Creole Studies – Phylogenetic Approaches
Creole Studies – Phylogenetic Approaches

Edited by

Peter Bakker
Aarhus University

Finn Borchsenius
Aarhus University

Carsten Levisen
Roskilde University

Eeva Sippola
University of Bremen

John Benjamins Publishing Company
Amsterdam / Philadelphia
# Table of contents

Preface ......................................................... IX

**CHAPTER 1**  
Introduction .............................................. 1  
*Carsten Levisen, Eeva Sippola and Peter Bakker*

**CHAPTER 2**  
Key concepts in the history of creole studies .......................... 5  
*Peter Bakker*

**CHAPTER 3**  
Phylogenetics in biology and linguistics .................................. 35  
*Finn Borchsenius, Aymeric Daval-Markussen and Peter Bakker*

**CHAPTER 4**  
Methods: On the use of networks in the study of language contact ......................... 59  
*Peter Bakker, Eeva Sippola and Finn Borchsenius*

**CHAPTER 5**  
Creole typology I: Comparative overview of creole languages ......................... 79  
*Peter Bakker and Aymeric Daval-Markussen*

**CHAPTER 6**  
Creole typology II: Typological features of creoles: From early proposals to phylogenetic approaches and comparisons with non-creoles ......................... 103  
*Aymeric Daval-Markussen and Peter Bakker*

**CHAPTER 7**  
West African languages and creoles worldwide .................................. 141  
*Aymeric Daval-Markussen, Kristoffer Friis Bøegh and Peter Bakker*

**CHAPTER 8**  
The typology and classification of French-based creoles: A global perspective ............ 175  
*Aymeric Daval-Markussen*
CHAPTER 9
The simple emerging from the complex: Nominal number in Juba Arabic creole
_Yonatan Goldshtein_

CHAPTER 10
Dutch creoles compared with their lexifier
_Peter Bakker_

CHAPTER 11
Similarities and differences among Iberian creoles
_Eeva Sippola_

CHAPTER 12
Afro-Hispanic varieties in comparison: New light from phylogeny
_Danae M. Perez, Sandro Sessarego and Eeva Sippola_

CHAPTER 13
Cognitive creolistics and semantic primes: A phylogenetic network analysis
_Carsten Levisen and Kristoffer Friis Bøegh_

CHAPTER 14
Lexicalization patterns in core vocabulary: A cross-creole study of semantic molecules
_Carsten Levisen and Karime Aragón_

CHAPTER 15
The semantics of Englishes and Creoles: Pacific and Australian perspectives
_Carsten Levisen, Carol Priestley, Sophie Nicholls and Yonatan Goldshtein_

CHAPTER 16
Feature pools show that creoles are distinct languages due to their special origin
_Peter Bakker_

CHAPTER 17
Complementing creole studies with phylogenetics
_Eeva Sippola_
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>From basic to cultural semantics: Postcolonial futures for a cognitive creolistics</td>
<td>381</td>
</tr>
<tr>
<td></td>
<td><em>Carsten Levisen</em></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Linguistics and evolutionary biology continue to cross-fertilize each other and may do so even more in the future, including in the field of creolistics</td>
<td>385</td>
</tr>
<tr>
<td></td>
<td><em>Finn Borchsenius</em></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Epilogue: Of theories, typology and empirical data</td>
<td>389</td>
</tr>
<tr>
<td></td>
<td><em>Bettina Migge</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Languages index</td>
<td>395</td>
</tr>
<tr>
<td></td>
<td>People index</td>
<td>401</td>
</tr>
<tr>
<td></td>
<td>Places index</td>
<td>407</td>
</tr>
<tr>
<td></td>
<td>Subject index</td>
<td>409</td>
</tr>
</tbody>
</table>
Preface

The book Creole Studies: Phylogenetic Approaches is the major outcome of the Cognitive Creolistics project, sponsored by the Velux Foundation between 2013 and 2015, Grant 31971. We are very grateful for the confidence that Velux had in our project, and the possibilities it has provided to explore innovative ways to study creoles.

The main investigators of the project were Peter Bakker, Finn Borchsenius, Aymeric Daval-Markussen, Carsten Levisen (now Roskilde University), and Eeva Sippola (now University of Bremen). The project was hosted by Aarhus University, and we are grateful for the support, office space, equipment and administrative help provided, especially by Gitte H. Nielsen in the final period. Jakob Feldtfos Christensen of Aarhus University Research Support Office was also instrumental in the success of the project. Institute leader Niels Lehmann was also very supportive, and he was always able to find solutions for sudden logistic changes. Finally we thank the Aarhus University Research Foundation (Aarhus Universitets Forskningsfond) for financial support that made this publication possible. The support provided by the Velux Foundation enabled us to make this book available to the whole world, including countries where creoles are spoken, by providing free open access to the contents of the book. The datasets underlying the studies are also made available at this site: https://phylogenetic-creole-studies.blogspot.com

Three researchers were connected to the project as consultants during some of the project’s lifetime, Karime Aragón, Robert D. Borges and Johannes Kizach. We received an incredible help from a number of assistants with collection and assessment of data and administrative tasks. We mention here Aline Mähr, Andrea Ariza, Benjamin Riise, Ehm Hjort Miltersen, Emily Jørgensen, Jeroen Willemsen, Johannes Kizach, Jonas Steno Olsen, Jonathan Puntervold, Karen Kiil Brøcker, Karen Nissen Schriver, Kasper Fynh Jacobsen, Kristoffer Friis Bøegh, Louise Orvokki Helander, Maria Jørgensen, Nynne Horn, Pernille Linde, Ottó Eliasson, Phaedra Kagiafa, Rane Schjøtt, Sergio González de la Higuera Rojo, Tobias Gretenkort, Vinije Haabo, Yonatan Ungermann Goldshtein. We also profited from the excellent work of three interns, Ana Paulla Braga Mattos, Ana Nely Çoban and Michael Nguyen. Phyllis Buchanan and Thomas Widmann copy edited the book and took care of many of the technical aspects. Daniel Bergman drew the maps. Mange tak, gracias, gaán tangi, merci, taing mhór, obrigado, thank you to all of you.
Numerous researchers, too many to mention, contributed to our ideas through the workshops and conferences that we attended or organized. The generous grant from Velux enabled us to organize a number of smaller and bigger workshops, such as *Time and Space in Linguistics: Interdisciplinary Computational Approaches & Cross-Creole Comparisons*, *Tenth Creolistics Workshop: “Innovations”* – with special attention to parallels between creole and sign language creation; *Language Ideologies and Music in Contact Situations; Semantics across English and Creoles: Australian and Melanesian Perspectives; Workshop Contact, variation and change. Focus on Afro-Latin Varieties*. These also led to publications elsewhere, including thematic volumes.

We were also able to discuss matters with the guests we hosted, Hugo Cardoso (Lisbon University, Portugal), J. Clancy Clements (Indiana University, USA), William Foley (University of Sydney, Australia), Cliff Goddard (Griffith University, Australia), Stéphane Goyette (Ottawa, Canada), Sophie Nichols (Griffith University, Australia), Mikael Parkvall (Stockholm University, Sweden), Danae M. Perez (University of Zürich, Switzerland, and University of Bremen, Germany), Carol Priestley (Griffith University, Australia), Britta Schneider (Free University, Berlin, Germany), Armin Schwegler (University of California, Irvine, USA), Sandro Sessarego (University of Texas at Austin, USA and FRIAS Freiburg, Germany), Sophia Waters (University of New England, Australia). Molly Flaherty (University of Chicago, USA) and Philip Baker (Colombo, Sri Lanka) were our special guests at the 10th Creolistics Workshop in Aarhus in 2015.

Finally, we would like to thank Miriam Meyerhoff (Victoria University, Wellington, New Zealand), Umberto Ansaldo (University of Hongkong), Bettina Migge (University College, Dublin, Ireland) and two anonymous referees for their critical comments on an earlier version of this book, which led to improvements of the text. We thank Kees Vaes of John Benjamins B.V. for his support and we hope that the publication will bear fruit for creole studies.

Peter Bakker, Finn Borchsenius, Aymeric Daval-Markussen, Carsten Levisen, Eeva Sippola
The Cognitive Creolistics Group
This book is a first step in a new exploratory way of doing creole studies with phylogenetic networks. As a computational tool originally developed for evolutionary biology, the use of phylogenetic networks has been exported to other disciplines, including linguistics, and this volume is the first book-length treatment of the phylogenetic approach to creole studies and contact linguistics.

The volume shows how phylogenetic tools can successfully help elucidate creole studies, and help us to test both old claims and new hypotheses. With a strong evidence base, the networks enable the researcher to visualize both similarities and differences. Graphs make relationships and categories visible. These include connections which were well known or only dimly suspected, or perhaps not even expected at all. Needless to say, the adaptation of phylogenetic networks requires a certain amount of conceptual and methodological engineering, as shown in this volume, since linguistic features are products of human history and collective creativity, rather than objective, natural features of the world.

For decades, creole studies have been uniquely positioned in the interface between linguistics and sociohistorical studies. This position has made creole studies a vehicle for innovations and debates in ways which have often challenged and rejuvenated more traditional paradigms. Today, creole studies have become part of the mainstream – there are creole textbooks, handbooks, courses, research groups, journals, societies, etc. The danger for any mature research field is that it loses its original vibrancy and forgets to be open to new interdisciplinary ideas. The approach in this book is thus unapologetically interdisciplinary. With phylogenetic methods and tools in combination with state-of-the-art linguistic typology and cross-linguistic semantics, the contributors to our volume study both contemporary and historical varieties. We investigate lexifier-creole relations, creole-creole relations, creole-substrate relations, and the relations between creoles and non-creoles. We showcase how phylogenetic networks can be utilized in contact situations in Africa, Australia, the Caribbean, India, the Indian Ocean, the Pacific, South America, and South East Asia. We provide new empirical evidence for a "creole
typology”, but we also unravel historical affiliations and areal groupings, and ini-
tiate a new semantics-based approach to creole lexicons.

The volume is based on work by “Cognitive Creolistics”, a research group which was based at Aarhus University 2013–2015. The group was led by Peter Bakker, and funded by the Velux Foundation. The core research team consisted of the biologist Finn Borchsenius (Aarhus), and the linguists Eeva Sippola (now Bremen), Carsten Levisen (now Roskilde) and Aymeric Daval-Markussen. In addition, three research students contributed to the book: Karime Aragón, Kristoffer Bøegh, and Yonathan Goldsthein. Project-external researchers and guests of the project have also contributed their invaluable expertise in several chapters: Sophie Nicholls (Western Sydney), Danae M. Perez (Zürich), Carol Priestley (ANU, Canberra) and Sandro Sessarego (Austin, Texas). Finally, Bettina Migge (Dublin) gave an outsider’s perspective on the whole enterprise. We also received valuable help from our student assistants, too many to mention here.

Creole Studies: Phylogenetic Approaches is united by a shared methodological curiosity, rather than a theoretical orthodoxy. We approach and discuss central concepts such as “creole”, “language”, “phylogeny”, and “feature” from a variety of assumptions and perspectives. All the chapters in this book adhere to a data-
oriented, empirical style of research and reasoning, and all chapters address spe-
cific research questions which are of great relevance to creole studies.

The chapters fall into in four main subsections. Chapters 2–6 address concep-
tual and methodological issues in creole studies, with a focus on “creole typology”, and the controversial question of “creole distinctiveness”. Linguistic typologists
and biolinguists will find this section thought-provoking, illuminative, and rich in empirical evidence.

In Chapter 2, Bakker provides an overview of key concepts, issues and discus-
sions in creole studies. This chapter is recommended for newcomers to the field
of creole studies.

In Chapter 3, Borchsenius, Daval-Markussen, and Bakker compare and con-
trast the use of phylogenetic methods in biology and linguistics, discussing method-
odological and conceptual developments in both fields, as well as new syntheses.
Readers with an interest in biology and linguistics, and the interface between these
two disciplines are referred to this chapter.

In Chapter 4, Bakker, Sippola and Borchsenius explain the methodological
practices and problems in the utilization of phylogenetic networks in this book.
Linguists and creolists without prior knowledge of phylogenetic networks will find this chapter a useful starting point.

In Chapters 5 and 6, Bakker and Daval-Markussen discuss how their datasets
led them to accept the existence of what they call the “creole typological profile”.
They discuss the structural properties of creoles, taking a broad geographical and
genealogical range as their point of departure. They conclude that no single structural property is unique to creoles, and no property is common to all creoles, but that despite their internal diversity, creoles form a typological group – when compared with non-creoles in a large-scale comparison. Chapter 5 provides a thorough synchronic review of proposed creole structures, whereas Chapter 6 focuses on history and diachrony.

Chapters 7–12 are devoted to the study of specific language groups, their interrelationships and origins. Historical linguists, including scholars of colonial and postcolonial linguistics will find these studies carefully crafted and highly relevant to the on-going work on classification.

In Chapter 7, Daval-Markussen, Bøegh, and Bakker investigate the common belief that the African languages which were involved in the creation of creoles were typologically similar. However, it turns out that the African languages in question were typologically diverse and heterogeneous when one takes stable structural features into account. It also appears that the African languages generally do not cluster with creoles, whether they were developed in the Atlantic or elsewhere. Based on these findings, the authors argue that substrate influence was not a determining factor in creole genesis in the more stable properties of languages.

In Chapter 8, Daval-Markussen sheds light on French-related creoles. In a quest for understanding both their interrelatedness and origins, he explores their degree of radicalness, i.e. their difference from French, and the areal features of French-related creoles in the Caribbean, North America, the Indian Ocean, and the Pacific.

In Chapter 9, Goldshtein studies Juba Arabic Creole of South Sudan in contrast and comparison with non-creole varieties of Arabic as well as Nilotic languages, including substrate languages, as well as a small sample of creoles. Focusing on the nominal number system, he finds that Juba Arabic clusters with other creoles rather than with Arabic or Nilotic languages. This leads Goldsthein to argue against the Feature Pool Hypothesis, and in favor of Creole Distinctiveness.

In Chapter 10, Bakker compares standard Dutch with Virgin Islands Creole Dutch, Berbice Creole and Skepi Dutch, based on lexical, grammatical, and phonological features. On the basis of this multi-faceted analysis, it is concluded that the three creoles were created independently.

In Chapter 11, Sippola investigates the Spanish-related and Portuguese-related creoles, which are the oldest and most geographically dispersed among the creoles of the world. The study supports previous classifications, in which a major division between Atlantic and Asian creoles is encountered. She then turns to investigate the more intricate aspects of the interrelationships, including the role of adstrates, substrates, and areal linguistics.
In Chapter 12, Perez, Sessarego and Sippola study Afro-Hispanic varieties in South America, with a special focus on the classification of Afro-Yungueño spoken in Bolivia. Taking on the dialect versus creole debate, they explore how phylogenetic networks can help nuance the discussion on creoles, post-creoles and dialects, emphasizing the importance of interpretation and the perspectives resulting from feature selection.

Chapters 13–16 study creole lexicons from a semantic perspective. Lexical typologists, cross-linguistic semanticists, and sociolinguists with an interest in metalanguage, will find these chapters inspiring and innovative.

In Chapter 13, Levisen and Bøegh compare the expression units for basic semantic categories across creoles and lexifiers. They find that the expression of basic categories is more different between the major colonial languages and their lexifiers than expected. However, they also find that, from a large-scale perspective, creoles, in their core lexicons, do cluster with their lexifiers, but that English-related creoles make up the most heterogeneous group.

In Chapter 14, Levisen and Aragón explore the expression of different semantic domains, using semantic molecules for feature selection. Basic words from the domains of society, environment and the human body, as well as abstract terms are compared. The study indicates that domains generally differ in closeness and distance, and that the social domain stands out as the area where creoles differ the most from the colonial lexifier languages.

In Chapter 15, Levisen, Priestley, Nicholls and Goldshtein draw lexical networks of new lects and languages in the Pacific in order to explore their contemporary relationships, but also to discuss and deconstruct the metalinguistic terminology of what is meant by “language”, “dialect”, “variety”, “creole”, and “English”. They demonstrate how the combination of phylogenetic network techniques and semantic primes can be used to circumvent and nuance the current sociolinguistic debate on metalanguage.

The volume ends with five evaluations of the results in Chapters 16 to 20. Bakker discusses the results from a perspective of a linguistic typologist, Sippola as a creolist, Levisen as a semanticist and Borchsenius as a biologist. They provide concluding comments, and further reflect on the prospects and problems for further use of phylogenetic methods in creoles studies, contact linguistics, and beyond.

Finally, in her invited review of the volume, Migge critically engages with the findings, the method and the scope of the work, giving also suggestions for future work in Chapter 20.
CHAPTER 2

Key concepts in the history of creole studies

Peter Bakker
Aarhus University

This chapter serves as a brief introduction to creole studies. It mostly deals with issues that are discussed in the book, but not exclusively. It starts with an introduction of terminology, and key terms are printed in CAPITALS. The second part presents a number of observations and claims that have been made more or less explicitly in creole literature. Some key sources are given, and arguments are discussed briefly. This background knowledge should enable the reader to interpret the subsequent chapters, and take an informed position on the basis of the results presented there. We end by discussing a number of issues that could become important in the future of creole studies.

2.1 Introduction

In order to understand this book in full, an introduction to Creole Studies is needed. We will first give a very short history of creole studies, where we will also introduce a number of important terms (2.2). Thereafter we will give an overview of the main issues in the creole studies of the past fifty years, where some central concepts, ideas and terms will be introduced. We will focus on structural characteristics, including complexity, and sociohistorical discussions. This will include looking at the existence or non-existence of a pidgin before a creole, who the creators of the creoles might have been, where the process took place, and whether it happened gradually or quickly. Finally, the sociohistorical and linguistic properties will be linked, before concluding the chapter. The following chapters of the book will shed light on many, but by no means all, of the issues dealt with here, and this is discussed towards the end.
2.2 Creole studies

As a field of linguistics, creolistics is especially important because it builds bridges between many other branches of science. Creole studies has ramifications in the social sciences, including historical research, migration studies, post-colonial studies and social anthropology. Within linguistics, several fields are relevant, especially philology, typology, psycholinguistics, language acquisition, computational linguistics, historical linguistics, theoretical linguistics, socio-anthropological linguistics and applied linguistics. Creole language documentation dates back to the 1600s and 1700s. The first grammar of a creole, Virgin Islands Creole Dutch, was printed in Danish in the late 1700s (Magens 1770). The members of the pietistic religious group called the Moravian Brethren, in particular, left rich written sources on creoles from the early 1700s, among others from the Danish West Indies and Suriname. Research on creoles as a subdiscipline of scientific linguistics goes back to the late nineteenth century, when scholars like Lucien Adam, Adam Coelho, D.C. Hesseling, Addison van Name, Auguste de St-Quentin, Hugo Schuchardt, and Leite de Vasconcelos had developed an interest in language contact and studied the circumstances under which creole languages emerged and developed (Holm 1988; Meijer & Muysken 1977; Krämer 2013, 2014).

The traditional view of pidgin and creole genesis (as laid out for example in Mühlhäuser 1997; Romaine 1988) is that pidgins came about in contact situations where groups of people without a common language had to communicate with each other, and crude, simplified contact languages were created, first quite individual and normless jargons, later more systematic and collective pidgins (on pidgins, see Parkvall & Bakker 2013). When these simplified languages had to be used for other functions and when they were employed as languages of wider communication, or became mother tongues (nativization), the limited structures and lexicons of pidgins were expanded, thus developing into languages of full communicative functionality. Those fully developed languages are called creoles and the process of the development of jargons or pidgins into creoles is called creolization. Creoles are almost always native languages or mother tongues, in contrast to pidgins. Pidgin-derived contact languages that are spoken as both second and first languages are sometimes called expanded pidgins or pidgincreeoles (Bakker 2008), and due to their functional and structural expansion they are quite similar to creoles (e.g., Tok Pisin, Nigerian Pidgin English, Sango, Plains Indian Sign Language).

Some languages show significant simplification with regard to their source language, but not so much innovation and expansion. These are sometimes called semicreoles or creoloids or restructured varieties (Holm 2003), and Afrikaans and the language of Réunion are among them. Also Vernacular Brazilian Portuguese and Lingala may belong in this class.
One important strain of research in creole studies is the reconstruction of the generation of creoles. Linguistically, one can distinguish phases of reduction and/or simplification (in pidginization) and expansion and grammaticalization (in creolization). The reduction phase, if not documented, can be inferred from the fact that, for instance, nominal gender and number marking, as well as verbal inflection are virtually never inherited from the lexifiers in creoles. This can best be explained by an earlier simplification stage, as many of the lexifiers’ grammatical categories do not find their way into the pidgins (Bakker 2003; Roberts & Bresnan 2008), so the absence in creoles is explainable through pidginization.

In recent years, some creolists have tried to discredit the scenario of simplified pidgins developing into complex creoles (Chaudenson 1992; Mufwene 2000, 2008; Aboh 2015). They find no evidence for a pidginization phase for some creoles, but their claims are rarely, if ever, backed up by historical or linguistic data. The process from pidgin to creole has been documented, for instance, for Pacific Englishes (Tok Pisin), Chinook Jargon in North America, and Sango in Central Africa, but not in the Caribbean or in Indo-Portuguese creoles. As the existing creole languages like Tok Pisin (documented pidgin stage, 1800s) and Haitian Creole (no documented pidgin stage, 1600s) are quite similar in structure, some creolists have assumed a process of simplification and/or pidginization also in the Caribbean, the Indian Ocean and elsewhere. Thus, some of the shared linguistic properties of creoles can then be explained on the basis of an assumption of a pidgin stage, even when a pidgin stage is not historically documented.

Creoles came about in situations where speakers of several languages were brought together, often in rather adverse circumstances connected with colonialism and expansionism. From a geographical point of view, most creoles are spoken in areas of European expansion fairly close to the equator, from where tropical products were extracted. Some creoles arose in the context of religious missions, others emerged with colonial enterprises, indentured labor, plantation economies and similar forms of economic expansion. In some cases, creoles arose in multilingual urban settings, including multi-ethnic schools (Roberts 2005 for Hawai’i, Ehrhart & Revis 2013 for Tayo, New Caledonia, and Volker 1982 for Unserdeutsch in Papua New Guinea).

In creole studies, a number of regions are often distinguished on geographical grounds (see Map 2.1). The creoles in these different regions may also be influenced by several distinct languages, e.g. African, Dravidian, or Austronesian, and they often also show similarities within the region. The regions are: the Atlantic (split into Gulf of Guinea, West Africa, and Upper Guinea in Africa, and the Caribbean, the Guyanas and Central/North America in the Americas), the Indian Ocean, South Asia/India and finally the Pacific, including Melanesia, Northern Australia and South East Asia, including the Philippines. A few isolated creoles are also known, such as Palenquero Creole Spanish in Colombia and Nubi Creole Arabic in East Africa.
The basic lexicon of creole languages is typically (but not always) derived from a colonial language, including dialects of a European language. The language that is the source of the vocabulary is called the lexifier or superstrate. The lexifier is often included in the name of the creole language, e.g., Jamaican Creole English. Perhaps as many as 100 creoles are based on Arabic, Dutch, English, French, Portuguese and Spanish. There are several creoles with these European lexifiers in different parts of the world, and also a few with non-European lexifiers such as Yilan Creole Japanese, Grand Ronde Chinuk Wawa, and Sango in Central Africa.

The grammatical structure of a creole often deviates considerably from that of its lexifier. Creolists disagree about the reasons why there are differences between creole and lexifier structures. Some emphasize that creoles are continuations of the lexifiers (e.g. Chaudenson 1978; Mufwene 2008), and those who adhere to this idea are sometimes called superstratists in the literature. Others stress the influence of speakers of other languages (e.g. Holm 1988; Lefebvre 1998). For instance, Papuan or Oceanic languages in the case of Pacific creoles, Philippine languages in the Philippines and the mother tongues of Africans in the case of the Atlantic creoles. Such languages are called substrates if they are no longer spoken in the creole-speaking communities, and adstrates if they are present in the community even after creoles developed. In this book also the innovative term constrate is used, as a cover term for all languages spoken by the people in the development of a creole, viz. the superstrate, adstrate and substrate languages. In the Atlantic contexts, the substrate languages were mostly Bantu, from Central coastal Africa, but there were also significant numbers of speakers of West African languages, such as Kwa languages, including Gbe languages like Fon and Ewe. With regard to the Caribbean, discussions of substrates dominate, whereas in
Asian and West African contexts, it is not easy to distinguish between substrates and adstrates (see e.g. Smith 2012; Corum 2015).

The expansion of reduced, makeshift languages into fully fledged languages can be adduced from the striking structural discontinuity between the lexifiers and the creoles, especially the most radical ones. Observed grammatical similarities of creoles across lexifiers have been the focus of attention of creolists since the 1970s, and they have been linked to universals (Bickerton 1984), to substrate influence (Lefebvre 1998), to psycholinguistic effects in the creation of grammatical systems, influence from second language acquisition (Becker & Veenstra 2003; Mufwene 2010; Plag 2011) to population movements (Mufwene 2008) and to other factors.

The influence of substrate languages can manifest itself in different areas of the creole languages. Substrate influence can be observed in morphosyntactic structures, e.g., word order patterns or reduplication in morphology (Aboh & Smith 2015), or in phonology, notably in the presence of certain phonemes alien to the lexifier, certain types of vowel harmony or phonotactic patterns. Substrate influence is also found in the lexicon, e.g., the words for cultural inheritance and natural phenomena are often transmitted from African languages in Atlantic creoles (Bartens & Baker 2012; Parkvall 2000). A number of structural features found in creoles may be due to what is sometimes misleadingly called language universals. If traits are universal, in that they are found in all languages, they are obviously present in creoles as well as non-creoles. There may be relevant statistical typological universals (Velupillai 2012, Chapter 2), however, in which (very) high proportions of languages display certain properties. Here creoles may show different preferences compared to non-creoles, displaying different sets of typological properties. There are also patterns of change that may differ: so-called diachronic universals (going back to Greenberg 1966) such as patterns of grammaticalization, may lead to structural results which differ among creoles compared to non-creoles.

Creolists debate the relative weight of the influence of the lexifiers, the substrates, universals and innovations (Michaelis 2008). In the past, some took extreme positions and categorically denied, for instance, any influence of substrate languages (e.g. Bickerton 1986), or any form of innovation (Aboh 2009), but currently the vast majority of creolists interested in the genesis of creoles accept a multitude of influences. Probably a large majority of creolists support the influence of substrates (see Lefebvre 2011 for a recent overview).

Some scholars have emphasized that the languages of the earliest arrivals of immigrants often influenced the creoles more deeply. The founder principle in creole genesis (Mufwene 1996) stipulates that a selection advantage was proffered to founder populations in colonial contexts. This approach is connected to the feature pool hypothesis. The feature pool consists of the properties of language
varieties involved in a concrete contact situation put together. Creoles would select properties from this feature pool (Mufwene 2001, 2002; Aboh & Ansaldo 2007). The idea was inspired by Ian Hancock’s original componentiality hypothesis (1986). The feature pool hypothesis seems to leave no room for structural innovations, despite the fact that newly-formed grammatical constructions have been observed in great numbers in all creoles, for instance new ways of expressing tense and aspect, or relative clauses (see e.g. Givón 1982; Kuteva & Comrie 2012; Bakker 2016). The feature pool idea has been criticized by a.o. Plag (2011) and McWhorter (2012) for being contrary to observed properties of creoles and unfalsifiable.

Many creolists have not only relied on linguistic data, but also on a wide array of other sources. Studies on the genesis of creoles ideally make use of historical and demographic data in order to pinpoint the origin of slaves. Earlier texts from creole communities have been studied, using methods from philology.

Demographic disproportion is often invoked as a crucial factor in creolization. For instance, it is sometimes suggested that slaves in the earlier phases of the settlement lived in a more intimate relationship with the Europeans, in what is called société d’habitation in the French tradition, or homestead phase, where servants and slaves would acquire close to perfect French. But in the société de plantation or plantation society, when the masters had become a minority, the lexifier became less accessible to each new slave population. The slaves learnt an approximation of an approximation of the superstrate or lexifier, and this would explain the deviant structures observed in creoles compared to the lexifiers (Chaudenson 1978).

It has been suggested that when the shift from a homestead phase to a plantation phase occurred, the slaves did not have enough access to French, Dutch etc., because there were too few speakers of these languages in relation to the total population. Thus, speech forms that were increasingly different from the lexifiers emerged in the different phases.

Researchers have also suggested that the speakers of the lexifiers could have spoken in a simplified manner, providing foreigner talk as input, rather than the colloquial varieties spoken among themselves. Foreigner talk is the way people speak to others who do not share their language, purposely simplifying their language (Ferguson 1971; Clements 1992).

For a long time, creolists considered creoles to be a result of the failed attempts by subjugated populations to acquire the lexifier language. The assumption was at the time that the lexifier was the target language for people. Since Baker’s (1990) article, aptly called “off target”, most creolists no longer believe that the lexifier was the target language for the displaced or indigenous communities. Rather, they wanted to create a means of interethnic communication, without having to worry whether the utterances were grammatically correct from the perspective of
the lexifier. Successful communication was the main goal: obtaining food, trade goods or a variety of services was more important than acquiring the other’s language to perfection. This would help explain why creoles have remarkably different grammatical systems compared to those of the “target”.

Different types of creoles can be distinguished on the basis of the historical circumstances under which they emerged. Some creoles developed around trading sites, where Europeans often built fortified communities for trade with the local population. Creoles that developed locally, for instance around trading forts in West Africa between Europeans and indigenous Africans, are called endogenous creoles (Greek endo “inside, within”), and those that involve large numbers of displaced workers, often slaves or indentured laborers, are called exogenous (Greek exo “outside”) and quite a few of them are spoken in insular settings (Chaudenson 1978). Creoles that developed around forts (e.g. in West Africa) are called fort creoles. Other creoles developed in multiethnic workforces, often in agricultural societies for which labor was imported from other parts of the world, either through slavery or indentured labor. Most Caribbean and mainland American creoles have formed as so-called plantation creoles. In several cases, however, maroons (escaped slaves) fled the plantations and established autonomous communities. These creoles are called maroon creoles, and those communities that were established early in the history of the colonies tend to speak creoles that are more distant from their lexifiers than for example fort creoles are. Maroon creoles are mainly spoken in the Guyanas, especially in Suriname, and insular West Africa. Furthermore, there are creoles that emerged in educational environments, or in armies with soldiers from diverse linguistic backgrounds. These can be called respectively urban school creoles (for instance Nicaragua Sign Language, Rabaul Creole German, perhaps also Hawaiian Creole English and Tayo in New Caledonia) and army creoles (for instance the Arabic creoles of East Africa).

Histories of languages and their connections can also be reconstructed by studying the diffusion of particular words, or forms of words, from one language to another, or from one geographical region to another. Such words can be inherited from an earlier common source language, or borrowed. Structural similarities can be explained by parallel and independent generalizations, but for some very specific forms of words or idioms, especially if they are numerous enough, it can be argued that these specific forms derive from one and the same source. Philip Baker and others have studied this especially for the English and French creoles, and for the specific forms of African words found in the Atlantic creoles (Baker 1993; Baker & Huber 2001). Word forms that are common between creoles spoken on both sides of the Atlantic have led scholars of English creoles (e.g., McWhorter 1995; Smith 1987) and of Portuguese creoles (e.g., Jacobs 2012)
to suggest a common origin for several creoles with the same lexifier. One can attribute the presence of some common quirks to migrations of persons travelling between different regions of the world (Speedy 2013; Baker & Mühlhäusler 2013). Certain creolists who believe in a common origin of creoles from one lexifier in a specific region (e.g., Atlantic English, South Pacific English, Indian Ocean French) support the idea of monogenesis, while others posit the occurrence of multiple independent geneses, or polygenesis. In the 1960s and 1970s, a strong monogenesis theory was discussed (e.g. Whinnom 1965; Taylor 1971) in order to explain perceived similarities between creoles across lexifiers, with one origin of a creole and subsequent relexification (i.e., the massive replacement of vocabulary but not of grammar). The idea that all creoles regardless of lexifier have a common origin is now universally rejected.

2.3 Issues in creole studies

In this section we are going to introduce some of the main issues in modern creolistics by making a number of statements about the nature of creoles. We will mention some of the main protagonists and antagonists and their arguments. We will start with general issues such as the characteristics of creoles, sociohistorical issues, pidgins, the influence from other superstrates and adstrates, complexity, the agents of creolization, etc.

A small selection of references will be given in almost all cases. These references are not exhaustive but point to some publications that have contributed to the issues.

2.3.1 General characteristics

According to some, creole languages have some specific properties that distinguish them from other languages. Scholars have discussed the social, historical and also linguistic aspects of these, and the connections between them.

1. Creoles are natural languages, just like any other language

This statement is uncontroversial among linguists. The expressability of natural languages entails that any thought can be expressed in any language, and that includes creoles. Of course the vocabulary of any language will have gaps in that some languages can use one word where other languages need circumlocations, compounds, phrases or sentences, but the vocabulary can easily be adjusted. Due to some sociohistorical coincidences, and not because of linguistic restrictions, some languages fare better for discussions of medicinal plants in the rainforest,
others better for quantum mechanics, as some of its speakers have developed an interest in such matters. All natural languages can easily expand the vocabulary to express everything, and the grammatical means to do so.

The same is true for grammatical properties. Many grammatical distinctions that many people may think are present in all languages, may be non-existent in almost all other languages, or be expressed in more detail in others. There are languages without the word “to have” (for instance Russian, Finnish and Diu Creole Portuguese), without distinct pronouns for first and second person plural (Haitian Creole, English), without a past tense (40% of all languages according to Dahl & Velupillai 2013), without plural endings (10% according to Haspelmath 2013a), and without verbal conjugations (15% according to Dryer 2013). If creoles lack distinctions that are present in their lexifiers, that observation makes them neither richer nor poorer.

2. Creoles are the result of retention, loss, and reconstitution

The three keywords that define creolization may be retention, loss, and reconstitution. Part of the lexicon and part of the grammar are retained from a lexifier, but there is a clear loss of properties as well. Processes of simplification and reduction have initially led to the formation of rather basic codes of communication. The results were clearly insufficient for broader patterns of communication, and hence communicative shortcuts through grammaticalization must have been created to make an efficient communication system, and a complete and fully natural language. Semantic distinctions were reintroduced into the emerging creole – and these distinctions were different from those found in the lexifier. For instance, definite and indefinite articles from the lexifiers are lost, and demonstratives and the numeral ‘one’ are retained and were also used for the reconstitution of definite and indefinite articles (Heine 1997). For instance, in Saramaccan wan is used for specific and unknown (< English one), di for specific and known (< English this). The articles in many creoles may distinguish specific/non-specific and known/unknown contrasts rather than (in)definiteness.

3. Creoles are alike, simple and mixed

It has been claimed that creoles are similar to one another, that creoles are complex but less complex than most languages of the world, and that creoles are mixed in that they combine lexicon and grammar from other languages (cf. Muysken 1988; McWhorter 2001; Aboh 2015). All three claims have been under constant discussion. Until the 1990s the statement that creoles were ALIKE, i.e., SIMILAR TO ONE ANOTHER, was probably the least controversial. Taylor (1971), Whinnom (1965) and others had noted and described striking similarities between French, English, Spanish and Portuguese creoles, later also Dutch and Arabic creoles.
Comparing creoles, they observed, for instance, a preverbal system of tense/mood/aspect markers, the expression of nominal plurality by a third person pronoun, etc., which are all quite different for the lexifiers. They also noted that these languages had not inherited much of the morphology and close to none of the irregularities of the lexifiers.

Creole languages have very complex grammars, but it has been argued that they are less complex than other natural languages because of their special history (e.g. McWhorter 2001; Parkvall 2008). Some people have called creoles simple, or simplified (but see e.g. DeGraff 2003 for a critique), and from the perspective of the lexifiers, one can observe lexifier material that is not echoed in the creole. Due to their special history, they did not acquire either unnecessary or decorative features, or complex morphophonology or tone, all of which take time to develop.

Finally, it has been claimed that creoles are mixed (e.g. Thomason 2001; Aboh 2015). Creoles have been formed through contributions from several languages, often called superstrates or lexifier, substrates and adstrates (see above). Some scholars find creoles less mixed as they display fewer loans than the lexifiers. The lexicon of a creole has its origin mostly in one language, the lexifier. There are a few special cases of creoles with mixed lexicons, with many roots from languages other than the main lexifiers. We should perhaps call them secondary lexifiers, and not substrates in these cases. The three examples of such languages are: Angolar (Portuguese and a Bantu language Kimbundu), Saramaccan (English and Portuguese), Berbice Creole (Dutch and Eastern Ijo, a Nigerian language), perhaps also Palenquero (Spanish and Bantu). Most mixed-language creoles are spoken by descendants of maroons.

The other creoles have an overall lexicon that is less mixed than what is found in non-creoles, not only in languages heavily shaped by contact such as English, but also languages like Swedish or Japanese. Overall, some 70% of the English vocabulary in a middle-sized dictionary is borrowed from Latin and Romance language, but only 7% of the basic lexicon (Grant 2009). In creoles, the proportion of non-lexifier lexicon is probably less than 5% in the overall vocabulary. Generally, upwards of 95% of the lexical and grammatical roots of a creole can be traced to one language, called the lexifier.

The mixture may also be associated with a dichotomy between the lexicon and the grammatical system: almost all the basic vocabulary, and indeed the lexicon in general, comes from one language, but the grammatical system is assumed to be from another source, thus making a creole a mixed language. This idea is inherent in the relexification hypothesis, now called relabeling, see below under (11) (Lefebvre 1998, 2014).
4. **The definition of “creole language” is circular**

It is not easy to define a creole language (but see McWhorter 2005). One cannot take the folk label “creole” to define creole languages, as only a minority are in fact called creoles by their speakers. Some communities call their languages pidgins, but they are in fact creoles (e.g. Tok Pisin (“Talk Pidgin”), Hawaiian Pidgin English, Solomon Islands Pijin, Nigerian Pidgin, Fernandoo Poo Pichi). On the other hand, some population groups are called creoles (e.g. Alaskan Russians, Aleuts of the Commander Islands), but their language is not classified as a creole. If one takes certain social or demographic criteria as a point of departure, e.g. new languages spoken by displaced populations, then it is easy to see that this is not enough: we call Cuban Spanish a dialect of Spanish, even though it is spoken mostly by people whose ancestors did not have Spanish as a mother tongue when they arrived on the island (cf. Haspelmath 2013b). French is spoken by descendants of Celtic and other languages, which has left traces as substrate features, but French is not called a creole.

If we call certain languages creoles when they display a large number of changes in phonology, lexicon and syntax in comparison with their nearest (lexical) relatives, then this reasoning is circular: the restructuring and deviant nature within the group of languages with the same lexifiers, is already part of the definition of the creole. In other words, a creole is in principle a language associated with certain sociohistorical events, but if the linguistic structures after such events do not conform to ideas about what a creole language is supposed to be like, we do not call a language a creole – despite a certain social history shared with situations in which creole languages came into being. And that is circular.

There are several potential ways to avoid this circularity. First, one can simply take the existing set of languages generally recognized as creoles by specialists. This would then be a closed set, but generalizations are limited, and expansion of the set is impossible. A second way to avoid circularity would be to isolate some properties that distinguish recognized creoles from non-creoles – a contested idea. Finally, one can try to define a creole as a language that shares its lexicon with other languages, but not its grammatical system, which should be quite different from the other languages of the family, and not identical to other languages, and visibly grammaticalized (see Cabral 1995 and DeLancey 2014 for some languages in South America and Asia). None of these methods are generally accepted. Defining what a creole is remains difficult.
2.3.2 Sociohistory of creoles and creolization

Creole languages are often associated with specific socio-historical events, such as colonization, expansion, population displacement and slavery. Here we discuss the most conspicuous links between properties of creole languages and socio-historical situations that creoles have been associated with.

5. It is impossible to distinguish creoles from non-creoles on the basis of the structure alone, without knowledge of the sociohistory of a language community

Where we have creole languages, a set of social circumstances have played a role (see discussion in the previous section). There is, in other words, no doubt that creoles can be associated with certain sociohistorical events. Some scholars say that one cannot tell the difference between a creole language and a non-creole language based on structural traits alone (Ansaldo & Matthews 2007, e.g. p. 4, 14, DeGraff 2001, 2003, 2005a, b, 2009; Mufwene 2001). For example, Ansaldo & Matthews claim that “creole languages do not form a typological or otherwise structurally unique class distinguishable from non-creole grammars” (2007: 14) and DeGraff states that “Creole grammars do not form a typological class that is aprioristically and fundamentally distinguishable from that of non-Creole grammars” (2009: 137). For them, there is no class of creoles, just a set of languages with certain histories that are called creoles.

6. Certain sociohistorical circumstances have an effect on language structure, and specific circumstances can lead to the emergence of creoles

Multilingual workforces may lead to the emergence of a pidgin, especially when the different parties want to maintain a social distance. Languages spoken by few people have fewer phonemes (Hay & Bauer 2007) Languages learned by many as a second language will show the effects of simplification (Kusters 2003; Lupyan & Dale 2010). Languages spoken by isolated populations will grow in complexity (Trudgill 2009, 2011). In a similar vein, certain sociohistorical circumstances that are responsible for the emergence of creole languages can have an effect on structure. These factors seem to be the development of a new society, where communication in the standard language is not desirable or not possible, and where therefore an in-group language develops which becomes a shared language. This can trigger a development in which the medium of interethnic communication becomes the default means of communication in the community, in some cases even a mother tongue.
7. The ecology of each situation of each creolization event was different, and therefore requires extensive data collection from historical sources

It is uncontroversial that each situation of creolization was unique. The ecology is always different, but this statement should not lead us to abandon the search for historical data and linguistic facts. Mufwene (2001) points out that the ecology of the genesis of all creoles was different, and therefore it is impossible to generalize. With ‘ecology’ he means the socio-historical circumstances in which creoles develop, including the presence of different languages with different levels of prestige. Chaudenson (1978) has proposed that subsequent generations tried to approximate earlier generations’ speech, without backing the claims up with documental evidence. These approaches are especially futile if documentation is available, even if fragmentary. Baker and Fon Sing (2007), for instance, collected early texts from Mauritius, Arends and Perl (1995) texts from Suriname, and van Rossem and van der Voort (1996) for Virgin Islands Creole Dutch. Even though historical and linguistic sources may be scarce and defective, any theory on creole genesis should be compatible with the available facts, and with as many facts as possible that can be unearthed.

8. Creoles are created as an act of identity

Once different groups have been gathered together, perhaps far away from their homeland, the second generation of locally-born populations may form a new identity. The locally-born do not identify with the indigenous populations (if present) and neither with the colonial powers. Creole languages may be interpreted as an expression of such a new, local identity (LePage & Tabouret-Keller 1985; Baker 1990; Muysken 1981).

9. Creoles are or were spoken by suppressed populations

Some of the world’s creole languages are spoken by populations whose ancestors were deported to other locations as slaves or as indentured laborers. In many societies their descendants have now, after the colonial period, obtained independence. Most creole populations now live in postcolonial settings, and the status of their languages range from dominant in all domains (Papiamentu in the Dutch Antilles) via widely spoken but not visible or official (Saint Lucia French Creole) to moribund (Diu Creole Portuguese, Grenada French Creole). Bislama, the English creole of Vanuatu, is an official language but not widely accepted as such.
2.3.3 Development: Pidgin stage or not

The existence of pidgins that would have influenced creoles is contested.

10. *A creole does not emerge without a preceding pidginization stage*

Before the 1990s, a pidgin stage before a creole was uncontroversial (Mühlhäusler 1997; Romaine 1988). Until the 1980s, scholars of pidgins and creoles agreed that pidgins did not have full expressability, whereas creoles were complete and mature languages, and people agreed that a pidgin could become a creole through expansion. More recently, a pidgin stage preceding a creole has become a controversial issue. People have argued in different veins that some creoles were not preceded by a pidgin stage or they denied the existence of a stage in which there was some kind of a reduced version or approximation of another language which became a creole (Bollée 1977; Mather 2007: 410; Thomason & Kaufman 1988).

In the Pacific, successive stages of documentation (e.g. Mühlhäusler 1979 for Tok Pisin) clearly show that a rather simple communication system was used in initial stages, which was variable and extremely simplified. In the Americas and the Caribbean, however, where (pidginization and) creolization took place centuries earlier than in Melanesia, proof of a pidgin stage is often lacking. Therefore, the proof for a pidgin stage is speculative when documentation is lacking, and mostly argued on the basis of indirect evidence.

11. *Creoles emerge without a pidgin*

Three strains of creolists have proposed that creoles emerged without a pidgin. First, creolists like Arends (1989) and Hazaël-Massieux (2008) have been looking at historical documentation in locations where today a creole is spoken. Documentation of the language may show a development away from the lexifier to a more creole-like language. Opponents, however, have said that the documentation of earlier stages does not document the “deep” creole or the “basilect” but displays a variety affected by missionaries, by literacy or by normative views of the languages. Others have claimed that creoles are direct relexifications of the substrate languages, and hence a pidgin was not part of the history of such languages (Lefebvre 1998). Again others have claimed that creoles are the result of recombination of features from two or more languages (Mufwene 2008; Aboh & Ansaldo 2007; Aboh 2015), and hence a pidgin is unnecessary to explain the properties of creoles, as all their properties are argued to be traceable back to the contributing languages.
12. *A creole is a nativized pidgin*

In connection with the preceding paragraphs, one can also claim that every creole language must be a continuation of a pidgin. Here the definition of “pidgin” is important. For some creolists, a pidgin is a rather stable system, which has to be learnt, and one can be a good or bad speaker of the pidgin (Parkvall & Bakker 2013). Other creolists (notably Bickerton 1981) have assumed a very chaotic system of communication with no rules or stability for the stage preceding creolization, more akin to a jargon. Both views agree that creoles build on incomplete communication (pidgin), which was not designed to be a fully effective form for communicating. Both views of pidgins assume simplification of lexifier properties in the pidginization process.

2.3.4 Influences from input languages

The exact contributions of the lexifiers, substrates, adstrates and dialects of the lexifiers are still under debate (see also the statements under (3) above).

13. *A creole language is a structural continuation of an imported or indigenous language, where only the word forms have been replaced by words from another (mostly European) language*

One theory on the genesis of creoles entails that creoles are, for instance, African languages with a European coat on. The inner form or deeper meaning, or even the entire structure, would be identical to another, substrate language. Haitian Creole would have a French lexicon, but a grammatical system from the African language Fongbe of the Gbe cluster (Lefebvre 1998). The idea, known as the relexification (or relabeling) hypothesis, suggests that a creole, if analysed by structural properties alone, should show close to one-to-one similarity with the substrate/adstrate languages.

14. *Creoles are natural continuations of the lexifiers, which developed just like for example Latin developed into Italian or Spanish*

All languages evolve and change, which is a natural process. One theory of creole genesis entails that there is no reason to classify some results of linguistic changes as creole languages, and others merely as overseas varieties of a language, or as local developments from another language. The evolution from Latin to French or Spanish would be equivalent to the development of French to Haitian Creole or Mauritian Creole. French would be Latin with Celtic substrate influence, Spanish would be vulgar Latin with Iberian and Arabic influence, Haitian and Mauritian would be French with African influence, especially Kwa languages and Bantu languages. As the substrates, or the specific constellations of the substrates, are
different in all these cases, the resulting linguistic outcomes are different. The lexicon and grammar of Latin would be continued in French and Spanish, and the lexicon and grammar of French would be continued in Mauritian and Haitian (Mufwene 2008).

15. *In the case of creoles, transmission of the language from parents to children was broken at one point in time*

Normally, languages are transmitted from parents to children, whereby members of each generation change the language a little bit. Changes are minor and gradual but over several generations one can notice differences. Reading Shakespeare from the 1600s is possible for a contemporary speaker of English, reading Chaucer from the 14th century is difficult, and reading Beowulf from a few centuries earlier is virtually impossible. Some creolists argue that creolization differs from this kind of normal transmission (Thomason & Kaufman 1988). At some point in time, the transmission was broken in creoles, in that people did not want to or were not able to learn the lexifiers. Thus they had to create a creole, perhaps based on a pidgin. In this view, there was a catastrophic event at some point in time which made the normal parent-child transmission of the language impossible. People with a range of language backgrounds were brought together, and they had to create a vehicular language, understandable for all. This would necessarily have been a quick process, and that is called creolization. After the creation of the creole, the transmission would be like any other natural language. Some creolists, however, promote the view that the lexifier was just transmitted from generation to generation in the normal way, perhaps with some loss, or with influence from substrate languages (e.g. Ansaldo & Matthews 2001; DeGraff 2001: 291, 2009; Aboh 2016).

16. *Creoles are the result of a combination of features from several languages, notably the lexifier and its vernaculars, and different indigenous or transplanted languages*

It is clear that a creole language differs from its lexifier phonologically and syntactically, in some cases to an extent that the creole is not intelligible to the speakers of the lexifier. Some people have claimed that creoles are formed by a combination of the structural properties of the superstrate and substrate languages, as exemplified in Aboh & Ansaldo (2007) for Surinamese creoles. An arbitrary mixture of structural features would have taken place, where some feature is taken from one source language, and another from a second source.

If this is correct, each and every feature of a creole must be attributable to one or more of the source languages, with no room for innovation. It does seem to happen that creoles display features not found in any of the known contact languages. In such a case, the feature is assumed to originate in an unidentified
substrate language or an unknown dialectal variety of the lexifier (Mufwene 2001). Innovations are downplayed or not mentioned. This idea was empirically tested in Aboh and Ansaldo (2007), but not all creole features were found in the identified source languages, and not all the lexifier and substrate properties were found in the creoles. Features that were not transmitted from lexifiers or substrates into the creole, were, according to Aboh and Ansaldo (2007) “semantically vacuous” (p. 53, 56) and not “semantically active” (p. 47) or “not semantically relevant” (p. 54). For another detailed application of the model, see Aboh (2015). Others have pointed out that creole languages display many innovations (e.g. Bickerton 1981; Baker 2001; Bakker 2014a).

17. Creoles show no structural traits from the indigenous or transplanted languages, and hence no substrate influence at all
Some people, notably Derek Bickerton (1981, 1984), have claimed that creoles developed from scratch: creoles were created by children who imposed a system on the chaotic speech (a jargon, rather than a pidgin) around them, and no influence at all can be contributed to the languages spoken as first languages by their parents. Research shows that there is influence from the substrates (and adstrates) on creoles (e.g. contributions in Muysken & Smith 1986; Michaelis 2008; Lefebvre 2011; Muysken & Smith with Borges 2015), but virtually nothing among the most stable structural features (Daval-Markussen et al., Chapter 7, this volume) at least for the Atlantic creoles.

2.3.5 Complexity of Creoles
Some creolists claim that creoles are less complex than other languages, probably due to the preceding pidgin stage. Others contest this claim.

18. Creoles are, as a group, less complex than non-creoles
There is an old view in linguistics, or perhaps an axiom or a dogma, that all languages are equally complex, as to the natural communicative capacities of languages. It is not always easy to come up with an acceptable definition of complexity, however. Complexity can be measured in many ways, for instance the number of grammatical rules, the number of words or syllables needed to express a specific idea or a text, the presence or absence of more or fewer grammatical categories, etc. The idea that all languages are equally complex, according to some metric, has been challenged repeatedly in recent years (e.g. Kusters 2003; Dahl 2004; Givón 2009; Shosted 2006; Miestamo et al. 2009; Sampson et al. 2009).

It has been claimed that creole languages as a group are simpler than non-creoles (e.g. McWhorter 2001; Parkvall 2008). Other people disagree with
this statement, suggesting that there are hidden or overt complexities in creoles (DeGraff 2001; Arends 2001; Ansaldo & Matthews 2001; Aboh & Smith 2009; Faraclas & Klein 2009). Or they claim that the loss of some properties of the lexifier are compensated for by an increase in complexities elsewhere.

19. **Some creoles are more complex than their lexifiers**

It is obvious that creoles generally lack a number of properties found in the lexifier, such as the presence of grammatical gender, inflectional morphology and variable constituent order. However, there can be structural traits in creoles that are more complex than what is found in the lexifier. Chinuk Wawa, for instance, had more personal pronouns than its lexifier, as do Solomon Islands Pijin and Tok Pisin. Also, it was claimed that the long and short forms of the verbs in Mauritian creole are not predictable, and, amongst others, therefore more complex than in French. In general, however, supporters of this claim have not undertaken comparisons of the complexity of lexifiers and creoles with a view to providing empirical proof of this theory. Aboh (2015) is an attempt, and his view is that creoles are hybrid languages, and therefore more complex than each of the contributing languages.

2.3.6 **Creators of creoles**

Who created the creoles? Who was most responsible for the new language? Some people hypothesize that the adults were at the root of their creation, others assume it was the children or adolescents.

20. **Creoles are created by children**

When children are confronted with the speech of adults, who all speak the dominant language “incorrectly”, or who speak a highly variable jargon, or an incomplete pidgin, they need to create order in the chaos, and their generation will speak a homogenous creole. The children are the ones who create the structures of the creoles (Bickerton 1981, 1984, 1988).

21. **Creoles are created by adults**

When adults learn a second language, they will almost always have an accent and make errors in several areas of the language. If access to a second language is limited, or temporary, the result may be a version of a language quite different from the lexifier language (cf. Plag 2011 for thoughts about creolization and second language acquisition). In this way creoles can develop. In many settings where creoles emerged, adults were the largest group of the populations, and hence the driving force behind creolization (Shnukal & Marchese 1983; Jourdan 1989, 1991). In one
version, subsequent cohorts of people arriving in the new society, approximate the approximations of the previous generation, which, after some generations, resulted in a creole language (Chaudenson 1978).

22. **Creoles are created by adolescents**
Newborn children are not normally observed creating new languages, and neither are adults instrumental in the formation of new languages or purposeful distortion of languages. The sector of the population most creatively involved in languages, are adolescents (Kiessling & Mous 2004). They create and maintain secret languages, youth slang and other utterances of verbal creativity. Creole languages could have been created by adolescents, possibly when communicating with each other (Muysken 1981).

2.3.7 **Gradual or quick**
Some creolists have argued for a rapid genesis of creoles; others have argued for a gradual development. This question is also connected to who the agents of creolization were (see 14 and 15 above).

23. **Creoles develop slowly, over several generations**
If new groups of people keep on arriving in the overseas society, with each generation the lexifier language will be changed fairly radically, more so than with normal transmission between generations. The new acquirers will be adults, and thus the effects of imperfect second language learning will make themselves felt. Thus, creoles develop slowly, and it would take many generations for creoles to stabilize. Some of the arguments for this position are found in the early documentation of some of the creoles, where there seems to be much more variation and less stability than in later periods. This idea is sometimes called gradual creolization (Arends 1989, 1993; Bickerton 1991).

Arguments for gradual creolization can be found in the documentation of early sources, which often show less significant deviations from the lexifier compared to the later creoles (see e.g. Arends 1989, 1993; Bakker 2014b).

24. **Creoles develop quickly, within one generation**
Some scholars argue that creolization is a very rapid development, even as brief as one generation (Baker 2001; Bickerton 1984). Within just one or few decades, the creolization process would have led to a new, stable language. The idea of rapid genesis is often associated with a preceding pidginization period. The first generation of new arrivals, being displaced people speaking many different languages,
will be forced to create an adequate medium of interethnic communication. The first generation will have the most impact on the structure, as they are the founders (Mufwene’s “founder effect”, 1996). In Suriname five English-based creoles are spoken, despite the fact that the country was an English colony for at most a few decades, so it is assumed the language must have come together rapidly.

25. Creoles show grammaticalization paths at an accelerated pace

Usually, grammaticalization entails a change of meaning, in many cases a broadening, and at the same time a reduction of the form. An example is the development of the numeral meaning “one” to an indefinite article in Western European languages or into a marker of nonspecificity and newness in creoles (Givón 1981; De Mulder & Carlier 2011). The development of grammatical elements out of lexical elements is one aspect of grammaticalization. The genesis of grammatical elements (such as case markers, TMA elements, verbal conjugations and the development of adpositions out of nouns) is usually assumed to be a rather slow process, taking up to several centuries.

In the creolization process, a new grammar has to be developed. Thus, in a relatively short amount of time, new grammatical markers develop for a number of semantic and cognitive distinctions. Many of the new markers show up in the earliest documentation of language samples in the creole societies. For many markers, the lexical stem, with a specific meaning, and the grammaticalized form, with a generalized meaning, coexist. In Saramaccan, for example, the word *manu* means “man, husband”, but it developed into a derivational affix – *ma* where it has the function of an agentive suffix, relating to both men and women. Similarly, *tan* means “to stand” and developed into *ta*, the non-punctual aspect marker derived from it, as in *a ta waka* “he is walking”. Both the early appearance and the continued presence of the source lexical items are taken to be witnesses for a rapid process of grammaticalization in creoles (cf. Plag 2002; Bruyn 2011).

2.3.8 Location

The question of where the creoles developed has been a source of contention for several decades.

26. (Some) creoles developed on plantations

One important group of creoles consists of those associated with plantations. Many creoles of the Caribbean region and the Americas are considered plantation creoles. The work forces for many of these were often slaves. In the 1970s, the idea was that slaves did not have a language in common and thus had to develop a pidgin to communicate with each other and with the bosses. The pidgin subsequently
creolized when the slaves became parents. Others, however, have argued that the English based creoles have so many words, structures and grammaticalized items in common, that they must go back to one place, perhaps in Africa (McWhorter 1995, 1999). In that case, the pre-creole was possibly imported to the plantations from Africa, or an early settlement in the region, for instance Barbados or Saint Kitts.

Still, there are many creoles that developed outside of plantation settings, such as the Arabic creoles in Africa, the Portuguese creoles in India, Tayo in New Caledonia, and the creoles that developed in West Africa.

27. Creoles developed in urban settings
Some people have argued that at least some creoles developed in urban settings. It was in urban settings that people speaking a diversity of languages met and interacted. This has been argued for by a.o. Roberts (2005) for Hawaii. Also Jourdan (1989, 1991) associated the stabilization of pidgin in the Solomon Islands with urban settings, as do Shnukal & Marchese (1983) for Nigeria. Some creoles have been associated with schools, notably Tayo French creole in New Caledonia and Hawaiian Creole (Baker 2001).

28. Some creoles developed in other settings: forts, mission stations and maroon societies
Slaves who escaped from the plantations, may have done so before a creole language had developed on the plantations. Creoles spoken in such societies, are called maroon creoles, of which most of the Suriname creoles are the best known examples.

Some creoles developed around forts that Europeans established for instance in West Africa. Local populations would settle around them, and develop a new vernacular. The same happened around some mission stations.

2.3.9 Reasons for perceived similarities
If indeed creoles show more similarities with other creoles than with other languages, including the lexifiers and substrate languages, then this calls for an explanation. This quest for an explanation is one of the fascinating parts of creole studies, where different opinions are found.

29. The structures of creoles are similar because their substrate languages show similarities
Creolists working with Caribbean creole languages, which are spoken mostly by descendants of Africans, noticed similarities amongst the creoles and suggested that these similarities could best be explained by the fact that the ancestors of the first generations of speakers were speaking African languages (Holm 1988).
30. *The innovations in creoles reflect pragmatic aspects of language, i.e. language use*

The perceived common features in creoles are due to the increased language use in a setting where communicative pressures lead people to make their speech as effective and efficient as possible. The grammatical shortcuts devised are pragmatic solutions leading to successful communication (Sankoff & Laberge 1974; Shnukal & Marchese 1983; Bakker 1987).

31. *The structures of creoles reflect the more deeply engrained properties of the human brain through a bioprogram*

Bickerton (1981) explained recurring semantic distinctions in creoles by claiming that these properties were more deeply entrenched in the brain. These semantic distinctions (like realis/irrealis, or singular/plural) surface in creoles, when children grow up without a model language. The children impose the bioprogram features, which surface particularly in plantation creoles and maroon creoles.

These similarities, if one accepts them, could also be explained on the basis of pragmatic issues in language (Givón 1973, 1982; Bakker 1987) combined with cognitive limitations and salience in communication, leading to similar semantic distinctions in creoles (Givón 1973).

32. *The innovations in creoles reflect cognitive aspects of human communication, e.g. through saliency of meanings*

The idea is that the similarities between creoles have a common cause, and that it is not likely to be caused by some biological skill or through the specific languages in contact, as many structural features recur despite differences in lexifier and substrates. Universal cognitive capacities, used in order to streamline conversations, would be responsible for similarities in creole structures (Givón 1973, 1982).

2.3.10 Semantics

Lexical semantics has been understudied in creoles.

33. *Word meanings in creoles often deviate from the cognates in the lexifier because of influence from other languages*

The lexicon of creoles is derived from the so-called lexifier, but the fact that words in creoles and their lexifiers are cognates does not mean that they have the same meaning. Often the creole words have meanings closer to the semantic range in a substrate or adstrate language. For instance, Yilan Creole Japanese has one word for “hand, arm, elbow”, like the most important substrate language Atayal
(Austronesian), even though the lexifier Japanese has several words for the different body parts. Among the 75 APICS languages, no fewer than 43 (57%) make no distinction between “hand” and “arm”, whereas most lexifiers do distinguish the two.

Creoles tend to be high-context languages, with lesser emphasis on encoded meaning (semantics) and a higher emphasis on contextualized meaning (pragmatics). Creole semantics is characterized by hyperpolysemy. Since creoles have fewer word forms available, many lexemes consist of numerous lexical units. Finally, creole semantics reflect a tendency towards transparent packaging of meaning, partly through the use of multi-word units and partly through polysemy patterns.

2.4 Research on creole languages and the contributions to this book

The issues discussed in this chapter relate to the genesis of creoles, their subsequent development and the contemporary state of creoles. Thus, the following issues play a role with regard to the structure of creoles: the influences from the different languages in their genesis (substrates, superstrates), the influences of external factors (communication, cognition, biology) on creole genesis and the way they combine and mix. Whether creole languages form a distinctive type compared to non-creoles, and whether creoles descend from a pidgin or not is also discussed. All of these issues appear to be controversial at one level, except for one: all linguists who have studied creoles agree that creoles are natural languages, with the full range of expressibility of the other natural languages of the world.

References

Aboh, E. O. 2009. Competition and selection. That’s all! In Aboh & Smith (eds), 317–44. doi: 10.1075/cll.35.20abo


Chapter 3

Phylogenetics in biology and linguistics

Finn Borchsenius, Aymeric Daval-Markussen and Peter Bakker
Aarhus University

The main goal of this chapter is to introduce the reader to the parallels and commonalities that exist between the fields of biology and linguistics. Researchers from both fields faced similar problems when seeking to account for the descent and diversification of related entities (species, languages). Therefore, they often sought mutual inspiration and opted for similar solutions. This has resulted in a convergence of models and methods in both fields. This chapter is divided into two parts. Firstly, we review some of the methodological and conceptual developments that have occurred in biology since the emergence of the field of evolutionary biology. There will be an emphasis on the last decade, where a variety of computer-based analyses have been developed. To illustrate the benefits of these tools, phylogenetic methods are applied in the second part of the chapter to a group of high-contact languages (creoles), which have long defied attempts at classification due to their multiple ancestry.

3.1 Origin of phylogenetics in biology and linguistics

Phylogenesis can be defined as the evolutionary development and diversification of a group of organisms. Phylogenetics is the study of that process, aiming at drawing up phylogenies, usually in the form of trees showing the evolutionary relationships between the members of a group of living organisms, languages, or other entities subject to evolution through descent with modification.

Evolution as a theory explaining the diversity of living nature dates back to biologists such as Charles Darwin and Alfred Russel Wallace in the middle of the 19th century. Though Darwin founded evolutionary theory, he was, however, not particularly concerned with phylogeny; rather, he was interested in the process of change leading to the emergence of new species. In fact, just a single sketch of a phylogenetic tree appears in Darwin’s ground-breaking book On the Origin of Species by Natural Selection (1859: 116–117). Nevertheless, phylogenetic thinking was the logical outcome of Darwin’s evolutionary theory.
It is noteworthy to realize that linguists adopted phylogenetic thinking at a very early stage. In fact, some evolutionary concepts were used in linguistics before they were used in biology. By the end of the 18th century, linguists already used the concept of homology (i.e. similarities in form between words may point to a common origin) and had realized that diversity could be explained via descent with modification (e.g. Schleicher 1853). Inspired by Darwin’s publications, they quickly moved to making trees representing linguistic descent. In 1863, the German comparative linguist Schleicher published a paper depicting an Indo-European language tree (Schleicher 1863). Interestingly, he wrote the paper as an open letter to his friend, the evolutionary biologist Ernst Haeckel, who had introduced him to the Origin of Species earlier the same year. The concept of language trees depicting evolutionary processes and the diversification of languages from a common ancestral form, however, predates Darwinian thinking. The idea of a genealogical language tree has been accredited to the early work of Friedrich Schlegel, who introduced a language tree, or Stammbaum (family tree) approach in a publication as early as 1808 (Schlegel 1808). The first published manuscript depicting a phylogeny was published by Carl Johan Schlyter in 1827 (Collin & Schlyter 1827). Schleicher had also used language trees with branches six years before the publication of Darwin’s book (Schleicher 1853).

In their review of the early history of cross-fertilization and borrowing of ideas in the establishment of phylogenetic thinking in linguistics and biology, Atkinson and Gray (2005: 517) conclude that “Darwinian ideas of descent with modification were probably less revolutionary in linguistics than they were in biology. Phylogenetic thinking and methodology in linguistics had already developed rapidly before Darwin, and this continued throughout the 19th century”. Particularly noteworthy is the fact that linguists in the late part of the 19th century recognized and understood the importance of distinguishing between innovations (i.e. new characters not present in the ancestral form) and retentions (characters inherited from a common ancestor). Biologists today make a similar distinction, but they call them synapomorphies (shared innovations) and symplesiomorphies (shared retentions). In biology, this distinction was widely accepted until 70 years later (Hennig 1950), when it led to a revolution of systematic biology, setting the principles for algorithmic reconstruction of phylogenetic trees.

Following Hennig’s (1950) publication of Grundzüge einer Theorie der Phylogenetischen Systematik, the development of methods for reconstructing trees based on the principle of shared derived character states gained much interest. With access to computers becoming easier and more widespread, particularly from the 1980s, and due to the rapid increase of processor speeds, biologists soon engaged in developing computer-based algorithms for reconstructing phylogenetic trees (see Felsenstein 2004 for an extensive overview).
The fundamental principle for tree construction was parsimony, i.e. searching for the shortest possible tree with the minimum number of character changes across the entire tree that could also explain the variation in character states observed in the set of entities being analysed. By the 1980s, this type of phylogenetic analysis had become widespread and generally accepted as the basis of systematic biology. It was also proliferating into the field of biogeography, i.e. the study of the geographical distribution of living organisms across the Earth as a way of reconstructing the ancestral distribution areas of groups of organisms and the series of geological events that have determined their current distribution (e.g. Wiley 1988). With improved computer capacity, alternative principles to parsimony also began to gain terrain, most notably maximum likelihood, a principle derived from statistical theory, where one searches for the model that gives the highest probability of observing the actual data. Maximum likelihood estimation of phylogenetic trees is computationally more demanding, but it offers the possibility of implementing a more sophisticated, and hence, a more realistic model for evolution.

Access to ever-increasing quantities of DNA sequence data (see e.g. Marx 2013), in conjunction with faster computers and improved algorithms, has resulted in a move away from parsimony towards likelihood-based methods in biology. At the same time, there are fewer and fewer phylogenetic analyses based on the morphological characteristics of the studied organisms (e.g. whether a species has developed limbs or not).

While biologists from the very beginning of the silicon age incorporated computer-based methods in phylogenetic analysis, linguists have been more reluctant to do so. Early attempts to introduce numerical approaches to linguistic analysis based on distance matrices (e.g. lexicostatistics by Swadesh 1952, 1955) were heavily criticized (e.g. Bergsland & Vogt 1962) and are now largely discredited by most linguists – but perhaps not always for the right reasons (see for instance the papers in Wichmann & Grant 2012 for a rehabilitation). Atkinson and Gray (2005) discuss some of the problems and criticisms of distance-based methods in linguistics. This discussion can be seen as a parallel to the massive critique that was initially raised against numerical taxonomic methods and distance-based tree estimation methods in evolutionary biology. A fundamental problem with these methods in such evolutionary analyses is that symplesiomorphies (retentions) that do not carry any true signal of common descent are as important as synapomorphies (innovations). They are therefore likely to group unrelated old lineages together, although similarities between them are simply a result of the absence of derived features present in more recently-formed groups.

In linguistics, one of the points of critique has been that the conversion of lexical character data such as Swadesh lists to distance scores results in loss of information, thus reducing the capacity of the method to reconstruct evolutionary
history accurately (Steel et al. 1988). In contrast, tree-building methods based on parsimony or maximum likelihood incorporate character changes as an implicit part of the model. This allows the underlying assumptions concerning the precise evolution of characters associated with the hypothesized tree to be traced and evaluated. However, over the past ten years or so, the transfer of computer-based methods from evolutionary biology to linguistic studies has accelerated and is now diversifying rapidly. This has been termed the “new synthesis” of biology and linguistics (McMahon & McMahon 2003: 18–21) and incorporates a variety of methods and subfields in both evolutionary biology and linguistics. In what follows, we will provide an overview of the recent transfer of methods and discuss current trends in computational biology that could have future applications in linguistics.

### 3.2 Phylogenetic studies in linguistics

The initial wave of papers in the new synthesis of linguistics and biology focused on understanding the history of human populations using linguistic data and phylogenetic methods, both at the local level (e.g. Forster et al. 1998 for dialects of Alpine Romance languages) and for widespread language families such as Austronesian (e.g. Gray & Jordan 2000). Remarkably, some of these early linguistic studies were initiated by non-linguists (anthropologists, psychologists, geneticists). Initially, most early studies employed maximum parsimony for the estimation of phylogenetic trees (e.g. Gray & Jordan 2000; Holden 2002; Rexová et al. 2003; Dunn et al. 2005; 2008).

Similar to what is happening in evolutionary biology, we are seeing a strong tendency of moving towards likelihood-based methods for reconstructing phylogenies in linguistics. In particular, Bayesian analyses based on the Markov Chain Monte Carlo algorithms (Huelsenbeck & Ronquist 2001) have become increasingly popular. These methods use an iterative approach to sample a large number of almost equally likely models that can be summarized statistically afterwards. These have become a standard tool in many areas of computational biology. Papers employing such methods include, among others, Gray & Atkinson (2003), Dunn et al. (2008), Dunn (2009), Gray et al. (2009), Fortunato & Jordan (2010), Dediu (2011), Bowern & Atkinson (2012), Nurbakova et al. (2013) and Maurits & Griffiths (2014). Interestingly, however, a recent simulation study reported that maximum parsimony did a better job of finding correct trees than both maximum likelihood and distance-based methods (Barbançon et al. 2013).
3.3 Dated language phylogenies

Another significant development in biology has been the application of a “relaxed clock”. Over the last decades, evolutionary biologists have increasingly applied molecular clock approaches to date splits in phylogenies and to make inferences about the timing of evolutionary events, such as the emergence of major evolutionary lineages relative to, e.g. geological history (Rutschmann 2006). The underlying principle is that mutations happen stochastically at a certain mean rate (“strict clock”, the biological parallel to the “glottoclock” in linguistics). In this case, the number of mutations separating two evolutionary lineages would be directly proportional to the time passed since they diverged from their common ancestor. Biologists, however, soon realized that this assumption rarely holds true. Mutation rates vary greatly among different groups of organisms and across evolutionary time. This led to various methods of “rate smoothing”, where substitution rates are allowed to vary across the branches of the phylogenetic tree (the “relaxed clock”, Drummond et al. 2006).

Dating of phylogenies has been applied to linguistic data, which assumes that cognates, like bases in a DNA sequence, evolve at a relatively constant frequency over time. Morris Swadesh was a pioneer in this area (e.g. 1952; 1972). A problem with all dating methods is, however, how to estimate the overall mean substitution rate and hence to time-calibrate the clock. In addition, social and historical differences such as relative isolation, community size (Nettle 1999) and social upheaval (Thomason & Kaufman 1988; Dixon 1997) may influence rates of change. Biologists will typically use the age of fossils of ancestral species to calibrate their clock or, on shallow timescales, an average mutation rate that has been determined experimentally. In parallel to this, Gray and Atkinson (2003) put a time-constraint on the points (nodes) of initial divergence in all of the major languages based on known historical information on earlier stages of the languages. The results indicated that the expansions of the early Indo-Europeans were linked to the development of agriculture approximately 8000 years ago, a suggestion made by Colin Renfrew a decade and a half earlier (Renfrew 1987).

A similar approach was used by Gray et al. (2009) to test the pulse-and-pause hypothesis accounting for the expansion and dispersal of Austronesian settlers from Taiwan throughout Insular Southeast Asia and the Pacific. In this study, they calibrated the tree with archaeological date estimates and known settlement times. Their results are consistent with the idea of migration pulses followed by pauses obtained through archaeological and DNA studies.

Recently, a number of investigations have used a Bayesian framework for dating linguistic trees (BEAST; Drummond et al. 2012) to test hypotheses concerning the origin and spread of language groups and potential correlates in terms of
cultural events (Lee & Hasegawa 2011) or climatic changes (Honkola et al. 2013). Altschuler et al. (2013) used a linguistic clock model to date the writing of the Homeric epics. Phylogenetic dating has also played a role in the nascent field of extraction of stable typological features for the purpose of studying deep-ancestry relationships (Dunn et al. 2005; Greenhill et al. 2010). Stable features are those that can be shown to change only very slowly over time and they are assumed to be less prone to borrowing. For example, Pagel et al. (2013) used a combination of statistical modelling and a dated phylogenetic tree to infer the relationship between common usage of words and stability over a 15,000-year timeframe.

The relative stability of structural properties of languages has long been a controversial issue and, apart from a few serious attempts at making generalizations on the basis of empirical data encompassing a large number of languages from various families (e.g. Nichols 1992), no conclusive evidence demonstrating the evolutionary variability of structural features of language has been put forward. With the advent of quantitative investigations taking large amounts of linguistic data into account, the issue has received renewed interest over the last decade. The results of several recent studies based on the features described in WALS strongly suggest that some grammatical features evolve more rapidly than others (e.g. Parkvall 2008; Wichmann & Kamholz 2008; Wichmann & Holman 2009; Dediu 2011; Ellison 2014), although the exact ranking of the various features identified by these authors does not match entirely (see Dediu & Cysouw 2013 for a review).

In order to pinpoint those features that are thought to evolve at a different pace, the following logic was adopted. In cases where the various values attested for an individual WALS feature are relatively equally distributed across members of a language family (i.e. only a few differences can be observed across a single family), the low level of variation in the descent from ancestor to present-day languages suggests a relative stability for that feature. In other cases, however, the distribution of the various feature values is strongly uneven, which indicates that the rates at which the possibilities for that feature have evolved through time are different. The more variation we observe, the more likely this variation has its roots in independent innovations or contact effects, thus suggesting a considerable instability for that feature (see e.g. Wichmann & Holman 2009: 12).

Phylogenetic dating has been somewhat controversial in biology, in part because of the uncertainty linked to the use of fossils as calibration points. A fossil can, in principle, only provide a minimum age for a given lineage because the lineage in theory could be much older than the fossil record. Another problem using fossils is how to align a fossil of an extinct species to modern taxa, resulting in uncertainty as to where in the phylogeny the fossil, and hence, the calibration point should be placed. Finally, criticism has arisen from the frequent disagreement between the diversification time estimates based on respectively molecular
dating and fossil records, the so-called “rock-clock” problem. Nevertheless, dated phylogenies have become a standard tool in analyses of biogeographic patterns and community ecology at a range of spatial scales. The use of dated phylogenies has also raised discussion in linguistics (McMahon & McMahon 2006; Heggarty 2006), but the concept has nevertheless gained considerable currency.

3.4 Is linguistic evolution tree-like?

In linguistics, the idea that language diversification rarely proceeds through clean speciation events has long been acknowledged (i.e. lateral transfer between languages is known to occur pervasively at all levels of linguistic description in the process of descent with modification). Most notoriously, a preliminary step in historical linguistic analyses following the Comparative Method explicitly stipulates that known borrowings should be removed (e.g. Campbell 1998: 128). Therefore, the whole enterprise of reconstructing language families and depicting their interrelationships completely ignores an important aspect of language evolution: horizontal influences. Reticulate (i.e. non-tree-like) evolution in language can be attributed to contact-induced innovations. However, only a few alternative models to represent these non-tree-like developments have been proposed and today, linguists still widely favor the tree model to represent the evolutionary histories of languages, following the Stammbaum model (Schleicher 1863). Schmidt’s (1871) ‘wave theory’ stands out as a notable exception, because in that model, linguistic evolution is represented with concentric circles, thus reflecting the overlapping spread of linguistic features as a gradual process involving both vertical and lateral influences (illustrated in Figure 3.1).

Therefore, the validity of the traditional Stammbaum model used in historical linguistics has been questioned by sociolinguists (e.g. Labov 2007), and especially so by creolists (e.g. Mühlhäusler 1980: 34; Hancock 1987: 265–266; Thomason & Kaufman 1988; McWhorter & Parkvall 2002: 179–180). As Nichols & Warnow (2008: 762) point out, “[t]rees are often reasonable models of evolution, but sometimes a network model is more appropriate. For example, when creolization or language mixture occurs the correct graphical model will contain additional edges between branches in order to indicate the dual parentage”.

1. The standard tool for historical linguistic analysis, the Comparative Method, is a technique that enables the researcher to describe and reconstruct the evolution and development of sibling languages from a proto-ancestor. For a detailed account, the reader is referred to e.g. Campbell (1998) or any other introductory textbook.
The question of whether tree-like structures are an appropriate model of language evolution continues to be debated (e.g. Geisler & List 2013) and some authors argue that language evolution should essentially be regarded as reticulate rather than tree-like because of extensive borrowing between languages (e.g. List et al. 2014a; 2014b). Nevertheless, theoretical studies based on simulated data have shown that tree-like representations are quite robust in relation to lexical borrowing (Greenhill et al. 2009). Likewise, an empirical study found no evidence for the common claim that borrowing would be particularly high in prehistoric languages, such as those spoken in hunter-gatherer communities, which would compromise hierarchical relationships in the deep nodes of linguistic trees (Bowern et al. 2011).

The potential problems in applying tree-like evolution to linguistic data has led to the widespread use of biological computational methods designed to visualize conflicting signals in the data such as split networks or reticulate networks (Bryant & Moulton 2004; Huson & Bryant 2006). In particular, split networks have been applied to the study of creole languages (Bakker et al. 2011; Daval-Markussen 2011; 2013). The issue has also been subject to theoretical work (Nakhleh et al. 2005; Erdem & Ringe 2006; Kanj et al. 2008; Towner et al. 2012). However, it would perhaps seem that the choice is not between trees or other types of networks, but rather that the methods may complement each other in exploring the structure
and phylogenetic signal in the data (Gray et al. 2010). This viewpoint is in line with Huson and Bryant (2006), who suggest using split networks and associated tests for tree-like evolution as an initial data exploration procedure that can be used to determine the best path for further action, e.g. the construction of phylogenetic trees or perhaps reticulate networks.

3.5 Other lateral influences between biology and linguistics

Other recent applications of biological methods to linguistic analysis include phylogeographic and population genetics models. Phylogeography is a recent field of biology currently in explosive development (Avise 2000). It aims to study the movement of genes and populations across geographic scales. It thus deals with individuals and populations assumed to actively exchange genes through mating or migration. Nevertheless, it relies extensively on the use of phylogenetic methods including trees and other types of networks. In linguistics, phylogeographic methods have been applied to problems such as the origin of Native Americans (Sicoli & Holton 2014), the expansion of the Arawak language family in lowland South America (Walker & Ribeiro 2011) and the location of the Indo-European homeland (Bouckaert et al. 2012).

Methods from population biology originally developed to discover limits between distinct populations on the basis of recombining genetic markers have been applied to linguistic problems, such as explaining the language diversity of the Sahul region (between Australia and Papua, Reesink et al. 2009). Likewise, methods for studying population bottlenecks, such as those associated with founder effects when small groups of individuals colonize a new isolated area, have been applied. These have been used to test whether the decay in phoneme inventories in human languages following migration from Africa across Asia was the product of a serial founder effect, such as has been proposed for human genetic diversity (Atkinson 2011; but see also the critical reactions in Linguistic Typology 15 issue 2, and elsewhere).

Another interesting recent approach to the study of language evolution involves direct modelling (see Gavin et al. 2013). Although not strictly a phylogenetic method, this approach offers an interesting complement for testing some of the assumptions underlying language evolution and phylogensis. Jansson et al. (2015) applied a mathematical modelling approach to test whether Mauritian creole may have been created only from a mutual desire to communicate (see also Parkvall et al. 2013). Their results strongly support this assumption and further indicate that it may be possible for a creole to develop quickly after first contact. The issue of modelling language evolution is increasingly gaining interest, albeit
controversially. The subject was recently reviewed by Gong et al. (2014a), provoking numerous comments and discussion (Gong et al. 2014b).

Lateral influences between the fields of biology and linguistics are not unidirectional. Recent years have witnessed a steady increase in the amount of biological studies incorporating linguistic data. These include a large number of studies mapping linguistic or other cultural traits on molecular phylogenies in order to study correlations between genetic and linguistic groupings, or to test directly if language barriers have resulted in genetic diversification in human populations. Two studies specifically addressing this question, however, both failed to establish language as a direct driver of human population divergence (Donohue et al. 2012; Zhang et al. 2014). In the latter case, language diversification correlated instead with geographic distance. Other examples of biological studies using linguistic data include three phylogeographic studies of maize (Mir et al. 2013), chili (Kraft et al. 2014) and banana (Perrier 2011), where cognate words related to these plants were used together with molecular and archaeological data to establish the area of origin and subsequent spread of domestication. The general value of linguistic data for solving problems in other fields of science was discussed by Pagel (2009), who concluded that for many comparative questions of anthropology and human behavioral ecology, historical processes estimated from linguistic phylogenies may be more relevant than those estimated from genes.

While computational phylogenetic analysis and related methods of evolutionary biology are by now well-established tools in linguistic studies, other recent developments of potential importance for understanding the diversity of human language, in particular in ecological science are just beginning to be applied (Gavin et al. 2013). Ecologists are increasingly studying biological diversity across spatial scales using a conjunction of data including species distributions, traits and phylogenetic position, a field known as macroecology or ecoinformatics. Methods from these fields were applied by Amano et al. (2014) to analyse the geographical patterns and drivers of extinction risk in languages worldwide. The authors found that both small geographical extension of the language and small speaker population size were associated with rapid declines in speaker numbers, causing 25% of existing languages to be threatened with extinction. In another article, Turvey and Pettorelli (2014) used methods from macroecology to assess the geographical congruence between the occurrence of languages and mammals threatened with extinction, suggesting that similar drivers may be affecting both entities. Lee and Hasegawa (2011) used methods for studying biological diversity patterns to demonstrate that geographical proximity and, more importantly, isolation by a surrounding ocean independently explain a significant proportion of lexical variation across Japonic languages.
3.6 Creoles, stable features and their substrates and lexifiers

The effects of language contact can be paralleled to lateral gene transfer in biological evolution, a phenomenon whereby one species receives genetic material from another species that is prevalent in bacteria and related organisms lacking a cell nucleus (Enfield e.g. 2003 even pushed the analogy so far as to refer to language contact as ‘linguistic epidemiology’). Thus, the effect of borrowing might compromise genealogical language trees. In extreme cases, the degree of contact is such that whole linguistic systems are disrupted following a catastrophic event (also referred to as a ‘break in transmission’, see Thomason & Kaufman 1988). The creolization process is unique in the following senses: On the conceptual level, it is unlike what can be observed in normal evolutionary processes, namely that as in the case of creoles, several parent languages spawn a single daughter language. In addition, linguistically, it results in languages with a number of structural similarities. These are the result of the analogous circumstances under which they emerged, and this then implies that creoles share a typological profile. The question of the genetic affiliation of creoles and other high-contact varieties and their typological status are still very much debated (e.g. Thomason & Kaufman 1988; Owens 1991; Winford 2005).

The recent development of quantitative typological databases such as WALS (Dryer & Haspelmath 2013) has facilitated large-scale grammatical comparisons of languages from a variety of families around the world. This has helped address a number of questions which have been central concerns for linguists for a long time and has led to novel ways of tackling these problems. One issue that is particularly relevant in the context of creole studies concerns the existence of stable morphosyntactic features.

The process of creolization entails a strong reduction of the lexicon and the morphosyntactic systems of the languages present in the contact situation (as a consequence of going through a pidgin phase). The implication that linguistic features evolve at differential rates in the context of creole studies is that the features which are temporally most stable, and hence more resistant to borrowing and change, are precisely those that will be expected to be retained in creoles. Hall (1959) already demonstrated that lexical-phonological changes observed in Tok Pisin occurred much faster than in non-creole settings.

Creolists have proposed several theories in order to explain the presence of shared structural properties in creole languages, such as a relative absence of inflections or systems virtually devoid of gender agreement. Some argue that the lexifier language provided both the bulk of the lexicon and a good portion of the grammatical structures observed in creoles. According to this view, creole
languages are considered to be continuations of the lexifier and should therefore be treated as dialects of those. Hence, according to Mufwene (e.g. 2007), Germanic and Romance-based creoles worldwide should simply be regarded as the latest members of the Indo-European family. Others, on the other hand, emphasize the role of the substrate languages in creologenetic settings (e.g. Lefebvre 1998). The proponents of this view claim that creole structures can best be explained in light of the grammars of West African languages. A third view argues that similarities in creoles reflect universal cognitive mechanisms underlying heavy linguistic restructuring processes. Most creolists take an intermediate position.

These questions have been central to creole studies since their inception and we will suggest potential answers by using phylogenetic networks to represent the degree of similarity between the languages included in our samples. We will discuss several analyses comparing various languages involved in the formation of a sample of Atlantic and Indian Ocean creoles on the basis of a selection of stable features from WALS identified by Wichmann and Holman (2009). The samples that were used consist of creole languages, to which additional data on the lexifiers and substrates were encoded in order to shed light on the degree of similarity between the languages involved in the creation of creoles. If creoles are more closely related to their lexifier, then we should observe that the creole languages cluster according to their respective lexifier. This would suggest that the lexifiers have played a particularly important role in the creation of creoles. Conversely, if the creoles cluster closely with the substrates included in our sample, this would indicate that the original languages of the subjugated populations provided an important part of the grammatical structure of creoles. And if the creoles group together apart from both lexifiers and substrates, this would indicate that creoles are a relatively homogeneous group of languages from a typological perspective and have more commonalities between them than between any other group of languages considered – which would also emphasize the importance of the creolization process as responsible for the similarities between creoles, and not the types of languages that happened to be involved in the process.

A sample consisting of 26 creoles and their six lexifiers produced the network in Figure 3.2 following the Neighbor-Net algorithm (Saitou & Nei 1987) implemented in SplitsTree v. 4.1.13 (Huson & Bryant 2006). The lexifiers all appear in a single group on the left of the graph, with the Germanic languages in the upper part of the cluster and the Romance languages in the middle. The creoles are scattered across the rest of the graph, with no apparent patterns according to geography, lexifier or substrate languages.

The graph in Figure 3.3 was built using data on the 26 creoles and 48 substrate languages representing the different branches of the Niger-Congo family that
Chapter 3. Phylogenetics in biology and linguistics

**Figure 3.2** A network of 26 creoles and their six lexifiers.

**Figure 3.3** Network of 26 creoles and 48 Niger-Congo substrate languages.
were involved in the creation of Atlantic creoles (principally Adamawa-Ubangi, Atlantic, Bantu, Ijoid, Kru, Kwa and Mande). All the creoles are grouped in a single cluster at the top of the graph, again without showing any groupings which could be consistently explained either by geography or by their lexifier. Only Sango, a creole spoken in the Central African Republic and based on Adamawa-Ubangi languages, appears at the periphery of the creole cluster – but removed from the three other Adamawa-Ubangi languages present in the sample (Doyayo, Mbum and Mumuye). The genealogy of the included substrate languages is partially reflected in the various clusters.

These results suggest that the typological make-up of creoles is rather similar and that the selection of stable features adequately reflects the grammatical parts of the linguistic systems that were not greatly affected by the creolization process. Hence, the selected stable features which were also present in the lexifier and substrate languages were not continued in the creolization process.

3.7 Creoles and genetic affiliation: Stammbaum, convergence, contact

The genetic classification of creoles has been a long-debated issue, especially after the publication of Thomason & Kaufman (1988), who posited a non-genetic origin of creoles and other high-contact varieties due to a break in intergenerational transmission, resulting in languages that inherit their grammar and lexicon from different sources (Thomason & Kaufman 1988: 200). A number of papers addressing the question have been produced since (see e.g. Owens 1991; Dimmendaal 1995; Clements 2002; Cardoso 2008; Lefebvre 2011 among many others), but no consensus has been reached and the question of how best to classify creoles and represent their interrelationships remains unsettled. Hence, the application of phylogenetic networks to incorporate lateral transfer in contact varieties is a highly-appealing method for representing the reticulation events that characterize this type of languages. Evidently, this is also true for more conservative languages, since no language can be claimed to be completely immune to contact-induced change (see for instance Friðriksson 2008 for a study of stability and change in Icelandic).

As the results presented in the previous section suggest, the multiple parentage of creoles is not obvious when applying a selection of stable features. In order to assess the degree of similarity between creoles, we have produced a new graph showing the relationships between the 26 sampled creoles and presented in Figure 3.4.

The creoles are spread throughout the graph and do not pattern according to either lexifier, substrate or geography, apart from a few cases where diffusion is known to have played a role (e.g. the Indian Ocean creoles from Mauritius and
Réunion in the top left, or those spoken in Haiti and Martinique, on the upper right). Again, these results suggest that the classification of creoles does not depend on the genealogical background of the languages involved in their creation, but rather, that they should be considered a phylogenetic unit in their own right, albeit on different grounds than traditional language families (i.e. not due to inherited traits but rather because of common grammatical structures resulting from the similarity of circumstances that led to their emergence).

3.8 A cognitive account of creole genesis

In order to explain the similarities that can be observed between creoles, regardless of their lexifier or of the substrate languages involved in their creation, as shown above (see also Bakker et al. 2011), we have to understand the conditions surrounding the emergence and development of creole languages. Several factors have to be considered, but the most relevant factor is the pidgin ancestry of creole languages (Parkvall & Goyette forthcoming). Much of the grammatical apparatus that many
older languages display, was filtered through a bottleneck, which resulted in the disappearance of irregularities and other ‘ornamental’ features such as gender agreement, thus facilitating effective communication. Hence, the pidgin past of creoles is reflected linguistically in their synchronic structural make-up, in that several properties found in older languages are absent in creoles.

New insights on the exact nature of these features were recently gained from quantitative studies based on WALS (Daval-Markussen 2011; 2013). In this section, we will discuss the five features that were extracted as characteristic of creoles and provide a cognitively-oriented explanation as to their widespread occurrence in creoles worldwide.

Feature 38A (Indefinite Articles) in WALS describes the occurrence of indefinite articles and allows for five feature values (Dryer 2013a). Many creoles use the same word for the indefinite article as for the numeral ‘one’, corresponding to the second most-common value attested in the available sample of WALS languages (112 out of the 534 languages included, or 20%). Therefore, the occurrence of this feature value is relatively common in the languages of the world. Its presence in a vast majority of creoles can be explained as a direct outcome of the creolization process followed by the grammaticalization of the numeral ‘one’ for marking indeterminateness or non-specificity. Studies in grammaticalization have shown that languages that have an indefinite article equivalent to the numeral ‘one’ are believed to have gone through a similar developmental path (Givón 1981; 1984: 410–411; Heine 1997: 65–82; Heine & Kuteva 2007: 45–46).

The presence (obligatory or optional) or absence of numeral classifiers is covered by WALS feature 55A (Gil 2013). Only three feature values are possible and these are attested in 400 languages. The most frequent possibility in the sampled WALS languages is an absence of numeral classifiers and creoles align with this preference, as classifiers are not attested in creoles. This can be explained by the relative youth of creoles, since classifier systems tend to develop slowly over time (see Aikhenvald 2000) and are primarily found as an areal feature in languages of Meso- and South America and Southeast Asia (Aikhenvald 2000: 101).

WALS feature 69A describes the position of tense and aspect markers in a large sample of languages of the world (Dryer 2013c). One feature value relates to the absence of inflection, while the other values describe the various ways of marking tense and aspect distinctions morphologically. Of the 1131 attested WALS languages, only 152 (or 13%) do not display tense-aspect inflection altogether (two of which, incidentally, are the creoles Ndyuka and Sango). Languages that lack inflection are typically located in Southeast Asia and Africa. Due to the scarcity of languages worldwide displaying this preference and to its widespread occurrence in creoles, this phenomenon can be directly linked to the process of creolization. Indeed, it is a symptom of the morphological reduction alluded to above, in
connection with an initial phase of pidginization in the development of creoles, which together result in extremely analytic languages. In turn, in order to express tense-aspect distinctions, the most obvious choice would be to use lexical items (primarily verbs, adverbs and auxiliaries) carrying the semantics of the various aspectual nuances and to use them as verbal markers, following a grammaticalization path similar to the one that can be observed in standard non-creolizing language change.

With regard to WALS feature 112A, which looks at the distribution of negative morphemes in a sample of 1157 languages (Dryer 2013b), almost half of the languages (502, or 43%) exhibit a preference for a negative particle (the African creole Sango being one of them) rather than being marked morphologically or by double negation. Hence, this feature is not at all uncommon in the languages of the world and therefore cannot be entirely explained as being a unique result of the creolization process.

Finally, feature 117A in WALS relates to predicative possession and shows the distribution of five different strategies in 240 languages (Stassen 2013). A subset of the languages in the sample (63, or 26%) display a preference for marking possession with a verb meaning ‘to have’. These languages are predominantly located in Europe, and therefore this strategy is also found in most lexifiers. However, the limited size of the sample does not permit further generalizations that would rule out an influence from substrate languages.

These five properties are almost universally present in the creoles of the world and are rarely attested together in other languages of the world (out of the 228 languages in WALS for which at least four feature values are known, only Rapanui follows the same pattern as creoles). The presence of some of these features can be taken as a direct reflection of the pidgin past of creole languages (e.g. lack of inflection and absence of complex tone), the linguistic consequences of which can still be detected synchronically. Other properties seem to surface in grammaticalization processes linked to creolization. These structural properties can therefore characterize creoles on linguistic grounds only. Bakker (2014) used these features to identify creoles with non-European lexifiers, with some success.

3.9 Conclusions

We reviewed the most important conceptual and methodological advances in both disciplines, especially during the past decades, which witnessed an explosion of studies taking advantage of computers for drawing phylogenies. In linguistics, phylogenetic analyses are becoming increasingly popular and diverse in the scope of issues. The theory-driven part was followed by a series of phylogenetic analyses
assessing the status of creoles among the languages of the world and especially among the languages that were involved in their creation.

The many parallels between biology and linguistics resulted in an extensive cross-fertilization between the two fields and this is best reflected in the number of linguistic studies using tools developed by bioinformaticians to track down evolutionary processes. Adopting methods originally designed to study biological evolution with linguistic data has become more and more widespread in historical linguistic studies. Phylogenetics has now come of age and the implementation of a wide range of algorithms has helped to advance our understanding of linguistic evolution as a whole, addressing new questions and sometimes finding unexpected answers and new challenges. The increasing number of interdisciplinary papers involving biologists, computer scientists and linguists, among others, reflects a growing interest in studies integrating insights from their respective disciplines so as to explain the patterns of human migration and diversity that can be observed worldwide today. The time now seems ripe for narrowing down on contact varieties so as to gain a better understanding of the processes and mechanisms underlying language change and that are at work in language contact situations.

References


CHAPTER 4

Methods
On the use of networks in the study of language contact

Peter Bakker, Eeva Sippola and Finn Borchsenius
Aarhus University / University of Bremen / Aarhus University

This chapter provides an overview of the phylogenetic models used in this book. In the introduction, we present the aims and limitations of the chapter and clarify some basic concepts. After presenting the steps of linguistic phylogenetic analyses, we proceed to explain the different data types that are examined in this book, such as lexical and typological features, and describe how these are coded. In the final sections, we discuss the resulting network and tree models and how to interpret the graphic representations built using linguistic data. In short, this chapter enables the reader to interpret the graphs and to assess the validity of the results.

4.1 Introduction

This introductory chapter provides an overview of the methods applied in this book, focusing mainly on network models. Computational phylogenetic methods have been developed in evolutionary biology and have been applied for the past two decades in historical linguistic studies to track language developments, evolution and relationships (see also Chapter 3). We believe that they are also suitable for the comparative study of contact languages in general, and creoles in particular, and can respond to the challenges creoles present for models of language evolution in historical linguistics.

Phylogenies, or evolutionary trees, are basic structures that help to visualize differences between species and how they are related to one another. In phylogenetic studies today, computers assist researchers in systematizing massive data sets and in analysing the differences statistically (Felsenstein 2004: xix). Evolution of life forms and products of cultures, including language, can be visualized in trees of descent or other types of graphs. These trees map the phylogeny, i.e. the
development of entities that have evolved from a common ancestor, and they map the degree of similarity between those entities.

The chapters in this book study creole languages and their interrelationships using mainly a method called splits networks. Depending on the aims, these are often applied in cases where the linguistic history and the links between varieties are unclear or ambiguous, as is often the case with creoles.

Although the field of phylogenetic linguistics has attracted a fair amount of attention among linguists in recent years (Bouckaert et al. 2012; Dunn et al. 2008; Dunn 2014; Gray et al. 2007; Wichmann & Good 2014), only a limited number of works include a comprehensive introduction to the methods most commonly used by linguists (e.g. McMahon & McMahon 2005; Nichols & Warnow 2008). This chapter is intended as a basic introduction to the computational phylogenetic methods relevant for this book, especially splits networks, from a linguist’s point of view. For a more detailed explanation of the concepts and methods used in the field in general and recent developments, we refer the reader to Dunn (2014) and the references therein.

The remainder of the chapter is structured as follows. After presenting the basic steps required for performing linguistic phylogenetic analyses, we proceed to present the different data types that are examined in this book (such as lexical and typological features) and to describe how they are coded. In the last sections, we discuss the resulting network and tree models, how to read them and how to interpret the information they display.

4.2 Steps of analysis: Encoding, representation, and interpretation

The application of phylogenetic and computational methods to the study of languages can be divided into three stages (after Heggarty 2006: 184). These stages include the encoding step, the representation step and the interpretation step. In the encoding phase, the linguistic data are selected and transformed into an adequate input data format that will be used for phylogenetic analysis. For some of the studies in this book, the authors have collected, selected and encoded the linguistic properties they study. This can include data both from secondary sources such as reference grammars and findings from fieldwork. Other authors use data from typological databases compiled by others, such as the Atlas of Pidgin and Creole Structures (APiCS; Michaelis et al. 2013a, b) and the World Atlas of Language Structures (WALS; Dryer et al. 2013). Another important dataset is the one compiled in the volume edited by Holm and Patrick (2007), Comparative Creole Syntax, which includes a large grammatical dataset of 18 creole languages worldwide. These data attest the presence or absence of a range of typological
features in the selected creoles and can be summarized as numerical values. If a feature exists in the language, this will be encoded as ‘1’ and if the feature is absent as ‘0’ (e.g. presence of tone = 1, absence of tone = 0). Thus, the collected data are typically first converted into numerical values in a matrix, where the languages are arranged in rows and lexical or grammatical features in columns. These numbers in the matrix form the basis of the next step, which involves converting the input data to a graphical, more easily readable form. Figure 4.1. is an example of a matrix with binary values for 21 typological properties in 24 languages from the Amazon.

| Cayubaba | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Baure    | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |
| Mosetén  | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| Lakondé  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| Kwaza    | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| Chiquitano| 1 | 1 | ? | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | ? | 1 | 1 | 0 | 0 | ? | 0 |
| Itonama  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Gavio     | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| Quechua  | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Aymara   | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Leko     | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Chácobo  | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| Movima   | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| Mekens   | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | ? | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Aikanã   | 1 | 1 | ? | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | ? | ? | ? | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| Kanoë    | 1 | 1 | 0 | 0 | 0 | 1 | ? | ? | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| Wari’    | 1 | 0 | 1 | 1 | 1 | 1 | ? | 0 | 0 | 1 | 1 | ? | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| Yuki     | 0 | 1 | 1 | 1 | 1 | ? | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cavinéia | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |
| Uru      | ? | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Arikapú  | 1 | 1 | ? | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | ? | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Yurakaré | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Karo     | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Karitiana| 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | ? | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

(1) Subordination through nominalization; (2) = Cross-reference; (3) = Evidentiality; (4) = Vowel harmony; (5) = Head marking; (6) = Verbal number; (7) = Directionals; (8) = Polysynthesis; (9) = Postpositions; (10) = Inclusive/exclusive distinction; (11) = Alienable/inalienable distinction; (12) = Verb classification; (13) = Strictly Nom-Acc alignment system; (14) = Asymmetrical morphology; (15) = Nasal harmony; (16) = Nominal number; (17) = Head-Modifier; (18) = Possession marked on possessum; (19) = Classifiers; (20) = Switch reference; (21) = Grammaticalized system.

**Figure 4.1** Feature matrix for 21 properties of 24 Guaporé-Mamoré languages (Brazil, Bolivia), from Crevels & van der Voort (2008).
In the *representation* stage, these matrices are analyzed and transformed into a graph representing the relationships between the languages being studied. Depending on the assumptions underlying the researcher’s hypothesis, a method for phylogenetic reconstruction will be selected. Most of the networks presented in this book are produced using the Neighbor-Net method (Bryant & Moulton 2004), which uses the Neighbor-Joining (NJ) algorithm (Saitou & Nei 1987) to construct phylogenetic trees. NJ is a distance-based algorithm that starts off by calculating the cumulative differences between each pair of languages in the matrix. These distances will then serve as the numerical basis for constructing a phylogenetic network. Figure 4.2 shows a splits network produced by Neighbor-Net on the basis of the matrix of Amazonian languages in Figure 4.1.

![Figure 4.2 A splits network of 24 languages of Lowland Amazonia (based on data from Crevels & van der Voort 2008).](image-url)
In this book, most authors chose to visualize their data as unrooted splits networks using the Neighbor-Net method as implemented in the software package SplitsTree4 (Huson & Bryant 2006). SplitsTree4 is relatively easy to use and makes it possible to produce different types of graphs and perform various statistical calculations.

In the third step, the interpretation stage, the results are used to draw inferences on the relationships between the selected languages. Are the visualized results in line with what we know about the languages included in the comparison? Are there languages that end up in unexpected clusters? Are there regional or genetic clusters? In our experience and based on reports from the literature on studies using phylogenetic software, almost all results conform to expectations based on earlier, non-computational studies, but there can be surprises as well.

The programs and network models can help us visualize the degree of similarity between languages in the data, but linguists with detailed knowledge of the languages in question are always needed to sort out the significance of the results (see also McMahon & McMahon 2005: 154). Unexpected results could, for example, be caused by coding errors or by incorrect preconceptions. Indeed, computer programs mechanically follow the selected algorithms on the data input, whereas humans are influenced by their ideas and expert knowledge. Hence, the data selection and encoding procedure are crucial steps towards achieving reliable results.

### 4.3 Data types

When working with linguistic data in phylogenetic analyses, two main types of data can be used: lexical and typological data.

#### 4.3.1 Lexical data

Lexical data can be coded according to formal properties (e.g. phonological form) or cognate relationships (are forms with the same meaning derivable from one original form?) or semantic properties (e.g. does the word meaning “father” include “father’s brother”, or is there a separate word like “uncle”?). The most commonly used type of data in phylogenetic studies is comparative lexical data based on cognate relationships of the lexical items. Lexical similarity methods, based on formal properties of words, are commonly used in investigations of dialectal data, because the presumption of cognacy in dialects is well-justified (Prokić 2010; Dunn 2014). In general, these kinds of lexical similarity measures work best at relatively shallow time depths (Greenhill 2011). Although it might first seem that
formal lexical similarity measures would be suitable for investigating creoles (if they have the same lexifier) due to their shallow time depth, this is not directly the case. Any exclusively lexical measure would make creoles look more similar to their lexifier than they are, as the lexicon of a creole originates predominantly in one language (except a handful of creoles which have a mixed lexicon). However, the grammatical system of a creole differs in many respects from the structures observed in the lexifier.

Due to synonymy and to the closeness of lexifier and creole forms in the lexicon, it can sometimes be difficult to determine which items should be selected when meanings are compared (see also McMahon & McMahon 2005: 156 and Dunn 2014: 193 ff. for a similar discussion). To address the issue, different levels of granularity can be used. As coding in itself is a form of analysis, as is the selection of word meanings, it is always necessary to carefully describe the procedure followed. In this book, word forms are compared in some of the chapters, but obviously only in groups of languages with the same lexifier (Dutch, Spanish, French, Portuguese, English). In addition, some chapters also use lexical semantics rather than mere word forms as data input (Levisen & Bøegh this volume, Chapter 13–15).

4.3.2 Typological data

Obviously, the exclusive use of lexical evidence in the case of creoles, be it based on phonological forms or cognacy, does not tell the whole story about the history and origin of these languages. Structural and functional features are therefore used prominently in this book. Such grammatical properties have also been used by others in the study of the evolution and interrelationships of both contact (Bakker et al. 2011) and non-contact languages (Dunn et al. 2005, 2008; Cysouw & Comrie 2009; Longobardi et al. 2013; Bøegh et al. 2016). This type of data is particularly useful for the analysis of grammatical contact and admixture (Dunn 2014: 193). Besides, it should also be kept in mind that the set of structural features at play in a language is subject to evolution, just like the lexicon of a language. In this sense, computer programs should also be able to detect linguistic evolution on a typological as well as a lexical level.

Today, the existence of typological databases makes this task much easier. In 2013, a new database of contact languages with large amounts of data on typological features with copious examples was published in the Atlas of Pidgin and Creole Language Structures (APiCS, Michaelis et al. 2013a) and the accompanying online version (Michaelis et al. 2013b). APiCS includes comparable synchronic data on the grammatical and lexical structures of 76 pidgins, creoles, and mixed languages. It provides a great resource for large-scale comparisons of these languages. The atlas was compiled by a team of experts who filled in the feature values and provided
example sentences for each language. The accompanying examples make it possible to evaluate the feature value selections and thus assess the data. This procedure can be contrasted with previous typological enterprises, where experts coded many different languages for a few specific features. This was done when putting together WALS (Dryer & Haspelmath 2013), the next database we will present.

The World Atlas of Language Structures (Dryer & Haspelmath 2013) describes the distribution of 144 predominantly morphosyntactic features in 2880 languages, including all known language families and covering all continents. Due to the data collection procedure followed for producing this work (a single author filled out many data points for many languages for a few features), the degree of completeness of the database varies enormously (from every single data point to none). This work has been widely exploited and used for a variety of purposes, including phylogenetic studies (see for example Wichmann & Saunders 2007; Bakker et al. 2011).

Such databases can be exploited and expanded for conducting phylogenetic analyses. Although studying linguistic similarities between unrelated languages with typological features is rather uncontroversial (Dunn 2014), there are some challenges regarding the use of typological data in phylogenetic studies. Typological features evolve in a limited design space, which result in a higher probability of chance homology. In other words, features can have the same value due to chance, simply because there are only few possibilities available. Whereas words with the same meaning can be expressed quite differently between languages, both in the length and nature of the phonemes, features such as verbs and subjects combine in a limited number of patterns (e.g. VS, SV or variable). The likelihood that two words with the same meaning in two unrelated languages accidentally have the same form is quite small compared to accidental similarities in structures. Thus, if one encounters two languages from different parts of the world (and with no known contact history) with a VS order, these orders will most likely be the result of independent developments, rather than inheritance. But an independent development is less likely in cases where two languages call a dog hund.

As mentioned above, most of the chapters in this book use typological data to study the evolution of creole (and other contact) languages and their relationships to other varieties. These typological data may be of a grammatical or lexical-typological nature. Several contributors to this book work almost exclusively with structural data, especially when they are interested in general properties of creole languages.

Since both structural and lexical properties of languages evolve, both can therefore be used for detecting a phylogenetic signal. The study of lexical evolution is much older than the study of typological evolution, and the latter is as yet less well-understood. However, some historical linguists have claimed that the most stable typological properties of languages would reach a deeper time depth than would be achieved with the lexicon only (Nichols 1992, 1994; Wichmann 2015).
The reason that the use of typological data is much less common, may be that most of the input data in the studies in general come from established studies in historical linguistics, especially involving cognate data of common words which are quite readily available, such as Swadesh lists for several languages and families. Stable words are the most common source in standard procedures in establishing genetic relationships between languages. They are sometimes combined with grammatical structures, notably grammatical morphemes. However, it is not uncommon to also use data involving phonological similarity or typological features when traditional historical linguistic input in the form of word lists is unavailable or unfeasible (Dunn 2014). In addition, it has been shown that the most accurate classifications of languages into traditional language families are obtained through a combination of lexical and typological data (Bakker et al. 2009).

4.4 Data coding

As mentioned above, when gathering input data for a phylogenetic analysis, a data matrix is generally constructed, where rows represent the given taxa, i.e. languages, and the columns represent different linguistic features that are being compared (Nichols & Warnow 2008: 764). The features can be words, speech sounds, grammatical constructions, parameters, etc., which will be reduced to a single value, a simple number. It should be emphasized that more often than not, such a number summarizes a complex reality in one numerical value. Romance languages, for instance, are notorious for displaying some highly-frequent adjectives which appear before the noun, whereas all other adjectives follow the noun. Scoring the result for this feature with a value standing for ADJ-N, N-ADJ or “free” could all be considered correct and wrong at the same time, which makes it important to be transparent and consistent about the coding principles.

The coding of features can be done either in a binary manner (e.g. 0 and 1) or with multi-state values (e.g. 1, 2, 3, 4). Binary features attest, for instance, the presence or absence of a word, phoneme, semantic distinction, grammatical distinction or syntactic construction, or they can cover two opposites (e.g. verb-initial or not). Multi-state features allow the presence of various values (e.g. the six possible orders of S, V and O, perhaps with an additional feature for “free order”). Naturally, multi-state features can be transformed into binary features by collapsing or separating some values, depending on the overall design of the feature set.

Finally, another important issue related to the data collection procedure is how many data points are necessary in order to conduct meaningful analyses (i.e. filled-in values in your matrix). There is no general rule of thumb, but in principle, the more, the better, and a few dozen features, if chosen carefully, may be indicative of
classificatory patterns. One can also have as a purpose to isolate as few features as possible that would give certain results, e.g. creoles in one cluster, and non-creoles in another (Daval-Markussen 2011, 2013). Whereas it has been argued that 40 of the most stable words are sufficient to map most language families of the world accurately (Brown et al. 2008; Holman et al. 2008), an ideal or minimal number of typological features has not been subject to discussion.

4.5 Networks and trees

Trees are well-known from traditional historical linguistics studies in the form of graphic representations of language families, where an earlier protolanguage splits off into new languages. However, such trees showing bifurcating splits are only one type of graph among several. Huson and Bryant (2006: 254) define the cover term phylogenetic network as any graphic illustration that depicts evolutionary relationships between a set of taxa (languages in our case). In terms of graph theory, phylogenetic trees form a subset of phylogenetic networks. Thus, this definition encompasses both phylogenetic trees (e.g. language family trees) and splits networks (see Figure 4.3 for an overview of various tree representations).

![Diagram of different network types](figure4.3)

Figure 4.3 Diagram of different network types (based on Huson & Bryant 2006: 255).

There are explicit and implicit networks. In explicit networks, an actual evolutionary pathway is specified based on the assumption that the languages included in the sample are related. In implicit or abstract networks, there is no assumption of an evolutionary pathway and the network merely represents ambiguities or
conflict in the data (Huson et al. 2010: 70–71). The networks in this book are almost without exception implicit networks. An evolutionary inference is not necessarily made from the results.

Trees are graphs consisting of a set of ‘nodes’ (splits in a tree) and a set of ‘edges’ or ‘branches’ (see Figure 4.4 for a terminological overview), each of which connects a pair of branches (subfamilies) or taxa (languages). Two nodes share only one path, which means that the tree is acyclic (Nichols & Warnow 2008: 761). Figure 4.4 shows a rooted tree in which the bottom node (labelled “root”) represents the linguistic ancestor (or protolanguage) of the selected languages. The figure shows ancestral nodes for the groups \textit{abcd}, \textit{abc} and \textit{bc}. The tree can also be reversed, with the root on top, or sideways, with the root either on the left or on the right. In a rooted tree, an ancestor to all the sampled languages is assumed, whereas an unrooted tree simply maps similarities without assuming a common ancestor for the branches or languages. An unrooted tree is depicted in Figure 4.5.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{rooted_tree.png}
\caption{Example of a rooted tree (adapted from Nichols & Warnow 2008: 761).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{unrooted_tree.png}
\caption{Example of an unrooted tree, where \textit{w} and \textit{x} are closer to each other than \textit{w} and \textit{z}, or \textit{w} and \textit{y} (based on Nichols & Warnow 2008: 761).}
\end{figure}
These rooted and unrooted trees are unambiguous in the sense that they present a single solution for a classification or a visual representation of similarities.

In a splits network, on the other hand, there is more than one way to get from one terminal node (tip, or taxon) to another. A splits network is thus able to depict conflicts and ambiguity in the data, which can be, for example, due to reticulation events, missing data, etc. Reticulations could be interpreted as resulting from, for instance, borrowings, like the lexical borrowings from French into English, which link the Germanic language English with the Romance language French. Reticulations are visible in the box-shaped webbing that can be seen in networks with conflicting information.

Figure 4.6 (based on Lehtinen et al. 2014: 201) shows a Neighbor-Net network where some of the splits are marked with dashed lines. The graph also includes bootstrap values for the splits (see Section 4.6 for an explanation of the bootstrapping procedure). Languages a, b, c, d, and e are at the tips of the independent lines emerging from the central part of the web. The parallel lines marked with a dashed line (Split 1) separate languages a and b from the others. Also, the connection between a and c (Split 2) conflicts with the branch containing a and b and weakens their cluster.

![Figure 4.6 Neighbor-Net network with bootstrap values and two splits indicated with dashed lines (based on Lehtinen et al. 2014: 201).](image-url)
In Figure 4.7 we present a graph based on real linguistic data from Dutch-based creole varieties. The resulting network shows the interrelationships between Dutch and several of its creole varieties included in the comparison.

![Splits network of Dutch, Berbice Creole, and three varieties of Virgin Islands Creole Dutch](image)

**Figure 4.7** A splits network of Dutch, Berbice Creole, and three varieties of Virgin Islands Creole Dutch (see Bakker, this volume, Chapter 10, for a more detailed description of the varieties included in the comparison).

The main type of networks used in this book are splits networks. As already mentioned, splits networks include representations of incompatible and ambiguous signals in the data. This is also useful when working with high-contact varieties such as creoles, where multiple influences can be observed in the genesis, but normal change in their subsequent development. In splits networks, parallel edges, rather than single branches, represent the weighted splits that are calculated from the data. The nodes in these networks do not necessarily represent ancestral states of evolution, but they rather represent ambiguous signals in the data.¹ Splits networks provide therefore an implicit representation of several possible evolutionary histories.

---

¹ This is due to the fact that even if every splits network represents a unique collection of splits, the uniqueness does not hold in the other direction: a given collection of splits can have many different splits network representations (Huson & Bryant 2006: 256).
pathways (Huson & Bryant 2006: 255). In this way, they can depict a combination of tree-like signal and 'noise' in the data (Nichols & Warnow 2008: 764).

The historical development of a language, including contact events, can be interpreted from these graphs. However, it is important to note that the parallel lines (splits) that produce the webbing do not necessarily imply contact with a substrate or adstrate language in the case of creoles. They can also indicate homoplasy i.e. independent innovations, where people creating creoles invent parallel solutions to communicative challenges, for instance in the creation of a system of articles. The parallel lines may also be indicative of deviation due to insufficient or misrepresented feature coding. It can therefore be necessary to go back into the unprocessed data in order to explain the patterns which can be observed in a split network, especially the unexpected ones. This does not mean that unexpected results are wrong. Indeed, one's working hypothesis might be incorrect from the outset or it may be the case that the researcher’s expectations were misconceived before starting the phylogenetic analysis proper.

4.6 Interpreting the results

Network models have many advantages when studying contact languages. For example, the Neighbor-Net algorithm can represent conflicting signals in the data and provide a detailed snapshot of the selected languages for the properties considered. It gives an overall representation of the structure of the data and works as a guide for further analysis. The interpretation of the resulting networks is not always straightforward (Bryant & Moulton 2004) and depends on exactly how the splits were constructed (Huson & Bryant 2006: 256).

Let us illustrate this further by means of a number of graphs. Figure 4.1, given above, showed a matrix with binary data for 21 structural properties in 24 indigenous languages spoken in the border area of Bolivia and Brazil, which is extremely diverse linguistically. Zeros and ones indicate respectively absence and presence of a property, while question marks are used to represent uncertainties or unknown data points. By just inspecting the features, it is not easy to perceive which languages are more closely connected to others, either via common descent or through contact. This is where the computational approach can prove valuable.

Figure 4.2 showed a network based on these data. A substantial amount of internal webbing can be observed. This indicates conflicts in the signal, which may be due to contact, or independent developments – in both cases, linguistic traits that are not the result of direct descent with modification as is typical of inheritance. On the other hand, some independent lines (edges) are also quite long, as for example in the case of Arikapú and Wari’, which indicates that there are also considerable
differences between these two languages and the rest of the sample. In the bottom, we can see one node with Quechua/Aymara. Looking at the matrix, it appears that both Quechua and Aymara have exactly the same scores for all the scored features in this dataset, which places them at the same terminal node. The language closest to these two is Uru. Again, this language belongs to a different family than the others, but it is spoken in an area near Quechua and Aymara and has been influenced by the former two (Muysken 2000). This historical development is visible in the webbing and the nearby positions of these languages in the graph. On the left, we can see Aikanã and Kwaza with a lot of webbing. The two languages are isolates, but they are spoken in the same region and most Kwazas also speak Aikaná, so the languages have influenced one another, which again is visible in the form of dense webbing.

We can see that some edges (lines) are longer than others, either measured from an imaginary centre or measured from the splits. On the right side, for instance, the line for Arikapú is much longer than the one for Cayubaba and Gavião. This may indicate two things: either fewer features are known for the language with the shorter line than for the one with the longer line, or it may mean that the features that are deviant for Arikapú, are more often different from all the other included languages, i.e. Arikapú is more divergent than other languages in the sample.

Let us now look at a creole example. Based on the 97 morphosyntactic binary features described in Holm and Patrick (2007), we have data on 18 creoles based on Arabic, Dutch, English, French, Nagamese, Portuguese and Spanish and from all parts of the world. In Figure 4.8, we see less webbing and longer edges, which means that fewer properties included in the comparison are shared between these creoles than between the languages in the Amazonian example before. The long edges indicate high diversity, more than in the Amazonian case. The only two languages that are closely connected are Guinea-Bissau creole and Cape Verdean creole, forming a cluster in the upper right of the graph. These two languages are both historically related and areally connected, their position in a cluster together comes therefore as no surprise. The other languages that show webbing are often not related at all, for instance Nagamese (spoken in India), Nubi (KiNubi) (spoken in Eastern Africa), in the upper left, and Palenquero (South America). This shows that the adjacent languages in the network share more structural features than languages further away, but that does not necessarily imply that these languages share a common origin.

After a number of non-creoles were added to the data set by looking up feature values in reference grammars, we get the result presented in Figure 4.9. The non-creoles (indicated in CAPITALS) chosen were either the most analytic language in their phylum, or they were among the non-complex languages of the world (see Bakker et al. 2011 for details). The isolating and non-related (to each other) languages Mina (Afro-Asiatic), Koyra Chiini (Songhay or Nilo-Saharan), and
Mandarin (Sino-Tibetan), as well as relatively non-complex Pirahã (Mura-Pirahã or Amerind), Brahui (Dravidian) and Indonesian (Austronesian) are all located in the upper part of the graph, clearly separated from the creoles. Note also that the inclusion of these languages has a minor, albeit visible, effect on the internal distribution of the 18 creoles relative to one another (compare for instance the position of Angolar in Figs. 4.8 and 4.9). The 18 creole languages cluster together against the non-creoles, despite the diversity of creoles as a group.

A network can also be rooted, that is, one can suggest a certain language to be (close to) the ancestor of the sampled languages. In the case of creoles derived from a single lexifier, rooting the network with the lexifier is justified, based on the reasonable assumption that e.g. all French creoles derive from French. An example of a rooted phylogenetic network can be found in Figure 4.10.

How reliable are the positions of different languages and clusters in the different trees? A common procedure in phylogenetic studies is to calculate a statistical support measure for the various splits in a graph. The most commonly used
Figure 4.9 Neighbor-Net network with 18 creoles and six unrelated non-creoles.

method to calculate support values is called bootstrapping. An example of support values calculated using this method can be seen in Figure 4.6. Bootstrapping is a resampling procedure where new datasets are constructed by randomly resampling the characters in original dataset. The new dataset is then analysed using the same method as for the original dataset. Typically, this operation will be
repeated at least 1000 times (depending on the size of the dataset). For each split in the original graph, it is then calculated which proportion of the graphs produced by resampling, contain that particular split. A high value means that the split is supported by many characters in the dataset, whereas a low value means that the split depends on just a few critical characters and therefore might disappear if for example the dataset is enlarged by adding new features. Hence, the support value gives a statistical measure of the robustness of the result. In practice, a support value of at least 70% is considered necessary for robust results (Huson et al. 2010: 44).

A rooted network of French-based creoles with bootstrap values is presented in Figure 4.11. All the edges and splits have a numerical value between 6.7 (very low) and 100 (very high).

**Figure 4.10** A rooted Neighbor-Net network, with Dutch, placed closest to the root, and two creoles, Berbice Creole and Virgin Islands Creole Dutch, based on four different textual sources and three time periods (see Section 10.3 for a more detailed explanation of the varieties and the periods).
Figure 4.11 Rooted tree with French and a sample of French-based creoles with confidence values (for more details, see chapter 5).

We hope that these examples will help the readers in the study of this book by allowing them to assess the validity of the results presented throughout.

Notes

The datasets for this chapter can be found here: https://phylogenetic-creole-studies.blogspot.com
References


This chapter provides an overview of structural properties of creole languages based on widely different languages and spoken in a broad geographic range. We discuss phonology, morphology, syntax and lexicon. Phonologically, creoles tend to have average properties. Creoles are generally not endowed with a rich morphological apparatus, for instance contextual inflection is largely absent, but compounding, reduplication and some derivational processes are common. Syntactically, creoles are quite diverse in their nominal structures. Preverbal markers are common in the verb phrase. A relatively fixed constituent order is found in the basic transitive sentence, mostly SVO. These creole properties are compared with a sample of non-creole languages. The results suggest that lexically, creoles seem to be less mixed than European languages.

5.1 Introduction

Which properties do creoles share? In this chapter, we discuss recurrent properties of creoles observed in a wide range of creoles worldwide. These are generalizations, almost all of which have exceptions. Only a few studies based on extensive comparisons of creoles have been undertaken (e.g. Velupillai 2015). Comparative studies on creoles will be summarized in the next chapter.

5.2 Phonology

The consensus among creolists specialized in phonology is that there is nothing typologically special or distinctive about creole languages that sets them apart from non-creoles. For instance, Klein (2011: 155) finds that creole phonological properties “are quite unremarkable in comparison with non-creole languages around the world”, compared to the diversity observed in other languages of the world (Klein 2011: 155). Smith (2008: 103) agrees that the phonological systems
of creole languages “do not form any kind of unique type” (Smith 2008: 112). Likewise, Good (2013: 71) mentions that creole phonologies are “average”, and therefore unremarkable. Still, creoles may be special in that no extremes are found in this group of languages.

5.2.1 Creole segmental inventories

Thomas Klein has conducted a number of studies about the phonological properties of creoles. He used three different samples to study the phoneme inventories of (i) a number of African creoles (2006b), (ii) only creoles based on Portuguese and Spanish (2005), and (iii) a wide range of creoles worldwide (2006a). The creole data were compared with those from UPSID, the UCLA Phonological Segment Inventory Database (Maddieson 1984), which contains data on segmental inventories of 451 languages, or from the World Atlas of Language Structures (WALS, Haspelmath et al. 2005, 2008).

According to Maddieson’s (1984) criteria, 70% of the world’s languages fall under the category “average”. Hence, 15% on each side would be considered extreme. Two thirds of the languages in UPSID (based on Maddieson 1984 and related work) have five to seven vowels, and this can be said to be an average score. In WALS (Maddieson 2005a), vowel inventories can be small (4 or fewer), average (5–6), or large (7–14). Iberian creoles have rather average inventories: five to eight vowels (Klein 2005).

Maddieson’s criteria covering 70% of the world’s languages (Maddieson 1984: 7), established a range of between 20 and 37 phonemes for most languages. All creoles spoken in Africa have “typical” phoneme inventories (Klein 2006b), within this range.

As to manner of articulation of stops, Klein (2005, 2006b) concluded that about two thirds of the creoles are typical of “average” languages. The rest are “complex”, while none of the creoles in his sample display a simple inventory.

Klein’s sample of 23 creoles from all over the globe (Klein 2006a) confirms these findings. Again, the creoles typically fall within the range of what can normally be observed in the world’s languages. The one exception is Ndyuka English creole of Surinam, which has fewer than 20 phonemes, and is classified as “simple”, having 19 phonemes. With regard to vowels, 18 creoles fall within the average range and 5 out of 23 are complex, having 8 or 9 vowels.

One can conclude that the great majority of creoles have “typical” phoneme inventories (18 out of 23), falling within the range of 70% of the world’s languages.

Despite their scores within average ranges, some cross-linguistically uncommon phonemes are also encountered, e.g. co-articulated stops in Saramaccan, interdental fricatives in Angolar, and pre-nasalized consonants in several Atlantic creoles with English, Spanish and Portuguese lexifiers (see Parkvall 2000, chapter 3).
Creole segmental inventories fall within the range of “average” languages, within the range of what is attested for non-creoles, or in some few cases more complex than average. In other words, on the one hand creoles are average, but on the other hand they are also somewhat exceptional in that there are no extremes in that none falls within the 15% upper or lower range of non-creoles. An explanation may be that neither a small nor a large inventory contributes to clarity in communication in a situation of extreme language contact, as in the genesis of creoles. If a language has many phonemes, it is more difficult to learn, based on this parameter, as there are more distinctions. However, if there are few phonemes, it becomes difficult to distinguish between similar words, and therefore such a language is harder to learn because of this lack of differentiation (cf. Trudgill 2004: 316).

5.2.2 Creole phonotactics

The properties of creole phonotactics have not been studied systematically for a large number of languages. It is sometimes said that creoles tend to have a high frequency of CV syllables, with fewer consonant clusters or diphthongs (e.g. Holm 2000; Romaine 1988; Valdman 1978). No typologically unusual phonotactic patterns have been reported. Data from the *Atlas of Pidgin and Creole Structures* (APiCS, Michaelis et al. 2013) show that most creoles do allow for complex onsets (41 of 76 languages, i.e. 54%).

Saramaccan, for instance, does not preserve any consonant clusters from its lexifiers English and Portuguese, e.g. *toóbi* < English *trouble*, *buuká* “peel” from Portuguese *esbúgar* and *filigifěliti* < Dutch *vliegveld* “airport”. On the other hand, some Ibero-Asian creoles do preserve several clusters from their lexifiers, e.g. Chabacano (Sippola 2011; Lesho 2013). Berbice Creole displays different CV patterns for its two lexical components, Ijo and Dutch (Bakker this volume, Chapter 10).

All creole languages have CV patterns, and some quantitative studies show that CV syllables are the most common ones in creoles (as they are in non-creoles) such as Virgin Islands Creole Dutch and Saint Lucian French (Klein 2011: 175). Also, VC syllables are found in all creoles surveyed, and CVC syllables are common.

In a preliminary study of pre-publication APiCS data, Comrie (2011) reported complex codas for 7.1% of 70 APiCS languages (in the then forthcoming APiCS, Michaelis et al. 2013; the figure is 8% in the published edition), and for 30.9% of 485 non-APiCS languages, a substantial difference (see also Maddieson 2005c).

Combinations of vowels are reported from a modest number of creoles (Klein 2011: 175), but to date, no systematic survey on the topic has been conducted.
5.2.3 Creole suprasegmentals

Smith (2008: 112–117) discusses stress, pitch and tone in creoles. Berbice Creole has fixed penultimate stress, other creoles are pitch-based, such as Papiamentu, while Angolar and Saotomense are examples of tone languages. Most tonal patterns in creoles are lexical, not grammatical. Saramaccan is a relatively complex tone language (Good 2004, 2009; McWhorter & Good 2012), and split according to the lexical sources: English roots may have accent, but African words have tones. Saramaccan is the only creole language for which tone sandhi is reported (Ham 1999). Nevertheless, extensive tone inventories and phenomena such as downstep, contour tones and patterns of tonal interaction that are very common in African languages have not been reported for any creole, suggesting that creole tonal systems are simpler than those of African, Asian and Central American non-creole languages (McWhorter 2005, 2011: 76–78). Comrie (2011) points out that only 2.7% of APiCS languages have tones, against 16.7% of 524 WALS languages (Maddieson 2005b).


5.2.4 Summary of Creole phonology

Creole phonologies can be quite different from both the lexifier and the substrate, as e.g. Berbice Dutch, Tok Pisin and Annobon phonologies show. Creole phonological systems deviate from those of the lexifiers, sometimes radically, but they are never exceptional or typologically rare – and in that sense creoles may be different from non-creoles indeed. Creole phonological systems are not simpler than those of non-creoles, they tend to be average or sometimes more complex than average, and only in the area of tones and phonotactics are creoles on average less complex than non-creoles.

5.3 Creole morphology

The claim that creoles have little or no morphology (e.g. Bloomfield 1933; Seuren & Wekker 1986; McWhorter 1998, 2001) is widespread both within and outside creolist circles. Empirical cross-linguistic studies, however, show that creoles do indeed present morphological systems (Plag 2005). Plag (2008, 2009; see also Plag 2011) discusses morphology in creoles from an interlanguage perspective. He finds
that the lack of inflectional morphology has a parallel in early second language acquisition, and can be explained as a natural constraint on language acquisition effective in both situations.

Farquharson (2007) points out, for instance, the presence of a diminutive prefix in Haitian, nominalizer affixes in Berbice Dutch, Jamaican and Sranan, as well as reduplication, compounding and zero-derivation (roots used as nouns as well as verbs without change in form) in several creoles. In any case, claims that creoles lack morphology altogether are often met with indignant reactions (see e.g. the discussion between DeGraff 2001 and Lefebvre 2001, 2003), even though creoles are not unique (Bakker 2015): many non-creole languages also display little morphology, e.g. Mandarin. In fact, 15% of the world’s languages have no inflectional morphology according to WALS (Dryer 2013b). In that sense, creoles are not exceptional.

As the lexical source of many grammatical elements is often still clear in creoles due to recent grammaticalization, the status of some elements as bound, cliticized or free can be difficult to determine. Free and bound forms may coexist, sometimes in full (lexical) and reduced (grammatical) forms (e.g. Saramaccan tāki “say”, from which tā“that”, or tan “stand”, from which tā “non-punctual aspect”).

Words in creoles may be morphologically simplex even though they are derived from a complex word, or a phrase. Examples abound, like Sranan sidon “sit down”, Tok Pisin hamas “how much?”. This is also true for nouns that have assimilated etymological articles as part of the noun stem in French creoles. These forms are morphologically simplex constructions, as in Morisyen: latant “aunt”, mo latant “my aunt”, en latant “an aunt“ (see e.g. Baker 1984; Ladhams 2012).

5.3.1 Inflectional morphology

Creoles inherited virtually nothing inflection-wise from the lexifiers (e.g. gender, person, number, agreement) – which may have strengthened the impression of morphological scarcity in creoles. Nevertheless, inflectional morphology can be found in several creoles. The bulk of the inflectional apparatus of creoles is the result of grammaticalization processes. For instance, the pronoun him grammaticalized into a transitive suffix -im in Tok Pisin and other creolized forms of Melanesian Pidgin. e.g. mi rausim ju “I throw you out” (from German raus “out!”).

Some creoles inherited inflectional morphemes from the lexifier. Some Portuguese creoles in India show distinct verb forms marking tense and aspect, inherited from Portuguese: kato “sang”, katad “sung”, katan “singing” (Clements 1996; Luís 2010). Berbice Dutch has inherited a.o. the -te perfective, -a(re) imperfective and -apu plural suffixes from Ijo (Kouwenberg 1994).
Two kinds of inflection are distinguished, contextual and inherent (Booij 1996). Contextual inflection links the inflected element to other elements in the sentence, e.g. person agreement markers in the verb linking to nominal elements, agreement markers in adjectives linking to nouns or case markers in nouns linking to their function in relation to the verb. In contrast, inherent inflection is not required by a syntactic context, e.g. plural markers on nouns, comparative markers on adjectives or tense-mood-aspect in the verb. Only inherent inflection appears to be inherited from the lexifier in creoles (Luís 2010).

Apart from the scarce inherited morphology, some inflection in creoles results from grammaticalization. For instance, several French-based creoles have cliticized or morphologized person markers and possessive markers, as in Haitian *malo* for *mwê ale* “I go” (Valdman 1978: 86–89), and *granpapa-m* “my grandfather”, with syllabic or affixal consonants used for several subject and object pronouns. Several scholars have also analyzed the creole preverbal TMA markers as being clitics or even bound morphemes, rather than separate particles (e.g. Veenstra 1994), which would mean that verb forms are inflected with morphological person marking and TMA affixation. Berbice Creole has object clitics, from Eastern Ijo pronouns (Kouwenberg 1992b). Portuguese creoles may both inherit and develop inflectional morphology (Baptista 2011).

Saramaccan has inflected prepositions, *umi* “of me”, *fii* “of you”, *fëën* “of him/her”. It can also be argued that forms like Haitian French creole *l pa t la* “she was not there” are morphologically complex verb forms rather than four distinct free morphemes (Valdman 1978: 92). In non-creole languages, single consonants would never be considered separate words, but either clitics or bound morphemes.

In creolized Chinuk Wawa of Grand Ronde (an Amerindian-based creole), the first person personal pronoun *na’iga* coexists with a grammaticalized person marker *na-* in the verb (likewise, the other pronouns exist as bound and independent forms):

(1) **Na’iga buš-wi’k-na-la:’dwa**
    I indeed-NEG-1-go
    ‘I must not go’ (Jacobs 1932: 36)

Also some function words are complex. Zamboangueño (Philippine Creole Spanish) has complex demonstratives, e.g.: *de-este* “of this” (Lipski & Santoro 2007: 390–392). The personal pronouns in Solomon Islands Pijin (a creole, despite its name) combine the person markers *mi* and *iu* with *fala* (< *fellow*) to denote plural (*mifala*, *iufala*, and *iumi* for inclusive), and *olketa* (< *all together*) for third person plural. In addition, there are dual (*mitufala*) and trial forms (*mitrifala*), plus inclusive forms (e.g. *iumitrifala*), yielding in all 15 personal pronoun forms (Jourdan 2004: 707), thus exceeding the inventory of its lexifier English.
The evidence reviewed above suggests a complete absence of inherited contextual inflection in creoles, but some inherent inflection is inherited from the lexifiers or grammaticalized (see Holm 2008; Luis 2010). Still, compared to global samples of languages, creoles are found nearly invariably at one end of the spectrum, i.e. closer to languages with no inflectional morphology than to polysynthetic languages – the latter with perhaps up to 20 derivational and inflectional morphemes (Grant 2002, 2009).

5.3.2 Compounding and derivational morphology, reduplication, compounding, suppletion

Compounding is the most salient morphological process found in creoles (e.g. Valdman 1978: 154), even where it is absent in the lexifiers. For instance, French requires a connecting element like *de* ("of") or *à* ("to") between nouns (e.g. French *combat de coqs*, "rooster fight"), whereas French creoles do not (e.g. Reunion creole *batay kok*, id.).

Most descriptions of creoles contain lists of derivational morphemes, inherited and/or borrowed, e.g. Valdman (1978: 130ff), Mühlhäusler (1997: 214ff) and Steinkrüger (2003). In most creoles, the derivational morphemes have forms derived from the lexifier. A study of Saint Lucian Creole (Brousseau 2011) reveals 41 potential derivational morphemes, of which some 16 appear productive. In Chabacano (Philippines Creole Spanish), at least six of the eight forms of the prefixes have been borrowed directly from local Philippine languages, whereas 10 of the 12 suffixes come from the Spanish lexifier (Steinkrüger 2003). Spanish affixes only connect to Spanish roots, but Philippine affixes can be appended to any lexical item. The meanings of the affixes, here and elsewhere, are only rarely identical to those in the source languages, but obviously share some of the meaning base.

McWhorter (1998) claims that derivation in creoles is semantically regular, and overall more predictable, because it is relatively recent. Plag (2011) argues against this, pointing to some unpredictable meanings of derivations, even in the earliest Sranan sources (see also Steinkrüger 2003: 266 for Chabacano).

One process that is close to universal in creoles is reduplication (Kouwenberg 2003; Bakker & Parkvall 2005). This is clearly an innovation during the creolization process, as reduplication is almost completely absent from pidgins (Bakker 2003b) and from all European lexifiers. Baker (2003) conjectures that various patterns of verb reduplication entail a “tripling [of] the verb stock without the acquisition of any additional morphemes” (2003: 218). Kouwenberg and LaCharité (2004, 2011) show that, when looking into details, the morphological processes involving full or partial reduplication do not speak in favor of a uniform process: reduplication is used for expressing a range of meanings across creoles, not only intensification and
reiteration, but there are also less iconic meanings such as X-like and the derivation of adjectives from verbs in some creoles that have their root in West Africa. Aboh and Smith (2015) argue convincingly that such reduplications are based on Gbe substrate languages. Grant (2003) argues likewise for substrate/adstrate influence for reduplication in Philippine Creole Spanish.

Suppletion is marginal in creoles. Exceptionally, Wellens (2005) reports the existence of seven ways of forming plural nouns, including three affixes in Nubi Creole Arabic (but see also Chapter 9 for a detailed study of plural marking in another Arabic creole).

5.3.3 Creole morphology: summary

All creoles use compounding, probably all at least some derivation, virtually all reduplication, and a sizeable minority also inflection (Plag 2005). Contextual inflection is never inherited. The relative scarcity of morphology cannot easily be attributed to the pidginization process (see Chapter 1), since inflectional morphology is not so rare in pidgins (Bakker 2003a; Roberts & Bresnan 2008), but this may also be because many pidgins have more morphologically-rich lexifiers than creoles.

The idea that creoles have little morphology may also be attributed to a tradition of a certain analysis of creoles, as being languages with little morphology. Several elements can be analysed as clitics or bound morphemes rather than particles, auxiliaries or pronouns.

5.4 Creole constituent order

In this section, we will discuss different general aspects of constituent order in creoles.

5.4.1 Sentential constituent order

The order of constituents in creoles also deviates from any sample of the world’s languages, and from pidgins. According to Dryer’s sample of 1377 languages (2013a), 41% have Subject-Object-Verb (SOV), 35% have Subject-Verb-Object (SVO), 7% Verb-Subject-Object (VSO), 2% Verb-Object-Subject (VOS), 1% Object-Verb-Subject (OVS), 0.3% Object-Subject-Verb (OSV), whereas 189 (14%) languages lack a dominant order. SOV and SVO are clearly the most common, followed by languages lacking a dominant word order.
In contrast, almost all creole languages show a rather strict dominant SVO order. Data from APiCS (Michaelis et al. 2013) reveals that out of the 76 contact varieties considered in the survey, only three languages (Michif, Pidgin Hindustani and Singlish) do not display a dominant SVO order. This is even true for languages where the substrate languages and lexifier have SOV order. For instance, Dutch is underlyingly SOV in subordinate clauses, and in clauses with an auxiliary (SAuxOV), and in foreigner talk, but SVO in main clauses (Hammarström 2015), and Ijo is a typical SOV language, but Berbice Dutch, with its clear Eastern Ijo component, is nonetheless SVO (Kouwenberg 1992a). (Classical) Arabic is a verb-initial language, but Nubi Creole Arabic of Kenya and Uganda, and Juba Arabic Creole of Sudan have SVO as their basic word order (Heine 1982; Watson & Ola 1985; Wellens 2005).

When calling creoles typically SVO languages, that does not mean that the order is always SVO. Nouns and adverbs can often be fronted for emphasis, and so can verbs (in the Atlantic creoles the verbs are often repeated, as in pseudo-English: “(it is) go I go”).

There are three groups of exceptions to SVO order. In this first group, this is because of a recent shift. Some creoles have shifted to a different basic word order in recent times under the influence of adstrate languages. In the second case, a creole can develop verbal agreement which allows for a more flexible word order. Thirdly, a few creoles have always had dominant SOV order, like Taiwan Creole Japanese.

Creoles in the first group are primarily found in Asia. Some Portuguese creoles in India show newer SOV order beside SVO order (Clements 1991, 2001, 2007). Philippine languages are verb-initial, and according to Grant (2002), Philippine Spanish creoles moved from SVO to VSO, although according to others there is no unambiguous evidence that they ever were SVO (E. Sippola, p.c.). Sri Lanka Malay and Sri Lanka Portuguese have become SOV languages under the pressure of the local Sri Lankan languages Tamil and Sinhala (see Bakker 2000, 2006 for a brief comparison, based on data in Smith 1979a, 1979b, 1984; Adelaar 1991; Hussainmiya 1987; see also Nordhoff 2009, Smith 2013 and other sources).

The second exception is perhaps unique. Chinuk Wawa is the creolized form of Chinook Jargon (a pidginized form of Lower Chinook) as it developed at the Grand Ronde reservation in Oregon, and the word order is relatively free in this language. Word order in the creolized Grand Ronde Chinuk Wawa is free (both OV and VO occur). Subject clitics precede the verb, and object markers follow the verb (Jacobs 1932; see also Chinuk Wawa Dictionary Project 2012).

Finally, two pidgincreoles (i.e. languages that have recently creolized and co-exist with pidgin varieties) and one creole also follow an order that is different from SVO: Nagamese (North East India) conforms to the locally dominant type and has SOV word order (Sreedhar 1974: 176), and Hiri Motu (Papua New Guinea,
Dutton (1996) displays both SOV and SVO word order. Yilan (Creole Japanese from Taiwan) is SOV, but younger speakers also use SVO (Chien 2015).

Not surprisingly, creolists have often mentioned the robust SVO order among the recurring properties of creoles. For Bickerton (e.g. 1981: 17, 1984), SVO was a historical coincidence, and not a property dictated by the bioprogram. Seuren (1998: 292–293) called SVO word order typical of creoles: “If a language has a Creole origin it is SVO, has TMA [tense-mood-aspect] particles, has virtually no morphology”. Creoles, or their predecessors in the form of a pidgin, are languages without case marking. Neither does verbal morphology indicate the semantic roles of the noun phrase in a sentence. Therefore, the most natural way to distinguish semantic and syntactic roles is by means of a fixed word order, and the two noun phrases, subject and object, are separated by an intervening verb, hence SVO. Nevertheless, the non-SVO creoles do not have case marking or verbal means to indicate the semantic roles of NPs. Hammarström & Parkvall (2016) suggest that the majority of creoles display SVO simply because they inherited the most common constituent order from their lexifiers.

5.4.2 Verb phrase word order

Virtually all creoles display a number of preverbal elements marking tense, mood and aspect (TMA). There are few exceptions such as Yilan (Taiwan Creole Japanese), which indicates time with adverbs (which, in turn, could grammaticalize into TMA markers). Creole preverbal markers have been analysed as auxiliaries, particles, proclitics, predicate markers and prefixes. (Winford 1993, chapter 3; Arends 1989; Baker & Corne 1986; Steinkrüger 2006: 6). With the exception of Chabacano (Steinkrüger 2009), at least two and often three of them can be combined in order to express various nuances, and they always appear in the order tense-mood-aspect-stem (see Bickerton 1981, 1984; Bakker et al. 1994; Singler 1990). Usually, the negative particle precedes the TMA markers and follows the subject. Givón (1982) argues that the order is determined by increasing semantic scope: aspect on the verb, mood on the sentence and tense on the discourse.

Some creolists claim that creolization can be compared to processes leading to the genesis of dialects. Both would be natural developments of lexifiers, and therefore, creolization would in no way be deviant from regular processes of change (e.g. Mufwene 2008). Languages without a pidgin past may exhibit one or two preverbal TMA markers, but only creoles have the possibility to combine them. It should be said that the order of TMA elements does not follow the order of bound morphemes in non-creoles, where the order is mood-tense-aspect-STEM, or its mirror image (Bybee 1985; Bakker 2008). Creoles instead use tense-mood-aspect-STEM
(and no post-verbal markers, except in some Asian Portuguese creoles, Upper Guinea Portuguese creoles and Berbice Dutch).

Negation either precedes (almost all of) the TMA markers, or it is sentence-final, or a combination of both. Muysken & Law (2001: 52) mention TMA-NEG-V as typical for creoles, but that is an error.

Not only is the order considered fixed and typical for creoles, the meanings are too: the tense marker can be called “anterior”, the mood marker “irrealis” and the aspect marker “non-punctual”. There are, however, many exceptions among creoles, as mentioned above. See Singler (1990) for some additional examples.

5.4.3 Serial verbs

Some, but by no means all, creoles have serial verb constructions. We discuss them here because they are often discussed as typical for creole languages. These can be directional serial verbs (like *he walk come market* “he walked to the market”), instrumental (*he take knife cut bread* “he cuts the bread with a knife”), beneficiary (*she buy book give her* “she bought the book for her”), comparative (*he big pass him* “he is bigger than him”), or completive (*she cut wood finish* “she quit cutting wood”). Serial verbs are typically found only in creoles with substrate languages that also have serial verbs (Seuren 1990; Parkvall 2000: 70–77; contributions to Lefebvre 2011). Thus, verb serialization is not typical for creoles. In fact, most creoles display no serial verbs (see APiCS features 84–86). The three types of serials are found in between 43 and 52% of the APiCS languages.

5.4.4 Ditransitive constructions

Haspelmath and Michaelis (2003) studied ditransitive constructions in creoles and compared the results with a sample of languages of the world, especially the potential substrate languages. Verbal meanings like “send, give, show” can be expressed as direct object constructions (DOC) (“Lea sold Theresa a mango”), as indirect-object construction (IOC) (“Lea sold the mango to Teresa”) or as serial verb constructions (pseudo-English: “Lea sell mango give Teresa”). All three are attested in creoles, and they argue for the DOC construction to be substrate-influenced, as it only appears in Atlantic and Indian Ocean creoles, and not in the Pacific or Asia, thus reflecting the areal predominance of these structures in non-creoles.
5.4.5 Noun phrase word order

Whereas sentential constituent order is relatively homogeneous in creoles, noun phrase word order is usually close to the order in the lexifier (Baptista & Guéron 2007). Mühlhäusler (1986) is a major overview of ordering patterns of nouns and adjectives, showing that, generally, the Romance creoles inherit the dominant N-ADJ order of French, Portuguese and Spanish, whereas the Dutch, English and German creoles display the ADJ-N order typical of the Germanic lexifiers.

Mühlhäusler (1986) surveyed 28 pidgins and creoles. 13 of them displayed a "mixed system" (as in many Romance languages, where a subset of high-frequency adjectives are preposed to the noun, i.e. these languages are both ADJ-N and N-ADJ, depending on the specific adjective), 12 of them have the word order ADJ-N and three N-ADJ. The N-ADJ languages in his sample are the pidgincreole Hiri Motu, displaying the same order as in the lexifier Motu (where adjectives are not a class in themselves) and the Portuguese creoles São Tomé and Papiá Kristang, reflecting the dominant order of Portuguese.

In line with Mühlhäusler (1986), the survey of 18 creoles worldwide in Holm & Patrick (2007) reveals that all the Romance lexifier creoles have both N-ADJ and ADJ-N order, but also Tok Pisin and Nagamese.

Demonstratives and determiners follow or precede nouns, or both, and numerals precede nouns. APiCS data reveal that prenominal demonstratives are more common (56 APiCS languages) than the opposite order (34 APiCS languages), with few languages displaying both orders.

Parkvall (2000: 62–63) surveys adpositions and finds that the Atlantic creoles display prepositions rather than postpositions, in line with the Greenbergian implicational universal that VO languages tend to have prepositions rather than postpositions (Greenberg 1963). This is confirmed by APiCS data (Huber et al. 2013a), which show that all creoles among the APiCS languages have prepositions, and only a few also have postpositions.

However, despite this, some SVO creoles have postpositions rather than prepositions. In Holm and Patrick’s (2007) sample, Nagamese and Berbice Dutch are the only creoles with postpositions, but Berbice also has prepositions (Kouwenberg 1994: 192 ff.). The Surinamese creoles have both, or rather a combination of prepositions and postpositions (archaic Sranan na hosini – LOC house in). Considering the fact that postpositions are common in many substrate languages of the Atlantic creoles, pre-nominal adpositions may be a typical property of the lexifiers, and the SVO order can perhaps be directly connected to the pidginization or creolization process.
5.4.6 Attributive possession

Attributive possession concerns the expression of genitive constructions, in which a relationship of ownership between a nominal, usually animate, possessor (PR) and a possessed item (possessee, abbreviated PE), is sometimes expressed in combination with a marker (M). Three English examples are: (a) her car (PR-PE), (b) John’s nose (PR-M-PE), (c) the money of my friend (PE-M-PR).

APiCS data show that possessor-possessee order is less common than the reverse (Huber et al. 2013b), with 47 to 56 languages – but many allow both orders. Attributive possession has been studied in more detail for a sample of 44 pidgin and creole languages by Heine and Kuteva (2001). They found only seven types of attributive possession in pidgins and creoles, out of 14 theoretical possibilities: two constructions without a marker: PR-PE (type A), PE-PR (type B), three with a fixed connective marker: M-PR-PE (G), PE-M-PR (C) and its mirror image PR-M-PE (E) and finally two constructions with distinct (usually suppletive) markers for singular and plural possessors, PR-M-PE (D) and M-PE-PR (F).

Most individual creole languages make use of several different constructions. Types E, F and G are rare in creoles, and all creoles that have these constructions also have alternative ways of expressing attributive possession. The PE-PR (type B) construction is only found in Romance-based creoles in their survey. By far, the most common construction belongs to type C, found in 28 creoles, and in some pidgins, out of the sample of 44 languages.

Rijkhoff (2004: 194–196) investigated the order of possessor and possessee in a sample of 50 non-creole languages. 29 had the order PR-PE, 10 the order PE-PR and 11 a combined order PR-PE-PR. 18 of the languages had a marker, but it is not clear what the position is with regard to the PR and PE elements. It is clear that the frequencies are quite different: 28 out of 44 creoles (64%) have the order PE-PR, but only 22% of the non-creoles.

5.4.7 Predicative possession

By far the most common way of expressing predicative possession is with a verb meaning “to have, to possess”, although alternatively, some creoles can also use a construction involving a genitive phrase to the same effect (e.g. with “it is to me”, see Stassen 2013). Korlai Portuguese creole of India forms an exception, in that it only has a locative way of expressing possession, along the line of “near me is X” (Clements 2013).
A verb “to have” is only found in 26% of the world’s languages (Stassen 2013). In Asia a verb meaning “to have” is very rare. Here again, creoles differ from non-creoles in proportions of types.

5.4.8 Summary: Creole constituent order

Whereas NP order shows little or no deviation from the lexifier except sometimes in the position of demonstratives and adpositions with regard to the noun, the verb phrase is usually radically restructured: an order S-NEG-TMA-V-O predominates. There are also SOV and VSO creoles. Serial verbs, ditransitives and possessives show a variety of constructions, often linked to substrate influence.

5.5 The creole lexicon

The lexicon of creoles has not been studied systematically from a cross-linguistic point of view, and this section is therefore more a summary of some previous work, and contains some suggestions for future studies.

There are a number of questions that have been asked with regard to the lexicon of creoles. First, creolists have discussed whether the lexicon of creole languages is more or less mixed than the lexicon of non-creole languages. Second, there has been a discussion on the size of creole lexicons: are there more or fewer words or roots in creoles than in other languages? Related to this is the question of expansion of the lexicon: how can one create new meanings on the basis of a set of existing words and meanings? Third, there is an idea in the creolist world that a creole language can replace all of its lexicon, perhaps with the exception of a few traces, towards a different lexicon. This idea has been called the relexification hypothesis. There are in fact three distinct meanings of the term relexification, to be discussed below. Fourth, creolists have attempted to find traces of substrate languages in creoles, i.e., when discussing lexicon, vocabulary from – in the case of Atlantic creoles – African languages. Not all of those West African patterns are direct borrowings of words, they include collocations e.g. “eye water” for “tear”. Fifth, one can ask whether creoles behave like other natural languages in the range of basic meanings, in the sense of atoms and molecules, following the Natural Semantics Metalanguage framework (e.g. Goddard 2011). Next, Holm (1988, chapter 3) has pointed out that creoles often display archaisms, nautical terms, vulgar terms and traits from certain regions in the homeland of the lexifiers (Holm 1988: 100–101). Finally, the words may have quite different meanings from the cognates of the lexifier due to influence from the substrate languages. The meaning of the word for “hand” for instance may also be extended to mean “arm” (no fewer than 43
of 75 APiCS languages), or “father” may also mean “uncle, father’s brother”. All lexifiers have a word for “tear”, while 38 of 66 APiCS languages (55%) have a mono-morphemic word for “tear”. Mono-morphemic words are more common in Asian contact languages, bimorphemic ones in the Caribbean creoles.

5.5.1 The lexicon: Mixedness

If one counts all the dictionary entries of a language like Spanish, Swedish, or Dutch, borrowed vocabulary as a rule exceeds inherited vocabulary. Perhaps up to 80% of all dictionary entries are not inherited from the Proto-languages. To what extent this is also true for languages not influenced by foreign vocabulary of a learned nature or special register (Greek, Latin, Arabic, Persian, Russian, Chinese, depending on what part of the world it is spoken), or a lexicon obtained through trade and contact, is not clear. These same languages, with more borrowings than inherited words, however, show very little admixture in their basic vocabulary (see below).

It has been observed before that the proportion of loans in languages increases the larger and less basic the lexicon is (Grant 2010). This appeared to be true for all of the 41 languages in the typologically-broad sample used by Haspelmath and Tadmor (2009).

Here, creoles tend to differ from non-creoles. Mauritian creole derives only 10% of the lexicon from non-French sources (Baker 1994: 66), and in the basic lexicon it is 99% French. Tadmor (2009: 57) compares systematically a sample of 41 of the world’s languages, for which translation equivalents were sought for over 1000 words. The actual number of words ranged from 982 (Gawwada) to 2158 (Otomi). The one creole included, Seychellois, ended up as no. 36 of 41, with only 10.7% loanwords, only surpassed by Ket (9.7%), Manange (8.3%), Old High German (5.8%) and Mandarin Chinese (1.2%). If Seychellois is representative for creoles, this would indicate that creoles are less lexically mixed than most languages. The basic 200-word vocabulary list of most languages usually contains at least a few loanwords, but there are as a rule fewer loans in creoles than in non-creoles, despite the fact that creoles are often considered “mixed” due to their supposed multiple ancestry. Mauritian French creole, for instance, has no non-French items among the 200 list of basic vocabulary (see below).

5.5.2 The lexicon: Quantity of roots and words

It is notoriously difficult to count the exact number of words or roots for a language. Perhaps the best documented lexicon of any creole is that of Mauritian. The dictionary of Baker and Hookomsing (1987) contains 18,000 entries, of which at least 15,000 are individual words (around 13,500 are from French, Baker 1994).
On the other hand, it has been estimated that only 700 English roots ended up in Sranan English creole (Smith 2001: 53). Unfortunately, detailed studies are lacking, which prevents any further generalization on the topic.

5.5.3 Expansion of the lexicon

If a language has fewer roots, speakers can reach full expressivity by using morphological devices. These are found in many creole languages. Quite a few creoles make use of zero-derivation, resulting in one root functioning both as a noun and as a verb (this is of course also found in non-creole languages, e.g. English and Salish). For instance, Saramaccan dëê´wojo (itself a compound of “dry” and “eye”) means both “brazen” and “brazenness”. The root lânga “long” also means “to elongate”, and hânsa “pretty” (< handsome) means both “pretty” and “to make pretty”.

Serial verbs can also be used to expand the lexicon, e.g. Tok Pisin painim “to search” (from English find him), painim pinis “to find” (Mühlhäusler 1997: 158).

Where the lexifier often has separate roots, the creole may have a compound, e.g. phrases may also be used originating from circumscriptions, e.g. kuku ania gauna “pipe”, literally “smoke eat thing” in the Hiri Motu pidgincreole (Mühlhäusler 1997: 158).

Hancock (1974) considers creoles as displaying a “closed system”, which leads to such languages facing special challenges in lexical expansion. After pidginization, full expressability is reached through expansion of meaning, compounding and circumlocution.

5.5.4 Substrate

Almost all if not all creoles have a number of words from their so-called substrate languages, e.g. African languages in the case of Atlantic creoles, or Austronesian items in the case of Pacific creoles, and Dravidian or Indic in the case of Indian creoles. As discussed above, the number of words derived from a substrate may be very small in a creole. Baker and Hookomsing’s dictionary (1987) of Mauritian Creole contains 15,000 entries, and only 194 of these are African (out of 662 words that are not of French origin). Baker (2012a) estimates 10% of the vocabulary to be from sources other than French. Farquharson and Baker (2012) identified 306 words of African origin in Jamaican, and also Baker (2012b) identified more than 300 African words in Haitian. No total vocabulary is mentioned, but these numbers constitute probably just a few percent of the total lexicon of these languages. In some languages, these substrate words refer mostly to flora, fauna and religion (Bartens & Baker 2012) and in other languages, mostly maroon creoles, they may
enter the basic vocabulary (e.g. Angolar and Berbice Creole). As this is mostly of interest for historical linguistics and not so much for typology, we will not discuss the matter further here. For discussion, see e.g. Muysken & Smith (1986), Mufwene & Condon (1993) and Parkvall (forthcoming).

5.6 Conclusions

There are no linguistic properties that are unique to creole languages. Phonologically, creoles avoid extremes and are average languages. Creoles never inherit contextual inflection from their lexifiers, but they may inherit or develop inherent inflection. Compounding is common. Creoles tend to have a fixed order of constituents, often SVO, with the possibility of movement of the constituents. Despite the diversity of creoles, it is possible to present generalizations that cover most creole languages, including creoles of both the Atlantic, Indian Ocean and the Pacific, as well as creoles with non-European lexifiers (see also Bakker 2014).

References


Bakker, P. 2003a. Pidgin inflectional morphology and its implications for creole morphol-

Kluwer. doi:10.1007/0-306-48223-1_2


Bakker, P. 2008. The development of tense, mood and aspect in creole languages and the typol-


Chapter 5. Creole typology I


Maddieson, I. 2005c. Syllable structure. In Haspelmath et al. (eds), 54–57.


Chapter 5. Creole typology I

doi:10.1016/0024-3841(79)90005-6


Tadmor, U. 2009. Loanwords in the world’s languages. Findings and results. In Haspelmath & Tadmor (eds), 55–75.


CHAPTER 6

Creole typology II
Typological features of creoles: From early proposals to phylogenetic approaches and comparisons with non-creoles

Aymeric Daval-Markussen and Peter Bakker
Aarhus University

In the late 1950s, creolists started drawing lists of shared lexical and grammatical properties of creole languages. In this chapter, a number of proposals and features are presented. We show that few of the proposed common properties are in fact shared by all creoles. We also discuss typological studies in which samples of creoles are compared with samples of non-creoles. Creoles, despite their diversity, tend to cluster in a different pattern from non-creoles, even when non-European creoles are added, providing robust evidence for a creole typological profile. Finally, the chapter deals with the neglected area of diachronic aspects of creoles. No systematic surveys comparing language change in creoles and non-creoles have yet been conducted. The pre-existence of a pidgin phase is a plausible explanation for some of the major differences observed between creoles and non-creoles.

6.1 Introduction

This chapter examines work that has attempted to show that creoles share certain grammatical properties. Recent studies about a special typological profile for creoles are also discussed.

We will start with an overview of general claims made about creole structures, from the 1950s to the early 2010s (in 6.2). We will deal with several sets of proposed pan-creole features that would distinguish creoles from their lexifiers or from non-creole languages in general. We will limit the discussion to the most influential of those proposals. Thereafter, we will provide a general overview of creole structures, focusing on recurrent structures as proposed by a range of creolists (in 6.3). Here we discuss mass comparisons between creoles and non-creoles and the results of studies using computational techniques. We will attempt to come up
with a minimum set of features distinguishing creoles from non-creoles. In the final section, we will compare changes in creolization with changes in non-creoles (in 6.4) before concluding this chapter (in 6.5).

6.2 Claims about typological properties of creoles, 1950s–2000s

Different sets of features have been proposed as being typical of creoles, all suggesting the existence of a typological profile. Early attempts were characterized by a focus on issues of genetic relatedness (e.g. Taylor 1956; Hall 1958). Students of creoles in the 1950s sought to explain the observed similarities by a shared history of the creoles (see e.g. Hancock 1975 for a discussion). In a later phase, the focus moved away from shared history and turned to the socio-historical circumstances underlying creolization (Holm 1988, chapter 2). This included an acquisition history without assuming a common origin or the existence of a proto-creole ancestor. The search for structural similarities defining creoles as a class, in opposition to non-creoles, is a later development, initiated by McWhorter (1998), and further refined in later publications (McWhorter 2001; 2005; 2011).

The quest for common properties of creoles derived from several lexifiers can be traced back at least to Taylor (1956), who observed the recurrent order of preverbal elements in Haitian French Creole, Sranan English Creole and Dominican French Creole. In all three languages, the order is (1) negative – (2) PAST.TENSE – (3) FUTURE – (4) PROGRESSIVE – (5) STEM. A category ADVERBIAL is found in different varieties of Haitian and Dominican, and no data are provided for Sranan. The ordering of Tense-Mood-Aspect elements has long been a recurrent topic in creole studies.

We will now present the proposed sets of pan-creole features.

6.2.1 Taylor (1971)

One of the first scholars to propose a concrete list of typical creole features was Douglas Taylor (1971), who selected 15 features under 12 headings. These are the general creole properties according to Taylor:

1. The third person plural pronoun is also used as plural marker in noun (‘house them’/‘them house’ = ‘houses’)
2. The combination of PAST and FUTURE markers expresses conditional (‘been go’ = ‘would’)
3. The serial verb ‘give’ is used as dative preposition (‘I buy book give you’)
4. Question words are compounds (e.g. ‘what place’ instead of ‘where’)
5. Possessive absolute is expressed with a preposition plus pronoun (not ‘ours’, but ‘for/of us’), or with a nominal phrase
6. Demonstratives follow the noun (‘house this’)
7. Definite articles follow the noun (‘house the’)
8. Possessive pronouns follow the noun (‘house me/my’)
9. Reflexives are expressed with ‘my body’ (‘I see my body’ for ‘I see myself’)
10. The iterative/habitual meaning has the same form as the completive, progressive or future form
11. _na_ is a general locative preposition, meaning ‘in, on, at, by, from, to’
12. _ma_ is the form for the disjunctive ‘but’

Taylor had noted that several structural properties were shared between a number of creoles, but not with their lexifiers. Thus, all the chosen features deviate more or less from (at least some of) the lexifiers. Taylor’s creole “universals” are a mixture of results of specific grammaticalization paths (1, 2, 3, 4, 5, 9), recurrent constituent orders (6, 7, 8), semantic properties (2, 10) and lexical forms (11, 12).

Taylor tested his 12 features against a sample of 13 creoles. It appeared that none of the features is shared by all of the creoles in his sample (Taylor 1971: 295). The number of creole languages displaying these traits varies from 10 (features 1 and 2) to a mere three (feature 8). Higher scores are found in the lexical domain (feature 11 scores 12 and feature 10 scores 13). If we take a closer look at the languages, it appears that Haitian Creole conforms most to Taylor’s set (11 out of 13 features), followed by Gulf of Guinea, Suriname and Lesser Antillean French Creoles. Jamaican, Krio and Guinea Creole Portuguese share five traits and Indian Ocean French scores lowest with four.

Hancock (1975) tested Taylor’s features against 14 Iberian creoles/pidgins: Papiá Kristang, Macau, Java, Sri Lanka, Bombay, Ermitéñio from Asia, six creoles from Africa (Annobon, Sãotomense, Angolar, Principense, Guiné and Cape Verdean), as well as Palenquero, Saramaccan and the pidgin Lingua Franca. If we only look at Taylor’s 13 grammatical features, the Atlantic creoles all score between 7 and 9 creole features, except Cape Verdean/Guinea-Bissau with only 3. The Asian creoles are less creole-like and score much lower (between 3 and 6). Palenquero scores 3, and Lingua Franca 2.

Grant (2009) later used Taylor’s features against another sample of 10 European-lexifier creoles and found a similar spread, from Gulf of Guinea creoles with 12 features out of an expanded set of 16 to Berbice Dutch with merely two features, the others spread between these extremes. Strikingly, not a single of Taylor’s typological features is found in a majority of the creoles (the maximum being six out of 15).

One can conclude that the diversity among creoles on the basis of these recurrent properties is large: there are no clear regional or lexifier-based patterns, and the number of features for each creole differs considerably, including features only found in a minority of the sample.
6.2.2 Markey (1982)

Markey (1982) discussed the question whether Afrikaans is a creole and formulated 11 traits he considered typical for creoles. This was tested against one creole, Virgin Islands Creole Dutch. The features he considered are given below (rephrased).

1. Creoles have no nominal gender
2. Creoles have a plural marker derived from third person pronoun (“nominal number by anaphora”)
3. Creoles have SVO order
4. Pronouns distinguish number, but rarely or never gender or case (“anaphoric unity”)
5. Creoles lack nominal case
6. Creoles do not innovate comparatives (“non-indigenous comparatives”)
7. Creoles have three tense-mood-aspect markers
8. Creoles have no non-finite verb forms (participles)
9. Creoles lack inflectional passive constructions
10. Creoles have one copula for existence and possession, but another one for location
11. Negative markers precede the verb and are analytic, and they lack a number of other properties (details not relevant here)

Only one of Markey’s features is also found in Taylor’s list (feature 2 in Markey is identical to feature 1 in Taylor). Note that many of Markey’s features are negative, i.e. they are observations on features that creoles lack (1, 5, 8, 9) – these are of course statements that are inspired by their presence in the lexifiers, as there are many non-creole languages in the world that lack these same categories. Such a view of creoles was prevalent at the time, and such differences with the lexifier languages are of course remarkable, but they may give a distorted picture of creoles.

All of Markey’s features were found in Virgin Islands Creole Dutch, and only features 2 to 4 are attested in Afrikaans. Markey’s sample is of course too small to make broad generalizations on which features characterize creoles as a group, but it clearly presents a set of properties that were assumed to be typical for creoles.

Some of Markey’s features contradict what can be observed: for instance, both Afrikaans and Virgin Islands Creole Dutch form the comparative like Dutch (Stolz 1987), but many other creoles use serial verbs (like ‘Mary big pass John’) for this purpose. One third of the languages in the Atlas of Pidgin and Creole Structures, (Michaelis et al. 2013, APICS) form the comparative with a verb meaning ‘to exceed, to surpass’), a construction which is absent in the lexifiers.
6.2.3 Bickerton (1981, 1984)

In several publications, Bickerton has enumerated a number of features that recur in many creoles, and related them to biological specifications in the human brain, a so-called language bioprogram. Bickerton (1981) is the most elaborate study. Typical creole features according to Bickerton (1981) are as follows:

1a. Constituents can be moved to the front for emphasis (‘the house I buy (it)’)
1b. Fronted verbs leave a copy of the verb at the original site (‘buy I buy a house’)
2. Creoles have determiners that indicate specific versus non-specific, and pre-supposedly known versus unknown
3. Creoles have tense-mood-aspect systems that consist of a core of three elements with specific meanings (3(a) non-punctual, 3(b) irrealis, 3(c) anterior) that appear in a specific order before the verb and with specific meanings of their combinations
4. Creoles have two complementizers, distinguishing between realized and non-realized complements
5. Creoles have no relative pronouns (‘the house I buy big’), but realizes a pronoun in indefinite relative clauses
6. Creoles have multiple negation (‘nobody no see nothing’)
7. Creoles have separate copulas for existential sentences (‘here get mountain’), which is the same as for the possessive (‘she get car’)
8. Creoles have a copula in locative constructions (‘mother there/sit in garden’)
9. Adjectives are a type of verbs
9a. Therefore, the copula is not used with adjectives (‘he big’)
9b. But adjectives can be inflected as intransitive and transitive verbs (‘he big-s’, ‘he big-s me’)
10. Questions end in a question particle and/or have rising intonation. WH-words are fronted
11. Question words are often bimorphemic (‘what-place’ for ‘where’) if not taken directly from the lexifier
12. Creoles have no passive construction; instead often an active construction is used (‘they build house’), or no diathesis (‘rice eat’ = ‘the rice is eaten’)

In other publications, Bickerton suggested that serial verbs could be the result of the same universal processes occurring during creolization, even though not found in all creoles, whereas the common SVO order in creoles was suggested to be an historical accident (see also Veenstra (2008) for an assessment of the impact of Bickerton’s work).

Bickerton’s features are quite explicit and almost all of them connect specific semantic distinctions to specific syntactic properties. Remarkably, Bickerton’s features show only very little overlap with the ones given by Taylor and Markey (Taylor’s
feature 4 and Markey’s feature 10(b) correspond respectively to Bickerton’s features 11 and 7b). Several people have tried to apply Bickerton’s features to sets of creoles.

Bakker (1987) studied them for Saramaccan and found that almost all are true for Saramaccan (1–4, 7–12), but often Saramaccan makes more semantic or syntactic distinctions than Bickerton’s features specify.

Mühlhäusler (1997: 200–202) tested them against Tok Pisin and Unserdeutsch, a German-based creole of Papua New Guinea, as well as German and English, and concluded that 8 out of 12 features are found in Tok Pisin, but only three or four in Unserdeutsch, against two or three in German and English. All twelve are found in Hawaiian Creole.

Romaine (1988: 65) mentions unpublished work describing the language of Norfolk/Pitcairn. At least nine of Bickerton’s and Markey’s features are found in the semi-creole (among them 7, 9, 10, 11 and possibly 6).

Bender (1987) applied a similar set of Bickerton’s features (15 features with subsets, totalling 30 features) to three European languages (English, French, German), five African languages (Amharic, Kunama, Nubian, Songhay, Swahili) and three contact languages, Pidgin Italian of Eritrea and the creoles KiNubi and Cameroon English Pidgincreole (also called Camtok). All of the African languages scored high on their scale of ‘creoleness’ (between 9.25 and 12, except for Swahili with 5.25) while the European languages scored low (2.5 to 4.25). In the case of the contact varieties, the Italian pidgin and Kinubi scored low (respectively 3.25 and 4.5) and Camtok high (16). For some languages only few data were available.

Robertson (2006) used a Bickerton-inspired set of features to test the creole status of Berbice Creole. He mentions 11 features, but only discusses eight of them, and these eight are as valid for Berbice as they are for Caribbean creoles.

6.2.4 Baker (2001)

Philip Baker (2001), in a seminal but little-cited article, took issue with the claims that creoles do not descend from pidgins, as was advocated by e.g. Chaudenson (1992), DeGraff (2003: 398) and Mufwene (1996, 2015a). It is true that documents with pidgin data are often absent for early stages of Atlantic creoles in the Americas and the Caribbean (see also Chapter 1). Besides, language data are altogether very scarce from the early period in this region.

Baker isolated a number of structural features associated with pidgins, and used a sample of 16 creoles, all with a European lexifier. The 16 languages are (in alphabetical order): Antillais (of Guadeloupe and Martinique, collectively, French-based), Caribbean English Creoles (collectively), Guyanais (French-based), Haitian Creole (French-based), Hawai’i Creole (English-based), Indo-Portuguese (collectively), Louisiana Creole (French-based), Mauritian Creole (French-based),
Papiamentu (Spanish/Portuguese, Dutch Antilles), Pitcairnese (Pacific, English-based), Réunionnais or Reunion Creole (Indian Ocean, French-based), Santome or Sãotomense (Gulf of Guinea, Portuguese-based), Sranan (Suriname, English-based), Tayo (Pacific, French-based), Tok Pisin (New Guinea, English-based) and Zamboangueno or Zamboanga Chabacano (Philippines, Spanish-based).

The features Baker investigated are (simplified, renumbered and reformulated):

1. There is no nominal gender
   1a. Adjectives do not change their form according to gender
   1b. Verbs do not change their form according to gender
   1c. There are no gender-marked definite articles
   1d. There is no gender distinction in pronouns
2. There is no overt number marking, but it can be used optionally
   2a. Wherever the lexical source language has a pronoun which is ambiguously singular or plural, this is exclusively singular in the pidgin, some other form being adopted or constructed as the corresponding unambiguous plural pronoun
3. Tense, modality and aspect are expressed by independent morphemes, not by inflections
   3a. Tense, modality and aspect are expressed by independent morphemes, not by inflections
4. Major word classes lack inflections for case
   4a. Where pronouns have contrasting nominative oblique forms in the source language, only the forms used by the speaker when pointing to the person(s) represented by the pronoun will be found in the pidgin
   4b. Where pronouns have contrasting nominative oblique forms in the source language, only the forms used by the speaker when pointing to the person(s) represented by the pronoun will be found in the pidgin
5. Zero copula in declarative equative sentences
   5a. Zero copula in declarative equative sentences
   5b. Zero copula in declarative locative sentences
6. The definite article(s) of the lexical source language is replaced by the demonstrative(s) from the latter
   6a. The definite article(s) of the lexical source language is replaced by the demonstrative(s) from the latter
   6b. If the lexical source language distinguishes between the indefinite article and the numeral ‘one’, the latter is adopted as the indefinite article
7. Adjectival intensifier: if the usual adjectival intensifier in the lexical source language does not also mean ‘a large quantity’, it is replaced by a word meaning the latter

8. The negator can only occur predicate-initially
9. Interrogatives: monomorphemic interrogatives in the lexical source language are replaced by bimorphemic structures with literal meanings of the type ‘what place’
10. Absence of the most basic monosyllabic locative preposition in the lexical source language as an independent morpheme (English to, French à, Portuguese/Spanish a)
10a. Absence of the most basic monosyllabic locative preposition in the lexical source language as an independent morpheme (English to, French à, Portuguese/Spanish a)
10b. Absence of the most basic monosyllabic genitive preposition in the lexical source language as an independent morpheme (English of, French/Portuguese/Spanish de)
Where these features are found in pidgins, we selected them in order to test whether they have become part of the derived creole, and thus whether they are also typical for creoles. Note that there is a certain overlap between these features and the ones proposed by Taylor, Markey and Bickerton. Baker (2001: 42) observes that “24 Pidgin features in all are listed (...) under ten headings. All 24 are found only in Antillais and Tok Pisin but none of the other languages has fewer than 12 of these. Five of the features are common to all 16 languages (1b, 1c, 2b, 4a, 6b)”.

Réunionnais, Pitcairnese and Indo-Portuguese were the creoles with the fewest pidgin/creole features in the sample, and Caribbean English Creoles, Haitian, Mauritian, Sranan, Antillais and Tok Pisin scored highest. Baker (2001: 42) suggests that a reason for a lower pidgin/creole index could be that the European males in the first three communities all had female cohorts from other ethnic groups, a factor which may have facilitated the transmission of more properties of the lexifier. Baker (2001: 44) considers pidginization the only likely explanation for these observations regarding creoles.

6.2.5 Muysken & Law (2001)

Muysken and Law (2001) introduce issues in creole studies to theoretical linguists. They discuss a number of debates on what is typical for creoles and we can enumerate some of them here. We have selected those that can be formulated as a statement, in our own words.

1. Much of the lexicon is characterized by multifunctionality, i.e. many or all roots can be used as two parts of speech, e.g., nouns and verbs
2. Creoles have TMA particles preceding the verb
3. Adjectives behave like verbs
4. The distinction between nouns and adpositions is not always clear
5. Creoles have very little morphology on verbs
6. Creoles have little or no morphology on pronouns
7. Most creoles have the order Subject ± TMA elements ± NEG ± V ± Indirect Object ± Direct Object ± Adverbial and Prepositional complements
8. They tend to show predicate-clefting, but with much variation between creoles

Most of these have been mentioned before, but the multifunctionality is new, as is the at times problematic distinction between nouns and adpositions. Note that the word order observation is incorrect for the position of the negative element. These features were not tested against a creole sample.
6.2.6 Holm & Patrick (2007)

The 97 structural features selected in the *Comparative Creole Syntax* volume (Holm & Patrick 2007) are by far the most elaborate attempt to characterize creole languages. However, the features are somewhat biased towards Atlantic creoles (Holm & Patrick 2007: vi). The authors divided the creole properties into 20 categories (§§1–20). We reproduce them here.

1. *Unmarked verbs*
   In the Atlantic Creoles, verbs generally indicate tense and aspect not with inflections but rather with preverbal (in some cases postverbal) markers. The tense reference of verbs without such markers, §1, is normally to the time-reference point of the discourse – usually present or past – which is either understood or established at the outset. Whether the meaning of the verb is a state (stative), §1, 1–2, or an action (non-stative), §1, 3–4, seems often to be related to the likelihood of its indicating the past.

   (1) 1.1 Statives with non-past reference
   (2) 1.2 Statives with past reference
   (3) 1.3 Non-statives with non-past reference
   (4) 1.4 Non-statives with past reference

2. *Anterior (or past) tense*
   The marker for the anterior (and in some cases past) tense refers to something that took place before the time-reference point of the discourse, §2. Again, whether the verb indicates a state, §2.1, or an action, §2.2, seems related to the likelihood of its indicating the same time-reference point of the discourse, or a time before that. As can be seen from examples in the following chapters, sometimes the anterior marker can be used in a construction indicating the counterfactual or something that did not happen, corresponding in meaning to the English construction “We would have gone”, §2.3. Such cases may overlap with conditionals, §6.2. In some Atlantic Creoles adjectives, §2.4, and locative phrases referring to place, §2.5, act as a predicate taking preverbal markers such as anterior.

   (5) 2.1 Statives with past reference
   (6) 2.2 Non-statives with (past-before-) past reference
   (7) 2.3 Anterior (or past) = counterfactual
   (8) 2.4 Anterior (or past) with adjective
   (9) 2.5 Anterior (or past) with locative
3. **Progressive aspect**
The marker of progressive aspect indicates that an action is going on; §3.1 reports whether such a marker exists. Sometimes it can indicate future, e.g., “I’m leaving tomorrow”, §3.2. In some Creoles the anterior marker can precede the progressive marker to indicate the past progressive, e.g., “I was leaving”, §3.3. Sometimes, the progressive marker can be used before an adjectival predicate to indicate that a state is inchoative, or just coming into being, e.g. “It’s getting cold”, §3.4.

(10) 3.1 Indicating progressive  
(11) 3.2 Indicating future  
(12) 3.3 Anterior plus progressive  
(13) 3.4 Progressive with adjective = inchoative

4. **Habitual marker**
Markers of habitual aspect indicate that an action recurs, §4. Sometimes this is indicated by the absence of any aspect marker, §4.1, sometimes by the same form as the progressive marker, §4.2, and sometimes by a marker which is only used to indicate habitual aspect, §4.3. Such markers can often be preceded by the anterior tense marker to indicate a past habit, e.g., “He used to go there”, §4.4.

(14) 4.1 Zero marker for habitual  
(15) 4.2 Progressive marker for habitual  
(16) 4.3 Marker for habitual only  
(17) 4.4 Anterior plus habitual

5. **Compleitive aspect**
The marker of completive aspect shows that an action has been completed, §5. In some Creoles such markers can only occur without any other markers – either before or after the verb, §5.1, or possibly with an adjectival predicate, §5.2 – whereas in other Creoles the completive marker can co-occur with the anterior or other preverbal markers, §5.3.

(18) 5.1 Completive only (before/after V)  
(19) 5.2 Completive plus adjective  
(20) 5.3 Anterior (or other preverbal markers) plus completive

6. **Irrealis mode**
The marker of irrealis mode indicates that a state or action is not (yet) a part of reality, §6, i.e. it refers to the future – in some Creoles taking the same form as the progressive marker, §6.1 – or (used after the anterior marker) it refers to the conditional, e.g., “We would go / would have gone”, §6.2. This combination can sometimes also indicate the future in the past, e.g., “We said that we would go”, §6.3, or the future perfect, e.g., “We will have gone”, §6.4.
Chapter 6. Creole typology II

(21) 6.1 Future (= progressive marker)
(22) 6.2 Anterior plus irrealis = conditional
(23) 6.3 Anterior plus irrealis = future in the past
(24) 6.4 Anterior plus irrealis = future perfect

7. Other combinations of verbal markers
Besides the above combinations of verbal markers, some Creoles also feature the co-occurrence of irrealis and progressive markers to indicate continuous future, e.g., “We will be going”, §7.1. Some can precede the latter with the anterior marker to indicate a continuous perfect conditional, e.g., “We would have been going”, §7.2, while some have yet other combinations, §7.3, including modals.

(25) 7.1 Irrealis plus progressive
(26) 7.2 Anterior plus irrealis plus progressive
(27) 7.3 Other auxiliary-like elements

8. Complementizers
Some Creoles have complementizers that mark infinitives, as in English: “We want to go”, §8, while some do not, §8.1. Sometimes the word meaning ‘for’ can serve this function, §8.2, while in some Creoles this word can act like a modal indicating an arranged future or an obligation, e.g., “We are to go”, §8.3. Sometimes this word acts as a subordinator introducing a tensed clause, like English so that: “We did it so that we could go”, §8.4. Many Creoles have a complementizer introducing clauses which is derived from the lexifier’s word for the subordinator that, e.g., “It’s important that we go”, §8.5. Some Creoles have a distinct word to introduce clauses after verbs of saying, thinking, or feeling, e.g., “I know that we’re going”, §8.6. Some Creoles allow omission of subordinators, §8.7.

(28) 8.1 Zero infinitive marker
(29) 8.2 ‘For’ as infinitive marker
(30) 8.3 ‘For’ as a (quasi-) modal
(31) 8.4 ‘For’ introducing a tensed clause
(32) 8.5 Subordinator from superstrate ‘that’
(33) 8.6 Distinct subordinator after verb of speaking
(34) 8.7 Zero subordinator

9. Dependent clauses
Many Creoles have subordinate clauses that are not part of a larger matrix clause, §9.1, or that function within such a clause as a noun phrase, §9.2. Most also have relative clauses in which the relativizer or relative pronoun functions as the subject, §9.3, the direct object, §9.4, or the object of a preposition, §9.5. Some allow certain relativizers or relative pronouns to be omitted, §9.6.
10. **Negation**

Many Creoles have a single particle next to the verb or verb phrase to negate it, §10.1, but some also have another negator at the end of the utterance, §10.2. Some Creoles use negative forms of indefinite pronouns and quantifiers after a negated verb, e.g., “We don’t have none”, §10.3.

- 10.1 Single negation (verbal)
- 10.2 Discontinuous double negation
- 10.3 Negative concord

11. **Passive**

Not all creoles have a passive construction parallel to that in their lexifier, e.g., “That house was built last year”, §11.1; many instead use an active equivalent, e.g., “They built the house last year”, §11.2.

- 11.1 Passive construction
- 11.2 Passive equivalent

12. **Adjectives as verbs?**

In some basilectal Creoles certain non-verbal predicates behave syntactically like verbs, §12: adjectives may take preverbal markers, §12.1, as may nouns, §12.2, and locative phrases, §12.3. For emphasis, some Atlantic Creoles may have predicate clefting, in which a predicate that is either an adjective, §12.4, or a verb, §12.5, moves to the front of the sentence after a highlighter, §13.5, and is followed by the original sentence. Other verbal behavior of adjectives includes comparison as a verb in a serial verb construction, §14, with the word equivalent to ‘pass’, §12.6, as opposed to comparison as an adjective as in the lexifier language, §12.7.

- 12.1 Preverbal markers before adjectives
- 12.2 Preverbal markers before nouns
- 12.3 Preverbal markers before locatives
- 12.4 Predicate clefting: adjectives or adjectival verbs
- 12.5 Predicate clefting: other verbs
- 12.6 Comparison with ‘pass’
- 12.7 Comparison as in superstrate
13. **The copula**

The term ‘copula’ is used loosely to correspond to most uses of English *be*, § 13. Many Atlantic Creoles have different words for the copula before predicate nouns, e.g., “She is my sister”, § 13.1; before locative expressions, e.g., “She is in Kingston”, § 13.2; and before adjectives, e.g., “She is nice”, § 13.3. Some Creoles have highlighters, particles that appear before other words to draw attention to them; these can be used before interrogatives, § 13.4, or in other constructions, § 13.5. The existential meaning of ‘There is’ can be conveyed with the equivalent “(It) have”, § 13.6.

(53) 13.1 Equative copula
(54) 13.2 Locative copula (with expression of place)
(55) 13.3 Zero copula with adjective
(56) 13.4 Highlighter with question words
(57) 13.5 Highlighter with other structures
(58) 13.6 Existential (‘have’ = ‘there is’)

14. **Serial verbs**

Some Creoles have constructions consisting of two or more verbs with the same subject but not linked as they might be in a European language, i.e., with a conjunction or an infinitive marker, § 14. In some cases, one of the verbs means ‘go’ and indicates that the action of the verb takes place in a direction away from the speaker, § 14.1, or means ‘come’ indicating motion towards, § 14.2. Another such construction uses the verb meaning ‘give’ to indicate to whom, or for whom, an action was done, § 14.3. Another uses the verb meaning ‘say’ to introduce a clause after a verb of speaking or thinking, § 14.4. Another uses the verb ‘pass’ to indicate ‘more than’ after verbs, including adjectival verbs in a comparison, § 14.5. In some Creoles the verbs in such constructions can number three, § 14.6, or four or even more, § 14.7.

(59) 14.1 Directional with ‘go’
(60) 14.2 Directional with ‘come’
(61) 14.3 Serial ‘give’ meaning ‘to, for’
(62) 14.4 Serial ‘say’ meaning ‘that’
(63) 14.5 Serial ‘pass’ meaning ‘more than’
(64) 14.6 Three serial verb constructions
(65) 14.7 Serial verb constructions with 4 or more verbs

15. **Noun phrase**

In the noun phrase, many Creoles use nouns without a determiner to refer to the whole category, § 15.1. Some Creoles have an indefinite article, § 15.2, or a definite article derived from a demonstrative determiner in the lexifier language, § 15.3. Some Creoles use the personal pronoun meaning ‘they’ after nouns to indicate
plurality, § 15.4, and some use this plural marker after nouns referring to a person to indicate that person’s usual associates, § 15.5. Most Creoles have demonstrative determiners, § 15.6, but in a few they occur at the end of a noun phrase along with the definite article or plural marker, § 15.7, even when the noun phrase includes a relative clause, § 15.8. In some Creoles adjectives precede nouns, § 15.9, while in others they follow them, § 15.10. Creoles almost never have gender agreement between nouns and their modifiers, § 15.11.

(66) 15.1 Bare nouns (generic, definite)
(67) 15.2 Indefinite article
(68) 15.3 Definite article
(69) 15.4 Plural marker (= ‘they’)
(70) 15.5 Personal noun plus plural marker
(71) 15.6 Demonstrative
(72) 15.7 Demonstrative plus definite or plural
(73) 15.8 Relative clause followed by definite or plural marker
(74) 15.9 Prenominal adjective
(75) 15.10 Postnominal adjective
(76) 15.11 Gender agreement

16. Possession
Some Creoles express possession between nouns without any inflections: the noun that is possessor simply precedes the noun that is possessed, § 16.1. Other Creoles use a preposition meaning ‘of’ (which may be optional) after the noun that is possessed, e.g., “the house of Maria”, § 16.2. Some Creoles indicate possession with a possessive adjective or determiner after the possessor, e.g., “Purchas his pilgrimage”, § 16.3. Creoles may have possessive adjectives or determiners like “my” occurring before nouns, § 16.4, or possessive pronouns of a distinct form like “mine”, § 16.5. In some Creoles possessive adjectives may be replaced by possessive pronouns for emphasis, § 16.6.

(77) 16.1 Nouns: juxtaposition [possessor + possessed]
(78) 16.2 Nouns: preposition [possessed (of) possessor]
(79) 16.3 Nouns: possessive adjective [possessor HIS possessed]
(80) 16.4 Possessive adjectives: prenominal
(81) 16.5 Possessive pronouns: distinct
(82) 16.6 Possessive pronouns as emphatic possessive adjective

17. Pronouns: case distinctions
Creole pronouns may indicate distinctions of case (like subject “he”, object “him”, possessive “his”) less frequently than the pronouns in their lexifier languages do. This section surveys such distinctions in the pronouns meaning “I” in § 17.1, “you
[singular]” in §17.2, “he/she/it” in §17.3, “we” in §17.4, “you [plural]” in §17.5, and “they” in §17.6. Some Creoles have reflexive pronouns that differ from those in their lexifier languages, e.g., “He hurt his body” instead of “He hurt himself”, §17.7. In some Creoles question words like “where?” consist of two morphemes with compositional or semantically transparent meaning, e.g. “What place?”, §17.8. Creoles have relativizers or relative pronouns, but they may differ from those of their lexifiers, §17.9.

(83) 17.1 Personal pronouns: first person singular  
(84) 17.2 Personal pronouns: second person singular  
(85) 17.3 Personal pronouns: third person singular  
(86) 17.4 Personal pronouns: first person plural  
(87) 17.5 Personal pronouns: second person plural  
(88) 17.6 Personal pronouns: third person plural  
(89) 17.7 Reflexive pronouns: distinct forms  
(90) 17.8 Interrogative pronouns: some bimorphemic  
(91) 17.9 Relative pronouns

18. Coordinating conjunctions
Some Creoles have different words for ‘and’ when it joins whole sentences, §18.1, as opposed to parts of sentences, §18.2.

(92) 18.1 ‘And’ joining sentences  
(93) 18.2 ‘And’ joining sentence parts: different

19. Prepositions
Some Creoles have a preposition (or a postposition after a noun) with general meaning of location, e.g., ‘in’, ‘on’, ‘at’, ‘to’, or even ‘out of’, §19.1. Some require no prepositions before places after verbs of motion, e.g., “Matilda run ___ Venezuela”, §19.2.

(94) 19.1 General locative preposition (or postposition)  
(95) 19.2 Zero preposition with motion verbs plus place

20. Miscellaneous
Most Creoles have no inversion of the subject and the verb (or auxiliary) to indicate a question, §20.1. A number use “-o” at the end of a sentence to signal heightened relevance or connectedness between the speaker and the person spoken to, §20.2.

(96) 20.1 Word order: questions SVO  
(97) 20.2 Sentence-final – o
Holm and Patrick’s (2007) book contains 18 sketches of creoles where each chapter is written by specialists of the language, based on these 97 features. The results are summed up as binary scores in tables at the end of each chapter. Not all features have been interpreted in the same way and there are some obvious errors (as can be expected in a work of this calibre). This landmark study was the first investigation in which a broad range of features were gathered for a world-wide sample of creole languages.

The creoles selected are based on seven different lexifiers: Arabic (Nubi), Assamese (Nagamese), Dutch (Negerhollands, Berbice Dutch), English (Jamaican, Krio, Ndyuka, Tok Pisin), French (Dominican, Haitian, Seychellois), Portuguese (Angolar, Cape Verdean, Guinea-Bissau Creole, Korlai Creole) and Spanish (Palenquero, Zamboangueno). Two of them show noteworthy admixture from African languages, Angolar (Portuguese/Bantu) and Berbice Dutch (Dutch/Ijo), and one has been considered both or either Portuguese and/or Spanish (Papiamentu). There is also a wide geographical spread, including creoles from the insular Caribbean (Dominican, Haitian, Jamaican, Negerhollands, Papiamentu), from the Guianas (Berbice Dutch, Ndyuka), from Colombia (Palenquero), from the Atlantic side of Africa (Cape Verdean/Guinea-Bissau Creole, Krio, Angolar), the Eastern coast of Africa (Nubi in the interior, Seychelles in the Indian Ocean), Melanesia (Tok Pisin), India (Korlai Creole, Nagamese) and the Philippines (Zamboangueno).

None of the 18 creoles scored a “+” for all 97 features. Krio, Ndyuka, Guinea-Bissau, Haitian and Jamaican are the five creoles with the highest scores, 81, 77, 73, 73 and 73 respectively. This top-5 includes, impressionistically at least, both creoles that are structurally far removed from their lexifiers (Ndyuka) and others that are fairly mesolectal, like Cape Verdean. They cover three lexifiers and four regions. The five languages with the lowest scores are Palenquero (53), Zamboangueno (53), Korlai (54), Nubi (54) and Nagamese (57) – incidentally, all these languages came into being independently from all the others. These creoles and their features were tested against a sample of non-creoles in Bakker et al. (2011).

6.2.7 Szmrecsanyi & Kortmann (2009)

Szmrecsanyi and Kortmann have published several articles concerning the properties of English varieties (e.g. Szmrecsanyi & Kortmann 2009). They investigated the presence or absence of a considerable number of structural properties found in English and non-standard varieties of English. The varieties investigated were classified either as first language varieties, second language varieties, or pidgins or creoles. Their observations were based on the data collected by a range of authors
for the *Handbook of Varieties of English* (Kortmann et al. 2004). It appeared that creoles in general displayed a number of common properties seldom found in non-creolized varieties of English.

They propose some features that they have observed to recur in pidgins and creoles, and also in some forms of vernacular Englishes. The following non-standard features, for instance, are pervasive in vernacular Englishes (numbers relate to the feature number in the sources):

- [74] Lack of inversion in main clause yes/no questions, as in ‘you get the point?’ (attested in 89% of the varieties in the survey)
- [10] ‘me’ instead of ‘I’ in coordinate subjects, as in ‘my brother and me were late for school’ (87%)
- [49] ‘never’ as preverbal past tense negator, as in ‘he never came’ (87%)
- [42] Adverbs have the same form as adjectives, as in ‘he treated her wrong right from the start’ (85%)

More relevant for us, the following features are found in pidgins and creoles but not, or much less, in L1 varieties.

- [40] Zero past tense forms of regular verbs, as in ‘I walk’ for ‘I walked’ (attested in 91% of L2 varieties and 87% of pidgins/creoles, but only in 30% of L1 varieties)
- [73] Lack of inversion/lack of auxiliaries in wh-questions, as in ‘what you doing?’ (attested in 75% of L2 varieties and 100% of pidgins/creoles, but only in 65% of L1 varieties)
- [6] Lack of number distinction in reflexives, as in ‘they saw it themself’ (attested in 75% of L2 varieties and 80% of pidgins/creoles, but only in 50% of L1 varieties)

The features above were found to be much more frequent in English-based pidgins and creoles. Most interesting for us, Szmrecsanyi & Kortmann (2009) consider the following features typical for (English-lexifier) creole languages:

- [14] Absence of plural marking after measure nouns (‘three house’)
- [44] Multiple negation / negative concord (‘he no see nobody’)
- [3] Special forms or phrases for the second person plural pronoun (‘youguys’, ‘yufala’, etc.)
- [6] Lack of number distinction in reflexives (‘themself’)

Note that all of these phenomena are also found to some extent in second language varieties and, to a lesser extent, in first language varieties too. Several of these have also been observed for creoles based on languages other than English.
6.2.8 Cysouw (2009)

All previously-mentioned studies have taken features supposedly typical of creoles as a point of departure. Cysouw (2009) is an exception. He compared two databases, one made for a typological atlas of the languages of the world (*The World Atlas of Language Structures*, WALS; Haspelmath et al. 2005), and the other a preliminary dataset collected for an Atlas of contact languages (*The Atlas of Pidgin and Creole Language Structures*, APiCS; later published as Michaelis et al. 2013). Cysouw compared 46 features that were present in both datasets and included some 130 languages in the comparison. Cysouw found that the pidgins and creoles cluster together, but in that cluster we also find some non-creoles such as Yoruba (West Africa), some South East Asian languages (Vietnamese, Thai, Indonesian), but also all the Indo-European languages present in the sample.

Unexpectedly also, the intertwined language Media Lengua is found among the creoles. In another group of languages, we find more contact languages such as the intertwined languages Ma’a and Michif, the Bantu contact language Lingala and the two restructured forms of Portuguese and Malay spoken on Sri Lanka, as well as a pidgin, Yimas-Arafundi pidgin of New Guinea. These results do suggest that creoles stand out as a group, but that some relatively isolating languages of South East Asia and European languages share a fair number of properties with them too.

The results, based on a comparison of features found in WALS and the preliminary features from APiCS, were used by Cysouw (2009) to extract a partial typological profile of creoles. The following 13 features were argued to define creoles/contact languages and set these apart as a typological group. According to Cysouw (2009), APiCS languages have (numbers refer to WALS feature numbers):

1. Prepositions (not postpositions) (85)
2. SVO (not SOV) word order (81)
3. Mixed position of content interrogatives (not non-initial) (93)
4. Mixed order of noun and adjective (not Noun-Adj) (87)
5. Mixed order of noun and genitive (not Gen-Noun) (86)
6. Obligatory pronouns (no subject affix) (101)
7. Mixed morphological plural (no plural affix) (33)
8. Mixed negative marking (no negative affix) (112)
9. Comitative = instrumental (52)
10. ‘and’ = ‘with’ (63)
11. Indefinite article = ‘one’ (38)
12. Generic-noun based indefinite pronouns (46)
13. ‘have’ possession (117)
Remarkably, the creole features found by Cysouw are not among those traditionally mentioned in the creolistic literature (with the exception of 2, and 11). Besides, five features are characterized by being defined as “mixed”, i.e. variable with no dominant pattern across creoles. Remember, however, that not only creoles are found in APiCS, but also pidgins, mixed languages and converted languages. However, more than two thirds of the languages in APiCS are uncontroversially creoles.

Daval-Markussen (2013) did a similar test, but he limited himself to just the 14 features that were directly comparable between APiCS and WALS (see Section 6.3.3. for some results).

6.2.9 Mauritian Creole and proposed creole features (Grant & Guillemin 2012)

In most of the tests in which suggested creole features played a role, a proposed set was applied to a range of languages. Grant and Guillemin (2012) approached it from the other direction, as they combined a range of proposed sets of features and applied it to a single language, Mauritian Creole French (MC). If Mauritian is viewed as a typical creole, one can then see which of the sets of features fits best. This, of course, is very approximative, because, as we have seen, creoles are quite diverse.

For the Holm and Patrick (2007) features, “MC shows 69 positive creole features and 28 negative ones”, i.e. 71% of the features. It appears that African and Asian creoles score lower, and Caribbean creoles higher. This is not surprising, as the authors acknowledge a bias towards Caribbean English creoles in the compilation of the feature sets (Holm & Patrick 2007: v). A similar result is found for the Bickertonian features, summarized in 14 features by the authors. Nine out of 14 features score positively, i.e. 64%. Taylor’s 18 features (combining Taylor 1971 as discussed above with a few additional features from Taylor 1977) fare less well, with a score of 7 out of 18 (39%). However, two of them are purely lexical (\(na\) as a preposition, \(ma\) as a conjunction), so 44% would be a more fitting typological score – still not very high. Grant and Guillemin also take the 20 word order features discussed in Dryer (1992) into consideration, but since these are not proposed creole features (nor are they for that matter contrasted with non-creoles), we will not include them here.

6.2.10 Summary of structural overviews and conclusions

These overviews of features were not the only ones made at the time. Romaine cites an unpublished study in which 31 supposedly typical creole features (Romaine 1988:65–66), among which those proposed by Bickerton, were tested against 18
creole and non-creole languages. Fewer than five out of 31 were found in English, French, German and Swahili, and between 16 and 30 were found in the four selected creoles (Cameroon English Pidgincreole, KiNubi Arabic Creole, Hawaiian Creole English and Guyanese English Creole). All other languages (including a pidgin, namely the Simplified Italian of Eritrea) fall in between.

In all cases, features were selected that distinguish creoles from their lexifiers, building on the most striking differences. The perceived typological distance between the creoles and the European languages/lexifiers may be a consequence of the selection of features: those structures in which creoles deviate from their lexifiers receive most attention in all samples. According to these criteria, the special position of creoles among this limited sample of the languages of the world is clear.

These lists of features of creoles, and the studies applying them to other languages, suggest that creoles are not a homogeneous group. Several languages that are classified as creoles do not display many of the properties assumed to be typical for creoles. In short, creoles vary a lot, and a number of non-creoles, especially some languages spoken in Africa, share some of the proposed creole properties.

In the previous sections, we have surveyed a number of proposals for specific properties that have been suggested to be common to many creole languages. It is striking how diverse the proposed features are. Between some of the independently-developed lists, there is hardly any overlap. Another striking feature is that not a single of the proposed features is universal to all creoles, and in many cases not even found in all of the languages of the samples proposed by the authors themselves. Clearly, creoles are quite diverse in their structures but they still display typological similarities.

6.3 Mass comparisons of creoles and non-creoles

In recent years, the availability of massive databases has enabled researchers to deal with enormous amounts of data with the assistance of computers and advanced programs. In this section, we will deal with studies comparing creoles with non-creoles.

6.3.1 Holm & Patrick’s creole sample among the languages of the world

Bakker et al. (2011) took the 97 features from Holm & Patrick (2007; see also 6.2.6 above) and their coordinated and highly-comparable typological database with data from 18 creoles as a point of departure. Data from various sets of non-creole languages from different families were added, skewing the sample towards the most analytic
languages of the major families represented. In a later stage, they also included lexifiers and substrates. The results showed that there was no overlap at all between the creoles and the non-creoles in their sample. This points to the distinctive character of creoles as a typological group, based on features deemed typical for creoles.

Bakker et al. (2011) also took a reduced set of 69 features (called “creole features in the narrower sense”) which was extracted from the 97 morphosyntactic features described in Holm & Patrick (2007) by eliminating those which were not shared by at least two thirds of the creoles (see Bakker et al. 2011: 25 and the online appendix to the article, available at <https://cl.ly/0L0j1d21f32>). These results also support the idea that creoles form a special group of languages when compared with non-creoles, as all the sampled creoles consistently cluster together against the non-creoles, regardless of their lexifier or geographical location. This is the case despite the relative heterogeneity within the set of creoles. This diversity was also observed in the previous section in the wide range of proposed creole properties, few of which happened to coincide between authors.

Bakker et al. (2011) also studied different types of creoles, in order to see whether this, and other factors played a role. The other variables used in the study were age, social history/genesis, lexifier or area (see Bakker et al. 2011: 21–28). Thus, the authors demonstrated that none of the factors that have been claimed to influence the emergence and development of creoles was responsible for the clustering of creoles observed in their survey.

These results showed quite robustly that creoles, despite their heterogeneity, form a distinguishable typological group among the languages of the world.

### 6.3.2 WALS features: Non-creoles and creoles

Parkvall (2008) compared creoles and non-creoles in order to measure the relative complexity of creoles. He selected the languages from WALS for which at least 30 data points were known. The features were selected on the basis of the information they provided about complexity, such as presence or absence of X, or features with more or less Y. He added four features to those in WALS, ending up with 43 typological features. His study resulted in a varied sample of 155 languages, among which are found the two creoles Sango and Ndyuka. Parkvall added feature scores from 32 pidgins and creoles and a constructed language (Esperanto) to the database. He found that creoles and pidgins as a group are found to range among the languages with a lower complexity, measured according to his metric.

Bakker et al. (2011) took the same subset of 43 typological features from WALS that was used by Parkvall (2008) and produced a network including 188 languages based on the descriptions of the Parkvall’s structural features. A partial
reproduction of the network is presented in Figure 6.1. In this network, all the pidgins and creoles (whose names are indicated in upper case, preceded by a ‘C’ for creoles, a ‘P’ for pidgins, while non-creoles start with ‘N’ and are typed in lower case) form a clear cluster in one corner. The only non-creole found in this cluster is Hmong Njua (Nhmo), a Hmong-Mien (Miao-Yao) language spoken in Laos and neighboring countries. This study, based on different samples and different data, shows unequivocally that creoles as a group are distinct from non-creoles. This time, the creole features were not taken as a point of departure, but general features proposed by typologists, not creolists, and the results were quite similar.

![Network Diagram](image)

**Figure 6.1** Fragment of a network based on 43 features for 188 languages, among them 34 pidgins and creoles, showing the cluster consisting of all pidgins and creoles, as well as Hmong Njua (Nhmo).

6.3.3 WALS features and APiCS features compared

Daval-Markussen (2013) selected 14 features that were directly comparable between APiCS and WALS. The first number refers to APiCS features, the second to WALS. The 14 features were: 8–91A, 22–34A, 28–37A, 29–38A, 32–42A, 33–41A, 34–54A, 35–53A, 56–71A, 58–98A, 59–99A, 62–101A, 70–52A and 77–117A (see Daval-Markussen 2013 for details). In a phylogenetic network (reproduced in Figure 6.2) containing all the APiCS languages and the 55 WALS languages for
Chapter 6. Creole typology II

which the features were known, a rough cluster of APiCS languages mixed with European languages can be discerned, and most non-European languages show up in other parts of the graph, but with the inclusion of a few pidgins and creoles. In the resulting network, the creoles (colored in grey) cluster overwhelmingly together:

Figure 6.2 APiCS and WALS languages, based on 14 shared features.

Four creoles are found among the non-creoles: Korlai Creole Portuguese, Ngbandi-based Sango of Central Africa, and Tok Pisin and Bislama, two English-lexified creoles of the Pacific. French and English are found among the French- and English-lexified creoles, respectively. A network with 55 APiCS creoles for 36 features yielded similar results, but this time, the European languages (i.e. English, French, Greek, Spanish, Hebrew, Latvian, Russian and Finnish) ended up in the middle of the creole cluster.

In studies examining features not specifically selected for being typical of creoles, the creoles shared a number of features and clustered mostly separately, and in one case included the European languages. The reason for this result may be
that when supposedly typical creole features are selected, creoles tend to stand out, and so do European languages. When features are selected on different grounds, creoles sometimes stand out together with the lexifiers (except Arabic) and other European languages. This suggests that creoles are rather similar to the languages of the European Sprachbund.

6.3.4 The four WALS features that set creoles apart from non-creoles

In the preceding sections, we have discussed a number of studies that made use of many typological features at the same time, which, as a set, distinguished creoles from non-creoles. One can also ask how few features are sufficient to distinguish creoles from non-creoles.

According to Daval-Markussen (2013), four features would set apart creoles from non-creoles. These are the four WALS features (Haspelmath et al. 2005) that were shared by a majority or all of the creoles described in Holm and Patrick (2007), and by none of the non-creoles in WALS for which the features were known. We will discuss these four features, including how they fare among the non-creole languages of WALS and among the contact languages of APiCS. After that, we will show how they are sufficient to distinguish 18 creoles among some 200 languages of the world. The four features are:

WALS 38A: The indefinite article is derived from the numeral ‘one’ (value 2). Note that this is not a rare feature (cf. Givón 1981), as it is found in 19% of the world’s languages in WALS. In the European languages that have an indefinite article, it is either identical to or a reduced form of ‘one’. Note also that it actually implies two facts: (a) there is an indefinite article and (b) the article is derived from the numeral ‘one’. No exceptions are found among the CCS creoles.

In APiCS, 66 of the 76 languages have an indefinite article (87%). This contrasts sharply with the percentage of WALS languages which have an indefinite article (46% have an indefinite article, Dryer 2013a). The exceptions in APiCS include all pidgins (including Chinuk Wawa, but Grand Ronde Chinook Creole has in fact indefinite articles). The only creole language that has no indefinite articles is Juba Arabic (and Ambon Malay if that is indeed a creole). Hence, it is remarkable how many creoles have an indefinite article which is not derived from the lexifiers’ indefinite article, also when seen in comparison with non-creoles.

The occurrence of this feature in a majority of creoles can be explained as a direct outcome of the creolization process followed by the grammaticalization of the numeral ‘one’ for marking indefiniteness or non-specificity. Studies in grammaticalization have shown that languages that have an indefinite article equivalent
to the numeral ‘one’ are believed to have gone through a similar developmental path (Givón 1981; 1984: 410–411; Heine 1997: 65–82; Heine & Kuteva 2002: 45–46).

WALS 69A: The WALS languages have no Tense-Aspect inflection (value 5). Only 14 % of 1131 WALS languages have no inflection for tense or aspect. The creole value is not found in any Western European language, all of which have some form of tense-aspect inflection. The Western European value is the most common one in the languages of the world, as affixes are found in 59 % of WALS languages. No exceptions are found among the 18 creoles in Holm & Patrick (2007).

Of the 75 languages in APiCS, 67 have some form of tense-aspect marking, but no distinction is made between particles and affixes, and hence it is not directly comparable.

Due to the scarcity of languages worldwide displaying this preference and due to its widespread occurrence in creoles, this phenomenon may be linked to the process of creolization. Indeed, this could be a symptom of the morphological reduction occurring in connection with an initial phase of pidginization preceding the development of creoles, which result in analytic languages. In turn, the most obvious choice to express tense-aspect distinctions would be to grammaticalize lexical items (primarily verbs, adverbs and auxiliaries) carrying the semantics of the various temporal and aspectual nuances and to use them as verbal markers, following a similar grammaticalization path to the one that can be observed in regular language change.

WALS 112A: Negation in creoles is expressed with a negative particle (value 2). This is not rare or uncommon either in the languages of the world, and it is found in most but not all European languages – and in virtually all Western European languages. 43% of the world’s languages express negation with a particle (Dryer 2013b). Other ways to express negation are e.g. negative verbs (Finnish), or morphological negation (Turkish). Among the CCS (Comparative Creole Structures, Holm & Patrick 2007) creoles, Angolar and Palenquero (and Berbice) deviate in having double negation (value 6).

53 of 76 APiCS languages make use of a negative particle. Some creoles also have a negative auxiliary verb as a possibility, beside the negative particle, and three creoles use double negation (Angolar, Fa d’Ambô/Annobon and Santome). Sri Lankan Malay and Sri Lanka Portuguese have a negative affix, but it can be questioned whether these are creoles or rather converted or metatypized languages (Ross 1996).

Hence, since this feature is not at all uncommon in the languages of the world, its widespread presence in creoles cannot be entirely explained as being a unique result of the creolization process.
WALS 117A: Predicative possession is expressed with a verb meaning ‘to have’ (value 5), and not with e.g. a locational or a genitive phrase (e.g. ‘it is at/with/of me’). A possessive verb is the most common strategy in the languages of the world (but in only 26% of WALS languages; Stassen 2013), and the only one found in the European lexifiers. However, some Celtic and Slavic languages in Europe use prepositional constructions with ‘at’. Sudanese Arabic, the lexifier of KiNubi, also uses a locational expression. Berbice Dutch and Krio deviate in having locational possession (value 1, according to Kouwenberg 2007). The data in Kouwenberg (1994: 136, 159, 607), however, suggest that value 5 should be the right one, as the language clearly has a possessive verb ‘to have’. This is also the value given in APiCS.

In APiCS, 66 out of 75 languages have a verb ‘to have’ (80%), and in 48 languages (64%), it is the only way to express predicative possession. Sango (Central Africa), Korlai (India, Portuguese-based) and Tayo (French creole from New Caledonia) are the only creoles in APiCS that diverge from the majority. Sango presents several possibilities, Korlai has only a locational possessive construction (also the converted language Sri Lanka Malay) and Tayo has a genitive construction.

If one applies these four features and looks for WALS languages for which the values of all four features are known, we end up with 116 non-creole languages. In the following graph, the abbreviations of creoles start with a lower-case ‘c’, languages starting with ‘X’ are non-creoles, ‘S’ are substrate languages and initial ‘L’ indicates lexifiers. Figure 6.3 (reproduced from Daval-Markussen 2013) shows that all creoles are grouped together in the lower right.

The algorithm used for producing the graph (Neighbor-Joining, Saitou & Nei 1987) looks for languages most similar to others, and these end up in adjacent branches. The resulting graph can also be displayed as a network so as to reflect conflicts (i.e. incongruent values) in the data set. Still, in Figure 6.3, not all creoles are grouped on a single branch: Angolar and Palenquero appear on a separate sub-branch (because of the double negation). The 18 creoles cluster together, but two non-creoles are also found in the cluster, Motu (an Austronesian language of Papua New Guinea) and Tahitian (an Austronesian language of Oceania). By adding feature 55A (absence of nominal classifiers) to features 38A (indefinite article derived from the numeral ‘one’) and 69A (no tense inflection), the results show a separate group of creoles. However, there is one exception as well, Rapanui (Austronesian, Oceanic), the language of Easter Island, clusters with the creoles. This is shown in Figure 6.4, which sets 19 creoles apart from 154 non-creoles, based on three features.

Thus, there is strong evidence that a limited number of features (three to five) can set creoles apart from most of the world’s non-creole languages. As none of these features is rare by itself on a worldwide scale, we cannot exclude a priori
Figure 6.3 A tree of 18 creoles and 116 non-creoles with four multi-state features.

Figure 6.4 A tree of 19 creoles and 154 non-creoles with three multi-state features (from Daval-Markussen 2013).
that there are other non-creole languages in the world that share the same configuration of these features. Yet, only one (in Figure 6.3) or two (in Figure 6.2) of the languages in WALS for which these features are known have identical scores as the creoles in the sample. The fact that four creoles have one feature that differs from the others is apparently not sufficient to split up the creole cluster. Thus, it may be the case that these three features are sufficient to distinguish creoles from almost all non-creoles with high reliability. However, most creoles in the sample have a European lexifier. We will now add six creoles of non-European origin and include them in the graph.

The non-European creoles that were selected for addition are Chinuk Wawa, Hiri Motu, Nheengatu, Plains Indian Sign Language, Rao and Yilan Creole Japanese (see Bakker 2014a for background). Table 6.1 shows the WALS values for these five features in all six languages.

Table 6.1 Six non-European creoles/pidgincreoles and their scores for five WALS features (pluses indicate values identical to the most common values for creoles).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Creoles</th>
<th>Chinuk Wawa</th>
<th>Hiri Motu</th>
<th>Nheengatu</th>
<th>PISL</th>
<th>Rao</th>
<th>Yilan</th>
</tr>
</thead>
<tbody>
<tr>
<td>38A Indefinite article = ‘one’</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>69A No tense-aspect inflection</td>
<td>+</td>
<td>+</td>
<td>–/+</td>
<td>+ (value 1)</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>112A Negative particle value</td>
<td>+</td>
<td>+</td>
<td>– (value 1, or 2)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>117A ‘have’ possession value</td>
<td>+</td>
<td>+</td>
<td>– (value 1)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>55A No numeral classifiers</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1 shows that none of the six new non-European creoles score unambiguously like the creoles of previous samples. For WALS feature 112A on negation, Nheengatu and Yilan differ from other creoles in that affixes are used rather than a negative particle (value 2). All four of the proposed features of the four-feature test have exceptions in one or two creoles. In some cases, there can be an explanation for the exception. For example, the overall absence of articles in sign languages makes the pidgincreole PISL fit in with other sign languages. For discussion of the details and the exceptions, see Bakker (2014b).
We can add the new creoles to the tree presented in Figure 6.4. Figure 6.5 is a tree of the 18 creoles from the original sample with the six non-European creoles and 116 non-creoles based on four multi-value features. The abbreviations of creoles start with ‘c’, lexifiers with ‘L’, substrates with ‘S’ and other non-creoles start with ‘X’. The abbreviated languages can be identified through Section 3 of the online appendix of the original article <https://cl.ly/0L0j1d2l1f32>. Four out of the six added creoles (the exceptions being PISL and Rao) are found in the creole cluster on two adjoining branches.

In Figure 6.6, we have applied the three-feature test with the original creoles and the six non-European creoles. One branch appears to contain all the original creoles plus Rapanui, whereas Yilan appears separately, as well as PISL and Rao. The bootstrap support values (the numbers on the branches) are very low, which means that the results are limited in their reliability. This is in all likelihood connected with the fairly low number of features selected here. Note, however, that the confidence level of the creole branch is 70, which is much higher than most others.
In short, the inclusion of the six non-European creoles has confirmed their creole status. Plains Indian Sign Language and Rao end up outside the creole cluster (because of one deviant feature each), but Chinuk Wawa, Hiri Motu, Nheengatu and Yilan clearly belong to the creole cluster. However, we cannot be sure that these four features are sufficient to distinguish creoles from non-creoles. We know that they distinguished the 18 sampled creoles as a set from the sampled 116 non-creoles. In order to test this more robustly, we have to use the full set of features which set the creoles apart from non-creoles. Unfortunately, not all data are known for a sufficiently large set of languages, so that is not possible at this stage.

6.3.5 Surinamese creoles, the lexifiers and the Gbe and Kikongo substrates

In a voluminous book, Muysken and Smith (2015) present the results of a long-term and detailed study of the possible impact of the Kikongo and (especially) Gbe languages on the Suriname creole languages. The results are mixed. In some areas, the contributors to the book found clear substrate influences (e.g. in non-iconic reduplications), but in other areas (e.g. verbal semantics), the influence was negligible. Muysken (2015) collected the data from the different authors, covering in all 82 features and 10 languages: Kikongo, three Gbe languages, early Sranan, Saramaccan and Sranan (three creoles) and three European languages.
English, Dutch and Portuguese. His Neighbor-Net graph shows four clear clusters: Kikongo, the European languages, the three creoles and the three Gbe languages. The creoles are neither particularly close to the lexifiers, nor to the substrates.

6.3.6 Summary mass comparisons

In all studies using large quantities of data in which creoles and non-creoles are compared, the two groups end up in different clusters. There are only a few exceptions. In some cases, European lexifiers end up with the creoles. Rarely, a non-creole ends up among the creoles. Even though no perfect separation can be consistently observed between creoles and non-creoles, all results point to a distinct typological profile of creoles. A core handful of features may be enough to distinguish creoles from non-creoles.

6.4 Diachrony and creoles

The preceding sections all dealt with the typology of creoles from a synchronic perspective. A study of changes in which creoles and non-creoles are compared has yet to be undertaken. In this section, we speculate about the role of diachrony for the explanation of the observed structural differences between creoles and non-creoles.

As yet, there do not seem to be changes that are typical or unique for creoles. This is certainly true for phonological changes, and most likely also for semantic, syntactic and morphological changes. Many creoles follow grammaticalization paths that are also found in non-creoles (Bruyn 1995). Heine and Kuteva (2002) list some creole grammaticalizations in their overview, and none of them seems exceptional. However, there are three aspects that require discussion in this context. First, the fact that all or most creoles have their origin in severely reduced varieties such as pidgins, creates a need for the speakers to develop a viable grammatical system, so that the language becomes a fully adequate communicative system. This process may make people adopt a number of morphosyntactic properties that lead to a special typological profile for creole languages, despite their diversity.

Another aspect of special interest is the speed of linguistic change. It has been repeatedly claimed that some changes take place faster in creoles than in non-creoles (Hall 1966), especially the speed of grammaticalization (see also Chapter 1). This is presumably a consequence of the necessity of developing a grammatical frame on the basis of a preceding pidgin, in a short time and in the absence of strict linguistic norms.
Finally, change in creoles differs from change in non-creoles in that there is potentially a strong influence from the first language(s) of the creators of the creoles on the newly-developing code. This has traditionally been called “substrate influence”. We will discuss these aspects below.

6.4.1 Pidgins

Textbooks routinely mention that creoles have developed from a pidgin (Hall 1966, Mühlhäusler 1997; Romaine 1988), but some creolists have raised objections to this (see chapter 2, this volume). Typically, however, no empirical evidence is brought forward to support such claims. Shared structural properties of creoles, both diachronic and synchronic, which cannot be traced to the lexifier or substrate languages, are impossible in their view (cf. Aboh & Ansaldo 2007).

For the Atlantic region, a pidgin stage, or at least a stage in which a European language was severely reduced in lexicon and grammatical means, is assumed, not documented, in contrast to the Pacific. As shown above, the structures of creoles show only limited continuation from the structures of the lexifiers, suggesting that they indeed descend from a severely reduced language, i.e. a pidgin (Baker 2001), jargon or some form of “basic variety” (Klein & Perdue 1997; cf. also Becker & Veenstra 2003), rather than being a descendant of the lexifier. Two well-documented facts point in this direction: the extremely limited inflectional apparatus of creoles, and the fact that creoles, despite their diversity, rank among the least complex languages (Parkvall 2008; McWhorter 2001). Also, in cases where a pidgin stage prior to creolization has been documented, we know what the outcome is, e.g. the English pidgincreoles of the Pacific, Chinook Creole and Sango. Still, the creoles with and without documentation of a pidgin stage are remarkably similar in their structures, so that an assumption of a previous pidgin stage is by far the simplest explanation. These similarities between expanded pidgins and creoles seem to be recognized even by Mufwene (2015a, b), even though his discussion of a possible creole typological profile only cites opponents of the idea (2015a).

6.4.2 Grammaticalization

In creoles, a number of semantic distinctions typically develop, apparently triggered in the expansion phase from a reduced system to a fully-fledged language. Many of the grammatical categories of the lexifier are not transmitted to the creole, and instead, a number of other distinctions tend to develop (see e.g. Taylor 1971; Bickerton 1981; Markey 1982 and others dealt with earlier in this chapter). For instance, TMA distinctions which appear preverbally, a sentence-final completive
marker, a preposition meaning ‘for’ developing into a complementizer and sometimes a mood marker, demonstratives become articles expressing specificity rather than definiteness, an indefinite article derived from the numeral ‘one’, a reflexive expressed with ‘body’, a fixed constituent order (commonly SVO), fronting for emphasis, a plural marker derived from the third person pronoun (Janson 1984), a locative/existential copula derived from ‘to have’ and bimorphemic question words (‘what-place’ rather than ‘where’). Even though all of these phenomena are also encountered in non-creoles, and few creoles show all or most of them, the combination of these features is the cornerstone of the creole typological profile.

6.4.3 Phonological processes

In most creoles, marked segments of the lexifiers have been modified, such as the front rounded vowels /y/ > /i/, /ø/ > /e/, voiced fricatives to unvoiced ones such as /z/ > /s/ and /v/ > /b/, interdentals to dentals, shift between liquids (see Parkvall 2000, ch. 3 for the Atlantic creoles). The loss or reinterpretation of marked contrasts (e.g. English tense/lax; nasal/non-nasal) is reported regularly, which may lead to a relative scarcity of marked phonemes in creoles (Smith 2008: 104). Even though the changes themselves are quite natural and common, the loss of distinction between tense and lax vowels is “not a change that occurs in any historical dialect of English” (Smith 2008: 121), suggesting that phonological changes in creolization can be different from changes in non-creoles. On the other hand, some cross-linguistically uncommon phonemes are also encountered, e.g. co-articulated stops in Saramaccan, interdental fricatives in Angolar, pre-nasalized consonants in quite a few Atlantic creoles with English, Spanish and Portuguese lexifiers. Typically, these appear in African words AND in lexifier words (ndeti ‘night’ in Saramaccan, ntela ‘star’ in Annobon, from Portuguese estrela), suggesting substrate influences in this realm (see Parkvall 2000, ch. 3).

6.4.4 Speed of change

Processes of phonological change seem to happen at an accelerated rate in the process of creolization, when compared to change in other natural languages. If one tries to measure how long ago the split took place between e.g. Cameroon English Creole and English, or Sranan and English, the date of the split is always much deeper than what the historical facts tell us (Hall 1959; Hall 1966: 122–124; Pentland 1979). Hall (1959) used different glottochronological calculations between Neo-Melanesian (Tok Pisin) and English. These resulted in a date of split of between one and three millennia, even though the real split occurred just a century
and a half ago. This suggests that the phonological processes are common, but that the speed is indicative of a catastrophic event such as pidginization and subsequent creolization, with considerable substrate influences. Parkvall (2000) suggested that most creole properties could have jelled within the first 25–50 years. Jansson et al. (2015) made a computer simulation of the genesis of Mauritian French creole, based on actual population data from the settlement period. Their conclusion was that 15 years could be enough for a creole to crystallize. A demonstration of the higher speed of grammatical change in creoles compared to non-creoles, after such an initial period, is still wanting.

6.4.5 Summary: Change

From a diachronic perspective, creole languages differ from non-creoles in that the quantity of changes observed in the creole vis-à-vis their lexifiers wrongly suggest a considerable time depth of split between the lexifier and the creole. In creolization, phonological changes and grammaticalization processes seem to take place at an extreme speed, but this remains to be demonstrated. The assumption of a pidgin stage (cf. Parkvall et al. 2014), even where it is not documented, would explain the many grammatical differences between a lexifier and its creoles.

6.5 Conclusions

In this chapter, we have shown that creole languages represent a quite diverse group of languages, something which is also clear when consulting APiCS (Michaelis et al. 2013). They are so diverse that no single linguistic property appears to be unique to creoles and that scholars who have attempted to come up with lists of typical creole properties ended up with very different lists. In some cases, the proposed creole properties were not even found in a majority of the creoles in their own sample. This diversity contrasts remarkably with the results of studies in which creoles are compared with a sample of non-creole languages of the world. Here, the results always point to a distinct typological profile for creoles. This is the case when existing feature sets supposedly typical for creoles are selected, and also when features are selected that would characterize the languages of the world. In all cases, the creoles and non-creoles are separate groupings. In some selections, European languages appear to end up with creoles, however. These are minor exceptions, as the overwhelming majority of studies involving mass comparisons group creoles separately from non-creoles.
In an attempt to isolate the minimum number of features that would group creoles apart, three, four or five may be enough to distinguish creoles from non-creoles in a statistically significant way – at least for the available sample of non-creoles. In short, there is a growing body of evidence that creoles form a typologically distinct type among the languages of the world.

Note

The datasets for this chapter can be found here: https://phylogenetic-creole-studies.blogspot.com

References


In creole studies, there has been a long tradition of discussing the respective contributions of African languages in the genesis of creoles spoken primarily in the Atlantic and the Indian Oceans. Creolists have often assumed that the African languages that contributed to the creation of creoles were typologically rather uniform (e.g. Singler 1988). Very few studies have compared more than a handful of languages when searching for African influences in creoles (e.g. Parkvall 2000). In this chapter, we investigate the claim that the African languages involved in the colonial settings, and hence in the creation of various creole languages, were typologically homogeneous. Moreover, we assess the extent to which a selection of proposed stable features is shared (i) between creoles with and without African influence, (ii) between creoles and West African languages known to have influenced these creoles, and finally, (iii) between creoles and their lexifiers.

### 7.1 Introduction

In the mid-19th century, the immense linguistic diversity that characterizes the African continent was brought to the fore by the publication of the German missionary Koelle’s monumental *Polyglotta Africana* (1854), which provided word lists in more than 100 languages, mainly from the vast Niger-Congo phylum or superfamily. This was one of the first serious attempts at classifying African languages, and this work laid the foundations for a long tradition of comparative studies, based mostly on lexical evidence (such as Westermann 1927). This tradition culminated with Greenberg’s (1963) genealogical classification of African languages, which continues to be one of the most commonly used classifications today. Although most major higher-level branches of Niger-Congo are agreed upon today, the exact relationships between those, and between many lower-level branchings and individual African languages, are still debated, as the many
controversies and constant revisions in the classifications testify (e.g. Williamson & Blench 2000; Sands 2009; Dimmendaal 2011). Extensive contact in various regions of Africa has occurred over at least a few millennia, resulting in widespread shared lexical and typological features due to long-term convergence, which in the long run has obscured long-distance linguistic relationships (see e.g. Heine & Nurse 2008).

Ever since the beginnings of creole studies in the late 19th century, scholars studying the emergence and development of creole languages have sought to attribute the similarity of creoles to the influence of the contributing languages present in the relevant contact situations (see e.g. Krämer 2013 for a recent review on early creolists such as Adolfo Coelho, Charles Baissac, Auguste de Saint Quentin and Hugo Schuchardt in the late 1800s). While some have emphasized the importance of the lexifier language (for instance Chaudenson 1992; Mufwene 2008), others have accentuated the role of substrate languages in the formation of these new vernaculars (e.g. Lefebvre 1998; 2011). Creolists have often assumed that the languages of the subjugated populations originating from West Africa were typologically rather uniform (e.g. Mufwene 1986; Singler 1988; Holm 1988: 79–89) and this has been seen as a crucial factor in explaining the relative homogeneity of Atlantic creoles. The idea of the homogeneity of African languages is at least as old as Vinson (1884: 346), who considered the languages of enslaved Africans to be “apparentées ou du moins de même forme” (related or structurally similar).

In this chapter, we will evaluate the validity of the claim that the African languages that were involved in the colonial settings, and hence in the creation of most creole languages, were typologically similar. We will first analyse a sample of West African and Central Southwest African languages based on a selection of typological features claimed to be relatively stable over time (Wichmann & Holman 2009). Since these features were most likely inherited rather than borrowed, these are assumed to represent well-suited markers of distant genealogical relationships, even under catastrophic events of the kind encountered in interrupted intergenerational transmission (for a discussion on deep genetic relations, see Hock & Joseph 2009, chapter 17; Campbell 2003). The settings into which creoles emerged and developed were characterized by extreme contact, which leads to contact-induced changes through shift according to Thomason & Kaufman’s scale gauging the effects of contact (Thomason & Kaufman 1988: 50). Because of this, a set of stable features will be used to assess the extent to which these specific features have been transferred to creole languages with European lexifiers. If on the one hand, many substrate features are found in creoles, then these languages can be considered continuations of the African languages (as suggested in Lefebvre 1998). If on the other hand, European structures are present in creoles, then creoles can
be viewed as continuations of the lexifiers (e.g. as claimed by Chaudenson 1992). We will test this by comparing their distribution in samples of creoles influenced by West African languages to creoles not known to have been influenced by these languages, for instance creoles spoken in Asia and Oceania. Finally, we will include the lexifiers of the creole languages selected in our sample in order to assess their influence on the new vernaculars.

7.2 Stable features

When studying creolization, we assume that a break in intergenerational transmission has occurred during the development of the languages involved in the contact situation. This development goes through a phase of substantial reduction of the grammatical system, followed by a phase where grammatical categories are created or recreated. This implies that the various areas of the grammar will evolve at various speeds. This assumption can be directly tested by comparing groups of languages known to have influenced each other in the course of their evolution (in this case creoles vis-à-vis their lexifiers and substrates) based on a selection of stable typological features.

While changes in languages (e.g. in the fields of phonology, morphology, syntax, semantics) have been shown to proceed at different speeds for various parts of speech and areas of grammar (e.g. Pagel et al. 2013), the idea of differential rates of development in different areas of grammar remains a controversial issue. The question has seen a renewed interest in the past decade, partly due to the advent of computational methods in historical linguistic studies. Several recent investigations (Maslova 2004; Cysouw, Albu & Dress 2008; Parkvall 2008; Wichmann & Holman 2009) have provided robust evidence that the evolution of different grammatical features does indeed occur at different paces, even though the results presented in these studies do not always agree on which features are least subject to change (an evaluation can be found in Dediu & Cysouw 2013).

In their investigation of differential stability in structural properties of language, Wichmann and Holman (2009) identified a total of 33 features from the World Atlas of Language Structures (henceforth WALS, Dryer & Haspelmath 2013) as being ‘very stable’ over time, with stability values (SV) ranging from 50.6% to 80.8% according to their metric C. Stability is measured on a scale ranging from minus 100 to plus 100; the lower the figure, the less stable; the higher the figure, the more stable. For our analysis, we retained these most stable features, but with some modifications. Some features are interdependent (e.g. OV languages tend to be postpositional, hence including both features would be redundant) and we
wanted to minimize any potential bias towards a particular area of grammar. Hence, interdependent features were filtered out from the data set and replaced with other 'stable' features from Wichmann and Holman (2009), i.e. features with a SV from 34.1% to 48.3%, related to the same areas of grammar as the original features (for a full overview of the features, see Table 7.1).

The following features were taken from WALS (with the original WALS feature numbers): WALS feature 32A ('Systems of Gender Assignment') was omitted because it relates to gender systems and therefore overlaps with 31A ('Sex-based and Non-sex-based Gender Systems'). Features 81A ('Order of Subject, Object and Verb') and 84A ('Order of Object, Oblique, and Verb') were redundant because of the presence of 83A ('Order of Object and Verb'), since all three features relate to OV-word order. We included instead 82A ('Order of Subject and Verb'), the only remaining stable feature relating to word order. Feature 87A ('Order of Adjective and Noun') was removed and replaced with 88A ('Order of Demonstrative and Noun') due to its strong cross-linguistic correlation with 83A ('Order of Object and Verb'). Finally, 137A ('N-M Pronouns') was omitted since it relates to lexicon and the focus of the present study is typological. Moreover, features for which insufficient data were available were also replaced (i.e. those for which less than 25% of the information was known). As a result, the very stable features 21A ('Exponence of Selected Inflectional Formatives'), 42A ('Pronominal and Adnominal Demonstratives'), 47A ('Intensifiers and Reflexive Pronouns'), 61A ('Adjectives without Nouns'), 79A ('Suppletion According to Tense and Aspect'), 99A ('Alignment of Case Marking of Pronouns') and 121A ('Comparative Constructions') were removed and replaced with the stable features 27A ('Reduplication'), 33A ('Coding of Nominal Plurality'), 48A ('Person Marking on Adpositions'), 65A ('Perfective/Imperfective Aspect'), 69A ('Position of Tense-Aspect Affixes'), 94A ('Order of Adverbial Subordinator and Clause') and 104A ('Order of Person Markers on the Verb'). The stability value of the selected features ranges between 35.7 and 80.8 and all belong to the categories 'stable' or 'very stable' as defined by Wichmann and Holman (2009: 43).

The final sample of 30 features covers all areas of grammar, though not all areas are equally represented (see Table 7.1 for details). Removed features are indicated by a strike-through.

The features were originally devised in the context of a study of West African languages (Bøegh et al. 2016). That study also included languages from other phyla. Here, we focus exclusively on Niger-Congo languages, as these cover the coastal areas from which slaves were imported.
Table 7.1 Overview of the 42 selected WALS features (the twelve removed features are indicated with a strike-through); see the online chapters of WALS (wals.info) for more detailed feature descriptions

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>SV</th>
<th>Areas of Grammar from WALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>9A</td>
<td>The Velar Nasal</td>
<td>54.6%</td>
<td>Phonology</td>
</tr>
<tr>
<td>10A</td>
<td>Vowel Nasalization</td>
<td>57.0%</td>
<td></td>
</tr>
<tr>
<td>18A</td>
<td>Absence of Common Consonants</td>
<td>55.3%</td>
<td></td>
</tr>
<tr>
<td>21A</td>
<td>Exponence of Selected Inflectional Formatives</td>
<td>55.1%</td>
<td>Morphology</td>
</tr>
<tr>
<td>27A</td>
<td>Reduplication</td>
<td>36.2%</td>
<td></td>
</tr>
<tr>
<td>28A</td>
<td>Case Syncretism</td>
<td>67.4%</td>
<td></td>
</tr>
<tr>
<td>29A</td>
<td>Syncretism in Verbal Person/Number-Marking</td>
<td>70.8%</td>
<td></td>
</tr>
<tr>
<td>30A</td>
<td>Number of Genders</td>
<td>72.9%</td>
<td>Nominal Categories</td>
</tr>
<tr>
<td>31A</td>
<td>Sex-based and Non-sex-based Gender Systems</td>
<td>80.8%</td>
<td></td>
</tr>
<tr>
<td>32A</td>
<td>Systems of Gender Assignment</td>
<td>66.9%</td>
<td></td>
</tr>
<tr>
<td>33A</td>
<td>Coding of Nominal Plurality</td>
<td>41.3%</td>
<td></td>
</tr>
<tr>
<td>39A</td>
<td>Incl./Excl. Distinction in Independent Pronouns</td>
<td>64.6%</td>
<td></td>
</tr>
<tr>
<td>40A</td>
<td>Incl./Excl. Distinction in Verbal Inflection</td>
<td>65.0%</td>
<td></td>
</tr>
<tr>
<td>42A</td>
<td>Pronomininal and Adnominal Demonstratives</td>
<td>51.7%</td>
<td></td>
</tr>
<tr>
<td>44A</td>
<td>Gender Distinctions in Independent Personal Pronouns</td>
<td>50.6%</td>
<td></td>
</tr>
<tr>
<td>47A</td>
<td>Intensifiers and Reflexive Pronouns</td>
<td>56.8%</td>
<td></td>
</tr>
<tr>
<td>48A</td>
<td>Person Marking on Adpositions</td>
<td>40.6%</td>
<td></td>
</tr>
<tr>
<td>57A</td>
<td>Position of Pronominal Possessive Affixes</td>
<td>55.0%</td>
<td></td>
</tr>
<tr>
<td>61A</td>
<td>Adjectives without Nouns</td>
<td>54.2%</td>
<td>Nominal Syntax</td>
</tr>
<tr>
<td>63A</td>
<td>Noun Phrase Conjunction</td>
<td>54.3%</td>
<td></td>
</tr>
<tr>
<td>65A</td>
<td>Perfective/Imperfective Aspect</td>
<td>36.0%</td>
<td>Verbal Categories</td>
</tr>
<tr>
<td>66A</td>
<td>The Past Tense</td>
<td>52.4%</td>
<td></td>
</tr>
<tr>
<td>69A</td>
<td>Position of Tense-Aspect Affixes</td>
<td>47.3%</td>
<td></td>
</tr>
<tr>
<td>73A</td>
<td>The Optative</td>
<td>56.7%</td>
<td></td>
</tr>
<tr>
<td>79A</td>
<td>Suppletion According to Tense and Aspect</td>
<td>52.4%</td>
<td></td>
</tr>
<tr>
<td>81A</td>
<td>Order of Subject, Object and Verb</td>
<td>53.3%</td>
<td>Word Order</td>
</tr>
<tr>
<td>82A</td>
<td>Order of Subject and Verb</td>
<td>35.7%</td>
<td></td>
</tr>
<tr>
<td>83A</td>
<td>Order of Object and Verb</td>
<td>66.8%</td>
<td></td>
</tr>
<tr>
<td>84A</td>
<td>Order of Object, Oblique, and Verb</td>
<td>55.1%</td>
<td></td>
</tr>
<tr>
<td>85A</td>
<td>Order of Adposition and Noun Phrase</td>
<td>70.8%</td>
<td></td>
</tr>
<tr>
<td>86A</td>
<td>Order of Genitive and Noun</td>
<td>65.3%</td>
<td></td>
</tr>
<tr>
<td>87A</td>
<td>Order of Adjective and Noun</td>
<td>50.6%</td>
<td></td>
</tr>
<tr>
<td>88A</td>
<td>Order of Demonstrative and Noun</td>
<td>42.4%</td>
<td></td>
</tr>
<tr>
<td>89A</td>
<td>Order of Numeral and Noun</td>
<td>54.9%</td>
<td></td>
</tr>
<tr>
<td>90A</td>
<td>Order of Relative Clause and Noun</td>
<td>54.5%</td>
<td></td>
</tr>
<tr>
<td>94A</td>
<td>Order of Adverbial Subordinator and Clause</td>
<td>44.5%</td>
<td></td>
</tr>
<tr>
<td>99A</td>
<td>Alignment of Case Marking of Pronouns</td>
<td>51.1%</td>
<td>Simple Clauses</td>
</tr>
<tr>
<td>104A</td>
<td>Order of Person Markers on the Verb</td>
<td>37.2%</td>
<td></td>
</tr>
<tr>
<td>118A</td>
<td>Predicative Adjectives</td>
<td>74.3%</td>
<td></td>
</tr>
<tr>
<td>119A</td>
<td>Nominal and Locational Predication</td>
<td>70.9%</td>
<td></td>
</tr>
<tr>
<td>121A</td>
<td>Comparative Constructions</td>
<td>56.0%</td>
<td></td>
</tr>
<tr>
<td>137A</td>
<td>N-M Pronouns</td>
<td>53.9%</td>
<td></td>
</tr>
</tbody>
</table>
7.3 Methods and sampling

In order to evaluate the degree of similarity between the languages included in our samples, we decided to draw phylogenetic networks using the software SplitsTree v.4.13.1 (Huson & Bryant 2006). The Neighbor-Net algorithm (Bryant & Moulton 2004) is particularly well-suited to our endeavor, since it has the advantage of depicting lateral influences, a factor of crucial importance in the emergence and development of contact varieties, and particularly of creoles.

We have aimed at collecting maximally inclusive language samples for the analyses conducted in this study. However, the number of languages spoken on the West African coast is so high that we had to limit our comparison to a subset of these languages. We focus on those languages that belong to the different subgroups of the Niger-Congo family that can safely be assumed to have been involved in the creation of creoles. We have selected at least one language from each of the

![Map indicating locations of selected African languages.](image)

Figure 7.1 Map indicating locations of selected African languages.
14 relevant branches of Niger-Congo (see Figure 7.3) in order to encapsulate the diversity of substrate languages. In order to include as many relevant languages as possible for the analyses, we selected African languages potentially involved in the creation of at least one creole according to suggestions in the literature (cf. Parkvall 2000). With regard to lexifiers, we have included data on all the lexifiers of the creoles in our sample. The approximate locations of the languages can be found on the map in Figure 7.1.

The branches represented are: Adamawa-Ubangi, Atlantic, Cross-River, Dogon, Edo, Gur, Ibo, Ijoid, Kru, Mande, (New) Kwa, Nupoid, Yoruboid, as well as the Bantu groups A, B, H and R (according to the broad geographic divisions provided by Guthrie 1948). The classification of Niger-Congo languages adopted here follows that proposed by Williamson & Blench (2000: 18) and is represented schematically in Figure 7.2.

![Figure 7.2 Major Niger-Congo splits according to Williamson & Blench (2000).](image)

The diversity of creole languages is represented by a total of 45 creoles derived from 12 different lexifiers and spoken on all continents (except Europe, where no creole is known to have emerged). Hence, a wide range of substrate languages was included in the data set, so as to compare the degree of similarity between these languages and the selected creoles.
Data collection was achieved in several ways. Initially, we proceeded by looking up the feature values in WALS (Dryer & Haspelmath 2013). In several cases however, too many data points were missing from WALS. In those cases, the data set was completed by finding the feature values in secondary sources, mostly in published reference grammars or grammar sketches. All additional feature scores were checked against the relevant source(s) by at least two people. The full list of languages used in this study can be found in Table 7.2, where the sources are also given. The mutual relationships between the various languages are indicated with the initial letter of the abbreviation: upper-case L and S stand respectively for ‘lexifier’ and ‘substrate’, and lower-case c refers to creoles. The last letter in the abbreviations for creoles refers to the lexifier: a Arabic, d Dutch, e English, f French, g German, j Japanese, m Malay, p Portuguese, s Spanish, and o for “other”.

Table 7.2  Languages included in this study and their mutual relationships, affiliation, location and sources consulted

<table>
<thead>
<tr>
<th>Language</th>
<th>Abbr.</th>
<th>Affiliation</th>
<th>Spoken in</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akan</td>
<td>Sakn</td>
<td>Niger-Congo, New Kwa (Nyo)</td>
<td>Ghana</td>
<td>Osam (1994); WALS</td>
</tr>
<tr>
<td>Ambon Malay</td>
<td>cAMBm</td>
<td>Creole</td>
<td>Indonesia</td>
<td>Van Minde (1997)</td>
</tr>
<tr>
<td>Angolar</td>
<td>cANGp</td>
<td>Creole</td>
<td>São Tomé</td>
<td>Maurer (1995)</td>
</tr>
<tr>
<td>Annobon</td>
<td>cANNp</td>
<td>Creole</td>
<td>Annobón</td>
<td>Segorbe (2007)</td>
</tr>
<tr>
<td>Arabic, Egyptian</td>
<td>Saeg</td>
<td>Afro-Asiatic, Semitic</td>
<td>Egypt</td>
<td>WALS</td>
</tr>
<tr>
<td>Arabic, Sudanese</td>
<td>Lapd</td>
<td>Afro-Asiatic, Semitic</td>
<td>Sudan</td>
<td>Bergman (2002); WALS</td>
</tr>
<tr>
<td>Bambara</td>
<td>Sbam</td>
<td>Niger-Congo, Mande</td>
<td>Mali</td>
<td>Brauner (1974); WALS</td>
</tr>
<tr>
<td>Batavia Creole</td>
<td>cBATp</td>
<td>Creole</td>
<td>Indonesia</td>
<td>Maurer (2011)</td>
</tr>
<tr>
<td>Batavia Tugu</td>
<td>cTUGm</td>
<td>Creole</td>
<td>Indonesia</td>
<td>Maurer (2011)</td>
</tr>
<tr>
<td>Bena</td>
<td>Sblu</td>
<td>Niger-Congo, Bantu G</td>
<td>Tanzania</td>
<td>Morrison (2011); WALS</td>
</tr>
<tr>
<td>Berbice</td>
<td>cBERd</td>
<td>Creole</td>
<td>British Guiana</td>
<td>Kouwenberg (1994)</td>
</tr>
<tr>
<td>Ternate Chabacano</td>
<td>cCHAs</td>
<td>Creole</td>
<td>Philippines</td>
<td>Sippola (2011)</td>
</tr>
<tr>
<td>Chinuk Wawa</td>
<td>cCHIo</td>
<td>Creole</td>
<td>USA</td>
<td>Zenk (2012)</td>
</tr>
<tr>
<td>Dagbani</td>
<td>Sdgbl</td>
<td>Niger-Congo, Gur</td>
<td>Ghana</td>
<td>Hudu (2014); Hyman &amp; Olawsky (2004); WALS</td>
</tr>
<tr>
<td>Dan</td>
<td>Sdan</td>
<td>Niger-Congo, Mande</td>
<td>Côte d’Ivoire, Guinea, Liberia</td>
<td>Doneux (1968); Welmers (1973)</td>
</tr>
<tr>
<td>Language</td>
<td>Abbr.</td>
<td>Affiliation</td>
<td>Spoken in</td>
<td>Source(s)</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>----------------------</td>
<td>------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Defaka</td>
<td>Sdef</td>
<td>Niger-Congo, Ijoid</td>
<td>Nigeria</td>
<td>Bennett et al. (2012); Jenewari (1983); WALS</td>
</tr>
<tr>
<td>Diola</td>
<td>Sdio</td>
<td>Niger-Congo, Atlantic</td>
<td>Gambia, Senegal</td>
<td>Sapir (1965); WALS</td>
</tr>
<tr>
<td>Doyayo</td>
<td>Sdoy</td>
<td>Niger-Congo, Adamawa</td>
<td>Cameroon</td>
<td>Hewson (2010); WALS</td>
</tr>
<tr>
<td>Dutch</td>
<td>Ldut</td>
<td>Indo-European, Germanic</td>
<td>Netherlands, Belgium</td>
<td>WALS</td>
</tr>
<tr>
<td>Efik</td>
<td>Sefi</td>
<td>Niger-Congo, New Niger-Congo, Benue-Congo (Cross-River)</td>
<td>Nigeria</td>
<td>Una (1900); WALS; Welmers (1973)</td>
</tr>
<tr>
<td>Ega</td>
<td>Sega</td>
<td>Niger-Congo, New Kwa</td>
<td>Côte d'Ivoire</td>
<td>Bole-Richard (1983); Salffner (2004); WALS</td>
</tr>
<tr>
<td>Engenni</td>
<td>Segn</td>
<td>Niger-Congo, New West Benue-Congo (Edoid)</td>
<td>Nigeria</td>
<td>Thomas (1978); WALS</td>
</tr>
<tr>
<td>English</td>
<td>Leng</td>
<td>Indo-European, Germanic</td>
<td>UK, etc.</td>
<td>WALS</td>
</tr>
<tr>
<td>Ewe</td>
<td>Sewe</td>
<td>Niger-Congo, New Kwa (Left Bank)</td>
<td>Ghana, Togo</td>
<td>Duthie (1996); WALS</td>
</tr>
<tr>
<td>Ewondo</td>
<td>Sewo</td>
<td>Niger-Congo, Bantu A</td>
<td>Cameroon</td>
<td>Essono (1994); Redden (1979); WALS</td>
</tr>
<tr>
<td>Fon</td>
<td>Sfon</td>
<td>Niger-Congo, New Kwa (Left Bank)</td>
<td>Benin</td>
<td>Lefebvre &amp; Brousseau (2002); WALS</td>
</tr>
<tr>
<td>French</td>
<td>Lfre</td>
<td>Indo-European, Romance</td>
<td>France, etc.</td>
<td>WALS</td>
</tr>
<tr>
<td>Fulfulde</td>
<td>Sfua</td>
<td>Niger-Congo, Atlantic</td>
<td>Cameroon</td>
<td>Stennes (1967); WALS</td>
</tr>
<tr>
<td>Gà</td>
<td>Sga</td>
<td>Niger-Congo, New Kwa (Ga-Dangme)</td>
<td>Ghana</td>
<td>Kotey (1969); Kropp-Dakubu (2006); WALS</td>
</tr>
<tr>
<td>German</td>
<td>Lger</td>
<td>Indo-European, Germanic</td>
<td>Germany</td>
<td>WALS</td>
</tr>
<tr>
<td>Grebo</td>
<td>Sgrb</td>
<td>Niger-Congo, Kru</td>
<td>Liberia</td>
<td>WALS</td>
</tr>
<tr>
<td>Guinea-Bissau</td>
<td>cGUIp</td>
<td>Creole</td>
<td>Guinea-Bissau</td>
<td>Kihm (1994)</td>
</tr>
<tr>
<td>Gullah</td>
<td>cGULe</td>
<td>Creole</td>
<td>USA</td>
<td>Cunningham (1970)</td>
</tr>
<tr>
<td>Guyane</td>
<td>cGUYf</td>
<td>Creole</td>
<td>French Guiana</td>
<td>Saint Jacques Fauquenoy (1972)</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Language</th>
<th>Abbr.</th>
<th>Affiliation</th>
<th>Spoken in</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haitian</td>
<td>cHAIf</td>
<td>Creole</td>
<td>Haiti</td>
<td>Valdman (1988)</td>
</tr>
<tr>
<td>Hawai‘i</td>
<td>cHawe</td>
<td>Creole</td>
<td>Hawai‘i</td>
<td>Vellupilai (2003)</td>
</tr>
<tr>
<td>Hiri Motu</td>
<td>cHIRo</td>
<td>Creole</td>
<td>Papua New Guinea</td>
<td>Dutton &amp; Voorhoeve (1975)</td>
</tr>
<tr>
<td>Igbo</td>
<td>Sigb</td>
<td>Niger-Congo, New West Benue-Congo (Igboïd)</td>
<td>Nigeria</td>
<td>Emenanjo (1987); WALS</td>
</tr>
<tr>
<td>Ijo-Kolokuma</td>
<td>Sijo</td>
<td>Niger-Congo, Ijoid</td>
<td>Nigeria</td>
<td>WALS; Williamson (1965)</td>
</tr>
<tr>
<td>Jamaican</td>
<td>cJAME</td>
<td>Creole</td>
<td>Jamaica</td>
<td>Patrick (2007)</td>
</tr>
<tr>
<td>Jamsay</td>
<td>Sjms</td>
<td>Niger-Congo, Dogon</td>
<td>Mali</td>
<td>Heath (2005); WALS</td>
</tr>
<tr>
<td>Japanese</td>
<td>Ljap</td>
<td>Japonic</td>
<td>Japan</td>
<td>WALS</td>
</tr>
<tr>
<td>Juba Arabic</td>
<td>cJUBa</td>
<td>Creole</td>
<td>South Sudan</td>
<td>Watson &amp; Ola (1984)</td>
</tr>
<tr>
<td>Kande</td>
<td>Skbs</td>
<td>Niger-Congo, Bantu B</td>
<td>Gabon</td>
<td>Grollemund (2006); WALS</td>
</tr>
<tr>
<td>Karipúna</td>
<td>cKARf</td>
<td>Creole</td>
<td>Brazil</td>
<td>Tobler (1983)</td>
</tr>
<tr>
<td>Kigirîyama</td>
<td>Snyf</td>
<td>Niger-Congo, Bantu E</td>
<td>Kenya</td>
<td>Lax (1996); WALS</td>
</tr>
<tr>
<td>Kikuyu</td>
<td>Skik</td>
<td>Niger-Congo, Bantu E</td>
<td>Kenya</td>
<td>Bergvall (1987); Mugane (1997); WALS</td>
</tr>
<tr>
<td>KiNubi</td>
<td>cKINA</td>
<td>Creole</td>
<td>Uganda, Kenya</td>
<td>Wells (2005)</td>
</tr>
<tr>
<td>Kisi</td>
<td>Skis</td>
<td>Niger-Congo, Atlantic</td>
<td>Guinea, Liberia,</td>
<td>Childs (1988); WALS</td>
</tr>
<tr>
<td>Kongo</td>
<td>Skon</td>
<td>Niger-Congo, Bantu H</td>
<td>Democratic Republic of Congo</td>
<td>Dereau (1955); WALS</td>
</tr>
<tr>
<td>Korlai</td>
<td>cKORp</td>
<td>Creole</td>
<td>India</td>
<td>Clements (1996)</td>
</tr>
<tr>
<td>Koromfe</td>
<td>Skfe</td>
<td>Niger-Congo, Gur</td>
<td>Burkina Faso, Mali</td>
<td>Rennison (1997); WALS</td>
</tr>
<tr>
<td>Krio</td>
<td>cKRIe</td>
<td>Creole</td>
<td>Sierra Leone</td>
<td>Peace Corps (1985)</td>
</tr>
<tr>
<td>Kriol</td>
<td>cROPe</td>
<td>Creole</td>
<td>Australia</td>
<td>Sandefur (1979)</td>
</tr>
<tr>
<td>Louisiane</td>
<td>cLOUF</td>
<td>Creole</td>
<td>USA</td>
<td>Neumann (1985)</td>
</tr>
<tr>
<td>Makhuwa</td>
<td>Smua</td>
<td>Niger-Congo, Bantu P</td>
<td>Mozambique</td>
<td>van der Wal (2009); WALS</td>
</tr>
<tr>
<td>Malay</td>
<td>Lmly</td>
<td>Austronesian, Malayo-Sumbawan</td>
<td>Malaysia</td>
<td>WALS</td>
</tr>
<tr>
<td>Martinican</td>
<td>cMARf</td>
<td>Creole</td>
<td>Martinique</td>
<td>Funk (1953)</td>
</tr>
<tr>
<td>Language</td>
<td>Abbr.</td>
<td>Affiliation</td>
<td>Spoken in</td>
<td>Source(s)</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>----------------------</td>
<td>------------------------------------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Mbum</td>
<td>Smbm</td>
<td>Niger-Congo, Adamawa</td>
<td>Cameroon, Central African Republic, Chad</td>
<td>Santandrea (1964); WALS</td>
</tr>
<tr>
<td>Mooré</td>
<td>Smoo</td>
<td>Niger-Congo, Gur</td>
<td>Burkina Faso</td>
<td>Nikiema (2001); WALS</td>
</tr>
<tr>
<td>Morisyen</td>
<td>cMORf</td>
<td>Creole</td>
<td>Mauritius</td>
<td>Police-Michel et al. (2012)</td>
</tr>
<tr>
<td>Motu</td>
<td>Lmot</td>
<td>Austronesian, Oceanic</td>
<td>Papua New Guinea</td>
<td>Lister-Turner &amp; Clark (1931); WALS</td>
</tr>
<tr>
<td>Mpongwe</td>
<td>Smpo</td>
<td>Niger-Congo, Bantu B</td>
<td>Gabon</td>
<td>Ambouroue (2007); WALS</td>
</tr>
<tr>
<td>Mumuye</td>
<td>Smum</td>
<td>Niger-Congo, Adamawa</td>
<td>Nigeria</td>
<td>Shimizu (1983); WALS; Wolff (1959)</td>
</tr>
<tr>
<td>Nagamese</td>
<td>cNAGO</td>
<td>Creole</td>
<td>India</td>
<td>Boruah (1993)</td>
</tr>
<tr>
<td>Negerhollands</td>
<td>cNEGd</td>
<td>Creole</td>
<td>Virgin Islands</td>
<td>de Jong (1926)</td>
</tr>
<tr>
<td>Ngoni</td>
<td>Sngo</td>
<td>Niger-Congo, Bantu N</td>
<td>Mozambique, Tanzania</td>
<td>Ngonyani (2003); WALS</td>
</tr>
<tr>
<td>Nheengatu</td>
<td>cNHEo</td>
<td>Creole</td>
<td>Brazil</td>
<td>da Cruz (2011)</td>
</tr>
<tr>
<td>Nigerian Pidgin</td>
<td>cNIGe</td>
<td>Creole</td>
<td>Nigeria</td>
<td>Faracas (1996)</td>
</tr>
<tr>
<td>Nupe</td>
<td>Snup</td>
<td>Niger-Congo, New West Benue-Congo (Nupoid)</td>
<td>Nigeria</td>
<td>Kandybowicz (2008); Kawu (2002); WALS</td>
</tr>
<tr>
<td>Obolo</td>
<td>Sobo</td>
<td>Niger-Congo, New Benue-Congo (Cross-River)</td>
<td>Nigeria</td>
<td>Rowland-Oke (2003); WALS</td>
</tr>
<tr>
<td>Palenquero</td>
<td>cPALS</td>
<td>Creole</td>
<td>Colombia</td>
<td>Schwegler (1998)</td>
</tr>
<tr>
<td>Papiá Kristang</td>
<td>cPAKp</td>
<td>Creole</td>
<td>Malaysia</td>
<td>Baxter (1988)</td>
</tr>
<tr>
<td>Papiamentu</td>
<td>cPAPs</td>
<td>Creole</td>
<td>Aruba, Bonaire, Curaçao</td>
<td>Munteanu (1991)</td>
</tr>
<tr>
<td>Pichi</td>
<td>cPICe</td>
<td>Creole</td>
<td>Bioko</td>
<td>Yakpo (2009)</td>
</tr>
<tr>
<td>Plains Indian Sign Language</td>
<td>cPISo</td>
<td>Creole</td>
<td>USA</td>
<td>Text fragments, Bakker (ms.)</td>
</tr>
<tr>
<td>Portuguese</td>
<td>Lpor</td>
<td>Indo-European, Romance</td>
<td>Portugal, etc.</td>
<td>WALS</td>
</tr>
<tr>
<td>Principense</td>
<td>cPRIp</td>
<td>Creole</td>
<td>Principe</td>
<td>Maurer (2009)</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Language</th>
<th>Abbr.</th>
<th>Affiliation</th>
<th>Spoken in</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sango</td>
<td>cSANo</td>
<td>Creole</td>
<td>Central African Republic</td>
<td>Pasch (1997)</td>
</tr>
<tr>
<td>Saramaccan</td>
<td>cSARe</td>
<td>Creole</td>
<td>Suriname</td>
<td>Rountree &amp; Glock (1982); McWhorter &amp; Good (2012)</td>
</tr>
<tr>
<td>Sena</td>
<td>Ssen</td>
<td>Niger-Congo, Bantu N</td>
<td>Mozambique</td>
<td>Funnell (2004); WALS</td>
</tr>
<tr>
<td>Shona</td>
<td>Sshn</td>
<td>Niger-Congo, Bantu S</td>
<td>Zimbabwe</td>
<td>Fortune (1955); WALS</td>
</tr>
<tr>
<td>Spanish</td>
<td>Lspa</td>
<td>Indo-European, Romance</td>
<td>Spain, etc.</td>
<td>WALS</td>
</tr>
<tr>
<td>Sri Lanka Portuguese</td>
<td>cSLPp</td>
<td>Creole</td>
<td>Sri Lanka</td>
<td>Jackson (1990)</td>
</tr>
<tr>
<td>Supyire</td>
<td>Ssup</td>
<td>Niger-Congo, Gur</td>
<td>Mali</td>
<td>Carlson (1994); WALS</td>
</tr>
<tr>
<td>Swahili</td>
<td>Sswa</td>
<td>Niger-Congo, Bantu G</td>
<td>Kenya, Tanzania, Uganda</td>
<td>WALS</td>
</tr>
<tr>
<td>Tayo</td>
<td>cTAYf</td>
<td>Creole</td>
<td>New Caledonia</td>
<td>Ehrhart (1993)</td>
</tr>
<tr>
<td>Temne</td>
<td>Stne</td>
<td>Niger-Congo, Atlantic</td>
<td>Sierra Leone</td>
<td>Kamarah (2007); WALS</td>
</tr>
<tr>
<td>Tok Pisin</td>
<td>cTOKe</td>
<td>Creole</td>
<td>Papua New Guinea</td>
<td>Dutton (1973)</td>
</tr>
<tr>
<td>Tommo So</td>
<td>Sdds</td>
<td>Niger-Congo, Dogon</td>
<td>Burkina Faso, Mali</td>
<td>McPherson (2010); WALS</td>
</tr>
<tr>
<td>Tupinamba</td>
<td>Ltup</td>
<td>Tupi, Tupi-Guarani</td>
<td>Brazil</td>
<td>Barbosa (1956)</td>
</tr>
<tr>
<td>Umbundu</td>
<td>Sumb</td>
<td>Niger-Congo, Bantu R</td>
<td>Angola</td>
<td>Schadeberg (1990); WALS</td>
</tr>
<tr>
<td>Unserdeutsch</td>
<td>cUNSG</td>
<td>Creole</td>
<td>Papua New Guinea</td>
<td>Volker (1983)</td>
</tr>
<tr>
<td>Vai</td>
<td>Svai</td>
<td>Niger-Congo, Mande</td>
<td>Liberia, Sierra Leone</td>
<td>WALS; Welmers (1976)</td>
</tr>
<tr>
<td>Wolof</td>
<td>Swlf</td>
<td>Niger-Congo, Atlantic</td>
<td>Gambia, Senegal</td>
<td>Pichl (1957); WALS</td>
</tr>
<tr>
<td>Yao</td>
<td>Syao</td>
<td>Niger-Congo, Bantu P</td>
<td>Malawi, Mozambique</td>
<td>Sanderson (1922); WALS</td>
</tr>
<tr>
<td>Yilan</td>
<td>cYILj</td>
<td>Creole</td>
<td>Taiwan</td>
<td>own corpus</td>
</tr>
<tr>
<td>Yoruba</td>
<td>Syor</td>
<td>Niger-Congo, New West Benue-Congo (Yoruboid)</td>
<td>Benin, Nigeria</td>
<td>WALS</td>
</tr>
<tr>
<td>Zamboangueño</td>
<td>cZAMS</td>
<td>Creole</td>
<td>Philippines</td>
<td>Forman (1972)</td>
</tr>
<tr>
<td>Zulu</td>
<td>Szul</td>
<td>Niger-Congo, Bantu S</td>
<td>South Africa</td>
<td>Doke (1927); WALS</td>
</tr>
</tbody>
</table>
7.4 African languages and their connections

In this section, we will evaluate the degree of similarity between West African languages and assess the extent to which these languages are typologically homogeneous based on the selection of stable typological features described in 7.3. The graph in Figure 7.3 shows the results in the form of a Neighbor-Net network.  

![Graph showing Neighbor-Net network of 36 Niger-Congo languages.]

**Figure 7.3** Neighbor-Net of 36 Niger-Congo languages.

1. The following color codes were selected so as to differentiate the languages according to genealogy in this and subsequent graphs (note that some families mentioned here only appear in later sections of this chapter): blue = Austronesian, Oceanic; brown = Tupi, Tupi-Guarani; orange = Indo-European, Romance; dark gray = Niger-Congo, Bantu; olive = Indo-European, Germanic; green = Niger-Congo, Ijoid; steel blue = Niger-Congo, Kru; light blue = Afro-Asiatic, Semitic; light cyan = Niger-Congo, Mande; light gray = Japonic; light green = Niger-Congo, Igboi; light orange = Niger-Congo, New Kwa; light pink = Niger-Congo, Adamawa-Ubangi; light yellow = Niger-Congo, Edoid; lime = Niger-Congo, Gur; magenta = Niger-Congo, Atlantic; dark green = Niger-Congo, Dogon; purple = Niger-Congo, Yoruboid; red = Niger-Congo, Nupoid; teal = Austronesian, Malayosumbawan; yellow = Niger-Congo, Cross-River.
The general picture that emerges from Figure 7.3 is that the languages form patterns according to their proposed genealogical relationships, although a number of inconsistencies can be observed. Several groups traditionally recognized in the literature are replicated: we find on the top left of the graph the two Ijoid (Sijo, Sdef), two Dogon (Sjms, Stms) and four Kwa languages (Sga, Sewe, Sakn, Sfon) clustered in separate groupings. Another split is visible in the lower left part of the graph, which includes the Gur languages (Sdgb, Skfe, Smoo), as well as the only representative of the Kru branch (Grebo, Sgrb, a language of Liberia). In the top of the graph, we find the Kwa languages distributed over two separate but contiguous clusters, and with two Adamawa languages among them, Doyayo (Sdoy) and Mbum (Smbm). On the lower right side of the graph, a cluster of Bantu languages interspersed with Atlantic languages can be seen. The only two Cross-River languages, Efik (Sefi) and Obolo (Sobo), appear at the outskirt of the Bantu/Atlantic cluster on the right. The Adamawa languages do not appear in a single cluster but are instead scattered throughout the graph. Also, five Nigerian languages are found in the top right, viz. Efik (Sefi), Igbo (Sigb), Nupe (Snup), Obolo (Sobo) and Yoruba (Syor).

Looking at geography as a potential factor explaining the distribution of the West African languages included in this section, no particular patterns can be discerned. Several languages that are grouped together are spoken in neighboring countries or contiguous areas, such as the three languages spoken in the southeastern part of Nigeria, Efik (Sefi), Igbo (Sigb) and Obolo (Sobo). It is therefore difficult to tell whether the various groupings are conditioned by geography rather than by genealogy. Overall though, the majority of languages do not appear to pattern along areal lines, as suggested for instance by the group that includes the Bantu and Atlantic languages, spoken respectively in the southwestern and western parts of Africa. Instead, the overall positions of the specific clusters in the network match well with Heine’s (1975; 1977) typological classification, based on word order patterns only.

We can also add the lexifiers of the creoles to these African languages. It is sometimes claimed that West African and Western European languages are not that different, and that therefore Atlantic creoles may show similarities. We added not only the European lexifiers, but also the non-European lexifiers: two Arabic varieties, Egyptian (Saeg) and Sudanese (Lapd), Japanese (Ljap), Austronesian Motu (Lmot) and Malay (Lmly), Tupi-Guaraní Tupinamba (Ltup) and Ngbandi (Lngb) from Central Africa, as well as three Germanic languages, Dutch (Ldut), English (Leng) and German (Lger), and three Romance lexifiers, French (Lfre), Portuguese (Lpor) and Spanish (Lspa).
The results can be seen in Figure 7.4. This graph shows that only Japanese and Tupinamba are found in the vicinity of the African languages, whereas the Arabic and European languages form a narrow cluster in the bottom. Furthermore, the European languages and the two varieties of Arabic are less diverse (i.e. the network is denser, suggesting a higher degree of similarity) than the African languages, at least for our choice of stable features. Note also that the addition of lexifiers only affects the clustering of African languages to a limited extent.

Several lower-level genealogical groupings within Niger-Congo known from the literature were replicated more or less faithfully in Figures 7.3 and 7.4 despite inconsistencies. We observed that European languages are more similar to Arabic than to Niger-Congo languages. The idea that the linguistic landscape in West Africa is characterized by a high degree of homogeneity is thus falsified. African languages display high variation and diversity. This is contrary to what is traditionally assumed in creole studies. In the following section, we will look at the influence that the African languages have exerted on a selection of creole languages.
7.5 West African languages and Atlantic creoles

In this section, the focus will be on the links between African varieties and a subset of creole languages, namely the Atlantic and Indian Ocean creoles. Indeed, since we are interested in tracking potential substrate patterns derived from African languages in creoles, we have included here a number of creoles which are known to have emerged in the wake of the European colonial expansions and which involved significant numbers of displaced African populations. We have therefore included a number of Bantu languages from the eastern and southern parts of Africa (thus adding the Bantu groups E, G, N, P and S to our sample, after Guthrie’s 1948 grouping) known to have been present in contact settings that led to the emergence of Atlantic and Indian Ocean creoles. The results are presented in Figure 7.5.

![Figure 7.5 Neighbor-Net of 23 creoles and 46 Niger-Congo languages.](image-url)
The effect of adding creoles and Bantu languages to the sample did alter the results of the internal groupings of the African languages to some extent, compared to the earlier classification of the 46 Niger-Congo languages presented in 7.3 and 7.4. The main clusters observed in the previous sections are replicated and in some cases, even enhanced. For instance, the expanded group of Bantu languages can immediately be identified at the bottom left, although two non-Bantu languages are present in the cluster, Diola-Fogny (Sdío, an Atlantic language spoken in Senegal) and Ega (Sega, a Kwa language of Côte d'Ivoire). Likewise, the links between the languages of the Dogon, Ijoid, Kwa and Mande clusters are now also more tightly connected. Moreover, the Atlantic languages are no longer located within the Bantu cluster (with the exception of Diola-Fogny). The Bantu languages in our sample show a relative homogeneity, especially when viewed from a wider perspective against the diversity of Niger-Congo languages, which occupy a major part of the typological space in Figure 7.5.

Adding the 23 African-influenced creoles to the African languages leads to a network in which the creole languages clearly differ from the African languages. The creoles form a tight cluster, suggesting that creoles are less diverse than Niger-Congo languages, and showing that they are typologically distinct from non-creoles, including the Niger-Congo languages. Almost all the creoles are gathered in a single, dense cluster in the upper right corner.

However, two creoles are located outside the creole cluster. Sango (cSANo, spoken in the Central African Republic) is found at the periphery of the creole cluster on the right of the graph. On the left, Jamaican (cJAME) appears in a group with the two Ijoid languages Ijo (Sijo) and Defaka (Sdef). A single non-creole language, Wolof (Swlf, an Atlantic language of Senegal), is found at the boundary of the creole cluster, in a group with Portuguese-based Guinea-Bissau Kriyol (cGUIp) and Guyanese French Creole (cGUYf). Interestingly, Wolof happens to have been involved in the creation of several creoles, including Kriyol (e.g. Kihm 2011). A case could be made here for substrate influence.

Within the creole cluster, there does not appear to be a single factor explaining the distribution of the languages, as these do not group uniquely according to either lexifier, geography or substrate influence. This suggests on the one hand that the influence of (West) African substrates on the creoles has been relatively minimal, at least with regard to the stable properties selected here, despite claims to the contrary (e.g. Lefebvre 1998; 2011). For instance, Haitian Creole (cHAIf) and Fon (Sfon) do not appear particularly close to one another, and neither do Berbice creole (cBERd, on the top right) and its major contributor Ijo (Sijo, on the left), despite the close ties between those (Smith et al. 1987; Kouwenberg 2012). Ijo is in fact found close to Jamaican (cJAME), but Kikongo and Kwa languages are in fact more likely sources of substrate influence for this creole (Farquharson & Baker 2012; Patrick 2007: 127).
On the other hand, the claim that creoles form a typologically coherent group of natural languages, is corroborated here. This position was put forward by McWhorter (1998) and Bakker et al. (2011), namely that creoles represent a synchronically distinguishable sub-group among the world’s languages.

7.6 West African languages and Asian creoles

We will now turn our attention to a broader sample of creoles encompassing varieties of Asian creoles, which can safely be assumed not to have received a particularly strong input from African languages, or none at all. We added them to the data set and built another phylogenetic network. The results are presented in the graph in Figure 7.6.

When adding the Asian creoles to the data set, the distribution of the African languages is quite similar to that reported on in the previous sections. The group of Bantu languages in the bottom is more pronounced than in Figure 7.5 and all the groups of Niger-Congo languages identified above are preserved. For instance, the Dogon, Ijoid and Gur languages are present here as well. Some of the inconsistencies observed in the previous sections can also be seen in Figure 7.6, such as the intrusion of Diola-Fogny (Sdio, Atlantic) and Ega (Sega, Kwa) within the Bantu group.

Similarly, Wolof (Swlf) appears further inside the core creole cluster and the African-lexifier creole Sango (cSANo, to the right) is now clearly located outside the creole cluster. The languages of the Adamawa group are still spread throughout the network. On the left, another creole (Hiri-Motu, cHIRo, a Motu-based creole of Papua New Guinea) can be found on the other side of the separation between creoles and non-creoles. Otherwise, all the creoles are located in a single cluster (note that Jamaican now appears within this group). Strikingly, the most ‘African’ of all creoles, the English-based maroon creole Saramaccan (cSARe, top right), is almost as far removed from the African languages as can be. It is also interesting to note that the creoles that appear closest to the African languages are not necessarily creoles which have been influenced by African languages. For instance, on the left, we find Hiri Motu (cHIRo, Motu-based, spoken in Papua New Guinea) grouped with Dan (Sdan, a Mande language of Côte d’Ivoire), and such an influence can hardly have occurred. This shows that the influence of the substrates can barely be detected when stable features are selected.

In the next section, the lexifiers of the creoles studied here will be added to our sample so as to assess the extent of their influence in the geneses of the various creoles.
Figure 7.6 Neighbor-Net of 45 creoles and 46 Niger-Congo languages.
7.7 Creoles and their lexifiers

In this section, we have included all the lexifiers of the creole languages selected in the previous sample. The sample consisting of creoles and their lexifiers was used to evaluate the degree of structural influence that the lexifiers have exerted upon these creoles. Several researchers maintain that creoles are mere continuations of the lexifiers and that they therefore are structurally and genealogically related to their superstrates (Chaudenson 1992; Mufwene 2008). If this is correct, this would mean that French creoles should cluster with French, and the other creoles should also cluster with their respective lexifier. The results are presented in Figure 7.7.

![Figure 7.7 Neighbor-Net of 45 creoles and 12 lexifiers.](image)

We can immediately observe a clear cluster in the top right corner that includes all the Indo-European lexifiers, with the three Romance varieties French (Lfre), Portuguese (Lpor) and Spanish (Lspa) on the left, and the three Germanic languages Dutch (Ldut), English (Leng) and German (Lger) on the right. Three acrolectal English-based creole varieties, Gullah (cGULe), Hawaiian (cHAWe) and Vincentian (cVINe) are located at the periphery of this cluster. On the other side of this cluster, we find the only known German-lexified creole, Unserdeutsch.
This creole is known to share only few typical creole features with Atlantic creoles (Mühlhäusler 1997: 200–202). Note also that it appears closer to the Romance lexifiers than to the Germanic languages. Thus, the results suggest that, structurally, these creoles are relatively similar to the Indo-European lexifiers when the stable features selected here are taken into consideration. However, other acrolectal creoles derived from these European languages can be found elsewhere in the graph (such as French-based Reunion creole (cREUf), at the opposite end of the graph from French in the lower left).

Several non-European lexifiers included in the data set appear in a cluster together with their lexified creole. Thus, Malay (Lmly, Austronesian, Malayo-Sumbawan) clusters with Tugu Malay Creole (cTUGm), Hiri Motu (cHIRo) shows up with Motu (Lmot, Austronesian, Oceanic), and Ngbandi (Lngb, Niger-Congo, Adamawa) is located in a cluster with Sango (cSANo). As for the remaining three lexifiers present in the sample, Sudanese Arabic (Lapd, Afro-Asiatic, Semitic), Japanese (Ljap, Japonic), and Tupi (Ltup, Tupi-Guaraní), these do not group up with the creoles they lexified (KiNubi (cKINa), Yilan Creole (cYILj), and Nheengatu (cNHEo) respectively). Also, the Spanish-based creoles Papiamentu, Palenquero and Zamboangueño (cPAPs, cPALS, and cZAMS) appear at the opposite end of the graph from Spanish, and so do all the French-based creoles, which also appear far from their lexifier. Note also that the creoles do not cluster consistently according to their lexifiers, which should have been the case if they were continuations of the superstrates.

The presence of several lexifiers in the direct vicinity of the creoles they contributed to clearly suggests that the structural influence of lexifiers has been non-negligible in the context of creolization, apart from the obvious lexical contribution. Nevertheless, about half of the lexifiers do not show any close typological ties to their respective creoles. Note also that Saramaccan (cSARe), which is often argued to be the most typical creole, is located furthest away from the cluster of European lexifiers. It is apparently as different from its European lexifiers as from the African languages (as was also observed in Figure 7.7).

These results, based on a selection of stable features, provide support for a position that at least some creoles are structural continuations of the superstrate (Hiri Motu, Sango, Tugu Malay), as advocated by e.g. Chaudenson (1992), or that some European lexifiers also underwent changes akin to creolization (see e.g. Bailey & Maroldt 1977; Goyette 2000; see also Collins 1980 for Malay). Still, the conservativeness of stable features in creolization is not the case for all creoles, for instance the Spanish creoles, Yilan, Nheengatu and KiNubi, which are quite different. Some European-based creoles can be structurally close to their lexifiers, especially acrolectal creoles, but others are also very different (e.g. Saramaccan). Thus, the fact that the creoles do not cluster systematically according to their lexifiers indicates that creoles cannot be considered continuations of the superstrate in a strict sense.
7.8 Transmission of stable features in creoles and non-creoles

The main focus of this section will be on the features that were found to be among the most stable in the languages of the world according to the investigation undertaken by Wichmann and Holman (2009). We made a reasoned selection of the features which show the highest degree of stability, factoring in interdependency (see 7.3). Among the languages considered so far, we observed that language groups traditionally identified in the literature were replicated, but only to a certain extent. We expect that including all the languages studied in the previous sections will strengthen the results, as genealogical groupings will be reinforced by adding new members to the sample.

We noted above that creoles tend to form relatively homogeneous clusters, thus emphasizing the shared typological properties of creoles among the languages of the world, even when compared with supposedly typologically similar West African languages and the lexifiers involved in the creation of the creoles in our sample. Selecting these features should reinforce these clusters on the assumption that the inherited material will more likely be retained despite creolization and result in genealogical groupings. In this section, we will first include all the languages selected in the previous sections in a single graph, and then, we will decrease the number of stable features for the same set of languages in order to get an impression of a “best fit”.

The first graph, presented in Figure 7.8, was built by using the same 30 stable features as selected for the previous sections (see Table 7.1 for details) with all the creoles, their lexifiers and substrates.

As observed previously in Figure 7.7, the Indo-European lexifiers appear in a single cluster (in the top left), again with a clear distinction between Romance and Germanic languages. These are surrounded by acrolectal creoles and by the two Semitic languages present in the sample, Egyptian and Sudanese Arabic (Saeg and Lapd respectively). The presence of all the Indo-European languages in the same cluster as the two varieties of Arabic may point to a deep genealogical relationship between Semitic and Indo-European, as suggested by Nostraticists (e.g. Dolgopolski 1998). Or it may be indicative of a bias in the selection of WALS features towards European languages, which all match the Standard Average European profile (originally put forward by Whorf (1941) and further developed in Haspelmath 2001). Alternatively, this might reflect the fact that long-term contact between these languages has taken place and hence resulted in languages with a similar morphosyntactic make-up.

The patterns of African groupings observed in the previous sections are also replicated, both with regard to known genealogical groupings (e.g. the Bantu languages in the upper right), and with regard to individual creoles placed in the same group as their respective lexifier (e.g. Motu (Lmot) and Hiri Motu (cHIRo)).
in the lower part of the graph) or altogether outside any meaningful cluster (Sango, cSANo, which is found on the right, with Yoruba (Syor) and the two Cross-River languages Efik (Sefi) and Obolo (Sobo), also spoken in Nigeria). A further noteworthy observation is the fact that all the non-Indo-European lexifiers (with the exception of Sudanese Arabic mentioned above) are located outside the creole cluster. Moreover, they either appear in-between clusters or at the periphery of the groups they belong to: Malay (Lmly in the top part) and Motu (Lmot in the bottom) are found in the transition zones between the African languages and the creoles. Only Japanese (Ljap, on the bottom right) and Ngbandi (Lngb, on the right) appear clearly inside African language clusters.

The creoles rarely cluster with their respective lexifier (or according to geography for that matter). For instance, Juba Arabic (cJUBa, in the lower left) does not appear particularly close to its lexifier, Sudanese Arabic (Lapd), nor do the creoles derived from the Indo-European languages appear in the creole cluster. Thus,
these results can again be considered as evidence against the claim that creoles are simply continuations of their lexifier.

We will now use fewer features for the analysis so as to see whether this procedure increases the links between the genealogical units in the sample. The 20 features retained were the ones with the highest stability values (see Table 7.1). The results are shown in Figure 7.9.

As can be seen from the graph in Figure 7.9, reducing the number of features resulted in a clear cluster comprising only the Indo-European and Semitic languages in the middle of the creole cluster (on the left, divided into Germanic, Romance and Semitic). The groups in the creole cluster are now more densely linked and contain fewer splits and shorter branches. Moreover, the gap between creoles and
non-creoles is now more pronounced. Remarkably, only four creoles are located outside the core creole group (Hiri Motu, cHIRo, KiNubi, cKINa, Sango, cSANo and Ambon Malay, cAMm). Again, the non-Indo-European lexifiers are scattered among the non-creole African languages. On the lower left, the Atlantic languages appear at the outskirts of the cluster of Bantu languages, this time with the addition of Mumuye (Smum, Adamawa), while Diola-Fogny (Sdio, Mande) and Ega (Sega, Kwa) are now located each on one side at the periphery of the Bantu cluster. Also, the Kwa and Gur groups (at the top) are now less salient than in previous graphs.

Thus, selecting fewer features resulted in less clear-cut genealogical groupings of the African languages when compared to the results presented in previous sections and other established classifications. Still, no substrate influence can be detected in specific pairings of creoles with other creoles, or creoles with their substrates.

Geography does not correlate with the observed discrepancies. For instance, the various sub-groups within the creole cluster are not conditioned by their geographical location, as no single group contains only creoles spoken in the same region. The Atlantic/Bantu cluster does not reflect geography either.

In order to find out whether the reduced number of features accounts for the discrepancies observed in Figure 7.9 and whether this obscured or enhanced the phylogenetic signal, we built a graph with the 10 most stable features identified by Wichmann and Holman (2009). The results are presented in Figure 7.10.

The graph in Figure 7.10 shows that reducing the number of features to 10, clearly disrupted many of the genealogical groupings observed in the previous sections. The creoles are now scattered around the graph in a seemingly random fashion and the same applies to the majority of the languages in our sample. However, some clusters of related languages remain, or were even reinforced, for instance, we can find the cluster of Bantu languages on the left with Atlantic languages at the periphery. Neither genealogy, nor geography (nor lexifier or substrate in the case of creoles), seems to account for the distribution of the various languages across the graph. The cluster grouping the Indo-European and Semitic languages no longer appears in the creole cluster, although they are still found in a separate cluster. The links between those are either disrupted or more distant. Likewise, in the lower right, we can find three Asian creoles (Hiri Motu, cHIRo, Nagamese, cNAGo and Yilan, cYILj) in a cluster with Dogon and Mande languages.

The only other relatively coherent groups that can be identified are the Gur, Kwa and Ijoid languages on the right. Even the languages from the various branches of Niger-Congo with only two representatives, such as Ijoid on the right, are more distanced. Hence, the operation of reducing the number of stable features did not reinforce the genealogical patterns. On the contrary, a lower number of features led to less coherent genealogical patterns. Alternatively, the features may not all reflect stability. In fact, in a study by Dediu & Cysouw (2013), different
methods for measuring the stability of linguistic features were compared. Their results were sometimes contradictory with regard to the stability values of individual features (although they also show an overall agreement as to which features belong to either the more or the less stable category). Another possible reason could be that the degree of contact in the period of creole formation was so extreme that it entailed the complete disruption of grammatical categories and therefore of even the most robust features of language. At any rate, the set of 30 features we retained for this study seems to provide the most satisfactory results as far as genealogical groupings are concerned.

7.9 Conclusions

The major aim of this chapter was to investigate whether the African languages involved in the genesis of creole languages were typologically homogeneous, and hence represent an explanatory factor for observed similarities between creoles. Also, we examined whether West African languages had a structural impact on the Atlantic creoles. Using phylogenetic networks, we found that the African
languages were quite diverse typologically. Most genealogical groupings of African languages identified in the literature appeared as units in our results, but with a number of exceptions. These exceptions could be due to errors in the authors’ scoring of features, contact influences or as a result of heavy contact, which affected even the most stable features. The results might also provide a better classification than one based solely on lexical features or on intuition.

When adding creole languages to the African languages in our analysis, we found that substrate languages did not align with the creoles they share an ancestry with. In fact, the only case where substrate influence could reasonably be invoked in our results is in the case of Guinea-Bissau Kriyol, which repeatedly appears directly next to one of its main substrate languages, Wolof. However, Wolof is still currently spoken alongside Kriyol, which means that the influence could have occurred at a later stage in the history of the creole. Alternatively, Wolof is a creole, but it has not been identified as such.

This strongly suggests that substrate influence was not a particularly decisive factor in creole genesis. If so, one would have expected to see Berbice (a Dutch-based creole of Guyana) to appear in the vicinity of Ijo (a language of Nigeria), due to the known tight links between the two languages (Smith et al. 1987), and not right next to Jamaican – which, again may have been influenced by Ijo, but to a lesser degree, as compared to e.g. Kikongo (Farquharson & Baker 2012) or Kwa languages (Patrick 2007: 127). Such systematic patterns linking creoles to their dominant substrates (or lexifiers for that matter) were not systematically observed in our results. This strongly suggests that substrate influence in creole genesis has not been as intense as suggested by some authors, at least in the case of the stable features selected here. Likewise, the results did not differ significantly when creoles with a non-African component were included in the analyses: creoles still appeared in clusters separated from non-creoles and African-influenced creoles did not appear closer to the African languages than Asian creoles did.

Including lexifiers resulted in patterns suggestive of a strong connection between European languages and creoles, thus potentially lending support to proponents of a superstratist account of creole genesis. However, the presence of Arabic in this grouping cannot be explained in this way. The results may also point to creolization as a factor playing a role in the genesis of Western European languages (as suggested by Mufwene 2008) and Arabic (see Versteegh 1984).

When we included all the languages involved in the creation of creoles selected for our sample, we found that, on the one hand, several African groups traditionally identified in the literature were replicated and, on the other hand, that the creoles consistently formed relatively coherent clusters. Finally, the last step in this study, i.e. reducing the number of stable features, did not result in a clearer phylogenetic signal. Several authors have come up with different sets of stable features
(summed up in Dediu & Cysouw 2013). It remains to be tested whether the selected features are indeed the most appropriate for testing genealogical continuity.

The implications of this study for creolistics in general is that the influence of the various languages involved in the process of creolization is, when selecting stable features, more prominently reflected in the lexifiers than in the African substrate languages. The European and Arabic lexifiers appeared to be morphosyntactically closer to the creoles, but not especially close to those in which creation they were involved. Similarly, many clusters in our results could not be directly connected with either genealogy or geography, or substrate influence in the case of creoles. For this selection of stable features, the substrates were more diverse than the lexifiers. Hence, the common assumption in creolistics that the substrate languages were typologically rather uniform was not found to be correct in the present investigation. Finally, our analyses provided further support to the idea that, from a structural point of view, creoles form a synchronically distinguishable sub-group among the world’s languages, including those from Western Africa which were directly involved in their creation.

Acknowledgements

We are grateful to our assistants and colleagues who helped us with scoring missing feature values and checking them, including Andrea Ariza, Benjamin Riise, Karen Kiil Brøcker, Sergio González de la Higuera Rojo, Yonatan Ungermann Goldshtein, and especially Robert Borges.

Note

The datasets for this chapter can be found here: https://phylogenetic-creole-studies.blogspot.com

References


Maslova, E. S. 2004. Динамика типологических распределений и стабильность языковых типов (Dynamics of typological distributions and stability of language types). Вопросы языкознания 5: 3–16.


The main aim of this chapter is to examine the internal relationships of several French-based creoles and to provide a classification of these languages, a problem that has been left unsolved (or only partially solved) by creolists so far. We show that the issue can be tackled with a phylogenetic approach based on a selection of lexical and structural features, resulting in suggestive patterns on the connections between the various creoles in the sample. An advantage in using phylogenetic methods is that it also allows us to assess the relative degree of divergence of a creole from its lexifier. Therefore, French was added to the sample of French-based creoles. The results obtained using this method are largely in accordance with previous attempts from various scholars and show that the application of phylogenetic tools to the study of contact languages can complement traditional methods.

8.1 Introduction

The study of French-based creole languages worldwide has brought about questions and problems similar to those encountered by scholars investigating creoles based on English and other colonial languages, such as identifying the number of independent geneses that have taken place or explaining the observed similarities shared by these languages. Can this relative homogeneity be explained because of the existence of a Portuguese-based proto-pidgin ancestor spoken along the West African coast? This explanation has been suggested so as to account for the striking similarities found in, especially, Atlantic creoles (e.g. Whinnom 1965). When and where did the various varieties emerge? How similar are the French-based creoles of the Atlantic region compared to those spoken in the Indian Ocean? These questions have been central to the field of creole studies since its inception as a sub-discipline of linguistics, but few have been answered conclusively. Applying tools borrowed from evolutionary biology, we will use this survey to shed new
light on the emergence and development of French-based creoles and provide an investigation of their interrelationships.

We will compare a selection of lexical and typological properties found in a sample of 12 French-based creoles distributed across the globe. After providing some general background information about the French-based creoles (FBCs) selected for this study, we will present the methodology and discuss the selection procedure followed to gather the sample of French-based creoles on the basis of which the analyses were conducted. In turn, this will enable us to provide a classification of the French-based varieties. A convenient feature of the methodology employed is that it allows to measure the relative degree of distance or divergence between the studied entities, and in our case, will permit us to assess the degree of divergence of the French-based creoles from their lexifier, French.

8.2 Previous work on the classification of French-based creoles

Like a majority of known creole languages, the French-based creoles emerged as a consequence of the European colonialist expansions of the 16th to the 19th centuries, where large numbers of people were brought together, often forcefully, to act as labor force in the colonies. The French joined the colonial enterprise somewhat later than their Portuguese and Spanish counterparts. However, they managed to develop plantation economies in a variety of territories worldwide, especially in the Caribbean and the Indian Ocean. The geographical location of the languages included in this study can be found on Maps 8.1 and 8.2.

Map 8.1 The French-based creoles of the Circum-Caribbean region included in this study.
The issue of the classification of French-lexified varieties has received relatively little attention in the literature, especially when compared to the English-based creoles (see for instance Hancock 1969, 1987; McWhorter 1995; Baker 1999; Baker & Huber 2001; Daval-Markussen & Bakker 2011; Holbrook 2012). One reason for this is perhaps the obvious geographical gap between the two main major areas where the languages are spoken: two oceans and a continent separate the French-based creoles of the Americas and the Caribbean from those of the Indian Ocean and the Pacific. This geographical gap also corresponds to clearly delineated linguistic areas for French creoles. Besides, in contrast to their English-based counterparts, the question of whether this group of creoles descends from a single proto-pidgin ancestor is less controversial: the three main areas where French-lexicon creoles are spoken (i.e. the Caribbean, the Indian Ocean and the Pacific) are so distant geographically that they are less likely to have been transplanted from one continent to another, in contrast to what was often the case within these territories (e.g. inter-island migration in the Caribbean). Thus, the situation is far less complicated with regard to the relationships between these languages, which is reflected in the proposed classifications, as we will show below. Finally, the colonial ideologies under the French and British were quite different. This may have resulted in the belief that is commonly encountered in studies of French-based creoles that they are not as divergent from their lexifier as English-based creoles are.

The most seminal work on French-based creoles was undertaken by Goodman (1964), which represents one of the first attempts to propose a classification based on a systematic comparison of actual linguistic features found in these languages, rather than basing it on geographical criteria, resulting in a classification of
convenience (which is what for instance Vintila-Radulescu 1976 does). Goodman investigated the distribution of 40 grammatical and lexical items in several French-based creoles (all of which are included in our sample). He proposed to divide those into five groups: (i) Louisiana French Creole, (ii) Haitian (consisting of a Northern, a Southern and a Western dialect), (iii) a group of Antillean creoles (itself further subdivided into Guadeloupe and Windward varieties, comprising all the other Caribbean creoles), (iv) Guyanais (spoken in French Guiana), and finally (v) a group consisting of the Indian Ocean creoles (Reunion Creole, Morisyen and Seselwa). Goodman opts for a monogenetic solution to explain the similarities between the French-based creoles across these geographical zones (1964: 130–131). This graph is reproduced in Figure 8.1. Note the division between the *ka* (preverbal marker) and *bay* (‘to give’) in the American varieties, reflecting the shared features between these languages.

![Figure 8.1](image-url)  
*Figure 8.1* The classification of French-based creoles proposed by Goodman (1964: 16).

Valdman (1978) represents the most comprehensive survey of French-based restructured vernaculars, but no classification as such is proposed in his book. Instead, the author chooses to rely on the five groups proposed by Goodman (1964) and to tabulate attestations relating the presence or absence of certain structural features in the various languages. Although his work is based on Goodman’s, Valdman argues that on linguistic grounds only, varieties such as Louisiana and Guyanais should belong to the Indian Ocean group because of several shared features, such as the position of determiners (Valdman 1978: 370–371). Some of the languages considered by Valdman are included in our survey. In contrast to Goodman, Valdman dismisses altogether a monogenetic explanation for the origin of these creoles (1978: 372–373).
The investigation undertaken by Baker (1987) does not directly address the question of classification, but rather represents an effort towards disproving the monogenetic hypothesis advocated by, among others, Goodman (1964). The author analysed 100-word Swadesh lists in 12 French-lexicon creoles (Louisiana, Haiti, Guadeloupe, Dominique, Martinique, St-Lucia, Trinidad, Guyanais, Reunion Creole, Morisyen, Rodriguais and Seselwa) in order to evaluate the possibility of independent geneses for each of the 12 varieties selected (all the languages selected by Baker are dealt with in this chapter, with the exception of Rodrigues, an offshoot of Morisyen). In this fashion, Baker posited that at least five independent geneses have taken place and proposed a classification where each group is represented by a separate branch. Hence, he divides the French-based creoles so that Louisiana and Guyanais both form separate branches and appear at each end of a group comprising all the Antillean varieties (Dominican, Guadeloupe, Haitian, Martinique, St. Lucia and Trinidad). The Indian Ocean creoles appear in two separate groups. The first group consists of the Isle de France creoles (Morisyen, Rodriguais and Seselwa), which have emerged as independent varieties genetically unrelated to Reunion Creole (Baker & Corne 1982), the only member of the second subgroup. Dotted lines connecting Reunion Creole to Morisyen and Seselwa illustrate the mutual influences between these varieties (see Figure 8.2 for a reproduction of Baker’s graph).

![Figure 8.2](image)

Figure 8.2 The classification of French-based creoles proposed by Baker (1987:75).

The approach chosen by Parkvall (1995) differs from the others in several respects. The aim of his work was first and foremost to investigate the issue raised by Baker (1987, discussed above), and Parkvall also tried to assess whether the Atlantic French-based creoles came about as the result of mono- or polygenesis. The author
bases his analysis on the distribution of 47 lexical and morphosyntactic features in a variety of French-based creoles. Parkvall convincingly correlates the origin of substratal and superstratal features in the various creoles to the assumed age of the creoles investigated. This leads him to conclude that for older French-lexified creoles, a majority of slaves came from Lower Guinea and that Normandy was the most likely place of origin of the French settlers. Besides, Parkvall posits that eight separate geneeses occurred for the French-based creoles: according to him, the creoles of St. Kitts, Tortuga (or Northern Haiti), Martinique, Guadeloupe, Haiti, Guyane, Louisiana and Mauritius emerged independently from one another. He does not propose a classification in the traditional sense and chooses instead to assign each language a place on a continuum reflecting the degree of divergence between the various creoles in relation to their lexifier, French.

Finally, we should briefly mention that the only known French-based creole spoken in the Pacific region was only discovered a few decades ago, and therefore it was not included in the classifications of the above-mentioned authors. Recent work on the sociohistory and demographics underlying the emergence of this creole by Speedy (2007a, 2007b; Ehrhart 2012) has shown that many colonists from Réunion Island arrived to New Caledonia in the 19th-century and may have influenced Tayo linguistically (Speedy 2013, see also Section 8.4 below).

8.3 Methods and sample

The phylogenetic analyses presented in this chapter were conducted using the software SplitsTree v.4.13.1 (Huson & Bryant 2006), following the Neighbor-Net algorithm (Bryant & Moulton 2004). This algorithm builds on the Neighbor-joining method of clustering (Saitou & Nei 1987), based on a distance matrix as input and this allows the user to compute phylogenetic trees or networks. Since we are not only interested in studying genealogical relationships, but also in tracking down possible reticulation events (i.e. language contact) in the historical development of the included French-based varieties, constructing phylogenetic networks to graphically represent the interrelationships between the languages of our sample was the most obvious choice.

The data were gathered by having native speakers of 12 French-based creoles translate the 50 sentences found in Hancock’s questionnaire (1987, see also Daval-Markussen et al. 2011) into their respective language. Most of the questionnaires were returned by several individuals, but in some cases, the translations were provided by a single consultant. The translations were realized in the standard written language and are provided in the online Appendix for the reader to consult (at <https://phylogenetic-creole-studies.blogspot.com>). Therefore, this sample is not meant to be representative
of the selected languages, but rather represents a convenience sample devised to enable exploratory research in the interrelationships of French-based creoles.

The French-based varieties included in our study are presented in Table 8.1.

Table 8.1 The 12 French-based creoles selected for this study with their abbreviations and geographical location.

<table>
<thead>
<tr>
<th>Language</th>
<th>Abbreviation</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominican</td>
<td>Dom</td>
<td>Caribbean</td>
</tr>
<tr>
<td>Guadeloupe</td>
<td>Gwa</td>
<td>Caribbean</td>
</tr>
<tr>
<td>Guyanais</td>
<td>Guy</td>
<td>South America</td>
</tr>
<tr>
<td>Haitian</td>
<td>Hai</td>
<td>Caribbean</td>
</tr>
<tr>
<td>Martinique</td>
<td>Mar</td>
<td>Caribbean</td>
</tr>
<tr>
<td>Morisyen</td>
<td>Mau</td>
<td>Indian Ocean</td>
</tr>
<tr>
<td>New Orleans Creole</td>
<td>Lou</td>
<td>North America</td>
</tr>
<tr>
<td>Réunion Creole</td>
<td>Reu</td>
<td>Indian Ocean</td>
</tr>
<tr>
<td>Saint Lucia Creole</td>
<td>StL</td>
<td>Caribbean</td>
</tr>
<tr>
<td>Seselwa</td>
<td>Sey</td>
<td>Indian Ocean</td>
</tr>
<tr>
<td>Tayo</td>
<td>Tay</td>
<td>Pacific</td>
</tr>
<tr>
<td>Trinidad French Creole</td>
<td>Tri</td>
<td>Caribbean</td>
</tr>
</tbody>
</table>

A total of 66 lexical and grammatical features were encoded in a binary fashion for the 12 French-lexicon varieties by identifying points of divergence in the translated sentences. For instance, Feature 1 describes the presence vs. absence of definite articles, while Feature 57 relates to the realization of the completive aspect either with a particle meaning ‘to finish’ or ‘already’. These data were then used to build a character matrix, which was formatted as a NEXUS file for input in SplitsTree (see Appendix). In order to comply with Hancock’s original approach, when several choices were available for a language, the most basilectal translation (as judged by the consultant, a specialist of the language or the author himself for languages with which he is familiar) was selected for the encoding process.

8.4 Classifying French-based creoles

In this section, we report on two phylogenetic analyses conducted with data describing the distribution of 66 lexical and grammatical features in 12 French-based creoles and in their lexifier, French. The results are presented in the graphs in Figures 8.3 and 8.4.

In the first graph, all 12 creole languages were included for the analysis. Two major clusters are immediately evident in the graph in Figure 8.3. On the left, there
is one group including the Caribbean creoles of Trinidad, Guadeloupe, Martinique, Dominica, Saint Lucia and, a little further apart, Haitian. There is another group consisting of the Indian Ocean varieties (Mauritius, Réunion and Seychelles) on the right. These clusters happen to correspond to groups of languages that have been argued to be related (e.g. Goodman 1964). The Lesser Antillean varieties Tri, Gwa, Mar, Dom, and StL form a subgroup of a larger cluster including Hai and Guy at its periphery. The Indian Ocean creoles Reu, Mau, and Sey form a neat cluster in the opposite end of the graph. Interestingly, Tay appears in the direct vicinity of Reu, and a direct link between the two languages has been suggested on the background of historical documentation (Speedy 2007a, 2007b, 2013). Surprisingly though, Lou is located close to Tay in the upper part of the graph, but nothing in their respective histories suggests any form of contact between the two territories. According to Baker (1987), the creoles of Haiti, Guyana and Louisiana came about as the results of separate geneses, which may be reflected by the intermediate positions of these languages in the graph. Besides, the graph presented by Baker (1987: 75, see Figure 8.2) is practically identical to the one presented in Figure 8.3, where the creoles of the Caribbean (with Dom and StL branching off from the related Gwa and Mar with Tri as an off-shoot) and the Indian Ocean (Reu at the periphery of the cluster and Mau and Sey closer to one another) appear as two
distinct groups, while the mainland varieties of Louisiana and Guyana emerged independently (and therefore appear outside either of the clusters). The limited number of languages included in this study can be a factor explaining why Lou and Tay appear relatively close together in the network in Figure 8.3. Another relevant factor might be the relatively acrolectal nature of the creoles in this part of the graph: Reu is often treated as a semi-creole in the literature (e.g. Burns 1954) and the two other creoles, Lou and Tay, both exhibit a lower degree of restructuring as compared to e.g. Haitian. The relatively younger age of Tayo (which is believed to have emerged in the late 19th century on a mission school, e.g. Ehrhart 2012) can perhaps be another explanatory factor.

A quick look into the population histories of the various territories represented offers possible explanations for the connections that can be observed between the selected creole languages in Figure 8.3.

For instance, in the Caribbean, Trinidad and Haiti were the first territories to be colonized by Europeans (the Spanish in this case) in the mid-1500s. However, these two colonies were not actively settled and exploited by the French until a century later in the 1670s. On the other islands of the Lesser Antilles, where French-based creoles are spoken today (Dominica, Guadeloupe, Martinique and St. Lucia), a slave trade and plantation economy did not develop before the late 1650’s and even later for Dominica, which was permanently settled in 1763 (Holm 1989: 371–372). In the case of Trinidad, the island was never actually a French possession, and the French-based creole that is spoken there today was transplanted by settlers from the neighboring islands Dominica, Grenada, Guadeloupe and Martinique in the mid-18th century. Besides, while there was regular contact between the Lesser Antilles islands throughout their history, Haiti and Trinidad have been more isolated since the 1700s (see e.g. Burns 1954).

In the Indian Ocean, the colonization endeavor did not start before the late 1600s and the creole languages of the area have been shown to have emerged following two separate geneses, one on Reunion Island, the other on Mauritius (Baker & Corne 1982), although mutual influences have occurred between the two languages in the course of their histories (see Chaudenson 2010 for a different view). The Seychelles archipelago was settled directly from Mauritius in the 1770s and therefore, Seselwa represents an off-shoot of Morisyen, diffused to the islands during this period (Chaudenson 2013). Finally, as previously mentioned, Tayo was recently discovered and represents the only known French-based creole spoken in the Pacific region. The colonization of New Caledonia did not begin before the 1840s and the local creole language, although resulting from a separate genesis which took place in a missionary school setting, was influenced by the creole spoken by indentured laborers who migrated from Reunion Island in the late 19th century (Speedy 2007a; Ehrhart 2012).
In the next graph, we have included French, the lexifier language, in the sample of 12 creoles. The binary values for the 66 FBC features were encoded for French and the data were added to the dataset. The results are presented in Figure 8.4.

![Figure 8.4 Neighbor-Net of 12 French-based creoles and their lexifier French with 66 features.](image)

The network in Figure 8.4 is virtually identical to the one presented in Figure 8.3, and the two major groups identified above are replicated: we find the Caribbean creoles on the left and the Indian Ocean varieties on the right. The topography of the internal groupings in the two networks remains similar to that observed in Figure 8.3: the clusters of the Lesser Antilles creoles appear in the same location in Figures 8.3 and 8.4, and so do the Indian Ocean creoles. Hence, adding French to the sample did not drastically affect the results. It is worth noting that French actually appears between Lou and Tay on one side and Reu on the other, and all three creoles are known for being particularly acrolectal varieties. Therefore, this corroborates the idea that the unexpected clustering of Lou with Tay could actually be conditioned by the fact that these languages are structurally relatively close to one another and to French. In the next section, we will attempt to offer a measure of the relative radicalness of the creoles in our sample in relation to their lexifier, French.
8.5 Measuring radicalness

Already in the early days of creolistics, scholars such as Van Name (e.g. 1870: 128) were aware of the fact that creoles can be placed on a continuum, ranging from varieties extremely close to the lexifier to deep creoles, and that this variation could possibly be explained as resulting from the influence of the lexifier on the creole. The linguistic consequences of decreolization are now accepted by virtually all creolists (although see Kihm 1984; Schwegler 2000; Mufwene 2001) and various clines of creoleness have been proposed (Alleyne 1980; Schneider 1990; Parkvall 2000; Baker 2001). Although the exact factors (and their interplay) affecting the degree of divergence from the lexifier are still far from being clearly understood, we know that the more radical creoles tend to be the ones that have not remained in close contact with their lexifier for a long time after their formative period (e.g. the English-based maroon creoles of Suriname). This is in accordance with the concept of decreolization. Bloomfield (1933: 474–475) proposed that the longer a creole co-exists in a diglossic situation with their lexifier, the more lexifier-like this creole will become over time. Hence, a crucial factor determining the relative degree of radicalness of a creole is linked to the demographic evolution of the population (Bickerton 1981).

Situations where exposure to the lexifier was limited or even non-existent during the formative years of a creole tend to produce more radical results (e.g. Haitian or the Suriname creoles). At the same time, the degree of radicalness of a creole can be seen as a function of demographic growth: the quicker the numbers of imported slaves grow, the more radical the resulting creole tends to be (see e.g. Parkvall 2000).

In order to assess the degree of radicalness of the selected French-based creoles in relation to their lexifier, French, we used the same sample as for Figure 8.4 and specified French to be at the root of the network. The resulting rooted network in Figure 8.5 shows the French-based creoles with French as the outgroup.

The results confirm the suggestion we made above to explain the cluster including Lou, Reu and Tay, namely that all three languages appear closest to the root, thus reflecting their relative acrolectality. However, Reu and Tay now appear further removed from one another. Although several authors (e.g. Alleyne 1980: 18; Bickerton 1984: 182; Lefebvre 1998) have attempted to evaluate the relative radicalness of individual French-lexicon creoles, most of these attempts were based on impressions. For instance, Lefebvre (1998: 4) claims Haitian to be a radical creole, and Louisiana, Tayo and Réunion are traditionally regarded as the least radical ones. Apart from Parkvall (2000), no other study has, to our knowledge, yet proposed an empirically-grounded cline of radicalness for French-based creoles.1

1. Note that Parkvall’s study also takes into account creoles based on other lexifier languages (Dutch, English and Portuguese).
Therefore, our results will only be evaluated against the results given by Parkvall (2000: 188). The cline of radicalness for the French-based creoles, based on the distance from the root (French), is presented in Table 8.2.

Table 8.2 Degree of radicalness of the 12 creoles included in the present study compared to Parkvall (2000).

| Degree of radicalness | Parkvall 2000 | hai > gwa, guy > lou > mar > mau | Present study | hai > dom > tri > stl > guy > mar > sey > gwa > mau > lou > tay > reu |

Most creolists will no doubt agree on the position of the various creoles on the cline of radicalness thus obtained through the phylogenetic analysis. This cline also corresponds fairly well to our own intuitions and is also largely in accordance with the one proposed by Parkvall (2000), even though a few discrepancies can be observed. For instance, Haitian was in both cases correctly identified as being the most radical French-based creole, followed by Guyanais, also in both studies.
However, when it comes to the lower end of the scale, creoles such as Gwa and Mau appear to be less radical than what would be expected.

Since the network analysis provides numerical values for the degree of divergence between the various languages included, we can also provide their relative distance from the lexifier. Those are given in Table 8.3.

<table>
<thead>
<tr>
<th>Language</th>
<th>Distance from root</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hai</td>
<td>0.56</td>
</tr>
<tr>
<td>Dom</td>
<td>0.54</td>
</tr>
<tr>
<td>Tri</td>
<td>0.53</td>
</tr>
<tr>
<td>StL</td>
<td>0.52</td>
</tr>
<tr>
<td>Guy</td>
<td>0.48</td>
</tr>
<tr>
<td>Mar</td>
<td>0.46</td>
</tr>
<tr>
<td>Sey</td>
<td>0.46</td>
</tr>
<tr>
<td>Gwa</td>
<td>0.44</td>
</tr>
<tr>
<td>Mau</td>
<td>0.39</td>
</tr>
<tr>
<td>Lou</td>
<td>0.37</td>
</tr>
<tr>
<td>Tay</td>
<td>0.31</td>
</tr>
<tr>
<td>Reu</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Hence, the cline of radicalness we obtained for the selected French-based creoles is largely in accordance with earlier claims concerning the relative radicalness of French-based creoles. The acrolectal varieties Lou, Reu and Tay appear at the lower end of the cline closest to the root, French, with Gwa and Guy occupying intermediate positions, and Haitian appears as the most divergent creole with regard to French, thus confirming the results of Parkvall (2000).

8.6 Discussion

The study of the origins and interrelationships between creole languages often involves investigations into the demographic composition of the territories under scrutiny, as the languages represented in each particular contact situation have to be accounted for so as to gauge their potential influence. However, the task of back-tracking population movements during this period, especially in insular settings, often proves to be daunting, as documentation is seldom complete and leaves many questions open. For instance, the transplantation of a language from one territory to another is known to have taken place for a number of creoles (see the papers in Huber & Parkvall 1997 for a survey), but in many instances, documentary evidence is simply lacking, thus making the task of ascertaining the origin of
certain shared features more difficult. In other words, can shared features between two (or more) creoles be attributed to a common origin, or did they come about as the result of diffusion or substratal influence, or do they simply represent cases of spontaneous parallel innovation in different places due to universal mechanisms underlying the process of creolization?

Several factors which have had an influence on the demographic history of the territories investigated, and hence on the linguistic outcome of the restructuring processes that have occurred, have to be kept in mind.

Firstly, the settlement history during the period of European colonial expansion is full of examples of migration between the islands, which was relatively easy and therefore frequent. Land owners often moved from one territory to another, principally due to wars between colonial powers and land shortage, and usually took their slaves with them (see e.g. Burns 1954), meaning that diffusion of a creole from one island to another was very likely – as was the case in Trinidad for instance. In other cases, however, population movements between islands were not recorded and in many cases, it is difficult to tease apart diffused features from independent parallel innovations.

Secondly, slaves were sold and moved to other islands independently of their owners. Moreover, slave ships were often taken by pirates so that their cargo would be sold back to local traders. Piracy and smuggling were highly lucrative activities during this period and were therefore appealing to a broad range of people. As the illegal movements of goods and people between islands was rarely, if ever, documented, we can only guess at the extent of these practices (see however Allen 2001 for an attempt at quantifying those) and their linguistic consequences.

Adding to the picture the incessant wars between the colonial powers further complicate the matter. Notoriously, the island of St. Lucia has for instance changed hands 14 times in the course of its history (Mitchell 2010:8). In such chaotic surroundings, the task of identifying separate geneses of creoles and of finding diffused linguistic features between these varieties becomes extremely complex and arduous.

### 8.7 Conclusions

The goal of this chapter was to conduct a typological study of a selection of French-based creoles worldwide using tools originally designed for studying the evolution of biological organisms. We have shown that the application of phylogenetic networks to linguistic data could add to our knowledge on the emergence and development of creole languages by focusing only on a subgroup of the world’s creole languages, namely the French-based varieties.
We selected a set of 66 lexical and typological features and the results we obtained were shown to be compatible with other proposals put forward by authors who have studied the relationships and affinities of French-based creoles.

Building on the same data set and analysis, we assessed the relative degree of radicalness of the creoles included in our sample. The distances between languages measured in this way reflected our own intuitions with regard to the degree of radicalness of the creoles in our sample and was in overall accordance with that of other authors. Thus, the results presented confirm that the effects of decreolization can be observed in creoles and that these indeed form a continuum ranging from very acrolectal varieties to deep basilects.

This study of French-based creoles was inspired by an earlier study of English-based creoles using similar methods (Daval-Markussen & Bakker 2011). In both cases, the computer-generated results were close to earlier classifications. The fact that the results obtained through the application of computational tools corroborate earlier proposals for classification shows that these tools are ideally suited to linguistic analyses, particularly for investigating cases where creolization has occurred.

Since we have not been able to interpret all our results in a satisfactory manner and since many questions remain open after this study, further work including other creoles derived from French (such as Carriacou or Rodriguais) and based on a selection of different features is required. Such an investigation will have to be more exhaustive in order to reveal further insights on the emergence and development of French-based creoles, as well as on their interrelationships.

Notes

The datasets for this chapter can be found here: https://phylogenetic-creole-studies.blogspot.com

References


This chapter examines the system of nominal number in Juba Arabic. Its complexity is compared with its superstrate language, Sudanese Arabic, and its substrate, the Nilotic language Bari. It also compares its similarity with a number of creoles, Nilotic languages and Arabic dialects. Both Arabic and Nilotic languages are notorious for their complex systems of nominal number, making Juba Arabic an interesting case for the study of creole complexity. Following Bakker et al. (2011), I use computational phylogenetic tools to look at the linguistic similarity of the languages under examination. The findings show that Juba Arabic is simpler than its constrictes, and that it is more similar to unrelated creoles than the languages from which it emerged.

9.1 Introduction

One of the major debates within the study of creole languages has been that of ‘creole exceptionality’. Some creolists claim that the notion ‘creole’ is merely a sociohistorical term, arguing that the emergence of a creole language is a process of gradual development like any other process of language change (Mufwene 2000, 2001, 2008; Aboh 2009; Ansaldo et al. 2007). Others claim that creole languages are distinct from non-creole languages with regards to grammatical structure and complexity due to them having emerged through a process of pidginization (McWhorter 1998, 2001, 2011; Parkvall 2008; Parkvall & Goyette forthcoming).

The results of recent large scale empirical studies using computational phylogenetic tools (Bakker et al. 2011; Daval-Markussen 2013; Daval-markussen, Bøegh & Bakker this volume) all seem to support the view that creoles constitute a typologically distinct group of languages. These studies have all been based on mass comparisons of grammatical features from typological databases such as WALS (Haspelmath et al. 2005) and APiCS (Michaelis et al. 2013), covering morphology and syntax, and to a lesser extent phonology. The present study uses some of the same tools to closely examine one particularly interesting part of the
The feature that this study focuses on is the system of nominal number in Juba Arabic and its relation to Arabic as well as a sample of languages from the Nilotic language family. Both the Arabic lexifier and Bari, the Nilotic substrate of Juba Arabic, are notorious for the complexity of their systems of nominal number (Dimmendaal 2002: 219; Neme & Laporte 2013). Therefore, it is particularly interesting to examine what happens in this part of the grammar in the languages that emerged from the linguistic contact between the two; especially taking into account McWhorter’s (2001) claim that “creole grammars are the world’s simplest grammars.”

Section 2 presents the two theories of creole development tested in this study. Section 3 contains the methodological preliminaries, giving an account of the phylogenetic tools used in this study as well as the methods of sampling and data collection. Section 4 contains an analysis of the linguistic data. Section 5 presents the results of the analysis and a discussion of the related findings. Section 6 discusses the theoretical implications of the findings and proposes a number of questions for further research. Finally, Section 7 concludes the study.

9.2 Theoretical preliminaries

Up until this day, there is little agreement among linguists on the definition of what constitutes a creole language and whether or not they make up a distinct typological group. A creole language is usually defined as a language that came into existence at a specific point in time in a situation of contact between groups that did not share a common language. These contact languages often have most of their lexicon from the language whose speakers were in a position of relative power over the other group(s) in the contact situation. This language will be referred to as the lexifier or the superstrate language. The lexifier is often the language of the colonial nations, whose influence spread with colonialist expansion. The language(s) spoken by the non-dominant group will be referred to as the substrate. In the case of Juba Arabic, the lexifier is Arabic and the substrate is the Nilotic language Bari. The term constrate will be used to when both lexifier and substrates are referred to. The role of the substrate and lexifier in the emergence of creoles and the typological status of these languages are still major issues of debate (Muysken & Smith 1994, 2015; Michaelis 2008, Bakker & Daval-Markussen 2014).

The typological status of creoles and the effects of the sociohistorical and cognitive processes involved in the emergence of creoles is a major topic of debate within the field of creole studies. Theoretically, this study focuses on two opposed theories on the typological profile and genesis of creoles: the feature pool hypothesis and the theory of creole distinctiveness.
9.2.1 The Feature Pool Hypothesis

Mufwene (2000) argues that creolization merely is a sociohistorical process. The processes of linguistic change in creolization are no different from the ones going on in language change in general. All changes in languages are explained as the result of competition between features found in different idiolects, dialects or languages, and in a situation of language contact the pool of competing features becomes larger and the competition more complex, but the basic mechanisms of selection are the same (Mufwene 2001: 6). Creoles are therefore not entirely new languages but merely seen as some of the latest languages or dialects to emerge within the family of their respective lexifier language (Mufwene 2008: 48). Aboh (2009, 2015), following the work of Mufwene, argues that no process of simplification occurs in the area of morphosyntax in creolization and that the notion of simplicity is irrelevant to understanding creole structure and genesis. Due to the larger pool of features, it would, on the contrary, be expected, that the languages are more complex and show a larger variety of structures.

The Feature Pool Hypothesis is testable as pointed out and demonstrated in McWhorter (2012: 175), i.e. features found in both the lexifier and the substrates should also exist in the creole, and features found in neither should not be present in the creole. Creole languages would then be expected to be part of the same language family as the lexifier, though with some influence from the substrate being present. McWhorter does not believe these statements about the creoles to be true and therefore considers the Feature Pool Hypothesis to be falsified.

9.2.2 Creole distinctiveness

Contrary to those advocating the Feature Pool Hypothesis, some creolists argue that creoles form a typologically distinct group. One of the first attempts to present a set of features that would separate all creoles from non-creole languages was undertaken by McWhorter (1998, revised in 2011). He claims that three features suffice to distinguish creoles from non-creoles: one phonological regarding tone, one semantic regarding derivation, and one morphosyntactic regarding inflection. Regarding morphosyntax he proposes that creoles have:

> little or no inflectional affixation, and among unbound inflectional markers, none of contextual inflection, or of inherent inflection (cf. Booij 1993 red.) of the paradigmatically complex sort.  
(McWhorter 2011: 111)

McWhorter (2001) makes another claim regarding creole distinctiveness, namely that the world’s simplest grammars are creole grammars. The claim that a typology of creoles could be based on a measure of simplicity was examined in a large-scale quantitative analysis by Parkvall (2008) and Parkvall & Goyette (forthcoming),
which confirms McWhorter’s statement that creole grammars are simpler than those of non-creoles. Studies of the typological status of creoles using mass comparison of large quantities of grammatical data powered by computational methods suggest that creoles form a typological group of languages, distinct from non-creoles methods (Bakker et al. 2011; Daval-Markussen 2013).

These studies all suggest that creoles form a distinctive group due to the sociohistorical context leading to the emergence of the creoles. Creolists such as McWhorter (2001) and Parkvall and Goyette (forthcoming) claim that it must have been due to the creoles’ past status as pidgins, whereas Bakker et al. (2011: 36) explain it more broadly as an effect of the rapid creation of a new language for means of communication between individuals without a common language. Bakker (2014: 445) points out that the loss of grammatical categories is not the only process happening in creolization. Rather, the reconstitution of lost grammatical distinctions through grammaticalization plays a major role in making creoles typologically distinctive.

9.2.3 Hypotheses

Based on the above-mentioned theories it should be possible to form two sets of hypotheses that can be tested using data from the languages under examination; one following the Feature Pool Hypothesis, and one following the Creole Distinctiveness Hypothesis.

1. Following the Feature Pool Hypothesis, one would expect: (a) all features found in Juba Arabic to also be found in one of its constrates; (b) all features found in both constrates to be present in Juba Arabic; (c) all features absent in both constrates to also be absent in Juba Arabic; (d) Juba Arabic to clearly group with the Arabic language sample in the phylogenetic tree based on structural features; (e) the creole languages of the sample to group along with their lexifiers.

2. Following the Creole Distinctiveness Hypothesis, one would expect: (a) the grammar of Juba Arabic to be simpler than the constrates, showing for instance little or no affixation; (b) the inflection found in Juba Arabic to not be of the paradigmatically complex sort; (c) Juba Arabic to group with the creole sample in the phylogenetic tree; (d) the creoles to form a linguistic group distinct from the non-creoles, not necessarily grouping along with their lexifier; (e) to find features in Juba Arabic that are not necessarily present in either constrate.
9.3 Methodological preliminaries

9.3.1 Phylogenetic trees and linguistics

The networks and trees in this study are produced using a distance-based method. This method works in two steps: First, a distance matrix is produced based on how much the languages differ with regards to the received data input. A distance of 1 from one language to another indicates that all features are different while a distance of 0 indicates that all features are shared. Once a distance matrix has been produced, various algorithms can be employed in order to produce a tree or network. For the large scale examination of all languages in this study, the Neighbour-joining algorithm is used. Neighbour-joining operates by computing a transformation of the produced distance matrix and pairing the languages that minimize the transformed distance (Nichols & Warnow 2008: 775). For close-up analyses of smaller numbers of languages, the Neighbor-Net algorithm is used (Bryant & Moulton 2004). Neighbor-Net has the advantage that it can depict lateral influences, which is an important factor in situations of contact. Since this study is purely cladistic, meaning that it examines the languages based structural similarity rather than historical family relations, all trees will be unrooted. This study uses the software SplitsTree (Huson & Bryant 2006).

There are various ways of testing the results produced by this kind of computational tool. One way is to benchmark the findings against widely accepted language classifications. Nichols and Warnow (2008: 777–779) suggest that the generated trees should contain established language groups, that no subgroups should be missing, and that no language from one group should be found in a group to which it does not belong, in order for the results to have a high level of validity. This of course presupposes that previous classifications are correct.

Another useful tool to test the validity of the proposed tree is by making a ‘bootstrap resampling analysis’, where the tree is generated for example 1000 times from a randomly selected subset of the data. If a single or few idiosyncratic feature(s) are responsible for the occurrence of a split in the tree, then this feature will be absent in a number of the resampled analyses. The analyses then measure the overall support for each split. If the split occurs in 100 percent of the resampled analyses, there is complete support for the given split, if it occurs 90 percent of the times, there is 90 percent support for the split (Dunn et al. 2008: 720).

However, none of these tests guarantee accuracy. If the analytical model used is inappropriate or the data is wrong or biased, the results will be unreliable in any event.
9.3.2 Sample

The selection of languages, their usual classification, and the main literature consulted on the languages is briefly summarized in Table 9.1. Since the main focus of the study is the historical affiliation of Juba Arabic creole, the sample has been chosen from three different language groups: a group of languages including and related to its lexifier; its substrates and constrates; and finally one consisting of a number of creole languages.

The language Juba Arabic emerged from contact between Arabic speaking soldiers from Northern Sudan and Bari speaking tribes after the soldiers settled in what is today the Juba area of South Sudan. Early forms of Pidgin Arabic have been reported to emerge in the area sometime between the 1850s and 1880s (Miller 2006). Today Juba Arabic serves the function of forming a major part of the national identity in that it distinguishes South Sudan from the Arabic speaking north (Mahmou 1983; Miller 2014).

The first group of languages that will be compared with Juba Arabic in this study are its lexifier, Sudanese Arabic, as well as Syrian Arabic and Classical Arabic. Syrian Arabic is chosen because it is a dialect that is spoken quite far from Sudan and historically it had very little contact with Sudanese Arabic. Classical Arabic is chosen because it is temporally removed from Sudanese Arabic and known to be structurally more complex than the modern dialects (Owens 2006: 9). This should give a rather broad spread of Arabic dialects, which is favorable in order to assert whether Juba Arabic should be counted as a dialect of Arabic. My main source for Sudanese Arabic is Bergman (2002). 1

The substrate language of Juba Arabic, Bari, is a Nilotic language of the Eastern Nilotic branch. The Nilotic sample contains one additional language from the Eastern branch, Maasai, and the Nuer and Dinka languages of the Western branch (cf. Dimmendaal 2007 on linguistic family relations within the Nilotic family). For general theoretical concerns regarding number marking in Nilotic languages, the analysis relies heavily on Dimmendaal (2002). Again the choice of languages from different branches of the same family will present a certain spread, which will help to validate the model of analysis. If the subfamilies of a greater language family are reproduced as sub-branches of the same branch, this will be an argument for the validity of the model of analysis.

Creole languages form the last sample. In order to get reliable and generalizable results regarding the nature of creole structures one needs a representative group of languages. The choice of languages in this study is based on lexifiers, with the goal

---

of including languages with non-Standard Average European lexifiers as well as one with a non-Indo-European lexifier (Bakker 2014: 443 on problems of sampling with regards to typological studies of creoles). This sampling method is ideal for testing the lexifier oriented Feature Pool Hypothesis versus the theory of creole distinctness because creoles have emerged from contact between languages that are structurally very different from each other. If they end up in the same group in the network, it can therefore not be due to common origin (i.e. a similar pool of features) but must be because creoles constitute a typologically distinct group of languages. The chosen languages are: Saramaccan, an English-based creole of Surinam; Karipúna, a French-based creole of Brazil; Korlai, a Portuguese-based creole of India; Hiri Motu, a pidgincreole based on the Austronesian language Motu of Southeast Papua New Guinea; and Nagamese, an Assamese-based language of India. Nagamese came about as a means of interethnic communication in trade rather than, as many of the others, in a situation of colonial expansion. It is known to be structurally fairly different from other creoles and is sometimes not classified as a creole, nevertheless it clearly groups with the creoles in the studies by Bakker et al. (2011).

Table 9.1 Language sample.

<table>
<thead>
<tr>
<th>Language</th>
<th>Affiliation</th>
<th>Spoken in</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juba Arabic</td>
<td>Creole</td>
<td>South Sudan</td>
<td>Watson (1984), Manfredi (in press), Miller (2006), Nhial (1975)</td>
</tr>
<tr>
<td>Syrian Arabic</td>
<td>Arabic</td>
<td>Syria</td>
<td>Cowell (1964)</td>
</tr>
<tr>
<td>Classical Arabic</td>
<td>Arabic</td>
<td>not spoken</td>
<td>Fischer (2002)</td>
</tr>
<tr>
<td>Bari</td>
<td>Eastern Nilotic</td>
<td>South Sudan</td>
<td>Mitterrutzner (1867), Müller (1864), Yokwe (1987), Spagnolo (1933)</td>
</tr>
<tr>
<td>Maasai</td>
<td>Eastern Nilotic</td>
<td>Kenya, Tanzania</td>
<td>Hollis (1905), Tucker &amp; Mpaayei (1955)</td>
</tr>
<tr>
<td>Nuer</td>
<td>Western Nilotic</td>
<td>South Sudan</td>
<td>Frank (1999), Crazzolara (1933)</td>
</tr>
<tr>
<td>Dinka</td>
<td>Western Nilotic</td>
<td>South Sudan</td>
<td>Andersen (2014), Nebel (1948)</td>
</tr>
<tr>
<td>Saramaccan</td>
<td>Creole</td>
<td>Surinam</td>
<td>McWhorter &amp; Good (2012), Bakker et al. (1994)</td>
</tr>
<tr>
<td>Korlai</td>
<td>Creole</td>
<td>India</td>
<td>Clements (1996, 2007)</td>
</tr>
<tr>
<td>Karipúna</td>
<td>Creole</td>
<td>Brazil</td>
<td>Tobler (1983)</td>
</tr>
<tr>
<td>Nagamese</td>
<td>Creole</td>
<td>India</td>
<td>Sreedhar (1985), Boruah (1993), Bhattacharjya (2007)</td>
</tr>
</tbody>
</table>
9.3.3 The data

In order for the analysis to be conclusive, the data needs to be sufficient, both qualitatively as well as quantitatively. It is difficult to ascertain how much data is needed, and of what type (cf. Bakker 2014). Following recent linguistic works using computational phylogenetic tools, the data used in this study is purely structural data (e.g. Bakker et al. 2011; Daval-Markussen 2013; Dunn et al. 2005, 2008). Whereas these studies mostly use the quite broad features of typological databases such as WALS (Haspelmath et al. 2005) and APiCS (Michaelis et al. 2013), the dataset used in this study was, for the most part, not predefined. Instead each data point is motivated through a close reading of the data. Since this is a cladistic study testing the typological affiliation of Juba Arabic with regards to marking of and agreement with nominal plurality, all structural distinctions made in the grammars of the languages of the chosen sample will be accounted for in a separate typological feature. The features exemplifying agreement with nominal number have been chosen by examining the typical targets of agreement (cf. Corbett 2002). The term ‘nominals’, in this account, covers both nouns and independent pronouns (Corbett 2002: 55; Daniel 2013).

This gives a relatively clear account for comparison of the ways the languages mark and agree with nominal plurality, in order to test whether the features present in Juba Arabic are taken from the lexifier or the substrate, or whether innovations are found, and if so, whether these innovations seem to be typical for the creole sample. Features will be coded as either present or absent for each language by means of a ‘1’ or a ‘0’ respectively unless otherwise noted. When the correct coding of a feature in a language is unknown, it will be coded with ‘?’. When a means of number marking is optional in a language, it will be coded as present.

9.4 The pool of features

9.4.1 Number affixing

All features dealt with here concern number affixing added to an unchanged inflectional base (cf. Corbett 2002: 139). Cases where affixing co-occurs with stem-change, as the Nuer example (1) below, are accounted for in F17.

(1) lɛp → lɛɛ\-ni
needles\sg\-gen needles\pl;gen-pl

Frank (1999:51)
### Table 9.2 Features concerning number affixing.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>plural suffixing</td>
<td>Plural marked by a suffix contrasting with an unmarked singular noun.</td>
</tr>
<tr>
<td>F2</td>
<td>replacive suffixing</td>
<td>Both singular and plural marked by contrasting suffixes. Replacive patterns are common in the Nilotic languages (Dimmendaal 2000:214).</td>
</tr>
<tr>
<td>F3</td>
<td>plural prefixing</td>
<td>Plural marked by a prefix contrasting with an unmarked singular noun.</td>
</tr>
<tr>
<td>F4</td>
<td>singulative suffixing</td>
<td>Singular marked by a suffix contrasting with an unmarked plural noun.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>This includes singulative marking of collective nouns.</td>
</tr>
<tr>
<td>F5</td>
<td>dual suffixing</td>
<td>Dual marked by a suffix.</td>
</tr>
<tr>
<td>F6</td>
<td>Transfixing</td>
<td>Number marked by replacive transfixing. A transfix is understood as a discontinuous affix placed within root, not as discrete entities but as one single meaningful unit, as shown in examples (2) and (3) below.</td>
</tr>
</tbody>
</table>

(2) $s^\text{a}\text{d}i^\text{q} \rightarrow a^\text{s}^\text{d}i^\text{q}\text{a}$  
friend$^{sg}$ → friend$^{pl}$  
’a friend’ → ‘friends’

(3) $k^\text{a}lb \rightarrow k^\text{i}la^\text{b}$  
dog$^{sg}$ → dog$^{pl}$  
’a dog’ → ‘dogs’

Sudanese Arabic and Bari both show a large variety of features relating to affixation. Sudanese Arabic has plural suffixation as well as replacive patterns, singulative suffixes on collective nouns, dual suffixing, and transfixing, and Bari has plural suffixation as well as replacive patterns, plural prefixing, and singulative suffixing. The only feature present in Juba Arabic is plural suffixation.

The system is most definitely simpler in the creole than in the constrates with regards to structural variation in affixing. If we examine the system of plural suffixing in Juba Arabic in comparison with its constrates, this picture only gets clearer. In Juba Arabic we find the plural suffixes $-\text{át}$ and $-\text{ín}$, stemming from the Arabic suffixes $-\text{āt}$, and $-\text{īn}$, where the long vowels have been exchanged with a short vowel with a pitch accent, and the plural suffix $-\text{jín}$ from the Bari plural suffix $-\text{jīn}$. In Juba Arabic the suffix only marks plurality and the choice of suffix seems lexically motivated with some variation accepted. The suffix $-\text{ín}$ is used mostly on animate nouns and adjectives, and $-\text{jín}$ only occurs with a number of substratal and adstratal nouns (Manfredi 2014).

In Arabic, besides marking plurality, the suffixes also mark gender, making them paradigmatically more complex than in Juba Arabic. If we compare this with the plural suffixes in Bari, the complexity comes from the variety of possible suffixes. In Juba Arabic there are three possible suffixes, whereas in Bari there are more than 20 possible suffixes. It should be noted while examining comparative
complexity with regards to affixing that a major contribution to the complexity of
the Arabic system of nominal number is the existence of more than 300 available
transfixing patterns (Neme & Laporte 2013).

Table 9.3 Coding of affixing.

<table>
<thead>
<tr>
<th></th>
<th>JA</th>
<th>SU</th>
<th>SY</th>
<th>CL</th>
<th>BA</th>
<th>MA</th>
<th>NU</th>
<th>DI</th>
<th>SA</th>
<th>KO</th>
<th>KA</th>
<th>NA</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>F2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>F3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

9.4.2 Number and noun stems

Table 9.4 Features concerning number and noun stem.

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| F7 | suppletion | Suppletion as a means of number marking is here understood as cases in which the difference between singular and plural is expressed by a complete change of the word. A criterion for establishing a feature as a case of paradigmatic suppletion and not as different lexical items is that there must be a plural paradigm in the language that one would expect nouns to be governed by. When the noun does not adhere to the expected paradigm and there is a different word expressing the plural meaning of the word, it will be considered suppletion (cf. Ljuba 2013).
| F8 | tonal change | Number marked by tonal change, e.g. Bari Ṽuṭū – Ṽuṭu ‘human – humans’ (Spagnolo 1933: 50).
| F9 | vowel height | Number marked by raising or lowering of vowel height, e.g. Bari kele – kala ‘teeth – tooth’ (Spagnolo 1933: 50).
| F10 | vowel length | Number marked by changed vowel length, e.g. Dinka djiir – djiir ‘grasshopper – grasshoppers’ (Andersen 2014: 230).
| F11 | Breathiness | Number marked by change in breathiness, e.g. Nuer thiñ – thin ‘breast – breasts’ (Frank 1999: 29).
| F12 | Diphthongization and monophthongization | Number marked by diphthongization of monophthongs, or monophthongization of diphthongs, e.g. Nuer tuŋ – tuŋŋ ‘spoon – spoons’ or thɔŋ – thɔɔŋ ‘leopard – leopards’ (Frank 1999: 29).
| F13 | partial irregular stem change | This feature covers a group of seemingly irregular changes of the last part of the stem, such as Bari kɔŋe – kɔnyen ‘eye – eyes’ (Spagnolo 1933: 50), or Maasai, e.g. giteŋ – gishu ‘ox – oxen’ (Hollis 1905: 32). Dimmendaal (2002) argues that this group forms a coherent pattern.
Stem-internal change as a means of number marking is a pattern found in great variety across the Nilotic sample. In both Sudanese and Juba Arabic suppletion is the only process found involving stem change. The few nouns with suppletive patterns of number marking found in Juba Arabic are suppletive forms that were either taken over from the lexifier or originated from transfixing patterns in Arabic, which implies that the change is now non-transparent and can only be regarded as suppletive (Manfredi in press). Again, number marked by stem-internal change is not very widespread across the creole sample.

Table 9.5 Coding of number and noun stems.

<table>
<thead>
<tr>
<th></th>
<th>JA</th>
<th>SU</th>
<th>SY</th>
<th>CL</th>
<th>BA</th>
<th>MA</th>
<th>NU</th>
<th>DI</th>
<th>SA</th>
<th>KO</th>
<th>KA</th>
<th>NA</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td>F7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

9.4.3 Collectives

Collective nouns are understood here as nouns that when unmarked for number refer to the plurality of something making up an internally coherent group and this is referred to as one. Singularization of collectives was dealt with in F4. Features concerning collectives are presented in Table 9.6.

Table 9.6 Features concerning collectives.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F14</td>
<td></td>
<td>plural of paucity</td>
<td>The plural of paucity is a subtype referring to a smaller collective than the unmarked one.</td>
</tr>
<tr>
<td>F15</td>
<td></td>
<td>plural of abundance</td>
<td>The plural of abundance is a type of pluralized collective noun referring to a larger quantity than the unmarked collective.</td>
</tr>
<tr>
<td>F16</td>
<td></td>
<td>plural of collectives</td>
<td>The plural of collectives is a type of pluralized construction that refers not to one bigger collective but to a number of different ones.</td>
</tr>
</tbody>
</table>

We demonstrate the different collectives using examples Bari and Syrian Arabic in (4) and (5).

(4) dābbānāt
    flie.paucity
    ‘some flies’
The pattern that we saw in the two previous paragraphs holds for the data relating to collectives as well. No subgroupings of collective nouns are found in Juba Arabic even though a number of these are found in both constrates. Bari has both a plural of abundance and plural of collectives. Sudanese Arabic has a plural of paucity but due to insufficient description it is not possible to settle whether it has a plural of abundance. In either case it would again be an argument in favor of a process of simplification. If the feature is present in Sudanese Arabic it would be a case of a feature being present in both constrates but absent in the creole, whereas if it is absent in Sudanese Arabic it would be a case of competing features where absence (i.e. simplicity) is chosen over presence (i.e. complexity). Inflectional subgroupings of collectives seem to be absent across the creole sample.

**Table 9.7** Coding of collectives.

<table>
<thead>
<tr>
<th>JA</th>
<th>SU</th>
<th>SY</th>
<th>CL</th>
<th>BA</th>
<th>MA</th>
<th>NU</th>
<th>DI</th>
<th>SA</th>
<th>KO</th>
<th>KA</th>
<th>NA</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td>F14</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F15</td>
<td>0</td>
<td>?</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**9.4.4** Other types of number inflection of nouns

**Table 9.8** Other features concerning number inflection.

| F17 | co-occurrence of affix and stem change |
| F18 | plural word |
| F19 | reduplication |

Co-occurrence of any one of F1-F6 with any one F7-F14.
A plural word is here understood as an otherwise semantically empty word marking the following noun or noun phrase as being plural. See examples from Juba Arabic in (6) and (7).
Reduplication is here understood as partial or full repetition of the noun

(6) nas silá
    pl weapon
    ‘weapons’

(7) nas ardá
    pl termite
    ‘termites’
The existence of a prenominal word marking plurality is the only feature regarding inherent inflection found in Juba Arabic that does not exist in any one of the con- struct languages. The Juba Arabic word *nas*, also meaning ‘people’, can function as a plural word. When functioning as a plural marker the word is semantically empty and does not carry any meaning of ‘personhood’ or animacy, as in (6) and (7).

### Table 9.9 Coding of other inflectional features.

<table>
<thead>
<tr>
<th></th>
<th>JA</th>
<th>SY</th>
<th>SU</th>
<th>CL</th>
<th>BA</th>
<th>MA</th>
<th>NU</th>
<th>DI</th>
<th>SA</th>
<th>KO</th>
<th>KA</th>
<th>NA</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td>F17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F18</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>F19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

### 9.4.5 Pronominal number

Table 9.10 presents other features concerning pronominal number. Since pronouns are inflected for number in all languages in the sample, the system will not be evaluated on the presence or absence of number inflection but on whether number inflection of independent personal pronouns is marked by affixation or suppletion, coded with ‘1’ and ‘0’ respectively. F23 regarding dual pronouns will be coded normally for presence or absence. A pronoun will be regarded as suppletive when there is no recognisable recurring stem, and as pluralized by an affix, when it can be analysed as an affix added to a recurring stem (cf. Daniel 2013).²

### Table 9.10 Other features concerning pronominal number.

<table>
<thead>
<tr>
<th>F20</th>
<th>first person pronoun</th>
<th>Number inflection of the first person independent pronoun.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F21</td>
<td>second person pronoun</td>
<td>Number inflection of the second person independent pronoun.</td>
</tr>
<tr>
<td>F22</td>
<td>third person pronoun</td>
<td>Number inflection of the third person independent pronoun.</td>
</tr>
<tr>
<td>F23</td>
<td>dual pronoun</td>
<td>A pronoun used for two referents exactly.</td>
</tr>
</tbody>
</table>

Looking at F20–F23 I find it hard to argue that either suppletion or affixation is simpler or more complex than the other, taking the Juba Arabic second person pronouns as an example. Juba Arabic third person pronouns are pluralized by means of an affix -*kum*, that is not found anywhere else in the language. In a case such as Nagamese, in which all pronouns are pluralized with the same suffix as

---

2. ‘Suppletion’ in my terminology is called ‘person-number stem’ and ‘affixation’ is called ‘person-number affix’ by Daniel (2013).
the nouns, a strong case can be made for this being simpler than suppletion, since
the system is completely consistent.

Table 9.11  Personal pronouns of Sudanese Arabic, Juba Arabic, and Bari.

<table>
<thead>
<tr>
<th></th>
<th>Sudanese Arabic</th>
<th>Juba Arabic</th>
<th>Bari</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first person</td>
<td>ana</td>
<td>ana</td>
<td>nan</td>
</tr>
<tr>
<td>second person</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>masculine</td>
<td>inta</td>
<td>itá</td>
<td>dɔ</td>
</tr>
<tr>
<td>feminine</td>
<td>inti</td>
<td></td>
<td></td>
</tr>
<tr>
<td>third person</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>masculine</td>
<td>hu</td>
<td>úo</td>
<td>nye</td>
</tr>
<tr>
<td>feminine</td>
<td>hi</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Plural</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first person</td>
<td>nihna</td>
<td>anina</td>
<td>yi</td>
</tr>
<tr>
<td>second person</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>masculine</td>
<td>intu</td>
<td>itakum</td>
<td>ta</td>
</tr>
<tr>
<td>feminine</td>
<td>intan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>third person</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>masculine</td>
<td>hum</td>
<td>úmon</td>
<td>sɛ</td>
</tr>
<tr>
<td>feminine</td>
<td>hin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sources: Nhail (1975: 82) and Manfredi (pc) for Sudanese Arabic and Juba Arabic, and Spagnolo (1933: 80) for Bari.

When examining the pronominal system in Juba Arabic and its constrates (see Table 9.12), it is clear that the pronouns come from Sudanese Arabic. A difference, though, is that second and third person pronouns in Sudanese Arabic are inflected for gender whereas in Juba Arabic they are not. This is in line with the prediction by McWhorter (2011) that there will be no inflection in creoles of the paradigmatically complex sort.

Table 9.12  Other features concerning pronominal number.

<table>
<thead>
<tr>
<th></th>
<th>JA</th>
<th>SU</th>
<th>SY</th>
<th>CL</th>
<th>BA</th>
<th>MA</th>
<th>NU</th>
<th>DI</th>
<th>SA</th>
<th>KO</th>
<th>KA</th>
<th>NA</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td>F20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F21</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>F22</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>F23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
9.4.6 Number agreement

Table 9.13 collects the features concerning number agreement. In order to be sure of comparing like with like, the analysis is focused on the typical targets of agreement (cf. Corbett 2002: 178). The element determining the agreement is referred to as the ‘controller’ and the element whose form is determined by the agreement is referred to as the ‘target’. In the case of the indefinite and definite articles (F28 and F29), they will be coded as number marking being absent if they are either missing or not marked for number.

Table 9.13 Features concerning number agreement.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>F24</td>
<td>demonstratives</td>
</tr>
<tr>
<td>F25</td>
<td>singular indefinite article</td>
</tr>
<tr>
<td>F26</td>
<td>definite article</td>
</tr>
<tr>
<td>F27</td>
<td>adjectives</td>
</tr>
<tr>
<td>F28</td>
<td>verb agreement with subject</td>
</tr>
<tr>
<td>F29</td>
<td>verb agreement with number of the object</td>
</tr>
<tr>
<td>F30</td>
<td>split nominal number with numerals</td>
</tr>
</tbody>
</table>

Both F24 and F30 are present in both constrates but not in the creole. F24 (demonstratives) follows the proposed tendency for creoles to have little or no contextual inflection (cf. McWhorter 2011) and is thus a case of a feature having vanished leaving the system simpler in the creole than in the constrates. In the case of F30 the situation is slightly different. The Arabic split number systems
differ significantly from the one found in Bari. In Arabic, nouns after the numerals 3–10 are followed by a plural noun, whereas nouns following numerals above 10 are singular in form. In Bari nouns following numerals 1–19 are singular, nouns following numerals 20–99 are plural, and for nouns following numerals 100 and above either form is possible. No such system is found in Juba Arabic, a fact which can only be interpreted as simplification.

The existence of an indefinite article derived from the numeral ‘one’ has been proposed as a pan-creole feature by Daval-Markussen (2014). Having an indefinite article not distinct from the word for ‘one’ is cross-linguistically quite common among those languages that have an indefinite article (Dryer 2013b). This feature is found in both Sudanese Arabic and Juba Arabic (Manfredi p.c.) so our data does not point to it being an innovation particular to creoles. The data is not sufficient to be conclusive on this point.

Adjectives agreeing for number constitute a feature found in Juba Arabic as well as both constrates but in no other creole of the data. This also seems to contradict the statement by McWhorter that there should be no contextual inflection in creoles. If we compare the systems of adjectival number agreement between Juba Arabic and its constrates, it is again clear that there is some sort of simplification at play. In both Sudanese Arabic and Bari, number agreement on adjectives is obligatory whereas in Juba Arabic it is optional. In Arabic number agreement on adjectives is also affected by gender and animacy.

Table 9.14 Coding of agreement.

<table>
<thead>
<tr>
<th></th>
<th>JA</th>
<th>SU</th>
<th>SY</th>
<th>CL</th>
<th>BA</th>
<th>MA</th>
<th>NU</th>
<th>DI</th>
<th>SA</th>
<th>KO</th>
<th>KA</th>
<th>NA</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td>F24</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>F25</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>1</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>F26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F27</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F28</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F30</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

9.5 Phylogenetic analysis

Figure 9.1 shows a network based on all languages and all data points processed with the Neighbor-Net algorithm.
What we see in Figure 9.1 is a clear tripartite split of the languages into the following clusters: one containing the Arabic languages in the bottom left of the figure; one containing the Nilotic languages in the upper left of the figure; and one containing all creole languages including Juba Arabic in the rightmost cluster of the figure.

Before entering into the discussion of the theoretical implications of the outcome of this analysis, I will evaluate the validity of the tree based on the criteria explained in Nichols & Warnow (2008: 777–779). These criteria are based on the relation between established language groups. Before discussing the creoles, we will look at the well-established language groups. Both the group of Arabic dialects and the Nilotic language group come out as distinct groups. Furthermore, the subdivision of Nilotic into Eastern Nilotic, containing Maasai and Bari, and Western Nilotic, containing Nuer and Dinka (cf. Dimmendaal 2007), stands out as well, and no language that should be in one group is found in another. This supports the method of sampling and data collection. If we examine the cluster at the right side of the network, we see that it contains all the creole languages and that they do not seem to group as would be expected if the creoles were merely a continuation of the lexifiers. Figure 9.2 shows a Neighbor-Net containing all creole languages including Juba Arabic.
Figure 9.2 Neighbor-Net including 6 creoles with 30 features.

In Figure 9.2 the seeming lack of clear clustering is striking. From the point of view of the Feature Pool Hypothesis the creoles would be expected to group according to their lexifiers. The Indo-European-based creoles would be expected to group subdividing into a Romance branch containing Karipúna and Korlai, and an English-based branch containing Saramaccan. If this were to be explained by the Portuguese adstratal influence, one can ask why Saramaccan is closer to the French-based Karipúna than the Portuguese-based Korlai. The Arabic and Motu-based creoles would then be expected be complete outliers. On the contrary we see that Saramaccan and Karipúna form a cluster, Korlai being closer to the branch leading to Hiri Motu than to its nearest Indo-European-based neighbor, and the Assamese-based Nagamese being a bit closer to the branch leading to Juba Arabic than to its nearest Indo-European-based neighbor. This seems to go along the lines of the theory of creole distinctiveness.

Figure 9.3 shows the relation between Juba Arabic and the group of Arabic dialects.

Figure 9.3 Neighbor-Net over three Arabic dialects and Juba Arabic.
Figure 9.3 gives a quite clear picture that from a structural point of view Juba Arabic, on the far right of the network, is not part of the Arabic dialect cluster on the far left of the cluster, in contrast to what would be expected from the Feature Pool Hypothesis.

The trees and networks shown so far in this section have all been based on all features applied to all or to a subset of the languages. In what follows we will examine the linguistic affiliation of the language based on a subset of the data by first producing a tree based solely on features relating to number marking, as shown in Figure 9.4, and then on the features related to agreement 5.5.

![Figure 9.4 Neighbor-joining tree based on features related to number marking (F1–23).](image)

The tree produced based solely on the features related to inherent inflection, F1-F23, gives an overall picture that is similar to the tree based on all features shown in Figure 9.1. The creoles still form a distinct group, and the other established group as subgroups are reproduced. Figure 9.5, based only on the rather small sample of features related to agreement, shows a slightly different result.

In Figure 9.5 we see that the established groups and subgroups have collapsed. Dinka is found with Syrian and Classical Arabic, and Nuer and Bari are closer than the more closely related Bari and Maasai. This is due to an insufficient number of data points and/or the possibility that features based solely on agreement patterns are not sufficient to distinguish the established groups and subgroups from each
other. Nevertheless, we still see that the creole languages form a group distinct from the non-creoles based only on the data related to agreement. In spite of the insufficiency of the data on agreement patterns, this strengthens the case for creole distinctiveness.

9.6 Theoretical implications and questions for further research

The analysis clearly shows that Juba Arabic displays a far smaller variety of features marked ‘present’ than both constrates. Of the features coded for presence/absence only five features were present in Juba Arabic compared with 12 in Sudanese Arabic and 13 in Bari. The only two features related to inherent inflection shared by Juba Arabic and both constrates are F1 plural suffixing, and F7 suppletion. From the data available it is not possible to compare the full extent of suppletive patterns and their contributions to the complexity of the respective languages. When analysed, it is clear that the system of plural suffixing in Juba Arabic is simpler than in the constrates, showing neither the paradigmatic complexity found in Arabic nor the vast possible variety of forms as found in Bari. Even the much sparser selection of features related to agreement shows a similar overall picture, with two features being present in Juba Arabic, three in Bari and five in Sudanese Arabic, even though the data on agreement is not sufficient to draw more complete conclusions.

The findings of this study thus make a case against the claim by Aboh (2009: 340), who holds that
The notion of simplicity is completely irrelevant to the understanding of the structure and genesis of creole languages (…). This is because creole languages are linguistic hybrids. They emerged from the recombination of linguistic features from different languages.

It has been proposed by Aboh and others that this restructuring happens as part of a regular unbroken transmission of the lexifier languages with features from the substrates entering the slowly transformed language (Mufwene 2001, 2008; Aboh 2009). Neither the notions of restructuring nor hybridization seem to adequately explain the analysed data here. This is clear when examining the entire set of features. Of the seven features present in both constrates, only three are present in Juba Arabic. In the cases where a feature is present in just one of the constrates, only two out of 13 are found in Juba Arabic. These findings support what was found by Parkvall (2008), who showed that, based on a broad set of typological features, creole languages as a group stand out from non-creoles as being structurally less complex. The theory that simplification is crucial to understanding creole structures due to a shared past as structurally simple pidgins (cf. McWhorter 1998, 2011; Parkvall & Goyette forthcoming) seem more plausible to explain the findings of this study.

The results of the phylogenetic analysis carried out in Section 5 comply well with similar studies using computational phylogenetics to study creole typology (Bakker et al. 2011; Daval-Markussen 2013; Daval-Markussen et al., Chapter 7, this volume). Like in these studies, creoles are clearly identified as clustering with neither the lexifier nor the substrates, but instead making up a group of languages distinct from the non-creoles. Applied to the data of this study, the method has shown to be sufficient to establish the languages as belonging to the groups predicted by recognized genealogical grouping.

Owens (2001) claimed that the Arabic creoles had been neglected by both arabists and creolists, especially with regards to analyses of the comparative grammatical structure and possible restructuring of features from Arabic. Though a lot has been published since then, such as Wellens’ grammar of Nubi (2005), the special issue of the Journal of Pidgin and Creole Languages on Arabic-based Pidgins and Creoles (Manfredi & Tosco 2014), and the soon-to-appear grammar of Juba Arabic (Miller & Manfredi forthc.), the question of their structural relation with the Arabic dialects and with each other has not yet been seriously addressed. To include creole languages such as Nubi, which it has been proposed have common origins with Juba Arabic going back to an Arabic military pidgin (Miller 2007) in a study like the present one, could make a great contribution to both the study of creole languages and the study of Arabic dialectology. The same holds for Arabic-based pidgins such as Turku (Tosco & Owens 1993), Pidgin Madame (Bizri 2010) and Gulf Arabic Pidgin (Naess 2008), as well as various African varieties of Arabic.
9.7 Conclusions

The comparative study shows that Juba Arabic is significantly simpler than the constrates, and that innovation has taken place. The phylogenetic analysis established creoles as a group of languages distinct from the non-creoles. Also, the creoles did not group according to the lexifier. These findings make a clear case against the Feature Pool Hypothesis and the notion that creolization would be a matter of restructuring and hybridization (Mufwene 2001, 2008; Aboh 2009). Rather, the findings comply very well with the Creole Distinctiveness Hypothesis: due to their past as pidgins, the creoles are simpler than non-creoles (McWhorter 2001; Parkvall 2008; Parkvall & Goyette forthcoming.) and group together as a distinct typological group (Bakker et al. 2011).

References


In this chapter, we compare lexical and grammatical data from three Dutch-based creoles: Virgin Islands Creole Dutch, Berbice Creole and Skepi Dutch of Guyana. We consider the lexicons, as well as phonological and typological patterns, and both synchronic and diachronic comparisons are made. Typologically, Berbice Creole appears closer to Dutch than the other two creoles, but lexically it is the farthest from Dutch. For some of the analyses, phylogenetic software was used to visualize connections and distances between each of the creoles and the lexifier Dutch. The conclusion is that the three creoles came into being independently. From a diachronic perspective, the documented forms of Virgin Islands Creole Dutch show mostly a gradual move away from Dutch and toward a more creole-like profile in the 20th century.

10.1 Introduction

In this chapter, we compare three creole languages that have a Dutch lexical base. The creoles are Skepi Dutch (Skepi) of Guyana, Berbice Dutch (Berbice) of Guyana, and Virgin Islands Creole Dutch, the latter based on data from the 18th through to the 20th century. The approximate locations of the three creoles can be found on Map 10.1. This study goes beyond Stolz (1986), Bruyn & Veenstra (1993), and Robertson (1989) in that both lexical and grammatical features are discussed from synchronic and diachronic perspectives.

We first introduce the three creoles in Sections 10.2 to 10.4. In Sections 10.5 and 10.6, we compare the three creoles, both from a lexical and a structural-typological point of view. As in the other chapters, we use computational tools developed in evolutionary biology to make automatized comparisons between the different languages, including network analysis and measures of distances.
10.2 18th-Century Virgin Islands Creole Dutch and 20th-Century Virgin Islands Creole Dutch

Virgin Islands Creole Dutch (VICD) has been documented for a period of 250 years, which is quite exceptional for a creole language. There are written documents for VICD covering a period of 250 years, as well as later tape recordings (Van Rossem & van der Voort 1996). The language was formerly spoken in the Danish West Indies, from 1917 the U.S. Virgin Islands. The first documents available in the language are letters written in VICD by slaves in the 1730s. The last speaker, Alice
Stevens, died in 1987. She had learned the language from her grandparents, and had worked with linguists who made tape recordings of her speech. Van Rossem and van der Voort (1996) produced a valuable anthology of texts covering two and a half centuries of VICD documents. That book has since been supplemented by publications of new text materials and analyses by Stein (2010a, b), Sprauve (2010), van Sluijs (2011) and Sabino (2012), among others.

It has long been known that there are considerable differences between 18th-century VICD and 20th-century VICD. Linguists have dealt with this discrepancy between older and newer periods of documentation in very different ways. Bickerton (1981: 74–75) did not include VICD in his studies of creole languages because he did not want to rely on written sources alone. Sabino (2012) does not consider the older missionary sources as completely authentic because they would be affected by grammatical prescriptivism, and therefore the older sources differ from the 20th-century sources. The latter are the only ones that Sabino considers reliable. Kouwenberg (2013: 884) does not consider some of the early texts “true” creole either. Unfortunately, a number of studies on VICD blend early sources with newer sources, as if the structure of the language had remained unchanged over a period of 250 years (e.g. Van Diggelen 1978; Bakker et al. 1994; De Kleine 2007). This gives a potentially distorted picture of the structural features of the language. In this chapter, earlier and newer sources are kept separate. Some creolists have attempted to map the diachronic developments (Stolz 1987; Muysken 1995, 2003; Hinskens 1998; van Sluijs 2014; see also Van Rossem & van der Voort 1996) and they observed considerable differences between older sources and 20th century sources.

Diachronic data in VICD suggest a gain through time of features commonly associated with creoles (see below), but also a development away from Dutch in the direction of a more typical creole profile. There are exceptions. Bimorphemic question words of the type what-reason for why or what-place for where have long been seen as typical for creoles (see Bickerton 1981; Muysken & Smith 1990; Holm & Patrick 2007). VICD, however, had more bimorphemic question words in the 1700s than were encountered in 20th-century texts, as can be seen in Table 10.1. Bimorphemic forms not directly derived from Dutch are shown in bold.

On the other hand, the absence of productive reduplication in earlier sources and its presence in later sources, indicates a gradual creolization (cf. Sørensen & Bakker 2003 for VICD). In this paper, we will distinguish two or three different time periods and treat Virgin Islands Creole Dutch spoken during each of these periods as a distinct variety or language.
Table 10.1 Question words in two different varieties of Virgin Islands Creole Dutch. Innovated bimorphemic forms are shown in bold.

<table>
<thead>
<tr>
<th>Modern Dutch Question words</th>
<th>Meaning equivalents English Question words</th>
<th>Older VICD Question words in Oldendorp 1767/8 (five or six bimorphemic, bold)</th>
<th>Modern VICD Question words in De Jong 1926 (two bimorphemic, bold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wat</td>
<td>what</td>
<td>wagoed, wat</td>
<td>wat, wa, awa</td>
</tr>
<tr>
<td>wie</td>
<td>who</td>
<td>wie</td>
<td>awi, widi (older: danawie)</td>
</tr>
<tr>
<td>waar</td>
<td>where</td>
<td>waar, na wat plek, wat plek</td>
<td>wà, wà api, wà apisi, api, apè</td>
</tr>
<tr>
<td>wanneer</td>
<td>when</td>
<td>wanneer, wattied, wol as</td>
<td>[not attested]</td>
</tr>
<tr>
<td>hoe</td>
<td>how</td>
<td>hoeso</td>
<td>huso (&lt; Dutch hoezo)</td>
</tr>
<tr>
<td>hoeveel</td>
<td>how much/how many</td>
<td>hoeveel, hoe moeschi</td>
<td>hovel, huwel, huwè</td>
</tr>
<tr>
<td>waarom</td>
<td>why</td>
<td>voor wagoed, wat maak</td>
<td>wama, awama</td>
</tr>
</tbody>
</table>

10.3 Berbice Creole

Dutch settlers and plantation owners had established themselves on the Wild Coast of South America from the early 1600s, following several decades of Dutch trade in the region (Hulsman 2009). The different colonies led to the emergence of several creoles, of which two have been documented, one of them being Berbice Creole, also called Berbice Dutch. Until recently Berbice has been spoken by people settled at Wiruni Creek, and formerly also along the Berbice River in Guyana, and perhaps elsewhere. Robertson (1983) mapped the known locations of Dutch creole speakers in Guyana in the 1800s, based on historical documentation.

There are several aspects of Berbice Creole in which it deviates from all other known creoles of the world in a unique way. An overview of the major deviant structural aspects of Berbice compared to other creoles can be found in Kouwenberg (1992). Most importantly, the basic lexicon of Berbice is derived from two different main sources, which is rather unusual among creoles. In most other creoles, African items are almost exclusively found in more peripheral areas of the vocabulary, such as religion, flora and fauna. Studies on African vocabulary in creoles can be found in articles in Bartens and Baker (2012) for a range of creoles. For other creoles the African elements fit the metaphor of “substrate”, in that only more in-depth studies dig up African features, whereas they are quite a basic part of Berbice Creole. With 38% of the Swadesh list from Eastern Ijo (Kouwenberg 1992: 264) rather than the main lexifier Dutch, only Saramaccan (English and
Chapter 10. Dutch creoles compared with their lexifier

(223)

Portuguese) and Angolar (Portuguese and KiMbundu) have comparable numbers of words from a second lexifier in the Swadesh list.

All languages are lexically blended to some extent, creoles probably less so than non-creoles, but the nature of the lexical mixture in Berbice creole is also unique in that it is mixed in basic lexical domains such as persons, kinship, and body parts (Kouwenberg 1994, 2012). No other creoles, and probably no non-creoles either, display etymologies from two different sources in these very basic semantic categories. The only exceptions are mixed languages of the Michif type (Bakker 1997), which came about in exceptional, but quite different, circumstances as well.

Berbice Dutch is also unique in its sources of bound morphemes. Creoles have often been characterized as languages with little morphology. Seuren (1998: 292–293), for instance, claimed: “If a language has a Creole origin it is SVO, has TMA [tense-mood-aspect] particles, has virtually no morphology.” Similarly, McWhorter (1998, 2005) claimed that creoles are languages with little or no inflection. Here, Berbice differs from most creoles in that it has at least three affixes of clearly inflectional nature: an imperfective suffix \( (a(r\epsilon)) \), a perfective suffix \(-\iota\epsilon\), and in nominal morphology a plural affix \(-apid\). In addition, there are some derivational suffixes and several processes of reduplication in Berbice Creole.

Furthermore, Berbice Creole has a few cliticized personal pronouns of Ijo origin that occur quite frequently (Kouwenberg 1993, 1994), and object clitic pronouns are quite unusual among creoles. Asian Portuguese creoles (especially Indo-Portuguese of Korlai and Diu) also have inflectional morphemes, and Palenquero Spanish creole uses a Bantu plural prefix, but Berbice is unique in the number of inflectional morphemes and the fact that the morphemes are found in both nominal and verbal domains. All of the inflectional and derivational suffixes are from Ijo, not from Dutch. Of course, the functions of these elements in Ijo and Berbice are not completely identical. For instance, the plural marker is generalized from a marker limited to human nouns in Ijo to one also marking nonhuman animates and inanimate nouns in Berbice (Kouwenberg 2012: 141). There are no Dutch affixes. The Ijo suffixes are used with both Ijo and Dutch nouns. The number and nature of African-derived affixes has no parallel in any other creole.

A third point in which Berbice differs from all creoles is the fact that the African component is from Ijo, and only from Ijo. Ijo is a group of languages spoken in southern Nigeria, formerly considered part of the Niger-Congo stock, but Dimmendaal (2011) considers the Ijoid languages as a separate family (see also Bøegh et al. 2016, for the deviant typological properties of Ijo within West Africa). In other Caribbean or South American creole languages, we encounter no words that are unambiguously derived from this language group. The overview in Bartens and Baker (2012: 281–282) lists percentages of African words in 22 creoles. Berbice has 100% from Ijo, most creoles have 0% from Ijo, and three languages have 1% or
2% of the African component from Ijo. However, scrutiny of the book chapters in Bartens & Baker searching for Ijo words reveals that many of the possible Ijo words have identical forms and meanings in other non-Ijoid African language groupings as well. In short, there may not be a single word that is unambiguously from Ijo in the other Atlantic creoles, and this contrasts greatly with Berbice.

Finally, even though the Berbice language has components of a European language (Dutch), an African language (Ijo) and an Amerindian language (Lokono, of the Arawakan family), the last generation of speakers were almost all Amerindians or of mixed Amerindian descent. How have these Amerindians adopted, or developed, a language with African and European elements? We provide no solution to these mysteries here. For discussion about the latter, see Kouwenberg (2009, 2012, 2013, 2015; Bakker forthcoming).

10.4 Skepi Dutch Creole

Skepi Dutch Creole developed along the Essequibo River in what is now Guyana, in a section formerly colonized by the Dutch and later by the British. The published materials are limited to a handful of sentences (Robertson 1983) and a list of 200 words (Robertson 1989). Therefore, we are only able to make a more detailed lexical comparison between Skepi and the other creoles. As for the grammar, the few sentences available for scrutiny (see the appendix in Bakker 2014, based on Robertson 1983) suggest that it is a creole language, structurally similar to other creoles, with a preverbal TMA system, SVO word order, no gender distinction, no inflectional morphology in verbs or nouns, and no inversion – all of which are features that Skepi shares with most other Atlantic creoles but not with Dutch. Lexically, it does not have the African component that is so prevalent in Berbice. We will give one historical example sentence here, which was found in a letter written in Skepi in 1780. The interpretation is mine, but we are much obliged to Margot van den Berg (2013) and Cefas van Rossem for the groundwork. See also Kouwenberg (2013) and <http://stemmenvanafrika.nl/skepi-nederlands-is-dit-geen-moye-taal/> and <http://www.hum.leidenuniv.nl/onderzoek/brieven-als-buit/over/brief-van-de-maand-december-2013.html>. The Skepi sentence is given in italics, and the word glosses in modern Dutch.

(11) *en sok kum kloeke dagka van Noom di sitte bi Warme lantta en zoek komen mooi dag van Oom die woont in warm land
en als um kom Weeran bi Bikkelante
en als hem kom terug in wittenland
*ham sel brangk van die 4 blabba moye Goeto
hem zal breng van die 4 kind mooie goed.*
‘en probeer op een mooie dag naar oom te komen die in een warm land woont en als hij teruggaat in Nederland zal hij voor de vier kinderen mooie goederen(dingen) meebrengen’ (modern Dutch)
‘…and try to come one beautiful day to Uncle who lives in a warm country, and when he comes back to the country of the White people, he will bring nice things for the four children.’

All words but two are from Dutch. Blaba could be an Akan word for child. Bikke is a widespread African word for European, found in French and English creoles, originating in Benuic (e.g. Igbo) and Ijoid languages, cf. Ijo beke (see Bartens & Baker 2012: 128). Kouwenberg (2013) suggests an Eastern Ijo etymon and thus a possible indication of contact between Berbice Creole and Skepi.

10.5 The Dutch creoles: Lexical comparison

In this section, we compare the Dutch creoles, using computational tools developed for tracking evolution, where appropriate (SplitsTree; Huson & Bryant 2006). The results are visualized in trees or networks.

Creoles can be characterized as languages restructured and expanded after a stage of rather severe simplification. Under such a characterization, there are exactly three Dutch-based creoles: Skepi, Berbice, and Virgin Islands Creole Dutch. Afrikaans is, in our view, not a result of expansion after severe reduction, and therefore not taken into account here. VICD differs so greatly at the various stages of its documented history that we take two (sometimes three) different varieties of it into account: late 18th-century and 20th-century VICD. When the earliest documents, that is, the slave letters from the 1730s, provide sufficient information, we also consider early 18th-century VICD as a distinct variety from later and much richer 18th century material.

We compare the three creoles on the basis of etymological source, both on the basis of all words from a 200-word selection, and the Dutch etymons only. Also, we consider lexical phonology, concretely phonotactics. Finally, a comparison is made on the basis of typological properties in the realm of morphology and syntax, where one can abstract away from the forms themselves. This has been done earlier by Daval-Markussen and Bakker (2011) for tracking the connections among the Atlantic English creoles, and in this volume by Sippola for Iberian creoles (Chapter 11) and by Daval-Markussen for French creoles (Chapter 8).
10.5.1 Origin of the roots

For the lexical comparison, we used Robertson's (1989) word list of ca. 200 English meanings collected for Skepi, as the point of departure. Robertson added the meaning equivalents of the same (English) meanings in Berbice and VICD. We have supplemented Robertson's list with additional Berbice and VICD words from other sources, notably Kouwenberg (1994) for Berbice, and De Josselin De Jong (1926) for VICD, in which more than a few additional words were located and added to the lexical database. Some English meanings have several translation equivalents, therefore the total number of words for each language is slightly more than 200.

Note that these are not Swadesh lists but rather a selection that covers some basic terms, some more specialized vocabulary, and some local lexicon. The criteria for compilation may be the words remembered by the final generation of semi-speakers of Skepi. The Berbice data were obtained from fieldwork and VICD equivalents (from De Josselin De Jong 1926) have been added by Robertson, supplemented by me. Note also that the numbers differ from those given in Robertson (1989), mostly due to the fact that words were added and more information on the etymology of Berbice items have become available.

Table 10.2 shows a quantification of the etymological origins of the retrieved words for the three languages. Table 10.2 makes it immediately clear that, as far as the origin of the lexicon is concerned, Skepi and VICD are much more similar to one another than they are to Berbice. In Berbice, only 63% of the 211 words of the 200-meaning list are from Dutch, compared to 91% and 90% in Skepi and VICD, respectively. The Berbice list has 30% West African words (Smith et al. 1987), compared to only 2% to 4% in the other creoles. (see Mufwene 1993 and Bartens & Baker 2012 for assessments of African components of creoles).

<table>
<thead>
<tr>
<th></th>
<th>Dutch origin</th>
<th>West African origin</th>
<th>Amerindian origin</th>
<th>English origin</th>
<th>Portuguese origin</th>
<th>Unknown origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berbice (211 words)</td>
<td>63% (132)</td>
<td>30% (63)</td>
<td>3% (6)</td>
<td>2% (4)</td>
<td>0% (1)</td>
<td>2% (5)</td>
</tr>
<tr>
<td>Skepi (211 words)</td>
<td>91% (192)</td>
<td>2% (4)</td>
<td>4% (8)</td>
<td>2% (5)</td>
<td>1% (2)</td>
<td>0% (0)</td>
</tr>
<tr>
<td>20th century VICD, (211 words)</td>
<td>90% (189)</td>
<td>3% (6)</td>
<td>2% (4)</td>
<td>2% (5)</td>
<td>0% (1)</td>
<td>3% (6)</td>
</tr>
</tbody>
</table>

Graphically, the lexical distances on the basis of the etymologies of roots between the three creoles can be presented as in Figure 10.1.
Chapter 10. Dutch creoles compared with their lexifier

Figure 10.1 Comparison of lexicon of Skepi, Berbice Creole, and Virgin Islands Creole Dutch, based on etymologies of a word list with 211 items.

Figure 10.1 does not represent an evolutionary pattern; the computational technique is only used as a means to make a graphic representation of lexical distances. The length of the branches (or edges) in Figure 10.1 represents the lexical distance between the three Dutch creoles in this respect. VICD and Skepi are much closer to one another because they have a similar number of roots from Dutch and other languages. Berbice creole appears quite distinct from the other two, obviously because of the presence of many West African roots.

10.5.2 Comparison of the Dutch roots

The next step in our lexical comparison is to compare the etymological sources of the Dutch words in the three creoles, based on shared etymologies. If a Skepi word, a VICD word and a Berbice word all have the same Dutch word as their source, the three words are counted as the same, even if their form may be somewhat different. If the languages use different Dutch words for the same meaning, then the creole words are counted as two or three distinct ones. Where no Dutch word was available for a particular meaning in Robertson’s list in one of the languages (this would most often happen in Berbice creole, when an Ijo word was used), that word was not included in the count.

132 of the meanings in the revised list yield words with Dutch etymologies in all three languages, and in 39 of the cases (30%), a different Dutch word is the etymological source in at least one of the languages, against 92 words (70%) with the same etymology. Examples include the word for “back”: in Skepi it is lent, from Dutch lende “side (of body)”, in Berbice atrí from Dutch achter “behind” and in VICD rigi from Dutch rug “back” (body part). Here all three forms have a different etymological source. Where Berbice Creole and VICD have a form for “beneath, under” from Dutch onder (ondro and onda respectively), Skepi has a form derived from beneden “down”, dialectal bene: bene. And VICD is the deviant one in “cup”: Skepi and Berbice have mok resp. mok(u) (Dutch mok “mug”), VICD has kopi (Dutch kop “cup” or diminutive koppie, kopje).
Taking all the 144 available words with a Dutch etymology from the word lists of the three creoles (i.e., including those words where more than one translation equivalent of the English meanings is available), it appears that VICD is the most deviant one of the three. Table 10.3 shows that VICD has 62% (107) common Dutch roots with Berbice and 67% (97) with Skepi, and Berbice is the language with most common roots with Skepi (78%, or 113). For five words, all three languages have different etymologies. This result is not surprising considering the different histories of the three colonies. The Berbice and Skepi speaking regions were geographically close. Contacts of both settlers and colonized people may be expected between settlements.

Table 10.3 Shared Dutch etymologies (common roots) between the three Dutch creoles

<table>
<thead>
<tr>
<th></th>
<th>Berbice</th>
<th>Skepi</th>
<th>Virgin Islands Creole Dutch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berbice</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skepi</td>
<td>78% (113)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virgin Islands Creole Dutch</td>
<td>62% (107)</td>
<td>67% (97)</td>
<td></td>
</tr>
</tbody>
</table>

10.5.3 Phonotactics of Dutch and Ijo words in Dutch creoles

The phonotactic patterns compared in this section are based only on the list of words generated from the 200 meanings found in Robertson (1989). For this study, presented in more detail in Bakker (2014), we surveyed only clusters found in this restricted data set, thus avoiding a bias toward the better-documented Berbice Creole and VICD, for both of which at least five times as many words are attested as for Skepi. As in the previous sections, Robertson’s (1989) list is supplemented with Berbice forms from Kouwenberg (1994) and VICD forms from De Josselin De Jong (1926).

The phonotactic patterns can be taken into account by looking at initial, medial and final clusters and combinations of vowels. We treat the Ijo and Dutch components from Berbice Creole separately; they appear to obey quite different phonotactic rules.

The Ijo words in Berbice Creole only allow one final consonant, the nasal -n, whereas the Dutch words in Ijo allow 11 final consonants or clusters. These clearly go back to Ijo and Dutch sources (cf. Harry 2003 for Ijo). The Skepi words show 21 final consonants and clusters, and VICD 16 different ones. There is only limited overlap between the two, as is visualized in Figure 10.2, where the length of the branches indicates the quantity of clusters not found in the other creoles.
Chapter 10. Dutch creoles compared with their lexifier

There are in total 22 initial clusters in the three creoles. Berbice allows 15 clusters, six are found in Ijo words, 11 in Dutch words, four of them only in the Ijo words, nine only in Dutch words, and three shared. Skepi has 15 initial clusters, and VICD 17 clusters. The Dutch lexical elements in the three creoles show much overlap, but again the Ijo elements deviate mostly from the others, suggesting that Ijo elements preserve Ijo phonotactic constraints.

Medial clusters are most numerous, but several of them also cross syllable boundaries. There are 45 medial clusters in the dataset, and each of the four creoles has a similar number not found in the other creoles. Berbice has 22 medial clusters, and both Ijo and Dutch elements have 13, of which only four are shared. Skepi and VICD have 24 and 29 medial clusters respectively. There is remarkably little overlap between the three creoles: only the clusters -mp- and -nd- are found in all three creoles.

Vowel sequences are a bit trickier, as it is not clear whether they represent diphthongs or separate syllables. There are only nine in all, four in VICD, six in Skepi, seven in Berbice, two of them shared between Ijo and Dutch elements. Again, we observe little overlap.
We can conclude that there is surprisingly little overlap between the phonotactic structures of the three Dutch creoles. This is also visible in the graph in Figure 10.3, which visualizes the relations between all of the 103 phonotactic features. The Dutch elements in Berbice appear to be the least deviant of the others. This lack of overlap can again be interpreted as pointing to a separate genesis for the three Dutch creoles, and different sets of dominant substrate languages for each of them. The fact that the Ijo and Dutch elements in Berbice Creole behave in very different ways, suggests that these elements follow their own norms according to source languages. These Ijo patterns are fairly close but by no means identical to modern Kalabari Ijo patterns (cf. Harry 2003), but we do not discuss possible changes and adaptations here. Perhaps this suggests a certain level of bilingualism among preceding generations at some point in time.

It has to be kept in mind, that this analysis is not based on a full database of the languages, but only on the final consonants and consonant clusters and vowel combinations found in a 200-word list with 211 words for each of the three creoles. As the sample is restricted to 200 words, some forms will certainly be missing. For example, Berbice also has initial spr- (springhan, but compare pringi from Dutch springen) and str- (strafu), medial -gr- (swagri) and -sl- (stisli) and the vowel sequence -uei in duei, none of which happen to be present in the 211 item word list, and therefore not taken into account.
For more detail, more analysis, and the datasets, see Bakker (2014). We have not dealt with epenthetic vowels inserted in etymological consonant clusters, paragogic vowels at the end of words, and consonants inserted before words that etymologically start with vowels.

In general, Berbice has the lowest number of phonotactic possibilities (the lowest number of final consonants, final consonant clusters, medial clusters, and vowel combinations, especially in the Ijo component). In some areas, VICD has more possibilities than Skepi (a wider range of final consonants, word-initial consonant clusters, vowel combinations), but sometimes Skepi has a wider range, e.g. word-medial and word-final consonant clusters. An explanation on the basis of different substrates would be one to explore, along the lines of Smith (1987) for the Surinamese creoles. It is clear that the Ijo influence on the Dutch component in Berbice was limited, because the phonotactic patterns for Ijo- versus Dutch-related words differ in Berbice Creole.

10.6 Typological comparison

10.6.1 Three varieties of Virgin Islands Creole Dutch and Berbice Creole

In this section, we compare morphosyntactic typological data for the Dutch creoles, where we also add Dutch and for VICD we include data from different time periods. Due to lack of information, Skepi is not taken into consideration here.

If one wants to study the similarities between languages that go beyond particular word forms, one must take typological features into consideration, such as the order of nouns and modifiers, the presence of prepositions versus postpositions, the presence or absence of gender, and the like. Examination of creoles (as defined independently by creolists) in terms of features – identified by typologists as either being typical for creoles, or as characteristic of some of the languages of the world – invariably points to the structural distinctness of creoles vis-a-vis noncreoles (see Bakker et al. 2011 for creoles worldwide; Szmrecsanyi & Kortmann 2009 for English creoles, and many of the studies in this book).

Phylogenetic models can help to show links between different creoles. We have selected a number of features in which the Dutch creoles are known to differ from one another. We limit our analysis in this section to morphosyntactic features.

For the typological comparison, we selected a number of features where previous research had shown that different values were used in different time periods for VICD. Hence, the selected features can shed light not only on typological connections with other Dutch creoles and Dutch, but also on temporal developments of VICD.
The features selected should not be taken to be representative of a typological profile of any sort. Note also that any numerical value of a typological feature summarizes a complex reality in a single value, but that does not exclude the possibility that some structural features also have alternative constructions with lower frequency. For Dutch, one could argue, for instance, that there is also reduplication, though not productive (tiktak, imitation of sound of a clock), or a possibility for double negatives (nooit niet litt. “never not”), but such minor patterns are not taken into consideration. Here we focus on the most general values of the selected features.

For VICD, we used four different types of sources from three time periods. For the first (the earliest) period, we used slave letters written between 1732 and 1760. For the second period, we used two types of sources from the 1760s: Oldendorp’s (1767–1768) dictionary, including the example sentences therein, and Oldendorp’s

Table 10.4 Structural features for four sources of Virgin Islands Creole Dutch, for Dutch, and for Berbice.

<table>
<thead>
<tr>
<th>Feature</th>
<th>VICD Slave letters, 1730s–1760s</th>
<th>VICD Oldendorp Dialogues, 1760s</th>
<th>VICD Oldendorp Dictionary, 1760s</th>
<th>VICD De Jong 1926</th>
<th>Berbice Creole</th>
<th>Dutch</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 with = and</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2 plurals – s/ en</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3 zero plural</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4 Dutch verbal morphology</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5 adjectives variable Ø/-e</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6 inherited verb-particle</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7 bimorphemic question words dominant</td>
<td>?</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8 bimorphemic question words present</td>
<td>?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9 infinitive om</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10 infinitive voor</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11 infinitive voor</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12 focusmarker</td>
<td>?</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>13 copula with adjectives</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14 preverbal Negation</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15 preverbal TMA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16 prepositions</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>17 postpositions</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>18 double negative</td>
<td>?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>19 possession N van/fan N</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>20 possession N “his” N</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21 possession N N</td>
<td>?</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>22 serial verbs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
dialogues between masters and slaves (Stein 2010b). The reason for taking these two sources from the same decade separately is the impression that these sources were grammatically different. One of them may be closer to the vernacular of the slaves, which could thus be verified. For the third period (the 20th century), De Jong’s (1926) primary materials were investigated.

We selected features based on suggestions adopted from Stolz (1986) and Sabino (2012), among others, not all of which are logically independent of one another. They are listed in Table 10.4. For example, features (2) and (3) both relate to plural marking, one relating to the presence of some plural suffixes, the other to the absence or presence of zero marking of plurality. For a more detailed description of the features, we refer to Bakker (2014). Here we just present a Table with 22 features.

Table 10.4 shows the scores given for the different features found in the sources from the three different varieties of VICD, as well as in Dutch and Berbice. Here, yes scores 1, no scores 0, and a question mark indicates that no answer could be provided based on the available data.

The data was analyzed using the computational tools devised for phylogenetic research (Huson & Bryant 2006). The results are shown in Figure 10.4. A network

![Figure 10.4](image-url)  

Figure 10.4 A network of Dutch, Berbice Creole, and three varieties of Virgin Islands Creole Dutch, based on typological data.
approach was chosen as it allows one to show not only the inherited features but also, potentially, borrowing and contact.

Figure 10.4 shows that the language of the slave letters is the closest to Dutch, followed by the language of Oldendorp’s materials (two different kinds; 1767–1768; 2010), and followed by the variety of 20th-century language documented by Josselin De Jong. This study shows that the differences between the two types of texts from Oldendorp from the 1760s are very small. These results again support the observation that the three varieties of VICD are quite distinct languages. Note that Berbice and 20th-century VICD, appear to be at roughly equal distances from Dutch.

A rooted tree, based on the same data, with Dutch as the first taxon, is shown in Figure 10.5. Figure 10.5 could be interpreted as a diachronic development away from a form of Dutch and toward 20th-century VICD, in an evolutionary framework, where Berbice develops parallel to VICD.

![Figure 10.5](image_url)  
**Figure 10.5** Rooted tree with Dutch, three varieties of Virgin Islands Creole Dutch, and Berbice Creole, based on typological data.
Figure 10.5 attests that Berbice and 20th century VICD are the furthest removed from Dutch, VICD of the slave letters is the closest to Dutch, and the VICD varieties documented by Oldendorp fall in between.

10.6.2 Skepi, Berbice, and 20th-Century Virgin Islands Creole Dutch: Grammatical traits

As previously mentioned, very little is known about the structural features of Skepi. Not even a dozen sentences have been published. Still, even when it is difficult to generalize, a few properties can be extracted from this limited material. For example, one can draw generalizations with regard to the question word system based on one question word only. With these cautionary remarks, we present the known structural properties of Skepi in Table 10.5, and their equivalents in Berbice, in VICD and in a category “most creoles”, roughly based on the data in Holm & Patrick (2007). Here too, binary scores are applied, for presence (1) or absence (0).

Table 10.5 Structural features of Skepi compared with structural features of other creoles and Dutch.

<table>
<thead>
<tr>
<th>Property</th>
<th>Berbice</th>
<th>Skepi</th>
<th>VICD</th>
<th>Most creoles</th>
<th>Dutch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Verbal inflection</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Nominal inflection</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Case in pronouns</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Bimorphemic question words</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Preverbal tense, mood, aspect</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Preverbal negation</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>SVO</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Inversion</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Serial verbs</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 10.6 A network of structural properties of the Dutch creoles.
By feeding these data into the Splits Tree software, the results shown in Figure 10.6 were obtained. Figure 10.6 shows that Skepi and VICD are structurally similar to one another, in fact identical. Longer branches represent more distance between languages. Both Skepi and VICD are close to other creoles, whereas Berbice is closer to Dutch than the other creoles, no doubt because of the presence of inflection in both Dutch and Berbice and the absence of preverbal negation in both. The closeness of Berbice and Dutch, however, is also somewhat misleading, because the concrete inflectional morphemes in the two languages are from different source languages, Ijo and Dutch respectively. The forms of the morphemes have not been transmitted from Dutch to Berbice Creole.

10.7 Conclusions

We have compared the Dutch creoles along a number of lexical and grammatical parameters: etymological composition of the lexicon, words inherited from Dutch, phonotactic differences and similarities, and typological features. Even if one disregards the African component of Berbice, the creoles are quite distinct from Dutch, and from one another.

Taking into account only the Dutch cognates shared by Virgin Islands Creole Dutch, Berbice, and Skepi, there is no reason to assume that any one of these creoles is derived from any of the others. The phonotactic patterns of the three creoles are quite different, and a significant minority of meanings across the three languages (between 22% and 38%) have different Dutch etymons. It is most likely, perhaps even obvious, that all of them emerged independently from one another. None of the three go back to the same creolized or pidginized ancestor. This conclusion is in line with findings of Silvia Kouwenberg (2009, 2013) with regard to Skepi and Berbice Creole, and it also appears likely for VICD as well.

From a typological point of view, Skepi and Virgin Islands Creole Dutch are quite close to one another and to other (Atlantic) creoles, but they are distant from Dutch, whereas Berbice is placed in between them. However, on the basis of a different, more elaborate set of typological features, Berbice is most distant from Dutch. These differences are exposed through the selection of particular features. The data from the three varieties of Virgin Islands Creole Dutch suggest a development over time away from Dutch and toward a more creole-like profile. This could also be due to variation, where earlier sources were based on less basilectal features.

This all indicates that an independent genesis of the three creoles is the most likely scenario, with only very moderate and superficial contact between Berbice and Skepi.
The results corroborate what has been suggested before about the possible connections and the differences among the Dutch creoles, and the development of Virgin Islands Creole Dutch. This is the first structural comparison of Dutch creoles that includes the limited Skepi data and suggests a more typical creole profile for this language.

Note

The datasets for this chapter can be found here: https://phylogenetic-creole-studies.blogspot.com

Acknowledgments

This chapter is based on an earlier article, which contains more detailed information and data about the topics dealt with here: Bakker, Peter. 2014. Three Dutch creoles in comparison. *Journal of Germanic Linguistics* 26(3): 191–222. The interested reader is referred to this article as well.

References


CHAPTER 11

Similarities and differences among Iberian creoles

Eeva Sippola
University of Bremen

This study provides an account of the classifications of Iberian creoles based on a phylogenetic network analysis of typological and lexical data. It maps the shared linguistic traits of these creoles and compares the differences between them. The results of the typological analysis support previous classifications, showing a clear division between Atlantic and Asian creoles, and the main subgroups are divided geographically. However, connections between the subgroups are less clear due to the degree of detail of the typological feature values. The similarities consist of typologically unmarked features that are shared with the lexifiers. The differences are a result of areal substrate and adstrate influence.

Keywords: Iberian creoles, areality, groupings, similarities, differences

11.1 Introduction

Creoles with a lexicon drawn from Spanish and Portuguese represent the oldest, most diverse, and most geographically dispersed creoles. This chapter provides an account of the classifications of this subgroup of contact varieties based on a phylogenetic network analysis of typological and lexical data. It aims to map the shared linguistic traits of the Iberian creoles and compare the differences between them.

Previous comparative studies of Iberian creoles (e.g. Cardoso et al. 2012; Clements 2009; Hagemeijer 2011; Jacobs 2012) have identified clusters based on geography and on structural and lexical features. However, most of these studies have concentrated on individual languages, areal subclusters, and features characteristic of a subgroup of creoles. To date, few studies have taken a wide comparative focus on these languages (for exceptions, see Hancock 1975; Ferraz 1987). The diversity of the varieties and the complex linguistic situations in which the creoles formed have made it challenging to offer a comprehensive account of shared features, although differences have been pointed out. To this we can add problems concerning the scarcity of the data available for many creole varieties.
Recently published databases and newly available computational methods offer the possibility of a wide comparative approach such as the one taken in this study. This study analyses both grammatical and lexical traits of 19 Iberian creoles. The data come mainly from the *Atlas of Pidgin and Creole Language Structures Online* database (Michaelis et al. 2013). The analysis is based on statistical modelling and computational tools for quantitative typology that are suitable for language contact situations, such as the phylogenetic programs SplitsTree and Mesquite (Daval-Markussen & Bakker 2011, 2012; Huson & Bryant 2006; Maddison & Maddison 2014). More generally, the phylogenetic classification of English-based Atlantic creoles by Daval-Markussen & Bakker (2011) has served as an inspiration for this paper. The focus of this study, however, is on the shared features and differences of languages from a wider geographical area, covering both Atlantic and Asian creoles.

Section 11.2 offers comments on the methods used in this chapter. Section 11.3 gives an overview of previous classifications of Iberian creoles and discusses some of the shared and differing features that have been identified among them. Sections 11.4 and 11.5 present the language sample and the typological features used in the analysis. Sections 11.6–11.8 cover the results, including the areal clusters of the creoles and the shared features and differences among them, followed by a discussion in Section 11.9. The conclusion follows in Section 11.10.

### 11.2 Methods

Both distance-based and character-based methods of phylogenetic analysis are used in this study. The software SplitsTree4 (Huson & Bryant 2006) is used to provide an overview of the classification of the Iberian creoles, their groupings, and the number of conflicting signals. The program creates phylogenetic networks using a distance-based method called Neighbor-Net (Bryant & Moulton 2004). First, the primary typological and lexical feature values are converted into a numerical matrix. Then the software is used to calculate the difference between each language in the sample by means of combining all the differences and similarities between the languages into a single distance measure. The resulting unrooted network (split graphs) is thus only based on language similarities and differences, and does not embody phylogenetic models as such (Wichmann et al. 2011: 207; Verkerk 2014: 12–13). In this case, the internal nodes of the graph represent the conflict between different splits in the data analysis. The conflict of the signal is produced in the contradictory groupings of taxa. The internal nodes do not indicate ancestors, and they do not point to specific features that are responsible for the conflicting signals (List et al. 2013: 145).
In addition to an impressionistic interpretation of the networks, we are interested in measuring the reticulation in them in a principled way. Consequently, auxiliary methods tested in this study include the calculation of delta scores in order to detect the involvement of each variety in the conflicting signal and to assess if this method can be applied to our empirical sample. According to Gray et al. (2010: 3925), the delta score for each taxon measures to what extent each language connects with the conflicting signal. It ranges from 0 to 1, and equals zero if the language is not involved in any conflicting signal (Wichmann et al. 2011: 210 give a more specific explanation of the calculation of the score).  

As a consequence of using distance-based methods, we cannot identify the characters or features that cause conflict in the networks (List et al. 2013: 145). Moreover, we cannot be sure that the clustering patterns are due to inherited features or common ancestry and not parallel development or borrowing. Therefore, a supplementary analysis was carried out using maximum parsimony implemented in PAUP* Phylogenetic Analysis Using Parsimony 4.10.b (Swofford 2002). Maximum parsimony analysis constructs the tree that requires the fewest changes in the values of each feature to evolve from an ancestor to all the observed descendants or language varieties (Dunn 2009: 1667). The strict consensus tree (here interpreted as a classification of synchronic features) was used to calculate ancestral states in order to identify those features that appear to be unique to certain clusters and those that have likely evolved independently or that have been subject to lateral transfer/borrowing. This mapping of the features was carried out with the help of the program Mesquite Tools for Phylogenetic Trees (Maddison & Maddison 2014).

11.3 Previous classifications

Iberian creoles have been studied since the early days of creolistics. Earlier research has focused mostly on individual languages, connections between a restricted number of languages, and areal subclusters. Here we will present the relevant findings for our comparative purposes. For historical and current information on Iberian creoles (speaker numbers, period of formation, and other sociohistorical information) we refer the reader to the descriptions in Volume II of APiCS (Michaelis et al. 2013).

1. A Q-residual score is a similar measure for calculating departures from a strict tree. However, it is dependent on scaling of the distances meaning that the lengths of terminal branches play a role. Gray et al. (2010: 3925) consider it to provide a more accurate value closer to the residual in standard statistics, but Wichmann et al. (2011: 212) note that the distinctiveness of one or more taxa in relation to the others should not affect the measure.
11.3.1 Early studies

Early creolists such as Adolpho Coelho and Hugo Schuchardt published work on Portuguese-based and/or Spanish-based creoles as far back as the decade between 1880 and 1890. Coelho can be characterized as an early universalist, pointing towards the mechanisms of imperfect acquisition in creole formation (Coelho 1881: 67, 69). Coelho supported his theory by pointing out similarities among Portuguese-lexifier creoles, such as the progressive preverbal marker ta and a number of shared lexical items, such as miste ‘need’ and papia ‘speak’ (Coelho 1881: 70; Holm 1988: 28). The father of modern creolistics, Schuchardt, published widely on Iberian creoles. In Kreolische Studien (Schuchardt 1882a, 1882b, 1883a, 1883b, 1884, 1888d, 1890) and other writings on Portuguese contact varieties (Schuchardt 1888a, 1888b, 1888c, 1889a, 1889b, 1889c), Schuchardt studied creoles from places ranging from São Tomé to the Philippines, grouping them based on their substrate languages and geographical areas. However, he tried to account for particular creole features individually, as a result of how they arose in each contact situation (Holm 1988: 30). The third early creolist of special relevance for Iberian creoles is Rodolfo Sebastião Dalgado, who collected Indo-Portuguese texts around the turn of the 20th century. He was the first to propose a theory of partial reciprocal transfusion for Portuguese-based creoles, where different varieties exerted mutual influence on each other. His hypothesis maintains that the Asian creoles were able to develop and maintain significant similarities to each other because of the economic and political links between the territories under Portuguese rule, which facilitated cultural interchange and population movements between them (Cardoso 2010).

The rapid growth of creole research that began in the 1960s included several studies on Iberian varieties and their connections to other creole varieties. A hypothesized early Portuguese pidgin played a significant role in the monogenetic theories of creole formation (e.g. Taylor 1971; Thompson 1961; Whinnom 1956, 1965). According to these views, the common structures shared by creoles were due to their origin in a Portuguese proto-pidgin that was used in West Africa and in Portugal and later carried to different parts of the world along with colonial expansion. This Portuguese pidgin would have been later relexified during the development of the English, French, Spanish, and Dutch creoles. The linguistic evidence of this variety is scarce, but varieties of this kind can nevertheless be shown to have existed based on observations preserved in literary works of the period (see, for example, Naro 1978; Kihm & Rougé 2013).

Another example of the monogenetic approach can be found in Whinnom’s account (1956) of the formation of the Philippine Spanish creoles, also known as Chabacano. He postulated that the Chabacano varieties were based on an earlier Asian Portuguese pidgin, which was later relexified by Spanish and taken to the
Philippines. It is true that certain similarities can be identified between Chabacano and the Portuguese-based creoles of Southeast Asia, especially those in Malacca and Macau. For example, these creoles have similar pronouns, a formally and functionally similar object marker *ku/ko or kon/kun(g)*, and a similar verbal marker *kaba* with completive functions. However, these shared syntactic structures are largely universal tendencies in creoles that have two closely related lexifiers, and no irrefutable linguistic evidence exists to account for a historical connection between a previous Portuguese pidgin and the Chabacano varieties spoken today (Clements 2009: 48, 53; Lipski 1988: 37). For example, Chabacano words that have been assigned a Portuguese origin, such as *agora* ‘now’ and *onde* ‘where’, are also attested in Old Spanish, and others, such as *endenantes* ‘before’ and *boneca* ‘doll’, are also found in contemporary dialects of Spanish (Lipski 1988: 34–35). Items that are more difficult to classify as Portuguese, Spanish, or substrate derived are *na* ‘LOC’, the concessive *maskin* ‘even if’ (but see Vazquez Veiga & Fernández 2012), *quilaya* ‘how’, and *cosa* ‘what’ (Lipski 1988: 35–36).

Similarly, arguing for a monogenetic base in the formation of Papiá Kristang, Hancock (1975) gave an overview of the different historical and grammatical connections between 14 Iberian-based creoles and a list of shared features and differences between them. The following creoles were included in the comparison: Papiá Kristang, Macau creole, Java creole, Sri Lanka Portuguese (Ceylon), Bombay creole, the creole of Annobon, the creole of São Tomé, Angolar, Principense, Guinea-Bissau Kriyol, Cape Verdean, Papiamentu, Palenquero, Ermiteño (a variety of Chabacano), Saramaccan, and the pidgin Lingua Franca. Hancock’s overview of the features was based largely on Taylor’s (1971: 294–295) list of creole features. He identified shared Asian and African features, such as the formation of a past participle with – *du*, the prepositions *na* and *ku*, and similarities across the TMA markers. He also observed differences in the combinatory possibilities of the TMA markers and in the use of the third person plural pronoun, which can function as a post-nominal pluralizing morpheme in some of the Atlantic varieties (Hancock 1975: 221–222). Hancock (1975: 214–215) stated that “the different characteristics of each Lusoasian, Lusoafri can and Luso Caribbean Creole may be attributed to each becoming depidginized in different linguistic environments, the common core remaining unaffected in each”.

11.3.2 Atlantic and Asian creoles

Using early creolists’ classifications as a starting point, Ferraz (1987: 348) made a first general classification of the Portuguese-based creoles based on structural and lexical features. He established two distinct groups among the Portuguese creoles of West Africa and four groups in Asia. In West Africa, he identified the Upper Guinea
creoles and the Gulf of Guinea creoles. In Asia, the varieties were divided into
two Indo-Portuguese groups, Malayal-Portuguese and Sino-Portuguese. Although
Ferraz (1987: 339) reasoned that the Portuguese creoles originated independently,
he also pointed out the close historical contact between the creole varieties during
their formative stages, especially in Asia. This contact during the formative states
led to shared grammatical features and lexical items in the Indo-Portuguese varie-
ties (see Dalgado’s theory discussed in 11.1). In Africa, Ferraz (1987: 348) described
the similarities of the creoles as having arisen from parallel development, with
influence from different substrates in the two or four areas mentioned. 2

The distinction between the Gulf of Guinea and Upper Guinea creoles is based
on the features listed in Tables 11.1 and 11.2 (Ferraz 1987: 342–348).

Table 11.1 Differences between the Gulf of Guinea and Upper Guinea creoles (Ferraz

<table>
<thead>
<tr>
<th>Feature</th>
<th>Upper Guinea creoles</th>
<th>Gulf of Guinea creoles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonetics and phonology</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depalatalization</td>
<td>no depalatalization</td>
<td>depalatalization</td>
</tr>
<tr>
<td></td>
<td>before /a, o, u/;</td>
<td>before /i, ĩ, j/</td>
</tr>
<tr>
<td></td>
<td>palatal consonants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>before /i, ĩ, j/</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word endings</td>
<td>no constraint</td>
<td>no constraint against</td>
</tr>
<tr>
<td></td>
<td>against consonants</td>
<td>consonants</td>
</tr>
<tr>
<td>Vowel harmony and sandhi rules</td>
<td>no vowel harmony</td>
<td>vowel harmony and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sandhi rules present</td>
</tr>
<tr>
<td><strong>Grammatical items and functions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impersonal pronoun a</td>
<td>not present</td>
<td>present</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word order</td>
<td>demonstratives and</td>
<td>demonstratives and</td>
</tr>
<tr>
<td></td>
<td>possessives precede</td>
<td>possessives follow the</td>
</tr>
<tr>
<td></td>
<td>the noun</td>
<td>noun</td>
</tr>
<tr>
<td>Diminutives</td>
<td>preserved</td>
<td>not preserved</td>
</tr>
<tr>
<td></td>
<td>meste</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Particle of obligation 'must'</td>
<td></td>
<td>se la/θε la/ θε ra/ya</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantifier muito</td>
<td>preceding the referent</td>
<td>following the referent</td>
</tr>
<tr>
<td>Serial verbs in complement clauses</td>
<td>no serial verbs in</td>
<td>'come', 'go', and 'speak'</td>
</tr>
<tr>
<td></td>
<td>complement clauses</td>
<td></td>
</tr>
</tbody>
</table>

2. This means that the languages in each subgroup show resemblances because they developed
together conditioned by similar social and historical processes, controlled by a centralized ad-
ministration, but with local differences, for example, in the number of speakers of the substrate
languages.
Table 11.2 Lexical differences between the Gulf of Guinea and Upper Guinea creoles (Ferraz 1987: 346–348).

<table>
<thead>
<tr>
<th>Lexical item</th>
<th>Upper Guinea creoles</th>
<th>Gulf of Guinea creoles</th>
</tr>
</thead>
<tbody>
<tr>
<td>'abdomen'</td>
<td>unreduced forms of Ptg. barriga</td>
<td>reduced forms, such as bega and bwega</td>
</tr>
<tr>
<td>'sky'</td>
<td>no prothetic vowel seu</td>
<td>prothetic vowel ɔ́pe, ɔ́se</td>
</tr>
<tr>
<td>'person, someone'</td>
<td>closer to Ptg. pessoa: ps00/pusua, arge/alge</td>
<td>nge, ninge (&lt; Ptg. ninguém 'nobody')</td>
</tr>
<tr>
<td>'to play, to enjoy oneself'</td>
<td>brinka &lt; Ptg. brincar</td>
<td>ɔ́nga and similar forms &lt; Ptg. folgar</td>
</tr>
<tr>
<td>'pretty, beautiful'</td>
<td>bunitu</td>
<td>g(l)avi, fumozu &lt; Ptg. formoso, nyuka</td>
</tr>
<tr>
<td>'quickly'</td>
<td>tpress &lt; Ptg. depressa</td>
<td>reduplicated forms, e.g. nja-nja</td>
</tr>
<tr>
<td>'son, daughter'</td>
<td>fiye, fiya, m(e)mine, m(e)mina, and similar forms</td>
<td>minu, mina, na</td>
</tr>
<tr>
<td>'stone'</td>
<td>words from Ptg. pedra</td>
<td>budu, ubudu</td>
</tr>
</tbody>
</table>

Turning to Ibero-Asian creoles, Ferraz again pointed out local Asian developments in grammatical and lexical structures (1987: 349–355). Features present only in Asian Portuguese varieties, and not in Africa, include:

1. Particular possessive constructions, such as Pedro sua casa ‘Pedro’s house’ and eu sua casa ‘my house’, which are not found in Africa.
2. Modifiers preceding the noun.
3. -mente as a productive suffix in adverbs of manner.
4. Lo (Indo-Portuguese) and logo (Malayo-Portuguese) as future markers.
5. Ja as the perfective marker.
6. Reduplication expressing plurality.
7. Por/para before a pronominal object in the Indo-Portuguese creoles, and ku in Malayo-Portuguese.
8. Portuguese ter ‘to have’ acquires the meaning ‘to be’ in the creoles.
9. Constructions derived from Old Portuguese laia ‘way, manner’.

In addition to these grammatical elements, Ferraz (1987: 354–356) provided lists of lexical items characteristic of some or all of the Portuguese varieties in Asia. Table 11.3 presents archaic and nonstandard Portuguese items from Asian creoles not found in West African creoles, while Table 11.4 lists words of substrate origin.
Table 11.3 Archaic and nonstandard Portuguese lexical items in Asian creoles.

<table>
<thead>
<tr>
<th>(Old) Portuguese item</th>
<th>Creole item</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) tarde ‘(in the) afternoon’</td>
<td>SLP(^3) a’tarde, PK a’tadi</td>
</tr>
<tr>
<td>a/(à) noite ‘(at) night’</td>
<td>PK a’nuti, TUGU a’notti, MAC a’note</td>
</tr>
<tr>
<td>frio, frialdade ‘coldness’</td>
<td>SLP frui’dade, BAT friu’dadi</td>
</tr>
<tr>
<td>adem ‘duck’</td>
<td>NOR ‘ada, DAM ad, SLP ‘ade, BAT adi, PK ‘a(r)di</td>
</tr>
<tr>
<td>na mais ‘only’</td>
<td>SLP na’ma(i)s, MAC na mas</td>
</tr>
<tr>
<td>secura ‘thirst, dryness’</td>
<td>NOR sekur, SLP sekura/sekur, BAT si’kura</td>
</tr>
</tbody>
</table>

Table 11.4 Substrate-derived items in Asian creoles.

<table>
<thead>
<tr>
<th>Substrate item</th>
<th>Creole item</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘clay’ &lt; Konkani-Marathi mati (Dalgado 1921:42)</td>
<td>DAM/goa ‘mate, MAC ‘mati</td>
</tr>
<tr>
<td>‘flower’ &lt; Indic phul ‘to bloom’</td>
<td>NOR/MAHE/BAT/BAC ‘fula, SLP ‘fu:la</td>
</tr>
<tr>
<td>‘frog’ &lt; Konkani manduk ‘edible frog’ (Dalgado 1921:27)</td>
<td>SLP, PL, BAT, MAC man’duku/o</td>
</tr>
<tr>
<td>‘gift’ &lt; Malay sãguvåt</td>
<td>DAM sa’gwat, SLP sago’ate/sago’ate, PK sa’gwati</td>
</tr>
<tr>
<td>‘jacket’ &lt; Malay baju</td>
<td>BAT, PK, MAC ‘baju</td>
</tr>
<tr>
<td>‘sad, sadness’ &lt; Malay saying ‘regret’, ‘pity’, ‘sorrow for’, ‘affectionate pining’, ‘love’</td>
<td>SLP sa’yan, MAC sa’yang, PK sa’yang ‘love’</td>
</tr>
<tr>
<td>‘walking stick’ &lt; Malay rotan</td>
<td>NOR rta, SLP, MAC ‘rta</td>
</tr>
<tr>
<td>‘washer man’ &lt; Malayalam mannattan</td>
<td>NOR mai’nat, SLP, MAC mai’nato, BAT mai’natu</td>
</tr>
<tr>
<td>~mannatti (Dalgado 1921:12)</td>
<td></td>
</tr>
<tr>
<td>‘watermelon’ &lt; Arabic battikh (Dalgado 1921:191)</td>
<td>NOR pa’tek, SLP, MAC pa’teka</td>
</tr>
</tbody>
</table>

Focusing on shared features, Holm (1988–1989: 266–267) examined the verbal markers of the Portuguese-based creoles and concluded that the differences in the patterning of the verbal markers lend support to the African subgroupings established by Ferraz (1987). However, he also pointed out that the creole varieties are interrelated not only through their European lexicon, but that they also share common syntactic features that do not derive from European Portuguese. Holm (2009) examined five additional features of Portuguese-based creoles in Asia that he proposed could originate from the general Portuguese pidgin developing in Africa in the 15th century. These are the verb ‘to go’ (vai/vay), which is derived from an inflected form in Portuguese, and other functional correspondences among African and Asian varieties, such as the completive kaba, the coordinating conjunction ku, the preposition na, and the negator nunca.

3. Abbreviations: BAT = Batavia Creole, DAM = Daman Creole, MAC = Makista, NOR = Northern Indo-Portuguese creoles, PK = Papiá Kristang, SLP = Sri Lanka Portuguese, TUGU = Tugu Creole.
11.3.3 Local developments

Today, we have access to a growing number of studies focusing on Iberian creoles. Taking a general approach, Clements (2009) examined two phenomena across the Iberian creoles. These present differences between the African and Asian varieties and among them: affixal tense-aspect markers and oblique, direct object, and indirect object markers. Clements (2009: 50) argued that the presence of affixal features depends on the degree of typological similarity of the languages present in the contact situation, markedness, and social conditions. Suffixal morphology is more likely in creoles that have formed in a more homogeneous contact situation and that have had a stronger presence of the target language, Portuguese (Clements 2009: 55). The oblique and object markers pattern differently between the African and Asian creoles. In the African varieties, the markers follow the predictions of the causal order hypothesis (Croft 1991: 185), in which antecedent roles (e.g. cause, comitative, instrumental) are coded similarly and differently from subsequent roles (e.g. benefactive, dative/recipient), but due to substrate influence, Asian creoles violate this hypothesis (Clements 2009: 67).

Several other studies have been made focusing either on smaller subgroups of Iberian creoles or on certain features. In the Atlantic area, a large number of shared structural and lexical items of African, West Atlantic origin point to a common origin for the Upper Guinea creoles, although the exact location of this origin is unknown (Baptista, Mello & Suzuki 2007; Jacobs 2012: 8; Quint 2000: 99–117). Jacobs (2009, 2012) has established a historical relationship between the Caribbean Papiamentu and the Upper Guinea creoles. This connection is based on the correspondences in morphology and several grammatical categories, such as pronouns, prepositions, question words, conjunctions, reciprocity and reflexivity, TMA markers, and auxiliary verbs. He also points out differences that separate Papiamentu from the Gulf of Guinea creoles and Palenquero (Jacobs 2012: 261). The African connection of Palenquero has also attracted attention, but the origins of this creole are not clear (Jacobs 2012: 13; Lipski 2008; Schwegler & Green 2007). In the Gulf of Guinea, a common origin is generally assumed for the varieties, based on lexical and structural comparisons (e.g. Ferraz 1979; Hagemeijer 2011; Parkvall 2000: 133).

The Asian creoles have received less attention from a comparative perspective. The interrelationships between the Indo-Portuguese varieties and the other Portuguese creoles in Asia have been noted on the basis of early studies (see, for example, Baxter 1996; Clements 2009; and the contributions in Cardoso et al. 2012). Similarly, historical connections on the Malacca-Macau and Malacca-Java axes have been established. Features that are shared along the Malacca-Macau axis include SVO order, accusative-dative markers derived from Portuguese com, negative perfective markers of the type nenang inda (PK) and inda nunka (MAC), vowel + [ŋ] as a reflex of Portuguese nasal vowels, and Malay lexical items (Batalha
In the Malacca-Java area, shared features are verb serialization to express directional motion and give-serialization, and the 1sg. poss minya (for further information, see Baxter 1996: 304). The idea of a possible relationship or shared origin between Chabacano and the Portuguese creoles of Asia surfaces occasionally (for example, Steinkrüger 2006), although a local, independent development of the Chabacano varieties is more widely accepted today (Fernández 2004, 2007, 2012; Lipski 1988, 1992).

To conclude, we notice that previous research has thus established two focal points for Atlantic creoles that have stimulated discussion about the interrelationships between areal subclusters: West Atlantic in the Upper Guinea area and the Bantu and Kwa substrate area in the Gulf of Guinea. In Asia, the emerging picture is on the one hand of substrate influence, and on the other, of partial diffusion between the varieties. In the following sections, we will examine whether computational methods and typological data show support for the groupings and hypotheses presented in the previous literature, and if so, to what extent.

### 11.4 Language sample

The analysis focuses on the 19 Iberian creole varieties included in the APiCS database. The varieties are (ordered geographically) Palenquero, Papiamentu, Cape Verdean Creole of Brava, Santiago Cape Verdean Creole, São Vicente Cape Verdean Creole, Casamancese Creole, Guinea-Bissau Kriyol, Angolar, Fa d’Ambô, Principe, and Santome in the Atlantic area (see Map 11.1), and Diu Indo-Portuguese, Korlai, Sri Lanka Portuguese, Papiá Kristang, Batavia Creole, Cavite Chabacano, Ternate Chabacano, and Zamboanga Chabacano in Asia (see Map 11.2).

![Map 11.1 Atlantic creoles in the sample.](image-url)
Map 11.2 Asian creole languages in the sample.

11.5 Feature data

The main feature sample for this study is based on the 130 APiCS features. This is the most comprehensive database that includes Iberian creoles. The 130 features are mostly structural and drawn from different areas of grammar: phonology, morphology, syntax, and lexicon. Although the selection is one of convenience, it is partly based on the features included in the World Atlas of Language Structures and on features that are typical of contact languages. The features are listed on the APiCS online database, and the division between feature categories employed in APiCS is presented in Table 11.5. In the database, the lexicon and phonology are underrepresented compared to morphology and syntax.

In APiCS, a feature has multiple values (between two and nine, depending on the feature). These values were manually coded into a matrix, using numerical coding between 1–6, meaning that the maximum number of values that could be present for a feature was six (features 94 Instrument relative clauses and 104 Focusing of the noun phrase). Additional information on the same features was added for Spanish and Portuguese in order to include the lexifiers in the comparison.

---

4. The difference between the original maximum, nine, and six is due to the fact that not all the APiCS languages are included in this study.
Table 11.5 Percentages of feature categories in APiCS.

<table>
<thead>
<tr>
<th>Category</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal categories</td>
<td>24 18.5%</td>
</tr>
<tr>
<td>Clausal syntax</td>
<td>19 14.6%</td>
</tr>
<tr>
<td>Argument marking</td>
<td>16 12.3%</td>
</tr>
<tr>
<td>Verbal categories</td>
<td>14 10.8%</td>
</tr>
<tr>
<td>Phonology</td>
<td>13 10.0%</td>
</tr>
<tr>
<td>Word order</td>
<td>12 9.2%</td>
</tr>
<tr>
<td>Lexicon</td>
<td>11 8.5%</td>
</tr>
<tr>
<td>Complex sentences</td>
<td>8  6.2%</td>
</tr>
<tr>
<td>Negation, questions, focusing</td>
<td>7  5.4%</td>
</tr>
<tr>
<td>Nominal syntax</td>
<td>6  4.6%</td>
</tr>
<tr>
<td>Total</td>
<td>130</td>
</tr>
</tbody>
</table>

APiCS offers a solid typological feature set designed especially with creoles in mind. The datasets have been prepared by experts in each language, and the feature values are often based on extensive fieldwork. The atlas offers a wide comparative picture, without the limitations of focusing only on specific areal clusters or lexifier features. Among the Iberian creoles, the sample includes current varieties and one historical variety, namely Batavia Creole. The data on extinct Batavia Creole are based on late 19th and early 20th century texts. The information on the features and languages included in the atlas is rather comprehensive. Information from the APiCS database is missing in only 2.3% of the traits on average. The great majority of the datasets have been completed in their entirety. Angolar, Sri Lanka Portuguese, Diu, and Fa d’Ambó are missing more information than the other languages, most probably due to the precarious situation of these varieties, but even the extinct Batavia Creole has a completion rate of 76.9%. However, as is to be expected, even this database presents some lacunas. There is a lack of information on the highly endangered Indo-Portuguese creoles, such as the Malabar creoles, which have not been described or documented comprehensively so far (but see Cardoso 2013), and Macau Creole Portuguese (Makista). In addition, no Afro-Portuguese or Afro-Hispanic varieties are included (see Perez et al. this volume, Chapter 12, for information on Afro-Hispanic varieties). These are clear limitations, but nevertheless, our study is the largest comparative picture offered so far.

Typological features have been avoided in comparative linguistics with a diachronic approach because they have few possible values and are prone to chance resemblance, their values are linked by implicational universals, and their values are often functionally motivated externally (Croft 2008: 230). However, typological features nevertheless offer a useful basis of comparison along predefined feature values, and are therefore most suitable for synchronic clustering/similarity measurements (Haspelmath 2010), such as the ones we are aiming for in our study.
11.6 Areal clusters

From the analysis of the 130 features from the 19 creole varieties in SplitsTree, we obtain the following graph. It is evident from Figure 11.1 that a signal of similarity can be established in these data. Languages that we know to be closely related from their external history, such as the Cape Verdean varieties or the Chabacano varieties, appear closer together in the graph. This means that languages that are closely related show similar grammatical and lexical patterns.

![Graph](image)

Figure 11.1 A split graph of a Neighbor-Net analysis of the 130 APiCS features in 19 Iberian creoles.

The general picture that can be observed from the graph is the division into Atlantic and Asian creoles. The split between these two main geographical areas is visible in the narrow area in the middle of the graph, with the Atlantic creoles on the right of the graph and the Asian creoles on the left. The main subclusters that emerge from the analysis are composed of the Gulf of Guinea creoles and Chabacano. Other, weaker clusters are formed by the Indo-Portuguese creoles, the Cape Verdean varieties, and the Malayo-Portuguese creoles. In addition, Guinea-Bissau Kriyol and Casamancese Creole appear between the Cape Verdean and the Gulf of Guinea clusters. However, there are also divergences from the pattern; Palenquero and Papiamentu show long independent lines on the Atlantic side, and thus do not form a Caribbean cluster.

The delta scores of the graph in Figure 11.1 are given in Table 11.6 for each of the 19 Iberian creoles. The average delta score is 0.3194. The lowest scores are among the languages that cluster together and are more separate from the webbing,
such as the Chabacano varieties and the Gulf of Guinea creoles. Languages that have a higher delta score are Palenquero, Papiamentu, and Guinea-Bissau Kriyol. In this particular case, it seems that these higher scores are connected to conflicting feature groups in the sample. This means that, for a part of the feature sample, these are similar to certain languages that form areal clusters, while for another part they pattern similarly to other languages in other areal groups. Therefore, the graph in Figure 11.1 should be interpreted in the first place as a map of typological types (Verkerk 2012: 73), which are based on areal features conditioned largely by substrate and adstrate influence.

Table 11.6 Delta scores for the Neighbor-Net analysis of 130 APiCS features in 19 Iberian creoles.

<table>
<thead>
<tr>
<th>Language</th>
<th>Delta score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zamboanga Chabacano</td>
<td>0.26162</td>
</tr>
<tr>
<td>Ternate Chabacano</td>
<td>0.26823</td>
</tr>
<tr>
<td>Principense</td>
<td>0.28064</td>
</tr>
<tr>
<td>Angolar</td>
<td>0.28412</td>
</tr>
<tr>
<td>Cape Verdean of Santiago</td>
<td>0.29853</td>
</tr>
<tr>
<td>Korlai</td>
<td>0.29901</td>
</tr>
<tr>
<td>Cavite Chabacano</td>
<td>0.29988</td>
</tr>
<tr>
<td>Santome</td>
<td>0.30374</td>
</tr>
<tr>
<td>Cape Verdean of Brava</td>
<td>0.31481</td>
</tr>
<tr>
<td>Sri Lanka Portuguese</td>
<td>0.32276</td>
</tr>
<tr>
<td>Fa d’Ambô</td>
<td>0.33137</td>
</tr>
<tr>
<td>Cape Verdean of São Vicente</td>
<td>0.33955</td>
</tr>
<tr>
<td>Diu</td>
<td>0.34191</td>
</tr>
<tr>
<td>Casamancese</td>
<td>0.34275</td>
</tr>
<tr>
<td>Papiá Kristang</td>
<td>0.34866</td>
</tr>
<tr>
<td>Batavia</td>
<td>0.35010</td>
</tr>
<tr>
<td>Papiamentu</td>
<td>0.35586</td>
</tr>
<tr>
<td>Guinea-Bissau Kriyol</td>
<td>0.35616</td>
</tr>
<tr>
<td>Palenquero</td>
<td>0.36984</td>
</tr>
</tbody>
</table>

Further testing was also done by selecting only one variety from each regional area and bi-clan (see Michaelis 2014), i.e. a combination of a lexifier clan and a substrate clan, but the results were very similar, although the number of languages was often very small. The Atlantic-Asian divide stays clear, as does the division into South Asian and Southeast Asian creoles, albeit to a somewhat lesser extent. Independent lines are long for all the varieties showing divergence on the level of individual languages. However, among the Atlantic creoles, regional clusters are not clear. The inclusion of the lexifiers, Portuguese and Spanish, does not affect
the clustering in Figure 11.1. The lexifiers are placed in the middle of the Atlantic-Asian divide, and their inclusion increases the webbing on the Atlantic side of the graph. This suggests that the lexifier, which in this case is Portuguese, shares mixed feature sets with the different Atlantic creoles, and these shared features do not follow areal clusters. As the results are not significantly different, we have decided not to reproduce the graphs here.

Subsets of features were also analysed. Phonological features from the APiCS database produced a mixed picture, due to the low number of features (13). Similarly, lexico-semantic information from the APiCS database did not produce any relevant clustering, as the number of the features (9) was too small. Networks with 108 structural/functional features from the APiCS database gave results that reproduce the groupings in Figure 11.1. These were chosen by excluding the lexical and phonological features. In a similar vein, a selection of 21 features from the nominal domain (features 2–7 and 9–10, focusing on word order, and features 22–34 and 37–40, focusing on nominal categories and nominal syntax) gives very similar groupings, with the exception of clustering Cape Verdean creoles on the same side with Asian creoles, as shown in Figure 11.2. This is probably due to a larger number of features that the Asian creoles share with the lexifier in this domain. When it comes to the divided clustering of Upper Guinea creoles, the results can correspondingly be interpreted as stronger adstrate influence on the continental Upper Guinea creoles Casamancese creole and Guinea-Bissau Kriyol.

![Figure 11.2](image-url)  

Figure 11.2 A splits graph of a Neighbor-Net analysis of 21 APiCS features of the nominal domain in 19 Iberian creoles.
11.7 Shared features

In addition to the Neighbor-Net analysis that reveals the areal clusters in our sample, we can conduct a more fine-grained analysis to reveal similarities and differences on the feature level in order to see which features are typical for each cluster. The features are mapped using parsimonious trees created in PAUP*, which are later analysed with Mesquite in order to detect cases of independent innovation or differences through the reconstruction of a given linguistic feature. Although we are not directly interested in the reconstruction of ancestral states, the tool is useful for detecting shared features in the established clusters.

The results of the analysis show shared features across different parts of grammar. We will go through them starting with features that are common to all the Iberian creoles in the sample, and thereafter we will present the features that are subject to some exceptions.

Table 11.7 presents the features that are shared by all the varieties in the sample. Iberian creoles present no dual pronouns, they have no syncretism in independent personal pronouns nor applicative constructions, and they mark instrumental expressions by adpositions.

Table 11.7 Shared features among 19 Iberian creoles.

<table>
<thead>
<tr>
<th>No. in APiCS</th>
<th>Shared features</th>
<th>Feature category</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>No dual in independent personal pronouns</td>
<td>Nominal categories</td>
</tr>
<tr>
<td>16</td>
<td>No person syncretism in independent personal pronouns</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>No antidual of paired body-part terms</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>Instrumental expressions by adpositions</td>
<td>Argument marking</td>
</tr>
<tr>
<td>91</td>
<td>No applicative constructions</td>
<td>Clausal syntax</td>
</tr>
</tbody>
</table>

These features are also commonly shared among creoles in general. For example, applicative constructions are extremely rare among creoles and can only be found in the African contact varieties and the non-creole Michif (Haspelmath et al. 2013). Antidual of paired body-part terms is a feature found only among the French creoles of the Indian Ocean, Haitian Creole, Nengee, and Creolese, and not pervasively in any of them (Michaelis & the APiCS Consortium 2013a).

Two of the shared features pertain to independent personal pronouns. Dual personal pronouns are atypical of creoles in general, and this feature is observed only in Pichi in the Atlantic area, and in a number of creoles spoken in the Pacific and Australia (Haspelmath & the APiCS Consortium 2013a). Here, an observation should also be made about grammaticalization processes that are in progress, which the data included in typological databases do not depict. For example, in Ternate Chabacano, the expression of duality is possible by adding the numeral
dos ‘two’ to the pronoun, as in mótru-dos ‘we.dual’, although the construction is not fully grammaticalized. Another shared feature from the area of personal pronouns is the lack of person syncretism, reflecting yet another general tendency in creoles. This feature is typical only in the Indian Ocean French creoles in the second and third person. It is also observed in some other varieties, such as Haitian Creole and Nengee in the first and second person (Haspelmath & the APiCS Consortium 2013b).

In the case of the instrumental expressions, the picture is less clear for the Iberian varieties due to different optional strategies that are common among the Atlantic creoles. Several creoles in the Caribbean and in West Africa use serial verbs to express instrumental relations. Actually, both adpositional marking and serial verb strategies are possible in the Gulf of Guinea creoles, with the exception of Fa d’Ambô, and in some Upper Guinea creoles, such as Casamancese Creole and the Santiago and Brava varieties of Cape Verdean Creole (Maurer & the APiCS Consortium 2013).

An even larger number of features is shared if exceptions are permitted. Table 11.8 lists the features for which only one variety in our sample presents an exception. Differences are generally not representative of only one variety, although Sri Lanka Portuguese stands out in several cases.

Table 11.8 Features shared across 18 of the 19 Iberian creoles.

<table>
<thead>
<tr>
<th>No. in APiCS</th>
<th>Shared features</th>
<th>Feature category</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Order of adposition and noun phrase</td>
<td>Word order</td>
</tr>
<tr>
<td>6</td>
<td>Order of cardinal numeral and noun</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Order of relative clause and noun</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Position of indefinite article in the noun phrase</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>No inclusive/exclusive distinction in independent personal pronouns</td>
<td>Nominal categories</td>
</tr>
<tr>
<td>36</td>
<td>No sortal numeral classifiers</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Comparative adjective marking</td>
<td>Nominal syntax</td>
</tr>
<tr>
<td>97</td>
<td>‘Want’ complement subjects left implicit</td>
<td>Complex sentences</td>
</tr>
<tr>
<td>99</td>
<td>Verb doubling not possible in temporal clauses</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>Savvy word exists</td>
<td>Lexicon</td>
</tr>
</tbody>
</table>

Sri Lanka Portuguese is distinct in not allowing the same strategies as the other Iberian creoles for five features, for which we find parallels in the adstrate languages. Generally, Iberian creoles all have prepositions, and in all of them the relative clause can be placed after the noun. In Sri Lanka Portuguese, however, the order features are exceptional and do not follow the general tendencies. The adpositions are postpositions, and the relative clause precedes the noun (Smith
Korlai Portuguese also uses postpositions to a considerable degree, but it allows prepositions as well, and has correlative clauses as the predominant relative clause type (Clements 2013). Also, contrary to the prevalent strategy among the Iberian creoles, Sri Lanka Portuguese does not mark the adjective in comparatives (Cardoso 2012: 110–111). This strategy is also possible in the Gulf of Guinea creoles, with the exception of Angolar, and in Ternate Chabacano. Among creoles in general, non-marking of the adjective in comparative constructions can be observed in African and Australian varieties (Michaelis & the APiCS Consortium 2013b). Another difference is that Sri Lanka Portuguese uses a desiderative particle in same-subject complement clauses of ‘want’, while most creoles use a verb in a construction where the complement subject is left implicit (Michaelis, Haspelmath & the APiCS Consortium 2013). Lastly, regarding numeral classifiers, which are not found in Iberian creoles, Sri Lanka Portuguese occasionally uses *pesaam* ‘person’ as a classifier for humans, reflecting a pattern found in the adstrate Tamil/Sinhala.

In addition, Zamboanga Chabacano differentiates between inclusive and exclusive personal pronouns, which is an adstrate feature. Principense presents postposed indefinite articles, which are pre-posed in other Iberian creole varieties. The Santiago variety of Cape Verdean Creole allows verb doubling constructions to be used to express temporal adverbial clauses. Lastly, Angolar does not have a word like *save* or *sabi* meaning ‘know’ (or something similar), deriving ultimately from forms of Portuguese or Spanish *saber*, which can be found in all other Iberian creoles.

The list of shared features that are found in the majority of the Iberian creoles is even longer, but variation comes increasingly into play when looking at other features. Table 11.9 presents the features that are shared by most of the varieties, but which nevertheless present two or three exceptions. Clear cases of exceptions are the following: Fa d’Ambô, Guinea-Bissau Kriyol, and Casamancese Creole have an indefinite article distinct from the numeral ‘one’ (feature 29); Principense and Santome verbal particles follow a Mood-Tense-Aspect order, but this feature does not apply for other Iberian creoles (feature 44); and Fa d’Ambô and Papiamentu have a neutral alignment in the case marking of personal pronouns (feature 59), whereas the rest of the Iberian creoles has an accusative alignment.

Many creoles permit several strategies for expressing a function, which makes the picture somewhat more complicated. For example, Fa d’Ambô is the only creole in the sample that only allows degree words following adjectives (feature 8), but this strategy is also found as a parallel to the shared pattern found in several other creoles in different geographical areas. These include the other Gulf of Guinea creoles, Casamancese Creole, Guinea-Bissau Kriyol, Cape Verdean Creole of Santiago, Korlai, Papiá Kristang, and Palenquero. On the other hand, Fa d’Ambô and Angolar do not mark the possessor noun phrase in an adnominal possessive construction (feature 38). Other varieties, such as Principense, Palenquero, and
Table 11.9 Features that are shared by most of the Iberian creoles.

<table>
<thead>
<tr>
<th>No.</th>
<th>Shared features</th>
<th>Feature category</th>
<th>Exceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Order of degree word and adjective</td>
<td>Word order</td>
<td>Fa d’Ambô</td>
</tr>
<tr>
<td>12</td>
<td>Position of interrogative phrases in content questions: initial</td>
<td></td>
<td>Diu, Korkai, Sri Lanka</td>
</tr>
<tr>
<td>20</td>
<td>Pronoun conjunction: singular pronoun overtly conjoined with other conjunct</td>
<td>Nominal categories</td>
<td>Angolan, Chabacano</td>
</tr>
<tr>
<td>29</td>
<td>Indefinite articles identical to ‘one’</td>
<td></td>
<td>Fa d’Ambô, Guinea-Bissau, Kriyol, Casamancese Creole</td>
</tr>
<tr>
<td>38</td>
<td>Adpositional or case marking of possessor</td>
<td>Nominal syntax</td>
<td>Fa d’Ambô, Angolan</td>
</tr>
<tr>
<td>44</td>
<td>Order of tense, aspect, and mood markers does not apply</td>
<td>Verbal categories</td>
<td>Santome, Principense</td>
</tr>
<tr>
<td>59</td>
<td>Accusative alignment of case marking of personal pronouns</td>
<td>Argument marking</td>
<td>Fa d’Ambô, Papiamentu</td>
</tr>
<tr>
<td>67</td>
<td>Experiencer in subject position in constructions with ‘like’</td>
<td></td>
<td>Cape Verdean Creole of São Vicente, Casamancese Creole, Korkai, Sri Lanka</td>
</tr>
<tr>
<td>81</td>
<td>Motion-to and motion-from differentiated</td>
<td>Clausal syntax</td>
<td>Cavite and Zamboanga, Chabacano</td>
</tr>
</tbody>
</table>

Cape Verdean Creole of Santiago, have this strategy as an option as well, but the creoles that do not belong to the Gulf of Guinea cluster tend to use adpositions or case to mark the possessor. The latter strategy is typical for the rest of the Iberian creoles. Similarly, although interrogative phrases in content questions are generally initial among Iberian creoles (feature 12), the Indo-Portuguese varieties present a clear exception, and a number of other Portuguese-lexifier creoles also permit the placement elsewhere.

Generally, Iberian creoles overtly conjoin singular personal pronouns and full noun phrases, as in ‘Mary and I’, or ‘you and your brother’ (feature 20). Only Angolan does not permit this strategy, and variation is also found in Cavite and Ternate Chabacano. Differentiation between motion-to and motion-from (feature 81), on the other hand, presents a clear case of variation that is reflected in the Chabacano lects described in APiCS, where the values of identity (with na), overlap (na, para na), and differentiation (na, di) are assigned to different varieties.

An overview of the shared features and their comparison with areal clusters and lexifier features reveals that most of the shared varieties are also shared with the Portuguese and Spanish lexifiers. A clearly distinct feature that differentiates the creole varieties from the lexifiers, but presents a lot of variation among the creoles,
is the position of TMA marking (feature 43). It is a well-known fact that creoles are set apart from their Spanish and Portuguese lexifiers due to the differences in the verbal system. However, this feature is also a nearly universal creole trait.

11.8 Areal differences

The features that define the areal clusters in Figure 11.1 can be identified in Mesquite. The same areal clusters can also be identified in the PAUP analysis that creates the parsimonious tree in Figure 11.3. The lexifiers stay apart from the creoles, and the main geographical division between Atlantic and Asian is clear.

![Figure 11.3 Unrooted Maximum Parsimony tree of the Iberian creoles.](image_url)

However, only two clusters get sufficient statistical support: the Gulf of Guinea creoles and the Chabacano varieties of the Philippines. Tables 11.10 and 11.11 list the features (with their corresponding APiCS feature number and the assigned values for these varieties) that appear only in these clusters and not in any other Iberian creoles in our sample.
Chapter 11. Similarities and differences among Iberian creoles

Table 11.10 Gulf of Guinea features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>No marking of possessor noun phrases</td>
</tr>
<tr>
<td>80</td>
<td>Serial verb constructions to express 'motion-from'</td>
</tr>
<tr>
<td>82</td>
<td>Serial verb construction (without a preposition) to express the transitive motion construction with 'push'</td>
</tr>
<tr>
<td>83</td>
<td>Serial verb construction (without a preposition) to express the transitive motion construction with 'pull'</td>
</tr>
<tr>
<td>100</td>
<td>Bipartite negative marker</td>
</tr>
<tr>
<td>101</td>
<td>Bipartite negative marker, before verb and after object</td>
</tr>
<tr>
<td>107</td>
<td>Optional vocative marker following noun</td>
</tr>
</tbody>
</table>

The use of serial verb constructions to express motion is characteristic of the Gulf of Guinea cluster, and it appears only very occasionally in other contact languages of the area that have a non-Iberian lexifier, e.g. Ghanaian Pidgin English, Nigerian Pidgin, and Sango. No marking of possessor noun phrases is more common in the Atlantic creoles, whereas optional vocative markers following the noun seems to be an areal feature. Bipartite negative marking is also possible in Palenquero, although it is not the only strategy available. Hagemeijer (2011: 140) connects these features and some others to the Western Bantu (Kikongo) and Niger Delta adstrates.

Table 11.11 Chabacano features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VSO</td>
</tr>
<tr>
<td>61</td>
<td>Verb-initial theme – recipient</td>
</tr>
<tr>
<td>102</td>
<td>Negation and indefinite pronouns</td>
</tr>
</tbody>
</table>

The features that are characteristic of the Chabacano creoles are related to verb-initial word order, which is a Philippine feature. Similarly, indefinite pronouns are formed by an existential construction, which is the most common strategy in the negative paradigm. This strategy also reflects adstrate structures (see also Sippola 2012 for a comparison of the variation in the Ibero-Asian creoles).

A look at the other features also shows some that are shared within areal groups. The Southeast Asian creoles, Chabacano, Papiá Kristang, and Batavia creole, share purely aspectual TMA systems (feature 49) and give-serials, where ‘give’ appears in first position expressing the recipient only (feature 86). The Indo-Portuguese varieties share postposed adpositions in NPs (feature 4), non-initial interrogative phrases (feature 12), and the possibility of using locational constructions to express predicative possession (feature 77). In continental Upper Guinea, the same marker is shared for the progressive and the future (feature 47).
11.9 Discussion

The results of our analysis confirm several areal clusters that were identified in previous research, such as the Cape Verdean varieties, the Gulf of Guinea creoles, Indo-Portuguese, Malayo-Portuguese, and the Chabacano varieties. Other Upper Guinea creoles appear more independently. That is also the case for the Caribbean creoles Papiamentu and Palenquero. This can be expected, as they are the only ones in South America and the Caribbean. Regarding the method, it is interesting to see that both the Neighbor-Net and the maximum parsimony analysis produced very similar results. When comparing these results with earlier classifications, we notice that the same lines of groupings have been established with the help of individual features or sub-selections of them in previous research. When comparing, for example, Ferraz’s (1987) features in Table 11.1 to the APiCS data, it is clear that the features partially match. Some features from Ferraz’s analysis, such as the diminutives, the particle of obligation, the unmarked personal pronoun for the Atlantic creoles, the –mente adverbials, and the forms of the TMA and object markers for Asian creoles are missing from the APiCS data, in addition to the lexical information. However, the features in APiCS include a much wider range of structural and functional information. Clements’ (2009) affixal TMA marker and grammatical features are also present in the APiCS data, but they did not appear as decisive for individual groupings.

The shared core of the Iberian creoles is a combination of typical creole traits and lexifier influence, as the naming of this group suggests. The features are to a great extent shared with the lexifiers. However, previous features, mostly based on the form of lexical and functional items, were not included in the APiCS sample (for example, the features in Holm 2009). Despite this, the number of shared features is considerable at 26 out of 130, when two to three exceptions are permitted. It will be interesting to see how different lexifier groups behave in this respect. When the shared features are divided according to their categories, word order and nominal categories are found to be salient, as can be seen in Table 11.11. As the original division into different categories of grammar was not even, we did further testing with regression analysis to see if the overall number of features in each category predicts the number of shared features in a meaningful manner. This was not the case, as a 1% change in the original number of features in a category leads to a 1.1% increase in the shared features in the same category. However, this might be due to the fact that we have a small number of data points.

In a recent phylogenetic study on the typological status of creole languages, Bakker et al. (2011) used 97 features selected in Comparative Creole Syntax (Holm & Patrick 2007) to compare creoles from different lexifier and areal groups, including seven Iberian creoles. Their results (Bakker et al. 2011:29) did not provide
Chapter 11. Similarities and differences among Iberian creoles

Evidence for claims about the decisiveness of lexifier or substrate influence for the groupings of the creoles. Although there was some partial clustering according to the lexifier among the seven Iberian creoles compared, when contrasted with creoles with a different lexifier, the clustering was not clear. The only Iberian varieties that were clearly connected in their network analysis were Cape Verdean and Guinea-Bissau Kriyol. In our data, the clustering of Cape Verdean and Guinea-Bissau Kriyol is not as evident, as seen in Figure 11.1. While their study was the first one to apply phylogenetic methods to a creole sample including Iberian varieties, it did not identify the specific features that characterize different clusters. In the section focusing on the differences, we have identified areal features which are also found in the substrate and adstrate languages of these creoles, showing that these do play a role.

Another point to keep in mind is that the sample as a whole has not gone through an assessment of areality in the typological sense (Bisang 2010: 419). This means that the results shed light on the typical features of this subgroup, which might or might not correspond to those of the creoles (see Bakker & Daval-Markussen this volume, chapter 6). Geography or areality is here understood as “individual sets of linguistic features which are pervasively observable in a particular world region or a part thereof against all others” (Kortmann 2013: 166). It is clear, however, that several features in the proposed areal clusters are not independent of one another, as shown for example by the fact that several features were about the use of serial verbs in motion constructions in the Gulf of Guinea cluster, and about the verb-initial order in the Philippines.

<table>
<thead>
<tr>
<th>Category</th>
<th>Shared features (max. 3 dev.) / all</th>
<th>% of shared features</th>
<th>Shared in the feature category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal categories</td>
<td>7 / 24</td>
<td>26.9%</td>
<td>29.2%</td>
</tr>
<tr>
<td>Clausal syntax</td>
<td>2 / 19</td>
<td>7.7%</td>
<td>10.6%</td>
</tr>
<tr>
<td>Argument marking</td>
<td>3 / 16</td>
<td>11.5%</td>
<td>18.8%</td>
</tr>
<tr>
<td>Verbal categories</td>
<td>2 / 14</td>
<td>7.7%</td>
<td>14.3%</td>
</tr>
<tr>
<td>Phonology</td>
<td>0 / 13</td>
<td>0.0%</td>
<td>0%</td>
</tr>
<tr>
<td>Word order</td>
<td>6 / 12</td>
<td>23.1%</td>
<td>50%</td>
</tr>
<tr>
<td>Lexicon</td>
<td>1 / 11</td>
<td>3.8%</td>
<td>9.1%</td>
</tr>
<tr>
<td>Complex sentences</td>
<td>2 / 8</td>
<td>7.7%</td>
<td>25.0%</td>
</tr>
<tr>
<td>Negation, questions, focusing</td>
<td>1 / 7</td>
<td>3.8%</td>
<td>14.3%</td>
</tr>
<tr>
<td>Nominal syntax</td>
<td>2 / 6</td>
<td>7.7%</td>
<td>33.3%</td>
</tr>
<tr>
<td>Total</td>
<td>26 / 130</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>
11.10 Conclusions

Computational analysis of typological data supports previous classifications of Iberian creoles that were established in early creole studies and refined in subsequent comparative work. The results show a clear division into Atlantic and Asian creoles, and the main subgroups were divided geographically. However, connections between the subgroups are less clear due to the degree of detail of the typological feature values, as clusters reflect mainly areal traits in these languages. The synchronic data used in this study thus provides no support for theories about the connections between Papiamentu and the Upper Guinea creoles, or the interrelations between the Chabacano varieties and the Portuguese creoles of Asia.

A mapping of the features responsible for this classification confirms that many similarities among the Iberian creoles are shared with the lexifiers. They also present universal tendencies that are further conditioned by the individual contact situations. Differences and characteristic traits of the Gulf of Guinea and Chabacano creoles, as well as Indo-Portuguese and the continental Upper Guinea creoles, are clearly defined by the adstrate and substrate languages, and thus lead to the areal patterning.

Notes

The datasets for this chapter can be found here: https://phylogenetic-creole-studies.blogspot.com

References


Whinnom, K. 1956. *Spanish Contact Vernaculars in the Philippine Islands*. Hong Kong: Hong Kong University Press.

Afro-Hispanic varieties in comparison
New light from phylogeny

Danae M. Perez, Sandro Sessarego and Eeva Sippola
University of Zürich and University of Bremen / University of Texas at Austin and Freiburg Institute for Advanced Studies / University of Bremen

This study examines the potential of phylogenetic analysis in the classification and comparison of Spanish contact varieties, with special focus on three Afro-Hispanic varieties spoken in South America. Our analysis is based on typological and dialectal comparative data. The results of the phylogenetic analysis show clear clusters of standard varieties and creole varieties, while the placement of other contact varieties in the network has to be interpreted in the light of the data collection methods and socio-historical information about the varieties.

Keywords: Afro-Hispanic, classifications, Spanish-based creoles, Spanish varieties

12.1 Introduction

This chapter explores the use of phylogenetic analysis in the comparison and classification of different Spanish-related varieties, with special focus on three Afro-Hispanic dialects. These are dialects or contact varieties spoken by communities of African descent in South America that exhibit a particularly high degree of variation on a continuum between rather standard-like and more vernacular features. This variation, added to the socio-historical evidence of intensive language contact, gives rise to discussions on how they should be classified (cf. Lipski 2005) and on whether or not certain Afro-Hispanic varieties stem from a creole ancestor.

This study analyzes both the phonological and morphosyntactic traits of a number of Afro-Hispanic varieties and Spanish-based creoles, as well as other varieties of Spanish spoken on four different continents. The data come from fieldwork corpora (collected by Perez and Sessarego), typological databases (Dryer & Haspelmath 2013; Michaelis et al. 2013), and published literature (Lipski 1985a, 1985b, 1987a, 1988, 1994a, 2007a, 2007b, 2008; Penny 2004; Quilis & Casado-Fresnillo 2008). The analysis is based on statistical modelling and computational
tools for quantitative typology that are suitable for studying language contact situations (Daval-Markussen & Bakker 2011, 2012; Huson & Bryant 2006).

One of the aims of the study is to contribute to the classification of different African-descendant varieties of Spanish. In particular, we will compare Afro-Yungueño Spanish (AY) (Bolivia), Afro-Choteño Spanish (AC) (Ecuador), and Afro-Chinchano Spanish (ACh) (Peru), as well as a selection of regional dialects of Spanish and Spanish-based creoles. This analysis will then allow us to graphically picture the distribution of these varieties in a phylogenetic network, which has been built according to the features we selected, and to identify some traits that characterize different varieties.

A second aim of the study is to evaluate the implications of data selection and feature value assignment when studying lesser-documented varieties for which no large-scale variationist analyses or extensive spoken or written corpora exist. We use different sets of data of Afro-Yungueño to examine to what extent they determine the final picture of the phylogenetic analysis. In doing so, we will discuss some of the strengths and limitations of the phylogenetic method in the context of Afro-Hispanic varieties. In Section 12.2, we provide an overview of the Afro-Hispanic varieties included in the comparison and outline the current state of research. Section 12.3 explains the methodology and the feature selection used in the comparison. Sections 12.4 to 12.6 cover the analysis, the results, and the discussion. Finally, conclusions follow in Section 12.7.

12.2 The Afro-Hispanic varieties

The colonial expansion of the Spanish language produced a large number of individual settings in which different dialects of Spanish came into contact with each other as well as with a wide range of other languages. The varieties that emerged from these settings are generally categorized on the basis of two different approaches. They are either classified as national or regional varieties, such as Colombian, River Plate, or Caribbean Spanish, or as typologically similar varieties that have certain dialectal features in common. Among these features are, for example, the use of vos instead of tú as the second person singular pronoun, or the maintenance or merger of the distinction between /j/ and /ʎ/ and the velarization of the final /n/ (Lipski 1994a). In cases of extreme language contact, new languages, i.e., Spanish-based creoles

---

1. Due to the wide range of definitions of the term creole (for different definitions and characteristic traits, see Bickerton 1981; Seuren & Wekker 1986; McWhorter 1998; Bakker et al. 2011; Mufwene 1997; DeGraff 2005; Thomason & Kaufman 1988; Romaine 1988) and the need for adopting a somewhat stable basis of comparison between language varieties, we will put aside terminological debates and for convenience use the terms creole and dialect broadly as two varieties that diverge from the lexifier or standard variety to either a greater or a lesser extent, respectively (cf. also McWhorter 2000: 10).
emerged: Palenquero (Colombia), Papiamentu (Dutch Antilles) and Chabacano (Philippines). These varieties are generally distinguished by their lack of noun–adjective and noun–verb agreement and the fact that they also present restructured verbal systems that make use of preverbal TMA markers. These features show a considerable departure from the lexifier and justify the classification of these varieties as creoles (Michaelis et al. 2013; Holm 1988–1989).

Even if we adopt a view that classifies Spanish varieties as either Spanish-lexified creoles or as dialects of Spanish, certain contact vernaculars may still require intermediate categories. We know from the Caribbean setting, for instance, that in some cases it is difficult to reach a consensus about where to draw the line between a dialect and a creole (e.g. Holm 2004; on the dialect vs. creole debate, see also Levisen et al. this volume, Chapter 15). These challenges are particularly pronounced in those varieties of Spanish that display a high degree of variation as vernacular features coexist with more acrolectal variants. In addition, many peripheral contact varieties of Spanish spoken by marginalized communities are still underresearched, and there is a lack of data that would be necessary to document their inventories of structural and lexical features (cf. Lipski 2005: 8).

Afro-Hispanic varieties make a case in point. Lipski (2005) compared several Afro-Hispanic varieties and observed that many of them, in particular the Caribbean varieties, experienced parallel restructuring, which could be the result of the extended contact between Caribbean Spanish and certain creole languages spoken in the region. Sessarego (2013a, 2013b), however, has argued that the majority of the Afro-Hispanic varieties of the Americas that deviate systematically from other native varieties of Spanish may be seen as the result of first-language acquisition (nativization) of advanced second-language grammars, without necessarily following the prototypical creole life cycle of pidginization preceding creolization, with eventual decréolization (cf. Mühlhäusler 1997).

One of these varieties is Afro-Yungueño Spanish (AY), a contact variety spoken by a small community of African descent in the valleys of Los Yungas in Bolivia. While the total number of Afro-Bolivians amounts to almost 20,000, the most basilectal variety of AY is retained among approximately 200 speakers in the villages of Mururata, Chijchipa, and Tocaña in the Nor Yungas Province, 120 km from La Paz. AY is surrounded by Andean Spanish and Aymara, and its lexicon is principally derived from these two languages. AY was not described until the late 2000s (Lipski 2007a, 2007b, 2008), and even though a considerable body of

---

2. Some doubts have been raised about the Hispanic origins of Papiamentu and Palenquero, and several scholars have suggested that they should be better analyzed as Portuguese-based creoles that subsequently underwent a process of Spanish relexification (cf. Goodman 1987; Martinus 1989; Schwegler 1993, 1999, 2014; McWhorter 2000; Jacobs 2009). Nevertheless, for the sake of simplicity, we will not go into the details of these debates, and we will classify them as Spanish creoles.
literature on AY has now been produced, no unanimous classification of its varietal status has been reached. On the one hand, Lipski (2008) and Perez (2015) have both argued that AY is likely to stem from a creole; Lipski (2008: 186) concluded that AY originates from a 16th century Afro-Hispanic pidgin, while Perez (2015) proposes that an 18th century Portuguese-based contact variety is likely to have been involved in its formation. On the other hand, Sessarego (2013b: 364) classified AY as a dialect of Spanish that resulted from “untutored second language acquisition processes, which could crystallize and survive in Los Yungas”; in other words, he argued that AY does not stem from a creole. The data presented in these debates so far has mainly focused on patterns of morphosyntactic simplification, such as the lack of gender and number agreement, as in *lu casa viejo* [pl. house old] vs. Sp. *las casas viejas* [ART-F.pl. house-PL old-F.pl] ‘the old houses’, or the reduced verbal paradigm based on the third person singular, as in *nohotro yora* [1pl cry] vs. Sp. (*nosotros*) *lloramos* [(1pl cry-prs.1pl) ‘we cry’]. In light of this disagreement on the varietal status of AY, new data is needed to further investigate the issue.

Sessarego (2013a, 2015) further contributed to these debates by adding data from two dialects spoken by people of African descent in Coastal Peru and Highland Ecuador. One of them is *Afro-Choteño* Spanish (AC), an Afro-Hispanic vernacular spoken in several rural villages scattered across the provinces of Imbabura and Carchi, Ecuador. This community consists of approximately 12,000 people, who are the descendants of the slaves taken to this region to work on the Jesuit sugarcane plantations during colonial times. Nevertheless, only some 1,000 elderly Afro-Choteños still speak the traditional vernacular, while younger members of the community tend to speak dialects that approximate the local standard, Highland Ecuadorian Spanish. Several articles by Lipski (1987b, 2009) and Schwegler (1999, 2014) proposed an analysis of the main grammatical features found in AC to shed light on its origin. In particular, Schwegler (1999, 2014) ascribed potential Afro-Portuguese creole roots to AC. He claimed that the existence of the third person pronoun *ele* (from Port. *ele* ‘he’) in this variety suggests that the slaves who entered Chota Valley in colonial times could have spoken a creole-like Afro-Portuguese contact variety (in partial support of the Monogenetic Hypothesis, cf. Granda 1968, 1970). Lipski (2009) provided a different account of the presence of *ele* in AC by seeing this element as the result of a paragogic process of final – *e* insertion affecting several items across the AC lexicon (e.g., *mujere* from Sp. *mujer* ‘woman’, *ayere* from Sp. *ayer* ‘yesterday’, etc.). In a recent study, based on

3. The latter claim is in line with McWhorter’s (2000) hypothesis that the Spanish creoles in the Americas descend from a Portuguese-based ancestor. Given its relatively recent inclusion into the discussions, however, AY was not considered by McWhorter, who simply mentions the absence of a creole in Bolivia as the result of unfavorable socio-historical conditions (McWhorter 2000: 34).
field research carried out in the Afro-Choteño villages. Sessarego (2013a, 2014b) supported Lipski’s account; moreover, he provided an analysis of the available socio-historical evidence for AC, which indicates a low possibility of stable creole formation or the introduction of a creole in the region.

Afro-Chinchano Spanish (ACh) is an Afro-Hispanic vernacular spoken in rural areas of the province of Chincha, Department of Ica, Peru, more precisely in the villages of San Regis, San José, El Guayabo, and El Carmen. This dialect is spoken by a few hundred people, mainly elderly Chinchanos who descend from the slaves taken to this region during the 17th century to work on Jesuit sugarcane plantations. Romero (1987, 1988, 1994) and Lipski (1994b, 2005) described the main features of Afro-Peruvian grammar by analyzing literary works, theatrical texts, and traditional songs from the 17th–20th centuries that reported Peruvian bozal speech. In addition, Cuba (2002) offered the first synchronic account of Afro-Peruvian Spanish by investigating the speech of the Afro-Chinchano communities mentioned above. More recently, Sessarego (2014a, 2015) has provided a new account of ACh grammar where no creole features are reported. In addition, he has shown socio-historical evidence for the Afro-Chinchano region that does not support an earlier creole phase for the variety.

Map 12.1 Afro-Hispanic varieties included in the study.

4. Tumbabiro, Carpuela, Chota, Santiago, Chalguayacu, Chamanal, Concepción, Caldera, and Cuajara.
Based on these socio-historical data, and on the overall grammatical similarities between AC, ACh, and standard Spanish, Sessarego has concluded that these Afro-Hispanic vernaculars, along with AY (cf. Sessarego 2011, 2013b, 2014c), may be seen as the result of the nativization of advanced second-languages, rather than as former creoles (cf. Sessarego 2013c).

In brief, we have seen that the debates about the status and possible earlier creole stages of Afro-Hispanic varieties in South America are still ongoing. The main aim of this chapter is to add new input to these debates.

12.3 Sample and methodology

The analysis focuses on AY, AC, and ACh, which are compared with a number of varieties of Spanish and Spanish-based creoles. The Afro-Hispanic varieties were selected according to the availability of contemporary data collected during fieldwork (Lipski 2007a, 2007b, 2008; Sessarego 2011, 2013a, 2015; Perez unpublished field data). Data for the Spanish creole varieties were extracted from the APiCS database (Michaelis et al. 2013) and include Palenquero and Papiamentu in Latin America and Cavite Chabacano, Ternate Chabacano, and Zamboanga Chabacano from the Philippines. The Spanish dialectal data were compiled from secondary sources (mainly Lipski 1985a, 1985b, 1987a, 1988, 1994a; Penny 2004; Quilis & Casado-Fresnillo 2008). These varieties can be divided into two main types: standard varieties, such as Argentinian, Bolivian, Cuban, Peninsular, and Mexican Spanish, and varieties influenced by second-language speaker effects, such as those spoken in the Andean Highlands, Equatorial Guinea, the Philippines, and the Sabine River region (United States). Their principal adstrate languages are Quechua/Aymara for Highland Andean Spanish, Bantu languages for Spanish in Equatorial Guinea, Tagalog for Philippine Spanish, and English for Sabine River Spanish.5 In addition, data points have been filled in for Judeo-Spanish,6 a variety spoken by Sephardic communities. These varieties are similar to the Afro-Hispanic varieties in that their typological profile has been claimed to show influence from other non-Romance languages, or to show influence from processes of second-language acquisition, language attrition, and language shift. We decided to include these varieties in the sample in order to provide a wider base of comparison and contrast for the Afro-Hispanic varieties.

5. Sabine River Spanish was also in contact with French at an earlier stage (Lipski 1988:6).
6. The data cannot be taken to be characteristic of one specific variety, but rather of an imagined standard described in Penny (2004).
Our sample is based on a combination of features from three sources: general typological features listed in the *World Atlas of Language Structures* (WALS, Dryer & Haspelmath 2013), typical contact language features from the *Atlas of Pidgin and Creole Structures* (APiCS, Michaelis et al. 2013), and features typical of Spanish dialectal variation (Lipski 1994a). The main feature set for this study is based on the 48 shared APiCS and WALS features. This is the most comprehensive typological database that includes both contact varieties (in APiCS) and the Spanish lexifier (WALS), and thus it offers a solid typological feature set designed especially for contact languages (see also Bakker et al. 2013 for similar typological datasets). The 48 structural features are mainly drawn from morphology and syntax (see Table 12.1).

In addition to the APiCS/WALS features, we designed a set of 24 features that commonly vary across Spanish dialects: at the phonological level, we selected the aspiration of /s/, the neutralization of phonemic distinctions, such as *yeísmo* (/ʎ/ > /j/), and alterations in unstressed vowels (e.g. /e/ > [i]; /o/ > [u]), among others (see Appendix 12.1 for the complete list of features). At the morphosyntactic level, we selected dialectal features that are typical of second-language and learner varieties of Spanish, such as variable subject–verb and noun–adjective agreement, the use of disjunctive object pronouns instead of clitics, the variable use of the copulas *ser* and *estar*, and non-standard uses of common prepositions (Lipski 2007a: 176).

The analyzed feature sets and data entries are listed in the appendix. The division between the APiCS feature categories in our sample is presented in Table 12.1. In our feature set, the lexicon is underrepresented compared to other areas of grammar due to the unavailability of comparable data. Phonology and nominal categories have the highest numbers of features.

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonology</td>
<td>16</td>
</tr>
<tr>
<td>Nominal categories</td>
<td>15</td>
</tr>
<tr>
<td>Word order</td>
<td>9</td>
</tr>
<tr>
<td>Clausal syntax</td>
<td>7</td>
</tr>
<tr>
<td>Argument marking</td>
<td>7</td>
</tr>
<tr>
<td>Nominal syntax</td>
<td>6</td>
</tr>
<tr>
<td>Verbal categories</td>
<td>5</td>
</tr>
<tr>
<td>Negation, questions, focusing</td>
<td>3</td>
</tr>
<tr>
<td>Complex sentences</td>
<td>2</td>
</tr>
<tr>
<td>Lexicon</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>72</td>
</tr>
</tbody>
</table>

Table 12.1 Feature categories in the sample.
Feature values were manually coded into a matrix. The most common value range for a feature was binary (50 features, e.g. “Is there inclusive/exclusive distinction in independent personal pronouns?”), followed by features that originally had multiple values, but resulted in one shared value for all the varieties (for shared features for Iberian creoles in APiCS see also Sippola this volume, Chapter 11). Only four features had multistate values with four value options. Testing in which these multivalue features were transformed into binary features did not provide significantly different results.

The two principal AY datasets were individually and independently prepared by Perez and Sessarego and are based on language samples collected during fieldwork. Nevertheless, it is important to point out that their methodologies and research focuses varied. Sessarego’s main focus was on phonological and morphosyntactic structures (2014c), and he conducted the interviews himself and carried out follow-up conversations with the informants in order to complement the data by eliciting specific structural features. Perez, by contrast, aimed at documenting AY by working with community members who made the recordings themselves. In addition, she collected field notes during participant observation. The third fieldwork corpus on AY was taken from the data provided by Lipski’s (2008) book, which includes data from interviews on oral history provided by an Afro-Bolivian historian. The datasets on Afro-Choteño and Afro-Chinchano Spanish were collected and prepared by Sessarego with similar goals and methods as for his AY corpus (2014a, 2014b).

The creole datasets in APiCS were prepared by experts in each language, and the feature values are often based on fieldwork data. Finally, the Spanish standard and dialectal varieties included in the comparison are of a rather different data type. Dialect atlases containing comparative language data from the Spanish-speaking world are rare or only partially compiled, and the features focus mostly on the lexicon and phonology (see, for example, Prieto et al. 2010–2014). We therefore decided to rely on secondary sources, collecting information from published articles and books. To evaluate the validity of the data from lesser-studied varieties, they were complemented by our own observations and fieldwork data from the Philippines and South America.

Distance-based methods of phylogenetic analysis have been employed in the study of contact varieties (Daval-Markussen & Bakker 2011, 2012; other chapters in this volume), although conflicting views on the use of such tools have also been presented (DeGraff 2012; DeGraff et al. 2013). Here the software SplitsTree4 (Huson & Bryant 2006) is used to provide an overview of the classification of the varieties in phylogenetic networks using Neighbor-Net (Bryant & Moulton 2004). As explained above, the primary typological feature values are first converted
into a matrix. Then the software is used to calculate the difference between each language in the sample by means of combining all the differences and similarities between the languages into a single distance measure. The resulting unrooted network is based on language similarities and differences and does not embody language evolution as such (Wichmann et al. 2011: 207). In this case, the internal nodes of the graph represent the conflict between the different splits in the data analysis. Contradictory groupings of taxa produce a conflict of the signal and a high amount of webbing in parts of the network. This method therefore provides an overall picture of the similarities and differences between the varieties studied, and thus it is a useful tool for evaluating the typological status of languages that have not yet been conclusively classified.

Typological features are generally structural characters that have few possible values, which leads to high rates of chance resemblance. In addition, their values are often linked through dependencies and implicational universals, and many values have a functional, external motivation (Croft 2008: 230). This is one reason why comparative linguistics with a diachronic approach has mainly focused on sound changes and lexical differences. However, typological features nevertheless offer a useful basis of comparison along predefined feature values and are therefore most suitable for synchronic clustering/similarity measurements (Haspelmath 2010), such as in the ones used in Sippola (this volume, Chapter 11) and the ones we are aiming for in our study.

12.4 Results of the phylogenetic network analysis

The first analysis of the 48 shared APiCS/WALS features is based on a balanced sample of Afro-Hispanic varieties, standard varieties of Spanish, and other dialects, as shown in the following graph made in SplitsTree (Figure 12.1). The data used for Afro-Yungueño in the network is from Sessarego (Afro-Yungueño S). The three creole varieties are located on one end of the network with long independent lines, which reflect the particular language contact scenarios, i.e. the geographical distance and typological divergence, of these varieties vis-à-vis the others. All the other varieties are on the opposite end of the network, and interestingly, the Afro-Hispanic varieties are located separately close to the part of the network where standard varieties appear.

In a nearly identical analysis, for which Perez’s data on Afro-Yungueño (Afro-Yungueño P) were used instead of Sessarego’s, we obtain the network in Figure 12.2. Here Afro-Yungueño has moved closer to the creole cluster of the network. It is close to Palenquero in particular and is clearly separate from Chota
Let us now increase the number of features in the sample by adding the 24 Spanish dialectal features into the network analysis. We get 12.3a with Sessarego’s data and 12.3b with Perez’s data. We can conclude that the dialectal features do not significantly affect the clustering of the Afro-Yungueño datasets used in this study. Nevertheless, it is evident that the two datasets place Afro-Yungueño again in two nearby, but somewhat different, positions of the network. According to Sessarego’s data in Figure 12.3a, Afro-Yungueño is closer to other Africanized varieties that are not generally held to have creolized, but rather experienced the effects of second-language acquisition. Perez’s data in Figure 12.3b, by contrast, indicate that Afro-Yungueño is closer to the Spanish-derived creoles than to most Afro-Choteño S
Afro-Chinchano S
Afro-Yungueño S
Spanish in Equatorial Guinea, Sabine River Spanish
Philippine Spanish
Cuban Spanish
Peninsular Spanish
Mexican Spanish
Zamboanga Chabacano
Palenquero
Papiamentu

Figure 12.1 Afro-Hispanic varieties (dataset from Sessarego), standard varieties of Spanish, other Spanish dialects, and creoles in a 48-feature Neighbor-Net.
other varieties. Furthermore, the longer line in 12.3b suggests that, on the basis of Perez’s data, Afro-Yungueño shows more unique traits and is typologically more distant from other varieties of Spanish than it is according to Sessarego’s data.

By increasing the number of varieties to 21, including all three different data-sets for Afro-Yungueño (data from Lipski 2008, Perez, and Sessarego), and running an analysis of the 72 features, we obtain Figure 12.4. Languages that were classified as standard varieties of European and Latin American Spanish appear on the bottom side of the network in a clear cluster that shows short independent branches for each variety. In other words, they are relatively similar to one another. On the upper side of the network, we find the creole varieties. However, as in the previous networks, these are not as tightly clustered together as the standard varieties of Spanish since considerable structural differences prevail. Between these two ends, we observe both Afro-Hispanic varieties and other varieties that underwent deeper restructuring due to the effects of second-language speakers, i.e. the varieties spoken in Equatorial Guinea and the Philippines, as well as Andean Spanish with

Figure 12.2 Afro-Hispanic varieties (datasets from Perez on Afro-Yungueño and Sessarego on Afro-Chinchano and Afro-Choteño) compared with standard varieties of Spanish, other Spanish dialects, and creoles in a 48-feature Neighbor-Net.
its significant adstrate influence. In other words, we find a gradual classification with creoles located on one end of the network, standard varieties located on the other end, and contact-influenced varieties located in different positions between these two groups. This corroborates claims that Africanized and other contact-influenced varieties have undergone certain degrees of restructuring, yet their typological status appears to still diverge significantly from the three Spanish creoles considered here. Holm (1992, 2004), for example, reached similar conclusions and proposed the terms *semicreole* and *partially restructured language variety* for this type of contact variety.

**Figure 12.3a** Sessarego’s data on Afro-Hispanic varieties compared with standard varieties of Spanish, other Spanish dialects, and creoles in a 72-feature Neighbor-Net.
With regard to Afro-Yungueño, it is striking that its position in the network in Figure 12.4 reflects the different current claims about its varietal status. Sessarego’s dataset is again farther away from the Spanish-based creoles, which is in line with his claim that Afro-Yungueño should not be considered a creole. Perez’s dataset, conversely, places Afro-Yungueño typologically closer to the other creole varieties, which supports her claim of Afro-Yungueño being a creolized variety, or a descendant of one. Similarly, the dataset provided by Lipski (2008) is in close proximity to Perez’s dataset, as is his claim that Afro-Yungueño is typologically “the most radically restructured variety of Spanish spoken natively, not only in the contemporary Spanish-speaking world, but in all of the known history of the Spanish language” (Lipski 2008: 186). The positions of the three different datasets in the network thus represent the individual claims advanced by each researcher, which ultimately also underscores the solidity of the conclusions drawn from the different datasets and analyses.

Figure 12.3b  Perez’s data on Afro-Yungueño compared with Sessarego’s Afro-Hispanic varieties, standard varieties of Spanish, other Spanish dialects, and creoles in a 72-feature Neighbor-Net.
Figure 12.4 A Neighbor-Net network based on 72 features including three datasets on Afro-Yungueño, other Afro-Hispanic varieties, standard varieties of Spanish, creoles, and other Spanish dialects.

12.5 Discussion of classifications and characteristic traits

As the former section has shown, phylogenetic networks provide a useful tool for visualizing the different analyses of the typological status of a variety based on different sets of data. The varying position of the Afro-Hispanic varieties between the first-language dialects and the Spanish-lexified creoles illustrates the reason for the discussions of their respective statuses. This is also the case for the three datasets on Afro-Yungueño: Lipski and Perez claim that it could be classified as a creole, and the phylogenetic analysis where Afro-Yungueño is positioned closer to the creole varieties supports this classification. On the other hand, the network
also underpins Sessarego’s claim that Afro-Yungueño should be categorized as a nativized advanced second-language variety (cf. Sessarego 2013c). These results suggest that both the nature of the data and their analysis determine the classification of a variety. Indeed, in some cases the same feature may be interpreted in different ways according to the authors’ view; see, for example, Perez’s (2015) and Sessarego’s (2016) contrasting analyses of the status of Afro-Yungueño *pue* (< Spanish *pues* ‘thus, so’). On the basis of the data provided here, we can nevertheless propose features that are essential in the typological classification of the Afro-Hispanic varieties.

In general, the Afro-Hispanic varieties shared many feature values with standard varieties of Spanish. These include, for example, the word order features (APiCS features 1–8), lack of inclusive/exclusive distinction, features about indefinite pronouns and articles, demonstratives, numerals, alignment of case marking of personal pronouns, ditransitive constructions with *give*, etc.

In addition, a number of morphosyntactic and phonological features proved relevant for the classification of Afro-Yungueño, Afro-Choteño, and Afro-Chinchano. The morphosyntactic structures shared by all three varieties were non-inverted polar questions, on the one hand, and more importantly, divergences from Spanish in the NP on the other, i.e. the presence of isolated plural markers instead of plural suffixes. AY was further set apart by a certain degree of restructuring of the pronominal system, i.e. the absence of the Spanish binary politeness distinction in second-person pronouns (Sp. *tú/vos* – *usted* as *oté* only), the loss of gender distinction in personal pronouns (Sp. *él/ella* as invariant *ele*), and the reduction of the lexifier’s three-way distance contrast in demonstratives (Sp. *este* – *ese* – *aquel*). To observe the importance of the pronominal system in the features is interesting since to date the discussions on the pronominal systems of these varieties have been marginal and mainly focused on lexical rather than structural aspects (e.g. Lipski 2005: 250–251, Lipski 2008: 98). In sum, the main features that clustered the Afro-Hispanic varieties together belong to the structures of the phonological inventory and the NP.

The phonological features shared by the three Afro-Hispanic varieties studied here are the preference for open CV syllables (either by adding paragogic vowels or omitting syllable-final consonants) and for the replacement of the *f* by *[xw]* or *[hw]*. These features were present in all three datasets (Lipski’s, Perez’s, and Sessarego’s). However, it is a well-known fact that the alteration of *f* is a common phonological feature found across rural Spanish varieties (see, for example, Lipski 2005: 240–241; Lloyd 1987: 515), and the presence of this element in several Afro-Hispanic dialects may simply indicate their status as isolated, non-urban contact varieties (for more discussion on this point, see Section 12.6).
The differences between the three datasets on Afro-Yungueño are more complex. They mainly concern the verbal system and the prepositional phrase. For example, Lipski (2008: 132) and Perez (2015) analyse certain structures of the verbal system and negation as more divergent from the lexifier than Sessarego does. Regarding the verb phrase, for example, they consider *suppletion according to tense*, i.e. the use of only one form per tense as in *nohotro come* [1pl eat-PRS] ‘we eat’ and *nohotro comió* [1pl eat-PST] ‘we ate’, as categorical. As for negation structures, Lipski (2008: 138) and Perez both hold that *negative* as well as *bipartite negative particles* coexist, thus recognizing double negation as part of this variety’s typological features. These differences explain the varying positions of Afro-Yungueño in the different graphs. In addition, Perez’s data indicate further divergences that are not provided in the other datasets. Among these features are the presence of the interrogative particle *pue* in questions, for example, or the non-existence of reflexive pronouns, which entails that *reciprocal and reflexive constructions* are identical in Afro-Yungueño. These features set Afro-Yungueño further apart from the other varieties and account for the longer line in the phylogenetic analysis resulting from Perez’s data (see Figure 12.4).

In summary, the features that determined the typological classification of Afro-Hispanic varieties in our study were the open CV syllables, the reduction of the VP and the NP (above all the pronominal system), and non-inverted questions. Additional traits of restructuring, such as changes in the use of prepositions, led to a further convergence between the Afro-Hispanic varieties and the creoles.

Second-language varieties, in fact, display similar features that set them apart. Variation in number and gender agreement, for example, is also present in these varieties. However, given their position closer to the standard (probably due to the influence of the respective national standard varieties, which do not deviate considerably from other standards), these varieties group together, apart from the languages we labelled as “creoles”.

These results are in line with Lipski’s (2008: 183–184) observations on the features that he considers relevant to determine the creole status of certain Iberoromance contact varieties. His comparison included 13 morphosyntactic features: null definite articles, plural particles, TMA particles, invariant verb forms, non-inverted questions, double negation, negator from *no/não, to have* as an existential verb, postposed noun phrases as possessives, subject pronouns used as objects, serial verbs, predicate clefting, and, finally, the absence of grammatical gender. Out of this list, Palenquero would display twelve, while Cape Verdean Creole and Afro-Yungueño would present nine features, which led Lipski (2008: 183) to conclude that Afro-Yungueño is typologically close to Cape Verdean Creole Portuguese. Our study included features from Lipski’s (2008: 184) list from the outset, i.e. the absence of gender distinctions, invariant verb forms,
non-inverted questions, and double negation. Others were included in the APiCS values (see Appendix 12.1).

However, the results of the phylogenetic analysis do not address the issue of the origins of Afro-Yungueño, i.e. whether it is more likely to stem from a 16th century Hispanic pidgin (Lipski 2008), an 18th century Portuguese contact variety (Perez 2015), or from an advanced second-language variety (Sessarego 2013b). The question cannot be answered without socio-historical evidence about the contact situation (Lipski 2005: Chapter 9, Lipski 2008: 44–45; Singler 2008).

12.6 Reflections on the method

The analysis of Afro-Hispanic varieties through a network model yielded some very interesting results. On the one hand, we were able to shed light on the classification of these varieties, as it has become evident that the varieties are placed on a continuum in which clear-cut classifications can be difficult to reach. On the other hand, we showed that different authors’ datasets on the same variety, in this case Afro-Yungueño, can result in different outcomes. We will briefly discuss the possible reasons behind the differing results and assess their implications for the method used. Our aim is to reveal some of the pitfalls in the study of lesser-documented varieties through large-scale computational methods.

The clustering of Afro-Yungueño as either a creole-like or a dialectal variety in the networks can be due to differences at the level of (a) data collection and data type, (b) grammatical analysis, (c) feature selection and value assignment, and (d) the mathematical model chosen in the phylogenetic analysis. It is especially relevant to keep in mind the first three levels of analysis, which range from the speech level to the value assignments for features, in the case of lesser-studied and under-documented varieties for which no corpora or dialectal databases exist.

From a methodological point of view, cross-variety comparisons should be based on the principle of accountability based on corpora or fieldwork data where the described lect can be extracted and defined with precision regarding, for example, the axes of formality, register, and extra-linguistic factors, such as age, sex, gender, and education. However, for practical reasons an idealized comparison variety is often created in large-scale studies. Especially in the case of lesser-studied varieties, such detailed descriptions simply do not exist, and comparative studies have to rely on whatever information is available.

A second factor relating to many levels of analysis is how different corpora are combined in comparative studies. The nature of the reference point varies also in our study: the varieties studied are abstractions, combining very different data types (for discussion in sociolinguistics, see Poplack & Levey 2010: 395). Although
a large part of our data comes from either fieldwork data (Perez, Sessarego) or questionnaire-based typological atlases (APiCS, WALS) that were compiled from descriptive materials, naturalistic corpus data, and expert or native speaker knowledge, variation was not taken into account in the feature value assignment. The variation challenge can be lessened by using two datasets from the same community, as we have shown. However, even if close collaboration has been sought in the process of compiling the data for this study, different authors analyze their data and assign the feature values differently due to theoretical and methodological divergences. An approach that includes checks on the reliability of analysis could help descriptive studies to overcome these challenges. In the present study, the datasets varied considerably regarding the method of data collection and the purpose, as well as the analytical framework within which the data were analyzed. In some cases, a certain feature was noticed by all three researchers, yet not all of them classified it in the same way; for example, the expression of pronominal subjects in Afro-Yungueño was observed by Lipski, Perez, and Sessarego, yet only Lipski (2008: 100) and Perez see this feature as categorical in the basilect. On the basis of these criteria, the final networks now suggest that the proportion of basilectal data, or their analysis as such, may be higher in Perez’s dataset (see Sessarego 2016 for additional divergences between their analyses).

Regarding the feature selection, as mentioned in the previous section, the presence of /xw/ or /hw/ for /f/ makes us further reflect on the nature of phylogenetic research and on the risks that a superficial reading of the results may lead to. As this feature is a common rural trait across Spanish dialects, the use of /xw/ or /hw/ per se does not add much to investigating the creole status of Afro-Yungueño, Afro-Choteño, and Afro-Chinchano or their supposed African origins. Along the same lines, a superficial analysis of cross-linguistic vowel systems may lead somebody to conclude that Japanese and Spanish, for instance, have much in common, since they share an almost overlapping vocalic triangle. When accounting for similarities and differences, this might be the case, but from a diachronic perspective, these languages do not share a common genetic origin, and such a comparison may thus yield little insight.

Beyond the bare linguistic data as such, the nature of the language types can vary and also affect the results. Our comparison included several different variety types, i.e. standard varieties, regional dialects, and language attrition situations (Sabine River, Philippines). These sociolinguistic realities range from monolingual situations (Peninsular Spanish) to highly multilingual environments (Equatorial Guinea, Philippines). The results of our study show tendencies regarding the varieties that correlate with these sociolinguistic settings. The creoles and standard varieties resulted in clearly distinct groups, while the other varieties appeared in different positions at the center of the network. Distinguishing between processes
of language-internal variation and contact-induced change is a challenging task 
(Heine & Kuteva 2005: 8; Poplack & Levey 2010: 392), which is likely to cause 
“noise” in the network and its interpretation. Similar developments and grammati- 
cal features can be observed as a result of all these processes in different languages. 
This is a fact that might be further accentuated in language attrition and heavy 
contact situations, such as in the case of the Afro-Hispanic varieties in this study.

Interestingly, no areal signal was detected beyond these creole and standard 
clusters. Peninsular Spanish, for example, clustered closely with Latin American 
varieties, and the Philippine creoles and Philippine Spanish appear at different 
ends of the network. This is not surprising considering the typological similarity 
between standard and dialectal varieties and also due to the fact that the sample 
was not geographically normalized between different geographical areas. A similar 
study on a large number of English-based varieties by Kortmann (2013) produced 
a clearer areal signal. However, Kortmann’s study used a higher number of features 
and included more unified feature values, which may account for the divergent 
results. We strongly believe that, although a general trend in classification of the 
varieties can be induced from our results, caution should be exercised in their 
interpretation – especially on the level of individual varieties – as the above men- 
tioned factors have a considerable effect on their classification.

12.7 Conclusions

In this chapter we have shed light on the classification of some Afro-Hispanic vari- 
eties, with special focus on Afro-Yungueño. Furthermore, we have discussed some 
of the possibilities and challenges of phylogenetic analysis in the study of closely 
related varieties, such as regional varieties of Spanish and Spanish-based creoles. 
Our phylogenetic analysis of selected morphosyntactic and phonological traits 
supports the classification of the Afro-Hispanic varieties as extreme non-standard 
varieties of Spanish. However, whether to classify Afro-Yungueño as a creolized 
variety of Spanish that is decreolizing towards a more standard-like variety, or 
as a dialect of Spanish that was never a creole, depends on the dataset and the 
analytical framework used, and the findings should always be corroborated by 
socio-historical data.

In addition, we have assessed relevant factors that influence the outcomes of 
phylogenetic studies when linguistic data from large-scale variationist studies or 
dialectal surveys are not available. Our analysis showed that differences in data 
types and in analysis, as well as the level of categorical abstraction in typological 
and dialectal features, can produce varying results. The phylogenetic network pro- 
vides a useful tool to visualize how these differences are structured.
Notes

The datasets for this chapter can be found here: https://phylogenetic-creole-studies.blogspot.com

References


Appendix 12.1 Feature lists

<table>
<thead>
<tr>
<th>Number</th>
<th>APiCS</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Order of subject, object, and verb</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Order of possessor and possesum</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Order of adjective and noun</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Order of adposition and noun phrase</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Order of demonstrative and noun</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Order of cardinal numeral and noun</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>Order of relative clause and noun</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>Order of degree word and adjective</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>Position of interrogative phrases in content questions</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>Gender distinction in personal pronouns</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>Inclusive/exclusive distinction in independent personal pronouns</td>
</tr>
<tr>
<td>12</td>
<td>18</td>
<td>Politeness distinctions in second person pronouns</td>
</tr>
<tr>
<td>13</td>
<td>21</td>
<td>Indefinite pronouns</td>
</tr>
<tr>
<td>14</td>
<td>22</td>
<td>Occurrence of nominal plural markers</td>
</tr>
<tr>
<td>15</td>
<td>23</td>
<td>Expression of nominal plural meaning</td>
</tr>
<tr>
<td>16</td>
<td>28</td>
<td>Definite articles</td>
</tr>
<tr>
<td>17</td>
<td>29</td>
<td>Indefinite articles</td>
</tr>
<tr>
<td>18</td>
<td>32</td>
<td>Pronominal and adnominal demonstratives</td>
</tr>
<tr>
<td>19</td>
<td>33</td>
<td>Distance contrasts in demonstratives</td>
</tr>
<tr>
<td>20</td>
<td>34</td>
<td>Adnominal distributive numerals</td>
</tr>
<tr>
<td>21</td>
<td>35</td>
<td>Ordinal numerals</td>
</tr>
<tr>
<td>22</td>
<td>36</td>
<td>Sortal numeral classifiers</td>
</tr>
<tr>
<td>23</td>
<td>38</td>
<td>Marking of possessor noun phrases</td>
</tr>
<tr>
<td>24</td>
<td>42</td>
<td>Comparative standard marking</td>
</tr>
<tr>
<td>25</td>
<td>54</td>
<td>Suppletion according to tense and aspect</td>
</tr>
<tr>
<td>26</td>
<td>56</td>
<td>The prohibitive</td>
</tr>
<tr>
<td>27</td>
<td>58</td>
<td>Alignment of case marking of full noun phrases</td>
</tr>
<tr>
<td>28</td>
<td>59</td>
<td>Alignment of case marking of personal pronouns</td>
</tr>
<tr>
<td>29</td>
<td>60</td>
<td>Ditransitive constructions with 'give'</td>
</tr>
<tr>
<td>30</td>
<td>62</td>
<td>Expression of pronominal subjects</td>
</tr>
<tr>
<td>31</td>
<td>70</td>
<td>Comitatives and instrumentals</td>
</tr>
<tr>
<td>32</td>
<td>71</td>
<td>Noun phrase conjunction and comitative</td>
</tr>
<tr>
<td>33</td>
<td>72</td>
<td>Nominal and verbal conjunction</td>
</tr>
<tr>
<td>34</td>
<td>73</td>
<td>Predicative noun phrases</td>
</tr>
<tr>
<td>35</td>
<td>76</td>
<td>Predicative noun phrases and predicative locative phrases</td>
</tr>
<tr>
<td>36</td>
<td>77</td>
<td>Predicative possession</td>
</tr>
<tr>
<td>37</td>
<td>88</td>
<td>Intensifiers and reflexive pronouns</td>
</tr>
<tr>
<td>38</td>
<td>89</td>
<td>Reciprocal constructions</td>
</tr>
<tr>
<td>39</td>
<td>91</td>
<td>Applicative constructions</td>
</tr>
<tr>
<td>Number</td>
<td>APίCS</td>
<td>Feature</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>40</td>
<td>92</td>
<td>Subject relative clauses</td>
</tr>
<tr>
<td>41</td>
<td>97</td>
<td>’Want’ complement subjects</td>
</tr>
<tr>
<td>42</td>
<td>100</td>
<td>Negative morpheme types</td>
</tr>
<tr>
<td>43</td>
<td>102</td>
<td>Negation and indefinite pronouns</td>
</tr>
<tr>
<td>44</td>
<td>103</td>
<td>Polar questions</td>
</tr>
<tr>
<td>45</td>
<td>108</td>
<td>Para-linguistic usages of clicks</td>
</tr>
<tr>
<td>46</td>
<td>112</td>
<td>’Hand’ and ‘arm’</td>
</tr>
<tr>
<td>47</td>
<td>120</td>
<td>Tone</td>
</tr>
<tr>
<td>48</td>
<td>122</td>
<td>Nasal vowels</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number</th>
<th>Dialectal feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>Loss through heavy aspiration of final /s/</td>
</tr>
<tr>
<td>50</td>
<td>Verbal infinitive (or base form) ends in /r/</td>
</tr>
<tr>
<td>51</td>
<td>Yeismo</td>
</tr>
<tr>
<td>52</td>
<td>Phoneme /i/ &gt; [hw] before unrounded vowels</td>
</tr>
<tr>
<td>53</td>
<td>Neutralization of flap /r/ and trill /r/ distinction</td>
</tr>
<tr>
<td>54</td>
<td>/r/ &gt; /hr/</td>
</tr>
<tr>
<td>55</td>
<td>Paragogic vowels added</td>
</tr>
<tr>
<td>56</td>
<td>Alterations of the type unstressed final [e] &gt; [i]</td>
</tr>
<tr>
<td>57</td>
<td>Non-fricative V[d]V, or [b], [g]</td>
</tr>
<tr>
<td>58</td>
<td>Velarization of final /n/</td>
</tr>
<tr>
<td>59</td>
<td>Assimilation of final /l/ and /r/</td>
</tr>
<tr>
<td>60</td>
<td>/θ/ exist as a distinct phoneme from /s/</td>
</tr>
<tr>
<td>61</td>
<td>Alteration/assimilation of /p/ and /f/</td>
</tr>
<tr>
<td>62</td>
<td>Velar fricative /s/</td>
</tr>
<tr>
<td>63</td>
<td>Verbal agreement altered</td>
</tr>
<tr>
<td>64</td>
<td>Noun-adjective agreement altered</td>
</tr>
<tr>
<td>65</td>
<td>Omissions or heavy alterations of articles</td>
</tr>
<tr>
<td>66</td>
<td>Invariant plurals</td>
</tr>
<tr>
<td>67</td>
<td>Use of the 3rd person form of the Spanish verb for other uses</td>
</tr>
<tr>
<td>68</td>
<td>Omissions and alterations to pronouns</td>
</tr>
<tr>
<td>69</td>
<td>Use of vosotros for the 2nd person plural</td>
</tr>
<tr>
<td>70</td>
<td>Use of vos prevalent</td>
</tr>
<tr>
<td>71</td>
<td>Use of lo as pronN OD</td>
</tr>
<tr>
<td>72</td>
<td>Palatalization of [s] before [i]</td>
</tr>
</tbody>
</table>
CHAPTER 13

Cognitive creolistics and semantic primes
A phylogenetic network analysis

Carsten Levisen and Kristoffer Friis Bøegh
Roskilde University, Aarhus University / Aarhus University

This study presents a semantics-driven lexical comparison of 20 creole languages and five European lexifier languages. Breaking new ground into understanding creole semantics, it utilizes insights from both cognitive semantics (in particular, the Natural Semantic Metalanguage approach) and phylogenetic approaches to linguistics comparisons. We provide an extensive study of label-meaning correlations as a way of exploring the relationship between word labels and word meanings across creoles and lexifiers. We conclude that creoles are not simply “versions” of their lexifier languages, and that it is misleading to say that creoles are “based” on European languages in their basic lexical-semantic configuration. At the same time, we find that creoles do relate more closely to their historical lexifiers than to other creoles, and that the lexical-semantic perspective adds a new dimension to the typology of creoles, nuancing the pictures from grammar-based comparisons.

Keywords: lexical semantics, cognitive creolistics, semantic primes, cross-creole comparison, creole-lexifier comparison

13.1 Introduction

This chapter compares a sample of creole languages and their lexifiers from the perspective of the lexicalization of basic semantic categories. We rely on the findings from several decades’ worth of semantic research, which has explored the shared semantic core of human languages. The Natural Semantic Metalanguage research programme (or the NSM approach) has found that words such as ’I’, ’you’, ’big’, ’small’, ’think’, ’feel’, ’good’, ’bad’, ’above’, ’below’, etc., are semantic primes, i.e. they are simple, indefinable elements of meaning, which, presumably, have been lexicalized in all languages of the world (Goddard & Wierzbicka 1994, 2002, 2014a, 2014b; Wierzbicka 2006; Peeters 2006; Goddard 2008). NSM research has been utilized in a number of ways in linguistics, cognitive studies,
Carsten Levisen and Kristoffer Friis Bøegh

and cross-cultural studies, but to the best of our knowledge, our study is the first to apply the NSM approach in a cross-linguistic study of creoles. In our work on creole semantics, we acknowledge especially the pioneering work of Stanwood (1997, 1999) on the semantics of Hawai‘i Creole English, Bartens and Sandström’s (2006) work on semantics in Iberoromance creoles, and Priestley on Tok Pisin (2008, 2012, 2013). We share the view expressed by Bartens and Sandström (2006):

Since the proposal of the NSM framework is to establish a natural semantic metalanguage consisting of a set of semantic primes common to and translatable into all languages, in short, the core of any language, the framework can be argued to be of certain interest to the debate on the origin and nature of creole languages whether these semantic primes consist of items retained from the lexifier language, or whether they have been taken from the sub- and/or adstrate. (p. 33)

Following the lead of these and other semantic studies, our aim is to undertake a new, large-scale semantic study of the lexicalization patterns for basic concepts across creoles, and to utilize phylogenetic networks in order to form and test hypotheses about their similarities and differences. The phylogenetic network technique allows us to visualize the relations between creoles and between creoles and their European lexifiers, and to add a perspective from lexical semantics to the grammatically-focused phylogenetic studies of creoles that already exist.

The scholarly investigation of creole languages has for a long time been dominated by two discourses: The complexity discourse, essentially the question of whether creole morphosyntax in general can be described as “simpler” than non-creole morphosyntax (see e.g. DeGraff 2001, 2005; McWhorter 2001, 2011; Bakker et al. 2011), and the stratum discourse which has investigated in detail the etymons in creole words and grammars from the perspective of superstrate influences and substrate influences (see e.g. Michaelis 2008; Holm & Intumbo 2009; Lefebvre 2011). These two important discourses in creolistics are both structuralistic in their ideological foundations and tend to be more focused on form than on meaning. By contrast, our study is meaning-based, and we seek to open a third discourse, namely a cognitive-semantic perspective, which is different from the complexity and stratum discourses. One of the core challenges for a meaning-based approach to creole semantics is to establish a stable etic grid upon which emic patterns can be read and interpreted. From a cognitive perspective, there is a clear advantage to working with natural language categories, rather than using “convenience categories” created by researchers. The crux of the matter is to establish a framework of comparison which reflects cognitively authentic categories. One line of lexical research has sought to anchor the etic grid in a language-external reality by using an extensionalist methodology, for instance by asking people to name pictures, describe video clips and similar. This extensionalist paradigm

...
has been advocated especially by the Nijmegen school of psycholinguistics (see e.g. Levinson et al. 2003; Majid, Enfield & van Staden 2006; Majid, Jordan & Dunn 2015). This paradigm has merits, but it cannot stand alone. As critiques of the methodology has noted, words represent ideas, and ideas cannot be captured as essences on videos or pictures. From a cognitive viewpoint, it must be maintained that the meanings of words embed in them a certain perspective and interpretation of the world, and that words are names for concepts, not directly names for “things in the world”. Also, semantic-conceptual constructs have developed in diverse ways across natural languages, which makes many word meanings in natural languages incommensurable (for discussion, see e.g. Wierzbicka 2007; Goddard & Wierzbicka 2014a; Enfield 2015). In order to make valid comparisons and clear hypotheses, we need to find an etic grid of concepts shared by natural languages, on which we can base our comparison. This is where the NSM approach offers itself as a well-tested tool for doing comparative cross-linguistic semantics (Goddard 2008), and where creole languages offer a uniquely interesting input to the theories of cognitive semantics and empirical lexical-semantic research (cf. Bartens & Sandström 2006).

13.2 Exponents of semantic primes across creole languages

The current version of the Natural Semantic Metalanguage research operates with 65 semantic primes (Goddard & Wierzbicka 2014a). The hypothesis is that all languages (or nearly all languages) have lexicalized these meanings in the form of an identifiable semantic unit. These lexicalizations are called exponents of the semantic primes. An exponent of a prime is a lexical structure which can take the form of a single word label, a phraseme label, or a bound morpheme label. The English exponents are presented in Table 13.1.

Compared with the Swadesh lists (Comrie & Smith 1977; Bowern 2008), a classic compilation of words and meanings, utilized for historical comparative research (Swadesh 1971: 283), there are a number of advantages to using exponents of semantic primes in large-scale lexical comparisons. Firstly, the basic vocabularies of the “Swadesh list tradition” were established impressionistically, reflecting partly a universal and partly a Eurocentric perspective on semantics.¹ Swadesh lists do not advance any cognitive claims, and they are not based on systematic

---

¹ For instance, the Swadesh list contains non-universal European color terms like yellow, red, and green (Wierzbicka 2013; Levisen, Sippola & Aragón 2016), European ethnogeographical terms like mountain (Bromhead 2011; Wierzbicka 2013), and English body part terms like neck which do not have any precise semantic equivalents across the world’s languages (Levisen 2015).
Table 13.1 Semantic Primes, English exponents (Goddard & Wierzbicka 2014a; Wierzbicka 2014).

<table>
<thead>
<tr>
<th>Substantives</th>
<th>Relational substantives</th>
<th>Determiners</th>
<th>Quantifiers</th>
<th>Evaluators</th>
<th>Descriptors</th>
<th>Mental predicates</th>
<th>Speech</th>
<th>Actions, events, movement</th>
<th>Location, existence, specification</th>
<th>Possession</th>
<th>Life and death</th>
<th>Logical concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>I~ME, YOU, SOMEONE, SOMETHING, PEOPLE, BODY</td>
<td>KIND, PARTS</td>
<td>THIS, THE SAME, OTHER~ELSE</td>
<td>ONE, TWO, MUCH<del>MANY, LITTLE</del>FEW, SOME, ALL</td>
<td>GOOD, BAD</td>
<td>BIG, SMALL</td>
<td>THINK, KNOW, WANT, DON’T WANT, FEEL, SEE, HEAR</td>
<td>SAY, WORDS, TRUE</td>
<td>DO, HAPPEN, MOVE</td>
<td>BE (SOMEBODY), THERE IS, BE (SOMEONE/SOMETHING)</td>
<td>BE (SOMEONE’S)</td>
<td>LIVE, DIE</td>
<td>WHEN~TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME</td>
</tr>
<tr>
<td>WHERE~PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE, TOUCH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOT, MAYBE, CAN, BECAUSE, IF, VERY, MORE, LIKE~AS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Primes exist as the meanings of lexical units (not at the level of lexemes)
Exponents of primes may be words, bound morphemes, or phrasemes
They can be formally, i.e., morphologically, complex
They can have combinatorial variants or allolexes (indicated with ~)
Each prime has well-specified syntactic (combinatorial) properties.

In a recent article on semantic fieldwork, Goddard & Wierzbicka (2014b:90) comment on these unfortunate omissions:

The absence of words like DO, HAPPEN, WANT, LIKE, BEFORE, AFTER, ABOVE, BELOW, MAYBE, and VERY seems unfortunate, given that linguists have such a high interest in agents, undergoers, volition and control, similarity, time reference, spatial relations, epistemic certainty, and gradability, and, in our experience, it is usually not difficult to identify lexical exponents of these concepts.
Another advantage of the prime-exponent comparison is the cognitive viewpoint that basic semantic units like I, YOU, BIG, SMALL, THINK, FEEL, GOOD, BAD, ABOVE, BELOW, are vital aspects of human language, and hence of the utmost importance in language creation. As such, these units might shed light on vital aspects of the semantics of creolization, both in terms of contemporary configurations of creole languages and from the perspective of historical developments.

13.2.1 Materials

As evidence for our lexicalization of basic semantic categories, we draw on a range of materials. Some previous NSM work has dealt directly with the topic, including Stanwood (1997, 1999), Bartens & Sandström (2006), Priestley (2008, 2012, 2013), Levisen & Jogie (2015), Levisen et al. (this volume, Chapter 15), and we draw on these sources. Secondarily, we draw on material from the APiCS database (Atlas of Pidgin and Creole Structures; Michaelis et al. 2013), a new and extremely helpful source for cross-creole studies. The database covers mainly classic questions in typological linguistics, but it also includes some lexical entries. For our purpose, the single most valuable asset of the database is the searchable example sentence database in the online version, which we have used to find evidence for lexicalizations. Finally, we have also consulted grammars, dictionaries and language experts to make the sample as complete as possible. 

The percentage of coverage in our final data set is 92.1, with 128 out of 1,625 characters (i.e. exponents of primes) missing in the matrix. While future research might reanalyse some of the values, and also fill more gaps, we are comfortable that our analysis will yield results that can progress the current state of creole semantics and advance the field of cognitive creolistics.

We have included 25 languages in our study: Five European lexifiers: Dutch, English, French, Portuguese and Spanish, and 20 creoles (see Table 13.2 for details). We selected the five European languages which are the most relevant in relation to the creation of creoles, but also with regard to the colonial and continuous post-colonial influences of languages in the world. The 20 creoles in the sample have all traditionally been classified as “creoles”, “semi-creoles”, or similar. Our selection criteria were practical: we wanted to rely on the semantically most well-described creoles, but we also sought to balance the selection, by having a global reach with creoles taken from all the main continents, as well as having creoles which relate to the main colonial European lexifier languages.

---

2. We acknowledge that, ideally, all exponents of creole primes should have been devised by speakers of the languages concerned who had been trained in linguistic semantics, and that all exponents should have been tested against the tools developed for doing semantic fieldwork (see Goddard & Wierzbicka 2014b: Appendix).
<table>
<thead>
<tr>
<th>Name</th>
<th>Main lexifier relation</th>
<th>Associated place/nation/area</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayisyen</td>
<td>French</td>
<td>Haiti</td>
<td>Fattier (2013)</td>
</tr>
<tr>
<td>Bislama</td>
<td>English</td>
<td>Vanuatu</td>
<td>Crowley (1990); (Levisen et al., this volume, ch. 15).</td>
</tr>
<tr>
<td>Dutch</td>
<td>–</td>
<td>Netherlands/Belgium</td>
<td>Elicitation (K.F.B)</td>
</tr>
<tr>
<td>English (Anglo)</td>
<td>–</td>
<td>United Kingdom, USA, Australia, etc.</td>
<td>Goddard &amp; Wierzbicka (2014a).</td>
</tr>
<tr>
<td>French</td>
<td>–</td>
<td>Hawaii</td>
<td>Stanwood (1997, 1999); Velupillai (2013); Levisen et al. (this volume, ch. 15).</td>
</tr>
<tr>
<td>Hawai’i Creole</td>
<td>English</td>
<td>Jamaica</td>
<td>Bible Society of the West Indies (2012); Chang (2014); Farquharson (2013).</td>
</tr>
<tr>
<td>Jumiekan</td>
<td>English</td>
<td>Jamaica</td>
<td>Bible Society of the West Indies (2012); Chang (2014); Farquharson (2013).</td>
</tr>
<tr>
<td>Morisyen</td>
<td>French</td>
<td>Mauritius</td>
<td>Baker &amp; Kriegel (2013); Grant &amp; Guillemin (2012).</td>
</tr>
<tr>
<td>Palenquero</td>
<td>Spanish</td>
<td>Columbia</td>
<td>Bartens &amp; Sandström (2006); Hualde &amp; Schwegler (2008); Schwegler (2013).</td>
</tr>
<tr>
<td>Papiamentu</td>
<td>Spanish</td>
<td>Aruba, Bonaire, Curaçao</td>
<td>Goillo (1972); Kouwenberg &amp; Murray (1994); Bartens &amp; Sandström (2006); Kouwenberg (2013b).</td>
</tr>
<tr>
<td>Saramaccan</td>
<td>English</td>
<td>Suriname</td>
<td>Aboh, Veenstra &amp; Smith (2013); McWhorter &amp; Good (2012).</td>
</tr>
<tr>
<td>Spanish</td>
<td>–</td>
<td>Spain</td>
<td>Travis (2005); Peeters (2006).</td>
</tr>
<tr>
<td>Sranan</td>
<td>English</td>
<td>Suriname</td>
<td>Winford &amp; Plag (2013).</td>
</tr>
</tbody>
</table>
In our presentation of the languages, we have made a compromise between “reformist” values in naming (cf. Alleyne 2014), and reader-friendliness. We have used lexifier-free names where possible, either through the use of place-based names (Trinidadian, Jumiekan) or by using other conventionalized language names (Tok Pisin, Bislama). We have also followed creole spelling rules for creole language names in cases where an official (or semi-official) spelling exists. At the same time, we acknowledge that the “place” of a language is not an actual place, but an associative place, i.e. there are numerous speakers of Jumiekan in London, or speakers of Kabuverdiano in Lisbon, Rotterdam (for a discussion of “the language concept”, see Levisen et al. this volume, Chapter 15).

All languages in our selection are presented on the same level, both the so-called lexifiers and the creoles. We do not have any non-contemporary exponents in our sample, except for Virgin Islands Creole Dutch and Berbice Dutch, two Dutch-related creole languages which went extinct in the late 20th century. We have not included any non-standard European dialects or historical colonial versions of the lexicalizations of primes (see discussion in Section 5).

### 13.3 Character coding and phylogenetic algorithm

To account for the lexical semantics of words and meanings, we have developed an approach to coding called mess-values, which stands for Meaning Equivalence and Structural Similarity. The idea is simple: We rely on a set of fixed meaning units (semantic primes) and we explore how these meaning units have been lexicalized across the languages in our sample (the exponents of the same primes). The mess-values format is not limited to, for instance, a maximum of 3, 6, or 9 states for a given character. Each character state represents the unique lexical structuration of a semantic prime in a given language.

We acknowledge that coding in itself is a form of analysis, and we would like to describe carefully how we did it. The premise for our coding practice is that all languages in the sample share some basic semantic meanings, which are then lexicalized in various ways. mess-values allow us to trace differences and similarities in lexicalization patterns across our selection of languages, based on a solid semantic (etic) grid. Table 13.3 exemplifies our coding of the exponents of the semantic prime if.

3. Reformist values, in Alleyne’s sense, means to avoid naming practices which are reflective of colonial prejudices, in the description of languages. In Alleyne’s proposal, this means to adapt a place-based naming practice. The more radical aspect of Alleyne’s proposal, which we have not followed, is to abandon the name “creole” from our description.
Table 13.3  The prime IF coded for four languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>Exponent</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td><em>if</em></td>
<td>1</td>
</tr>
<tr>
<td>Jumiekan</td>
<td><em>ef</em></td>
<td>1</td>
</tr>
<tr>
<td>Tok Pisin</td>
<td><em>sapos</em></td>
<td>2</td>
</tr>
<tr>
<td>Bislama</td>
<td><em>sipos</em></td>
<td>2</td>
</tr>
</tbody>
</table>

In our coding, English and Jumiekan will be given the same value (value 1), whereas Tok Pisin and Bislama will be given another value (value 2). Whereas exact equivalence is the target for ‘meaning’, the lexical structures are based on relative similarity, i.e. we accept minor differences such as *ef* and *if*, and *sapos* and *sipos*.

We acknowledge that some coding judgments can be difficult to make, especially in the case of structural family resemblances where language A is similar to B, and B to C, but where A is only vaguely similar to C. To address the issue, we have designed *mess*-values with different levels of granularity, from rough to fine. Rough values will be the most inclusive, and almost akin to etymological values. But values can also be fine-grained, for instance in order to trace micro-differences, or sociolinguistic details (see also Levisen et al., this volume, Chapter 15). In the present chapter, we have aimed at a medium-level granularity. This seems to be the most useful option for the study of lexicalization patterns across creoles world-wide. Using a medium granularity approach, French *grand* ‘big’ and Spanish *grande* ‘big’ will be given the same value (despite their difference in pronunciation), but Dutch *groot* ‘big’ will be given a different value despite its etymological relation with *grand/grande*.

The most important aspect of our coding practice is that it allows us to break free from both etymological values and structure-only values. By giving semantics the central role, and by focusing on the pairing of semantics with lexical labels, we add a new, important dimension to both creole lexical studies and the study of the semantics of creolization. The Bislama word *sipos* ‘if’ – while obviously structurally taken from the English word *suppose* – does not mean ‘suppose’, and in fact, there are no words in Bislama sharing the same meaning encoded in the English verb *to suppose* or the phrasal structure *I suppose*, both of which are culture-specific constructs of Anglo English discourse (Wierzbicka 2006). However, the semantic prime *IF* exists in both languages, lexicalized as *if* and *sipos*, respectively, and this is why they can be meaningfully and readily compared.

In our choice of algorithm, we have opted for a networks approach, more specifically Neighbor-Net (Bryant & Moulton 2004), which we use as a tool for forming hypotheses, and for visualizing results. Neighbor-Net produces unrooted networks where languages are distanced through splits between branches and nodes in relation to each other (and not to a common root), with branch
lengths reflecting levels of similarity in the data. This approach is well-suited for our purposes, as our unrooted networks make no direct claims about evolutionary histories. Neighbor-Net ensures a maximum relevance in the provided data, in that instead of producing a bifurcating structure (a “compromised” tree), the algorithm represents all of the data by consistently joining taxa under nodes. A tree-like network indicates a consensus for splits in the material, meaning that if the languages under comparison cluster in narrow branch-like formations, this is intuitively interpretable, whereas networks with extensive web formations require further inspection. The network structures are interpreted from the following criteria: (1) exponents of primes are retained from a, potentially common, source; (2) there are innovations in the lexicalization of primes, possibly shared innovations between a subset of the taxa; (3) exponents of primes are similar by chance; (4) characters are misrepresented, including gaps in the collected material.

13.4 Results

In the following, we will present the results of our study showing different networks. We begin by visualizing the relationship between the exponents of primes in five European lexifiers, English, Dutch, French, Portuguese and Spanish. If we look at the five modern standard languages, they show up in a network as portrayed in Figure 13.1:

![Network of contemporary Dutch, English, French, Portuguese and Spanish](image)

**Figure 13.1** Network of contemporary Dutch, English, French, Portuguese and Spanish, based on MESS analysis of semantic primes.
Figure 13.1 replicates some well-known facts about the relationship between the
colonial European languages, in particular the Romance-Germanic divide, and
the relative closeness of Spanish and Portuguese compared to the other lexifier
languages (i.e. it shows support for the concept of ‘Iberoromance Creolistics’ rather
than a strong division into ‘Portuguese-based vs. Spanish-based’ approaches to
lexicalization). In the lexical structuration of basic semantic categories, English
and French are not particularly close, and in fact, Dutch and English are in equal
distance from French. Given what we know about the extensive French influence
on English lexicon, one could perhaps have expected a closer affinity. In the lexi-
calization of basic semantic categories, however, such contact effects do not show
up in noticeable ways.

We will now add other languages to the picture, namely the languages which
are usually classified as “English-based creoles” in order to see how they cluster.
The result can be seen in Figure 13.2:

![Figure 13.2 Network of English, English-related creoles, Dutch French, Spanish and
Portuguese, based on mess analysis of semantic primes.](image)

The analysis shows a fairly uneven distribution of the so-called English-based
creoles. Jumiekan, Trinidadian, Nicaraguan Creole English, Nigerian Pidgin and
Hawai‘i Creole English all cluster in the relative proximity of English, but, impor-
tantly, they are all different from English, i.e. they all differ in their labelling of
basic semantic categories. Secondly, we see that there is a Melanesian grouping
consisting of Tok Pisin and Bislama, which differ considerably from English. Thirdly, we also see a Surinamese grouping consisting of Saramaccan and Sranan, which forms yet another independent group, somewhat closer to the Romance and Romance-related languages.

Phylogenetically speaking, we see three well-formed clusters. A “core” English creole cluster, an Anglo-Melanesian cluster, and an Anglo-Surinamese cluster. The core English creole cluster is even more pronounced in Figure 13.3, where outlier branching has been omitted from the phylogeny. We see that the cluster contains the lexifier as well as a geographically-diverse grouping of creoles. They are set in a robust, out-branching tree-like structure. Comparing the divide between the Iberoromance languages and the various English-based creoles, we see that, while the English-based creoles do form a distinct cluster, the final spread of the taxa indicates variation in the lexicalization of primes that should not be neglected (i.e. they are distinctive languages). While Hawai‘i Creole English and English are set closely together (in the “core” cluster), Jumiekan and Nigerian Pidgin cannot, based on the network, be classified as “very close to English” (on the relatedness between Englishes and English-related creoles, see also Levisen et al., this volume, Chapter 15).

In Figure 13.2, the two clusterings of Tok Pisin and Bislama on the one hand, and of Saramaccan and Sranan on the other hand make it clear that the standard term “English-based” is somewhat problematic, semantically speaking, because it glosses over important differences within this so-called “English-based” group of languages.
Firstly, note that the entire central structure of the network becomes “thicker” compared to that in Figure 13.3 (where it is more tree-like), when the two outliers were included. This suggests that the outliers have drawn webs to both the English-creole “core” group as well as to other colonial language groupings (that we know to have also contributed to the creoles, but to a much lesser extent).

Secondly, including the two outliers does not result in the English-creole “core” cluster being more markedly linked to any other grouping of languages in our sample. Saramaccan and Sranan show up as the least English-like of the so-called “English-based” creoles. The Anglo-Surinamese and Anglo-Melanesian clusters are distinct from other groupings. What is more, they display substantial internal diversity (note especially the distance between Tok Pisin and Bislama, which have often been classified as dialects of “the same” language (for a discussion, see Levisen et al., this volume, Chapter 15).

The phylogenetic networks also allows us to place focus on one language and to grade the relative basic lexico-semantic relatedness of this language to others. We have tested this with English in the centre, giving binary values (binary mess-values): either “like English” or “not like English”. In the network below, we compare English with the other languages in the English-related creole cluster, the Anglo-Surinamese cluster and the Anglo-Melanesian cluster. If we do this, we get the result portrayed in Figure 13.4. The result is virtually identical to the result of standard mess-values coding. The fact that both types of analysis leads to a similar result is indicative of the robustness of our analysis.

![Figure 13.4](image_url) Network of English and English-related creoles based on binary mess analysis of semantic primes.
When we include Dutch-related creoles (Figure 13.5), we get a Dutch-creole cluster, with clear out-branchings in its internal structure.

![Network diagram](image)

**Figure 13.5** Network based on mess analysis of semantic primes (Dutch-related creoles added)

We see a division between Dutch and Afrikaans in one tightly-knit branch, as we would expect, given that Afrikaans is commonly classified as a semi-creole (see e.g. Holm 2001). Virgin Islands Dutch Creole, a creole with mostly Dutch-derived exponents of primes, is distinctly branching off further down in the network structure, and Berbice Dutch is branching off in a manner that reflects its status as an atypical Atlantic creole, in that a major part of its basic lexicon (like its system of inflectional verbal morphology) were originally derived from the Ijoid languages of southern Nigeria (Kouwenberg 1994; see also Bakker this volume, Chapter 10).

Adding now French-related creoles to the picture, we get to the following graph (Figure 13.6):
Roughly speaking, the French-related creoles mirror the structure of the English and English-related creoles’ “core” cluster mentioned above, in that we see a clear divide between the French/French-related cluster and the other groupings. The internal structure of the cluster is characterized by relatively long end-branches, as is the case for Dutch-related creoles. What we can infer from this spread is, as previously mentioned, that we find variation internally among the French-related creoles. From the treatment of creoles in Francophone creolistics, one would perhaps have expected to see French-related creoles being more “alike” and also closer to French than creoles of other colonial-lexical relations. This was not supported by our analysis. However, our study does suggest one major difference between English-related and French-related creoles if we include the Neo-Melanesian and the Surinamese creoles as “English-based”, as is usually done in the literature: There are no parallels to the Anglo-Melanesian or the Anglo-Surinamese clusters in the world of French and French-related creoles. In conclusion, our study suggests that the traditional lex-based theory, which operates with the concepts of “English-based” and “French-based” creoles needs to be nuanced, primarily because these concepts seem to hide the fact that creoles related to English are much more diverse in terms of lexical semantics than French-related creoles are (for further discussion, see Section 13.5).
Figure 13.7 Network based on MESS analysis of semantic primes (Iberoromance-related creoles added).

In Figure 13.7, the full sample is included. The Iberoromance creole cluster bears some resemblance to the French-related creole cluster, but with more intricate internal divisions. The two European languages, Spanish and Portuguese, form the tightest unit. On one side of this group, we find creoles spoken on islands off the West African coast, Kabuverdiano (close to Portuguese and Spanish) and Sao-Tomense (which is on the periphery of the cluster). On the other side, we see the Atlantic creoles Papiamentu and Palenquero. These two creoles branch out quite independently from the others, drawing no clear web formations towards the lexifiers.

13.5 Discussion

We will begin our final discussion by showing a network without any of the colonial languages, asking the question: could we do creolistics without “the lexifiers” as the focal point?

Based on the evidence, we must answer both ‘yes’ and ‘no’ to the question. From a macro-typological reading of the network, we can see that the creoles are indeed grouped into lexifier branches: English, French, Dutch and Iberoromance. This result is interesting, for two reasons. Firstly, it shows that, unlike in grammatical
typology, where creoles cluster together, rather than around the lexifiers (cf. Bakker et al. 2011), the opposite seems true for lexical typology, at least with regards to the label-meaning configurations in the core lexicon. However, it must be noted that the so-called “English-based” group of creoles is the most problematic for lex-based theory. In particular, Anglo-Melanesian creoles and Anglo-Surinamese creoles are too far removed from say, Trinidadian or Hawai’i Creole English semantics to be meaningfully portrayed as belonging to the same “type” of languages.

The other main insight emerging from our study is that we get a much more diverse result by using semantics as a yardstick in lexical comparison instead of the etymology-driven “stratum paradigm”. If the creoles were truly “based” on English, French etc., we would have expected a very different picture, where these most basic concepts were lining up more or less in the exact same way as their lexifiers – essentially replicating Figure 13.1, with very little difference between lexifiers and creoles, or even none at all.

Consider for instance, Bislama, which, seen from a purely etymological perspective, has exponents of semantic primes which are almost exclusively “lexified” by English. Except for save ‘know’, and save ‘can’, the exponents of semantic
primes in Bislama can all be ultimately traced back to English. By contrast, if we add the semantic factor into our lexical analysis, we see that the pairing of labels and meanings is substantially different in Bislama than in Anglo English. This discrepancy suggests that there is considerable cryptodiversity at play, i.e. there is a split between the history of “word labels” and the history of “word meanings” (on cryptodiversity in creoles, see also Levisen & Jogie 2015). The paradigm case is Bislama *sipos* v. English *suppose*, which look alike in terms of their labels, but differ in terms of their meaning. In other words, the semantic factor is pointing away from “Europe”, and we would expect that the tendency would be even more explicit if we moved away from basic lexicon and analysis domains which are known to be more culturally-specific, i.e. words relating to cultural values, emotions, visual categories, ethnogeographical concepts etc. (see e.g. Priestley 2008; Bromhead 2009; Nicholls 2013; Levisen & Priestley 2017).

In summary, the creoles of our sample show two trends in basic lexical semantics: firstly, in a macro-perspective, creoles generally align with colonial languages, or their so-called “lexifiers”, although the “English-based” classification seems somewhat problematic. Secondly, there is considerable diversity across creoles of the same lexifiers, when the interface between lexical structuration and meaning is used as the yardstick. The concrete implication to current creolistic theory and analytical practice would be to adopt a more nuanced classification system than the lex-based tradition, where languages are talked about and thought of as “French-based” or “Portuguese-based”. In particular, the lex-based tradition does not take into account well enough the semantic factor in lexical description. Instead of the “base” metaphor, the most viable alternative would be to go back to the “relation” metaphor, as it was used by e.g. Hancock (1985) and to talk about languages related to, but not based on, English. The “relation” metaphor also allows us to better grade the nature of relationship compared with the categorical “base” metaphor, and this is helpful in cross-creole comparisons as well as in the study of different semantic domains in creoles (see also, Levisen & Aragón this volume, Chapter 14).

James Matisoff, one of founding fathers of areal semantics, has memorably stated that “Franglais does not have much in common with Taglish or Japlish” (2004: 383). If this is true for world Englishes, then it is even more significant and important for creolistics and the study of English-related creoles. The ongoing work on semantic diversity across global Englishes (see e.g. Wolf & Polzenhagen 2009) serves as a major inspiration for creole semantics, and future work needs to explore further the areal semantic perspective, moving beyond the narrow scope of “substrate influences” and into a fully-fledged areal lexical typology, which can account for both traditional languages, creole languages and global Englishes, and the way these languages and varieties construe their social worlds with words.
13.6 Concluding remarks

The phylogenetic network-assisted approach to lexical semantics revealed that we need to qualify and nuance the way we describe creoles. Neighbor-Net analysis of mess-values allows us to compare the lexical exponents of basic semantic categories across creoles and their lexifiers. Different from grammatical typology, where creoles stand out as a group, the study of the basic lexicon shows a different picture, where the creoles cluster with their lexifiers. More precisely, the semantics-based network suggests that there are at least six basic creole groupings: English-related, French-related, Dutch-related, and Iberoromance-related creoles, and two separate clusters, Anglo-Surinamese creoles and Anglo-Melanesian Creoles. Moreover, the semantic-based networks showed that there is considerable diversity within creoles of “the same lexifier” in the most basic semantic categories.

This chapter has only dealt with basic semantic categories, which means that we cannot make any claims about the semantic situation in general. However, we have reason to believe that the difference and diversity in meaning configurations would be even more substantial if we moved outside the core lexicon. Future research in creole semantics should focus on ways in which insights from areal semantics could be integrated into the phylogenetic study of words and meanings.

Acknowledgments

We would like to thank Alexandra Andersson, Aymeric Daval-Markussen, Carol Priestley, Cliff Goddard, Eeva Sippola, Karime Aragón, and Peter Bakker for helpful suggestions, comments, and for help with providing materials for the comparison. Needless to say, any errors or misinterpretations of linguistic categories or materials remain our own.

Notes

The datasets for this chapter can be found here: https://phylogenetic-creole-studies.blogspot.com

References

doi:10.1075/jpcl.26.1.02bak
doi:10.1093/molbev/msh018
doi:10.1016/0024-3841(77)90063-8


Velupillai, V. 2013. Hawai'i Creole structure dataset. In Michaelis et al. (eds).


CHAPTER 14

Lexicalization patterns in core vocabulary
A cross-creole study of semantic molecules

Carsten Levisen and Karime Aragón
Roskilde University, Aarhus University / Aarhus University

The study of semantic domains is important for creolistics, given the complex label-meaning configuration in creoles vis-à-vis the European lexifiers. Due to the lexical semantic creativity in the creolization process as well as the subsequent developments and contacts with lexifiers, substrates, and other contact varieties, each domain seems to have its own history, its own configuration. Comparing words in four different semantic domains, we contrast the labels and lexicalizations of social concepts, body-part terms, environmental concepts, and abstract categories, with an anchor in core categories in conceptual semantics. Based on findings from cross-linguistic investigations into core vocabularies: semantic molecules and semantic primes, we study the lexicalizations of the following meanings in creoles: ‘children’, ‘women’, ‘men’, ‘mother’, ‘father’, ‘wife’, ‘husband’ (social molecules), ‘head’, ‘eyes’, ‘ears’, ‘mouth’, ‘nose’, ‘hands’, ‘legs’ (body-part molecules), ‘sun’, ‘sky’, ‘ground’, ‘water’, ‘fire’, ‘day’, ‘night’ (environmental molecules), and ‘not’, ‘maybe’, ‘can’, ‘because’, ‘if’, ‘very’ and ‘more’ (semantic primes – abstract concepts). We utilize phylogenetic networks to compare and contrast lexicalization patterns between domains. Our study suggests that these core semantic-conceptual constructs tend to cluster with their lexifiers, but that there are important differences across domains – the label-meaning configurations of the social domain stand out as the most diverse, and the environmental domain as the most homogenous.

Keywords: Semantic domains, semantic molecules, cross-creole comparison, creole-lexifier relations

14.1 Introduction

Recent work in cross-linguistic semantics, areal semantics, and lexical typology has made new inroads into the study of semantic domains (see e.g. Wierzbicka 1999; Enfield 2003; Matisoff 2004; Vanhove 2008; Newman 2009; Goddard & Wierzbicka 2014a; Koptjevskaja-Tamm 2015). These studies seem highly relevant

DOI 10.1075/z.211.14lev
© 2017 John Benjamins Publishing Company
to the discipline of creolistics, where new inputs from cognitive semantics and contemporary lexical studies are much needed. Following the cognitive turn in semantics, our approach is conceptual (i.e. non-referentialist) and synchronic. We see words as reflective of shared conceptualizations, shared ideas and shared ways of interpreting the world through language (for conceptual theories of words and meanings, see e.g. Wierzbicka 1992, 2013a; Fauconnier 1999; Sharifian 2011; Goddard & Wierzbicka 2014a).

One of the key insights emerging from areal semantics is that semantic conceptualizations and lexicalization patterns can cut across “language” boundaries. Different semantic domains in a language might reflect different histories. Some domains of a language might be almost identical to those of another language, whereas others might be radically different. Based on this premise, our study sets out to explore how lexicalization patterns of different semantic domains vary across creoles and lexifiers. Using phylogenetic networks, we trace the lexicalization patterns for basic and universal (or near-universal) semantic-conceptual categories in different domains across a selection of creoles. Our aim is to explore the potentially distributed histories of lexicalization patterns across creoles. We look at basic concepts in four different semantic domains: human social concepts, body-part terms, environmental concepts, and abstract logical terms. In our research design, we work with an etic grid of non-arbitrary semantic categories, i.e. with meanings that are globally lexicalized (universals, or near-universals), and our aim is to find out how these concepts have been lexicalized across creoles. We are informed by two types of semantic research, firstly, on semantic molecules, and secondly, on semantic primes (see Goddard & Wierzbicka 2014a, 2014b; Goddard in press; Levisen & Bøegh, this volume, Chapter 13). These two new aspects of semantic theory seem particularly useful in semantic fieldwork (Goddard & Wierzbicka 2014b), as well as in lexical studies of creoles and beyond.

14.2 Creoles, lexifiers, and semantic domains

The development of creole lexicons was, historically speaking, a series of big-scale trial-and-error experiments, where speakers took over, adapted, regurgitated, invented and reassembled meanings, structures, and pairings of meanings with available labels. In our research, we are especially interested in these new pairings of lexical labels with meanings. What is sometimes lost in creolists’ discourse of “substrate” and “superstrate” influences, is the core insight that lexical structure (i.e. labels) and semantic-conceptual configurations can point in different directions in terms of “influence”. In other words, labels derived from European words by no means guarantee “European semantics”.

In creolistics, lexical structures and word forms have often taken priority over meaning. Consider for instance the traditional take on “pequenino research” (see e.g. Huber 2013: 436), where the primary question is whether “a word derived from pequenino exist or does not exist in the language” and on whether the pequenino-like word is disyllabic, trisyllabic, or tetrasyllabic (pikni, piknini, pikinini). Our study differs in that it takes a meaning-first approach. We ground our study in basic semantics, and we ask how universally lexified concepts, such as ‘children’, are expressed across creoles. For example, the two look-alikes Bislama pikinini and Trinidadian pikni are not equivalents. Only Bislama pikinini means ‘children’. The Trinidadian word pikni means something like ‘black children (offensive term)’, and thus, the two words pikni and pikinini stand for two different semantic conceptualizations. From a meaning-first perspective, the lexically comparable words are: Bislama pikinini and Trinidadian chirren (for pickaninny as a racist concept in the Americas, see also Hill 2008: 59).

Our approach bears somewhat more resemblance with Borges’ (2013) phylogenetic study of the semantics of kinship in Surinamese creoles. The main difference to this study is that our approach targets natural semantic categories, i.e. we do not make use of any speaker-external categories such as “-1 generation”, “alter-dependent”, “relative age”, but take our point of departure solely in natural language categories such as ‘father’, ‘mother’, ‘men’, ‘women’, ‘children’. These words are examples of semantic molecules, configurations of meanings which are central in language, and which function as building blocks in semantic concept formation (for semantic molecules, in kinship semantics, see Wierzbicka 2013b). 1 In English for instance, words like toy or tummy are semantically dependent on the molecule ‘children’. Semantic molecules like ‘father’, ‘mother’, ‘men’, ‘women’, and ‘children’ appear to be universal or near-universal. They play a role in the concept formation of any language (Goddard 2011: 375ff.). Some semantic molecules are culture-specific or area-specific, such as ‘God’ [m], ‘country’ [m], and ‘money’ [m], which have played an important role in the semantics of European languages. Other molecules are presumably universal or near-universal. In our study, we will rely on the latter group, the universal or near-universal semantic concepts. We have chosen molecules from three different domains: the social domain, the domain of the human body, and the environmental domain. We have selected seven molecules from each of these domains, 2 choosing the best candidates for universally shared meanings.

1. Semantic molecules are comparable with “intermediate concepts” in the Moscow School of Semantics (see e.g. Apresjan 1992, 2000).
2. In molecule theory, the set of social molecules include also ‘to be born’ [m]. Body-part molecules include also ‘face’ [m], ‘teeth’ [m], ‘breasts’ [m], ‘fingers’ [m], ‘fingernails’ [m], ‘skin’ [m], ‘blood’ [m], and ‘poo’ [m].
In addition to our selection of social molecules, body-part molecules, and environmental molecules, we have added seven semantic primes. These seven primes are all highly “abstract”. They relate to important discourses in linguistics: ‘not’ to “negation”, ‘maybe’ to “potentiality”, ‘can’ to “possibility”, ‘if’ to “conditionals”, ‘because’ to “causation”, ‘very’ to “intensification” and ‘more’ to “augmentation”. Taken together, these four series of word meanings in natural language make up the etic grid for our cross-domain lexicalization analysis. The total list of concepts, is summarized in Table 14.1.

Table 14.1 Selected semantic molecules and semantic primes.


Where there were several possible contenders, contexts of use had to be examined to determine the candidate that had the prime meaning found in other languages. Careful examination was also necessary since in some cases the words that are primes and molecules also have other meanings. The aim was to locate the exponents of semantic primes and the best representatives of semantic molecules.

14.3 Words and coding

Our study includes 15 creole languages and 5 European languages (see Table 14.2). Our material comes from a variety of sources, primarily from research on the lexicalization of semantic molecules and primes, from the APiCS database (Atlas of Pidgin and Creole Structures; Michaelis et al. 2013), from experts (see acknowledgments), and various reference grammars, dictionaries, and Bible translations. The European languages in our sample are English, Dutch, French, Spanish and Portuguese – the five main lexifier languages in the history of European colonization.

Our coding is based on MESS values, which stands for “meaning equivalence and structural similarity” (on MESS values, see Levisen & Bøegh this volume, Chapter 13). The advantage of MESS values is that they are based on the relation between meaning and labels, rather than on purely structural, or etymological principles. Consider the following “test sample” in Table 14.3 of Trinidadian, Nigerian Pidgin, Portuguese, Bislama, and English lexicalizations of the concept of ‘children’.
Table 14.2 The sample.

<table>
<thead>
<tr>
<th>Name</th>
<th>Lexifier relation</th>
<th>Associated places/nations/areas</th>
<th>Consulted material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayisyen</td>
<td>French</td>
<td>Haiti</td>
<td>Fattier (2013); Hall (1953); Valdman et al. (2007); Vilsaint &amp; Berret (2005).</td>
</tr>
<tr>
<td>Chabacano</td>
<td>Spanish</td>
<td>Philippines</td>
<td>Escalante (2005); Forman (1972); Lipski &amp; Santoro (2007); Llamado (1969); Sippola (2011, 2006); Steinkrüger (2013).</td>
</tr>
<tr>
<td>Dutch</td>
<td>–</td>
<td>Netherlands/Belgium</td>
<td>Levisen &amp; Bøegh (this volume, Chapter 13).</td>
</tr>
<tr>
<td>English (Anglo)</td>
<td>–</td>
<td>Britain, USA, Australia, etc.</td>
<td>Goddard &amp; Wierzbicka (2014a).</td>
</tr>
<tr>
<td>Jumiekan</td>
<td>English</td>
<td>Jamaica</td>
<td>Blair (2013); Durrleman-Tame (2008); Farquharson (2013); Yakpo (2012).</td>
</tr>
<tr>
<td>Morisyen</td>
<td>French</td>
<td>Mauritius</td>
<td>Baggioni (1990); Baker &amp; Kriegel (2013); Grant &amp; Guillemin (2012); Hearn (1885); Carpooran (2011).</td>
</tr>
<tr>
<td>Palenquero</td>
<td>Spanish</td>
<td>Colombia</td>
<td>Hualde &amp; Schwegler (2008); Moñino &amp; Schwegler (2002); Schwegler (2013).</td>
</tr>
<tr>
<td>Papiamento</td>
<td>Spanish</td>
<td>Aruba, Curacao</td>
<td>DeBose (1975); Dijkhoff (1993); Goilo (1962); Kouwenberg (2013); Kouwenberg &amp; Murray (1994); van Putte &amp; Garcia (1990).</td>
</tr>
<tr>
<td>Portuguese</td>
<td>–</td>
<td>Portugal</td>
<td>Peeters (2006); Levisen &amp; Bøegh (this volume, Chapter 13).</td>
</tr>
<tr>
<td>Spanish</td>
<td>–</td>
<td>Spain</td>
<td>Travis (2002); Peeters (2006); Levisen and Bøegh (this volume, Chapter 13); Aragón (2016, 2017).</td>
</tr>
<tr>
<td>Tayo</td>
<td>French</td>
<td>New Caledonia</td>
<td>Corne (1997); Ehrhart (1993); Ehrhart &amp; Revis (2013).</td>
</tr>
<tr>
<td>Trinidadian</td>
<td>English</td>
<td>Trinidad</td>
<td>Mühleisen (2013); Winer (1993).</td>
</tr>
<tr>
<td>Virgin Islands</td>
<td>Dutch</td>
<td>Danish Virgin Islands</td>
<td>Hesseling (1905); Josselin de Jong (1926); Sabino (2012); Sluijs (2013); van Rossem &amp; van der Voort (1996).</td>
</tr>
<tr>
<td>Dutch Creole</td>
<td>English</td>
<td>Trinidad</td>
<td>Mühlisen (2013); Winer (1993).</td>
</tr>
<tr>
<td>Yumplatok</td>
<td>English</td>
<td>Australia</td>
<td>Shnukal (1988).</td>
</tr>
</tbody>
</table>
Table 14.3 Lexicalization of ‘children’ in five languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pikinini</td>
<td>Children</td>
</tr>
<tr>
<td>Children</td>
<td>Piknini</td>
</tr>
<tr>
<td>Bikinini</td>
<td>Crianças</td>
</tr>
<tr>
<td>Chirren</td>
<td>(Nigerian Pidgin)</td>
</tr>
<tr>
<td>(Trinadian)</td>
<td>(Portuguese)</td>
</tr>
</tbody>
</table>

In our analysis, these five languages conform to three basic lexicalization types, as in Table 14.4.

Table 14.4 Lexicalization types – ‘children’ in five languages.

Type 1: Children (English), chirren (Trinadian)
Type 2: Pikinini (Bislama), piknini (Nigerian Pidgin)
Type 3: Crianças (Portuguese)

The implication for the creation of a phylogenetic network is that English and Trinadian are given the same MESS value for the feature ‘children’, Bislama and Nigerian Pidgin are seen as different values, and Portuguese yet a different value.

To further illustrate our coding practice, consider the five colonial languages, and their solution to the lexicalization of the semantic concepts of ‘children’ and ‘women’, as Table 14.5 illustrates.

Table 14.5 Lexicalization of ‘children’, and ‘women’ in five European languages.

<table>
<thead>
<tr>
<th>‘children’</th>
<th>‘women’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinderen (value 1)</td>
<td>Vrouwen (value 1)</td>
</tr>
<tr>
<td>Children (value 2)</td>
<td>Women (value 2)</td>
</tr>
<tr>
<td>Les enfants (value 3)</td>
<td>Les femmes (value 3)</td>
</tr>
<tr>
<td>Crianças (value 4)</td>
<td>Mulheres (value 4)</td>
</tr>
<tr>
<td>Niños (value 5)</td>
<td>Mujeres (value 4)</td>
</tr>
</tbody>
</table>

As we can see, all these European languages have different lexicalization strategies for the concept of ‘children’. For the concept of ‘women’, Portuguese and Spanish are similar in lexical labels, which means that they will receive the same coding value.
14.4 Results

Having illustrated our method, we will now present four phylogenetic networks based on the pairings of labels and meanings in four different domains.

14.4.1 Abstract concepts

To recap, the abstract concepts selected for the study were the semantic primes ‘not’, ‘maybe’, ‘can’, ‘because’, ‘if’, ‘very’ and ‘more’. In the following, we will present the lexicalization patterns one by one, and visualize with a phylogenetic network how the exponents of these primes relate to each.

The lexicalization of ‘not’ is diverse in the sample, but there is a tendency towards a no or no-like lexical label across different languages (1). Another recurring lexicalization structuration of the same prime is a pa or pa-like label (2). Finally, there are some lexicalizations in the sample that stand out (3).

1. No-type labels
   - Chabacano no
   - Bislama no
   - Jumiekan no
   - Nigerian Pidgin no
   - Palenquero nu
   - Papiamento no
   - Spanish no
   - Tok Pisin no
   - Yumplatok no
   - Virgin Islands Creole Dutch no

2. Pa-type labels
   - Ayisyen pa
   - Morisyen pà
   - Tayo pa

3. Other solutions
   - French ne...pas
   - Kabuverdiano ka
   - Papiá Kristang ńgka
   - Trinidadian eh

The lexicalization of ‘maybe’ is also very diverse in the sample. Lexical structures like talves (4), mebi (5), petet (6) and ating (7) are the major candidates for semantic primes across creoles and lexifiers, but there are other solutions (8).
(4) Talves-type labels
   Kabuverdiano talves
   Papiamento talbes
   Portuguese talvez
   Spanish tal vez

(5) Mebi-type labels
   English maybe
   Nigerian Pidgin mebi
   Trinidadian maybe

(6) Petet-type labels
   Ayisyen petèt
   French peut-être

(7) Ating-type labels
   Bislama ating
   Tok Pisin ating

(8) Other solutions
   Chabacano sigúro
   Morisyen kikfwa
   Virgin Islands Creole Dutch tunes

The lexicalization of ‘can’ is also diverse in the sample, with pode-like (9) and
kan-like (10) lexicalizations as the most salient. We also see language-specific
solutions (11).

(9) Pode-type labels
   Chabacano pudi
   Kabuverdiano pode
   Papiá Kristang podi
   Portuguese poder
   Spanish poder

(10) Kan-type labels
   Dutch kan
   English can
   Virgin Islands Creole Dutch kan

(11) Other solutions
   Bislama save
   Morisyen kapav
   Tayo kone
   Tok Pisin inap
The lexicalization of ‘because’ is diverse in the sample, with many bikos-like (12), paske-like (13) and porque-like lexicalizations (14), but there are a number of other ways attested as well (15).

(12) Bikos-type labels
   English because
   Nigerian Pidgin bikôs
   Trinidadian because
   Yumplatok bikos

(13) Paske-type labels
   Ayisyen paske
   French parce que
   Morisyen paskì
   Tayo paske

(14) Porque-type labels
   Portuguese porque
   Spanish porque

(15) Other solutions
   Bislama from
   Chabacano kasi
   Jumiekan kaa
   Kabuverdiano pamodi
   Tok Pisin long dispela
   Virgin Islands Creole Dutch fa

Compared with the other abstract concepts, the lexicalization of ‘if’ is not diverse in the sample. Si-like (16) and if-like (17) structures are common, and sapos-like structures in Melanesian creoles (18).

(16) Si-type labels
   Ayisyen si
   Chabacano si
   French si
   Kabuverdiano si
   Morisyen si
   Palenquero si
   Papiamento si
   Portuguese se
   Spanish si
   Tayo si
The lexicalization of ‘very’ is extremely diverse in the sample. For starters, the modern European languages have five different labels for this prime:

(19) **European solutions**
   - Dutch *heel*
   - English *very*
   - French *très*
   - Portuguese *muito*
   - Spanish *muy*

Only a few European-like pairings are found in the creoles (20), the tendency being that creoles vary greatly in their lexicalizations of this prime (21).

(20) **European-type solutions**
   - Ayisyen *tré*
   - Kabuverdiano *mutu*
   - Palenquero *mu*

(21) **Other solutions**
   - Bislama *tumas*
   - Chabacano *bung/byen*
   - Jumiekan *bad*
   - Morisyen *bukù*
   - Nigerian *wèlwèl*
   - Papiamento *masha*
   - Papiá Kristang *bomong*
   - Tayo *tro*
   - Tok Pisin *tumas*
   - Trinidadian *rel*
   - Virgin Islands Creole Dutch *mushi*
   - Yumplatok *mata*
By contrast, the lexicalization patterns for the augmenter ‘more’ are much less diverse. These largely follow colonial lines with either a Germanic base (22), or an Iberoromance base (23), or a French base (24). The French-related creoles are, (at least in a fine-grained coding sense), all somewhat different from contemporary French.

(22) a. **Me-type labels**
   Dutch *meer*
   Virgin Islands Creole Dutch *me*

   b. **Mo-type labels**
   Bislama *moa*
   English *more*
   Jumiekan *muo*
   Nigerian Pidgin *mÔa*
   Tok Pisin *moa*
   Yumplatok *mo*

(23) **Mas-type labels**
   Chabacano *mas*
   Kabuverdiano *más*
   Palenquero *má*
   Papiamento *mas*
   Portuguese *mais*
   Spanish *más*

(24) a. **Ply-type labels**
   French *plus*

   b. **Plis-type labels**
   Ayisyen *plis*
   Morisyen *plis*
   Tayo *plis*

These results take us to the following network in Figure 14.1, based on a shared semantic core of “abstract concepts” and their different lexicalizations.

The first interesting result emerging from the phylogenetic analysis is that the MESS values of only seven features based on primes show a very similar pattern to that of 65 features (for a featurization of all semantic primes, see Levisen & Bøegh, this volume, Chapter 13) – from a big picture perspective, the creoles team up with their lexifiers. At the same time, none of the languages have similar values in all categories, which would have been the case if we had measured the lexicalization of the same semantic concepts in Mexican Spanish v. European Spanish, or Australian English v. British English (see Levisen et al., this volume, Chapter 13). The implication is that all languages in our sample are substantially different in the pairing of lexical structure and these basic concepts.
What the network also reveals is that Morisyen is the creole in the sample which is most different from French in this core lexical domain, and Tok Pisin and Bislama (along with Jumiekan) are the creoles which are furthest from English. We can also see that Spanish and Portuguese stand out from the group of those with European lexifiers, in that they belong to the same cluster. If we look at the internal split in the Iberoromance/Iberoromance-related creoles, we find an interesting picture: there is a division between Spanish/Spanish-related creoles and Portuguese/Portuguese-related creoles, but Spanish and Portuguese (the modern languages), are relatively closer to each other than the Spanish-related creoles and the Portuguese-related creoles are to each other.

14.4.2 Social molecules

We now move to a different aspect of semantics, namely basic social semantics. How do speakers of creoles construe labels for basic concepts in this semantic domain, and does the network differ from the abstract domain? In this subsection, we will first present the results in the form of a phylogenetic network, and then we will discuss the similarities and differences.
It is obvious that the phylogenetic network for basic social meanings is different from the network for abstract semantics. We can still see a basic divide between Romance-related and Germanic-related languages, and especially the French/French-related branch stands out clearly. However, there is no longer a Dutch/Dutch-related v. English/English-related divide, and generally there is a more “interactive” structure to network. On the individual language level, we can see that Yumplatok forms a unit with Bislama and Tok Pisin – and somewhat surprisingly, also with Virgin Islands Creole Dutch – in its basic social language. This group of languages has branched out even more decisively from the group of English-related creoles. The most closely related languages in the social-semantic network are: Tok Pisin and Bislama, Nigerian Pidgin and Jumiekan, Portuguese and Kabuverdiano.

One of the most interesting words in our sample is the semantic molecule ‘children’. The molecule is the collective socio-biological concept expressed in English via a plural form *children*. The lexicalization of this concept is interesting even within European lexifiers, given that all five languages differ:
Within the creoles, there are numerous *pikinini*-type lexicalizations (26). The word figures on Baker and Hubers’ list of “world-wide” lexical features in creoles (2001: 208) and ultimately, this label goes back to a Portuguese word that meant ‘small’ or ‘very small’. In Ayisyen and Tayo we find a comparable lexicalization pattern in *peti* ‘small’ (27), but if we compare the two Melanesian creoles Bislama and Tayo, we see that only Tayo maintains a polysemy pattern: *peti*₁ ‘small’ and *peti*₂ ‘children’, whereas Bislama, has *smol* for ‘small’ and *pikinini* for ‘child’.

### (26) *Pikinini*-type solutions
- Bislama *pikinini*
- Jumiekan *pikni*
- Nigerian Pidgin *pikín*
- Tok Pisin *pikinini*
- Yumplatok *piknini*

### (27) *Pitit*-type solutions
- Ayisyen *pitit*
- Tayo *piti*

Apart from *pikinini*, there is a great variety of different lexicalizations across creoles, some of which have been regurgitated, shortened, reduplicated, borrowed, or in other ways display differences to the solutions of lexicalization in the European lexifier.

### (28) *Other solutions*
- Chabacano *kriyatúra*
- Kabuverdiano *fidju*
- Morisyen *zanfan*
- Palenquero *moná*
- Papiamento *muchanan*
- Papiá Kristang *krenkrensa*
- Virgin Islands Creole Dutch *kin*

---

3. The Spanish molecule is expressed via the lexical structure *niños* which covers the collective socio-biological irrespective of gender, but Spanish also has an elaborate gender-driven lexical semantics with a wealth of lexical structures, including *niños* (collective, males) *niñas* (collective, females), *niño* (individual, male) and *niña* (individual, female).
At a first glance, the labels for the two collective socio-biological meanings ‘women’ and ‘men’ seem closer to the European labels than ‘children’. However, when we take a closer look at the labels, they cluster together in ways that make the creoles group closer together with each other than with the colonial languages.

(29) a. **Men-type solutions**
   English *men*

   b. **Man-type solutions**
   Jumiekan *man*
   Nigerian Pidgin *man*
   Tok Pisin *man*
   Trinidadian *man*
   Yumplatok *man*

(30) a. **Women-type solutions**
   English *women*

   b. **Uman-type solutions**
   Jumiekan *uman*
   Nigerian Pidgin *uman*
   Yumplatok *oman*

(31) a. **Mulheres-type solutions**
   Portuguese *mulheres*
   Spanish *mujeres*

   b. **Muhé-type solutions**
   Chabacano *muhér*
   Palenquero *muhé*
   Papiamento *muhé*
   Papiá Kristang *muhé*

When we talk about the difference in lexical structure between English *men* (collective) and Jumiekan *man* (collective), or Portuguese *mulheres* (collective) and Papiamento *muhé* (collective), an objection could be made, of course, that this can be explained via difference in the reduction of structures and labels which is characteristic of creoles vis-a-vis non-creoles. In our view, all differences should all be modelled, and this is better done if we take a holistic lexicogrammatical perspective, rather than one that leaves out the grammatical facts from the analysis.

Semantically, the pairs ‘mother’ and ‘father’, as well as ‘wife’ and ‘husband’, are relational concepts, which means that they differ from the collective and socio-biological concepts ‘children’, ‘women’, and ‘men’.

In some languages, speakers have invented elaborate ways of encoding endearment or age-relational aspects in addition to such basic words, but in this study
we are only after the most basic concepts. We have attempted to locate the most basic relational category for “someone’s mother”, “someone’s father”, “someone’s husband”, and “someone’s wife”. It appears that the labels *mama* (for mother), and *papa* (for father) are widespread as the lexical structuration for this molecule. This means that there are many false friends of a subtle nature that one needs to be aware of. In English translation, the Bislama phrase *mama blong mi* does not mean my ‘my mama’, but ‘my mother’.

(32) **Mama-type solutions**

Ayisyen *māmā*
Bislama *mama*
Morisyen *maman*
Palenquero *mama*
Papiamento *mama*
Tok Pisin *mama*
Virgin Islands Creole Dutch *mama*
Yumplatok *ama (mama)*

(33) **Papa-type solutions**

Ayisyen *papa*
Bislama *papa*
Morisyen *papa*
Palenquero *pāpā*
Tok Pisin *papa*
Virgin Islands Creole Dutch *popa*
Yumplatok *papa/baba*

With regards to the final pair, ‘wife’, and ‘husband’, it seems clear that these terms might only be near-universals, given that culturally-specific components of meaning may have spilled over into the word meaning. The claim here is simply that the core components in ‘wife’ and ‘husband’ are cross-linguistically lexicalized.

One potentially interesting finding emerging from our sample is that the ‘woman’-‘wife’ polysemy is much more prevalent that ‘man’-‘husband’ polysemy. In our sample, the latter appears exclusively in English-related creoles in the Australia-Pacific region.

---

4. In English, for instance, words like *dad*, and *daddy* provide quite different conceptual takes on relational semantics, than *father*. Likewise French *maman* encodes a different semantics than the basic concept of *mère*, etc.
Chapter 14. Lexicalization patterns in core vocabulary

(34) ‘Woman’–‘wife’ polysemy in creoles
   Bislama woman
   Chabacano muhér
   Kabuverdiano mujer
   Morisyen fam
   Palenquero muhé
   Tayo fam
   Tok Pisin meri
   Yumplatok oman

(35) ‘Man’–‘husband’ polysemy in creoles
   Bislama man
   Tok Pisin man
   Yumplatok man

Label-sharing is a notoriously treacherous field, when we search for “cultural information” in language. From a cognitive semantic perspective, it is important to maintain that polysemy patterns do not necessarily carry any cultural message—or more often, they merely reflect historical pathways which are not or no longer a part of speakers’ linguistic worldviews (for discussion, see Wierzbicka 2007: 50–51; Deignan 2003). In English, “my woman” (in reference to ‘my wife’) is inherently sexist, or at least “macho”, but this is not so in the creoles. The Bislama phrase woman blong mi does not mean “the woman who belongs to me”, but “my wife”. What is more, all the European languages, except for English, display a similar polysemy pattern, and Dutch even has a ‘man’–‘husband’ polysemy.

14.4.3 Body-part molecules

The human body is one of the most intensely studied semantic domains. In fact, human languages have been called “body-centric” because of the tendency to use body words in concept building, either via polysemy or via phraseology (for the semantics of the human body, see Andersen 1978; Senft 1998; Enfield & Wierzbicka 2002; Majid et al. 2006; Wierzbicka 2007, 2013a; Majid 2010; Zouhair & Yu 2011; Kraska-Szlenk 2014; Levisen 2015). We will now go through the major findings in the lexicalization of the following fix points: ‘Head’, ‘eyes’, ‘ears’, ‘nose’, ‘mouth’, ‘hands’, ‘legs’.

5. In Urban Bislama waef blong mi is now widespread. This change might reflect the ongoing contact with English, and perhaps an accommodation to the avoidance of the English problem term my woman.
All seven words are conceptualized as “parts of the body”. Note that the molecules reflect natural semantics. This means that we are studying the concept of ‘nose’ (not ‘noses’), and the concept of ‘eyes’ (not ‘eye’). It is well-known that the word for ‘hands’ can sometimes mean something else in addition to ‘hands’ – in many languages the same word can also be used to talk about something like “arm”, but although “hands” as a semantic unit is clearly delineated, “arms” are not a cross-linguistically conceptual universal or even near-universals (Wierzbicka 2007, for discussion see also Koptjevskaja-Tamm 2008). Apart from the selected seven molecules, the list of body-part molecules also includes ‘face’, ‘legs’, ‘teeth’, ‘fingers’, ‘fingernails’, ‘breasts’, ‘skin’, ‘blood’, and ‘poo’, concepts which will not be taken into consideration in the present study.

The body network unfolds in the following way:

Figure 14.3 Phylogenetic network for lexicalization of body-part molecules.

6. The English word poo has an attitudinal semantic component, which is not universal.
On first impression, the body network provides a middle ground between logical concepts and social concepts in the way the languages relate to each other. The clear-cut grouping of creoles with their “lexifiers” has been restored, but with a slightly boxier structure, and therefore a less sharp overall distinction than in the network for logical concepts.

Perhaps the most interesting difference can be found inside the clusters. English and Jumiekan are outliers in the body network, and, the Australian/Melanesian creoles are now located together with Trinidadian and Nigerian Pidgin as part of the core cluster of the English-related creoles. In the French/French-related cluster, Morisyen again stands out as the most different from the other members of the group. Spanish and Portuguese are again closer to each other than to the Iberoromance creoles, and Papiamento is the most peripheral member. There is a truly “world-wide” effect at play: Palenquero (Colombia) and Chabacano (Philippines) have identical values – and Tok Pisin (Melanesia) and Trinidad (Caribbean) are very closely related in the lexicalization of basic body meanings.

We will now present two case studies based on two molecules ‘head’, and ‘eyes’. These two molecules have been selected because they show differences in their lexicalization patterns. The universal status of ‘head’ is not fully agreed on in the literature. Burenholt (2006: 169), for instance, has claimed that there is no word for ‘head’ in Jahai, a Mon-Khmer language spoken in Malaysia (see also Majid 2010: 64). Wierzbicka, not convinced by the evidence provided for the Jahai case, criticized that the lack of a systematic polysemy analysis and the exotisizing nature of the claim (Wierzbicka 2007). Based on linguistic and comparative analysis, Wierzbicka found in her metastudy (2007) on body-part semantics, that ‘head’ is indeed a universal – or rather a near-universal concept, and she also provided a convincing re-analysis the Jahai case (2013a: 40–42). To clarify Wierzbicka’s claim, consider her definition of the word head, phrased in semantic primes and molecules:

[A] Semantic explication for head
a. one part of someone’s body
b. it is above all the other parts of the body
c. it is round [m]
d. when someone thinks about something, something happens in this part of this someone’s body

The key idea is that the first three components (a–c) are shared and universal aspects and her hypothesis is that in all languages there is a head-like concept meaning “a body part, which is above the other parts of the body and which is round” (for further evidence on the ‘roundness’ factor, see also Majid 2010: 59). At the same time, she contends that component (d), ‘when someone thinks about something,
something happens in this part of this someone’s body’ is a European idea, which is not necessarily shared across languages and cultures (Wierzbicka 2007).

Evidence from creoles suggests that Wierzbicka’s hypothesis is plausible. We have had no difficulty in localizing ‘head’-meanings in text materials and descriptions. At the same, we also found a minor variation in the configuration of ‘head’ concepts. Take for instance hed ‘head’, which in traditional Melanesian creoles relates to ‘wanting’, rather than ‘thinking’ (for instance Bislama stronghed lit. ‘strong head’ means ‘someone who does as he or she want’s, and not ‘someone who thinks well’. For a study on hed in Tok Pisin, see also Kulick (1992). Despite these minor differences in conceptualization, head and head-like semantic configurations tend to be “superstratally” oriented. We find the following seven types of lexicalization patterns in our language sample:

(36) **Hoofd-type labels**
Dutch hoofd

(37) a. **Hed-type labels**
Bislama hed
English head
Nigerian Pidgin hed
Trinidadian head
Tok Pisin hed

b. **Ed-type labels**
Jumiekan ed
Yumplatok ed

(38) **Kabesa-type labels**
Chabacano kabésa
Kabuverdiano kabésa
Spanish cabeza
Palenquero kabesa
Papiamento kabes
Papiá Kristang kabesa
Portuguese cabeça

(39) a. **Tet-type labels**
Ayisyen tèt
French tête
Tayo tet

b. **Latet-type labels**
Morisyen latet

(40) **Kop-type labels**
Virgin Islands Creole Dutch kop
Consider now the concept of ‘eyes’. To our knowledge, there are no claims against the universality of ‘eyes’ in the literature. Wierzbicka’s (2007: 9) definition for this allegedly shared human meaning unfolds as follows:

[B] Semantic explication for *eyes*

a. two parts of someone’s body
b. they are on one side of the head [m]
c. because people’s bodies have these two parts, people can see

The concept of ‘eyes’ not only relies semantically on the concept of ‘head’, but also on the semantic prime ‘two’. In English, this inherently dual concept is realized with the label *eyes*. When we look at the lexicalization patterns across creoles and lexifiers, we find a richer diversity than with regards to ‘head’.

(41) **Ogen-type labels**
    Dutch *ogen*

(42) **Eyes-type labels**
    English *eyes*
    Trinidadian *eyes*

(43) **Ai-type labels**
    Bislama *ae*
    Nigerian Pidgin *ay*
    Tok Pisin *ai*
    Yumplatok *ai*

(44) **Oi-type labels**
    Kabuverdiano *ói*

(45) **Yai-type labels**
    Jumiekan *yai*

(46) **Zye-type labels**
    Ayisyen *je/zye*
    French *yeux*
    Tayo *sje/sjø*

(47) **Lyzye-type labels**
    Morisyen *lyzye*

(48) **Ojos-type labels**
    Chabacano *óhus*
    Portuguese *olhos*
    Spanish *ojos*

(49) **Oho-type labels**
    Palenquero *oho*
Carsten Levisen and Karime Aragón

Wowo-type labels
Papiamento wowo

Ólu-type labels
Papiá Kristang ólu

Hogo-type labels
Virgin Islands Creole Dutch hogo

Etymologically speaking, the labels all seem traceable back to European words. From a contemporary cognitive perspective, the interesting thing is to see how some labels have very similar or identical labels, whereas others are marginally more European-like.

14.4.4 Environmental molecules

The semantics of environmental concepts links with the study of ethnogeography, but meanings which we will explore in this section have a semantic importance that goes way beyond landscape terms (Bromhead 2011a, 2011b). We are researching the lexicalization of the concepts of ‘sun’, ‘sky’, ‘ground’, ‘fire’, ‘water’, ‘day’, and ‘night’, which are usually termed ‘environmental molecules’ (Goddard & Wierzbicka 2014b). As with the concept of ‘head’, we acknowledge that there might be minor differences in the full semantic conceptualizations of ‘sun’, ‘water’, ‘fire’ etc., across linguistic communities because of differences in cosmologies and worldviews in speakers. We do not want to disregard such differences, but it appears that at least all the core components in the lexicalized concept are shared and therefore these molecules remain optimal for cross-linguistic semantic studies.

The environmental network (see Figure 14.4) tells a new story. It sharply distinguishes Germanic-type and Romance-type lexical structures, but the most significant difference is the role of English/English-related languages in the network. English, it seems, is much closer to its related creoles in environmental language than in the other domains. Also, taken as a whole, the English-related creoles cluster closer together than the other clusters in the network. Generally speaking, there is much less variety in the lexicalization of environmental molecules, than say, in logical concepts, social molecules or even body-part molecules. The concept of ‘sun’, for instance, has very few lexicalizations:

Son-type labels
Bislama san
Dutch zon
English sun
Jumiekan son
Nigerian Pidgin son
Tok Pisin *san*
Trinidadian *sun*
Virgin Islands Creole Dutch *sun*
Yumplatok *san*

(54) **Sol-type labels**
Chabacano *sol*
Kabuverdiano *sol*
Portuguese *sol*
Spanish *sol*

(55) **Soley-type labels**
Ayisyen *soléy*
French *soleil*
Morisyen *soley*
Tayo *solej*

(56) **Solo-type labels**
Papiamento *solo*

---

**Figure 14.4** Phylogenetic network for environmental meanings.
The concept of ‘night’, which semantically speaking, roughly means ‘the time where people cannot see the sun, and where (most) people sleep’, also has few lexicalization patterns, and again, the English-related creoles cluster around the same label.

(57) **Nacht-type labels**
    Dutch *nacht*

(58) **Nayt-type labels**
    Bislama *naet*
    English *night*
    Jumiekan *nait*
    Nigerian Pidgin *nayt*
    Tok Pisin *nait*
    Trinidadian *night*
    Yumplatok *nait*

(59) **Nuí-type labels**
    French *nuit*

(60) **Noche-type labels**
    Kabuverdiano *noti*
    Portuguese *noite*
    Spanish *noche*

(61) **Anoti-type labels**
    Chabacano *anoche*
    Papiamento *anochi*
    Papiá Kristang *anoti*

(62) **Lanwit-type labels**
    Ayisyen *lannwit*
    Morisyen *lanwit*
    Tayo *lanwi*

(63) **Dungku-type labels**
    Virgin Islands Creole Dutch *dungku*

14.5 **Discussion**

Every study has its limitations, and especially studies that seek to break new ground. We would like to address two important potential objections to our approach and analysis, one criticism which could be raised by semanticists, and one criticism which might be raised by creolists.
Firstly, there might be an objection to our focus on pairing (i.e. pairing of lexical structure with semantic meaning), and the use of MESS values. We move in an area where cognitive semantics and structural lexicology overlap, and an anti-structuralist semanticist might find our attempt to reconcile meaning and structure in this way somewhat provoking. Our reply would be that semantic concepts are not just concepts of the mind, but embodied in the concrete speech practices of various groups of people. As Enfield (2015: 176) contends: “Linguistic meaning is causally grounded in the making-public of concepts. To be linguistic concepts, they have to be aired, and shared.” Also, the real-life “contact potential”, is governed to a high-degree by the actual labels, and the semantic (re)interpretation of these. However – and this is important – we do not claim that semantics is structure, only that semantic categories are linked to structure and structures. We also acknowledge that MESS values have a limited scope and use for semanticists, in that that they can only be applied in cases where there is semantic equivalence. All languages, including all creole languages, have numerous cultural keywords, and culture-specific semantic categories, which have emerged through the discourse and history of the particular community, and these cannot meaningfully be compared within the design of our current research. In Bislama for instance, there is a rich discourse of *kastom*, roughly ‘traditional culture’, a “celebratory” concept and a cultural keyword which is used to talk about the past as a vital part of modern life (Levisen & Priestley 2017). It would be plainly wrong to equate this concept with English *custom* (the etymon), since the two words differ considerably in both semantic meaning and discursive functions. Culture-specific words make up of one of the unresolved (and possibly irresolvable) issues in phylogenetic approaches to natural language semantic in general. How to meaningfully do cross-linguistic semantics in domains where there may be no etic grid (i.e. no shared meaning-base to rely on), requires a different way of measuring than MESS values. Given the rapid progress in current conceptual semantics, our prediction is that that phylogenetic methods can be used in the future when we have better semantic descriptions of semantic domains in many lects and languages. What we need is careful semantic explications of creole word meanings, which can then be compared, contrasted, and eventually brought into large-scale phylogenetic analysis.

Creolists might say: What about dialects other than the ones represented in this study? And what about historical varieties? Here, the answer is more straightforward and positive. It is possible, at least theoretically, if the evidence is available, to apply MESS values to any kind of “dialect of French”, or to any historical variety of, say, “Melanesian Pidgin”, etc. We could add these languages/varieties to make a more advanced and sophisticated network of domains. In the chapter
on Englishes and Creoles in the Pacific (Levisen et al., this volume, Chapter 15), the question of languages, dialects and varieties will be discussed in more detail, with the assistance of the phylogenetic network technique.

14.6 Concluding remarks

With regards to basic concepts, creoles cluster around the European lexifiers in core vocabulary of all semantic domains: social concepts, body-part terms, environmental concepts, and abstract terms, but there is a domain effect: the social domain is least European-like, and the environmental domain is the most European-like in terms of the labelling of basic concepts. The differences in the networks across domains suggest that semantic domains reflect different contact histories and relations to not only lexifiers, but also in creole-creole relations. We have made advances in the application of phylogenetic techniques to the crossover between lexical studies and conceptual semantics, in a way that has yielded new results and new understandings.

Notes

The datasets for this chapter can be found here: https://phylogenetic-creole-studies.blogspot.com

Acknowledgements

We would like to thank Alan Baxter, Ange Jessurun, Armin Schwegler, Carol Priestley, Diana Guillemín, Dominique Fattier, Eeva Sippola, Joseph Farquharson, Karin Speedy, Kathleen Taylor, Melissa Jogie, Nick Faraclas, Robbert van Sluijs, and Robin Sabino for their kind help with locating words and meanings in creoles. Needless to say, any mistakes or misinterpretations are our own.

References


Josselin de Jong, J. P. B. de 1926. Het huidige Negerhollandsch (teksten en woordenlijst) [Verhandelingen der Koninklijke Academie van Wetenschappen te Amsterdam 26]. Amsterdam: Koninklijke Academie van Wetenschappen te Amsterdam.

Swolkien, D. 2013. Cape Verdean Creole of São Vicente structure dataset. In Michaelis et al. (eds).
This chapter provides a lexical-semantic comparison of a selection of Englishes and English-related creoles in the Australia-Pacific area. Faced with the conundrum in sociolinguistic classificatory practice and its contested categories: “language”, “creole”, “dialect”, “variety”, and “English(es)”, we will attempt to circumvent the problematic of metavocabulary by taking a new, two-pronged approach. Firstly, we rely on semantic primes, simple words meanings such as I, you, people, body, big, small, know, think, see, hear as our comparandum, and compare and contrast the lexicalizations of these basic meanings across our sample. Secondly, we utilize phylogenetic networks for visualizing our results and as a tool for forming new hypotheses. Our results provide counter-evidence to the claim that Melanesian and Australian creoles are “varieties of English”. In our sample, we find three basic types of relations. “Shared-core” types (Australian English v. New Zealand English); “closely related core” types (Hawai‘i Creole v. Anglo Englishes); and “distantly related core” types (Tok Pisin v. Anglo English, Kriol v. Anglo English, or Yumplatok v. Anglo English). We measure our results against Scandinavian languages in order to explore the language-dialect question, and against Trinidadian – a Caribbean creole, in order to explore the extent of lexical-semantic areality. We conclude that current sociolinguistic metavocabulary is inadequate for representing the complexity of the new ways of speaking in the Australia-Pacific region, and we suggest a principled areal-semantic investigation of words based on semantic principles.

**Keywords:** Creole semantics, semantics of Englishes, areal semantics, language ideology, sociolinguistic classification

DOI 10.1075/z.211.15lev
© 2017 John Benjamins Publishing Company
15.1 Introduction

This chapter sets out to explore the similarities and differences in the lexical structures (labels) of shared semantic categories in Englishes and Creoles across the Australia-Pacific region. The aim is to contrast and compare the following areally (and historically) related Englishes and creoles of the Australia-Pacific region: Tok Pisin, Bislama, Pijin, Hawai‘i Creole English, Roper River Kriol, Yumplatok (Torres Strait creole), Australian English and New Zealand English. We include in our study also Urban Tok Pisin (as spoken in Port Moresby) and Urban Bislama (as spoken in Port Vila).¹ Using phylogenetic networks, we visualize similarities and differences in order to shed new light on the semantic and lexical relations between these ways of speaking. Despite their dominantly English-origin lexicons, the contemporary creoles and Englishes in the region are known to be semantically diverse (Jourdan 2002, 2008; Siegel 2008: 1–2; Nicholls 2013; Levisen & Priestley 2017). Many semantic categories are culture-specific and incommensurable – and therefore also “unfeaturizable”. To help overcome this problem, our comparison relies only on meanings which are cross-linguistically shared. The theory-base for the selection of features is based on semantic primes, i.e. on lexical units which have been found to be cross-linguistically expressible in the form of concrete and discrete lexical units, and which therefore can serve as a viable foundation for comparison and “featurization” (on Natural Semantic Metalanguage (NSM) theory in general, see Wierzbicka 2013; Goddard & Wierzbicka 2014; on the application of NSM with phylogenetic network methods, see Levisen & Bøegh this volume, Chapter 13). The list of semantic primes contains simple meanings such as I, you, people, body, know, think, see, hear (Anglo English exponents) or Bislama exponents of the same primes: mi, yu, man, bodi, save, tingting, lukim, harem (for the full list, see Section 15.3).

Our aim is to model difference and similarity from a label-meaning correlation approach: The English label people and the Tok Pisin label manmeri are two very different lexicalizations of the same prime; English body and Tok Pisin bodi are two very similar lexicalizations of the same prime. Taking each semantic prime as a feature, we ask how basic meanings have been lexicalized across English-speaking and creole-speaking communities. Combining insights from conceptual semantics with phylogenetic techniques, our study can help explore important questions of areal influences, local innovations, colonial and postcolonial influences, and global linguistic flows.

¹. The urban varieties are spoken in other towns as well, not only in the capitals. The capitals, however, seem to be centres for innovation, and for the sake of simplicity, we will in this paper focus on the urban lects spoken in Port Moresby and Port Vila.
Based on our findings, we will critically engage with the claim that creoles are “varieties of” their lexifiers, a claim which seems increasingly popular. For instance, in the recent high-profiled De Gruyter Mouton Handbook called *Varieties of English*, creoles were included as “varieties of English”. With the work of Mufwene (2001) as backing, the editors of the Pacific and Australasia volume, Kortmann and Schneider (2008) defend this view by saying that they “adopt the trend of recent research to consider them [creoles] as contact varieties closely related to, possibly to be categorized as varieties of, their respective superstrate languages. (p. 3). In these superstratist theories of creolization, creoles are essentially seen as “kinds of French”, “kinds of English”, etc. (Mufwene 2001; Chaudenson 2001). The evidence from lexical semantics presented in this study seems to point in the opposite direction. Our exploration of creoles in the Australia-Pacific region allows us to discuss and provide new empirical evidence which can help us explore terms such as “language”, “dialect”, “creole”, and “English”, all hotly debated key words in contemporary linguistics and sociolinguistics (see e.g. McWhorter 2005; Goddard 2009; Pennycook 2006, 2010; Blommaert 2013; Bakker et al. 2011; Alleyne 2014).

15.2 The “language” concept and its critics

In the era of postmodern sociolinguistics, the notion of “language” has increasingly come under fire. Postmodern scholars see “language” as a pernicious concept which must be rejected because of its essentialist orientation, its built-in nation-state semantics, and its reification of verbal realities. In this discourse, new terms, such as “languaging”, (Makoni & Pennycook 2007; Jørgensen 2008), “super-diversity” (Blommaert & Rampton 2011; Blommaert 2013), and “metrolingualism” (Pennycook & Otsuji 2015) have been coined to supersede the modernists’ concerns with “languages” and “dialects”. In modern sociolinguistics, “language” was often explained via Weinreich’s dictum: a language is a dialect with a navy and army² (see e.g. Romaine 2000; Edwards 2013: 9, 120; Spolsky 2014: 141). This dictum reflected the idea that a certain dialect, through historical circumstances had gained prestige and power, and had arisen to become the “language” of a particular nation state. This view contrasted with earlier Romantic and essentialist thinking in which dialects were seen as a being in a naturally subordinate relationship to a national language.

However, the discourse of subordination (a dialect of a language), continued in modern sociolinguistics, within a new “variety” concept (a variety of English), a concept which packaged the subordination in a more benign way. The cultural

---

². Weinreich did not coin this saying himself. He made it known to a broader public, and it is commonly associated with this name (Edwards 2013: 120).
logic embedded in the *variety* concept seems to be that it is “good to have variety”, and that “the more variety the better” – whether it be “varieties of ice-cream” or “varieties of English”. The *variety* concept and its hidden subordinating discourses, have not been scrutinized in the same way as concepts of “language”, “creole”, “dialect”, or “English”, but it seems clear that “variety” emerged within the cultural discourse of a capitalist, and liberal-multicultural Anglo world, with its specific conceptual take on verbal realities.

Studies in creoles and contact languages (see e.g. Holm 1988, 2000; Singler 1990; Siegel 2008; Thomason 2001; Mufwene 2001; McWhorter 2005; Lefebvre 2004, 2011, 2015; Velupillai 2015) have largely adapted a modernist understanding of “language”, drawing on both the concept of “variety” and “dialect”. To account for creole-specific sociolinguistic situations, creolists have further incorporated the concepts “basilect”, “mesolect”, “acrolect”, originally proposed by Stewart (1965: 15) and Bailey (1971) (see also Sebba 1997) and Velupillai (2015: 212–215). On the current creolistics scene, there is no consensus about what *creole* means, or how it should be defined. Some scholars have argued that the creole concept needs to be deconstructed (see e.g. Ansaldo, Matthews & Lim 2007), or discarded altogether (Alleyne 2014), whereas others have argued on structural-typological grounds, that it should be vindicated (McWhorter 1998; Bakker et al. 2011). In the related field of World Englishes, the concept of “variety” has been utilized extensively, together with the pluralization of English – *World Englishes, Global Englishes, Indigenous Englishes, Postcolonial Englishes, New Englishes*, etc. (Kachru 1992; Schneider 2007; Kirkpatrick 2007; Mesthrie & Bhatt 2008; Sterzuk 2011; Canagarajah 2013; Lefebvre 2015).

“Creoles” and “Englishes” bring to the discussion some important linguistic paradoxes: in some parts of the world, speakers are fighting to have their creole recognized as a real “language”, whereas other creole speakers are fighting against the use of the term “language” in the description of their way of speaking, because they consider it to be bad, broken, or of a colonial heritage (for discussion, see e.g. Sakoda & Siegel 2003: 1; Alleyne 2014). The view from contemporary sociolinguistics, that “a language” in itself is a fiction, which needs to be deconstructed and resisted in order to understand the nature of “human languaging”, adds yet another layer to the paradox.

The postmodern critique of “language” has some merits. It is historically true to say that the semantics of the English word *language* is compromised by Anglocentrism, and that the family of European concepts: English *language*, French *langue*, German *Sprache*, etc. are conceptually wedded to the nation state. From a postcolonial perspective, it is also clear that the English concept of *language* has been embedded in both colonial and imperial British discourses, and most recently, in an Anglo-international neo-colonial discourse.
On the other hand, there are some weaknesses in the postmodern position. Firstly, the specific attack on “language” seems to lose sight of the bigger picture, namely that neither of the words *language*, *creole*, *dialect*, *variety*, or *English(es)*, are neutral, or culture-independent concepts. Rather, all these metalinguistic categories are English words with an Anglo semantics, and as such they are reflective of Anglo ideals and views of the world. Other linguistic communities have developed different conceptual tools for thinking and talking about different kinds of people and their different ways of speaking. The language ideologies and metavocabularies embedded in Pacific and Australian creoles differ substantially from English (and European) words and ideologies. Consider for instance, the Roper River Kriol phrase *Im sabi tok blekbala* ‘He/she speaks blekbala’). This construct, *tok blekbala*, cuts across the English folk categories *language*, *dialect*, and *creole*. The idea behind the phrase is that the person understands the cultural-interactional strategies associated with being *blekbala*, literally ‘Black fellow’, and the many associated *blekbala* ways of speaking. In Tok Pisin, the *wantok* construct, literally ‘one talk’, emphasizes the importance of people who share the same words and *tok ples* ‘talk place’ emphasizes the association of words with place and land. In Tok Pisin, as well as in Papuan languages, it is common to have one word for meanings such as ‘words’, ‘talk’, and a ‘particular ways of speaking associated with a specific group of people’. Thus for instance, *Koromu sakine* could be translated as ‘Koromu words’, ‘Koromu talk’ or ‘Koromu way of speaking’. Such major differences in semantic conceptualizations show how close the link between word meaning and language ideologies are, and how problematic it is to use *language* – and its Anglo meanings, in the description of speakers everywhere.

Against the more radical side of the postmodern conceptions, we would also like to maintain that speakers everywhere seem to have a metalinguistic awareness of “other people’s words”. Consider again an example of a typical instance of metalinguistic discourse in Tok Pisin:

(1) *Nahe, em i tok ples bilong mipela. Tepere, em i tok ples bilong ol lain bilong maunten.* ‘Nahe (yam), is our word/way of speaking. Tepere (yam), that is the mountain people’s word/way of speaking.’

(Tomas 2004: T7.4.6)

There seem to be a semantic-conceptual recognition of both the words and the ways of speaking of other groups of people, and suitable metavocabularies for these perceived differences have been developed. The key challenge in contemporary sociolinguistics etc. is to reconcile the sociocognitive facts from speakers’ own words, with the justifiably critical approach to English words and Anglo ideologies of *language*.

In the following we will test how the pairing of lexical structures and core meanings compare in a selection of ways of speaking, usually referred to as “creoles” and “Englishes” by linguists. In our attempt to re-think the problems with
classification and metalanguage, we are inspired by work on cross-linguistic semantics and the application of phylogenetic networks in semantics. Phylogenetic networks allow us to take a fresh look at the issues, breaking free of some of the most difficult issues relating to the dialect-variety-language complex and the ideologized positions on the creole question. With a foundation in conceptual semantics and a strong theory-base (the NSM primes), we have secured a non-arbitrary comparandum, against which differences and similarities can be measured.

15.3 Words and coding

We have selected ten named ways of speaking – which in English are commonly labelled “languages”, “Englishes”, “creoles”, “varieties”, or “dialects”, or a combination of these. In our selection we have emphasized that each of the ways of speaking should have psychological reality for speakers, i.e. they should be named or otherwise recognized in discourse as social/linguistic facts. Table 15.1 shows the ten ways of speaking, and some core publications associated with them.

Table 15.1 Sample.

<table>
<thead>
<tr>
<th>Name</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawaii Creole</td>
<td>Sakoda &amp; Siegel (2003); Velupillai (2003).</td>
</tr>
<tr>
<td>Roper Kriol</td>
<td>Sandefur (1979); Munro (2011); Schultze-Berndt, Meakins &amp; Angelo (2013).</td>
</tr>
<tr>
<td>Solomon Pijin</td>
<td>Jourdan (2002); Beimers (2009).</td>
</tr>
<tr>
<td>Tok Pisin</td>
<td>Mühlhäusler, Dutton &amp; Romaine (2003); Smith (2003).</td>
</tr>
<tr>
<td>Urban Tok Pisin**</td>
<td>Devette-Chee (2011).</td>
</tr>
<tr>
<td>Yumplatok (Torres Strait Creole)</td>
<td>Shnukal (1988).</td>
</tr>
</tbody>
</table>

* Autoglossonym: Bislama blong Vila.
** Autoglossonym: Tok Pisin bilong Mosbi.

We recognize that there are other relevant Englishes and creoles in the region, for instance Aboriginal Englishes, Pacific Englishes, Fiji English, Maori English, Pitkern, Norf’k, Nauruan Pidgin English, and Gurindji Kriol (Tryon & Charpentier 2004; Sharifian 2006; Eades 2013; Zipp 2014; Mühlhäusler & Nash 2012; Meakins 2013), as well as historical lects such as New South Wales Pidgin (Troy 1994). There are also ethnolects spoken in urban areas of Australia, such as for instance Lebanese Australian English (discussed by Moore 2008: 194), and sociolects such as Frankofon Bislama, (described by Kanas 2001). All these ways of speaking deserve to be studied, and could be included in future comparative studies.
In our research design, we did not want to assign any special status to the “Anglo Englishes” (i.e. Australian English and New Zealand English) in the way we treated words, meanings, and the pairing of lexical structures with meanings. Also, we did not want to make any assumptions about British English being the “mother language” or being the superordinate of all the other ways of speaking, given that “ancestry”, in any case, is a notoriously difficult and multifaceted concept to operate with in creolistics as well as in the study of Global Englishes.

The words compared in this study are the 66 semantic primes (see the full list of exponents for Anglo English and Bislama in Table 15.2 and Table 15.3).

Table 15.2 Semantic primes – Anglo English exponents (Goddard & Wierzbicka 2014).

<table>
<thead>
<tr>
<th>Exponents</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I<del>ME, YOU, SOMEONE, SOMETHING</del>THING, PEOPLE, BODY, KIND, PART</td>
<td></td>
</tr>
<tr>
<td>THIS, THE SAME, OTHER<del>ELSE</del>ANOTHER</td>
<td></td>
</tr>
<tr>
<td>ONE, TWO, SOME, ALL, MUCH<del>MANY, LITTLE</del>FEW</td>
<td></td>
</tr>
<tr>
<td>GOOD, BAD, BIG, SMALL</td>
<td></td>
</tr>
<tr>
<td>KNOW, THINK, WANT, DON’T WANT, FEEL, SEE, HEAR</td>
<td></td>
</tr>
<tr>
<td>SAY, WORDS, TRUE</td>
<td></td>
</tr>
<tr>
<td>DO, HAPPEN, MOVE</td>
<td></td>
</tr>
<tr>
<td>BE (SOMEWHERE), THERE IS, BE (SOMEONE/SOMETHING), (SOMETHING) IS (SOMEONE’S)</td>
<td></td>
</tr>
<tr>
<td>LIVE, DIE</td>
<td></td>
</tr>
<tr>
<td>WHEN~TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT</td>
<td></td>
</tr>
<tr>
<td>WHERE~PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE, TOUCH</td>
<td></td>
</tr>
<tr>
<td>NOT, MAYBE, CAN, BECAUSE, IF, VERY, MORE, LIKE</td>
<td></td>
</tr>
</tbody>
</table>

Table 15.3 Semantic primes – Bislama exponents (Levisen 2016).

<table>
<thead>
<tr>
<th>Exponents</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI, YU, MAN, SAMTING, OL MAN, BODI, KAEN, PAT</td>
<td></td>
</tr>
<tr>
<td>HEMIA, SEMAK, NARAFALA</td>
<td></td>
</tr>
<tr>
<td>WAN, TU, SAMFALA, OLGETA, PLANTI<del>FULAP, HAMAS…NOMO</del>SMOLSMOL</td>
<td></td>
</tr>
<tr>
<td>GUD, NOGUD, BIFALA, SMOL</td>
<td></td>
</tr>
<tr>
<td>SAVE, TINGTING, WANTEM, NO WANTEM, HAREM, LUKIM, HAREM</td>
<td></td>
</tr>
<tr>
<td>TALEM, TOKTOK, TRU</td>
<td></td>
</tr>
<tr>
<td>MEKEM, HAPEN, MUV</td>
<td></td>
</tr>
<tr>
<td>STAP (LONG WAN PLES), IKAT, I (WAN MAN/WAN SAMTING), (SAMTING) BLONG (WAN MAN)</td>
<td></td>
</tr>
<tr>
<td>LIV, DED</td>
<td></td>
</tr>
<tr>
<td>TAEM, NAO, BIFO, AFTA, LONGTAEM, SOTTAEM, SAMTAEM, WAN TAEM NOMO</td>
<td></td>
</tr>
<tr>
<td>PLES, IA, ANTAP, ANDANIT, LONGWE, KOLOSAP, SAED, INSAID, TAJEM</td>
<td></td>
</tr>
<tr>
<td>NO, ATING, SAVE, FROM, SIPOS, TUMAS, MOA, OLSEM</td>
<td></td>
</tr>
</tbody>
</table>

Exponents of primes exist as the meanings of lexical units, not at the level of lexemes. Exponents of primes may be words, bound morphemes or phrasemes. Exponents of primes can be formally complex. Exponents of primes can have language-specific combinatorial variants (allolexes indicated with ~). Exponents have well-specified syntactic (combinatorial) properties.
For Bislama and each of the lects in our study, the full set of primes had to be
determined by careful analysis. Where there were several possible contenders,
contexts of use had to be examined to determine the candidate that had the prime
meaning found in other languages. Careful examination was also necessary since
in some cases the words that are primes also have other meanings. In others there
are several forms, or allolexes, of the same prime (cf. Goddard & Wierzbicka 2002;
Peeters 2006; Goddard 2011). Thus data from texts and dictionary examples were
checked in order to determine the lists of equivalent semantic primes (apart from
a few words where this hasn’t been established to date).

Our study relies on MESS values (see Levisen & Bøegh this volume, Chapter 13),
which entails that we focus on semantically equivalent terms and their lexicali-
izations. Unlike in other studies (see Levisen & Bøegh this volume, Chapter 13;
Levisen & Aragón this volume, Chapter 14), where the level of granularity targets
the medium level, (i.e. not etymology-based, but inclusive in its sameness judg-
ments of lexical structures), the coding in this chapter aims at fine-level granular-
ity. To exemplify, consider the concepts ‘I’, ‘you’, and ‘someone’, in a fine coding
vis-à-vis a medium-granularity coding, as can be seen in Table 15.4.

Table 15.4 Types of coding.

<table>
<thead>
<tr>
<th>Language</th>
<th>Exponents of primes</th>
<th>Fine-grained coding</th>
<th>Medium granularity coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian English</td>
<td>I-ME</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Bislama</td>
<td>MI</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Australian English</td>
<td>YOU</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Bislama</td>
<td>YU</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Australian English</td>
<td>SOMEONE</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Bislama</td>
<td>MAN</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

While the English exponent of the prime is I with the allolex me, the Bislama expo-
nent of the prime is mi (no allolexy). In a fine-grained coding, these lexicalizations
will show up as two different values, but in a medium-granularity coding, they will
be given the same value, based on the similarity between me and mi. For contrast,
consider you and yu for you which are, both from a fine and medium-granularity
coding, considered similar and someone and man for someone, which both from
a medium and a fine-grained coding practice would be considered different. We
opt for the more fine-grained approach in this chapter in order to understand the
dynamics and interplay of the creoles and Englishes of the area at a detailed level
(see Levisen & Bøegh this volume, Chapter 13, for a more extensive explanation
of the methods).
15.4 Results

In the following sections we will present the results. First we will visualize the network of Englishes and Creoles in our sample, and then, we will compare the network with a network based on Scandinavian lexical semantics. The reason for this is to provide a comparative backdrop. Danish, Norwegian and Swedish are often treated as textbook examples of the Weinreichian “language-dialect” scenario (see e.g. Edwards 2009: 64; Levisen 2012: 20–24). This will allow us to base our comparative analysis with a very well-studied group of languages/dialects. We will also compare our results with Trinidadian Creole, in order to bring in a perspective that can help us determine the degree of areality involved in our sample (i.e. to what degree there is a Pacific factor in basic lexical semantics).

15.4.1 The Australia-Pacific Network

Based on the coding described above, the phylogenetic network for lexical semantics looks as in Figure 15.1:

Figure 15.1 The Australia-Pacific Network.

The first important result is that two linguistic lects show up in exactly the same spot in the network: Australian English and New Zealand English. This is a significant finding. We know that there are structural differences between Australian and New Zealand English, and that there are some lexical differences, and presumably
also some (minor) semantic differences between these two lects taken as a whole (Burridge & Mulder 1998). What the network suggests to us is that the pairing of basic core meanings and lexical structure works in exactly the same way across these two lects. As such, Australian English and New Zealand English can be said to be identical in their core, and because of this, these two ways of speaking form a special relationship, which is unlike any other relation in our sample.

The second important result is that of the alleged “English-based creoles”, only one creole, Hawai‘i Creole, clusters with English. This means that in an Australia-Pacific context, Hawai‘i Creole is a standout lect in two ways: (a) it relates closer to English than any of the other creoles, and (b) it does not relate closely to any of the other creoles of the region.

The third important insight: Tok Pisin, Pijin and Bislama are often talked about as “dialects” of the same Neomelanesian “language” (Keesing 1988: 3). Regardless of how we define “dialect”, what emerges from the network is that these three lects are not related to each other in the same way as Australian English and New Zealand English are related in their core lexical semantics. In other words, if we call Australian English and New Zealand English “dialects” or “varieties” of English, then we might have to find other more meaningful terms with which we can describe the so-called “Neomelanesian dialects”. We need to distinguish between lects with a shared core – but some different cultural keywords (Australian English v. New Zealand English), and lects where even the core differs in its lexical semantic configuration (Tok Pisin, Pijin and Bislama). Certainly, there is a branch-effect that binds together the three Melanesian lects, and moreover, we can see that it mirrors the geographical and “locational” aspect of the languages: Papua New Guinea, Solomon Islands, and Vanuatu, where Solomon Islands Pijin takes a position in the “middle” between Tok Pisin and Bislama. Pijin shares some MESS-values with Tok Pisin, which it does not share with Bislama, and some MESS-values with Bislama, which it does not share with Tok Pisin. The Solomon Islands Pijin separation line is slightly shorter than that of Bislama and Tok Pisin. This means that Tok Pisin is the furthest away from English, further than Bislama, which again is further from English than Solomon Islands Pijin.

The fourth important insight: The two urban lects, which we have called Urban Tok Pisin (Tok Pisin bilong Mosbi) and Urban Bislama (Bislama blong Vila), link closely with traditional Tok Pisin and Bislama. The relative closeness of these lects is expected, given that they are a part of a new kind of diversification. The main

3. The different cultural keywords of Australian English and New Zealand English reflect different colonial histories (i.e. Aboriginal words and concepts in Australian English, and Maori words and concepts in New Zealand English, and other words which have emerged through differences in the cultural discourses of these two national communities. Cf. e.g. the “dreamtime” concept in Australian English, or the “hakka” of New Zealand English.
issue here is that the urban lects are showing signs of separation from the non-Urban, traditional lects – and that this separation is happening even at the level of the lexicalization of core semantic categories. This again suggests to us that we are dealing with a different type of distinction compared to the Australian English and New Zealand English situation, where we found an identical core.

The fifth insight is that the two Australian creoles, Yumplatok and Roper Kriol, do not cluster together. We can see that Yumplatok is relatively speaking closer to the two Anglo English varieties (Australian English and New Zealand English) than Roper River Kriol, and that they are markedly different from the Hawaiian type. Of all the creoles in the sample, this makes Roper Kriol the most isolated lect in the sample.

Based on the phylogenetic network analysis of creoles and Englishes in the Pacific and Australia, we can distinguish between three main different types of relations. These are:

1. The Australian English v. New Zealand English type: This relation is based on an identical core, i.e., there is nothing that distinguishes the two lects in core lexical semantic configuration.
2. The Anglo English v. Hawaiian type: This relation is constituted by a high degree of overlaps in the core, but with numerous significant differences.
3. The Anglo English v. Roper Kriol/Tok Pisin/Yumplatok type: This relation is constituted by a clear distance between the cores in lexical semantic configuration.

We can further observe two more types:

4. The Tok Pisin, v. Pijin v Bislama type: These relations make up a cluster, but with a clear separation between the lects within it.
5. The Tok Pisin v. Urban Tok Pisin type: This relation is based on a slight diversification of lexical cores, but in a way which is “deep” and core-changing, unlike for instance the Australian English – New Zealand “same core” situation.

Having accounted for the types of relations, we can also, based on these phylogenetic networks, propose a separation of the lects in our sample into four different groups (see Table 15.5).

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
<th>Group D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian English</td>
<td>Yumplatok</td>
<td>Pijin</td>
<td>Tok Pisin</td>
</tr>
<tr>
<td>New Zealand English</td>
<td>Roper Kriol</td>
<td></td>
<td>Bislama</td>
</tr>
<tr>
<td>Hawai‘i Creole</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If we remove the two Anglo Engishes, Australian English and New Zealand English, from the network, we get the following network in Figure 15.2.
This picture does not change our previous interpretation, but it helps us to underline two aspects of creoles in Australia-Pacific, namely that there is a Melanesian cluster, whose members are relatively different from each other (except for the two urban lects, which cluster with the traditional lects).

For the sake of comparison, we will now add a Caribbean creole, Trinidadian, to the network, to test how a non-Pacific and non-Australian creole will relate to the creoles in the network. The Trinidadian exponents of semantic primes are taken from Levisen and Jogie (2015).

As we can see in the network in Figure 15.3, Trinidadian links most closely with Hawai‘i Creole. This has implications for our understanding of areality. It appears that there are certain lects (type 2, high degree of overlaps in the core), which do not reflect a major areal component in terms of their basic/core lexical semantics. In the network, there is a gulf between Pacific-Hawaii and Pacific-Melanesia, but not between Pacific-Hawaii and Trinidadian. It also suggests to us that lects which we classified as type 3 (distantly related lexical cores), are more prone to incorporating areal features. We will, in the following sections, dig into one of such areally defined lexical structurations.
A case study in ‘people’

The concept of ‘people’ is a semantically simple, but lexically fascinating concept among the young lects in Australia and Melanesia. There are three main solutions to the lexicalization in our sample: pipel, man, and manmeri. All three labels are presumably English-lexified. The history of pipel and man are self-explanatory. Manmeri is perhaps the most interesting, given that it is formally complex as it is a compound consisting of the two words man and meri ‘woman’. These two words may have English lexical roots since it is sometimes said that meri derives from ‘Mary’ (spelt this way in Hall 1943) or even ‘marry’, Tok Pisin marit in maritman ‘married man’ and haus marit ‘married quarters’ (Mihalic 1971: 134). However, in 1903 Hagen (1903: 248–249) recorded that on the Rai Coast of New Guinea in 1876 the Russian scientist explorer Miklouho-Maclay brought with him a housekeeper from Palau Island called Mira. Local people (in 1903 at least) referred to her as his ‘marry’ using the “Pidgin English name for coloured, i.e. swarthy woman”. In addition, Nevermann surmised (1929: 253–254, cited and translated by Mühlhäuser
1982a: 102–103) that meri could derive from Tolai mari ‘to love’ or mári ‘pretty’ and Mühlhäusler (1982a: 102–103) proposes that meri is a strong candidate for lexical conflation since there are similar forms in at least two contact languages, Tolai/Kuanua and English. Schuchardt (1883: 166) lists woman for ‘woman’, Hagen (1903: 249) gives marry for ‘woman’ and Hall’s (1943) dictionary shows that man-meri ‘people’ (spelt man mary) was in use in the 1940s. From 1969 manmeri was promoted by the Australian administrator in place of kanaka in an attempt to remove “colonialist and racist expressions” (Mühlhäusler 1982b: 111–112). Though we may not know how long this term has been in use, we can see that other lects and languages in Papua New Guinea have similar lexical compounds. Examples from the northern island of Karkar to the Highlands of Papua New Guinea, include Waskia karimet ‘people’ based on kari ‘man’ and imet ‘woman’ (Barker & Lee 2008), Koromu henatamaite ‘people’ based on hena ‘woman’ and tamaite ‘man’ (Priestley 2009, forthcoming), West Kewa onaa ‘people’, based on ona ‘woman’ and aa ‘man’ (Franklin & Kirapeasi 2009), and Kamano vea’ ‘people’ based on vhe ‘man’ and a ‘woman’ (Payne et al. 2005).

(2) Pipel
   People (Australian English)
   People (New Zealand English)
   Pipo (Hawai’i Creole)
   Pipul (Roper River Kriol)

(3) Man
   Man (Yumplatok)
   Man (Bislama)
   Man (Pijin)

(4) Manmeri
   Manmeri (Tok Pisin)

One of the difficulties in the analysis of ‘people’ lies in the fact that the word pipol exists in all the languages of the sample, but it has different functions and meanings and multiple histories. First, there is some evidence of Anglicization in general to account for, and especially in the Urban lects, where pipol appears to be taking over more and more. Another interesting aspect is the trans-creole contact between Melanesia and the Caribbean, facilitated though reggae and dancehall music. Words such as piipl (for ‘people’), an iconic word in Jamaican reggae, appears to be mimicked by Melanesian artists, from where it has the potential to influence the creole languages (Levisen, 2017). If we peel off the layer of recent Anglo and Jamaican influences, it is also clear that the word pipol is not a new phenomenon in a code such as Bislama. It appears to have existed as an auxiliary lexical tool to be used for disambiguation. The background for this is that the word
man has a two-way polysemy. It can mean ‘people’, but it can also mean ‘a man’. The following polysemy analysis is based on Bislama, but we believe that it can be used as to account for the other man-based polysemies in the region.

(5)  a. Man\(^1\) ‘People’
    b. Man\(^2\) ‘A man’

(6)  a. \(Ol\ man ia oli \ldots\)
    b. \(Man ia hemi\ldots\)

In this case, there is grammatical help in the disambiguation process because man ‘people’ attracts the marker predicate marker \(oli\), whereas man ‘person’ attracts the predicate marker \(hemi\) or \(i\). There is however, no technique for distinguishing ‘people’ and the collective concept of ‘men’, on any grammatical grounds. The intended meaning must depend on a contextualized interpretation, and this is where the auxiliary term \(pipol\) comes in handy. \(Ol\ man ia (oli) \ldots\) can mean either ‘these people’ or ‘these men’. The move towards replacing one of the senses of man with a new lexical item in urban lects is attractive, and a classical example of the way in which speakers of creoles can develop their structure-meaning “pairing strategies”. The potential problem for effective communication is that man in Bislama can also mean ‘someone’. This gives the following range of polysemous meanings:

(7)  a. Man ‘people’
    b. Man ‘a man’
    c. Man ‘men’
    d. Man ‘someone’

Tok Pisin has solved the problem in a different way, and in a less polysemy-based way:\(^4\)

(8)  a. Manmeri ‘people’
    b. Man ‘a man’
    c. Man ‘men’
    d. Wanpela ‘someone’

This takes us back to the manmeri structure. From an outsider perspective, it is tempting to apply an etymological interpretation of the word, i.e. to re-invent it interpretatively as ‘men and women’. This, however, would be a major mistake. The fact that manmeri can refer to children as well, points towards the fact that manmeri truly means ‘people’ and not ‘men and women’ (or even worse: men

\(^4\) In both Bislama and Tok Pisin the word man is used for singular and plural. It is common in languages of Melanesia and Australia to have the same word for singular and plural uses of a noun. Plurality may be indicated in other parts of the noun phrase or sentence.
and Maries). Colonial or neo-colonial attitudes to creole semantics are likely to lead to cognitively inauthentic readings, which reflect etymologies and historical processes, rather than meanings.

15.5 Discussion

At this point we will allow ourselves a short digression into Scandinavian phyllosemantics. Given the textbook status of the Scandinavian situation, we believe that a comparison between Scandinavian and Australian/Pacific Englishes and Creoles can help shed light on the nature of similarity and difference. Firstly, we will add Danish to the existing network. When we do that, we get the following picture in Figure 15.4:

Figure 15.4 The Australia-Pacific Network, Danish Added.
As we can see, Danish lexicalizations are not totally unrelated to the languages in the network (such as for instance Chinese or Russian lexicalizations would have been). In fact, several lexicalizations are shared between Danish and English, such as for instance *kan* (Danish) and *can* (English) (for the prime CAN). In the network, we see that Danish is somewhat further removed away from English than any of the other lects in the network. The most interesting finding, seen from a big picture perspective, is to compare the difference between Danish and English in relation to the Tok Pisin v. English or the Roper River Kriol v. English situation. The visualization seems to suggest that if we consider Danish and English to be “different languages”, then we should also consider Roper River Kriol and English, and Tok Pisin and English to be different “different languages” (if we still want to stick to with the traditional language concept, that is).

To further test the classical European situation with Australia/Melanesia, we will as our next move add a new lect, Jutlandic, which is commonly described as a “dialect of Danish” in the literature (Pedersen 2003). We find few, but real, lexicalization differences in the basic semantics of standard Danish v. Jutlandic. To exemplify:

(9) Semantic Prime ‘I’ ‘VERY’ ‘NOT’
    Danish      jeg     meget        ikke
    Jutlandic   a       møj         æ

Also, we will add the two other Scandinavian standard languages, Norwegian (bokmaal) and Swedish. The addition of three Scandinavian “languages”, and one “dialect of Danish” gives us the picture in Figure 15.5.

Unsurprisingly, the network recognizes these four Scandinavian lects as belonging to a separate cluster. Danish, Jutlandic and Norwegian are close, with Swedish as the outlier within the cluster. We also see a Weinreichian effect, given that the distance between Danish and Jutlandic (no navy, no army) and Danish and Norwegian (national army, national navy), is approximately the same. The addition of Scandinavian lects, helps to advance further our understanding of the Australia-Pacific situation, and to further question the discourse of “dialect (of)”, “language” and “variety (of)” in linguistics and public discourse. Firstly, we note that, taken as a whole, the Scandinavian cluster resembles the Melanesian creole cluster in its structure. Secondly, the Danish v. Norwegian relation is structurally quite similar to the Hawaiian v. English type relation.

It is also worthwhile paying attention to how the “old” Scandinavian situation, does not resemble the “young” Australia-Pacific network. For instance, nothing in the Scandinavian cluster resembles that of the Tok Pisin v. Urban Tok Pisin type of relation. Neither can any lects in the Scandinavian cluster or in the Melanesian cluster be likened to the Australian v. NZ English relation. There is an important lesson to be learned from this: the difference in the two young Englishes,
Australian English and New Zealand English, does not display any “kernel difference” in the way the Scandinavian lects do.

Having reviewed the Pacific-Australian sample in comparison with a Scandinavian sample, the evidence from the phylogeny points in one direction. It does not seem viable to call Tok Pisin, Bislama, Kriol or Yumplatok “varieties of English”. In fact, based on our evidence, one could just as well call English “a variety of Danish”. If the distance between Danish and English (in terms of the parameters given in this study) is approximately similar to that of Tok Pisin and English, why then, is it still possible – and even “trendy” – to talk about Tok Pisin as

Figure 15.5 The Australia-Pacific Network, Scandinavian Network Added.
a variety of English? Our study suggests that it is meaningful to classify Australian English and New Zealand English as two kinds of the same things (Englishes). However, Tok Pisin, Bislama, Solomon Islands Pijin, Roper Kriol, and Yumplatok are not “kinds of English”. The problem runs deep, even outside the decidedly superstratist positions. Take for instance the tradition of thinking and talking about Tok Pisin, Bislama, Roper Kriol, etc. as “English-based” creoles, which is just as problematic as the “varieties of English” discourse. English is not a “base” for contemporary Kriol, Bislama and Tok Pisin semantics, just as it is not a base for Swedish or Danish. It is true that in many cases the origin of lexical structures (word labels) were taken from English, but from a meaning-perspective, it is just as unhelpful to call Tok Pisin a variety of English, as it is to call English a variety of Tok Pisin. In fact, the conceptual semantics hidden in statements such as “Tok Pisin (Kriol, etc.) is a variety of English” or “Kriol is based on English”, paves the way for a continued conceptual colonialism in our metalanguage, subjugating the semantics and the lexical structures of Tok Pisin, Roper Kriol, Bislama, etc., to the English language. The positive ring of the **variety** discourse (“variety is good”, “the more variety, the better”), is hiding the subjugating semantics, which does the job of maintaining the status of English as the general product, which then comes in different wrappings.

Our study has laid a foundation for doing phylosemantic studies of creoles and Englishes, but there are numerous aspects which we have not been able to explore in this chapter. Firstly, more lects could be included (see Section 15.3), including also words and meanings from the so-called “substrate languages”. Second, the challenge for future studies in phylosemantics is to find ways to combine the study of basic/core meanings with more complex and more culture-specific meanings (for ideas, see our introduction). Areality is worth exploring in much more detail than our case studies in this chapter have allowed, and the areal aspect should be studied in three ways: areal lexical labels, areal pairings of labels and meanings, and area-specific semantic categories. It is important to emphasize that our work does not give “the” semantic perspective on Englishes and creoles. We have only investigated the lexicalization of core semantics categories, and we have reason to believe that the language-specificity and the aspects of areality would shine through more clearly, if more specific semantic domains were compared. The extent of such diversity remains to be studied.
15.6 Concluding remarks

Phylogenetic networks can help us realize patterns, form new hypotheses, and help us circumvent some of the problematic terms in contemporary sociolinguistics: language, creole, dialect, variety, and Englishes. We have demonstrated how we can start rebuilding typologies based on semantics, and the lexicalization of basic meanings. Based on our results, we conclude that there are three basic types of relations in the Australia-Pacific region (a) identical-core relations, such as Australian English v. New Zealand English, (b) closely related cores, such as the relation between Anglo Englishes and Hawai‘i Creole, (and e.g. Trinidadian, if we expand our reach), and (c) distantly related cores, such as Anglo Englishes v. Tok Pisin, Bislama, Pijin, Yumplatok or Roper River Kriol. The three Melanesian creoles Tok Pisin, Bislama and Pijin resemble each other in a way which is comparable to different Scandinavian languages/lects, but they do not resemble English in basic lexical semantics much more than e.g. Danish resembles English. The “trendy” view of creolization, now influencing World Englishes studies, was not supported by our study. We argue that the current metavocabulary in sociolinguistics is likely to lead to a colonial understanding (the subjugation of creoles to their so-called lexifiers) in ways which must be both questioned and resisted on linguistic and sociocognitive grounds.

Acknowledgements

We would like to thank participants in the workshop “Semantics across Englishes and Creoles”, Aarhus University, October 2014, for helpful ideas and suggestions. More specifically, we would like thank Anna Löfgren for consultancy support on Swedish, Augustron Asial and Karolin Mereisi for consultancy support on Bislama, Elena Govor for translation of Hagen from Russian, and Sairam Tomas for consultancy support on Tok Pisin and Koromu.

Notes

The datasets for this chapter can be found here: https://phylogenetic-creole-studies.blogspot.com
References


doi: 10.3366/edinburgh/9780748625291.001.0001


doi: 10.1017/CBO9781139164153


doi: 10.1080/14790710802387562


doi: 10.1017/S0261444800006583


doi: 10.1075/slcs.70


doi: 10.1075/tsl.95


doi: 10.1515/9783110294651


CHAPTER 16

Feature pools show that creoles are distinct languages due to their special origin

Peter Bakker
Aarhus University

This book presents the main results of the Cognitive Creolistics project, sponsored by the Velux Foundation (2012–2015), to whom we express our gratitude. In the project we wanted to obtain a better idea of the typological and historical connections between creole languages, as well as a better idea of the cognitive foundations behind the typical properties behind creole morphosyntax.

How can we identify a creole language? We found that a limited set of four or five features would characterize all creoles, but no non-creoles (Chapter 4) – at least not among the approx. 100 languages for which we had data available. If these four or five multi-value features do indeed characterize all and only creoles, this can be used a test. One could potentially use this set of features to flag up languages as creoles that were not previously identified as such. Assuming that a special process of creolization is responsible for their collective emergence, then the presence of a negative particle, a verb ‘to have’, an indefinite article derived from the numeral ‘one’, and a lack of tense/aspect inflection characterize creoles.

In this book, we can also see new challenges and new horizons. The three levels of research are reflected in the three parts of the book: The first part (Chapters 5 to 7) compares samples of creoles world-wide with each other and with samples of non-creoles world-wide. In many cases, the non-creoles include the languages that are known to have been spoken by the population groups who contributed to the formation of the creoles. The second part (Chapters 8 to 12) explores historical relations between creoles with the same lexifiers and the possible historical relations between them. The third part (Chapters 13 to 15) relates to the micro-level: semantic connections between creoles (sometimes form-based) and the languages from which the creole lexicons derive. In all cases, phylogenetic programs have been used to shed light on the issues at hand.

The chapters are highly empirical and based on feature pools of varying sizes. One theoretical musing about feature pools was that creoles, and all other results of language contact, are the result of more or less arbitrary combinations of structural features of the languages in contact. The results presented in this book do not lend
support to those suggestions, as the creoles are more similar to one another than to substrates or superstrates, or their combinations. Creoles always have properties not found in any of the contributing languages.

This book is probably among the most balanced in its use of many different creoles. It has chapters focusing on French-derived creoles (6), Arabic-derived creoles (7), Dutch-derived creoles (8), Spanish-Portuguese derived creoles (9,10) and English based-creoles (11, 12, 13). Non-European based creoles are also included, as part of the larger samples in the world-wide survey (5).

Software for phylogenetic developments was created to map biological evolution. Language evolution and the evolution of life can both be seen as the result of continuous modification by descent. There are important differences in the rate of change, the nature of change and the quantity of horizontal transmission. Biologists have used these techniques not only for DNA sequences and on the molecular level, but also on very different, superficial levels such as looking at the presence of wings, feathers, fins and similar. All forms of life on earth seem to go back to one origin, but it is far from certain that all languages go back to one origin. Based on the strict principles of the comparative method, Harald Hammarström identifies over 400 language families and isolates (Glottolog 2.7 lists 430 in February 2017). This number may eventually be reduced, but it is still a staggering number: it is unlikely that language was invented from scratch more than 400 times.

Working with creole languages and phylogenetic networks, we use software developed for biological purposes. We utilize it sometimes to discover historical relationships between creoles, or creoles and non-creoles – which is along the lines of the original purpose of the software. We also use it to track morphosyntactic or semantic similarities that are often not due to shared inheritance. This was not the original aim of these programs, but they have proved to be a really useful tool, for instance in working with a large amount of data, and in the visualization of similarities and differences between the languages involved in creole genesis, and other non-creole languages.

Is there a creole typological profile? I think the results point in different directions. For phonology (even though not investigated in this book), there is nothing special about creoles – except perhaps that they are average. For morphology and syntax, we can be confident in our conclusions that creoles indeed have a special place in the typological space of the languages of the world. This was a conclusion reached earlier by Szmrecsanyi & Kortmann (2009) on the basis of varieties of English (including vernacular varieties), by Bakker et al. (2011/2013) on the basis of linking different sets of existing world-wide databases of creoles and non-creoles, and of gathering comparable information on additional languages, by Bakker (2014b) who added non-European creoles to the set, and by Muysken (2015) on the basis of a project on the roles of African languages on Caribbean creoles, where creoles also cluster apart from the lexifiers and substrates.
Nevertheless, these conclusions have been criticized. Critical voices pointed to errors in assigning specific features, attacked the feasibility of the program, claimed that the selection of features is biased towards European or creole properties, expressed serious doubts about the quantity of features (too few, too many), conjectured that this work had an Indo-European bias, or a European bias in the lexifiers, or objected that it had only one set of biclans (all European lexifiers and Niger-Congo substrates). See Bakker (2014a, b) for discussion of criticisms.

The studies in this book generally lead to the same conclusions as the earlier ones, and most of the objections (to the extent that it was possible) have been addressed in this book. Attacking the program goes against decades of successful work in biology, and one decade of successful work in linguistics by now. We have also included non-European creoles, to the extent that they exist. The authors have used different sets of features in this book – almost all of them based on lists compiled by others (e.g. APiCS; stability studies based on WALS; lists of number distinctions in Nilo-Saharan and Arabic varieties; independently defined semantic primes, translations of a set of sentences, in Chapters 7–9, 11–15), and not specially designed to prove a specific idea. Sometimes features were also used that were triggered by availability of data collected for their study (8–10, 12–15).

Influence from substrate languages could be detected in a number of features in the case of Iberian creoles. The same is true for a number of semantic patterns (13–15), but hardly any African influence could be observed when looking at the most stable features (Chapter 7).

Chapter 7 (and future work by Daval-Markussen, in press) also shows that the inclusion of creoles with different biclans, still finds creoles distinct from non-creoles, so the distinctness is not the result of the specific biclans in earlier studies.

In the book, the authors reach a number of seemingly contradictory conclusions. On the one hand, creoles appear to be incredibly diverse in their structural properties (e.g. not only verb-medial, but also verb-initial and verb-final). On the other hand, all the tests including samples of creoles, non-creoles and a large set of features, lead to results where creoles and non-creoles form separate clusters. Only very rarely are non-creoles found among creoles, or the other way around, with only a few percent of overlap. When focusing on stable features, however, we find that Western European and Semitic languages are a central part of the creole cluster (Chapter 7). This can be interpreted in different ways. First, creoles continue many lexifier structures. This interpretation, however, collides with the observation that other lexifiers such as Japanese and Tupinamba are not part of the creole cluster, even though the Yilan and Nheengatu creoles clearly are (Figure 7.8). Second, it could be that the European and Middle Eastern languages have undergone language contact effects similar to what is found in creolization. Creolization has been invoked as having an effect at least on English, French and colloquial
Arabic. A third interpretation may be that it is just by chance – a possibility, but unsatisfactory.

In my view, empirical studies based on large pools of features have firmly established that creoles inhabit a special place in the typological space of the languages of the world. No opponents have proposed alternative selections of features to contest this. No one has offered any other form of proof of the opposite, nor have they even attempted to do so.

Another important result is the reduction of features that set the creoles apart from non-creoles. Distinct sets of creoles and non-creoles have resulted from studies using 46 features (Cysouw 2009 based on preliminary data), or 76 (Szmrecsanyi & Kortmann. 2009), 97 (Bakker et al 2013, based on Holm & Patrick 2007), 43 (Bakker et al 2013, based on Parkvall 2008), 83 (Muysken 2015) features, all based on different datasets and/or different features. But also selections of 30 features (Chapter 7 and Chapter 9, which deal with totally different sets), 48 features (Chapter 12), 72 features (Chapter 12), perhaps also 130 (Chapter 11), and 65 (Chapter 13) or even merely four or five features (Daval-Markussen 2013, and Chapter 6 in this volume) can distinguish creoles from non-creoles in the available samples. McWhorter (1998) proposed as few as three features as being found in all and only creoles. This includes one phonological feature, and all of these relate to absence rather than presence. The only successful challenge is Harald Hammarström, who proved that it is possible to select 12 features with 12 creoles and 12 non-creoles, in which creoles do not form a separate group. Any other selection of existing data sets, based on existing lists of features, sets creoles apart, as shown with different sets in e.g. Bakker et al. (2013).

It should be pointed out that lexical semantics (Chapters 13, 14 and 15) yields results that most often link the creoles to the lexifiers, but in some cases I think the authors have only measured the forms of the words with specific meanings (Chapter 13 and especially Chapter 14), rather than the range of meanings of specific primes.

Are creoles a distinctive subgroup? I see myself as a careful and conservative scientist who takes empirical observations as a point of departure. As a scientist, I am more interested in the observable facts and theories explaining them, and not in politics or ideologies. If all tests point to creole distinctiveness, we have to accept it as an empirical fact. This is not a revolutionary statement. In fact, this is what all creolists were in agreement about until the 1990s, until a small group of non-empirical ideologists were able to change public opinion among creolists to an idea that creoles cannot be structurally distinguishable from non-creoles, without backing this hypothesis up at any point with facts about creole and non-creole grammars.
References


Complementing creole studies with phylogenetics

Eeva Sippola
University of Bremen

Creole studies is a field devoted to the empirical description and study of creole languages and to the theoretical issues related to their formation and development, which are closely tied to theories of language acquisition and change. In addition to descriptions of individual creoles and their cultural and sociohistorical contexts, the field has since its beginnings been interested in the nature of the interaction between the substrates, lexifiers, and universal forces that shape creole languages (Kouwenberg & Singler 2008). In this book, an important addition to creole studies is made by large-scale typological comparisons of a wide range of creoles and related varieties, such as restructured Afro-Hispanic varieties, from different geographical regions. This kind of comparison has not been undertaken in such a great volume until recently, although smaller scale comparisons have been made (Holm & Patrick 2007). The comparative perspective was more prominent in the early years of modern creolistics (Taylor 1956; Hancock 1975, etc.), and it is making a comeback with this volume and some other recent works of a similar nature (Bakker et al. 2011). The results obtained in the studies included in this volume shed light on the current state of creole studies and indicate some promising lines for future research.

In this commentary, I focus on three different aspects that are shared by most of the contributions in this volume and assess their place within creole studies. First, I will discuss the attempt to establish a typological profile for creoles. Second, I will take a more general approach in examining the usefulness of phylogenetic tools for the study of creoles. Third, I will focus on methodological aspects of the approach and their connections to the collection and analysis of linguistic data.

The first part of the book addresses the question of whether there is a creole typological profile. It shows some evidence that only a handful of typological structural features are sufficient to group creole languages separately from other languages, although there are exceptions (Daval-Markussen & Bakker this volume, Chapter 6). However, these features are also fairly common in languages that
are not considered to be creoles. For example, the indefinite article is derived from the numeral ‘one’ in 19% of the world’s languages in WALS, and furthermore, this is the case for all the European lexifiers. Categorically this seems to be a matter of lexifier influence, but in reality, the indefinite articles are only rarely derived directly from the lexifiers, and the authors explain their development in terms of independent grammaticalization in the creoles. Similarly, the lack of Tense-Aspect inflection is found in a great majority of the creoles and in 14% of the world’s languages, but contrary to the previous example, the lack of tense/aspect inflection is not found in their lexifiers. The use of a negative particle is also common both for creoles and other languages of the world. When examining this logic of grouping and classifying creoles, it thus seems that they form a group which shares some of their characteristic features with “older” languages. In other words, the creole group can also include languages not traditionally classified as creoles, and some other groups can include creoles, as can be seen in Figure 6.3. These overlaps raise a question about the suitability of the network model when searching for a typological profile and suggest that more work is required for a more precise method of classification which ideally could be applied to a variety of situations, taking into account the languages involved in the birth of the new varieties. In addition, although Bakker and Daval-Markussen point towards independent grammaticalization processes in creoles in connection with the features they have identified as significant for the creole group, the exact developments that lead to such a profile require further study.

It should also be kept in mind that several authors have expressed concern about the ideological motivations behind searching for an “exceptional” creole profile (DeGraff 2001, 2003, 2005; Ansaldo et al. 2007). It is therefore important to explicitly state the reasons that justify the search for a typological profile of creoles. The aim in the first part of this volume apparently is to reveal processes of language creation and change with a focus on linguistic “products of high-contact environments in specific sociohistorical settings” (Ansaldo & Matthews 2007:4). Whether these processes are connected to the expansion of a reduced pidgin to a fully-fledged language, a faster speed of grammaticalization, sociohistorical constraints, or the influence of the L1 of the people participating in the creation process of a creole, is a question that remains to be answered. At any rate, the chapters in this volume provide an important step in the process of trying to find answers to these questions using a wide set of different creoles from around the world.

Secondly, the volume demonstrates the usefulness of computational tools for providing a comprehensive picture of relationships between creoles with the same lexifier. Phylogenetic analysis makes it possible to bring together earlier work on individual languages or subgroups of creoles and get an overview of the situation. In the chapters dealing with the Dutch creoles (Bakker this volume, Chapter 10),
French creoles (Daval-Markussen this volume, Chapter 8), Iberian creoles (Sippola this volume, Chapter 11), and Afro-Hispanic varieties (Perez et al. this volume, Chapter 12), it is clear that phylogenetic tools are a useful way to corroborate or reject hypotheses made in previous research, especially when studying a selected group of languages that share a lexifier and some sociohistorical contexts. In all these chapters, the results actually support previous classifications and explanations about the development of these varieties. Some authors choose to go into detail about the birth of these varieties, while others are more cautious with regard to claims about their origin. More detailed phylogenetic studies including both different lexifiers and substrate languages would probably provide interesting insights into the relations between social forces and language structure.

In Daval-Markussen’s study of French creoles (Chapter 8), besides reproducing the established geographical groupings of these languages, the degree of similarity and difference to the lexifier was developed into a measurement of radicalness. Similar notions can also be applied to other creole groups when comparing them with the lexifier languages. Rooting the network of Iberian creoles used in Sippola (this volume, Chapter 11) with the addition of the lexifiers we get Figure 17.1, which shows the Gulf of Guinea creoles as the most distant from the lexifiers. Other languages, such as the Cape Verdean and Indo-Portuguese varieties, appear closer to the root, reflecting and confirming their closeness and continued contact with the lexifier.

The classification of restructured varieties in Perez et al. (this volume, Chapter 12) sets the creole comparisons into a wider perspective, by investigating them together with other postcolonial varieties. We could apply the degree of radicalness also to this study and note that the Atlantic creoles seem to be more radical than the Spanish varieties and the Afro-Hispanic varieties studied.

Many of the chapters (especially those by Levisen, Sippola, and their co-authors, Chapter 11–15) also provide support for general claims in creole studies. Creole languages are diverse, but show clear influence both from the lexifier and the substrate or adstrate languages. As shown in these chapters, the lexifier influence is not constrained to the lexical form, but can be observed also in structural features. The substrate influence is evident in some structural features, but also in the semantics of the lexical items. On the other hand, the different types of creoles do not cluster together in the networks if we try to apply the division of these languages in different settings according to their sociohistorical formation, such as fort, maroon, or plantation creoles. However, the measurement of radicalness might otherwise reflect the sociohistorically based classification of creoles. Creoles with prolonged contact with the lexifier seem to show fewer independent developments and more shared features. Sociohistorical circumstances have an effect on language structure and functions, as has been shown in previous studies (e.g.
Eeva Sippola

We can conclude that creoles are mixed, partially similar to each other (especially within areal subgroups), and diverse according to the specific environments and sociohistorical settings.

Regarding the methodology of large-scale typological comparisons, the *Atlas of Pidgin and Creole Language Structures* has provided researchers with an excellent database for testing their theories and analysis. Yet, despite the good quality of individual datasets in APiCS and in the questionnaires used, the variable nature of the data is evident, and due caution should be exercised in using it for phylogenetic studies. Some datasets are based on varieties that can be characterized as acrolectal (e.g. Cavite Chabacano), while others are collected from basilectal sources. Also, typological categories in these kinds of atlases are predefined according to a limited set of feature values and as such do not always capture variation in the actual language data. What should be done in a situation where two different data collection strategies were used, and no variationist studies exist for the language in question? How do the preselection of interesting features for creole studies and the genealogically unbalanced language sample affect the outcome of the comparison? The choices made, obviously influence the results on a larger scale,
and as the APiCS editors have pointed out, the properties of the languages in the atlas should not be translated into statements such as “creoles generally have…” (Michaelis et al. 2013: xxxvii), as pidgins, mixed languages, and other contact languages are also included in the database. In addition, the language data in these large comparisons probably includes a number of misinterpretations or coding errors, which might further complicate the picture. Consequently, it is necessary to always go back to the language data to interpret the results. This is exemplified clearly in Perez et al. (this volume, Chapter 12), which is the first study in which two experts on the same language scored the same features, but in different ways. One of the results of their analysis is that the classification of the varieties under study depends on the dataset and the analytical framework used. The authors note in the conclusion of their chapter that the findings should always be corroborated by socio-historical data.

Using phylogenetic networks to visualize relations between different languages and varieties is an applied solution that does not give definite answers about the languages themselves. It is therefore important to remember that detailed grammatical, lexical, and sociohistorical studies are essential for understanding the individual formation and development of each creole and its speech community. These two approaches, computational mass comparisons and careful grammatical and sociohistorical studies of individual varieties, in tandem can provide answers to central questions in creole studies on different levels. No subfield of linguistics alone will provide answers to the questions of language creation or change, but different fields can complement each other.

References


In this section, I would like to sum up the main insights and achievements of the semantics section of the book (Chapters 13–15) and give suggestions for future work. I would also like to critically reflect on the limits, problems and obstacles for utilizing phylogenetic analysis, and at the same time highlight some of the promises, and potentials for integrating lexical and semantic evidence in the study of postcolonial linguistic communities.

Unlike animal species with legs, wings, trunks, and claws, the semantics of natural languages, has no visible structures. As Saint-Exupéry’s Little Prince says: L’essentiel est invisible pour les yeux. Meanings, perhaps the single most important aspect of language – at least seen from the perspective of speakers – are invisible. They are coined to serve as a conceptual currency in a community, but only the labels of words, not the meaning of words, are directly observable. The fact that meanings are invisible does not make them less real to speakers than directly observable lexical labels, but the tools with which we approach the invisible, namely our metalanguage, is bound to play a much more important role for our analysis.

The distinction between labels and meanings is vital for a cognitive creolistics. Creoles tend to be hyper-polysemous, and as a consequence, the same lexical labels can have relatively many related meanings. Many words have gone through “earthquake lexicalizations”, resulting in a high false-friend rate across creoles, and in lexifier/creole relations. For many years, creolists have asked the question: “what is the lexifier?” We now know that another question is even more important: “what is the semantifier?” – or how is semantics constituted by speakers in postcolonial communities. To answer this question, we need in-depth analyses based on semantic fieldwork. Also, we need a metalanguage which is capable of representing meanings with a fine-grained resolution. It is important for future research to resist the temptation of doing data collection without a strong theory base. If we don’t have such a foundation, we easily end up comparing apples and oranges, or even worse: we end up finding apples and oranges in ethnolinguistic
Carsten Levisen

communities which have no interest in fruit. APiCS, for instance, does not get its semantics section entirely right, when it asks whether ‘green’ and ‘blue’ are the same in language X, Y, Z. (Huber and the APiCS Consortium, 2013). The concepts of ‘blue’ and ‘green’ are not self-evident or “natural”, but categories of visual semantics in certain European languages, which for historical reasons have found these discrete visual categories useful. To ask if speakers of other languages also distinguish blue and green, is essentially to ask: Does your semantic categorization mirror that of ours? A Russian creolistics team might have asked if language X, Y, Z distinguish goluboj roughly ‘light blue’ and sinij, roughly ‘dark blue’, because this distinction is important in Russian semantics. A research team from Melanesia would have proposed many visual descriptors unknown to Westerners’ semantics: for instance, they might have asked if the visual qualities of pigs can be described via the visual appearance of certain sea shells. In this part of the world, this is a very relevant visual-semantic feature (for recent account on the semantics of color and visuality in creoles, see Levisen, Sippola & Aragón 2016).

So, what should we do to make progress in the study of phylosemantics in a postcolonial linguistic framework? Firstly, I believe we should deliberately move slowly, and try to avoid making quick data collections based on semantically ill-formed questions, and/or European-tinged semantic categories. We should acknowledge the problem of terminological ethnocentrism, a problem which is closely linked with the metalanguage question. Two of the “quick” roads, which must be avoided, are to uncritically incorporate dictionary data, or to succumb to a version of naïve referentialism in our data collection. The aim of a traditional bilingual dictionary is to translate word Y in language X into the closest category of a European colonial language Y, and as such dictionaries are, to a very large extent, tied up with European metalanguages. Depending on whether the dictionary is French-based or English-based, the “results” will often differ, simply because English and French words differ in semantics – and word Y in language X is likely to mean something different than the English or the French word.

The other danger is to ignore the advances in cognitive semantics and cultural semantics, and return to a referentialist agenda in semantics, which seems to lend itself more easily to quick-and-dirty-data collection. It does not help to ask “what is your word for ‘brother’, ‘arm’, ‘color’, ‘love’, ‘anger’, ‘river’, etc.?”, given that many languages do not operate with these conceptual constructs (cf. Wierzbicka 2013). Using semiotic tools in the form of videos or pictures, does not solve the problem either, if the underlying non-linguistic representational agenda is still to mirror the same Anglo/English semantic categories i.e. ‘brother’, ‘arm’, ‘color’, etc. If we impose categories on a creole or any other postcolonial lect or language, because this category is relevant in European semantics, then, at best we close our eyes to semantic diversity, but at worst we also exercise conceptual colonialism on other
speakers’ semantics. This is precisely why insights from cross-linguistic cognitive and cultural semantics need to be considered carefully, and incorporated into the design features of future phylosemantic studies, and this is also why we need to do “slow research” in this area.

The semantics chapters in this book (13–15) all took their point of departure in lexical units rather than whole lexemes, following a semantics tradition from Cruse (1996: 77–78). We have primarily worked with simple meanings (semantic primes), such as ‘I’, ‘you’, ‘something’, ‘someone’, ‘good’, ‘bad’, ‘big’, ‘small’, and relatively low-complexity meanings (semantic molecules), such as ‘men’, ‘women’, ‘children’, ‘day’, ‘night’, ‘head’, ‘eyes’, etc., all solidly researched concepts in cross-linguistic semantics. In Chapter 13, Levisen and Bøegh laid the groundwork for doing basic conceptual semantics in creoles and lexifiers, without assigning any a priory primacy to the lexifier label, but treating the lexical semantics of all lects and languages in our sample as constitutive. This work could be strengthened further by adding to their sample various relevant dialectal, ethnolectal and sub-stratal exponents of the same primes. In Chapter 14, Levisen & Aragón took a first step into a slightly more complex area of semantics (semantic molecules), showing that domain factors can be important for phylosemantics. The chapter also revealed some of the problems of comparing meanings across creoles and lexifiers: What if only 3 out of 4 semantic components of a word meaning match between two languages, as for instance in the case of head in Anglo English, and hed in traditional Melanesian pidgins/creoles. On the other hand, the head-hed problem might open up new avenues for phylosemantics, helping us to move into a higher-resolution semantics, where components of semantic units, rather than whole semantic units, can be used as features in comparative research. In Chapter 15, Levisen et al questioned current dogmas in creolistics and sociolinguistics, showing how phylogenetic networks can raise awareness of classificational practices, and help us deconstruct powerful, but non-universal ideas, such as ‘language’, ‘creole’, ‘English’, ‘dialect’ and ‘variety’, and take a fresh look at the issues. Now that we have carefully surveyed the basics of lexical semantics, it is time to expand our horizons, and capture the more areal and culture-specific aspects of meaning, such as cultural keywords, areal concepts, polysemy patterns, and other culture-dependent configurations (see Levisen & Jogie 2015; Levisen 2016).

In this book, we have seen how traditional structural-typological features can be paired with lexical-semantic features – and how historical and contemporary features can be brought together. It is important to continue to develop multi-dimensional creolistics, because, as we have seen, different areas of linguistic evidence – grammatical, lexical, semantic, pragmatic features – may point in different directions, and even different kinds and types of semantics point in different directions. All natural language perspectives are equally important if we want to
understand what constitutes the verbal practices in postcolonial communities. It is important for phylogenetic research to keep refining the “feature concept”. We need to test a variety of ways in which featurization can be used as a technique to help us answer our research questions, and help us to form new questions for creolistics research, and more generally for a postcolonial linguistics.

References

Linguistics and evolutionary biology continue to cross-fertilize each other and may do so even more in the future, including in the field of creolistics

Finn Borchsenius
Aarhus University

Linguistics and biology have a long tradition of cross-fertilization and borrowing of ideas (Atkinson & Gray 2005). Even before the birth of evolutionary theory in biology in the middle of the 19th century, linguists had adopted phylogenetic thinking, and by the end of that century, linguists were conceptually ahead of biology when it came to understanding and reconstructing the process of phylogenesis. The advent of computerized analytical methods in the late 20th century, coupled with an exponential growth in the amount of primary data, particularly DNA sequences, led to an explosive development of computational methods in evolutionary biology. Many of these new methods have been adopted by linguists. At the same time, linguistic data are increasingly being used in biological studies, relating to human evolution and other fields such as the domestication history of important crop plants (e.g. Mir et al. 2013).

As evidenced by the review presented in chapter four of this book, as well as the multitude of other chapters where phylogenetic analysis is employed, computational approaches have become a standard tool for understanding the evolution of languages. The transfer of methods developed by evolutionary biologists is facilitated by the increasing amount of linguistic data available in international databases such as WALS and APiCS. The emergence of a multitude of freely available scientific software tools for performing such analyses has further enhanced the process. In biology, the use of phylogenetic information has long since expanded from evolutionary biology to a number of related fields, such as ecology and biogeography. Decreasing costs and higher degrees of automatization in modern DNA analysis have been a major driver in this process, facilitating great advances towards the completion of the tree of life. In linguistics, data similar in kind to DNA information are not available. The coding of cognates is a much more
laborious way of gathering data compared to automated DNA analysis. Linguistic
data is analysed in much the same way as the morphological features of plants
and animals were in earlier days, before DNA sequencing became a standard tool.
Nevertheless, the available amount of data is steadily increasing, as is the number
of methods that are being applied.

Creoles represent a particular challenge in terms of phylogenetic analysis, as
one may reasonably argue that a horizontal transfer poses a particularly inherent
problem in creolization. The birth of a creole language may indeed be compared
to a hybridization situation in biology, where two sets of hitherto separate traits
are brought together. Creolization also typically involves a reduction phase, where
structural and lexical diversity are strongly reduced. That situation might be paral-
leled with the phenomena of population bottlenecks or founder effects in biology.
These effects describe situations where genetic diversity is suddenly and strongly
reduced, either because of a severe reduction in number of individuals due to
disastrous events, or the foundation of a new isolated population by just a few
immigrants, e.g. birds arriving on an oceanic island. Methods have been developed
to analyse what happens to the genetic structure during such events in biology,
but it is not obvious if, or how, such methods could be transferred to creolistics.

The specific characteristics of the creolization process have led to particular
care being taken when applying phylogenetic analysis, and rightly so. For exam-
ple, creoles typically cluster together in a phylogeny because of similarities in
structure, even though they are independently derived. Also, horizontal trans-
fer is assumed to be particularly prevalent in creolization. Therefore, methods
like split networks, designed to visualize conflicting signals in the data, are often
preferred in analysis of creole language evolution (e.g. Bakker et al. 2013; Daval-
Markussen 2014). However, it should be remembered that (1) after the initial re-
duction phase, creoles evolve like any other language, which means that models
assuming a tree-like evolution may reasonably be applied once the creole has been
born, and (2) horizontal transfer and parallel evolution are an inherent part of all
evolution, also in biology. Lateral transfer is especially prevalent in bacteria and
related organisms lacking a cell nucleus, but everywhere in the tree of life parallel
evolution of similar structures takes place. In studies of DNA sequences there are
only four possible states of each character, leading to numerous similar characters
in two DNA strands being the results of parallel evolution rather than common
ancestry. However, in biology, methods are largely robust to these effects, and this
also seems to be the case when applied to linguistic data (Greenhill et al. 2009).

Seen from the perspective of a biological phylogeneticist, tree-like representations
could be used far more in the study of creole languages. Tree-based methods have
the advantage of providing simple ways of calculating statistical support values
and representing them using trees. In general, lateral transfer is likely to result in low resolution and support values, rather than erroneous tree topology. The most obvious pitfalls, such as inferring that all creoles have a common ancestor just because they cluster in a phylogenetic tree, can easily be avoided by methodological understanding and careful interpretation of results. Phylogenetic analysis, as well as other computational methods, should thus hold a great potential for advancing our understanding of the creolization process, if used with caution.

In biology, one of the strongest trends in the last two decades has been a move towards integrating large sets of data from different sources, such as analysing genetic, spatial, paleontological and climatic data together. For example, the distribution of a certain group of plants or animals may be studied taking into account both physical conditions (e.g. climate), ecological traits (e.g. habitat) and phylogenetic relationships. At the same time the current distribution may be aligned with what it was in the past, as evidenced by fossil record, past climatic conditions and major geological events to understand historical legacies and evolutionary scenarios underlying current patterns. A similar tendency can perhaps be expected in linguistics. A few papers have used phylogeographic methodology where traits and spatial distribution data are analysed together (e.g. Walker & Ribeiro 2011). Others have used an approach typical of the biological discipline termed macro ecology, which aims to understand biological processes, such as evolution or extinction, on grand spatial scales. An example is a recent attempt to understand language extinction patterns and associated drivers on a global scale (Amano et al. 2014). I predict that there is vast potential for introducing such integrative and large-scale analytical approaches in linguistics, combining phylogenetic information with geospatial and environmental information, in order to understand if the human natural environment leaves a predictable imprint on language structure.

Science is constantly on the move and interdisciplinarity is on the rise. The longstanding cross-fertilization between linguistics and evolutionary biology is an excellent example of how the meeting of different scholarly traditions can result in new achievements and mutual enrichment.

References


This book examines the place of creoles from a typological perspective using modern phylogenetic modelling tools. Exploring the similarities and differences that exist among creoles and between creoles and their input languages, the authors aim to generate new insights into persistent and at times hotly debated topics such as creole genesis and the relationships among creoles and between creoles and other languages, most specifically their input languages. The volume casts a very wide net. It investigates creoles from the Atlantic, Asian, Pacific regions and Africa, and also considers creoles associated with a range of different lexifiers (Arabic, Dutch, English, French and Iberian), common superstrates and some of the substrate inputs. But diversity does not stop there either. The authors also examine data from a range of linguistic levels, including phonological, morphological, syntactic and semantic phenomena.

The first four chapters set the stage for the research reported on in this volume. Chapter 1 introduces readers to the topics to be investigated in this volume via a broad style overview of what the research team consider to be the most important issues in research on creoles. For instance, the chapter sketches some positions on the linguistic processes and agents of creole genesis, opposing views on what constitutes a creole and whether they are a typologically distinct group of languages, and presents some of the proposals as to the reasons for their alleged uniqueness. The second chapter provides a very insightful introduction to current phylogenetic work, including discussion of its origin, development and methods. Chapter 3 ‘scores’ creoles from different regions and with different European and substrate input languages using a range of linguistic features (phonology, morphology, constituent ordering, lexicon) while Chapter 4 critically examines prior typological research on creole languages. Chapters 5 to 13 report on the different types of phylogenetic analyses carried out by the research team. The relative homogeneity of African input languages to Atlantic creoles and their impact on creoles is examined in Chapter 5 while Chapters 6, 8, 9 and 10 examine the relationships
among creoles that were influenced by French, Dutch, and Iberian languages, respectively and between them and their main European input language. Chapter 7 investigates the similarities and differences between Juba Arabic, one of the Arabic contact varieties, and its input languages. Finally, Chapters 12 and 13 explore the representation of lexical concepts across a broad range of creoles (Chapter 12) and among creoles for whom English was an important input language (Chapter 13).

Based on the results of the different scoring exercises, the authors arrive at a number of conclusions about the typology of creoles:

a. there are no linguistic properties that are unique to creole languages
b. creole grammars are not inherently simpler than those of other languages
c. however, there are broad generalizations that cover most or all creoles
d. the broad similarities between creoles are due to processes of simplification that took place prior to their genesis
e. most creoles with the same European inputs cluster together but they do not form a homogeneous group; sets of subclusters can be identified
f. creoles do not cluster with their substrate inputs

The team’s use of phylogenetic research methods can be taken to provide an empirically grounded insight into longstanding hunches and issues invoked by people working on creoles and those from related fields. Two things that I keep hearing from linguists working on other languages, including institutions that fund research on language documentation, for instance, is the either openly worded or implicitly presented assumption that all creoles are alike and are naturally simpler or unmarked linguistic entities. ¹ The analyses presented in this volume visually represented by the numerous phylogenetic trees clearly belie these received notions, showing that there is, in fact, quite a bit of variety among the languages referred to as creoles. Equally interesting is the fact that this diversity is not simply due to differences in lexifiers, but also occurs among creoles that were influenced by the same European language. Leaving aside for a moment the empirical difficulties inherent in defining notions such as simplicity and complexity, Chapter 3 in particular highlights the fact that creoles are not inherently lacking in complexity but, like all languages, show different degrees of complexity across different areas of grammar. While I don’t have hopes that the rest of the linguistic community will, on foot of this book, magically deconstruct their long-standing views about contact languages that continue to be deeply rooted in the colonial linguistic enterprise (e.g. Mühleisen 2002), it definitely presents another important piece of evidence towards confirming that creoles are natural languages.

¹. And thus, due to their alleged lack of unique features, are of lesser interest to structural, documentation and typological research.
Reading the different contributions, I was also struck by a number of issues. For reasons of space, I will limit my discussion to only a few that seem most pertinent to me right now. My main caveat concerns the choice of features. While the team made a concerted effort to draw on a wide range of linguistic features, I cannot stop but feel uneasy when looking at the various phylogenetic trees and the wide-ranging conclusions that are drawn based on them. For once, while the authors identified various problems with previous, in many cases more localized typological approaches, I am not convinced that the analyses in this volume resolve the main issue that plagues all typological studies: representativeness. The selected features, though enriched by features used for broad cross-linguistic comparisons (e.g. WALS), remain abstract and in many ways closely overlap in type and kind with those used in previous attempts. As a result, the authors are not able to report any ground breaking new results as such, but are mostly confirming or adding some (minor?) detail to previous typological REPL based however on newer data sets for the creoles (e.g. APiCS database (Michaelis et al. 2013)) and using enhanced modelling methods.

Second, and more fundamentally problematic are the assumptions underlying this enterprise. I find it surprising or actually disconcerting that after decades of research on a wide range of aspects relating to language it is apparently still considered perfectly normal to assume that the essence of a language can be captured based on a narrowly defined set of structural features. While it is true, and regrettably so in my opinion, that the bulk of linguistic research continues to be preoccupied with a narrow set of phonological, morpho-syntactic, semantic and lexical etc. phenomena mostly derived from a decontextualized set of sentences or word lists obtained using behaviourist data collection methods (e.g. elicitation), I am not sure that it is simultaneously warranted to assume that these features are in any way central to defining languages and central to defining relationships between languages. This kind of reasoning seems to be highly problematic and indicative of a lack of critical reflection among linguists about their endeavors, including its historical roots and developments. If we are interested in getting new insights into the relationships between languages, should we not first deconstruct common assumptions about what constitutes a language and consequently our own endeavors (linguistic research), second consider what is central or indicative about human communication systems and how we can study them and third

---

2. See Kouwenberg (2010a & b) for a discussion of the usefulness of WALS features (Dryer et al. 2013) in typological research on creoles.

3. The interested reader might start with reading some recent publications about what has come to be known as colonial linguistics (e.g. Errington 2008; Makoni & Pennycook 2007) and to consider recent discussions in the area of language documentation (e.g. Childs et al. 2014).
invest more time and effort in the reconstruction of the contact settings (and the histories of speakers of languages)? I submit that this is particularly important when dealing with non-European languages such as creoles that have had to suffer from centuries of wilful and accidental distortion by mostly Euro-American (trained and untrained) linguists. My feeling is that uncritical feature crutching is perpetuating and maybe also infusing new life into received notions about language simply by using more sophisticated tools to do it. Just to clarify: I have no problem with feature crutching, but you have to critically assess what you are crunching and to what end. For instance, why do pragmatic and sociolinguistic aspects of language never feature in typological work?

Every time I engage with speakers of Maroon languages in Suriname, French Guiana or beyond in both mundane and more formal settings in urban and rural contexts through oral communication or e-media, I am always struck by the vast amount of meaning making that takes place that has to date remained firmly below the surface of most of our research activities. In fact, research into these aspects of language which require more than just passing knowledge of a few decontextualized features continues to be overtly and covertly discouraged because it is unsettling as it requires engaging with and developing new kinds of research modalities and much closer and democratic collaboration with the speakers of the languages that we study. But I believe that embarking on this unknown adventure will fundamentally expand our current thinking about creole languages and languages in general and will help us to rise above impoverished feature matrixes and dubious word lists. In order to get there, however, we need to critically examine Chomskyan notions of (linguistic) competence that continue to dominate much of what’s going on in the field of linguistics and transition towards Hymesian notions of communicative competence and beyond.

Chapters 12 and 13 make an attempt at breaking away from the stranglehold of typical structural linguistic concerns by focusing on semantics or what the authors call patterns of lexicalizations. Instead of using the Swadesh list, the authors are basing their work on current lexico-semantic approaches. However, oddly enough, the comparisons focus for the most part on the nature of the etymological shapes of words and their most basic meaning rather than on a full analysis of the actual meanings and functions of the words (Huttar et al. 2007). So again, in order to enable broad comparisons, the languages have been reduced to simplified objects. Can we not imagine a more sophisticated way of doing comparisons that will not require erasure of potentially important information?

4. In this regard, it is equally important to take on board Kouwenberg’s (2010a & b) remarks on whether it makes sense to compare creoles to existing typological databases that are unsuitable for dealing with contact languages.
A final point about features relates to the notion of stable linguistic features invoked in Chapter 5. Here a selection of creoles and African languages (and European languages) are compared “to test the claim that the African languages involved in the colonial settings, and hence in the creation of various creole languages, were typologically homogeneous.” However, instead of selecting a range of features, curiously, the authors base their comparison on what they describe as features that are “relatively stable over time” because they “were most likely inherited rather than borrowed, these are assumed to represent well-suited markers of distant genealogical relationships”. If creoles arose from creative and in many ways idiosyncratic situated interactions between different sets of linguistic practices employed by human agents to fill their communicative needs, as most people working on creole genesis nowadays believe, then why would either stable or unstable linguistic features be privileged over other features in creole genesis? It seems to me that both features have an equal chance of being propagated in this kind of a setting and thus the assumption that the presence or absence of certain kinds of linguistic features is indicative of a certain kind of relationship between creoles and their input languages is problematic.

There are also other kinds of reductionisms. For instance, in some places languages are depicted as a living organisms and sociolinguistic processes such as language contact are likened to biological processes (“lateral gene transfer”). Are we regressing to the view that languages are bounded objects and that signs of variation, change and ‘outside’ influence are a sign of degeneration? Even if the book uses analytical tools derived from biology, there is no need for such unwarranted metaphoric extensions. Languages are very different from cells and bodies and any attempt at reheating overcooked parallels between them are not only running the risk of misrepresenting language related matters but are also propagating views from the colonial and nationalist era. Another kind of reductionism relates to the presentation of research on creoles such as creole genesis, for instance. A rather complex and multifaceted area of investigation is reduced to a few somewhat eccentric theoretical approaches (superstratist versus relexification accounts; abrupt versus life-cycle accounts of creole formation) in order to generate a few succinct hypotheses that can serve as a basis for comparative exercises. This seems rather crude especially because there is little reference to existing critiques about these approaches, making it unclear what the research in this volume actually contributes to the discussion. And on the other hand, because there are few references to more nuanced approaches, it represents research in this area as somewhat simplistic.

My final point relates to the interpretation of phylogenetic trees. On the one side, these visual representations of relationships between languages are great because they allow us to grasp a great amount of complexity pretty much
instantaneously. However, isn’t this also part of the problem? When you look at the trees in more detail, it is in fact not quite clear how to interpret them reliably because the multitude of intersecting lines is actually quite hard to separate out. The more I look at the trees in the book, the less clear I am about how to derive significance from these webs of lines.

In summary, while I commend the authors on the huge amount of work that they put into the analyses that figure in this volume, I am uncomfortable with the uncritical glossing over a number of pertinent issues and the lack of willingness to push beyond the narrowly defined limits of current typological work.

References


Languages index

A
Adamawa 48, 147, 149, 151, 153–154, 158, 161, 165
Adamawa-Ubangi see Adamawa
West African languages 8, 46, 141–168
Afrikaans 6, 106, 225, 298, 305
Afro-Asiatic languages 72, 148, 153, 161
Afro-Chinchano Spanish 270, 273, 276, 278
Afro-Choteño Spanish 270, 272, 276, 278–282
Afro-Hispanic 4, 252, 269–287, 375, 377
Afro-Hispanic pidgin 272
Afro-Peruvian Spanish 273
Afro-Portuguese 252, 272
Afro-Yungueño Spanish 270–271, 277–287
Aikanã 61–62, 72
Akan 47, 146, 148, 225
Alpine Romance languages 38
Ambon Malay 49, 126–127, 148, 165
Amerindian 84, 224, 226
Amharic 108
Andean Spanish 271, 274, 279
Anglo-Surinamese 303–304, 306, 308, 310
Angolar 14, 47, 49, 73–74, 80, 82, 95, 105, 118, 127–128, 135, 148, 223, 245, 250, 252, 254, 258–259
Annobon 47, 49, 82, 105, 127, 135, 148, 245
Annobonesee see Annobon see also Fa d’Ambô
Antillais 108, 110, 179
Arabic creoles 3, 8, 13, 25, 86, 213
Arabic languages 209, see also Egyptian Arabic, Gulf Arabic Pidgin, Juba Arabic (Creole), KiNubi Arabic Creole, Nubi Creole Arabic, Sudanese Arabic, Syrian Arabic
Arikapú 61–62, 71–72
Army creoles 11
Asian Portuguese 89, 223, 244, 247
Assamese 118, 199, 210
Atayal 26
Atlantic English creoles 12, 225
Atlantic French 179
Aymara 61–62, 271, 274
B
Bambara 47, 146, 148
Bantu 8, 14, 19, 48, 118, 120, 147–154, 156–158, 162, 166, 223, 250, 261, 274
Batavia Creole 49, 148, 248, 250–252, 254, 261
Batavia Creole Portuguese 148
Bena 47, 148
Benuere 225
Bislama blong Vila 350, 354
Bolivian Spanish 282
Brahui 73, see also Dravidian
Brava Cape Verdean Creole 250
British English 325, 351
| C | Cameroon English Creole 135, see also Camtok
Cameroon English Pidgincreole 108, 122, see also Camtok
Camtok 108
Cape Verdean 105, 118, 245, 250, 253, 262–263, 377
Cape Verdean Creoles 72, 255, 257–258, 284
Cape Verdean (Creole) of Brava 250, 254
Cape Verdean (Creole) of Santiago 250, 254, 258–259
Cape Verdean (Creole) of São Vicente 250, 254, 259
Cape Verdean Creole Portuguese 284 see also Kabuverdiano
Caribbean creoles 25, 93, 108, 121, 178, 182, 184, 262, 345, 356, 370
Caribbean English creoles 108–110, 121
Caribbean Papiamentu 249
Carriacou 189
Casamancese Creole 250, 257–259
Cavite Chabacano 250–251, 254, 274, 282, 378
Cayubaba 61–62, 72
Celtic 15, 19, 128
Central Southwest African 142, see also West African languages
Chavacano 49, 343, see also Chabacano
Chinese 93, 361
Chinook Creole 126, 134, see also Chinuk Wawa
Chinook Jargon 7, 87, see also Chinuk Wawa
Chinuk Wawa 8, 22, 84, 87, 126, 130, 132, 148
Creolese 256
Cross-River languages 154, 163
Cuban Spanish 15, 278, 280–282
D | Dagbani 47, 146, 148
Danish 6, 220, 253, 360–364
Defaka 47, 146, 149, 157
Dinka 198–199, 202, 209, 211–212
Diola 129, 131, 146, 149, 155–159, 163–166
Diola-Fogny 157–158, 165
Diu Creole Portuguese 13, 17
Diu Indo-Portuguese 250–251, see also Diu Creole Portuguese
Dogon 147, 149–150, 153–154, 157–158, 165
Dominican French Creole 104, 118, 176, 179, 181
Doyayo 47, 48, 146, 149
Dravidian 7, 73, 94
Dungku 338
Dutch creoles 13–14, 70, 72, 82–83, 87, 89–90, 105, 118, 128, 167, 219–237, 244, 298–299, 305–307, 310, 370, 376, see also Virgin Islands Creole Dutch
E | Edoid 147, 149, 153
Efik 146, 149, 153–156, 159, 163–164, 166
Egyptian Arabic 47, 148, 154, 162
Engenni 146, 149
English creole of Vanuatu 17, see also Bislama
Ermiteñó 105, 245
Esperanto 123
Ewe 8, 149
Ewoondo 146, 149
F | Fa d’Ambó 127, 250, 252, 254, 257–259
see also Annobon
Finnish 13, 125, 127
Fon 8, 149, 157
Fongbe 19
Franglais 309
Frankofon Bislama 350
G | Gà 149
Gã-Dangme 149
Gavião 61–62, 72
Gawwada 93
Gbe 8, 19, 76, 132–133
German 83, 90, 108, 122, 148–149, 154, 160, 348
German creoles 90, 108
Germanic languages 46, 69, 90, 149, 153–154, 160–162, 164, 302, 325, 327, 336
Old High German 93
Ghanaian Pidgin English 261
Global Englishes 309, 348, 351
Grebo 47, 146, 149, 154
Greek 11, 93, 125
Grenada French Creole 17
Guiné 105
Guinea Creole Portuguese 105
Guinea-Bissau Creole see Guinea-Bissau Kriyol
Gulf Arabic Pidgin 213, see also Arabic creoles
Gullah 47, 49, 149, 160
Guarani Kriol 350
Guyanese English Creole 122
Guyanese French Creole 157
Haitian French Creole 84, 104, see also Haitian Creole
Hawaiian Creole see Hawai’i Creole
Hawaiian Creole English see Hawai’i Creole
Hawaiian Pidgin English see Hawaiian
Hebrew 125
Highland Andean Spanish 274
Highland Ecuadorian Spanish 272
Hiri Motu 87, 90, 94, 130, 132, 150, 158, 161–162, 165, 199, 210
Hmong-Mien 124
Hmong Ngua 124
Iberian creoles 80, 105, 225, 241–264, 276, 371, 377–378
Ibero-Asian creoles 81, 247, 261
Iberoromance creoles 284, 294, 302, 307, 310, 325–326, 333
Icelandic 48
Igbo 146, 150, 154, 225
Igboi 147, 150, 153
Eastern Ijo 14, 84, 87, 222, 225
Ijo-Kokoloua 150, 153–157, 159, 163–164, 166
Indian Creoles 94
Indian Ocean creoles 8, 46, 48, 89, 95, 156, 178–179, 182, 184
Indian Ocean French 12, 105, 257
Indian Sign Language see Plains Indian Sign Language
Indigenous Englishes 348
Indo-European 36, 43, 46, 120, 149, 151–153, 160–165, 199, 210, 371
Indo-Portuguese 7–8, 108, 110, 244, 246–249, 252–253, 259, 262, 264, 377
Indonesian 73–74, 120
Italian 19, 108, 122
Jahai 333
Jamaican 82–83, 94, 105, 118, 150, 157–158, 167, 358,
Jamaican Creole English 8, see also Jumiekan
Jamsay 47, 146, 150
Japanese 14, 27, 148, 150, 154–155, 161, 163, 286, 371, see also Yilan Creole
Japanese, Taiwan Creole Japanese
Japlish 309
Japonic languages 44, 150, 153, 161
Juba Arabic Creole 3, 87, 126, 150, 163, 193–214, 260
Judeo-Spanish 274, 282
Jutlandic 361
K
Kabuverdianu 47, 49, 150, see also Cape Verdean
Kalabari 230
Kande 47, 146, 150
Karipuna 47, 49, 150, 199, 209–212
Ket 93
Kigiriyama 47, 146, 150
KiKongo 47, 150, 167, 167, 261
Kikuyu 47, 146, 150
KiMbindu 14, 223
KiNubi 47, 49, 72, 108, 128, 150, 161, 165
KiNubi Arabic Creole 122
Kisi 47, 146, 150
Kong 47, 146, 150
Konkani 248
Konkani-Marathi 248
Korlai Creole 49, 73–74, 118, 125, 128, 150, 199, 209–214, 223, 251, 254
Korlai Portuguese Creole 91, 258
Koromfe 47, 146, 150
Koromu 349, 358, 364
Koyra Chiini 72, 74
Krâyol la Lwizyàn 298, 306–308, see also Louisiana creole
Krio 105, 118, 128, 150
Kriol 49, 150, 345, 362–362, see also Roper River Kriol
Kriyol 157, 167, see also Guinea-Bissau Kriyol
Kru 48, 147, 149, 153–154
Kuanua 358
Kunama 108
Papuan 8, 349
Peninsular Spanish 274, 278–282, 286–287
Persian 93
Peruvian Spanish 273
Philippine languages 8, 85, 87
Philippine Creole Spanish 8, 84, 86, 244
Philippine Spanish 274, 278–282, 287
Pichi 15, 47, 49, 151, 152
Pidgin Hindustani 87
Pidgin Italian 108
Pidgin Madame 213
Pijin 346, 354–355, 358, 364 see also Solomons Islands
Pijin Pirahã 73, 74
Pitcairnese 109–110, see also Pitkern
Pitkern 350, see also Pitcairnese
Plains Indian Sign Language 6, 8, 130, 132, 151
Plantation creoles 24, 26, 377
Portuguese pidgin 244–245, 248
Principeño 47, 49, 105, 151, 245, 250, 254, 258–259
Q Quechua 61–62, 72, 274
T Tagalog 274 Taglish 309 Tahitian 128 Taiwan Creole Japanese 87–88 Tamil 87, 258
Temne 47, 146, 152
Ternate Chabacano 148, 250, 251, 254, 256, 258–259, 274
Thai 120
Tok Pisin bilong Mosbi 350, 354
Torres Strait Creole 346, 350, 354
Trinidad French Creole 176, 181–182
Trinidadian Creole 353, 357
Tugu Malay Creole 161
Tupi 152–154, 161
Tupi-Guarani 152–154, 161
Tupinamba 152, 154–155, 371
Turkish 127
Turku 213

U
Umbundu 47, 146, 152
Unserdeutsch 7, 49, 108, 152, 160
Upper Guinea creoles 8, 245–247, 249, 255, 257, 262, 264
Upper Guinea Portuguese creole 89
Urban Bislama 331, 346, 350, 354, see also Bislama
Uru 61, 72

V
Vai 152
Vernacular Brazilian Portuguese 6
VICD, see Virgin Islands Creole Dutch
Vincentian 47, 49, 152, 160
Virgin Islands Dutch Creole cf. Virgin Islands Creole Dutch

W
Wari’ 61–62, 71
Wolof 47, 152, 157–158, 167
World Englishes 309, 348, 364

Y
Yao 47, 146, 152, 155–156, 159, 163–164, 166
Yilan 131–132, 152, 161, 165, 371, see also Yilan Creole Japanese
Yilan Creole Japanese 8, 26, 88, 130
Yimas-Arafundi Pidgin 120
Yoruba 47, 125, 129, 131–132, 146 152–156, 159, 163
Yoruboid 147, 152–153

Z
Zamboangueno 73–74, 84, 109, 118, 152, 161
Zulu 47, 146, 152
People index

A
Abah 7, 9–10, 13–14, 18, 20–22, 86, 134, 193, 195, 212–214, 298
Adam 6
Adamson 82
Addison 6
Aikhenvald 50
Albu 143
Allen 188
Alleyne 185, 299, 347–348
Altschuler 40
Ambouroue 151
Andersen 199, 202, 331
Andersson 310
Angelo 350
Ansaldo 10, 16, 18, 20–22, 134, 193, 348, 376
Aragón 2, 4, 295, 309–310, 315, 319, 352, 382–383
Arends 17–18, 22–23, 88
Ariza 364
Asial 364
Atkinson 36–39, 43, 385
Avise 43
B
Baggioni 319
Bailey 161, 348
Baissac 142
Bajao 319
Baptista 84, 90, 249, 298, 319
Barbançon 38
Barbosa 152
Berg 224
Bergman 148, 198–199
Bergland 37
Bergvall 150
Berger 148
Bhattacharyya 199
Biberauer 298
Bickerton 9, 19, 21–23, 26, 88, 107–108, 110, 121, 134, 185, 221, 270
Bisang 263
Bizzri 213
Blair 319
Blench 142, 147, 298
Blommaert 347
Bloomfield 82, 185
Boegh 2–4, 64, 141, 144, 193, 223, 293, 316, 318–319, 325, 346, 352, 383
Bole-Richard 149
Bollée 18
Boooj 84, 195
Borchsenius 2, 4, 35, 59, 385
Borges 21, 168, 317
Borouah 151, 199
Bouckaert 43, 60
Bowern 38, 42, 295
Brauner 148
Bresnan 7, 86
Brocke 168
Bromhead 295, 309, 336
Brousseau 85, 149
Brown 67
Bryant 24, 133, 219
Burenholt 333
Burrell 21, 30, 354
Bybee 88
C
Cabral 15
Campbell 41, 142
Canagarajah 348
Cardoso 48, 241, 244, 249, 252, 258
Carlin 24
Carlson 152
Carpooran 319
Casado-Fresnillo 269, 274
Chang 298
Charpentier 350
Clements 10, 48, 83, 87, 91, 150, 199, 241, 245, 249, 258, 262, 378
Coelho 6, 142, 244
Collin 36
Collins 161
Comrie 10, 64, 81–82, 295
Condon 95  
Corbett 200, 203–204, 207  
Corne 88, 179, 183, 319  
Corum 9  
Cowell 199  
Crazzolara 199  
Crevels 61–62  
Croft 249, 252, 277  
Crowley 298, 319, 350  
Cruse 383  
Cruz 151  
Cuba 273  
Cunningham 149  
Cysouw 40, 64, 120–121, 143, 165, 168, 372  

deVette-Chee 350  
Dicks 319  
Dijkhoff 319  
Dimmendaal 48, 142, 194, 198, 201–202, 209, 223  
Dixon 39  
Doke 152  
Dolgopolki 162  
Doneux 148  
Donohue 44  
Dress 143  
Drummond 39  
Dunn 38, 40, 60, 63–66, 197, 200, 243, 295  
Durrelman-Tame 319  
Duthie 149  
Dutton 88, 150, 152, 199, 350  
E  
Eades 350  
Edwards 347, 353  
Ehrhart 7, 152, 180, 183, 298, 319  
Ellison 40  
Emenanjo 150  
Enfield 45, 295, 315, 331, 339  
Erdem 42  
Escalante 319  
Essono 149  
F  
Faraclas 22, 151, 298, 319, 340  
Farquharson 83, 94, 157, 167, 298, 319, 340  
Fattier 298, 319, 340  
Fauconnier 316  
Felsenstein 36, 59  
Ferguson 10  
Fernández 245, 250  
Ferraz 241, 245–249, 262  
Fischer 199  
Fon Sing 17  
Forman 152, 319  
Forster 38  
Fortunato 38  
Fortune 152  
Frank 199–200, 202  
Franklin 358  
Fridhiksson 48  
Funk 150  
Funnell 152  
G  
García 319  
Gavin 43–44  
Geisler 42  
Gil 50  
Givón 10, 21, 24, 26, 50, 88, 126–127  
Glock 152  
Goilo 298, 319  
Goldshtein 3–4, 168, 193, 345  
Gong 44  
González de la Higuera Rojo 168  
Good 60, 80, 82, 152, 199, 298  
Gooden 82  
Goodman 177–179, 182, 271  
Gordon 350  
Govor 364  
Goyette 49, 161, 193, 195–196, 213–214  
Granda 272  
Grant 14, 37, 85–87, 93, 105, 121, 298, 319  
Gray 36–39, 43, 60, 243, 385  
Greenberg 9, 90, 141, 340  
Greenhill 40, 42, 63, 386  
Griffiths 38  
Grollemund 150  
Guéron 90  
Guillemín 121, 298, 319, 340  
Guthrie 147, 156  
H  
Haeckel 36  
Hagemeijer 241, 249, 261, 298  
Hagen 357–358, 364  
Hall 45, 104, 133–135, 319, 357–358  
Ham 82  
Hammarström 87–88, 370, 372  
Hancock 10, 41, 94, 104–105, 177, 180–181, 241, 245, 309, 375  
Harry 228, 230  
Hasegawa 40, 44  
Haspelmath 13, 15, 29–30, 45, 57, 65, 80, 89, 93, 120, 126, 143, 148, 162, 193, 200, 252, 256–258, 269, 275, 277  
Hay 16, 350  
Hearn 319  
Heath 150  
Heggarty 41, 60
Heine 13, 50, 87, 91, 127, 133, 142, 154, 287
Hennig 36
Hesseling 6, 319
Hewson 149
Hill 317
Hinskens 221
Hock 142
Holbrook 177
Holden 38
Hollis 199, 202
Holm 6, 8, 25, 60, 72–73, 81, 85, 90, 92, 104, 111, 118, 121–123, 126–127, 142, 183, 221, 235, 238, 244, 248, 262, 271, 280, 294, 305, 348, 372, 375
Holman 40, 46, 67, 142–144, 162, 165
Holton 43
Honkola 40
Hookomsing 93–94
Huber 11, 90–91, 177, 187, 317, 328, 382
Hudu 148
Huelsenbeck 38
Hulsmann 222
Hussainmiya 87
Hutter 392
Hyman 148

I
Intumbo 294

J
Jackson 152
Jacobs 11, 84, 87, 241, 249, 271
Janson 135
Jansson 43, 136
Jenewari 149
Jessurun 340
Jogie 297–298, 309, 340, 356, 383
Jordan 38, 295
Jørgensen 347
Joseph 142, 340
Josselin de Jong 226, 228, 319
Jourdan 22, 25, 84, 346, 350
K
Kachru 348
Kamarah 152
Kamholz 40
Kanas 350
Kandybowicz 151
Kanj 42
Kaufman 18, 20, 39, 41, 45, 48, 142, 270
Kawu 151
Keesing 354
Kiessling 23
Kihm 149, 157, 185, 244
Kirapeasi 358
Kirkpatrick 348
Klein 22, 79–81, 134
Klingler 298
Koelle 141
Koptjevskaja-Tamm 315, 332
Kortmann 118–119, 231, 263, 287, 347, 370, 372
Kotey 149
Kouwenberg 82–85, 87, 90, 128, 148, 157, 221–226, 228, 236, 298, 305, 319, 375, 391–392
Kraft 44
Krämer 6, 142
Kraska-Szenk 331
Kriegel 298, 319
Kropp-Dakubu 149
Kulick 334
Kusters 16, 21
Kuteva 10, 50, 91, 127, 133, 287

L
Laberge 26
Labov 41
LaCharité 82, 85
Ladhams 83
Lang 319
Laporte 194, 202
Law 89, 110
Lax 150
Lee 40, 44, 358
Lefebvre 8–9, 14, 18–19, 21, 46, 48, 83, 89, 142, 149, 157, 185, 294, 348
Lehtinen 69
LePage 17
Levey 285, 287
Levinson 295
Lim 348
List 42, 242–243
Lister-Turner 151
Ljuba 202
Llamado 319
Lloyd 283
Löfgren 364
Longobardi 64
Luis 83–85
Lupyan 16

M
Maclagan 350
Maddieson 80–82
Maddison 242–243
Magens 6
Mahmoud 198
Majid 295, 331, 333
Makoni 347, 391
Manfredi 199, 201, 203, 206, 208, 213
Mansilla 319
Marchese 22, 25–26
Markey 106–108, 110, 134
Marx 37
Maslova 143
Mather 18
Matissoff 309, 315
Matthews 16, 20, 22, 348, 376
Maurer 148, 151, 257
Maurits 38
McMahon 38, 41, 60, 63–64
Meakins 350
Meijer 6
Melio 249
<table>
<thead>
<tr>
<th>Name</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saunders</td>
<td>65</td>
</tr>
<tr>
<td>Schadeberg</td>
<td>152</td>
</tr>
<tr>
<td>Schlegel</td>
<td>36</td>
</tr>
<tr>
<td>Schleicher</td>
<td>36, 41</td>
</tr>
<tr>
<td>Schlyter</td>
<td>36</td>
</tr>
<tr>
<td>Schmidt</td>
<td>41–42</td>
</tr>
<tr>
<td>Schneider</td>
<td>185, 347–348</td>
</tr>
<tr>
<td>Schuchardt</td>
<td>6, 142, 244</td>
</tr>
<tr>
<td>Schultze-Berndt</td>
<td>350</td>
</tr>
<tr>
<td>Schweger</td>
<td>151, 185, 249, 271–272, 298, 319, 340</td>
</tr>
<tr>
<td>Sebba</td>
<td>348</td>
</tr>
<tr>
<td>Segorbe</td>
<td>148</td>
</tr>
<tr>
<td>Senft</td>
<td>331</td>
</tr>
<tr>
<td>Sessarego</td>
<td>2, 4, 269, 271–274, 276–281, 283–286</td>
</tr>
<tr>
<td>Seuren</td>
<td>82, 88–89, 223, 270</td>
</tr>
<tr>
<td>Shakespeare</td>
<td>20</td>
</tr>
<tr>
<td>Sharifian</td>
<td>316, 350</td>
</tr>
<tr>
<td>Shimizu</td>
<td>151</td>
</tr>
<tr>
<td>Shnukal</td>
<td>22, 25–26, 319, 350</td>
</tr>
<tr>
<td>Shosted</td>
<td>21</td>
</tr>
<tr>
<td>Sicoli</td>
<td>43</td>
</tr>
<tr>
<td>Siegel</td>
<td>319, 346, 348, 350</td>
</tr>
<tr>
<td>Singler</td>
<td>88–89, 141–142, 285, 290, 348, 375</td>
</tr>
<tr>
<td>Sørensen</td>
<td>221</td>
</tr>
<tr>
<td>Spagnolo</td>
<td>199, 202, 206</td>
</tr>
<tr>
<td>Speedy</td>
<td>12, 180, 182–183, 340</td>
</tr>
<tr>
<td>Spolsky</td>
<td>347</td>
</tr>
<tr>
<td>Sprauve</td>
<td>221</td>
</tr>
<tr>
<td>Sreedhar</td>
<td>87, 199</td>
</tr>
<tr>
<td>Stanwood</td>
<td>294, 297–298</td>
</tr>
<tr>
<td>Stassen</td>
<td>51, 91–92, 128</td>
</tr>
<tr>
<td>Staudacher-Valliamé</td>
<td>151</td>
</tr>
<tr>
<td>Steel</td>
<td>38</td>
</tr>
<tr>
<td>Stein</td>
<td>221, 233</td>
</tr>
<tr>
<td>Steinkrüger</td>
<td>85, 88, 250, 319</td>
</tr>
<tr>
<td>Stennes</td>
<td>149</td>
</tr>
<tr>
<td>Sterzuk</td>
<td>348</td>
</tr>
<tr>
<td>Stevens</td>
<td>221</td>
</tr>
<tr>
<td>Stewart</td>
<td>348</td>
</tr>
<tr>
<td>Stolz</td>
<td>106, 219, 221, 233</td>
</tr>
<tr>
<td>Suzuki</td>
<td>249</td>
</tr>
<tr>
<td>Swadesh</td>
<td>37, 39, 66, 179, 222–223, 226, 295–296, 392</td>
</tr>
<tr>
<td>Swoford</td>
<td>243</td>
</tr>
<tr>
<td>Swolkin</td>
<td>319</td>
</tr>
<tr>
<td>Szmrecsanję</td>
<td>118–119, 231, 370, 372</td>
</tr>
<tr>
<td>T</td>
<td></td>
</tr>
<tr>
<td>Tabouret-Keller</td>
<td>17</td>
</tr>
<tr>
<td>Tadmor</td>
<td>93</td>
</tr>
<tr>
<td>Taylor</td>
<td>12–13, 104–107, 110, 121, 134, 244–245, 340, 375</td>
</tr>
<tr>
<td>Thomason</td>
<td>14, 18, 20, 39, 41, 45, 48, 142, 270, 348</td>
</tr>
<tr>
<td>Thompson</td>
<td>244</td>
</tr>
<tr>
<td>Tobler</td>
<td>150, 199</td>
</tr>
<tr>
<td>Tomas</td>
<td>349, 364</td>
</tr>
<tr>
<td>Toronzioni</td>
<td>151</td>
</tr>
<tr>
<td>Tosco</td>
<td>213</td>
</tr>
<tr>
<td>Towner</td>
<td>42</td>
</tr>
<tr>
<td>Travis</td>
<td>298, 319</td>
</tr>
<tr>
<td>Troy</td>
<td>350</td>
</tr>
<tr>
<td>Trudgill</td>
<td>16, 81</td>
</tr>
<tr>
<td>Tryon</td>
<td>350</td>
</tr>
<tr>
<td>Tucker</td>
<td>199</td>
</tr>
<tr>
<td>Turégano Mansilla</td>
<td>319</td>
</tr>
<tr>
<td>Turvey</td>
<td>44</td>
</tr>
<tr>
<td>V</td>
<td></td>
</tr>
<tr>
<td>Valdman</td>
<td>81, 84–85, 150, 178, 319</td>
</tr>
<tr>
<td>Van der Voort</td>
<td>17, 61–62, 220–221, 298, 319</td>
</tr>
<tr>
<td>Van der Wal</td>
<td>150</td>
</tr>
<tr>
<td>Vandeputte-Tavo</td>
<td>350</td>
</tr>
<tr>
<td>Van Diggelen</td>
<td>221</td>
</tr>
<tr>
<td>Vanhove</td>
<td>315</td>
</tr>
<tr>
<td>Van Minde</td>
<td>148</td>
</tr>
<tr>
<td>Van Name</td>
<td>185</td>
</tr>
<tr>
<td>Van Rossem</td>
<td>220–221</td>
</tr>
<tr>
<td>Van Sluijs</td>
<td>221</td>
</tr>
<tr>
<td>Vasconcelos</td>
<td>6</td>
</tr>
<tr>
<td>Vazquez Veiga</td>
<td>245</td>
</tr>
<tr>
<td>Veenstra</td>
<td>9, 84, 107, 134, 219, 298</td>
</tr>
<tr>
<td>Velupillai</td>
<td>9, 13, 79, 298, 348, 350</td>
</tr>
<tr>
<td>Verhaar</td>
<td>319</td>
</tr>
<tr>
<td>Verker</td>
<td>242, 254</td>
</tr>
<tr>
<td>Versteegh</td>
<td>167</td>
</tr>
<tr>
<td>Vilsaint</td>
<td>319</td>
</tr>
<tr>
<td>Vinson</td>
<td>142</td>
</tr>
<tr>
<td>Vintila-Radulescu</td>
<td>178</td>
</tr>
<tr>
<td>Vogt</td>
<td>37</td>
</tr>
<tr>
<td>Volker</td>
<td>7, 152</td>
</tr>
<tr>
<td>Voorhoeve</td>
<td>150, 199</td>
</tr>
<tr>
<td>W</td>
<td></td>
</tr>
<tr>
<td>Walker</td>
<td>43, 387</td>
</tr>
<tr>
<td>Wallace</td>
<td>35</td>
</tr>
<tr>
<td>Warnow</td>
<td>41, 60, 66, 68, 71, 197, 209</td>
</tr>
<tr>
<td>Watson</td>
<td>87, 150, 199</td>
</tr>
<tr>
<td>Weinreich</td>
<td>347</td>
</tr>
<tr>
<td>Wekker</td>
<td>82, 270</td>
</tr>
<tr>
<td>Wellens</td>
<td>86–87, 150, 213</td>
</tr>
<tr>
<td>Welmers</td>
<td>148–149, 152</td>
</tr>
<tr>
<td>Whinnom</td>
<td>12–13, 175, 244</td>
</tr>
<tr>
<td>Whorf</td>
<td>162</td>
</tr>
<tr>
<td>Wichmann</td>
<td>37, 40, 46, 60, 65, 142–144, 162, 165, 242–243, 277</td>
</tr>
<tr>
<td>Wiley</td>
<td>37</td>
</tr>
<tr>
<td>Williamson</td>
<td>142, 147, 150</td>
</tr>
<tr>
<td>Winer</td>
<td>298, 319</td>
</tr>
<tr>
<td>Winford</td>
<td>45, 88, 298</td>
</tr>
<tr>
<td>Wolf</td>
<td>309</td>
</tr>
<tr>
<td>Wolff</td>
<td>151</td>
</tr>
<tr>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Yakpo</td>
<td>151, 319</td>
</tr>
<tr>
<td>Yokwe</td>
<td>199</td>
</tr>
<tr>
<td>Z</td>
<td></td>
</tr>
<tr>
<td>Zenk</td>
<td>148</td>
</tr>
<tr>
<td>Zhang</td>
<td>44</td>
</tr>
<tr>
<td>Zipp</td>
<td>350</td>
</tr>
<tr>
<td>Place</td>
<td>Pages</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>A</strong></td>
<td></td>
</tr>
<tr>
<td>Central Africa</td>
<td>7–8, 48, 125, 128, 151, 154, 157</td>
</tr>
<tr>
<td>East Africa</td>
<td>7, 11, 72, 118</td>
</tr>
<tr>
<td>South Africa</td>
<td>152, 298</td>
</tr>
<tr>
<td>West Africa</td>
<td>7–9, 11, 25, 46, 86, 92, 94, 118, 120, 141–168, 223, 226–227, 244–249, 257, 307</td>
</tr>
<tr>
<td>Amazon</td>
<td>61–62, 72</td>
</tr>
<tr>
<td>Andean Highlands</td>
<td>274, 282</td>
</tr>
<tr>
<td>Angola</td>
<td>152</td>
</tr>
<tr>
<td>Asia</td>
<td>15, 43, 87, 89, 92, 105, 143, 245–250, 264</td>
</tr>
<tr>
<td>South Asia</td>
<td>7</td>
</tr>
<tr>
<td>South East Asia</td>
<td>1, 7, 39, 50, 120, 245</td>
</tr>
<tr>
<td>Atlantic</td>
<td>3, 11, 21, 95, 118, 134, 141, 147, 149–150, 152–153, 175, 249–250, 253–256, 260, 389</td>
</tr>
<tr>
<td>Australia</td>
<td>1, 7, 43, 150, 256, 298, 319, 350, 355, 357, 359, 361</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td></td>
</tr>
<tr>
<td>Barbados</td>
<td>25</td>
</tr>
<tr>
<td>Belgium</td>
<td>149, 298, 319</td>
</tr>
<tr>
<td>Berbice River</td>
<td>222</td>
</tr>
<tr>
<td>Bolivia</td>
<td>4, 61, 71, 270–272</td>
</tr>
<tr>
<td>Brazil</td>
<td>61, 71, 150–152, 199</td>
</tr>
<tr>
<td>British Guiana</td>
<td>148</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>149–151</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
</tr>
<tr>
<td>Carchi</td>
<td>272</td>
</tr>
<tr>
<td>Caribbean</td>
<td>1, 3, 7–8, 18, 24, 118, 176–177, 181–183, 253, 257, 262, 333, 358</td>
</tr>
<tr>
<td>Ceylon</td>
<td>245</td>
</tr>
<tr>
<td>Chad</td>
<td>151</td>
</tr>
<tr>
<td>Chichipa</td>
<td>271</td>
</tr>
<tr>
<td>Chinchilla</td>
<td>273</td>
</tr>
<tr>
<td>Chota Valley</td>
<td>272–273, 277</td>
</tr>
<tr>
<td>Colombia</td>
<td>7, 118, 151, 176, 220, 271, 273, 319, 333</td>
</tr>
<tr>
<td>Commander Islands</td>
<td>15</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td></td>
</tr>
<tr>
<td>Danish West Indies</td>
<td>6, 220</td>
</tr>
<tr>
<td>Democratic Republic of Congo</td>
<td>150–151</td>
</tr>
<tr>
<td>Dominica</td>
<td>182–183</td>
</tr>
<tr>
<td>Dutch Antilles</td>
<td>17, 109, 271</td>
</tr>
<tr>
<td><strong>E</strong></td>
<td></td>
</tr>
<tr>
<td>East Africa</td>
<td>see Africa</td>
</tr>
<tr>
<td>Easter Island</td>
<td>128</td>
</tr>
<tr>
<td>Eastern Africa</td>
<td>see Africa</td>
</tr>
<tr>
<td>Ecuador</td>
<td>270, 272</td>
</tr>
<tr>
<td>El Carmen</td>
<td>273</td>
</tr>
<tr>
<td>El Guayabo</td>
<td>273</td>
</tr>
<tr>
<td>Equatorial Guinea</td>
<td>274, 278–279, 281–282, 286</td>
</tr>
<tr>
<td>Eritrea</td>
<td>108, 122</td>
</tr>
<tr>
<td>Essequibo River</td>
<td>224</td>
</tr>
<tr>
<td>Europe</td>
<td>51, 128, 147, 309</td>
</tr>
<tr>
<td><strong>F</strong></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>149, 298, 319</td>
</tr>
<tr>
<td>French Guiana</td>
<td>149, 178, 220, 392</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td></td>
</tr>
<tr>
<td>Gabon</td>
<td>150–151</td>
</tr>
<tr>
<td>Ghana</td>
<td>148–149</td>
</tr>
<tr>
<td>Grand Ronde</td>
<td>8, 84, 87, 126</td>
</tr>
<tr>
<td>Grenada</td>
<td>183</td>
</tr>
<tr>
<td>Guadeloupe</td>
<td>108, 176, 178–183</td>
</tr>
<tr>
<td>Guaporé-Mamoré</td>
<td>61</td>
</tr>
<tr>
<td>Guiana</td>
<td>118</td>
</tr>
<tr>
<td>Guinea</td>
<td>148, 150, 180, 249–250, 261, 279, see also Equatorial Guinea, New Guinea, Papua New Guinea, Guinea-Bissau</td>
</tr>
<tr>
<td>Guyana</td>
<td>7, 11, 167, 182–183, 219, 222, 224, 298, see also Guyane, French Guiana</td>
</tr>
<tr>
<td>Guyane</td>
<td>149, 180, see also French Guiana</td>
</tr>
<tr>
<td><strong>H</strong></td>
<td></td>
</tr>
<tr>
<td>Haiti</td>
<td>49, 150, 180, 182–183, 298, 319</td>
</tr>
<tr>
<td>Highland Ecuador</td>
<td>272, see also Ecuador</td>
</tr>
<tr>
<td>Highlands of Papua New Guinea</td>
<td>358, see also Papua New Guinea</td>
</tr>
<tr>
<td><strong>I</strong></td>
<td></td>
</tr>
<tr>
<td>Ica</td>
<td>273</td>
</tr>
<tr>
<td>Imbabura</td>
<td>272</td>
</tr>
<tr>
<td>India</td>
<td>1, 25, 72, 83, 87, 91, 118, 128, 199</td>
</tr>
<tr>
<td>Indonesia</td>
<td>148, 177, 251</td>
</tr>
<tr>
<td>Insular Southeast Asia</td>
<td>39</td>
</tr>
<tr>
<td>Location</td>
<td>Pages</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Java</td>
<td>105, 249–250</td>
</tr>
<tr>
<td>Karkar</td>
<td>358</td>
</tr>
<tr>
<td>Kenya</td>
<td>87, 150, 152, 199</td>
</tr>
<tr>
<td>La Paz</td>
<td>271</td>
</tr>
<tr>
<td>Laos</td>
<td>124</td>
</tr>
<tr>
<td>Lesser Antilles</td>
<td>183</td>
</tr>
<tr>
<td>Liberia</td>
<td>148–150, 152, 154</td>
</tr>
<tr>
<td>Lisbon</td>
<td>299</td>
</tr>
<tr>
<td>London</td>
<td>299</td>
</tr>
<tr>
<td>Los Yungas</td>
<td>271–272</td>
</tr>
<tr>
<td>Louisiana</td>
<td>179–180, 182, 298</td>
</tr>
<tr>
<td>Lower Guinea</td>
<td>180, see also</td>
</tr>
<tr>
<td>Guinea, Upper Guinea</td>
<td></td>
</tr>
<tr>
<td>Lowland Amazonia</td>
<td>62</td>
</tr>
<tr>
<td>Malaysia</td>
<td>150–151, 251, 319, 333</td>
</tr>
<tr>
<td>Mali</td>
<td>148–150, 152</td>
</tr>
<tr>
<td>Martinique</td>
<td>49, 108, 150, 176, 182–183</td>
</tr>
<tr>
<td>Mauritius</td>
<td>17, 48, 151, 180, 182–183, 298, 319</td>
</tr>
<tr>
<td>Melanesia</td>
<td>7, 18, 118, 333, 356–359, 361, 382</td>
</tr>
<tr>
<td>Mozambique</td>
<td>150–152</td>
</tr>
<tr>
<td>Nor Yungas Province</td>
<td>271</td>
</tr>
<tr>
<td>Normandy</td>
<td>180</td>
</tr>
<tr>
<td>North America</td>
<td>3, 7, 181</td>
</tr>
<tr>
<td>Northern Australia</td>
<td>7</td>
</tr>
<tr>
<td>Northern Haiti</td>
<td>180</td>
</tr>
<tr>
<td>Northern Sudan</td>
<td>198</td>
</tr>
<tr>
<td>Oceania</td>
<td>128, 143</td>
</tr>
<tr>
<td>Oregon</td>
<td>87</td>
</tr>
<tr>
<td>Palau Island</td>
<td>357</td>
</tr>
<tr>
<td>Peru</td>
<td>270, 272–273</td>
</tr>
<tr>
<td>Philippines</td>
<td>7–8, 118, 148, 152, 244–245, 251, 260, 263, 271, 274, 276, 279, 286, 319, 333</td>
</tr>
<tr>
<td>Pitcairn</td>
<td>108</td>
</tr>
<tr>
<td>Port Moresby</td>
<td>346</td>
</tr>
<tr>
<td>Portugal</td>
<td>151, 244, 298, 319</td>
</tr>
<tr>
<td>Port Vila</td>
<td>346</td>
</tr>
<tr>
<td>Rai Coast</td>
<td>357</td>
</tr>
<tr>
<td>Reunion</td>
<td>see Reunion Island</td>
</tr>
<tr>
<td>Reunion Island</td>
<td>6, 151, 170, 180, 183</td>
</tr>
<tr>
<td>Rotterdam</td>
<td>299</td>
</tr>
<tr>
<td>Sabine River</td>
<td>274, 286</td>
</tr>
<tr>
<td>Sahul region</td>
<td>43</td>
</tr>
<tr>
<td>Saint Kitts</td>
<td>25</td>
</tr>
<tr>
<td>San José</td>
<td>273</td>
</tr>
<tr>
<td>San Regis</td>
<td>273</td>
</tr>
<tr>
<td>Senegal</td>
<td>149, 152, 157</td>
</tr>
<tr>
<td>Seychelles</td>
<td>118, 182–183</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>150, 152</td>
</tr>
<tr>
<td>Solomon Islands</td>
<td>25, 354, 365</td>
</tr>
<tr>
<td>South Africa</td>
<td>see Africa</td>
</tr>
<tr>
<td>South America</td>
<td>1, 4, 15, 43, 50, 72, 181, 222, 262, 269, 274, 276</td>
</tr>
<tr>
<td>South Asia</td>
<td>see Asia</td>
</tr>
<tr>
<td>South East Asia</td>
<td>see Asia</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>see Asia</td>
</tr>
<tr>
<td>Southeast Papua New Guinea</td>
<td>see Papua New Guinea</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>120, 152, 248, see also Ceylon</td>
</tr>
<tr>
<td>Sudan</td>
<td>87, 148, 198–199</td>
</tr>
<tr>
<td>Northern Sudan</td>
<td>198</td>
</tr>
<tr>
<td>South Sudan</td>
<td>3, 150, 194, 198–199</td>
</tr>
<tr>
<td>Suriname</td>
<td>6, 11, 17, 24, 109, 152, 185, 220, 298, 392</td>
</tr>
<tr>
<td>Syria</td>
<td>199</td>
</tr>
<tr>
<td>Taiwan</td>
<td>39, 88, 152, 251</td>
</tr>
<tr>
<td>Tanzania</td>
<td>148, 151–152, 199</td>
</tr>
<tr>
<td>Tocaña</td>
<td>271</td>
</tr>
<tr>
<td>Tortuga</td>
<td>180</td>
</tr>
<tr>
<td>Trinidad</td>
<td>179, 182–183, 188, 298, 319</td>
</tr>
<tr>
<td>U.S. Virgin Islands</td>
<td>220</td>
</tr>
<tr>
<td>see also Virgin Islands</td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>87, 150, 152</td>
</tr>
<tr>
<td>Upper Guinea</td>
<td>7, 250, 261, see also Guinea</td>
</tr>
<tr>
<td>USA</td>
<td>148–151, 298, 319</td>
</tr>
<tr>
<td>Vanuatu</td>
<td>17, 298, 319, 354</td>
</tr>
<tr>
<td>Virgin Islands</td>
<td>151, 220, 305</td>
</tr>
<tr>
<td>West Africa</td>
<td>see Africa</td>
</tr>
<tr>
<td>West Indies</td>
<td>6, 139–140, 190, 220, 239, 268, 298, 311, 313, 343, 380</td>
</tr>
<tr>
<td>Western Africa</td>
<td>168</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>152</td>
</tr>
</tbody>
</table>
Subject index

A
acquisition 104
history 104
of language 6, 9, 12, 22, 83, 244, 271–272, 274, 278, 375
of morphemes 85
acrolectal creole variants 160–162, 173, 184–185, 187, 189, 271, 378
act of identity 17
adolescents 22–23, 88, 272
adstrate features 258
influence 86, 241, 254–255, 280
languages 8, 19, 26, 71, 87, 257, 263–264, 274, 377
structures 261–262
adstrates 3, 8–9, 12, 14, 19, 21, 261, 294
adults 20–23, 25, 221, 272–273
African vocabulary 11, 82, 92, 94, 135, 222–224, 226, 236
allolexy 352
Amerindian 84, 224, 226
ancestor 15, 17, 25, 36, 39–41, 60, 68, 73, 104, 175, 177, 236, 242–243, 269, 272, 387
ancestral states 70, 243, 256
Anglicization 30, 358
Anglocentrism 348
languages 27, 81–82, 89–90, 93, 120–121, 124–128
Non-APiCS languages 81
features 3, 50, 254, 261, 263, 356
semantics 309–310, 315–316, 345
areality 241, 263, 345, 353, 356, 363
article
definite article 13, 105, 109, 115–116, 181, 207, 284, 291
indefinite article 13, 24, 50, 109, 115–116, 120, 126, 128, 130, 135, 207–208, 257–259, 291, 369, 376
B
basic lexicon 8, 14, 93, 222, 305, 309–310
basilect 18, 114, 181, 189, 236, 271, 286, 348, 378
BEAST 39
bi-clan(s) 254, 371
bimorphemic question words 107, 109, 117, 135, 221, 232, 235
binary features 66, 72, 276
bioprogram 26, 88, 107
ears 315, 318, 331
eyes 202, 315, 318, 331–333, 335, 383
hands 315, 318, 331–332
head 315, 318, 331, 333–336, 383
legs 315, 318, 331–332, 381
mouth 315, 318, 331
nose 315, 318, 331–332
bootstrap 69, 74–75, 131, 197
borrowings 11, 14, 36, 40–42, 45, 69, 85, 92–93, 142, 234, 243, 328, 385, 393
branches 6, 36, 39, 41, 46, 68–70, 128, 131, 141, 147, 154, 164–165, 179, 198, 210, 227–228, 236, 243, 279, 300–301, 305–307, 327, 354
C
capitalism 348
caller 36–38, 74–75, 181, 243, 277, 297–298, 301, 386
circularity 15
clan 254

Subject index
cognition 27
cognitive authenticity 294
creolistics 2, 293, 297, 369, 381
semantics 293–295, 316, 331, 339, 382–383
viewpoint 295, 297
colonial influences 297, 346
languages 4, 8, 175, 304, 307, 309, 320, 329, 382
comparative method 41, 370
complexity 5, 12, 16, 21–22, 123, 193–194, 201–202, 204, 212, 294, 345, 385, 390, 393
componentiality hypothesis 10
computational approaches 71, 264, 379, 385
biology 38
linguistics 6
tools 1, 189, 193, 200, 219, 225, 233, 242, 269, 376
constrates 8, 193–194, 196, 193–194, 196, 198, 201, 204–208, 212–214
contact languages 6, 20, 35, 59, 64, 71, 93, 108, 120, 126, 175, 194, 251, 261, 275, 348, 358, 379, 390, 392
corpus 152, 269–270, 276, 285–286
distinctiveness 2–3, 194–196, 199, 210, 212, 214, 372
genesis 3, 6–7, 9–10, 17, 19, 23, 27, 49, 81, 136, 141, 166–167, 194, 213, 370, 389, 393
creole-lexifier 293, 315
semantics 27, 293–294, 297, 309–310, 345, 360
creoloids 6
cross-crorele comparisons 293, 297, 309, 315
cross-linguistic comparisons 80, 82, 92, 144, 198, 286, 294, 296, 315, 383, 391
semantics 1, 4, 295, 336, 339, 350, 383
cryptodiversity 309
D
data coding 66
data types 59–60, 63, 276, 285, 287
dating of phylogenies 39–41
definite articles see article
delta scores 243, 253–254
demographics 10, 15, 180, 185, 187–188
demonstratives 13, 84, 90, 92, 105, 109, 135, 144–145, 207, 246, 283, 291
demonstrative determiners 115–116
diaplectic 3, 133
concept of dialect 4, 271, 345, 383
non-standard dialects 118, 287, 299
distance matrix 37, 180, 197
distinctness 199, 231, 371
distortion 23, 392
DNA 37, 39, 370, 385–386
E
ears see body parts
encoding 60, 63, 181, 329
endogenous creoles 11
environment 4, 11, 245, 286, 376, 378, 387
environmental concepts 315–318, 336, 340
molecules 315, 318, 336
ethnogeography 336
ethnolectos 350
etymology 223, 225–228, 231, 236, 300, 308, 318, 336, 352, 359–360, 392
European lexifiers 8, 51, 85, 108, 128, 130, 133, 142, 154, 161, 199, 293–294, 297, 301, 315, 326–328, 340, 371, 376
evolution 35, 37–38, 41, 43, 45, 52, 59, 64–65, 70, 185, 188, 225, 385–387
language evolution 19, 41–43, 59, 64–65, 143, 277, 370, 386
evolutionary biology 11, 35–39, 44, 59, 175, 219, 385, 387
theory 35–37, 40–41, 45, 52, 59, 67, 227, 234, 385, 387
exogenous creoles 11
expanded pidgins 6, 134
explicit networks 67
extensional methodology 294
eyes see body parts
Subject index

F
feature pool 9–10, 369

f
hypothesis 3, 9–10, 194–196, 199, 210–211, 214
featurization 325, 346, 384
fieldwork 60, 226, 252, 269, 274, 276, 285–286, 296–297, 316, 381
foreigner talk 10, 87
fort creoles 11
founder effects 24, 43, 386
founder principle 9

G
gender agreement 45, 50, 83–84, 116, 284, 292
genesis 23–24, 27, 49, 70, 88, 123, 167, 183, 195, 230, 236, 390
global linguistic flows 346
glottoclock 39
gradual creolization 23, 221
grammatical
features 40, 61, 105, 143, 181, 193, 219, 246, 262, 272, 287
typology 307, 310
grammaticalization 7, 9, 13, 24, 50–51, 83–84, 105, 126–127, 133–134, 136, 196, 256, 376
granularity 64, 300, 352


H
hands see body parts
head see body parts
historical
linguistics 6, 41, 52, 59, 66–67, 95, 143
relations 249, 369–370
sources 6, 17, 23, 221, 232–233, 274
history of creoles 5, 15, 18, 64, 77, 104, 167, 225, 253, 276, 315
homestead 10
homogenous language 22, 315
homoplasy 71
horizontal transfer 386
hybridization 32, 213–214, 386
hyperpolysemy 27
implicit networks 67–68
indefinite articles see article
Indo-European lexifiers 160–163, 165, 199
inherited elements 7, 11, 14, 36, 49, 83–86, 88, 93, 142, 162, 232, 234, 216, 243, 393
innovation 1, 6, 9–10, 20–21, 26, 36–37, 40–42, 71, 85, 188, 200, 208, 214, 256, 301, 346
interrogatives 109, 115, 117, 120, 259, 261, 284, 291
jargons 6–7, 19, 21–22, 87, 134
kinship semantics 223, 317
language change 42, 51–52, 103, 127, 193, 195
contact 6, 45, 52, 59, 81, 180, 195, 242, 269–270, 277, 369, 371, 393
evolution 19, 35, 41–43, 45, 52, 59, 64–65, 70, 143, 188, 225, 277, 370, 385–387
trees 36, 45

L
language change 42, 51–52, 103, 127, 193, 195
trees 36, 45
lex-based theory 306, 308–309
lexical

data 63, 241
features 167, 241, 245, 271, 328
labels 300, 316, 320–321, 363, 381
structuration 299, 302, 309, 330, 356
lexicization
types 320
basic lexicon 8, 14, 93, 222, 305, 309–310
lexifier influence 262, 376–377
linguistic worldviews 331, 336

M
manmeri research 346, 357–359
distance matrix 37, 180, 197
maximum likelihood 37–38
medium of interethnic communication 10, 16, 24, 199
reductionism 393
reduplication 9, 79, 82–83, 85–86, 144–145, 204, 221, 223, 232, 247
reference grammar 60, 72, 148, 318
reggae 358
relabelling 14, 19
relaxed clock 39
relexification 12, 14, 18–19, 92, 244, 271, 293
restructured languages 6, 92, 120, 178, 225, 271, 280–281, 375, 377
retention 13, 36–37
reticulate networks 41–43, 67
S
salience 26
sampling 74–75, 146, 194, 197, 199, 209
school creoles 11
semantic domains 4, 309, 315–316, 321, 326, 331, 339–340, 363
equivalence 295, 339
explications 333, 335, 339
fieldwork 381
molecules 4, 315–318, 327, 383
of creolization 297, 300
semicores 6, 108, 183, 280, 297, 305
sociohistory 1, 5, 12, 15–16, 180, 193–196, 243, 375–379
sociolects 350
spelling 299
split networks 42–43, 71, 386
SplitsTree 46, 63, 146–181, 197, 225, 242, 253, 276–277
Stammbaum 36, 41, 48
superdiversity 347
superstrate influences 294, 316
languages 8, 20, 193–194, 347
superstrates 8, 10, 12, 14, 27, 113–114, 160–161, 316, 370, 389
superstratists 8, 167, 347, 363, 393
suppletion 85–86, 144, 202–203, 205–206, 212, 284, 291
suppressed populations 17
Swadesh list 37, 66, 179, 222–223, 226, 295–296, 392
T
target language 10, 249
taxon 68–69, 234, 243
tone 14, 51, 61, 82, 195, 202, 292
trade 11, 93, 188, 199, 222
transmission 20, 48, 110, 162, 370
broken 20, 45, 48, 142–143
normal 20, 23, 213
building 38
typology 1, 6, 242, 270
creole typology 2, 79–95, 103–137, 175–189, 195, 213, 293, 389–390
grammatical typology 308–310
lexical typology 308–309, 315
typological data 63–66, 231, 233–234, 250, 264, 275
typological databases 45, 60, 64, 122, 193, 200, 256, 269, 275, 392
typological distance 122
typological features 40, 59–60, 64–67, 123, 189, 242, 252, 264, 275–277, 284, 383
typological properties of languages 9, 61, 65, 104, 162, 176, 223, 225
U
This book launches a new approach to creole studies founded on phylogenetic network analysis. Phylogenetic approaches offer new visualisation techniques and insights into the relationships between creoles and non-creoles, creoles and other contact varieties, and between creoles and lexifier languages. With evidence from creole languages in Africa, Asia, the Americas, and the Pacific, the book provides new perspectives on creole typology, cross-creole comparisons, and creole semantics. The book offers an introduction for newcomers to the fields of creole studies and phylogenetic analysis. Using these methods to analyse a variety of linguistic features, both structural and semantic, the book then turns to explore old and new questions and problems in creole studies. Original case studies explore the differences and similarities between creoles, and propose solutions to the problems of how to classify creoles and how they formed and developed. The book provides a fascinating glimpse into the unity and heterogeneity of creoles and the areal influences on their development. It also provides metalinguistic discussions of the “creole” concept from different perspectives. Finally, the book reflects critically on the findings and methods, and sets new agendas for future studies. Creole Studies has been written for a broad readership of scholars and students in the fields of contact linguistics, biolinguistics, sociolinguistics, language typology, and semantics.