Testing the Creole Language Bioprogram: A Phylogenetic Network Analysis

Rachel Rubin

s1486152

Leiden University

Abstract

Recent debates in the field of creole linguistics (creolistics) over the validity of a typological class of creole languages have inspired the adoption of new techniques to empirically test the most prominent theories of creole origins and creole typology. Phylogenetic computational tools, i.e. phylogenetic trees and networks, have been utilized in linguistics to model evolutionary scenarios and to predict genetic relationships between languages, and more recently in creolistics to identify typological and genetic relationships between pidgin and creole languages. Following several recent analyses seeking to validate the creole typological class and to test theories of creole genesis and classification using phylogenetic computation (Bakker et al. 2011; Daval-Markussen & Bakker 2012; etc), the present thesis continues this line of research by applying the same methods in order to test an older, frequently overlooked theory of creole genesis—Derek Bickerton's Language Bioprogram Hypothesis (LBH).

Bickerton's LBH attributes creole genesis to an innate human Bioprogram for language, invoked in situations where linguistic input is insufficient due to the influence of certain extralinguistic factors, resulting in a break in transmission of the lexifier language. This thesis seeks to test the accuracy of a set of LBH features in distinguishing between creoles and non-creoles in a phylogenetic network analysis, supplementing the results with a multiple regression analysis testing the correlation between the degree of creoleness and sociohistorical factors predicted by Bickerton's theory. Following the specifications of the LBH, the network analysis was expected to show a clear distinction between creoles and non-creoles, as well as patterns within the cluster of creoles associated to the presence of certain extralinguistic factors.

The output of the analysis indicates the validity of a creole typological class, yet the complete distribution of the languages in the networks cannot be entirely accounted for by the predictions of the LBH. While some of the clusters in the network can be attributed to prolonged contact with the superstrate language, which proved to be a significant factor in the regression analysis, other groupings are less predictable. The similar patterning of creoles and languages with low complexity scores according to Parkvall's (2008) metric in the phylogenetic network analysis indicates the compatibility of Bickerton's LBH and the proposal of a structurally less complex, synchronic class of creole languages. The variability in the patterning of creoles within the creole cluster in the phylogenetic network analysis is thought to be related to the unique

combination of extralinguistic factors influencing the development of each individual creole. The structural variation among the class of creole languages is expected to increase with time.

Keywords: pidgins; creoles; typology; creole genesis; Language Bioprogram Hypothesis; phylogenetic networks

Contents

Chapter 1: Introduction

Since the beginning of academic interest in creole languages, two themes have been at the forefront of the study: the origin of creole languages, and their classification; these themes continue to dominate the field and will be the focus of this thesis. In the present day, the debates over the origin and classification of these languages have manifested themselves in a heated debate over the existence of a creole typology, and therefore whether or not creoles can be considered a separate class of languages. This recent trend in researching creole typology has in a way reinvigorated the unresolved debates within the field, with the implementation of new techniques and information aiding in the continued development of decades-old theories. This thesis seeks to embrace these new techniques in continuation of research into creole origins and classification.

Creole typology research has reached a critical period where the leading theories are so diametrically opposed from one other that the field would benefit significantly from the confirmation, or invalidation, of those theories. Whereas many creolists support the concept of a creole typology and are actively conducting research to discover which features distinguish this class of languages from others, there are still a significant number of creolists who are against the idea that creole languages differ systematically from non-creoles and whose academic works reflect that sentiment (Chaudenson 2003; DeGraff 2003; Mufwene 2000). In order to continue with research in this field, and for that research to be both conclusive and impactful, it is important to determine whether or not the object of study—creole languages—indeed warrants an independent field of study. One of the aims of this thesis will be to address the validity of the class of creole languages as an independent linguistic entity.

Beginning with the earliest versions of the theories of creole genesis and classification, the history of the study of creoles as a distinct class of languages will be reviewed, following the development of those theories into their current forms. With a better understanding of how the prominent theories in the field came to be, the portrait of the current divide within the field is much clearer. Because the earlier development of theories of creole origin and classification relied largely on incomplete historical information and comparative analyses of creole languages, the field required the adoption of new technology and techniques for the advancement of theory. Thus, in order to empirically validate their claims regarding the genesis and classification of creoles, some researchers have turned to phylogenetic tools usually employed in the biological

sciences to determine genetic relationships, which have proven their usefulness to the field of linguistics (Barbançon et al. 2013; Nakhleh et al. 2005) and specifically, creole linguistics, hereafter creolistics (Bakker et al. 2011; Bakker & Daval-Markussen 2012; etc.). These tools have provided researchers of creoles the ability to empirically test theories that have been developing over the history of the study. The analysis conducted in this thesis will take advantage of such phylogenetic tools in an attempt to address both the question of how creoles came to be and the issue of the genetic relatedness of creoles to one another and to non-creoles by testing one of the prominent, yet frequently disregarded theories of creole genesis, Derek Bickerton's Language Bioprogram Hypothesis (1981;1984). The Language Bioprogram Hypothesis, hereafter LBH, is a theory of creole language genesis that is the result of the culmination of Universal theories of grammar in the second half of the 20th century. This theory, which has long been dismissed by the majority of the community, will be tested using the recently adopted phylogenetic tools to determine whether or not the theory is viable as an explanation for the development of creoles, as well as an answer to several important questions regarding the origin, classification, and typology of these languages.

Following methodology from previous phylogenetic analyses of creoles (Bakker et al. 2011; Bakker & Daval-Markussen 2012; etc.), three phylogenetic network analyses will be conducted using sets of features predicted by the LBH, determining whether or not those features are capable of clustering the languages in the sample together in the resulting phylogenetic networks; positive confirmation of this hypothesized clustering in the network would partly validate the predictions of the LBH. Following the phylogenetic network analysis, a multiple linear regression analysis will be conducted to determine whether or not sociohistorical circumstances are a significant factor in predicting the degree to which a language reflects the creole typology specified by the LBH, which can potentially provide an account for the observed variation within the proposed class of creole languages. The research aims of this analysis are, thus, to use phylogenetic computation to test the validity of the LBH as a theory of creole genesis, supplementing the phylogenetic network analysis with data from a subsequent regression analysis which will aid in determining whether or not the theory allows for the internal variation within the class of creole languages that has been so extensively observed in the data. This thesis seeks to apply the latest techniques applied in genetic and comparative

linguistic research to an older, somewhat overlooked theory of creole typology—scrutinizing Bickerton's creole prototype in the context of modern phylogenetic computation.

Because the LBH predicts that the extent to which a language will reflect the creole typology is related to extralinguistic factors (concerning sociohistorical circumstances), the phylogenetic network and regression analyses are expected to show that creole languages pattern together when factoring in the features proposed by the Bioprogram, and that the degree to which these languages reflect the Bioprogram is correlated to the presence of certain extralinguistic factors. Any variation observed in the patterning of creole languages in the networks is expected to be accounted for by the influence of said extralinguistic factors, as speculated by Bickerton. The importance of researching the origin and classification of creole languages reaches far beyond just the field of creolistics—research into creole genesis can have significant implications for the origin of all human language, and research into the classification of creoles can result in a better understanding of genetic relationships of languages and of language contact phenomena in general.

Chapter 2: Development of Theories of Genesis in Creolistics

To better understand the current state of the field and the theories that dominate it, it is best to review the history of said theories and the context of their development. This review of the relevant literature will begin with an overview of the origin and development of the various theories of creole genesis and classification, with an important look at the social and historical context that shaped them. The frequently overlapping theories concerning pidginization, creolization and the genetic relationships between creoles and non-creoles will be reviewed and discussed, along with the ways in which these theories have fallen in and out of favor in the field, often in a cyclical manner. It will be shown that the questions behind these theories have manifested in present debates within the field regarding the existence of a creole typology, which in turn will be tested in the present research. The intention is also to suggest that almost none of these theories are mutually exclusive, adding to the complex nature of the question of creole genesis and classification.

2.1 A Note on Sociohistorical Context

It is important to note that the context surrounding the study of creole languages, including social attitudes, economic and colonial interests, and general linguistic theory, has greatly influenced theories concerning the languages, their speakers, and their origins. The entire study of creole languages has been affected by social and historical context, always conforming to current social attitudes or specific trends at a period in the history of general linguistics, such as the increasing trend towards Universal theories of grammar in the mid 20th century. Because pidginization and creolization are as much sociohistorical phenomena as they are linguistic phenomena, social and historical context is as important to the study of these languages as the analysis of linguistic data. By examining the changes in this context throughout the history of the field of creolistics, it is possible to achieve a better, more complete understanding of the development of the theories within the field, past and present. While some theories deserve to be revisited, such as Bickerton's LBH in the present analysis, other theories that were shaped by negative social attitudes should remain in the past; the field should continue to focus on the linguistic data and the sociohistorical conditions that are associated with them.

Of great import to the current field, as well as to the current analysis, is the historically negative perception of creoles and their speakers, largely based on their race and indigenous

languages (DeGraff 2003:391). Attitudes towards language and race played a significant role in shaping the early studies and documentation of pidgin and creole languages. Concerning general attitudes towards language at the beginning of the study of pidgins and creoles, purity was the standard to which language was held. The perceived beauty and purity of languages such as Greek and Latin was widely upheld as the standard, and languages that deviated from the standard were considered bastardizations. Thus, pidgin and creole languages were mainly ignored as an object of study, often documented only for their role in trade, colonization, and religious missions. As Holm explains, "Pidgins and creoles were largely ignored by earlier linguists not only because of [the] misunderstanding of their identity, but also because of the prevailing notion of what language was and why it was worth studying" (1988:1).

Western attitudes towards race largely condemned non-whites, including the attitudes of most of those who began with the documentation of these languages; in general, creole speakers were perceived as intellectually inferior, and specifically linguistically inferior, to the white European. The social attitudes surrounding the early documentation of pidgin and creole languages strongly influenced the intentions in documenting these languages and in how they were perceived. Because pidgin and creole speakers were considered to be of an inferior race, the languages that they spoke, both natively and pidgins/creoles, were too deemed inferior and unworthy of scholarly inquiry. Regardless, these early accounts have provided the linguistic community with early documentation of pidgin and creole languages, which otherwise would likely have gone undocumented. As many theories within the field have proven to be somewhat cyclic, in that they often reoccur in cycles through the generations, so too have the intentions of the field of creolistics come again into question. Whether or not the foundation of the field is valid, that is, whether or not pidgins and creoles warrant an independent field of study based on a unique typology and a common sociohistorical context, has been a common topic of debate over the last decade, as will be discussed further in Chapter 5. Because this thesis joins a collection of recent works in creolistics seeking to validate the field and its object of study, it is therefore relevant to review the historical and social context that has brought the study to this point, and to question whether or not the study really has moved beyond its biased origins to focus on linguistic and sociohistorical evidence, as opposed to submitting to popular social attitudes.

2.2 In the Beginning: Creole Genesis

Because the common opinion of creoles and their speakers at the beginning of their study was generally that these languages are failed attempts by an inferior people at learning a European language, it follows that the earliest theories of creole genesis belonged to the Superstratist school of thought, which supposes that creoles are variants of their lexifier, i.e. the (often European) superstrate language; this contributes to the popular belief that such languages are not to be thought of as languages in their own right. These negative opinions of creoles stem from racist social attitudes towards these languages and their speakers. The early documenters of pidgin and creole languages mainly attributed the origin of these languages to the intentional simplification of the European language on the part of the Europeans, a theory which will later be referred to as the 'baby-talk theory' among creolists, and the inferior language abilities of the pidgin and creole speakers (DeGraff 2003:394). At this point in time, there was no significant theorizing regarding the different processes involved in pidginization and creolization and the roles of these processes in the development of creoles.

The emergence of scholarly interest in pidgin and creole languages was followed by attempts to more seriously theorize about their origins, which would eventually become one of the most notable debates in the field of creolistics. Addison Van Name, an American trained in philology, is credited with the first attempts at a scientific study of creole languages, beginning with his 'Contributions to Creole Grammar' (1869-70) which is a comparative study of creole languages (Holm 1988:24). Many of the observations and hypotheses that Van Name recorded regarding these languages were insightful for such an early stage in the scholarly study of pidgins and creoles and were oftentimes indicative of future debates and theories. Additionally, Van Name seemed to understand that a relationship exists between the processes of pidginization and creolization, well before the proposal of a theory of the lifecycle of pidgins by Hall (1962). Van Name also theorized about the composition and formation of creoles, postulating that creole languages are a reduction of the lexical source language as well as products of human innovation and restructuring, introducing new ideas about creole genesis (Holm 1988:26). Although considered mainly a Superstratist, Van Name indicated the significance of universal processes of language acquisition, as well as the sociohistorical forces behind creolization (Holm 1988:26). As it quickly became evident that Superstratism does not sufficiently account for a range of

features present in Creole languages, the theory was disregarded by many in favor of substrate influence or universal processes of language acquisition.

Theoretical concerns relating to pidgin and creole origins emerged as a central theme in pidgin and creole studies in the final decades of the 19th century. Ideas about creole genesis at this time centered around two theories at opposing ends of a spectrum, namely the Universalist and the Substratist schools, which to this day remain two of the most prominent positions on creole genesis in the field. The Universalist approach, attributed to the Portuguese philologist Adolfo Coelho, centers around the idea that features of creole languages are the result of universal language tendencies, specifically patterns of adult second language acquisition, allowing for absolutely no substrate influence. Conversely, the Substratist approach, attributed to French philologist Lucien Adam, proposes that the influence of substrate languages on the acquisition of European languages is responsible for creole genesis. These theories will continue to be present throughout the development of the field of creolistics. At the time, the study of language was "dominated by the comparative and historical paradigm. Among the opinions widely held by the scholarly establishment of the day were the following: […] the idea of linguistic evolution: highly inflected languages were often regarded as developmentally more advanced; that languages change from within, following natural laws: both language mixing and 'man-made' changes were regarded as marginal [..]" (Mühlhäusler 1997). Because of the focus on the comparative and historical paradigm in general linguistics at the time, there was yet to be a connection between theories in the broader study of language and the study of pidgins and creoles, especially considering that the historical model of language evolution did not apply well to these languages.

Likely the most important early creolist was Hugo Schuchardt, another contributor to the debate over creole origins, often referred to as the father of creole studies (Holm 1988:29). Schuchardt's position on the origin of creole languages fell on the spectrum somewhere between Coelho's extreme Universalist position and Adam's Substratist theory. Although he preferred to explain creole features using Substratism, he acknowledged that some aspects of creole languages are better explained through universal patterns of language acquisition. Schuchardt is considered by many modern creolists to have hinted towards ideas that would not be developed until much later in the field, namely the parallel development of creoles (monogenesis vs. polygenesis) and the creole continuum (Holm 1988:28), both of which will be discussed in the

following section. However, it should also be noted that Schuchardt's work has received some criticism, specifically regarding the lack of a consistent model resulting from his observations (Mühlhäusler 1997) and his common connection to 'baby-talk theory,' which hypothesizes that slave masters resorted to the use of simple language in order to communicate with intellectually inferior slave populations, resulting in the lack of a correct model of the European language to replicate (Holm 1988:33). Despite the strong negative connotations associated with 'baby-talk theory', recent accounts of creole genesis equating creoles to 'approximations' of European superstrates are reminiscent of this earlier assumption about creolization; these recent accounts will be reviewed in section 2.4 of Chapter 2 and additionally in Chapter 5. The basis that Schuchardt and his contemporaries created for the field of creolistics was instrumental in the development of the theories to be introduced in the following sections of this chapter.

2.3 A New Perspective on Genesis: Classification of Creoles

In the mid-1900s, there was a revival of interest in the study of pidgin and creole languages; during this period came significant developments in theories of creole origins. John Reinecke, an American creolist, is credited with substantial developments in the study of the sociolinguistic aspects of creole languages and the ways in which sociohistorical factors contributed to creole genesis. Reinecke approached creole languages from an alternative perspective, preferring to first classify creole languages according to sociohistorical factors, then proceeding to analyze linguistic features of creoles in order to identify any patterns of correlation between sociohistorical and linguistic features. His approach resulted in the identification of certain groups of creoles that shared common sociohistorical backgrounds: "Thus, he arrived at the following categories, each defined in detail and illustrated with a number of examples: (1) plantation creole dialects (e.g. those of the Caribbean area and the islands off West Africa), settlers' creole dialects (e.g. creole Portuguese in Guiné-Bissau and asia), and (3) trade jargons (i.e. pidgins)" (Holm 1988:40). Similarly, Bickerton also developed a sociohistorical classification for creoles, distinguishing between 'plantation creoles' and 'fort creoles'; his sociohistorical classification of creoles is later incorporated into his Universal theory of creole genesis, which provides a more comprehensive explanation of the relationship between sociohistorical circumstances and creolization. This change in direction towards sociolinguistic analysis is reflected in the greater field of linguistics, in which sociolinguistic theories began to

emerge in the early 20th century and became a significant part of Western linguistics in the 1960s; it is also indicative of the recognition of the innate connection between social factors and language, and more specifically, between sociohistorical factors and creolization.

The renewed interest in understanding pidgin and creole origins developed into an interest in the issue of the genetic classification of pidgin and creole languages, which has occupied a central position in creolist theory to the present day. The two most important concepts relating to the classification of pidgins and creoles that emerged from this period were introduced by Hall and Taylor. Whereas Hall proposed that creole languages be classified according to genetic relationships with their lexifier languages, Taylor believed that pidgins and creoles are best classified within a separate class of languages, and that the Stammbaum model of the genetic relatedness of languages is not applicable to pidgin and creole languages: "Taylor questioned the suitability of the concept of genetic relatedness when it came to pidgins and creoles, particularly since Meillet has specified that such relatedness implied continuity" (Holm 1988:44). The question of substrate influence was again brought to the forefront of the debate; however, instead of validating the theory that certain creole features can be traced back to substrate influences, it was a question of how relevant those influences are in the genetic classification of creoles. Hall's insistence on retaining the Stammbaum model and the classification of creole languages as genetic descendents of their European lexifier languages is representative of a recurring rejection in general linguistics of the validity of these languages as linguistic systems independent of their European lexical source languages.

Counter to Hall's belief that creole languages should be classified according to genetic relationships with their lexifier languages, Taylor posed several challenges to the idea of the genetic continuity of creoles, suggesting that observed changes in the genetic relationships of creole languages indicates a more complex image of creole classification than previously thought (Mühlhäusler 1997). This leads to the development of an important concept, namely relexification. Taylor's observation of the changing relationships between creoles and their European lexifiers over time provides evidence for the process of relexification, meaning that a new lexical source language is introduced and assumes the role of the previous lexifier language. An example of this often discussed in the literature is the proposed relexification of Papiamentu from a Portuguese lexifier to a Spanish lexifier (Holm 1988:312), resulting in an uncertainty over the lexical base of the creole. Relexification became popular among other creolists and remains

an often-cited theory which continues to contribute substantially to the understanding of creole origins to the present day. Another important theory introduced during this period was the theory of the lifecycle of creoles proposed by Hall (1962), which suggests that a creole is necessarily preceded by a pidgin or jargon phase, solidifying the speculation of several aforementioned early creolists such as Schuchardt—this theory became widely accepted in creolistics, but has been the source of significant debate in modern creolistics, to be discussed in the following section.

At the first international conference on creole studies, monogenesis theory was proposed by English linguist Robert Le Page (Holm 1988:46), which actually initiated the discussion that spawned the concept of relexification. The theory postulates that creole languages can be traced to a common origin, "the Portuguese-based pidgin that arose in the fifteenth century in Africa, perhaps from the Lingua Franca, and that was eventually relexified […]" (Holm 1988:46), with relexification responsible for the differences in lexical inventories between creoles. Returning to the aforementioned Papiamentu example, the earlier, Portuguese-based Papiamentu may have easily been identified as a descendent of the Portuguese-based proto-pidgin assumed under monogenesis, but was later relexified by Spanish. Although innovative and popular among creolists at its outset, monogenesis was later determined unlikely to account for the existence of all creoles, owing to the extreme unlikelihood that specific pidgin and creole languages are descendants of this 'proto-pidgin,' such as pidgins and creoles with African lexifiers "in whose genesis no European language was involved" (Holm 1988:47); this essentially invalidated the theory in its most extreme form. Doubts were expressed concerning the scope of the theory, specifically to what extent creole languages share a common ancestor, and the influence of substrate languages—the possibility of one 'proto-pidgin' as the origin of all creole languages was highly contested. It is more plausible that certain groupings of creoles with similar structures may be descendents of a common ancestor, perhaps in the case of the French-based creoles or the Atlantic creoles, a theory which many creolists presently hold to be true. According to Bakker (2002:72), "[Monogenesis] is a theory that now belongs to the historiography of creolistics, but it should otherwise now be disregarded, as it has been falsified in at least two different ways." Polygenesis, or the parallel development of creoles, is in effect any theory of the origin of creole languages that involves more than one originating pidgin or creole.

One last theory to introduce before considering the current theoretical state of the field is

the creole continuum, which is a theory that developed in response to inquiries regarding the connection between creole languages and post-creole varieties. The concept centers on the idea that "in situations in which a creole coexists with its lexical source language and there is social motivation for creole speakers to acquire the standard, […] the speech of individuals takes on features of the latter—or avoids features of the former—to varying degrees" (Holm 1988:52). Therefore, a continuum would exist, where at one end are the acrolectal varieties of the creole which are closest to the lexifier language, at the other end are the basilectal varieties of the creole which are the most conservative forms of the creole, and in between are the mesolectal varieties. It is said that a speaker occupies a range on the continuum, able to communicate in varieties more basilectal and more acrolectal than the speaker's native variety. One of the reasons that the concept of the creole continuum was rapidly adopted by many creolists is because it serves as a model for the common process of decreolization, by which creole speakers lose creole features and gain standard features resulting in a variety that more closely resembles the lexifier. These cases are important to consider, as varieties that have undergone decreolization are often wrongly used as evidence of the genetic continuity of creoles and their superstrates. The continuum was initially described using linguistic features to define the structure of the continuum, mainly in order to avoid the circularity of creating a continuum of social statuses defined by social factors, and was then ordered according to a hierarchy, from linguistic features associated with acrolectal varieties, to linguistics features associated with mesolectal and basilectal varieties. Interestingly, "after these lects were arranged by linguistic criteria only, their speakers were found to fall into a corresponding social hierarchy" (Holm 1988:57), providing evidence for the direct correlation between social and linguistic features of creoles. The creole continuum remains a useful tool in linguistic and sociolinguistic analyses of creole languages and societies, and will resurface in the discussion of this analysis in Chapter 5 in an attempt to account for variation in the results of much recent research in creolistics, including the results of the present analysis.

2.3.1 Problems with the Stammbaum Model

      Over the years, many have commented on the inadequacy of the Stammbaum model of the genetic relatedness of languages to account for the genetic affiliations of pidgin and creole languages. From the beginning of the discussion of the classification of pidgins and creoles, there has been a divide between those who support genetic continuity between the languages and their

lexifiers and those who reject it. The history of perceiving creoles through the perspective of their lexifiers encouraged many to support theories of genetic continuity, but others identified patterns of structural differences between creoles and their lexifiers which lead to the proposal of a break in transmission. The representation of language families as trees in the Stammbaum model has been contested by those who support the proposal of a break in transmission during the process of creolization. Thomason and Kaufman (1988:152) take both linguistic and sociohistorical data into consideration in their analysis of the transmission of European languages to plantation colonies. They found that "Given the paucity of definite, generally agreed-upon features of European-language grammar in the least decreolized creoles […], we see no real room for doubt that these languages resulted from a sharp break in transmission" (Thomason & Kaufman 1988:152). In their opinion, this sharp break in transmission excludes such languages from consideration for genetic classification. In effect, the proposed break in transmission would render any tree model of creole language evolution unrepresentative of creole origins and relationships.

2.4 Current State of Theories in Creolistics

In the current state of affairs, no one theory is generally considered to account entirely for the origin of pidgins and creoles; rather, most consider a combination of factors to be responsible for pidginization and creolization. Considerable evidence has been provided against each theory individually as a comprehensive theory of creole genesis, but combining certain aspects from the various theories has resulted in a deeper understanding of pidgin and creole origins. Increased emphasis on and improved access to historical information, in combination with new comparative methods in creolistics, has enabled advancements in research into creole genetic relationships.

Because convincing arguments have been made for a break in transmission occurring in at least some cases of creolization (Thomason & Kaufman 1988:151), the Superstratist position seems unlikely to accurately predict the occurrence of pidginization and creolization or the classification of pidgins and creoles. However, there are still several relatively recent publications calling for genetic continuity of creoles and lexifiers (Chaudenson 2003; DeGraff 2003), reminiscent of early Superstratist theories, yet with entirely contradictory motives. Approximation theory is one such account, proposing that "early plantation slaves spoke not

creoles, but close approximations of the lexifiers," owing to greater contact between slaves and European colonist populations than previously thought, and that "as this influx mounted, new slaves gradually came to be exposed less to whites' native variety of the lexifier than to slaves' approximations thereof, this becoming their primary model" (McWhorter 1998:789). Thus, under this account, creoles are distant approximations of their European lexifiers. Though historical records for certain instances of creolization may corroborate the conditions assumed under Approximation Theory, McWhorter (1998:800) has indicated that even in situations where there is evidence that slave populations did acquire a close approximation of a European language, there is also evidence of a creole language being spoken alongside the European variety. It is mainly in light of the aforementioned arguments that many current theories of creole genesis and classification avoid reverting to the lexifier languages for a classification model.

On the other hand, Substratism has manifested with each new generation of creolists and is supported, to varying extents, by most recent accounts of creole genesis (Holm 1988:43;65). As several previous analyses have shown (Lefebvre 2011), many features observed in creole languages can be identified in one or more of their respective substrate languages, indicating a potential source for the features in question. Furthermore, features shared by several creoles have been found to be present in their common substrates (Lefebvre 2011), providing additional support for the Substratist position. Even in accounts of creole genesis that do not rely entirely on substrate influence (Thomason & Kaufman 1988), the role of substrates in the formation of creoles is acknowledged and accounted for. However, Substratism as a comprehensive theory of creolization, meaning that the features in creoles not attributed to the lexifier language can be traced back to substrate influences, is not presently a common position in the field. A combination of substrate influence and universal patterns of adult second language acquisition— a position mildly popular among early forward-thinking creolists such as Schuchardt and Van Name—became quite popular as a theory of creole genesis. As Universal theories of grammar increased in popularity in the late 20th century, Universal theories of creole genesis also began to emerge. Bickerton's Language Bioprogram Hypothesis (1981;1984) quickly gained recognition for its extreme Universalist position and its strong implications for the origin of language in general. Because Bickerton's theory predicts a set of common features in creoles, and because he allows for no substrate influence, criticisms became increasingly frequent; these discussions will be reviewed further in Chapter 3. However, many others have taken more mild Universalist

positions (Parkvall 2008; McWhorter 1998;2001). In fact, most current accounts of creole genesis attribute creolization to a combination of universal processes of language acquisition, substrate influence, and sociohistorical circumstances.

Monogenesis, as discussed above in section 2.3, became quickly unpopular with the identification of strong evidence that the processes of pidginization and creolization occurred parallel to the proposed original instance; which supposedly lead to the diffusion and subsequent relexification of a Portuguese proto-pidgin. However, through the proposal of a single originating pidgin responsible for the development of all creole languages, many other theoretical advancements to the field were made. The concept of relexification, which was essential to the theory of monogenesis, proved to be extremely applicable to instances where colonies were taken over by colonial powers from a new linguistic background, resulting in a creole with remnants of multiple lexifier languages. Additionally, with the rejection of a common ancestor of all creole languages, there came a general consensus in the field that similar processes of pidginization and creolization must have been initiated simultaneously in various locations, indicating common sociohistorical circumstances necessary for these processes to occur. Theories proposing several groups of related creoles each originating from a shared ancestor, such as the Atlantic English-based creoles, became popular, further complicating the issue of pidgin and creole classification. Depending on one's position on the spectrum between monogenesis and polygenesis, creoles could be said to originate from a single ancestor and therefore constitute a single genetic family of languages, or creoles could each have developed individually, suggesting a complete absence of genetic relationships between these languages. The most widely shared belief in modern creolistics regarding the classification of creoles among themselves is relatively moderate, connecting several creoles through common ancestors, but holding firmly to the belief that there have been multiple, possibly innumerous, instances of pidginization and creolization throughout history.

Two trends in modern creolistics that are central to the present analysis are creole typology research and the Creole Exceptionalism debate, which will be reviewed individually in Chapters 4 and 5, respectively. While creole typology research seeks to validate the typological class of creole languages, identifying a common set of features that distinguishes creoles from non-creoles, the Creole Exceptionalism debate began with several creolists arguing against the existence of a creole typology. This strong opposition within the field has preoccupied much

current of the recent discourse. Additionally, most figures in the field have put increased emphasis on the importance of supplementing linguistic and sociolinguistic research in the field with historically accurate information. As pidginization and creolization are as much social and historical processes as they are linguistic processes, it is impossible to gain an understanding of the mechanisms behind these processes without applying accurate historical information to the research. Whereas those arguing for a separate class of creole languages use sociohistorical information to predict the circumstances surrounding pidginization and creolization, and to identify theories of genesis that best match the historical data, those who protest the existence of a typological class of creole languages apply sociohistorical information to their arguments seeking to dispute common theories of pidginization and creolization. McWhorter (1998:800) analyzed sociohistorical information to determine that certain commonly accepted predictions of the conditions present during pidginization and creolization were not borne out historically. Many common theories of pidginization and creolization predict that slave populations did not have access to the European superstrate and therefore resorted to developing contact varieties, yet McWhorter insists that this was often not the case. He argues that historical evidence shows that some slave populations may have had greater access to the superstrate than previously predicted, and that such slave populations may have been able to acquire a second language variety of the superstrate. Despite these historical findings, McWhorter still discovered evidence that pidginization and creolization occurred in these speech communities; such findings have led to the discovery that these linguistic processes are social processes as well, and that the presence of a creole identity may be a significant force behind pidginization and creolization. However, other figures in the field have used the same kind of demographic information presented by McWhorter to suggest that there is no evidence of a historical or linguistic need for pidgins and creoles (Chaudenson 2003; DeGraff 2003), and on that basis reject the concept of a break in transmission and the resulting typological distinction between creoles and non-creoles. This analysis aims to address both the question of creole genesis and the discussion over the existence of a creole typology, because the confirmation of a set of creole features may indicate a common origin of these languages. The role of creole identity in pidginization and creolization will be considered in this analysis as another factor to consider in the debate over creole genesis, and can potentially aid in resolving discrepancies between theories of pidginization and creolization and the sociohistorical data. It is likely that a comprehensive theory of pidginization, creolization,

pidgin and creole classification, and creole identity is required to answer the major questions in the field.

Chapter 3: Language Bioprogram Hypothesis

Universal theories of creole genesis grew in popularity in the late 1970s (Bickerton 1976:176), following the trend towards theories of universal grammar in general linguistics. Perhaps the most comprehensive Universal theory of creole genesis, and the main focus of the present study, is Derek Bickerton's Language Bioprogram Hypothesis (LBH), which can be considered an extreme manifestation of the Universalist position introduced in the previous Chapter. The LBH clearly has roots in Chomsky's generative grammar, but differs from Chomsky's UG (Universal Grammar) in that the LBH does not predict that universals will necessarily be shared by all languages (Bickerton 1981:298), but rather that a language will reflect the proposed universals more in relation to the quality of input in child language acquisition, which is in turn influenced by sociohistorical conditions. Additionally, Bickerton explains that his hypothesis suggests that "the infrastructure of language is specified at least as narrowly as Chomsky has claimed" (1984:173), indicating that, while the LBH predicts a very specific set of biological properties of language, the theory does not contradict the assertions of generative grammar. This agreement between the two theories is demonstrated in Figure 1 below, altered from Bickerton (1981:298). It is important to note that, at least according to Bickerton himself, the LBH is not a theory of creole genesis, but a theory of the biological foundations of all human language (Bickerton1986;1981), which has significant implications for the success of the theory within the field of creolistics, as well as for the overall reach of the hypothesis. In theory, the only difference that he proposes to distinguish between creoles and non-creoles is a sharp break in transmission, affecting the quality of the input for the next generation of language learners, invoking the Bioprogram. Although admittedly unconcerned with creole typology (1986:9), the common biological origins of creoles, and all languages, in combination with the shared sociohistorical conditions of creoles resulting in a break in transmission, contribute to Bickerton's support of creoles as a valid synchronic class of languages. However, his strong focus on developing a general theory of the origins of human language appears to result in a lack of complete understanding of pidgin and creole origins, which will be discussed further in section 3.1.
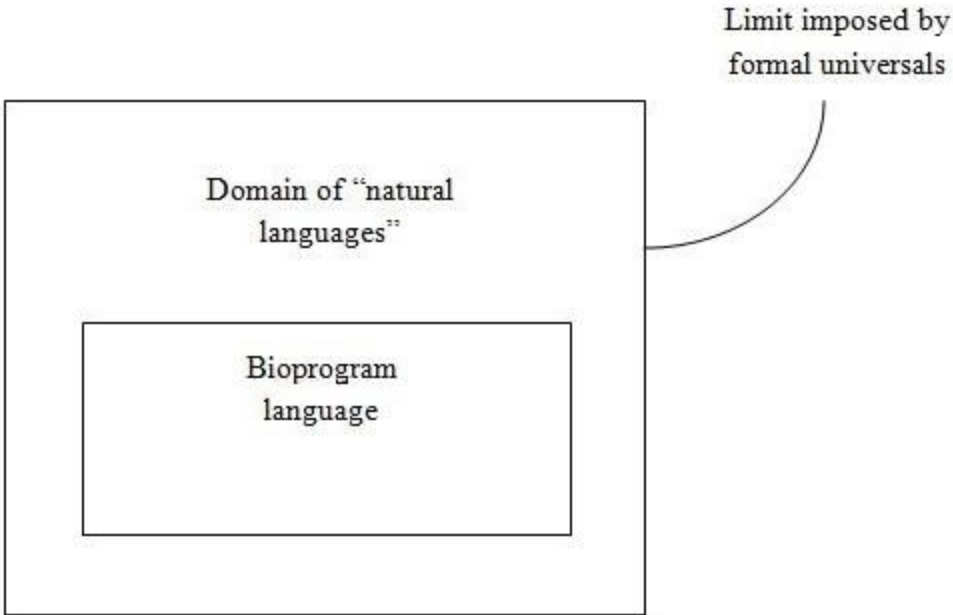
Figure 1: *Relationship of Bioprogram to formal universals. Altered from Bickerton (1981:298)*

Bickerton's LBH (1981;1984) posits that creole languages originate from the innate human biological program for language, invoked in children who receive pidgin languages from their parents and society as linguistic input and proceed to apply their innate language faculty (what Bickerton refers to as the Bioprogram) during the language acquisition process, rendering a structurally complete, functioning first language—the resulting creole. As mentioned above, similar concepts can be identified in Chomsky's UG theory, which also proposes certain universal properties of language that are a function of the human language capacity, indicating that certain linguistic structures are innate; however, the LBH is more specified than UG, and can therefore be more easily falsified.

Because "the LBH claims that the innovative aspects of creole grammar are inventions on the part of the first generation of children who have a pidgin as their linguistic input, rather than features transmitted from preexisting languages," (Bickerton 1984:173), the theory relies heavily on, and at the same time provides strong support for, the pidgin-creole lifecycle. Bickerton maintains the necessity of the pidgin-creole lifecycle by indicating systematic variation between pidgins and creoles, specifically concerning syntactic structure, which pidgins often lack. The LBH provides an explanation for the patterns of the development of similar syntactic structures in creoles from pidgins which lacked any such features. An appealing aspect of the LBH is that it provides an account of structural variation between creoles as well as structural similarities shared among creoles. Many theories over the course of the history of creolistics (discussed in

Chapter 2) have attempted to account for the numerous features shared by creoles from different superstrate and substrate backgrounds, originating in distant areas of the world, with often no apparent historical connections. Monogenesis is one such theory that attempted to account for the aforementioned similarities by positing a common genetic ancestor (Holm 1988:31;46), yet the LBH is supported by considerably more linguistic and historical data. Because the theory predicts that the quality of the input from the pidgin directly corresponds to the degree of intervention of the Bioprogram, a spectrum can be established for the degree of creoleness of these languages—languages with less input from the pidgin correspond to creoles exhibiting more Bioprogram features. The quality of the linguistic input provided by the pidgin is heavily influenced by the sociohistorical conditions under which the pidgin originated; despite this specification of the theory, Bickerton dismisses the idea that "there is [any] direct interaction between environmental factors of any kind and language itself" (1986:7), instead proposing that sociohistorical conditions affect the severity of the break in transmission, resulting in varying degrees of Bioprogram involvement, indirectly influencing the linguistic outcome. Pidginization, under the LBH, results from a sort of approximation of the European lexifier, which differs from Approximation theory discussed above in 2.4 in that Bickerton still maintains a break in transmission following the development of the pidgin, therefore distinguishing between the processes of pidginization and creolization: "Increasingly, throughout the growth period, newcomers must have acquired whatever knowledge of superstrate they could gather, not from native speakers. nor even from those who had learned from native speakers, but from those who had learned from non-native speakers" (Bickerton 1986:12). The sociohistorical involvement in the process of pidginization, meaning the affects of certain sociohistorical factors on the resulting pidgin, contributes to Bickerton's sociohistorical classification of creoles into three subgroups: plantation creoles, fort creoles, and maritime creoles. These distinctions are important to his theory because the sociohistorical conditions present in the development of plantation creoles are most likely to result in Bioprogram intervention, in his opinion. Bickerton has therefore proposed a social matrix of plantation creoles (1986:10), which indicates the combination of sociohistorical conditions most likely to produce more structurally impoverished pidgins, resulting in creoles that more closely resemble the Bioprogram. Table 1 below demonstrates variation among the three subgroups of creoles according to two sociohistorical conditions, indicating a sort of hierarchy of languages most likely to reflect the Bioprogram. In section 3.2, it

will be shown that the sociohistorical features proposed by Bickerton to result in a more structurally impoverished pidgin do seem to correspond to creoles exhibiting more Bioprogram features.

	Plantation Creoles	Fort Creoles	Maritime Creoles
Permanent Population Displacement	+	-	-
First Generation Nativization	+	+	-

Table 1: *Sociohistorical classification of creoles according to two sociohistorical factors.*

Bickerton has provided substantial linguistic evidence for his theory, which has strong implications for the nature of all languages, not only creoles. However, there have also been valid criticisms of his theory, specifically the following argument posed by Bakker (2002:86): "In short, the bioprogram features, whose usefulness I do not contest, are not biological. They do not surface when children, deaf or hearing, are forced to create a new linguistic system. The bioprogram features come to the surface when a number of people, of whatever age, but probably/preferably young, have to communicate by means of different home language systems." This criticism of the LBH resembles the issue of scope in monogenesis theory, where the ideas presented by the theory are not necessarily invalid, but the extent to which the theory can explain language origins may be somewhat overreaching. Additional criticisms of Bickerton's LBH, mostly concerning his strong focus on language origins and therefore his lack of attention to creoles and their speech communities, will be presented in the following section.

3.1 Biological Basis for Language Genesis

Because the LBH is a "unified theory of language acquisition, creole language origins, and general language origins" (Bickerton 1981:297), it has implications that reach far beyond the field of creolistics. It is central to the theory that the existence of the Bioprogram is necessitated by the mechanism of human language genesis, a consequence of the developmental path that human evolutionary traits have followed. Bickerton (1981:295) reviews the process of natural selection as he believes it applies to human language, originating in an earlier, related species as basic mental representations and lexicalizations, those with greater skill in the language faculty consistently selected for reproduction; this accounts for the continued development of human language. "Since language-skilled individuals possessed a higher potential for survival, they

would produce more offspring than other individuals, and the capacities that had arisen in them by random variation would be preserved and transmitted intact to their descendants" (Bickerton 1981:295). Because human traits that have developed via natural selection function in terms of very specific biological features and capacities, it is more likely that the human language capacity too developed in terms of concrete features as opposed to an abstract ability for language. However, this alone does not account for modern human language, which appears to have evolved far beyond what Bickerton claims to be the biological specifications of language continued through natural selection. If not, it would be expected that languages still largely reflect said biological specifications. Bickerton thus proposes that the cultural evolution of language, in contrast with the biological evolution of language presented above, occurs at a much faster rate and at a more abstract level, therefore remaining separate from the biological capacity for human language (or the Bioprogram). While the Bioprogram still represents the basis for natural human language, cultural developments and innovations have resulted in modern languages which no longer resemble their biological origins. Language now appears to be much more of a cultural phenomenon than a biological phenomenon in many ways, with very few commonalities between them all, i.e. language universals. However, the theory claims that severe conditions, such as the sharp break in transmission occurring during the process of creolization, can invoke the Bioprogram to help reconcile the impoverished input of the pidgin with the biological requirement for certain linguistic functions. Creoles, according to Bickerton's theory, may well present a rare insight into the origins of human language. Creoles are therefore not the main object of study of the LBH, but rather a coincidence in our history that provide the possibility of reviewing the history of language evolution, without which the development of Bickerton's theory would never have been possible.

One particularly controversial aspect of the LBH is that Bickerton considers creolization to be a purely linguistic process (1986:7), which has been strongly contested by others in the field (Thomason & Kaufman 1988). Although Bickerton adamantly denies the influence of environment on language, other research has shown that this is not the case. In fact, much of Bickerton's own work shows connections between sociohistorical conditions and pidginization, which in turn should influence the process of creolization. It is not plausible to consider the linguistic process of creolization outside of the context of the sociohistorical process of creolization, because the genesis of creole languages and the genesis of creole societies and

identities are simultaneous and deeply interconnected. Additionally, evidence from other works in the field (McWhorter 1998; Chaudenson 2003) indicates that many of Bickerton's assumptions about the sociohistorical conditions surrounding pidginization and creolization are misguided, specifically concerning population demographics. To begin with, the large discrepancy between superstrate and substrate populations that the LBH relies on has been found to be an inadequate account of population demographics in some creole communities. Therefore, it is likely that not all pidgins and creoles result from a large disparity between speech communities, indicating other forces driving pidginization and creolization. One such force is the development of a creole identity in creole speech communities, which could potentially be an important factor in creolization; this does not rule out Bioprogram intervention, but rather indicates additional motives for the need for creolization other than purely communicative need. Because the LBH aims at a greater understanding of language in general, it often overlooks the class of pidgin and creole languages as individual objects of study, resulting in a lack of understanding of certain processes at work during pidginization and creolization.

However, unlike many other theories of creole genesis, the LBH has the potential to accommodate new findings in recent research to become a more comprehensive theory of creole genesis, while still serving as a model for the origins of all human language. Although some of Bickerton's assumptions about the sociohistorical conditions during pidginization and creolization may not have been entirely accurate, there still appears to be a strong correlation between the sociohistorical conditions he proposes, and the presence of features predicted by the Bioprogram; therefore, the present analysis continues the attempt to validate this long neglected theory. The LBH is also able to account for the presence of substrate influence in creoles, likely resulting from heavy substrate influence during pidginization and mainly excluding structural transfer. Bickerton's LBH is a plausible theory of the linguistic process of creolization that could benefit from adaptation to account for the sociohistorical process of creolization.

3.2 Specifications of the Bioprogram

As a theory of creole genesis, the LBH predicts sociohistorical factors that contribute to the conditions necessary for pidginization and creolization as well as linguistic structures that languages exhibiting Bioprogram intervention should posses. The sociohistorical conditions under which pidginization and creolization occur, or the social matrix of creoles described by

Bickerton, refer largely to population demographics which rendered communication difficult, similarity and number of substrate languages also affecting the quality of communication, and the length of the duration of the pidginization and creolization processes (Bickerton 1984; 1986; Holm 1988). The Pidginization Index (PI) is a metric developed by Bickerton (1984) to measure the degree of impoverishment of a pidgin, determined by a combination of sociohistorical factors relating to the access of substrate speakers to the superstrate language; this metric is used to determine which creoles would have required more input from the Bioprogram (adapted from Bickerton 1984:178):

$$PI = Y \: x \: \frac{P}{R}$$ where: Y = number of years between colonization and Event 1[1]
P = total number of substrate population at Event 1
R = yearly average of post-Event 1 immigrants

The number of years between colonization and Event 1 is expected to be significantly lower in impoverished pidgins, owing to the proposed connection between the ratios of superstrate to substrate speakers, the length of time during which the pidgin formed, and the degree of impoverishment of the creole; the greater the disparity between substrate and superstrate populations, the less input there would be for the subsequent pidgin. The speed at which the substrate population increases following Event 1 is expected to directly correspond to the level of dilution of the superstrate language, and therefore to the degree of impoverishment of the proceeding pidgin. The above formula should therefore equate to the degree of impoverishment of the pidgin and thus the degree to which the Bioprogram will intervene in the process of creolization. Similarly, Thomason & Kaufman (1988) discuss the influence of plantation demographics on the resulting pidgin and creole languages: "We would expect such population differences to affect the structures of the resulting creoles, and, though evidence for variation according to substrate populations is still scanty, we have good evidence of structural differences according to the proportion of TL [target language] speakers" (1988:155). Though Bickerton's Bioprogram proposes a specified relationship between sociohistorical circumstances and creolization outcomes unprecedented in creolistics, his sentiments are echoed by other figures in the field. Two additional sociohistorical factors are considered by Bickerton to be influential in creating the necessary conditions for pidginization and subsequent creolization,

---

[1] Event 1 refers to the point in time at which substrate and superstrate populations reached equal numbers.

namely, withdrawal of the superstrate and maroonage. Early withdrawal of the superstrate language, as in the case of the early withdrawal of English colonizers from Suriname[2], is thought to correlate to less contact with the superstrate language and therefore a more conservative creole. Moreover, maroonage, "the creation of communities of escaped slaves" (Bickerton 1984:178), is expected to produce the most conservative of creoles, due to the complete absence of superstrate influence after departure from the plantations, and therefore corresponding to creoles that exhibit the most Bioprogram features, such as Saramaccan Creole. These sociohistorical factors will be assessed for correlations to Bioprogram features using a multiple linear regression analysis in the analysis.

The linguistic features attributed to the Bioprogram are entirely structural in nature, as Bickerton believes that syntax is the only level at which substrate and other influences cannot penetrate (Holm 1988:145). One of the central features proposed to represent the creole typology is the lack of inflection exhibited by creole languages, which will be discussed in detail in the following section. Although not in itself a significant aspect of the LBH, lack of inflection is directly related to the manifestation of analytic TMA (tense-modality-aspect) marking systems in creole languages. These TMA systems show surprisingly little variation among geographically distant creoles from differing lexical bases (see Table 2, altered from Bickerton 1984; 2008). Another feature shared by many creoles is the tense reference of unmarked stative verbs in comparison with unmarked nonstative, or dynamic, verbs. A pattern has emerged in these creoles showing that unmarked stative verbs indicate nonpast reference, whereas unmarked nonstative verbs indicate past reference. Other structural features specified by the LBH include serial verb constructions (such as instrumental serialization), the use of separate clauses for causal constructions, non-inverted word order in wh-questions, and marking of specific indefinite reference by the numeral 'one' and nonspecific reference with a null marker (Bickerton 1984:185-6). A structural feature of the Bioprogram given significant attention by Bickerton (1984:180-2) is the multiple functions of the constituent *fu* in several creole languages; the prediction is that the languages that most reflect the Bioprogram will employ more functions of the constituent *fu* than creoles that developed from a richer pidgin.

Arguably one of the strongest arguments in favor of the LBH is the apparent validity of the proposed relationship between the degree of impoverishment of a pidgin, quantified in

---

[2] To be replaced by Dutch colonists.

sociohistorical factors, and the presence of Bioprogram features. Table 3 below demonstrates the correspondence between, on a basic level, sociohistorical conditions and linguistic features. The table shows the four functions of the constituent *fu* in creole languages, ranging from least frequent on the left to most frequent on the right; as predicted by the LBH, the languages exhibiting the least common functions of *fu* correspond to the creoles originating from the most impoverished pidgins, and therefore closest to the Bioprogram. This relationship between sociohistorical circumstances and linguistic features is one of the findings of Bickerton's work motivating the current research, as it is not only indicative of the shared typology of creole languages, but also of the potential common origins of creoles and all languages.

	Saramaccan Creole	Fa d'Ambu	Guyanese Creole	Hawaiian Pidgin (Creole)	Haitian Creole
-anterior	zero	zero	zero	zero	zero
+anterior	bi	bi	bin	bin	te
+irrealis	o	ske	go	go	av
-punctual	ta	xa	a	stay	ap
+anterior +irrealis	bi-o	bi-ske	bin go	-	t'av
+anterior -punctual	bi-ta	bi-xa	bin a	bin stay	t'ap
+irrealis -punctual	o-ta	ske-xa	go a	go stay	av-ap
+anterior +irrealis -punctual	bi-o-ta	bi-ske-xa	bin go a	bin go stay	t'av-ap

Table 2: *Strikingly similar TMA systems of several creoles from four lexical base languages and distant locations of origin.*

fu can be tensed	*fu* introduced tensed complements	*fu* is a modal	*fu* marks + certain complements
Saramaccan Creole	Saramaccan Creole	Saramaccan Creole	Saramaccan Creole
	Haitian Creole	Haitian Creole	Haitian Creole
	Providencia/San Andres Creole	Providencia/San Andres Creole	Providencia/San Andres Creole
	Sranan	Sranan	Sranan
		Guyanese Creole	Guyanese Creole
		Mauritian Creole	Mauritian Creole
		Jamaican Creole	Jamaican Creole
			Hawaiian Creole
			Papiamentu

Table 3: *The various functions of the constituent 'fu' in several creole languages.*

Chapter 4: Creole Typology

The most recent manifestation of the debate over the origins and classification of creole languages is found in creole typology research. Although pidgins and creoles have been studied independently from other languages since the foundation of the field of creolistics, it is now coming into question whether or not there exists a linguistically valid reason to study these languages as an independent, typological class of languages, such as for example the Romance languages. A typological class of languages is expected to share a set of features that can distinguish that class from languages that do not share the same set of features; traditionally, this is because of a shared origin or history, and typological classification is often said to be obligatorily connected to genetic classification. The discussion of creole typology is therefore quite controversial because the origins and genetic classification of creoles remain to the present day uncertain, and the internal variation among the group of creole languages is considerable. One's position on the origin and classification of creoles greatly informs one's stance on the creole typology debate, as the three concepts are highly interconnected. As Lefebvre (2011:4) explains, Superstratists are most likely to reject any proposal of a creole typology, as this would indicate discontinuity with the lexifier and the typological profile of the lexifier; Universalists, such as Bickerton, are much more inclined to support the existence of a creole typology, tied to their unique origins; and Substratists, including Lefebvre herself, have shown mixed opinions regarding creole typology. In order to argue that there is no typological profile of creole languages, but that creole languages exhibit typological features of their substrate languages, Lefebvre cites specific substratal origins for the features exhibited by creoles. She indicates that differences in features of creoles are due to differences in substrates between those creoles. However, others (McWhorter 1998; 2001; Parkvall 2008) argue that, although there has been undeniable substrate influence in the processes of pidginization and creolization, the simplification that occurred during these processes resulted in a synchronic typological class of creole languages distinguishable from other languages by a set of significantly less complex features. A crucial aspect of the present analysis is to vindicate the existence of a creole typological profile, which should correspond to the shared origins of these languages, by testing the claims of the LBH regarding the shared origins and common structural features of creoles.

A common characteristic attributed to the creole typology is the systematically less complex nature of the features present in these languages (McWhorter 1998; 2001; Parkvall

2008). Seeing as complexity, or simplicity, alone is not a feature, individual features of creoles are often compared to the same features in non-creole languages in order to determine which languages systematically use less complex methods of representing those features. The analyses conducted by McWhorter (1998; 2001) and Parkvall (2008), to be reviewed in the following sections, are representative of such comparisons of creole and non-creole typologies. Specific features frequently attributed to the creole typology include SVO word order, lack of inflection, a system of internally structured preverbal TMA markers, serial verb constructions, less marked phoneme inventories, and overall less marked structures, some of which imply the presence of the others. It is important to review previous analyses of creole typology because of the strong likelihood that Bioprogram features to a great extent overlap with features attributed by others to the creole typology.

4.1 McWhorter's Creole Prototype

McWhorter's collection of works on identifying and defining what he refers to as the creole prototype are some of the earliest attempts to propose a creole typology. McWhorter's first work on identifying the creole prototype (1998) focuses mainly on identifying a set of features that distinguishes creoles from all other languages, his prototype, which clearly corresponds to the concept of a typological class of languages; he also rather successfully argues against the Superstratist position and its stance on the creole typology. McWhorter (2001) continues his work on the prototype, developing a set of diagnostics of complexity and applying these diagnostics to creoles and non-creoles alike, hoping to identify a pattern of simplicity in the features of creoles. These works constitute a large part of the relevant literature on the creole typological profile.

4.1.1 McWhorter (1998)

After establishing that the Superstratist position rejects creoles as a typological class (1998:788), McWhorter begins his analysis by providing sociohistorical evidence counter to those claims, specifically to emphasize the apparent inaccuracy of sociohistorical information provided by Superstratist accounts to explain the impossibility of a break in transmission. He proceeds to outline his analysis, consisting of a sample of eight creole languages exemplifying what he calls the creole sociohistorical profile: "natively spoken languages that were created via rapid adoption as a lingua franca by slave populations five hundred years ago or less"

(1998:791). The hypothesis of his analysis is that creoles adhering to this creole sociohistorical profile should exhibit a shared set of features owing to the extreme simplification of the languages involved in the pidginization and creolization processes due to a sharp break in transmission. Although each feature of creole languages may be individually identifiable in non-creoles, a set of features capable of distinguishing creoles from non-creoles indicates a typological class. McWhorter proposes the following three features as constituting the creole prototype because they combine "low perceptual saliency with low import to basic communication, [encourage] learners acquiring the language rapidly and informally to bypass acquiring them" and "they only develop internally as the result of gradual development over long periods of time:" *minimal use of inflectional affixes*, *tone carries a low functional load*, and *limited derivation* (1998:792). The position taken in the article is that creoles constitute a synchronic class of languages, sharing a set of significantly less complex features owing to the sharp break in transmission that occurred during their genesis; the only difference between creoles and non-creoles is age, because non-creoles have had much longer to accrue the complexities that creoles lack. It can therefore be inferred that creoles, over time, will also accrue such complexities, eventually beginning to resemble languages that do not share the creole sociohistorical profile, after which they would be defined purely as a sociohistorical class (McWhorter 1998:799). McWhorter then discusses the motivation he believes to have driven the development of the three aforementioned features in creoles sharing the creole sociohistorical profile. Discarding inflections is a method of achieving optimal learnability and processability, whereas the opacity of tone carrying a functional load would render such a feature impractical in emergency acquisition processes; lastly, creoles have not been in existence long enough to have experienced the semantic drift required for semantically opaque derivation (1998:793-8). These motivations behind the structural outcomes of pidginization and creolization, according to McWhorter, account for the lack of inflection, low functional load of tone, and limited derivation observed in creole languages adhering to the creole sociohistorical profile. The arguments presented in his analysis also provide convincing evidence counter to the Superstratist position.

4.1.2 McWhorter (2001)

A continuation of his 1998 work, McWhorter (2001) continues his research into the creole prototype and furthers his aim of validating the synchronic typological class of creole

languages. His (2001) analysis, however, seeks to assess the complexity of creole features in comparison with non-creole features with the expectation that creole languages are systematically less complex than non-creoles. His assertion that creoles exhibit less complex features than non-creoles stems from the conclusion drawn in McWhorter (1998) that complexities in language accrue over time, a process that is currently taking place in the case of languages to which the creole sociohistorical profile applies. The three features found in McWhorter (1998) to constitute the creole prototype can thus be reconsidered through the perspective of language complexities, and in this context they provide support for the proposed lack of complex features in creole languages. McWhorter draws attention to the basic similarities between his theory of the lack of complexity in creoles and Bickerton's LBH, namely that both theories propose that creoles represent a sort of underlying layer of human language (2001:126). McWhorter importantly notes that the long-held belief that all languages are of equal complexity is baseless, arguing that because a language has developed complexities that serve legitimate functions in the language does not necessarily indicate that such complexities are necessary developments, especially if such complexities are absent in related varieties (2001:130). Whereas languages not adhering to the creole sociohistorical profile likely reach the limit of human propensity to process language and therefore level in terms of complexity, creoles are much younger and have yet to develop the same level of complexity; this is where creole languages differ from others typologically. The functional definition of complexity used in his analysis is that "an area of grammar is more complex than the same area in another grammar to the extent that it encompasses more overt distinctions and/or rules than another grammar" (2001:135). On that foundation, McWhorter developed four diagnostics of complexity with which to compare creole features to non-creole features: a less marked phoneme inventory, less syntactic rules, less overt and grammaticalized expressions for semantic and/or pragmatic distinctions, and less inflection (2001:136). While inflection itself is no more complex than free morphemes, it is often associated with more complicated morphophonological processes. Additionally, inflection can be considered more complex because of the complicating effects that inflection has on grammars over time and because some inflection, such as gender marking, does not correspond to necessary concepts in the language (2001:137-8).

The Northeast Caucasian language Tsez and Saramaccan Creole are compared in terms of the four diagnostics of complexity, with the finding that the grammar of Tsez is more complex in

all areas according to the metric, with the exception of determination marking; McWhorter attributes these differences in complexity to the fact that Saramaccan Creole has not existed long enough to share the same complexities observed in the grammar of Tsez (2001:141). Furthermore, the languages in the previous studies of the creole prototype (McWhorter 1998) contrast with Tsez in the same ways as Saramaccan Creole, even the French plantation creoles, despite some evident decreolization due to prolonged contact with the superstrate (2001:143). Comparison of a creole and a much older, analytic grammar shows that creoles are less complex even when inflection is not a relevant factor (2001:144). The overall finding of the analysis is that even where creoles do exhibit overt marking of grammatical structures, it is almost always less complex (McWhorter 2001:162). What is extremely interesting to mention in light of the present research is that McWhorter notes the existence of creoles whose sociohistories have lent them moderate inflection and various other elaborations absent in typical creoles (2001:162). This directly reflects Bickerton's assertion that sociohistorical conditions can influence the input for the processes of pidginization and creolization, resulting in variation in the manifestation of typically creole features, as well as the subsequent influence of sociohistorical factors such as prolonged contact with the superstrate. While McWhorter's works on the creole prototype (1998; 2001) have been crucial to the foundation of creole typology research, the future of this line of research lies in computational data analysis and should rely on these types of intuitive qualitative analyses as a suggestion for further research.

4.2 Parkvall's Simplicity Metric

McWhorter's (2001:162) conclusion that he would expect that on a spectrum of complexity, there would be a subset of creoles at the lower end, foreshadowed the research conducted by Parkvall on the systematic simplicity of creole grammars. Parkvall, like McWhorter (2001), considers the creole typological profile to consist of overall less complex constructions (2008:265). However, he too makes a point of arguing that structural simplicity does not indicate expressive simplicity, as there is evidence that languages can possess more or less complex structures than other languages but no evidence that this increases or reduces expressive capacity (2008:266). In fact, Parkvall goes on to suggest that structurally simpler languages can be seen as more efficient languages because they perform the same functions as complex languages with much less effort (2008:268). His analysis also repeats the sentiment that

synthesis, i.e. inflection, in itself is no more complex than analytical structures, but the resulting effects of synthetic forms add complexities to the grammar. The analysis factors in 47 features from the World Atlas of Language Structures (WALS) and eight supplementary features from additional sources, taking care to neglect features whose simplicity is often compensated with complexity in another area of the grammar. Parkvall's work, similar to McWhorter (2001) but unlike McWhorter (1998), measures the complexity of creoles in comparison with non-creoles according to a set of 55 features, but does not attempt to identify a specific set of features belonging to the creole typology. The analysis found that creoles do pattern differently from non-creoles in terms of complexity; statistical analysis showed that no other grouping of languages, not even artificial groupings, patterned as low on the complexity scale as creoles languages, regardless of sociohistorical background. The results provide direct counterevidence to any arguments against the typological similarity of creoles. Parkvall reiterates the fact that it is possible for languages to be analytic and complex, but that is not what the data show for creole languages. Furthermore, the fact that pidgins most closely resemble creoles typologically, and that expanded pidgins trend somewhere in between pidgins and creoles, provides additional support for the existence of a pidgin-creole lifecycle, which is vital to Universal theories of creole genesis such as the LBH (Parkvall 2008:281). The analysis also comes to the conclusion that complexity correlates with age, and creoles, owing to a break in transmission, are not old enough to exhibit such complexities. These results are compatible with McWhorter's (1998; 2001) findings as well as with Bickerton's LBH, which proposes that the presence of Bioprogram features is a direct effect of the break in transmission during creolization, and that languages evolve away from the Bioprogram, in accumulating more complex structures over time.

4.3 Comparative Creole Syntax

Comparative Creole Syntax (Holm & Patrick 2007) is a compilation of descriptions of structural features in 18 creole languages, with the aim of creating an unprecedented comparison of structural features among various creole languages of differing lexical bases. The comparison analyzes 97 features in 18 creoles from eight lexifier languages, assigning each language a binary value for each feature. Unlike the analyses discussed above that aim to provide support for a typological class of creole languages, Comparative Creole Syntax (CCS) is purely

descriptive in nature, recording extensive data on the structure of creole languages for the purpose of comparison, and to inspire further research. This kind of comparative descriptive work is crucial for the continuation of creole typological research, as large amounts of data across creole languages are required for successful quantitative data analysis. As preverbal TMA marking systems are central to the LBH, the first seven groupings of features in CCS, listed below in Table 4, will be particularly useful to the present analysis.

Unmarked Verbs	Anterior (or Past) Tense	Progressive Aspect	Habitual Aspect	Completive Aspect	Irrealis Mode	Other Combinations of Verbal Markers
statives with non-past reference	statives with past reference	indicating progressive	zero marker for habitual	completive only (before/after V)	future (= progressive marker)	irrealis plus progressive
statives with past reference	non-statives with (past-before-) past reference	indicating future	progressive marker for habitual	completive plus adjective	anterior plus irrealis = conditional	anterior plus irrealis plus progressive
non-statives with past reference	anterior (or past) = counterfactual	anterior plus progressive	marker for habitual only	anterior (or other preverbal markers) plus completive	anterior plus irrealis = future in the past	other auxiliary-like elements
non-statives with non-past reference	anterior (or past) with adjective	progressive with adjective = inchoative	anterior plus habitual		anterior plus irrealis = future perfect	
	anterior (or past) with locative					

Table 4: *CCA TMA features*

4.4 APiCS

The Atlas of Pidgin and Creole Language Structures (APiCS) is an online database that records features of pidgin and creole grammars, providing a mechanism for comparison across pidgin and creole languages, similar to WALS in function and design. Like CCS (2007), APiCS does not aim to make any assumptions about creole origins, classification, or typology, but rather serves to document all known features of creoles for use in other analyses, and is a specifically beneficial contribution to creole typology research. This database is part of a growing trend of synchronic documentation of language structures for descriptive and comparative purposes. A

key issue with relying on data from APiCS for research into creole typology and genesis is that synchronic data is not always representative of the origins of creoles, especially considering the proposal of a synchronic class of creole languages; and as discussed in the above analyses (McWhorter 1998; 2001; Parkvall 2008), it is expected that creole languages will develop more complexities with time. Additionally, prolonged contact between a creole and its superstrate language often triggers decreolization, resulting in a creole continuum, in which case the data not might be entirely representative of the language as a whole.

Chapter 5: On Creole Exceptionalism

Those who contest the validity of a typological class of creole languages (DeGraff 2003;2005; Chaudenson 2003; Mufwene 2000) argue that the belief that creoles are in some way different from other languages is "Creole Exceptionalism," and is based on linguistic exoticism rather than linguistic fact. These creolists refer to their position as "Uniformitariansm," implying that they intend to treat all languages in a uniform manner, as opposed to baselessly designating a group of languages as typologically distinct, or worse, less complex. Chaudenson (2003) and DeGraff (2003;2005) propose that creolization is a sociohistorical process, not a linguistic process. Sociohistorical conditions, such as colonialism and slavery, initiated the process of creolization which resulted in new varieties of European languages, according to the so-called 'Approximation Theory' of creolization. Chaudenson (2003) specifically emphasizes the need to focus on sociohistorical facts in order to better understand creole languages, and claims that popular theories of creole genesis are founded on inaccurate sociohistorical data (2003:124). There are, in his opinion, no differences between languages that undergo creolization and other language contact phenomena (2003:125); regional varieties of European lexifiers from 200-300 years ago are most likely the source of the proposed creole features (2003:126). Chaudenson's own theory of creolization closely resembles Bickerton's (1986) theory of pidginization, both referring to a "social matrix of creolization" occurring in a "plantation society" which results in a language variety that is an approximation of an approximation of the dominant European language, for Bickerton the pidgin and for Chaudenson the creole (2003:127-8).

DeGraff (2003) provides a slightly less sociohistorical and slightly more heated position on the Creole Exceptionalism debate. DeGraff (2003) reviews the field of creolistics in the context of the racist origins of the study of creoles and their speakers and insists that the distinction between creoles and non-creoles is founded in the belief that creoles cannot be equal to their European superstrate languages (2003:311). DeGraff thus shares Chaudenson's (2003) belief that creolization is a sociohistorical process "with no operational structural criteria" (2003: 391). Furthermore, he argues that creole genesis cannot be considered different than other language genesis, as the human language capacity remains the same as ever (2003:391). However, supporters of the creole typological class often associate the features of the creole typology with the relative youth of creole languages and the fact that certain complexities in language develop over long periods of time (McWhorter 2001; Parkvall 2008); in this

perspective, the human language capacity remains the same, but creoles constitute a synchronic class of languages, representing an earlier stage in language evolution than non-creoles. DeGraff goes on to propose that Creole Exceptionalism implies structural inadequacy in creoles which is associated with expressive inadequacy (2003:392). However, advocates of the creole typology stress the expressive equivalence of all languages, regardless of structural complexity (McWhorter 1998; 2001; Parkvall 2008). In fact, most of the field shares DeGraff's sentiments regarding the systematic marginalization of creoles in the early years of the field, including his condemnation of theories such as 'baby-talk theory'. Although he provides evidence that Haitian Creole exhibits some features more complex than their French equivalents, DeGraff is only proving the importance of massive data analysis by demonstrating that small-scale qualitative analysis is not always sufficient for large-scale comparison. Most importantly, as Parkvall (2008) concludes from his complexity rankings of a variety of the world's languages, it is creoles that stand out, not the languages of people of color in general (2008:282).

Bickerton (2004), weighing in on the debate, argues that the Uniformitarian position fails to account for well-documented similarities among creole languages (2004:831). In opposition, he restates his belief that exceptional circumstances result in exceptional consequences, referring to the severe sociohistorical conditions surrounding pidginization and creolization and the resulting break in transmission and emergence of creole languages (2004:831); creoles are therefore, in Bickerton's opinion, exceptional. Bickerton also notes the circularity of theories within the field, seeing that "[t]hirty years ago, it was widely held that to regard Creoles as mere dialects of their European lexifiers, rather than languages in their own right, was demeaning to creole speakers. Now, apparently, some see it as demeaning to Creole speakers if their languages are treated as novel and distinct entities, rather than as dialects of some European language" (2004:831). Indeed, Bakker shares a similar sentiment in his aptly titled article, "Creolistics: Back to Square One?" (Bakker 2014b), in which he argues against the Uniformitarian position and for a distinct class of creole languages. Rather than Creole Exceptionalism, Bakker prefers "Creole Distinctivism," reflecting his view that creoles are linguistically distinct from non-creoles, but not particularly exceptional (2014b:179). He insists that phylogenetic research verifies the distinctiveness of creoles based solely on linguistic data (2014b:181), which signifies that creoles are ordinary languages distinguishable from other languages by a set of features. McWhorter (2001) and Parkvall (2008) have suggested that creole languages share a set of

distinct features often relating to lower complexity scores, owing to a break in transmission, resulting in pidginization and creolization. Bakker also provides support for the pidgin-creole lifecycle, observing a pattern of highly synthetic substratum and superstrates structurally reducing to highly analytic pidgins which. in turn, develop into creoles with increasing complexity (2014b:185); from this it can be inferred that a break of transmission occurred followed by increasing development. He continues to reject the Uniformitarian "Feature Pool Idea," which proposes that features of creole languages are the result of a combination of features selected from a pool of superstrate and substrate features, by restating the fact that there are often features in creoles that cannot be traced to superstrate or substrate influence (2014b:183). In fact, phylogenetic research conducted by Daval-Markussen & Bakker (2012) supports Bakker's skepticism of the Feature Pool Idea (see section 6.2.1, Chapter 6). Importantly, Bakker highlights the weaknesses of qualitative analyses, frequently encountered in creole typology arguments, in comparison with the analysis of massive data (2014b:181), and indicates that the future of creolistics should rely more on empirical research and massive data analysis.

The present thesis is intended to contribute to the research validating the typological class of creole languages. Bickerton's LBH is a theory compatible with recent research in creolistics—it accounts for the patterns of structural development observed from the pidgin stage to the creole stage, the simplicity of creole features, and the deviation of certain creoles from the creole typological profile.

Chapter 6: Phylogenetic Computation

There has been a recent increase in the use of phylogenetic computation, a tool borrowed from the biological sciences, to map the evolution of languages based on comparative analysis of sets of data points for each language (Nichols & Warnow 2008:760). Access to these computational tools has revolutionized studies of phylogeny and language evolution in linguistics, providing a method for quantitatively measuring relationships between languages. Huson & Bryant (2006:254) question the suitability of the application of phylogenetic trees to complex evolutionary scenarios, proposing instead that in cases of "hybridization, horizontal gene transfer, recombination, or gene duplication or loss," phylogenetic networks are more suited to account for the data. Their logic clearly stretches to the use of phylogenetic computational methods to model language evolution, as languages rarely evolve separately and independently from one another. The authors thus introduce the software SplitsTree4 (Huson & Bryant 2006), the phylogenetic computation software used in the present analysis, as a comprehensive framework for estimating phylogenetic trees and networks (2006:254), specifically taking lateral contact into account in their model. The three main types of phylogenetic network discussed in their article are: phylogenetic trees, split networks, and reticulate networks, the latter two deviating from the standard tree model of evolution. While a reticulate network "represents evolutionary histories in the presence of reticulate events such as hybridization, horizontal gene transfer, or recombination," and represent a modern way of perceiving evolutionary relationships and developments, split networks are "obtained as a combinatorial generalization of phylogenetic trees and [are] designed to represent incompatibilities within and between data sets" (Huson & Bryant 2006:254). Figure 2 below, originally from Huson & Bryant (2006:255), schematically represents the differences between the several types of phylogenetic networks.
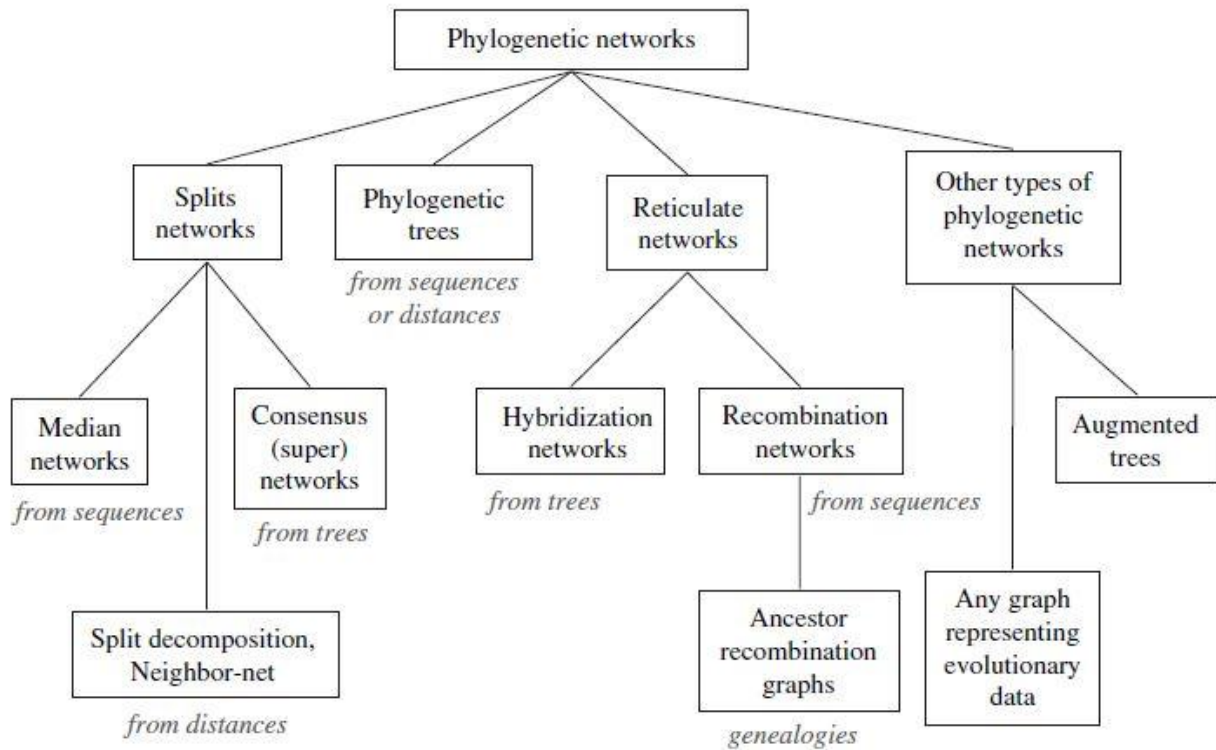
Figure 2: *Diagram of the various phylogenetic network models. Originally from Huson & Bryant (2006:255)*

Split networks provide 'implicit' representations of evolutionary relationships, as they do not require ancestral nodes and do not necessarily directly map such evolutionary relationships, as opposed to reticulate networks, which provide 'explicit' models of evolution, including the effects of lateral contact (Huson & Bryant 2006:255;263). Split networks therefore appear to be able to provide the best model of the creole typological class, as they do not explicitly map evolutionary relationships, which is quite complicated in the case of creoles, but rather measure distance between nodes based on incompatibilities in the data. In these split networks, each edge (that which corresponds to a branch in a phylogenetic tree) represents a split in the data, and the length of an edge is proportional to the weight of the split (Huson & Bryant 2006:256). Many of the themes discussed in Huson & Bryant (2006) relating to the necessity of new models of phylogenetic computation capable of accounting for lateral contact apply directly to the field of linguistics and the appropriateness of the Stammbaum model to account for language evolution.

6.1 In Linguistics

There are numerous researchers in linguistics adopting phylogenetic computational techniques to gain more insight into linguistic phylogeny. Historical comparative research often

takes advantage of these tools to predict evolutionary relationships, overlap, and divergences between certain languages based on sets of characters, often cognates. While Nichols & Warnow (2008:762) acknowledge that in certain instances, such as cases of pidginization, creolization, and other forms of intense language contact, phylogenetic networks are more suitable for representing the data, phylogenetic computation is mainly used in the form of phylogenetic trees within linguistic research in order to more accurately model linguistic phylogeny. Nichols & Warnow (2008) also stress the importance of the distinction between implicit phylogenetic networks, which represent graphically how the data do not fit into a tree model, and explicit phylogenetic networks, which model an explicit evolutionary scenario (2008:763); in the case of the majority of phylogenetic computation in the field of linguistics, explicit models are more commonplace, as the aim of most linguistic phylogenetic computation is to model specific language evolution scenarios. The research conducted by Nichols & Warnow (2008) is an important step in further incorporating phylogenetic tools into linguistic research, surveying the various methods of phylogenetic computation and types of linguistic data used as input to estimate linguistic phylogeny (2008:760). An effective method of testing the accuracy of phylogenetic methods discussed in Nichols & Warnow (2008:782) is generating models of already established genetic relations in order to determine whether or not the models reflect what is already well known about the data. Similarly, Barbançon et al. (2013) attempt a comparison of the various phylogenetic reconstruction methods employed by linguists by testing the ability of each method to accurately model data sets for which the trees are already well established (2013:146). Their research is entirely focused on phylogenetic trees modeling language evolution scenarios, and therefore explicit phylogenetic models. Barbançon et al. (2013:146) employed a parametric model of linguistic evolution developed in Warnow et al. (2006), as linguistic character evolution is not properly modeled by biological evolution models; the linguistic model allows for borrowing between lineages, but still cannot account for the development of contact varieties such as pidgins and creoles.

The input for phylogenetic computational models is in the form of data matrices, consisting of languages, characters, and states; characters take the form of any linguistic feature, states being the various forms that a character can take. Each language can thus be defined in terms of its states for each character (Barbançon et al. 2013:147). According to Barbançon et al. (2013:147), in the current state of the field, "linguistic characters are of three types: lexical,

phonological, and morphological." They go on to explain that not enough is known about syntactic change to justify the inclusion of syntactic characters in phylogenetic computation. However, in the case of phylogenetic computational models of creole typology, modeling genetic descent in the network is largely irrelevant and futile, therefore rendering the exclusion of syntactic characters unnecessary.

In fact, Donohue et al. (2011) argue against the use of typological features as input for split phylogenetic networks with the intention of modeling linguistic phylogeny. Although split networks are not often the primary phylogenetic computational method employed by linguists, Donohue et al. (2011) suggest that linguistic phylogenetic research using split networks and typological features as character input result in clusters representing human geography rather than linguistic phylogeny (2011:369). Because split networks model distance based on splits (or divergences) in the data between languages, and typological features of languages are the result of both phylogeny and areal diffusion, the analysis conducted in Donohue et al. (2011) found that where phylogeny did not account for the clusters in their network, geography did (2011:377). Split phylogenetic networks taking typological data as input are therefore more suitable for showing typological and contact relationships between languages, as opposed to phylogenetic relationships. Because it is not always possible to say with certainty which clusters in a split network are the result of phylogeny and which are the result of geographical diffusion, the authors propose that such models are not applicable for interpreting phylogeny. However, for the present analysis, their results are quite positive, reinforcing the validity of typological groupings predicted by split phylogenetic networks. The present analysis makes no claims about phylogeny based on clusters in the networks, but rather seeks to test the validity of the predictions of the LBH regarding the common origins of creole languages by indicating whether or not the languages included in the network cluster together based on the set of characters predicted by the LBH. If successful, meaning that the languages do cluster according to their states for the characters predicted by the LBH, it can be inferred that creole languages do share a creole typological profile specified by the LBH.

6.2 In Creolistics

The differences in the use of phylogenetic computation in linguistics and creolistics stem from the lack of cohesion between creole genesis and traditional ideas about linguistic phylogeny

and language evolution. The purpose of phylogenetic computation in linguistics is to reconstruct genetic affiliations and evolutionary scenarios, whereas creole research adopts phylogenetic techniques to analyze structural differences and similarities among creoles and between creoles and other languages. These differences in research goals are reflected in the different phylogenetic computational methods employed by historical linguists and creolists, the former preferring phylogenetic trees and explicit phylogenetic models and the latter turning to implicit phylogenetic models, specifically split phylogenetic networks.

6.2.1 Daval-Markussen & Bakker (2012)

Daval-Markussen & Bakker (2012) also use phylogenetic networks in order to support the theory of a creole typology; however, their approach involved creating phylogenetic networks combining different language samples and a set of 97 creole features adopted from CCS (2001) in order to identify different patterns of clusters. The authors list four major theories accounting for creole genesis, namely: 1) the Superstratist approach; 2) the Substratist approach; 3) the Feature Pool Hypothesis; and 4) the Universalist position (Daval-Markussen & Bakker, 2012:90). Four phylogenetic networks were generated to test the four theories of creole language development in order to determine which of the theories best accounts for the shared structural features of creoles.

The first phylogenetic network, testing the Superstratist approach, included 18 CCS creoles and seven lexifier languages, assuming that if the structural similarities were adopted from the lexifier languages, the creoles would be expected to cluster with their respective lexifiers. However, the study found that the seven lexifiers cluster together on one side of the network, separate from the creoles, indicating that "the superstrates have had a rather limited influence on the grammatical makeup of the incipient creoles at the time of restructuring" (Daval-Markussen & Bakker 2012:91). The second phylogenetic network, testing the Substratist position, included 18 CCS creoles and 19 substrate languages, under the assumption that if creole structural features can be attributed to substrate influence, then the creoles would be expected to cluster with their respective substrates. The network shows a clustering of creole languages apart from their substrates, while "[s]everal West African languages often mentioned in the context of creoles are found in the vicinity of the creole cluster and form a transition zone between creoles and non-creoles" (Daval-Markussen & Bakker, 2012:92). The third phylogenetic network,

testing the Feature Pool Hypothesis, uses 18 CCS creoles, three substrates, and one lexifier, assuming that under the Feature Pool Hypothesis, the creole Seychellois (for which two of the substrates and the lexifier language were selected) would cluster with its associated lexifier and substrate languages. The results show that, despite the fact that Seychellois was expected to cluster with the three languages that contributed to its formation (under the Feature Pool Hypothesis), the creoles once again clustered together away from the non-creoles. Finally, the fourth phylogenetic network generated for the study, testing the Universalist position, uses 18 CCS creoles, 19 substrates, seven lexifiers, and eight non-creoles. The results seem to indicate that the Universalist position is most strongly supported by the phylogenetic network: "[T]he software was able to detect a clear phylogenetic signal in only a few cases, which in itself is not surprising, since the features were originally selected as representative of the Atlantic creoles. Hence, the results lend support to the Universalist position" (Daval-Markussen & Bakker, 2012:93). This study has been able to show that, despite creating phylogenetic networks taking various samples and feature sets as input, creole languages consistently cluster together, against the predictions of three of the major theories of creole genesis. The relative success of the Universalist position in the phylogenetic networks will be especially relevant to the present analysis.

6.2.2 Bakker et al. (2011)

The focus of the research conducted by Bakker et al. (2011) was to identify a set of features capable of distinguishing creoles from non-creoles in a phylogenetic network, to determine a way of measuring the degree of creoleness, as well as to identify whether or not extralinguistic factors are connected to the degree of creoleness of the languages. The authors began their analysis by selecting the sample of 18 creoles from CCS (Holm & Patrick 2001) and the 97 features recorded in CCS in order to determine the degree of creoleness of each of the included languages; a multiple regression analysis was then conducted to determine if any extralinguistic factors (e.g. geographical location, age, etc.) contributed to the degree of creoleness. The results of this first test showed that there appears to be no correlation between the degree of creoleness and the specified extralinguistic factors. However, it is also possible that those 97 features from CCS included in their study do not accurately define the creole typology, and are therefore not representative of the degree of creoleness of the languages. To rectify this,

Bakker et al. then dropped down to 69 of the original 97 features based on the criterion that a feature had to be present in at least two thirds of the languages of the CCS sample. With these changes in place, the study found that: "[N]o matter whether we use all of the CCS features or only the subset of 'creole features in the narrower sense,' we find no statistically significant effects of age, area, lexifier or type on the number of features present in a given creole" (2011:32). On the one hand, these results suggest that creole languages share a set of similar structural features that cannot be attributed to extralinguistic factors relating to their development, while on the other hand, these results lend nothing to an explanation of the shared structural features of creoles.

Additionally, Bakker et al. generated phylogenetic networks in order to determine how structurally close the creole languages are to one another. Through this method, they found "that creoles are very similar to each other with regard to the number of creole features they instantiate. The statistical analysis thus adds fuel to the idea that creoles form a typologically rather coherent group of languages" (Bakker et al. 2011:32). Furthermore, the authors found that despite sharing a similar number of creole features, none of the languages (apart from the historically connected Guinea Bissau Creole and Cape Verdean) appeared to be particularly closely connected to any other creoles; this suggests that, although many of these languages share lexifier languages, substrate languages, location of origin, and sociohistorical backgrounds, the 18 creoles included in the sample are all equally structurally distinct from one another. This equality in structural distance between the languages will play a significant role in the discussion of the present analysis in Chapter 9. Although this study has contributed to the theory of a creole typology, the results have somewhat complicated the identification of an underlying cause of the shared structural features of creole languages.

6.2.3 Daval-Markussen (2013)

Daval-Markussen's analysis considers sets of creole features proposed in previous studies and, using those sets of features, conducts phylogenetic network analyses to determine which sets of features result in the closest clustering of creoles. The first phylogenetic network generated in the analysis uses Bakker et al.'s (2011) 69 binary 'creole features', the 18 CCS creoles, and 32 non-creoles. The results of the first phylogenetic network indicate that creoles do appear to share a similar typological profile (Daval-Markussen 2013:281). Daval-Markussen

then went on to generate a phylogenetic network using the features from CCS, but reduced those 97 features down to a mere 18, resulting, as was expected, in "a stronger phylogenetic signal, thus reinforcing the position of creoles as an independent group" (Daval-Markussen 2013:283). The author concludes that, because a limited set of features is able to distinguish creole languages from non-creoles, the support for a distinct typological class of creole languages is reinforced. The final phylogenetic network that Daval-Markussen produced in his analysis includes a sample of 19 creoles, 154 non-creoles, and factors in three multi-state features: "we decided to produce a final tree based on three multi-state features, the two pan-creole features identified above [*indefinite = 'one'* and *no tense-aspect inflection*] as well as Feature 55A in WALS, which deals with numeral classifiers (upon suggestion by John McWhorter, personal communication, 14 Nov. 2010)" (Daval-Markussen 2013:290). In this last phylogenetic network, all of the creoles clustered on one node; this result has significant implications for the existence of a creole typological class.

6.2.4 Bakker (2014a)

In defense of the contributions of phylogenetic computation to creole typology research, Bakker (2014a) responds to the criticisms that have presented about this recent trend in the field, largely pioneered by Bakker himself. In his response, Bakker claims that there is no human being capable of processing the amount of data required to accurately draw conclusions about the existence of a typological class of creole languages (2014a:347).One criticism worth noting, as it is a topic that has presented several times throughout the review of the relevant literature, is that quantitative data cannot replace qualitative data. However, Bakker argues that there "are no *a priori* reasons to doubt the superiority of either quantitative or qualitative data. However, more data tend to be better than fewer data" (2014a:438). In fact, the present analysis continues the trend of phylogenetic computation in creolistics because of the convincing results of previous studies employing the same techniques. Also relevant to the present analysis is the sixth criticism addressed by Bakker, specifically that some data from Holm & Patrick (2001), which has been used as a source of data in several of Bakker's previous analyses, is inaccurate. Although Bakker agrees that some values for certain features of creoles found in CCS may be debatable, he stresses the importance of such comparative works compiled by experts and native speakers for massive data analysis, as no one person has sufficient knowledge of all languages involved in

these studies (2014a:439). Another common critique of creole typology research in general is that creole languages are concluded to be simpler than non-creoles in many of these studies. While there has been convincing evidence provided in support of the systematic simplicity of creole features in other creole typology studies (McWhorter 1998; 2001; Parkvall 2008), phylogenetic computation in creole typology research does not draw any such conclusions; even non-creoles with similarly low complexity scores pattern differently from creoles in previous phylogenetic network analyses (2014a:441), indicating that simplicity is just one aspect of the creole typological profile. One criticism proposing a bias in favor of creoles with European lexifiers in phylogenetic analyses of creoles initiated an important discussion concerning proper sampling techniques in creole research. Because most creoles originated out of colonialism and slavery, many creole languages have European lexifier languages, resulting in necessarily disproportionate samples. However, this criticism inspired Bakker to test whether or not creoles with non-European lexifiers also pattern with creoles from previous phylogenetic research. Bakker developed three criteria from which he drew his sample of creoles with non-European lexifiers: "a lexicon that contains far from all of the roots of another language (lexifier, superstrate); a grammatical system that is only partly inherited from the lexifier (some substrate, some superstrate aspects); a grammatical system that is partly the result of traceable grammaticalisation of lexical items (innovation, reconstitution)" (Bakker 2014a:445). His hypothesis is that languages compatible with the three creole criteria, and with a non-European lexifier, will show evidence of the creole typological profile by patterning with the other creoles from previous research. The four features used as input in the phylogenetic network are: indefinite article derived from numeral 'one', no tense aspect inflection, negation expressed with negative particle, and possession expressed with verb meaning 'have' as opposed to with location or genitive (Bakker 2014a:450); these features, individually, are found in non-creoles around the world, but together have proven sufficient to distinguish creoles from non-creoles. The results of the study show that several creoles with non-European lexifiers pattern with other creoles, and away from non-creoles, when factoring in the four aforementioned features, indicating that the creole typological profile applies not only to creoles with European superstrates.

Chapter 7: Methodology

This section explores the different steps taken in the present research to test Bickerton's LBH and the predictions that it makes about creole languages. In the present study, three phylogenetic network analyses were conducted, along with one multiple regression analysis. The following sections will outline the setup, execution, and output of the several analyses, including the selection of language samples and feature sets, the organization of the data into matrices and later into nexus files, and the implementation of software to perform analyses on the data provided.

7.1 Language Sample

Choosing a language sample for creole language research can be particularly challenging, and the selection of a sample of creole languages should be justified in each study, in terms of the specific research goals of the study in question. Bakker (2014:443) presents an accurate overview of the complexities of sampling, and particularly of diversity samples in creolistics, which aim at collecting a sample that is equally representative of the language diversity in the world. Because creoles often arose under specific sociohistorical circumstances, namely European colonization, a diversity sample of creole languages would not be very representative of the linguistic diversity in the world. One aspect of sampling creoles made apparent in Bakker (2014:443) is the many aspects involved in the selection of these samples, namely lexifier languages, sociohistorical origin, geographical proximity, and combinations of lexifier and substrate languages, or 'bi-clans'. For the present analysis, an attempt was made to evenly distribute the sample across several lexical base languages, sociohistorical profiles, and geographical locations. In fact, because of the strong connection between sociohistorical conditions and linguistic features predicted by Bickerton's LBH, it is important to include creole languages that deviate from the creole sociohistorical and typological profiles. The idea is that, by including languages of differing origins, the typological distances between languages shown by the phylogenetic network analyses will be accounted for by sociohistorical factors shown in the multiple regression analysis. The main sample of the present research includes 16 languages from six lexical base languages. Additionally, conducting data analyses such as the present analysis requires languages for which a large amount of data is already recorded; therefore, all of

the creole languages included in the sample for the present analysis are documented in APiCS, which serves as a sufficient source of creole typological data. The creole languages and some of the justifications for the decision to include certain languages in the study are presented below.

Jamaican Creole: an English-based creole originating during the English colonization of the Caribbean islands. Due to prolonged contact with its English lexifier, Jamaican Creole exhibits evidence of the presence of a creole continuum, defined in Chapter 2. Significant substrate languages include Akan, Koongo, and Gbe (Farquharson 2013 in APiCS; Holm 1988).

Hawaiian Creole: an English-based creole developed on the Hawaiian Islands in plantations during American control of Hawaii. Despite the fact that the preceding pidgin varied significantly in relation to the various ethnic groups of its speakers (Bickerton 1984:174), the creole exhibits a high degree of uniformity. The major substrate influences include Hawaiian, Cantonese, and Portuguese (Viveka Velupillai 2013 in APiCS). The inclusion of Hawaiian Creole is quite interesting for testing Bickerton's LBH, as the islands are geographically quite distant from other creole-speaking locations and, despite having non-West African substrate influence, the language exhibits the same typically creole features often attributed to West African substrate languages.

Sranan: an English-based creole originating in the former English, then Dutch, colony of Suriname. Although originally spoken in plantations during the colonial era like many creoles, the early withdrawal of the English superstrate language following Dutch acquisition of Suriname may have resulted in a lack of decreolization and therefore potentially the retention of more conservative creole features. Additionally, much of the Surinamese plantation population was of South Asian descent, providing further substrate influence not present in the development of other Caribbean creoles (Winford & Plag 2013 in APiCS).

Saramaccan: an English- and Portuguese-based creole language, originating in a community of maroon slaves in the former English, then Dutch, colony of Suriname. Maroonage, according to Bickerton (1984:178), presents an even more severe case of early withdrawal of the superstrate language, resulting in more typically creole features and less typically (European) superstrate features (Aboh & Veenstra 2013 in APiCS).

Creolese: an English-based creole originating in the English plantation colony of Guyana. Like Jamaican Creole, Creolese also exhibits a creole continuum, with extreme variation between the varieties closest to and farthest from the lexifier language, owing to prolonged contact with the English superstrate. Significant substrate influences include Twi, Akan, Dutch, Hindi, and Arawak. Similar to the situation in Suriname, many laborers in Guyanese plantations were imported from South Asia, resulting in typologically diverse substrate influence (Devonish & Thompson 2013 in APiCS).

Berbice Creole Dutch: a Dutch-based creole originating in the privately-owned Dutch colony of Berbice in present-day Guyana. As opposed to the above mentioned creoles which originated in large plantation colonies, Berbice Creole Dutch originated in the small Dutch colony of Berbice, in what is now Guyana. Because the colony was ceded to the British around 150 years after its foundation, Berbice Creole Dutch can be considered among the group of creoles with early withdrawal of a superstrate language. However, because of the small nature of the Berbice plantation colony, there was likely more intimate contact between the labor population and the European language-speaking population (Kouwenberg 2013 in APiCS; Holm 1988:329).

Negerhollands: a Dutch-based creole originating in the former Danish West Indies, corresponding to the present-day U.S. Virgin Islands—St. Thomas, St. John, and St. Croix. Despite being under Danish rule, many of the plantations on the islands were Dutch-operated, and Dutch was therefore the operating language on many of the plantations in the Danish West Indies. The colony eventually came under U.S. control, and the English-based creole developed alongside Negerhollands eventually replaced the language altogether on the islands, and it eventually fell into extinction (Sluijs 2013 in APiCS; Holm 1988:325).

Haitian Creole: a French-based creole spoken in the French plantation colony of Haiti, on the island of Hispaniola. The Haitian Creole speech community in Haiti is the world's largest creole language-speaking community. Despite prolonged contact with French, which remains an official language and the standard language of Haiti, the majority of the population has little access to the language and thus only a minority of the population is bilingual in French and Haitian Creole. However, prolonged contact with the superstrate has resulted in decreolization and a

subsequent creole continuum. Despite sharing a French superstrate language with Seychellois, Réunion Creole, and Mauritian Creole, these three French-based creoles in the Indian Ocean appear to share many more features among themselves than they share with Haitian Creole; this should present an interesting dynamic in the phylogenetic network analysis. The main substrate influences come from the Kwa languages and the Bantu languages (Fattier 2013 in APiCS).

Seychellois: a French-based creole originating in the French plantation colony of the Seychelles, which is located off of the eastern coast of Africa, in the Indian Ocean. Because of the British takeover of the islands shortly after French colonization, early withdrawal of the superstrate also applies to Seychellois; this is likely why a very small percentage of the population is bilingual in French. However, close contact with the islands of Mauritius and Réunion may account for structural similarities between the three languages, owing to areal diffusion. The population demographics of the Seychelles under French rule are also compatible with the predictions of the LBH (Michaelis & Rosalie 2013 in APiCS; Holm 1988).

Mauritian Creole: a French-based creole originating in the Dutch, then French plantation colony of Mauritius. Mauritius, like Guyana and Suriname, had many laborers imported from South Asia, and therefore has various substrate influences with distant genetic relationships. Along with the Seychelles and Réunion, the British took over control of Mauritius not long after French colonization, resulting in a somewhat early withdrawal of the superstrate language. However, unlike the case of the Seychelles, French remained the operating language in the colony, resulting in maintained contact between the creole and the superstrate and continued bilingualism in Mauritian Creole and French (Baker & Kriegel 2013 in APiCS; Holm 1988).

Réunion Creole: a French-based creole originating in the French plantation colony of Réunion. Réunion Creole has likely had the closest relationship to French than any of the other French-based creoles; this is because of the fact that France regained control over Réunion after a period of British control of the island, which eventually transformed from a colony into an overseas department. Levels of bilingualism in French and Réunion Creole are therefore high, and it is to be expected that Réunion Creole shares many features with its lexifier. It is interesting to include the three French-based creoles in the Indian Ocean in the study because their social histories would predict different structures according to the LBH, yet their geographical proximity

indicates the likelihood of common structural features among the three languages. Holm, in fact, argues that Réunion Creole is likely a variety of the French superstrate, never completing the process of creolization. (Bollée 2013 in APiCS; Holm 1988:392).

Santome: a Portuguese-based creole originating on the island of São Tomé and Príncipe, which was used as a hub for the slave trade. The majority of the population speaks Portuguese, and smaller percentages speak one of several creole languages. The Portuguese-based creoles spoken on the west coast of Africa are thought to be the oldest creoles originating during the colonial period. Including these languages in the research could have significant implications for the proposal of a synchronic class of creole languages. Following predictions made by Bickerton and others (Parkvall 2008; McWhorter 2001), these creoles should be both older and structurally more complex due to their age. The major substrate influences are Edo and Kikongo (Hagemeijer 2013 in APiCS; Holm 1988).

Guinea-Bissau Creole: a Portuguese-based creole originating in the West African country of Guinea-Bissau. Because the Portuguese settlements on the coast of Guinea-Bissau were hotspots for the exportation of slaves during the height of the slave trade, Guinea-Bissau Creole is one of the oldest creole languages emerging out of colonialism and the slave trade. There has been considerable decreolization of Guinea-Bissau Creole due to prolonged contact with Portuguese, which remains the official language of Guinea-Bissau as well as the language of prestige (Intumbo, Inverno, & Holm 2013 in APiCS; Holm 1988).

Palenquero: a Spanish-based creole originating in a community of maroon slaves in Colombia. There is substantial disagreement in the field concerning the origins of Palenquero, and specifically over whether Spanish or Portuguese is the superstrate language of the creole. Because many support the idea that the language formed in a maroon slave community, it likely experienced severe early withdrawal of the superstrate language. However, the Palenquero-speaking community was eventually incorporated into the Spanish-speaking Colombia, and the speakers have therefore been bilingual in Palenquero and Spanish for a substantial amount of time (Schwegler 2013 in APiCS; Holm 1988:310).

Papiamentu: a Spanish- and/or Portuguese-based creole originating in the Dutch ABC Islands, i.e. Aruba, Bonaire, and Curaçao. The economy of the islands during the height of colonialism relied heavily on plantations and the slave trade, both obviously dependent upon large slave populations. There is some doubt about the lexifier language of Papiamentu, as it exhibits lexical items and other typological features traceable to both Spanish and Portuguese; however, many have suggested that the similarities between Papiamentu and other Portuguese creoles indicates Portuguese origins, and that the Spanish influence must have occurred later in development (Kouwenberg 2013 in APiCS; Holm 1988:312-13).

Nubi Creole Arabic: a Sudanese Arabic-based creole, with substrate influence from Egyptian Arabic, Swahili, and English. The language has had prolonged contact with and significant influence from Arabic, being spoken in a community constantly surrounded by superstrate and substrate influences. However, the Nubi community presents an interesting case of creole identity as a force driving the preservation of a creole language, which can also act as a force against decreolization, to be discussed further in Chapter 9 (Luffin 2013 in APiCS; Holm 1988:573).

 The second phylogenetic network analysis, computed using only CCS features, requires a sample with recorded data points for all of the relevant CCS features for each language. Therefore, the language sample for the second phylogenetic network excludes the languages from the sample of the first analysis that are not recorded in CCS. The languages included in the second sample are therefore: Berbice Creole Dutch, Guinea-Bissau Creole, Haitian Creole, Jamaican Creole, Nubi Creole, Negerhollands, Palenquero, Papiamentu, and Seychellois. The third and final phylogenetic network analysis compares the creole language sample from the first phylogenetic network with several non-creoles to determine whether or not the structural features predicted by Bickerton's LBH are capable of distinguishing creoles from non-creoles. The non-creoles included in the network are classified as one of the following:

> *Superstrates*: English, French, Spanish, Portuguese, Dutch, Egyptian Arabic[3]
> *Substrates*: Swahili, Fongbe, Arawak, Hindi, Ijo, Malagasy, Wolof
> *Low Complexity Score*: Pirahã, Vietnamese, Indonesian

---

[3] WALS did not have records on Sudanese Arabic, the lexifier language of Nubi Creole Arabic; Egyptian Arabic was chosen as a replacement because of the proposed structural similarities between the two varieties.

Three languages exhibiting extremely low complexity scores according to Parkvall's (2008) metric, namely Pirahã, Vietnamese, and Indonesian, have been included in the analysis to test whether or not creoles pattern similarly to non-creoles with similarly low complexity scores with LBH features as input; if these languages do pattern similarly in the network, it could indicate that the LBH features are compatible with Parkvall's simplicity theory of creole typology. Following the predictions of the LBH, the hypothesis is that given the linguistic features specified, creole languages conforming to a certain sociohistorical profile will pattern together in the phylogenetic network, apart from their superstrates, substrates, and even distinct in some way from non-creoles exhibiting similarly low complexity scores.

7.2 LBH Features

Though the overall goal was to choose an unbiased set of features representative of the specifications of Bickerton's LBH, each of the three phylogenetic network analyses compares a slightly different set of features, owing to less access to important data for some sets of languages. The first phylogenetic network used the following set of features, altered from their original format to accept only binary values, for easier processing later in the analysis:

APiCS #1: SVO word order
APiCS # 23: plural word/particle
APiCS # 29: indefinite article is identical to numeral 'one'
APiCS # 43: TMA markers in a leftward position in relation to the verb
APiCS # 47: marks progressive
APiCS #48: nonpunctual aspect marker for both habitual and progressive aspect
APiCS # 49: mixed temporal-aspectual system
APiCS # 51: stative verbs with present reference and dynamic verbs with past perfective
        reference are both unmarked (or identically marked)
APiCS # 73: copula present in predicative noun phrases
APiCS # 74: zero copula in predicative adjective constructions
APiCS #75: copula present in predicative locative phrases
APiCS # 84: directional serial verb constructions with 'come' and 'go'
APiCS # 100: negative particle
CCS # 8.3: 'For' is a (quasi-) modal
CCS # 8.4: 'For' introduces tensed clauses
CCS #13.1: equative copula in predicative noun phrases
CCS # 13.2: locative copula in predicative locative phrases
WALS # 69A: no tense-aspect inflection

These features correspond to the major linguistic specifications of Bickerton's LBH, specifically SVO word order (APiCS #1), analytical grammatical markers (APiCS #23,48,100), system of preverbal TMA markers (APiCS #43,47,48,49,), lack of tense-aspect inflection (WALS 69A), three-way copula distinction (APiCS #73,74,75;CCS # 13.1,13.2), distinction between interpretation of stative and dynamic verbs (APiCS #51), indefinite articles derived from numeral 'one' (APiCS #29), serial verb constructions (APiCS #84), and the various grammatical functions of 'For' (CCS #8.3,8.4). Although there were additional features in CCS compatible with the specifications of the LBH, values for these features were not available for several languages in the sample, and these features were therefore excluded from the first phylogenetic network analysis. However, the set of 18 features is sufficiently representative of the main predictions of the LBH.

Still wanting to somehow include the rest of the CCS features compatible with the LBH features, the second phylogenetic network analysis uses the set of all relevant CCS features and the subset of creole languages from the first analysis that are also recorded in CCS, as mentioned in 7.1. The set of CCS features used in the computation of the second phylogenetic network is the following, adapted to accept only binary values for ease of computation:

CCS #1.1: statives with non-past reference
CCS # 2.2: non-statives with (past-before-) past reference
CCS #3.1: indicating progressive
CCS #4.2: progressive marker for habitual
CCS #6.1: future (=progressive marker)
CCS #7.2 : anterior + irrealis + progressive
CCS # 8.3: 'For' as a (quasi-) modal
CCS # 8.4: 'For' introduces a tensed clause
CCS # 10.1: single negation (verbal)
CCS # 11.2: passive equivalent
CCS #12.1: preverbal markers before adjectives
CCS #13.1: equative copula with noun phrases
CCS #13.2: locative copula with predicative locative phrases
CCS #13.3: zero copula with predicative adjectives
CCS #14.1: directional with 'go'
CCS #14.2: directional with 'come'

Although several of these features overlap with the feature set from the first phylogenetic network, such as the distinction between interpretation of stative and dynamic verbs (CCS #1.1,2.2), the various grammatical functions of 'For' (CCS #8.3,8.4), and the three-way copula

distinction (CCS #13.1,13.2,13.3), this feature set also addresses some additional specifications of the LBH, including anterior + irrealis + progressive constructions (CCS #7.2), single negation (CCS #10.1), passive equivalents (CCS #11.2), and preverbal markers before adjectives, indicating the verb-like function of adjectives (CCS #12.1).

Gathering a feature set for the final phylogenetic network analysis was a bit more challenging, as no database has sufficient data for both creoles and non-creoles that is relevant to this analysis. However, many APiCS features list a WALS equivalent; it is therefore possible to compare languages recorded in both APiCS and WALS based on one set of features. The first step was to find APiCS features from the first phylogenetic network analysis that have WALS equivalents listed, after which the features were restated to allow for binary values applicable to both creoles and non-creoles. The resulting feature set is as follows:

WALS #21B: monoexponential TAM markers
WALS #33A: plural word/particle
WALS #38A: indefinite article identical to numeral 'one'
WALS #69A: no tense-aspect inflection
WALS #81A: SVO word order
WALS #112A: negative particle
WALS #119A: different encoding of nominal and locative predication

Although few in number, this feature set captures the features most central to the LBH: preverbal TMA marking system (WALS #21B), no tense-aspect inflection (WALS #69A), SVO word order (WALS #81A), analytical grammatical markers (WALS #33A, 112A), and a three-way copula distinction (WALS #119A). The feature set is representative of Bickerton's LBH, and though some of the non-creoles lacked data for several features, each language has data for at least four of the seven features. The three feature sets presented in this section will be used as input, representative of the specifications of the LBH, for each of the three phylogenetic network analyses. In the following section, values will be assigned to the features in each feature set for all languages in the corresponding language sample; these values will be organized into data matrices which will serve an important function in setting up the phylogenetic analyses.

While that covers the linguistic features included in the analysis, the regression analysis aims to identify patterns between linguistic and sociohistorical factors, which will require a set of sociohistorical factors to test. Because not all creole languages included in the analysis have

sufficient sociohistorical information on record, a small set of well-documented sociohistorical features was chosen:

Early withdrawal of superstrate
Prolonged contact with superstrate
Maroonage
Plantation creole

The first three features address the issue of contact between the superstrate and the creole, which is directly related to the access that creole speakers had to linguistic input from the lexifier language. Bickerton predicts that the less access available to the lexifier language, the more a language will rely on the Bioprogram for structural features; therefore, maroonage should be associated with the most conservative creoles and prolonged contact should indicate less intervention from the Bioprogram. The fourth feature, *plantation creole*, has to do with the environment in which the creole originated. The features are assigned a value of either 1 to indicate the presence of the feature or 0 to indicate the absence of the feature for each language in the sample.

7.3 Assigning Feature Values

In massive data analysis, it is important to rely on databases for recording data points, as no one person has enough knowledge to accurately report feature values for an entire sample. For the present analysis, data recorded in APiCS, CCS, and WALS were used to assign values to the three feature sets used in the phylogenetic network analyses. It is still important to review several data sources for comparative purposes if possible, as well as to review all of the information those sources provide. The main objective is, of course, to extract the correct information from the sources. For example, many languages in APiCS are given multiple values for one feature because, in very rare cases, other forms may be used in place of the most common form; in such cases, the more common form is taken as representative of the language for that particular feature. To prepare for the software analysis, data matrices were created in excel for each of the three phylogenetic network analyses, and one was created for the multiple regression analysis; these matrices serve as an efficient method of assigning feature values for each language in a format that can easily be transformed into input for the subsequent software analysis. The four matrices can be found appendix A. This section will review specific instances

where a decision had to be made concerning the assignment of certain feature values. The rest of the values can be assumed to have been accepted directly from the source, i.e. APiCS, CCS, or WALS.

The first feature that came into question was APiCS #23 concerning plural marking, which can take several forms in creole languages. Because Bickerton's LBH specifies analytical grammatical markers, such as negative and plural markers, the value of the feature can be assigned either 1 for the presence of analytical grammatical marking of plurality or 0 for the absence of such markers. However, in the data, it appears that while several main strategies of plural marking occur in creoles, reduplication as a plural marking strategy co-occurs with most of these strategies, and only five times does reduplication occur as a main plural marking strategy. For the current analysis, it is clear that where reduplication is used as a plural marking strategy by the languages in the sample, it is always under specific conditions and serves only as a complementary plural marking strategy. Therefore, only the main plural marking strategies were considered when assigning a value to APiCS #23.

APiCS #43, which has to do with the position of TMA markers in relation to the verb, can be assigned either 1 for TMA markers that occur to the left of the verb or 0 for TMA markers that occur anywhere else. The issue that arose with assigning values for the languages in the sample to this feature was that several of the languages in the sample exhibit a combination of TMA marking and tense-aspect inflection, technically resulting in TMA marking to the left and to the right of the verb. However, tense-aspect inflection is addressed with another feature in this analysis, i.e. WALS #69A, and can therefore be disregarded in the consideration of APiCS #43; therefore, only the TMA markers themselves will be considered for the value of this feature. For instance, Berbice Dutch Creole exhibits tense-aspect suffixation, but its analytical TMA markers occur only to the left of the verb, which results in a positive valuation for this feature in Berbice Creole Dutch.

Another problematic feature is APiCS #48, which addresses whether or not these languages have a nonpunctual aspect marker that marks both habitual and progressive aspect. For many of the languages in the sample, APiCS has on record that a separate habitual marker exists that is never used to mark progressive aspect. However, many of these habitual markers are used only in rare cases or in specific tenses, and function alongside a nonpunctual marker that is used to mark both habitual and progressive aspect; in such cases, the languages are deemed

compatible with the LBH for APiCS #48. According to APiCS, habitual aspect in Hawaiian Creole is unmarked in the present tense, which would be incompatible with the specifications of the LBH. However, Bickerton (1984:175) provides examples from Hawaiian Creole as spoken in 1900 which show evidence of an aspect marker *stei*, used "as a marker of nonpunctual (durative or iterative) aspect," which serve as evidence against the value assigned for this feature in Hawaiian Creole in APiCS:

(1) samtaim dei stei kam araun, polis
  'Sometimes the police used to come around.'

(2) wan taim wen wi go hom inna nait dis ting stei flai ap
  'Once when we went home at night this thing was flying about.'

Hawaiian Creole (Bickerton 1984:175)

Examples (1) and (2) above show the nonpunctual marker *stei* indicating both habitual and progressive aspect, respectively. In light of these examples, Hawaiian Creole is believed to have a nonpunctual marker used for both habitual and progressive aspect, for the purposes of this analysis. It is possible that in modern usage of Hawaiian Creole, the nonpunctual marker *stei* is no longer common among speakers; however, Bickerton's LBH is primarily concerned with the genesis of creoles, and therefore the structural profile of creoles at the time of their genesis.

APiCS features #73, 74, and 75 deal with the three-way copula distinction attributed by Bickerton to the Bioprogram (2008:38). Specifically, APiCS #74 deals with the zero copula predicted by Bickerton to present in predicative adjective constructions. Several languages in the sample use a copula in these constructions in very rare circumstances, but more frequently use the zero copula; in these cases, the present analysis assigns these languages the value of zero copula. Conversely, other languages use both a copula and the zero copula equally in predicative adjective constructions, and in such cases, the present analysis assigns these languages the value of explicit copula for this feature.

Hawaiian Creole presents yet another problem to the assignment of feature values with APiCS #84, which deals with serial verb constructions. While APiCS records that Hawaiian Creole does not exhibit 'come' and 'go' directionals, again counterevidence to these data can be found in Bickerton (1984:175):

(3) dei kam in da mawning taim go skul
'They came to school in the morning.'

(4) da frs jaepani keim ran awei fram jaepan kam
'The first Japanese who arrived ran away from Japan to here.'

Hawaiian Creole (Bickerton 1984:175)

In example (3) above, *go* marks the direction of the preceding verb *kam* 'come' in the serial verb construction, which is in the direction of the school. Example (4) shows the second verb in the serial construction, *kam* 'come', indicating the direction of the preceding verb, *ran awei* 'run away'. In light of the examples provided by Bickerton, Hawaiian Creole will be considered among the languages in the sample with 'come' and 'go' directionals.

Only four of the CCS features require discussion, as the values for the other CCS features were accepted as they were originally recorded in the source. The first two features, CCS # 8.3 and 8.4 are concerned with the grammatical functions of *fu* (or other variants thereof) in creole languages. Bickerton (1984:182) proposes four grammatical functions of the constituent *fu* that are typical to the creole typology, and which follow a hierarchy, so that more uses of the constituent observed in a language corresponds to the degree to which the language reflects the Bioprogram. The four grammatical functions of *fu* are: "*fu* can be tensed; *fu* introduces tensed complements; *fu* is a modal; *fu* marks + certain complements." CCS #8.3 and 8.4 correspond to '*fu* is a modal' and '*fu* introduces tensed complements', respectively. For most of the languages in the sample, it is clear whether or not *fu* serves these functions in the grammar. However, in the case of Berbice Creole Dutch, both functions are observed, but are very rare in usage. Because the LBH predicts the presence of these features at the genesis of creoles, and because the theory is compatible with Parkvall (2008) and McWhorter's (2001) proposal that creoles will continue to acquire and adapt features over time, the presence of these features in the sample languages is enough confirmation for the present analysis. The presence of CCS features #8.3 and 8.4 in Guinea-Bissau Creole is similar to the situation in Berbice Creole Dutch. In CCS (2001:63), it is recorded that *fu*, i.e. *pa* in Guinea-Bissau Creole, is not a true modal verb, yet example (5) below indicates that the constituent can serve the same function as in the other languages; this feature, CCS #8.3, is therefore considered to be present in the language.

(5) es   kusa   li   i   pa   kumsa janan desdi gosi
     this   thing here COP for   begin at.once since now
     'This thing should begin right now.'

Guinea-Bissau Creole (Kihm 1994:74 in Holm & Patrick 2001:63)

The other two CCS features that need to be mentioned, #13.1 and 13.2, address the three-way copula distinction specified by the LBH. As opposed to APiCS #73, 74, and 75, which are only concerned with whether or not the languages employ a copula in predicative noun phrase constructions, predicative adjective constructions, and predicative locative constructions, CCS #13.1 and 13.2 deal more specifically with whether or not the copula used in predicative noun phrase constructions differs from the copula used in predicative locative constructions, and therefore, whether or not the language differentiates between the copulas. In the case of one language, Haitian Creole, the value assigned for CCS #13.1 does not match the data recorded in CCS, but rather the data recorded in APiCS, which provide convincing arguments for the presence of a copula in predicative noun phrase constructions. Whereas CCS reports that Haitian Creole has no equative or locative copula, the APiCS entry on Haitian Creole considers the constituent *se* found in predicative noun phrase constructions to be a copula, as shown in example (6).

(6) Malis se   yon   doktè
    Malis SE   INDF   doctor
    'Malis is a doctor.'

Haitian Creole (Dejean 1982:15; DeGraff 1995:70 in APiCS 2013)

The WALS features used in the third phylogenetic network analysis were assigned the same values as in the WALS database for the non-creole languages, and for the creole languages, the same values were assigned as the corresponding APiCS values in the first dataset.

In order to value the sociohistorical features, various sources were used to find accurate information for each of the languages in the sample. Sources include APiCS, Holm (1988), and CCS. The features for the regression analysis were also formatted to accept binary values, for ease of computation and interpretation; the values were recorded in a data matrix in Excel, to be found in appendix A.

7.4 Phylogenetic Network Analysis

The next step in the analysis is to use the data matrices created when assigning feature values to prepare the input for the phylogenetic network software, SplitsTree4 (Huson & Bryant 2006). The format for the input for SplitsTree4 (Huson & Bryant 2006) is the "Nexus" file format, which is outlined in a manual by Schnoebelen (2009). The binary values for each feature set for each of the languages in the sample were converted into a "Nexus" file format; this step was repeated for each of the three data matrices corresponding to the three phylogenetic network analyses. The features were formatted in terms of binary values in order to simplify the equation to languages that either conform to or deviate from the LBH in relation to each individual feature. The "Nexus" files were then opened in the SplitsTree4 (Huson & Bryant 2006) software, forming networks using the NeighborNet algorithm, which creates split phylogenetic networks. A bootstrapping analysis can also be run to determine the strength of the relationships observed in the network (Schnoebelen 2009:10). This process is also repeated for each of the three datasets. The resulting phylogenetic networks can now be analyzed for typological relations and other relevant patterns.

7.5 Regression Analysis

In order to measure the significance of certain sociohistorical factors on the emergence of typically creole features in the sample languages, a metric needs to be established for the presence of LBH features in the sample languages. For the present analysis, the degree to which a language reflects Bioprogram features, or the *degree of creoleness*, will be measured in the percentage of the features for which the values are compatible with the specifications of the LBH. For instance, Jamaican Creole is compatible with the specifications of the LBH for 16 out of the 18 features of the first dataset of the analysis, and therefore has a value of 89 for *degree of creoleness*. The values for the sociohistorical factors and the *degree of creoleness* for each of the languages in the regression analysis sample, i.e. the same sample as the first phylogenetic network analysis[4], were recorded in a data matrix in Excel. The Excel file was then imported into SPSS, the software used for the statistical analysis. A multiple linear regression analysis was performed on the data using *degree of creoleness* as the dependent variable and *early withdrawal of the superstrate*, *prolonged contact with the superstrate*, *maroonage*, and *plantation creole* as

---

[4] With the exception of Réunion Creole and Santome because these languages were missing values for several features contributing to the degree of creoleness metric, which is central to the regression analysis.

the independent variables. The output will be analyzed for any significant sociohistorical factors influencing the degree to which the languages conform to Bioprogram features.

Chapter 8: Results

The results of the several analyses described in the previous chapter will be reported and interpreted in the following sections, in preparation for the discussion of the results in relation to the predictions of Bickerton's LBH. The output of the phylogenetic network analyses, in the form of phylogenetic networks, was examined for clusters, which can indicate relationships between languages in the sample; in the case of the current analysis, typological relationships. Additionally, using the bootstrap function specified for 1000 replications, it is possible to measure the strength of the splits in the network responsible for distinguishing between creoles and non-creoles, or between the creoles in the sample. The output for the regression analysis will be interpreted for significance, with the aim of identifying any sociohistorical factors that are directly or inversely related to the dependent variable, the *degree of creoleness* metric. The goal of the analysis of the data is to identify any patterns relevant to the predictions made by the LBH or to creole typology in general.

8.1 Phylogenetic Network Analysis

The first two phylogenetic network analyses were computed using only creole language samples, and will therefore only represent relationships between creoles based on the set of LBH features. These networks will test whether or not the languages form significant groupings, or if the group of languages shows no internal relationships. For these first two network analyses, the output is expected to show certain groupings of languages which correlate to languages that reflect the Bioprogram to varying degrees based on sociohistorical circumstances; this hypothesis is compatible with the predictions of the LBH. The output for first two phylogenetic network analyses is presented below in Figures 3 and 4, respectively.
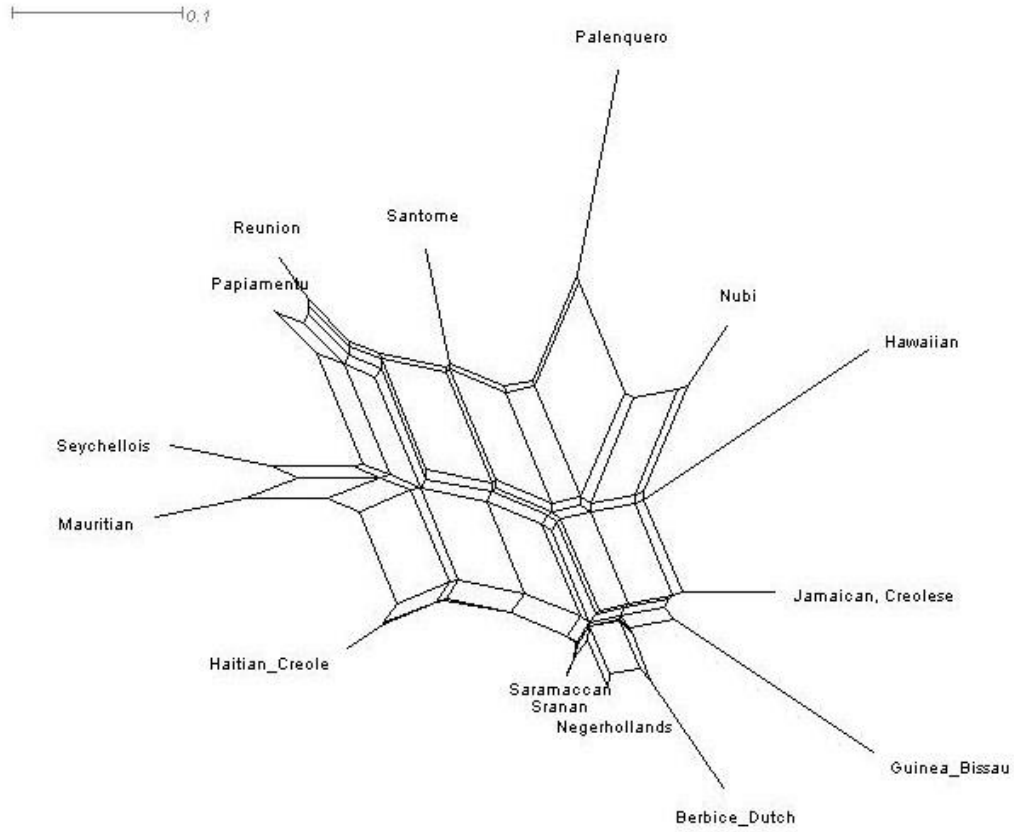
Figure 3: *Output of the first phylogenetic network analysis*
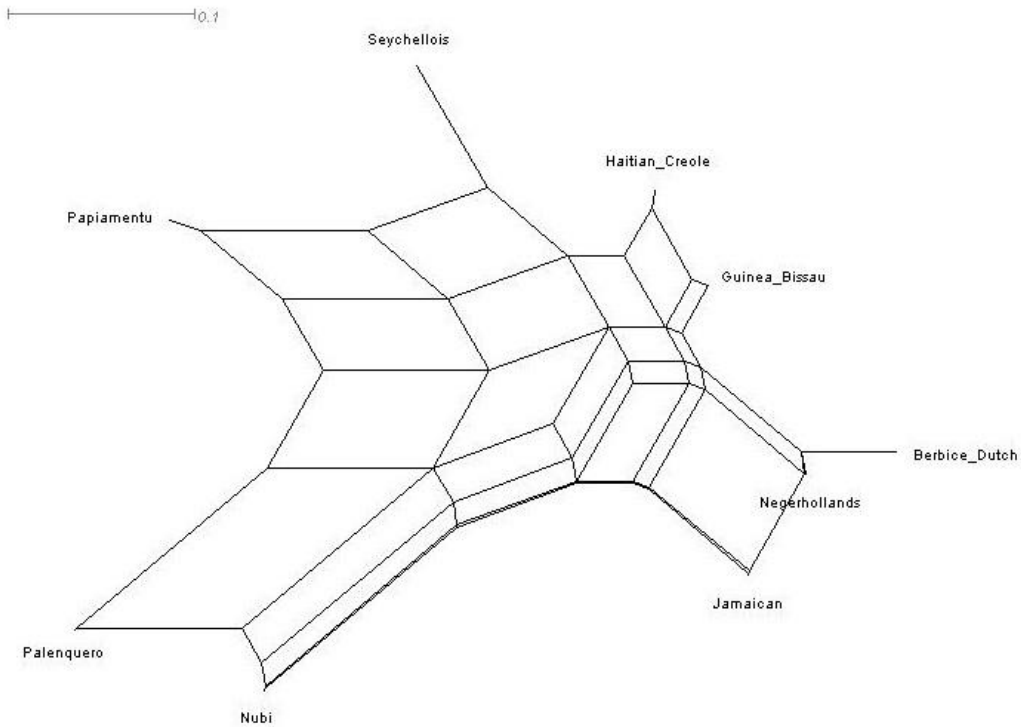
Figure 4: *Output of the second phylogenetic network analysis*

What the first network shows is relatively equal distances between the languages in the sample, according to the LBH feature set. There do not appear to be any clearly distinct groupings of languages, except for maybe Seychellois and Mauritian Creole, which is not entirely surprising considering geographical proximity and historical contact. An unexpected grouping is Réunion Creole and Papiamentu, which pattern together in the top left of the network. In the bottom right corner of the network, there appears to be a cluster of languages, including Saramaccan, Sranan, Negerhollands, Berbice Creole Dutch, Guinea-Bissau Creole, Jamaican Creole, and Creolese. Indeed, when the bootstrapping analysis was performed, it showed a pretty significant split distinguishing the bottom half of the network from the top half, consisting of the above-mentioned cluster and Haitian Creole, which patterns slightly to the left of the other languages in the grouping.

The second phylogenetic network analysis took less data as input, i.e. fewer languages and features, and the output therefore shows a much less convoluted network. The LBH predicts the same outcome for the second network as it did for the first—significant groupings of languages based on varying degrees of creoleness. Several clear groupings are immediately apparent, including Palenquero and Nubi Creole Arabic, Papiamentu and Seychellois, Haitian Creole and Guinea-Bissau Creole, and Berbice Creole Dutch, Negerhollands, and Jamaican Creole. After running the bootstrapping analysis, a significant split between the languages in the sample became apparent, separating Seychellois, Papiamentu, Palenquero, and Nubi Creole Arabic from the other languages in the sample. The potential relevance of these divisions will be discussed in Chapter 9. It is, however, clear that the second phylogenetic network shows less equality in distance between the languages in the sample.

The third phylogenetic network analysis had the largest input of the three phylogenetic analyses in terms of languages, and therefore resulted in the most complex network of the three, shown below in Figure 5.

Figure 5: *Output of the third phylogenetic network analysis*

Figure 6: *Bootstrapping analysis performed on the output of the third phylogenetic network analysis*

The first clear trend in the network is the split in the middle distinguishing the creole languages from the other languages in the sample, with the exception of Vietnamese, Wolof, and Fongbe, which pattern along with the creoles in this instance. The fact that a clear distinction presents between creoles and non-creoles in the language sample alone provides support for the existence of a creole typology. While no split alone accounts for just the creole languages in the sample, another split in the network does distinguish the three non-creoles that pattern with the creoles, i.e. Vietnamese, Wolof, and Fongbe, from all of the other languages in the sample; although these languages pattern with creoles in the network, there is also something that distinguishes them from the creole languages in the sample. Another split in the network separates Palenquero and Berbice Creole Dutch from the grouping of Vietnamese, Wolof, Fongbe, and the creole languages. The one grouping that does consist of only creole languages includes Jamaican Creole, Creolese, Saramaccan, Sranan, Haitian Creole, Negerhollands, Seychellois, Mauritian Creole, Réunion Creole, Santome, Papiamentu, and Palenquero, which does exclude Berbice Creole Dutch, Nubi Creole Arabic, Hawaiian Creole, and Guinea-Bissau Creole. Potential explanations for the different groupings of creoles in this phylogenetic network will be explored in the discussion of the results. Looking at the interaction between the creoles and non-creoles in the network, the creole languages definitely pattern separately from several of their superstrates and substrates, with the exception of Fongbe and Wolof, which may have implications for Superstratist and Substratist theories of creole genesis. Additionally, of the three non-creoles that do pattern with the creole languages in the network, one language ranks extremely low in Parkvall's (2008) simplicity metric; this could contribute to Parkvall's (2008) proposal of the systematic simplicity of the creole typology. The trends identified in the three phylogenetic network analyses will be discussed in terms of their implications for the LBH and creole typology in general in Chapter 9.

8.2 Regression Analysis

The multiple linear regression model is intended to identify any relationships between a continuous dependent variable, in this case the *degree of creoleness*, and several predictor variables. For the present analysis, a multiple linear regression analysis was conducted to identify any relationships between the percentage of features for which the languages are compatible with the Bioprogram, and sociohistorical factors predicted by the LBH to influence

that percentage; the features are all listed above in Chapter 7. The interpretation of the output from the regression analysis showed that one factor, *prolonged contact with the superstrate*, was a significant factor in relation to *the degree of creoleness*. The output is presented in Table 5 below.

	t	Sig.
Early Withdrawal of the Superstrate	-1.456	.179
Prolonged Contact with Superstrate	-2.498	.034
Maroonage	.372	.719
Plantation Creole	-.108	.916

Table 5: *Output of the multiple linear regression analysis*

Although the ANOVA shows that the model as a whole cannot account for variation in *degree of creoleness* between the languages in the sample, indicated by a p-value of .261, the coefficients indicate that *prolonged contact with the superstrate* is inversely related to the *degree of creoleness*, with a p-value of .034, which is significant at p≤.05. Apart from *prolonged contact with the superstrate*, only *early withdrawal of the superstrate* approached significance in the model, with a p-value of .179. These two factors can potentially account for some of the variation in the phylogenetic networks.

Chapter 9: Discussion

Several trends observed in the results of the phylogenetic and regression analyses require further discussion in order to understand their implications for the field of creolistics and to contextualize the results in terms of the goals of this thesis. The restated research goals of the present analysis are:

- ➢ Validate the proposal of a creole typology
- ➢ Test whether or not creoles form significant groupings in phylogenetic networks based on LBH features
- ➢ Establish relationships between sociohistorical circumstances and the presence of Bioprogram features

The networks generated in the three phylogenetic network analyses do show some significant groupings that warrant discussion, in addition to the output of the multiple regression analysis, which indicates an inverse relationship between *degree of creoleness* and *prolonged contact with the superstrate*. These results need to be considered in the context of the research goals, restated above, in order to draw any meaningful conclusions concerning the present research. The first step is to analyze the direct implications of the results of the analyses for the future of Bickerton's LBH and of creole typology research. It is important to identify any patterns in the way that the languages cluster in the networks, for example in relation to sociohistorical factors, geographical contact, or genetic affiliations, though it is likely to be some combination of those factors. It would be particularly interesting if there is any interaction between the groupings in the network and the results of the regression analysis, which will be addressed further on in the discussion.

Other important topics of this discussion include the implications of the results for the creole typology, a reconsideration of Bickerton's LBH, and potential accounts of the variation observed in the group of creole languages. Some groupings in the network analyses indicate that the creole typology may indeed be connected to Parkvall (2008) and McWhorter's (1998; 2001) suggestion of the relative simplicity of the creole typology. Section 9.1 will address the creole typological profile in the context of the results of the present analysis, with the aim of drawing conclusions regarding the relationship between the LBH, Parkvall and McWhorter's simplicity metrics, and the creole typology. After contemplating the relevance of the results of the analyses for creole typology research, Bickerton's LBH will be reconsidered in light of the results.

Although the results of the several analyses are not compatible at every level with Bickerton's theory of creole genesis, trends can be inferred from the results and implemented in the reformation of the LBH to account for recent developments in creolistics. The compatibility shown in the phylogenetic networks between Bickerton's LBH and Parkvall's proposal of the systematic simplicity of the creole typological profile can be incorporated into a comprehensive theory of creole genesis and classification. Lastly, it is important to discuss the interaction between the linguistic and extralinguistic factors associated with creole languages, as it was predicted by the LBH and plays a large role in interpreting the output of the analyses. The variability in the data shows most strongly that various processes are at work during creolization and afterward that have influenced creoles in ways distinct from other languages, yet also unique to each creole or group of creoles. Hopefully, the discussion of the results brings new insight to the debate over creole origins and classification.

9.1 Implications for the LBH and Other Theories of Creole Genesis

According to the specifications of the LBH, the phylogenetic networks analyses were expected to show significant clusters of creole languages associated with certain explanatory factors. In a general sense, the languages in the first phylogenetic network are relatively equidistant from one another, showing structural variability among the class of creole languages. The similar distances between the languages in the network indicate that, although the creole typological profile may be unique to creole languages, creoles themselves vary considerably in terms of how many of these features are exhibited and which features specifically; these findings reflect similar findings in current phylogenetic research (Daval-Markussen & Bakker 2012; Bakker et al. 2011), as discussed in Chapter 6. When considering the few clear groupings in the network, it appears that several factors do correlate with the clustering of the languages. Seychellois and Mauritian cluster together, with Haitian Creole and Réunion creole patterning close by, representing both the French-based creoles and the French-based creoles spoken on the Indian Ocean islands, as a subset of the French-based creoles. Indeed, creoles in contact for long periods of time, such as the French-based creoles in the Indian Ocean and the English-based Atlantic Creoles, do often share many features. The first phylogenetic network analysis also seems to interact with the results of the regression analysis, with the languages assigned a positive value for *prolonged contact with the superstrate* patterning at the periphery of the

language cluster, which itself appears to be correlated to *degree of creoleness*. The network therefore shows that the languages in the sample cluster according to both *degree of creoleness* and *prolonged contact with the superstrate*, reflecting the pattern identified in the multiple regression analysis. Though *prolonged contact with the superstrate* does influence *degree of creoleness*, it does not account for the full distribution of creole languages in the network. With the clear grouping of languages with recorded geographical contact, it appears that the interaction of various sociohistorical factors exclusive to each language is responsible for the unique development of each individual creole, and therefore the structural variation observed in the creole class of languages.

The second phylogenetic network also shows a division between creoles with a low score for *degree of creoleness*, i.e. Seychellois, Nubi Creole Arabic, Palenquero, and Papiamentu, and those with a higher score. However, there is no apparent explanatory factor able to account for the distribution of languages in the network, as both groupings include languages with and without *prolonged contact with the superstrate* and close geographical contact with other creole languages; the languages in both groupings also come from various lexifier-bases and have differing substrate influences. The results of this network analysis again support the structural variation within the class of creole languages, and the complexity of the process of creolization. The relatively equal distances observed between creoles in phylogenetic network analyses in both the present analysis and previous research (see Chapter 6) serves as counterevidence against theories of monogenesis, as well as any direct correlations between *degree of creoleness* and sociohistorical factors, at least in terms of Bickerton's LBH predictions.

The final phylogenetic network analysis, taking into consideration both creoles and non-creoles, provides the most insight into the creole typological profile. The clear division in the network between the creoles and non-creoles, with the exception of three non-creoles that pattern with the creole languages, is an interesting result for the LBH and the creole typology, indicating at least a unique typological class of creole languages and at most providing evidence against other theories of creole genesis. Additionally, the three non-creoles that pattern with the creole cluster can be distinguished from the rest of the sample with one split in the network, suggesting that there is a difference between these languages and the creoles in the sample. The fact that the creoles do pattern differently from other languages according to the specifications of Bickerton's LBH validates both the idea of a typological class of creole languages as well as the accuracy of

the LBH in accounting for the creole typology. These results contribute to several other works which have been able to distinguish between creoles and non-creoles according to a chosen set of features (Daval-Markussen & Bakker 2012; Daval-Markussen 2013). Furthermore, the fact that the creoles in the network pattern distinctly apart from several of their lexifier and substrate languages provides concrete evidence counter to the Superstratist and Substratist theories of creole genesis, reflecting the results of Daval-Markussen & Bakker (2012). In fact, only Berbice Creole Dutch patterns anywhere near a European language, though not its own lexifier language, and only two languages representative of substrate influence pattern similarly to the creoles— Fongbe and Wolof. Regardless of which features are chosen as input for the phylogenetic network analysis, Superstratist and Substratist positions should predict patterning with the superstrates and substrates, respectively; this is not what is shown in the networks generated during this analysis or others (see Chapter 6).

9.1.1 Simplicity of the Creole Typological Profile

The only non-creoles that pattern similarly to the creoles in the network include one language ranked quite low on Parkvall's (2008) simplicity metric, Vietnamese, and two substrates that reflect several of Parkvall's metrics of simplicity, Fongbe and Wolof. Parkvall's simplicity metric (2008) was calculating using numerous typological features, yet he stresses the lack of inflection exhibited by creoles. Parkvall expresses that it is not inflection itself that renders a language complex (see Chapter 4, section 4.2), but the effects associated with inflection, including morphophonemic processes. Inflection as a diagnostic of simplicity, and simplicity as descriptive of the creole typology, is compatible with the specifications of Bickerton's LBH. Several of the features associated with the LBH can be traced back to a lack of inflection, including analytical grammatical markers (negative and plural markers), lack of tense-aspect inflections, and preverbal TMA markers. Fongbe and Wolof are the only languages in the sample, apart from the creole languages and Vietnamese, that exhibit no tense-aspect inflection, a distinction that divides the phylogenetic network into two distinct groups. Parkvall suggests not that creoles must be the only languages at the lower end of the complexity scale, but that the grouping of creoles at the bottom of the scale distinguishes them as a typological class. The results of the third phylogenetic network analysis provide support for Parkvall's classification of

creoles based on low complexity scores as well as for Bickerton's LBH specifications as a set of creole typological features, which likely overlap.

There are also significant implications of the regression analysis for the validity of the LBH, which predicts that several sociohistorical factors are influential in the degree to which languages reflect the Bioprogram, or the creole typological profile. Although three of the four predictor variables showed no significant relationship to *degree of creoleness*, *prolonged contact with the superstrate* did show a significant inverse relationship to *degree of creoleness*. Creole languages associated with *prolonged contact with the superstrate* could also be expected to have increased more complex structures, as increased input and pressure from the dominant language can be associated with the adoption of more complex, superstrate features. Furthermore, the fact that the data in the first phylogenetic network appear to interact with the results of the regression analysis adds even more support to Bickerton and Parkvall's theories.

9.2 Implications for the Creole Typology

Again, the results of the third phylogenetic network analysis provide the strongest support for the creole typological class. Other analyses have employed phylogenetic computational techniques to test the creole typology, finding that creoles pattern away from superstrates, substrates, based on sets of features consisting of as few as just three features (Daval-Markussen & Bakker 2012; Daval-Markussen 2013). The third phylogenetic network generated in the present analysis indirectly distinguished creoles from non-creoles, based on a set of featured predicted by Bickerton's LBH. The fact that several different analyses, using different, though often overlapping, sets of features frequently observed in creole languages, found a division between creoles and non-creoles provides strong support for the existence of a typological class of creole languages. Yet the internal variation among the creole languages observed in the phylogenetic networks indicates that the relationships between creole languages and extralinguistic factors are complicated and likely unique to the origin and development of each creole or group of creoles. The patterning of creoles with Vietnamese, a language with a low complexity score, and two substrates that exhibit lack of tense-aspect inflection, Wolof and Fongbe, may indicate overlap between Bickerton's LBH and Parkvall (2008) and McWhorter's (2001) typological profile of creoles based on systematically simple structures. As discussed in the previous section, Bickerton's LBH and the creole typology in general can be defined in terms

of less complex structures, relating to circumstances of creole genesis which will be discussed in a subsequent section. Interestingly, the results of the phylogenetic network analyses conducted by Daval-Markussen & Bakker (2012) also show several West African substrate languages patterning around the periphery of the creole cluster (see Chapter 6, section 6.2.1). This could indicate either the significance of substrate influence in the development of creoles, or the presence of structural similarities between West African languages and the creole typological profile.

Despite the fact that simplicity can contribute to an understanding of the creole typology, it is not a comprehensive description of the creole typological profile, as the current analysis and previous analyses have found a distinction between creoles and non-creoles with equally low complexity scores (see Chapter 4, section 4.2). The circumstances surrounding creole genesis, and the resulting simplicity of creole structures, may contribute to the creole typology, but it does not define it entirely; this contributes further to the idea that multiple processes are involved in the formation of the creole typology, including simplification, geographical diffusion, superstrate and substrate influence, and various sociohistorical factors. While the present research contributes to the evidence in support of a creole typology, it also shows the compatibility between the specifications of Bickerton's LBH and the creole typological profile. Although the results of the analysis show a typological relationship between the creole languages, it does not account for most of the structural variation observed among the class of creole languages. *Prolonged contact with the superstrate* did show significance in relation to *degree of creoleness* according to the regression analysis, yet there is much more variation in the first two phylogenetic network analyses that cannot be accounted for by the LBH. A reconsideration of the LBH in light of the results of the present analysis follows in the next section.

9.3 Reconsidering the LBH

Although the LBH is linguistically compatible with the results of the analysis, only one of the four sociohistorical factors predicted by the LBH showed significance in relation to *degree of creoleness*. This section attempts a reconciliation of the LBH with the knowledge that has been gained through recent phylogenetic research of the creole typology. In Chapter 4, the idea of defining the creole typology based on the simplicity of creole features was presented in Parkvall (2008) and McWhorter (1998; 2001) (see sections 4.2 and 4.1, respectively); some of

the main points of their proposals will be reviewed here for the discussion. Parkvall and McWhorter suggest that the circumstances surrounding creole genesis triggered the process of creolization, which included the simplification of superstrate features, and perhaps the innovation of new, less complex features. Their theory is that creoles are a synchronic class of languages, exhibiting less complex features because complex features accrue over time, and creole languages are still too young to be as complex as languages that had no break in transmission. This account of creole genesis is compatible with Bickerton's LBH, which proposes that a break in transmission triggered creolization, relying on intervention from the Bioprogram depending on the quality and quantity of the input provided by the pidgin. Because the Bioprogram is supposed to have originated back at the genesis of human language, and only resurfaces when input is required (see Chapter 3), it can easily represent the simplification processes assumed by Parkvall and McWhorter to constitute a large part of the process of creolization. Both theories assume that creoles are a synchronic class of languages, which acquire complexities over time, growing to resemble non-creoles. However, in phylogenetic network analyses, lack of complexity cannot account entirely for the distribution of creoles, as creoles pattern somewhat differently from non-creoles with similarly less complex features (Parkvall 2008:282). The Bioprogram features are actually a more specific description of the creole typology, considering that the creoles did pattern differently in the network analyses according to features specified by the LBH. Parkvall and McWhorter make no claims regarding the internal variation within the class of creole languages, leaving the question of the relationship between the creole typology and sociohistorical factors unanswered. The regression analysis showed *that prolonged contact with the superstrate* is inversely related to the *degree of creoleness*, accounting for some of the variation observed in the networks, specifically in the first phylogenetic network analysis. However, other trends possibly explaining the distribution of the languages in the networks include geographical diffusion, common superstrate and substrate influence, early withdrawal of the superstrate, and the creole identity. During the origin and early development of creoles, many factors likely contributed to the resulting creole language, including extent and duration of contact with the superstrate, number and diversity of substrates, population demographics, and the formation of a creole identity. During later stages of development, other extralinguistic factors continue to influence the development of the language, specifically prolonged contact with the superstrate, geographical diffusion, and association with a

creole identity. It is likely that the typological profile of each creole language, or group of creole languages, has been influenced by a unique combination of the above mentioned extralinguistic factors. These conclusions do not exclude the LBH as a theory of creole genesis, but rather indicates that Bickerton's theory oversimplifies the relationship between sociohistorical factors and the presence of Bioprogram features.

9.4 Social History and Internal Variation

This section focuses specifically on the sociohistorical factors that have contributed to the genesis and development of creole languages, proposed in the previous sections to account both for the distinction between creoles and non-creoles and the internal variation among the class of creole languages. The extralinguistic factors considered by this analysis to have significantly influenced the various stages of creole development, from pidginization to creolization to decreolization, will be reviewed in relation to the specific effects that each factor can have on creole languages; this examination of extralinguistic factors associated with creole development is intended to contribute to an understanding of the complex nature of creolization and the various processes that condition it. Beginning at the beginning, and thus with the process of pidginization, as this analysis follows Bickerton's LBH which assumes the existence of a pidgin-creole lifecycle, there are several extralinguistic factors that can influence the process and, most importantly, the outcome of pidginization. Social conditions and linguistic input are two features most commonly associated with the early conditions necessary in speech communities for the initiation of pidginization. Many in the field have written of the social conditions necessary to initiate the processes of pidginization and creolization (Bickerton 1984;2008; Chaudenson 2003), and some have gone as far as to classify creoles according to the social circumstances under which they originated. Examples of such sociohistorical classifications of creoles include Bickerton's distinction between plantation, fort, and maritime creoles (see Chapter 3) and, to a lesser extent, Bickerton and Chaudenson's social matrices of creolization. Plantation creoles are thought to be the most compatible with the social matrix of creolization, providing the perfect social conditions for pidginization: massive importation of labor from linguistically-distinct communities, limited contact between labor population and colonist population, and the need for a means of communication. Therefore, from the beginning of the establishment of what would become creole speech communities, there are several different sociohistorical factors that can

result in multiple distinct outcomes of pidginization. The second extralinguistic factor contributing to the different outcomes of pidginization is the number of substrate influences and the degree of mutual intelligibility between those substrates. The idea is that fewer substrate influences, and a higher degree of mutual intelligibility between those substrates, results in easier communication among these speech communities, and therefore less immediate need for simplification of superstrate or substrate varieties. In fact, Thomason and Kaufman (1988:157-8) discuss the lack of diversity in the substrates of Guinea-Bissau Creole and Santome as a potential factor in the low degree of creoleness of the resulting creoles; they also go on to discuss the number of dominant-language speakers as a contributing factor to the variability in degree of creoleness observed in the French-based creoles. It follows that, based on the compatibility of the social history of a creole language with the social matrix of creolization, i.e. the conditions triggering pidginization and creolization, in combination with the amount of substrate influence and the degree of mutual intelligibility between said substrates, a spectrum of outcomes of pidginization can be expected. Until this point, Bickerton's account of pidginization and the social conditions of early creole speech communities is compatible with the above explanation of the influence of extralinguistic factors on the process of pidginization. Because the LBH predicts that sociohistorical conditions surrounding pidginization influence the quality of the pidgin, the expected results of pidginization according to Bickerton can also be seen in terms of a spectrum, predicting a trend of plantation creoles with high diversity in substrate influence at the lower end. Though it can already be assumed that the spectrum of possible outcomes of pidginization results in a high degree of variability among the class of creole languages, still more extralinguistic factors influence the process of creolization as well as the subsequent stages of creole development.

Several extralinguistic factors are thought to play a role in the next step of creole genesis, namely the early withdrawal of the superstrate, relexification, prolonged contact with the superstrate, genetic affiliations, geographical diffusion, and the creole identity. Although these factors influence the development of creoles at different stages and in different ways, and though several of these have proven to be insignificant in terms of *degree of creoleness* in the regression analysis, they have all been argued to have influenced creoles to an extent, and will therefore be reviewed for their contributions to the variation observed among the creole languages. Likely the first factor to influence the process of creolization is whether or not the language experienced

early withdrawal of the superstrate language. In many communities, access to the superstrate was withdrawn not long after creolization, often "because of political change, as with the switch from English to Dutch rule in Suriname; […] the result is to cut off any further influence from native speakers of the dominant language such as would otherwise continue to occur, albeit on a reduced scale and in an indirect manner" (Bickerton 1985:178). Though *early withdrawal of the superstrate* did approach significance in the regression analysis, only *prolonged contact with the superstrate* proved to be a significant factor in relation to *degree of creoleness*. However, the fact that *early withdrawal of the superstrate* does not predict the degree to which a creole will reflect the creole typology does not necessarily indicate that this factor has no influence on creolization. In fact, it would be expected that, depending on which languages were involved in the genesis of a particular creole, the early withdrawal of the superstrate would result in varying levels of dependence on the process of creolization to satisfy various different gaps in the grammar.

The process of relexification is often connected to early withdrawal of the superstrate, as many sociohistorical factors are connected to and interdependent upon one another. When relexification occurs, a new superstrate is introduced to the speech community, providing new superstrate input for the creole speakers and often disguising evidence of earlier processes that took place during the development of the language. Not surprisingly, *prolonged contact with the superstrate* was the only extralinguistic factor to show a significant relationship to the degree of creoleness, likely because this factor does not affect the input provided to the creole speakers, but rather refers to the relationship between the creole and the superstrate after creolization. Prolonged contact with the dominant language is said to trigger the process of decreolization (Bickerton 1984:178), which is a term used to describe the process of the loss of typically creole features and the adoption of more superstrate features in response to social pressure from the dominant population (Holm 1988:9). The relationship *between prolonged contact with the superstrate* and *degree of creoleness* is therefore much less convoluted than the relationship between *early withdrawal from the superstrate* and *degree of creoleness*, as *prolonged contact with the superstrate* is associated with later stages of creole development and cannot result in the same variation observed in creoles with early superstrate withdrawal.

Other extralinguistic factors that can contribute to the variation observed among the class of creole languages in the phylogenetic network analyses include geographical diffusion, genetic affiliations, and the creole identity. It has been proposed several times throughout the history of

the study of creoles that genetic affiliations exist among some subgroups of creole languages, such as the French-based creoles and the English-based Atlantic Creoles; this is referred to as polygenesis, and was discussed in more detail in Chapter 2. Such theories may be corroborated by several of the groupings of supposed genetically related creoles observed in the phylogenetic network analysis (see Chapter 8), such as the French-based creoles in the Indian Ocean. Other processes that take place after genetically affiliated creoles have been separated from one other may mask the common origins of those languages, resulting in an inaccurate picture of creole genesis. It is important to consider the fact that some creoles may be genetically affiliated, and therefore share a set of features, which can result in certain unexplained trends in phylogenetic network research. For example, the proposed genetic relationships within the English-based Atlantic creoles (Daval-Markussen & Bakker 2011) may account for the clustering of Jamaican Creole and Creolese in the heart of the creole cluster in the third phylogenetic network.

Geographical diffusion has likely contributed to the distribution of creole typological features among the class of creole languages, yet is often overlooked by analyses of creole typology and genesis. Many of the proposed genetic relationships between creoles may also be a misinterpretation of the geographical diffusion of features, as many creoles in close contact share a set of features, such as the French-based creoles in the Indian Ocean. It is not entirely relevant to the present analysis whether or not shared features occur in creoles because of genetic affiliations or geographical diffusion, but that something other than the human capacity for language is responsible for the presence of specific features in creole languages. However, both of these extralinguistic factors can account for the grouping of certain creoles in the phylogenetic network analysis, such as the grouping of the French-based creoles in the first network, or the grouping of the Atlantic English-based creoles in the third network. In fact, very few of these extralinguistic factors are mutually exclusive, meaning that nearly any combination of these factors can be responsible for the features observed in creole languages, and therefore the inexplicable variability observed in the networks. In the same way that Thomason & Kaufman believe that "the structure of the emerging creole will be a function of the structures of its developers' native languages" (1988:153), this discussion proposes that pidgins and creoles are also a function of their sociohistorical circumstances. The last extralinguistic factor to take into account is the presence and importance of the creole identity in creole speech communities. Though not exactly quantifiable, many have reported on the importance of identity in the

formation of creole languages, and especially the differing roles of the creole identity in situations where the dominant language remains in contact with the creole, as opposed to situations where the superstrate was withdrawn. The creole identity will be discussed in more detail in relation to the creole typology in the following section. Although this exact account of pidginization, creolization, and the processes that influenced the origin and development of creoles may not be perfectly compatible with several other theories of creole genesis, the influence of the above-discussed extralinguistic factors can be identified in any comprehensive theory of creole genesis.

9.4.1 The Creole Identity

Because creole speakers were forced to rapidly disassociate from their native languages in favor of a contact variety, i.e. the pidgin or creole language, they also disassociated with part of their native culture in favor of a creole identity. As language and culture have been said to be deeply intertwined, speech communities often correlate to shared cultural identities: "Important as European expansion, slavery, the plantation system, and the deracination of Africans unquestionably are, it is now generally recognized that the conjoint participation of different peoples, not least those from Africa, produced from a very early time a distinctive African-American or creole culture" (Bolland 1998:1). McWhorter's (1998:800) claim that, in some creole speech communities, close approximations of the lexifier were spoken alongside early creoles indicates that creolization is a social process as much as it is a linguistic process. The creole identity can be seen as a force that both drives and inhibits the process of decreolization. Take, for instance, the Haitian Creole speech community in the former plantation colony of Haiti, which exhibits a complex post-colonial relationship with its superstrate language, French. One could argue both that the creole identity inhibited the decreolization of Haitian Creole by encouraging members of the speech community to identify with the creole language as opposed to the dominant European language, and that the creole identity is the force driving decreolization, as members of the community acknowledge the power and privilege provided by access to the superstrate language and Western culture. Bolland (1998:2) refers to the creole-society model as "a significant ideological moment in the decolonization process of the Caribbean." Because of the inherent relationship between society, language, and culture, Bolland's connection between the creolization and decolonization of a society can be extended to

language. The creole identity can function differently in each creole speech community, likely in relation to other contributing sociohistorical factors, specifically in relation to various combinations of substrate languages, cultures, and identities. However, it is clear that the formation of a creole identity influences the development of the associated creole in ways not yet completely understood. Future research could address the creole identity individually in order to understand its influence on creole development.

9.5 Creole Sociolinguistics and the Creole Continuum

Not many analyses of creole typology and classification address the sociolinguistic situation in creole speech communities, despite the fact that creoles have developed sociolinguistically as much as structurally. Decreolization in several creole speech communities due to prolonged contact with the superstrate language has resulted in an interesting sociolinguistic phenomenon, namely the creole continuum (see section 2.3). With the superstrate language held as the standard for a prolonged period of time, creole speakers begin to use less typically creole features and more features associated with the European superstrate language. This process results in a spectrum, or continuum, of varieties in intermediary positions between the so-called 'deep creole' and the superstrate. These intermediary varieties have been shown to be correlated to social factors, including age, profession, and social class. The internal variation observed in each of these creoles complicates creole typology research, as the basilectal varieties are quickly becoming less representative of creole languages. Following the proposal of a synchronic class of creole languages (Bickerton 1981; McWhorter 1998; 2001; Parkvall 2008; Thomason & Kaufman 1988), it would seem as though the synchronic class of creole languages has been disappearing since its origin, constantly evolving in relation to the many extralinguistic influences associated with creole origins. As the world grows more connected, and social gaps in creole societies shrink, the sociolinguistic situation in these societies will continue to change, affecting creole languages in the same ways social factors influence non-creole languages.

Chapter 10: Conclusion

This analysis set out to address two pressing issues in the field of creolistics, namely the existence of a typological class of creole languages and the source from which the shared set of typological features stem. The long history of the debate over creole genesis in creolistics has recently been subjected to quantitative methods relatively new to the field; the present thesis is a continuation of these efforts. In the review of the relevant literature, several trends were identified in the theories that have presented over the years, specifically Superstratist, Substratist, and Universalist theories of creole genesis. These three theories have circulated in the field for decades, resurfacing in the discourse on creole classification and, more recently, creole typology. Modern creole typology research seeks to identify a set of typological features that set creole languages apart, as well as languages that fit the resulting typological profile. In order to avoid the circularity of identifying languages that match a profile derived from analyzing features of those languages themselves, Bakker (2014) proposes defining creoles non-linguistically in addition to only considering compatibility with a typological profile. Phylogenetic computational techniques, used to predict genetic relationships in the biological sciences, have been adopted into the field of linguistics, and more recently creolistics, in order to help in defining linguistic genetic relationships. Previous phylogenetic network analyses of creole languages have tested the most prominent theories of creole genesis, i.e. Superstratism, Substratism, and Universalism, with the aim of identifying trends in the clustering of creoles in the network (Daval-Markussen & Bakker 2012). While Superstratist and Substratist theories predict clustering of creoles with their superstrates and substrates, respectively, creole languages consistently patterning away from superstrate and substrate languages would support Universal theories of creole genesis. The results of these analyses have indicated that creoles do pattern away from superstrate and substrate languages.

Bickerton's Language Bioprogram Hypothesis presents a comprehensive Universal theory of creole genesis and creole typology, proposing a human Bioprogram for language, which predicts both a set of typological features associated with creole languages and a common structural base for all creoles, and in fact for all human language. Because Bickerton's theory accounts for the structural similarities between creoles as well as the internal variation observed within the class of creole languages, there was great potential for this theory to explain trends in previous research and to shed light on this controversial topic in the field of creolistics. The

present analysis was designed to test the accuracy of the LBH in accounting for the distribution of creoles in phylogenetic networks. The first two phylogenetic network analyses computed the values of LBH features for only creole languages, showing a relatively high degree of variability in the output. The third phylogenetic network analysis, considering both creole and non-creole languages, shows a clear division between the creoles and the other languages, with some interesting clusters instantly identified. Bickerton's LBH features were successful in accounting for the creole typological profile, but did not have much explanatory power in terms of the variability within the sample of creole languages, as the multiple regression analysis shows only one significant predictor variable, namely *prolonged contact with the superstrate*. The results also have significant implications for the creole typology, as the distinction in the third network between creoles and non-creoles, and the additional distinction between the creoles and the languages of low complexity, indicates that some set of features included in the analysis is unique to creoles.

The discussion of the results considered the various extralinguistic factors potentially responsible for the unexpected patterning of languages within the creole cluster, arriving at the conclusion that there is likely variability in the combination of extralinguistic factors influencing creolization in each creole community, associated with the structural variation observed in the network analyses. Variation in the combination of social factors surrounding creole origins and subsequent development results in variation in the individual structure of each language, while the languages as a group still adhere to the creole typological profile. While some of the contributing factors show a more predictable relationship to the *degree of creoleness*, such as *prolonged contact with the superstrate*, other factors, like the presence of a creole identity, can behave differently across the many creole speech communities. While the results of the phylogenetic network analysis support the creole typology, this analysis only contributes to a definition of the creole typological profile; it does not attempt to define it entirely.

10.1 Further Research

The future of creole typology research depends on both quantitative methods, such as phylogenetic computation, and qualitative analysis. In order to gain a deeper understanding of the combination of extralinguistic factors that influence the structure of creole languages, more focus must be placed on the history of these languages and their speakers. However, in order to

better understand the creole typological profile and its origins, more emphasis must be placed on massive data analysis. Future research is encouraged to take a diachronic perspective, using phylogenetic computation to map the relationships between creoles over time according to a set of features representative of the creole typological profile. Because creoles have been argued to constitute a synchronic class of languages originating around 500 years ago, it is expected that the structural profiles of the languages will have become increasingly distinct from one another, and also the Bioprogram, over time. Extralinguistic factors such as geographical proximity and the creole identity are also expected to have influenced the development of creoles since their genesis. Though strong evidence has been presented in support of a synchronic class of creole languages, including in the current research, it is no longer practical to conduct synchronic research in order to solve a diachronic dilemma. Lack of interest in the early documentation of pidgin and creole languages has resulted in a gap in the historical record of many creole languages and societies; this presents a substantial methodological issue concerning the potential for diachronic quantitative research in creolistics. However, several languages with substantial historical documentation appear to be suitable candidates for this future line of research, namely Sranan, Negerhollands, and Papiamentu, among others. Continuation of this line of research will lead to more insight into creole languages, their origins, and potentially the origins of human language.

References

Avram, A.A. (2013). The distribution of diagnostic features in English-lexifier contact languages: Virgin Islands Creole. *Acta Linguistica Hafniensia*, *45*(2), 206-227.

Baker, P., & Huber, M. (2001). Atlantic, Pacific, and world-wide features in English-lexicon contact languages. *English World Wide*, *22*(2), 157-208.

Baker, P., & Kriebel, S. (2013). Mauritian Creole structural dataset. In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/55).

Bakker, P. (2002). Some future challenges for pidgin and creole studies. In Gilbert, G. (Ed.) *Pidgin and Creole Linguistics in the Twenty-First Century* (69-92). New York: Peter Lang.

Bakker, P., Daval-Markussen, A., Parkvall, M., & Plag, I. (2011). Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages, 26*(1), 5–42. Reprinted in 2013 in: Bhatt, P. & Veenstra, T. (Eds.), *Creole Languages and Linguistic Typology* (9-45). Amsterdam: John Benjamins.

Bakker, P. & Daval-Markussen, A. (2012). Explorations in creole research with phylogenetic tools. Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH, 89-97.

Bakker, P. (2014a). Creolistics: back to square one. *Journal of Pidgin and Creole Languages, 29*(1), 177-194.

Bakker, P. (2014b). Creoles and typology: problems of sampling and definition. *Journal of Pidgin and Creole Languages*, *29*(2), 437-455.

Barbançon, F., Evans, S.N., Nakhleh, L., Ringe, D., & Warnow, T. (2013). An experimental study comparing linguistic phylogenetic reconstruction methods. *Diachronica*, *30*(2), 143-170.

Bickerton, D., & Escalante, A. (1970). Palenquero: a Spanish-based creole of Northern Colombia. *Lingua*, *24*, 254-267.

Bickerton, D. (1973). The nature of a creole continuum. *Language*, *49*(3), 640-669.

Bickerton, D. (1975). *Dynamics of a Creole System*. Cambridge: Cambridge UP.

Bickerton, D. (1976). Pidgin and creole studies. *Annual Review of Anthropology*, *5*, 169-193.

Bickerton, D. (1981). *Roots of Language*. Ann Arbor: Karoma.

Bickerton, D. (1982). *The Lexical Learning Hypothesis and the Pidgin-Creole Lifecycle*. University of Duisburg.

Bickerton, D. (1984). The language bioprogram hypothesis. *The Behavioral and Brain Sciences, 7*, 173-221

Bickerton, D. (1990). *Language and Species*. Chicago: U of Chicago P.

Bickerton, D. (2004). Reconsidering creole exceptionalism. *Linguistic Society of America*, *80*(4), 828-833.

Bickerton, D. (2008). *Bastard Tongues*. New York: Hill and Wang.

Bickerton, D. & Szathmáry, E. (Eds.) (2009). *Biological Foundations and Origin of Syntax*. Boston: MIT Press.

Bickerton, D. (2013). The evolution of language. In D. Reisburg (Ed.), *The Oxford Handbook of Cognitive Psychology* (524-539). Oxford: Oxford UP.

Bolland, O.N. (1998). Creolisation and creole societies: a cultural nationalist view of Caribbean social history. *Caribbean Quarterly, 44*(1/2), 1-32 Retrieved from http://www.jstor.org/stable/40654019.

Bollée, A. (2013). Réunion Creole structural dataset. In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/54).

Borges, R. (2013). The role of extralinguistic factors in linguistic variation and contact-induced language change among Suriname's Coppename Kwinti and Ndyuka Maroons. *Acta Linguistica Hafniensia*, *45*(2), 228-246.

Calvin, W.H., & Bickerton, D. (2000). *Lingua Ex Machina: Reconciling Darwin and Chomsky with the Human Brain*. Boston: MIT Press.

Chaudenson, R. (2003). Creolistics and sociolinguistic theories. *International Journal of the Sociology of Language, 160*, 123-146.

Currie, T.E., Greenhill, S.J., & Mace, R. (2010). Is horizontal transmission really a problem for phylogenetic comparative methods? A simulation study using continuous cultural traits. *Philosophical Transactions of the Royal Soceity*, *365*, 3903-3912.

Daval-Markussen, A.,& Bakker, P. (2011). A phylogenetic networks approach to the classification of English-based Atlantic creoles. *English Worldwide*, *32*(2), 115-146.

Daval-Markussen, A. (2013). First steps towards a typological profile of creoles. *Acta Linguistica Hafniensia, 45*(2), 274-295.

DeGraff, M. (2003). Against creole exceptionalism. *Language*, *79*(2), 391-410.

DeGraff, M. (2005). Linguists' most dangerous myth: the fallacy of Creole Exceptionalism. *Language in Society, 34*(4), 533-591.

Devonish, H., & Thompson, D. (2013). Creolese structure dataset. In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/5).

Donohue, M., Musgrave, S., Whittin, B., & & Wichmann., S. (2011). Typological feature analysis models linguistic geography. *Language*, *87*(2), 369-383.

Dryer, M.S., & Haspelmath, M. (Eds.). (2013). The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info).

Enoch, A.O., Veenstra, T., & Smith, N.S.H. (2013). Saramaccan structure dataset. In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/3).

Farquharson, J.T. (2013). Jamaican structure dataset. In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/8)

Fattier, D. (2013). Haitian Creole structure dataset. In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/49).

Guthrie, M. (1948). *The Classification of the Bantu Languages*. London: Oxford UP.

Hagemeijer, T. (2013). Santome structure dataset. In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig:

Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/35).

Holm, J., & Patrick, P.L. (Eds.). (2001). *Comparative Creole Syntax: Parallel Outlines of 18 Creole Grammars*. London: Battlebridge.

Holm, J. (1988). *Pidgins and Creoles* (Vols. 1-2). Cambridge: Cambridge UP.

Huson, D.H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*, 254-267.

Huson, D.H., & Bryant, D. (2013). *User Manual for SplitsTree4 V4.13.1*.

Intumbo, I., Inverno, L., & Holm, J. (2013). Guinea-Bissau Kriyol structure dataset. In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/33).

Kouwenberg, S. (2013). Berbice Dutch structure dataset. In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/28).

Kouwenberg, S. (2013). Papiamentu structure dataset. . In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/47).

Ladd, R.D., Roberts, S.G., & Dediu, D. (2014). Correlational studies in typological and historical linguistics. *Annual Review of Linguistics*, *1*, 221-241.

Lefebvre, C. (Ed.). (2011). *Creoles, Their Substrates, and Language Typology*. Amsterdam: John Benjamins.

Luffin, X. (2013). Kinubi structure dataset. . In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/63).

McWhorter, J.H. (1998). Identifying the creole prototype: Vindicating a typological class. *Language*, *74*(4), 788-818.

McWhorter, J.H. (2001). The world's simplest grammars are creole grammars. *Linguistic Typology*, *5*, 125-166.

Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.). (2013). Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info).

Michaelis, S.M., & Rosalie, M. (2013). Seychelles Creole structural dataset. In: Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/56).

Mufwene, S. (1996). Creolization and grammaticization: what creolistics could contribute to grammaticization. In Philip Baker and Anand Syea (Eds.) *Changing Meanings, Changing Functions: Papers relating to grammaticalization in contact languages*. London: U of Westminster P. 5-28.

Mufwene, S. (2000). Creolization is a social, not a structural, process. In Neumann-Holzschuh, I. & Schneider, E. (Eds.), *Degrees of Restructuring in Creole Languages* (65-83). Amsterdam: John Benjamins.

Mühlhäusler, P. (1997). *Pidgin and Creole Linguistics*. London: U of Westminster P.

Nakhleh, L., Ringe, D., & Warnow, T. (2005). Perfect phylogenetic networks: a new methodology for reconstructing the evolutionary history of natural languages. *Journal of the Linguistic Society of America*, *81*(2), 382-420.

Nichols, J., & Warnow, T. (2008). Tutorial on computational linguistic phylogeny. *Linguistics and Language Compass*, *2*(5), 760-820.

Parkvall, M. (2006). Was Haitian ever more like French? In Deumert, A. & Durrleman-Tame, S. (Eds.) *Structure and Variation in Language Contact* (315-335). Amsterdam: John Benjamins.

Parkvall, M. (2008). The simplicity of creoles in a cross-linguistic perspective. In M. Miestamo et al. (Eds.), *Language Complexity: Typology, Contact, Change* (265-285). Amsterdam: John Benjamins.

Parkvall, M., Jansson F., & Strimling, P. (2013). Simulating the geneis of Mauritian. *Acta Linguistica Hafniensia*, *45*(2), 265-273.

Schnoebelen, T. (2009). A how-to guide for using phylogenetic tools on linguistic data
(SplitsTree, MrBayes). Stanford: Stanford UP.

Schwegler, A. (2013). Palenquero structure dataset. . In: Michaelis, S.M., Maurer, P.,
Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures
Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online
at http://apics-online.info/contributions/48).

Sluijs, R. van (2013). Negerhollands structure dataset. In: Michaelis, S.M., Maurer, P.,
Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures
Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online
at http://apics-online.info/contributions/27).

Taylor, D. (1971). Grammatical and lexical affinities of creoles. In D. Hymes (Ed.),
Pidginization and Creolization of Languages (293-296).Cambridge: Cambridge UP.

Thomason, S.G., & Kaufman, T. (1988). *Language Contact, Creolization, and Genetic
Linguistics*. Berkeley: U of California P.

Velupillai, V. (2013). Hawai'i Creole structure dataset. In: Michaelis, S.M., Maurer, P.,
Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures
Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online
at http://apics-online.info/contributions/26).

Winford, D. & Plag, I. (2013). Sranan structure dataset. In: Michaelis, S.M., Maurer, P.,
Haspelmath, M., & Huber, M. (Eds.) Atlas of Pidgin and Creole Language Structures
Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online
at http://apics-online.info/contributions/2).

Appendix A: Data Matrices

Data matrix from regression analysis

	early withdrawal of superstrate	prolonged contact with superstrate	maroonage	plantation creole	degree of creoleness
Jamaican	0	1	0	1	89
Hawaiian	0	1	0	1	67
Sranan	1	0	0	1	100
Saramaccan	1	0	1	0	100
Creolese	0	1	0	1	89
Berbice Dutch	0	0	0	1	83
Negerhollands	0	0	0	1	94
Haitian Creole	0	1	0	1	83
Seychellois	0	0	0	1	67
Mauritian	0	1	0	1	61
Guinea-Bissau	0	1	0	0	78
Palenquero	1	1	1	0	61
Papiamentu	1	1	0	1	67
Nubi Creole	0	1	0	0	78

Data matrix from first phylogenetic network analysis

	APiCS													CCS				WALS
	1	23	29	43	47	48	49	51	73	74	75	84	100	8.3	8.4	13.1	13.2	69A
Jamaican Creole	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0
Hawaiian Creole	1	0	0	1	1	1	1	0	0	0	1	1	1	0	0	0	1	0
Sranan	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
Saramaccan	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
Creolese	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0
Berbice Dutch	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1
Negerhollands	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0
Haitian Creole	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	0	0
Seychellois	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	0	0	0
Mauritian Creole	1	1	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0	0
Réunion Creole	1	1	1	1	1	0	1	0	1	1	1	0	1	?	?	0	0	0
Santome	1	1	1	1	1	1	1	1	1	1	1	1	0	?	?	0	0	0
Guinea-	1	0	0	0	1	0	1	1	1	0	1	1	1	1	1	1	1	0

Bissau																		
Palenquero	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	1	1	1
Papiamentu	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	0	0	0
Nubi Creole Arabic	1	0	1	1	1	1	1	1	1	0	1	0	1	0	0	1	1	0

Data matrix from second phylogenetic network analysis

	CCS															
	1.1	**2.2**	**3.1**	**4.2**	**6.1**	**7.2**	**8.3**	**8.4**	**10.1**	**11.2**	**12.1**	**13.1**	**13.2**	**13.3**	**14.1**	**14.2**
Berbice Dutch	1	1	1	1	0	0	1	1	1	1	1	1	1	0	1	1
Guinea-Bissau	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1
Haitian Creole	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	1
Jamaican Creole	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1
Nubi Creole Arabic	1	1	1	1	1	1	0	0	1	1	0	1	1	0	0	0
Negerhollands	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1
Palenquero	1	1	1	0	1	1	0	0	1	1	0	1	1	1	0	0
Papiamentu	1	1	1	0	1	1	0	1	1	1	0	0	0	1	1	1
Seychellois	1	1	1	1	0	1	0	1	1	1	1	0	0	0	1	1

Data matrix from third phylogenetic network analysis

	WALS						
	21B	**33A**	**38A**	**69A**	**81A**	**112A**	**119A**
Jamaican Creole	1	1	0	0	1	1	1
Hawaiian Creole	1	0	0	0	1	1	1
Sranan	1	1	1	0	1	1	1
Saramaccan	1	1	1	0	1	1	1
Creolese	1	1	0	0	1	1	1
Berbice Dutch	1	0	1	1	1	1	1
Negerhollands	1	1	1	0	1	1	1
Haitian Creole	1	1	1	0	1	1	1
Seychellois	1	1	1	0	1	1	0
Mauritian Creole	1	1	1	0	1	1	0
Réunion Creole	1	1	1	0	1	1	0
Santome	1	1	1	0	1	1	0
Guinea-Bissau	1	0	0	0	1	1	1
Palenquero	1	1	1	1	1	1	1
Papiamentu	1	1	1	0	1	1	0
Nubi Creole Arabic	1	0	1	0	1	1	1
English	1	0	0	1	1	1	0
French	0	0	1	1	1	1	0

Spanish	0	0	1	1	1	1	1
Portuguese	?	0	1	1	1	1	?
Dutch	?	0	0	1	0	1	0
Egyptian Arabic	0	0	0	1	1	1	0
Swahili	1	0	?	1	1	0	0
Fongbe	?	1	0	0	1	0	?
Arawak	?	0	1	1	1	1	?
Hindi	0	0	0	1	0	1	0
Ijo	?	0	0	1	0	0	?
Malagasy	1	0	0	1	0	1	1
Wolof	?	0	0	0	1	0	1
Pirahã	1	0	0	1	0	0	?
Vietnamese	1	0	?	0	1	0	1
Indonesian	1	0	0	1	1	0	1

Appendix B: Phylogenetic Networks

Bootstrapping analysis performed on the first phylogenetic network
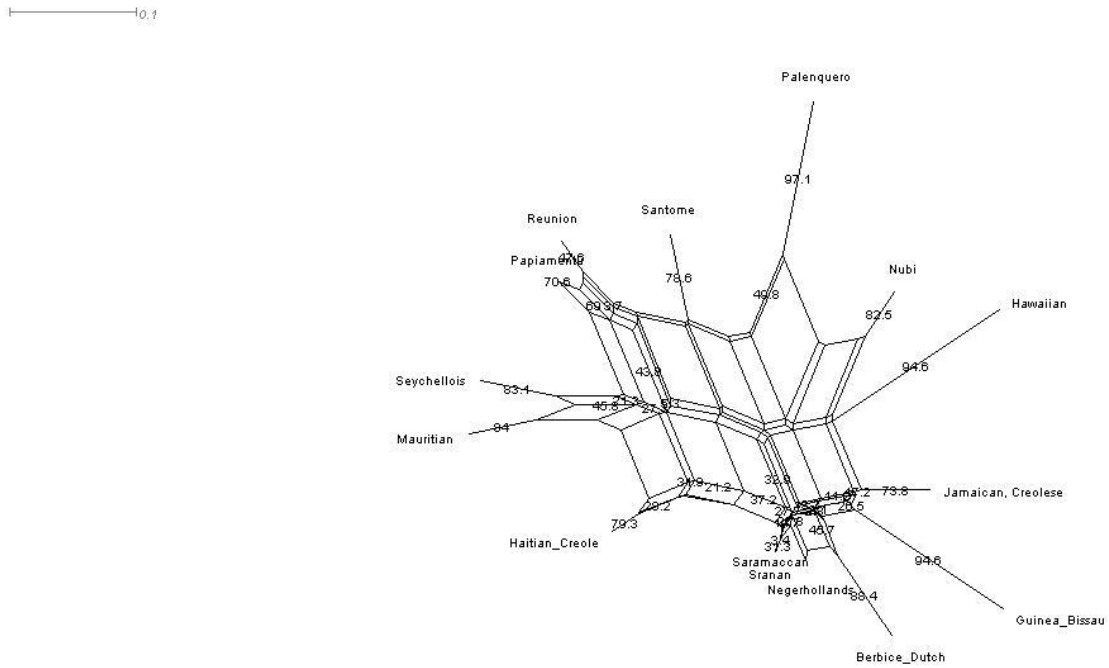
Bootstrapping analysis performed on the second phylogenetic network