A QUESTION OF LANGUAGE VITALITY?

-

ON INTERROGATIVES IN AN ENDANGERED CREOLE

Kathrin Brandt

Inaugural-Dissertation zur Erlangung des Doktorgrades der Philosophischen Fakultät
der Universität zu Köln
im Fach Englische Philologie

vorgelegt von
Kathrin Brandt

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

This work would not have been possible without the generous help and support of numerous people to whom I would like to extend my gratitude. Needless to say, any errors and shortcomings in this thesis are my own.

First and foremost, I am incredibly thankful to my supervisor, Professor Dany Adone, who fostered my interest in linguistics in general and Creole languages and syntactic structures specifically, ever since my undergraduate studies. It was her encouragement that led me to pursue a PhD, her advice that brought me to Louisiana and her support and understanding when the circumstances changed that allowed me to complete this dissertation. Furthermore, my gratitude goes to Professor Marlyse Baptista for taking on the role as my second advisor, providing support from afar and accommodating my less than ideal time-line.

The fieldtrips for this study would have been impossible without financial support and I am thankful to have received funding from a.r.t.e.s. international and the Dr. Mocker Stiftung. Fieldwork in Louisiana was an incredibly enriching and positive experience thanks to the wonderful people who welcomed me and my questions and were so very giving, generously sharing their knowledge, their stories and their culture. I am deeply indebted to all Louisiana Creoles who spoke with me including but not limited to those who provided the data on which this dissertation is grounded. My heartfelt thanks go to the whole community and the interviewees in particular, among them, Joanne Babineaux, Helen Boudreaux, Corinne Cailler, Brian Costello, Aaron Flegence, Albert Guerin, Kirby Guerin, Adrian Huval, Gail Hurst, Mary Landry, Ginger Latilois, Laura Melancon, Georgia Thibadeaux, Rita Marks, Laura Joseph, Alphonse Roberts, Thomas Roberts, and Herbert Wiltz. It is my hope that with this and future works I can repay their kindness in some small and insufficient way.

I would like to express my gratitude to all who helped me find my footing in Louisiana. This includes Professors Tom Klingler, Ingrid Neumann-Holzschuh and Kevin Rottet who not only wrote the seminal works that this dissertation relies on heavily but also offered feedback and advice in the early stages of the project. Tom Klingler took the time to meet with me during my first week in Louisiana and directed me to the first consultants for which I am very thankful. No interviews would have been conducted without the many people who generously referred me to speakers. For this, our discussions of Creole culture and language in Louisiana and many other things, I thank Dr. Laina Padgett, Dr. Amanda LaFleur, Dr. Angelique Bergeron, Dr. Deborah Clifton, John LaFleur, Mavis Arnaud Frugé, George Marks, Bob Chateaulin, Carole Lancon and Lynn

Breaux. A special thank you goes to Pity and Karl Koenig and little Schumi, who made Louisiana a home away from home. As fully assimilated foreigners, they introduced me to so many community members and to the Louisiana way of life. They welcomed me into their home and, importantly, taught me how to cook Gumbo. I am very glad to have met them and their friendship is an unexpected and invaluable takeaway from my time in Lafayette.

This thesis has benefited immensely from the proofreading and constructive criticism provided by Dr. Oliver Mayeux. I cannot thank him enough for his diligence and patience and his enthusiasm for my work at times when my own was waning. At University of Cologne, Dr. Melanie Brück has been a pillar of support throughout the whole process and I am thankful to her for many different things. She has been a great friend providing encouragement and cheer and, as an equally great colleague, took on extra work to free up more writing time for me. She also gave feedback on chapter drafts even while she was doing her own fieldwork. Dr. Astrid Gramatke made numerous valuable comments on each chapter which made this dissertation so much better. I am additionally grateful to her not only for her friendship but also for serving as my sounding board for the syntactic analysis and for many deep and enjoyable discussions about syntactic theory and Creole languages. I am so grateful for the incredible help with editing given by Sandra Ringsmut and Mareike Plenk whose companionship over coffee and candy made the long days in the office feel considerably shorter and the task so much less daunting. Special thanks are also owed to Sandra for her friendship and the many ways in which she has supported me over the years.

At last, I would like to take this opportunity to thank my family who made this dissertation possible in countless tangible and less tangible ways. I could not have completed this thesis without their moral support and belief in me or without the many hours they spent on babysitting duty to allow me more time at the desk. A very special thank you goes to my parents, who gave me my love for language through nature and nurture, supported me throughout my studies, and taught me that any goal is achievable. Finally, my deepest thanks go to Stefan, for his patience while I concentrated on this project, for taking such good care of me during stressful times and for pushing me to keep going; and to David who brings me such joy and reminds me of what really matters every day.

1 INTRODUCTION

Departing from the conviction that linguistic research has much to gain from the integration of different research fields and theoretical approaches, the subject matter of the present work is deliberately interdisciplinary. Concisely said, it investigates the endangered French-lexifier Creole language of Louisiana (LC) in its ecology while focusing on wh-constructions from a generative perspective. Consequently, the discussion rests on the pillars of Creole studies, the study of language endangerment and minimalist syntax. The former two have some common ground in the emphasis they put on mechanisms and effects of language contact on societal and individual levels. Creolists are concerned with the new linguistic systems that arise from intense language contact while scholars working with endangered languages are concerned with their documentation and revitalization in addition to the study of the processes at work when a language is lost. Nevertheless, they share an interest in the ways in which language contact influences linguistic behavior, functional domains and the linguistic structure itself. For minimalist syntax this is no more than a minor concern, as the theory aims to account for all cross-linguistic structures in as uniform a fashion as possible. These positions naturally appear to be at odds with each other, and they do in fact take very different approaches but that does not mean that they are necessarily contradictory. The stance taken here is that these viewpoints can greatly inform each other.

To this effect the purpose of the present study is threefold:

I. To document LC's current sociolinguistic profile and language ecology.
II. To account for some ways in which language contact and obsolescence have sociolinguistic and linguistic effects.
III. To provide an in-depth documentation and a generative analysis of a part of LC grammar.

The following sections briefly introduce the fields of inquiry the present work builds upon and the ways in which they are relevant for LC and the analysis conducted here.

1.1 Creole studies

Creole languages are among the planet's youngest languages but they are fully-fledged linguistic codes that fulfill the complete range of purposes required of any L1. They are native languages and can be used in all domains and functions although they are less established than older languages (Adone 2012, 4). Consequently, they are not always employed to their full potential (e.g. excluded from use in official domains) as many of

them hold low prestige (Webb 2013), and they retain a high degree of structural variation and flexibility that is uncommon among heavily standardized languages. The latter property makes them particularly interesting from the standpoint of generative syntax but also potentially difficult to account for, which may be one reason why the amount of generative work conducted on Creole language is limited compared to the overall body of research.

Creole studies are naturally situated in the field of contact linguistics which emphasizes the potential languages have to impact one another as first recognized in the seminal work of Weinreich (1953). Creoles develop from intense language contact. When the need for communication drives members of different linguistic groups to develop a linguistic system this often results in a Pidgin, and Creoles have been argued to emerge from the structural extension of Pidgins, with or without nativization, for example by Mühlhäusler (1980) among many others. Whether this is always the case or not is a long-standing debate in the field as historical evidence of pre-creole Pidgins is usually scarce (Bakker 2008, Meyerhoff 2008).

A related debate revolves around Creole genesis in terms of the origin of linguistic forms and structures. It is uncontroversial that most Creoles emerged as a consequence of European expansion between the 17th and 19th century and that they derive much of their lexicon from the language of the dominant group, most often the colonizing party, the superstrate language that serves as the lexifier. The other languages involved in the contact situation are subsumed under the notion of the substrate language(s). Following the distinctions made in Arends, Muysken, and Smith (1994) approaches to Creole genesis can broadly be classified as emphasizing superstrate influence, substrate influence or linguistic universals. Relatedly, Creole formation is linked to either second language learning with adults in a formative role or first language acquisition with children as the creative agents. Virtually all creolists nowadays recognize that "no single mechanism fully accounts for pidgin and creole genesis" (Kouwenberg and Singler 2008, 7) and research into Creole genesis has largely shifted to determining the contributions made by substrates, superstrates and universals and their interaction in the formation of a new language. Notably, the least amount of attention tends to be afforded to universals in the discussion and when they are addressed scholars are usually quick to clarify that they do not mean to promote a Bickertonian view. Bickerton's highly controversial Language Bioprogram Hypothesis (LBH) was first proposed in Bickerton (2016, [1981]) and later refined and weakened (Bickerton 1984, 1988). For an overview of his claims and counter positions see Veenstra (2008a). While many aspects of the LBH clearly could

not be upheld, the central and still unrefuted but also unproven claim remains that Creole languages more closely resemble "the heart of language" (McWhorter 2002, 39) than older languages. This view connects Creole studies to basic assumptions made in the generative approach about Universal Grammar and the Faculty of Language. The present study is not prominently concerned with the genesis of LC in particular or Creole languages in general but recognizing this shared theoretical underpinning is one motivation for choosing the generative framework for the syntactic analysis.

Another large question in the field is whether Creole languages are exceptional; a notion that relates to the Creole prototype proposed by McWhorter (1998), their supposed simplicity (e.g. Parkvall 2008) and the structural properties Creoles all over the world seem to have in common (e.g. Bakker et al. 2017). Historically, many features have been classified as typically Creole but the many exceptions to each claim and the lack of agreement on particular characteristics have raised doubt whether Creoles do in fact form a typological class beyond their similar circumstance of emergence. I return to the issue briefly in Chapter 4.4.

Concerning the development of Creoles in later stages of their existence, much work has been devoted to the concept of decreolization (Mühlhäusler 1980, Bickerton 1980) which postulates a Creole-specific form of language change in which contact to the lexifier causes the Creole to shed some of its typical Creole properties in favor of structural approximation to the lexifier often resulting in a Creole continuum. This approach has been shown to be inaccurate in several points (Baptista 2015, Siegel 2010) and unable to capture some of the realities of language change in Creoles, as it disregards language internal developments and effects of contact to non-lexifier languages (Aceto 1999, Snow 2000, cf. the discussion in Mayeux 2019 for LC). According to De Camp (1971) the other paths open to Creoles are language loss and evolving to a 'normal' language. Brandt and Gabel (2017) propose a more multifaceted account of the possible paths Creoles undergo during their development.

Within the field of Creole studies Louisiana Creole (LC) is a special and therefore particularly interesting case. LC developed during the 18th and 19th century in what today constitutes the South of Louisiana. It is a French-lexifier Creole that has been in constant language contact with varieties of its lexifier as well as the English language which became increasingly more dominant over time. This linguistic diversity in conjunction with social and political dynamics and identities developed in the context of colonial society, Americanization, and racial segregation created a highly complex multilingual environment. All Creole languages are contact languages by nature and they usually co-

exist with other languages in the community but LC has been under more pressure from higher status languages than most. Due to its unique situation, it is said to have always been more similar to its lexifier than other French-based Creoles (Klingler and Neumann-Holzschuh 2013) but also to decreolize towards French (Neumann 1985) and to be under heavy English influence while also undergoing language change caused by language obsolescence. Uncovering such effects on a deep structural level is one main goal of this study.

1.2 Language endangerment

Creole languages are in no way exempt from the current linguistic crisis that has been on the forefront of much linguistic work since the 1990s. Chapter 2 below is devoted to introducing the reader to the field, which is why some very short remarks to situate the present research in the larger paradigm suffice here. LC is considered severely endangered (Klingler 2003a, Klingler and Neumann-Holzschuh 2013) and is most appropriately categorized as moribund, stage 8a on the EGIDS classification scale (Lewis and Simons 2010).

Foremost, the field of linguistics responds to the increased loss of languages with efforts to document endangered varieties within the limited timeframe during which this remains possible, with the intention of providing material on which to base maintenance or revitalization efforts (Austin and Sallabank 2011). Revitalization of LC is in its infancy but it can rely on a number of excellent works that provide a solid foundation including grammars, text collections, orthographies and dictionaries such as Neumann (1985), Neumann-Holzschuh (1987), Klingler (1992, 1996, 2003a), Valdman (1998), Valdman and Rottet (2010), and Landry et al. (2016) and the small but active community of language activists keeps adding learning material. This is not to say that the amount of documentation is sufficient which it is clearly not. More research into LC is desperately needed before it disappears, which is why this thesis attempts to contribute to this body of documentation.

At the same time, the present work is an example of another type of response to language endangerment to be found in the field. This type of research focuses on unearthing, discussing and exemplifying the processes and mechanisms at work in language endangerment and language shift to foster a better understanding of the complex interaction of factors in those scenarios (Swiggers 2007). It is in this spirit that a sociolinguistic and a structural component are included in this study in order to depict the full range of effects following from language endangerment and language contact to

be observed in LC. The sociolinguistic part aims to show how far language shift has progressed and how it affects language proficiency, functional domains, language use and language attitudes and vice versa. The structural part focuses on wh-questions as a cross-linguistically very well researched syntactic construction that has received no particular attention in the field of language endangerment, as it is not considered an area that is prone to language change. Studies on the structural effects of language shift usually concentrate on morphosyntax and the reduction of morphological paradigms (e.g. Schmidt 1985) or the loss of grammatical properties (e.g. Rottet 2001). The present study aims to show that language endangerment influences the linguistic system on a deep structural level resulting in considerable differences between I-grammars even with regard to question formation and the structural strategies accepted by individual speakers. Furthermore, previously undocumented complex questions are found to reflect the successive cyclic nature of syntactic movement.

1.3 Minimalist syntax and wh-questions

As mentioned, the theoretical framework used for the syntactic analysis is the minimalist program as the most recent form of generative syntax. Since its beginning with Chomsky (1957) the goal of generative syntax has been to provide an explanatory and comprehensive account of language structures. Operating under the basic assumption that all languages share a common core of language universals ('principles') but differ in the ways these are realized ('parameters') the approach postulates an innate language faculty containing the Universal Grammar that allows children to acquire their native language(s) quickly and easily despite the 'poverty of stimulus argument' (Chomsky 1995, 1986b). This refers to the limits of the input the child receives through child-directed speech which can never contain all possible sentences of the language as the generative power of any language is considered infinite (i.e. an unlimited number of grammatical clauses can be produced in each language) and is additionally fraud with the errors that are typical of performance but not competence.[1] Competence, the tacit knowledge of one's native language and the related ability to distinguish between acceptable and unacceptable structures therein, is impossible to observe directly but at the same time the actual target of generative research (Chomsky 1965). Under this view any situation which may cause speakers to rely on their innate capacity or universal language structures more strongly than adult speakers of established, standardized language usually do, can potentially offer valuable insights. The most obvious field to

[1] For details on the underlying concepts in generative grammar as well as a history of the theory, the reader is referred to Freidin and Lasnik (2006).

investigate is of course the acquisition of a first language by children and much generative research has been and still is conducted in this area. Less obvious fields that come to mind are Pidgins and Creole languages and language obsolescence. These areas are less clearly related to the concept of UG because they involve an already societally and individually established language rather than the formation of an I-grammar, effects from contact with other languages, sociolinguistic dynamics and a variety of influencing factors that muddy the waters. Nevertheless, for both research fields the point can be made that the structures or structural changes observed in these languages have the potential to reflect universal aspects of language more closely than other either more established or safer languages do. It follows that an endangered Creole language like LC is a very interesting candidate for a generative analysis.

Minimality has become a central concept in linguistics since the emergence of the Minimalist Program (MP) in the 1990s (Chomsky 1995). In minimalist syntax, language is conceived of as an efficiently organized cognitive system, as is expressed in the Strong Minimalist Thesis (SMT) (Chomsky 2000, 95) which states that "language is an optimal solution to legibility conditions." The SMT presupposes that the faculty of language interacts with other cognitive systems which must accept an utterance as legible. These cognitive systems require a linguistic structure to abide by certain (as of yet not fully defined) conditions which are met by human language in the most optimal way because language is designed as the optimal solution to them. Any system approaching such "good design" (Chomsky 2000, 95) should be as economical as possible, e.g. require the least amount of computational effort possible. In this spirit, minimalist syntax strives to eliminate all stipulations, motivate assumptions empirically and generally simplify the descriptive system that had gained considerable complexity in the late days of Government and Binding Theory.

Cross-linguistic and especially intra-language variation and imperfections in the supposedly optimal system present the obvious challenges to such an approach. Among the clearest empirical issues is the notion of syntactic movement, the deplacement of constituents from their base-position. Crucially, the Minimalist Program re-conceptualizes movement as internal merge[2], thus eliminating an independent syntactic operation. Instead, the MP relies on the Copy Theory of Movement (Chomsky 1995, Corver and Nunes 2007), which states that internal merge involves the copying, re-merger and subsequent deletion of a constituent which creates a chain. In accordance

[2] Movement is considered an instance of internal merge in this thesis but for the sake of readability and convention the two terms are used interchangeably.

with 'Last Resort', internal merge only applies to prevent the derivation from crashing and is assumed to occur successive cyclicly, that is to say in a series of short moves rather than one long step. Empirical arguments for this notion are discussed in Chapter 3 and Chapter 8.

Wh-movement, the phenomenon under investigation in this study, is among the best-documented and most-researched syntactic movement operations. On the one hand, this thesis intends to contribute to the growing body of research on the matter in smaller, less well-known languages. On the other hand, it investigates which structural features are observable in the special case of LC and how they can be accounted for in terms of the proposed analyses and theoretical formatives such as the Copy Theory of Movement and successive cyclicity. A previously undocumented long-distance construction, the 'ki-insertion clause' proves particularly relevant in this regard. In addition, an existing proposal by Rottet (2006) concerning the distribution of wh-pronouns is put to the test.

1.4 Research questions and the organization of this study

The previous sections have established the framework in which this work is conducted and discussed some goals and objectives, which are repeated below to equip the reader with a clear account of the research questions and aims pursued in this thesis.

1. In which domains and contexts is LC used?
2. How is language endangerment reflected in linguistic behavior?
3. Which attitudes can be found among LC speakers towards their language?
4. Which distribution of wh-expressions can be observed and is the distinction between subject and object wh-pronouns made by Rottet (2006) borne out in more recent data?
5. Which strategies can be employed for question formation in LC with regard to e.g. root clauses, complex questions, island effects and multiple wh-questions?
6. Are there any asymmetries or variation patterns in the data and how can these be accounted for?
7. How can the given structures be analyzed in the framework of the MP?
8. Does the LC data support a successive cyclic analysis of Wh-movement?
9. Which structural consequences (if any) in the formation of wh-questions follow from the close contact to English and LC's endangered status?

This thesis is organized as follows. The first part provides the theoretical background. The reader is introduced to the research field of language endangerment in Chapter 2, which briefly lays out the relevance of linguistic diversity, models of language loss and factors in the assessment of language vitality as well as some structural

consequences of language obsolescence that are documented in the literature. Chapter 3 provides an overview of the analysis of wh-questions in the minimalist framework, identifying five major research areas. Concluding the first section, the history of Louisiana, its synchronic linguistic profile, LC genesis and LC grammar are briefly described in Chapter 4.

The second part presents the research conducted for this study. The methodology is described in Chapter 5, including the field methods, such as an elicitation judgment task, the Manipulated Translation Acceptability Task (MTAT), which was developed for this study, the reasoning behind taking a qualitative approach and a brief note on the notation of LC as no official orthography exists. Chapter 6, Chapter 7 and Chapter 8 contain the data analysis and discussion. Chapter 6 presents the results from the sociolinguistic questionnaire and discusses their implications in light of the endangerment of LC in accordance with the research questions 1-3 above. After some discussion of the demographics and the speech community, it is shown how language proficiency declines intergenerationally and as language attrition on the individual level. Patterns of language use and functional domains are discussed and related to demographic variables, namely age, sex, race and hometown before language attitudes are reported. The linguistic data is the focus of Chapter 7. Two short sections provide some information on previous research into interrogatives in Creole languages and LC specifically. On the basis of the data collected on two fieldtrips in 2014 and 2015, strategies for question formation in LC and their limitations are discussed. Chapter 7.4 is concerned with variation in the data putting a special focus on the distribution of *ki*-final wh-pronouns (cf. Rottet 2006) and effects of language endangerment. These Chapters essentially address the research questions 4-6 and discuss some effects of language endangerment and language contact in accordance with RQ 9. Lastly, the generative analysis is conducted in Chapter 8 in the spirit of research questions 7, 8 and 9, arguing for a wh-copying analysis of *ki*-insertion clauses and thus providing evidence for the Copy Theory of Movement. Chapter 9 provides the conclusion.

2 LANGUAGE ENDANGERMENT

One objective of this study is to contribute to the body of documentation on Louisiana Creole. This goal is a linguistic one but it takes a different, broader perspective than research solely focused on studying linguistic structure. Such an approach needs to be understood in the context of the current global "linguistic crisis of unprecedented scale" (Crystal 2000, viii). Linguists have reacted to this crisis with a new interest in endangered languages since the 1980s (beginning with the pioneering work by Dorian (1973), (1981)), which developed into the related fields of DOCUMENTARY LINGUISTICS and LANGUAGE REVITALIZATION. This chapter provides a brief overview of language endangerment, the reasons why it comes to pass as well as why it matters, the effects it can have on a social and a structural level and the best practices for assessing a language's vitality. This chapter does not provide an account of language revitalization practices, as the revitalization of Louisiana Creole is still in its infancy. Some challenges inherent in fieldwork on endangered languages are discussed in Chapter 5.

2.1 The linguistic situation

Estimates on the number of languages spoken on the planet differ, sometimes substantially, for several reasons. First, the number depends on one's exact definition of a language. In this respect, difficulties in distinguishing between languages and socio- or dialects can account for major deviations (Chambers and Trudgill 1998, Romaine 2000, Wardhaugh 2008). Second, language change adds another level of complexity, as over time dialects can evolve into independent languages, languages can vanish from the repertoire and new languages can arise. Third, information on languages and speakers can be unreliable due to a lack of detailed surveys (projects like, Asher and Moseley (2007), Eberhard, Simons, and Fennig (2019), and Moseley (2010) are a notable exceptions but still far from comprehensive) and because people tend to over- or under-report their competence in a language, depending on social and political factors among others (Grenoble 2011). *Ethnologue* currently lists 7111 living human languages (Eberhard, Simons, and Fennig 2019), though the information found here is also known not to be fully reliable. Estimates between 6000 and 7000 languages are generally accepted, e.g. (Austin and Sallabank 2011, Nettle and Romaine 2000, Crystal 2000).

Remarkably, only 4% of these languages are spoken by 96% of the world's population, while 4% of the world's population speak 96% of the languages, according to Crystal (2000, 15). Austin and Sallabank (2011) claim that the 20 largest languages are spoken by 50% of the populace. This obviously 'uneven distribution' (Krauss 1992,

Grenoble 2011) of speakers among languages corresponds to an imbalance in power, with the speakers of majority languages holding more social and economic influence. As a result, more and more speakers of small languages with just a few hundred or thousand speakers SHIFT to the locally dominant language, a trend accelerated by a globalized society and modern communicative tools. Their native language, in turn, is supplanted and becomes endangered in their community. Endangered languages can be defined as languages that "may no longer continue to exist after a few more generations as they are not being learnt by children as first languages" (Austin and Sallabank 2011, 1). They are in danger of becoming extinct, i.e. not having any (native) speakers[3]. Classifying the degree of language endangerment is an issue that is addressed below, but linguists agree that at least half of the currently known languages are in danger of being completely lost in the near future. Less conservative estimates range up to 90% (Krauss 1992, 7) or even 95% (Krauss 2007, 3) of languages being lost or about to be lost by 2100.

2.2 In favor of linguistic diversity

Linguists of course perceive such prospects as catastrophic, as they are deeply appreciative of linguistic systems and interested in their diverse means of expression. From a scholarly perspective, linguistic diversity is essential in that many endangered languages contain rare or unique features that have not yet been sufficiently studied (Palosaari and Campbell 2011). On the emotional level, the loss of a language equals the loss of something interesting and beautiful for anybody fascinated with languages, akin to the loss of biological diversity (Hale 1992). People who do not share this valuation of languages and linguistic diversity may not agree that it matters when languages die, as long as people are able to communicate. Indeed, the point has been made that one common language would facilitate communication and as a consequence all forms of global dealings among people. Nevertheless, there are many reasons why addressing language loss is necessary.

Thieberger (1990, 334) list seven main arguments in favor of language maintenance: the preservation of national linguistic resources, social cohesion, identity, diversity, cultural maintenance, individual well-being and social justice.[4] These points are

[3] Note that some debate surrounds the issue of when a language should be considered dead. Cases where neither speakers nor documentation remain are uncontroversial. Some scholars argue that languages die when they are no longer learned by children, others claim that they are dead when just one speaker is left (Crystal 2000) and for others the existence of documentation and/or knowledge of isolated phrases suffices to assume some viability, when the language is claimed by a community (cf. the distinction made between dormant and extinct languages in Lewis and Simons (2010)).

[4] For studies on the relation between linguistic diversity and social justice see Piller (2016) and Zuckermann, Shakuto-Neoh, and Quer (2014)

more closely interconnected than listing them may suggest. This section presents the general reasoning in favor of linguistic diversity.

One major argument for maintaining linguistic diversity is that languages are intrinsically intertwined with cultures. This interdependence has been evident in the struggles within linguistics to define the field's object of study, Language (e.g. the works of Saussure, Chomsky, Whorf). The relation between a language and its environment has been the object of linguistic investigation at least since Haugen and Dil (1972) extended the ideas of Boas, Sapir and Whorf and coined the notion of LANGUAGE ECOLOGY, and is at the heart of ANTHROPOLOGICAL LINGUISTICS (Duranti 1997). Therefore, it can be said that the assumption that language and culture depend on and shape each other has become a central concept in modern linguistics. The following statement made by the Australian government in 2012 illustrates this: "A key understanding [...] is that Indigenous language is inseparable from culture […]." (Committee on Aboriginal and Torres Strait Islander Affairs 2012) as does the claim made by the UNESCO (2003, 1), "each and every language embodies the unique cultural wisdom of a people. The loss of any language is thus a loss for all humanity." Consequently, many groups see the loss of their language as entailing the loss of their culture, at least in part. As such, the situation constitutes a threat to linguistic, intellectual and cultural diversity.

It is well known in biology that diversity is a prerequisite for development and successful evolution. This also holds true for language and culture according to Crystal (2000). It can be argued that the potential of human thought, knowledge and expression is embedded within linguistic diversity; consequently, our possibilities for future development are diminished through its loss. Moreover, from the point of view of linguistics, diversity is also required to provide comprehensive data against which linguistic theories can be checked. The verification or falsification of hypotheses is dependent on the availability of diverse linguistic data. Each language has the potential to offer invaluable insights for linguistic theory that may never be gained without access to that language (Hale 1998). Mithun (1998, 189) points out that in the diversity of linguistic structure we find the "full creative capacities of the human mind". It is a key objective of linguistics to account for what is possible in human language and what is not, hence revealing our cognitive potential.

Furthermore, what is lost is not limited to a specific linguistic system and a few cultural traditions, but also includes a vast amount of indigenous and non-canonical KNOWLEDGE that is encoded in language and usually passed on orally. When communities shift to a dominant language and stop teaching their native language to their children,

they usually stop to transmit culture specific knowledge as well (Nettle and Romaine 2000) or no longer have the adequate means to do so (Jocks 1998). The new majority language is often unable to adequately express cultural concepts, such as the cultural meaning associated with certain relations and encoded in kinship systems in the Aboriginal cultures of Australia (Adone p.c.). This loss of knowledge includes, but is not limited to, the peoples' history (Crystal 2000). In cases where knowledge is transmitted through stories and songs, as is true for the Aboriginal cultures of Australia for example (James 2016), those cannot be passed on without knowledge of the language. Even written records, should they exist, are of little use to later generations unless they are familiar with the language. This indicates a serious "effect […] on human intellectual life" (Hale 1998, 193). This substantial gap in human knowledge worldwide does not only concern history but the whole of INDIGENOUS EPISTEMOLOGY, which

> refers to a cultural group's ways of thinking and of creating and reformulating knowledge […]. More specifically, indigenous epistemology involves cultural models for thinking and acting and cultural ways of conceptualizing and constructing knowledge about the human and natural worlds. (Gegeo and Watson-Gegeo 1999, 25).

This clearly goes beyond factual knowledge and touches upon notions of IDENTITY. It is a common position that identity, be it personal, ethnical, in-group or national, is created through language (Joseph 2004). Even in a more moderate view, language serves as an important marker of identity that is tightly interwoven with notions of culture, nationality, heritage and community. The serious implication is that language endangerment threatens the speakers' identity on a personal and a community level. The critical mass of science on this aspect is insufficient, as it is difficult to investigate directly. Hallett, Chandler, and Lalonde (2007) study the correlation between the use of native language and youth suicide rates among indigenous people in Canada. They find that youth suicide rates could be predicted on the basis of language use. The availability of their indigenous language seems to have an even stronger impact than other factors influencing cultural continuity, such as education, land rights etc. It is of course very hard to prove how this relates to notions of identity, though it seems to be a very plausible assumption that it does. Even if one is not prepared to draw this conclusion, the results inevitably show a link between native language and well-being, including mental health.

In the light of these arguments, language endangerment clearly is a much bigger issue than is suggested by the lack of public awareness. Linguistics, as a discipline, was thus faced with the question of how to define its role in this situation. As is described in Austin and Sallabank (2011), this resulted in many publications, a surge of funding for language documentation projects which is now coming to an end, as well efforts to support LANGUAGE MAINTENANCE, keeping a language from becoming endangered, and

LANGUAGE REVITALIZATION, reversing the process of endangerment. The most famous examples are the Maori and Hebrew language respectively. One concern with such efforts was, and still is, the complexity of the processes at work in language endangerment and potential eventual loss. Consequently, comprehending the mechanisms in place constitutes a first step.

2.3 Models of language loss

Language endangerment may result in language loss/death though this is not necessarily the case and, in some cases, the process can be halted or reversed. Nevertheless, models of language endangerment consider language loss to be the endpoint of that process and are therefore models of language loss. To avoid the endpoint of language loss, usually actions have to be taken that change some of the factors causing language endangerment.

Sasse (1992b) developed one of the earliest theories of language death. He identifies three factors investigated in contemporary studies of endangered languages. EXTERNAL SETTINGS put a language and its speakers under pressure. Accordingly, speakers change their SPEECH BEHAVIOR. Finally, the changed patterns of language use are known to result in changes in the language system, the STRUCTURAL CONSEQUENCES. In his terminology, language shift constitutes the first, extended phase of language endangerment in which speakers give in to external pressure and begin to use a dominant language rather than their native one. In the second phase, the language DECAYS, meaning that as it is used considerably less, speakers' proficiency declines and they use "pidgin-like simplification" that leads to "serious linguistic disintegration" (Sasse 1992b, 15). The final stage in his model is LANGUAGE DEATH, when the language is no longer spoken.

LANGUAGE SHIFT today is usually taken to refer to the process of one language replacing another as the main medium of communication of a community (Potowski 2013). This process involves the continuous (though non-linear) progression from the start of linguistic pressure to its endpoint in language loss, including the possible structural consequences of this process. As such, it can include what Sasse calls 'language decay'. Nevertheless, Sasse's account outlines the relevant factors and stages that are typically involved in a gradual language loss scenario, the most common form of language death (Rottet 2001)[5].

Language shift necessarily requires LANGUAGE CONTACT, namely the presence of a socially more attractive language that is adopted over time. It follows that circumstances creating language contact generally provide external settings favorable to language shift.

[5] Sudden language death scenarios, on the other hand, involve the rapid loss of all speakers e.g. through death (genocides, natural disasters) (Campbell and Muntzel 1989).

Alternative outcomes are of course possible, for example cases of stable bi- or multilingualism within communities or the emergence of new linguistic systems (i.e. Pidgins, Creoles, Mixed languages). The latter is an outcome of language contact but not necessarily an alternative to language shift. The new languages either replace the former native language(s) in the case of mixed societies such as colonial slave systems, as the main communicative system (e.g. some Creoles), which constitutes language shift, or the two co-exist in a more or less stable form of bilingualism (e.g. Pidgins).

Rottet (2001) lists external factors that are or were historically likely to cause language shift. One of the most consequential among them is European colonization, which triggered the loss of many indigenous languages in the Americas, Australia and Asia alike. While this typically involved active repression of the local languages that were perceived as inferior by the new ruling class, different forms of migration and changing populations can have similar, albeit slower effects. Minority languages tend to be associated with old ways of life, they do not represent an orientation towards advancement (Rottet 2001). It follows that economic and social advantages stand to be gained from speaking the more prestigious language in a modern society. While language contact is far from a recent phenomenon, arguably as old as Language itself according to Thomason (2001), Crystal (2000) argues that the effects were strengthened through globalization as well as modern communication and transportation tools which increased mobility and contact beyond a local scale. The influence of dominant languages and cultures consequently has a far greater reach now than in the past, allowing for new dimensions of cultural and linguistic assimilation.

As Sasse (1992b) shows, these factors ultimately affect speech behavior. A common view is that they influence language policies, attitudes and ideologies, which in turn have an effect on language use, language choice and language transmission (Bradley 2002, Rottet 2001, Dorian 1998, Grenoble and Whaley 1998a). For example, endangered languages are often actively repressed through laws or rules that forbid the use of a minority language either in particular contexts (e.g. in education) or altogether. Even in the absence of active policies opposing the use of a language, social dynamics can cause people to abandon their native language or dialect. This happens when outsiders or members of the community have negative attitudes towards the language, for example viewing the language as having no value in terms of future prospects or identity and as marking people as uneducated and socially inferior (Bradley 2002). Notions of linguistic purity and monolingual ideologies that are prevalent in many Western societies contribute to such beliefs and a negative perception of bilingualism. The nationalist

movements in Europe during the 18th and 19th century advocated for similar views which became deeply rooted in many Western cultures. Under the catch phrase 'One Nation, One Language' an ideological position was developed claiming that a common language forms a common national identity, with Germany, France and Russia providing some well-known examples of such policies (Gardt 2000). This type of thinking is still influential, as is for example reflected in the English-only movement in the USA (Dorian 1998, Ricento 2013).

External factors put speakers under PRESSURE to learn the dominant, more prestigious language of their surroundings. For Crystal (2000), this is the first stage of a language loss scenario. Once the community members have some competence in the L2, the stage of BILINGUALISM begins. Without any change of the circumstances or some form of intervention, the use of the original language(s) will continue to decline. This development can last a long time or progress rather quickly, largely depending on the particular situation and language attitudes, which in turn are influenced by language policies and ideologies. This does not involve a clearly defined linear progression but a complex process involving socio-historical conditions, linguistic behavior, linguistic as well as social structure, ideologies and attitudes (Tsitsipis 1998).

Due to the continued presence and availability of another language, people change their linguistic behavior. Usually, the more prestigious language is used in all formal capacities with the non-dominant language becoming restricted to the private sphere to differing degrees. Such a separation of languages through use in different FUNCTIONAL DOMAINS does not always follow the same patterns. Cases where the minority language disappears from HIGHER DOMAINS, i.e. formal situations, politics and public life, and becomes associated with LOWER DOMAINS, i.e. the home and the private sphere, follow the top-to-bottom pattern of language shift. East Sutherland Gaelic is a typical example according to Dorian (1981). Ancient Hebrew (before the 'revival'), Modern Standard Arabic or many Aboriginal languages of Australia, on the other hand, are cases of the opposite bottom-to-top pattern (Campbell and Muntzel 1989, Blair and Fredeen 1995). Here, the language is retained in some very high, formal, ritual or religious functions while vanishing from peoples' everyday lives. Languages that occupy separate domains can co-exist in DIGLOSSIA where each language has its established social functional (Fishman 1967, 2000, Ferguson 2000) in a manner similar to the roles fulfilled by standard languages and dialects respectively[6]. In an unstable bilingual situation, the

[6] The original notion of DIGLOSSIA as proposed by Ferguson (1959) is concerned with high (H) and low (L) varieties of the same language. The extension to historically unrelated languages was advanced by Fishman (1967) among others.

functional distinction is less clear and is eroded over time, with speakers using one language in more and more functional domains. Unstable bilingualism is an intermediate state, existing in transition (Fishman 2000, Rottet 2001) and ultimately leads to a reduction of social functions that are fulfilled by the less powerful language. This is a common occurrence in language shift situations where the endangered language vanishes from various areas of life over time.

The loss of intergenerational transmission is another crucial development in language shift. Models of language endangerment or death generally acknowledge either a lack of language acquisition by children as a sign of language endangerment or the necessity of child language acquisition in order to maintain it (UNESCO 2003, Krauss 1992, Wurm 2002). A breakdown of transmission, which prevents children in the community from acquiring a language as their L1, results in a change of speaker demography and a decline of the total number of speakers. Only older community members speak the language, until those speakers pass away. When children grow up with the dominant language as their L1 they do not acquire enough competence in the non-dominant language to pass it on to their children. Furthermore, they will use it very rarely and their attitudes towards that language are often negative to the point of feeling ashamed of it (Dressler 1982, 328). Consequently, they do not transmit the minority language to their children and the community becomes monolingual in the majority language.

Fishman (1972, 1980) developed a language assimilation model for immigrants in the United States, who became monolingual in English by the third generation. An analysis of US census data by Alba et al. (2002) confirmed this pattern, with the exception of Spanish speaking immigrants half of whose third generation appears to preserve their native language. Sometimes intergenerational transmission stops abruptly (Dorian 1986) leaving the community with members who are fluent speakers as well as a younger generation that has no competence in the language at all. This often involves a conscious decision on the part of the parents not to raise their children speaking the non-dominant language in favor of perceived greater competence in the majority language and the associated socio-economic advantages. For many languages the process is more gradual. Usually, an intermediate generation grows up bilingually, using each language in the allotted domain at least for a time. Rottet (2001, 29) distinguishes between cases of switched language dominance and cases of non-reciprocal language use. Both can be found while language shift is in progress and children are exposed to the endangered language to some extent but it is not the undisputed medium of communication nor used

in all domains. Even children who grow up primarily speaking the minority language may change their choice of preferred language once they go to school if another, socially more dominant language is the medium of instruction and communication in this setting. Increased contact with people outside the community or community members who have already shifted languages can have a similar effect. The speakers may compartmentalize according to functional domains for a varying period of time but their use of the majority language will likely increase and take over domains successively. In cases where a language is of very low prestige or perceived as backward younger generations may refuse to speak it. When approached in that language by elders they often reply in the competing language. This may be due to a lack of competence in the case of passive/receptive bilinguals or be a deliberate choice, reflecting negative attitudes towards the language.

Such processes create a diverse range of speakers, allowing for much more variety than is usually found among speakers of established languages. The speakers of non-endangered languages are usually either native, and hence fluent, or L2 learners of varying proficiency. Endangered languages differ in that their native speakers are not necessarily fluent. This observation goes back to Dorian (1977) who introduced the category of SEMI-SPEAKERS[7] to refer to speakers who were not fluent in Gaelic but could understand rather well and even speak it to a varying extent. These speakers were fully accepted community members and some made regular use of their mother tongue, but fluent speakers perceived many structures they produced as mistakes. The term has since fallen out of use, with such speakers included under the slightly broader category of HERITAGE SPEAKERS instead, that is commonly used to refer to speakers who were not fully socialized in their home language and show different levels of competence in it as a result (Polinsky and Kagan 2007). Dorian's typology of speakers has been extended and revised, for example by Campbell and Muntzel (1989) who added the category of REMEMBERERS i.e. formerly fluent speakers who suffer from L1 attrition but may remember their language by making a conscious effort. For other speakers, the lack of proficiency is due to incomplete acquisition rather than attrition because they did not acquire the language to a sufficient degree or at all. This depends largely on the amount of exposure to the obsolescent language during childhood (Sasse 1992a). Grinevald and Bert (2011) note that speakers do not only vary in terms of their individual competence but also in their language attitudes and usage patterns, their level of linguistic security and the degree of first language acquisition they attained, which relates to the state of the language at the

[7] Please note that the term has been criticized for implying a cognitive deficit in those speakers. I will not adopt it for that reason.

time of a speaker's birth. As a consequence, speakers can have various different levels of competence and language attitudes. GHOST SPEAKERS for example are speakers who deny knowing the language despite clear competence. NEO or NEW SPEAKERS on the other hand did not acquire the language as their L1 but learn it as an L2 later, either out of a general interest (e.g. linguists) or to claim their linguistic and cultural heritage. Fluent speakers, in the typology used by Grinevald and Bert (2011), can be further divided into younger and older fluent speakers, a distinction first made by Dorian (1981). Both use the language on a regular basis and are highly proficient, but the latter speak a more conservative variety than the former who tend to use a changed form. Despite the terminology this is less a matter of age than a question of at which point in the development of the language speakers acquired it.

2.4 Assessing language vitality

Diachronically, gradual language death or language shift entail change. The language in question starts out as the community's main language, a stable linguistic system. Changes in the social, sociolinguistic and structural make-up are often subtle in the beginning and become more pronounced over time. From a synchronic perspective language endangerment is hence a matter of degree. In between the two opposite poles of vibrant and extinct, many languages exist in an intermediate state. At any given point in time, different languages occupy different places on this continuum, their endangerment of varying severity. This naturally gives rise to the question of how to assess language vitality and how to classify languages accordingly. Several approaches have been made to the issue, focusing on different factors that were considered the most influential on language vitality. The related field of language maintenance and revitalization is concerned with measures than could be taken at different stages to stop language shift or even reverse processes of language endangerment and loss. There have been many approaches to achieving such change, often initiated or carried out with and by the community itself [8]. While this is a large and interesting field, it concerns the topic of this thesis only where it intersects with language documentation and is hence not discussed in any detail.

Regarding the assessment of language viability, Dixon (1991) focuses on Aboriginal languages of Australia. He postulates five stages of language endangerment, mainly relying on the absolute number of fluent, native speakers. In stage one, the

[8] For more information on strategies for language maintenance and revival as well as case studies the following (among many others) provide insightful contributions: Hinton, Huss, and Roche (2018), Austin and Sallabank (2011), Grenoble and Whaley (1998b), Bradley and Bradley (2002)

language is not yet endangered but "is used as the first language by a full community of at least some hundreds of people and is used in every aspect of their daily lives" (Dixon 1991, 237). In stage two the language already is an L2 for some community members while most still acquired it as their native language. L1 speakers are few and rather old at stage three. English, the invading language in the case of Australia, is the native language for most community members at this stage. When L1 speakers of the endangered language die, nobody has full knowledge of the language anymore. Some speak a "modified version of X, with simplified grammar" (Dixon 1991, 237), that is a changed form of the language. At stage five, language shift is complete, with the whole community speaking English.

While he approaches the issue from the angle of the number of speakers, there are clear overlaps with the issues of speaker demography, functional domains and intergenerational transmission. He explicitly refers to the age of fluent speakers and mentions the use in "every aspect of their daily lives" (Dixon 1991, 237) as a characteristic of a non-endangered language (stage 1). It is implied that this functional distribution will change in the course of language endangerment. Intergenerational transmission necessarily underlies the distinction between L1 and L2 speakers Dixon postulates from stage two.

Fishman (1991) proposes another classification model, the GRADED INTERGENERATIONAL DISRUPTION SCALE (GIDS). His is a sociocultural scale that is primarily concerned with the loss of functional domains and resulting disruptions, which he splits into eight stages. Fishman considers those stages to be "quasi-implicational" (Fishman 1991, 87), indicating that higher stages presuppose the developments lower on the scale. His focus is narrower in that he includes neither a state of completed language shift, nor a stage prior to endangerment. Instead, his classification begins with the language (X) being used in public domains to some extent but not primarily. At the same time, his focus is broader in that he includes measures to be taken for X to reach the next lower stage, reviving the language. Table 2-1 contains the eight stages, with the description changed to the corresponding status description for stage six (Fishman 1991, 92) where Fishman employs the goals for maintenance efforts as a header.

Table 2-1 Classification by Fishman (1991)

Stage 1	Some use of Xish in higher level education, occupational, governmental, and media efforts
Stage 2	Xish in lower governmental services and mass media but not in the higher spheres of either
Stage 3	Use of Xish in the lower work sphere (outside of the Xish neighborhood/ community) involving interaction between Xmen and Ymen
Stage 4	Xish in lower education (types a and b) that meets the requirements of compulsory education laws
Stage 5	Xish literacy in home, school and community, but without taking on extra-communal reinforcement of such literacy
Stage 6	Xish is normal language of informal, spoken interaction between all three generations of the family, with Yish reserved for matters of greater formality and technicality than those that are the common fare of daily family life
Stage 7	Most users of Xish are a socially integrated and ethnolinguistically active population but they are beyond child-bearing age
Stage 8	Most vestigial users of Xish are socially isolated old folks and Xish needs to be re-assembled from their mouths and memories and taught to demographically unconcentrated adults

Fishman is most concerned with the functional domains in which X and Y are used respectively. The first five stages focus on the usage of the endangered language in public domains and the media as well literacy in X. This is a factor not considered in Dixon's approach and it may be of great importance for some endangered language. It is not however a necessarily relevant factor for all languages, as many indigenous languages are and have always been oral and their speakers seek to preserve them this way. For indigenous languages spoken in remote places, as for example the Aboriginal languages of Australia or the Amazon, what Fishman describes as stage six constitutes a stable and desirable state (Adone p.c.). Upon closer inspection, his scale includes many factors beyond functional domains, such as the age and social integration of speakers.

Wurm (2002, 14) lists "hallmarks for various levels of […] endangerment". His five stages are based the decline of use of the language and ultimately state that the older the speakers are whose usage decreases, the more severe is the case of language endangerment. A language is only potentially endangered if children's use decreases but seriously endangered if the decline happens for middle-aged adults and fluent speakers are aged 50 and above according to Wurm (2002). While his account is very brief, it nevertheless contains notions of speakers' age, usage patterns and intergenerational transmission.

Table 2-2 Classification by Wurm (2002)

Potentially endangered	Child speakers declining
Endangered	Few speakers left, no children
Seriously endangered	No good speakers younger than 50
Moribund	Few old speakers
Extinct	No fluent speakers

From the approaches presented above it appears that multiple, potentially interacting factors contribute to a language endangerment scenario. All researchers consider the speakers' age and the amount of usage of the language in their models. Implicitly, they also refer to intergenerational transmission, that is to what extent children in the community acquire the language. Only Fishman (1991) includes functional domains.

In 2003, an expert group on behalf of UNESCO developed the most comprehensive system for assessing language vitality to date (UNESCO 2003), which is currently in use. The following section describes the proposed system based on the document published in 2003. They list nine factors that serve as indicators for and contribute to the level of language vitality/ endangerment. The first six relate to observable factors that may lead to language endangerment. Points seven and eight are concerned with language attitudes and factor number nine can be relevant to decide how urgently a specific language should be documented. These factors are:

1. Intergenerational Language Transmission
2. Absolute Number of Speakers
3. Proportion of Speakers within the Total Population
4. Shifts in Domains of Language Use
5. Response to New Domains and Media
6. Availability of Materials for Language Education and Literacy
7. Governmental and Institutional Language Attitudes and Policies, Including Official Status and Use
8. Community Members' Attitudes towards Their Own Language
9. Type and Quality of Documentation

(UNESCO 2003)

The authors emphasize that none of the indicators can be meaningful if regarded in isolation. Only considering all influential factors for any particular language can create a

clear and comprehensive picture of its viability. For each factor a scale indicating levels of endangerment from stable (grade 5) to extinct (grade 0) is provided.

INTERGENERATIONAL TRANSMISSION refers to children learning the language in their homes. This is closely linked to the speakers' age, as speakers necessarily grow older on average when no younger generations speak the language anymore. Consequently, the associated scale is not concluded when children no longer learn the language, at which point it is considered definitely endangered (grade 3), but continues on the basis of the speakers' age. This is commonly perceived as the most crucial factor in language vitality.

Krauss (1992, 2007) considers languages that have more than 10,000 speakers to be safe. This view is shared by Nettle (1999) who claims that about 60% of the world's languages have fewer speakers. The ABSOLUTE NUMBER OF SPEAKERS can be an unreliable indicator for three reasons. First, a small language can be more stable than the numbers may suggest as is illustrated by some small languages spoken in Papua New Guinea that are used by all generations and are in fact growing despite having just a few hundred speakers. Second, 'large' languages can be lost quickly when children do not acquire them as can be seen in the continuous downwards trend found for a number of Celtic languages (Barrena et al. 2007). Third, it is close to impossible to reliably count speakers of endangered languages due to the diversity of speaker types (Grinevald and Bert 2011) and the unreliability of their self-assessments outlined in Chapter 2.3 above. Nevertheless, a language spoken by millions is much more likely to remain stable than one spoken only by a handful of people and Barrena et al. (2007) find that languages spoken by fewer than 1000 individuals tend to have decreasing rates of transmission and usage.

The absolute number of speakers provides the most insight when it is considered alongside the PROPORTION OF SPEAKERS WITHIN THE TOTAL POPULATION. 100 speakers can form the complete population of an isolated village or they can be a very small minority within a larger population that they identify with. The former group would under most circumstances be more likely to retain their language. When the speakers of a minority language are part of a large community of majority language speakers they have more incentive to shift languages in order to function as community members. Such group membership can be based on "ethnic, religious, regional or national" (UNESCO 2003, 9) identification. A six-stage scale of language vitality is provided encompassing scenarios where all community members speak the language (grade 5 - safe) to the point where none do (grade 0 - extinct).

The fourth factor in the UNESCO model is concerned with SHIFTS IN THE DOMAINS OF LANGUAGE USE in the tradition of Fishman. The basic assumption is that how much and with whom a language is used will influence the rate of intergenerational transmission. Diglossia (stage 4 in the UNESCO model) is not necessarily considered a threat to the non-dominant language. The tipping point is instead reached at stage three ('dwindling domains'), where the language begins to recede from the home domain.

The fifth factor is also related to the notion of functional domains. A language's RESPONSE TO NEW DOMAINS AND MEDIA determines if and to what extent the language is used in new domains, such as education, work environments etc., and adjusts to accommodate new fields (of work, technology or thought) in its vocabulary. The use of the language in new domains can be decisive regardless of whether traditional domains are lost or not. This is closely connected to language use in new media, which can be crucial as for example in the case of television, which can contribute substantially to the spread of the dominant language due to its wide distribution and reception.

Finally, the AVAILABILITY OF MATERIALS FOR LANGUAGE EDUCATION AND LITERACY generally gives a better chance of maintenance to an endangered language, as literacy is associated with progress in most societies. The effect is especially positive where the written language is used in schools, as a medium of instruction, and in administration. Many minority languages have a long-standing oral tradition and some communities do not wish to change that fact. A language that is not written is excluded from many important domains in modern societies, such as literature, law, policymaking, governance and news as well as social media. The associated scale is graded as shown in Table 2-3.

Table 2-3 UNESCO on the availability of written materials

Grade	Availability of Written Materials
5	There is an established orthography and a literacy tradition with grammars, dictionaries, texts, literature and everyday media. Writing in the language is used in administration and education.
4	Written materials exist, and at school, children are developing literacy in the language. Writing in the language is not used in administration.
3	Written materials exist and children may be exposed to the written form at school. Literacy is not promoted through print media.
2	Written materials exist, but they may only be useful for some members of the community; for others, they may have a symbolic significance. Literacy education in the language is not a part of the school curriculum.
1	A practical orthography is known to the community and some material is being written.
0	No orthography is available to the community.

While the first six factors are observable and measurable to some degree, language attitudes are a subtler but nonetheless powerful force, as mentioned above. The ad hoc expert group distinguishes between GOVERNMENTAL AND INSTITUTIONAL ATTITUDES (factor 7) on the one hand, and COMMUNITY LANGUAGE ATTITUDES (factor 8) on the other. Such attitudes can concern bi- or multilingualism in general, the perception of the endangered language as pleasing or difficult and the acceptance of public use of a minority language by both the minority and the majority groups (Bradley 2002, 1). In the case of governmental attitudes these translate implicitly or explicitly into language policies, such as the (lack of) official recognition of the non-dominant language and its use in education and other public domains. Governmental attitudes thus result in either support or suppression of the minority language as well as several intermediate states as for example passive assimilation (grade 3). This refers to the preference for the dominant language in public domains although there are no active policies to this effect because the governing body is indifferent towards minority languages. The scale proposed by the expert group ranges from equal support over different degrees of assimilation to prohibition (UNESCO 2003, 14).

Community attitudes are likely to be influenced by governmental and majority attitudes. But even the most positive policies cannot maintain a language if it is rejected by the speech community. Within the community the native language can be an integral, highly valued and important part of their identity or it can be perceived to impede their efforts for economic and social success. In cases where the non-dominant language is stigmatized due to having very low prestige and/or its use is actively prohibited, speakers have been known to be ashamed of their language and refuse to speak it (Ravindranath 2009). Such attitudes have clear and obvious consequences for the vitality of the language, as they have a large impact on language transmission and usage. The UNESCO scale is based on the proportion of community members that hold positive or negative attitudes towards their language.

The last factor refers to the level of existing language documentation. Language data in the form of texts and recordings, their transcriptions and analysis and, in the best case, grammars and dictionaries are important tools in projects concerned with language maintenance or revitalization. The TYPE AND QUALITY OF DOCUMENTATION determines which steps can reasonably be taken in order to improve the situation of an endangered language. They also offer communities the opportunity to remember forgotten knowledge and connect with their heritage in the case of complete language loss. With

the help of a six-grade scale the state of documentation can be classified from undocumented to superlatively documented languages.

On the basis of these nine factors and the respective classification the overall vitality of a given language can be assessed. It must be noted that the authors explicitly caution against the simple addition of the numbers a language occupies on each scale (UNESCO 2003, 17). The situation is more complex than will be suggested by any generalizing model of classification and each case is unique in its specific dynamics and results. The factors, scales and their interpretation hence have to be adjusted to each particular language, context and purpose. Even within one community the process of language endangerment and its effects are not uniform: "language death is a gradual and varied process, with the general trends affecting different members or sections of a community in a different way" (Crystal 2000, 89). In this spirit, no overall scale for the level of language endangerment is provided. Earlier attempts of such classifications, as described above, usually contain five or more levels of language endangerment (e.g. Kincade 1991, Wurm 1991). A similar, more recent model is provided by Krauss (2007) and takes into account many of the factors described by the UNESO ad hoc expert group. Table 2-4 gives an overview of the proposed levels of language vitality/ endangerment. Note that more fine-grained classifications employing a combination of levels, e.g. *a*-*b* or +/- distinctions (following the pattern used for grading in the school systems), are possible in this model and allow for a more precise reflection of mixed stages of language endangerment within one community.

Table 2-4 Framework for classifying languages

'safe'			*a+*	
e n d a n g e r e d	*stable*		*a*	all speak, children & up
	i n d e c l i n e	*unstable; eroded*	*a–*	some children speak; all children speak in some places
		definitively endangered	*b*	spoken only by parental generation and up
		severely endangered	*c*	spoken only by grandparental generation and up
		critically endangered	*d*	spoken only by very few, of great-grandparental generation
extinct			*e*	no speakers

(Krauss 2007, 1)

According to Krauss (2007), the SAFE category, *a+*, refers to languages that are not only still acquired by children but also are likely to continue serving as mother languages for future generations. He estimates that only 5% of all languages currently fall into that category and states that no language with fewer than 10.000 speakers should be labeled safe. EXTINCT languages, on the other hand, classified as '*e*', are languages that, beyond a lack of speakers, also lack any possibility to obtain documentation. Cases in which a few words and phrases are remembered or where written sources exist although the language is no longer spoken are on the margins of that category and classified as '*e+*'. All languages that fall in between these two extremes are considered endangered to varying degrees. That includes so-called STABLE languages, category *a*, which are spoken in the home domain and acquired by children regardless of the amount of external pressure from a dominant language that may be increasingly used in public domains such as education and media. While such languages are endangered, they are not yet caught in a spiral of decline. That changes when children cease to use the language as a standard medium of communication, e.g. in cases when only some children use it regularly or they only use it with certain interlocutors but not among each other. Krauss (2007) considers such languages to be UNSTABLE (category *a-*). This category also includes ERODING languages that are spoken by all children in some parts of an extended speech community but only by some in other parts. When children no longer learn the language as an L1 and

fluent speakers are of the parent generation and older, it is categorized as DEFINTELY ENDANGERED (category *b*). The next class, '*c*', contains SEVERELY ENDANGERED languages which are characterized by a lack of competence in the parent generation as well, preventing them from passing the language on to their children even if they wish to do so. Fluent speakers are to be found in the grandparent and older generations in these cases. Finally, languages whose very few speakers belong to the great-grandparent generation are classified as CRITICALLY ENDANGERED, category *d*.

Lastly, Lewis and Simons (2010) combine the UNESCO system with Fishman's GIDS and the scale employed by Ethnologue to form the E(xpanded)GIDS. This scheme contains more finely grained categories primarily at the ends of the spectrum to account for the many different situations languages can be in which are considered safe in the UNSECO model. At the other end of the continuum, a distinction is made between DORMANT languages that are no longer spoken but are claimed by a community where they serve as identity markers as part of their heritage, and languages that a fully EXTINCT. This system allows a better assessment of which languages may become vulnerable in the future.

Table 2-5 Expanded Graded Intergenerational Disruption Scale

Level	Label	Description	UNESCO
0	International	The language is used internationally for a broad range of functions.	Safe
1	National	The language is used in education, work, mass media, government at the nationwide level.	Safe
2	Regional	The language is used for local and regional mass media and governmental services.	Safe
3	Trade	The language is used for local and regional work by both insiders and outsiders.	Safe
4	Educational	Literacy in the language is being transmitted through a system of public education.	Safe
5	Written	The language is used orally by all generations and is effectively used in written form in parts of the community.	Safe
6a	Vigorous	The language is used orally by all generations and is being learned by children as their first language.	Safe
6b	Threatened	The language is used orally by all generations but only some of the child-bearing generation are transmitting it to their children.	Vulnerable
7	Shifting	The child-bearing generation knows the language well enough to use it among themselves but none are transmitting it to their children.	Definitely Endangered
8a	Moribund	The only remaining active speakers of the language are members of the grandparent generation.	Severely Endangered
8b	Nearly Extinct	The only remaining speakers of the language are members of the grandparent generation or older who have little opportunity to use the language.	Critically Endangered
9	Dormant	The language serves as a reminder of heritage identity for an ethnic community. No one has more than symbolic proficiency.	Extinct
10	Extinct	No one retains a sense of ethnic identity associated with the language, even for symbolic purposes.	Extinct

(Lewis and Simons 2010, 110 adapted from Fishman 1991)

2.5 Structural consequences

Language endangerment is characterized by variability, as is clear from the above. This is not limited to the diversity of factors involved in its emergence to varying degrees or to the diverse speaker types that develop and the corresponding differences within the speech community. As Sasse (1992b) recognizes, language endangerment can also cause changes in the structure of the minority language, a process he refers to as *language decay*. These structural consequences can affect all descriptive linguistic levels, from phonology to syntax. As the language becomes more endangered, speakers in turn become less fluent and develop linguistic insecurity which Rottet (2001, 25) describes as a breakdown of linguistic norms making speakers lose their sense of 'correctness'. This is one important reason for the amount of variability found on the structural level in an endangerment situation. Speakers are no longer certain which form is grammatical and often use different words and constructions in free variation, that is their choice does not reflect any social variables, as is usually the case with regional, social or gender variation (Palosaari and Campbell 2011).

Often the first observable effect is frequent code-switching. Code-switching is common among proficient bilinguals and does not necessarily indicate a language endangerment situation. Neither does it imply a lack of competence in either language involved. Myers-Scotton and Jake (1995) show that it obeys structural restrictions. One language, which they term the MATRIX LANGUAGE, provides grammatical/functional morphemes while the EMBEDDED LANGUAGE only provides lexical content morphemes in MIXED CONSTITUENTS. In the case of endangered languages, the dominant language serves as the matrix language, hence providing the grammatical frame but also many lexemes. Lexical borrowings are not at all unusual in 'healthy' languages and can enrich the vocabulary. In the case of endangered languages "the influx of non-native forms tends to be pervasive" (Aikhenvald 2012, 101) and goes hand in hand with a decline in the speakers' lexicon in the non-dominant language. As the availability of native lexemes is reduced, non-native lexemes are used for greater ease of communication. This development goes beyond code-switching and reduces speakers' knowledge of the language as well as its value as a means of communication. It illustrates one main characteristic of structural change that occurs because of language endangerment, namely the LOSS of sounds, morphs, lexemes or structures. At the same time new elements can be introduced from a dominant language, e.g. vocabulary items and structural features. It appears, on the surface, as if the changes to be found in endangered languages

are to some extent 'random', with some elements being lost while others are added to the language.

It is true that language change in an endangered language is hard to predict but some observable tendencies are described in the literature. Indeed, loss remains the overarching theme on a structural level. REDUCTION is usually taken to refer to the decline in native vocabulary described above. SHRINKAGE, on the other hand, refers to the loss of styles and registers in a language. This can be most clearly illustrated by some languages of Aboriginal Australia, where the cultural taboo relation between a person and their in-laws is reflected in a specific style of speech to be used around them. Dixon (1980) reports that in Dyirbal, an indigenous language spoken in Queensland, North East Australia, this style is grammatically almost identical to regular speech but makes use of a completely separate vocabulary. Tsunoda (2006) suspects a similar system for the neighboring language Warrungu which cannot be confirmed, as the remaining speaker did remember the existence of such a style but had no knowledge of the words. Tsunoda (2006, 98) takes this to indicate that such styles are lost relatively early in the process, as the same speaker was generally fluent in Warrungu.

On the structural level, (morpho-) syntactic structures can be subject to loss as well. Schmidt (1985) provides an example from Dyirbal, where ergativity was lost in younger speakers' language (Young Dyirbal, hereafter YD). She reports the REDUCTION of allomorphy in Dyirbal case morphology before the ergative was lost completely. Schmidt (1985) describes five stages until the complete absence of ergative marking, beginning with the presence of nine allomorphic conditions, which were gradually reduced to just one, the suffix *-gu*. When found in young speakers' Dyirbal, this suffix no longer marked ergative case. She found that younger speakers would follow the pattern of the dominant language, English, relying on word order for the identification of subjects rather than the ergative-absolutive structure of traditional Dyirbal. Since subjects of intransitive verbs are not distinguished from those of transitive verbs in this new system, she assumes an underlying nominate-accusative structure despite the continued use of the affix *-gu*. Similarly, the locative suffix, which has seven allomorphs in traditional Dyirbal, was reduced to just one or two in younger people's speech and ultimately replaced by a prepositional construction similar to the ones found in English (Schmidt 1985).

American Finnish (AF) has a different socio-linguistic background, being a contact variety of a stable language spoken by immigrants in the USA. The structural effects observed by Larmouth (1974) nevertheless are quite similar, as speakers seem to loose morphological case markings on adjectives, using case affixes only on the associated

nouns. The loss of morphological case marking appears to be a possible structural consequence of language endangerment, especially for peripheral cases as Schmidt (1985) documents for Dyirbal, where the morphological marking of locative, allative, possessive and instrumental was replaced by the generalized ergative suffix and analytic constructions. In this language the whole ergative-absolutive system seems to be lost, thus affecting not only the morphological form but also the structural make-up of the language.

Often, the loss of one form or structure does not constitute a full loss, as the grammatical distinction remains at least for an intermediate period. Where several means exist to express the same thing, speakers tend to favor just one in an endangered language. This can mean the loss of alternative expressions or their gradual assimilation to another, preferred construction. In other words, alternations in the linguistic system are eliminated in favor of REGULARIZATION. Dorian (1973), (1980) describes this SIMPLIFICATION for East Sutherland Gaelic by referring to the passive construction. Passive is marked in this variety of Gaelic by two possible constructions, the *bith*('be')-passive and the *dol*('go')-passive. The former construction employs a subject pronoun and a preposition while the latter does not use an overt pronoun but involves lenition to mark phi-features. Dorian noticed that some speakers would use the characteristics of the *bith*-passive in a *dol*-construction, adding a pronoun and preposition. This trend clearly corresponded to the age of the speakers. Older speakers produced the 'correct' traditional form of the *dol*-passive. Only speakers over the age of 80 did so 100% of the time with the changes becoming more frequent the younger the speakers were. By creating this structure analogous to the *bith*-passive construction, young speakers simplified their system as it allowed them to disperse with the subtle and difficult phonological change that is required to indicate phi-features of the subject in the subject-less construction. Phonological simplification can be observed in the loss of phonological distinctions as suggested by Anderson (1983). Palosaari and Campbell (2011, 112) give the example of the indigenous Central American language Pipil spoken in El Salvador that had contrastive long/short vowels, a distinction which many speakers have lost.

Simplification and regularization, as described above, do not necessarily contradict the notion of variability as previously mentioned. While it is true that the former suggests the loss of alternative forms and the latter implies the co-existence of several forms, which appear in free variation, simplification happens over time. It appears plausible that individual speakers would go through a stage in which they do not know or follow the rules of distribution for allomorphs and allophones or the 'correct'

properties of syntactic constructions. They may use them in free variation as described by Palosaari and Campbell (2011) before ultimately settling on a pattern or losing the grammatical category completely.

Campbell and Muntzel (1989) note that the INTRODUCTION OF NEW MATERIAL into the language is not limited to vocabulary items, as described earlier in this section. Instead, they argue that structural features of the dominant language can enter the endangered language's grammar. Their point is illustrated with an example from American Finnish, which employs overt agents in impersonal passives and non-finite complements of nouns where standard Finnish does not. The replacement of the ergative by the nominative-accusative system in Dyirbal as described by Schmidt (1985) could be argued to prove a similar point, though not as clearly as there is no overt marking of a new case system. The clearer case in YD is that of English prepositions replacing locative case marking, which also illustrates a development towards analytic language structure rather than syntheticity. The new function YD speakers assign to some bound morphemes, such as the comitative suffix *–bila* that now serves to mark instrumental case as well, represents another form of INNOVATION to be found in the changed grammar of an endangered language (Schmidt 1985). The replacement of structural features will likely go hand in hand with an intermediate period of variability in which both structures are used within one community or even by the same speaker.

Every case of language endangerment is unique, as the cultures, people, languages and circumstances involved are diverse. Nevertheless, some changes have been observed in several endangered languages and Rottet (2001) proposes a list of possible generalizations for changes that cannot be directly attributed to a contact language. The development from a synthetic to a more analytic linguistic system (1) seems to prevail and has been described above for YD. The reduction of inflectional morphology (2) is also clearly illustrated by YD and AF above. The elimination of redundancy (3) is observable for example in the loss of adjective-noun agreement in AF. Fantasy morphology (4) is a rarely observed phenomenon described by Sasse (1992a) for which he provides an example from Arvanitika. Speakers who lack the knowledge of the correct form make up morphemes that are not at all part of the language of fluent speakers but sound like a potential element of the language. The loss of avoidance styles in Indigenous Australian languages illustrates the reduction of stylistic variation (5) and phonological changes (6) were described above for Pipil. Extensive non-meaningful variation (7) refers to the use of alternate forms and structures where their distribution is conditioned neither by linguistic (e.g. allomorphic) nor social variables. Lastly, Rottet lists the loss of productive

word formation (8) and the avoidance of multiclausal sentences (9). It remains unclear if these effects occur in any particular order cross-linguistically hence suggesting an increasing degree of pathology associated with their appearance. Tsunoda (2006) for example concludes from the study of Warrungu that stylistic variation, point five in Rottet's account, is lost at an early stage of language endangerment. Productive word-formation, his point 8, can only be replaced by borrowing from another language to account for new domains or speaking about things that the language has no words for in another language. Such behavior has been reported for multiple endangered languages and whether it constitutes the loss of productive word formation would depend on whether they coin any new words at all. The complexity and heterogeneity of language endangerment make it unlikely that structural effects will appear in a rigid chronological order.

Finally, it should be made clear that many structural changes occurring in endangered languages are not unusual per se. They do not uniquely result from language endangerment. Rather, these are natural changes one would expect to find in any language, such as the incorporation of lexemes from a contact language, the loss of phonological distinctions or the reduction of allomorphy. Crystal (2000) attests a different "extent, range, rate and quality" to the changes occurring in endangered languages as opposed to natural language change. In his view, the processes themselves remain similar but they occur quicker, more frequently, at several linguistic levels at the same time and are more radical than usual. Rottet (2001) agrees that the changes found in endangered languages are regular processes that occur in a higher than usual density. He distinguishes between changes that are the result of language contact and hence externally caused, and internally caused process of restructuring and simplification. Teasing apart those motivations is difficult in any language contact study and especially so in the case of endangered languages where contact often is a big part of what causes endangerment and effects cannot be linearly traced back to a single cause. It is hardly possible to determine with certainty which changes would have occurred had the language not been endangered but all other factors remained the same.

3 SYNTACTIC THEORY OF INTERROGATIVES

In terms of syntactic phenomena this study is concerned with the formation of interrogatives. The ability to ask questions and inquire after information seems to be found in all languages although they employ different structures to do so. Those structures are commonly accounted for as the results of movement operations, namely WH-MOVEMENT since the early days of generative grammar (e.g. Chomsky 1977). The present chapter briefly presents the central notions and developments in the extensive body of research on the issue in order to provide a broad theoretical basis for the more specific analysis conducted in this thesis. Data collection for the present study aimed to uncover the behavior of LC with regard to the issues presented in this overview and results are discussed in Chapter 7. The generative analysis in Chapter 8 hones in on the more specific phenomenon of medial wh-constructions.

To assume movement operations underlying the formation of wh-clauses, such as questions and relative clauses, has been standard since Chomsky (1964, 1977). Wh-movement is a form of A'-MOVEMENT, i.e. movement of arguments or adjuncts into a position that allows either type of constituent, as opposed to A-positions which can only be filled by argument constituents. Both target phrases and maximal projections rather than heads (e.g. Radford 2004). The basic concept behind the notion of wh-movement is that a wh-expression, such as *who*, *what*, *why* but also *which student* etc., is based-generated in the position occupied by the respective non-interrogative constituent and then is moved, or COPIED and DELETED in more minimalist terms, through the derivation until it occupies a sentence-initial position within CP. This is illustrated in (1), (2) and (3).

(1) John will learn *Spanish*.

(2) John will learn *which language*?

(3) *Which language* will John learn?

The relation between the italicized items in in (1), (2) and (3) is hard to deny, especially due to so-called echo-questions, such as (2). This is even more obvious with morphologically case marked pronominal expressions, as in (4), as well as in more overtly case marking languages, such as German.

(4) Whom did she kiss?

The accusative marking on the wh-expression in (4) is generally taken to be assigned by the verb to its complement, the direct object *him/ whom*. Its presence on the sentence initial wh-expression indicates that accusative case was assigned inside the verbal domain before movement to the left periphery of the clause.

Many analyses have been proposed to account for this phenomenon. Since this thesis works within the generative framework, the following is limited to accounts proposed therein. Conceptually, the perception of movement operations within generative grammar changed significantly over time. Initially, movement was a major syntactic operation represented as the principle MOVE α (Chomsky 1993a) and considered an "imperfection" (Chomsky 1995, 317). Chomsky's view changes in more recent work, where movement is re-defined as a form of the operation MERGE, namely INTERNAL MERGE[9], which refers to the action of re-merging an item that is already included in the derivation. EXTERNAL MERGE, in contrast, refers to the merger of an item that is not yet included in the derivation, although it must be part of the NUMERATION in some recent minimalist accounts (Chomsky 2000). Both merger operations apply at the ROOT, the highest node, in accordance with the EXTENSION CONDITION (Chomsky 1993b, 1995). Moving a constituent is considered costlier than merging a 'new' constituent in terms of computational effort, since it involves the formation of a chain (see Chapter 3.1 below). Movement consequently occurs only in accordance with the economic principle of LAST RESORT to prevent the derivation from crashing. Since Chomsky (1993b), it can only be triggered by "morphological checking requirements" (Chomsky 1995, 257). The argument, usually referred to as MERGE OVER MOVE, is in line with the view that movement constitutes an imperfection and it fits well with the notion of numerations, as it can be argued that movement only takes place when no matching item is left in the numeration to be externally merged. On the other hand, it is considerably weakened with the redefinition of movement as a simple merge operation as well as the most recent view that the two options are an optimal solution to a "fundamental duality of semantics" (Chomsky in Cheng and Corver 2006, x). While movement in syntax is accepted, its precise theoretical and conceptual implementation has been controversial in recent approaches due to the reformulations occurring in the minimalist program and phase theory especially (Boeckx 2008).

Empirically, A'-movement has proven especially challenging, as it cannot be tied to agreement in a principled manner, which leaves it without a clearly identifiable morphological driving force, and also because is subject to considerable cross-linguistic variation. Wh-movement is the most common and most well-researched form. Different eras in the generative endeavor have yielded various approaches to wh-movement. They all face some of the same challenges. The major issues are:

[9] Please note that movement is regarded as a form of merge rather than a separate syntactic operation here. Nevertheless, the term *movement* is used for convenience and convention.

1. To provide an analysis that is powerful enough to be explanatory but flexible enough to allow for cross-linguistic variation.
2. To identify the positions involved in the movement, pertaining to notions of locality, cyclicity and the fine structure of the clausal left periphery.
3. To account for constraints on wh-movement, such as island effects which are introduced below.
4. To determine a trigger for movement, either as a property of the wh-word itself, a property of one or more of the involved positions or as the result of some other syntactic operation or general rule.
5. To explain when and why the resulting clauses are interpreted as interrogatives as well as how the same is achieved in polar questions.

3.1 Accounting for cross-linguistic variation

In Government and Binding Theory (GB), movement was assumed for any number of phenomena and motivated by Move α (Chomsky 1993a). The argument of uniformity (ibid.) was powerful and where movement operations could not be clearly identified, they were often postulated as LF movements to achieve an analysis similar to that of observable structures. For wh-constructions this mainly concerned the analysis of WH-IN-SITU structures, defined by Pesetsky (1982, 586) as "a wh-Phrase in an A-position". In English, wh-in-situ only occurs with echo questions (2) or multiple wh-questions (5), in which only one wh-constituent is fronted[10].

(5) Who said *what*?

Concerning so called wh-in-situ languages, Huang (1982b) argues for an English-like LF structure of questions in Chinese. Wh-expressions always remain in-situ in this language, which at first glance is difficult to unite with the movement analysis commonly used in generative accounts. Huang (1982b) bases his analysis on the scope properties that force an indirect question interpretation of (6) but a matrix question interpretation of (7). Since the two examples only differ in the choice of verb ('ask' vs. 'believe'), he concludes that the wh-expression must have scope over the subordinate clause in (6) and the matrix clause in (7).

(6) Zhangsan wen wo [shei mai-le shu]
ask me who bought books

'Zhangsan asked me who bought books.'

(7) Zhangsan xiangxin [shei mai-le shu]
believe who bought books

'Who does Zhangsan believe bought books?' (Huang 1982b, 371)

Consequently, he postulates two LF structures in which the wh-word is moved to the left periphery of the subordinate clause in (6) and to the left periphery of the matrix clause in

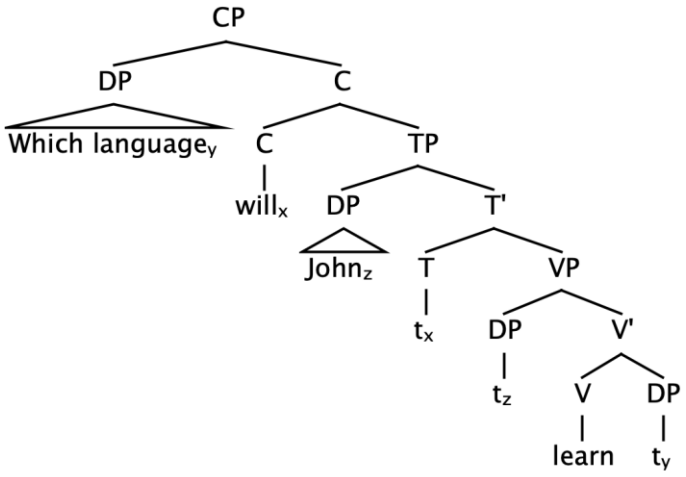
[10] Note that some languages front more than one wh-constituent with interesting implications for the analysis such as those outlined in Bošković (2007).

(7). This is a typical example for the logic applied in this early framework and it holds appeal as it allows a more or less uniform analysis for typologically different languages.

The structure of a wh-question can hence generally be depicted as in (8) (the object question from (3) serves as an example).

(8) Typical structure (of an object question) under the GB approach

[$_{\text{Spec CP}}$ Which language$_y$ [$_{\text{C}}$ will$_x$ [$_{\text{TP}}$ John [$_{\text{T}}$ t$_x$ [$_{\text{VP}}$ learn t$_y$]]]]]?



In the GB model, extracted constituents leave behind TRACES that occupy the vacated position, so it remains filled und unavailable to further operations. In later, more minimalist accounts (Chomsky 1993b, 1995) the notion of traces is replaced by that of COPIES of the moved item, which are left behind and usually deleted at PF resulting in pronunciation of the highest copy. In principle, any copy which is not deleted can be spelled out allowing for a unified account of the English and Chinese data above. Movement occurs in the exact same way with deletion of different copies causing pronunciation in different positions. Such covert movement within the COPY THEORY OF MOVEMENT was most explicitly developed by Pesetsky (2000). Both traces and copies form a CHAIN headed by the moved constituent, allowing for interpretation in the former position(s) and sharing the head's properties (e.g. binding). Reintges, LeSourd, and Chung (2006) operate within the frame of the Copy Theory and consider data from wh-agreement morphology in Coptic Egyptian and Passamaquoddy as further evidence of an underlying movement operation in wh-in-situ constructions. While the scope effects observed by Huang (1982b) might be accounted for under a different analysis, special morphological agreement shows that movement takes place in overt syntax. The only structural difference to languages with wh-movement then lies in the choice of copy in the A-bar chain that is pronounced.

The Copy Theory has been widely adopted since it can better account for scope effects as in (6) and (7) above, so-called reconstruction effects, as in (9) below where the interpretation of the anaphor depends on semantic reconstruction of *they* in a position c-commanding *each other*, and instances of wh-agreement, as in Irish briefly described below and discussed in more detail in Chapter 8. The chain links all positions through which the constituent moves, which allows the lower, silent copy of *they* in (9) to bind the anaphor.

(9) They are likely to appear to each other to be happy. (Lasnik 2006, 207)

3.2 Landing sites

In early accounts, before the idea of split CPs took root with Rizzi (1997), most movements involved just one trace per constituent, as in (8) above. Long-distance movements in which a constituent is moved out of its clause as in (11) and (12) are the exception and introduce the next central issue (point 2. above), namely the identification of the positions involved in wh-movement:

(10) $[_{CP}$ Who$_i$ $[_C$ did$_x$ $[_{TP}$ Mary t_x kiss $t_i]]]$?

(11) $[_{CP}$ Who$_i$ $[_C$ did Paul say $[_{CP}$ t_i $[_{TP}$ Mary kissed $t_i]]]]$?

(12) $[_{CP}$ Who$_i$ $[_C$ did Paul say $[_{CP}$ t_i $[_{TP}$ John thinks $[_{CP}$ t_i $[_{TP}$ Mary kissed $t_i]]]]]]$?

(13) Which man do you think was/*were kissing Mary at the party?

One might argue that the object of the verb *kiss*, *who*, is either base generated in its surface position or moved directly there in (11), (12) and (13). Base generation in the matrix clause is ruled out by the general analysis of wh-constructions outlined above as well as the agreement relation between the wh-expression and the auxiliary in (13). The auxiliary *(to) be* has to agree in person and number with the wh-expression *which man* for the sentence to be grammatical. The fronted wh-expression is not in a suitable position to enter into an agreement relation, which is established between heads and specifiers in Government and Binding Theory (Chomsky 1986a), as it is in another clause altogether. Hence, this indicates that the wh-expression must have occupied the subject position of the subordinate clause at one point of the derivation, where it could agree with the auxiliary. Please note that under phase-based accounts, it would also be feasible to assume an agreement relation that occurs independently of wh-movement as A-chains are separate from A-bar chains according to Chomsky (2008). Nevertheless, the wh-expression would have to be situated within the c-commanding domain of the auxiliary at

some point during the derivation to allow for probing and the resulting agreement relation and formation of the A-chain. Base-generation in the surface position would be ruled out under such an account as well.

Regardless of whether it passes through the subject position, the structure indicated in the examples (11) and (12) suggests that *who* lands in the intermediate CP projection(s) before moving on. This is in line with the assumption that wh-movement applies on a LOCAL level i.e. within one clause (Lasnik 2006). The concept is formalized in the MINIMIZE CHAIN LINKS CONDITION by Chomsky and Lasnik (1993) which requires chain links to be as short as possible, indicating that a moved constituent stops at each possible landing site along the way. Long distance movement as a consequence is SUCCESSIVE CYCLIC, occurring in fact as a series of short movements. The evidence for successive cyclic movement is reviewed in more detail in Chapter 8 but an overview is given here.

Evidence for the hypothesis can be found in various languages. German for example makes use of partial wh-movement (McDaniel 1989), in which the wh-object is only moved part-way into the intermediate CP and *was* ('what') serves as a scope-marker in the matrix clause and indicates an interrogative interpretation.

(14) *Was* glaubst du *wen* Mary geküsst hat?
What believe-2P.SG you who-ACC Mary kissed has?
'Who do you believe Mary kissed?"

According to McCloskey (2001) Irish selects a specific complementizer only for subordinate clauses out of which a wh-expression was extracted. His data contains this complementizer in two subordinate clauses, indicating successive cyclic movement through both CPs. Such cross-linguistic data forms the basis for a uniform analysis that is extended to languages in which the effects are less immediately observable, e.g. English. Some evidence for the underlying structure can be found in acquisition data from English, as children have been observed to repeat a wh-expression in the intermediate position (Roeper and de Villiers 2011) or strand part of the wh-phrase there.

(15) *Who* do you think *who's* in there, really really really? (Amber 4;6)
(Roeper and de Villiers 2011, 210)

(16) *What* do you want *else* to eat? (Boeckx 2008, 27)

The wh-scope marking construction as illustrated in (14) and in particular wh-copying, illustrated here in (15) with data from English first language acquisition are central to the discussion of the LC data (cf. Chapter 8).

Conceptually, there are different generative accounts of locality. The notion is closely related to that of ISLANDS, or BARRIERS in older terminology, PHASES in the most

recent formulation. While all of these have different theoretical implications, they are all concerned with movements across some kind of clausal or phrasal boundary and the ways in which these are restricted. These restrictions are discussed under Chapter 3.3. The notion of cyclic movement can be deduced from each account. The implication is maybe the clearest in the case of phases. The basic idea behind the notion of phases is to reduce computational effort in narrow syntax and dispense with LF and PF as separate components by assuming that the derivation is TRANSFERRED to the interfaces in pieces or phases (Chomsky 2008). Only material that escapes the transfer domain and is situated at the edge of a phase remains available for further operations according to the PHASE IMPENETRABILITY CONDITION (PIC) (Chomsky 2001, 2008). This includes wh-movement. Under the assumption that CPs are phases and the Minimize Chain Links condition, it follows naturally that wh-moved material would target the CP-edge until it reaches its ultimate landing site. Phase theory leaves many open questions though and for the purpose of this thesis it suffices to accept the underlying notion of locality. The technicalities remain unresolved in the theory's newest formulation which is why I follow Boeckx (2008, 48ff.) in rejecting phase theory in its current form despite intriguing theoretical premises.

Accepting the notion of successive cyclic movement on the basis of the cross-linguistic evidence, which this thesis confirms with data from LC, one must ask next which position precisely is targeted in each step. The GB account in (8) presents a simplified sentence structure that has been challenged on solid empirical grounds, introducing more complex internal structures for each of the three major layers, namely VP, IP and CP (Kayne 1984, Pollock 1989, Rizzi 1997 among many others, for a summary see Cinque and Rizzi 2008). The most relevant for wh-movement clearly is the internal structure of CP as re-defined by Rizzi (1997) in his SPLIT CP HYPOTHESIS. He shows that the CP layer can contain several functional projections instead of one unique C-projection to account for the distribution of topicalization, cf. (17) and (18), and focalization, cf. (19) and (20).

(17) *That comment*, you should not accept from a student.

(18) Class must be an environment [<u>in which</u>, *that comment*, you should not accept].

(19) *His wife*, he never would have addressed in such a way.

(20) I believe [<u>that</u> *his wife*, he never would have addressed in such a way].

As illustrated by (18) and (20), the topicalized constituent and the focused constituent can each be preceded by another element, e.g. a relative pronoun and an

overt complementizer. If there was a single CP projection, it would be headed by the complementizer, such as *that* in (20), and possibly preceded by an element in the specifier of CP. Such an analysis is clearly incompatible with the data in (17) - (20). Rizzi (1997) concludes that separate projections for topicalized and focused material are needed, namely FocP and TopP. He further assumes a top-most projection ForceP which determines sentence force and holds the complementizer which identifies a clause as e.g. declarative or interrogative.[11] Data from Italian prompts him to include another projection, FinP directly above IP. The preposition/particle *di* introduces non-finite control clauses. While, at first glance, it appears to be the non-finite counterpart of the finite complementizer *che* (cf. (21) and (22)) it is ordered differently with respect to dislocated constituents, as illustrated in (23) - (26).

(21) Credo *che* loro apprezzerebbero molto il tuo libro.

(22) Credo *di* apprezzare molto il tuo libro.

(23) Credo *che* il tuo libro, loro lo apprezzerebbero molto.

(24) *Credo, il tuo libro, *che* loro lo apprezzerebbero molto.

(25) *Credo *di* il tuo libro, apprezzare molto.

(26) Credo, il tuo libro, *di* apprezzare molto.

(examples taken from Rizzi 1997, 288)

While the topicalized expression *il tuo libro* must follow the finite complementizer *che*, it must precede the non-finite *di* for the sentence to remain grammatical. It follows that *che* and *di* occupy different positions within the CP. According to Rizzi (1997), *che* is a complementizer heading ForceP, while *di* marks the clause as non-finite and heads FinP. Topic and Focus projections are situated in between ForceP and FinP. The clausal left periphery is hence structured as follows:

(27) Fine structure of the clausal left periphery after Rizzi (1997, 297)

[ForceP [Force^0 [TopP [Top^0 [FocP [Foc^0 [TopP [Top^0 [FinP [Fin^0 [IP]]]]]]]]]]]

Cinque's cartographic approach to sentence structure takes this line of thinking a step further and assumes a distinct position for each possible class of items. Split projections and the cartographic approach are well-suited to describe discourse effects and language specific intricacies such as the Italian patterns illustrated above but their universal application to all languages, which is a theoretical premise in the view of Cinque and Rizzi (2008), can be regarded critically. For languages in which the observed

[11] This has interesting implications for the position occupied by wh-expressions which are tied to force marking by Cheng (1997). This is discussed in Chapter 3.5

effects are absent, the consequence is a large amount of stipulated empty projections which does not agree well with the minimalist endeavor. LC is a case in point as nothing in the data from the present study necessitates the assumption of a split CP. This thesis consequently abides by the more minimal notion of the CP layer although extensions on cross-linguistic grounds are possible.

3.3 Constraints on wh-movement

Despite the fact that wh-movement can cross clause boundaries as illustrated in Chapter 3.2 above, Ross (1967) first observed that it is not generally unbounded. Instead, it appears to be constrained as there are several types of constituents and clauses that do not allow for the extraction of wh-elements from inside of them depending on their different functions and positions. Some of these restrictions appear to be language specific and it is one objective of the present study to determine if and to which extent they are in effect for LC.

In his seminal work, Ross (1967) identifies several types of constituents out of which a wh-element cannot be extracted. He terms those constituents ISLANDS, which are governed by ISLAND CONSTRAINTS. Among them are NP complement islands, as illustrated by the contrast between (28) and (29), relative islands as in (30) and (31) and sentential subject islands as in (32) and (33).

(28) Who$_i$ did Paul claim he kissed t_i?

(29) *Who did Paul make the claim that he kissed?

The empirical observation is that fronting of a wh-element from inside a complex NP is prohibited. This is known as the COMPLEX NP CONSTRAINT (CNPC). Ross (1967) includes relative clauses such as (30) and (31) under this constraint though later analyses tend to separate them out based on a distinction made between complements and adjuncts. It remains equally ungrammatical to extract a wh-constituent from inside a relative clause.

(30) Paul knows the man who kissed Mary.

(31) *Which girl$_i$ does Paul know the man who kissed t_i?

Additionally, he identifies the SENTENTIAL SUBJECT CONSTRAINT, which forbids wh-movement from within a clause in subject position.

(32) That he kissed Mary surprised nobody but Paul.

(33) *Who$_i$ did that he kissed t_i surprise nobody but Paul?

Though not included in Ross (1967), it also proves ungrammatical to front wh-elements from within an adjunct clause of various kinds as observed by Huang (1982a).

(34) Paul is/ was/ would be jealous because/ when/ if Peter kissed Mary.

(35) *Who$_i$ is Paul jealous because Peter kissed t_i?

(36) *Who$_i$ was Paul jealous when Peter kissed t_i?

(37) *Who$_i$ would Paul be jealous if Peter kissed t_i?

Chomsky (1973) recognized WH-ISLANDS as an additional type of island phenomenon, which later became influential in the formulation of general principles and locality requirements on movement operations, such as the minimal link condition or the attract closest principle.

(38) How$_i$ does Paul think Peter won over Mary t_i.

(39) Paul wonders who$_j$ Peter won over t_j that way.

(40) *How does Paul wonder which girl$_j$ Peter won over t_j t_i?

(41) ??Which girl$_j$ does Paul wonder how$_i$ Peter won over t_j t_i?

Generally, an asymmetry is perceived between (40) and (41) for example by Huang (1982a), Lasnik (2001). Neither is well-formed, but clauses in which the adjunct, *how* in the above examples, is fronted over the argument, *which girl* in the above examples, seem to be completely unacceptable while the reverse order, as in (41), is less severely degraded.

In Government and Binding Theory island effects were generally accounted for by SUBJACENCY (Chomsky 1973), which is an early formulation of what was later stated as BARRIERS and most recently as PHASES, ultimately constituting a form of locality and/or cyclic movements. Islands were accounted for in terms of BOUNDING NODES, typically taken to be NP and TP. Movements that crossed more than one bounding node were considered ungrammatical which accounted for most of the island effects outlined above. Rizzi (1982) showed that the identity of bounding nodes is language specific, identifying CPs rather than TPs as the relevant category for Italian. Subjacency did not suffice to account for the asymmetry between (40) and (41) as it is violated in both cases and hence cannot provide an explanation for their difference in acceptability. With Chomsky (1981) and Huang (1982a) the Empty Category Principle served as an additional constraint, which required traces of A-bar movements, including wh-movement, to be locally governed (i.e. roughly, c-commanded by a lexical head or antecedent). (40) violates the ECP in addition to violating the Subjacency constraint, which makes it even less acceptable than (41) which only violates Subjacency (Lasnik 2001).

According to (featural) relativized minimality as developed by Rizzi (1990 and subsequent works) island effects are in fact intervention effects caused by movement over a constituent with a similar characteristics which in current theories means a

similar featural configuration. By his assessment it is not the nature of some boundary they pass as such that results in grammatical violation but the interplay between the moved constituent and a constituent crossed in the movement. He claims that a constituent which is similar to the moved item can intervene in the formation of the movement chain preventing the system from co-indexing the lower and higher copies. The larger the similarities, the graver the violation which is how markedness asymmetries are explained in this model. Originally, this was expressed in terms of Relativized Minimality, Government and the Empty Category Principle with the result that "antecedent governing cannot take place inside the domain of a potential antecedent governor" (Rizzi 1990, 5). This is formally expressed in RELATIVIZED MINIMALITY (Rizzi 1990, 7),

(42) X α-governs Y only if there is no Z such that
  i) Z is a typical potential α-governor for Y,
  ii) Z c-commands Y and does not c-command X.

Minimalist approaches have since restated notions of government in a feature valuation framework and the adapted form of relativized minimality refers to the intervention of items with the same features, i.e. [wh]-features which prevent the A'-chain from forming properly in the case of wh-islands.

3.4 Movement triggers

Concerning point 4, the motivation for movement, the general trend in Government and Binding Theory was to postulate a rule, such as "move wh-phrase into COMP" (Chomsky 1977, 85), that required the wh-phrase to occupy a position with the CP. This early account was not feature driven in the minimalist sense but it already assumed a [wh]-feature on the wh-word itself that would make the item a suitable candidate for attraction into the CP projection in accordance with the rule. In a similar vein, a functional head X^0, such as the complementizer C^0 or COMP could contain a feature [+wh], which would cause a matching [wh]-constituent to be moved into the CP-projection in later, feature driven approaches. The complementizer receives a phonetically empty form in those instances to capture the distribution of complementizers like *that* or *if*, which cannot co-occur with a fronted wh-word or phrase as is illustrated in example (43) and (44).

(43) I wonder if Paul is meeting the president.

(44) I wonder who/ which man *if Paul is meeting.

(45) Who is Paul meeting?

Despite this distribution, wh-elements are not analyzed as occupying the head C position, or any other head position, since they can be phrasal, but rather the associated specifier. Instead the null-C analysis is widely adopted for subordinate clauses. In English main clauses, the auxiliary is raised into the head C position, as was assumed above (Radford 2004).

Features serve as the main motivators for syntactic operations in the Minimalist Program. Under the latest Minimalist approach (Chomsky 2001 and subsequent publications), they are heavily formalized and can be either VALUED or UNVALUED[12]. Agreement occurs between a PROBE, containing unvalued features, e.g. [u-number] and [u-person] on finite verbs and auxiliaries, and a GOAL, containing matching valued features, e.g. [Sg-number] and [3rd-person] on a subject (pro-)noun like *John*/ *he*. The probe, as the higher element in the structure searches its c-commanding domain for a matching goal. Through FEATURE VALUATION, a process that roughly corresponds to what was formerly referred to as CHECKING (Chomsky 1995, 228f), the goal assigns matching values to the probe's formerly unvalued features and vice versa, hence creating agreement, for example between subjects and verbs or adjectives and nouns.

Chomsky (2001) regards Agree as a component of Move, which implies that all movement involves feature checking in the spirit of earlier work as mentioned above. This intertwines the two operations in a way which is theoretically elegant for head- and A-movement. It has been adopted in much generative work because it offers a common trigger, namely the need to value unvalued features, for two major syntactic phenomena (cf. e.g. Epstein and Seely 2002, Müller 2004). This homogenous picture faces several empirical challenges and is abandoned to a certain extent in Chomsky (2008). Among the challenging data are instances of LONG-DISTANCE AGREEMENT, such as in Icelandic, where a matrix clause verb can agree for number with the embedded nominative subject and assign case to the dative experiencer which is raised into the matrix TP specifier at the same time. Another issue is MULTIPLE AGREEMENT in which one probe has features valued by several distinct goals as illustrated in (46), taken from Chomsky (2001, 16), where the expletive subject checks T^0's [person] and [EPP] feature but not the number feature which is instead valued through a second agreement relation with the object.

(46) There is likely to arrive a man.

[12] Please note that I choose to focus on feature valuation rather than their interpretability. The usual distinction between un-/ interpretable features presupposes a notion of semantic interpretability and/or some form of look ahead to LF that appears unnecessary for a formalized account. Their corresponding valuation status on the other hand is more tangible and conceptually useful.

In general, agreement, hence the valuation of features, which occurs without (overt) movement and movement which occurs but is not clearly driven by the need to check morphological features are difficult to account for under the approach taken in Chomsky (2001). Since the issue is complex but rather marginal to the context of this thesis, the reader is referred to Grewendorf and Kremers (2009) for a discussion of the Icelandic data and possible analyses, a more general discussion of the relation between movement and agreement can be found in Bobaljik and Wurmbrand (2003, 2005).

For wh-movement the introduction of probe-goal relations calls the role of the [Wh]-feature as the trigger for movement - as assumed in earlier works - into question. The [Wh]-feature on C requires a wh-expression, but it is not necessarily the driving force behind the movement operation, especially as it is not expressed morphologically on probe or goal in English. Instead, wh-movement is triggered by the additional presence of an [EPP]-feature on C that makes an overt specifier obligatory, in analogy to the [EPP] on T, namely the EXTENDED PROJECTION PRINCIPLE which states that each T head requires a subject in spec-TP position (Chomsky 1982, 10). This remains a standard solution, which is not without benefits as it provides a uniform structure of TP and CP. Nevertheless, [EPP] of interrogative C is largely stipulated to provide a feature-driven account for Wh-movement and is based on little independent empirical or conceptual grounds. It is conceptually very different from other features, such as phi-features which are present on both probe and goal and valued through agreement. The [EPP]-feature is a remnant from GB approaches in which such relations were less central and its theoretical adequacy is doubtful (Bošković 2002, 2007).

In most recent version of the theory, namely in PHASE-BASED accounts, the EDGE FEATURE is more commonly found as a movement trigger. Though it, to a lesser extent, faces the same problems as the [EPP] solution with regard to having a special status in terms of feature valuation, it is conceptually better motivated in the context of phase theory. As described above, pieces of a syntactic derivation are spelled out by phase and become inaccessible to syntax after this in accordance with the PIC. Transfer of the complete structure that exists at this point would result in either a finished clause, in which case nothing would be gained in terms of computational load, or a crash of the derivation as it would be unfinished when pronounced. To avoid this, Chomsky (2000) proposes an escape hatch for items which are required for the remaining derivation. This so-called EDGE consists of the (phase-) head and its specifier and is not transferred alongside any arguments or adjuncts left inside the DOMAIN. Items that are needed in the course of the derivation move to the edge to escape Transfer. This movement is made

possible by the edge feature [EF] according to Chomsky (2008), a feature that is present on all constituents and allows for merger in the most general sense. This feature becomes the main trigger for A'-movement which then occurs independently of agreement. A-movement on the other hand is still driven by the presence of unvalued features and a generalized form of the [EPP]. In slightly earlier accounts, Chomsky assumes that the presence of unvalued features is necessary for an item to be ACTIVE and hence available for any syntactic operation, be it Move or Agree. In this view, a [Wh]-feature on the wh-expression would remain unvalued and be checked in the final position where the interpretable [Q]-feature checks the Q-operator on C°, as described above. Recall that this PIC-account is not primarily feature driven, but establishes the inaccessibility of the phasal domain from which the wh-expression optionally 'flees' via the edge. This account needs a generalized [EPP] feature to make this initial movement possible. The triggering issue remains largely unresolved though. The [EF] in its most recent form applies regardless of unvalued features.

Please note that phase theory is included here because it has received much attention in recent generative research and cannot be ignored in an overview of the current theoretical framework. However, I only refer to it in the analysis in a very indirect sense, namely in discussing locality, the copy theory of movement and cyclicity. Although these notions can also be derived from phase theory in principle, the details of the account face many challenges, conceptually as well as empirically. No version of the theory can currently provide a coherent analysis that does not run into intrinsic paradoxes or make countless stipulations. (cf. e.g. Boeckx 2008, Boeckx and Grohmann 2007, Klein 2012) For this reason, the analysis is not framed in a phasal context although the suggested account can in principle be integrated in such an approach.

3.5 Clause typing

That clauses must be typed is a long-standing assumption since Chomsky and Lasnik (1977, 445) who state that "each clause must be identified as declarative [...] or interrogative (a direct or indirect question)". Sentence force can be marked in various ways but is universally present. One way to identify a clause as an interrogative is the presence of an interrogative particle or marker, which is situated in C position. According to Cheng (1997) such a marker can be identified for yes-no questions in all wh-in-situ languages, where it serves to type the question as such. In some wh-questions the related wh-marker can be zero. A clause can hence be typed as a wh-question either by the presence of an overt or covert wh-particle or by overt wh-movement, as in English

(Cheng 1997). It is assumed that the presence of a wh-element in the specifier of CP marks a clause as interrogative. This generalization is known as the CLAUSAL TYPING HYPOTHESIS,

> Every clause needs to be typed. In the case of typing a wh-question, either a wh-particle in C^0 is used or else fronting of a wh-word to the Spec of C^0 is used, thereby typing a clause through C^0 by Spec-head agreement. (Cheng 1997, 29)

In more minimalist terms, the idea is that the wh-constituent checks the [wh]-feature on C^0, hence marking the C head as [+wh]. One issue with this under a probe-goal approach (outlined in Chapter 3.4) is that it requires the [wh]-feature on the C head to be unvalued, or the clause would be typed as a wh-question regardless of checking. That would mean that the [wh]-feature on the wh-expression is valued, excluding it as a motivator for the movement operation and requiring another unvalued feature to render the constituent active, the mechanics of which are discussed in Chapter 3.4. Chomsky (2000) assumes an operator feature [Q] on both C^0 and the wh-element. This feature is valued on the wh-expression and unvalued on C^0 and it is responsible for clause typing. C^0 hence contains [Q-u, Wh] until the wh-expression checks the [Q] feature while it values its [wh]-feature at the same time. This may be a possible solution though the independent motivation for both features is unclear and the issue of movement triggers remains as is discussed above. Setting the issue aside, wh-moved elements would be situated in the specifier of ForceP under a split CP analysis, where they type the wh-question.

MULTIPLE WH-FRONTING LANGUAGES, such as Serbo-Croatian, Russian or Bulgarian, present another challenge. While English only fronts one wh-element in constructions like (11) above and Chinese or Japanese as wh-in-situ languages spell out all wh-elements in the base position, these languages front all wh-phrases, resulting in structures like (47).

(47) Ko koga voli? (Bošković 2002, 352)
Who whom loves?

Even under a non-feature-based account, it cannot be argued that both expressions front for reasons of clause typing. In fact, such an analysis would be incompatible with the claim in Cheng (1997) that the use of a wh-particle and wh-movement are mutually exclusive. She argues that instead just one wh-element is moved to mark sentence force while any additionally fronted element is moved due to a licensing requirement. She claims that wh-words in these languages are not inherently interrogative and hence need to be licensed by a [+wh]-element in C^0 position. Such licensing processes are mostly dispensed with in minimalist accounts. Instead, it is often argued that the different wh-expressions target different positions, either within an

extended internal CP structure in the spirit of Rizzi (1997), or as adjuncts to IP (Rudin 1988).

The empirical picture is even more complex, since languages which front more than one wh-element do not behave in a uniform way. Superiority effects, based on the SUPERIORITY CONDITION (CHOMSKY 1973, 246), can be observed under different circumstances in different languages, such as in Bulgarian questions with two fronted wh-elements and Serbo-Croation long-distance questions (but not regular short-distance wh-questions). To account for the relative order of wh-elements in multiple fronting languages is a challenge to syntactic theory, as it has to consider their relative base positions and the order of their movements in accordance with principles such as ATTRACT CLOSEST, each of which must be independently motivated to yield the correct surface order under the correct conditions. Several proposals have been made, see e.g. Bošković (1997, 2002) and the contributions in Boeckx and Grohmann (2003). Mutiple wh-fronting is not grammatical in LC and hence the topic is marginal to the discussion here. It does, however, raise questions about the precise position targeted by wh-movement in a cartographic approach and it provides a possible link between wh-movement and more discourse driven operations, such as topicalization and focalization. Bošković (2002) assumes that all but the first wh-element in Bulgarian move because they are inherently focused, hence occupying SpecFocP in a split CP analysis. In a similar spirit, the specifier of FocP has been proposed as the de facto locus of wh-phrases in matrix question.

Rizzi (2001) suggests two separate heads namely Force^0, which can be null and is declarative when overt in Italian, and Interrogative^0, which hosts the interrogative complementizer *se* in embedded sentences. A focused constituent cannot co-occur with wh-elements in Italian matrix clauses, which is why Rizzi (1997) argues that these move into SpecFocP and hence compete for this position with focused material. Note that this already implies that wh-elements are not primarily responsible for clause typing, as they occupy the focus position below ForceP. At the same time, it ties wh-movement to focalization as mentioned above. In his paper from 2001, the order of *se* and wh-phrases with respect to focused constituents leads Rizzi to conclude that the interrogative complementizer is situated below its declarative counterpart but above wh-elements in matrix clauses, which in turn are positioned above wh-elements in embedded clauses. This implies that wh-elements occupy different positions in different clause types arguing against a unified account and common trigger. The newly coined projection occupied by *se* is termed IntP. In his analysis a few wh-adverbs are exceptional in that they can be

base-generated in SpecIntP which is in line with their distribution. These special wh-elements remain involved in interrogative clause typing.

Taking such reasoning one step further, Aboh and Pfau (2011) argue that wh-words are not involved in clause typing at all but instead move for independent reasons. Sharing the observation by Cheng (1997) of overt and covert (=null) question particles, they conclude that these are situated in InterP, which goes back to Aboh (2004) and corresponds to the IntP from Rizzi (2001). Their split CP structure is similar to Rizzi's.

(48) ForceP>InterP>TopicP>FocusP>FinitenessP (Aboh and Pfau 2011, 96)

The position of wh-elements in the specifier of FocP hence provides a distinct projection in which interrogative force is encoded: either via yes/no question particles in InterP for which the authors provide cross-linguistic evidence from Gungbe and Nweh, or via the presence of a wh-element in SpecFocP in wh-questions. They note that Foc^0 is "not inherently interrogative" (Aboh and Pfau 2011, 94) as it can attract a number of categories. Taking this into account as well as rejecting the notion that interrogative force could be located in two different positions, they propose a unified account that leaves interrogative clause typing solely to InterP for polar as well as wh-questions, in embedded as well as matrix clauses. Wh-elements then front for reasons fully independent of clause typing which they relate to their DP internal structure, namely the presence or absence of [Q] and focus projections/ features that make them targets for attraction into the CP layer.

For the purpose of this thesis, the discussion is merely something to be aware of, since a more minimal version of the CP is assumed that does not contain several functional projections. This choice is made simply because it is not necessary for the present account to postulate a complex CP-internal structure. This does not indicate that such a structure cannot underlyingly be present in LC and evidence may arise from research on other syntactic configurations than the ones considered in this thesis. Concerning clause typing, I follow the Chomskyan approach to assume a [Q] operator feature which must be checked to construct a question but which is not considered the cause of movement in the present account. It is argued that this feature is underspecified and can be set to the values [wh] or [y/n] or, cross-linguistically at least, [focus]. The trigger for wh-movement in the chosen view is the unvalued [wh]-feature on the wh-phrase that causes the constituent to move up the derivation until it finds a suitable goal in the interrogative C head.

4 LOUISIANA CREOLE – HISTORY, SOCIOLINGUISTIC SETTING AND LINGUISTIC PROFILE

4.1 History

This chapter provides some historical background limited to issues of relevance for the development of the Creole language and culture in Louisiana. As described in Hall (1992), Louisiana's history is complex and less well known than that of most of the states that make up the USA. In large parts due to the landscape and climate, colonization of Louisiana was difficult and progressed slowly from the beginning. The social structure remained distinct from most colonies for the whole colonial period, resulting in circumstances that differed from what was typically found in other colonies where a highly stratified society was the norm. "In Colonial Louisiana, there was more chaos than order" (Hall 1992, xiii), encouraging an unusual amount of contact between the different factions.

Settlement in the territory that became Louisiana began long before colonial powers arrived. Evidence of early inhabitation dates back to 3500 BCE. These are considered the ancestors of the indigenous groups that were native to the area when European influence began, the largest of which were the Natchez and the Choctaw. The first Spanish explorers visited in the 16th century, later France approached the Mississippi Delta area from their Canadian colony (Acadia). René Robert Cavelier Sieur de La Salle claimed the area for France in 1682, naming it LA LOUISIANE after the French king Louis XIV, but actual settlements only started in 1699 when France sent Pierre le Moyne d'Iberville, who build the first fort in Biloxi. This fort was supposed to secure the French rule over a vast region, effectively blocking in the British settlements to the East. Despite these ambitious intentions, in 1706, only a total of 85 French or Canadian settlers were recorded in a census (Hall 1992, Klingler 2003a). The very low number of French people settling under the harsh conditions Louisiana's landscapes had to offer lead to unique socio-historical circumstances. The situation did not allow for a typical colonial plantation society as was developing elsewhere at the time (Marshall 1997). Some consequences this had for the social, cultural and linguistic development in Louisiana are outlined below.

France privatized the task of populating and managing the difficult colony in 1712. In 1718, Bienville, who was managing Louisiana on behalf of the Company of the West Indies and acting as governor at the time, established New Orleans. The Company

was supposed to import 3,000 slaves and 6,000 colonists (Rottet 2001). The latter goal effectively turned Louisiana into a penal colony, as volunteers were rare in France. Since most of the deported were criminals and came from urban areas (Rottet 2001), very few of them had the agricultural knowledge or discipline necessary for survival in their new surroundings. As a consequence mortality rates were high not only during the voyage but also in Louisiana itself (Hall 1992). As Klingler (2003a) points out, the settlers often depended on Native Americans, who knew the land, for survival. The Company of the West Indies brought in farmers from Germany and Switzerland to help settle and cultivate the land. This was the first bigger group of immigrants who were able to actually harvest food to sustain the local population. The Mississippi banks on which they settled are today known as the German Coast. Their numbers were not very significant, with around 330 people in 1720, and they were "fully assimilated by the French population" (Klingler 2003a, 5) meaning that their cultural and linguistic impact was rather low. Ultimately, the Company failed in their endeavor of populating Louisiana and turning it into a flourishing colony. They returned the land to France in 1731. There is no census available from that time but data from 1726 leads Hall (1992) to conclude that there were less then 2000 Europeans of all stations living in the colony, with their number drastically declining over the next years. The Natchez Revolt of 1729 alone reduced the number of white settlers by roughly 10% according to Hall (1992).

As a consequence, the black and Indian population outnumbered the colonizers quickly. In 1706 Iberville lead a raid on the British and captured 3,178 African slaves. This remained an isolated incident until 1719 when the first slave boats arrived and actual large-scale slave import from West Africa began under the rule of the Company of the Indies. There was a brief but intense period of slave trade, with France rather than the French Caribbean, until the reestablishment of French rule in 1731 (Klingler 2003a). Marshall (1997) and Hall (1992) agree that the number of slaves roughly equaled that of the free population in 1727 and grew quickly after that. Just five years later, in 1732, there were more than twice as many slaves than free settlers. Nevertheless, labor shortage remained a large problem for colonial Louisiana. The plantations along the Mississippi, e.g. the German Coast were growing and starting to move beyond simply sustaining the local population (Klingler 2003a). Most slaves worked here, where plantations nevertheless remained small compared to other colonies. Other enslaved people worked in New Orleans. Besides the slave owning urban settlers and the plantations by the banks of the Mississippi, many settlers were poor farmers working the land to sustain themselves. It follows that the distribution of work forces was unequal in

geographic as well as social terms preventing a typical colonial society from fully forming. (Marshall 1997). As a result of the low numbers and the distribution of people across the land, there was intense contact between whites and blacks – enslaved and free[13], French and Africans, colonists and indigenous people. Hall (1992) clearly shows how deeply afraid the French were that the different native nations and black slaves might unite against the colonizers. They tried to separate the two groups as much as possible to the point where it was forbidden to own Indian and African slaves at the same time. They also enticed the Indian settlements to catch runaway slaves in order to avoid understandings between the two. These efforts were successful to a degree, nevertheless numerous black, Indian and black-Indian conspiracies were reported or suspected during the 1730s (Hall 1992). Due to the low number of settlers, the harsh conditions and the role played by Native Americans as well as free Blacks and African slaves, Louisiana had a disposition towards the mixing of people and cultures. Even enslaved people had some contact with people from other estates with resulting marriages and festivities as well as trade arrangements between whites, blacks and Indians alike (Marshall 1997). The situation naturally caused interracial contact to a degree that was unusual in colonial societies.

During the second half of the 18th and the subsequent 19th century the population became even more diverse. To put an end to the so-called SEVEN-YEAR WAR between France and Britain, all French territories east of the Mississippi, including parts of Louisiana, were allotted to Britain in the treaty of Paris 1763. As a secret side-agreement, the rest of Louisiana fell under Spanish rule, with the exception of New Orleans, which remained under French control for a limited period of time. The change in ownership was formally announced in 1764 and would last until 1800. The treaty included the French territories in Canada, which were now under British rule. Many French Canadian settlers refused to swear allegiance to the British crown and hence denounce their Catholic religion (Rottet 2001). British military campaigns had been frequent in French Canada during the war and the British began the expulsion of French Canadians in 1755. The process became known as LE GRAND DÉRANGEMENT or THE GREAT UPHEAVAL. The British deported French settlers from Novia Scotia on a large scale, first to other British colonies in the Americas, later even back to France (Neumann 1985). Most of the Acadians did not stay in the places to which they were re-settled, often because they were not very welcome there or

[13] Note that Louisiana actually had a number of free Blacks, some of whom were community members and some of whom were living within Indian settlements. According to Marshall (1997) free black men served as a police force for slaves until they were involved in a revolt in 1795. During the late 18th century Maroon communities of creole slaves formed in direct proximity to the plantations, some stayed even within the grounds. The runaways adapted to life in the swamp and kept in contact with slaves on the plantations, in fact forming another group that, though not technically free, did not live in slavery (Hall 1992).

had little in common with the non-French speaking, non-Roman-Catholic population in the British settlements. The deported Acadians slowly made their way to different places they felt were more suited for them (Holm 1989). Although the British never sent any French Canadians to Louisiana, they started arriving there in numbers by the mid-1760s. According to Valdman and Rottet (2010) approximately 3,000 people from Acadia migrated to Louisiana. The Spanish rulers at the time relocated the newcomers along the Mississippi, regardless of personal or economic consequences.

During the Spanish period, Louisiana's population grew rapidly. The Acadians were the largest group of free whites to arrive, though immigration directly from Europe took place as well (Rottet 2001). After the Acadians settled, the numbers of free settlers and slaves were roughly equal again (Marshall 1997). The slave population increased drastically as well. Hall (1992, 279) states that "it is obvious that large numbers of slaves were imported into Spanish Louisiana […] It is impossible to determine through studying slave-trade documents, where these slaves came from and in what proportions." It is hence unclear which cultural and linguistic influences were present beyond the French and Acadian ones, though it seems safe to say that Louisiana was quite diverse at this point, though routed in French based language and culture. By 1803 Louisiana counted ca. 50,000 inhabitants (Rottet 2001), as compared to 5,562 Europeans and 5,499 Africans slaves that are reported for the end of the French era in 1763[14] (Kilman cited in Neumann 1985). Marshall (1997) argues that most of the enslaved population originated in West Africa and had a West African language as their L1.

After the British-American War of Independence, the Mississippi River constituted the Western US border. The Spanish returned Louisiana to Napoleon in 1800 though they remained the governing party until 1803 when Louisiana became part of the United States. This step facilitated immigration to Louisiana and further diversified the population, as Americans from other states as well as Europeans made their homes there. French dominance in the region began to fade in the 19th century. The French and their descendants were no longer in the majority by 1820 (Rottet 2001). In 1809/1810, a large group of refugees from St. Domingue came to the area. These were roughly 10,000 people who were already fully socialized in a French-based Creole language and culture. The newcomers were of mixed racial and social status, whites and blacks, in part free, in part slaves. Their arrival indicated the high time of the plantation system in the area (Marshall 1997, Rottet 2001). This has raised the question whether the Creole language in

[14] Please keep in mind that different surveys covered different parts of what is Louisiana today. The numbers from different sources are comparable only to some degree. They can offer a rough orientation but no definite and ultimate count. For example, Hall (1992) lists 4,598 slaves and 3,654 free for the year 1763.

Louisiana developed regionally or due to the presence of Haitian Creole (HC) (Valdman 1997). The genesis question is briefly discussed below. Here it suffices to say that most scholars do not consider Louisiana Creole a result of HC being spoken in Louisiana.

With the inclusion of Louisiana in the United States a shift in language, culture and identity began. The Civil War constituted a turning point after which Americanization began to outweigh the state's French heritage, linguistically as well as culturally (Marshall 1997). The Acadians were influential in colonial Louisiana and generally held in high esteem but lost their social standing in this new era (Rottet 2001). In the late 19th century, a hundred years after their arrival, their descendants no longer self-identified as Acadians. Instead a local identity had formed under the abbreviated label CAJUN, from *acadien*, which had a much lower economic and social standing and was associated with "rural, uneducated populations" (Rottet 2001, 57). With the Americanization of the society and matching policies came a period of devaluation of all forms of French in Louisiana. As associations with the term *Cajun*[15] tended to be negative to a degree, this naturally implied negative perceptions of their language in the general public, which was furthered by US language policies. The English-only-policy, which was implemented after the civil war, effectively banned all forms of French from schools, infrastructure and institutions. Consequently, French vanished from the schools during the 1920s to 1950s because children were punished for speaking anything but English, as well as from church (Dubois, Leumas, and Richardson 2018). The former main language of the state was "almost fully absent from politics by the early 20th century" (Rottet 2001, 59) and from public life in general. This was in part due to the policies banishing it from official domains and in part due to the stigmatization following from them and the low social standing of what had become the Cajun minority. Rottet (2001) moreover observes a certain isolation of the French speaking population in the 20th century that contributed to turning all forms of French in Louisiana into oral languages of the homes. LC, as the language associated with slavery and common among uneducated poor citizens, fared even worse. Although no distinctions were made officially, LC was and still is the least prestigious language spoken in Louisiana and its use would have been even less socially acceptable than French. During the Second World War, soldiers' knowledge of French proved useful and with the industrialization of agriculture Cajuns began to work in larger companies. Both factors served to reduce resentments and isolation, but life in Louisiana

[15] Please note that cultural, ethnic and linguistic labels are interwoven in Louisiana. Most francophone whites will identify as Cajuns, regardless of heritage or the language they speak, while blacks are often considered Creoles. Nevertheless, the linguistic distinction does not follow racial lines and not all francophone Louisianans are actually descendants of Acadian settlers. This is further discussed in Chapter 4.3

was irrevocably on the path of Americanization. As francophone citizens became less isolated, they further assimilated to mainstream society and language rather than promoting the use of French. The growing importance of radio and TV was another factor in establishing English as the language of everyday life (Rottet 2001).

4.2 The genesis of Louisiana Creole

When considering the historical development and multifaceted contact among peoples described above, the means of communications during colonial times arises as a major issue. Clearly, some form of French first became the main language used in everyday communication. The colonizers spoke varieties of French, which were far from standardized even in France at the time (Valdman 1998). Especially with the addition of deported criminals, different dialects and sociolects must have been present. The imported slaves, on the other hand, were native speakers of West African languages and exposed to nautical vernacular French on the ships (Marshall 1997). Once in Louisiana, they would be faced with a large amount of variation among the French speech they encountered as well as several other languages, such as the mother tongues of fellow slaves and the native population. With the arrival of the Acadians another form of French entered into the mixture. Due to the socio-historical circumstances described above, the different groups remained in close contact. Consequently, the situation naturally gave rise to a speech continuum (Valdman 1998) and a large amount of variation, regionally as well as socially.

It follows that the genesis of Louisiana Creole is hard to pinpoint. As Valdman (1997) points out, the slow development of the early French settlements argues against a radical Creole genesis. In fact, Webb (2013) uses LC as the prototypical example of gradual Creole genesis in which a Creole language developed from a French-based Pidgin. While this seems to be the common assumption, very little is known about this Pidgin. Marshall (1997) claims that LC developed in the early 18th century. This is based on several records of the presence of a Creole language at the time, such as in LePage du Pratz's *Histoire de la Louisiane* from 1758, the records of a conspiracy trial from 1731 cited by Hall (1992, 110) or in the records of a murder trial in 1748 cited by Hall (1992, 114&178) and analyzed in depth by Klingler (2003a). Actual texts in LC are only passed on from the mid 19th century. A collection can be found in Neumann-Holzschuh (1987), the oldest of which is a Creole translation of a fable from 1846. The main question is whether LC developed locally and is hence indigenous to Louisiana or was introduced to

the region by the presence of Creole speaking settlers, namely the Haitian refugees who arrived in 1809.

Klingler (1997) examines the samples that Hall (1992) provides from 1748 and hypothesizes that early LC was indeed spoken at the time. Whether this was LC, the previous Pidgin or some in-between system, it was an unstable and highly variable form which nevertheless became widely spoken as a means of communication between different social and ethnic groups (Marshall 1997). Holm (1989) claims that white children born in the colony would learn LC growing up and continue to use it with slaves and free blacks and Hall (1992) provides several quotes indicating that white Creoles would learn LC from their nurses and keep using it throughout their lives. Harrison (1882, 287) writes

> The Creole children, entrusted from infancy to the care of negro *mamans*, learn the patois before they learn the regular French [...]. All the *petits blancs* or 'poor white trash' of the urban and plantation population speak the same patois simultaneously with the French.

Racist and derogatory sentiments aside, this indicates that LC was widely known by the late 19th century at least[16]. Klingler (1992) argues that LC was used in the St. Martinville area to communicate with the Haitian refugees when they arrived in the early 19th century. These claims and pieces of evidence indicate that LC was in fact commonly spoken and most likely used in Louisiana before HC was present there and hence is not a result of that contact. Hull (1993) finds further demographic support for this assumption and concludes that "a study of the origins of the slave population [...] in Louisiana strongly suggests that Creole there was not primarily introduced by slaves brought from the Lesser Antilles or Haiti, though such contacts occurred at a later date" (Hull 1993, 392).

Hall (1992) claims that LC originated in Senegal where a relexified version of a Portuguese-based pidgin developed that came to Louisiana with the slaves, 2/3 of which embarked there (Klingler 2003a). The suggestion is to some extent in line with Hull's (1979) hypothesis on the origin of French-based Creoles, which allows sub- and superstrate influence for each colony but always on the basis of a "common grammatical, phonological and lexical core in virtually all Creoles which is attributed to PigFr"[17] (Hull 1979, 211). Hall (1992) bases her hypothesis on the presence of terms of African origin in LC. While it is possible that LC originated in West Africa, I am not aware of any substantial evidence in favor of this view. Klingler (1997) shows that there is little to no

[16] Please note that *Creole* here refers to the locally born children of the French upper class. Some interpretations of the term are discussed below.
[17] PigFr refers to a French Pidgin that was, according to Hull (1979), a deliberate calque of Portuguese Pidgin in Benin rather than Senegal.

linguistic evidence for African influence on LC grammar. The only features that seem to resemble structures found in African languages are post-nominal determiners and the agglutination of French determiners to nouns. Both are often found in Creole languages and could be explained by the generalization of initial errors made under "averse learning conditions" (Valdman 1998, 18). Since lexical borrowings are quite common in language contact situations, the presence of lexical items with African origins in LC appears plausible as well. To determine with certainty whether LC descends from a West African French Pidgin would require extensive research and is not the subject of this thesis. A definite claim can hardly be made due to the lack of early language data and detailed historical accounts. Ultimately, general beliefs about the monogenesis or in situ genesis as well as the substrate vs. superstrate influence underlie each position. Therefore, I follow the argumentation of Klingler (2003a) and conclude the historical remarks with the least controversial stance that LC is indigenous to Louisiana where it developed gradually during the 18th century and coexisted with different varieties of French from the start.

4.3 The sociolinguistic profile of South Louisiana

Though English is undoubtedly the main language used for everyday purposes in South Louisiana today, the territory is still far from monolingual. Due to its past, Louisiana remains culturally unique "even today" (Klingler 2003a, xxv). While Louisiana has Spanish-speaking and Vietnamese-speaking communities of immigrants, aside from English, it is French which is the most present language especially in the South of Louisiana (where 'French' refers to Standard French (SF), Louisiana Creole (LC) and Cajun French (CF)).

4.3.1 Louisiana Creole

LOUISIANA CREOLE, the subject of this thesis, is the highly endangered French-based Creole language spoken in South Louisiana. The language differs structurally from French as spoken in Louisiana or elsewhere (cf. Chapter 4.4). It shares much, though not all, of the regional French lexicon but exhibits several structural features typically found in Creole languages. These include a (near) absence of morphological markings, the use of TMA markers, analytic language structures and postponed definite articles (Valdman and Klingler 1997)[18]. As Klingler and Neumann-Holzschuh (2013) point out, LC is structurally closer to French than many other French-based Creoles. One likely important factor is the

[18] For a detailed overview of proposed typical creole features and a data-driven discussion of the question if creoles form a typological class see Daval-Markussen and Bakker (2017).

continued coexistence with the local variety of French. LC is characterized by variation though, as is discussed below, and influence from Louisiana Regional French is most strongly attested in white speakers (Klingler and Neumann-Holzschuh 2013).

Creole has a number of meanings in Louisiana, which can make much discussion ambiguous. Historically, *Creole* is generally understood to refer to persons of European descent born in the colonies as opposed to those settlers who migrated from elsewhere. In many colonies, locally born black people - usually slaves - were referred to as *Black Creole* or *Creole slaves*. As Dominguez (1977) describes in detail, Creole identity was of little importance in Louisiana until the late 18th century when the refugees from Saint Domingue arrived who commonly used the term. Louisiana Creoles began to identify as such in order to separate themselves from the Americans who were migrating to Louisiana. They thus added a cultural dimension trumping the main criterion of descent. *Creole* now referred to French-speaking and French-educated Louisiana natives. It did not pertain to a racial distinction. For example, the *gens de couleur* (free people of color) were considered Creoles as well as the white descendants of French plantation owners. As a consequence, *Creole* denotes the mixed-race population of Louisiana in another common interpretation (Valdman and Rottet 2010). With the Jim Crow laws perceptions changed and because of the inclusiveness of the Creole identity in Louisiana, white Louisiana Creoles faced suspicions about racial impurity. In reaction, white Creoles excluded anyone from the category who could not claim a 'pure' bloodline. *Creole* in Louisiana now was supposed to refer solely to locally born, French-speaking whites. But "the colored Creole population continued to identify itself as Creole in spite of the written literature that denied them that right" (Dominguez 1977, 595). Consequently, two groups developed which identified as Creoles by different criteria. Both required French ancestry and language/culture but one was an exclusively white elite while the other primarily included light skinned blacks. For a detailed discussion the reader is referred to Dominguez (1977). Today, being Creole in Louisiana is largely associated with the black community and French heritage, though it is not necessarily claimed as an identity label (Melancon 2000). Nevertheless, some white people of French/Spanish descent are claiming a Creole identity today, and it is often acknowledged that Louisiana Creoles are "racially diverse people" (LaFleur and Costello 2012, 8) who share a culture, food and language.

4.3.2 Louisiana Regional French

CAJUN FRENCH is the variety derived from the French spoken by French settlers and the Acadians and it is the label most locals apply to the regional variety, although it is not uncontroversial. First, it can be difficult to distinguish between the different varieties spoken in Louisiana, as some speakers have elements from CF as well as LC in their speech and many locals as well as the general public tend not to make a distinction between the two in the first place. Despite the detailed descriptions offered for example by Neumann (1985) and Klingler (2003a), no clear linguistic boundary has been drawn between the local languages. Within each language a large amount of variation is present, further impeding clear distinctions. Hence, an uncontroversial term, such as LOUISIANA FRENCH (hereafter LF), which can refer to any local variety, is often preferred by linguists (Valdman and Rottet 2010). Second, the term *Cajun* as derived from 'Acadian' is subject to some debate and consequently best avoided, although it is omnipresent in South Louisiana and is associated with French-based culture, language and Louisiana cuisine worldwide. Its meaning remains opaque, covering linguistic, ethnic, racial and cultural traits. Used as the most common label for local French-based culture and identity in South Louisiana today, it still implies a perceived cultural, culinary and linguistic Acadian heritage. While it is undisputed that these settlers had an influence on the cultural and linguistic development in the French triangle, the people of Louisiana at the time were far from a homogenous group and stances differ on the amount of influence they exercised. Activists and scholars around Lafayette, the University of Louisiana and CODOFIL (see below) advocate the French-Canadian connection but others oppose the emphasis on Acadian influence and stress locally developed characteristics (John LaFleur p.c.). Historically, *Cajun* was a derogatory, highly offensive term used for uneducated locals. Since the 1970s it has been effectively rebranded to refer to all Francophone white Louisianans, creating a homogenous and marketable unification of the French speaking population. Trépanier (1991, 164) refers to this process as "Cajunization of the French Louisiana identity". She argues that "this choice [of *Cajun* as a label] can only be interpreted as the desire for the French Louisiana elite to assure for the region a 'white' identity." According to her, it did not take into account the diverse heritage of French-oriented white people in Louisiana and it still excludes black and Indian people of French culture (Trépanier 1991). As such, distinctions along racial lines are superimposed on linguistic and cultural distinctions, creating blurry demarcation lines and terminological confusion.

It is true that many people from this region can trace their ancestry to the exiled French Canadians. It is equally true that strong French influences were in place before they arrived and that many locals find their roots with other groups of settlers, such as the French and in some cases Spanish colonizers or their West African slaves. Irrespective of historical accuracy (or possibly lack thereof) *Cajun* or *Acadian* are terms commonly used to refer to the people, language and culture of the region. I follow Klingler (2003b) in terming the local variety of French usually referred to as *Cajun French*, LOUISIANA REGIONAL FRENCH (hereafter LRF), as not all speakers self-identify as Cajuns. Some may hence be offended by the label, others may not even be aware of its inaccuracy for their heritage and black speakers will not consider themselves to be Cajuns regardless of their lineage.

4.3.3 Multilingual South Louisiana

Commonly, a LANGUAGE CONTINUUM is assumed for French Louisiana encompassing: Louisiana Creole, Louisiana Regional French and Standard French (Valdman 1997, Klingler 2003a). Such a continuum is often taken to imply a hierarchical order from low to high prestige codes. For Louisiana, Standard French (SF) is considered the high prestige language alongside English. Local varieties such as LRF or LC are held in much lower regard. For LRF attitudes have changed in the past 40 years, as people embraced it as a marker of local identity. LC, on the other hand, has always been at the bottom of the hierarchy. Taking colonial history into account, this is unsurprising, as LC would be associated with rural, uneducated populations and slavery and was stigmatized in relation to English and LRF as well as SF. This low prestige is reflected in the terms used for LC even today. Unlike linguists, speakers and locals rarely refer to the language as *Louisiana Creole*. It was instead labeled *the broken French*, *français nègre, Mo kouri mo vini* (Neumann 1985) or *Negro French* (Holm 1989), indicating the perception of an inferior, corrupted linguistic code as well as racial association, which are discussed below. Even today, LC as *the broken French* is often contrasted to *the better French* or *the good French*, which is taken to refer to LRF or standardized forms that are taught in schools. Klingler (2003b) stresses the terminological confusion with regard to language labels in Louisiana, which makes the continuum approach debatable and Melancon (2000) argues against a continuum altogether. She points out that that neither variety is clearly defined in terms of linguistic features. Further, the role played by SF is less clear than suggested by placing it at the top of the continuum. While learning SF has become fairly popular and may well have a negative impact on local forms of French, this does not make it the variety to which that native speakers of LC or LRF aspire. Consequently, it is

misleading to assume that SF is the most prestigious language. Melancon (2000) argues in addition that the mixing of people, languages and cultures described in Chapter 4.1 did in fact not take place in the 20th century when division along racial lines was strong. On the other hand, Klingler and Neumann-Holzschuh (2013) point out that

> Contact between speakers of Louisiana Creole and Louisiana Regional French was likely to have been much more intensive in the late nineteenth and early twentieth centuries, when poor whites and ex-slaves (and their descendants) often worked side by side in sugarcane and cotton fields as sharecroppers or hired hands.

Landry (2016, 251) shows how a more or less unified mixed-raced Creole population "became racialized over time and separated into racial groups". In rural Louisiana that is still the case to some extent in many places, as a racial divide persists in many towns, which have black neighborhoods in which poverty and many related social problems are even more prominent than in the white parts of town, which are often rather poor as well. The extent of contact between the racial groups appears to differ from place to place. For example, I received a surprising amount of attention from passersby when interviewing an older black man on his front porch in Henderson. One white person even felt the need to stop his car to observe us. When asked about this behavior the interviewee explained it was to be expected that "many white folks don't like a young white lady like yourself sitting with a black guy like me". Claims of a continuum where LC decreolizes towards LRF, and LRF develops towards SF are indeed doubtful in light of these social circumstances found in Louisiana. Moreover, if locals take pride in their language, wishing to promote and maintain LRF as is commonly reported, it is questionable why they would attempt to approach SF in their speech. The continuum approach consequently has to be taken with a grain of salt although most of these reservations seem to argue more against the influence of SF on LRF and LRF on LC rather than the existence of a continuum per se. Especially in light of the LC-variety spoken by whites in the Henderson area, which is characterized by the presence of LRF features, contact and its effects on LC cannot be denied. Distinguishing LC from LRF is also more feasible than Melancon's arguments make it out to be. While speakers often exhibit linguistic features of both varieties, those features can usually be identified as typically LC or LRF. A speaker who is fluent in both languages was able to differentiate clearly between the two during fieldwork. The confusion is less about a linguistic demarcation of codes, although it appears to be an impossible task to fully and clearly establish a distinction as well; it is more a matter of speakers not abiding strictly by one or the other. This is a consequence of language contact as is discussed below. Despite the possible objections to a continuum approach, it is certain that historically and linguistically different varieties exist as well as a good amount of variation.

The dominance of English and related policies as described above had an impact on French-based language varieties in Louisiana. The intense stigmatization and active repression experienced by a whole generation of speakers most certainly caused negative consequences for LC and contributed to the decline in usage and eventual language endangerment. Speakers of both LC and LRF report being (corporally) punished for speaking their native language as children in school, often regardless of the fact that they had no knowledge of English, although some creolophones claim that punishments were especially severe for speaking LC as compared to LRF (Rita Marks, p.c.). A decline was sure to follow and Louisiana French rapidly lost ground, most likely on the path to loss as an L1 in Louisiana.

In the late 20th century, the situation changed significantly, in large parts due to the foundation of CODOFIL ('Council for the Development of French in Louisiana') in 1968 (Rottet 2001). Their efforts started a remarkable renaissance of French-based language and culture in Louisiana. Public awareness of local French-based language and culture increased significantly and their maintenance and preservation became important issues. People started to make an effort to this effect. In this spirit, numerous activities were and still are carried out, including many French tables where people get together to speak, teach or learn their (heritage) language, culture days, festivals, music events, the establishment of museums, French immersion schools (since 2011) and culture centers and many more, to the point where now many street signs are bilingual and French has become an accepted and beloved part of Louisiana. This went hand in hand with a change in attitudes that leaves most Cajuns today proud of their heritage, language and culture (Klingler 2003a). While efforts were largely directed towards SF in the beginning, the preservation of LRF was desired, and ultimately campaigned for, by the populace (Marshall 1997). The role of Standard French as promoted by CODOFIL has been criticized, since the focus on international French, for example of benefit for the growing tourism industry, can be argued to be counterproductive for the maintenance of local varieties. Tourism is a strong and important industry in Southern Louisiana with more than 51 million visitors in 2018[19]. Americans, Canadians and Europeans visit for Mardi Gras, one of the numerous festivals or the year-around impressive landscapes, unique cuisine and ever-present local music. Today, this provides an important motivation for maintaining Louisiana's cultural and linguistic identity as distinct from the US at large. The positive effect this has had on the status of French in Louisiana is

[19] Louisiana Department of Culture (2019), http://www.crt.state.la.us/tourism/louisiana-research/ (last accessed 22. August 2019)

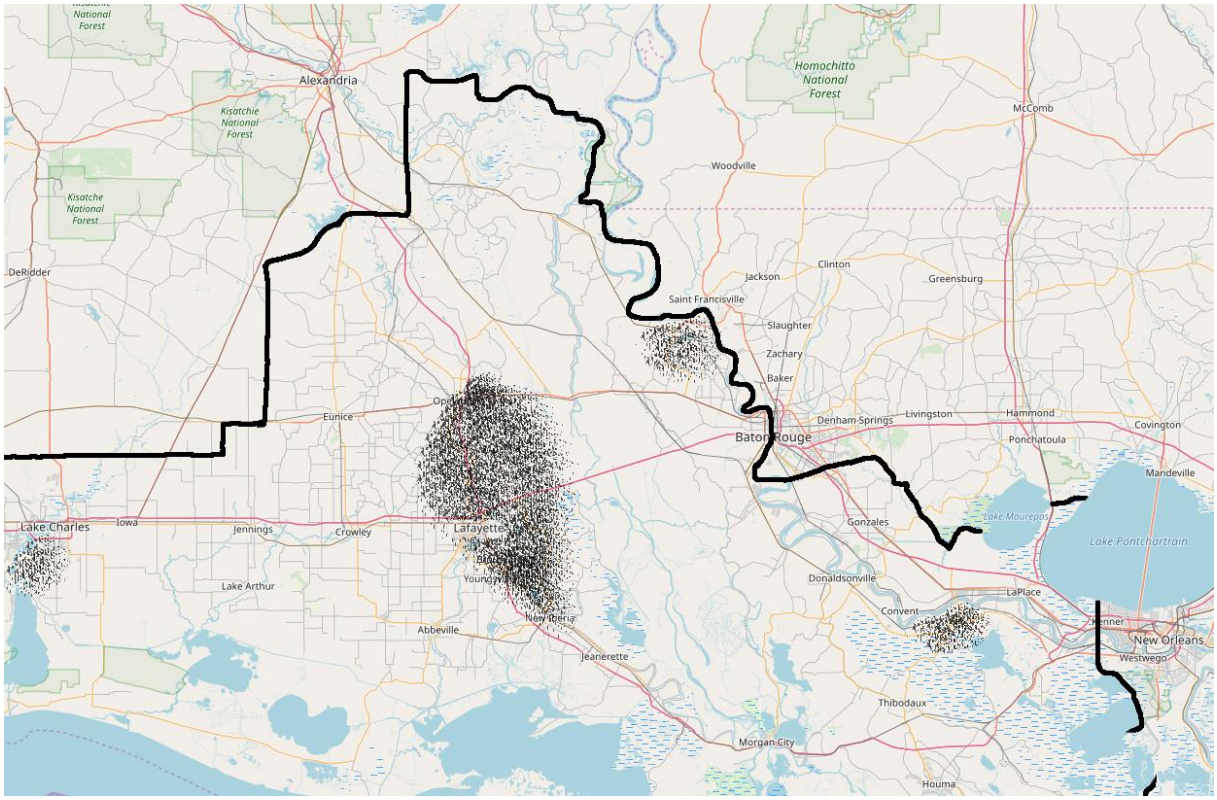
obvious. At the same time many native speakers of LRF are aware of differences between their local forms of French and the one spoken in France and Canada respectively and it is their local way of speaking they aim to preserve. The discussion has been ongoing since the 1970s and still is lively at many French tables throughout the region and many speakers are worried that international French might hinder more than help their goals. This is especially true because English remains most children's first language. French, French-Canadian or Belgian teachers then expose them to SF in immersion classes, instead of them acquiring a local variety. Klingler (2003a) cautions that this may bring about a shift towards SF. While the usefulness of SF in this situation is certainly debatable, Klingler and Neumann-Holzschuh (2013) deem the success of implementing it in schools limited and argue that LRF is in fact far more present in Louisiana. LRF is nowadays very visible in the public eye, though not necessarily prominent as a first language, and appears to become increasingly important. Klingler (2003a), like many others, is optimistic for its future, though Rottet (2001) remains skeptical, arguing that LRF is declining as a L1.

In contrast, little to no efforts were made to preserve LC. Consequently, the situation of LC is much more dire than it is for LRF. Valdman (1997) already identifies LC as the most endangered language in Louisiana and Klingler (2003a) predicts that it will be lost within 20 to 30 years. According to Marshall (1997) speakers simply had no interest in maintaining LC. While that may be true to some extent, it may also be a result of the lower social and economic status of those speakers. A few activists made some isolated efforts mainly in the 1980s, such as the Creole Inc. magazine and the *Creole Linguistics* column. Unfortunately, none of these were continued into the present. In recent years community members have again begun to advance LC on a small scale. Creole speaker and scholar Christophe Landry et al. (2016) reports "hundreds of people all over the world with a desire to learn Louisiana Creole". Using social media and a few publications, he offers an opportunity to do just that. While those efforts as well as the prominence of LC Zydeco music contribute to some revalorization of LC, the reclamation movement is currently based online and thus disconnected to a degree from the offline speech community. The overall public attention devoted to the language is still marginal. Only now some steps are taken to include the Creole language and people at least officially, for example by changing the name of a major music festival from the *Festivals Acadiens* to the *Festivals Acadiens et Créoles*. The current situation of LC is discussed in more detail below.

4.3.4 Louisiana Creole as an endangered language

Louisiana French is spoken within the so-called French-triangle, indicated by the line in Figure 4-1. LC, more specifically, is limited to a few select areas marked in grey.

Figure 4-1 Southern Louisiana - the French triangle and LC speaking areas



© OpenStreetMap

In this map, adapted from Neumann (1985), the main areas in which LC can be heard are highlighted. Klingler (2003a) largely agrees and list the German Coast and Acadian Coast, namely St. James and St. John Parishes, Pointe Coupee Parish around False River, and the area West of the Atchafalaya River Basin, namely the Bayou Teche region, St. Martin Parish, Lafayette and St. Landry Parish. This area has the most speakers according to him, though there are some notable regional differences between the Creole spoken in the last area and the Mississippi region. This in fact already touches upon two major descriptive difficulties. On the one hand, to identify and count speakers and on the other hand, to account for what Klingler (2003a, xxvi) refers to as "regional variety" within LC and "important variation tied to social groups". A third difficulty logically follows in determining language vitality, since this is an endangered language.

Speakers of LC are more difficult to identify than may be apparent. Firstly, the language is rarely spoken in public. One reason, according to Rottet (2001), is that it is highly stigmatized. Consequently, speakers prefer not to use LC where they might be heard by outsiders because they feel ashamed or fear negative reactions. Klingler

(2003a) reports that some speakers will not admit to speaking LC even when asked, similar to what Grinevald and Bert (2011) describe as ghost speakers. This attitude appears to be changing to some extent, as several speakers reported that they had been made ashamed of their language when they were younger but now took some measure of pride in it. This is discussed in Chapter 6.1. Secondly, LC is commonly associated with the descendants of slaves (Rottet 2001), hence the typical speaker is perceived to be black (Klingler 2006). In this, the confusion of labels in Louisiana can be very misleading. *Creole* and *Cajun* as terms both refer to languages as well as cultural traits and ethnicity. They are also interwoven with notions of race. Klingler (2003b, 89) shows "that the use of language labels is highly variable and often has more to do with ethnic identity than language use". Consequently, a francophone black person in Louisiana may identify as Creole though they speak Cajun French and white persons may consider themselves Cajuns, even though they speak Louisiana Creole and/or are not descendants of the Acadian settlers. Furthermore, most speakers of LRF and LC refer to their language simply as 'French' and would consequently check that box on a census questionnaire for example, which is why census data can be unreliable (Klingler 2003a). In sum, identifying speakers of LC in the field can be challenging, estimating their overall numbers even more so.

Nevertheless, linguists and demographists alike attempt to gauge the numbers of LC speakers. One much quoted figure is that put forward by Neumann (1985, 20). She estimates 60,000 to 80,000 speakers, 10,000 to 20,000 of which she supposes to be white, indicating that LC, though mainly spoken by black people, is not actually limited to one racial group. Valdman (1998) figures 20,000 to 30,000 speakers of LC although another decade later, Klingler (2003a) assumes less than 50,000 speakers. He adjusts that number significantly downwards in Klingler (2006), when he estimates about 10,000 to 20,000 speakers. The US Census Bureau reports 6,297 speakers of 'Creole French' in Louisiana for 2006-2008, and for 2009-2013 lists 6,706 speakers of 'French Creole'. While the previous figures are estimates, these represent an actual count. Unfortunately, the census does not take into account the different labels and resulting confusion described above. In addition, the census asks for the language spoken a home. For many native speakers of LC this is no longer the case, for example in cases when a spouse passed away or speakers live alone. It also does not include speakers of LC outside of the state, such as in Texas where LC language islands exist (Klingler 2003a). One may consequently expect that this count might be on the low side. Please note that the most recent census data seem to show an increase rather than a decrease, counting

403 more speakers during 2009-2013 than during 2006-2008. Assuming that these are actually speakers of LC, it would imply a positive trend, though on a small scale. Hopefully, this may be a result of the recently made language maintenance efforts. Although it appears unlikely that LC has turned into the main language of a household in this time frame in many cases, some speakers may have used their language more often, hence opting to select it in the census. Of course the questionnaires may also have reached more LC speakers by chance or speakers of other French-based Creole languages, such as Haitian Creole, make a stronger appearance in those numbers than they did three years prior, as the census lists 'French Creole' and does not specify further. Neumann-Holzschuh (2015) refers to 4,000 to 7,000 speakers, which amounts to less than 10 percent of her estimate 30 years prior. In the most recent work on LC, Mayeux (2019, 59) estimates 3,500 to 6,000 speakers. All of these numbers have to be considered carefully, bearing in mind that they are best guesses which may sometimes be rather generous, or in the case of the census data not fully reliable for other reasons. Despite these reservations they show an obvious decline in the number of speakers over the years, which indicates the highly endangered status of LC.

This decline surfaces in other characteristics as well. One of those is the age of speakers. As Rottet (2001) points out, LC is spoken mainly by older people. Klingler (2003a) agrees that most speakers are above the age of 60. A few younger speakers took part in the present study, but native speakers indeed tend to be from the grandparent or often great-grandparent generation. This of course indicates that younger generations either did not acquire LC or if they did, do not use the language anymore. Furthermore, Marshall (1997) reports that almost all speakers of LC are bilingual, with English the other language in their repertoire. Neumann (1985) already observed this trend but still included a group of speakers over the age of 60 who were monolingual in LC. The younger group from 30 to 60 years of age is reported to be bilingual even then. In the more recent literature bilingualism is assumed to be the norm in the LC speech community, as there appear to be no more monolingual speakers (Klingler and Neumann-Holzschuh 2013). Most likely the speakers interviewed by Neumann (1985) either learned English as it grew more and more dominant or passed away. There may certainly be isolated occurrences of monolingual LC speakers unknown to researchers but this would be an exception.

Relatedly, the notion of SPEECH COMMUNITY must be applied carefully in Louisiana. Usually, this term refers to bodies of people who are in frequent and regular interaction using a common linguistic code, the usage of which differs significantly from that of

another community, thus setting them apart. It is most commonly used for smaller communities such as villages where a particular language or dialect is spoken but has also been applied to nations or subgroups of e.g. immigrants. One the one hand, most of South Louisiana is made up of rural, close-knit communities. The social networks are dense, formed in small towns, churches and through cultural connections and common interests. People know each other and often rely on friends and neighbors. On the other hand, these communities do not necessarily equal speech communities. LC speakers live among LRF speakers and the English-speaking majority. Strong connections to people are in place and they do not pattern according to linguistic groupings. LC speakers are concentrated in the areas listed above but they remain in the minority in these areas. They do not necessarily live in close proximity to one another and do not belong to an ethnically or culturally separate group that could be considered the *reference population* in the sense of the UNESCO (c.f. UNESCO Ad Hoc Expert Group on Endangered Languages 2003). As Klingler (2003a, xxxii) puts it, the "community is fragmented". This certainly contributes to the loss of LC, as a language will naturally be used less within a community if not all members have competence in this language. It will also be less powerful as an identity marker. English offers a comfortable alternative, which excludes nobody in the case of Louisiana and is hence used most often. Using LC instead for many speakers seems to be a conscious choice rather than the default option.

Following from the community structure, the linguistic situation in Louisiana is largely shaped by LANGUAGE CONTACT. This is true historically, as was described above. As Klingler and Neumann-Holzschuh (2013) put it:

> Sustained contact may have been one avenue for the spread of Louisiana Creole among some elements of the white population (Neumann 1984, 1985), and it may also explain why the Louisiana Creole spoken today, especially in the Bayou Teche area, generally shows more French-like features.

Valdman (1998, 3) even attests a "symbiotic relationship" between LC and other local French-based varieties. This situation explains why LC differs less from French than many other French-based Creoles. Today, speakers of LRF, LC and English interact on a daily basis and typically speak at least two of the three languages with SF adding another layer. Concerning LC, the effects are most commonly phrased in terms of decreolization.

DECREOLIZATION usually refers to a linguistic development in the direction of the lexifier. As discussed for example by Siegel (2010), the usefulness of the concept is debatable. Mainly, the term sets observable change in Creole languages apart from causes and effects of language change in other languages, although it remains unclear in which ways the processes differ, except for involving a Creole language and its lexifier. To arrive at a meaningful definition, one would first have to identify typical Creole/basilectal

features that give way to their acrolectal counterparts. While attempts at listing typical Creole features have been made, most prominently by McWhorter (2005 and subsequent works), cases of decreolization are not generally associated with those specific features. Further, one needs to define what is targeted by decreolization, namely which changes are referred to precisely. Some consider purely linguistic changes to be the hallmark, such as grammatical structures or phonological change. A more sociolinguistic approach includes language use and the speech community. In the case of LC, the presence of LRF, SF as well as English provides additional complexity as described above. It is uncontroversial that LC has changed significantly since the earlier records, especially in the Bayou Teche area, and many of these changes replace features that were considered typical for Creole languages in the past. Some concrete linguistic features as well as the notion of Creole prototypicality are briefly addressed in Chapter 4.4 below. Neumann (1985) treats these changes as effects of decreolization while Mayeux (2019) argues for a non-Creole specific account. He concludes his empirical study by stating that the processes at work in LC are best accounted for in frameworks of language change and language contact, as decreolization cannot account for contact-induced change caused by a dominant language other than the lexifier and also disregards internally motivated processes. Valdman (1997) argues that LC develops towards French but is unsure if this is a process of decreolization or caused by language contact to LRF or "resulting from language loss" (Valdman 1997, 13). All three are possible factors, which are likely to exert an influence on LC. But, as argued by Melancon (2000), it remains unclear what would motivate speakers to change their speech to resemble SF or LRF in light of the social structures, especially if "LC serves mainly as a symbol of identity for a community that has shifted to English" (Valdman 1998). Decreolization and contact to LRF/SF would hence be less likely triggers of change, though by no means impossible. It can be argued that many speakers of LC are not fluent in the language anymore, which results in structural as well as lexical gaps in their linguistic knowledge. It would then be feasible to assume that some of those gaps are spontaneously filled with material from contact languages as linguistic insecurity facilitates language change towards a contact variety. This view makes the influence of LRF more plausible than it appears to be from the language attitudes cited by Melancon (2000). As a consequence of the fragmented speech community and the presence of LRF in the region, many speakers are exposed to LRF more than SF.

Following that line of thought, one would also expect structural change to introduce English-like constructions rather than or at least alongside SF- or LRF-

elements despite the larger linguistic distance, as English is the dominant language. Recall that such change has been observed in language endangerment scenarios. For example young speakers replaced the Dyirbal ergative system with a nominative-accusative one and followed the English pattern by introducing prepositional phrases instead of locative case markers (Schmidt 1985). Additional effects one would expect from the contact to English are frequent use of code-switching and extensive borrowing. Both have been observed in the literature (Neumann 1985, Klingler 2003a) and during fieldwork for this study. Another less frequent consequence are lexical calques. Surprisingly, very little research is concerned with the effects following from the intense language contact with English, beyond general claims that LC is in contact with and influenced by LRF and English. The studies that exist tend to focus on lexical effect since these are the most obvious (Klingler, Picone, and Valdman 1997). The present study is only marginally concerned with these effects though preposition stranding is discussed as a notable example in Chapter 7.3.

Lastly, modern LC is characterized by a large amount of VARIATION. This variation has never been studied systematically but was frequently observed (Klingler 2003a, Valdman 1998, Marshall 1997, Klingler 2000). Regional or geographic variation is the most obvious, as the Creole spoken in the Bayou Teche region differs from that spoken in the Mississippi area in many (subtle) ways. Generally, these are regarded as two separate varieties, Teche Creole (TLC) and Mississippi Creole (MLC) (Speedy 1995, Klingler 2000, Mayeux 2019, Klingler and Neumann-Holzschuh 2013). It was proposed that they may even have developed independently (Klingler 2000, Speedy 1995) but this notion has now been refuted (Klingler 2000, Rottet 2006). TLC is considered less basilectal than MLC, due to more intense contact with SF and LRF and resulting effects of decreolization. While it is debatable if decreolization is an applicable concept here, language contact and change clearly seem to be at work. For a detailed discussion see Mayeux (2019). The MLC variety, on the other hand, seems to have preserved more 'original' Creole characteristics but at the same time seems to be more severely endangered. Field workers commonly report that creolophones in this area are more difficult to find, which fieldwork for this study confirmed. On the linguistic level, lexical differences are the most noticeable at first glance. MLC also maintains more features that are found in the 19th century texts and hence considered basilectal. Examples are the consistent use of post-nominal determiners and the lack of gender marking on the indefinite singular article. While the variation between long and short verb forms can be observed in both modern varieties, it

is used systematically to mark tense-aspect distinctions only in TLC[20] (Klingler and Neumann-Holzschuh 2013, Klingler 2003a, Neumann 1985). Valdman (1997) further lists variation due to social variables, situational variation, generational variation and variation as a means of group identification. For many speakers, the use of LC is limited to a small, specific group of interlocutors. It stands to reason that these patterns of interaction give rise to variation which Klingler (2003a, xxvi) connects to social groupings, as an in-group variety may likely develop. What Valdman (1997) takes to be generational variation may in addition be an intergenerational decline in competence and as such a consequence of language endangerment. That would be in line with Marshall's (1997, 345) assessment that "the vast amount of variation present in LC is a reflection of this linguistic insecurity". Since LC was never a standardized language, the decline in usage and related loss of linguistic security among speakers paired with the increased amount of language contact naturally gives way to much variation, even within a single speaker's usage.

Considering the complexity of the situation in Louisiana as well as the objections to the concept of decreolization in general, I follow Mayeux (2019) in taking a broader perspective that regards language change as a consequence of language contact rather than to approach observable change purely from a decreolization point of view. While it is tempting to explain unexpected similarities and blurry demarcation lines by decreolization or a development towards another code, such claims are not entirely based on social, demographic and linguistic facts. Note that influence of the lexifier or a variety of the lexifier is by no means excluded in this approach, it simply does not appear necessary to establish a separate concept for such contact effects.

4.4 Selected aspects of Louisiana Creole grammar

This section provides a brief sketch of LC grammar, focusing on features that set the language apart from varieties of French and cause it to pattern with other Creole languages, as well as features pertaining to movement operations due to this study's concern with wh-questions.

Creole exceptionalism and the idea of Creole languages as a distinct typological class based on their structural make-up remain a major controversy in the field. One issue is that exceptionalism is often linked to notions of simplicity and inferiority, thus taken to imply that Creoles are abnormal. This view is strongly opposed by DeGraff (2003) and Mufwene (2000) among others, who focus on their non-exceptional status in

[20] See the description below for details and examples.

the sense of equality, taking a superstrate approach in considering Creoles descendants of their lexifiers. The contemporary mainstream view is that Creoles are fully-fledged languages, not exceptional in terms of their value, be it positive or negative.[21] The question remains if they form a group that is distinct from other languages. Some scholars see the greatest commonality among Creole languages in their shared "specific sociohistorical settings" (Ansaldo, Matthews, and Lim 2007, 4) or stress, "how much structural variation there is to be found among those languages" concluding that "indeed, the notion of prototypicality of Creole grammar is all but abandoned" (Kouwenberg 2010a, 184). Others insist, "Creoles are a synchronically identifiable type of language" regardless of historical development (McWhorter 2012, 377).

Recently, Bakker et al. (2017, 370) provided extensive empirical evidence that "Creoles are more similar to one another than to superstrates and substrates". They show that Creoles tend to cluster together with respect to feature pools of varying size, as previously observed under the premise of structural complexity/simplicity by Parkvall (2008). While studies have long attested typically Creole characteristics for the respective language(s) under observation there is no clear consensus as to which features these are. Researchers have proposed overlapping but by no means identical lists of features that synchronically set Creole languages apart from other linguistic codes (among others Bickerton 2016, [1981], McWhorter 1998, 2005, Holm and Patrick 2007) all of which have been critically received in the field. Daval-Markussen (2013) points out three features which are to be found in most Creoles and no other languages in this configuration, namely the presence of an indefinite article derived from the numeral one, predicative possession marked by a verb 'to have', and absence of bound tense or aspect inflection. While this thesis does not take a typological or even cross-linguistic position, it is assumed that Creole languages exhibit certain structural similarities. Each of these characteristics is by no means restricted only to Creoles, nor are they taken to be necessary or sufficient indicators of the existence of a typological class[22] but they remain remarkable commonalities most of which are to be found in LC as well. It is notable that some of these features are less absolute in LC than some other Creoles, either due to LC's status as a conservative Creole that has always been structurally closer to its lexifier than other Creole languages or as a consequence of language change that may be termed decreolization (Neumann-Holzschuh 2000).

[21] Under the assumption that Creole languages are younger than any other languages currently in use, some scholars maintain their special value to linguistics though. See Krämer (2013) for reflections on the desire of creolistics to define creole languages as especially worthy of studying.
[22] On the difficulty of including Creole languages in traditional typological accounts see Kouwenberg (2010a, 2010b).

Concerning the three features identified by Daval-Markussen (2013), LC patterns with other Creoles, as it has an indefinite article *en/enn* derived from French *un*, employs the verb *gen* to mark predicative possession and generally expresses tense, mood and aspect through free markers rather than bound inflectional affixes. Each of these, with the exception of possession marking, is described in more detail in the respective sections below.

4.4.1 The verbal system

Absence of inflectional morphology is a common feature in analytic languages, which is not restricted to Creoles but characteristic for them nonetheless. Typically, these languages employ unbound preverbal markers to express tense, mood and aspect instead. This may well be the feature that has been perceived as most typically Creole in linguistics. It appears in Bickerton (2016, [1981]), McWhorter (1998, 2005) and any introduction to the structure of Creole languages as the textbook feature. In depth studies of individual languages have challenged the across-the-board applicability of this feature, however, as some Creole languages make use of inflectional morphology to some extent and the overall picture appears to be more complex than the generalization suggests. LC is a case in point as is discussed below.

The free markers employed in LC are *te*, *ape*, *(s)a/ (v)a/ ale*, *se* and sometimes *bin*. *Te* (also *t*) is a tense maker that expresses past or anteriority (cf. (49)) while *ape* (also *ap, pe* or *e*) marks progressive aspect (cf. (50), (51)). *Sa, (v)a* and *ale* are future markers whose distributions carry different nuances but are not fully systematic (cf. (52), (53)). Rottet (1992) takes *sa* to be a future tense marker that carries perfect aspect with action verbs and *va/a* to be irrealis mood markers with a future interpretation. *Se* is used for conditionals, marking irrealis mood (cf. (54)). The intersection of future tense and irrealis mood has been a common perspective in Creole studies since Bickerton (2016, [1981], 1984) subsumed the categories under the irrealis label. The different markers can be combined to express compound features for example the past or future progressive (cf. (55)).[23]

(49) Ye te kouche tou-sel. (Klingler 2003a, 237)
3PL PST sleep alone
They slept by themselves.

[23] The interpretation of the different preverbal markers is actually more variant and complex than depicted here. It may depend on context and/or the type of verb which is used and is subject to regional as well as idiosyncratic variation. *Ale* for example is not attested for the Pointe Coupee variety and there *te* can indicate irrealis as well. *Ape* can also serve as a marker of habitual actions and so on. For more detailed accounts, cf. Neumann (1985), Klingler (2003a) and Klingler and Neumann-Holzschuh (2013).

(50) To konpran byen sa m'ape di twa? (Neumann 1985, 210)
2SG understand well REL 1SG PROG say 2SG

Do you understand what I'm telling you?

(51) Dolo-la e galope. (Klingler 2003a, 256)
Water-DEF PROG run

The water is running.

(52) Li sa pare pou vou dèmen. (Klingler 2003a, 260)
3SG FUT ready for 2SG tomorrow

She will be ready for you tomorrow.

(53) Nou va fe la rekol, mwa e twa. (Neumann 1985, 215)
1PL FUT make DET harvest, 1SG and 2SG.

We will (do the) harvest, you and I.

(54) Si mo se konnen, mo se mannde la [sic.] plen […] (Klingler 2003a, 261)
If 1SG COND know, 1SG COND ask 3SG? a lot

If I had known, I would have asked her lots of things […]

(55) To t'ap travaye pou le Kadjen. (Neumann 1985, 210)
2SG PST PROG work for DET.PL. Cajun.

You were working for the Cajuns.

LC had just one invariant verb form in the nineteenth century records, the long form (Neumann 1985). This form is used invariantly with and without preverbal markers. In the absence of TMA markers, stative verbs are interpreted as present tense and non-stative or dynamic verbs are ambiguous in that they can express habitual or universal present as well as simple past. The anteriority marker *te* marks past tense with stative verbs and pluperfect with dynamic verbs. This system largely represents what would be expected from a basilectal Creole according to Bickerton (2016, [1981]). In the 20th century this system was in principle maintained for verbs belonging to class II in the classification by Neumann (1985) but for a restricted class of verbs a short form has developed which is used to mark TMA distinctions in a regular way in the Teche variety. (Neumann 1985)

(56) Manje ('eat'): long form; manj ('eat'): short form

The short form is used to indicate habitual present, universal present and imperative as well as with the impersonal construction *(i)fo* ('Il faut') (Neumann 1985). The long form, when it appears without any additional markers, expresses past tense/ perfective aspect. This is a tense/aspect distinction which is marked on the verb, namely the habitual expressed by the short form, and a completed past action expressed by the

long form. In combination with preverbal markers of any kind the long form is used as well.

(57) chòp-la frèm a siz-èr. (Neumann 1985, 196)
shop-DET.DEF close at six-hour

The shop (always) closes at six o'clock.

(58) chòp-la freme a siz-èr. (Neumann 1985, 196)
shop-DET.DEF close VS at six-hour

The shop closed at six o'clock.

(59) chòp-la te freme kon mo vini (Neumann 1985, 196)
shop-DET.DEF PST. close VL when 1SG come

The shop was closed when I came.

The distribution of distinct verb forms is similar but less clearly cut in Pointe Coupee according to Klingler (2003a). For example, either form is used in free variation after *te* and the distribution in the imperative appears to correspond to the use of formal or informal forms of address.

4.4.2 The nominal system

Many nouns in LC feature agglutinated elements (e.g. articles) either as a single consonant or a syllabic element (Klingler 1997) (e.g. *dolo* – water, *larjan* – money). Syllabic agglutination is not attested in LRF but is commonly found in French-based Creole languages, most frequently in but not limited to the Indian Ocean Creoles (Baker 1984, Grant 1995). These elements do not carry any grammatical meaning but are part of the lexeme. Irrespective of agglutination, definite and indefinite determiners occur in LC although these can sometimes be difficult to distinguish from agglutinated articles in spoken LC when just one of the two is present. I follow Klingler (2003a) in assuming agglutination only in the presence of an additional determiner or adjective for nouns that also occur without the agglutinated element .

The singular indefinite determiner *en*/*enn* is derived from the French numeral *un*. Recall that this is one of the features attested for all Creole languages in Daval-Markussen (2013). For plural indefinites LC also has partitive *de*. Neumann (1985, 107) distinguishes a masculine (*en*) and a feminine (*èn/enn*) form for the Bayou Teche variety, although she cautions that both forms occasionally appear with nouns that carry the other gender in SF or LRF. Since nominal inflection is absent in LC, there is no other indicator for grammatical gender with the exception of natural gender, which is often indicated by lexical pairs (eg. *Maman*/*Papa*). Gender marking on the indefinite determiner is one of the features in which LC resembles French more closely than other French-based Creole languages. According to (Klingler 2003a) the different variants

occur in Pointe Coupee parish as well but *en* is much more frequent than any other form and no gender distinction is apparent.

(60) En tas kafe (Klingler 2003a, 171)
DET.INDEF. cup coffee

A cup of coffee.

The definite article appears most frequently as *la* or *l* in the singular and *le* in the plural. The nominal, like the verbal domain, has undergone a significant structural change from the basilectal 19th century texts, namely in the distribution and the position of the definite article. Like the systematic distribution of short and long verb forms, this development is most pronounced in the less basilectal Teche variety as well. In the historical records only specific nouns appear with a determiner. In modern LC all plural nouns tend to be used with a determiner, regardless of their specificity although the distinction is largely maintained for singular nouns (Klingler and Neumann-Holzschuh 2013). Historically, definite determiners are only attested post-nominally taking the forms *la* for singular or *la-ye* in the plural. In Pointe Coupee parish, speakers still mainly use postnominal determiners, namely *la* for singular nouns and *ye* for plural nouns. Prenominal determiners, *la*, *l* and *le* are also attested but less frequent (Klingler 2003a, 174). The Teche variety on the other hand relies mostly on prenominal determiners, which are distinguished according to gender as well as number. LC has a masculine form *l*/*lə,* a feminine form *la* and a plural form *le*, closely resembling the LRF and SF system (Klingler and Neumann-Holzschuh 2013). Neumann (1985, 109) considers this "an important indicator of decreolization"[24]. The postnominal form *ye* has become rare in the Teche variety (Neumann 1985, Mayeux 2019).

(61) Post-nominal plural determiner in the Pointe Coupee variety:

Ye loure bari-ye. (Klingler 2003a, 174)
3PL roll barrel-DET.DEF.PL

They rolled the barrels.

(62) Post-nominal singular determiner in the Pointe Coupee variety:

Chyen-la trape lode lapen-la. (Klingler 2003a, 173)
Dog-DET.DEF.SG catch scent rabbit- DET.DEF.SG

The dog caught the rabbit's scent.

(63) Pre-nominal masculine determiner in the Breaux Bridge variety:

Li mande pou ki l kouto. (Neumann 1985, 109)
3SG Ask for who DET.DEF.MASC knife.

He asked who the knife is for.

[24] My translation.

(64) Pre-nominal feminine determiner in the Breaux Bridge variety:

La fiy vini reste avèk mwa isi.
DET. DEF. FEM. girl come stay with 1SG.OBJ. PRN here

The girl came to live here with me. (Neumann 1985, 110)

(65) Pre-nominal plural determiner in the Breaux Bridge data

Li fini manje le pakonn. (Neumann 1985, 110)
3.SG. finish eat VL DET.DEF.PL pecan

He has finished eating the pecans.

LC's pronominal system features number and person distinctions as well as separate forms for the formal and informal second person, but no gender distinctions. While certain forms are used in emphatic/ independent or object functions but not as subject pronouns, the paradigm as a whole is characterized by much homophony. The subject pronouns are *mo, to/vou, li, nou/nouzòt, vouzòt,* and *ye*. Variations occur for the 2nd person forms. As object pronouns, the first-person singular takes the forms *mon, mwa* and the second-person appears as *twa*. The independent or emphatic pronouns are largely identical to the object pronouns. Possessive pronouns are similar to the subject pronouns although *mon* can appear in the first-person singular and the third person singular possessive pronoun is *so* rather than *li*.

4.4.3 Constituent structure

Bakker et al. (2017, 87) attest that "almost all" Creole language exhibit basic SVO word order. In this, LC is no exception which is unsurprising as the lexifier also has SVO. Further, the order (past) tense, (irrealis) mood, (imperfect) aspect, verb stem is generally attested for Creole languages and acknowledged even by scholars who otherwise stress the diversity of Creole TMA systems (Winford 2018). Nevertheless, this order paints only part of the picture, ignoring a multitude of other functional categories that also occur in Creole TMA systems. From the limited amount of studies available on the matter, Winford concludes that Creole TMA systems tend to abide by the universal hierarchy in Cinque (1999). Cinque argues for a semantic distinction between several mood, tense and aspect phrases that appear in a fixed order cross linguistically which he takes to be determined by phrase structure. For example, evidential or epistemic mood markers precede future tense markers but the irrealis marker occurs between future and anterior tense in his model. LC abides by this as well, at least to the extent that is observable from the combinations of TMA markers in spontaneous data. Foremost, *te* can be followed by *ape* to form the past progressive and *ape* can sometimes be preceded by *se* in order to refer to "irrealis progressive action" (Klingler and Neumann-Holzschuh 2013). This

indicates that both tense and mood precede aspect marking in LC. For the order of tense and mood, the universal TMA assumption is made but to the best of my knowledge not evidenced by data in which tense and irrealis marking co-occur. Note that *te* serves as the irrealis marker almost as frequently as *se* in the data of Klingler (2003a). According to Winford (2018), this is not uncommon for Creole languages which often tend to express the hypothetical via the past marker. This may suggest that the two categories are not as clearly separated, in LC specifically but potentially in different Creole grammars, as the TMA labels imply. If they were in complementary distribution, which further research would need to confirm, one could argue that they in fact occupy the same position instead of postulating two separate projections. The internal order of preverbal markers remains unaffected by this observation.

The position of the negative marker[25] *pa* is another feature in which LC resembles SF to some extent. *Pa* can appear both pre- and postverbally. In the absence of a tense marker it stands before verbs in the past but after the verb in all other tenses. The position of *pa* hence is a major indicator for the correct tense interpretation of unmarked class II verbs (cf. (66), (67)) (Klingler and Neumann-Holzschuh 2013). With respect to verbs with two forms, *pa* appears after the short but before the long form (cf. (68), (69)) (Klingler 2003a).

(66) Mo bwa pa diven. (Neumann 1985, 322)
1SG drink NEG wine

I don't drink wine.

(67) Mo pa bwa diven. (Neumann 1985, 322)
1SG NEG drink wine

I didn't drink any wine.

(68) Li pètèt lenm pa sa. (Klingler 2003a, 321)
3SG maybe like/love NEG that.

He may not like it.

(69) Pouki to pa manje tò Gumbo? (Klingler 2003a, 321)
Why 2SG NEG eat VL 2SG.POSS Gumbo

Why didn't you eat your Gumbo?

Furthermore, the negator appears in a fixed position with respect to TMA markers. It always follows *te*, *se* and *sa* (cf. (70)) but appears before *ale*, *ape, a* and *e* (cf. (71)). *Pa* does not co-occur with *va* or *a* according to Neumann (1985, 323). Instead the marker *ale* appears after *pa* in negated future sentences.

[25] For the present study the grammatical category of TMA and negative markers is of little concern but the question of their (non)morphological and grammatical status is revisited frequently in studies on creole languages (e.g. Gabel and Brandt 2017, Henri 2016, Michaelis 2015, Henri and Kihm 2015). For a discussion of the grammatical category of *pa* in LC as either a negative adverbial marker, a clitic or an affix see Henri (2016).

(70) Li te pa kante kant. (Klingler 2003a, 324)
3SG PST NEG tell story

He didn't tell stories.

(71) No, li pa e dormi. (Klingler 2003a, 324)
No, 3SG NEG PROG sleep.

No, he is not sleeping.

This distribution baers some resemblance to the position of *pas* in French which follows finite but precedes infinite verb stems. The distribution of *pa* seems to be similar in Pointe Coupee and Breaux Bridge, although Klingler (2003a) observes more variation, partially attributable to the fact that the use of the long and short form is not as regular in the Pointe Coupee variety in the first place. The patterns described above lead Rottet (1992, 274) to propose the underlying structure (72).

(72) [TP [MoodP [NegP [AspP [VP]]]]]

To accommodate the fact that negation appears pre- and postverbally affecting the tense and/or mood interpretation as well as the complementary distribution of TMA markers and short verb forms, Rottet (1992) proposes a verb movement analysis for mesolectal LC, namely the Teche variety documented by Neumann (1985). He argues that the short stem can appear before *pa* because it moves from V^0 to TP where it precedes NegP. Occupying the TP head position, the short verb form consequently blocks the appearance of an unbound tense maker. At the same time the presence of any intervening non-affixal mood or aspect markers, as heads of their projections, blocks successive cyclic movement through the lower projections, hence preventing verb movement. He assumes a zero affix that triggers the movement process and causes the verb to appear in its short form. Despite a few weaknesses, such as considering *pa* a negative adverb in SpecNegP rather than the head Neg^0 to account for verb movement over NegP[26], the analysis accounts for the present complexities and distinguishes basilectal Creoles that do not exhibit verb movement from mesolectal or decreolizing ones that may show such behavior. This distinction fits well with the fact that neither the short form nor post-verbal *pa* are attested in the 19th century texts, which are assumed to represent basilectal LC (Rottet 1992). For the present thesis, the verb movement account also serves to show that syntactic movement operations are possible in Creole languages in general and LC specifically.

Another movement operation, that is more commonly attested for Creoles, is focus fronting. Bickerton (2016, [1981]) postulates this as a universal Creole feature.

[26] Klingler (2003a, 320) considers *pa* a negative adverb as well.

Usually, linguists distinguish between focus fronting and cleft constructions, though both technically can be analyzed as involving syntactic movement. Clefts typically contain a copula and the background clause often is relativized (cf. (73)) while focus fronting does not employ either strategy (cf. (74)) though it is attested with focus particles in languages that employ such markers.

(73) Cleft Construction: It is Paul who we talked about.

(74) Focus Fronting: Paul, we talked about.

Creole languages tend to use cleft constructions either as alternative or preferred constructions. Of the language listed in APICS 92% use a cleft construction either including a copula, a focus marker or no marker and 35% use fronting of the focused noun or noun phrase which usually also includes a focus marker (Maurer and Consortium 2013). The analysis of LC in this respect is complicated by homophony or unclear grammatical status of some elements. For example, Clements and Mahboob (2000, 479) attest that LC does not exhibit a focus or question particle *ki*, probably due to the presence of *ki*-final question words. Rottet (2006), on the other hand, analyzes the same *ki* as a focus particle used on interrogative subjects. In a similar manner, *se* can function as a copula in LC but also as an emphatic particle. Consequently, a construction such as example ((75)) could be interpreted as either a cleft sentence or a case of focus fronting depending on the analysis of *se*.

(75) Se Paul nou pale pou.
COP/FOC Paul 1PL talk about.
(It is) Paul we talked about.

Regardless, it is uncontroversial that LC has a means of focusing which involves the left periphery rather than in situ focusing for example by means of a focus particle. This is another instance of syntactic movement in the language and one that is closely linked to the analysis of wh-questions which is addressed in this study.

4.4.4 The lexicon

The lexicon of LC is only marginally relevant to the present study. For the sake of completeness, it should be mentioned that etymological research can be of great benefit to determine LC genesis and the historical connections to other languages, Creoles and French varieties as well as substrates. LC appears to resemble other French-based Creoles in that most of the lexicon is derived from old forms of French. Neumann-Holzschuh (2015) identifies lexical items that occur in Louisiana as well as Acadia but not Quebec and, vice versa, in Louisiana and Quebec but not Acadia. She further points out many items LC shares with LRF and several words that originate from either African or

Native American substrates, concluding that the LC lexicon reflects the diverse history of linguistic contact in the region.

5 METHODOLOGY

The data for this study was collected during two fieldtrips. These fieldwork phases took place for eight weeks in the summer of 2014 and five weeks in spring 2015. Materials were prepared beforehand and adapted during fieldwork. In this chapter, the methods used in the field and for the analysis are described.

5.1 Field methods

All participants signed a consent form (Appendix A), which was developed for this study and presented to them before any recordings were made. The form gave them several choices regarding data use and anonymity. The intended purpose(s) and all options were also presented orally to all interviewees in order to avoid misunderstandings and potential disadvantages due to illiteracy.

The goal was to collect two main bodies of data: sociolinguistic and language/ linguistic data. The sociolinguistic data was obtained through semi-structured interviews which were held in English and followed a SOCIOLINGUISTIC QUESTIONNAIRE which was inspired by the work of Rottet (2001), Melancon (2000) and Graham (1981). This questionnaire was designed to collect biographical information, as well as linguistic biographies as listed in Table 5-1.

Table 5-1 Biographic/linguistic biographic information collected in the sociolinguistic questionnaire

Biographical Information	Linguistic Biography
1. Name	1. Languages spoken
2. Sex	2. Languages understood
3. Year of Birth	3. First language(s)
4. Hometown	4. Formal French education (& duration)
5. Place of residence (& duration)	5. Best language (most comfortable, at home)
6. Other places lived (& duration)	6. Language spoken most
7. Occupation	7. Best language grandparents
8. Marital status	8. Best language parents
9. Education	9. Languages spoken/understood by children
10. Ethnicity (self-assessment)	10. Best language children
11. (number and age of) children	11. Languages spoken/understood by grand-children
12. (number and age of) grandchildren	12. Best language grandchildren

Further, participants were asked to self-assess their language proficiency through stating their ability to perform each of ten tasks requiring increasing proficiency in LC according to Melancon (2000). Participants were then asked to translate "I am going to go." as well as "I went to the store." into LC. These sentences translate distinctively into LC and LRF as pointed out by Graham (1981, 21) for the former. In LC the first translation is: "Mo (ape/pe/e) kouri." while speakers of LRF would say "Je va aller." or "Je vais aller." if they tend towards Standard French. The second clause can be translated as "Mo kouri/gòn o grosri" as opposed to the LRF "Je viens d'aller au grocery". This serves to identify speakers who claim to speak LC but whose language is closer to LRF linguistically, despite their self-identification[27]. It was expected that creolophones would use the pronoun *mo* rather than *je*, isolating TMA markers such as *ape/pe, a* or *te* rather than inflectional endings and the typical LC verb *kouri* or the English borrowing *gòn* rather than the LRF form *aller*. The third part of the questionnaire is concerned with language use and is largely modeled after the relevant part of Rottet (2001) questionnaire although Klingler (2003b) and Mayeux (2019) use similar criteria. The development of language use is first addressed on the basis of interlocutors in the past and present. For different interlocutors, participants were asked which language they used, first as children then currently. Speakers indicated which language they spoke as children with their parents, grandparents, siblings, friends and classmates in school as well as friends outside of school. The question was to be answered with: "LC, mostly LC, LC/English, mostly English, English". The same question was then asked for current interlocutors, such as parents, siblings, spouse, children, grandchildren, friends/neighbors, co-workers or shopkeepers. Similarly, fifteen different usage contexts (e.g. "I speak LC on the street for a chat", "I speak LC in my home", "I dream in LC") were then given and participants were asked to state how often they use Louisiana Creole in such situations. Options were: "Always, often, sometimes, rarely, never". This indicates domains of language use. Finally, the questionnaire contained several open questions concerning language attitude, such as "In your opinion, should people try to keep Louisiana Creole alive? Why?" or "Do you feel that speaking Louisiana Creole is important to who you are?".

The questionnaire was designed this way for several reasons. First, to include such an extensive sociolinguistic questionnaire in a largely syntactic study at all seemed important because LC is an endangered language. As such, establishing some familiarity between interviewer and interviewee before asking for acceptability judgments provided

[27] Please see Chapter 4.3 and Klingler (2003b) for more information on identity and language labels in Louisiana.

a methodological advantage, as it was expected to reduce nervousness and insecurity. More importantly, the data collected here is valuable in its own right as it allows some insight into speakers' attitudes, usage patterns and linguistic biographies. Given the current level of endangerment of LC, conducting fieldwork and neglecting such information would have been a large oversight. Furthermore, connections between the demographic and sociolinguistic data obtained here and the linguistic data from the second part of the study arise during the data analysis in Chapter 7, which is valuable to the discussion and allows insights into the effects sociolinguistic factors have on linguistic structure.

Moreover, the questionnaire covers several areas of interest, often addressing the same notions from different angles. For example, domains of language use cannot be fully set apart from the respective interlocutors, questions concerning interlocutors during childhood are necessarily connected to parts of the linguistic biography, such as questions concerning the language(s) spoken by parents and grandparents and which language was generally used at home. Each part offers additional and new information, which can be related to other questions to create a complex profile rather than singular context-free glimpses into sociolinguistic behavior. Of course, relations between demographic and sociolinguistic data can in principle be established, indicating, for example, if certain language attitudes or usage patterns are typical of age or ethnic groups, or of creolophones who grew up monolingually and only learned English later in life. While the collected information offers many opportunities for interesting analyses, measures of statistical significance cannot be obtained due to the sample size. This is an unfortunate but hardly avoidable consequence of studying aspects of an endangered language in depth, which is further addressed below.

When it comes to linguistic analysis, SPONTANEOUS SPEECH is often considered the most salient data source for linguistic features of which speakers are unaware. Unguided, casual speech has been the focus of much sociolinguistic research since William Labov (Labov 2006 [1966] and subsequent work). In this spirit, recording natural speech was one goal that was established before fieldwork. In the field, two main factors made conducting such recordings difficult. First, the target structures were wh-questions. Naturally, these only arise during conversation and are seldom found in narrations or similar modes that require just one speaker. Additionally, complex wh-questions of e.g. the long-distance variety are rare in natural discourse altogether and are usually elicited which is why Dąbrowska (2010) uses this sentence type to test the adequacy of linguists' and native speakers' judgments. Secondly, the structure of the speech community

described in Chapter 4 does not offer many opportunities to record conversations between LC speakers. For the majority of interviews only one speaker was present. Family members often were in the room but not fluent enough in LC to hold a conversation if they had any knowledge of the language at all. The default for casual conversation was almost universally English. As a consequence, the amount of spontaneous data obtained during fieldwork (ca. 2h) is sufficient, especially considering the study's focus on syntactic structures. Complex syntactic structures or marginal, marked (in the Chomskyan sense) language features are rarely found in spontaneous speech which is why a variety of linguistic research methods are designed to target such structures deliberately. During fieldwork, two of these methods were employed, namely GRAMMATICALITY/ACCEPTABILITY JUDGMENTS and QUESTIONS AFTER STORIES.

The main body of linguistic data was collected through ACCEPTABILITY JUDGMENTS. Acceptability judgments are commonly used in research on syntax, where it is argued that they reflect the speakers' linguistic knowledge. Culicover (2014, 266) calls this the "classical approach to discovery of the grammar". The idea is to capture native speakers' "reported perception of acceptability" (Schütze and Sprouse 2013, 28) of particular structures by presenting them with example sentences, which are then judged (un)acceptable. Many researchers include a scale to identify degrees of well-formedness. Like most methodologies, acceptability judgments have their shortcomings, some of which are addressed below, but their considerable advantage is that they give insight into possible linguistic structures that do not occur reliably in natural speech (Schütze and Sprouse 2013) and which the speakers cannot consciously describe or articulate, as they are unaware of these rules governing their grammar. By virtue of their passivity they are also free of production errors, which might result in false conclusions, and they allow a focus on structural features that is hard to obtain in production tasks (Tremblay 2005). Further, the occurrence of a certain structure in a linguistic corpus does not necessarily guarantee its grammaticality, as it may be judged unacceptable even by the speaker who produced it due to production errors (Schütze and Sprouse 2013).

One critical point is that all of the above properties have led to a common association of grammaticality judgments with linguistic competence as is described in Schütze (2016). To assume that judgments offer direct insight into a speaker's competence is, however, a misconception: modern linguistics has never claimed to access the underlying system but solely its output (Schütze and Sprouse 2013). This can be seen in Chomsky's early work where he argues for a distinction between grammaticality and acceptability (Chomsky 1965, Miller and Chomsky 1963). This is why *acceptability* rather

than *grammaticality* is used to refer to the judgment task here, as the results represent observable behavior reflecting properties of mental faculties rather than direct observations of those systems (Schütze 2016).

Acceptability judgments have further been criticized as being unreliable as they are based on intuitions, which have no scientific value. Sampson (2007) even claims that speakers are not in fact capable of judging acceptability in the sense that they reject sentences which do not conform to their mental grammar. Rather, he argues, they simply reject sentences or structures, which are unfamiliar to them and would likely change their assessment with continued exposure. While such reasoning is not false *per se*, it seems to simply describe language change, on the one the hand, which necessarily requires the grammar of a language to adapt over time, and, on the other hand, I-grammars, namely the more or less subtle differences in the grammars of individual speakers of the same language. For this reason, the present study does not follow Sampson's recommendation to avoid judgment tasks altogether but relies on judgments made by multiple speakers instead, to avoid the possible bias inherent to introspection which is no longer considered methodologically appropriate. A related issue is that acceptability judgments are to some extent artificial and require metalinguistic awareness which can be considered problematic. If so, any reported data must be seen critically as it differs from directly observable behavior that is not mediated through a conscious decision on the participant's part. Such concerns seem to be largely unfounded as reported perceptions have been widely used for the construction of theories in several fields (Schütze and Sprouse 2013). Tremblay (2005, 134) points out that grammaticality judgments remain "influenced by performance factors" since competence remains unobservable. According to her, results can still be influenced by extra-linguistic factors, such as the mode and context of presentation,[28] as well as extra-grammatical factors, such as prosodic and pragmatic factors.

To counteract such effects as much as possible a number of measures were taken. To address the extra-linguistic influences, the timeframe for the interviews was reduced whenever possible by meeting interviewees twice or more. This also had the side effect of creating a certain level of familiarity with myself as the interviewer, likely resulting in more honest responses even when they appeared to go against my expectations. This proved to be a concern, as speakers would initially accept sentences they later rejected, most likely in an effort to be polite. To achieve a less formal setting for this rather

[28] Other conceivable factors would be the speaker's state of mind, concentration, number of sentences already presented, or linguistic flexibility, i.e. the willingness to accept or produce non-standard constructions.

formal[29] mode of data collection, interviews were always held in a location of the speakers' choice, most often their homes, workplace or somewhere else within their comfort zone. This was also an ethical imperative, as many interviewees were elderly and I had to be mindful of their health and state of mind. Interviews were stopped or postponed whenever participants seemed to overextend themselves or appeared uncomfortable with the situation. Since LC is a predominately oral language, effects arising from oral presentation of stimuli sentences were impossible to avoid fully.

In order to ensure that the list would not contain any lexical, pragmatic or grammatical errors that would lead speakers to reject an example on grounds other than the syntactic construction, the LC sentence list had to be devised with the help of a native speaker. But, in the initial stages of fieldwork, no willing co-researcher could be found. Since field time was limited, interviews had to commence nonetheless and the methodology needed to remain consistent throughout the study to allow for comparative analysis. This left a list of English sentences and the need for methodological adjustments.

The set of sentences to be judged consisted of 75 sentences, including 12 distractors. These contained a range of constructions to be tested. The structures included the following main types. Root questions (76), (77) and finite as well as non-finite embedded interrogatives (78)(79), including distinctions between subject, object and adjunct questions, multiple wh-questions (80) with different fronting patterns, (multiple) long-distance wh-questions (81), (82) and partial wh-fronting questions (83), which are ungrammatical in English,

(76) Who arrived?
(77) Who did John hit?
(78) I know who did this.
(79) What do you refuse to tell him?
(80) Who saw what when?
(81) Who do you think did this?
(82) Who did you say saw what when?
(83) *What do you say who hit whom?

The most obvious step was to present speakers with the English sentences and ask for translations while attempting to deduce which were acceptable in LC. Such an approach must be employed very carefully, as simple direct elicitation, i.e. translation,

[29] Please note that formality is here taken as a characteristic of the situation and a feature assigned to the judgment method as compared to other methods employed in linguistic research. It does not refer to the formal/informal distinction as applied in experimental cognitive science, where the field methodology used for this study would be considered semi-formal at most.

has several drawbacks that are widely addressed in the field of language documentation (Frawley 2003, Hellwig 2010). The largest concern is translatability. Translations can only be meaningful where concepts and structures are similar in the two languages and cultures involved. The larger the differences are between the language under investigation and the working language used for translation, the more effort and reflection is required on the consultant's part to provide an equivalent sentence. The task becomes even more cumbersome when the items to be translated are marginal or even unacceptable in the working language. Such metalinguistic undertakings are often foreign to consultants, who show different levels of inclination towards them. While these tasks are easily performed by some non-linguists, they remain confusing and bothersome to most. Translatability often pertains to semantic differences but the main issue for this study was that consultants might produce, accept or reject syntactic structures on the basis of English grammar rather than LC, which can reasonably be expected to be different. Another concern was put forward by Duranti (1981), who stresses that the interview situation remains a social, communicative event which will shape the linguistic output and result in translations in a formal style, counter to the intention of studying vernaculars. For the present study this was merely something to be aware of since (a) the focus was on the range of possible syntactic constructions where style variation is less relevant than in phonology or lexicology, and (b) LC appears to be a language of the home, mainly restricted to private domains and consequently is not very formal in any context which is to be expected as (c) dying languages have fewer styles according to Rottet (2001) among others. Nevertheless, if a continuum approach is taken, the elicited data may possibly reflect a more acrolectal variety of LC than other means might have yielded.

The main difficulty hence remained to obtain meaningful acceptability judgments. To ask interviewees to translate a construction like the multiple fronting wh-question in (84) which is ungrammatical in English in the first place, and then ask their opinion on the resulting LC sentence seemed methodologically unsound for obvious reasons. It was expected that speakers would struggle with producing sentences that are ungrammatical in LC and would quickly tire of being asked to do so. Likely, consultants would feel pressured when continuously presented with complex constructions that are questionable or ungrammatical in English, which they should translate into LC, a language that is commonly perceived as 'simple' by its speakers.

(84) *Who what when saw?

To avoid such effects, the MANIPULATED TRANSLATION ACCEPTABILITY TASK (MTAT) was developed to fit this fieldwork situation. This design requires flexibility from the interviewer but yields meaningful acceptability judgments on a large number of constructions. For each type of wh-question an English sentence was chosen as a base-line, the direct translation of which was least likely to be unacceptable in LC on the basis of what has previously been documented about LC question formation (for example in: Neumann 1985, Klingler 2003a, Klingler and Neumann-Holzschuh 2013) and the behavior of typologically similar languages such as other Creoles and French (for creole typology and the relatively close resemblance of LC to French see Bakker et al. 2017) but also what is possible in English as the working language, because asking participants to translate an ungrammatical sentence seems ill-advised. Consultants are then asked to translate this example. The interviewer in turn notes down their reply and follows up with a question that begins with "Could I also say…". These contain the same sentence in the interviewee's choice of words but with different word orders or slight syntactic modifications. For example, the base line would be a subject question with a fronted wh-expression such as (85), which would likely be translated as indicated. I would then follow up with (86):

(85) Who arrived? – Ki arive?

(86) To mean the same thing, could I also say "Arive ki?"

If the answer was positive, follow-up questions were asked such as

(87) Which one would you usually say? Would you use this? Have you heard others say it like this?

These questions served to determine the degree of acceptability on a Likert-type scale, where 1 means "fully acceptable", 2 means "okay but slightly strange", 3 stands for "maybe acceptable", 4 means "I would not use it, others might", and 5 indicates full unacceptability. They could also confirm if the base sentence truly was the most unmarked option. In several cases these follow-up questions were needed in order to receive a clear answer, as consultants tended to allow too much leeway by accepting a highly marked sentence rather than indicating the unacceptability. This tendency was likely caused by the fact that I often presented myself as an interested learner of LC in order to make the process appear less constructed. To encourage interviewees to participate and allow them to be more comfortable with the situation this was a very useful strategy. A negative side effect was that they would be more tolerant of marginal constructions which was counteracted as described above.

Before the first fieldtrip, it was assumed that acceptability judgments might have their limitations in the area of multiple long-distance wh-questions. This was borne out as consultants tended to find them confusing and often did not provide a translation and/or state an opinion on their acceptability. The main body of judgment data was hence complimented by a task that was originally designed to research the availability of long-distance wh-constructions in child language by de Villiers and Roeper (1996) who refer to the method as QUESTIONS AFTER STORIES. 16 creolophones participated in this elicitation task. Consultants were presented with a few pictures taken from de Villiers and Roeper (1996) which formed a brief story containing two events (see appendix B). The interviewer told the story in English with the help of the pictures. Consultants then retold the story in LC, providing the vocabulary they were most likely to use and an additional corpus that bears no relevance to the study of wh-questions but may still be useful for future research. This step also served to make sure that the story was clear and was remembered fully. The interviewees were then asked an ambiguous long-distance question as given in example (88).

(88) Ekan li di so Papa komen li fe mal a limèm?
When PRN say POS Papa how PRN make pain to PRN REFL

When did he tell his father how he hurt himself?

The consultants' answers would indicate whether a long- or short distance reading of *ekan/when* was preferred in the presence of an intervening wh-expression in the embedded question as either the time of telling (event 2) or the time of hurting (event 1) could be given as a reply. To extend this data set, during the second field trip the process was repeated with a second story build from the cards provided by Sardinha (2011) in which a lady realizes that her neighbor was in an accident and is in the hospital. She asks the neighbor's mother if she should retrieve his mail and take care of his plants.

(89) Ki li monde komen ede?
Who PRN ask how help

Who did she ask how to help?

The ambiguous question in (89) again allows for two readings. In the short-distance interpretation *ki* ("who") refers to the mother as the object of the verb *monde* ("ask") while the long-distance reading allows the object of the verb *ede* ("help"), namely the neighbor as the answer. In contrast to (88) this is an object question rather than an adjunct question. This may have implications for extraction possibilities.

5.2 Consultants, analysis and the qualitative approach

Data for the present study was collected during two fieldtrips to Louisiana. The fieldwork for this study concentrated on the Bayou Teche and False River regions where chances of meeting and interviewing creolophones were expected to be highest. There are isolated reports of Louisiana Creole being spoken on the German Cost and in the Lake Charles area (Neumann 1985, p.c. Clifton 2014). These are not confirmed in the recent literature and no speakers could be found in these areas during the course of this study. This does not necessarily mean that the language is no longer used there, as it is a language rarely spoken in public and fieldwork did not concentrate on these areas. Nevertheless, considering the overall state of LC, it appears unlikely that unknown large LC-speaking communities exist. No clear statement can be made with regard to families, social groupings or households that may still be using LC on a regular basis unnoticed by the general populace as well as linguistic fieldworkers. If so, they could not be won for participation in the present study.

In total, 22 consultants are considered in the study, some of which completed the study design only in part. Multiple conversations were recorded with speakers that did not result in an interview to be used comparatively, for a number of reasons. For example, these discussions often covered small parts of the questionnaire but were not concluded for reasons such as interviewee fatigue, time constraints, obvious discomfort of the interviewee etc. These are not included in the consultant figures but still inform the overall picture. Two data sets are disregarded despite completion of all tasks. One was provided by the only speaker who is aged less than 45 years and belongs to a group of language activists led by Christophe Landry. He is a new speaker according to speaker typologies such as Grinevald and Bert (2011) and discussed more concretely for the Louisiana case by Mayeux (2019). To prevent skew, this data is not included in the analysis. Another interviewee claimed to speak LC but her speech was heavily influenced by LRF to the point where it was unclear which language she was using. In the hopes of adding interviews with similar speakers, both data sets were recorded as potential points of comparison but are ultimately excluded from the analysis. In sum, 22 sociolinguistic interviews, 17 manipulated translation acceptability tasks and 14 Questions after stories from 22 interviewees remain included for analysis. Additionally, approximately 2 hours of spontaneous data were recorded. The data set is hence sufficiently large for the purposes of this study. While a larger number of participants in principle improves any research, the achieved number of interviews is actually higher than expected beforehand,

considering the level of language endangerment. Commonly, studies on endangered languages have to rely on a handful of speakers.

Due to the community structure and the low status of LC discussed in Chapter 4 identifying creolophones and gaining them as consultants proved to be a major task during fieldwork. My status as an outsider, a young white researcher from Europe without any community ties or competence in the language, likely was an impeding factor. As a starting point, a couple of speakers as well as a few locals actively involved in French maintenance in the region were contacted, many of whom were kindly referred by Thomas Klingler. Beyond that initial referral, fieldwork was conducted without local institutional support, although several researchers from local universities and institutions were consulted during the fieldtrips and provided valuable insights and suggestions. Most contacts were made through immersion in the local Creole culture, by attending French tables, Cajun and Zydeco dances, cultural and community events and reaching out to any community member recommended by someone at one of those events. Some speakers would suggest others who might be interested in participating. Ultimately, the most helpful contacts in terms of introductions to speakers were the social circles around NUNU Art and Culture Collective in Arnaudville, the secretary of a seafood company in Henderson and Dr. Angelique Bergeron in the False River area, among many others. Since funding was very limited, participants could not be reimbursed for their time, which can be inappropriate depending on the cultural context anyway (Adone 2008 Adone p.c.). Considering that most of the consultants are employed despite their advanced ages (see below) or highly involved with taking care of (grand-) children and family members, their free time is rare and precious, which was an additional consideration in scheduling interviews.

Interviews were generally conducted one-on-one, although a few of the recordings were pair interviews, which sometimes yields just one answer for two speakers. It is also possible that in some of those cases one person's point of view influenced the other consultant's reply. For the sociolinguistic questionnaire these are treated as two separate data sets where possible since social and linguistic backgrounds are distinct. This is done despite the possibility that participants may influence each other because many questions relate to information that differs from speaker to speaker and these are unlikely to cause social pressure to reply in a certain way. Only the questions on language attitudes should be considered carefully in this regard. Nevertheless, this proved no cause for major concern as the overall picture in that section appeared to be quite homogeneous. For the acceptability judgment task and the picture

stories pair-interviews are considered one data set as answers did not diverge in those instances.

The overall number of participants as well as the amount of structures under consideration does not allow for a meaningful quantitative analysis in the sense of inferential statistics. While some descriptive statistics is employed for example to depict the acceptance rates for grammatical structures, no claims are made about statistical significance of tendencies and preferences or their statistic correlations to demographic and sociolinguistic factors. The size and structure of the present data set does not allow for such an analysis[30], nor was this the intention of this study. Because the research questions for this study are non-comparative and concerned with LC's syntactic structure, a qualitative approach was expected to yield more insights than a purely quantitative analysis, especially considering the endangered status of the language and the related linguistic insecurities among the speakers (Blainey 2017). There certainly is a lot of merit to quantitative studies but they categorize and generalize by nature, breaking down the presence or absence of observable differences and effects to *P* values. This categorical application of statistics is currently under debate even in the natural sciences (e.g. Hurlbert, Levine, and Utts 2019, Wasserstein and Lazar 2016). The present study seeks to identify subtler and possibly even individual differences and to describe one specific aspect of the linguistic system within the current linguistic ecology. From this point of view, a qualitative analysis is the better fit. It is only in relating social variables (such as age, place of birth or linguistic background) to linguistic variables (such as the acceptance of multiple wh-fronting or long-distance readings or the preference for preposition stranding) that the qualitative analysis may fail to make claims of equal reliability to a quantitative-oriented approach. This study aims to provide a mixed approach in the sense of Angouri (2010) by combining judgment, experimental and interview data, from the maximal number of participants that could be achieved in the given frame. The sample size and interview types are more in line with a qualitative research tradition but acceptability judgments are also often used in quantitative approaches. The present study meets a number of criteria applied for formal research methodology, as opposed to informal, and therefore possibly less reliable, applications of research methods. This includes the number of speakers, which is larger than ten, the reliance on naïve speakers rather than linguists and the use of a labeled scale instead of binary response options (Schütze and Sprouse 2013). Despite these efforts towards

[30] Please note that Sprouse, Schütze, and Almeida (2013) predict a scaled judgment task as employed in this study to yield statistical power (80% likelihood of correctly predicting a difference) for an empirical coverage of 80% for a sample size of 35 participants. In comparison, a simple Yes/No task reaches the same reliability with only 15 participants.

quantitative methodology, the data set is not large enough to achieve statistical power, the usefulness of which for the present study is debatable as described above. Nevertheless, it allows for an analysis and a comparison between speakers and groups of speakers according to social or sociolinguistic variables. In this way, tendencies can be documented and such factors are considered and discussed with a focus on the variables which Rottet (2001) identifies as most relevant in the Louisiana context. Although his work is concerned with LRF in a partly Native American community, there are many similarities in terms of the historical setting and the endangered status of the language. He finds ethnicity, sex and age to be the crucial social variables, more so than residence, socioeconomic class or occupation. Following this approach, all noticeable linguistic or sociolinguistic features are presented with reference to their occurrence or acceptance within each racial, age and gender group despite the lack of statistical significance. Race especially can be expected to be relevant for linguistic behavior as several studies of LC report (Klingler 2003a, Neumann 1985, Mayeux 2019) and age is always an important factor in endangered languages.

This thesis, in the tradition of much work in modern sociology and Critical Race Theory (Delgado and Stefancic 1998), and counter the pervasive scientific racism that is to be found in many early works on Louisiana and any other colony, considers race as socially constructed category. In this view, racial distinctions rely on arbitrary characteristics such as skin color that are perceived to be meaningful by society which results in the formation of social groups whose circumstances, experiences and cultures differ in real and systematic ways (Gallagher 2004). Race is a socially and culturally essential concept in the U.S. and the resulting inequalities are still very much in effect (Bhopal 2018) especially in the rural parts of Southern states such as the area where this study was conducted. The binary distinction between black and white speakers used in this thesis is of course a stark simplification of the diversity found in Louisiana. In no way does this reflect any intention to abide by or promote racist notions, nor does it indicate that all people of color are associated with a black identity. Nevertheless, it reflects the divide that has shaped social structure and the life experiences of the people interviewed for this study most of whom attended segregated schools as children. Race is deliberately chosen as an analytical category over the related concept of ethnicity, because ethnicity as a cultural and linguistic distinction is a very complex issue in South Louisiana. For example, many black consultants would not self-identify as African Americans since they do not consider themselves Africans, nor are they simply Americans. When asked for their ethnicity, many consultants had to put considerable thought into the issue and

replies like “a black French Louisiana man” or “a black Creole American” were not uncommon.

Returning to the relevance of socio-demographic factors for LC, Klingler and Neumann-Holzschuh (2013) observe that the speech of white creolophones resembles LRF more closely and thus represents a more acrolectal and/or decreolized variety. Age, on the other hand, is likely an important factor because of the level of language endangerment. The present analysis additionally takes hometowns into account, as interviews were conducted in two different areas and regional variation is to be expected (Klingler and Neumann-Holzschuh 2013). Further, variables which are not listed by Rottet (2001) but become apparent from the sociolinguistic interviews include L1, proficiency level in LC and knowledge of LRF/SF. I also follow Rottet’s approach in comparing the results between age groups, as far as is possible given the low number of young speakers.

One aim of this study is to detect changes in the grammatical system as well as the sociolinguistics of LC and to determine the cause of this change if possible. A necessary prerequisite for the identification and description of linguistic change is knowledge of the previous linguistic norm to which the current data is compared (Rottet 2001). In the case of endangered languages that oftentimes poses significant difficulties, as they are commonly previously undocumented. Luckily, in the case of LC, excellent documentation exists in Neumann’s seminal work from 1985. Unfortunately, this work is recent enough to only allow for limited conclusions as a time span of thirty years is rarely sufficient to document language change. Since changes in endangered languages are expected to proceed at a much faster pace than usual (Crystal 2000), Neumann’s grammar still remains a valuable resource for comparison. Further, historical sources are considered, mainly the collection of early LC texts also provided Neumann-Holzschuh (1987). It must be noted that neither source is primarily concerned with wh-questions. Consequently, no points of comparison can be expected for the more marginal or rarely used construction types. In these instances, an analysis of what is possible in the current grammar is provided and the oldest fluent speakers are considered as a possible base line of the norm as suggested by Rottet (2001). In addition to historical and field data examples from published works and the LOUISIANA CREOLE DIACHRONIC CORPUS (LCDC)[31] are included where suitable.

[31] I am very grateful to Oliver Mayeux for making the Louisiana Creole Diachronic Corpus (LCDC) database available for this study.

5.3 Notation

No official writing or spelling system exists for Louisiana Creole. Few speakers interviewed for this study claimed to write this largely oral language and many were surprised to see the interviewer writing down their LC utterances. Writing LC has consequently always been a mainly, though not exclusively, academic undertaking. LC speech was historically also written in court protocols (Hall 1992) and literary works (e.g. Cable, Faulkner, Mercier etc.). Different publications employed different systems. Some are more closely aligned with French orthographic conventions; others take a phonemic approach (Klingler 1996, Neumann 1985). With the emergence of Creole orthographies for more firmly established Creole languages the reliance on French spelling became less common, as it was often perceived not to sufficiently acknowledge distinct Creole identity and culture.

Neumann (1985) employs a largely IPA-based notational format. Her spelling makes use of a mix of graphemes and IPA symbols in order to capture the phonological representation and variation of LC while maintaining a certain level of standardization for practical purposes. For example, Neumann uses an IPA symbol to indicate nasalization of a vowel, such as *ẽ* for [ɛ̃]. Later approaches chose to indicate nasalization by adding the letter 'n', thus eliminating the need for an IPA keyboard but adding the difficulty to distinguish between pronounced and unpronounced nasal consonants.

Landry et al. (2016) published the most recent and least academia-oriented orthography guide. Their suggestion "abandoned use of French, English, and Haitian orthographies all together to craft a writing system that is both familiar and accessible to Americanized Louisiana Creoles" (Landry et al. 2016, 1). They primarily aim to provide guidelines which allow community members to write their language in order to strengthen LC. This is a system developed by the new generation of heritage speakers, who are community members trying to promote and strengthen LC. The system they provide is unique to LC, remains as simple as possible and relies on characters and rules which are familiar to LC speakers from either American English or French spellings that can be found throughout Louisiana. Some of their choices may not be ideal from a linguist's perspective as some phonemic details are necessarily omitted or artificial distinction maintained for the sake of familiarity but they focus on distinctiveness for the purpose of language maintenance as well as community usability.

Klingler (1996) argues convincingly for a phonemic spelling system. His approach is specifically designed to account for the pronunciation of LC, which serves to set it apart from LRF and other French varieties. From a linguistic perspective there is clear merit to

such an approach. Unfortunately, it never became established among LC community members. In the *Dictionary of Louisiana Creole* (Valdman 1998) his orthography is implemented. In the more recent *Dictionary of Louisiana French* (Valdman and Rottet 2010), the editors compromised, as a phonetically based spelling system was largely rejected by the non-linguists on the project (Klingler p.c.). They instead opt to make reading the words as easy as possible for French speakers, from Louisiana as well as any other nationality. Consequently, their notations are very close to International French conventions with the exception of local variants that are given as alternatives. It must be noted though that this dictionary refers to all forms of French spoken in Louisiana, which means LRF, rather than LC, primarily, although LC words are included.

For his 2003 monograph, Klingler adopts his earlier notation that is "very close to that of the Institut Pédagogique National, which serves as the semiofficial spelling of Haitian Creole" (Klingler 2003a, 144). It remains largely but not fully phonemic, thus balancing readability and practicality with linguistic accuracy and phonetic detail. From a linguistic perspective, this appears to be the best practice, which is why the notations in here are based on this system. Occasionally, I further simplify the spelling where it reflects phonemic variation that is not significant for the present analysis. For a detailed description of the graphemic representation of phonemes and allophones please see Klingler (2003a, 143-153). For the purpose of this thesis, the allocation of word boundaries is the most relevant, although the spelling of nasal sounds for example cannot be circumvented in any writing of LC. Klingler assumes agglutination of prenominal elements such as *l-* or *di-* even in cases where their appearance is optional. Consequently, these are spelled as part of the noun, as in *latet* ("head"). He further hyphenates compounds only if the elements occur together very frequently and are certain to form a lexical unit. This includes *sa-ki* ('what'). I follow Klingler's notation and extend it to any pronominal wh-expressions. Nevertheless, the status of *ki* and *sa* is to be determined in the present discussion and such spelling does not indicate any conclusions of the morphosyntactic analysis. Nasal vowels are indicated by adding the grapheme *n* when followed by a consonant or word final. The digraphs *an*, *en*, *on* and *œn* hence represent the nasal vowels of LC. Only in intervocalic or initial position does V+*n* not indicate nasalization but an oral vowel plus the nasal consonant. A nasal vowel before [n] is spelled as the digraph V+*n* followed by another grapheme *n*.

6 SOCIOLINGUISTIC FINDINGS

This chapter aims to identify speakers in terms of social variables (Mesthrie et al. 2000) and to describe their linguistic background before analyzing patterns of language use and levels of language proficiency. This is important to better understand the current linguistic situation in South Louisiana and serves to provide a basis for the linguistic study. Getting a clearer picture of who speaks LC today constitutes a first step in grasping reasons for the ongoing language shift, the level of endangerment of LC as well as the role the language plays in multilingual speakers' identities and its role in local communities. As a second step, language attitudes found among the speakers and their self-reported views on the status of LC are discussed.[32] The data presented here is related to the linguistic data during the following discussion.

6.1 Demographics

Speakers consulted for this study are demographically structured as follows. The majority of consultants were above the age of 75 at the time of the interview, while only six speakers were younger than 60 years.

Table 6-1 Age of participants

age	
46-60	6
61-75	6
75+	10
Total	**22**

The way interviewees are arranged according to age here allows for a few relevant distinctions. First, the oldest speakers were born before World War II, in the year 1940 or earlier. The War is commonly considered a turning point for the cultural and linguistic developments in Louisiana (Rottet 2001). The year of birth for speakers in the second oldest category is between 1941 and 1954/55, which means they were born during World War II or within the years directly after the war in which social effects are said to have taken root. Speakers born and raised after World War II fall into the youngest category. This is reflected in the intervals chosen here while the arrangement simultaneously maintains an even spacing of age groups in steps of 15 years.

[32] Parts of the data discussed in this chapter were presented at the SPCL winter meeting in conjunction with the LSA in Washington D.C. 2016 as well as the CELC Postgraduate Workshop (July 2016). I am grateful to the audiences for helpful comments and critical questions.

Table 6-2 Gender, ethnicity and place of residence

	f	m	Total
black	**5**	**8**	**13**
Bayou Teche	4	6	10
False River	1	2	3
white	**7**	**2**	**9**
Bayou Teche	7	1	8
False River		1	1
Total	**12**	**10**	**22**

As is indicated in Table 6-2, gender parity among the consultants was not fully reached but with 12 female and 10 male participants the distribution is relatively balanced. 13 participants are black and 9 speakers are white. Despite the fact that white speakers of LC make up 40,91% of the sample as opposed to 59,09% black participants a slight bias exists towards white speakers. It is unclear what the de facto numerical relation between speakers of the two ethnicities is, but recall that Neumann (1985) estimates 80% black and 20% white speakers. Despite the larger number of black consultants, it follows that white interviewees are overrepresented in this study. The only way to avoid this under field conditions would have been to forgo opportunities for interviews simply because the speakers were white or to exclude their data from the analysis. Neither option is chosen here; instead race is considered a possible social factor that may correlate to linguistic behavior as is discussed above.

In sum, most interviewees consulted for this study were from the grandparent or great-grandparent generation. 16 out of 22 participants were above the age of 60 at the time of the interview. Creolophones from both ethnicities as well as genders were included. The least represented demographic category is, somewhat ironically considering the traditional overrepresentation of NORMs[33] (Chambers and Trudgill 1998) in studies on linguistic variation, that of white males, with only two interviewees from that category.

Table 6-3 depicts some of the main demographic features of the speakers who were interviewed for this study. These are discussed in detail in the sections below.

[33] Nonmobile Older Rural Male

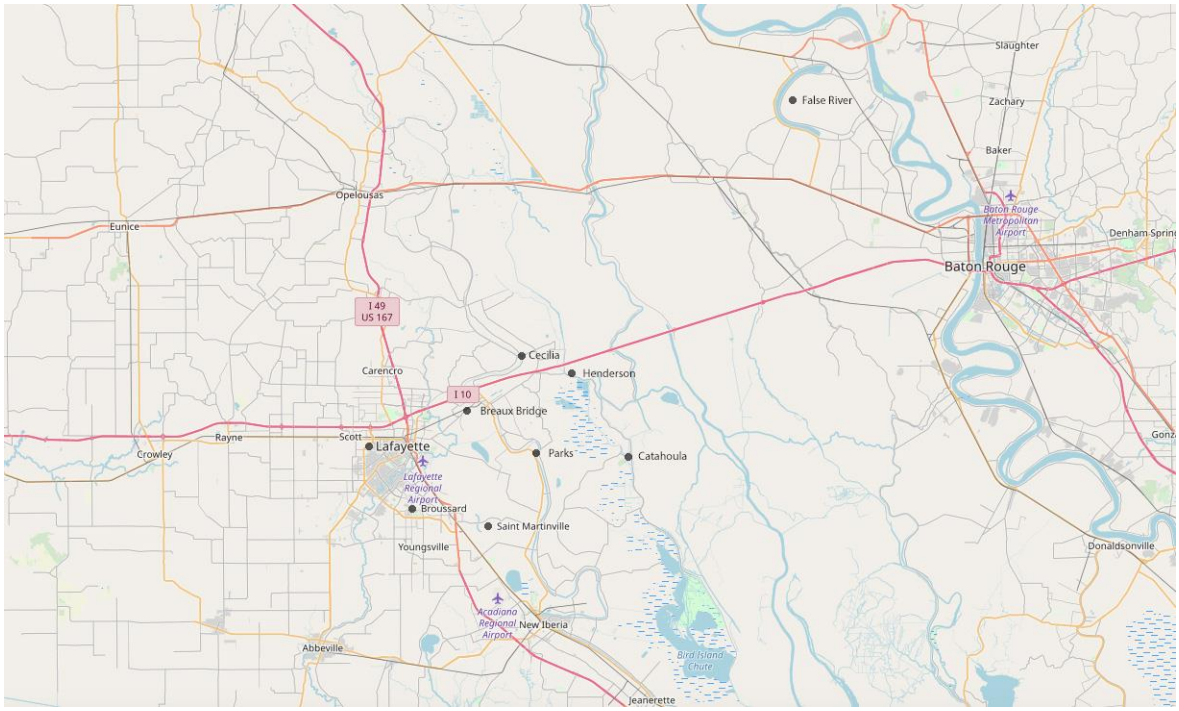
Table 6-3 Demographics of participants

Corpus ID	Sex	Birth Year	Race	Hometown	Residence	Variety	Occupation	Education
HW	m	1951	b	Breaux Bridge	Lafayette	TLC	teacher French & Spanish	tertiary
RM	f	1936	w	Cecilia	Arnaudville	TLC	housewife	high
JN	m	1933	b	Broussard	Broussard	TLC	farmer, butcher, restaurant owner	none
GL	f	1952	w	Henderson	Henderson	TLC	retired teacher, tourist guide	tertiary
TR	m	1955	b	Henderson	Henderson	TLC	construction, truck driver	regular
RBM	m	1963	b	Cecilia	Cecilia	TLC	labor at seafood factory	high
CC	f	1940	b	Bon Breaux	Lafayette	TLC	retired housekeeper	basic
LJ	f	1960	b	Breaux Bridge	Breaux Bridge	TLC	housekeeper	regular
JB	f	1953	b	Arnaudville	Cecilia	TLC	house wife, care taker	basic
MB	f	1936	b	Broussard	Broussard	TLC	ex-police officer, city council	high
ML	f	1928	w	Cecilia	Cecilia	TLC	farm	regular
MR	f	1937	w	Cecilia	Cecilia	TLC	housekeeper (part time)	high
LM	f	1937	w	Arnaudville	Cecilia	TLC	telefon operator, grocery store vendor	high
GT	f	1945	w	Henderson	Henderson	TLC	housewife	high
BC	m	1966	w	New Roads	New Roads	MLC	historian, geneologist	tertiary
AH	m	1966	w	Parks	Cecilia	TLC	IT, company owner	tertiary
KG	m	1937	b	The big Island (New Roads)	Port Allan	MLC	farmer, truck driver, museum	basic
GH	f	1949	b	The big Island (New Roads)	New Roads	MLC	retired director of NGO	tertiary
AF	m	1935	b	St. Martinville	St. Martinville	TLC	teacher, school board member	tertiary
AG	m	1939	b	The big Island (New Roads)	Irbanville	MLC	truck driver	regular
AR	m	1957	b	Henderson	Henderson	TLC	maitenance, transportation supervisor	high
HB	f	1939	w	Catahoula	Butte LaRose	TLC	housekeeper, truck driver	basic

6.1.1 Region

The majority of participants is from the Bayou Teche area and consequently speaks a variety of TLC. They stem from a number of communities, which are situated mainly within and around St. Martin Parish. These include Breaux Bridge, Catahoula, Cecilia, Henderson, St. Martinville, and Parks. Lafayette and Broussard are located in the neighboring Lafayette Parish. Most consultants still reside within the region if not necessarily the same community. Only four consultants are from Pointe Coupee Parish, namely the False River region, which belongs to the Mississippi area where the MLC variety is spoken. Most interviewees remained in the general area they were born in all their lives. They often moved to other towns and communities in the region but few left their parish or even the state. For the purpose of this study, only those cases in which somebody left the francophone triangle are distinguished. This is true for six of the participants who spent a significant amount of time elsewhere, usually out of state. AG has one of the longest periods of absence with 20 years spent in New York but others also lived elsewhere for considerable amounts of time.

Figure 6-1 Hometowns



© OpenStreetMap

Similar to the distribution of black and white speakers in the sample, a bias is inherent in this regional distribution of participants. Clearly, the data for this study must favor the Teche variety as more participants came from this region. This is not unexpected as LC is the most vital in this area as observed by Klingler and Dajko (2006, 16) who claim

that “it is in the Teche region that the most numerous and, on average, the youngest LC speakers are to be found” and confirmed by Mayeux (2019). As this was expected, efforts concentrated on the Bayou Teche region, which may well have compounded the effect. Consequently, the larger proportion of interviewees from this region mirrors the actual distribution of speakers in Louisiana. This circumstance justifies the inclusion of the MLC data in the sample. Additionally, regional variation is always considered as a possible factor. Again, one may rightfully argue that the present imbalances in the sample are due to opportunistic data collection. Originally proposed as a necessary but far from sufficient characteristic of a good language documentation corpus by Woodbury (2003, 13), the opportunistic approach naturally meets resistance from a corpus linguistics perspective, where great care is taken to ensure a balanced data set. While fieldworkers always strive to obtain data from as many social circles and as wide a range of demographics as possible, I believe that the reality of fieldwork with endangered languages cannot afford to fully circumvent such effects. This is in line with the “fundamental importance of taking a pragmatic stance” which Himmelmann (2006, 4) postulates as the premise for his theoretical and methodological guidelines for language documentation. The advantages of a larger sample must be weighed against those of a more homogenous data set. Where the decision is made to include potentially unbalanced data, as is the case here for regional varieties, the main objective must be to remain aware of this decision and consider it in all generalizations made from the data. This is commonly done in studies conducted in Louisiana (Klingler 1992, 2003a, Neumann 1985, Mayeux 2019, and for LRF , Blainey 2017, Dajko 2009) and the present analysis follows the practice concerning sex, age, race and region.

6.1.2 Education and occupation

In Table 6-3 each participant’s current and often past occupation is listed. Details are omitted for the sake of readability. Clearly, speakers are in a wide range of professions. Most of them were retired at the time of the interview though many held jobs to make ends meet. Contrary to expectations, not all of them worked in low-income professions. There are a number of teachers and other academics as well as civil servants and businessmen. When the professions are broadly classified into manual labor/low income and higher income/higher status occupations, the following picture emerges:

Table 6-4 Occupation of participants

	Blue-collar	House	White-collar
black	8	1	4
white	4	2	3
Total	**12**	**3**	**7**

A majority worked or still works in blue-collar professions, but seven participants worked in higher status professions which equals 31,82 % of the sample. Three women were stay-at-home mothers. Considering white and black speakers in employment separately, the percentage of blue-collar workers is 57,14% and 66.6% respectively. This may imply a racial factor, as a larger proportion of white creolophones works in higher status professions. On the other hand, the figures would be reversed had all data sets been included, as one white speaker with low income and one black speaker in a white color profession were removed from the corpus (cf. Chapter 5.2). Consequently, no generalizations can be made from this ethnic distribution. 36,84% of the consultants in employment were white-collar professionals. Considering the popular perception that uneducated working-class people speak LC, this number is unexpectedly high, although these participants often come from low-income families. Their parents and grandparents were usually farmers or performed manual labor of some kind and as children the interviewees were often required to work on the fields. This also interfered with their formal education. Several consultants report that they could not go to school on a regular basis because they had to help in the fields or around the house, or work in some other way to support their families. In many cases these were large families in which the children were expected to contribute. For some this meant that they had to leave school as teenagers, either because they were needed elsewhere, had to earn an income or because their sporadic early education had not prepared them sufficiently to continue their studies. In some cases, education was pursued to a larger extent because it was valued highly in the home, which is not necessarily always the case in the community. In fact, I encountered the attitude that children should not seek to surpass their parents and/or that education was just time-consuming and worthless anyway. This does not only occur among creolophones but seems to be a commonly held view among the poor rural population in South Louisiana. Other interviewees were intrinsically motivated to further their education either from genuine enjoyment or with economic success in mind.

Table 6-5 sums up the education level reached by LC speakers in this study. To account for individual biographies while maintaining comparability, five generalized

categories are used. *None* refers to speakers who never attended school at all. *Elementary education* refers to participants who did not go to school very regularly and did not or just barely continue beyond elementary school level. *Regular* education level is the label used for students who went to high school but did not graduate. High school graduates are labeled *high* and college students receive the label *tertiary*. No further distinction is made between undergraduate and graduate degrees.

Table 6-5 Education level of participants

	Education level
none	1
elementary	4
regular	4
high	7
tertiary	6
Total	**22**

As is illustrated above, the majority of participants (13 out of 22) hold a high school degree or higher. Only four out of 22 participants did not receive any education beyond the elementary school level. This again contrasts starkly with the common perception of LC-speakers as uneducated. It must also be noted that a higher education level does not necessarily correspond to higher incomes. Only among the college graduates did all interviewees end up in white-collar professions. A high school degree paved the road to such an occupation only for one participant and this involved a lot of effort and determination on her part. The other six either worked in blue-collar jobs or in the home. Although 22 does not constitute a representative sample, this poses the question of whether formal education below the college level is advantageous at all in this area and, if so, why these advantages do not seem to extend to creolophones. While language stigma may be a possible factor here, it is unlikely to be the decisive one, mainly due to the fact that LC-speakers are bi- or multilingual as a rule and all of them speak English. The employment situation in rural South Louisiana is difficult and few higher-level positions are available. Despite this, people feel highly attached to their home regions and often do not want to move away or even to the city for better employment opportunities. Louisiana's school system notoriously scores very low in national comparisons[34], so that it is questionable how much their degree would have been worth had they left. Coming from poor families, they

[34] most recently by McCann (2019), https://wallethub.com/edu/states-with-the-best-schools/5335/ (last accessed 24 August 2019)

were often taught to strive for stability and survival rather than ambitious goals. Many hence perceive a steady income as a hallmark of economic success.

6.2 Language competence and multilingualism

As is typical of endangered languages, speakers show different levels of language proficiency. Broad self-assessments have been shown to be unreliable in this context as they largely depend on language prestige and language attitudes (Grenoble 2013). As mentioned, a brief translation task is employed in the sociolinguistic questionnaire following Graham (1981). This task determined whether someone spoke LC at all or just considers himself or herself Creole but is in fact an LRF speaker (cf. Chapter 5.1). Beyond this step, no testing of actual language proficiency was conducted for two reasons. On the one hand, neither are testing materials available nor do I feel entitled or sufficiently proficient in LC to determine speakers' competence levels. On the other hand, such a language test, even had it been available, would run counter all considerations for creating a suitable interview environment and certainly would put insecure speakers under pressure, which was to be avoided as much as at all possible. Consequently, reliance on self-assessment could not fully be avoided. To achieve a more detailed and hopefully less skewed picture, consultants were asked whether they had the ability to perform ten linguistic tasks taken from Melancon (2000, 102), which require increasing proficiency levels. Answers are coded with 1- fully able, 2- with some difficulty and 3- unable. Speakers' proficiency level is classified according to the assessments they provide. Group A indicates fluent speakers who can perform all tasks without or with little difficulty. Group B indicates speakers who have difficulties with two or more tasks. One speaker is placed within Group A despite indicating he would have difficulties with two tasks, because these concerned matters at the bottom of the task ranking, which require low proficiency. This speaker proved himself quite fluent during the following interview and spontaneous conversations but apparently did not recall the vocabulary needed for the low-level tasks immediately. Group C contains speakers, who judged themselves unable to perform one or more of the tasks. These are the most insecure and least proficient speakers in the sample. In sum, 14 speakers fall into proficiency Group A, five in Group B and three in Group C. Table 6-6 to Table 6-9 indicate the distribution of proficiency levels over race, gender, region and age group.

Table 6-6 Proficiency by race

	A	B	C
black	9	3	1
white	5	2	2
Total	**14**	**5**	**3**

Table 6-7 Proficiency by gender

	A	B	C
f	7	3	2
m	7	2	1
Total	**14**	**5**	**3**

Table 6-8 Proficiency by region

	A	B	C
Bayou Teche	11	4	3
False River	3	1	0
Total	**14**	**5**	**3**

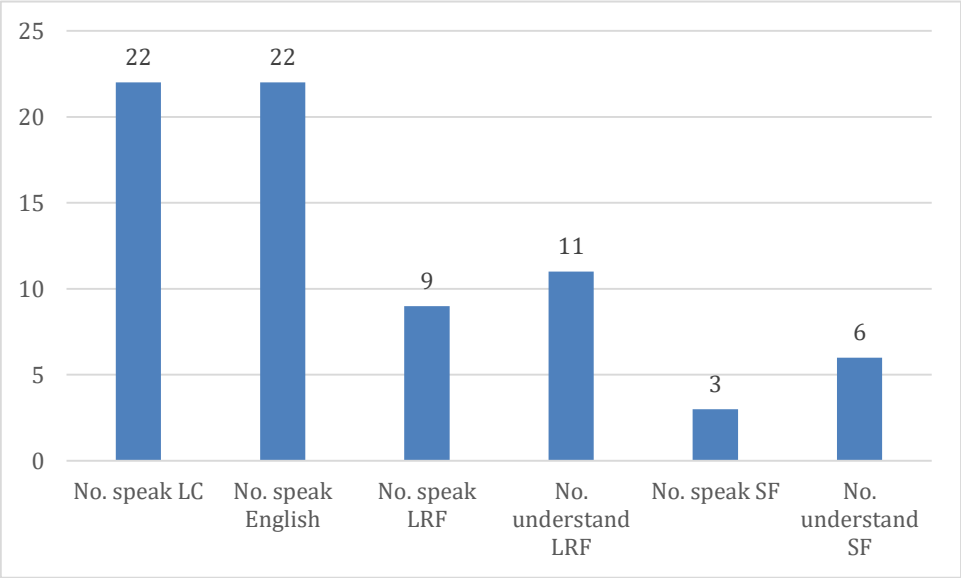
Table 6-9 Proficiency by age group

	A	B	C
46-60	3	2	1
61-75	3	1	2
75+	8	2	
Total	**14**	**5**	**3**

For most of the demographic factors no obvious relation to the level of language proficiency can be established. Black speakers tend to be – or at least consider themselves – more proficient than white speakers, which fits the idea of LC as a marker of identity for the former group. Men seem to be slightly more proficient than women in this sample but the difference is a minor one. Most obviously the difference between age groups appears potentially significant. As mentioned, no statistic testing was conducted due to the sample size, but with eight out of ten speakers, the oldest age group clearly has the highest ratio of fluent speakers. This would be expected as proficiency usually decreases alongside language vitality. The regional comparison has little relevance as the number of speakers from the Teche region is much larger than that from the False River area. All Group C speakers are from Bayou Teche but this is likely to be a coincidence considering the overall numbers.

As mentioned, creolophones in Louisiana nowadays always speak English and possibly other languages as well. Monolingual speakers do not seem to exist anymore. In addition, interviewees claimed varying levels of competence in both LRF and SF. Please note that the following are self-assessments that were not confirmed through any testing.

Figure 6-2 Languages known



While all participants speak LC and English, only some of them claim knowledge of either LRF or SF or both. Only three speakers have productive language skills in SF, although six claim to have receptive competence. The speakers of SF learned the language in classroom contexts, usually in high school or college. Those with passive knowledge do not necessarily rely on remnants of French learned in school, though this is true for some participants. Others simply 'picked up' some understanding of SF from their exposure to LC and LRF as well as French media, mainly radio programs. Since no formal testing of this self-estimated ability was conducted, the actual level of passive SF-competence may well vary strongly. For LRF the numbers are notably higher, with nine interviewees claiming to speak LRF and half of the participants being able to understand the local variety. Since LRF and LC share large parts of their lexicon, some degree of intelligibility is to be expected. Most of the interviewees who speak both LC and LRF can clearly distinguish between the varieties, especially those who grew up bilingually with both languages or married a speaker of LRF.

Most speakers who took part in this study grew up with LC as their L1. Child bilingualism with either English or LRF occurred in six cases but the majority of consultants acquired LC as a first and dominant language. The two participants who acquired English as their L1 learned LC as children from their grandparents.

Table 6-10 L1 of participants

L 1	
English	2
English|LC	3
LC	14
LC|LRF	3

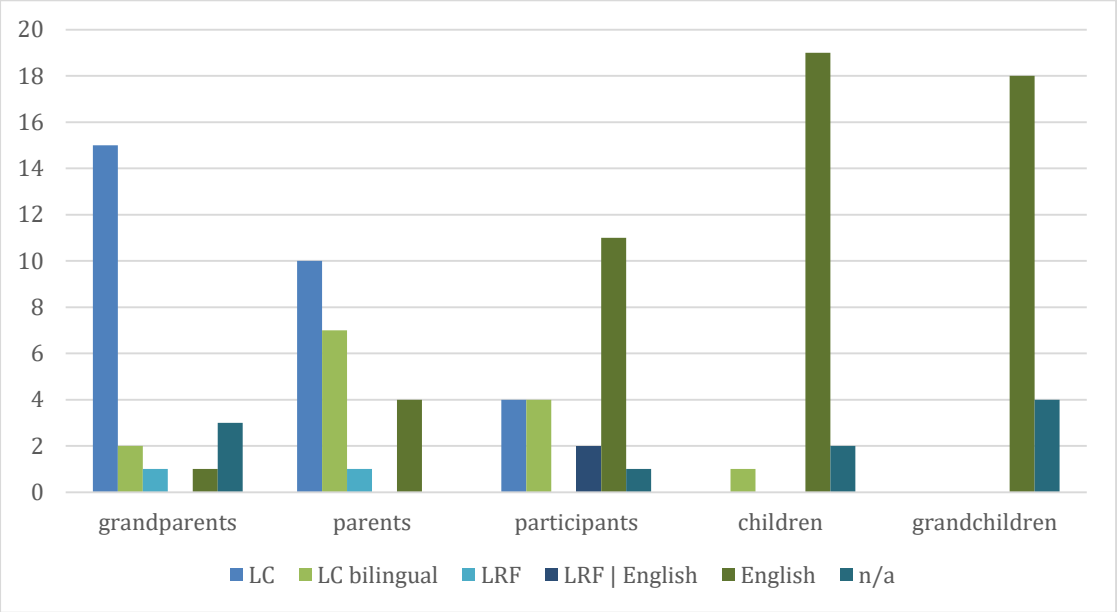
The reader may be misled to infer that the speakers were mainly monolingual throughout their childhood, since only six individuals report being raised bilingually. However, English already was established in their environment at the time, and was asserted in school at the latest. LRF also had its place in Southern Louisiana society and many were familiar with it to some extent. Commonly, LC was the dominant but not the only language during early childhood. For 11 of the consultants a monolingual LC speaker was present in the home which made it necessary for LC to be the main medium of communication. Other households were bilingual to varying degrees. Some speakers acquired LC from one grandparent only and never used it with any other family members; other households used English or LRF and LC in what interviewees often described as a mixture. It is assumed that this pertains to language choice as well as code-switching practices.

Seven of the participants had no or very little knowledge of English before they entered school. School often was a traumatic experience for these children because they were expected to speak only English regardless of their mother tongue. Consultants remember feeling deeply insecure and lost in the unfamiliar setting and language. Many faced admonishment or even corporal punishment for using their native language in school and consequently "didn't dare speak French or Creole in school" (AF). Those with older siblings who were already schooled in English often learned some of the language from them. Those who grew up with English as their first, or one of their first languages, obviously knew English before they entered schooling and were consequently spared this experience. Two speakers who grew up monolingually with LC also claim to have gained sufficient competence in English before they went to school. Surprisingly, eight interviewees made no clear statement regarding their English skills before school. Most stated that they could not remember having encountered any problems but that English was not the language of the home. They commonly assumed to have picked up English on the fly or from friends. The present data does not contain information on the specific schools that were attended. Possibly, some elementary schools coped with the multilingual

circumstances better than others, resulting in a smoother transition and hence a less memorable learning experience.

Interviewees were asked which is their 'best language' in the sense of the language in which they feel most comfortable. The same question was also posed with regard to their grandparents, parents, children and grand-children. The following chart illustrates this intergenerational development. *LC bilingual* as a category includes cases of more or less balanced bilingual speakers of LC and English or LC and some form of French as well as cases in which one parent/ grandparent spoke LC and the other LRF or English.

Figure 6-3 Transgenerational 'best' language



For 15 participants LC clearly is the best and main language of their grandparents. In sum 19 participants' grandparental generation provided linguistic influence out of which 17 spoke LC best if one includes the two bilinguals in the grandparental generation. Three participants grew up without grandparents in their life. The one participant whose grandparents spoke mainly English reports that LC was used as a secret code to discuss matters that the children were not supposed to know about. This apparently motivated them to learn as much of the language as they could.

In the parent generation LC still seems relatively strong but the proportion of bilinguals who use LC alongside English, as well as the number of speakers whose best language is English is much higher. In several cases, LC was used among the older generations but not used with the children as often. AF for example remembers that his mother would use LC with her siblings but not her children because she "was convinced that French was gonna mess you up in learning English". A reverse situation was in place for

RM whose parents and grandparents were primarily LRF speakers, but who would use LC with the other children in the neighborhood. She was not sure how this came to be although it is most likely to be attributed to the fact that they were living in Cecilia, "I mean people from Cecilia speaks [*sic*] Creole". She acquired both LRF and LC and can clearly differentiate between the two codes. Six participants grew up with a parent who was fully or almost monolingual in LC and spoke no or very little English. These are among the 10 cases in which LC remained the best language in the parent generation. In four instances the parents spoke English and LC equally well. Three of the households in the *LC-bilingual* category used LC alongside LRF or another French variety. One was a Cajun family in a LC-speaking community, one parent couple consisted of a French-Canadian father and an LC-speaking mother and for the third family, the grandparents consisted of one LRF and one LC speaker which resulted in mixed codes in the family.

The participants themselves largely consider English their best language. Only four consultants chose LC. Notably, LC also is the best language of the parent generation for each of these individuals. Another four claim that they are equally comfortable in LC and English. Their parents fall into the *LC bilingual* category in two cases, one of which includes an LC-monolingual parent, and one had parents who spoke LC best. The last interviewee's parents actually speak English but he was mainly raised by his LC speaking grandmother. Two of the three speakers who were raised bilingually with LRF and LC consider LRF their best language today. All remaining interviewees agree that English is their best language.

While language shift can be clearly observed in the development over these three generations, the crucial tipping point seems to be reached with the consultants' children. 20 of the 22 interviewees have children and among them only one son would consider LC his best language. Even in this family he is the only one among the siblings to speak the language. Otherwise, English is the clearly dominant language for this generation. The trend continues with the interviewees' grandchildren. This data clearly portraits the loss of Louisiana Creole over generations. For the children and grandchildren, a closer look is worthwhile as the data presented so far is solely concerned with the 'best' language. Therefore, consultants were asked which languages their children and grandchildren respectively speak and understand.

Figure 6-4 Language competence other than English in children and grandchildren

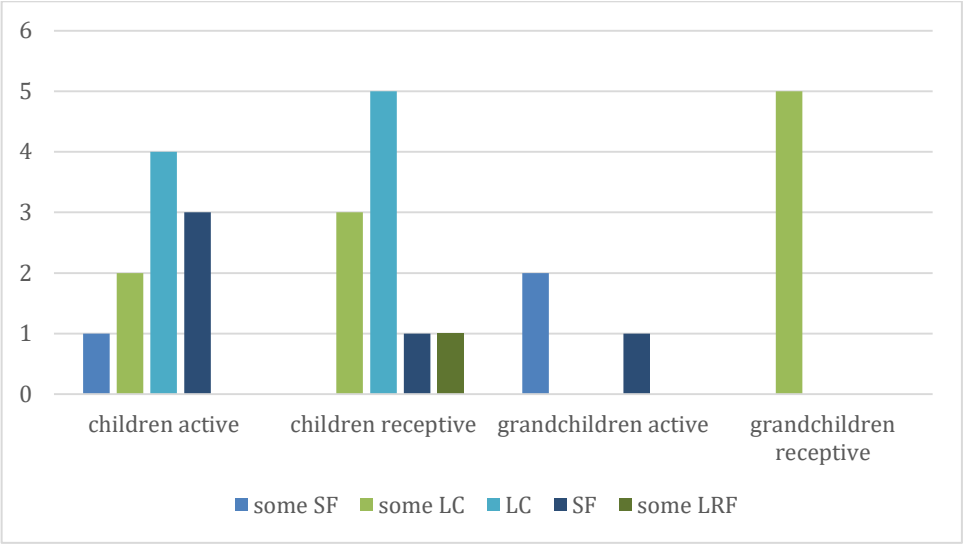
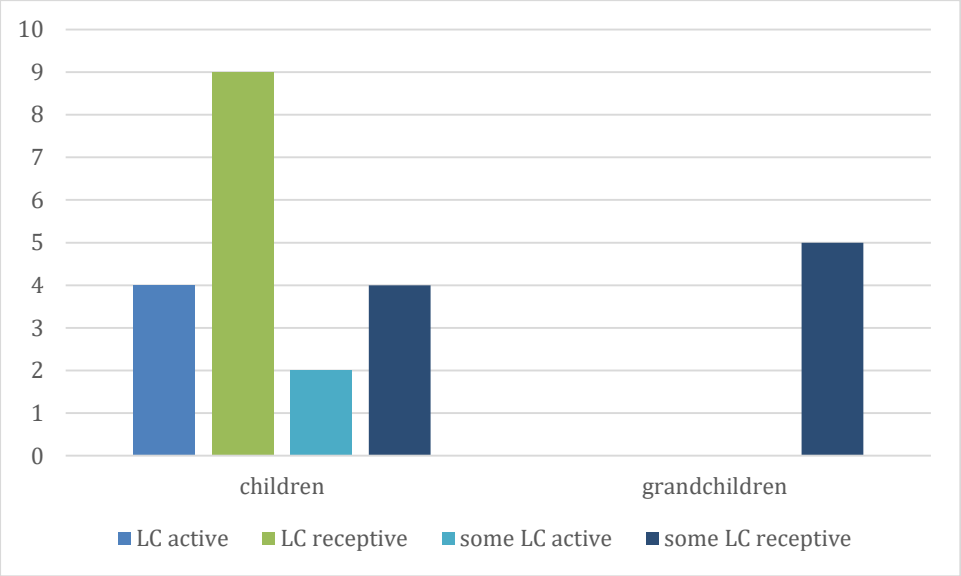


Figure 6-4 only indicates the degree of active and passive competence in LC, LRF or SF. English is not considered, as it is the dominant language for all Louisianans in these generations. *Receptive* competence refers to children or grandchildren who are able to understand the respective languages when addressed in them but never speak the language themselves. Naturally, those with active competence also have receptive skills in the same language. These are omitted from the receptive section in the graph to depict the different proficiency levels involved. One important conclusion to draw from this data concerns SF. On the one hand, the number of children and grandchildren with any SF proficiency is rather low, with four families that have speakers in the child generation and three families with some competence in the grandchild generation. This is in line with the assessment made by Melancon (2000, 102) of the low impact SF-schooling has had in Louisiana. On the other hand, if grandchildren possess any language competence in a French variety at all, it appears to be SF. A total of five grandchildren were enrolled in French immersion classes at least for a short time.

The main point here is that knowledge of LC appears to further diminish from the child to the grandchild generation. While four children have active language skills in LC and an additional five understand the language fully, none of their children are able to speak or understand LC on that level. Moreover, speakers with reduced competence levels who can speak or understand just a little bit of LC appear in the child generation. Similarly, five participants' grandchildren only have some receptive skills in LC, usually because they are occasionally addressed in LC when their elders are agitated or giving them instructions. This is further illustrated in Figure 6-5 below. In this chart, in contrast to Figure 6-4, the active speakers are included in the number of children with receptive skills, which is why

their number seems to be higher. Since the focus here is solely on LC, this serves to depict the full spectrum and clearly illustrates the rapid decline of the language. It is a straightforward assumption that the grandchild generation will not transmit any LC to their children, unless massive language maintenance efforts lead to heritage language learning.

Figure 6-5 LC competence in children and grandchildren



6.3 Language use

While language proficiency levels and intergenerational transmission are important indicators of language vitality (cf. UNESCO 2003), patterns of language use and the usage of languages in functional domains must be considered as well to achieve a full impression. To this end, the questionnaire asked for interlocutors in the past and present, providing a list and five answer options from LC to English.

6.3.1 Interlocutors

Table 6-11 gives an overview of the languages used with groups of interlocutors as children. The total of replies included here frequently is higher than the number of respondents. The same 22 consultants were interviewed for the language use section but sometimes they provide more than one answer because language use differs not only with groups of interlocutors but often also within them. For example, some participants used LC with one parent and English with the other as children or they would use LC with particular friends or siblings but not with that group in general. In such cases both answers are salient and are included in the analysis. Considering the language used with parents for example, 22 interviews were conducted. Three interviewees used LC to differing degrees with their mother and father respectively, with one parent in the *LC* and one parent in the *mostly LC* or

the *LC | English* section. Another interviewee spoke mostly English with the mother and LC with the father. It follows that in sum 26 data points exist for the 22 interviewees in this section.

Table 6-11 Language use in childhood by interlocutors

	Grandparents	Parents	Siblings	Classmates/ friends in school	Friends outside school
LC	17	11	7	-	6
mostly LC	1	6	3	2	4
LC | English	1	3	3	2	3
mostly English	2	3	6	5	3
English	-	2	3	13	5
n/a	1	1	1	-	3
Total	22	26	23	22	24

The language use patterns as children reveal two main tendencies. First, the older the interlocutor, the more likely LC appears to be used. This usage pattern fits the language proficiency levels and multilingual capabilities across generations as discussed above. The grandparent generation was the most fluent in LC and the least fluent in English, which is why LC was the main medium of communication with them. Second, usage of LC among the younger generations declines and seems to depend on the domain of use. Due to the dominance of English in school contexts, participants would use more English than LC with classmates and friends in school. With friends outside of school and siblings on the other hand, the picture is more diverse; LC and English both play a role although LC appears to be slightly more prominent.

Since language use changes over the generations, age might be a factor. One might expect that the older interviewees had higher levels of LC use in their childhood than younger consultants. Below, the results are patterned by age group to give an impression. With each table a graph is included for visualization. Again, where several answers apply, they are included and added to the total.

Table 6-12 Usage with grandparents in childhood by age group

	46-60	61-75	75+	Total
LC	4	5	8	17
mostly LC		1		1
LC | English	1			1
mostly English	1		1	2
n/a			1	1
Total	**6**	**6**	**10**	**22**

Figure 6-6 Usage with grandparents in childhood by age group

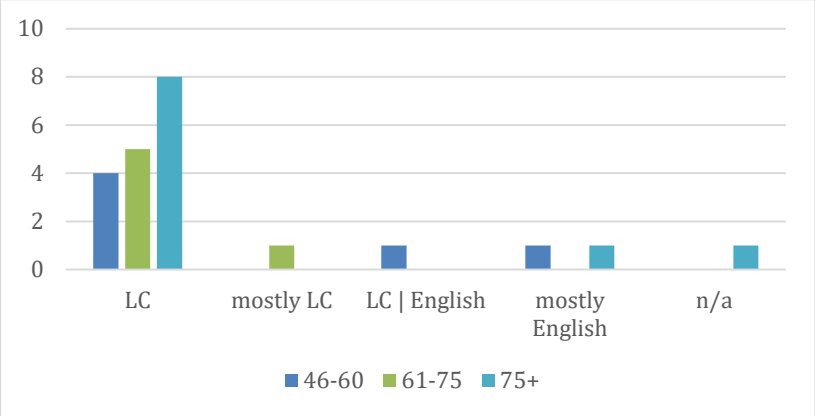
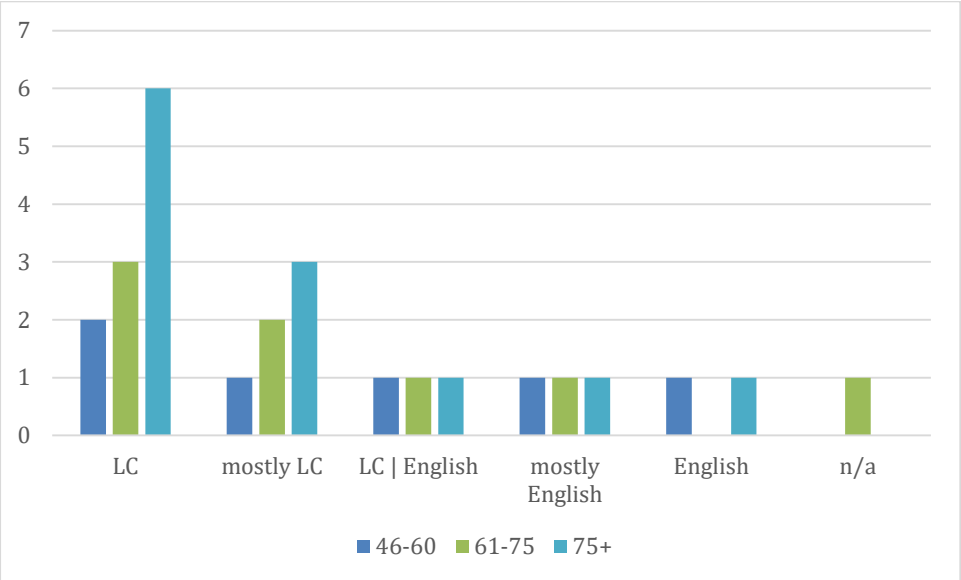


Table 6-13 Usage with parents in childhood by age group

	46-60	61-75	75+	Total
LC	2	3	6	11
mostly LC	1	2	3	6
LC | English	1	1	1	3
mostly English	1	1	1	3
English	1		1	2
n/a		1		1
Total	**6**	**8**	**12**	**26**

Figure 6-7 Usage with parents in childhood by age group



For the usage with the grandparent and the parent generation, age does not appear to matter very much. LC was the most frequently used language with parents and

grandparents during childhood regardless of the age of the interviewees. Usage with grandparents was more common than with parents through all age groups, which is fully in line with the overall results in Table 6-11. The number of participants who were aged 75 and older at the time of the interview is the highest, which is why their relative numbers are higher. But when it comes to interlocutors of the same generation as the participants, age appears to make a difference.

Table 6-14 Usage with siblings in childhood by age group

	46-60	61-75	75+	Total
LC	-	2	5	7
mostly LC	-	2	1	3
LC | English	2	1	-	3
mostly English	3	1	2	6
English	1	-	2	3
n/a	-	1	-	1
Total	**6**	**7**	**10**	**23**

Figure 6-8 Usage with siblings in childhood by age group

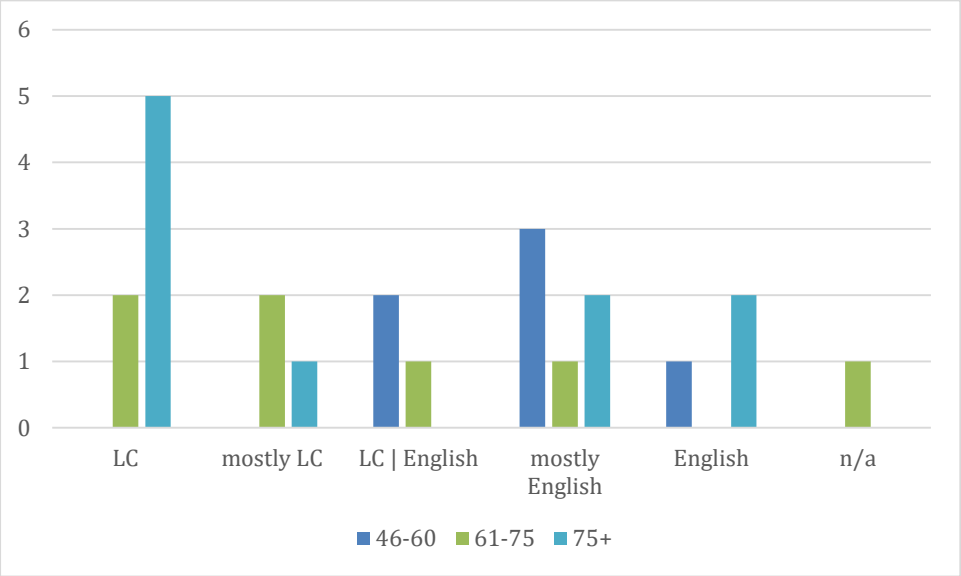


Table 6-15 Usage with classmates in childhood by age group

	46-60	61-75	75+	Total
mostly LC			2	2
LC | English			2	2
mostly English	1	2	2	5
English	5	4	4	13
Total	**6**	**6**	**10**	**22**

Figure 6-9 Usage with classmates in childhood by age group

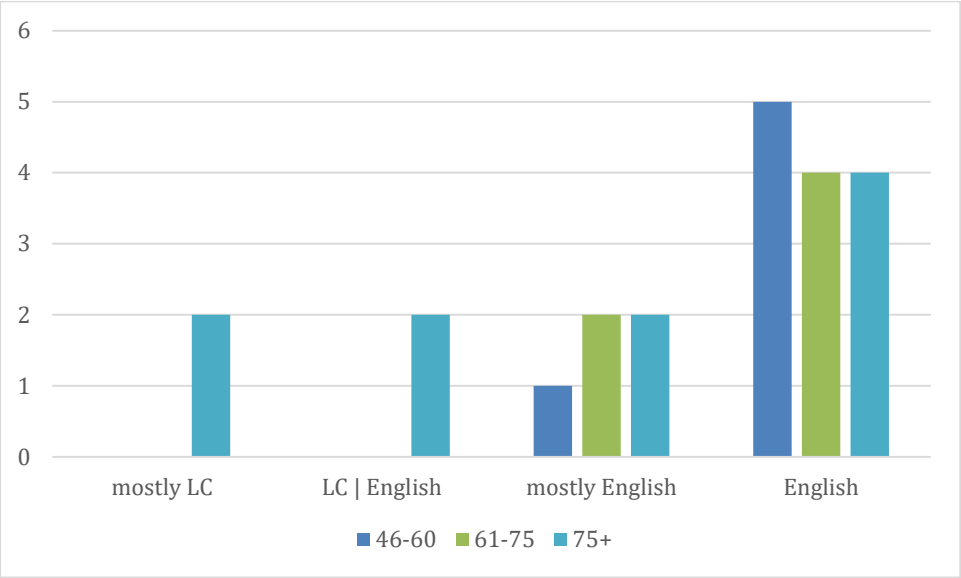
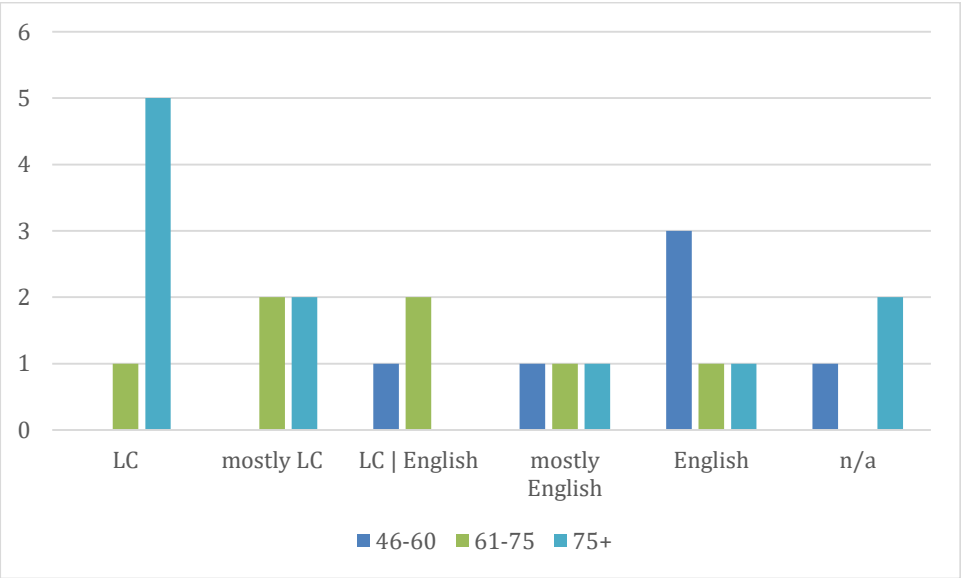


Table 6-16 Usage with friends in childhood by age group

	46-60	61-75	75+	Total
LC		1	5	6
mostly LC		2	2	4
LC | English	1	2		3
mostly English	1	1	1	3
English	3	1	1	5
n/a	1		2	3
Total	**6**	**7**	**11**	**24**

Figure 6-10 Usage with friends during childhood by age group



The use of LC with siblings is more frequent among the oldest age group although two interviewees used English with their siblings at the time. This was the case, for example, for younger siblings who learned English from the older ones. Among the youngest participants, nobody used only or mostly LC. Two of six interviewees in this group used English and LC bilingually with their siblings as children, while English was the main language among siblings for the remainder. The usage in the 61-75-year-old group seems to resemble that of the older age group, as LC was used often. Five out of seven replies from the age group claim to have spoken solely or mostly LC with (one of) their siblings.

The same pattern continues concerning the use of LC and English respectively with friends outside of school. Both languages were used in this context to varying degrees. The proportion of LC use is highest in the oldest group of consultants with 63,6% of predominant or exclusive use of LC and just one speaker in the *English* as well as the *mostly English* category. The middle age group slightly favors LC over English but the shift appears to be emerging in the increased use of bilingual modes, as there are two consultants in the *LC | English and* the *mostly LC* category each. One interviewee chose the *LC*, *mostly English* and *English* category respectively. Overall, this is a fairly even distribution with a tendency towards LC. This reflects the ongoing language shift as English seems to be in the process of establishing itself as the medium of casual conversation at the time. A comparison with the data from the youngest group strengthens the impression, as they clearly favored the use of English with their friends.

English dominated language use in school contexts for all age groups as expected. Only for the oldest group did LC constitute an option at all in this domain. It must be noted that the question does not refer to the languages spoken in class but the ones used with classmates and friends in school in general. Nevertheless, English was the main language used. Four participants from the oldest group used LC in school, though never exclusively. Even that practice no longer applied for the next age group and the youngest participants spoke almost exclusively English in school.

The comparison of language use patterns in childhood according to age groups in sum reveals that language shift was already ongoing during the interviewees' childhood. The data clearly shows that the language was vanishing from school, the most formal domain in their lives, and was strongest in interactions with the grandparent generation. Although it often was the main language of the home, some decline is already visible in the use among siblings and friends. Especially the youngest group of participants differs from

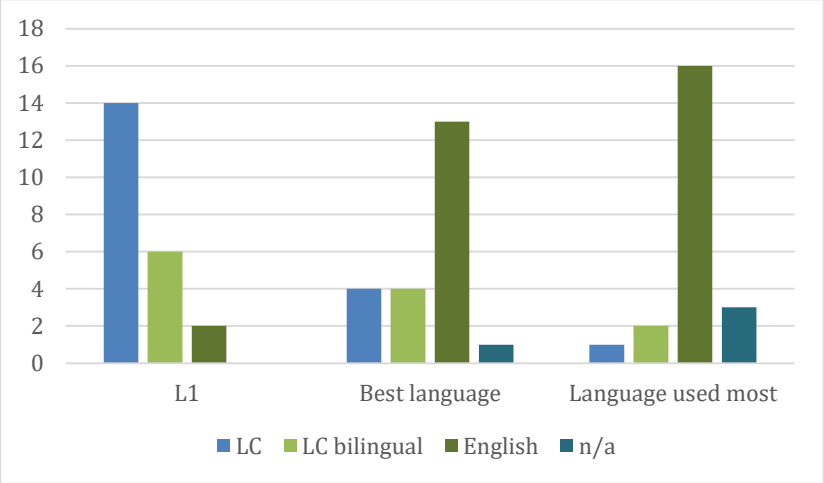
the older ones, indicating a generational split occurring for speakers who were born in the late 1950s and thereafter.

To complete the picture this data is contrasted with the language use patterns at the time of the interviews. Language use at large has clearly shifted almost completely to English as the main language of everyday use. 74% of the interviewees said that English was the language they used most in everyday life. Just one person claimed that LC was the most-used language and the remainder chose English alongside either LC or LRF. When both LC and English were chosen, LC came with a qualifier, such as "I use Creole as much as possible" or "We speak Creole at home".

The distribution of the language used most mirrors the distribution of the best language among the interviewees to some extent. LC is a bit stronger in that category, as people who considered it their best language nevertheless do not claim to use LC the most. Oftentimes, a speaker who is equally comfortable in both English and LC chose just English in the *used most* section and speakers who selected LC as their best language chose *LC | English* when asked which language they used most.

The shift is even more clearly illustrated when comparing the *L1*, *best language* and *language used most* among the interviewees. The decline of LC is clearly visible as the same creolophones who grew up speaking only LC now consider English their best language in many cases and, even if they do not, it is the language used most. For the sake of clarity, the data for this chart is simplified in the following ways. In two instances interviewees considered LRF and English their equally best language or the language they use most. These are subsumed under *English* in Figure 6-11 below, as this thesis is concerned with LC rather than the role of LRF. As described above six speakers grew up bilingually with LC and either English or LRF. These are also summarized in one bilingual category below. For the best language and the language used most, the number in this category solely represent bilingual use of LC and English as no one chose LC and LRF in these sections.

Figure 6-11 L1, best language, language used most



This figure offers an interesting profile of the LC situation, as it includes a diachronic, psycholinguistic and sociolinguistic dimension. LC is prominent as an L1. The decline in the other two categories shows the diachronic development of language loss and language shift towards English. These two columns are concerned with synchronic perspectives. Here the general trend continues but a slight difference is revealed between the psychological role of LC, the mother tongue, and its actual use. First, it is remarkable how many interviewees consider English their best language despite the fact that it is not their L1. This reveals a possible shift in actual competence in both LC and English or at least represents the perceived level of proficiency speakers assign to themselves and the psychological reality of the two languages in their minds. Despite the fact that they grew up speaking LC, often exclusively, English is now cognitively dominant. This may well go hand in hand with certain levels of L1 attrition for LC. English is even more dominant in actual language use. While LC still plays a more or less major role in the minds of eight interviewees, only three assign it some relevance in their daily usage. This indicates the complex interplay between the psychological reality and cognitive representation of the respective language on the one hand, its domains and functions on the other hand, as well as perceived or actual language proficiency. The data further shows that the decline of LC is not limited to intergenerational developments which was already illustrated by the proficiency data above. The comparison indicates a severe decline in the use and the psychological prominence of LC over the lifespan of the interviewed creolophones.

Of the socio-demographic factors considered here[35], race seems to have the strongest impact on which language is considered the best language. The distribution is almost the same for men and women, as well as across age groups. LC is slightly better

[35] Whenever sex, age, race and region are considered as possible socio-demographic factors of influence, numbers are reported here only when these factors appear to have an observable effect.

represented in the oldest age groups, with one more person choosing LC. Two white female speakers from the oldest group consider LRF and English their best languages. Interestingly, just one white speaker considers LC the best language in the repertoire but three black speakers chose this option and an additional four interviewees selected LC alongside English which no white interviewees did. This seems to indicate that the language has a higher psychological value for black speakers since no clear difference can be seen between black and white speakers when it comes to the language used most. These results are in line with earlier assertions (Neumann 1985, Dubois and Melançon 2000, Klingler 2003a, Mayeux 2019) that LC functions as an identity marker mainly for black Creoles in Louisiana and that white speakers have a more ambivalent relation to the language. No other category seems to be relevant for the language used most either, with the possible exception of the regional distribution which is discussed below. The lack of a difference between socio-demographic groups is an indicator of language endangerment and shows the advanced stage of language shift to English as it is used the most across the board rather than by certain groups of speakers.

Table 6-17 Best language by race

	black	white	Total
LC	3	1	4
LC | English	4		4
LRF | English		2	2
English	6	5	11
Total	**13**	**8**	**21**

Obviously, this remains a cursory glance at the ends of a varied spectrum as these questions aimed for absolute categories, like the best language or the one used most. It is likely that LC plays a role that is not captured by these approaches. LC may not be the dominant language in daily use but that does not mean that it is not spoken. Below, a more distinct consideration of factors that may be relevant for language use offers a more detailed impression. Rottet (2001) points out that language choice is influenced by factors such as the interlocutor, the setting, the domain, the function, the channel and the topic of conversation. The already extensive nature of the sociolinguistic questionnaire did not allow for an in-depth investigation of all these factors. Instead, I focus on the interlocutors, which Rottet (2001, 104-108) identifies as the most relevant category for the LRF context and which also appear to be very important for LC. The chosen groups of interlocutors serve as representatives of certain domains at the same time. In addition, functions are

contained in the usage statements. The channel or medium of communication, and the topics of conversation are only marginally touched upon.

Interlocutors obviously have changed over time. Most grandparents and parents passed away and none of the interviewees still attend school. To achieve maximum comparability the groups of interlocutors remain the same wherever possible, namely *parents*, *siblings* and *friends/ neighbors*. In addition, *spouses* are included as representatives of the home domain and most frequent interlocutors. *Children* and *grandchildren* similarly belong to the private domain though rarely the same household anymore. Moreover, they serve to indicate intergenerational developments of language usage and transmission. *Co-workers* are included in the classmates' stead, to represent the more formal domain in everyday life. Lastly *shopkeepers* add a domain of casual but public interaction. This selection of interlocutors covers the main functional domains, ranging from the home, and the private over the public to the more formal, as well as all generations of community members. In relation to the usage patterns during childhood a fairly complete idea of the overall development can be formed.

Table 6-18 Language use by interlocutors

	Parents	Siblings	Spouse	Children	Grand-children	Neighbors/ friends	Co-workers	Shop-keepers
LC	-	3	-	-	-	-	-	-
mostly LC	2	2	2	1	-	4	1	-
LC | English	-	4	3	2	1	4	1	1
mostly English	3	11	7	4	3	10	2	5
English	-	6	6	14	13	4	11	8
LRF	-	-	2	-	-	2	-	-
SF	-	-	-	-	-	-	1	-
n/a	17	1	2	2	5	2	7	8
Total	**22**	**27**	**22**	**23**	**22**	**26**	**23**	**22**

Clearly, the role of LC is much diminished as compared to the interviewees' childhood. Exclusive use of Louisiana Creole is attested only among siblings and only in three cases. Note that the total of replies in this column is 27, indicating that language use occasionally depends on the particular brother or sister one interacts with. This holds for one of the three instances that use LC with their siblings. Comparing the use of languages with parents is not very helpful, since just five consultants' parents are still alive. These naturally are among the younger participants. Two of them claim to speak mostly LC with

their parents. For one of them, LC used to be the only language used with the parents during childhood. Consequently, this constitutes a very subtle shift towards English. Surprisingly, the other interviewee claimed to have spoken mostly English with his parents as a child. This is either a mistake in data processing or due to a conscious effort on the family's part to use LC in order to maintain the language or an inconsistency in self-assessment most likely caused by the interview situation and my obvious appreciation of LC. The three consultants who currently speak mostly English with their parents also favored English with them as children. One of them used LC and English with the parents during childhood also indicating a slight shift towards English.

With their spouses, interviewees also tend to use more English than LC. Two couples use mostly LC and three use both languages among themselves. In two cases the interviewees married an LRF speaker and this local variety became their main language. The majority uses or used mostly or only English with their spouse. The decline over generations is clear in that LC is rarely used with children or grandchildren. With the exception of one son who is the only child in the family mostly addressed in LC, English is dominantly used with the child generation. This is fully expected from the results on language proficiency where LC knowledge was shown to be much diminished in the child- and grandchild generations.

In casual private settings outside the family, namely with neighbors and friends, four interviewees claim to use mostly LC though nobody uses LC exclusively. It is important to note that two of these interviewees gave a second reply, which was *English* or *mostly English*. Their LC interactions are limited to specific creolophone friends and neighbors. LC is not the general code used with all interlocutors in the friends-and-neighbors-group but is consciously spoken to particular individuals. In one of the remaining cases it is surprising that the interviewee claimed to use mainly LC only with this group of interlocutors and English with all others with the exception of siblings where the bilingual option was chosen. This seems to indicate that the use of LC is probably also restricted to specific people, possibly a relatively large group, in this case, as it seems unlikely that LC would not be used with most family members or in shops but used as the default with friends and neighbors in general. More details would be necessary in order to make stronger claims on this issue.

In the more public or formal domains, such as with shopkeepers and co-workers, English is even more dominant. For both categories a large number of interviewees (eight for shopkeepers and seven for co-workers) did not give any information, either because they do not work anymore or because the question simply was not discussed or answered in the interview. Considering this, English clearly is the main language in these fields, with

11 and eight selections in the two categories respectively. In the cases where LC was spoken to some extent in the workplace, there was an LC-speaking co-worker rather than a common use of the language in the workplace.

This data points towards the important role of particular interlocutors in the usage of LC. Nevertheless, other factors may play a role in the distribution of language choice. This includes the suggestions made by Rottet and described above but may also involve social factors such as age, race, gender and place of residence. Recall that these are to be considered as possible influences as the sample is somewhat biased in each of these categories.

For the interlocutors at the time of the interviews age seems to play a minor role. Notably, but not very surprisingly, all three interviewees who use LC with their siblings or one sibling are from the oldest age group. Among the youngest age group LC is unusual between siblings and just one person uses LC and English bilingually in this context. Similarly, the use of LC with spouses is strongest in the 75+ category. The two couples who use mostly LC fall into this age group as do two of the three LC | English bilingual couples. The age patterns concerning language use with friends and neighbors are a bit unexpected as illustrated in Table 6-19. Here, the oldest group seems to favor English. Only in the two younger age groups did interviewees claim to use mostly LC with particular friends and neighbors. This may indicate a conscious effort made by younger speakers to maintain LC or simply a lack of available proficient LC interlocutors in the oldest category because many friends and acquaintances passed away, or both. For the other interlocutors age makes little difference, which is to be expected as the overall dominance of English is asserted for children, grandchildren, co-workers and shopkeepers, so results do not vary much.

Table 6-19 Language use with friends/neighbors by age group

	46-60	61-75	75+	Total
mostly LC	2	2		4
LC | English	1		3	4
mostly English	2	3	5	10
English	1	2	1	4
LRF			2	2
n/a	1		1	2
Total	**7**	**7**	**12**	**26**

Race also appears to have no major impact on the patterns of language use with different interlocutors. For the interlocutors at the time of the interviews results are more or less balanced with just minor deviations such as that two white speakers use LC with

siblings but just one black speaker claimed to do so. The most notable effects arise with interlocutors during childhood. Unexpectedly, LC appears to be more commonly used among the white speakers at the time. For example, seven white but only four black speakers used LC with their parents or one parent as is indicated in Table 6-20. In a similar vein, five white speakers claimed to have used LC with their siblings, as opposed to two black speakers. The same trend is found in the results concerning friends outside school, where five white participants but just one black interviewee selected LC.

Table 6-20 Language used with parents during childhood by race

	black	white	Total
LC	4	7	11
mostly LC	3	3	6
LC | English	3		3
mostly English	3		3
English	1	1	2
LRF		1	1
n/a	1		1
Total	**15**	**12**	**27**

These figures run counter to previous expectations, as the notion of LC as a language largely spoken by the black population would suggest a more common use among black speakers. Since the difference is no longer perceptible in the language pattern at the time of the interviews, this indicates that language shift was progressing faster and earlier in the black LC speech community. Possibly, some black speakers in the parent-generation felt more economic pressure to assimilate and considered knowledge of English an opportunity for their children to gain socioeconomic success. This is suggested by some comments made by interviewees about their parents' insistence on good English skills and their fear of a negative impact on those skills through usage of LC. While this is not a systematically studied aspect in the questionnaire, it is a plausible explanation for the pattern that white speakers more consistently used LC during their childhood than black speakers. Another factor may be the identity marking function of the language at the time. Since LC is less commonly associated with the white population, white speakers could have attributed more conscious value to using the language then black speakers. It served as a clear indicator of membership in some specific rural communities and hence may have been an in-group marker. Counter to such reasoning, the data on the best language presented above indicates that black speakers identify more strongly with LC and consider it more of an identity marker. From the data obtained in this study the reasons cannot be clearly identified and

most likely a mix of factors was at work. By the time of the interviews the shift to English had progressed far enough that no distinction remains and English is the dominant language for all speakers.

The picture that emerges when the results are sorted by the interviewees' sex is also most striking during childhood. It seems that girls were more often spoken to in LC than boys. This is true for all interlocutors in the home domain, such as parents, grandparents and siblings, as is illustrated by the tables below. Among friends, either in or outside of the school context, no difference between boys and girls is obvious. A likely reason for this is that women were expected to work around the house and the fields and raise children rather than acquire another profession, which was why the use of LC was not perceived as a possible hindrance. At the time of the interviews the difference had again almost fully disappeared. Only in use with siblings do women show a slight preference for LC as compared to male interviewees.

Table 6-21 Language used with parents during childhood by sex

	f	m	Total
LC	8	3	11
mostly LC	4	2	6
LC | English	2	1	3
mostly English	2	1	3
English		2	2
LRF		1	1
n/a		1	1
Total	**16**	**11**	**27**

Table 6-22 Language used with grandparent during childhood by sex

	f	m	Total
LC	11	6	17
mostly LC		1	1
LC | English		1	1
mostly English	1	1	2
LRF		1	1
n/a		1	1
Total	**12**	**11**	**23**

Table 6-23 Language used with siblings during childhood by sex

	f	m	Total
LC	6	1	7
mostly LC	3		3
LC | English	1	2	3
mostly English	2	4	6
English	1	2	3
n/a		1	1
Total	**13**	**10**	**23**

The distribution of language use with different interlocutors does not differ very much between the two regions in which the data was collected. Since the sample from the Bayou Teche region is much larger than that from the False River area, a direct comparison is difficult. Nevertheless, a few points catch the eye. First, the relative proportion of interviewees who spoke mostly or primarily LC with their parents is higher in the Bayou Teche (68,18%) than the False River (40%) region. More importantly, speakers in the False River area claimed to not or barely have used LC with their siblings as children. Only one of them uses English as well as LC with his siblings today. Beyond this instance none of the four interviewees from that region uses LC on a regular basis at all nowadays. Overall, this confirms the observation by Rottet (2001) as well as Klingler and Neumann-Holzschuh (2013) that language loss is more advanced in the Mississippi region. This is further confirmed in the *best language* and *language used most* sections, where the interviewees from the False River area favored English even more clearly then the interviewees from the Bayou Teche area.

Lastly concerning interlocutors, all interviewees were asked if there was a group or groups of people with whom they regularly spoke LC or which consisted mainly of LC speakers. Replies varied from plain denial, as the interviewees did not have such interlocutors in their life, over lists of efforts made to speak more LC, such as visiting French tables or visiting with elders in nursing homes, to naming family members and groups of friends that communicate in LC. All answers had in common that speakers referred to particular interlocutors rather than whole categories. Nobody claimed to use LC with their whole family or all their friends. Instead, interviewees stated that they used LC with e.g. “all of my 1st cousins” (CC), “my friends in Cecilia, my sister-in-law and […] one cousin” (RM), “only with Ms. Laura” (MR) or “my aunts and uncles” (RBM).

6.3.2 Functions

Interviewees were further presented with 15 statements concerning language use, such as "I use LC in my everyday life", "I use LC to tell jokes" and so on (for the complete list see the questionnaire in the appendix). These statements cover different domains of use (e.g. home, work, official business) as well as different functions (e.g. humor, singing, storytelling). The speakers were asked to assess if those statements were true for them on a scale ranging from 1-always, 2-often, 3-sometimes, 4-rarely to 5-never. Occasionally, the interviewees fluctuated between two answers or provided different specifications for different interlocutors. As with the interlocutors above, both answers are included in these cases, which is why the total can sometimes exceed 22.

One thing that stands out in this data is that despite the general low rates of use of LC that are described above, the picture here appears a bit more diverse. The first statement, "I have regular transactions in LC in my everyday life.", received more positive replies than expected. This statement was rated to be true sometimes, often or always 16 times out of 22 replies as is shown in Table 6-24. Consequently, LC is still used on a more or less regular basis by many speakers although it is not the language they use most and language choice is highly dependent on the interlocutor.

Table 6-24 Frequency for i. "I have regular transactions in LC in my everyday life."

always	2
often	7
sometimes	7
rarely	5
never	1
Total	**23**

For the present analysis, I have opted to represent replies as numbers in order to calculate the means and thus make the different statement somewhat more comparable. Where replies fell between two points on the scale because interviewees could not decide on a definite answer, an intermediate value was recorded as x.5. For Table 6-24 this equals a mean value of 2.82 on the 1-5 scale, where 1 is *always* and 5 is *never*. Table 6-25 below lists the results for all statements in this manner. The first noticeable fact is that the mean value is above 2.5 for all domains, functions and channels represented in the list. This indicates that use of LC on average ranges somewhere between *sometimes* and *rarely* in the strongest categories and is used even less in other domains or functions. This overall result supports the assessment of LC as a critically endangered language. A more detailed

consideration of the values in Table 6-25 reveals in which domains and functions LC is maintained best and where language shift has progressed the furthest.

Table 6-25 Average frequency rating of usage statements

Statement	Mean value
i. regular use	2.82
ii. home	3.95
iii. work	4.47
iv. running errands	3.95
v. official business	4.9
vi. think	2.89
vii. dream	3.24
viii. phone	3.14
ix. chat on the street	3.34
x. sing	4.35
xi. joke	3.55
xii. count	3.67
xiii. curse/cuss	2.75
xiv. secret code	2.76
xv. story telling	4.2

The more formal and/or public domains are covered by statements iii. “I speak LC at my work”, iv. “I speak LC running errands.”, v. “I conduct official business speaking LC (institutions, landlord etc.)” and ix. “I speak LC on the streets for a chat.”. The mean values for these categories are very high. Statement v. even receives an average rating of 4.9, which means official business is just shy of never being conducted in LC. The value for use of LC in the work domain is slightly lower but still very high with 4.47. Of the four, statements iv. and ix. have the lowest results with 3.95 and 3.34 respectively. This still means that the interviewees rarely use LC when running errands and a little less than sometimes for chats. Nevertheless, LC is more frequently used in this capacity than in the more formal domains, showing that though use is rare in all public domains, more casual public communication is more likely to contain LC. Oftentimes interviewees indicated that they would be comfortable chatting casually in LC but only did so with a few select interlocutors they were liable to meet by chance and who were fluent enough to hold a conversation.

On the other hand, statements ii., vi., vii., and xii. represent the private domain. This includes the home domain, where ratings are not much better than in the public one, namely 3.95. This is mostly due to a lack of interlocutors in the home when interviewees live alone or their spouses and/or children do not speak LC. Thinking (vi.), dreaming (vii.)

and counting (xii.) are functions that are subconsciously performed to a degree. The results from those statements hence provide insight into the psychological role of LC as opposed to its factual use in communication. Thinking and dreaming take place within a speaker's mind and without any interlocutor that may influence language choice. They consequently reveal how relevant or irrelevant LC has become in the interviewees' cognitive and psychological reality. The values for these functions at large are lower than those associated with actual speech acts. Thinking and dreaming range around the realm of *sometimes* with 2.89 and 3.14 respectively. Several interviewees were pleasantly surprised to realize that their dreams sometimes used LC, especially when they featured childhood memories or deceased family members such as parents or grandparents. This suggests that the choice to use LC is determined by the interlocutor even within dreams. Counting, with a mean result of 3.67, tends more towards *rarely* than *sometimes*, most likely because it is an activity that one often performs more deliberately than thinking or dreaming. Furthermore, counting is a rather formal process and often learned in or associated with school, which is clearly an English domain.

Rottet (2001) identifies humor as one of the main functions for which the endangered variety of LRF is used in the community in which he conducts his fieldwork. Even younger speakers who rarely use LRF at all would joke in their heritage language. Interestingly, this does not appear to be true for LC. Joking (statement xi.) scores at 3.55, placing humor somewhere between *sometimes* and *rarely*. Of course, this is rather frequent as compared to the use of LC for example in formal domains but it does not stand out as a major function for which the language is employed.

In a similar manner and counter to previous expectations, LC does not seem to fulfill much of a cultural maintenance function. The classic applications associated with such matters would be storytelling and singing songs, especially traditional stories and songs. These exist for Louisiana although they do not have the same spiritual significance they often hold for indigenous cultures and languages. Most famous are the humorous stories of "Bouki e Lapen" which are folk tales that were passed on orally and documented in written form by Louisiana scholar Alcée Fortier in 1894. In the foreword he already notes that the transmission of these tales seemed to have stopped with the Civil War and few people either remembered them or were willing to share them. Interviewees often referred to these tales when asked about storytelling and were certainly familiar with them but very rarely claimed to pass them on to the next generations. In accordance with this, storytelling is among the four functions that score above a value of four, which equals *rarely*, at 4.2. Similarly, singing songs has a value of 4.35 which places it among the four least frequent

functions. This is surprising, as music, especially Zydeco, is widely considered a main vehicle and motivation for the maintenance and popularity of LC. Interviewees commonly expressed the view that Zydeco had been keeping LC alive despite the lack of speakers and intergenerational transmission. Nevertheless, very few of them sing songs in LC themselves.

The highest levels of use and corresponding lowest mean values are found with two functions that are also performed within the private domains. First, one of the highest rates of LC usage is associated with cursing (xiii.) which holds an average value of 2.75. Speakers often admitted to doing this when agitated. Since manners and politeness hold high cultural value in Southern Louisiana, some interviewees were insistent that they never curse in any language, which received no value but was registered as n/a in order to avoid skewing the results. Nevertheless, it is quite possible that some consultants chose to reply *never* and were not referring to language choice but the action of cursing in general. Consequently, the frequency of using LC for this function among the speakers who actually do curse may be even higher than suggested by the mean value, which is within the range of *sometimes*.

The second equally frequent function is the use of Louisiana Creole as a secret code which is employed to keep others in the vicinity from understanding (xiv: 2.76). This usually applies to children without any knowledge of LC, which their elders use to their advantage when discussing matters they do not consider appropriate for the children. Some interviewees experienced the same behavior from their parents and grandparents but were exposed to LC often enough to acquire it to some extent, which is why the secret code strategy ultimately failed in their case. Since LC is not spoken nearly as much anymore, the younger generations do not pick up enough of it to be able to gather what their parents, grandparents or other relatives are saying. This result is in line with what Rottet (2001) found for LRF, as he identifies secrecy as the second major function besides humor. This appears to be similar for LC although results differ concerning other activities like joking and singing where LC is not maintained to a noticeably higher degree than in other functions. Lastly, speaking LC on the phone has a value of 3.14, which shows that the use of LC is not restricted to face to face communication.

Please note that all of these results, with the occasional exception of the work domain in statement iii., refer to the situation at the time of the interviews. In the discussion, interviewees often remarked on the changes they experienced with respect to language use during their lifetime. For example, AF remembered speaking LC at home with his wife in the past and KG used to always think in LC but rarely does so anymore. Similarly, GL stated that she makes a conscious effort to sometimes think in LC as a way of practicing and actively maintaining her heritage language at least in her mind. Also note, that

statement vi. “I think in LC” received particularly many replies. Several interviewees were unsure how often they use LC in their thoughts and could not decide on a definite answer, which is why the total of data points for this category is 26.

For each of the statements an analysis was conducted with regard to the four demographic factors that were identified as a potential bias for the present sample. Generally, there appear to be no major obvious influences of demographic categories on these choices. Nevertheless, a few noticeable differences exist such as between age groups for example.

Concerning regular use of Louisiana Creole, the middle age group surprisingly seems to make slightly less use of LC than the oldest and younger age groups. This pattern repeats with regard to use in the home and the work place and reappears when it comes to dreaming as well as chatting, singing, joking and counting in LC although the differences are often rather small. Either, a comparatively large number of interviewees aged 61-75 stated to never use LC for the respective function and/or a comparatively low percentage is represented in the highest categories which were selected for the respective function. As examples, see the table for dreaming below, where four out of six respondents from the middle age group selected *never* and the table for regular use where one person chose *always* but the remainder selected *sometimes* or less often than that.

Table 6-26 Frequency for vii. “I dream in Louisiana Creole” by age group

	46-60	61-75	75+	Total
always	2		2	4
often	3			3
sometimes		1	1	2
rarely	1		4	5
never	1	4	1	6
n/a		1	2	3
Total	**7**	**6**	**10**	**23**

Table 6-27 Frequency for i. "I have regular transactions in LC in my everyday life" by age group

	46-60	61-75	75+	Total
always		1	1	2
often	3		4	7
sometimes	3	4		7
rarely		1	4	5
never			1	1
Total	**6**	**6**	**10**	**22**

Race as a possible influencing factor seems to be relevant only for a small number of functions. Although the numbers sometimes appear dissimilar at first glance, the percentage of interviewees within each racial group who selected the two highest frequencies for each statement often is very similar. As an example, consider the use of LC in thinking in Table 6-28. The distribution appears to be quite different, which is true when considering each frequency level separately as well as the total of responses. But at closer inspection 33,3% of the white and 35,7% of the black interviewees think in LC always or often, revealing a clear similarity. Percentages for the home as well as the work domain or the use of LC as a secret code do not differ much when white and black creolophones are considered separately.

Table 6-28 Frequency for vi. "I think in Louisiana Creole" by race

	black	white	Total
always	1	2	3
often	4	1	5
sometimes	4	4	8
rarely	4	1	5
never	1	1	2
Total	**14**	**9**	**23**

On the other hand, a considerable difference can be observed in a few other functions, such as casual chats. 77,8% of the white interviewees stated they did so often or sometimes but only 37,5% of the black interviewees. Both groups sometimes use LC when running errands but roughly the double proportion of black creolophones never speaks LC in this context. White speakers also seem slightly more inclined to use LC on the phone often, although a larger part of black interviewees sometimes does so. In contrast, black speakers more often joke, count and curse in LC. Furthermore, storytelling in LC is only attested for black creolophones. No white interviewees ever do apparently, but two black interviewees stated that they often told stories in LC, another one selected *sometimes* and

four at least rarely employ LC to tell stories. In sum, the pattern concerning race is not very clear. In many functions no notable difference appears but where one is present a tentative conclusion may be that black speakers tend to restrict the use of LC to more to functions within the private domain (e.g. joking, counting, cursing) than white speakers who more often use the language in public. The use in traditional cultural functions such as storytelling and singing, though generally low, is more frequent in the black community according to this data.

The third demographic variable, the interviewees' sex, appears to make an obvious difference in a few functions but matters not at all in many others. It seems to be largely irrelevant for statements i., ii., v., x. and xv. In the workplace men more often claimed to sometimes or rarely use LC. When it comes to running errands, women are the most likely to use the language, with 58,3% stating to sometimes use LC in this context as opposed to 10% of the men. One reason likely is that woman are the ones who run errands in a relatively traditional gendered division of tasks, though this does not account for over 50% of the male interviewees stating that they never use LC running errands. There is no clear difference between men and women concerning thinking in LC, although women more often claim to always do so. This also holds for dreaming. Although the overall distribution for dreaming in LC at least sometimes or more often is rather balanced, women chose *always* more often than men. The picture is less nuanced with chats on the street, speaking on the phone and the use of LC as a secret code. In these functions, women seem to use LC more frequently than men. On the other hand, joking, cursing and counting are on the opposite end of the spectrum, as men clearly favor these functions in LC. Apparently men and women typically, but by no means exclusively, use LC in different functions. Figure 6-12 and Figure 6-13 represent functions more often performed in LC by men while Figure 6-14 and Figure 6-15 illustrate more typically female functions.

Figure 6-12 Frequency for xi. "I joke in LC" by sex

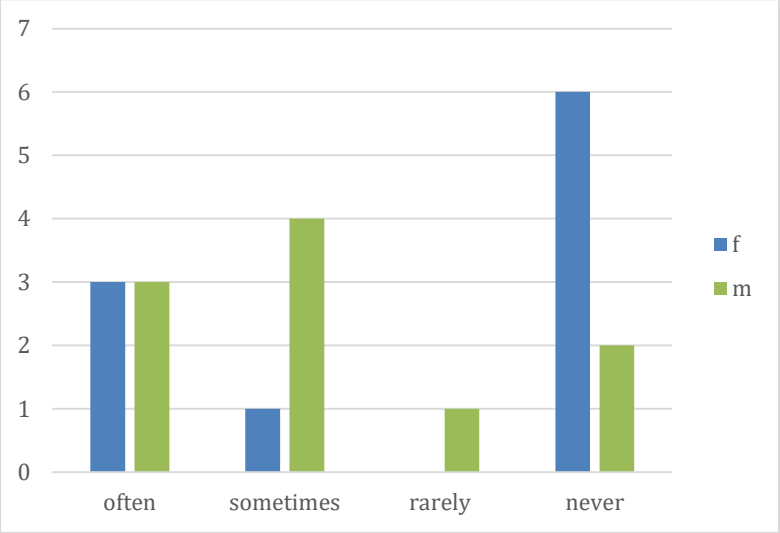


Figure 6-13 Frequency for xiii. "I curse/cuss in LC" by sex

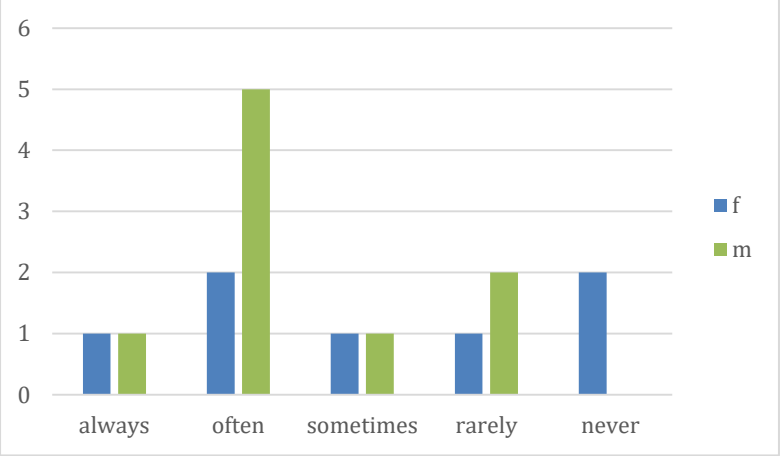


Figure 6-14 Frequency for ix. "I use LC on the street for a chat" by sex

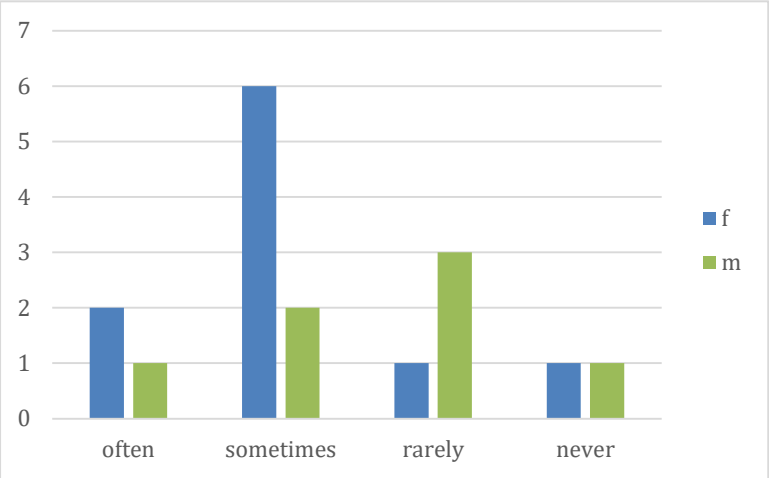
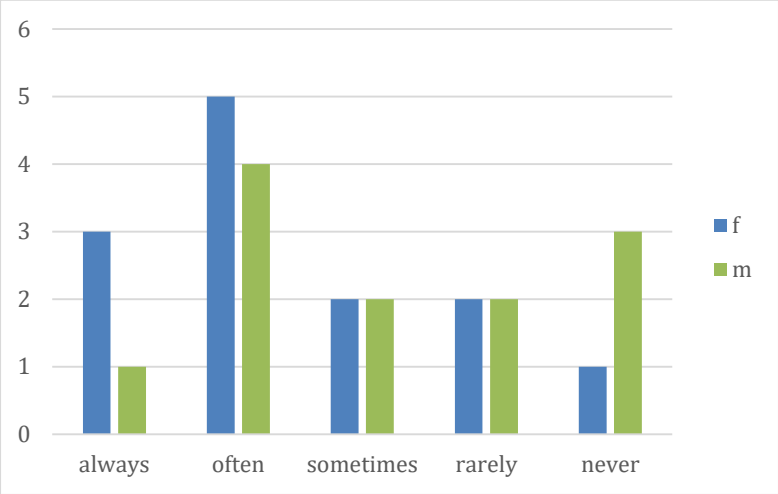


Figure 6-15 Frequency for xiv. "I use LC to keep others from understanding what I am saying" by sex



Lastly, the impact of regional factors remains difficult to judge because of the low number of interviewees from the False River area. The distribution of the frequency in which the different functions are performed in LC seems to be roughly parallel to that for the Bayou Teche region though on a generally lower level. Rarely does a reply from the False River creolophones appear within the highest frequency for a respective function. Most notably, there are a few statements for which the interviewees from the False River area select relatively high frequencies as compared to those from the Bayou Teche. This firstly concerns the workplace, where two respondents from False River stated to sometimes use LC. Since this equals 50% of the False River sample, it suggests a surprisingly high level of use in the work domain in this region. Counter to such reasoning, my impression is that these are individuals who happen to work with a fellow speaker and try to use the language occasionally. The view formed from conversations with creolophone

and non-creolophone locals is that LC is not generally used in the workplace and certainly not more often in the Mississippi region than the Bayou Teche area. Support is lent by the fact that the other two interviewees from that area stated that they never use LC at work and that interviewees from False River never use LC for official business. Other functions in which False River frequencies are relatively high are joking, cursing and the use of LC as a secret code. All four interviewees stated that they often cursed in LC and that they used it to keep others from understanding. These functions have in common that they belong to the private domain and the recipients are meant to be very limited. Speakers would rather curse in LC than English because it appears less rude to them when the people around them are unlikely to know what they are saying. Consequently, this is not necessarily an expression of the cognitive role of LC as the native language that makes an appearance in moments of agitation, as one might think. Since creolophones from False River do not use LC particularly frequently for thinking, counting and dreaming, such subconscious psychological relevance of LC is not likely to be the cause. Rather, the common thread seems to be secrecy in the sense that speakers know that only a select few will understand what they are saying when they speak LC. That secrecy would be a main function for which LC is used by speakers in the False River region fits the lower level of language vitality attested for the larger Mississippi area as observed for example by Mayeux (2019).

In sum, language use has changed significantly from the interviewees' childhood to the time of the interviews. English has replaced LC in most functional domains with the exception of particular interlocutors among friends and family members who speak LC. This is especially true for the public domain where LC is barely used anymore while it was not uncommon to speak it during childhood. In the formal work domain English dominates today and was largely dominant in childhood as well, because English was enforced at school. Generally, LC is more often used with older or elderly interlocutors and rarely with the younger generations. This trend was already recognizable when the interviewees were children but has since solidified to the point where it is highly unusual to speak LC with a young person. Similarly, the language is not usually used with strangers. Demographic variables influence the usage patterns, main domains and functions to some extent. Most noticeably, LC was more widely used among the white speakers than the black interviewees during their childhood. By the time of the interviews, language shift had progressed enough to negate the difference. A few functions appear to be more typically performed in LC depending on gender, race or age and language shift appears to be more pronounced in the False River region than the Bayou Teche area but, overall, the speakers in this sample use LC nowadays as a language of the private domain to be spoken with familiar interlocutors.

6.4 Language attitudes

Language attitudes as held by the government and the community are among the factors UNESCO suggests for assessing language vitality and the likelihood of language maintenance. Governmental attitudes of course go hand in hand with language policies, such as the suppression or support of a minority language and they often influence community attitudes. It is a commonly held view that community attitudes are central to language vitality, and Bradley (2002) even refers to languages attitudes as "the key factor in language maintenance". Considering the overall vitality of LC as it appears in the usage data as well as the history of stigmatization that creolophones experienced over the years, one would expect a rather poor attitude towards LC to be prominent.

On the institutional and governmental level this expectation is fulfilled to some extent. As described in Chapter 4, language policies never favored LC in any way and historically fostered a positive attitude towards monolingual use of English. In this respect, a paradigmatic change occurred in the late 20th century, which made the presence of French in Louisiana highly relevant in terms of culture, identity and economics, and stressed the Cajun regional identity and ties to Acadia. LC benefits from this change to some degree but by far not in the same way that SF and LRF do. These varieties are actively promoted through institutional efforts such as CODOFIL, immersion schools etc. Similar measures were never taken to support LC. It can be argued that LC is now tolerated or even looked at favorably from the governmental side simply due to the more positive perception of multilingualism in general and French specifically. But the lack of institutional support for the language in light of the previous negative attitudes held by officials and the public alike does not in any way create a framework for language maintenance. Consequently, it would be a stretch to attest positive governmental and institutional attitudes towards LC.

Community attitudes, on the other hand, seem to have changed more drastically than expected from the overall vitality of LC. As discussed in Chapter 4.3.4, researchers have long attested stigmatization of LC and reluctance to use the language publicly if at all (Rottet 2001). Judging from the lack of intergenerational transmission, the low usage rates and the reduction of functional domains, one would still expect negative attitudes. But when asked how they felt about their language, interviewees expressed 100 percent positive attitudes. Since the attitude questions were asked openly it is natural for answers to differ in wording but the sentiments expressed were always positive. "Love", "pride" and "heritage" feature among the most frequently used terms. HW for example claims that "the way words are expressed actually gives some excitement, some joy" and AF stresses that speaking LC with elderly community members is important because they "would just love it". Several

interviewees added that they used to feel ashamed of their mother tongue when they were younger but no longer felt that way. KG for example remembers that "they taught that the Creole people were the most uneducated people in the world", which is why he deliberately prevented his children from learning the language and never used it in front of them. Generally, such views are no longer in place. LM for example feels that "they can make fun of my language all they want. My language is my language." Nevertheless, the same speakers who now value their language so highly rarely use LC in public as described above and very few of them transmitted LC to their children. The one son who is fully fluent in LC actually refuses to speak it with his mother or anyone else in public. Despite assertions to the contrary, this behavior seems to indicate either that attitudes towards LC are more negative than they appear in the interviews or that a gap exists between language attitudes and linguistic behavior. Possibly, speakers changed their thinking but did not adjust their behavior accordingly. This appears likely because LC plays such a marginal role in South Louisiana and the speech community is spread so thin. By the time the community developed a more positive view of LC, creolophones had already taken to using English in many domains, children had become young adults who did not know their heritage language and the surrounding majority community was English speaking. In the wake of the efforts targeted at SF and LRF, the LC speaking community appears to have found a new sense of nostalgic appreciation for their language but language shift has already progressed quite far. At this point, positive attitudes alone cannot suffice to maintain the language. Rather, organized large-scale efforts are required to achieve LC learning by the child and grandchild generation in order to give LC the chance to become one of the native languages in the community again in the future. I believe that many locals would be interested and willing to learn or contribute but at the moment the movement has not yet spread very far beyond the online community of new speakers.

The interviewees generally have mixed expectations for the future of LC. Of the 21 interviewees who were asked if they thought LC would stay alive, seven expected that it would, four said that they did not know and ten assumed that LC would die out. These interviewees usually felt like MR, "I'm glad I was a part of it […] but I can see it going", HW, "We're losing it because we have not promoted it" or CC, "We have lost it. It's not too many people here who speak it". That a third of the interviewees would expect LC to stay alive is encouraging if a bit surprising. Six of these seven interviewees have in common that at least one of their children and sometimes their grandchildren have at least some knowledge of LC, although that may be just passive competence. In other words, the speakers who transmitted LC to the next generation to some extent appear to hold more positive

expectations about the language's future. Their personal experience of a certain degree of language maintenance within their families as well as the presence of LC in Zydeco music and the generally increased interest in LC were given as reasons for to the impression that LC will be maintained. Other interviewees, whose younger family members have some competence in LC as well, disagreed. They often claimed that they belonged to the last generation of speakers and that LC was going to die with them despite the limited knowledge of their children or grandchildren, as the younger generation rarely used the language and only ever with their elders. KG was clearly disappointed that it "don't look like the young kids are interested in it" and TR also thought that the young people "don't care, not [for] the Creole". This reflects some level of irritation as well as resignation not only with the younger community members but also with the fact that LC does not play a large role in French maintenance in Louisiana. In contrast, JN and MB, who believe that LC will be maintained, assumed that the younger generation would become more interested in LC as they grow older, because "there is a certain age that speaks it. The older they get, the more they're gonna learn, the young ones". This view is shared by RBM who feels that LC will remain "If they study it and keep it going. They have to do it!" AF for example, is convinced that LC will be maintained, although he himself did not teach the language to his children which he considers his "biggest sin". Nevertheless, he believes that the efforts currently made to preserve French in Louisiana will bear fruit for LC as well. Regardless, he considers SF more important for the younger generation and Louisiana at large. Similarly, GT feels that LC and SF share enough vocabulary that children who learn SF in school are well prepared for taking up LC. These speakers allow that LC might give way to a variety of French but are largely unconcerned about it either because "everything has to develop" (GT) or SF is simply seen as more advantageous. Such a view is not uncommon and even creolophones who actively promote LC repeatedly emphasized how their language allowed them to speak to Canadians, French people or Creoles from the Caribbean. Intelligibility and cross-cultural communication are thus seen as positive features and valued higher than distinctiveness. This view is opposed by those members of the speech community who are more concerned with heritage and cultural preservation. KG for example criticizes that "You can go to school to learn French. But there's nobody to teach the Creole".

All interviewees who expected LC to die out expressed the wish that their predictions would not become true but did not truly believe that LC would be maintained by the younger generations. Notably, all seven speakers who held positive views on LC's survival are from the Bayou Teche area. Interviewees from the False River Region do not share their optimism. One person from False River acknowledged an increased public and

academic interest in the language but remained uncertain about its future prospects, the remainder assumed that it would soon die out although they are "pressing now to try to preserve that. We have to be more persistent." (GH)

Regardless of what prospects interviewees ascribed to LC, they almost uniformly agreed that something should be done to keep LC alive. Two speakers were indifferent but the remainder thought that efforts should be made to maintain the language. The reasons why they "wish[ed] that everybody would speak it" (AF) often are to be found with notions of identity and heritage. 17 out of 21 interviewees considered LC to be important for who they are as a person, a part and marker of their identity. HW for example stresses LC's uniqueness and says that "it's an easier way of expressing. To show that comradery […] some feeling of family, some tie. We're still who we are." In a similar vein, GT argues

> I was afraid it was gonna die. But after so long, it's still here, I don't think it will. We're just too stubborn. It's familiar and it's comfortable. For people who want to be different or who want to be identified as unique or are glad to be what they are. You know, it makes you what you are. For me that's why I like it. You are who you are.

This sense of identity is largely attributed to the role LC played in the consultants' childhood, when many of them were socialized in LC. In the words of HB,

> I learn to live the Creole language, I learn to cook the Creole language, I learn to love my family the Creole language. […] I'm proud of that, that's how I learned to live.

Hearing and speaking LC brings back memories of simpler times for some speakers, like HB who felt that "[It] brings me back, it's something I wanna hang on to. […] When I talk Creole, I'm picking cotton again, walking barefoot." While these associations would have been rather negative in the past, when working on the fields meant that the family was poor and living off the land, they now carry an undertone of nostalgic almost longing appreciation. Many creolophones I spoke with acknowledge the role of LC as a formative part of their lives. But such sentiments tend be oriented toward the past. Consequently, some speakers stated that speaking LC used to be important for their identity but has ceased to do so. Now the importance of LC lies in a certain awareness and relevance of one's cultural and linguistic heritage rather than a major contribution to their current life and activities which is in accordance with the low rates of use of LC in everyday life.

Furthermore, interviewees were asked whether they considered speaking LC necessary in order to be considered a Creole person. Only five creolophones felt that was the case. Eleven speakers disagreed, usually because they considered *Creole* to be an ethnic category that is defined by French and sometimes African descent. TR for example is convinced that "You can be born a Creole and still not speak French." The remaining five interviewees did not have a clear opinion on the matter. These are interesting results, since it was often said that the answer would depend on what one considers a Creole person to

be. It would depend on the definition of *Creole*. There seems to be little agreement on the meaning of being Creole in the speech community (cf. the discussion of the term in Chapter 4). Even among those who said that being Creole required speaking Creole, few strong convictions were in place but speakers rather agreed that this seemed sensible. Some had looked up a definition before the interview and much discussion ensued around questions of ancestry and race. AH for example said "I don't even know if I'm defined as Creole" because he is white. A racial distinction persists in the community and is ascribed to the language as well. JB for example insists that the LC used by the white inhabitants of Cecilia "it don't sound like us. I don't know if it's because we black and they white. […] To me they don't sound like us." The diffusion of the meaning of *Creole* as a term further complicates matters as is evident from LJ's statement: "The white people speak a different Creole from us. They call it Cajun." It appears that *Creole*, most likely due to its ambivalent and sometimes contrary interpretations, does not serve as a strong identity label in the area. Despite its constant presence in New Orleans where regional cuisine, music and culture serve as tourist attractions and are labeled *Creole*, this role is fulfilled by the Cajun label in more rural areas. Creole has a multitude of meanings but it is not a frequently applied name for the people or the culture or even the language itself, which is most often referred to as *French*. Many Louisianans associate *Creole* with the Caribbean people, languages and culture instead.

English, which has become the dominant language in the speech community, is not resented for its role in replacing LC. Quite the contrary, all interviewees agreed that it was a good and necessary thing to speak English as citizens of the USA but also in a globalized world. This does not mean that they would be content to see LC fully replaced by English. Many creolophones hold the view that multilingualism is advantageous in many ways and want to see it promoted. As bilinguals themselves, they "recognize the importance of bilingualism" (GL) and believe that "you should learn different languages" (RBM). Some see advantages for cognitive development, as they perceive the minds of bilinguals to be more flexible, others stress the importance of knowing several languages for business and tourism. Knowledge of any form of French is considered especially valuable in the Louisiana context and many interviewees emphasized the importance of French being taught in schools. In this they do not necessarily refer to LC but any form of French, though many consider LC special in terms of heritage and their cultural identity. As HB puts it, "When you can speak Creole along with the other Frenches you're rich!" The distinction between SF, LRF and LC can be fuzzy at times, as mentioned above, and the general public in Louisiana tends to just consider all of these 'French'. Creolophones are generally aware of the

distinction but also value the similarities between LC and varieties of French as well as other French-based Creole languages. For example, RBM claims that, "If it's the real French I may not understand everything but a little bit".[36] The consensus seems to be that through knowledge of LC many languages and dialects can be understood to some extent and cross-cultural communication is facilitated in many places.

Within Louisiana interviewees perceive LC to be more or less homogenous. Some are aware that the language differs from community to community but they often consider these differences to be minor ones that mainly concern phonology and the occasional word. The existence of two main regional varieties as proposed and largely accepted in academia (Klingler 2003a) was not mentioned once during the discussions and creolophones commonly do not appear to be aware of it. Arguably, most of them have little to no contact to speakers from the other region. Most interviewees were unable to judge if younger speakers spoke a different form of LC simply because they knew no young speakers. The LC spoken by someone from the child or grandchildren generation appears to differ very little, usually because these speakers acquired the language from their elders. Certainly, the new speakers who learn LC via social media currently are the most active in language maintenance and their speech features some structural differences. However, Mayeux (2019) argues that at present these do not suffice to define a distinct variety of 'Neo-Louisiana Creole'. Of the creolophones interviewed for this study, none were in contact with any of those new speakers.

[36] Also note the reference to the *real French* and the hierarchical perception inherent therein.

7 LINGUISTIC DATA ANALYSIS

The discussion in this chapter centers on question formation in LC. Before the results from the present study are presented and discussed, a brief overview of related research is given.

7.1 Previous research on wh-questions in Creole languages

Wh-questions are a major field of study in generative syntax and a large amount of research has been concerned with such structures in specific languages as well as cross-linguistically (cf. the groundbreaking works of Ross 1967, Chomsky 1977, Huang 1982b, Cheng 1997, Rizzi 1997, Cinque 1990, as well as more recent proposals such as Cable 2010, Bošković 2002, and the contributions in Cheng and Corver 2006). Naturally, the topic has received some attention from creolists as well, albeit on a much smaller scale. Most general works on Creole languages and grammar sketches of particular Creoles tend to mention question formation, which includes wh-questions, but do not discuss it in any detail. A number of studies have concerned themselves with interrogative structures specifically.

Clements and Mahboob (2000) assume that question words in most cases developed during Creole genesis and are relatively resistant to later changes. They presuppose that wh-expressions in pidgins and Creoles are bimorphemic due to a high level of semantic transparency as compared to monomorphemic expressions (e.g. *what person* rather than *who*, *what place* rather than *where*) (Clements and Mahboob 2000, 461). This indeed appears to be a cross-linguistic tendency in Creole languages although exclusively monomorphemic question words are attested for approximately 30% of the contact languages in APICS (Michaelis et al. 2013). Clements and Mahboob (2000) take a cross-linguistic approach comparing several Creoles to their lexifiers and conclude that the interrogative structures found in Creole languages are results of MUTUAL LINGUISTIC ACCOMMODATION (Thomason and Kaufman 1988) interacting with universal tendencies. They predict the presence/absence of a question or focus particle as well as the position of the wh-word on the basis of the languages that were present at the time of Creole formation as well as their respective importance at the time.

Wh-movement is generally attested for a number of Creoles, although the exact analysis remains controversial. Hirokuni (2006, 280) considers wh-movement a central operation in syntax and gives examples of wh-initial clauses from Hawaiian Creole, Guyanese Creole, Jamaican Creole and Indo-Portuguese Creole. The paper stresses the cross-Creole similarity between those languages and argues that Hawaiian Creole's main substrates do not exhibit the same pattern. However, the conclusion that this supports genesis accounts that rely on language universals over substrate theories disregards the

possibility of English superstrate influence. In the APICS (Haspelmath and Consortium 2013b) the position of wh-elements is considered in feature 12 which distinguishes between clause-initial position and non-initial position either referring to in-situ expressions or to constituents that are moved to another non-initial position. In the APICS data, initial wh-expressions indeed appear to be the dominant strategy but half of the languages allow another structure, such as in situ phrases. A number of Creoles also rely exclusively on non-initial interrogative expressions. This seems to be especially common in the African and Asian Creoles.

Support for a wh-movement analysis for Creole wh-initial questions in Creole languages comes from Du Plessis (1977), who argues in favor of successive cyclicity with data from Afrikaans, where the preposition can be stranded not only in its base position but also in intermediate landing sites. In a similar vein, Muysken and Law (2001) attest that HC is sensitive to wh-movement of subjects as a morpheme *ki* is inserted only when the subject of an embedded clause is wh-moved. The further implications of these findings are discussed in Chapter 8 below.

Recall that in the spirit of minimalist syntax, movement is in fact an instance of internal merge, which is considered more costly than external merge (Chomsky 2000). Movement hence is a last resort operation that requires more processing efforts than merger of a new constituent that is not yet part of the derivation (cf. Chapter 3). Continuing along this line of thought, it stands to reason that movement is a comparatively complex syntactic operation. Complexity remains a buzzword in creolistics, where much debate revolves around the notion of supposedly simple Creole grammars (among many others: McWhorter 2001, Parkvall 2008, DeGraff 2001, Lefebvre 2001). Parkvall (2008, 255), following McWhorter (2001), considers an "expression [...] more complex than another if it involves more rules, i.e., if it requires a lengthier description" which they take to be independent from expressive possibilities offered by the language. Combining both lines of reasoning, syntactic movement of any kind should be considered a complex linguistic feature as it certainly fits the definition compared to alternative strategies such as in situ constructions or functional particles. Surprisingly, this connection is not usually made and I am unaware of any studies on the (lack of) complexity of Creole grammars that consider movement operations. At the same time, focus fronting appears as a typical Creole feature in Bickerton (2016, [1981]) despite its obvious complexity.

One related debate in Creole studies is concerned with the analysis of focus- or wh-initial sentences. These could either be instances of fronting, which would imply syntactic movement, or they could be considered bi-clausal cleft constructions, in which the focus

constituent would be base-generated in the initial position. Durrleman and Shlonsky (2013) building on Durrleman (2008) discuss the matter extensively for Jamaican Creole (JC) but propose an extension to other Creole languages. Their main interest is in the Jamaican particle *a* which appears before a focused or wh-constituent as in (90) and (91).

(90) A Jiemz$_i$ im$_{*i/k}$ muma lov bad. (Durrleman and Shlonsky 2013, 92)
A James 3SG mother love bad

It's JAMES his mother loves a lot.

(91) A (h)uu im bring dat *fi/ fa? (Durrleman and Shlonsky 2013, 92)
A who 3SG bring that PREP

Who did s/he bring that for?

As illustrated by the second example, JC has a distinct form for stranded prepositions, which must be used before the silent copy of a wh-moved constituent. That is *fa* in (91), where the default form *fi* is not acceptable due to wh-movement. The authors take this as evidence in favor of a movement analysis, since the form of the preposition is determined by the extraction of the wh-phrase. This would not be feasible if the wh-expression was base-generated sentence-initially. They further discuss if such constructions can be analyzed as biclausal clefts which would imply that *a* is a copula in the examples above. Despite the existence of a homophone copula *a* in JC, Durrleman and Shlonsky (2013) conclude that the particle in focused or wh-constructions is not a copula since it does not require a subject, can be preceded by TMA markers and can occur with AP and PP predicates which is impossible for the copula. If *a* is no copula, the construction is less likely to be a cleft construction. In addition, the authors show that focus *a* can appear inside a relative clause, which clefts cannot, and that no subject-object asymmetry exists for the JC construction. Such an asymmetry is attested for example for French, where a cleft construction can be used to answer a subject question but not an object question. For these reasons, they prefer a focus movement analysis.

Their analysis centers on exhaustiveness, as the exhaustive interpretation is obligatory in context where the marker *a,* which they analyze as an exhaustiveness operator, is obligatory. Ultimately, they propose a split CP analysis for JC but suggest an Exhaustiveness Indicator Phrase (EIP) rather than a Focus Phrase, as this accounts better for the JC data. Wh-moved constituents do not always require an exhaustive answer in JC unless they are paired with *a*. When this is the case they are assumed to occupy SpecEIP, when they do not they are situated in a wh-projection within the left periphery. This projection is potentially part of a complex head that contains both [EI] and [wh] features. The idea is somewhat extended to other language types. Durrleman and Shlonsky (2013, 102) suggest a quantificational head in the left periphery that is "a critical probe for a

quantificational feature, a superclass feature that can be further specified [wh] or [focus]". Clause typing in their account is located in another higher projection that probes the [Q] feature on the wh-expression. The proposed analysis may need to be tested and extended in a more in-depth account containing detailed derivations. Nevertheless, it contains a number of interesting ideas that can be of use for the present endeavor.

First, it is noteworthy that JC marks wh-movement on the stranded preposition, hence providing clear evidence for such movement to the left periphery in a Creole language. Second, the analysis suggests the possibility that wh-expressions front due to their quantificational properties, which they share with focused elements, thus explaining the often complementary distribution of wh-phrases and focused phrases while still allowing separate features though not necessarily projections.

Veenstra (2008b) holds an opposing view. He claims that focus constructions in Creole languages are best analyzed as biclausal clefts rather than instances of internal merge. Since this study is concerned with wh-movement rather than focus constructions, his arguments pertaining to focus-clefts are not discussed in any detail. Suffice it to say that he finds the evidence from topic and focus placement insufficient and stresses that tense, mood and negation markers as well as modals can occur between the supposed focus marker and the focused DP in JC as well as HC, suggesting a biclausal structure with the full range of projections below the CP level. This is relevant for the present discussion if one assumes that focus movement and wh-movement are closely tied or potentially identical operations for which no indication appears in the LC data. This would then imply a potential analysis of wh-questions in Creoles as cleft constructions. For wh-questions that do not contain an overt question or focus particle, Veenstra does not explicitly suggest such an analysis. In fact, he rejects the general correlation between wh and focus. The questions that arise from the two approaches with regard to LC is if the LC data supports a movement or a biclausal analysis for wh-questions.

Alexandre (2012) provides an in-depth minimalist account of interrogative structures in Cape Verdean Creole (CVC), which, according to her analysis, exhibits overt wh-movement. Rejecting the cartographic approach in favor of a more minimalist analysis, she assumes that overt complementizers occupy C^0, which type the clause in accordance with the CTH. Clauses are typed as interrogatives by the Q-feature which is present on *ki* (for direct questions) and *si* (for indirect questions). In addition, CVC has a formal feature [wh] that is present not only on interrogative complementizers but also with relative clauses and cleft constructions which are introduced by *ki*, *pa* or a zero constituent.

CVC employs a variety of strategies for question formation. Wh-movement with a null spell-out of the lower copy, a gap, is the standard strategy.

(92) Kusé ki txiga [~~kuse~~]? (Alexandre 2012, 79)
What that arrive GAP

What did arrive?

Other strategies available in CVC are pied-piping (cf. (93)) as well as stranding of prepositions and preposition chopping. Preposition stranding causes the appearance of *el* in the position of the lower copy (cf. (94)), which she analyzes as a spelled-out trace or imperfect copy of the wh-phrase that appears invariantly as the third person singular form. With the chopping strategy (cf. (95)), the stranded preposition is deleted at PF. Note that neither pied-piping nor chopping appear to be available for a certain class of prepositions she classifies as 'heavy'.

(93) Bu ka sabe ku kenha ki bu sa ta papie [~~ku kenha~~]? (ibid., 81)
2SG NEG know with who that 2SG PROG talk GAP

Don't you know with whom you are talking?

(94) Ken/kenha ki bu sa ta papia ku- el? (ibid., 83)
Who that 2SG PROG talk with-3SG

Who are you talking with?

(95) N purgunta ki skolas ki Maria ta trabadja. (ibid., 84)
1SG ask which schools that Maria IPFV work

I asked which schools Maria works. (at)

CVC also allows for in situ constructions in main clauses, although their interpretation appears to be mainly echoic. Alexandre (2012, 93) reports diverse judgments for embedded contexts.

As indicated by her proposal for the role of C^0, all examples contain an element *ki* glossed as *that.* Alexandre (2012, 94) takes this to be the complementizer which contains the formal features [+D] and [+T], referring to nominal properties and finiteness. It also has the ambiguous features [Wh] and [Q]. The latter would be responsible for clause typing as an interrogative. Assuming that the notion of feature ambiguity is used in the sense of Simpson (2000), the idea is that C^0 can potentially be specified for [+Q] and/or [+Wh]. Simpson (2000, 104) assumes that "C^0 is critically ambiguous prior to wh-movement in ranging over a variety of potential values—focus, +wh+Q, yes/no+Q etc—and that wh-movement into Spec of C^0 will function to disambiguate C^0". According to Cheng and Rooryck (2000), discussing wh-in-situ in French, disambiguation of C^0 can occur via a number of operations. Narrowing Simpson's proposal, they take the feature on C^0 to be a Q-feature which needs checking and can be further specified for [wh] or [yes/no]. The Q-feature in French can be checked via wh-movement or rising intonation which represents

the PF realization of an intonational Q-morpheme. For languages that contain Q-particles, these are also able to enter a checking relation with C^0, thus valuing the Q-feature. In their analysis, wh-movement into SpecCP results in a wh-question yielding the feature configuration [+wh, +Q]. Rising intonation has the default setting [y/n] unless an in situ wh-phrase remains within C^0's checking domain, which would result in a wh-interpretation due to an LF operation. This is how they account for the optionality of wh-movement in French. Note that an ambiguous head in the left periphery that can be specified for a number of feature settings is quite similar to what Durrleman and Shlonsky (2013) suggest for Italian-like languages, when they propose a quantificational feature that can be specified for [wh] or [focus]. In principle such underspecified heads only differ from many other variable underspecified features assumed in syntax (in the sense of Rooryck (1994)), such as phi-features, in that their settings are not as obviously related under the same superfeature as e.g. singular, dual or plural settings for [number] or the different [person] values. According to Alexandre (2012), wh-phrases in CVC move into the specifier position in order to disambiguate the [Q] and [wh] features via feature checking. The study's main focus is on the phenomenon exemplified in (94), the insertion of *el* after stranded prepositions. She proposes an extension of the Copy Theory of Movement in which the lower copy of the wh-phrase is defective so that it is not deleted but converted post-syntactically into the default pronominal form *el*. This is a PF rescue operation to avoid full preposition stranding.

Adone and Vainikka (1999, 78) study the acquisition of wh-constructions by children acquiring Morisyen, the Creole language of Mauritius (MC), under the premise of testing Bickerton's hypotheses, according to which Creole languages resemble unmarked UG structures more closely than other languages. It follows that children should more easily (i.e. with fewer errors) acquire Creole languages. Since this study is 20 years old theoretical positions have since evolved and MC itself has undergone language change towards French (Adone p.c.). A repetition of the study would certainly be fruitful to uncover language change in MC similar to the analysis conducted below with regard to the observations of LC by Rottet (2006). Nevertheless, the study's results relate to several other studies in creolistics and in FLA and can inform the present discussion keeping their frame of reference in mind. Adult MC appears to behave as is expected from a Creole. It exhibits the typical bi-morphemic wh-words and fronts the wh-expressions without auxiliary inversion. In MC a complementizer *ki* occurs, which the authors analyze in a similar manner to CVC *ki* (cf. Alexandre 2012 above) as an overt C-head (cf. (96)). Remarkably, they find that this complementizer cannot co-occur with a monomorphemic, fronted wh-word such as *kan* (=when) or *kote* (=where) as illustrated by (97). The authors hypothesize that such

monomorphemic expressions occupy C^0 rather than the specifier position, hence blocking the use of another C-head. They recognize that while this accounts well for the distribution of *ki*, it requires certain wh-adjuncts to be considered heads and raises the question of their origin. If they are base-generated in this position the analysis needs to account for the co-indexation of the vP adjunct position. They opt to assume that these expressions are regularly wh-moved but cliticize to C^0.

(96) Kisana ki to ti don sa liv la? (Adone and Vainikka 1999, 78)
Who that 2SG TNS give DEM book DET

Who did you give this book (to)?

(97) *Kan ki to pu vini? (Adone and Vainikka 1999, 79)
When that 2SG MOD come

When will you come?

The authors focus on long-distance movement operations and island effects in their study. They find that MC is less vulnerable to island effects than English. MC allows extraction of a wh-phrase not only out of an embedded clause as in (98) and across a wh-adjunct as in (99), which is marginal in English, but even across an intervening wh-argument as in (100), which is ungrammatical in English.

(98) Kinsala Pier fin rakonte ki li ti pu zwen ~~kinsala~~?
Who Pier ASP tell that 3SG TNS MOD meet

Who did Pier say that he should meet?

(99) Kinsala Pier fin panse kimanyer li ti pu zwen ~~kinsala~~ ~~kimanyer~~?
Who Pier ASP think how 3SG TNS MOD meet

Who did Pier consider how he should meet? (English: ??)

(100) Kinsala Pier fin panse kizafer Mari ti pu kapav done ~~kinsala~~ ~~kizafer~~?
Who Pier ASP think what Mari TNS MOD can give

Who did Pier consider what Mari can give (to)? (English: *)

Interestingly, this pattern is not borne out in the data they obtain from children. In contrast to adult MC speakers, children appear to obey an English-type argument-adjunct distinction with regard to wh-extraction islands. Adone and Vainikka (1999) conclude that this is likely to be the unmarked setting in UG but remain uncertain as to why the distinction is lost in adult MC. These empirical results also support the notion that the monomorphemic wh-expressions occupy C^0 as these are adjuncts and children allow wh-movement across them, which would be expected if the specifier was available as an intermediate landing-site. Whether this strategy is available due to the monomorphemic or the adjunct status of the items remains unresolved.

Moreover, MC allows for partial wh-movement[37], similar to constructions found in Germanic languages such as German. In such sentences the matrix wh-position is occupied by a wh-phrase that can be considered a semantically empty scope marker or potentially a moved constituent see McDaniel (1989) for an early analysis and the papers in Lutz, Müller, and Von Stechow (2000) for diverse approaches. Regardless, the second wh-phrase remains in an intermediate position, namely the CP of the lower clause. It is this lower question word that is interpreted and answered. This is illustrated by Adone and Vainikkas' (1999, 82) German example for partial movement of the object wh-phrase with the fronted expletive *was* (cf. (101)) as well as its MC counterpart in (102).

(101) Was glaubt Hans mit wem Jacob jetzt spricht?
Q believe H. with whom J. now speaks

lit. What does Hans believe who Jacob is talking with?

Who does Hans believe Jacob is talking with?

(102) Ki Zan ti krwar ar kinsala Mari ti pe koze?
Q Zan TNS believe with whom Mari TNS ASP talks

lit. What does Zan believe who Mari is talking with?

Who does Zan believe Mari is talking with?

Examples from Adone and Vainikka (1999, 82)

Both examples are fully grammatical. They are to be answered with who the embedded subject is talking to according to the matrix subject. In other words, in (101) the speaker asks for the identity of the person that Jacob is speaking to according to Hans. If Hans believes that Jacob is speaking to Paul, that is the required answer, even if Jacob is in fact speaking to James. The children studied by Adone and Vainikka (1999) preferred this type of construction, predominantly choosing a partial interpretation over a long-distance or even a short-distance interpretation of the questions with which they were presented. Since this pattern also has been observed in English first language acquisition, the authors conclude that it is likely to be the default strategy and hence one aspect in which MC indeed appears to rely on unmarked UG patterns.

In sum, Adone and Vainikka (1999) make some interesting predictions to keep in mind for the analysis of LC. For example, the question arises if LC allows wh-extraction across intervening argument and/or adjunct wh-expressions as well as partial wh-movement constructions in the manner of MC.

[37] The construction is more commonly referred to as WH-SCOPE-MARKING in more recent publications. PARTIAL WH-MOVEMENT has become the term for similar constructions in which wh-movement does not go beyond the embedded clause but that do not contain a wh-scope marker in the matrix CP (cf. Lutz, Müller, and Von Stechow 2000). This thesis is only concerned with wh-scope-marking constructions but occasionally refers to a 'partial' interpretation to indicate the semantic relevance of the wh-phrase in the embedded CP.

The review of previous research on interrogatives in Creole languages opens some lines of inquiry this study pursues further for the LC case, as listed for each publication above. The main issues to consider are, which (morpho-)syntactic structures can be used to form interrogatives in LC? Is movement involved in the derivation or are these biclausal cleft constructions? If wh-movement occurs, is this a successive cyclic operation? Does LC employ a focus/interrogative particle? What is the status of *ki*? Does LC show island effects and if so with respect to which kind of interveners? Are effects of language endangerment observable in question formation? Where data is not homogenous, (how) do the differences relate to demographic and sociolinguistic variables?

7.2 Previous research on interrogatives in LC

Interrogatives in LC are under-researched, especially considering the comparatively complex pronominal system in place in LC (Neumann 1985, 333). The major works on LC by Neumann and Klingler offer brief descriptions of the interrogative system but no detailed analysis, only listing the question words and strategies available in the language. Neumann (1985) lists interrogative pronouns in the Teche variety according to functional category. The subject pronouns *sa-ki, ki-ki, ki-se-sa-ki* and *ki-se-ki* apply regardless of animacy or humanity of the referent. Object questions on the other hand distinguish human and non-human referents in the choice of interrogative pronoun. For human referents Neumann lists *ki, ki-se, ki-se-sa, ki-ki/ke, sa* and *sa-ki/ke*. For object questions with non-human referents only a subset can be employed, namely *ki, ki-se, ki-se-sa* and *sa*, with *sa* being the most frequent. She further notes that *ki*, *ki-sa* and *ki-se* are neutral interrogative pronouns, resembling French *quoi*, but does not give any further analysis of their status. For copula questions most of the above can be used for both, human and non-human referents. *Ki* and *ki-ki* are only appropriate in a [+human] context and *ki-sa* and *sa-ki* can only be used with [-human] referents in copula questions. She also lists adjunct/adverbial interrogatives. Importantly, as mentioned above, she maintains that the list of interrogatives is incomplete. Notably, the interrogative pronominal system and the distribution of forms appear to be highly complex and varied which necessitates further research. According to Klingler (2003a), the Pointe Coupee variety also distinguishes between subject and object contexts and between human and non-human referents. The form *ki-moun* is used exclusively for human referents, subject or object. The forms *sa-ki* and *ki* can refer to human and non-human referents but *sa-ki* is limited to subjects, while *ki* by itself can also be used for objects. *Sa* as a monomorphemic interrogative pronoun only occurs in object contexts with non-human referents. *Lekèl* and *nekel* on the other hand only are used for non-human

referents, regardless of their grammatical function. The distribution of wh-pronouns as described by Klingler (2003a) and Neumann (1985) is summarized in Table 7-1.

Table 7-1 Interrogative Pronouns in LC

	TLC (Neumann 1985)		MLC (Klingler 2003a)	
	Subject	Object	Subject	Object
[+ human]	*sa-ki,* *ki-ki,* *ki-se-sa-ki,* *ki-se-ki*	*ki,* *ki-se,* *ki-se-sa,* *ki-ki/ke,* *sa,* *sa-ki/ke*	*ki-moun,* *sa-ki,* *ki (rare)*	*ki-moun,*
[-human]	*sa-ki,* *ki-ki,* *ki-se-sa-ki,* *ki-se-ki*	*sa* *ki,* *ki-se,* *ki-se-sa*	*sa-ki,* *ki,* *lekèl,* *nekel*	*ki (rare),* *sa,* *lekèl,* *nekel*

Interrogative pronouns in LC are monomorphemic with the exception of *ki-moun* ('who(m)') which is only attested for the Pointe Coupee variety. While this is not entirely unusual for contact languages, most Creoles have compound expressions "consisting of a generic noun and an adnominal interrogative word" (Haspelmath and Consortium 2013c). This is another aspect in which LC seems to follow the lexifier more closely than other Creole languages (cf. Chapter 4.4) such as HC, which has four compound and three monomorphemic expressions (Fattier 2013). Interrogative pronouns appear sentence-initially in LC, which is expected from the Atlantic Creoles. Both Neumann (1985) and Klingler (2003a) further mention the fact that polar questions can be formed simply by rising intonation without the presence of an interrogative marker or any changes in word order. This strategy is more common amongst Creole languages than other languages, although the use of a question particle also is not unusual for Creole languages (Velupillai 2015, 507). According to Veenstra (2008b) "rising intonation in yes/no questions instantiates a Q-morpheme", as discussed in the aforementioned work of Cheng and Rooryck (2000) (see Chapter 7.1 above).

LC is very briefly mentioned in Clements and Mahboob (2000). They find that, counter the predictions made by their model, LC does not exhibit a focus particle *ki*. This is ascribed to the word-final position of *ki* in several question words in the language. This

view is challenged in a study by Rottet (2006), who solely considers LC, analyzing 325 questions from 13 sources. Much of his data stems from older sources, the most recent is Ancelet (1994). Below a comparison of his results with the distribution of wh-expressions in the more recent data from the present study is conducted to uncover language change in the interrogative system. Rottet's analysis detects a distinct distribution of question words with and without word final *ki*. He maintains that the variation marks a regular grammatical distinction, namely that between subject and object questions. Final *ki* only appears in subject questions and not in object questions. Rottet (2006, 244) concludes "this language does use a focus particle *ki* with subjects but not with direct objects, thus bringing this variety in line with the prediction of Clements and Mahboob (1999)". In this view, the final *ki* in expression like *ki-sa-ki* or *ki-ki* is hence analyzed as a focus particle marking subjecthood. For object questions, this implies the presence of a focus marking zero constituent (∅) in the spirit of a uniform analysis. In light of these studies, the status of *ki* appears to be an important consideration in the analysis of the LC data. The distribution of *ki* in the present data is not regularly conditioned by grammatical function as illustrated below. I argue that despite the grammatical function it fulfills in Rottet's data, the final *ki* in the present data no longer systematically functions as a marker of subjecthood, nor is it a complementizer introducing a relative clause, which would indicate a cleft construction. Instead this element seems to have lexicalized to some extent and is now simply the optional final segment of a question word for many speakers although it retains some sense of a grammatical marker in that the distinction is still present as a tendency rather than a rule.

On the basis of the previous research on LC and interrogatives in other Creole languages, the analysis below sets out to

a. describe possible structures of interrogatives in LC

b. uncover the current distribution of wh-expression with special regard to final *ki*

c. discuss effects arising from language contact and language endangerment

d. provide a generative analysis

These issues are interconnected to such a degree that they cannot be treated fully separately in the analysis. The distribution of *ki* for example is necessarily linked to its grammatical analysis, which has implications for the generative account. Similarly, the description of marginal structures which are not standard strategies but are accepted by some speakers to some extent are relevant to capture the subtle effects of language endangerment and the influence of contact languages. Furthermore, a number of more detailed questions are implied in the above objectives, such as the syntactic analysis

questions listed in Chapter 7.1. which are concerned with syntactic positions, movement and locality.

7.3 Questions in LC

The present and the following chapters seek to provide the relevant data on a number of the aforementioned issues and research questions on the basis of the acceptability judgments obtained through the Manipulated Translations Acceptability Task (MTAT), the 'Questions after Stories' task and the spontaneous data collected during fieldwork. Discussion then moves to the implications of these data for the analysis. All examples presented below are elicited translations or modifications thereof that were presented for acceptability judgments. For practical reasons, sentences are consequently not explicitly marked as translations. Instead, examples that stem from recordings of spontaneous language receive the marking [SD].[38] Sentences that were not directly offered by the interviewees but constitute manipulations of their translations which were rated for acceptability, are marked as manipulated translations [MT]. First, the constructions available for question formation in LC are listed and exemplified. The structures which were univocally accepted or produced are assumed to represent the default strategies in LC. On the other hand, structures which were rejected across the board are taken not to be part of LC grammar but are briefly mentioned. Particularly interesting in the context of language endangerment and language change are marginal constructions that appear to be part of only some speakers' grammars. These are described where they emerge from the data. In the ensuing discussion in Chapter 7.4 I turn to the distribution of wh-pronouns, the status of *ki* and patterns that appear in relation to demographic variables. Finally, a generative analysis of the present data in accordance with the research questions listed above is conducted in Chapter 8. The focus of that discussion is a special type of long-distance question, termed *ki*-insertion-clause, which appears in the data.

As expected, a number of interrogative structures appear to be readily available in LC. Each of these options is described and illustrated below. Some variation is present in this data for example when it comes to the choice of wh-expression. In the present description the least complex or most common expressions are selected for presentation but meaningful variation is discussed in Chapter 7.4, particularly with regard to the choice of wh-words and the appearance of final *ki*.

[38] Several examples also are taken from other sources which are indicated accordingly. Examples from the Louisiana Creole Diachronic Corpus (LCDC) database referenced as: LCDC followed by speaker initials.

7.3.1 Root questions

Root questions which ask for subjects as in (103) and (104), objects as in (105), or adjuncts as illustraed (106) by the means of a fronted wh-expression are one default strategy.

(103) Ki arive/vini? (GL)
Who arrive/ come

Who arrived?

(104) Ki-k a arive avèk twa? (LCDCRM, Mayeux 2018) [SD]
Who ASP arrive VL with 2SG

Who is going to arrive with you?

(105) Ki Zan konye? (HW)
Who John hit

Who did John hit?

(106) Eou Zan kouri? (LM)
Where John go

Where did John go?

The structure exemplified above is akin to standard wh-movement questions as found in English and Creoles such as CVC (cf. Chapter 7.1 above, (Alexandre 2012). The wh-expression occupies a sentence initial position and leaves a gap in its base position. When the complement of a preposition is the target constituent, preposition stranding occurs. A translation that structurally corresponds to example (107) was preferred by all speakers except KG in this study over a sentence in which the preposition is pied-piped as in (108). Preposition stranding also occurs in spontaneous speech, as illustrated by (109) and (110) and in declarative clauses (cf. (111)).

(107) Ki ye parle avèk? (TR)
Who 3PL talk VL with

Who did they talk about?

(108) Avèk ki ye parle? [MT] (TR)
With who 3PL talk

Who did they talk about?

(109) Kwa-se t ape hale pou? (LCDCLM, Mayeux 2018) [SD]
What 2SG ASP shout for

What are you shouting for?

(110) Sa t ole ROB en DOLLAR STORE pou?
What 2SG want rob DET dollar store for

What do you want to rob a dollar store for? (LCDCLA, Mayeux 2018) [SD]

(111) Sa-fe ye te pa gen pèrsonn pou parle avèk
DEM-make 3PL TNS NEG have somebody for speak with

So they didn't have anyone to speak with (LCDCGB, Mayeux 2018) [SD]

This is noteworthy because preposition stranding is neither attested in the historical data nor in Neumann (1985) or Klingler (2003a). Klingler (2003a, 333) explicitly observes "when interrogative pronouns or noun phrases consisting of an interrogative adjective and a noun are objects of prepositions, the preposition precedes them". His examples are cases of preposition pied-piping such as (112).

(112) Dan lekèl to kouri? Klingler (2003a, 333)
In which 2SG go

Which one did you go into?

Neumann-Holzschuh (2009) identifies the occurrence of stranded prepositions in Louisiana French as a contact induced phenomenon. Since French disallows preposition stranding, this cannot be an effect of decreolization but must be either internally motivated or attributed to the influence of English. Considering the high frequency of preposition stranding in American English, especially in casual speech, and the fact that all LC speakers are bilingual speakers of this variety, this is a clear structural effect of language contact with English. The new data from the present study indicates that preposition stranding is becoming the norm in LC since almost all consultants preferred stranding over pied-piping. In fact, pied-piping appears to be restricted for some of the interviewees.

Just four speakers actively used a pied-piped preposition in the MTAT. Three of these fully accepted a version of their sentence in which the preposition was stranded. Solely KG found (113) unacceptable. He is the only speaker who produced more than on clause with a pied-piped preposition although preposition stranding occurs in his sample as well.

(113) Ki-moun ye te parle pou? (MT) (KG)
Who 3PL TNS speak about

Who did they talk about?

All other interviewees provided translations with stranded prepositions. The manipulated versions with pied-piped prepositions were accepted by eight speakers, considered slightly marked by two and strongly marked by three speakers and two interviewees fully rejected the resulting clause. In embedded clauses, the pied-piped prepositions were more readily accepted, likely because they no longer appeared sentence-initially. One speaker still found the sentence (114) unacceptable, two gave it a medium acceptability rating and another three found it to be just slightly odd but acceptable. The results are similar for the second example, which was supposed to represent a non-finite embedded clause (*Who did you avoid talking to last night*?) but was rarely translated as such. This pied-piped structure like (115) received four lightly marked, one medium

marked and one strongly marked acceptability rating while the remainder fully accepted this sentence.

(114) Mo konne avèk ki li parle. (TR)
1SG know with who 3SG speak

I know who he/she talked to.

(115) Avèk ki to pa parle yeroswar?
With who 2SG NEG speak last night

Who didn't you talk to last night?

(intended: Who did you avoid talking to last night)

The question whether this allowance or disallowance of pied-piped prepositions is tied to any socio-demographic or sociolinguistic factors is discussed in Chapter 7.4.2 below.

7.3.2 Embedded questions

LC has embedded interrogatives such as indirect questions. These can ask for the subject (cf. (116)), object (cf. (117), (118)) or adjunct (cf. (119)) of the embedded clause respectively. Prepositions are stranded in embedded interrogatives as well, as is illustrated by (120). The wh-expression is fronted to the beginning of the embedded clause where it is pronounced. Note that most examples in the present corpus happen to contain the verb *konne* (=know). While this is the most frequently used verb in such constructions they are not principally limited to matrix clauses that use *konne* but can also appear with other verbs expressing thought or speech such as *mande* (=ask) or *di* (=say).

(116) Mo konne ki fe sa. (LM)
I know who do that.

I know who did this.

(117) To konne sa mo di? (LCDCLD, Mayeux 2018) [SD]
You know what I say

Do you know what I say?

(118) Mo konne ki li konye. (HW)
1SG know who 3SG hit

I know who(m) he hit.

(119) Mo pa konne kofe li gòn. (JB)
1SG NEG know why 3SG leave

I don't know why he left.

(120) Mo konne ki li parle avèk. (JN)
1SG know who 3SG talk. with

I know who she talked to.

Interviewees uniformly rejected clauses in which the interrogative pronoun was moved out of the embedded clauses into an initial position in the matrix clause, as in (121).

(121) *Ki mo konne fe sa. (AH)
Who 1SG know do this

I know who did this.

7.3.3 Complex wh-questions

When an interrogative expression that originates in an embedded clause is raised to the initial position of an interrogative matrix clause, this is a complex wh-question. A long-distance operation is ungrammatical for questions embedded in non-interrogative matrix clauses such as (121) above. On the other hand, such structures are often acceptable when the matrix clause is interrogative. Clauses of this type are attested for LC in the present data, although they are not part of any previous descriptions. It is possible to pose subject (cf. (122)), object (cf. (123)) and adjunct (cf. (124)) questions this way. Notably, *ki* is frequently inserted in an intermediate position situated at the beginning of the embedded clause (cf. (125)). The status of this element in terms of its grammatical and functional category, its optionality and distribution and any theoretical implications, is hitherto unresolved and is subject to analysis below.

(122) Ki vou krwa fe sa. (GL)
Who 2SG believe do this

Who do you (form.) think did this?

(123) Ki to di John lèm? (AF)
Who 2SG say John love

Who did you say John likes?

(124) Komen li di y aranje l traka? (RM)
How 3SG say 3PL solve DET trouble

How did he say they solved the problem?

(125) Ki to krwa <u>ki</u> fe sa. (CC)
Who 2SG believe KI do this

Who do you think did this?

7.3.3.1 Wh-scope-marking

In the category of long distance wh-questions, wh-scope-marking must also be considered. Recall that this strategy, which moves the wh-pronoun to an initial position in the embedded clause and inserts a potentially expletive interrogative expression in the beginning of the matrix clause, is attested for MC ((Adone and Vainikka 1999), cf. (101), (102) above). Such constructions obtained mixed results from LC speakers in the acceptability judgments conducted for this study. To ensure that the two wh-expressions would not be co-indexed, interviewees were given manipulations of their own translation of (126) as well as (127) that structurally resembled (128) and (129), which is of course

ungrammatical in English, with a distinct initial wh-expression, as exemplified in (130), (131) and (132), (133).

(126) Who do you think did this?

(127) Who did you say hit whom?

(128) What do you think who did this?

(129) What did you say who hit whom?

(130) Ki to krwa (ki) fe sa? Manipulated to:

(131) Sa to krwa ki fe sa?
WH 2SG believe who do this

Lit.: What do you think who did this?

(132) Ki to di konye ki? Manipulated to:

(133) Sa to di ki konye ki?
WH 2SG say who hit who?

Lit.: What did you say who hit who?

Two interviewees translated (126) as (131) immediately, which means that they actively produced a wh-scope-marking clause. The wh-scope-marking construction in (131) was fully acceptable to six interviewees but fully rejected by four speakers[39]. Three of these four corrected the sentence in (131) by changing the initial item to *ki*, creating a construction identical to or closely resembling (125) above and allowing co-indexation of the two wh-expressions. The fourth choose *kwa* as the initial expression, creating example (134) which then received a slightly less acceptable rating than the original translation.

(134) Kwa to krwa ki fe sa. (GT)

The same sentence received an intermediate acceptability judgment (3 on the scale) by one other speaker and a third commented that this was acceptable but an LRF way of speaking rather than LC. This is likely due to the presence of *kwa* which is perceived as an LRF lexical item. Nevertheless, it clearly is part of some speakers' lexicon as it was mentioned as a marginal possibility by three white speakers and one black speaker. Two consultants even used the form *kwa-se* as is discussed in Chapter 7.4.1 below.

For (127), there were no spontaneous translations as a scope-marking construction and only three speakers fully accepted the respective manipulation. Four found it slightly marked and three gave it a medial acceptability rating. Just two fully rejected this clause.

[39] Note that the investigation of island effects below provides additional support for the presence of a wh-scope marking construction in LC grammar.

7.3.3.2 Wh-islands

Island effects, which are a major consideration in the study of wh-questions, also occur in long-distance contexts. The study design considers wh-islands and the potential asymmetry between argument and adjunct interveners, strong adjunct islands and the that-t effect.

Wh-Island effects arise in complex sentences with more than one interrogative pronoun or phrase. Asking for the object as well as the adjunct in the same clause is a marginal operation for English, where this is best accomplished through an in-situ echo question with rising intonation such as (135). Each constituent can be asked for separately, as illustrated by (136) and (137). When both wh-elements are moved to higher positions, the sentence becomes ungrammatical. Moving the adjunct across the argument wh-phrase is considered the more severe violation of syntactic rules. This is illustrated in examples (38) - (41) from Chapter 3, repeated here as (136) - (139).

(135) Paul thinks Peter won over which girl how?

(136) How$_i$ does Paul think Peter won over Mary t_i.

(137) Paul wonders who$_j$ Peter won over t_j that way.

(138) *How$_i$ does Paul wonder which girl$_j$ Peter won over t_j t_i?

(139) ??Which girl$_j$ does Paul wonder how$_i$ Peter won over t_j t_i?

Two examples of this type were included in the study design. The first uses a long-distance subject question that contains an in-situ wh-adjunct (*how*) in the subordinate clause as the base line. In the MTAT, context was provided to make a long-distance question with two wh-expressions accessible. Interviewees were presented with the scenario in (140) and asked to translate the final long-distance question posed by speaker C. The translation then was manipulated into a version with the wh-adjunct in the initial position of the embedded clause, which creates an adjunct wh-island. A subject and an object question were formed this way along the lines of (141) and (142). To form an adjunct question the positions of the wh-elements were reversed to create a subject wh-island across which the adjunct wh-expression is extracted as in (143). Example translations and their manipulated forms are given in (144) to (147).

(140) A: How did Paul help Peter yesterday?
B: By finding him a place to stay.
C: Who did you say helped Peter how?

(141) ??Who did you say how helped Peter?

(142) ??Who did you say how Paul helped?

(143) *How did you say who helped Peter?

(144) Ki to di ede Peter komen? (GL)

(145) Ki to di komen ede Peter? [MT] (GL)

(146) Ki to di komen Paul ede? [MT] (GL)

(147) Komen to di ki ede Peter? [MT] (GL)

This task posed a considerable challenge for most interviewees, similar to the formation of multiple wh-questions discussed below. AG insisted that it was impossible to form such a question in LC and several other speakers expressed their hesitation about providing a structure like (144). They often amended that they would not use such a sentence although it was not incorrect. Consequently, a number of avoidance strategies emerged in the resulting translations. AF inserted *e* ('and') to create the coordinate structure in (148). HW chose a non-finite subordinate clause to express the embedded wh-adjunct *how* in the form of *to do what* (cf. (149)). Five interviewees offered an alternative structure containing a subject pronoun in the subordinate clause as in (150).

(148) Ki to di ede Pierre e komen? (AF)
Who 2SG say help Peter and how

Who did you say helped Peter and how?

(149) Ki to di ede Peter a fe kwa? (HW)
Who 2SG say help Peter to do WH?

Who did you say helped Peter to do what?

(150) Ki-se to di mwa komen ye ede Peter? (LM)
Who 2SG say 1SG how 3PL help Peter

??Who did you tell me how they helped Peter?

These gap filling constructions supposedly have the same meaning, although thethird person plural pronoun *ye* must be in conflict with the usage of the subject wh-phase *ki-se* if both have the same referent. This referent has to be unknown for the wh-question but is necessarily known when a personal pronoun is used. Possibly, this could be analyzed as a partial wh-question, where *ki-se* serves as a wh-scope marker and the content question is provided in the intermediate position. These sentences would then not represent wh-islands but rather instances of wh-scope marking, which appears to be a marginal but possible strategy for LC as discussed above.

Considering the difficulties inherent in obtaining translations for these sentences in the first place, constructions with more than one wh-expression that is not in-situ must be regarded as marked if not highly marked for LC. Consequently, manipulating these translations in various ways to receive a rating for the equivalents of (145) - (147) above was not a feasible strategy in every interview. Of the 18 interviewees, seven were not given a manipulated translation or did not provide an acceptability rating. The remaining 11 speakers uniformly rejected the fronting of a subject wh-expression across the adjunct wh-

expression, as in (145). This clearly indicates that extracting a subject over an adjunct wh-island is ungrammatical in LC. The wh-adjunct creates an island out of which the argument wh-phrase cannot be extracted.

The extraction of an object wh-phrase across the intervening wh-adjunct has even lower response rates, which is mainly due to the decision not to pursue this line of questioning when interviewees became uncomfortable with it. Only five judgments were made. Remarkably, one (MR) of those found (151) fully acceptable and two (TR and GT) considered the manipulated translation in (152) possibly acceptable (a 3 on the scale). The other two rejected the manipulated translations as ungrammatical.

(151) Ki t a di komen Paul ede? [MT] (MR)
Who 2SG TNS say how Paul help

Who did you say how Paul will help?

(152) Ki to krwa komen Paul ede? [MT] (TR)

These results seem to tentatively indicate that extracting an object wh-expression across a wh-adjunct constitutes a less severe rule violation than extraction of the subject in the same environment. The number of obtained judgments is too low to make a more definite claim and the complete rejection by two interviewees argues against a regular grammatical pattern.

When it comes to the extraction of an adjunct wh-expression across an intervening wh-argument like in (143) and (147) above, three speakers (MR, JB, TR) fully accepted a manipulated translation similar to (153). Nine interviewees considered these manipulations unacceptable under a long-distance reading and six did not provide a rating. These data suggest that the construction is ungrammatical for most LC speakers although not more severely so than argument extraction. Notably, two of the interviewees who found the construction acceptable also accepted the extraction of the object across and wh-adjunct (cf. (151) and (152) above). Apparently, the grammar of a select few participants is less sensitive to island effects than LC grammar at large. The question if this effect may be correlated to any socio-demographic factors is addressed in chapter 7.4.2 below.

(153) Komen to kra ki ede Peter? [MT] (TR)
How 2SG believe who help Peter

How do you think who helped Peter?

In conclusion, LC shows clear sensitivity to wh-islands in finite clauses. Extraction of arguments across intervening wh-adjuncts is ungrammatical although three interviewees accepted the extraction of an object across an adjunct wh-phrase. An argument-adjunct asymmetry cannot be observed as extracting the adjunct wh-expression across and argument wh-expression was judged equally unacceptable.

The second example made use of a non-finite embedded clause with a fronted wh-adjunct across which an object wh-expression was extracted. After some context was provided in the explanation that a friend had patiently taught me to cook gumbo, the translation of (154) was required. Again, this task turned out to be challenging for the interviewees. The expected translation as in (155) occurs only three times out of the 14 interviews in which a translation was provided. The other ten interviewees chose a number of different strategies. Five interviewees included the answer to one of the questions in the translation as in (156).

(154) What did she explain how to cook?

(155) Sa Lisa eksplike komen kwi? (HW)
What Lisa explain how cook

(156) Ki Lisa eksplike komen fe en gòmbo? (GL)
WH Lisa explain how make a gumbo

What did Lisa explain how to make a gumbo?

The conflict in (156) is very similar to that in (150) above. The initial wh-expression asks for the identity of a referent that is given in the embedded clause, essentially this constitutes gap filling. This creates a twofold problem for the analysis under the assumption that this is a long-distance object question. On the one hand, the semantic conflict between information that is given (*gòmbo*) and the same information that is inquired after (*ki*) in the same sentence is even more obvious than in (150) above where the second constituent was a pronoun (*ye*) and hence more referentially ambiguous than the DP *en gòmbo*. On the other hand, syntactically, the wh-expression *ki* is assumed to be base-generated in the object position in the embedded clause and wh-moved to the initial position in the matrix clause from there. Clearly, this cannot account for (156) as the object position in the complement clause is occupied by the co-referential expression *en gòmbo*. Possibly, one might assume that the DP and the wh-expression represent different and distinct copies of the same chain both of which are pronounced. This is a highly problematic approach that would violate the basic assumptions behind copy theory in that the two elements of the chain are co-indexed but clearly different lexical items thus hardly constituting a chain. The approach further requires both 'copies' - which they are clearly not - to be pronounced, which is potentially possible but debated (cf. e.g. Nunes 2004). Even under these theoretically problematic assumptions the LF conflict would remain and should make the sentence uninterpretable and hence unacceptable. This suggests that a wh-scope-marking analysis is the more suitable candidate. Under this analysis, no co-referentiality remains between *ki* and *en gòmbo.* Instead the initial wh-expression is considered either an expletive wh-pronoun or a

phonetic realization of the [Q]-feature, hence an interrogative marker[40], and the second wh-expression *komen* that is moved to the initial position of the embedded clause is interpreted as the content question. The object DP *en gòmbo* is simply complement to the verb *fe* and bears no relation to the matrix clause *ki*.

Four participants fronted the wh-adjunct over the wh-object as in (157). At first glance, this pattern seems to suggest that fronting of an adjunct over an argument wh-phrase is more acceptable out of non-finite than finite embedded clauses. This can be the case only if a long-distance reading of *komen* applies. A last related strategy which speakers used to translate (154) was the repeated insertion of the same wh-expression, most commonly *komen* although KG used *sa/saki* in both positions, effectively asking *what* but not *how*. Instead of a long-distance question containing two distinct wh-expressions, these participants formed a sentence with several instances of the same wh-word as in (158). It is unclear if both wh-expressions are instances of the same adverbial question or if they refer to different information. Potentially, the initial *komen* could inquire after the manner of teaching (=patiently) and the second *komen* could refer to the manner of cooking, the recipe. This strategy occurs five[41] times with this example.

(157) Komen li eksplike twa kwa pu kwi? (JN)
How 3SG explain 2SG WH to cook

How did she explain what to cook?

(158) Komen li eksplike li komen kwi? (ML)
How 3SG explain 3SG how cook

How did she explain her how to cook?

To decide on an analysis of (157) and (158) above, it is essential if a long-distance reading is applied to the initial *komen* or if it refers to the manner of explaining. Under a long-distance interpretation, (157) shows a decreased level of sensitivity to wh-island effects in non-finite contexts and (158) can be considered a spelled-out movement chain which would lend support to the copy theory of movement. To clarify, interviewees were presented with a manipulation of their translation that structurally corresponds to (159) which several consultants had originally provided already (cf. (157)), or asked which answers were possible to (158) if they had produced such a sentence.

(159) How did she explain what to cook?

(is 'with rice and Okra' a possible reply?)

e.g. Komen li eksplike twa ki pu kwi? [MT](GT)

[40] Note that the former implies an analysis as two distinct questions while the latter suggests (LF) movement (cf. Lutz, Müller, and Von Stechow 2000). The analysis of such constructions is further discussed in Chapter 8 below.

[41] Ten interviewees used a strategy other than the expected translation. Two of those offered two alternative translations which used different strategies, and three used two strategies within the same sentence ("Komen Lisa a eksplike komen fe du Gòmbo?" RM), which is why the total number of alternative structure occurrences listed here exceeds ten.

Only three interviewees allowed a long-distance reading in which *komen* relates to the manner of cooking. Seven consultants refused a long-distance interpretation. Eight speakers either did not reply because they found the sentence confusing and had no spontaneous answer or were not given the question as this was added during fieldwork after the issue became apparent. Most interviewees preferred to answer the embedded questions (an equivalent of *what* in most cases) rather than the matrix question, hence replying "*komen kwi gòmbo*". This pattern is similarly attested by Roeper and de Villiers (2011) in children acquiring English as their L1. This, together with the high rate of non-replies, shows once again that having two distinct wh-expressions within one question is marginal for LC. The long-distance interpretation is not readily available in this example. It follows that on the basis of this data the extraction of a wh-argument from a non-finite embedded clause across an intervening wh-island is no more grammatical than it is from a finite embedded clause.

Since the data is sparse it should be complimented by further examples. Further investigation of the question whether LC allows long-distance readings despite a wh-intervener, where a short-distance interpretation is available, is in order. Recall that long-distance questions without an intervening wh-expression are fully acceptable as described above. Since the MTAT was expected to reach its limitations in this matter, the data on long-distance questions containing two wh-expression - one of which occupies a matrix clause CP position and the other a similar position in the embedded clause - is complimented by the 'Questions after Stories' task. As described in Chapter 5.1 this task is designed specifically to investigate if a long-distance interpretation is available. The interviewees were presented with two picture stories and asked a long-distance question with an intervening wh-phrase after each. The first question, here given as (160) extracts an adjunct from the embedded clause and the second, given as (161), extracts an argument across an adjunct. The corpus contains 14 recordings for the first and 12 recordings for the second story.

(160) Ekan li di son Papa komen li fe mal a limèm?
When 3SG say POS Papa how 3SG make pain to 3SG REF

When did he tell his father how he hurt himself?

(161) Ki-se li mande komen ede?
Who 3SG ask how help

Who did she ask how to help?

The results for the first story and the related question in (160) are almost uniform. All but one interviewee chose a short-distance reading apparent in their reply. They answered with the time of telling, namely "*se swa kan li benyen*" (=that night when he was

bathing) or something similar. One speaker (AF), who also speaks SF, chose a long-distance reading, answering with the time of hurting. For the second story and the question in (161), the trend is similar. Most replies reflect a short-distance interpretation of *ki*, answering that she had asked the mother. One person (JB) applied a long-distance interpretation and said that she was trying to help her neighbor. Two interviewees answered the embedded rather than the matrix question, saying that she intended to collect the mail and water the plants. Since this was already observed concerning (159) above, it confirms the marginal status of such construction as does the fact that two speakers found this sentence difficult to process. One decided on a short distance interpretation and the other did not give an answer. Despite these deviations, the overall result also indicates a clear preference for a short-distance reading although long-distance readings appear possible in principle. This is clear as well from the data on long-distance questions without a wh-intervener presented above.

In conclusion, the data for this study shows that LC has wh-island effects in finite and non-finite environments. In contrast to English, there is no observable argument-adjunct asymmetry. Due to the ungrammaticality of the construction speakers found alternative structures to express the intended content, such as conjunctions or the simplification of the sentence to contain just one question. There appears to be a general preference for such questions over clauses that ask two questions within one sentence. Although wh-scope-marking is not clearly attested as a regular strategy for LC, speakers often fell back on a partial interpretation when faced with questions containing two wh-expressions. This result is in line with the prediction made by Adone and Vainikka (1999) that a partial movement strategy constitutes the unmarked option in UG. Despite its on-the-fence status as a strategy for LC, it seems to exist as a possible fallback option for the interpretation of clauses that cannot be processed otherwise.

7.3.3.3 Adjunct islands

Island effects, similar to those observed when a wh-expression is extracted from an embedded clause across another wh-expression, arise when a wh-expression is moved outside of an adjunct subordinate clause which is introduced by a subordinating conjunction. This is illustrated in (162) and (163).

(162) John left the party because he was avoiding who?

(163) *Who John left the party because he was avoiding __?

Like with wh-islands, consultants were presented with the context in (164). They were then asked to translate the subject, object and adjunct questions given in (165) - (167) below.

(164) John is angry because Paul and Mary had an argument and Paul hurt Mary with his words.

(165) John is angry because who hurt Mary?

(166) John is angry because Paul hurt who?

(167) John is angry because Paul hurt Mary how?

From the resulting LC sentences, the subject, object and adjunct respectively were extracted out of the adjunct clause and moved to the initial position of the matrix clause. The results of the acceptability ratings for these clauses are summarized in Table 7-2. The results are not uniform for either type of extracted constituent although the present study finds the effect most obviously weakened for subjects in LC while ratings for the extraction of adjuncts and objects are more negative. Overall, adjunct islands appear to have a weaker effect in LC than in English but they are present at least for some speakers.

Table 7-2 Acceptability judgments for extraction out of adjunct islands by syntactic function

	1	2	3	5	n/a	Total
adjunct extraction out of adjunct clause	3	2	1	6	6	18
object extraction out of adjunct clause	2	1	2	6	7	18
subject extraction out of adjunct clause	6	2	1	3	6	18

The translation of the subject question in (165) usually resembled (168). The manipulation of this translation resulted in a sentence along the lines of (169). This was judged fully acceptable by six interviewees and fully unacceptable by three speakers.

(168) John fashe paskè ki se ki fe du mal a Mary? (ML)
John angry because who make pain at Mary

John is angry because who hurt Mary?

(169) Ki se ki John fashe paskè fe du mal a Mary? [MT] (ML)

Sentence (166) was commonly translated as (170). Fronting of the wh-object to the CP of the matrix clause as in (171) received low acceptability ratings. Two interviewees found the manipulated translation fully acceptable but six judged it fully unacceptable. Seven did not provide a rating, two judged the manipulated translation to maybe be acceptable and one found it slightly odd but not ungrammatical.

(170) John fashe parskè Paul fe mal a ki? (HB)
John angry because Paul make pain at who?

John is angry because Paul hurt who?

(171) (A) ki John fashe parskè Paul fe mal ___? [MT] (HB)

Lastly, the adjunct question was translated in the fashion of (172). The manipulation in (173) received rather negative judgments, similar to the subject fronting construction described above. Nevertheless, half of the interviewees considered the sentence at least

possibly if not fully acceptable. If the adjunct island was in full effect, that would not be the case.

(172) John fashe paskè Paul fe Mary mal komen? (JN)

(173) Komen John fashe parcek Paul fe mal a Mary? [MT] (JN)

7.3.3.4 The that-t effect

The that-t effect refers to the impossibility of extracting a subject wh-constituent from an embedded clause across an overt C-head, namely the subordinator *that* in English. This is a language-specific phenomenon attested in English where the presence of *that* in a sentence like (174) and (175) makes the clause ungrammatical despite the optionality of the subordinator *that* in many contexts. The effect only occurs when subjects are extracted but not with wh-movement of objects as in (176) which is known as the SUBJECT-OBJECT ASYMMETRY. Originally, this was framed as a consequence of the direct adjacency of the subject wh-trace (or copy in more minimalist terms) and the complementizer, which is why it is called the that-trace effect. Accounting for the phenomenon has proven challenging for generative syntax.

(174) Who_j did you say t_j kissed Mary?

(175) *Who_j did you say that t_j kissed Mary?

(176) Who_i did you say that John kissed t_i?

To test whether similar effects arise in LC, examples with fronted wh-subjects, wh-objects and wh-adjuncts were manipulated to contain overt complementizers in the embedded clause. For subjects, the category for which the effect is present in English, no such pattern becomes apparent. Interviewees accepted both (177) and (178) or variants thereof.

(177) Ki-moun to di lèm John? (AG)

(178) Ki-moun to di ke lèm John? [MT] (AG)

Of the 18 interviewees, eight spontaneously produced a sentence with an intermediate *ki* when asked to translate '*Who did you say likes John*?' although the English base sentence does not contain *that*. This indicates a preference for an overt complementizer despite its optionality only if one assumes *ki* to represent a complementizer in this instance. Only GL rejected a manipulation of the original translation due to the presence of an overt *ki* as unacceptable. HB found the insertion of *ke* slightly less acceptable than the same sentence without the intervening complementizer.

For the corresponding object-question, '*Who did you say John likes?*', HW was the only interviewee who inserted *ki* in the intermediate position (cf.(179)). When *ki* was substituted for *ke* the sentence became less acceptable for him although it remained

grammatical. GL and GT did not accept *ki* or *ke* in an intermediate position, requiring the complementizer to be covert. It follows that GL's rejection of (178) is not limited to the extraction of subjects across on overt complementizer but is extended to objects. Overall, the insertion of an intermediate complementizer appears to be acceptable with wh-moved objects but less common than with wh-moved subjects.

(179) Ki to di (ki) John lèm? (HW)

(180) Kimon vou di ke John lèm? [MT] (KG)

The movement of an adjunct wh-phrase is not at all hindered by the presence of an overt complementizer. 14 interviewees translated *How did he say that they resolved the problem*? in a manner identical or similar to (181). The speakers who did not include *ke* fully accepted a manipulated version of their translation which contained the complementizer although three of them clarified that their first choice would be the sentence without an overt complementizer. Strikingly, all but one speaker chose *ke* rather than *ki* in this context although several speakers remarked that the two are usually interchangeable.

(181) Komen li di <u>ke</u> ye aranje le traka? (JB)

The results retain a certain level of ambiguity due to the homonomy of *ki* which can mean *that* or *what/who.* As a consequence, it is unclear if interviewees interpreted the intermediate *ki* in (178) to be a declarative complementizer, *that*, or a relative pronoun or wh-expression, *who* or even just an interrogative marker similar to CVC above. If it was a wh-expression of some kind, it could also be interpreted as a spelled out intermediate copy of the wh-movement chain which would more likely be in the specifier of C^0 rather than the head position. It is impossible to determine which grammatical status the speakers assigned to *ki*.

On the one hand, the use of *ke* as *that* is rare according to Klingler (2003a). On the other hand, Klingler and Neumann-Holzschuh (2013) consider *ki* the exception and *ke* the regular complementizer. The predominance of actively produced *ke* in the adjunct questions supports this view of *ke* as the standard complementizer as does the fact that five speakers inserted *ke* into (182) despite the absence of a complementizer in the English sentence they were asked to translate. The sentence originally served to test if wh-in-situ is allowed in embedded contexts, which it is not, but showed this unexpected side effect.

(182) Ki se John kra <u>ke</u> Mary rencontre? (RM)
Who John believe that Mary meet
Who does John believe Mary met?

This would argue for a distinct treatment of *ki* in the subject and object questions such as (179) that make use of intermediate *ki* instead of *ke* exclusively. Either this is not a variant of the declarative complementizer at all or something about the extraction of a subject from the embedded clause triggers the use of the form *ki*. The well-known qui/que alternation of French (Kayne 1976, Rizzi 1990) comes to mind that appears to be dependent on the extraction of the subject from the embedded clause. Also, recall that insertion of *ki* in the intermediate CP layer commonly occurs with long-distance questions but *ke* never appeared in these contexts. The choice of *ki* vs. *ke* and the identity of *ki* in intermediate positions in long-distance questions are discussed in Chapter 8 below.

7.3.4 Wh-in-situ

Wh-in-situ questions are found in a number of languages as a valid question formation strategy. The wh-expressions in these languages occupy the same position as their non-interrogative subject, object or adjunct counterparts. Languages with wh-movement often allow such constructions only in an echoic reading, as in the English example (183) below where the question does not ask for truly new information but rather confirmation of what was previously said.

(183) A: I saw a pink alligator in the yard.
B: You saw <u>what</u> in the yard?

French, the lexifier of LC, in fact is particularly interesting in that it appears to have optional wh-movement in the sense that wh-expressions can remain in situ in a non-echoic fashion and can also be fronted as is exemplified by (184) and (185) both of which are possible and fully grammatical in French (Adli 2006).

(184) Où tu vas? (Adli 2006, 164)
Where 2SG go2SG

(185) Tu vas où?
2SG go2SG where

This phenomenon has triggered a plurality of studies and proposed analyses, aiming to uncover distribution patterns and constraints and provide explanatory accounts (Bošković 1998, Cheng and Rooryck 2000, e.g. Adli 2006, Mathieu 2009, Déprez, Syrett, and Kawahara 2013), some of which can be extended to LC such as Cheng and Roodryck's notion of the intonational Q-morpheme mentioned above. Considering that LC is a French-based Creole language, it is one objective of the present study to determine if the intriguing French pattern is replicated to any extent in LC. The previous literature on LC devotes little attention to the topic but generally assumes that LC allows in-situ questions only in echo readings, just like English and most wh-movement type languages.

Overall, the present data does not suggest that LC allows (optional) in-situ questions in a principled manner. The acceptability judgments do not indicate that constructions such as (186) - (188) below, are a generally valid strategy for LC question formation with the exception of the echoic interpretation which is accompanied by rising intonation (cf. Cheng and Rooryck 2000, Déprez, Syrett, and Kawahara 2013 for rising intonation in French wh-in situ).

(186) *Arive ki? [MT] (JB)
Arrive who

Who arrived?

(187) ?? John konye ki se? [MT] (LM)
John hit who

Who did John hit?

(188) ? John kouri eou? [MT] (TR)
John go where

Where did John go?

But, surprisingly, these examples are not rejected across the board. For root questions, acceptability ratings differ notably depending on the syntactic function of the wh-element. The indication of different degrees of markedness in these examples is meant to represent the acceptability ratings in Table 7-3. Leaving subject wh-phrases in situ is clearly unacceptable. For objects and adjuncts, the picture is less clear. Nine out of 14 interviewees who rated this particular sentence still found (187) or a variant thereof unacceptable. Note that the elicitation material contained two object questions and an in-situ variation of both, resulting in a higher total number of tokens. The second sentence was a variant of (189), depending on the exact translation provided by each participant. There is no observable difference between the two object clauses in terms of their acceptability rating, which is why the results are summarized under the object label in Table 7-3 below.

(189) ?? To war ki se? [MT] (RM)
2SG see who

Who did you see?

Table 7-3 Acceptability for wh-in-situ in root questions by syntactic function

	1	2	3	4	5	n/a	Total
Adjunct	5	2	4	2	3	2	18
Object	5	3	4	2	17	5	36
Subject		1		2	15		18
Total	**10**	**6**	**8**	**6**	**35**	**7**	**72**

Eight of the 36 instances of in-situ object questions were judged fully acceptable or just slightly odd (1 or 2 on the scale). 19 times such a construction was judged to be not at all or barely acceptable. For adjunct wh-expressions the acceptance rates are even higher, amounting to more positive (seven instances for 1 or 2 on the scale) than negative (five instances for 4 or 5 on the scale) evaluations. In other words, approximately 50% of the in-situ object questions and only approximately 30% of the adjunct-in-situ questions were deemed unacceptable across interviewees. This is a surprise, as one would expect a number closer to 100% from a non-in-situ language, which LC is considered to be. In part these results may be due to an echoic interpretation the interviewees assigned to the in-situ-questions, despite my best efforts not to pronounce them with the intonation and stress placement associated with echo questions. This certainly was the case for GT and AF who explicitly specified that *To war ki?* and *John konye ki?* were acceptable only with rising intonation under an echoic interpretation. But this explanation does not account for the distribution of acceptability across syntactic functions. There is no reason to assume that adjunct questions would be more readily interpreted in an echoic fashion than object questions and subject questions in turn. Interestingly, AF also offered an in-situ translation of *Who have you seen*? spontaneously (*To war ki*?) which cannot be attributed to an echo-reading as there was no presupposed context whatsoever.

The data on embedded interrogatives suggest that in-situ expressions are considerably less acceptable in these clauses than in root questions. One long-distance question containing wh-objects was manipulated into an in-situ variation such as (190) with the additional variation of adding an overt complementizer to introduce the embedded clause. This makes no observable difference yielding a total of 34 sentences for the 18 participants.

(190) John kra Mary rencontre ki? [MT] (MR)
John believe Mary meet who
Who does John believe Mary met?

13 times interviewees rejected the in-situ construction and twice it was considered highly marked (4 on the scale). An additional three speakers assigned the construction medium acceptability and only five in-situ sentences were found fully or mostly acceptable. Of those five, two clearly used an echoic interpretation. In 13 instances interviewees did not provide any judgment for the construction. This is an indication that these examples were perceived as bothersome during the interview which implies that they certainly do not represent a standard strategy.

While these data do not clearly mirror the French optionality of wh-in-situ, LC is possibly approaching the French pattern in this case. A subject-object asymmetry with

regard to in-situ wh-expressions is attested for French (Obenauer 1994) and they are prohibited in embedded contexts as well (Shlonsky 2012, 245). LC shows similar tendencies although these are not regularized considering that in-situ constructions regardless of their syntactic function are not acceptable to all speakers. If analyzed in such a way, this constitutes an instance of SF or LRF influence on LC, which some creolists would consider a consequence of decreolization. As argued in Chapter 4.3.4 this thesis takes the broader viewpoint of either contact-induced or internally motivated language change rather than a decreolization approach. This is best confirmed through consideration of demographic and sociolinguistic data. If acceptability of in-situ constructions is caused by contact to SF or LRF, it should correspond to variables such as competence in LRF or SF, and relatedly race and possibly place of residence as this may influence degrees of exposure to forms of French. This analysis is conducted in Chapter 7.4.2 below. If no connection to these factors can be established, an alternative account could be incomplete creolization. Under this view, LC is not developing towards the French pattern but never fully departed from it. Keeping in mind that LC is more French-like than other French-based Creoles and that it developed gradually with a large amount of internal variation throughout all stages of its existence, this is a valid explanation. It could be argued that in-situ constructions would have been documented earlier in this case but since they clearly do not constitute the default strategy, it is to be expected that they would only appear in a targeted study and are unlikely to occur in broader language documentation projects.

7.3.5 Multiple wh-questions

Multiple wh-questions are questions containing more than one wh-expression. Languages have different strategies to implement such questions syntactically. Some languages leave all wh-expressions in-situ (e.g. Japanese), some front just one wh-phrase or word (e.g. English) and some front several wh-expression (e.g. Bulgarian, Serbo-Croatian, Russian), which can lead to superiority effects. Such examples, where several fronted wh-expressions must appear in a fixed order have proven challenging for syntactic theory which is hard pressed to account for the cross-linguistic variation in a principled manner (Bošković 2002, Richards 1999).

For example, Russian allows a free order of initial wh-expressions as illustrated by (191), while Bulgarian requires a fixed order that corresponds to the ordering in base-positions as in (192) and Serbo-Croatian shows superiority effects only in certain contexts such as embeddedness and in the presence of on overt complementizer (cf. (193)). The position of the second wh-phrase after the overt C head *li* in (193) is especially challenging for an analysis that relies on movement to a specifier position in the left periphery.

(191) a. Kto kogo ljiubit? (Bošković 2002, 354)

b. Kogo kto ljiubit?

'Who loves whom?'

(192) a. Koj kogo običa? (Bošković 2002, 354)
Who whom loves

b. *Kogo koj običa?

(193) a. Ko li koga voli? (Bošković 2002, 353)
Who C whom loves?

b. * Koga li ko voli?

In the light of such data and the related debate about appropriate ways to account for this variation in analyzing wh-movement, a secondary objective of the present study is to investigate if LC data could contribute to the discussion.

There is little to be gained for the debate from LC data, as multiple fronting clearly is not a possibility in LC grammar. Instead LC corresponds to the English pattern of fronting just one wh-expression. The MTAT was designed to account for all possibilities in this regard. The same multiple wh-question (an equivalent of *Who saw what when*?) was supposed to be judged with the fronted wh-subject, wh-object and wh-adjunct respectively as well as with multiple fronted wh-expressions in various orders. All these possible configurations were included first as root questions and repeated in a long distance question (an equivalent of *Who did you say saw what when*?). Notably, this task was not completed in a single interview.[42] This shows that multiple wh-questions are marginal in LC at best. Consultants sometimes struggled with questions containing two wh-expressions, which are part of the study design mostly in the context of wh-islands, and struggled especially with three wh-expressions within the same sentence. Many speakers found the sentence difficult to translate and chose an avoidance strategy instead of the word-for-word translation in (194).

(194) Ki (se) war sa/ki ekan?
Who see what when

Most often, a connector such as *e* or *epi,* both of which mean *and,* was inserted between the words for *what* and *when* to create a conjunction structure. Some speakers chose to represent the meaning in two sentences, such as AF who provided (195). This example also contains a declarative expression corresponding to *something* instead of *what.* This occurs twice in the corpus and effectively reduces the number of wh-expressions in the

[42] Note that these very complex sentences were expected to potentially pose difficulties beforehand. Consequently, there was no point in confronting interviewees with each ordering once it was established that multiple fronting was not an option.

sentence. CC and LJ chose to insert another verb to make the sentence more accessible as illustrated in (196).

(195) Ki war kelk- shoj? Ekan li war sa? (AF)
Who see something? When 3SG see it/what?

Who saw something? When did he/she see it/what?

(196) Ki ki war sa k'arrive (ekan)? (CC/LJ)
Who see what happen when

Who saw what happen (and when)?

On the basis of such translations, meaningful manipulations and judgments of their acceptability are barely achievable and of questionable usefulness and reliability. The few manipulation judgments obtained through this task seem to suggest that fronting the subject wh-expression is the default process and fronting the object wh-expression is more marked while fronting the adjunct wh-expression in the presence of a subject and an object wh-expression is fully ungrammatical. Multiple frontings are also rejected across the board.

7.3.6 Polar questions

Lastly, polar questions, although they are no wh-questions, are included here for the sake of completeness, as they do of course constitute a question formation strategy for LC. As mentioned above, polar questions in LC, like many other Creoles (cf. Haspelmath and Consortium 2013a), are commonly formed with declarative sentence structure and rising intonation. Alternatively, the marker *èskè* can be used according to Klingler (2003a).

(197) Èskè to monnde to tant pou la gòm-la? (Klingler 2003a, 329)
ESKE 2SG ask. 2SG aunt for DET gum-DET

Did you ask your aunt for the chewing gum?

Usage of *èskè* is rare in Klingler's data and not attested at all in the present corpus. The LCDC contains only one instance of *kèskè*. Possibly, the marker is being lost as a consequence of language endangerment which tends to reduce redundancies such as marking of the interrogative sentence force through intonation and a sentence initial free interrogative particle, which was always optional. The most likely hypothesis is that the free marker was an acrolectal strategy, since it is not attested in the historical data, that is being lost in the process of "stylistic shrinkage" (Campbell and Muntzel 1989, 195) which is typical of endangered languages as described in Chapter 2.5. Under such an analysis the attestations in Klingler's corpus represent sporadic occurrences of such features in the speech of some select creolophones. Mayeux (2019) observes similar tendencies for other morphosyntactic features in LC. Nevertheless, it suggests that an interrogative marker is in principle a strategy that has been used by LC speakers, which can be one consideration in determining the role of *ki* below.

In sum, the most common strategy is what Alexandre (2012) terms the gap strategy. That the wh-moved constituents leave behind a gap is most obvious with stranded prepositions, which clearly have the fronted wh-word as their complement. Preposition stranding was preferred over preposition pied-piping in the present data. LC also allows for embedded interrogatives in which the wh-expression is fronted within the embedded clause but not the matrix clause. Long-distance wh-movement is possible and shows sensitivity to wh-island effects without the argument-adjunct asymmetry attested for English. Adjunct island effects are observable but seem to differ with respect to syntactic functions and between speakers. There is no apparent that-t-effect in the data. For wh-scope-marking acceptability judgments are mixed but the strategy seems to be employed to interpret clauses that would otherwise be uninterpretable. Multiple wh-questions are very marginal and fully ungrammatical when more than one wh-expression is fronted. Wh-in-situ is attested in an echoic fashion and for some speakers as an acceptable question formation strategy.

7.4 Variation and the relation between sociolinguistics and syntactic structure

This chapter addresses irregularities and distinctive features that occur in the data and investigates whether these can be related to linguistic or socio-demographic variables. In Chapter 7.4.1, I begin by investigating the distribution of wh-pronouns relating to grammatical functions and humanity of the referent as these are the distinctive features identified by Neumann (1985) and Klingler (2003a). Further, the current data is tested against Rottet's (2006) hypothesis concerning the grammatical function of a final element *ki*. The findings are related to LC's level of vitality. The chapter also considers the distribution of intermediate *ki* in long-distance questions. Chapter 7.4.2 investigates the effects socio-demographic variables as well as speaker idiosyncrasies have on the choice of wh-pronoun, the acceptability of marginal structures such as wh-scope-marking and wh-in-situ constructions, sensitivity to island effects, rates of preposition pied-piping, and the appearance of clauses containing an intermediate *ki*.

7.4.1 The form and distribution of wh-pronouns and the status of *ki*

One main issue that arises in the discussion above when considering wh-questions in LC is the form and distribution of interrogative pronouns, as the pronominal system is not fully described and the distribution appears to be linked to grammatical factors such as syntactic functions according to Rottet (2006). Another related notion is the identity of *ki* within interrogative expressions.

Concerning the form and distribution of interrogative pronouns in LC, 379 questions from the present corpus as well as 57 from the LCDC are analyzed. The present corpus consists of 18 MTATs each of which contains 28 sentences for translation, 35 manipulations of such translations and 12 distractors. To determine the use of interrogative pronouns, all manipulated translations are excluded since a native speaker did not directly produce these. On the other hand, distractors are included when they happened to contain a wh-pronoun. Wh-adverbs, such as *komen* ('how') or *kofe* ('why') were excluded as well, as they never contain final *ki*, leaving only interrogative pronouns in a subject or object function. Because of the multiple and long-distance questions in the corpus, several sentences include two wh-pronouns which is why the number of tokens is higher than the total number of questions considered here. A total of 433 wh-pronouns from 379 questions out of the corpus of this study is considered and complimented by the 57 questions from the LCDC. From the LCDC data, only sentences that contain a wh-pronoun were selected and wh-adverbs were excluded as well.

This comparison yields the following set of wh-pronouns, *Ki, ki-k, ki-ki, ki-ke, ki-se, ki-se-ki, ki-sa, ki-s, kwa, kwa-se, kwa-ki, sa, sa-k, sa-ki, ke, ke-se, kimoun, ki-moun-ki*. In addition, the following occur but are likely not be subject/object wh-pronouns despite their use in the associated position, *li, se, kekun, kofa*.

Recall the form and distribution of wh-pronouns in LC as previously documented. The relevant table is repeated here as Table 7-4 to be compared with the data from the present study summarized in Table 7-5. Clearly, the range of interrogative subject and object pronouns used by LC speakers is wider than previously documented. At the same time, a number of items listed in the literature do not occur in the present corpus. These are bold printed in Table 7-4. The bold items in Table 7-5 indicate an item that appears in a previously undocumented function and items in bold italics are previously undocumented in LC in any syntactic function.

Table 7-4 Interrogative PRNs in LC

	TLC (Neumann 1985)		**MLC (Klingler 2003a)**	
	Subject	Object	Subject	Object
[+human]	sa-ki, ki-ki, **ki-se-sa-ki,** ki-se-ki	ki, ki-se, **ki-se-sa**, ki-ki/ke, sa, sa-ki/**ke**	ki-moun, sa-ki, ki (rare)	ki-moun
[-human]	sa-ki, ki-ki, **ki-se-sa-ki,** ki-se-ki	sa ki, ki-se, **ki-se-sa**	sa-ki, ki, **lekèl,** **nekel**	ki (rare), sa, **lekèl,** **nekel**

Table 7-5 Interrogative Pronouns in the study's corpus

	MTAT Corpus		LCDC	
	Subject	Object	Subject	Object
[+human]	**Ki**, Ki-k/ki/**ke**, **Ki-sa-k**, **Ki-se**, Ki-se-ki, ***Ki-moun-ki***, ***Kwa,*** ***Kwa-ki***, **Sa**, Sa-k/ki, ***Se-ki***	Ki, Sa, Ki-se, Ki-k/ki, **Ki-s/sa**, ***Kwa,*** ***Kwa-se,*** ***Kwa-ki,*** Sa-ki, ***Ke,*** ***Ke-se,*** Ki-moun, ***Ki-mo(u)n-ki***, ***Ke-mo(u)n***	Ki, Ki-k, Sa-k	_____
[-human]	_____	Ki, Sa, ***Ke-se,*** Ki-se, ***Kwa,*** ***Kwa-se,*** **Sa-k/ki**	**Ki**, ***Kwa***, Sa	Ki, Sa, **Sa-ki** Kèskè

The more complex forms *ki-se-sa* and *ki-se-sa-ki* that appear in Neumann (1985) are not attested in the present corpus. Neither are the forms *lekèl* and *nekel* from Klingler (2003a). One reason could be that these forms have fallen out of use and are consequently being lost in the process of language shift taking place in Louisiana. Another cause could be that the study design and data collection left gaps in the data for example with regard to one particular functional or demographic category, so that some forms just were not recorded despite continued use in the community. For the two forms attest in Klingler (2003a) but not the present corpus, this is a likely explanation, as the MLC regional variety which he documents is underrepresented in the present data. The first reason, on the other hand, appears to be the most plausible explanation for the absence of *ki-se-sa* and *ki-se-sa-ki.* These pronouns literally incorporate a rather complex meaning (*who/what-is-it-who/what*) resembling the semantic structure of a cleft construction but not making use of the syntactic construction itself. But, considering their interchangeable use with other pronominal forms, these items do not seem to transport a semantic distinction such as focusing. Neumann

(1985, 334) translates her example simply as *Qui appelle*? (=*Who called*?) rather than as a sentences along the lines of *Who was it that called*?.

(198) Ki-se-sa-k'ape pele? (Neumann 1985, 334)
Who ASP call

Who called?

The hyphenated spelling also indicates that she considers these items lexicalized. Under this assumption, using the longer forms of the interrogative pronouns offers no communicative benefits in terms of a semantic distinction over the use of synonymous shorter forms. This makes them typical candidates for the reduction process endangered languages tend to undergo. Naturally, these forms may still occur despite their absence from the present corpus. A targeted study would be needed to be fully certain they are no longer used. Nevertheless, the present corpus contains a solid set of subject and object interrogative pronouns with 180 tokens in the subject and 253 in the object category. Among the objects, 181 have [+human] referents and 72 referents are non-human. The corpus contains a gap in the field of [-human] subjects as these simply were not present in the example sentences used in the MTAT. This is due to the focus on syntactic structure rather than lexical form in the design of the MTAT. Since the 180 [+human] wh-pronouns in subject function do not contain a single one of these forms and their use was never restricted to the [-human] subject category, it seems safe to assume that the results would have been similar had such sentences been included.

A closer look at forms that are present in the current corpus rather than those that are absent reveals the introduction of new material and new functions as well as clear differences in frequency. The most striking change occurs in regard to *ki* as a wh-pronoun. Contrary to Klingler's observation that *ki* rarely appears by itself without additional material, *ki* is the most frequent wh-expression in the data collected for this study. It occurs as a single free morpheme 207[43] times. Contrary to Neumann's observation that *ki* only occurs in object function, 86 of these are used as a wh-subject and 121 are wh-objects. All subjects and 101 of the wh-objects have [+human] referents. This distribution still indicates an overall preference for the use of *ki* in object functions although one must consider effects of the study design. The corpus contains more objects than subjects and several questions include a subject and an object wh-pronoun. In those, the subject expression usually is fronted and the object expression remains in situ. It may be that, rather than a preference for expressing objects through *ki*, this distribution indicates a preference to express

[43] This number includes four instances where speakers offered *ki* or an alternative as well as five instances of *me ki,* where *me* is taken to be a sentence initial discourse particle rather than part of the wh-expression. Generally, items are included and counted separately whenever an interviewee offers two or more alternative expressions in the translation.

subjects or simply fronted wh-pronouns through more complex forms. In the LCDC *ki* appears 18 times, eight of which are subjects and ten are objects.

Note that *ki* is not the most frequent wh-pronoun in the LCDC sample, although it is the second most frequent one following *sa*. *Sa* occurs 30 times, which amounts to more than 50% of the wh-pronouns in the LCDC. All but four of those appearances are in object functions as in (199) and all have non-human referents. By comparison, *sa* is rare in the MTAT data where it is the third most frequent item with 33 occurrences. A strong similarity to the LCDC data is observable in the functional distribution, as *sa* is mainly used in object functions (30 occurrences), especially for [-human] referents (24 occurrences).

(199) Sa t ole? (LCDCMY, Mayeux 2018)
What 2SG want

What do you want?

In general, the LCDC data contains fewer types and tokens of wh-pronouns. A lower token frequency is to be expected, as this corpus contains spontaneous and often narrative data, which naturally contains few questions. With 21 speakers whose utterances are included, the potential for a similar amount of type-variation as found with the MTAT corpus is principally given but not borne out. Hypothetically, interviewees chose the most common expressions for the given purposes in the LCDC's natural conversation setting. In the MTAT, wh-expressions were required quite often, which may have prompted a higher level of diversity especially when more than one expression was required in the same sentence creating the need to differentiate between them. Note that the use of *ki* across grammatical functions and [+/-human] referents is nevertheless confirmed in the LCDC data as is the use of *sa-ki* for [-human] objects and the occurrence of *kwa* though not the related forms.

The second most frequent item in the MTAT corpus is *ki-se*, with a total of 73 occurrences. 45 of these are objects, most of which have [+human] referents (39 items) and 28 are subjects, showing a similar distribution to *ki* above. As mentioned, *sa* is in the third position concerning frequency. This is notable because *sa* appears only in Neumann (1985) where it is restricted to [-human] objects. While this still is the prevalent use of *sa* in the MTAT corpus, it is also attested in subject function in three instances and at least three times with a [+human] referent in object function. Sentence initial *sa* additionally occurs in clauses with another wh-expression in a lower position like (200). While these could be interpreted as being co-referential, which would make them instances of *sa* having a human referent, they are instead analyzed as wh-scope-marking clauses below.

(200) Sa John kra ki Mary MET? (HB)
WH John believe who Mary meet

Who does John believe Mary met?

The distribution of *sa* shows that the original use in [-human] object functions has been slightly extended to [+human] and subject contexts in a few instances. While it is still mainly used in its original function, it is crucially no longer unacceptable in other contexts. On a larger scale, *ki* is used in subject function as well and predominantly but not exclusively appears with [+human] referents. This development, too, is not unusual for an endangered language as speakers become insecure about grammatical restrictions and begin to diffuse and overgeneralize patterns. A pattern that seems to be preferred by many LC speakers when it comes to wh-pronouns is the choice of a monosyllabic form, with the exception of *ki-se* described above. This is attested in the frequency with which *ki* and *sa* appear in their base form and constitutes a lexical and possibly grammatical simplification process. The less frequent wh-expressions range from 18 occurrences for *sa-ki* to just one for example for *ke* and *ke-se* respectively. This is not to say that the distribution of the less frequent items is cannot be grammatically conditioned. *Ki-se-ki* for example only occurs in a subject function and *ki-s/sa* just as objects as do *ke* and *ke-se*.

The latter forms are not attested as wh-pronouns in earlier work. The complementizer *ke* appears in the C-head position of embedded clauses though. Either a homophonous wh-pronoun developed for some reason, the item is being borrowed from LRF or the occurrence of *ke* and *ke-se* is a result of linguistic insecurity. Due to the phonological similarity between *ki* and *ke*, and the fact that *ki* can be used as a wh-pronoun, a relative pronoun and a complementizer some speakers may be unsure of the distinction and begin to use the two items interchangeably. This development is further facilitated by the fact that *ki* and *ke* are both already attested in Neumann (1985) as the final element in a wh-pronoun. *Que* is attested as an interrogative pronoun in older forms of French and rarely appears in modern SF. Consequently, it could also be a form that is transferred from LRF which tends to use older words. The fact that *ke* and *ke-se* occur very rarely in the data with one instance of each item, which are importantly not produced by the same speaker, indicates that this is a mostly idiosyncratic phenomenon at the moment. But, if LC continues on its current trajectory, I hypothesize that a repetition of this study at a later date might produce a higher frequency of *ke* and *ke-se* (also potentially *ke-sa* and *ke-ki*) due to either continued contact to LRF or persistent free variation due to a breakdown of the linguistic system. Alternatively, these forms may vanish completely if endangerment results in a reduced pronominal system which relies solely on *sa* and *ki* for example. Note that the form *ke-mon* is also distinguished from the more standard MLC form *ki-moun* in Table 7-6 below.

While this constitutes phonological variation at this point, it can be argued to reflect a related process.

One type of item that newly appears in the present corpus is *kwa* with the associated forms *kwa-ki* and *kwa-se*. As mentioned above *kwa* is generally perceived to be an LRF lexical item and it does not appear in earlier documentation of LC. On this basis, it is all the more noticeable that *kwa* occurs in the MTAT as well as the LCDC corpus. Most likely, this is a contact induced borrowing that does not express any content or concept previously unavailable in LC. As such this represents an instance of the introduction of new material (Campbell and Muntzel 1989), the use of which falls into the category of extensive non-meaningful variation attested for endangered languages by Rottet (2001, 44). In the simple form *kwa* it is preferred to express objects especially of the [-human] variety but one subject and two [+human] objects also occur. With the addition of *ki* the use is restricted to [+human] referents, namely five subjects and three objects, while the form *kwa se* only appears with [-human] referents. This seems to indicate that the addition of *se* and *ki* respectively are responsible for the functional distribution rather than the core item itself and raises the question whether these additions are part of the lexical item or function as independent markers.

Kwa-se occurs in the present corpus in the context of (201). These occurred as translations of the English sentence "What do you think John hopes that Mary will tell him?". Clearly, this meaning is not fully captured by the LC translations containing *kwa se*. The reappearance of *se* in the embedded clause seems to suggest a conditional construction along the lines of example (54) in Chapter 4 above, repeated here as (202).

(201) Kwa se m ape swate Mary se di mon? (JN, CC)
WH (COND) 1SG PROG hope Mary IRR/COND say 1SG
What I am hoping (that) Mary would tell me?

(202) Si mo <u>se</u> konnen, mo <u>se</u> mannde la [sic.] plen […]
If 1SG COND know, 1SG COND ask 3SG? a lot
If I had known, I would have asked her lots of things […]
(Klingler 2003a, 329)

Since (201) hardly constitutes a typical conditional clause, the alternative analysis is to consider the first appearance of *se* a part of the interrogative expression *kwa-se* which is formed parallel to *ki-se*. The second *se* in the embedded clause would best be analyzed as an irrealis marker then and bear no direct relation to the matrix appearance. Considering the other wh-expressions in the present data, nothing points to a free marker *se* although it may well have started out as a focus marker for example.

Table 7-6 shows the distribution of wh-pronouns in the MATA corpus. *Se* occurs word-finally 78 times in the MTAT in the expressions *ki-se* (73 x), *kwa-se* (4 x) and *ke-se* (1 x). These appear as subjects (28 x) and objects (50 x) and 67 of these have [+human] referents which equals 86%. This last pattern is unexpected from the distribution of *kwa-se* that seems to be limited to [-human] referents. It follows from the data that no clear prediction about the syntactic function or the [+/-human] value can be made from the presence of *-se*. Consequently, *se* is no free marker of either objecthood or non-human referents at present and hence regarded as part of the wh-expression. This is not to say that it did not begin as a grammatical marker of some kind which has lexicalized.

Table 7-6 Frequency of wh-pronouns in the MTAT corpus

	Object			Subject	Total
Wh-Pronoun	**-human**	**+human**	**O total**	**+human**	
Ki	20	101	121	86	207
Ki-se	6	39	45	28	73
Sa	24	6	30	3	33
Ki-moun		12	12	6	18
Sa-ki	1	3	4	14	18
Ki-se-ki				11	11
Ki-ki		1	1	9	10
Kwa	7	2	9	1	10
Ki-k		1	1	7	8
Kwa-ki		3	3	5	8
Sa	4		4	3	7
Ke-mon		5	5		5
Ki-moun-ki		1	1	3	4
Kwa-se	4		4		4
Se	3		3		3
Ki-s		2	2		2
Ki-sa		2	2		2
Ke		1	1		1
Ke-se	1		1		1
Kekun		1	1		1
Ki-ke				2	2
Ki-sa-k				1	1
Kofa	1		1		1
Li		1	1		1
Sa-va-kwa	1		1		1
Se-ki				1	1
Total	**72**	**181**	**253**	**180**	**433**

The status of *ki* in the final position of the wh-pronoun is the more interesting case considering Rottet's (2006) analysis of *ki* in LC as a marker of focus and subjecthood and the general discussion around interrogative/focus markers in Creole languages. *Ki* occurs as the final element of a wh-expression such as *ki-ki, ki-se-ki, se-ki*, *ki-moun-ki* or *kwa ki* in 69 tokens in the MTAT corpus. This number includes clipped forms such as *ki-sa-k*, although these may strictly also allude to a final *-ke* form, but excludes the 207 occurrences of *ki* by itself. As the corpus contains 180 wh-subjects only 55 of which contain a final element *ki*, this is clearly not obligatory as a marker on wh-subjects in general. Neither is its use restricted to subject constituents as *ki* appears as the final element on 14 wh-object-pronouns. Of these nine have [+human] and five have [-human] referents. The LCDC contains *ki*-final forms as well, although infrequently with a total of four tokens. *Sa-ki* occurs three times, two of which are shortened to *sa-k* and there is one instance of *ki-k*. Three of these refer to [+human] subjects but the last one is in object function with a [-human] referent.

Table 7-7 Distribution of *ki* as the final element in wh-pronouns

	Object			Subject	Total
	-human	+human	O total	+human	
Final *ki* MTAT	5	9	14	55	69
Final *ki* LCDC	1		1	3	4
Total	**6**	**9**	**15**	**58**	**73**

These data do not abide by the regular grammatical distinction observed in Rottet (2006) although the same general trend to use *ki*-final wh-expressions for subjects is confirmed. His data stem from sources that are at least 20 years older than the present sample. That the grammatical distinction is not fully followed in the comparison with the MTAT data indicates that it is being lost. The regular grammatical marking is weakened in favor of increasingly free variation. If LC follows common patterns of reduction in endangered languages (Aikhenvald 2012), it is to be expected that the use of final *ki* in wh-expressions will become even less conditioned by grammatical functions and at the same time decrease in frequency until it ultimately is lost. From the current data it appears most likely that simple *ki* and *sa* will remain as the most commonly used interrogative pronouns although borrowings like *kwa* or English question words and other LC lexical items like *ke* may well be integrated into the system as an effect of language contact and language change.

Finally, translations occur in the MTAT that contain two wh-pronouns where the English sentence had just one such as (200) above or that contain three wh-expressions instead of two as in (205). These clauses are referred to as *KI*-INSERTION CLAUSES here. These are complex wh-questions that have a wh-expression in the sentence-initial position and potentially contain further wh-expressions in-situ in the embedded clause but additionally use a wh-pronoun, usually *ki*, in an intermediate position introducing the embedded clause. These sentences come in three types. In the first type the initial and the intermediate wh-pronoun are identical (cf. (203)). In the second type the initial wh-pronoun contains *ki* (eg. *ki-se*) and the intermediate expression is *ki* (cf. (204)). In the third type the initial and the intermediate wh-expression are phonologically unrelated (e.g. *Sa* and *ki*) as in (205).

(203) Ki to kra ki fe sa? (CC/LJ)
WH 2SG believe WH do this

Who do you think did this?

(204) Ki-se to kra ki laime John? (CC/LJ)
WH 2SG believe WH like John

Who do you think likes John?

(205) Sa to di ki ede Peter komen? (CC/LJ)
WH 2SG say WH help Peter how?

Who did you say helped Peter how?

The nature of those constructions and the question if they are to be treated uniformly or not is addressed in Chapter 8 below. Potentially, they can be regarded as instances of pronounced lower copies, most plausibly for the first type or questions containing an overt complementizer, most plausibly for the second type or instances of wh-scope-marking, most plausibly for the third type. I argue in favor of the first analysis below and a distinct treatment of wh-scope-marking. To provide some grounds for the discussion, the distribution of the different types and pronouns is presented here.

There are 46 examples of such constructions in the MTAT corpus and an additional ten examples with the intermediate expression *ke* rather than *ki*. I take *ke* to be the complementizer *that* and not a wh-expression, which is why these clauses are excluded in Table 7-8 below. Clearly, *ki* is the most frequent wh-expression to be inserted in the intermediate position. There is just one instance of *sa* and *sa-ki* respectively.

Table 7-8 Distribution of initial and intermediate wh-pronouns in *ki*-insertion clauses

	Intermediate wh-PRN			
Initial wh-PRN	**sa**	**ki**	**Sa-ki**	**Total**
Sa	1	4	1	**6**
ki		19		**19**
Ki-ke		1		**1**
Ki-se		13		**13**
Ki-se-ki		1		**1**
Ki-moun		2		**2**
Kwa-ki		1		**1**
Se		3		**3**
Total	**1**	**44**	**1**	**46**

The initial pronouns show much more variation, which is unsurprising considering the diverse possibilities offered by the pronominal system. The most frequent item in this position is *ki* as well. These 19 examples represent sentences of the type of (203) above. A total of 18 other expressions in the initial position contain *ki* either initially or finally and thus are instances of the second type illustrated by (204) above. Nine wh-pronouns do not contain *ki*, namely the six instances of *sa* and the three instances of *se.* These are the most likely to represent a wh-scope-marking construction as described in Chapter 7.3.3.1 above or a focus construction which is what *se* can be canonically used for. With 37 sentences, the majority of the *ki*-insertion clauses are subject questions. Those nine sentences that are object questions either use *ki* (2 x), *sa* (4 x) or *se* (3 x) initially. This distribution shows a clear preference for the insertion of intermediate *ki* in subject questions, especially when the initial wh-expression equals or contains *ki*. With the exception of the two object questions that use *ki* in the initial position, which can be analyzed as instances of the complementizer *ki*, all object *ki*-insertion clauses are of Type 3 and likely represent wh-scope-marking as mentioned. Consequently, the present data show a clear subject-object asymmetry. *Ki*-insertion clauses apparently are a valid strategy only for subject not for object extraction. Since all subjects in the MTAT have [+human] referents, a distinction according to the [+/-human] feature can only be possible for the nine object questions. Seven of these have non-human referents but two refer to a person. There is no clear conclusion to draw from the data in this respect, as both types of referents are attested.

In sum, the data depict a system of wh-pronouns that is considerably more diverse than previously documented. The monosyllabic wh-pronouns *ki* and *sa* are the most frequent ones in the MTAT and LCDC, contrary to earlier documentation. In part, the variety of wh-pronouns is due to a process of expansion through borrowing from LRF and the

addition of new functions for established LC items.[44] This creates much variation that is free with regard to the linguistic variables considered here. Previously observed distributions of wh-pronouns that reflect grammatical distinctions, such as subject or object status, are no longer fully in effect. While the present data still shows similar tendencies, these rules are frequently violated as a consequence of language endangerment. A previously undocumented construction is termed a '*ki*-insertion clause', because the intermediate expression in those sentences is almost uniformly *ki* while there is some variation concerning the initial wh-pronoun. For this construction a subject-object asymmetry is detected. The following section investigates if the distribution of wh-pronouns, *ki*-insertion and marginal constructions can be related to socio-demographic variables.

7.4.2 Variation and socio-demographic variables

As documented in the above sections a large amount of variation persists in the interrogative system of LC, not only in the types of forms and structures which are spontaneously produced but also in the ones that are accepted or not. Examples are the marginal constructions that are accepted and even produced by some speakers and/or in some circumstances but rejected by others, such as wh-in-situ sentences or a wh-scope-marking interpretation of long-distance questions. Such inconsistencies are typical of an endangered language and likely the result of grammatical rules breaking down due to linguistic insecurity as speakers use their language less and less and become unsure of grammatical properties (Rottet 2001). Most often this results in the loss of marginal, rarely used and/or complex constructions that are not shared with the dominant language, a process also known as negative borrowing (Aikhenvald 2012). When the linguistic feature cannot be simply omitted, it is replaced by either a corresponding structure from a more dominant language, which can be seen as the introduction of new material on a structural level, or the reduction of alternatives through the generalization of another language-internal pattern in the forms of over-generalizations, simplification and the leveling of paradigms (cf. Chapter 2). Before the original structure or item is fully lost, a period of much variation occurs during which speakers are not fully committed to either system and apply the different strategies more or less unsystematically (cf. e.g. Schmidt 1985's observations of YD). This is often perceived as a breakdown of the linguistic rules and system although it can in some cases be the beginning of a new, reduced system that is more easily maintained by the younger generations against the pressure of the dominant language. By the estimation of the present thesis, LC currently is in this state of variability in respect to its

[44] Since this is the first study that gathered data focused on wh-constructions, some of the observed wh-pronouns may have been present earlier and simply have gone undocumented.

interrogative system. The meaningless variation that is observed in this phase of language development is said not to be conditioned by social variables as claimed for example by Rottet (2001). In the present chapter, this assumption is tested against the MTAT[45] data from LC to determine whether the variation can be related to socio-demographic variables or is completely free. The factors considered here are age, race, sex, region, LC proficiency level, L1 and knowledge of LRF or SF. Like in Chapter 6, only factors that influence the distribution are discussed.

Relating to the previous chapter, I begin with the distribution of wh-pronouns in the MTAT corpus according to socio-demographic criteria. One main finding in Chapter 7.4.1 above is the preference for *ki* and *sa.* The use of these forms occurs across socio-demographic variables, though a preference for *sa* is visible among black speakers (24 tokens vs. 11 from white speakers). The most influential factor concerning the choice of wh-pronoun is the level of proficiency in LC. Interviewees from the least proficient Group C used almost exclusively *ki* with the exception of one instance of *ki-k* and *ki-ki* each. In Group B almost half of the tokens (39 out of 80) are *ki* and another 11 are *sa*. *Ki* remains the most frequent item in the most proficient group of interviewees but with a growing level of self-assessed proficiency, more diverse forms are used. Most likely, the more proficient speakers feel confident to vary the forms they utter, while the most insecure speakers fall back on the most basic forms that are unlikely to be considered wrong by their peers.

Considering the distribution of *kwa* and the related forms *kwa-ki* and *kwa-se*, which clearly constitute a borrowing from either LRF or SF, the results are surprising. The *kwa* forms do not predominantly occur in the speech of LC and LRF or SF bilinguals as expected. Instead, they appear most frequently with black speakers who claim to have no knowledge of either French variety. The eight occurrences of *kwa-ki* were produced by a single speaker, AR, a black man in the youngest age group from the Bayou Teche area. He only contributes one instance of simple *kwa*, the remainder are provided by other interviewees. These speakers are black as well with the exception of four tokens of *kwa* which were produced by three different white interviewees all of whom speak either LRF or SF.

This distribution does not argue against a borrowing analysis. Rather, it provides some insight into the kind of borrowing occurring here. The bilingual white speakers who use *kwa* may do so consciously or subconsciously in order to lighten the cognitive load of the different languages in their repertoire which is one of the main motivations for borrowings according to Matras (2009). They appear to do so rarely, as *kwa* is attested just

[45] This analysis does not refer to the LCDC since most of the structures under consideration here and some of the socio-demographic features are not attested for or in this data.

once for three of them and twice for the fourth, indicating that they tried to distinguish between their languages at least for the context of the MTAT. But the majority of tokens were produced by black speakers who do not speak LRF or SF. In those cases, the usage of *kwa* is much more likely to be a prestige loan, the inclusion of an item the speakers perceive to be a standard expression in higher status French-based varieties around them, that is integrated into the LC system to the extent that *kwa-ki* or *kwa-se* are created.

The use of the complementizer *ke* or the derived form *ke-se* as an interrogative pronoun on the other hand, is attested only for two white men of the youngest age group one of whom knows SF the other speaks LRF. It is impossible to fully tease apart if the instances of *ke* in wh-pronoun functions are due to transfer from LRF/SF or due to language-internal functional expansion or a combination of both. The fact that the two interviewees who used such forms are speakers of a French variety points to a language contact explanation although the transfer of the French use of *que* in wh-expressions could be argued to facilitate the functional expansion of *ke* in LC for those speakers as well.

In accordance with the pattern described above for the distribution of *ka* and *sa*, final *ki* in wh-expressions is barely attested with the least proficient group of interviewees. As illustrated in Table 7-9 below, the majority of *ki*-final wh-pronouns are attested for the most proficient group, with 51 of 69 items. Keeping in mind that this is the largest group with 13 of the 18 interviewees makes the difference to speakers in Group B less obvious, as the three speakers in this group used 16 final wh-expressions. Proportionally, they make more use of final *ki* in object functions which make up 31,25% of the *ki*-final items in this group as opposed to 17,6% in Group A. Recall that this is the non-canonical option according to Rottet (2006). This result indicates that the less fluent speakers who are still proficient enough to actively and confidently use the language retain different grammatical properties in this respect, and arguably in other aspects as well, and thus promote language change. As discussed in Chapter 6 above, a certain correlation exists between age and LC proficiency. The oldest interviewees are the most proficient ones on average but the competence levels of the middle and the youngest age group do not differ much. However, the youngest age group makes more use of *ki*-final wh-expressions in object functions (47,06%) than the middle age group (33,33%) while the rate is lowest in the oldest group (12,5%). This shows that the level of language proficiency and age interact in this feature. The least proficient speakers avoid the feature altogether. The more proficient and the older speakers are, the more likely they are to abide by the marking of the grammatical distinction through final *ki*. Hence, they are less likely to use a *ki*-final expression in object functions. The younger proficient speakers are more likely to do so, indicating that the

grammatical distinction is being lost. The fragmentation of the speech community and the low prestige of LC give way to such developments easily, as younger speakers are not likely to be confronted with older fluent speakers willing to correct them or even insistent about a 'correct' form of LC.

Table 7-9 Functional distribution of final *ki* by LC proficiency

Final ki	A	B	C	Total
O	9	5		14
S	42	11	2	55
Total	**51**	**16**	**2**	**69**

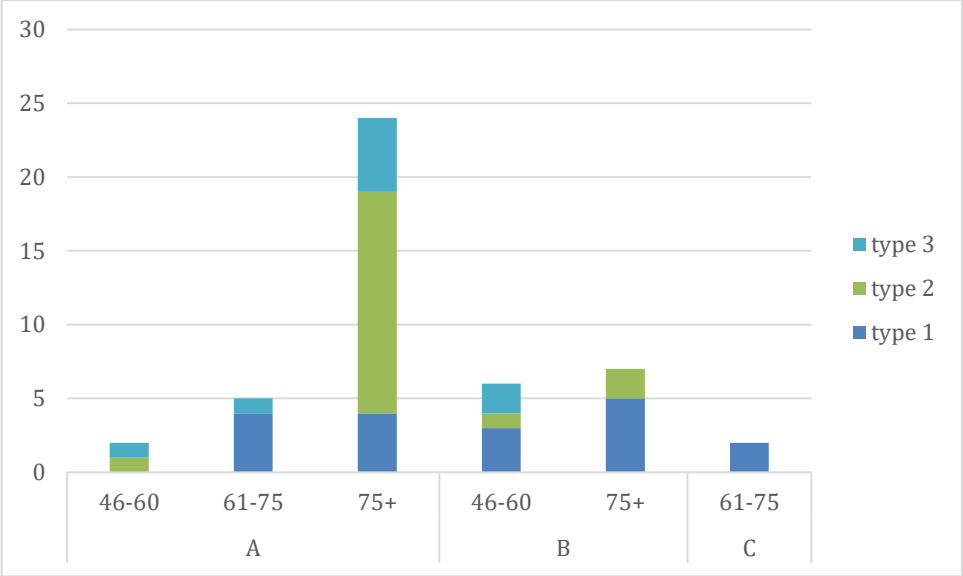
Table 7-10 Functional distribution of final *ki* by age group

Final ki	46-60	61-75	75+	Total
O	8	2	4	14
S	17	6	32	55
Total	**25**	**8**	**36**	**69**

Concerning the insertion of *ki* in an intermediate position in long-distance questions, socio-demographic factors seem to be of limited relevance. Recall that three distinct types are present in the data, which differ with respect to the initial wh-pronoun. Type 1 clauses use *ki* in the initial and the intermediate positions, Type 2 uses a complex wh-expression containing *ki* in the initial position and *ki* in the lower position and Type 3 employs a wh-expression that does not contain *ki* in the initial position, such as *sa*. Men tend to use these sentences more often than women in the MTAT corpus but age and the level of proficiency in LC are the most influential factors here as well. Type 2 *ki*-insertion clauses are most frequently attested for the oldest age group, which is unsurprising as this group shows the highest rate of complex wh-expressions, which are used in the initial position in this construction. This age group generally has the most appearances of intermediate *ki*, as does the most proficient group of interviewees, indicating that this is not a feature that occurs primarily with young, insecure speakers. The two characteristics are not necessarily related in the present sample anyway, but for the sake of the argument one might assume that the use of *ki* in an intermediate position is a result of linguistic insecurity, a repetition used by unpracticed speakers to help carry the meaning over clause boundaries. The distribution of *ki*-insertion clauses does not support such assumptions. The construction may well fulfill the function to maintain the link between the question word and the embedded clause but it is not a coping strategy of low-proficiency speakers. Instead it is most common among the

older and more proficient speakers, especially from the Bayou Teche region. Note that the difference between proficiency Group B and the least proficient Group C is again much more pronounced than the one between Group A and B. Again, the least proficient speakers avoid the construction altogether. This distribution is illustrated in Figure 7-1.

Figure 7-1 Distribution of *ki*-insertion clauses by LC-proficiency level and age group



In sum, the use of mono-syllabic and complex wh-expressions, the distribution of *ki*-final wh-pronouns and the appearance of *ki*-insertion clauses are influenced to some extent by the proficiency level and the age of the speakers as the most relevant socio-demographic and sociolinguistic factors. These are the variables one would expect to matter in an endangered language. For the use of newly coined or borrowed expressions, race and knowledge of contact languages seem to be more relevant although the use of contact items is clearly not limited to speakers with knowledge of the contact language. Ultimately, the observed variation seems to be relatively but not completely free. Essentially, speakers of LC make use of the full range of possibilities afforded to them by their individual linguistic biographies and repertoires creating a pronominal system which offers an extended range of alternative strategies, some of which are preferred by speakers from similar socio-demographic backgrounds. The conventions that define LC grammar are eroded in this process, as items are borrowed, distinctions are lost, and rules are weakened, although this does not occur in a uniform way for all speakers but is largely idiosyncratic.

The same effect of variation as a byproduct of the loss of grammatical rules is observable on the level of syntactic structures used to form interrogatives. Here it is even clearer that the variation in what is and what is not possible structurally is individually conditioned rather than dependent on larger socio-demographic categories.

For example, some of the interviewees show less sensitivity to island effects than others as described in Chapter 7.3.3 above. Particularly, MR allows both the extraction of a wh-argument across a wh-adjunct and vice versa (cf. examples (151) and (153) above). The former is judged possibly acceptable by TR and GT. TR fully accepts the latter as well, as does JB. Considering that all other interviewees firmly reject either of those structures, the question arises if these four speakers have anything in common that may relate to their tendency to allow extraction out of wh-islands or if these are purely individual traits of their respective I-grammars. No clear commonalities between these speakers arise from the sociolinguistic and demographic data collected for the study. Two of them are black, two are white, three are female, one is male. They are in different proficiency and age groups ranging from A to C and from the youngest to the oldest group. One spent many years in other states, the remainder never left South Louisiana. They received different levels of education although none of them attended college. Two work in the home and two work in blue collar professions. The only resemblances in their lives are that they come from the Bayou Teche area, that they did not go on to tertiary education and that they did not work in a white-collar profession. The most likely relevant factor is a regional one. All four participants live in Henderson or Cecilia, which are neighboring communities and although only two of them are friends they may all well know each other. Possibly, a decreased sensitivity to island effects is developing in this area, although this cannot be predicted from the present data, as the other six interviewees from the same area did not show the same effect. At this point, the phenomenon remains an individual one that is not visibly conditioned by sociolinguistic or demographic factors.

Concerning adjunct islands, the data presented above indicates a continuum of extractability based on to the functional category of the wh-moved constituent. Wh-subjects were allowed to be moved over an adjunct island more often than objects which were in turn very slightly more often accepted than wh-moved adjuncts (cf. Table 7-2 above). Linking this information to the individual speakers and the associated demographic and sociolinguistic information reveals some idiosyncratic patterns but no relevant socio-demographic variable. For example, TR, who is rather tolerant of extraction across wh-islands as shown above, also accepts all three types of extraction out of an adjunct island. LM fully accepts extraction of subjects and objects while JN and ML rate all three types with a 1, 2 or 3, remaining within the range of acceptable constructions but assigning different values to the different types. On the other end of the spectrum, HB, GT and HW, the only three speakers to reject the extraction of subjects from an adjunct clause, have clear island effects in place and allow neither extraction type. The overall pattern, that subject

extraction is more acceptable than the other two types, is based on the sum of given evaluations. This exact pattern is visible on an individual level just with one speaker (AH), who fully accepts the subject extraction clause and fully rejects the object and adjunct clauses. This clear-cut distribution does not occur with any other interviewees. The analysis of the acceptability judgments under consideration of sociolinguistic and demographic factors indicates no obvious pattern. These results show that despite the general tendency to be more tolerant of subject extraction from adjunct clauses, this does not constitute an absolute rule in either LC grammar at large or most individuals' I-grammars. Instead, very few speakers agree on the same acceptability ratings of all three circumstances, specifically agreement exists only among the ones that fully accept or fully reject all three. This illustrates the interpersonal variation found in the LC community quite clearly. This is not a classic case of free variation, where speakers use alternative means of expression freely and without any link to socio-demographic variables that are typically found in studies of linguistic variation. The present phenomenon is very similar because it shows free (i.e. not socially determined) variation in the acceptance of structural patterns. This is a case of variation in an endangered language that occurs on a deeply structural level.

Wh-in-situ questions were shown not be a standard strategy in LC above but still acceptable to some speakers with a preference for in-situ objects over in-situ subjects. This resembles the French pattern which is why competence in LRF or SF has to be considered as a sociolinguistic factor. Counter expectations this factor is not obviously influential as is illustrated by Table 7-11. Speakers with and without knowledge of LRF or SF reject in-situ wh-subjects. The acceptance rate for objects is higher overall but the hypothesis that knowledge of SF or LRF would correspond to the acceptance of leaving wh-objects in-situ cannot be confirmed without a doubt. The rejection rate for in-situ objects is highest among interviewees who do not speak LRF or SF and the one speaker who accepts in-situ subjects to some degree also falls into this category but half of the LRF speakers still find an object-in-situ construction unacceptable. SF speakers also are more or less evenly spread across the spectrum of acceptance. This data does not fully negate a contact explanation as speakers of LRF or SF are more likely to allow wh-objects in-situ than speakers who do not speak either variety of French but it does not fully confirm the hypothesis either. The alternative explanation that LC always had this underlying possibility which was just never documented because it is not the default strategy, appears equally plausible. Most likely, the pattern is inherently present in LC but knowledge of a French variety can increase acceptance due to structural similarities.

Table 7-11 Acceptability ratings of wh-in-situ questions by functional type and LRF/SF knowledge

	Acceptabilty Ratings						
	1	**2**	**3**	**4**	**5**	**n/a**	**Total**
LRF							
Adjunct	2	1	2	1	1		5
Object	1	1	2	1	6	3	10
Subject					7		5
LRF & SF							
Adjunct	1						1
Object	2						2
Subject				1			1
SF							
Adjunct		1		1			2
Object	1		1	1	1		4
Subject					2		2
No LRF/SF							
Adjunct	2		2		2	2	8
Object	1	2	1		10	2	16
Subject		1		1	6		8
Total	**10**	**6**	**8**	**6**	**35**	**7**	**72**

A clearer case of contact induced change is the preference of preposition stranding over preposition pied-piping. Recall that the data obtained for this thesis clearly illustrates this trend. This is a relatively recent and contact induced development that has now stabilized to the point that a few speakers actually disallow preposition-pied-piping which used to be the standard strategy. Klingler and Neumann-Holzschuh (2013) observe that pied-piping is a more acrolectal strategy. In terms of demographic and sociolinguistic factors, no clear pattern emerges to confirm this association. The six interviewees (RM, JN, CC, AH, AG and HB) who consider preposition pied-piping marked or even ungrammatical are from all demographic groups. Crucially, they are from all age and proficiency groups, indicating that this is not a process applied by younger, insecure speakers to compensate for a lack of knowledge of the correct construction. The two possibilities co-exist and are used in free variation with preposition stranding as the clearly preferred option. This shows how influential the English language is on a structural level, which is to be expected with the sociolinguistic findings from this study in mind.

Wh-scope-marking is discussed above in two contexts. On the one hand, it is considered a possible strategy in long-distance wh-questions that allows speakers to in fact avoid a long-distance operation by inserting an interrogative dummy expression in the matrix clause and interpreting the content wh-question locally in the embedded clause. At the same time, it is considered a sub-type of *ki*-insertion clauses above. The question arises

which analysis is most suitable for these constructions and if they are to be treated uniformly with other cases of *ki*-insertion or if possibly all *ki*-insertion types are wh-scope-marking constructions. This question is discussed in Chapter 8 below. The present section is concerned solely with the most classic example of wh-scope-marking sentences, repeated here as (206), (207) and their acceptability to speakers in relation to sociolinguistic and socio-demographic variables.

(206) Sa to kra ki fe sa? (TR)

(207) Sa to di ki konye ki? (CC) (MT)

Recall that wh-scope-marking constructions are previously undocumented for LC but were shown to exist in the language, as two speakers actively produced such a sentence. One of those speakers is TR who exhibited very little island effects above and who also used *ki*-final wh-expression with objects twice. Clearly, he has one of the most liberal grammatical systems among the consultants who participated in this study. It is notable that he is among the youngest speakers in this study too, although the patterns in his speech are not generally characteristic for his age group. TR is thus best characterized as a young fluent speaker whose grammatical system is different from older attested versions. The other speaker who actively produced a wh-scope-marking construction is AH. She did not abide by the subject/object distinction for *ki*-final wh-expression either but she is more conservative with regard to island effects.

For optimal comparability, all manipulated translations that are wh-scope-marking constructions should contain the same initial and intermediate pronouns. Since the sentences in which the clearest wh-scope-marking constructions occurred were originally not designed to test for this phenomenon this is not done uniformly in all the data. For the present analysis I only include sentences that are formed with *sa* and *ki* as in (206) above which results in 19 rated manipulated translations. The most obvious pattern in this data is the relative clear cut between speakers who do and speakers who do not accept this construction. Sentences like (206) above were rated fully acceptable in nine and fully unacceptable in five of those 19 instances. Four times the construction was judged to be slightly odd and once possibly acceptable, all of which refer to a variant of (207). In addition, three other speakers received a clause with an initial *kwa* instead of *sa*, which was rated as possibly acceptable.

The socio-demographic variables that appear relevant in this matter are age and race. Black speakers accept the structure more often than white speakers who in turn find it fully unacceptable more often. Furthermore, the younger speakers were more acceptant of it than the older ones. This is illustrated by Table 7-12 and Table 7-13.

Table 7-12 Acceptability of wh-scope-marking by age groups

	Acceptability Rating				
Age	**1**	**2**	**3**	**5**	**Total**
46-60	4	1		2	7
61-75	2	1			3
75+	3	2	1	3	9
Total	**9**	**4**	**1**	**5**	**19**

Table 7-13 Acceptability of wh-scope-marking by race

	Acceptability Rating				
Race	**1**	**2**	**3**	**5**	**Total**
b	6	1	1	1	9
w	3	3		4	10
Total	**9**	**4**	**1**	**5**	**19**

In conclusion, the distribution of the form of wh-pronouns is conditioned only in a very limited way by sociolinguistic variables, mainly the proficiency level in LC and knowledge of contact languages. The distribution of marginal syntactic structures is completely free. With the exception of wh-scope-marking questions, which seem to be more acceptable to younger and black speakers, these are cases of idiosyncratic variation that cannot be related to socio-demographic factors. This behavior at large is consistent with observations of endangered languages.

8 A GENERATIVE ANALYSIS OF WH-QUESTIONS IN LC

This chapter provides an analysis of some of the data presented above in the minimalist framework. The analysis of wh-questions in LC presented here aims to account for root questions using the gap strategy, embedded questions and long-distance questions. It does not focus on island effects or wh-in-situ constructions specifically because these represent marginal strategies represented in individuals' grammars rather than LC grammar at large as discussed above. There also is no reason to assume that their derivation would differ from standard accounts of the same structures in other languages.

Long-distance questions were not explicitly documented before this study but can be expected from a wh-movement system. The special case of *ki*-insertion clauses on the other hand is not a standard strategy, which makes it particularly interesting. One central question is if the three distinct types listed above are to be treated under a unified account. In the spirit of Felser (2004), Dayal (1994) and Bruening (2006), it is proposed that Type 1 and Type 2 can be accounted for under a wh-copying analysis but Type 3 is best considered distinctly as an instance of wh-scope-marking. The observable subject-object asymmetry in *ki*-insertion clauses sets LC apart from other wh-copying languages such as German[46] as does the pronunciation of the more complex wh-item in the matrix clause. Both issues are accounted for under the proposed analysis.

8.1 Root wh-questions

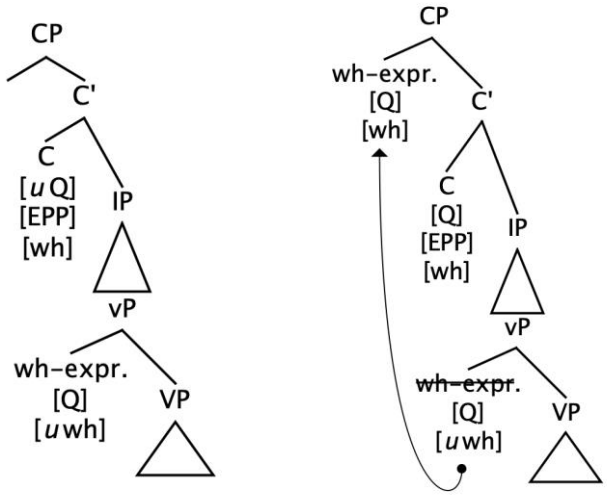
Root clauses as well as embedded wh-questions in LC can be represented in terms of the standard analysis of wh-questions in wh-movement languages such as English. The LC data presents no additional challenge to the analysis beyond those the theory faces in general. Following Chomsky (2000), wh-expressions have an uninterpretable, or rather unvalued, [wh]-feature as well as an interpretable [Q]-feature. The unvalued [wh]-feature renders the wh-phrase or expression active and hence available for internal merge. An interrogative C^0 on the other hand has an unvalued [Q]-feature and in the case of wh-questions an [EPP] feature. This [EPP] feature requires the CP-specifer to be filled, hence triggering movement of the nearest constituent that is active and can check the [Q]-feature on C^0. In some recent work the [EPP] feature has been replaced by the edge feature [EF] in the context of phases[47] but in principle both concepts are equally applicable here. Below, I adopt a proposal by

[46] An informal and unrepresentative survey among my German peers indicates more reservations towards object than towards subject wh-copying constructions as well. As a German native speaker, I am inclined to share that judgment which is not addressed in the literature. However, it is widely attested that acceptability of wh-copying in German is inconsistent across speakers with no relation to social, demographic or regional variables.

[47] As stated above, this thesis does not work within the phase theoretical framework due to the many inherent challenges the theory has yet to overcome (Klein 2012 and references therein). It is recognized that recent generative work usually operates within this framework which is why phasal counterparts of the proposed analysis are continuously referenced.

Bošković (2007) that dispenses with either trigger and relies solely on the presence of unvalued features to motivate movement. This idea is fully applicable to the derivation of root questions as well. The wh-expression is moved to the specifier position of CP either due to the presence of an unvalued feature or to satisfy the [EPP] and values its [wh]-feature under [Q]-agreement. This results in a configuration where the wh-expression occupies the specifier of CP, has a valued [wh]-feature and in turn values the [Q]-feature on C^0, yielding a wh-question. This is exemplified in (208).

(208) Wh-movement, standard account



LC root questions can be analyzed following this pattern. In contrast to other wh-movement languages, LC does not exhibit verb movement to C^0 (cf. German) or auxiliary inversion (cf. English) in question formation. Recall that LC, like many Creoles, does not employ any question specific word order but simply fronts the wh-expression to form a wh-question and marks yes/no questions only via intonation. In terms of the featural configuration, this means that no [TNS]-feature must be assumed to be present on C^0 to motivate verbal fronting. For yes/no questions, the analysis of French by Cheng and Rooryck (2000) can well account for the LC strategy. It is assumed that the [Q]-feature on the interrogative C-head is checked by an intonational [Q]-morpheme which sets the value to [y/n] resulting in a polar question.

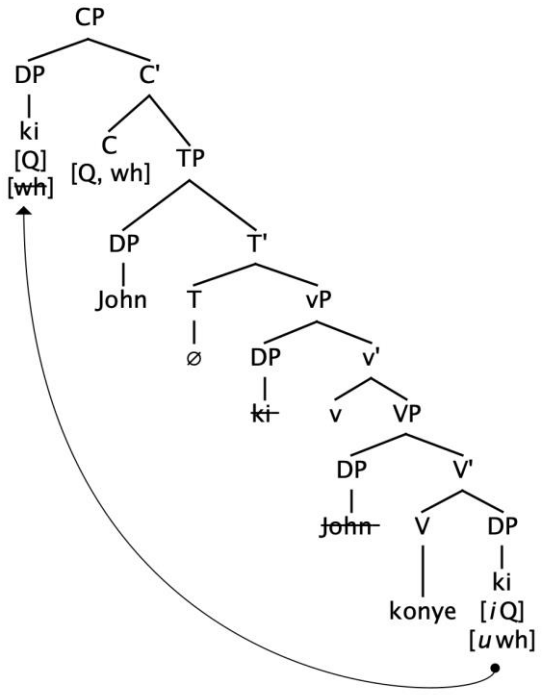
This derivation of root questions largely corresponds to the analysis Alexandre (2012) proposes for CVC. By extending the notion of feature disambiguation she applies to overt interrogative C heads to zero complementizers as occurring in LC main clauses, the analysis captures what is often glossed over as Q-agreement. In the analysis depicted in (208) above, the wh-feature on the wh-phrase is checked through an agreement relation

with the C-head that bears a [Q]-feature. How exactly this feature values the [wh]-feature on the moved constituent, given that this is considered distinct from the interpretable Q-feature on the same constituent, is not usually explicitly addressed[48]. Supposedly, [Q] binds a bound variable in the wh-phrase by ill-understood mechanisms. Under the disambiguation approach, all elements start out with the standard features but the Q-feature on C can and must not only be checked but further specified e.g. as [+wh] or [+y/n]. Formally, [wh] and [Q] remain distinct features, as questions can be [-wh] and wh-clauses can be non-interrogative (e.g. relative clauses) but setting the value of [Q] to [+wh] provides a formal valuation environment for the [*u*wh]-feature on the wh-phrase. Note that the [Q]-feature is not involved in triggering the movement of the wh-expression. It could be hypothesized that it is valued as [Q,wh] under long-distance agreement in cases of wh-in-situ constructions and as [Q,y/n] in polar questions where the intonational [Q]-morpheme provides the [y/n] value. The former could also be well-accounted for under a copy analysis with a pronounced root copy. The conditions for this pronunciation would have to be defined which is not the purpose of this work as the focus her is not on in-situ constructions.

This analysis accounts for LC root questions such as (103) to (113) above. An example derivation of the object question in (105) is sketched in (209) below. The wh-expression *ki* originates as the argument of the verb *konye*. From there it moves to the specifier of CP which attracts the nearest wh-expression via [EPP] to disambiguate [Q]. This checking relation values the [wh]-feature on *ki*, resulting in a grammatical wh-question. Note that *konye* is a class I verb according to Neumann (1985, 189) which means that it has a long and a short form. The long form represents the base form that is also attested of class II verbs. Since the distinction marks tense/aspect, it has been proposed by Rottet (1992) that the short forms of class I verbs actually undergo verb movement to Spec,IP as known from inflectional languages. This conclusion is generally supported here but bears no relevance to the present example. Also note that the deleted copy of *ki* in the specifier of vP is a necessary assumption under a phase-based approach to syntax. If such a view is taken, movement would proceed cyclically from the object position to the vP specifier to escape spell-out at phase level and move to CP from there. Since this is not the theoretical approach I depart from, it suffices to indicate the possibility. In further derivations this alternative is implied and occasionally mentioned but not represented in the tree structure.

[48] For a detailed discussion and a focus-based account see Miyagawa (2010).

(209) Ki John konye?[49]



8.2 Long-distance questions and successive cyclicity

The analysis of long-distance questions is conducted along the same lines with the additional layer of a second CP. The wh-expression is generally assumed to move through CP1 before raising to CP2 -the matrix CP- where it is pronounced. This is a long-standing assumption in generative grammar that is based on several pieces of empirical evidence. McCloskey (2001) argues for successive cyclic wh-movement through Spec,CP with data from Irish. The Irish complementizer system distinguishes several complementizers whose usage depends on the presence or absence of A-bar relations reaching into the embedded clause. The complementizer *aL*, McCloskey (2001) claims, must introduce any clause out of which wh-extraction has occurred. His example, here repeated as (210), illustrates that both clauses are introduced by the element associated with wh-movement. This suggests movement through both Spec,CPs.

(210) Rud a gheall tú a dhénanfa (McCloskey 2001, 68)
thing aL promised you aL do [COND-S2]

'something that you promised you would do'

[49] In addition, several syntactic processes are omitted here and in any structure given in this thesis for the sake of readability. For example, movement of the subject *John* indicated by a deleted copy on the VP level should in fact also proceed through vP, at least in a phase-based account that requires all constituents to relocate to the edge of vP before Spell-Out or to be pronounced in their base position. Hereafter, I include the vP only where it is necessary to the argument in order to keep the size of tree structures to a minimum. In the spirit of cross-linguistically universal analysis, it is nevertheless assumed to exist in LC.

McCloskey (2001) further argues that this also provides evidence for the feature driven nature of wh-movement, which is assumed in minimalist accounts. His argument goes as follows: If *aL* is a complementizer, there must be some property that distinguishes it from the other complementizers. This property would then be the presence of a wh-feature on C^0, as the only difference to the particle *go* is that *go* introduces clauses in which no wh-movement has applied. Hence, the feature that differentiates these two complementizers is also the feature that drives wh-movement (in combination with the [EPP] in McCloskey's (2001) analysis). The phenomenon has become known as WH-AGREEMENT and was documented for Chamorro (Chung 1994) as well as Passamaquoddy (Bruening 2004, 2006). In those languages agreement does not occur with an overt complementizer. Instead wh-movement results in a morphological marking on both the embedded and the matrix verb. The data in Muysken and Law (2001) from HC can also be considered wh-agreement, as HC marks the extraction of a wh-subject out of an embedded clause by insertion of a specific complementizer *ki*. This is why in (211) the wh-pronoun can only be interpreted as the subject of the embedded clause and in (212) it must be the object of the matrix verb.

(211) Kimoun Jan di ki ale nan mache?
who Jan say that go LOC market

`Who did Jan say is going to the market?' (=subj)

NOT: `Who did Jan tell to go to the market?' (=obj)

(212) Kimoun Jan di *ki ale nan mache?
who Jan say that go LOC market

`Who did Jan tell to go to the market?' (=obj)

NOT: `Who did Jan say is going to the market.' (=subj)

Examples from Muysken and Law (2001, 53)

Following the analysis of LC *ki*-insertion below, this thesis proposes an alternative analysis of the HC examples as wh-copying identical to the LC pattern rather than wh-agreement. The data still lends empirical support to successive cyclicity under this analysis, it simply represents another type of empirical evidence which comes from medial wh-constructions. These include wh-scope-marking constructions but especially the WH-COPYING STRATEGY. Such appearances of intermediate wh-expressions in child language are observed by Thornton (1991), (1995) for English but also for French and Dutch as reported in Roeper and de Villiers (2011). They occur to a lesser extent in L2 learners even when neither the L1 nor the target language use a medial wh-strategy as reported by Slavkov (2008) for French Canadians learning English as an L2.

In adult speech wh-copying is attested for a number of typologically diverse languages such as German (Felser 2004, Höhle 2000, Fanselow and Mahajan 2000 among

many others)[50], Passamaquoddy (Bruening 2004, 2006) and Romani (McDaniel 1986) but also Creole languages like Afrikaans (Du Plessis 1977). In these constructions, just like observed in the speech of children acquiring different languages, a wh-moved constituent is pronounced twice, in the root position, where it appears in a long-distance question, and in the intermediate position in the embedded clause, where a copy is conventionally assumed to be deleted in long-distance questions that do not experience wh-copying. The pattern is exemplified in (213). The present thesis proposes an extension of this analysis to account for LC *ki*-insertion clauses below, effectively adding LC to the list of languages that provide empirical support for copy theory and the notion of successive cyclicity.

(213) Warum glaubst du, warum sie das getan hat? (Felser 2004, 549)
Why believe 2SG why 3SG that done has

Why do you believe she did this?

Accepting the idea of successive cyclic wh-movement on the basis of the evidence presented above, a long-distance version of (209) is derived as in (214).

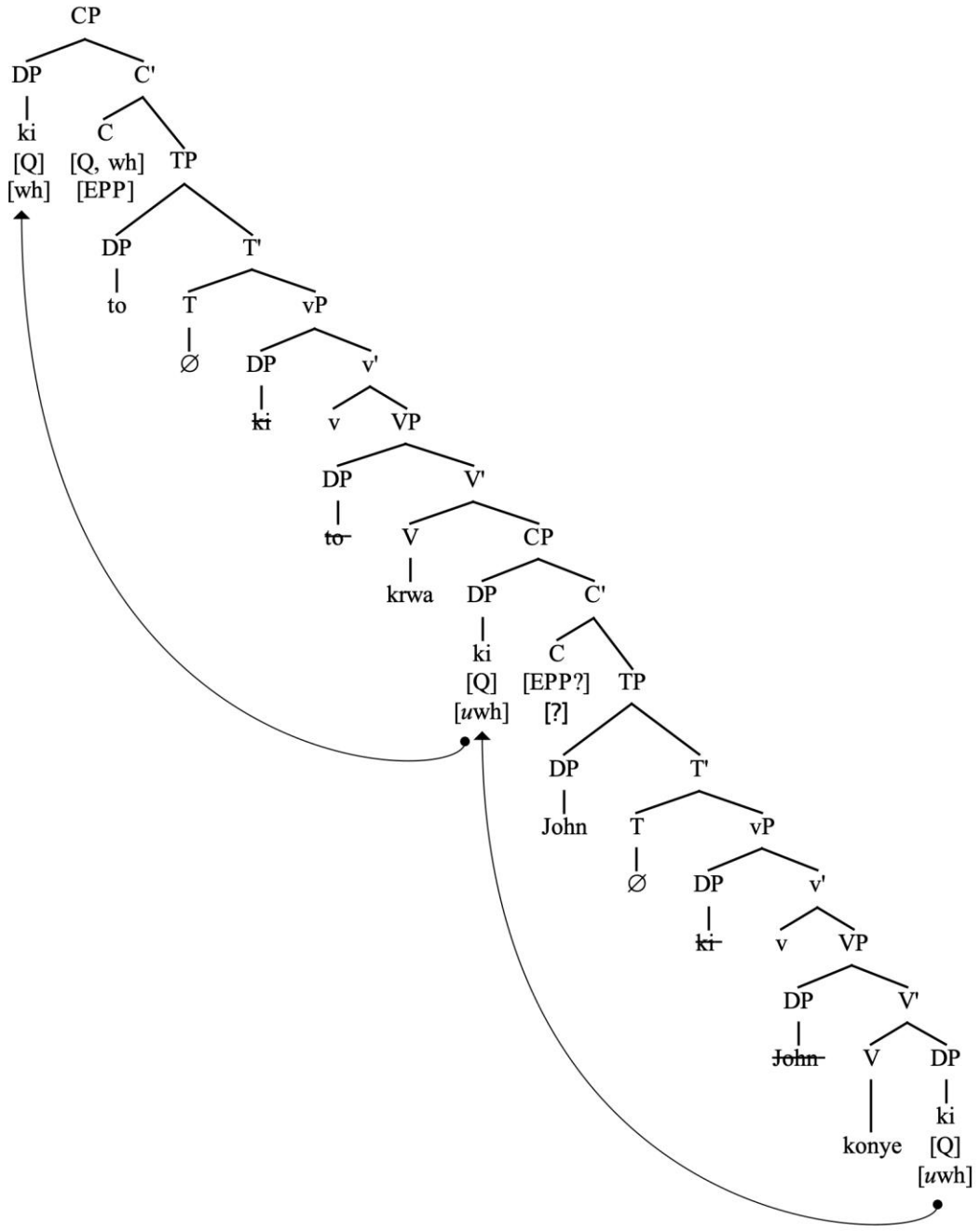
(214) Ki to krwa John konye?

(215) [$_{CP2}$ ki [$_{C}{}^{0}$ [to [~~to~~ krwa [$_{CP1}$ ~~ki~~ [$_{C}{}^{0}$ [John [~~John~~ konye ~~ki~~]]]]]]]]

One crucial and as of yet unresolved issue in minimalist syntax revolves around the features that motivate the intermediate movement step, which Felser (2004) refers to as "The triggering problem: On the assumption that agreement and hence movement is triggered by matching but uninterpretable features of the probe, what triggers movement of a *wh*-expression to the specifier of intermediate non-interrogative heads?" (Felser 2004, 547). In the example derivation in (216), this is indicated by the question marks in the features of the intermediate C head.

[50] There seems to be much agreement that wh-copying constructions can only be interpreted as evidence for the existence of a) copies and b) successive cyclicity. For opposing accounts of wh-copying see Schippers (2010) and Murphy (2016), who question its analysis as pronounced lower copies in Dutch and German respectively.

(216) Ki to krwa John konye?



The triggering problem constitutes a fundamental conceptual issue in syntactic theory. It hinges on the probe-goal relation proposed by Chomsky (2001), that requires movement to depend on feature checking. Note that attempts have been made to separate movement and agreement conceptually. For Chomsky (2001), Agree is a component of Move, so all movement must involve feature checking. In *On Phases* (Chomsky 2008) the relation is much more indirect, as agreement is concerned with Φ feature valuation and movement is triggered by [EF]. Bobaljik and Wurmbrand (2003, 2005) and Bošković (2007) propose a clear separation of the two operations while Miyagawa (2010) argues for a

unified account. If movement was in fact motivated independently from agreement and hence feature checking, the triggering problem could be solved easily by any other explanation as to why constituents move. Felser (2004) in fact propose such an alternative account which concentrates on the idea of active constituents, in the spirit of the notion of Phase Balance developed by Heck and Müller (2000) and Müller (2004). According Chomsky (2001), a constituent that has unvalued features is active and hence available for attraction into higher positions, where it can have this feature checked, i.e. agree with another constituent. If the derivation is assumed to proceed in phases, Felser (2004) argues, the wh-expression might be moved simply to avoid spell-out since it is still active which would cause the derivation to crash. According to her, the deletion of intermediate copies in long-distance questions is due to "PF-considerations -the requirement that for a wh-expression to be pronounced, its uninterpretable [wh] feature must have been checked or valued, though [Q]-agreement with a matching C-head- rather than being triggered by optionally added uninterpretable features of the intermediate phase head." (Felser 2004, 571). A similar more general account of syntactic movement is developed in Bošković (2007) where he assumes that all constituents containing an unvalued feature are probes. As probes, they move upward in the structure, from A' to A' position for wh-expressions, until they find a matching valued feature in their c-command domain which they can agree with. This account has the advantage that movement is driven by features but it does not require feature checking in intermediate landing sites. He works within the framework of phase theory but the account can just as well be applied without the notion of phasehood to which many objections have been raised (e.g. Boeckx and Grohmann 2007, Ott 2009). There are numerous issues to be worked out with both proposals revolving around the nature of [EF/EPP], the question what qualifies as a phase, how to account for empirical phenomena such as superiority effects or long-distance agreement and many more. In the words of Miyagawa (2010, ix), "one of the great mysteries of human language is the existence of movement operations." This thesis does not aspire to resolve the mystery. For its purposes the above considerations suffice to indicate the direction in which a solution is suspected.

8.3 Ki-insertion clauses, wh-copying and wh-scope marking

Returning to the analysis of LC, recall that in Chapter 7.4.1 above three types of *ki*-insertion clauses are identified. Type 1 refers to complex questions that employ the wh-phrase *ki* in the matrix as well as in the intermediate position. Type 2 uses a complex form of the wh-phrase containing *ki* in the matrix CP and *ki* in the embedded clause and Type 3 employs a phonologically unrelated form in the matrix CP and *ki* in the embedded clause. As stated above, Type 1 is the most frequent in the MTAT data with 19 occurrences closely followed by Type 2 with 18 instances. For Type 3, 9 instances are attested. These numbers refer only to the actively produced sentences in the corpus, all of which are translations of English examples that do not contain the intermediate wh-expression. They do not include judgments made for manipulated versions of long-distance constructions. However, almost all interviewees were fully acceptant of such structures indicating that *ki*-insertion is a valid strategy for LC.

This thesis proposes that Types 1 and 3 are instances of wh-copying and wh-scope marking respectively. Both phenomena are attested for a number of languages cross linguistically and they are often considered subtypes of the same construction but more recently tend to be perceived as two separate constructions. (e.g. Bruening 2006, Felser 2004).

The derivation of a wh-copying construction is generally assumed to equal the derivation of a long-distance question. It contains a wh-movement chain from the base position through the intermediate Spec,CP into the matrix Spec,CP. In phasal accounts Spec,vPs are additional intermediate landing sites adding more cycles to the derivation. The only difference between a long-distance wh-question and a wh-copying construction occurs when in the latter the copy in the lower CP is pronounced rather than deleted. Consider the LC examples where (217) represents a long-distance question and (218) a wh-copying construction of the same question. Both sentences are derived though successive cyclic movement as rudimentary sketched in (219) and formalized in (220) . Both constructions are fully acceptable to both interviewees and the other consultants and there is no difference in their interpretation according to the interviewees.

(217) Ki to kwa fe sa? (JB)

(218) Ki to kwa ki fe sa? (HW)

(219) [$_{CP2}$ Ki$_a$ to kwa [$_{CP1}$ ki$_a$ fe sa]]

(220) [CP2 WH … [CP1 WH….]]

Wh-scope marking looks quite similar but crucially does not use the same wh-expression in both places but usually relies on the least marked wh-pronoun (the equivalent of *what*) in the matrix position.

(221) [CP2 WH2… [CP1 WH1….]]

The phenomenon has been accounted for by many analyses that make two fundamentally different assumptions. Lutz, Müller, and Von Stechow (2000) distinguish approaches that involve DIRECT DEPENDENCY from INDIRECT DEPENDENCY approaches. The former would not necessarily differentiate between wh-scope constructions and wh-copying syntactically, treating both as special instances of wh-long-distance movement as they presume a direct relation between the matrix wh-expression and the embedded wh-expression, i.e. a chain or (LF) feature movement relation. Many scholars taking such a position consider WH2 to be an expletive that serves as a scope marker (Lutz, Müller, and Von Stechow 2000). This serves to account for the fact that the embedded wh-phrase in (222) has scope over the whole clause. The influential proposal by McDaniel (1989) establishes a chain between *was* and *wen*, where the former is a scope marker and the latter a full wh-phrase. The wh-phrase then undergoes LF-movement to the scope marker position. In a minimalist framework, chain formation as a formal operation can be replaced by movement. Cheng (2000) supposes that the initial scope marker is the phonetic realization of the [wh]-feature of the embedded wh-phrase that is raised to the matrix clause via feature movement and thus establishes the chain. The direct dependency approach is often taken in the analysis of German (e.g. McDaniel 1989, Cheng 2000, Höhle 2000, Brandner 2000) where wh-scope marking and wh-copying indeed share many properties.

(222) Was glaubt sie wen Paul getroffen hat?
WH2 believes she WH1 Paul met has
Who does she think Paul met (with)?

The indirect dependency approach was foremost developed by Dayal (1994). Under this view, the relation between the two wh-expressions exists only in the sense that WH2 relates to the CP1. The implication is that the matrix and the embedded clause both are fully-fledged wh-questions and the initial wh-expression originates within the upper CP, for example in object position as indicated in (223). The embedded wh-phrase serves as the semantic restriction on the proposition over which the matrix wh-expression ranges. Under this view, (222) has the semantic structure (224).

(223) [CP2 WH2… WH2[CP1 WH1…WH1]]

(224) What proposition (in the form of an answer to the question "Who did Paul meet?") does she believe?

This analysis has turned out to be more cross-linguistically appropriate (see e.g. Legate (2011) for Walpiri or Bruening (2006) for Passamaquoddy) although possibly the indirect dependency approach to scope-marking is more tenable for some languages (such as Hindi) than others (such as German). The focus of the present analysis lies on the wh-copying construction in LC rather than the wh-scope marking construction, because LC speakers clearly prefer the former. Nevertheless, it is important how wh-scope marking is analyzed because under the direct dependency approach it is not structurally distinct form the wh-copying strategy. German has provided the most arguments in favor of a direct dependency analysis and hence a unified treatment of wh-scope marking and wh-copying.

These conclusions are challenged in Felser (2004) who points out several ways in which the two differ in German. This includes the observation that wh-scope marking can occur with conjoined embedded wh-clauses which wh-scope-marking cannot as illustrated by (225) and (226).

(225) Ich frage mich was er meint [wer heute kommt] und [wann es
1SG wonder WH 3SG think who today comes and when it
losgeht]?
starts
I wonder what he thinks as to who will come today and when it will start?

(226) *Ich frage mich wer/wann er meint [wer heute kommt] und [wann
1SG wonder WH COPY 3SG thinks who today comes and when
es losgeht]?
it starts
*I wonder who/when he thinks who will come today and when it will start?

They further differ in their acceptability with a matrix raising verb like *scheinen* (=*seem*), which is ungrammatical for wh-scope marking but not for wh-copying (cf. (227) and (228)) as well as the option for full wh-phrases to appear in the intermediate position which is fine for a wh-scope marking but not for a copying construction (cf. (229) and (230)).

(227) *Was scheint es wen Hans geschlagen hat? (Höhle 2000, 251)
what seems it who Hans hit has

(228) Wen scheint es wen Hans geschlagen hat? (Felser 2004, 552)
who seems it who Hans hit has
Both: Who does it seem Hans hit?

(229) Was glaubst du welchen Mann sie liebt (Felser 2004, 552)
What believe you which man she loves

(230) *Wen glaubst du welchen Mann sie liebt?
Who$_{AKK}$ believe you which man she loves
Both: Which man do you believe she loves?

These properties alongside a few others listed in Felser (2004) and Bruening (2006) distinguish wh-copying from wh-scope marking in German, the language that presents the most arguments in favor of a unified analysis. In addition, as Bruening points out, some languages only employ one of the constructions and not the other, e.g. French which only uses wh-copying but not wh-scope-marking (cf. Reis 2000) and Hindi which uses wh-scope marking but not wh-copying (cf. Fanselow and Mahajan 2000). It can therefore be concluded, that they represent two separate constructions cross-linguistically and that assumption can be extended to LC in the spirit of uniform cross-linguistic analysis.

It follows that Type 1 *ki*-insertion clauses represent wh-copying and Type 3 *ki*-insertion clauses represent wh-scope marking. Recall that Type 3 is the least frequent and also the least well-received type in the MTAT data, which is why the following centers on the derivation of wh-copying in LC as the prominent strategy. Support for the existence of a copying strategy in LC comes from clauses that contain identical copies of the wh-expression in both position where these are not *ki*. In one instance HB finds the use of *ki-se* in both positions fully acceptable (***Ki-se*** *to kra* ***ki-se*** *fe sa?).* Furthermore, five consultants provided a sentence similar to (231) in order to represent *What did Liza show me how to cook?*, effectively reducing the number of questions asked in the sentence but still employing two wh-words, namely two instances of *komen*.

(231) Komen Liza eksplike komen fe du gòmbo? (RM)
How Liza explain how make Gumbo
How did she explain you how to cook Gumbo?

Having established that a wh-copying account is plausible for Type 1 *ki*-insertion clauses, their derivation is discussed in the following. According to Felser (2004, 562) a typical wh-copying derivation in German proceeds as in (232) to (235) where the wh-expression undergoes successive cyclic movement to the matrix Spec,CP, where the wh-feature is valued though agreement with [Q] of C^0. As argued above, I depart from Felser's analysis slightly by claiming that this involves disambiguation of [Q] in CP3, which is set to [Q, wh].

(232) Wen glaubst du wen Peter meint wen Susi heiratet?
Who do you believe Peter thinks Susi will marry?

(233) [$_{CP3}$ wen [$_{C'}$ glaubst] du…[$_{CP2}$ wen [$_{C'}$] Peter meint [$_{CP1}$ wen
who believe you who P. thinks who
[*u*wh, Q] [*u*Q] [*u*wh, Q] [*u*wh, Q]
[$_{C'}$] Susi wen heiratet]]]
S. who marries
[*u*wh, Q]

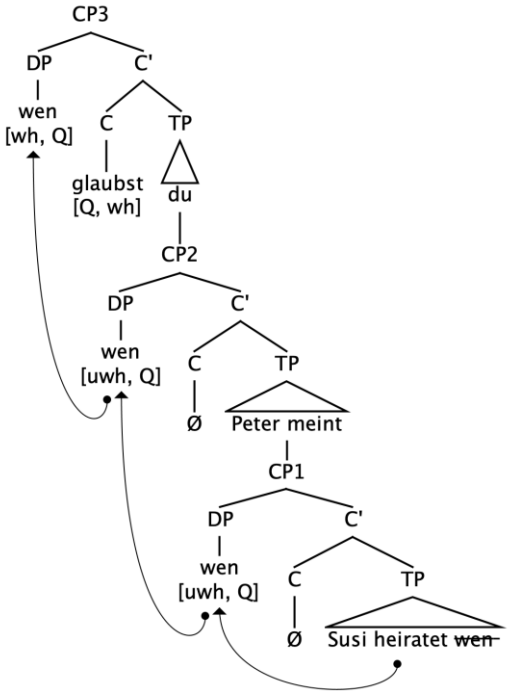
(234) [$_{CP3}$ wen [$_{C'}$ glaubst] du…[$_{CP2}$ wen [$_{C'}$] Peter meint [$_{CP1}$ wen
Who believe you who P. thinks who

[wh, Q] [Q, wh] [*u*wh, Q] [*u*wh, Q]

[$_{C'}$] Susi ~~wen~~ heiratet]]]
S. ~~who~~ marries

[*u*wh, Q]

(235) Wen glaubst du wen Peter meint wen Susi heiratet?

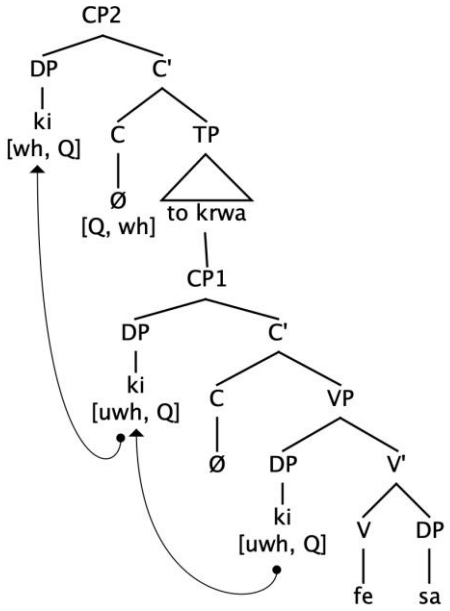


Spell-Out then applies to the entire derivation, which in Felser's account is based on the notion that phases depend on convergence. This remains a phase-based account in which no phase will be transferred to the computational systems unless all constituents therein contain only interpretable features. It could be argued that such an approach largely dispenses with phases as all features would have to be valued within a vP for example to achieve transfer of this supposed phase. At the same time this closely relates to the triggering problem, to which potential solutions are sketched above. Discussing possible configurations of phase theory is not among this thesis' goals but I still would like to suggest that under the account of Bošković (2007) the data can be analyzed with or without reference to phases. As long as the wh-feature on the object wh-phrase remains unvalued, it keeps probing for a matching goal and moves up the derivation into each available A'-position until the wh-feature is valued in the interrogative matrix CP. This can include for example the specifier position of vP, which is an edge, where the wh-phrase would escape transfer and remain available for further steps or under a convergence-based account, vP would not be transferred due to the unvalued wh-feature. Regardless, the suggestion by

Bošković (2007) seems better suited as a unified explanation for each intermediate movement step than stipulations of additional movement triggering features on the embedded declarative C-heads (e.g. [EPP] or [EF], cf. Chomsky (2000), Grewendorf and Kremers (2009)) or Felser's proposal which relies on numeration in the spirit of Heck and Müller (2000). Her idea is that the wh-expression moves to the next edge because there is a probe in the numeration with an uninterpretable feature that the constituent is able to check. Problemtically, there needs to be an awareness of what is in the numeration before these items are merged in this account. This implies a form of look-ahead that is wholly undesirable in a minimalist account and she resorts to a separate occurrence of PF and LF spell-out in which LF depends on convergence but PF transfer occurs by phase. The conclusion that cyclicity is due only to PF constraints that require the [wh]-feature to be checked for pronunciation is ultimately very similar to what Bošković (2007) proposes.

Applying this analysis to LC wh-copying is straight forward. Consider the wh-copying construction (236).

(236) Ki to krwa ki fe sa?



The wh-expression originates in the embedded clause as the subject of *fe*, hence in the VP or vP specifier position. It comes with the features [*u*wh, Q]. The presence of the unvalued wh-features drives movement to the next CP. Probing the c-command domain reveals no matching goal, so the constituent moves on to the next available A'-position which is the matrix Spec,CP, unless one assumes an intermediate landing site in Spec,vP of

the matrix clause under phase considerations. In the absence of empirical evidence for this step in the LC data, I do not.

One possible step is omitted here. The subject needs to move to the embedded Spec,TP (not included in the tree structure above) if one assumes LC T^0 to contain an [EPP] feature along the lines of English[51]. Considering that subjects appear before TMA markers, that seems plausible. The question remains if long-distance wh-movement passes through the A-position before moving on to the embedded CP or if these are two distinct movement operations and separate chains. Generally, A- and A-bar chains are considered distinct due to the *Generalized Inactivity Condition* (Chomsky 2008) that declares the root of a chain inaccessible for further operations which may motivate a phase-approach as the independent movement of two copies of the same subject wh-phrase can be well accounted for under parallel probing (Chomsky 2008) where T^0 and C^0 probe only when the C-phase is reached. For the purpose of this thesis, A-movement to the subject position is considered an unrelated operation the mechanics of which are subject to much other work (Lasnik 2006, Hornstein, Nunes, and Grohmann 2005, Manzini and Roussou 2000, Hornstein 1999) and not discussed here.

This account of Type 1 *ki*-insertion clauses as wh-copying distinguishes them from long-distance questions solely through the non-deletion of the copy in the intermediate CP which is a desirable outcome considering the optionality of *ki* in the data. Consultants often stated that one could insert or leave out the medial expression at will, which in no way indicates that these would be structurally different derivations. It also reflects the general avoidance of multiple wh-questions. While there are ways to express more than one question within one clause as presented above, LC speakers clearly prefer to include just one wh-question within a sentence. The preference for copying over wh-scope marking found in the MTAT data can also be explained on the basis of this observation if the wh-scope construction does not involve direct dependency. Two open questions remain under this approach that need to be addressed.

First, an account of Type 2 *ki*-insertion clauses, which contain *ki* in an intermediate position and a larger wh-constituent such as *ki-se* in the matrix CP (cf. Table 7-8 above) is not yet included in the proposal. Second, the MTAT data contain a subject-object asymmetry which is unaccounted for, as all spontaneously produced *ki*-insertion clauses are subject questions. The former issue, Type 2 *ki*-insertion clauses, is addressed first and illustrated by (204) repeated here as (237).

[51] It is debatable if a TP projection must even be assumed for clauses without overt TMA markers. I would argue in favor, as a zero TMA marker appears to have interpretative values in interaction with the verb resulting in either anterior or present tense interpretation and also on cross-linguistic grounds.

(237) Ki-se to kra ki lèm John? (CC/LJ)
WH 2SG believe WH like John
Who do you think likes John?

Principally, this can be subsumed under a wh-copying account, if one allows selective spell-out. The full wh-phrase would then be realized in the matrix position and a reduced form in the embedded clause. A selective spell-out is not uncommon in accounts of wh-copying and it is in line with assumptions such as the null hypothesis of Nunes (2004) COPY+MERGE THEORY OF MOVEMENT, which claims that "the simplest -therefore most desirable- version of the copy theory of movement should take traces and heads of chains to be subject to the same principles and be accessible to the same operations" (Nunes 2004, 2). Cheng (2000) suggests a separate spell-out of the wh-feature in German wh-scope marking constructions, applying a direct dependency analysis in which the wh-feature moves to the matrix CP and is spelled-out as the scope marker. Felser (2004) builds on Cheng's notion of a semantic division of wh-expressions into a wh-operator part and a core part (e.g. somebody, somewhere, something), which is familiar from Creole languages that often build their wh-expressions bimorphemically in this fashion and thus support the semantic compositionality. For future research a comparative study that investigates if other Creoles make use of wh-copying constructions and if and under which circumstances they spell out the two morphemes of their wh-phrases separately would likely provide great insight into the validity of this proposal. The problem for the present account lies in the position in which the operator and the core part appear. Felser (2004, 559) argues that the Q-operator is spelled out in the matrix clause and the content part in the lower position on the basis of German *was-für* split illustrated by the relation between (238) and (239).

(238) *Was für Bücher* hast du gekauft?

(239) *Was* hast du *für Bücher* gekauft?

Both: What kind of books did you buy?

Similarly, Bruening reconsiders his (2004) analysis of Passamaquoddy *tan* constructions and identifies them to be copying constructions as well "only now they are spelled out unequally: the lowest copy is spelled out fully, while the higher copy is spelled out in a more minimal way." (Bruening 2006, 40) He concludes that in both languages the higher copy is minimal and the semantic content is located in the lower copy. Consequently, the LC pattern runs counter the cross-linguistic evidence in that the more minimal expression always occurs in the lower position in the present data.

Since both copies are identical prior to spell-out according to the feature uniformity condition (Nunes 2004, 70), there is no reason to associate the larger constituent with one position in a principled manner, but the idea that the lower CP is the more contentful

position has great merit considering the rest of the LC data. Recall that speakers tend to answer the embedded questions for example with long-distance wh-islands such as the questions that were asked to determine if a long-distance reading of the initial wh-expression is available (e.g. *Ekan li di <u>komen</u> li fe mal a limèm?)*. They also generally prefer short-distance over long-distance readings wherever these are available. These facts point to a preference for interpreting content locally, in the lower CP as is in line with the cross-linguistic observations.

A generalization of the interpretational pattern suggested by Bruening (2006) and Felser (2004) thus seems appropriate for LC and desirable from a theoretical perspective. The spell-out pattern with the larger wh-phrase in the matrix positions contradicts this analysis at first glance, since the more minimal expression clearly is not located in the matrix CP in (237) above. However, the contradiction is just an apparent one, based on the amount of phonological material that appears in the matrix and embedded clause respectively. Bruening (2006) argues that the more CONTENTFUL item is spelled-out in the lower position. This does not necessarily have to equal the longer lexical item. *Ki-se*, *ki-se-ki* or similar wh-expressions may be more complex than a simple *ki*, but one would be hard-pressed to claim they are more contentful. LC-speakers who still use these more complex expressions, which seem to be disappearing from LC, make no difference between their interpretation and other, simpler, wh-pronouns. If any difference can be observed it is a grammatical one, for example in the spirit of Rottet's analysis of final *ki* in wh-expressions as a focus/subject marker. That distinction is being lost in LC as well since it is no longer systematically applied as shown above. However, some remnants of this grammatical function remain, even if this no longer constitutes an absolute rule but just a tendency for most speakers. For lack of a better expression, final *ki* retains a 'grammatical flavor'. The same could be said of *se* that has never been analyzed as a grammatical marker inside wh-expressions but serves as an irrealis/conditional marker or as a copula when it occurs independently. Under this view, it is unsurprising that the additional material in the wh-expression in Type 2 *ki*-insertion clauses appears in the matrix CP since it does not add semantic content to be spelled-out in the lower position but rather grammatical content to be associated with the root of the chain.

The bimorphemic MLC form *ki-moun,* on the other hand, seems to directly reflect Felser's operator-core split but is not separated upon spell out, i.e. *moun* is not spelled out in the embedded clause separate from *ki* in the matrix clause. This is a possible counter-argument. However, the expression occurs with an intermediate *ki* just once in the corpus and the item is a highly lexicalized one. In fact, considering its distribution as discussed in

Chapter 7.2 above, the point can be made that *moun* represents the lexicalized realization of a [+human] feature. In this view, *moun* belongs into the same 'grammatical flavor' category as the other lexicalized grammatical morphemes contained in LC wh-pronouns.

While it can be argued that the matrix wh-expression is not the more contentful one, nothing in the present data indicates that the lower copy is more contentful either. It follows that the conclusion to be drawn from this data must be more hesitant than what is proposed cross-linguistically. In the LC case, I suggest, like in German, the grammatical features are spelled-out in the matrix CP but the intermediate expression serves simply to maintain the cross-clausal dependency rather than representing the core part of the wh-expression as suggested by Felser (2004). It reminds the listener of the semantic connection between the matrix-initial wh-expression and the embedded proposition. This purpose is fulfilled easily by the most minimal and most semantically variable wh-expression LC has to offer, *ki*.

The second issue that needs to be accounted for is the observed subject-object asymmetry. In the MTAT data, *ki*-insertion occurs almost exclusively in subject questions. There are only two object questions that do not appear to be cases of wh-scope marking since they use an expression including *ki* in the initial position. One of these clearly is not a copying construction but an instance of the homophonous complementizer *ki* introducing the embedded clause, as it occurs in the translation of *What do you think **that** John hopes Mary will tell him?*. For the second instance given in (240), such an analysis is possible but no more or less plausible than a copying account.

(240) Ki to di (ki) John lèm? (HW)
Who 2SG say WH John like?

Who did you say John likes?

One might argue that wh-copying is not disallowed in object questions per se on the basis of this example. I do not believe that to be the case but for the purpose of this thesis, suffice it to say that *ki*-insertion, which is always optional, is strongly preferred in subject questions. Since this distinction is not attested in other wh-copying languages, it is not discussed in the literature on the phenomenon. There are parallels to other research fields however that are worth exploring.

The sensitivity to the syntactic function of the wh-moved constituent ties in with what Muysken and Law (2001) overserve for HC, where *ki* appears in an intermediate position only when the subject of the embedded clause is wh-moved. A unified analysis of both LC and HC in this regard suggests itself. They classify *ki* in the HC data as a complementizer that appears only in this particular grammatical environment. Although not explicitly stated in Muysken and Law (2001), this indicates an instance of

complementizer agreement that causes the embedded C-head to take a particular form once a wh-subject has passed through its specifier. The standard analysis for such a scenario is what McCloskey (2001) proposed for Irish, or the related suggestion by Nunes (2004) who argues that the wh-expression attaches to C^0 via adjunction and the resulting constituent is reanalyzed as a morphological unit, which is how the alternate spell-out form is generated. Such accounts can in principle be applied to all instances of wh-copying if one takes the lower copy to be a complementizer. A number of arguments can be raised against this analysis on theoretical and empirical grounds. Conceptionally, it necessitates the assumption that for every wh-constituent a homophone complementizer exists in the language (Bruening 2006)[52]. Empirically, as Felser (2004) points out, it cannot account for the co-occurrence of wh-copies and overt complementizers in some dialects of German, that do not generally allow doubly filled complementizers, or the fact that some speakers allow phrasal wh-elements such as wh-PPs to occur in the intermediate position, as these cannot be considered heads. Following these arguments, I have not taken a complementizer agreement approach to wh-copying above and on the basis of the limited data from Muysken and Law (2001) would suggest that it is not appropriate for the HC case either. Instead, if the subject-object asymmetry of LC can be accounted for, this analysis could likely be extended to HC.

A second parallel to the LC pattern exists in data from first language acquisition of e.g. English where wh-copying and wh-scope marking have been known to occur (Chapter 3.2 above, cf. Adone and Vainikka (1999) for similar observations in MC). Thornton (1991) and Thornton and Crain (1995) suggest a complementizer agreement analysis for the FLA data as well, which is rejected here on the same grounds as in the HC case above. Roeper and de Villiers (2011) notice that children acquiring their first language take longer to correctly produce long-distance subject questions than they do long-distance object questions. They report studies from English (Stromswold 1995) and French and Dutch (Jakubowicz and Strik 2008) acquisition. French and English seem to support the hypothesis that children have more difficulties with long-distance subject questions than their object counterparts. Partial wh-questions serve as an avoidance strategy and these are also more often attested in subject questions. The Dutch data is less clear. The children actually use a copying strategy in Dutch but no obvious subject-object pattern appears. Similar to what is observed here for LC, but with a higher frequency, a partial interpretation of wh-island questions also persists in FLA of several languages until the children are six or

[52] See Bruening (2006, 38) for additional support against this view from the distribution of a specific complementizer and wh-copies in Passamaquoddy.

seven years old (Roeper and de Villiers 2011, 229). The authors argue for a wh-scope-marking stage in FLA which is then replaced by cyclic movement and consider this further evidence for successive cyclicity.

Their explanation for the subject-object asymmetry remains cursory. Roeper and de Villiers (2011, 208) suggest that subject long-distance questions are more challenging than object long-distance questions despite the fact that objects actually cover the larger distance, due to intervening constituents. In particular, a wh-moved object crosses two DPs, the subjects of the embedded and the matrix clause (cf. (241)), and a wh-moved subject crosses just one DP, the subject of the matrix clause in (242).

(241) Who did Peter say Paul likes ~~who~~?
DP(WH-OBJ) DP(SUBJ) DP (SUBJ) ~~wh(OBJ)~~

(242) Who did Peter say ~~who~~ likes Paul?
DP(WH-SUBJ) DP(SUBJ) ~~DP(WH-SUBJ)~~ DP (OBJ)

The latter seems to pose more difficulty though, which they suggest depends on some feature that identifies the constituent's grammatical function. Island effects can be linked to items of the same nature passing one another in the derivation in the spirit of Relativized Minimality (Rizzi 1990), restated as Featural Relativized Minimality which claims that,

(243) In ...X...Z...Y...
A local relation is disrupted between X and Y when
1. Z structurally intervenes between X and Y
2. Z matches the specification in morphosyntactic features of X

(Villata, Rizzi, and Franck 2016, 78)

This may cause children to initially disallow the raising of one subject across another assuming subjecthood is somehow visible to the derivation.

This thesis argues that the most plausible account of the LC subject-object asymmetry must explore this line of argument. There is something about wh-subjects that makes LC speakers more likely to rely on a copying mechanism when wh-subjects are extracted. The derivation contains a wh-chain for both subject and object questions, so the cause of the asymmetry is unlikely to be found with the wh-movement steps themselves. Instead, the subject function seems to potentially cause the non-deletion of the medial copy. The number of ways in which a wh-subject differs from a wh-object in its feature configuration is limited. One option is [case], as subjects are marked for nominative in many languages. This idea is undesirable on theoretical grounds as subject case is assigned in the specifier position of the finite verb which is not part of the wh-movement chain and it is implausible for LC due to the lack of case marking. One could argue that subjects contain an

inherent focus marker of some kind which is absent in objects and used to be systematically realized in the occurrence of final *ki* in the wh-expression. It was discussed above that this distinction is being lost and, in fact, the pattern rather suggests a subject marker than a focus marker for which there is no independent evidence[53]. The most likely candidate in my view is the theta role. Theta roles are not usually represented as features but they are present in the derivation nonetheless and assigned within the VP. The lower as well as the matrix subject in a long-distance wh-question are assigned the agent role by the respective verb[54]. If the wh-subject is marked for [+agent] and has to cross another similarly marked DP in the matrix clause this can be considered a semantic intervention effect[55]. This is not a full island effect that would result in ungrammaticality but it is enough of an irritation that it allows spell-out in the lower CP for computational ease. By spell-out of *ki* in the lower copy, the movement chain is emphasized, indicating that the fronted wh-subject is in fact the agent of the embedded clause. It seems useful to think of medial *ki* in Type 1 and Type 2 *ki*-insertion clauses as optional realizations of the sematic features [Agent, wh]. This also resolves the question of how to account for the occurrence of the simple wh-phrase in the medial position rather than the matrix clause. The content that is spelled-out in this position is the agent role the wh-subject plays in the embedded clause. Further, no instances of fronted wh-adjuncts occur with intermediate *ki* in the data. This could potentially be a result of the elicitation design that did not offer many opportunities for such a construction. Should this be the regular pattern on the other hand, the present analysis accounts for this as well since adjuncts are not marked as agents any more than objects.

The proposed analysis resolves a number of seemingly unrelated issues that were touched upon in this thesis. First, as mentioned above it seems to be easily be extended to HC on the basis of the sparse available data. Thinking of the difference between (211) and (212) above as the presence/ absence of wh-copying represents a much simpler analysis than to establish a mechanism by which a certain complementizer is inserted only when an embedded subject is wh-moved, which would constitute complementizer agreement. Instead, it can be argued that HC shows a clear subject-object asymmetry with regard to wh-copying and, in contrast to LC, spell-out of the lower copy appears to be obligatory to distinguish the subject form the object question.

[53] Note that Rottet (2006) makes no distinction between the two, taking the occurrence of *ki* in subject wh-phrases as evidence for focus marking.
[54] The present study did not test for non-agentive subjects such as used in the rare passive construction described in Klingler (2003a, 316) and Neumann (1985, 279).
[55] Friedmann and Novogrodsky (2011) observe a similar effect in SLI

Secondly, recall that long-distance questions containing an overt complementizer in the embedded clause presented a somewhat unclear picture in the MTAT data. As described in the discussion of the that-t effect above (Chapter 7.3.3.4), eight consultants spontaneously inserted *ki* into the medial position in the elicited subject question although the English base sentence did not contain an overt complementizer (*Who$_i$ did you say t$_i$ likes John?*). Speakers did not do so with the corresponding object question (*Who$_i$ did you say John likes t$_i$?*). When an overt complementizer was present in the base sentence as in the adjunct question, interviewees chose *ke* rather than *ki*. This distribution is puzzling if one considers *ki* and *ke* both to be complementizers. It could only be accounted for under a complementizer agreement approach to explain the choice of *ki* over *ke* with wh-moved subjects which would still leave open why the effect does not occur with wh-moved objects or adjuncts. It further creates the need to establish agreement between a declarative C-head and the wh-phrase that unnecessarily complicates the system. Under the analysis proposed here the issue is naturally resolved because the medial *ki* in subject questions is not considered an overt complementizer but rather a wh-copy. Due to the subject-object asymmetry that strategy is unavailable in the object questions which is why interviewees did not spontaneously insert *ki* in those clauses. The complementizer is in fact translated as *ke* in the present data as exemplified in the adjunct question although the form *ki* as *that* does exist according to several consultants. Further study would have to confirm if is that truly is the case or if the perceived complementizers *ki* are actual instances of wh-copying.

One issue that is not fully resolved under the present analysis are the few instances of apparent wh-copies in non-subject contexts and/or not involving *ki*. In principle, such examples support a copying analysis for LC but at a closer look their analysis is challenging for the proposed framework. These examples occur in the MTAT data especially when consultants were asked to translate a complex sentence containing multiple wh-expressions like in like (231) repeated here as (244).

(244) Komen Liza eksplike komen fe du gòmbo? (RM)
How Liza explain how make Gumbo
How did Liza explain how to cook Gumbo?

Additionally, one time a subject question containing two instances of *ki-se* received a fully acceptable rating. The latter is easily accounted for under the copying analysis. The speaker simply accepted spell-out of full copies in both positions which is not too far from what the system currently allows in general and can be regarded as idiosyncratic variation. It actually represents a potential future development in the language considering its lack of vitality and the structural consequences arising from that circumstance. It is possible that

the reduction of the lower copy to *ki* and the grammatical subject-object distinction in wh-copying will be lost in LC's further development.

The doubling of the wh-adjunct *komen* above is less straight forwardly accounted for. Probably, constructions of this kind do not occur in spontaneous speech and despite their presence in the corpus their grammaticality is doubtful. This is due to the elicitation context. Such sentences were produced when interviewees were confronted with a translation they found difficult or even impossible. Wh-copying was chosen as a strategy to avoid wh-island effects by five speakers. Other speakers chose other strategies as described in Chapter 7.3.3.2 above. This is to say that the interviewees who produced such sentences were making an effort to include two question words as in the sentence they were given for translation but considered a direct translation inappropriate either due to island effects or processing issues. They then constructed a sentence that employs a familiar strategy (wh-copying) while asking just the question perceived to be the most important one (*komen*). The short-distance reading often assigned to *komen* in those sentences reflects the local interpretation of the embedded copy that is shared with wh-scope marking construction which are marginally possible in LC as well.

In sum, a wh-copying analysis can account well for *ki*-insertion clauses of Type 1 and Type 2. This appears to be the most minimal and therefore most suitable analysis which can additionally account for the observed subject-object asymmetry, is cross-linguistically confirmed and lends support to the copy theory of movement. One question that cannot be answered for a lack of historical data on complex questions is if this construction came to be as a consequence of language endangerment or if LC always had the option of wh-copying. Since French employes the strategy as well, it is possible that wh-copying has always been present but was not previously documented. It is equally possible that the weakening of grammatical rules has led to more liberal patterns of non-deletion and an increased need to clarify semantic relations in complex clauses.

9 CONCLUSION

This thesis set out to accomplish a combination of language documentation and analysis with a dedicated focus on the latter. Taking the endangered status of LC as a starting point, it has assessed sociolinguistic as well as structural notions and has honed in on an analysis of wh-questions in particular. In sum, it has been shown that language endangerment in LC has observable effects in terms of linguistic behavior but also on a deep structural level which is visible in the variation to be found in the distribution of wh-pronouns, produced sentences and acceptability judgments. Furthermore, several forms and structures present in the interrogative system of LC that were previously undocumented have been described. Most notably, a wh-copying construction has been revealed to provide evidence for the Copy Theory of movement.

9.1 Sociolinguistics

The demographic profile of participants in this study, if considered representative in a broad sense of the speech community, shows that the intergenerational transmission of LC has largely ended. The youngest consultants are over the age of 45, the majority was 75 years and older at the time of the interviews and language proficiency clearly declines steeply from the grandparent to the grandchild generation. Most interviewees stem from the Bayou Teche area which is where LC is considered to be the most vital (Klingler and Dajko 2006, Mayeux 2019) although language shift has progressed here significantly as well.

Concerning the sociolinguistic research three main interconnected questions (RQ 1-3 in 1.4) have been asked and answered. The first was concerned with functional domains and usage contexts, the second asked about the effects of language endangerment and the third inquired after language attitudes. The first two issues are interwoven to such a degree that they are best addressed together rather than separately.

Effects of language endangerment are clearly observable in a number of ways in the sociolinguistic data. To start with, language proficiency is highest on average among the oldest group of consultants and declines over the generations as illustrated in Figure 6-3 and Figure 6-4. Most interviewees grew up speaking LC but have now switched to English for almost all purposes. They simultaneously often consider English their best language despite the fact that it is not their L1, illustrating its cognitive dominance. This development is reflected in the interlocutors, past and present. As children the interviewees used LC with more distinct interlocutors than they do in the recent data although the preference for using LC with older generations indicates the ongoing decline of LC even then. This is most

pronounced in the youngest age group who were less likely to use LC with their peers and siblings than the older participants. Use of LC when it occurs at present was shown to depend heavily on the specific interlocutor alongside a number of usage contexts in which it more typically appears, such as casual conversation with certain people and the use as a secret code for example. Moreover, a preference for LC in cursing and thinking shows that it retains a cognitive or psychological value for some speakers that is no longer reflected in actual verbal use of the language. Generally, LC is restricted to private domains in which it is casually used with suitable interlocutors. This constitutes a diachronic development as it was more present in public domains during the interviewees' childhood. The reduction of functional domains is part of the process of language loss although LC cannot be argued to have been used in the most formal domains at any point. Unexpectedly, it is not prominently used in a cultural maintenance function, such as singing songs or telling folk-stories which one might expect from a cultural heritage viewpoint.

These results were checked against the socio-demographic variables of sex, age, race and region. Overall, socio-demographic factors appear to have little impact on the linguistic behavior covered in the questionnaire, which is common in endangered languages. It was confirmed that black speakers tend to identify more strongly with the language, as previously asserted by Neumann (1985), Dubois and Melançon (2000), Klingler (2003a), and Mayeux (2019) and reflected here in their higher rates of choosing LC as their 'best language'. At the same time, they are more reluctant to publicly use LC than white speakers on average but more likely to perform cultural functions like singing. Furthermore, women and men typically but not exclusively use LC differently. Men show a tendency to use LC for joking and cursing, while women tend to use LC for casual chats and in order to keep others from understanding conversations.

Concerning language attitudes, synchronically positive community attitudes have been contrasted with historically negative institutional attitudes that can now at best be classified as passive assimilation (cf. UNESCO 2003, 14) but used to be actively suppressive. Relatedly, community members' attitudes towards LC used to be much more negative in the past. The community seems to have developed a new sense of appreciation of the language that has replaced predominant notions of shame and rejection. The positive attitudes are reflected in a general supportive attitude towards the idea of language maintenance or revitalization. However, not all consultants are primarily concerned with LC in those matters, as intelligibility with other Creoles and varieties of French as well as cross-cultural communication are regarded positively and valued higher than distinctiveness. The actual expectations for the future development of LC among the speakers differ considerably. One

third of the interviewees believe that LC will remain alive while the remainder is pessimistic of its chances to be maintained.

In sum, oppression and the resulting endangerment of LC have left its mark on the speech community, their attitudes, their patterns of language choice, the functional domains and the level of language proficiency and intergenerational transmission to the point where large-scale efforts will be needed to change the tide for LC.

9.2 Interrogative System

The structural focus of the present thesis has been on question formation, especially wh-questions. This is the first study to focus on wh-questions in LC and, unsurprisingly, it uncovered a larger repertoire of wh-expressions and question formation strategies than was attested in previous less targeted documentation.

In answer to the fourth research question, it could be shown that LC speakers make use of a wide range of wh-pronouns including a number of previously undocumented expressions such as forms containing the LRF lexical item *kwa* (also *kwa-se, kwa-ki*), the form *ke-se* and the addition of *ki* to *ki-moun* forming *ki-moun-ki*. At the same time some complex expressions documented by Neumann (1985) and Klingler (2003a) are not represented in the data. Crucially, the short forms *ki* and *sa* are the most frequent wh-expressions in the present corpus. *Ki* is considered rare by Klingler (2003a) and restricted to object functions in Neumann (1985) but it occurs in all functions in the more recent data. *Sa*, on the other hand, is considered to be restricted to objects with [-human] referents by Neumann (1985) and Klingler (2003a) but appears with [+human] subjects and objects in the present data. Relatedly, several other forms could be shown to occur in previously unattested functions with respect to distinctions between subject and object forms as well as between [+/-human] referents. These developments indicate a loss of grammatical distinctions marked through the choice of wh-pronoun and thus represent several kinds of structural consequences to be found in endangered languages. The addition of an LRF-based wh-pronoun is an instance of introduction of new material, while the predominant use of the base forms *sa* and *ki* represent instances of simplification. At the same time the other available forms keep appearing and are in the process of losing their functional and grammatical distinctions hence implicating free variation.

In a similar vein, the predictions made by Rottet (2006) are not fully borne out in the present data. The presence of a final element *ki* in a wh-pronoun appears to be losing its subject marking function and occurs with objects and occasionally even non-human referents. Wh-subjects still are the preferred context for final *ki* but the distinction is no longer absolute. Furthermore, the *ki*-final forms only occur in the speech of the more

proficient speakers. Older proficient speakers are the most likely to abide by the distributional configuration observed by Rottet (2006) but the younger and/or less proficient the speakers are, the less likely they are to do so. This indicates a development that, I hypothesize, will probably lead to the loss of this grammatical distinction followed by the loss of the *ki*-final forms in favor of simple *ki* and *sa*.

Concerning syntactic strategies for question formation (RQ 5), a number of strategies and phenomena have been identified. The gap strategy of wh-movement has been shown to be the default strategy for root questions. A contact-induced development that is ascribed to influence from English as previously noted by Neumann-Holzschuh (2009) is preposition stranding. In the present data, this strategy is preferred over preposition pied-piping, counter the observations made in Klingler (2003), and for some speakers pied-piping has ceased to be a valid strategy. In a partial answer to RQ 9 about the structural effects of contact to English, it is concluded that LC has adopted the English preposition stranding strategy to a large extent.

Leaving a wh-expression in-situ is not a standard strategy, not even as a regular optional pattern as in SF, but acceptance levels vary with regard to grammatical functions. In-situ wh-subjects are clearly rejected but in-situ wh-objects appear to be more acceptable and leaving wh-adjuncts in-situ is even less marked. While this distribution suggests a development towards the French grammatical system, this connection could not be corroborated with the socio-demographic data, as proficiency in LRF or SF do not correspond to higher acceptance rates of wh-in-situ structures. The most plausible explanation therefore is that LC, as a Creole which is structurally closer to its lexifier than many others, always had the in-situ option under certain circumstances at least for some speakers. The construction was unlikely to appear in studies not targeted at wh-questions.

Further available structural strategies are embedded questions and complex wh-questions which were nor previously described. Within complex questions, LC has been shown to be sensitive to island effects, in particular wh-islands and to a lesser extent adjunct islands though no obvious that-t effect appeared. For wh-islands no argument-adjunct asymmetry could be detected but adjunct islands, which this study generally finds to be weaker than in English, show a difference between subjects, which are more easily extracted, and objects and adjuncts. The island effects are less severe or even absent for select speakers, indicating considerable differences between I-grammars. Generally multiple wh-constructions asking more than one question within the same sentence are marginal to LC if not fully ungrammatical. Although wh-scope-marking is not clearly attested as a regular strategy for LC, speakers often fall back on a partial interpretation

when faced with questions containing two wh-expressions. Fronting more than one wh-expression is prohibited and leaving a distinct wh-expression in an intermediate position or in-situ is avoided if possible. This relates to a preference for short-distance readings where one is available.

In conclusion, beyond the standard strategy for question formation, LC shows much structural variation which is independent of socio-demographic variables. Instead, the differences seem to occur between I-grammars, largely idiosyncratically, which is common for endangered languages and has been argued for with regards to Creole languages for example by Baptista (2017). That such differences occur in the field of wh-questions, which has never been identified as an area affected by a Creole continuum, indicates that effects of language endangerment persist on this deeply structural level.

9.3 Wh-Copying

Concerning the analysis of LC wh-questions in the MP (RQ 7), a standard analysis of wh-movement in the minimalist program has been shown to account for LC root clauses. This thesis has argued for two slight modifications of the account outlined in Chomsky (2000). First, following Bošković (2007), no [EPP] or [EF] feature is postulated on C-heads as the presence of an unvalued features is considered sufficient to motivate movement because the constituent probes its c-commanding domain until it finds a matching goal. Each wh-movement step is hence triggered by the presence of an unvalued [wh]-feature on the wh-item until the interrogative C^0 is c-commanded by the expression and values this feature. Second, following Simpson (2000) and Alexandre (2012) a disambiguation view of the interrogative C-head is taken which suggests that the [Q]-feature on C is set to the value [+wh] due to agreement with the wh-expression.

The focus of the further generative analysis has been on the issue of successive cyclicity (RQ 8) and the Copy Theory of Movement due to the identification of *ki*-insertion clauses as an unexpected construction in the data. These are long-distance questions which contain an intermediate item *ki*. This thesis has argued that these clauses are best analyzed in terms of wh-copying and wh-scope marking. Type 3 *ki*-insertion clauses, which use a distinct wh-expression (most commonly *sa*) in the matrix clause and *ki* in the initial position of the embedded clause have been argued to be instances of wh-scope marking under an indirect dependency analysis in the spirit of Dayal (1994). This is largely based on cross-linguistic grounds and the fact that the two wh-expressions do not experience co-indexing in LC.

Furthermore, Type 1 and Type 2 *ki*-insertion clauses are best analyzed as wh-copying, hence these represent wh-movement chains in which more than one copy is at

least partially pronounced. The copying analysis captures the optional nature of the intermediate instance of *ki* and it fits well with the avoidance of multiple wh-questions since the two wh-expressions are actually copies of the same item. Following the derivation of copying constructions proposed in Felser (2004), Type 1 *ki*-insertion clauses thus provide empirical support for the Copy Theory of Movement. Type 2 clauses differ in that their initial wh-expression is not *ki* but an expression containing *ki*. For these clauses partial spell-out of the lower copy must be assumed which runs counter cross-linguistic patterns where the higher copy tends to be more minimal rather than the lower one. On the basis of the notion of a more contentful lower copy suggested by Bruening (2006), it was argued that the additional material in the higher copy of LC does not in fact represent additional content. Instead the *se* and *ki*-elements in the complex wh-expression were considered to be remnants of a subject and potentially a focus marker which thus represent grammatical information. While their function as grammatical markers in LC appears to be lost as described above, they still retain a grammatical character which is fully in line with the assumption that the formal characteristics of the wh-expression are expressed in the matrix CP. The operator core split prosed by Cheng (2000) and Felser (2004) cannot be confirmed with the LC data. In its stead, the higher copy is taken to represent formal features while the lower copy expresses a semantic characteristic, most likely the agent theta role of the embedded subject. Spelling-out a minimal form of the lower copy hence allows speakers to affirm that the fronted wh-subject is the agent of the embedded clause which makes the sentence more accessible as a local interpretation becomes available. Simultaneously, this analysis accounts for the subject-object asymmetry observed with *ki*-insertion clauses. Intermediate *ki* appears only with long-distance subject questions because the agentive wh-expression crosses over the matrix clause subject which is an intervener in the sense of featural relativized minimality (Villata, Rizzi, and Franck 2016). This does not constitute a strong enough effect to result in ungrammaticality but it suffices to license the spell-out of the lower copy.

9.4 Suggestions for future research

A number of approaches for future studies emerge from this thesis. It would be informative to conduct similar research on other endangered as well as other Creole languages and compare the results to the present findings. Studies along these lines could reveal if similar variation concerning question formation strategies, interveners etc. exists in Creole grammars, which are known for their variability, or if the observed effects must be fully ascribed to the structural consequences of language endangerment. This can only be

answered with a comparative study and goes beyond the scope of this thesis. At the same time, the sociolinguistic results should also be compared cross-linguistically.

Another approach to the same question would be a longitudinal study that compares older forms of LC with the present and future data. While this is in principle a most promising approach, the historical data is too scarce to allow for claims about the marginal constructions discussed here. Nevertheless, a follow-up study on LC question formation would have the potential to reveal if and how the described developments proceed in the future especially among the new speakers who now begin using LC as their heritage language. Moreover, it could determine if some predictions made in this work, such as the eventual loss of complex wh-pronouns and the dominance of *ki* and *sa*, are in fact correct. In addition, further research into categories that are underrepresented in the current study would also be beneficial. This primarily concerns non-human subjects and long-distance adjunct questions. Any such study should include a more balanced data set to better account for regional variation if consultants for further study of the MLC variety can be found.

In the wider field of Creole studies, it would be especially promising to investigate the occurrence of wh-copying and wh-scope marking in other Creole languages and include data that puts the distinction between the derivations to the test on the basis of the examples from better researched languages like German. This should include the properties that lead scholars to distinguish between the two structures in the first place (Felser 2004, Bruening 2006, and contributions in Lutz, Müller, and Von Stechow 2000). As mentioned earlier, Creoles with bimorphemic wh-expressions seem especially well suited to test the operator-core split proposed for selective spell-out in wh-copying constructions.

APPENDIX

A. Informed Consent Form

University of Cologne

University of Cologne • Albertus-Magnus-Platz • 50923 Cologne

Department of English

Applied English Linguistics
Kathrin Klein

Phone US
Phone +49 221 470 3041
k.klein@uni-koeln.de

Informed Consent

Dear participants,

Thank you very much for taking the time to help. By signing this document you allow me to use your input for my work.

Please understand that,

1. You can withdraw your consent completely or in part anytime.
2. You can disallow publication of particular statements, discussion of certain topics etc. anytime and independently from the general consent you are asked to give below.
3. I agree to honor your contribution to the work appropriately.

Please choose which actions you would like to allow by circling the respective letter below.

I give Kathrin Klein the right to:

a. Make audio-recordings during our session(s).
b. Make video-recordings during our session(s).
c. Use these recordings as well as the form and content of what I say during the session(s) for publications of the following kind, unless I state otherwise (see 2. above):
    i. PhD thesis.
    ii. Conference papers/ talks.
    iii. Teaching.
    iv. Articles in books or journals.
    v. Future research.
d. Mention my name in publications in the following form:
    i. Full name.
    ii. Surname.
    iii. Alias/ Pseudonym.
e. Store the recordings and related information with an archive.

Louisiana,_____
Date

_____ _____
Name Signed

B. Sociolinguistic Questionnaire

1. Biographical Information
    1. Name
    2. Gender
    3. Year of Birth
    4. Hometown
    5. Place of Residence? (how long?)
    6. Other places you have lived (and length of stay)
    7. Occupation
    8. Marital Status
    9. Education
    10. Ethnicity (self-assessment)
    11. (number and age of) children
    12. (number and age of) grandchildren

2. Language
    a. Background
        1. Which languages do you speak?
        2. Which languages do you understand?
        3. First language used as a child?
        4. Did you learn French at school? How long?
        5. Which language do you speak best?/ Which language are you most comfortable with? Which language do you speak at home?
        6. Which language do you speak most?
        7. Which language did your grandparents speak best?/ Do/ Did your grandparents speak to you?
        8. Which language do/ did your parents speak best?/ Do/ Did you parents speak to you?
        9. What languages do your children speak/ understand?
        10. Which language do your children speak best?
        11. What languages do your grandchildren speak/ understand?
        12. Which language do your grandchildren speak best?

b. Competence

1. Can you do the following in LC?

Scale 1-3: 1 without difficulty, 2 with difficulty, 3 not at all.

i. Count to 10

ii. Name the days of the week

iii. Name the complete date.

iv. Order a meal at a restaurant

v. Give biographical information about yourself.

vi. Carry out an appropriate conversation at birthdays, church etc.

vii. Describe your hobbies in detail.

viii. Describe details from your employment, studies, social events to a native speaker.

ix. Describe your hopes and fears concerning the future in the correct tense.

x. Discuss a controversial subject.

2. Please translate to Louisiana Creole:

I am going to go.

I went he store.

3. Usage

Indicate on a scale of 1-5:

1 Creole; 2 mostly Creole; 3 Creole/ English; 4 mostly English; 5 English

1. When you were a child, which language did you speak to

i. Your parents

ii. Your grandparents

iii. Your siblings

iv. Your friends and classmates at school

v. Your friends outside school

2. Today, which language do you usually speak to

i. Your parents

ii. Your siblings

iii. Your children

iv. Your grandchildren

v. Your spouse

vi. Your neighbors/ friends

vii. Co-workers

viii. Shop keepers

3. Of the groups in 2, which consist mostly of creole speakers? Is there a group of people in your life you use LC with primarily?

4. Are the following statements true for you?

Scale 1-5:

1 always; 2 often; 3 sometimes; 4 rarely; 5 never

i. I have regular transactions in Louisiana Creole in my everyday life.

ii. I speak Louisiana Creole at my home.

iii. I speak Louisiana Creole at my work.

iv. I speak Louisiana Creole running errands.

v. I conduct official business speaking Louisiana Creole (institutions, landlord etc.)

vi. I think in Louisiana Creole.

vii. I dream in Louisiana Creole.

viii. I speak Louisiana Creole on the phone.

ix. I speak Louisiana Creole on the streets for a chat.

x. I sing songs in Louisiana Creole.

xi. I tell jokes in Louisiana Creole.

xii. I count in Louisiana Creole.

xiii. I curse in Louisiana Creole.

xiv. I speak Louisiana Creole to keep others from understanding what I'm saying.

xv. I tell stories in Louisiana Creole.

4. Language Attitudes
    1. For which purposes do you use Louisiana Creole and English respectively?
    2. How do you feel about Louisiana Creole? How do others around you think about it? Are you proud of speaking Creole?
    3. In your opinion, will Louisiana Creole stay alive in Louisiana? Please explain.
    4. In your opinion, should people try to keep Louisiana Creole alive? Why?
    5. In your opinion, is the Creole language of Louisiana the same everywhere it is spoken? Where is it different and in which way?
    6. Do young people speak the same Louisiana Creole as older people?
    7. Do you feel that speaking Louisiana Creole is important to who you are?
    8. Is speaking Louisiana Creole necessary to be considered a Creole?
    9. Do you feel that speaking Louisiana Creole sets speakers apart from others?
    10. In your opinion, is the English language a good thing for your community?

C. MTAT

Scale: 1- fully acceptable; 2- ok, but a bit strange; 3- maybe acceptable; 4- I would not say it, others might/ very marked; 5- completely unacceptably

Manipultaed Translation Acceptability Task

Participant:
Date:

How do you say….	**Could you also say… Manipulate translation to match:**
Who arrived?	Arrived who?
Who did John hit?	John hit who?
The town is where John arrived.	
Where did John go?	John went where?
Who told John?	
Who does John believe Mary met?	John believes Mary met who? John believes that Mary met who?
Why did John hit him?	
Who have you seen?	You have seen who?
	If/ that you have seen who?
About who(m) did they talk?	Who did they talk about?
This is who they talked about.	
I know who did this.	Who I know did this.

Who did this?

Who do you think did this?	What do you think who did this?
I know who(m) he hit.	
I know to who(m) she talked.	I know who she talked to.
	I know she talked to who.
Who she talked to is none of your business.	
I don't know why he left.	
Why did you avoid talking to him?	
Mary can tell him what she wants.	
What do you refuse to tell him?	
Who did you avoid talking to last night?	To who(m) did you avoid talking last night?
What do you think that John hopes Mary will tell him?	
Who did you say hit whom?	What did you say who whom hit?
Who saw what when?	What saw who when? When saw who what? Who what when saw?

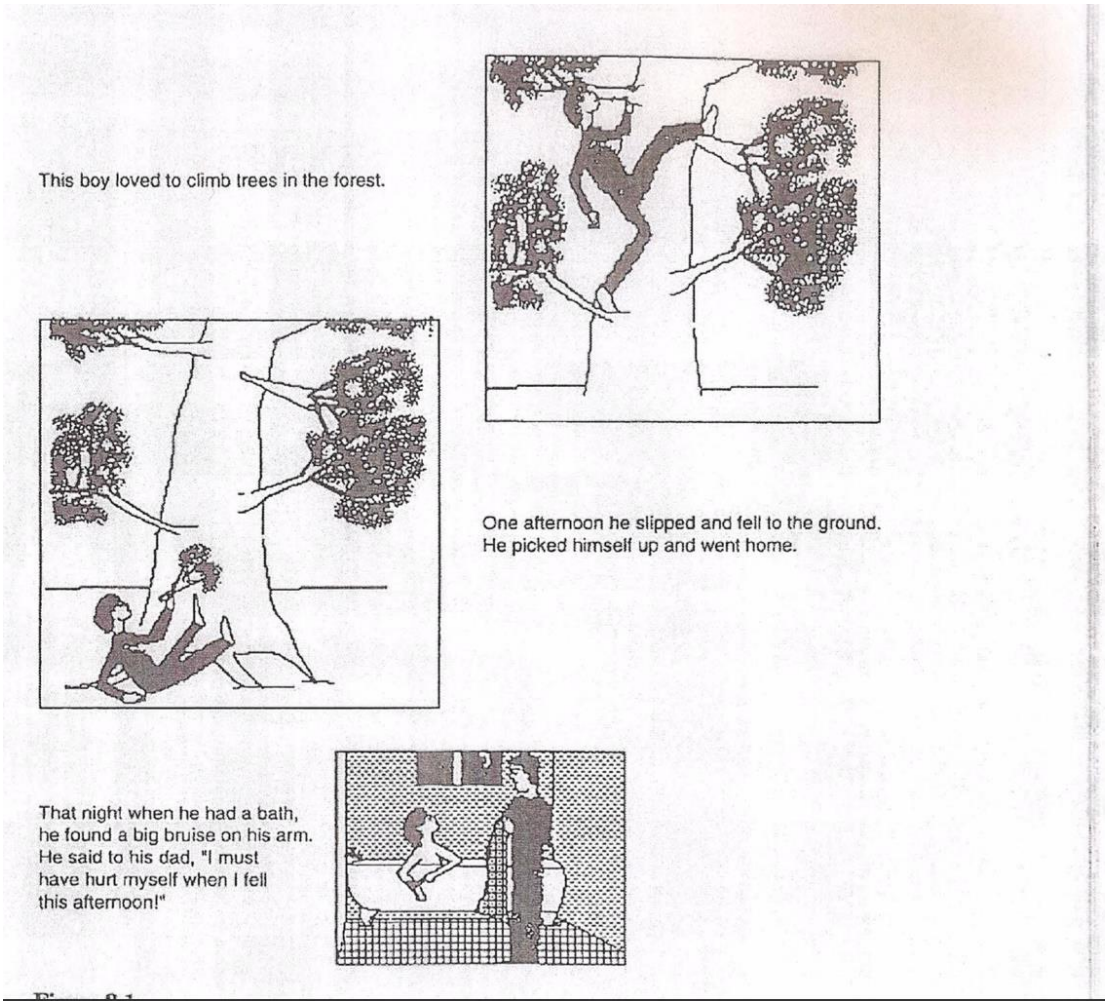
	What who when saw?
	When who what saw?
	Saw who what when?
Who did you say saw what when?	What did you say saw who when?
	When did you say saw who what?
	Who what when did you say saw?
	What who when did you say saw?
	When who what did you say saw?
	Did you say saw who what when?
	Who did you say what saw when?
	Who did you say that likes John?
Who did you say likes John?	Who did you say John likes?
Who did you say that John likes?	
Why does John only like difficult girls?	
How did he say that they resolved the problem?	How did you say they resolved the problem?
context: A: How did Paul help Peter yesterday?	

B: By finding him a place to stay. C:	
Who did you say helped Peter how?	Who did you say how helped Peter?
	Who did you say how Paul helped?
	How did you say who helped Peter?
Where is Peter staying?	
When did you say how Paul helped Peter?	Answer: yesterday possible?
I know how Paul helped Peter.	
context: Liza patiently taught me how to cook Gumbo.	
What did Liza explain how to cook?	
I know why Liza taught you to cook Gumbo.	
How did she explain what to cook?	Answer: patiently. With Ocras and rice also possible?
context: John is angry because Paul hurt Mary with his words.	
John is angry because Paul hurt who?	Who is John angry because Paul hurt?
John is angry because who hurt Mary?	Who is John angry because hurt Mary?

John is angry because Paul hurt Mary how?	How is John angry because Paul hurt Mary?
You think that who came to the party?	
Who do you think came to the party?	

D. Questions after Stories

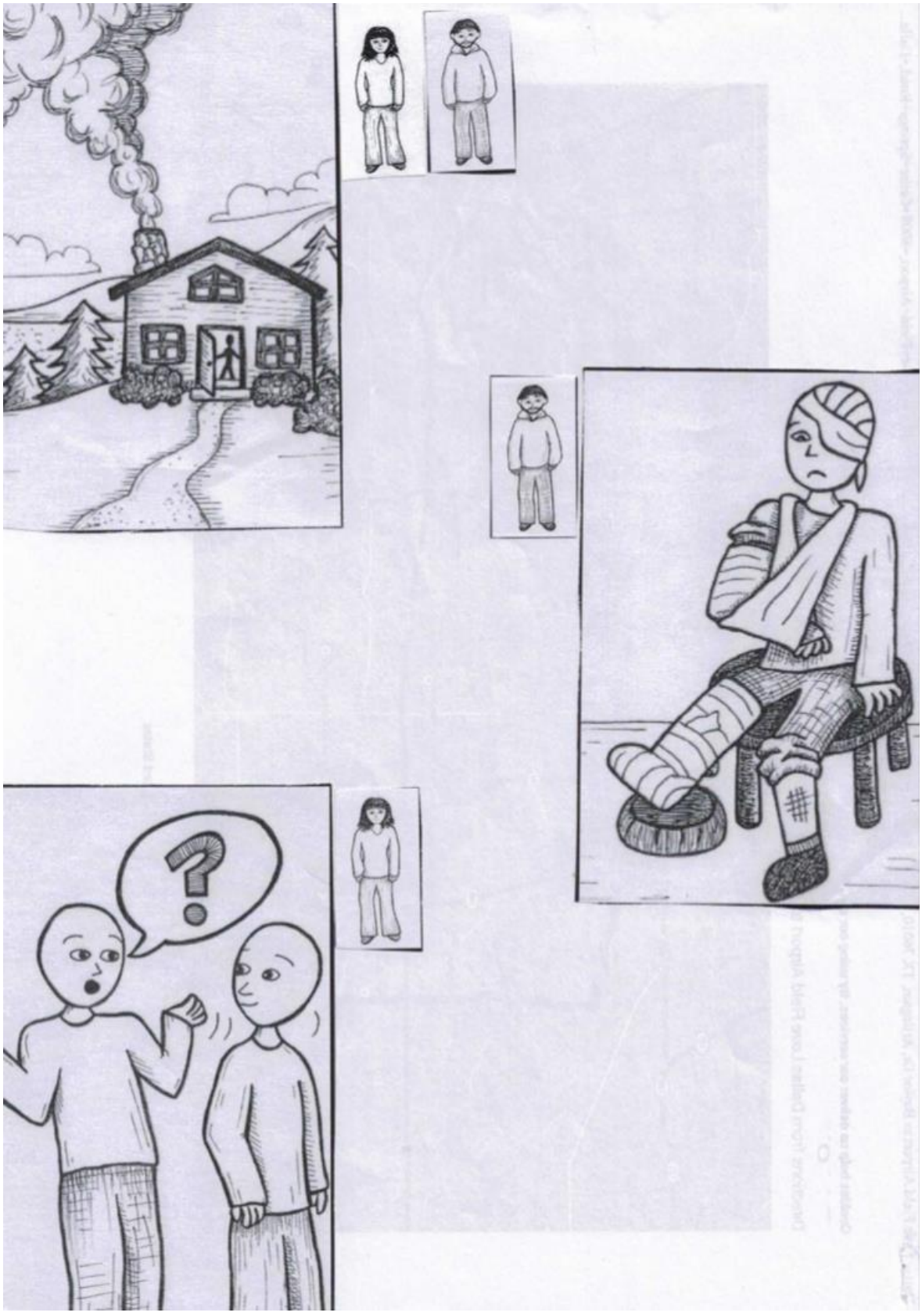
Picture Story 1



"When did he say how he hurt himself?"

"Ekan li di komen lie fe mal a limèm?"

Picture Story 2



"Who did she ask how to help?"

"Ki li mande komen ede?"

LIST OF REFERENCES

Aboh, Enoch O. 2004. *The morphosyntax of complement-head sequences: Clause structure and word order patterns in Kwa*. Oxford: Oxford University Press.

Aboh, Enoch O., and Roland Pfau. 2011. "What'sa wh-word got to do with it." *Mapping the left periphery: The cartography of syntactic structures* 5:91-124.

Aceto, Michael. 1999. "Looking beyond decreolization as an explanatory model of language change in creole-speaking communities." *Journal of Pidgin and Creole Languages* 14 (1):93-119.

Adli, Aria. 2006. "French wh-in-situ questions and syntactic optionality: Evidence from three data types." *Zeitschrift für Sprachwissenschaft* 25 (2):163-203.

Adone, Dany. 2008. "Fieldwork research: A Matter of Taking and Giving." In *Was ist linguistische Evidenz? Kolloquium des Zentrums Sprachenvielfalt und Mehrsprachigkeit, November 2006*, edited by Claudia Maria Riehl and Astrid Rothe, 109-124. Aachen: Shaker Verlag.

Adone, Dany. 2012. *The acquisition of creole languages: How children surpass their input*. Cambridge [u.a.]: Cambridge University Press.

Adone, Dany, and Anne Vainikka. 1999. "Acquisition of wh-questions in Mauritian Creole." In *Language Creation and Language Change. Creolization, Diachrony, and Development*, edited by Michael De Graff, 75-94. Cambridge, Mass: MIT Press.

Aikhenvald, Alexandra Y. 2012. "Language contact in language obsolescence." In *Dynamics of contact-induced language change*, edited by Claudine Chamoreau and Isabelle Léglise, 77-109. Berlin, Boston: De Gryuter Mouton.

Alba, Richard, John Logan, Amy Lutz, and Brian Stults. 2002. "Only English by the third generation? Loss and preservation of the mother tongue among the grandchildren of contemporary immigrants." *Demography* 39 (3):467-484.

Alexandre, Nélia. 2012. *The defective copy theory of movement: Evidence from wh-constructions in Cape Verdean Creole*. Vol. 41, *Creole Language Library*. Amsterdam, Philadelphia: John Benjamins Publishing.

Ancelet, Barry Jean. 1994. *Cajun and Creole folktales: The French oral tradition of south Louisiana*. Jackson: University Press of Mississippi.

Anderson, Janet I. 1983. "Syllable simplification in the speech of second language learners." *Interlanguage Studies Bulletin*:4-36.

Angouri, Jo. 2010. "Quantitative, Qualitative or Both? Combing Methods of Linguistic Research." In *Research Methods in Linguistics*, edited by Lia Litosseliti, 29-48. London: Bloomsbury.

Ansaldo, Umberto, Stephen Matthews, and Lisa Lim. 2007. *Deconstructing creole*. Amsterdam, Philadelphia: John Benjamins Publishing.

Arends, Jacques, Pieter Muysken, and Norval Smith. 1994. *Pidgins and creoles: An introduction*. Vol. 15, *Creole Language Library*. Amsterdam, Philadelphia: John Benjamins Publishing.

Asher, R. E., and Christopher Moseley, eds. 2007. *Atlas of the world's languages*. 2. ed. ed. London: Routledge.

Austin, Peter, and Julia Sallabank, eds. 2011. *The Cambridge handbook of endangered languages*, *Cambridge handbooks in linguistics*. Cambridge, New York: Cambridge University Press.

Baker, Philip. 1984. "Agglutinated French articles in Creole French: their evolutionary significance." *Te reo* 27:89-129.

Bakker, Peter. 2008. "Pidgins versus creoles and pidgincreoles." In *The handbook of pidgin and creole studies*, edited by Silvia Kouwenberg and John Victor Singler, 130-157. Wiley Online Library.

Bakker, Peter, Finn Borchsenius, Carsten Levisen, and Eeva Sippola. 2017. *Creole Studies – Phylogenetic Approaches*: John Benjamins.
Baptista, Marlyse. 2015. "Continuum and variation in Creoles: out of many voices, one language." *Journal of Pidgin and Creole Languages* 30 (2):225-264.
Baptista, Marlyse. 2017. "Competing I-grammars in creole genesis: A synchronic and diachronic view." *Journal of Pidgin and Creole Languages* 32 (2):398-415.
Barrena, Andoni, Esti Amorrortu, Ane Ortega, Belen Uranga, Esti Izagirre, and Itziar Idiazabal. 2007. "Does the number of speakers of a language determine its fate?" *International journal of the sociology of language* 2007 (186):125-139.
Bhopal, Kalwant. 2018. *White privilege: The myth of a post-racial society*. Bristol: Policy Press.
Bickerton, Derek. 1980. "Decreolisation and the creole continuum." In *Theoretical orientations in creole studies*, edited by Albert Valdman, 109-127. New York [u.a.]: Acad. Press.
Bickerton, Derek. 1984. "The language bioprogram hypothesis." *Behavioral and brain sciences* 7 (2):173-188.
Bickerton, Derek. 1988. "Creole languages and the bioprogram." In *Linguistics: The Cambridge Survey*, edited by Frederick Newmeyer, 267–84. Cambridge: Cambridge University Press.
Bickerton, Derek. 2016. *Roots of language*. Berlin: Language Science Press. Original edition, 1981.
Blainey, Darcie. 2017. "Sociolinguistic research with endangered varieties: The case of Louisiana French." *Canadian Journal of Linguistics/Revue canadienne de linguistique* 62 (4):576-595. doi: https://doi.org/10.1017/cnj.2017.30.
Blair, Heather, and Shirley Fredeen. 1995. "Do not go gentle into that good night. Rage, rage, against the dying of the light." *Anthropology & education quarterly* 26 (1):27-49.
Bobaljik, Jonathan D, and Susi Wurmbrand. 2003. "Long distance object agreement, restructuring and anti-reconstruction." *Proceedings of the North East Linguistic Socitey* 33:67-86.
Bobaljik, Jonathan D, and Susi Wurmbrand. 2005. "The domain of agreement." *Natural Language & Linguistic Theory* 23 (4):809-865.
Boeckx, Cedric. 2008. *Understanding minimalist syntax: Lessons from Locality in Long-Distance Dependencies*. Malden: Blackwell.
Boeckx, Cedric, and Kleanthes K Grohmann. 2003. *Multiple wh-fronting*. Vol. 64. Amsterdam, Philadelphia: John Benjamins Publishing.
Boeckx, Cedric, and Kleanthes K Grohmann. 2007. "Remark: Putting phases in perspective." *Syntax* 10 (2):204-222.
Bošković, Željko. 1997. *The syntax of nonfinite complementation: An economy approach*. Cambridge, Mass [u.a]: MIT Press.
Bošković, Željko. 1998. "LF movement and the Minimalist Program." *Proceedings of the North East Linguistic Socitey* 28:43-57.
Bošković, Željko. 2002. "On multiple wh-fronting." *Linguistic inquiry* 33 (3):351-383.
Bošković, Željko. 2007. "On the locality and motivation of Move and Agree: An even more minimal theory." *Linguistic inquiry* 38 (4):589-644.
Bradley, David. 2002. "Language Attitudes: the key factor in language maintenance." In *Language endangerment and language maintenance*, edited by David Bradley and Maya Bradley, 1-10. London: Routledge.
Bradley, David, and Maya Bradley, eds. 2002. *Language endangerment and language maintenance*. London: Routledge.
Brandner, Ellen. 2000. "Scope marking and clausal typing." In *Wh-scope marking*, edited by U. Lutz, G. Müller and A. von Stechow, 45-75. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Brandt, Kathrin, and Astrid Gabel. 2017. "Where do we go from here? On the Developmental Paths of two Creole Languages." SPCL Summer Conference, Tampere, Finland.
Bruening, Benjamin. 2004. "Two types of wh-scope marking in Passamaquoddy." *Natural Language & Linguistic Theory* 22 (2):229-305.
Bruening, Benjamin. 2006. "Differences between the wh-scope-marking and wh-copy constructions in Passamaquoddy." *Linguistic Inquiry* 37 (1):25-49.
Cable, Seth. 2010. *The grammar of Q, Q-particles, WH-movement, and pied-piping*. Oxford [u.a.]: Oxford University Press.
Campbell, Lyle, and Martha C Muntzel. 1989. "The structural consequences of language death." In *Investigating obsolescence: Studies in language contraction and death*, edited by Nancy Dorian, 181-196. Cambridge: Cambridge University Press.
Chambers, Jack K, and Peter Trudgill. 1998. *Dialectology*. Cambridge: Cambridge University Press.
Cheng, Lisa. 1997. *On the typology of wh-questions*. New York, London: Garland Publishing Inc.
Cheng, Lisa. 2000. "Moving just the feature." In *Wh-scope marking*, edited by U. Lutz, G. Müller and A. von Stechow, 77-99. Amsterdam, Philadelphia: John Benjamins Publishing Company.
Cheng, Lisa, and Norbert Corver. 2006. *Wh-movement: Moving on*. Vol. 42, *Current studies in Linguistics*. Cambridge, Mass.: The MIT Press.
Cheng, Lisa, and Johan Rooryck. 2000. "Licensing Wh-in situ." *Syntax* 3:1-19.
Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
Chomsky, Noam. 1964. *Current issues in linguistic theory*, *Janua linguarum Series minor*. The Hague: Mouton.
Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.
Chomsky, Noam. 1973. "Conditions on Transformations." In *A Festschrift for Morris Halle*, edited by Stephen Anderson and Paul Kiparsky, 232-286. New York: Holt, Rinehart & Winston.
Chomsky, Noam. 1977. "On Wh-movement." *Formal Syntax: proceedings of the 1976 MSSB Irvine Conference on the Formal Syntax of Natural Language*:71-132.
Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, Mass: MIT Press.
Chomsky, Noam. 1986a. *Barriers*. Cambridge, Mass.: MIT Press.
Chomsky, Noam. 1986b. *Knowledge of language: Its nature, origin, and use*. Westport, Conneticut, London: Praeger Publishers.
Chomsky, Noam. 1993a. *Lectures on government and binding: The Pisa lectures*. 7 ed. Berlin, New York: Mouton de Gruyter.
Chomsky, Noam. 1993b. "A minimalist program for linguistic theory " In *The view from building 20*, edited by Kenneth Hale and Samuel J. Keyser, 1-52. Cambridge, MA: MIT Press.
Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Mass.: MIT Press.
Chomsky, Noam. 2000. "Minimalist inquiries: the framework." In *Step by Step. Essays on Minimalist Syntax in Honor of Howard Lasnik*, edited by Roger Martin, David Michaels and Juan Uriagereka, 89-155. Cambridge, Mass.: MIT Press.
Chomsky, Noam. 2001. "Derivation by phase." In *Ken Hale: A life in language*, edited by Michael Kenstowicz, 1-52. Cambridge, MA: MIT Press.
Chomsky, Noam. 2008. "On phases." *Current Studies in Linguistics Series* 45:133.
Chomsky, Noam, and Howard Lasnik. 1977. "Filters and control." *Linguistic inquiry* 8 (3):425-504.
Chomsky, Noam, and Howard Lasnik. 1993. "The theory of principles and parameters." In *Syntax: An international handbook of contemporary research*, 506-569. Berlin: de Gruyter.

Chung, Sandra. 1994. "Wh-agreement and "referentiality" in Chamorro." *Linguistic inquiry* 25 (1):1-44.
Cinque, Guglielmo. 1990. *Types of Ā-dependencies*. Cambridge, Mass: MIT Press.
Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.
Cinque, Guglielmo, and Luigi Rizzi. 2008. "The cartography of syntactic structures." *Studies in linguistics* 2:42-58.
Clements, J Clancy, and Ahmar Mahboob. 2000. "WH-words and question formation in pidgin/creole languages." In *Language Change and Language Contact in Pidgins and Creoles*, edited by John McWhorter, 459-497. Amsterdam: John Benjamins.
Corver, Norbert, and Jairo Nunes. 2007. *The copy theory of movement*. Vol. 107: John Benjamins Publishing.
Crystal, David. 2000. *Language death*. Cambridge, UK; New York, NY: Cambridge University Press.
Culicover, Peter W. 2014. "The history of syntax." In *The Routledge handbook of syntax*, edited by Andrew Carnie, Yosuke Sato and Dan Siddiqi, 490-508. New York: Routledge.
Dąbrowska, Ewa. 2010. "Naive v. expert intuitions: An empirical study of acceptability judgments." *The linguistic review* 27 (1):1-23.
Dajko, Nathalie. 2009. "Ethnic and geographic variation in the French of the Lafourche Basin." PhD Thesis, Tulane University.
Daval-Markussen, Aymeric. 2013. "First steps towards a typological profile of creoles." *Acta Linguistica Hafniensia* 45 (2):274-295. doi: https://doi.org/10.1080/03740463.2014.880606.
Daval-Markussen, Aymeric, and Peter Bakker. 2017. "Creole typology II: Typological features of creoles: from early proposals to phylogenetic approaches and comparisons with non-creoles." In *Creole Studies – Phylogenetic Approaches*, edited by Finn Borchsenius Peter Bakker, Carsten Levisen and Eeva Sippola, 103-140. Amsterdam, Philadelphia: John Banjamins.
Dayal, Veneeta. 1994. "Scopemarking as indirect wh dependency." *Natural Language Semantics* 2:137-170.
de Villiers, Jill, and Thomas Roeper. 1996. "Questions after Stories: On Supplying Context and Elimiating It as a Variable." In *Methods for Assessing Children's Syntax*, edited by D. McDaniel, C. McKee and H. Smith Cairns, 163-187. Cambridge: MIT Press.
DeGraff, Michel. 2001. "On the origin of creoles: A Cartesian critique of neo-Darwinian linguistics." *Linguistic Typology* 5 (2/3):213-310.
DeGraff, Michel. 2003. "Against creole exceptionalism." *Language* 79 (2):391-410.
Delgado, Richard, and Jean Stefancic. 1998. "Critical race theory: Past, present, and future." *Current legal problems* 51 (1):467.
Déprez, Viviane, Kristen Syrett, and Shigeto Kawahara. 2013. "The interaction of syntax, prosody, and discourse in licensing French wh-in-situ questions." *Lingua* 124:4-19.
Dixon, Robert M. W. 1980. *The languages of Australia*, *Cambridge language surveys*. Cambridge, UK; New York: Cambridge University Press.
Dixon, Robert MW. 1991. "The endangered languages of Australia, Indonesia and Oceania." In *Endangered languages*, edited by Robert H. Robins and Eugenius M, 229-255. Oxford: Berg.
Dominguez, Virginia R. 1977. "Social Classification in Creole Louisiana." *American Ethnologist* 4 (4):589-602.
Dorian, Nancy C. 1977. "The problem of the semi-speaker in language death." *Linguistics* 15 (191):23-32.
Dorian, Nancy C. 1980. "Language shift in community and individual: The phenomenon of the laggard semi-speaker." *International Journal of the Sociology of Language* 25:85-94.

Dorian, Nancy C. 1998. "Western language ideologies and small-language prospects " In *Endangered Languages: Current Issues and Future Prospects*, edited by Lenore Grenoble and Lindsay Whaley, 3-21. Cambridge: Cambridge University Press.
Dorian, Nancy C. 1973. "Grammatical Change in a Dying Dialect." *Language* 49 (2):413-438.
Dorian, Nancy C. 1981. *Language death: the life cycle of a Scottish Gaelic dialect*. Philadelphia: University of Pennsylvania Press.
Dorian, Nancy C. 1986. "Abrupt Transmission Failure in Obsolescing Languages: How Sudden the 'Tip' to the Dominant Language in Communities and Families?" *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society February 15-17*:72-83.
Dressler, Wolfgang U. 1982. "Acceleration, retardation and reversal in language decay." In *Language spread. Studies in diffusion and social change*, edited by Robert L. Cooper, 321-336. Bloomington: Indiana University Press.
Du Plessis, Hans. 1977. "Wh Movement in Afrikaans." *Linguistic Inquiry* 8 (4):723-726.
Dubois, Sylvie, Emilie Gagnet Leumas, and Malcolm Richardson. 2018. *Speaking French in Louisiana, 1720-1955: Linguistic Practices of the Catholic Church*. Baton Rouge: LSU Press.
Dubois, Sylvie, and Megan Melançon. 2000. "Creole is, Creole ain't: Diachronic and synchronic attitudes toward Creole identity in southern Louisiana." *Language in Society* 29 (2):237-258.
Duranti, Alessandro. 1981. *The Samoan fono: A sociolinguistic study*: Department of Linguistics, Research School of Pacific Studies, The Australian National University.
Duranti, Alessandro. 1997. *Linguistic anthropology*, *Cambridge textbooks in linguistics*. Cambridge: Cambridge University Press.
Durrleman, Stephanie. 2008. *The syntax of Jamaican Creole: A cartographic perspective*. Vol. 127. Amsterdam, Philadelphia: John Benjamins Publishing.
Durrleman, Stephanie, and Ur Shlonsky. 2013. "Focus and wh in Jamaican Creole: Movement and Exhaustivenes." In *Beyond functional sequence*, edited by Ur Shlonsky, 91-106. Oxford [u.a.]: Oxford University Press.
Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2019. *Ethnologue: Languages of the World*. 22 ed. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.
Epstein, Samuel David, and T Daniel Seely. 2002. "Rule applications as cycles in a level-free syntax." *Derivation and explanation in the minimalist program*:65-89.
Fanselow, Gisbert, and Anoop Mahajan. 2000. "Towards a minimalist theory of wh-expletives, wh-copying, and successive cyclicity." In *Wh-scope marking*, edited by Uli Lutz, Gereon Müller and Arnim von Stechow, 195-230. Amsterdam: Benjamins.
Fattier, Dominique. 2013. Haitian Creole structure dataset. In *Atlas of Pidgin and Creole Language Structures Online*, edited by Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber. Leipzig: Max Planck Institute for Evolutionary Anthropology.
Felser, Claudia. 2004. "Wh-copying, phases, and successive cyclicity." *Lingua* 114 (5):543-574.
Ferguson, Charles. 2000. "Diglossia." In *The Bilingualism Reader*, edited by Li Wei, 65-80. London: Routledge.
Ferguson, Charles A. 1959. "Diglossia." *Word* 15 (2):325-340.
Fishman, Joshua A. 1967. "Bilingualism with and without diglossia; diglossia with and without bilingualism." *Journal of social issues* 23 (2):29-38.
Fishman, Joshua A. 1972. *The Sociology of Language*. Rowley, MA: Newbury.
Fishman, Joshua A. 1980. "Language maintenance." In *Harvard encyclopedia of American ethnic groups*, edited by S. Thernstrom, A. Orlov and O. Handlin, 629-38. Cambridge, MA.: Harvard University Press.

Fishman, Joshua A. 1991. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Vol. 76. Clevedon: Multilingual matters.
Fishman, Joshua A. 2000. "Bilingualism with and without diglossia; diglossia with and without bilingualism." In *The Bilingualism Reader* edited by Li Wei, 81-88. London/New York: Routledge.
Frawley, William John. 2003. *International encyclopedia of linguistics*. Oxford: Oxford University Press.
Freidin, Robert, and Howard Lasnik, eds. 2006. *Syntax: Critical concepts in linguistics*. London: Routledge.
Friedmann, Naama, and Rama Novogrodsky. 2011. "Which questions are most difficult to understand? The comprehension of Wh questions in three subtypes of SLI." *Lingua* 121 (3):367-382.
Gabel, Astrid, and Kathrin Brandt. 2017. "Particular Particles in two French-based Creoles." Eleventh Creolistics Workshop, Giessen, 23.-25. March 2017.
Gallagher, Charles Andrew. 2004. *Rethinking the color line: Readings in race and ethnicity*. New York: McGraw-Hill.
Gardt, Andreas (ed.). 2000. *Nation und Sprache. Die Diskussion ihres Verhältnisses in Geschichte und Gegenwart*. Berlin: De Gruyter.
Gegeo, David Welchman, and Karen Ann Watson-Gegeo. 1999. "Adult education, language change, and issues of identity and authenticity in Kwara'ae (Solomon Islands)." *Anthropology & education quarterly* 30 (1):22-36.
Graham, Mack. 1981. "Louisiana Creole survey methodology." *Notes on Linguistics* 20:18-21.
Grant, Anthony P. 1995. "Article agglutination in Creole French: a wider perspective." In *From contact to creole and beyond*, edited by Peter Bakker, 149-176. London, UK: University of Westminster Press.
Grenoble, Lenore. 2013. "The Arctic Indigenous Language Initiative: Assessment, Promotion, and Collaboration." *Proceedings of the 17th Foundation for Endangered Languages Conference*:80-87.
Grenoble, Lenore A, and Lindsay J Whaley. 1998a. "Toward a typology of language endangerment." In *Endangered languages. Language loss and community response*, edited by Lenore A Grenoble and Lindsay J Whaley, 22-54. Cambridge: Cambridge University Press.
Grenoble, Lenore A. 2011. "Language ecology and endangerment." In *The Cambridge Handbook of Endangered Languages*, edited by Peter K. Austin and Julia Sallabank, 27-44. New York: Cambridge University Press.
Grenoble, Lenore A., and Lindsay Whaley, eds. 1998b. *Endangered languages. Language loss and community response*. Cambridge [u.a.]: Cambridge University Press.
Grewendorf, Günther, and Joost Kremers. 2009. "Phases and cycles. Some problems with Phase Theory." *The Linguistic Review* 26 (4):385-430.
Grinevald, Colette, and Michel Bert. 2011. "Speakers and communities." In *The Cambridge Handbook of Endangered Languages*, edited by Peter Austin and Julia Sallabank. Cambridge: Cambridge University Press.
Hale, Ken. 1992. "Endangered languages: On endangered languages and the safeguarding of diversity." *Language* 68 (1):1-42.
Hale, Ken. 1998. "On endangered languages and the importance of linguistic diversity." In *Endangered Languages. Language loss and community response*, edited by Lenore A. Grenoble and Lindsay Whaley, 192-216. Cambridge: Cambridge University Press.
Hall, Gwendolyn Midlo. 1992. *Africans in Colonial Louisiana- The Development of Afro-Creole Culture in the Eighteenth Century*. Baton Rouge: Louisiana State University Press.
Hallett, Darcy, Michael J. Chandler, and Christopher E. Lalonde. 2007. "Aboriginal language knowledge and youth suicide." *Cognitive Development* 22 (3):392-399.

Harrison, J. A. 1882. "The Creole Patois of Louisiana." *The American Journal of Philology* 3 (11):285-296.
Haspelmath, Martin, and the APiCS Consortium. 2013a. Polar questions. In *The atlas of pidgin and creole language structures*, edited by Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber. Oxford: Oxford University Press.
Haspelmath, Martin, and the APiCS Consortium. 2013b. "Position of interrogative phrases in content questions." In *The atlas of pidgin and creole language structures*, edited by Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber. Oxford: Oxford University Press.
Haspelmath, Martin, and the APiCS Consortium. 2013c. Interrogative pronouns. In *The atlas of pidgin and creole language structures*, edited by Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber. Oxford: Oxford University Press.
Haugen, Einar, and Anwar S. Dil. 1972. *The ecology of language*. Stanford, Calif: Stanford University Press.
Heck, Fabian, and Gereon Müller. 2000. "Successive cyclicity, long-distance superiority, and local optimization." *Proceedings of WCCFL*.
Hellwig, Birgit. 2010. "Meaning and translation in linguistic fieldwork*." *Studies in Language* 34 (4):802–831.
Henri, Fabiola. 2016. "Conditioned reordering in Louisiana Creole: The case of sentential negation." SPCL Winter Meeting in conjunction with the LSA, Washington D.C., USA.
Henri, Fabiola, and Alain Kihm. 2015. "The morphology of TAM marking in creole languages: a comparative study." *Word Structure* 8 (2):248-282.
Himmelmann, Nikolaus P. 2006. "Language documentation: What is it and what is it good for." *Essentials of language documentation* 178:1-30.
Hinton, Leanne, Leena Huss, and Gerald Roche. 2018. *The Routledge handbook of language revitalization*. New York: Routledge.
Hirokuni, Masuda. 2006. "Microsyntax and macrodiscourse: "Da mistery" in Hawai'i Creole." In *Studies in Contact Linguistics: Essays in Honor of Glenn G. Gilbert*, edited by Linda L. Thornburg and Janet M. Fuller, 269-292. New York: Peter Lang Publishing Inc.
Höhle, Tilman N. 2000. "The w-… w-construction: Appositive or scope indicating?" In *Wh-Scope marking*, edited by Uli Lutz, Gereon Müller and Arnim von Stechow, 249-270. Amsterdam: Benjamins.
Holm, John A. 1989. *Pidgins and Creoles Vol. II - Reference Survey*. Cambridge: Cambridge University Press.
Holm, John, and Peter L Patrick. 2007. *Comparative creole syntax: Parallel outlines of 18 creole grammars*. London: Battlebridge Publications.
Hornstein, Norbert. 1999. "Movement and control." *Linguistic inquiry* 30 (1):69-96.
Hornstein, Norbert, Jairo Nunes, and Kleanthes K Grohmann. 2005. *Understanding minimalism*. Cambridge: Cambridge University Press.
House of Representatives Standing Committee on Aboriginal and Torres Strait Islander Affairs. 2012. *Our Land Our Languages*. Canberra: Commonwealth of Australia.
Huang, C.T. James. 1982a. "Logical relations in Chinese and the theory of grammar."PhD Thesis, Department of Linguistics and Philosphy MIT.
Huang, CT James. 1982b. "Move WH in a language without WH movement." *The linguistic review* 1 (4):369-416.
Hull, Alexander. 1979. "On the origin and chronology of the French-based creoles." In *Readings in Creole Studies*, edited by Edgar Polomé Ian Hancock, Morris Goodman and Bemd Heine, 201–215. Ghent: John Benjamins.
Hull, Alexander. 1993. "The Transmission of Creole Languages." In *Atlantic Meets Pacific: A Global View of Pidginization and Creolization; Elected Papers From the Society for Pidgin and Creole Linguistics*, edited by Francis Byrne and John A. Holm, 391-398. Amsterdam/ Philadelphia: John Benjamins.

Hurlbert, Stuart H., Richard A. Levine, and Jessica Utts. 2019. "Coup de Grâce for a Tough Old Bull: "Statistically Significant" Expires." *The American Statistician* 73 (sup1):352-357. doi: https://doi.org/ 10.1080/00031305.2018.1543616.

Jakubowicz, Celia, and Nelleke Strik. 2008. "Scope-marking strategies in the acquisition of long distance wh-questions in French and Dutch." *Language and Speech* 51 (1-2):101-132.

James, Bentley. 2016. "People of the Crocodile Islands & the Language of Spiritual Attachment." 15th Association of Australian Studies Conference "Nature and Environment in Australia", Cologne, 28.Sep. - 01.Oct.

Jocks, Christopher. 1998. "Living words and cartoon translations: Longhouse "texts" and the limitations of English." In *Endangered Languages. Language loss and community response*, edited by Lenore A. Grenoble and Lindsay Whaley, 217-233. Cambridge: Cambridge University Press.

Joseph, J.E. 2004. *Language and Identity: National, Ethnic, Religious*. Basingstoke: Palgrave Macmillan.

Kayne, Richard. 1976. "French relative 'que'." In *Current studies in Romance linguistics*, edited by F. Hensey and M. Luján, 255-299. Washington: Georgetown University Press.

Kayne, Richard S. 1984. *Connectedness and Binary Branching*. Vol. 16. Dordrecht, NL; Cinnaminson, USA: Foris Publication.

Kincade, Dale. 1991. "The decline of native languages in Canada." In *Endangered languages*, edited by Robert H. Robins and E.M. Uhlenbeck, 157-176. Oxford, New York: Berg.

Klein, Kathrin. 2012. "On the Edge: Movement in a phase based Theory of Syntax."unpublished Staatsarbeit, Englisches Seminar I, University of Cologne.

Klingler, Thomas A, and Nathalie Dajko. 2006. "Louisiana Creole at the periphery." *Creole Language Library* 28:11.

Klingler, Thomas A, Michael D Picone, and Albert Valdman. 1997. "The Lexicon of Louisiana French." In *French and Creole in Louisiana*, 145-181. New York: Plenum Press.

Klingler, Thomas A. 1992. "A descriptive study of the Creole speech of Pointe Coupee Parish, Louisiana with focus on the lexicon."PhD Thesis, Indiana University.

Klingler, Thomas A. 1996. "Comment écrire le créole louisianais?" *Plurilinguismes* (11):179-203.

Klingler, Thomas A. 1997. "Colonial Society and the Development of LC a." In *Language variety in the South revisited*, edited by Cynthia Bernstein, Thomas E. Nunnally and Robin Sabino, 140-151. Tuscaloosa: University of Alabama Press.

Klingler, Thomas A. 2000. "Louisiana creole: The multiple-geneses hypothesis reconsidered." *Journal of Pidgin and Creole Languages* 15 (1):1-35.

Klingler, Thomas A. 2003a. *If I could turn my tongue like that - The Creole language of Pointe Coupee Parish, Louisiana*. Baton Rouge: Louisiana State University Press.

Klingler, Thomas A. 2003b. "Language labels and language use among cajuns and creoles in Louisiana." *University of Pennsylvania Working Papers in Linguistics* 9 (2):77-90.

Klingler, Thomas A. 2006. "Louisiana Creole." In *Concise Encyclopedia of Languages of the World*, edited by K. Brown and S Ogilvie, 332-333. Oxford: OUP.

Klingler, Thomas A., and Ingrid Neumann-Holzschuh. 2013. "Louisiana Creole." In *The survey of pidgin and creole languages. Volume 2: Portuguese-based, Spanish-based, and French-based Languages*, edited by Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber. Oxford: Oxford University Press.

Kouwenberg, Silvia. 2010a. "Creole studies and linguistic typology: Part 1." *Journal of Pidgin and Creole Languages* 25 (1):173-186.

Kouwenberg, Silvia. 2010b. "Creole studies and linguistic typology: Part 2." *Journal of Pidgin and Creole Languages* 25 (2):359-380.

Kouwenberg, Silvia, and John Victor Singler. 2008. *The handbook of pidgin and creole studies*. Vol. 38. Malden, Mass: Wiley Blackwell.

Krämer, Philipp. 2013. "Creole exceptionalism in a historical perspective–from 19th century reflection to a self-conscious discipline." *Language Sciences* 38:99-109.
Krauss, Michael. 1992. "The World's Languages in Crisis." *Language: Journal of the Linguistic Society of America* 68 (1):4-10.
Krauss, Michael. 2007. "Classification and Terminology for Degrees of Language Endangerment." In *Language Diversity Endangered*, edited by Matthias Brenzinger, 1-8. Berlin, Germany: Mouton de Gruyter.
Labov, William. 2006 [1966]. *The social stratification of English in New York city*. Cambridge: Cambridge University Press. Original edition, 1966.
LaFleur, John, and Brian Costello. 2012. *Louisiana's French Creole Culinary & Linguistic Traditions*. Louisiana: Creole Bookstore Publishing.
Landry, Christophe. 2016. "A Creole Melting Pot: the Politics of Language, Race, and Identity in southwest Louisiana, 1918-45." Doctoral Thesis (PhD), University of Sussex.
Landry, Christophe, Cliford St. Laurent, Michael Gisclair, with contributions from, Oliver F. Mayeux, and Eric Gaither. 2016. *Kouri- Vini. A Guide to Louisiana Creole Orthography*. Louisiana: Louisiana Historic & Cultural Vistas.
Larmouth, Donald Wilford. 1974. "Differential interference in American Finnish cases." *Language*:356-366.
Lasnik, Howard. 2001. "Derivation and representation in modern transformational syntax." In *Handbook of syntactic theory*, edited by Mark Baltin and Chris Collins, 62-88. Malden, Oxford: Blackwell Publishers.
Lasnik, Howard. 2006. "Conceptions of the cycle." In *Wh-movement: Moving on*, edited by L. Cheng and N. Corver, 197-216. Cambridge, Mass.: MIT Press.
Lefebvre, Claire. 2001. "What you see is not always what you get: Apparent simplicity and hidden complexity in creole languages." *Linguistic Typology* 5 (2-3):186-213.
Legate, Julie Anne. 2011. "Warlpiri Wh-Scope Marking." *Syntax* 14 (2):97-121.
Lewis, M Paul, and Gary F Simons. 2010. "Assessing endangerment: expanding Fishman's GIDS." *Revue roumaine de linguistique* 55 (2):103-120.
Louisiana Department of Culture, Recreation and Tourism. 2019. "Louisiana Research." accessed 24. August 2019.
Lutz, Uli, Gereon Müller, and Arnim Von Stechow. 2000. *Wh-scope marking*. Vol. 37, *Linguistics Today*. Amsterdam: John Benjamins Publishing.
Manzini, M Rita, and Anna Roussou. 2000. "A minimalist theory of A-movement and control." *Lingua* 110 (6):409-447.
Marshall, Margaret M. 1997. "The Origin and Development of Louisiana Creole French." In *French and Creole in Louisiana*, edited by Albert Valdman, 333-349. Boston, MA: Springer US.
Mathieu, Éric. 2009. "Les questions en français: Micro et macro-variation." *Le Français d'Ici: Études Linguistiques et Sociolinguistiques de la Variation*:37-66.
Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
Maurer, Philippe, and the APiCS Consortium. 2013. "Focusing of the noun phrase." In *The atlas of pidgin and creole language structures*, edited by Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber. Oxford: Oxford University Press.
Mayeux, Oliver. 2018. Louisiana Creole Diachronic Corpus.
Mayeux, Oliver. 2019. "Rethinking decreolization: Language contact and change in Louisiana Creole." PhD Thesis, Theoretical and Applied Linguistics, Cambridge University.
McCann, Adam. 2019. "2019's States with the Best & Worst School Systems." https://wallethub.com/edu/states-with-the-best-schools/5335/, accessed 24. August 2019.

McCloskey, James. 2001. "The morphosyntax of wh-extraction in Irish." *Journal of Linguistics* 37 (01):67-100.
McDaniel, D. 1986. "Conditions on Wh-chains." PhD Dissertation, Linguistics, City University of New York.
McDaniel, Dana. 1989. "Partial and multiple wh-movement." *Natural Language & Linguistic Theory* 7 (4):565-604.
McWhorter, John. 2001. "The world's simplest grammars are creole grammars." *Linguistic typology* 5 (2):125-66.
McWhorter, John. 2002. "The rest of the story: Restoring pidginization to creole genesis theory." *Journal of Pidgin and Creole Languages* 17 (1):1-48.
McWhorter, John. 2012. "The nature of argument: Is the creole exceptionalism hypothesis dead?" *Journal of Pidgin and Creole Languages* 27 (2):377-387.
McWhorter, John H. 1998. "Identifying the creole prototype: Vindicating a typological class." *Language*:788-818.
McWhorter, John H. 2005. *Defining creole*. Oxford: Oxford University Press.
Melancon, Megan Elizabeth. 2000. "The Sociolinguistic Situation of Creoles in South Louisiana: Identity, Characteristics, Attitudes." PhD Thesis, LSU Historical Dissertations and Theses, Louisiana State University (7283).
Mesthrie, Rajend, Joan Swann, Ana Deumert, and William L Leap. 2000. *Introducing Sociolinguistics*. Edinburgh: Edinburgh Univ. Press.
Meyerhoff, Miriam. 2008. "Forging Pacific pidgin and creole syntax: Substrate, discourse, and inherent variability." In *The handbook of pidgin and creole studies*, edited by Silvia Kouwenberg, 48-73. Malden, Mass: Wiley-Blackwell.
Michaelis, Susanne Maria. 2015. "Inflectional complexity in creole languages: Evidence from the Atlas of Pidgin and Creole Language Structures." Summer meeting of the SPCL & ACBLPE, Graz, Austria, 07.-09. July 2015.
Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath, and Magnus Huber, eds. 2013. *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
Miller, George A, and Noam Chomsky. 1963. "Finitary models of language users." In *Handbook of Mathematical Psychology*, edited by Duncan Luce, Robert Bush and Eugene Galanter, 419-491. New York and London: John Wiley and Sons, Inc.
Mithun, Marianne. 1998. "The significance of diversity in language endangerment and preservation." In *Endangered Languages. Language Loss and Community Response*, edited by Lenore A. Grenoble and Lindsay Whaley, 163-191. Cambridge: Cambridge University Press.
Miyagawa, Shigeru. 2010. *Why agree? Why move? Unifying agreement-based and discourse-configurational languages*. Cambridge, Mass. [u.a.]: M. I. T. Press.
Moseley, Christopher. 2010. *Atlas of the World's Languages in Danger*. Paris: Unesco Publishing.
Mufwene, Salikoko. 2000. "Creolization is a social, not a structural process." In *Degrees of restructuring in creole languages*, edited by Ingrid Neumann-Holzschuh and Edgar W Schneider, 65-84. Amsterdam: John Benjamins.
Mühlhäusler, Peter. 1980. "Structural expansion and the process of creolization." *Theoretical orientations in creole studies*:19-55.
Mühlhäusler, Peter. 1980. "Structural Expansion and the Process of Creolization." In *Theoretical Orientations in Creole Studies*, edited by Albert Valdman and Arnold Highfield, 19-56. London: Academic Press.
Müller, Gereon. 2004. "Phrase impenetrability and wh-intervention." *Minimality effects in syntax* 70:289.
Murphy, Andrew. 2016. "What copying (doesn't) tell us about movement: Remarks on the derivation wh-copying in German." In *Replicative Processes in Grammar*, edited by

Katja Barnickel, Matías Guzmán Naranjo and Johannes Hein et al., 128-149. Leibzig: Universität Leibzig.
Muysken, PC, and Paul Law. 2001. "Creole studies. A theoretical linguist's field guide." *Glot International* 5:47-57.
Myers-Scotton, Carol, and Janice L Jake. 1995. "Matching lemmas in a bilingual language competence and production model: Evidence from intrasentential code switching." *Linguistics* 33 (5):981-1024.
Nettle, Daniel. 1999. *Linguistic Diversity*. Oxford: Oxford University Press.
Nettle, Daniel, and Suzanne Romaine. 2000. *Vanishing voices: the extinction of the world's languages*. Oxford: Oxford University Press.
Neumann, Ingrid. 1985. *Le créole de Breaux Bridge, Louisiane: étude morphosyntaxique, textes, vocabulaire*. Hamburg: H. Buske.
Neumann-Holzschuh, Ingrid. 1987. *Textes anciens en créole louisianais*. Vol. 8, *Kreolische Bibliothek*. Hamburg: Helmut Buske Verlag.
Neumann-Holzschuh, Ingrid. 2000. "Restructuration dans un creole "conservateur": Le cas du creole louisianais." In *Degrees of restructuring in creole languages*, edited by Ingrid Neumann-Holzschuh and Edgar Werner Schneider, 383-408. Amsterdam: John Benjamins Publishing.
Neumann-Holzschuh, Ingrid. 2009. "Contact-induced structural change in Acadian and Louisiana French: mechanisms and motivations." *Langage et société* (3):47-68.
Neumann-Holzschuh, Ingrid. 2015. "Entre la Caraïbe et l'Amérique du Nord: le créole louisianais et son lexique à la lumière de ses contacts linguistiques et culturels." New Orleans and the Global South. Caribbean, Creolization, Carnival. A Transarea Symposium, Cologne.
Nunes, Jairo. 2004. *Linearization of chains and sideward movement*. Vol. 43. Cambridge, Mass: MIT press.
Obenauer, Hans-Georg. 1994. "Aspects de la syntaxe A-barre: Effets d'intervention et mouvements des quantifieurs." PhD Thesis, Paris 8.
Ott, Dennis. 2009. "The conceptual necessity of phases: Some remarks on the minimalist enterprise." In *Explorations of phase theory: Interpretation at the interfaces*, edited by K. Grohman, 253-275. Berlin: Walter De Gruyter.
Palosaari, Naomi, and Lyle Campbell. 2011. "Structural aspects of language endangerment." In *The Cambridge handbook of endangered languages*, edited by Peter Austin and Julia Sallabank, 100-119. Cambridge; New York: Cambridge University Press.
Parkvall, Mikael. 2008. "The simplicity of creoles in a cross-linguistic perspective." In *Language complexity. Typology, contact, change*, edited by Matti Miestamo, Kaius Sinnemäki and Fred Karlsson, 265-285. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
Pesetsky, David Michael. 1982. "Paths and categories." PhD Thesis, Massachusetts Institute of Technology.
Pesetsky, David Michael. 2000. *Phrasal movement and its kin*. Cambridge, Mass: MIT Press.
Piller, Ingrid. 2016. *Linguistic diversity and social justice: An introduction to applied sociolinguistics*. Oxford: Oxford University Press.
Polinsky, Maria, and Olga Kagan. 2007. "Heritage languages: In the 'wild' and in the classroom." *Language and Linguistics Compass* 1 (5):368-395.
Pollock, Jean-Yves. 1989. "Verb movement, universal grammar, and the structure of IP." *Linguistic inquiry* 20 (3):365-424.
Potowski, Kim Author. 2013. "Language maintenance and shift." In *The Oxford Handbook of Sociolinguistics*, edited by Robert Bayley, Richard Cameron and Ceil Lucas, 321-339. Oxford: Oxford University Press.
Radford, Andrew. 2004. *Minimalist syntax :exploring the structure of English*. Cambridge: Cambridge University Press.

Ravindranath, Maya. 2009. "Language Shi and the Speech Community: Sociolinguistic Change in a Garifuna Community in Belize." PhD Thesis, Publicly accessible Penn Dissertations.

Reintges, Chris H, Philip LeSourd, and Sandra Chung. 2006. "Movement, Wh-agreement, and Apparent Wh-insitu." In *WH-movement: Moving On*, edited by Lisa Lai-Shen Cheng and Norbert Corver, 165-194. MIT Press, Cambridge, MA.

Reis, Marga. 2000. "On the parenthetical features of German was... w constructions and how to account for them." In *Wh-scope marking*, edited by Uli Lutz and Gereon Müller, 359-407.

Ricento, Thomas. 2013. "Language policy, ideology, and attitudes in English-dominant countries." In *The Oxford Handbook of Sociolinguistics*, edited by Richard Cameron Robert Bayley, Ceil Lucas, 525-544. Oxford: Oxford University Press.

Richards, N. 1999. "Feature cyclicity and ordering of multiple specifiers." In *Working Minimalism*, edited by S. Epstein and N. Hornstein, 127-158. Cambridge: MIT Press.

Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, Mass.: MIT Press.

Rizzi, Luigi. 1997. "The fine structure of the left periphery." In *Elements of grammar. Handbook in Generative Syntax*, edited by Liliane Haegeman, 281-337. Dordrecht: Kluwer.

Rizzi, Luigi. 2001. "On the position "Int (errogative)" in the left periphery of the clause." In *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*, 287-296. Amsterdam; New York: Elsevier.

Roeper, Tom, and Jill de Villiers. 2011. "The acquisition path for wh-questions." In *Handbook of generative approaches to language acquisition*, edited by Jill de Villiers, 189-246. Dordrecht: Springer.

Romaine, Suzanne. 2000. *Language in society: An introduction to sociolinguistics*. Oxford: Oxford University Press.

Rooryck, J. 1994. "On two types of underspecification: towards a feature theory shared by syntax and phonology." *Probus* 6:207-233.

Ross, John Robert. 1967. "Constraints on variables in syntax." PhD Thesis, Massachusetts Instutute of Technology.

Rottet, Kevin J. 1992. "Functional categories and verb movement in Louisiana Creole." *Probus* 4 (3):261-290.

Rottet, Kevin J. 2006. "Interrogative pronouns in Louisiana Creole and the Multiple Genesis Hypothesis." In *History, society and variation in honor of Albert Valdman*, edited by J. Clancy Clements, 235-249. Amsterdam: John Benjamins.

Rottet, Kevin James. 2001. *Language shift in the coastal marshes of Louisiana*. New York: P. Lang

Rudin, Catherine. 1988. "On multiple questions and multiple wh fronting." *Natural Language & Linguistic Theory* 6 (4):445-501.

Sampson, Geoffrey, R. 2007. "Grammar without grammaticality." *Corpus Linguistics and Linguistic Theory* 3 (1):1-32. doi: https://dio.org/10.1515/CLLT.2007.001.

Sardinha, Katie. 2011. "Story-builder: Picture Cards for Language Activities." www.story-builder.ca June 2014.

Sasse, Hans-Jürgen. 1992a. "Language decay and contact-induced change: Similarities and differences." In *Language death: factual and theoretical explorations with special reference to East Africa*, edited by Matthias Brenzinger, 59-80. Berlin, New York: Mounton De Gruyter.

Sasse, Hans-Jürgen. 1992b. "Theory of language death." In *Language death: factual and theoretical explorations with special reference to East Africa*, edited by Matthias Brenzinger, 7-30. Berlin; New York: Mouton de Gruyter.

Schippers, Ankelien. 2010. "Partial wh-movement and wh-copying in Dutch: Evidence for an indirect dependency approach." Annual Meeting of the Berkeley Linguistics Society.

Schmidt, Annette. 1985. "Young people's Dyirbal. An example of language death from Australia." *Cambridge Studies in Linguistics London*:1-252.
Schütze, Carson T. 2016. *The empirical base of linguistics. Grammaticality judgments and linguistic methodology*. Berlin: Language Science Press.
Schütze, Carson T., and Jon Sprouse. 2013. "Judgment data." In *Research Methods in Linguistics*, edited by Robert J. Podesva and Devyani Sharma, 27-50. Cambridge: Cambridge University Press.
Shlonsky, Ur. 2012. "Notes on wh in situ in French." *Functional heads: The cartography of syntactic structures* 7:242-252.
Siegel, Jeff. 2010. "Decreolization: A Critical Review." In *IUWPL9: New Perspectives on Language Contact and Contact-Induced Change*, edited by Megan E. Solon Clancy Clements, Jason F. Siegel, B. Devan Steiner, 83-96. Bloomington: IULC Publications.
Simpson, Andrew. 2000. *Wh-movement and the theory of feature-checking*. Philadelphia: John Benjamins Publishing.
Slavkov, Nikolay. 2008. "Medial wh-words and inversion phenomena in complex questions: the case of Canadian French speakers acquiring L2 English." *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference (GASLA 2007)*:218-232.
Snow, Peter. 2000. "Caribbean creole/non-lexifier contact situations: A provisional survey." *Journal of Pidgin and Creole languages* 15 (2):339-343.
Speedy, Karin. 1995. "Mississippi and Tèche Creole: Two separate starting points for Creole in Louisiana." *From contact to creole and beyond*:97-111.
Sprouse, Jon, Carson T. Schütze, and Diogo Almeida. 2013. "A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010." *Lingua* 134:219-248.
Stromswold, K. 1995. "The acquisition of subject and object wh-questions." *Language Acquisition* 4:5-48.
Swiggers, Pierre. 2007. "Two key concepts of language endangerment: language obsolescence and language death." *Linguistica* 47 (1):21-33.
Thieberger, Nicholas. 1990. "Language maintenance: why bother?" *Multilingua* 9 (4):333-358.
Thomason, Sarah Grey. 2001. *Language contact*. Edinburgh: Edinburgh University Press.
Thomason, Sarah, and T. Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
Thornton, R., and S. Crain. 1995. "Successful cyclic movement." In *Language Acquisition Studies in Generative Grammar*, edited by T. Hoekstra and B Schwartz, 215-252. Amsterdam and Philadelphia: John Benjamins.
Thornton, Rosalind. 1995. "Referentiality and wh-movement in child English: Juvenile D-Linkuency." *Language Acquisition* 4 (1-2):139-175.
Thornton, Rosalind Jean. 1991. "Adventures in long-distance moving: The acquisition of complex Wh-questions."
Tremblay, A. 2005. "On the use of grammaticality judgments in linguistic theory: Theoretical and methodological perspectives." *Second Language Studies* 24:129-167.
Trépanier, Cécyle. 1991. "The Cajunization of French Louisiana: Forging a Regional Identity." *The Geographical Journal* 157 (2):161-171.
Tsitsipis, Lukas D. 1998. *A linguistic anthropology of praxis and language shift: Arvanitika (Albanian) and Greek in contact*. Oxford: Clarendon Press.
Tsunoda, Tasaku. 2006. *Language endangerment and language revitalization: An introduction*. Berlin/ New York: Walter de Gruyter.
UNESCO, Ad Hoc Expert Group on Endangered Languages. 2003. *Language Vitality and Endangerment Document submitted to the International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages Paris, 10–12 March 2003*.

Valdman, Albert. 1997. *French and Creole in Louisiana*, *Topics in language and linguistics*. New York: Plenum Press.
Valdman, Albert. 1998. *Dictionary of Louisiana Creole*. Bloomington, Ind.: Indiana University Press.
Valdman, Albert (senior editor), and Kevin James Rottet. 2010. *Dictionary of Louisiana French. As spoken in Cajun, Creole, and American Indian Communities*. Edited by Assistant Editors:, Barry Jean Ancelet, Thomas A. Klingler, Amanda LaFleur, Tamara Lindner, Michael D. Picone and Dominique Ryon. Jackson: University Press of Mississippi.
Valdman, Albert, and Thomas A. Klingler. 1997. "The Structure of Louisiana Creole." In *French and Creole in Louisiana*, edited by Albert Valdman, 109-144. New York: Plenum Press.
Veenstra, Tonjes. 2008a. "Creole genesis: The impact of the language bioprogram hypothesis." In *The handbook of pidgin and creole studies*, edited by Silvia Kouwenberg, 219-241. Malden, Mass: Wiley-Blackwell.
Veenstra, Tonjes. 2008b. "Questions in Creoles: Implications for the Cartographic Approach." Formal Approaches to Creole Studies, CASTL, U Tromsø (N), 14.-15.11.2008.
Velupillai, Viveka. 2015. *Pidgins, Creoles and Mixed Languages: An Introduction*. Vol. 48. Amsterdam: John Benjamins Publishing Company.
Villata, Sandra, Luigi Rizzi, and Julie Franck. 2016. "Intervention effects and Relativized Minimality: New experimental evidence from graded judgments." *Lingua* 179:76-96.
Wardhaugh, Ronald. 2008. *An Introduction to Sociolinguistics*. Oxford: Blackwell Publishing.
Wasserstein, Ronald L., and Nicole A. Lazar. 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose." *The American Statistician* 70 (2):129-133. doi: 10.1080/00031305.2016.1154108.
Webb, Eric Russell. 2013. "Pidgins and Creoles." In *The Oxford handbook of sociolinguistics*, edited by Robert Bayley, 301-320. Oxford: Oxford University Press.
Weinreich, Uriel. 2010. *Languages in contact: Findings and problems*. New York: de Gruyter. Original edition, 1953.
Winford, Donald. 2018. "Creole Tense–Mood–Aspect Systems." *Annual Review of Linguistics* 4 (1):193-212. doi: https://doi.org/10.1146/annurev-linguistics-011516-034054.
Woodbury, Anthony C. 2003. "Defining documentary linguistics." *Language documentation and description* 1 (1):35-51.
Wurm, Stephen A. 1991. "Language death and disappearance: Causes and circumstances." In *Endangered languages*, edited by Robert H. Robins and E.M. Uhlenbeck. Oxford, New York: Berg.
Wurm, Stephen A. 2002. "Strategies for language maintenance and revival." In *Language Endangerment and Language Maintenance*, edited by D Bradley and M Bradley, 11-23. London: Routledge Curzon.
Zuckermann, Ghil'ad, Shiori Shakuto-Neoh, and Giovanni Matteo Quer. 2014. "Native Tongue Title: compensation for the loss of Aboriginal languages." *Australian Aboriginal Studies* (1):55.