)
- (18) *une chose qu'on peut espérer c'est que l'Université soit revalorisée*
(Apothéloz and Roubaud 2015)

Regardless of its status, the left segment of French pseudo-clefts always introduces an underspecified referent which is then taken up and identified as value in the right segment of the construction. On that basis, Apothéloz and Roubaud (2015)

⁷³ Pseudo-clefts are also sometimes designated as *wh*-clefts (e.g. Declerck 1988; Delin 1990; Lambrecht 2001; Collins 2006)

equate the semantics of pseudo-clefts to that of ‘construction désignative’ (*designating construction*) in which the left part lexically categorises a referent and the right part designates and further identifies it. In (26), for instance, what is conveyed is *la chose qui m’étonne = aucun éditeur n’a pensé à vous*. This description differs from that of PRICs whose specificational meaning is established between two distinct referents or state-of-affairs which are not equated. Thus, in an instance like (28), the subclause and the clefted constituent introduce two separate situations, i.e. *je m’intéresse au tarot* and *parce que je vais tirer les cartes*, which are not construed as a categorised/identified pair but as a value/variable combination. This is comparable to the specific relation established in regular reduced clefts like (29). Here too, the cleft allows to identify the value *parce que c’est pas intéressant* as that exhaustively satisfying the omitted variable *vous ne travaillez pas à l’extérieur parce que x*.

- (19) si je m’intéresse au tarot **c’est pas parce que je te vais tirer les cartes**
 j’en ai rien à foutre de tirer les cartes mais c’est pour la l’ initiation
 symbolique il y a derrière le tarot (CRFP)
 ‘if I’m interested in tarot **it’s not because I’m going to draw cards** I
 don’t give a fuck about drawing cards but it’s for the symbolic initiation
 that’s behind tarot’
- (20) A: je ne travaille pas euh je travaille que pour moi pour mes enfants
 B: oui
 A: je travaille pas pour l’extérieur c’est pas intéressant
 B: oui **c’est parce que c’est pas intéressant [que vous ne travaillez
 pas à l’extérieur]**
 A: non (CRFP)
 ‘A : I don’t work uh I only work for me for my children
 B: yes
 A: I don’t work for anyone external it’s not interesting
 B: yes **it’s because it’s not interesting [that you don’t work for
 anyone external]**
 A: no’

French PRICs therefore share their semantics with regular reduced *c'est*-clefts and not pseudo-clefts. From a pragmatic viewpoint, another difference between pseudo-clefts and PRICs lies in the presupposition triggered by the subclause. While the subclause of pseudo-clefts always presupposes the indisputable existence of the variable at stake, *if/si*-PRICs pose a condition on the truth of the variable it presupposes and conveys the speaker's stance on the probability of condition fulfilment. The condition encoded by the *if/si*-clause may thus be presented as a fact, as an open possibility, as an unlikely possibility or as a non-fact (Declerck and Seki 1990: 44). This opposition is illustrated in (30) and (31a). In (30), the subclause *ce que je préfère* presupposes that there exists something that the speaker prefers in the field in which he works, which is then specified as *l'animation de soirées*. By contrast, in (31a), the *si*-clause does not hold any presuppositional meaning. Its role is here to restrict the relation between the variable *je raconte cette histoire parce que x* and the value *parce que ça m'est arrivé le quatre avril* to the cases in which the speaker would in fact be telling the story. This is made explicit by the logical structure in (31b).

- (21) ouais ben **ce que ce que je préfère dans ce domaine-là c'est l'animation de soirées** (CRFP)
 'yeah well what I prefer in this field is hosting parties'
- (22) (a) si si je vous raconte cette histoire c'est parce que ça m'a touché (CRFP)
 'if I'm telling you this story it's **because it touched me**'
- (b) If $x = I$, if x is telling this story, if x does it for reason y , then $y =$ because it touched x

The comparison of French PRICs with reduced *c'est*-clefts and pseudo-clefts leads to the conclusion that their semantic and presuppositional behaviour matches that of the former and not the latter. This is line with Declerck and Seki's (1990) analysis of English PRICs.

In sum, I have argued in this section that English pre-modified reduced *it*-clefts, as described by Declerck and Seki (1990), have a French equivalent displaying the same syntactic, semantic and pragmatic characteristics. Drawing from the CRFP

data, I have corroborated Declerck and Seki's (1990) claim according to which PRICs should be analysed as a subtype of reduced clefts.

6.3. Information structure of reduced clefts

Having discussed the basic morphosyntactic characteristics of reduced clefts, I now focus on their referential (Section 6.3.1) and relational information structure (Section 6.3.2), before accounting for their discourse uses (Section 6.3.3).⁷⁴

6.3.1 The referential information structure of reduced clefts

Reduced clefts, whether PRICs or regular reduced clefts, are traditionally characterised as clefts whose cleft-relative clause is omitted due its high degree of recoverability. With regard to their overall information structural articulation, the different typologies addressing the use of reduced clefts define them as clefts whose clefted constituent is new and often contrastive while the cleft-relative clause is given. To test these predictions, I first examine the discourse-familiarity patterns of reduced *it*- and *c'est*-clefts, whose distribution is summarised in Table 4 below.

	Reduced <i>it</i> -clefts	Reduced <i>c'est</i> -clefts
New-given	54 (56.8%)	58 (65.9%)
New-new	—	—
Given-new	—	—
Given-given	36 (37.9%)	30 (34.1%)
Unclear	5 (5.3%)	—
Total	95 (100%)	88 (100%)

Table 4. Distribution of patterns of discourse-familiarity in reduced *it*- and *c'est*-clefts

⁷⁴ Parts of this subsection build on work presented in Bourgoin et al. (2021).

While the results of my analysis on the discourse-familiarity confirm the systematic discourse-givenness of the cleft-relative clause of both reduced *it*- and *c'est*-clefts, they reveal a more heterogeneous picture with respect to the informational status of their clefted constituent. In both English and French the value may be discourse-new or discourse-given, with the former being the most frequent category. This goes against the description of reduced clefts provided by some of the existing typologies of clefts (e.g. Prince 1978; Collins 2006) which assumed a systematic newness. To account for this difference let me consider (32)–(35) below.

- (23) is there any point in bringing anything or you know has the ship sailed on the second# and he ^said no **it's the tw\entieth**# [~~that it will sail~~]
(LLC-1)
- (24) A: A: ben je suis un peu spécialiste de la bande dessinée ici à la bibliothèque déjà et puis j'en ai fait un peu oui parce que ça me plaisait et puis j'ai essayé puis en fait en fait je me suis rendu compte que ça ne me correspondait pas c'était trop carré trop fermé pour moi
B: trop carré trop fermé
A: oui
B: c'est-à-dire [**c'est le cadre de l'image**]_{H/H} [~~qui est trop fermé~~]
A: eh ben oui (CRFP)
'A: well I'm a bit of a specialist of comic books here at the library first of all and then I also did a little bit of it yes because I liked it and then I tried then actually I realised that it wasn't for me it was too strict too narrow for me
B: too strict too narrow
A: yes
B: what do you mean **it's the frame** [~~that's too narrow~~']
- (25) A: yes well now one thing that I have been asked about presumably you didn't have any hand in this in putting me forward as a referee some publisher or other last year about a year ago asked me what I thought about a series publishing early grammarians
B: oh Perrins perhaps

A: because I forget who it was# but I didn't ^think **it was Perrins#**
~~[who asked me that]~~ it could be (LLC-1)

(26) A: hum hum et est-ce que vous ressentez une différence entre l'IUT et
 et l'université la façon de travailler les é- les étudiants que vous croisez
 dans

B: les non oui c'est sûr que bon j'y suis là il y a pas très longtemps ici
 mais à la fac quoi mais c'est vrai qu'à l'IUT c' était plus sympathique on
 a j'ai sympathisé plus vite qu' ici les étudiants ils étaient plus euh plus
 sympa euh bon [c'est peut-être parce que je suis pas là depuis
 longtemps]_{HH} ~~[que j'ai sympathisé moins vite ici]~~ ça va peut-être
 (CRFP)

'A: hm hm and do you feel like there's a difference between the UIT
 and and university in the way people work the s- the students you run
 into in

B: the no yes of course I'm here not too long ago ici but at university
 but it's true that at UIT it was more pleasant we I got on well more
quickly than here students they were more uh nicer well **it may be**
because I haven't been here for too long ~~[that I got on more slowly~~
~~here]~~'

(32) and (33) exemplify the new-given category while (34) and (35) are of the new-given type. In the first two examples, the clefted constituent is not only discourse-new, but it also holds explicit contrastive value. In (32), the value *the twentieth* is explicitly contrasted with the referent *the second*, both of which are related to the date of sailing. Though less marked, (33) also establishes an overt relation of contrast between *le cadre* and other potential values, e.g. the job market for comic books artist, the work environment, etc., exhaustively satisfying the variable *x est trop fermé*. Because the clefted constituents of (32) and (33) are discourse-new, contrastive, and prosodically focal, the two instances illustrate Prince's (1978) *focus-background* and Collins's (2006) *old-presupposition* clefts, which they argue are the only types which accept reduction. Compare them now to (34) and (35). In both cases, the values, i.e. *Perrins* in (34) and *j'y suis pas depuis très longtemps* in (35), have already been introduced in the preceding discourse. By taking them up

and relating them to the variables *x asked me that* and *j'ai sympathisé moins vite parce que x* respectively, both clefts allow to spell out the specification relation at play between the constituents. As such, what is new is the specification relation itself, which is made explicit and salient by the clefts. Interestingly, this use had already been identified for full *it*- and *c'est*-clefts (see Sections 3.3.2 and 4.3.2) which reveals an overlap in some of the functions displayed by full and reduced clefts.

The quaternary typology proposed in Section 2.4, which distinguishes between two types of given and two types of new, allows me to further characterise how discourse-familiarity is operationalised in reduced clefts. For this, I consider the discourse-familiarity categories found in the variable, whose distribution is shown in Table 5 below.

	<i>It</i> -clefts		<i>C'est</i> -clefts	
	Full	Reduced	Full	Reduced
Evoked	36 (25.2%)	70 (73.7%)	49 (12.5%)	77 (87.5%)
Inferable	42 (29.4%)	25 (26.3%)	126 (32.1%)	11 (12.5%)
New-anchored	22 (15.4%)	—	39 (10%)	—
Brand new	41 (28.6%)	—	178 (45.4%)	—
Unclear	2 (1.4%)	—	—	—
Total	143 (100%)	95 (100%)	392 (100%)	88 (100%)

Table 5. Distribution of discourse-familiarity types in the variable of reduced *it*- and *c'est*-clefts

As is revealed by the analysis, the variable of reduced clefts can be of two types of given. In the majority of tokens in both languages, the presupposed open proposition is evoked, which means that it is explicitly mentioned in the prior context. It is the case of (33) and (34) in which the fact that something is too narrow and that someone asked the speaker about a series publishing early grammarians is already asserted. By contrast, the fact that there is something the speaker did not like about Andrew Ray in (36) is not given as-is but can be inferred from her mentioning that she did not like him in general. Similarly, the very fact that the

speaker's newspaper is sent to customers in (37) implies that someone takes care of the packaging process.

(27) and I can't say I liked Andrew Ray very much he stood# a I ^think it was the way he \stood# [~~that I didn't like~~] he hasn't learnt to st to stand yet (LLC-1)

(28) le moyen le plus sûr de se procurer le journal est de s'abonner notamment quand on se se trouve en province et voilà et puis au moins on l'a dans sa boîte aux lettres (...) oui c' est du boulot aussi ouais c'est du boulot c'est de l'organisation la mise sous enveloppe la mise sous pli pour les abo- pour les abonnés [**c'est nous aussi**]_{HH} [~~qui la faisons~~] (CRFP)

'the safest way to get the journal is to subscribe for instance when one is is in the countryside and at least one receives it in one's mailbox (...) yes it takes work too but yeah it takes work it takes organisation putting them in envelopes putting them in envelopes for sub- for subscribers **it's us too** [~~who do it~~']

That the cleft-relative is more frequently evoked than inferable in both English and French suggests that the more salient the variable is in the context, the more easily deletable it is.

While it would be tempting to assume that reduced clefts are simply the unmarked option for clefts with discourse-given variable, the picture is in fact more complex. As shown by the 149 tokens of full clefts with an evoked/inferable variable, discourse-given variables are not necessarily omitted by speakers. Thus, a variable that is in focus⁷⁵ in Hedberg's (2000) terminology, i.e. representing a highly salient matter of current concern, like that in (36) and (37), can be omitted but a variable that is activated but not in focus will most likely yield a full *it*-cleft and

⁷⁵ Unlike in my analysis, Hedberg's (2000) use of the notion of focus is not related to prosody but to the description of the givenness hierarchy she draws up. 'In focus' is described as a cognitive status, i.e. an "attention state[s] that a speaker can assume the intended interpretation has for the addressee" (2000: 895) and is placed at the top of the hierarchy consisting of eight levels, namely focus > activated > familiar > uniquely identifiable > referential > type identifiable.

not a reduced one. This is exemplified in (38) where the reduced cleft *it was your wife's phone message* would not be enough to convey the specification relation between the value and the variable *x* *caused the doctor to bring the details of the nursing homes* as it is not in focus or salient enough.

- (29) A: yes but that would account for him bringing the details of the nursing-home
 B: well not necessarily no I phoned him on Friday morning and told him I wanted mother to go into a nursing-home
 A: yes
 B: and the doctor said he thought it was not necessary at all
 A: yes quite right you see what I'm (coughs) asking is this your suggestion is# that ^{H^}it was your w^{H\}ife's phone message# that ^{M^}caused the d^{M\}octor# to ^{M^}bring d^{M\}etails# of the n^{H\}ursing-homes# (LLC-1)

The choice of a full cleft with a discourse-given variable over a reduced form may be motivated by interactional and rhetorical reasons. In (39), for instance, the full cleft uttered by B comes as an answer to A's question, itself also in the cleft format, and takes up information regarding his role within the newspaper company that was previously mentioned. The cleft allows A to reuse both information that he has himself shared before, i.e. directing politicians, as well as information shared by his interlocutor, i.e. orienting them. The new information can thus be considered the tying of the two variables which share the same value.

- (30) A: c'est à moi en fonction de la spécialisation du politique de dire bon bah voilà il faudrait que vous ayez la réaction de cette personne
 B: ah donc allez voir cette personne c'est toi qui orientes
 A: voilà et **c'est moi qui voilà qui oriente et qui euh qui qui oriente et qui dirige le journaliste** (CRFP)
 'A: it's up to me depending on the politician's specialisation to say well there you go you should get this person's reaction
 B: ah so you go up to that person it's you who guide

B: exactly and it's me indeed who guide and who euh who who guide and who lead the journalist'

Overall, although my results are in line with the main descriptions on reduced *it*- and *c'est*-clefts with regard to the systematic discourse-givenness of their cleft-relative clause, they highlight the fact that different types of given may be selected, with evoked variables being more prevalent in both languages. The clefted constituent also displays informational variety and does not always introduce new information as suggested by the different typologies. The comparison with full clefts shows that the choice between a full and reduced variant does not solely depend on the discourse-familiarity of the cleft-relative clause but also stems from the different discourse functions. I develop this further in Section 6.3.3.

6.3.2 The relational information structure of reduced clefts

While the prosodic profile of full *it*- and *c'est*-clefts has received attention in the literature (see Prince 1978; Declerck 1984; Collins 2006; Doetjes et al. 2004; Mertens 2012 among others), the same cannot be said for their reduced variants. The few accounts addressing its use predict that the type of focus encoded is either narrow and borne by the clefted constituent (e.g. Prince 1978; Collins 2006) or all-focus in the case of explicative clefts (Rialland et al. 2002; Doetjes et al. 2004). I start by examining the prosodic patterns of reduced *it*-clefts, which are shown in Table 6.

Prosodic pattern	Reduced <i>it</i> -clefts
Focal + 0	89 (93.7%)
Focal + 0 + focal copula	1 (1%)
Non-focal + 0 + focal copula	5 (5.3%)
Total	95 (100%)

Table 6. Prosodic patterns of reduced *it*-clefts

With regard to the prosodically coded information structure, reduced *it*-clefts overall show much less variation in patterns due to the omission of the cleft-relative clause. The value is in the vast majority of cases focal, as illustrated in (40), except for a minority of cases in which the copula carries the information focus, as is the case in (41). Here, focus is shifted on the copula to emphasise the accuracy of the specification relation. The other occurrences of focal copula correspond to dependent interrogatives.

- (31) A: but you haven't gone have you
 B: well **it was a ^couple of years \ago# ^couple of y\ears ago# [~~that I went~~]** (LLC-1)
- (32) A: I've got a list of examiners nominated for MA examinations in English (...) for the candidates from Panamerican College and I'm trying to establish which branch they should be put under (...)
 B: no I'm I'm sure it's the modern English the modern language one
 B: **it \is the modern language# [~~that they should be put under~~]** well both of them should be on the modern board (LLC-1)

As far as the distinction between contrastive and presenting focus is concerned, I argue that reduced clefts with focal value instantiate the information structure of contrastive focus-presupposition. For instance, in (42), there is a contrastive focus on *wine* (as a member of the set *all sorts of things*).

- (33) A: cos all sorts of things go into port don't they I mean **it's ^not just w\ine# [~~that go into port~~]** it's
 B: like toenails [~~that go into port~~] (LLC-1)

There is no overt presupposition, but Halliday (1967a: 206) points out that information can be recoverable “anaphorically, by reference, substitution or *ellipsis* [italics mine]”. Like full clefts, reduced clefts express identifying-specificational semantics, which instruct the hearer to infer the variable from the preceding discourse. Hence, the presupposition in clefts like (42) can be viewed as anaphorically recoverable by ellipsis, and as part of the information structure.

These *it*-clefts are chosen when the speaker wants to combine the exhaustive specification relation conveyed by the grammatical organisation with the prosodically coded information structure of a contrastive focus relating to a presupposition, which is the unmarked information structure choice within a cleft environment.

Turning to *c'est*-clefts, the different prosodic patterns observed in the CRFP are summarised in Table 7.

Prosodic pattern	Reduced <i>c'est</i> -clefts
[<i>c'est</i> X] _{L-L-/HL-/H/H}	44 (50%)
[<i>c'est</i> X] _{HH/LH}	44 (50%)
Total	88 (100%)

Table 7. Prosodic patterns of reduced *c'est*-clefts

Like full *c'est*-clefts, the prosodic realisations of reduced *c'est*-clefts can be categorised according to the type of boundary tones. In the first category, the clefted constituent displays a terminal boundary tone which is either assertive, e.g. L-L- or HL-, or interrogative, e.g. H/H. This is exemplified in (43) in which the adjunct displays a terminal assertive tone at its right edge thus marking the end of the informational object. In Mertens's (2012) taxonomy, this pattern encodes broad focus scoping over the whole construction, which, given the omission of the cleft-relative clause, triggers a certain salience on the clefted constituent. In the case of the HL- tone, this salience is emphasised prosodically.

- (34) A: si l'a-- l'assistante sociale de secteur connaît déjà la famille elle suit la famille c'est elle qui me fait le dossier moi donc pas toutes pas toutes
 B: hum hum voilà c'est ça ce n'est pas vous qui faites forcément l'enquête

A: [*c'est uniquement quand l'assistante sociale de secteur ne connaît pas*]_{L-L-} [~~que je fais le dossier~~] (CRFP)

'A: if the so~ the local social worker already knows the family she follows the family it's her who writes the file for me so not all of them not all of them

B: hm hm there you go it's not necessarily you who investigate

A: **it's only when the local social worker doesn't know [~~that I do the investigation]~~'**

In the second pattern, the rightward syntactic boundary of the reduced cleft does not bear a terminal tone but a continuative one which implies that the information object is not complete yet. As discussed in Section 4.2.2, and as argued by Avanzi (2011), this can be analysed as a floor-holding strategy whereby the speaker indicates that s/he is going to add more information. It is the case in (44), in which the reduced cleft is immediately followed by a subordinate clause detailing the reason why the speaker only goes to the workshop once a week. Here, the syntactically-coded emphasis on the clefted constituent is not matched by prosodically-marked narrow focus. As a result, the clefted constituent shows a lower level of salience as reduced clefts of the first type.

- (35) donc euh l'atelier danse se se j'y vais une fois par semaine quelquefois deux mais souvent [**c'est qu'une fois**]_{HH} [~~que j'y vais~~] parce que je suis tellement prise par ailleurs que je peux pas tout faire
 'so uh the dance workshop d d I go there once a week sometimes two but often **it's only once** [~~that I go~~] because I'm so busy elsewhere that I can't do everything'

Thus, in comparison with reduced *it*-clefts, reduced *c'est*-clefts show more prosodic variation. More importantly, the syntactic and prosodic means to encode emphasis are not systematically mapped onto one another.

6.3.3 A functional perspective

Much like their overall description, the functional characterisation of reduced clefts has so far remained limited. Its use has mainly been illustrated with question/answer pairs or other contexts in which the variable is somewhat present in the prior context. The qualitative analysis of the LLC-1 and CRFP datasets allows me to refine this account. My findings show that there are three discourse contexts

in which speakers are particularly inclined to select reduced clefts, which, like full clefts, I categorise according to the localised or general function of the cleft.

Firstly, as already noted in the literature, reduced clefts are often used in answer to *wh*-questions, which presuppose a proposition and inquire into one of its elements. In (45), speaker B is asked about the reasons driving Chaucer's pardoner to behave the way he does. A answers first with the NP *not something outside himself*, indicating what value does *not* fill the semantic gap in the presupposed proposition, and then produces a reduced cleft specifying the values that do fill the gap in the open proposition, which itself does not need to be restated. A similar analysis can be made of the reduced *c'est*-cleft in (46). Here, speaker A explains that he wanted to go to boarding school as a teenager, after which speaker B inquires about the reasons behind it. A provides the answer by using a cleft whose variable, i.e. *je voulais partir en pension parce que x*, is already laid out in the corresponding question and is hence easily omittable.

- (36) A: and what is the pardoner being driven by
 B: well not not something outside himself really# **it's his ^{H^}own - desire for m^H/oney#**
 A: hm
 B: I suppose# ^{H^}and a s^H\Vort of power# [~~which he is driven by~~]
 (LLC-1)
- (37) A: et je voulais euh partir euh partir en pension en fait c'était une envie que j'avais d'e~ d'essayer ce que c'était
 B: pourquoi tu voulais partir en pension
 A: hum [**c'était pour prendre un peu de distance_{LL} avec euh mes parents certainement**]_{HL}-[~~que je voulais partir en pension~~] (CRFP)
 'A : and I wanted uh to go to boarding school actually it was a wish I had to tr~ to try it out
 B: why did you want to go to boarding school
 A: hm it was to distance myself a little bit from uh my parents most likely [~~that I wanted to go to boarding school~~']

Because this type of reduced clefts generally reacts to a specific proposition given in the prior context, I consider their use to be local.

Secondly, reduced *it*- and *c'est*-clefts are frequently used to establish an overt contrast between two or more values. For example, the two speakers in (47) are debating the question who was the most powerful person in the family. Speaker A first uses a *wh*-interrogative pseudo-cleft, which asks this question in an open-ended way (who is the person who has the ultimate say about things?) and then produces an interrogative *it*-cleft. The nuclear accent on the copula *be* conveys that speaker A questions speaker C's suggestion that it is the daughter who is the dominant person. Speaker A ends with a reduced interrogative *it*-cleft contrastively proposing *the father* as value. In French, reduced clefts are particularly productive in contexts in which two values are introduced one after the other, the first of which is discarded to make way for the second value successfully exhaustively satisfying the variable. The contrast is corrective here. This type of sequential polarity change accompanying the reduced cleft, which had already been identified as typical of French by Rouquier (2007), is exemplified in (48). The negated value *parce qu'elle est originale* is immediately followed by the accurate value packaged in an inferential cleft.

- (38) A: would you say that Dad is really the powerful person in the family
 C: I'm not sure that the girl isn't in a way [...]
 A: we were saying at least that I I feel it's very important to get a to get to know who is the person who really hm has the ultimate say about things in the home
 B: hm
 A: whether it^M/is this daughter# who's^H taken on the role of
 C: hm
 A: m^M/other# - or ^Hwhether it is f^Mather# [~~who's taken on the role of mother~~] because there were sort of hints about father being a fairly well suggesting that he was quite a severe man (LLC-1)
- (39) ils vous disent ah moi je préfère une édition originale mais [**c'est pas parce qu'elle est originale**]_{L-L} [~~qu'ils la préfèrent~~] c'est qu'elle est a-

elle est elle est imprimée avec un caractère agréable avec des marges agréables

'they tell you ah but I prefer an original edition but **it's not because it's original [~~that they prefer it~~]** it's that it's a~ it's it's printed with a pleasant special character with pleasant margins'

Like question/answer pairs, the establishment of an overt contrast by means of a reduced cleft demonstrates a narrow functional scope, which is attested cross-linguistically.

When their information is partitioned into a given-given sequence, reduced *it*- and *c'est*-clefts may also fulfil a textual role playing out at the discourse level. The cleft is in this case used to relate the value and variable, which are both recoverable from the preceding context but which had remained unassociated thus far. For that reason, it typically appears towards the end of a discourse segment and/or forms the conclusion to a speaker's thought process. (34) and (35), repeated as (49) and (50) below, illustrate this use for English and French respectively. In both cases, the information conveyed in both parts of the cleft is progressively built up in the leftward context and finally construed as value and variable via the cleft.

(40) A: yes well now one thing that I have been asked about presumably you didn't have any hand in this in putting me forward as a referee some publisher or other last year about a year ago asked me what I thought about a series publishing early grammarians

B: oh Perrins perhaps

A: because I forget who it was# but I didn't ^think **it was P/errins# [~~who asked me that~~]** it could be (LLC-1)

(41) A: hum hum et est-ce que vous ressentez une différence entre l'IUT et et l'université la façon de travailler les é- les étudiants que vous croisez dans

B: les non oui c'est sûr que bon j'y suis là il y a pas très longtemps ici mais à la fac quoi mais c'est vrai qu'à l'IUT c' était plus sympathique on a j'ai sympathisé plus vite qu' ici les étudiants ils étaient plus euh plus sympa euh bon **[c'est peut-être parce que je suis pas là depuis**

longtemps]_{HH} [~~que j'ai sympathisé moins vite ici~~] ça va peut-être (CRFP)

'A: hm hm and do you feel like there's a difference between the UIT and and university in the way people work the s- the students you run into in

B: the no yes of course I'm here not too long ago ici but at university but it's true that at UIT it was more pleasant we I got on well more quickly than here students they were more uh nicer well **it may be because I haven't been here for too long [~~that I got on more slowly here~~]**'

From this functional overview, I can conclude that reduced clefts are more than an informationally motivated variant of clefts and the unmarked option for clefts with discourse-given variable. This description would indeed leave the large proportion of full clefts with discourse-given, and even textually evoked, variable unexplained. Rather, as demonstrated in this section, reduced clefts are, like full clefts, a construction in their own right with their own potential for achieving specific rhetorical effects.

6.4 Concluding remarks

The goal of this chapter was to shed light on the hitherto overlooked reduced variant of *it*- and *c'est*-clefts. As such, I aimed at determining whether reduced clefts are a mere unmarked choice for clefts with a salient variable or, conversely, whether they form of subtype of clefts in their own right with specific functional properties. The content of this chapter can be summarised as follows.

First, the review of the existing literature and the building of the LLC-1 and CRFP datasets have highlighted the gap in the availability of methodology for the study of reduced clefts. While the different recognition tests found in the literature generally allowed me to successfully retrieve full clefts, they proved to be partly insufficient to exhaustively extract reduced clefts. For that reason, I have proposed three new criteria specific to reduced clefts which include high degree of predictability of the variable, presence of the state of affairs in the preceding

context, and presence of the non-qualifying set. Among these three criteria, only the absence of the first two is systematically treated as excluding evidence while the measure of contrast is viewed as corroborating evidence.

The analysis of the basic properties of reduced clefts, which, to my knowledge, had not been systematically examined in any previous studies, has revealed a number of things. First, reduced clefts are typically less frequent than full clefts. Though this trend was observed in both languages, the ratio between the two variants was found to be higher in French. Next, the morphosyntax of their clefted constituent generally differs from that of their full counterparts. This is true for English and French and for both the syntactic categories and grammatical functions instantiated in the clefted constituent. Investigating the general characteristics of reduced *it*-clefts in comparison with *c'est*-clefts also allowed me to conclude that English PRICs (Declerck and Seki 1990) have a French equivalent displaying the same syntactic, semantic and pragmatic characteristics.

Additional asymmetries were uncovered with the examination of the information structure of reduced clefts. With regard to the discourse-familiarity of their constituents, only two patterns are attested for reduced clefts, against four for full clefts. This is consistent with the different descriptions (e.g. Declerck 1988; Hedberg 1990; Mikkelsen 2007) and can be explained in terms of the systematic givenness, typically evoked in my data, of the variable which prevents them from displaying a new-new or given-new pattern. More surprisingly, not all discourse-given variables yield a reduced cleft, which shows that the choice between a full and reduced form does not solely depend on the distribution of information. Prosodically, reduced *c'est*-clefts were found to show more variation in their realisation than reduced *it*-clefts. Similarly to full clefts, reduced *it*-clefts are characterised by narrow contrastive focus on the clefted constituent while reduced *c'est*-clefts are typically realised as broad-focus structures. This confirms that the syntactic and prosodic means to encode emphasis do not systematically co-occur (Portes and Reyles 2022).

Finally, from a functional perspective, I have identified three contexts in which reduced clefts typically appear which include question/answer pairs, expression of overt contrast, and conclusion of a discourse segment, the last two of which are shared with full clefts.

In sum, reduced clefts do not appear to be a mere unmarked realisation of clefts with a discourse-given variable. In view of their morphosyntactic and information structural properties, reduced clefts are better described as a full-fledged category of clefts serving specific rhetorical purposes.

Conclusion

In this thesis, I have investigated the use of cleft constructions in English and French spoken data. In particular, I have examined the way their relational and referential information structure are operationalised. This involved analysing how the information conveyed in both parts of the cleft is partitioned into focal/non-focal and new/given. This analysis was carried out by adopting an in-depth qualitative-quantitative corpus-based approach.

In Chapter 1, I introduced the main concepts under study, viz. information structure and cleft constructions, by way of a comprehensive review of the existing literature. Among other things, I clarified the theoretical framework this study adheres to, which is that of Halliday's (1967a, 1994, and Greaves 2008, and Matthiessen 2004, 2014) functional linguistic theory. In Halliday's approach to language, grammatical meaning is seen as being symbolised by the grammar and prosody and involves conventionalised form-meaning pairings. As such, *it*- and *c'est*-clefts were investigated in accordance with the underlying grammatical and prosodic principles of English and French, which are detailed in Chapter 1.

Summarising the current state-of-the-art also entailed reviewing the different typologies of the information structure of *it*- and *c'est*-clefts which have been drawn up in the respective literatures. Doing so led to the identification of a number of shortcomings. First, the lack of agreement surrounding the nature and source of focus, e.g. syntactic, pragmatic, or prosodic, has created different, and at times contradictory, predictions with regard to its encoding in clefts. Some authors (e.g. Lambrecht 2001), consider it to be borne by the clefted constituent while some others have claimed that it can occur in different locations within the cleft in the case of *it*-clefts (Prince 1978; Declerck 1984; Geluykens 1988; Collins 2006; Huber 2006), or can scope over the whole construction in the case of *c'est*-clefts (Rialland et al. 2002; Doetjes et al. 2004; Mertens 2012).

Another limitation which was highlighted by the literature review concerns the competing approaches to discourse-familiarity. While some studies (e.g. Prince 1978; Declerck 1984; Huber 2006) adopt a cognitive approach in which the speaker's assumptions are taken into account, others (Geluykens 1988; Collins 2006; Rialland et al. 2002; Doetjes et al. 2004; Mertens 2012; Hasselgård 2014), favour a textual

approach in which a constituent is only considered discourse-given if the hearer can infer it from the preceding context. This discrepancy is of course somewhat problematic in that speakers may choose to be uncooperative and place themselves in violation of the Gricean maxims, which means that their assumptions may not in fact always inform the grammatical choices they make. Furthermore, corpus data, in which only the texts themselves are available, do not easily lend themselves to a cognitively-based analysis. For this reason, this study was conducted without considering the speaker's assumptions. By including the determiner structure of NPs more systematically into the analysis than in Prince's (1981) model, however, the textual approach I adopted still took into account their choices with regard to givenness vs. newness.

More generally, most of the typologies of *it*-clefts exemplify to some degree the conflation of the referential and relational layers of the information structure of clefts. As such, discourse-new constituents are then predicted to bear prosodic focus while discourse-given items are taken to remain unaccented, which goes against Halliday's (1967a) early predictions. This amalgam is however rejected by the taxonomies of *c'est*-clefts (Rialland et al. 2002; Doetjes et al. 2004; Mertens 2012) which, like Halliday (1967a), accept the de-coupling of discourse-familiarity and focus. This decoupling was largely corroborated by the findings of this study.

Along with the aforementioned issues, Chapter 1 also underlined a number of gaps which this study aimed at filling. These relate to the treatment of reduced clefts, i.e. clefts with an omitted cleft-relative clause, and to the extent of the cross-linguistic equivalence of *it*- and *c'est*-clefts, which only a handful of studies investigate.

All in all, the present research was driven by the following aims.

- (i) to provide a refined comparative account of the information structure of *it*- and *c'est*-clefts
- (ii) to characterise the interplay between referential and relational information structure and the interaction between syntax and prosody
- (iii) to shed light on the hitherto overlooked reduced clefts
- (iv) to offer a comprehensive methodology for the study of multi-faceted notions such as givenness/newness and prosodically coded focus

These aims were examined with an approach strongly grounded in empirical study of spoken contextualised data. More specifically, I carried out an in-depth and systematic analysis of the referential and relational information structure of clefts in extensive datasets, which had not been done before, and for which I proposed an exhaustive analytical framework.

The research questions tied with the information structure of clefts were examined in Chapter 3 and 4, which focused on *it*- and *c'est*-clefts respectively, and Chapter 5, which offered a contrastive perspective building on the main conclusions of the preceding chapters. To start with, I underlined the higher frequency of occurrence of *c'est*-clefts, which, in my data, were found to be twice as common as *it*-clefts. From a morphosyntactic viewpoint, I showed that the equivalence between *it*- and *c'est*-clefts is mostly restricted to their basic syntactic structure while the morphology of their copula and the form and function of their clefted constituent differ to some degree. As I have shown, *it*- and *c'est*-clefts differ in terms of the diachronic developments they have undergone, i.e. fixation of the copula's morphology, and the ways they have specialised for grammatical and cohesive functions.

In Chapter 3 and 4, I identified and classified the different patterns of discourse-familiarity observed in the data. Both English and French clefts were found to instantiate four types, viz. new-given, new-new, given-new and given-given, the last of which had tended to be neglected in many of the extant typologies, with exceptions such as Huber (2006), who recognised its existence but yet did not treat it as a full-fledged type, and Johansson (2002) and Hasselgård (2014). Cross-linguistically, the given-new type revealed itself to be more common in French than in English which I explained by the higher number of anaphoric clefted constituents in *c'est*-clefts.

The findings on relational information structure of *it*- and *c'est*-clefts likewise revealed some asymmetries between the constructions in the two languages. In particular, while prosodically coded focus is mostly narrow in *it*-clefts, bearing on either the clefted constituent, an element of the cleft-relative clause, or both, *c'est*-clefts significantly favour broad focus whose scope extends over the whole construction. Thus, both the relational and referential information structures of *it*- and *c'est*-clefts somewhat differ, which further adds to their distinct

operationalisation, which interacts with differences in their systems of information structural principles. This constitutes a significant novel finding which only the extensive contrastive analysis presented in this study has singled out.

Strikingly, neither *it*- or *c'est*-clefts display a prototypical articulation which dominates other marginal information structures. This breaks with descriptions in which clefts are treated as a unified class whose outliers remain minimal in their frequency of occurrence. I attribute this difference not only to the competing approaches to the notion of focus, i.e. formal-pragmatic for Lambrecht (1994, 2001) and functional for Halliday (1967a, 1994) and this study, but also to the extent of corpus verification. While it would be tempting to assume that the most striking instances of clefts, typically contrastive new-given clefts, are representative of the whole class, I have shown with the present study that speakers actually rely on an array of cleft structures taking on different information structures and discourse functions.

On the question of the relation between discourse-familiarity and prosodic focus, I have shown that newness and presence of prosodic focus do not systematically co-occur, and vice versa. On the contrary, discourse-given information can be part of the focus and discourse-new information can be deaccented, thus relegating it to the informational background. Strikingly, the instrumental analysis of English tokens has revealed the existence of a hierarchy of information foci, whereby multiple-foci *it*-clefts tend to have a higher-pitched prominence on the clefted constituent. To make sense of this finding, I argued that information foci are best conceived in terms of contrastive focus vs. presenting focus. When the information focus on clefted constituent is more prominent, then it illustrates marked contrastive focus while the information focus in occurring in the cleft-relative clause corresponds to an unmarked presenting focus, regardless of the discourse-newness or discourse-givenness of the constituent in focus. Taken all together, the discourse-familiarity results have uncovered an 'equi-probable' distribution (Halliday and James 1993) between given and new for both value and variable. These findings differ from that on focus assignment which have shown that there is a default option, i.e. typically, hierarchically primary, contrastive information focus on value. This shows that the referential and relational layers of information structure are manipulated as distinct systems by speakers.

In Chapter 5, I presented insights on the pragmatics of clefts which I examined from a contrastive perspective. For this, I considered two of the main pragmatic implications of clefts, viz. exhaustivity and contrast, which I then related to their general functional profile. *It*- and *c'est*-clefts have in common that they generally trigger an implicature of exhaustivity which may be cancelled or emphasised either through lexical means, prosodic means, or both. While cancellation was found to be equally instantiated in both languages, the results uncovered a slight preference for reinforced exhaustivity in French. As far as contrast is concerned, its realisation turned out to be more overtly marked in *it*-clefts, in which the clefted constituent typically bears contrastive focus. Thus, in English *it*-clefts, the contrastive meaning associated with identifying the correct value(s) from the set of potential candidates, is, in the unmarked case, further highlighted by contrastive focus. For *c'est*-clefts, on the other hand, contrast which arises from the extraction process and the specificational meaning is not as overtly marked prosodically. It is also wholly lost in the minority of cases in which the cleft selects an anaphoric and non-paradigmatic clefted constituent which acts as a discourse-marker. These particular clefts provide evidence for a more advanced level of grammaticalisation of *c'est*-clefts. The functional profile of *it*- and *c'est*-clefts shows that some discourse uses, especially those with a localised scope, involve contexts that exploit the inherent contrastive meaning of clefts, and nothing else. The other discourse uses exploit this contrastive, singling out, meaning in specific ways, e.g. to shift the topic, or they foreground the specificational relation itself, e.g. to wrap up a discussion or to reinforce cohesion.

Interestingly, the analysis of both *it*- and *c'est*-clefts revealed that the realisation of both constructions may not only be influenced by interactional aspects but may also contribute to the interactional context. This is exemplified by the use of onsets in English, i.e. the first prominent syllable of the information unit, and the initial accent in French. When these display a large pitch excursion size, they indicate some sort of disruption such as the introduction of a new topic or disagreement or, in the case of corrective focus, contradiction of the expectations previously generated. Discursively, this idea is also illustrated in French by the prevalence of continuation rises which signal that the speaker is holding the floor in order to add more information.

The last chapter, Chapter 6, was devoted to the reduced form of clefts, i.e. clefts in which the cleft-relative clause is not restated because of its salience in the context, whose description had so far remained limited and whose use had not been investigated empirically. Because the recognition tests available for full clefts were found to be insufficient for the retrieval of reduced ones, I proposed three criteria which are high degree of predictability of the variable, presence of the state-of-affairs in the preceding context, and presence of the non-qualifying values. The resulting dataset allowed me to make a number of claims. First, the reduced variant of clefts is less frequent than the full one, more so in French than English. Informationally, reduced clefts show less variation in both their prosodic and discourse-familiarity patterns, but the ones available do not mirror the distribution found in full clefts by any means. Furthermore, a discourse-given cleft-relative clause is not systematically deleted, as was confirmed by the relatively significant portion of full new/given-given clefts. This, along with the different discourse functions attested in the data, e.g. question/answer pairs, overt contrast, cohesion, shows that reduced clefts are more than an informationally motivated variant of clefts and the mere unmarked option for clefts with discourse-given variable. Rather, reduced clefts are a full-fledged subtype of clefts with their own potential for achieving specific rhetorical effects.

In sum, the present study has verified a number of recurrent claims in the literature on clefts mainly relating to their semantic, pragmatic and information structural characteristics. While some were largely corroborated by empirical evidence provided by the analysis of the LLC-1 and CRFP, e.g. information structure variance, higher frequency of French clefts, etc., others have displayed a need for fine-tuning, e.g. prototypicality, status of reduced clefts, etc. In my approach to clefts, they are shown to be specificational constructions enabling multiple and versatile information structural realisations, whose overall profiles specialise differently in individual languages. The account presented in this thesis thus resets thinking about cleft constructions and provides a basis for future research. An obvious domain for further research is the opposition between a preference for narrow focus in English clefts and for broad focus in French clefts. Having uncovered a specialisation of *c'est*-clefts for cohesive functions in which the pragmatic meaning and contrastive effect of clefts are lost, future research looking into the

diachronic path of this specialisation would be enlightening. In the same vein, supplementing my findings on spontaneous spoken data with a comparative analysis of discourse-familiarity and the discourse motivations of clefts in written material⁷⁶, in which anaphoric *it*-clefts with a cohesive function like that of *c'est*-clefts can be expected to be more frequent, would allow one to draw up a more exhaustive picture of the functional profile of *it*- and *c'est*-clefts. It is my hope that by triggering discussion, challenge and extension of research questions, this thesis will contribute to renewed research into clefts.

⁷⁶ Because of the nature of the data, this type of study would be carried out without investigating prosodically coded focus. The contrastive aspect would therefore rely on the referential layer of information structure and the functions of clefts only.

Glossary⁷⁷**ENGLISH INTONATION SYSTEM**

Compound tone	fusion of two tones yielding two tonics within a single tone unit whose existence is not accepted by all scholars
Contrastive (narrow) information focus	information focus which is informationally contrastive and which relates to information presented as presupposed from the preceding discourse
Information focus	focus coded by the placement of the nuclear accent on a specific syllable of the tone unit. The domain of the information focus typically includes the tonic syllable and the larger constituent it is part of
Nuclear accent	main pitch accent within the tone unit
Onset	first accented syllable of the intonation unit which may be a pitch accent preceding the nuclear accent or the nuclear accent itself if it comes first or is the only accent in the tone unit
Presenting (broad) information focus	information focus falling on the last lexical constituent of the tone unit, which it marks as the most salient new, but which does not mark any information as presupposed
Tonality	segmentation of discourse into tone units whose boundaries are typically accompanied by a number of pitch discontinuities such as pauses and final lengthening
Tone unit	basic unit within Halliday's (1967a) description of the English intonation system. A tone unit is a basic unit of intonation which is also a unit of information
Tones	different pitch movements occurring on the tonic syllable (e.g. Fall, Rise, Rise-Fall, Fall-Rise and Level tone)
Tonic syllable	syllable bearing the nuclear accent and displaying the main pitch change. The segment preceding the tonic is the pre-tonic or pre-nuclear segment and the one

⁷⁷ The glossary includes all terms relating to the English and French intonation systems which are discussed in depth in the present thesis.

following the tonic the post-tonic segment or post-nuclear tail

Tonicity placement of the nuclear accent within the tone unit

FRENCH INTONATION SYSTEM

Appendix final part of the utterance occurring after the final accent which is realised with an extra-low flat contour and a compressed register

Broad focus focus which occurs when the nuclear accent is located in its default position, i.e. at the end of the intonation unit

Continuative tones tones marking a major prosodic boundary but the continuation of the information object (e.g. /HH, \HH, HH, H/H, LH, HL)

Final accent mandatory stress located on the last full syllable of the intonation unit typically accompanied by a lengthening of the syllable by which it is carried and by pitch movement

Initial accent optional stress typically appearing on the first syllable of a word

Intonation Unit highest phonological constituent within the three-tiered intonation system of French built around prosodic stress

Narrow focus focus which occurs when the nuclear accent is moved from its otherwise default position at the end of the intonation unit to an earlier position

Terminal tones tones indicating the completion of a maximal prosodic unit and of an information object (e.g. L-L-, HL-, LL-)

Tones different pitch heights within Mertens's (2006) annotation system (e.g. L-L-, HH, HL-, H/H, LL\, l-l-, etc.)

References

- Akmaijian, Adrian. 1970. On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry* 1. 149–168.
- Akmaijan, Adrian. 1979 [1973]. *Aspects of the Grammar of Focus in English*. Cambridge, MA: Massachusetts Institute of Technology. Ph.D dissertation.
- Anthony, Laurence. 2019. AntConc (3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Apothéloz, Denis & Marie-Noëlle Roubaud. 2015. Constructions pseudo-clivées. *Encyclopédie Grammaticale du Français*. <http://encyclogram.fr>. (accessed online 11 December 2020).
- Auer, Peter. 2005. Projection in interaction and projection in grammar. *Text* 25(1). 7–36.
- Avanzi, Mathieu. 2011. Note sur la prosodie des clivées du type c'est X qu- V en français parlé. In Gilles Corminboeuf & Marie-José Béguelin (eds.), *Du système linguistique aux actions langagières. Hommages à Alain Berrendonner*, 113–124. Bruxelles: De Boeck/Duculot.
- Bateman, John A. 2017. The place of systemic functional linguistics as a linguistic theory in the twenty-first century. In Tom Bartlett & Gerard O'Grady (eds.), *The Routledge Handbook of Systemic Functional Linguistics*, 11–26. Oxon: Routledge.
- Belletti, Adriana. 2005. Answering with a “cleft”: the role of the null subject parameter and the VP periphery. In Laura Brugè, Guilana Giusti, Nicola Murano, Walter Schweikert & Guiseppina Turano (eds.), *Proceedings of the Thirtieth Incontro di Grammatica Generativa*, 63–82. Venezia: Cafoscarina.
- Beyssade, Claire, Barbara Hemforth & Jean-Marie Marandin. 2009. Prosodic markings of information focus in French. In Hi-Yon Yoo & Elisabeth Delais-Roussarie (eds.), *Proceedings of Interfaces Discours Prosodie IDP*, 109–122.
- Birner, Betty J., Jeffrey P. Kaplan & Gregory Ward. 2007. Functional compositionality and the interaction of discourse constraints. *Language* 83(2). 317–343.

- Blanche-Benveniste, Claire. 1990. *Le français parlé - études grammaticales*. Paris : Edition du CNRS.
- Blanche-Benveniste, Claire. 2002. Auxiliaires et degrés de « verbalité ». *Syntaxe et Sémantique* 3. 75–97.
- Blanche-Benveniste, Claire. 2006. Les clivées françaises de type : *C'est comme ça que, C'est pour ça que, C'est là que tout a commencé*. *Moderna Språk* 100. 273–287.
- Blanche-Benveniste, Claire, José Deulofeu, Jean Stéfanini & Karel van den Eynde. 1984. *L'Approche pronominale et son application au français*, Paris: SELAF.
- Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9-10). 341–345.
- Bolinger, Dwight. 1972. A look at equations and cleft sentences. In Evelyn S. Firchow (ed.), *Studies for Einar Haugen*, 96–114. The Hague: De Gruyter Mouton.
- Bolinger, Dwight. 1977. *Meaning and Form*. London and New York: Longman.
- Bouchard, Jacynthe, Fernande Dupuis & Monique Dufresne. 2007. Un processus de focalisation en ancien française : le développement des clivées. *Proceedings of the 2007 annual conference of the Canadian Linguistic Association*.
- Bourgoin, Charlotte. 2017. The Role of the English *It*-Cleft and the French *C'est*-Cleft in Research Discourse. *Discours* 21.
- Bourgoin, Charlotte. 2021. Towards a new typology of the referential information structure of specificational *it*-clefts. [Manuscript submitted for publication].
- Bourgoin, Charlotte & Kristin Davidse. in press. Making the case for distinguishing information structure from specification in English *it*-clefts. In Caterina Bonan & Adam Ledgeway (eds.) *It-clefts: empirical and theoretical surveys and advances*. Amsterdam: Mouton de Gruyter.
- Bourgoin, Charlotte, Gerard O'Grady & Kristin Davidse. 2021. Managing information flow through prosody in *it*-clefts. *English Language & Linguistics* 25(3). 485–511.
- Bourns, Stacey K. 2014. Contrasting *c'est*-clefts and *it*-clefts in discourse. In Stacey K. Bourns & Lindsay L. Meyers (eds.), *Perspectives on Linguistic Structure and Context*, 199–222. Amsterdam/Philadelphia: John Benjamins.
- Brazil, David. 1997. *The communicative value of intonation in English*. Cambridge: Cambridge University Press.

- Büring, Daniel. 1998. Identity, Modality, and the Candidate Behind the Wall. In Devon Strohovitch & Aaron Lawson (eds.), *Proceedings of SALT 8*, 36–54. Ithaca, NY: Cornell University.
- Büring, Daniel & Manuel Križ. 2013. It's That, and That's It! Exhaustivity and Homogeneity Presuppositions in Clefts (And Definites). *Semantics and Pragmatics* 6. 1–29.
- Byram Washburn, Mary, Elsi Kaiser & Maria Luisa Zubizarreta. 2014. The English It-Cleft: No Need to Get Exhausted. *Proceedings of the 'Questions in Discourse Conference' at Universität Göttingen*.
- Carter-Thomas, Shirley. 2002. Theme and Information Structure in French and English: A Contrastive Study of Journalistic Clefts. Paper presented at the *14th Euro-International Systemic Functional Workshop*.
- Carter-Thomas, Shirley. 2009. The French *C'est*-Cleft: Function and Frequency. In David Banks, Simon Eason & Janet Ormrod (eds.), *La linguistique systémique fonctionnelle et la langue française*, 127–157. Paris: L'Harmattan.
- Chafe, Wallace. 1974. Language and Consciousness. *Language* 50. 111–133.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In Charles N. Li (ed.), *Subject and Topic*, 25–56. New York: Academic Press.
- Chafe, Wallace. 1987. Cognitive Constraints on Information Flow. In Russell Tomlin (ed.), *Coherence and Grounding in Discourse*, 21–51. Amsterdam: John Benjamins.
- Chafe, Wallace. 1994. *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. The University of Chicago Press: Chicago.
- Chomsky, Noam. 1971. Deep structure, surface structure, and semantic interpretation. In D.D. Steinberg & L.A. Jakobovits (eds.), *Semantics*, 183–216. Cambridge: Cambridge University Press.
- Clark, Herbert H. & Susan E. Haviland. 1977. Comprehension and the Given-New contract. In Roy Freedle (ed.), *Discourse production and comprehension*, 1–40. Norwood: Ablex.

- Clech-Darbon, Anne, Georges Rebuschi & Annie Riolland. 1999. Are there cleft sentences in French? In Georges Rebuschi & Laurice Tuller (eds.), *The Grammar of Focus*, 83–118. Amsterdam: John Benjamins.
- Collins, Peter. 1991. *Cleft and pseudo-cleft constructions in English*. London: Routledge.
- Collins, Peter. 2006. *It*-clefts and *wh*-clefts: Prosody and pragmatics. *Journal of Pragmatics* 38(10). 1706–1720.
- Couper-Kuhlen, Elizabeth & Tsuyoshi Ono. 2007. ‘Incrementing’ in conversation. A comparison of practices in English, German and Japanese. *Pragmatics* 17(4). 513–552.
- Crystal, David. 1969. *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
- Cruttenden, Alan. 1997. *Intonation*, 2nd edition. Cambridge: Cambridge University Press.
- Daneš, Frantisek. 1987. On Prague School Functionalism in linguistics. In René Dirven & Vilém Fried (eds.), *Functionalism in Linguistics*, 3–38. Amsterdam: John Benjamins.
- Davidse, Kristin. 1999. Are there sentences that can be analyzed as there-clefts? In Guy A. J. Tops, Betty Devriendt & Steven Geukens (eds.), *Thinking English Grammar: To honour Xavier Dekeyser, Professor emeritus*, 177–193. Leuven: Peeters.
- Davidse, Kristin. 2000. A constructional approach to clefts. *Linguistics* 38(6). 1101–1131.
- Davidse, Kristin. 2014. On specificational *there*-clefts. *Leuven Working Paper in Linguistics* 4.
- Davidse, Kristin & Ditte Kimps. 2016. Specificational *there*-clefts: Functional structure and information structure. *English Text Construction* 9(1). 115–142.
- Davidse, Kristin & An Van linden. 2020. Revisiting ‘*it*-extraposition’: The historical development of constructions with matrices (*it*)/(*there*) *be* + NP followed by a complement clause. In Paloma Núñez-Pertejo, Mariá José López-Couso, Belén Méndez-Naya & Javier Pérez-Guerra (eds.), *Crossing Linguistic Boundaries: Systemic, Synchronic and Diachronic Variation in English*, 81–103. London: Bloomsbury Academic.

- Davidse, Kristin & Ngum Njende. 2019. Enumerative there-clauses and there-clefts: specification and information structure. *Acta Linguistica Hafniensia* 51(2). 160–191.
- Davidse, Kristin, Ngum Njende & Gerard O’Grady. 2022. Putting specificational and presentational there-clefts on the map - and redefining the field of clefts. *Leuven Working Papers in Linguistics* 9.
- De Cesare, Anna-Maria & Davide Garassino. 2015. On the status of exhaustiveness in cleft sentences: An empirical and cross-linguistic study of English *also-/only-*clefts and Italian *anche-/solo-*clefts. *Folia Linguistica* 49(1). 1–56.
- De Stefani, Elwys. 2008. De la malléabilité des structures syntaxiques dans l’interaction orale : le cas des constructions clivées. In Jacques Durand, Benoît Habert & Bernard Laks (eds.), *Congrès Mondial de Linguistique Française*, 703–720.
- Declerck, Renaat. 1983. Predicational clefts. *Lingua* 61(1). 9–45.
- Declerck, Renaat. 1984. The pragmatics of *it*-clefts and *wh*-clefts. *Lingua* 64. 251–289.
- Declerck, Renaat. 1988. *Studies on copular sentences, clefts and pseudo-clefts*. Dordrecht: Foris Publications.
- Declerck, Renaat. 1992. The inferential construction *it is that*-construction and its congeners. *Lingua* 87. 203–230.
- Declerck, Renaat & Shigeki Seki. 1990. Premodified reduced IT-clefts. *Lingua* 82(1). 15–51.
- Dehé, Nicole & Bettina Braun. 2013. The prosody of question tags in English. *English Language and Linguistics* 17(1). 129–156.
- Delais-Roussarie, Elisabeth, Hiyon Yoo & Brechtje Post. 2011. Quand frontières prosodiques et frontières syntaxiques se rencontrent. *Langue française* 170. 29–44.
- DELIC. 2004. Présentation du Corpus de Référence du Français Parlé. *Recherches sur le français parlé* 18. 11–42.
- Delin, Judy. 1990. A multi-level account of cleft constructions in discourse. *Proceedings of the 13th Conference on Computational Linguistics* 2. 83–88.
- Delin, Judy & Jon Oberlander. 1995. Syntactic constraints on discourse structure: The case of *it*-clefts. *Linguistics* 33(3). 465–500.

- Destruel, Emilie. 2012. The French *c'est-cleft*: an empirical study on its meaning and use. In Christopher Piñón (ed.), *Empirical issues in syntax and semantics, Volume 9*, 95–112. Paris: CSSP.
- Destruel, Emilie & Joseph P. De Veugh-Geiss. 2019. (Non-)Exhaustivity in French *c'est-Clefts*. In Christopher Pinon (ed.), *Empirical Issues in Syntax and Semantics 12*, 91–120. Paris: CSSP.
- Destruel, Emilie & Leah Velleman. 2014. Refining contrast: Empirical evidence from the English *it-cleft*. In Christopher Piñón (ed.), *Empirical issues in syntax and semantics, Volume 10*, 197–214. Paris: CSSP.
- Destruel, Emilie, David I. Beaver & Elizabeth Coppock. 2019. It's Not What You Expected! The Surprising Nature of Cleft Alternatives in French and English. *Frontiers in Psychology* 10. 1–15.
- Doetjes, Jenny, Rebuschi, Georges & Annie Riailand 2004. Cleft Sentences. In Francis Corblin & Henriëtte de Swart (eds.), *Handbook of French semantics*, 529–552. Stanford: CSLI Publications.
- Di Cristo, Albert. 1998. Intonation in French. In Robert Hirst & Albert Di Cristo (eds.), *Intonation Systems. A Survey of Twenty Languages*, 195–218. Cambridge: Cambridge University Press.
- Di Cristo, Albert. 1999. Vers une modélisation de l'accentuation en français: 1ère partie. *Journal of French Language Studies* 9. 143–179.
- Di Cristo, Albert. 2016. *Les musiques du français parlé*. Berlin: De Gruyter.
- Di Cristo, Albert. 2019. *Conditionnements de l'information et de marquages prosodiques dans les langues naturelles. Première partie : La structure informationnelle et ses déterminants*. Pre-print.
- Donnellan, Keith. 1966. Reference and definite descriptions. *Philosophical Review* 75(3). 281–304.
- Du Bois, John. 1980. Beyond definiteness: the trace of identity in discourse. In Wallace Chafe (ed.), *The pear stories: Cognitive, cultural and linguistic aspects of narrative production*, 203–274. Norwood: Ablex.
- Delahunty, Gerald P. 1982. *Topics in the Syntax and Semantics of English Cleft Sentences*. Irvine: University of California, Irvine. Ph.D. dissertation.
- Delahunty, Gerald P. 1995. The inferential constructions. *Pragmatics* 5(3). 341–364.

- Dufter, Andreas. 2008. On explaining the rise of *c'est*-clefts in French. In Ulrich Detges & Richard Waltereit (eds.), *The Paradox of Grammatical Change: Perspectives from Romance* (Current Issues in Linguistic Theory 293), 31–56. Amsterdam: John Benjamins.
- Dufter, Andreas. 2009a. Clefting and discourse organization: Comparing Germanic and Romance. In Andreas Dufter & Daniel Jacob (eds.), *Focus and Background in Romance Languages*, 83–121. Amsterdam: John Benjamins.
- Dufter, Andreas. 2009b. Beyond focus marking: fine-tuning the evolution of cleft types from Latin to Modern French. Paper presented as 31. *Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft DGfS, AG9*.
- Dufter, Andreas. 2015. Clefts and pseudo-clefts in French: formal and functional motivations. Invited talk at KU Leuven, Belgium.
- Emmott, Catherine. 1992. Splitting the Referent: An Introduction to Narrative Enactors. In Martin Davies & Louise J. Ravelli (eds.), *Advances in Systemic Linguistics: Recent Theory and Practice*, 221–228. London: Harold Pinter.
- Emmott, Catherine. 1997. *Narrative comprehension: a discourse perspective*. Oxford: Clarendon Press.
- Erteschik-Shir, Nomi & Shalom Lappin. 1979. Dominance and the functional explanation of island phenomena. *Theoretical Linguistics* 6(1/3). 41–86.
- Erteschik-Shir, Nomi & Shalom Lappin. 1983. Under Stress: A Functional Explanation of English Sentence Stress. *Journal of Linguistics* 19(2). 419–453.
- Esser, Jürgen. 1988. *Comparing reading and spoken intonation*. Amsterdam: Rodopi.
- Féry, Caroline. 2001. Focus and Phrasing in French. In Caroline Féry & Wolfgang Sternefeld (eds.), *Audiatur Vox Sapientiae*, 153–181. Berlin: Akademie Verlag.
- Filppula, Markku. 2009. The rise of *it*-clefting in English: areal-typological and contact-linguistic considerations. *English Language and Linguistics* 13(2). 267–293.
- Firbas, Jan. 1971. On the concept of Communicative Dynamism in the theory of functional sentence perspective. *Sborník prací filozofické fakulty brněnské univerzity* A19. 135–144.
- Firbas, Jan. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.

- Fontaine, Lise & David Schönthal. 2020. Referring and the nominal group: a closer look at the selector element. In Gordon Tucker, Guowen Huang, Lise Fontaine & Edward McDonald (eds.), *Approaches to Systemic Functional Grammar*, 174–190. Equinox.
- Garassino, Davide. 2014. Clefts sentences. Italian-English in contrast. In Anna-Maria De Cesare (eds.), *Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic*, 101–138. Berlin: Mouton de Gruyter.
- Garde, Paul. 1968. *L'accent*. Paris: Presses Universitaires de France.
- Geluykens, Ronald. 1988. Five types of clefting in English discourse. *Linguistics* 26. 823–841.
- Gentens, Caroline. 2016. The discursive status of extraposed object clauses. *Journal of Pragmatics* 96. 15–31.
- Givón, Talmy. 1975. Focus and the scope of assertion: Some Bantu evidence. *Studies in African Linguistics* 6(2). 185–206.
- Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study*, 1–41. Amsterdam and Philadelphia: John Benjamins.
- Givón, Talmy. 2001. *Syntax: An Introduction. Volume 2*, 2nd edition. Amsterdam: John Benjamins.
- Granger, Sylviane. 2010. Comparable and translation corpora in cross-linguistic research. Design, analysis and applications. *Journal of Shanghai Jiaotong University* 2. 14–21.
- Greenbaum, Sidney. 1996. *The Oxford English Grammar*. Oxford: Oxford University Press.
- Gundel, Jeanette. 1974. *Role of topic and comment in linguistic theory*. Austin: University of Texas Ph.D. Dissertation.
- Gundel, Jeanette. 1980. Zero NP-anaphora in Russian: a case of topic-prominence. *Proceedings from the 16th Meeting of the Chicago Linguistic Society. Parasession on Anaphora*. 139–146.
- Gundel, Jeanette. 1988. Universals of topic-comment structure. In Michael Hammond, Edith Moravcsik, & Jessica Wirth (eds.), *Studies in Syntactic Typology*, 209–239. Amsterdam: John Benjamins.

- Gundel, Jeanette. 1999. Topic, focus, and the grammar-pragmatics interface. *Proceedings of the 23rd Annual Penn Linguistics Colloquium. University of Pennsylvania Working Papers in Linguistics* 6(1). 185–200.
- Gundel, Jeanette. 2003. Information Structure and Referential Givenness/Newness: How Much Belongs in the Grammar? In Stefan Müller (ed.), *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar*, 122–142. CSLI Publications.
- Halliday, M. A. K. 1963. The tones of English. *Archivum Linguisticum* 15(1). 1–28.
- Halliday, M. A. K. 1967a. Notes on transitivity and theme in English, Part II. *Journal of Linguistics* 3. 199–244.
- Halliday, M. A. K. 1967b. *Intonation and grammar in British English*. The Hague: Mouton.
- Halliday, M.A.K. 1985. *An introduction to Functional Grammar*, 1st edition. London: Arnold.
- Halliday, M.A.K. 1994. *An introduction to Functional Grammar*, 2nd edition. London: Arnold.
- Halliday, M.A. K. & William Greaves. 2008. *Intonation in the grammar of English*. London: Equinox.
- Halliday, M. A. K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Halliday, M. A. K. & Z.L. James. 1993. A quantitative study of polarity and primary tense in the English finite clause. In John M. Sinclair (ed.) *Techniques of Description: Spoken and Written Discourse*. London: Routledge.
- Halliday, M. A. K. & Christian Matthiessen 2004. *An Introduction to Functional Grammar*, 3rd edition. London: Routledge.
- Halliday, M. A. K. & Christian Matthiessen 2014. *An introduction to Functional Grammar*, 4th edition. London: Arnold.
- Hamlaoui, Fatima. 2010. Focus, contrast, and the syntax-phonology interface: The case of French cleft-sentences. *Proceedings of the 18th International Congress of Linguists*.
- Hartmann, Jutta M. 2011. Focus, Predication and Specification: the Case of It-clefts. Paper presented at the *CRISSP Seminar Series*, KU Leuven, Belgium.
- Hartmann, Jutta M. 2016. Apparent predicational clefts. Paper presented at the *International conference on non-prototypical clefts*, KU Leuven, Belgium.

- Hasselgård, Hilde. 2004. Adverbials in it-cleft constructions. In Karin Aijmer & Bengt Altenberg (eds.), *Advances in Corpus Linguistics*, 195–212. Amsterdam & New York: Rodopi.
- Hasselgård, Hilde. 2014. It-clefts in English L1 and L2 academic writing: The case of Norwegian learners. In Kristin Davidse, Caroline Gentens, Lobke Ghesquière & Lieven Vandelanotte (eds.), *Corpus Interrogation and Grammatical Patterns*, 295–320. Amsterdam: John Benjamins.
- Hatcher, Anna G. 1948. From *Ce suis je* to *C'est moi* (the ego as Subject and as Predicative in Old French). *Publications of the Modern Language Association* 63(4). 1053–1100.
- Hedberg, Nancy. 1990. *Discourse Pragmatics and Cleft Sentences in English*. Minnesota: University of Minnesota Ph.D dissertation.
- Hedberg, Nancy. 2000. The referential status of clefts. *Language* 76(4). 891–920.
- Hedberg, Nancy. 2013. Multiple focus and cleft sentences. In Katharina Hartmann & Tonjes Veenstra (eds.), *Cleft Structures*, 227–250. Amsterdam: Benjamins.
- Hedberg, Nancy. 2021. Discourse Functions of Clefts and Pseudoclefts. Presented at the *Listen Seminar Series, KU Leuven, Belgium*.
- Hedberg, Nancy & Lorna Fadden. 2007. The Information Structure of It-clefts, Wh-clefts and Reverse Wh-clefts in English. In Nancy Hedberg and Ron Zacharski (eds.), *The Grammar-Pragmatics Interface: Essays in Honor of Jeanette K. Gundel*, 49–76. Amsterdam/Philadelphia: John Benjamins.
- Hjelmslev, Louis. 1961. *Prolegomena to a theory of language* (revised English edition of original Danish version, 1943). Madison: University of Wisconsin Press.
- Hengeveld, Kees. 1989. Layers and operators in Functional Grammar. *Journal of Linguistics* 25. 127–157.
- Hengevel, Kees & Lachlan McKenzie. 2018. Negation in Functional Discourse Grammar. In Evelien Keizer & Hello Olbertz (eds.), *Recent developments in Functional Discourse Grammar*. Studies in Language Companion Series 205. 18–45. Amsterdam: John Benjamins.
- Herment, Sophie. 2008. Interdépendances entre prosodie et syntaxe dans les clivées en “it”. Paper presented at the *14ème colloque de Villetaneuse sur l’anglais oral*, Villetaneuse, France.

- Herment, Sophie & Laetitia Leonarduzzi. 2012. The Pragmatic Functions of Prosody in English Cleft Sentences. Paper presented at *Speech Prosody 2012 Sixth International Conference*, Shanghai, China.
- Higgins, Francis R. 1979. *The Pseudo-Cleft Construction in English*. London: Routledge.
- Hirst, Daniel & Albert Di Cristo. 1984. French Intonation: A Parametric Approach. *Die Neueren Sprachen* 83(5). 554–569.
- Horn, Laurence R. 1981. Exhaustiveness and the semantics of clefts. *North East Linguistics Society* 11. 125–142.
- Huber, Stefan. 2006. The complex functions of it-clefts. In Valéria Molnár & Susanne Winkler (eds.), *The Architecture of Focus*, 549–578. Berlin: Mouton de Gruyter.
- Huddleston, Rodney. 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. Information Packaging. *The Cambridge grammar of the English language*. 1363–1447. Cambridge: CUP.
- Jespersen, Otto. 1927. *A modern English Grammar on historical principles. Part III, Syntax*. Second Volume. London: Allen & Unwin.
- Jespersen, Otto & Niels Haislund. 2007 [1954]. *A modern English Grammar on historical principles. Part VII, Syntax*. London: Routledge.
- Jespersen, Otto. 1984 [1937]. *Analytic Syntax*. Chicago: University of Chicago Press.
- Johansson, Mats. 2001. Clefts in contrast: A contrastive study of *it* clefts and *wh* clefts in English and Swedish texts and translations. *Linguistics* 39. 547–582.
- Johansson, Mats. 2002. Clefts in English and Swedish. A Contrastive Study of *IT*-clefts and *WH*-clefts in Original Texts and Translations. Lund: Lund University. Ph.D dissertation.
- Jun, Sun-Ah & Cécile Fougeron. 2000. A Phonological model of French intonation. In Antonis Botinis (ed.), *Intonation: Analysis, Modeling and Technology*, 209–242. Dordrecht: Kluwer Academic Publishers.
- Kiss, Katalin É. 1998. Identificational focus versus information focus. *Language* 74. 245–273.
- Kaltenböck, Gunther. 2005. *It*-extraposition in English: a functional view. *International Journal of Corpus Linguistics* 10(2). 119–159.

- Karssenberg, Lena. 2018. *Non-prototypical clefts in French: A corpus analysis of "il y a" clefts*. Berlin: De Gruyter Mouton.
- Karssenberg, Lena & Karen Lahousse. 2015. Two types of clefts Information Structure, existential presupposition & syntax. Paper presented at the 41 *Incontro di Grammatica Generativa*, Perugia, Italy.
- Katz, Stacey. 1997. *The Syntactic and Pragmatic Properties of the C'est-cleft Construction*. Austin: University of Texas at Austin Ph.D dissertation.
- Katz, Stacey. 2000a. A Functional Approach to the Teaching of the French *C'est-Cleft*. *French Review* 74(2). 248–262.
- Katz, Stacey. 2000b. Categories of *C'est-Cleft* Constructions. *Revue Canadienne de Linguistique* 45(3/4). 1001–1021.
- Kimps, Ditte. 2016. Non-prototypical English specificational cleft constructions and their prosodic patterns. Paper presented at *the International Workshop on non-prototypical clefts*, KU Leuven, Belgium.
- Krifka, Manfred. 2008. Basic Notions of Information Structure. *Acta Linguistica Hungarica* 55(3/4). 243–276.
- Kunstmann, Pierre. 1990. *Le relatif-interrogatif en ancien français*. Geneva: Droz.
- Lahousse, Karen & Marijke Borremans. 2014. The distribution of functional-pragmatic types of clefts in adverbial clauses. *Linguistics* 52(3). 793–836.
- Lahousse, Karen & Béatrice Lamiroy. 2017. *C'est ainsi que: grammaticalisation ou lexicalisation ou les deux à la fois ?* *Journal of French Language Studies* 27. 161–185.
- Lambrecht, Knud. 1994. *Information structure and sentence form*. Cambridge: Cambridge University Press.
- Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39(3). 463–516.
- Langacker, Ronald. 1987. *Foundations of Cognitive Grammar*. Volume I. *Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald. 1991. *Foundations of Cognitive Grammar*. Volume II. *Descriptive application*. Stanford: Stanford University Press.
- Langacker, Ronald. 2021. Functions and assemblies. In Kazuhiro Kodama & Tetsuharu Koyama (eds.), *The Forefront of Cognitive Linguistics*, 1–54. Tokyo: Hituzi Syobo.

- Laurent, Nicolas. 2018. Les prédications en c'est : une approche systématique. *L'Information Grammaticale* 158. 19–29.
- Lehmann, Christian. 2008. Information structure and grammaticalization. In Elena Seoane & María José López-Couso (eds.), *Theoretical Issues in Grammaticalization*, 207–229. Amsterdam: Benjamins.
- Martin, Jim R., 1992. *English Text: System and Structure*. Amsterdam: John Benjamins.
- Martin, Philippe. 1978. Perception des séquences de contours prosodiques de phrases synthétisées. Actes des 9èmes J. E. P., Lannion, France. 21–30.
- Matthiessen, Christian. 1992. Interpreting the Textual Metafunction. In Martin Davies & Louise Ravelli (eds.), *Advances in Systemic Linguistics: Recent Theory and Practice*, 37–81. London: Pinter.
- McGregor, William. 1997. *Semiotic Grammar*. Oxford: Clarendon.
- McGregor, William. 2021. *Neo-Firthian Approaches to Linguistic Typology*. London: Equinox.
- Mertens, Piet. 1987. *L'intonation du français. De la description linguistique à la reconnaissance automatique*. Unpublished Ph.D dissertation.
- Mertens, Piet. 1993. Accentuation, intonation et morphosyntaxe. *Travaux de Linguistique* 26. 21–69.
- Mertens, Piet. 2006. A Predictive Approach to the Analysis of Intonation in Discourse in French. In Yuji Kawaguchi, Ivan Fónagy & Tsunekazu Moriguchi (eds.), *Prosody and Syntax: Cross-linguistic perspectives*, 64–101. Amsterdam: John Benjamins.
- Mertens, Piet. 2008. Syntaxe, prosodie et structure informationnelle : une approche prédictive pour l'analyse de l'intonation dans le discours. *Travaux de Linguistique* 56(1). 87–124.
- Mertens, Piet. 2012. La prosodie des clivées. In Sandrine Caddéo, Marie-Noëlle Roubaud, Magali Rouquier & Frédéric Sabio (eds.), *Penser les langues avec Claire Blanche-Benveniste*, 127–139. Aix-en-Provence: Presses Universitaires de Provence.
- Mertens, Piet. 2019. *Phonétique, phonologie et prosodie du français*. Leuven: Acco.

- Mertens, Piet, Jean-Philippe Goldman, Eric Wehrli & Arnaud Gaudinat. 2001. La synthèse de l'intonation à partir de structures syntaxiques riches. *Traitement Automatique des Langues* 42(1). 142–195.
- Mikkelsen, Line. 2007. On so-called truncated clefts. In Ljudmila Geist & Björn Rothstein (eds.), *Kopulaverben und Kopulasätze: Intersprachliche und intrasprachliche Aspekte*, 47–68. Tübingen: Niemeyer Verlag.
- Miller, Jim. 1996. Clefts, particles and word order in languages of Europe. *Language Sciences* 18(1/2). 111–125.
- Muller, Claude. 2002. Clivées, co-référence et relativisation. In Georges Kleiber & Nicolas Le Querler (eds.), *Traits d'union*, 17–32. Caen: Presses Universitaires de Caen,
- Muller, Claude. 2003. Naissance et évolution des constructions clivées en « c'est...que... »: de la focalisation sur l'objet concret à la focalisation fonctionnelle. In Peter Blumenthal & Jean-Emmanuel Tyvaert (eds.), *La Cognition Dans le Temps*, 101–120. Tübingen: Max Niemeyer Verlag.
- Nelson, Gerald. 1997. Cleft constructions in spoken and written English. *Journal of English Linguistics* 25(4). 340–348.
- Nølke, Henning. 1983. Quelques réflexions sur la structure sémantique des phrases clivées en français moderne. *Modèles linguistiques* 5(1). 117–140.
- O'Grady, Gerard. 2010. *A grammar of spoken English discourse: The intonation of increments*. London: A&C Black.
- O'Grady, Gerard. 2013. *Key concepts in phonetics and phonology*. New York: Palgrave MacMillan.
- O'Grady, Gerard. 2014a. The use of key in projecting face threatening acts in televised political debate. *Text & Talk* 34(6). 685–711.
- O'Grady, Gerard. 2014b. An investigation of how intonation helps signal information structure. In Wendy Bowcher & Bradley A. Smith (eds.), *Systemic Phonology: Recent Studies in English*, 27–52. London: Equinox.
- O'Grady, Gerard. 2017. Intonation and systemic functional linguistics. The way forward. In Tom Bartlett & Gerard O'Grady (eds.), *The Routledge handbook of systemic Functional Linguistics*, 146–162. New York: Routledge.
- O'Grady, Gerard & Tom Bartlett. 2019. Linearity and tone in the unfolding of information. *Acta Linguistica Hafniensia* 51(2). 192–221.

- Patten, Amanda. 2012. *The English it-cleft: A constructional account and a diachronic investigation*. Berlin: De Gruyter Mouton.
- Pike, Kenneth L. 1954. *Language in relation to a unified theory of the structure of human behavior. Part I*. Glendale, CA: The Summer Institute of Linguistics.
- Pöldvere, Nele, Victoria Johansson & Carita Paradis. (forthcoming). *A guide to the London-Lund Corpus 2 of spoken British English*. Lund Studies in English. Centre for Languages and Literature, Lund University.
- Portes, Cristel & Uwe Reyles. Combining syntax and prosody to signal information structure: the case of French. Paper presented at *Speech Prosody 2022*, Lisbon, Portugal.
- Prince, Ellen. 1978. A comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54. 883–906.
- Prince, Ellen. 1981. Towards a taxonomy of given-new information. In Peter Cole (eds.), *Radical pragmatics*, 223–255. New York: Academic Press.
- Prince, Ellen. 1992. The ZPG letter: subjects, definiteness, and information status. In William C. Mann & Sandra A. Thompson (eds.), *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, 295–326. Amsterdam: Benjamins.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Reinhart, Tanya. 1981. Pragmatics and Linguistics: an analysis of Sentence Topics. *Philosophica* 27. 53–94.
- Reeve, Matthew. 2011. The syntactic structure of English clefts. *Lingua* 121(2). 142–171.
- Rialland, Annie, Jenny Doetjes & Georges Rebuschi. 2002. What is Focused in C'est XP qui/que Cleft Sentences in French? *Proceedings of the First International Conference on Prosody*. 595–598.
- Rochemont, Michael. 1986. *Focus in Generative Grammar*. Amsterdam: Benjamins.
- Rochemont, Michael. 2016. Givenness. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford Handbook of Information Structure*, 41–63. Oxford: Oxford University Press.
- Rooth, Mats. 1985. *Association with Focus*. Amherst: University of Massachusetts. Ph.D dissertation.

- Rossi, Mario. 1999. *L'intonation, le système du français: description et modélisation*. Paris: Ophrys.
- Rossi, Mario, Albert Di Cristo, Daniel Hirst, Philippe Martin & Yukihiro Nishinuma. 1981. *L'intonation. De l'acoustique à la sémantique*. Paris: Klincksieck.
- Roubaud, Marie-Noëlle & Frédéric Sabio. 2015. Les clivées en C'est là que, C'est là où : structures et usages en français moderne, *Repères DoRiF 6. Recherches sur la syntaxe verbale en français et en italien. Hommage à Claire Blanche-Benveniste*. http://www.dorif.it/ezone/ezone_issues.php (accessed online 28 March 2019).
- Roubaud, Marie-Noëlle & Frédéric Sabio. 2018. C'est comme ça que j'ai perdu mon papa ! Les constructions en c'est comme ça que en français parlé et écrit. Paper presented as *SHS Web Of Conferences 46*.
- Rouget, Christine & Laurence Salze. 1986. « C'est...qui, C'est...que, le jeu des quatre familles », *Recherches sur le Français parlé 7*. 117–139.
- Rouquier, Magali. 2007. Les Constructions clivées en ancien français et en moyen français. *Romania 125(497/498)*. 167–212.
- Rouquier, Magali. 2018. Les constructions clivées. *Encyclopédie Grammaticale du Français*. <http://encyclogram.fr> (accessed online 03 September 2020).
- Sabio, Frédéric & Christophe Benzitoun. 2013. Sur les relations entre syntaxe et discours : dispositifs de la rection et dispositifs macrosyntaxiques. *Studia Universitatis Babeş Bolyai - Studia Philologia, Universitatea Babeş-Bolyai 58(4)*. 97–110.
- Scappini, Sophie. 2007. Étude du dispositif d'extraction en « c'est...qu », différenciation entre une relative en « c'est...qu » et une proposition clivée. *L'Information Grammaticale 114*. 53–56.
- Scappini, Sophie. 2013. Un sous type de la construction clivée en “C'est...qu”: La structure d'enchaînement: “Et c'est pour ça que...” et d'autres exemples. *Studia UBB Philologia 58(4)*. 81–95.
- Schwenter, Scott & Richard WALTEREIT. 2010. Presupposition accommodation and language change. In Kristin Davidse, Lieven Vandelanotte & Hubert Cuyckens (eds.), *Subjectification, Intersubjectification and Grammaticalization*, 75–102. Berlin: Mouton de Gruyter.

- Simon, Anne-Catherine. 2004. *La structuration prosodique du discours en français*. Berne: Peter Lang.
- Sinclair, John. 1992. Trust the Text: The Implications Are Daunting. In Martin Davies & Louise J. Ravelli (eds.), *Advances in Systemic Linguistics: Recent Theory and Practice*, 5–19. London: Harold Pinter.
- Sornicola, Rosanna. 2006. Interaction of syntactic and pragmatic factors on basic word order in the languages of Europe. In Giuliano Bernini & Marcia L. Schwartz (eds.), *8 Pragmatic Organization of Discourse in the Languages of Europe*, 357–544. Berlin, Boston: De Gruyter Mouton.
- Svartvik, Jan. (eds.). 1990. *The London corpus of spoken English: Description and research*. Lund Studies in English 82. Lund University Press.
- Svartvik, Jan & Randolph Quirk. 1980. (eds.), *A corpus of English conversation*. Lund: C. W. L. Gleerup.
- Tench, Paul. 1990. *The roles of Intonation in English*. Bern: Peter Lang.
- Tench, Paul. 1996. *The intonation systems of English*. London: Cassell.
- Traugott, Elizabeth. 1982. From propositional to textual and expressive meanings: some semantic-pragmatic aspects of grammaticalization. In Winfred Lehmann & Yakov Malkiel (eds.), *Perspectives on Historical Linguistics*, 245–271. Amsterdam: Benjamins.
- Traugott, Elizabeth. 1989. On the rise of epistemic meanings in English: an example of subjectification in semantic change. *Language* 65. 31–55.
- Traugott, Elizabeth. 2010. (Inter)subjectivity and (inter)subjectification: A reassessment. In Kristin Davidse, Lieven Vandelanotte & Hubert Cuyckens (eds.), *Subjectification, Intersubjectification and Grammaticalization*, 29–74. Berlin: Mouton de Gruyter.
- Vaissière, Jacqueline. 1991. Rhythm, accentuation and final lengthening in French. In Johan Sundberg, Lennart Nord & Rolf Carlson (eds.), *Music, Language, Speech and Brain*, 108–120. London: Macmillan Press.
- Van den Eynde, Karel. 1995. Methodological reflections on descriptive linguistics. Knud Togeby's principles and the pronominal approach. In Lene Schosler & Mary Talbot (eds.), *Studies in Valency*, 111–131. Odense: Odense University Press.

- Van Praet, Wout. 2019. Focus assignment in English specificational and predicative clauses: intonation as a cue to information structure? *Acta Linguistica Hafniensia* 51(2). 222–241.
- Van Praet, Wout & Kristin Davidse. 2015. Revisiting the typology of English copular clauses: ascription and specification in categorizing and identifying clauses. *Leuven Working Papers in Linguistics*.
- Vander Klok, Jozina, Heather Goad & Michael Wagner. 2018. Prosodic focus in English vs. French: A scope account. *Glossa: a journal of general linguistics* 3(1). 1–47.
- Velleman, Dan, David Beaver, Emilie Destruel, Dylan Bumford, Edgar Onea & Liz Coopock. 2012. *It*-clefts are IT (inquiry terminating) constructions. *Proceedings of SALT 22*. 441–460.
- Verstraete, Jean-Christophe. 2007. *Rethinking the Coordinate-Subordinate Dichotomy. Interpersonal Grammar and the Analysis of Adverbial Clauses in English*. Berlin: Mouton.
- Ward, Gregory, Jeffrey P. Kaplan & Betty J. Birner. 2007. Epistemic would, open propositions, and truncated clefts. In Nancy Hedberg & Ron Zacharski (eds.), *Topics on the grammar-pragmatics interface: Papers in honor of Jeanette K. Gundel*, 77–90. Amsterdam: John Benjamins.
- Weinert, Regina & Jim Miller. 1996. Cleft constructions in spoken language. *Journal of Pragmatics* 25. 173–206.