John Benjamins Publishing Company

Creoles are typologically distinct from non-creoles*

Peter Bakker[a], Aymeric Daval-Markussen[a], Mikael Parkvall[a,b], and Ingo Plag[c]

[a]RC-GLU: Research Centre for Grammar and Language Use, Aarhus University / [b]Stockholm University / [c]Universität Siegen

In creolist circles, there has been a long-standing debate whether creoles differ structurally from non-creole languages and thus would form a special class of languages with specific typological properties. This debate about the typological status of creole languages has severely suffered from a lack of systematic empirical study. This paper presents for the first time a number of large-scale empirical investigations of the status of creole languages as a typological class on the basis of different and well-balanced samples of creole and non-creole languages. Using statistical modeling (multiple regression) and recently developed computational tools of quantitative typology (phylogenetic trees and networks), this paper provides robust evidence that creoles indeed form a structurally distinguishable subgroup within the world's languages. The findings thus seriously challenge approaches that hold that creole languages are structurally indistinguishable from non-creole languages.

1. Introduction

In creolist circles, scholars have defended two opposite opinions with regards to the synchronic structures of creoles. Some claim that they do not differ from non-creole languages (e.g. Hall 1966, DeGraff 2001, Ansaldo 2004, Chaudenson 2003: 57–62, Goury & Léglise 2005, Lefebvre 2000: 128–130, Mufwene 1999, 2007b, Winford 2005: 411, 2008), while others claim that creoles do form a special class with specific typological properties (e.g. McWhorter 2001, 2005, Parkvall 2008). Parkvall (2008) as a recent exception notwithstanding, the debate about the

---

* The full data for this paper can be found at https://www.researchgate.net/publication/235270676_Supplementary_Materials_Bakker_et_al._2011/file/d912f510b88714dba8.pdf

typological status of creole languages has severely suffered from a lack of systematic empirical study.

This paper will present a large-scale empirical investigation of the status of creole languages as a typological class on the basis of well-balanced samples of creole and non-creole languages. Using statistical modeling (multiple regression, e.g. Baayen 2008) and recently developed computational tools of quantitative typology (phylogenetic trees and networks, e.g. Dunn et al. 2008), we provide evidence that creoles do form a structurally distinguishable subgroup within the world's languages. Our findings thus seriously challenge approaches that hold (usually in the absence of wide-ranging empirical evidence) that creole languages are indistinguishable from non-creole languages.

We look at creoles from a typological point of view in this paper, where typology is taken in a Greenbergian sense, not in a parametric sense. Thus, we consider creoles in a comparative and synchronic perspective, and we do not make any claims as to whether the diachronic developments in creolization differ from those in language change in non-creoles.

We will first put our study in the perspective of discussions on creole properties in Section 2. We will briefly characterize creoles in Section 3. In Section 4 we will discuss the main computational techniques we use in our study. These techniques have been developed in evolutionary biology and have found recent application in linguistic studies of genetic relationships and areal phenomena. Our language sample and data will be discussed in Section 5. Section 6 presents the results of our investigations. The final section will present conclusions and some perspectives for future research.

2.   Previous work

The search for common properties of creole languages can already be noted in the writings of the forefathers of the discipline more than a hundred years ago (e.g. Coelho 1880, Schuchardt 1914). In more modern times, Muysken (1988) may be taken as a point of departure for the main issues in the discussion. He split up the debate into three issues:

– Are creole languages simpler than other languages?
– Are creoles more similar to each other than to other languages?
– Are creoles more mixed than other languages?

Basically, Muysken's answers were negative for all three questions. Despite this, 'creolistics' has continued as a distinct field within linguistics, and even those who claim that creole languages are no different from other languages, discuss 'creoles'

as a group. The issues raised by Muysken have remained a subject of debate, which we will not attempt to summarize here. We will only consider complexity and similarity, as the issue of mixedness is not relevant to our claims.

2.1 Complexity

The discussion on creole simplicity received a boost with McWhorter (2001), who claimed that 'the world's simplest grammars are creole grammars'. The reactions, as exemplified in the peer commentaries on McWhorter's article (in *Linguistic Typology* 5 (2/3), pp. 167–387) were mostly hostile and skeptical, and only few accepted the thesis. McWhorter provided a range of arguments here for his claim that creoles differ typologically from non-creoles, discussed in much more detail in McWhorter (2005), updating his works on this question since 1998. One could criticize McWhorter's method, however, because he mostly compares individual creoles with individual non-creole languages. If one selects as a non-creole a language that is renowned for its complexity, such as the Caucasian language Tsez in McWhorter's case, then one might arrive at the conclusion that virtually all languages, including creoles, will turn out 'simpler' than the language they are compared to. This does not necessarily mean that McWhorter's claim is incorrect, just that the conclusion could be an artefact of the selection. If one wants to establish, for instance, whether elephants are fast animals, it makes a big difference if one compares them with leopards or with snails, where the selection of exemplars would obviously lead to widely different conclusions. Thus, his argument usually appeals to gut feeling rather than being founded on quantified (or quantifiable) features.

What one needs is a balanced sample of the world's languages (along the lines of Rijkhoff & Bakker 1998) to compare a set of creoles to (cf. Hagège 2001: 174). To the extent that creolists have ventured into comparing creoles with other languages, they have compared them either with other individual creoles or with each other within groups of creoles of the same lexifier, or with supposed substrate languages, or with the lexifier languages, but until recently never with a sample of languages of the world. This limited scope has probably contributed to the (in our view incorrect) conviction that creoles constitute an arbitrary subset of the world's languages.

Here the views of typologists and creolists differ considerably. The claim that creole languages make up a typologically distinct grouping is, in our experience, not particularly controversial among typologists (even though this view is not often expressed in print). Typologists will include creoles (if not as a class of their own) with groupings of more isolating languages rather than with e.g. polysynthetic or agglutinative languages. However, the same claim triggers hostile and often emotionally laden reactions among creolists. It is therefore time to settle

the matter. All our evidence indicates unequivocally that creoles are typologically different from non-creoles, as will be shown below in our large-scale comparison.

The first empirical study that actually compared a sample of creoles with a large-scale sample of the languages of the world was Parkvall (2008). He, however, only looked at complexity issues. His conclusion was that creoles as a group are structurally complex (as are all human languages), but less complex than any typological, geographical or genealogical grouping of the languages of the world. In addition, he showed that it was impossible to come up with any systematic or non-rational grouping of languages (labeled 'silly' by the author, e.g. languages whose name start with a particular letter, or spoken by people who have a certain skin color) that would turn out less complex than the group of creoles. In Parkvall's calculation, based on quantifiable features in the *World Atlas of Language Structures* (WALS), there were a couple of non-creoles that were less complex than some of the creoles, but these did not form a natural grouping. The conclusions are unavoidable that creoles form the least complex grouping of languages, but also that some individual non-creoles are less complex than some individual creoles. Creoles as a group all belong to the languages in the less complex part of the spectrum. Papiamentu was, in Parkvall's metric, the creole with the highest complexity score (0.32). Out of 153 non-creole languages, only 29 had a lower score than the most complex of 32 creoles, i.e. the creoles belong to the lowest 20% of the complexity spectrum.

Parkvall (2008) showed that, from the perspective of phonological and morphosyntactic complexity of the languages of the world, creoles as a group stand out as being less complex than non-creoles.[1] This does not mean that all creoles are less complex than all non-creoles in all respects. It has been shown, for instance, that segmental inventories in the phonological systems of creoles (and pidgins) are not significantly different from those of non-creoles (Klein 2005, 2006a, 2006b, this volume, Bakker 2004, 2009). Also creole numeral systems appear to be of the same complexity as the ones found in non-creoles (Hammarström 2008).

2.2 Similarity

The second question, about creoles being an identifiable subgroup in the sense that its members are more alike one another than other languages, is represented by McWhorter (1998). Among those who do not consider creoles a distinctive subgroup are Mufwene (e.g. 2007b), DeGraff (e.g. 2001) and Ansaldo (e.g. 2004). Explicit claims of this sort are found in e.g. Chaudenson (2003), Goury & Léglise

---

1. Only two creoles — Ndyuka and Sango — were included in the WALS sample, but Parkvall added data for more creoles from other sources. The additional creoles patterned like the two already present.

(2005), Hall (1966), Lefebvre (2000), Mufwene (1999), Winford (2005, 2008), to just mention a few. Simplifying his views considerably, Chaudenson (e.g. 2003) believes that (French-lexicon) creoles developed by successive generations who tried to imitate the imperfect French spoken in colonial situations, and this view has since been taken up by other creolists and gained a considerable following. With there being no identifiable break in transmission, creoles should, according to this school of thought, be regarded as daughters of their lexifier, and they are as compatible with a *Stammbaum* structure as any other language.[2] According to this reasoning, creoles should be as similar to the lexifier as non-creole dialects of the same lexifier.

Still, creoles have been claimed to be different from non-creoles in the synchronic structural properties, and in the type or rate of transmission between generations, in a diachronic sense. DeGraff has traced the history of this 'creole exceptionalism' and pointed out ideological backgrounds in a number of papers (e.g. 2001, 2003, 2005a), but usually without providing a reasonable number of linguistic examples to refute the claim: a handful of counterexamples are not sufficient to disprove McWhorter's claim (see also Farquharson 2007 for exceptions among creoles). Many of Mufwene's (e.g. 2001, 2003) works are similarly devoid of empirical linguistic documentation for the claim that creole languages are indistinguishable from 'ordinary' languages. Ansaldo and Matthews (2007: 4) say the time is ripe to 'dissolve the notion of "creole" as a particular type of language'. We do not consider creoles to be exceptional languages, but we will provide substantial empirical evidence for the position that creoles form an identifiable and distinct subgroup among the world's languages.

In those works that claim that creoles do not differ from other languages, empirical documentation for the claim and linguistic arguments are often lacking or at best scarce and limited. For instance, there are only two studies that investigate this question, in the framework of the 'feature pool' theory (Mufwene 2001) by using empirical data (Aboh & Ansaldo 2007, Bobyleva 2009). In a recent collection of articles (Ansaldo, Matthews & Lim 2007) devoted to attacking the idea of creole exceptionalism, the bulk of the papers either do not address the question, or fail to provide empirical data. Aboh & Ansaldo's paper is in fact one of the few papers

---

2. In many regards, Mufwene (2007a: 64, 2007b: 59, 61) goes further than most, in seeing creoles as dialects of their respective lexifiers, rather than as languages in their own right. This is potentially topped by DeGraff (2003: 402, 2005b: 338) and Ansaldo (2004), who do not content themselves by claiming that the label 'creole' fails to correspond to a typological profile, but in fact that this even 'cannot' and 'could not' be the case — in other words, no actual data is really needed, but the whole idea of creoles as (diachronically) descendants of 'broken transmission' and that they (synchronically) would represent a typological class can be refuted by means of 'theoretical observations' (DeGraff 2003: 398; Ansaldo 2004: 490).

that actually makes use of creole language data (but see Plag 2011 for a detailed critique of Aboh & Ansaldo's approach).

In this paper we address the question of whether creole languages display a distinct typological profile with a number of empirical studies based on different samples of languages and linguistic features. Our conclusion will be that they do indeed. In providing the evidence, our data shed light on two of the three questions alluded to above: with regard to complexity, creoles are distinct from non-creoles, and creoles are more similar to one another than they are to non-creoles — which means that they do indeed have a distinct typological profile, independent of the complexity issue.

3.   Creoles

Creole languages are a set of languages that came about under certain specific social circumstances. Creoles are sometimes characterized on the basis of socio-historical circumstances (e.g. nativized or vernacularized developments of pidgins, which are makeshift languages used in some contact situations), or on the basis of structural features, i.e. a language which, after having undergone restructuring from a pre-existing language, now has e.g. SVO word order, preverbal TMA marking and little or no morphology. One can also take the social criteria as primary, however, and then investigate to what extent they correlate with linguistic features, in order to avoid circularity in the definition of creoles (cf. Mufwene 1986, Bakker 2008). In practice, most linguists use both criteria.

Creoles are found in many parts of the world. Some creoles share quite specific features, such as shared specific borrowings from African languages (e.g. the pronoun *unu*, presumably from Igbo, in English creoles), specific sources for grammaticalization, e.g. copulas in English-based creoles from a determiner. These can be interpreted as to some creoles having a partially shared history (cf. Smith 1987, Baker & Huber 2001, McWhorter 2005, Daval-Markussen & Bakker, 2011, and earlier work). On the other hand, creoles are found in widely separate areas of the world, and with different sources of the lexicon, which suggests at least a number of independent geneses of creoles (see for instance the geographical and linguistic diversity of the creoles presented in Holm & Patrick 2007).

Creoles are often conveniently grouped on the basis of the main source of their lexicon, called lexifiers. A classification based on their sociohistory and the way they came into being makes a distinction between maroon creoles, fort creoles and plantation creoles. 'Maroons' refer to slaves who escaped from the plantations to establish their own groups away from areas exploited by European settlers. 'Fort creoles' are those thought to have emerged in or around European trading posts,

forts or comparable urban settlements. Their genesis often involved interethnic marriages. 'Plantation creoles', obviously, are those believed to have been born in plantation colonies.

The question asked here is: can it be proven that creoles form a specific structural type? Yes, we believe this can be done. In the next section we discuss the techniques used.

4. Phylogenetic trees

The traditional *Stammbaum* trees used in linguistics for displaying historical connections between languages are quite limited. Most importantly, they are only based on shared, historical features. For instance, English will unequivocally belong to the Germanic branch, even though the language is significantly influenced by other languages, especially French. Any contact influence is necessarily ignored, and double inheritance (mixture) is not possible in this model. For a long time, contact influence between languages, especially if they were not genetically related or typologically very distant, was denied or downplayed by historical linguists. Perhaps this denial of horizontal influence was caused by the exclusive focus on inherited material, where linguists endeavored reconstructing earlier stages. Also, linguists may have become prisoners of their tree metaphor: as branches of trees rarely if ever grow together, something like that could not happen with language trees either.

Whereas historical linguistics was a source of inspiration for the development of ideas of biological evolution in the 1800s (see Atkinson & Gray 2005),[3] today it is the biologists who take the lead in thinking about evolution and inspire linguists. In recent years, a number of algorithms have been developed by bioinformaticians to help visualize biological evolution (see e.g. Huson & Bryant 2006). The resulting phylogenetic networks have a number of advantages over the old evolutionary trees. First, they can account for horizontal relationships, i.e. contact phenomena. Second, the length of the branches of the trees reflect how many features two languages have in common. In some cases, rough absolute time dimensions can be inferred. Third, there is an obvious advantage with having a computer program produce a visual representation, since it can handle an amount of data much larger than any number of human beings ever could on their own during a lifetime. At

---

3. At this point, it is worth emphasizing that throughout this essay, evolution is understood as descent with modification and nothing more, contrary to the 19th century views, which are reflected in contemporary writings and so eloquently summed up in DeGraff 2001 — and to which none of the authors adhere.

the same time, the networks are completely objective and thus not influenced by any preconceptions and prejudices. The features all have the same weight, but the selection of the specific features will of course have an effect.

State-of-the-art algorithms now make it possible to draw trees that show not only inheritance, but also horizontal influence (contact, borrowing). A disadvantage is that only commonalities and differences are measured. It is impossible for the program to know whether these commonalities are inherited or borrowed.

4.1 Computational phylogenetics

During the past quarter of a century, the use of computers has greatly facilitated the work of biologists and enabled bioinformaticians to devise more powerful and sophisticated algorithms to provide a graphical representation of the most probable evolutionary history of species. The role of lateral processes (horizontal influence) has become increasingly evident in speciation events, and the tools to represent these graphically have consequently been adjusted. Networks rather than trees are used when reticulate events are believed to have occurred in a species' historical development. A phylogenetic network can be defined as any graph used to visualize evolutionary relationships (represented by edges or branches) between gene sequences, species, or, as in our case, languages (represented by nodes or taxa). Thus, a phylogenetic tree is one type of phylogenetic network, but one that does not take into account reticulation events. The main difference between a tree and a network is that there is only one possible path between two taxa in a tree (i.e. a strictly bifurcating mother–daughter relationship), whereas a network provides several different paths between taxa (i.e. a multifurcating relationship between a daughter and two or more parents). A splits network has the advantage of being able to depict conflicts in the dataset, for example due to reticulation events. The main methods used to detect a phylogenetic signal and to compute these phylogenetic networks are quantitative in nature and rely heavily on statistics. The most popular methods are Maximum Parsimony, which infers the most probable tree according to the evolutionary scenario involving the lowest number of changes. Other methods are Maximum Likelihood, a probabilistic approach searching for the most likely tree given the data, and models based on Bayesian inference, another approach based on probabilities, where the chance of a hypothesis being true is updated accordingly with the cumulation of observed evidence.

In order to tackle the problem of incompatibilities in a dataset in the form of conflicting signals (due for instance to reticulation events), the method of split decomposition was developed (Bandelt & Dress 1998). This method provides split-graphs or networks which are computed according to the compatibility of collections of splits found in the data. In this paper, the approach adopted follows

the network-based method of split-decomposition for computing phylogenies (Bandelt and Dress 1992), using the software SplitsTree (Huson & Bryant 2006).

4.2 Phylogenetic trees and networks in linguistics

The use of computational phylogenetics for the purpose of studying linguistic evolution has increased tremendously in recent years. Both lexical and structural studies using the tools and methods developed by bioinformaticians have proven to be successful (see Nichols & Warnow 2008 for an overview and evaluation of a number of linguistic studies and the various algorithms used; see also Dunn et al. 2008 for a non-technical description of the mathematical procedures behind the various algorithms).

Biologists have also used these tools to draw trees of certain language families, e.g. Gray & Atkinson (2003). On the basis of vocabulary, they drew a tree of Indo-European languages, which was quite close to the existing type of trees that linguists had produced, on the basis of their data and knowledge. Gray & Atkinson also added a time-line, which, however, resulted in a deeper time depth than what most historical linguists had assumed for Indo-European. Their tree seemed to reconcile views that estimate a time depth based on the languages without the Anatolian languages, and those who tended not to include those. More recently, a similar enterprise was undertaken for the vast Austronesian family, with a similar success rate (Gray et al. 2009).

April McMahon and her colleagues were among the first linguists to use such trees in a variety of applications. They used lexical data from Indo-European languages (McMahon & McMahon 2003, 2006), Australian Aboriginal languages (McMahon & McMahon 2006) and English dialects (McMahon et al. 2007).

Dunn et al. (2005) went a step further. Previous work focused on languages known beforehand to be genetically related. Dunn et al. (2005) applied it to a sample of languages spoken in Western Melanesia. Some of these belong to the Austronesian language family, others to a number of families called Papuan or Non-Austronesian. The Austronesian languages are clearly genetically related, whereas the Papuan languages are so different that their genetic relationship cannot be established.[4] The results of Dunn et al. (2005) are quite remarkable. They did not use lexical features, but exclusively structural-typological features, an innovation compared to earlier works. The program neatly splits the languages into an Austronesian set of languages and a set of groupings of Papuan languages, roughly corresponding to the groups established by the traditional comparative

---

4. In other words, 'Papuan' is merely shorthand for 'non-Austronesian languages of Melanesia', and is not intended to represent a genealogical grouping.

method. Despite some controversy over methodological issues and interpretation (Donohue et al. 2008), the results still stand as a landmark (see Dunn et al. 2008).

We take the model again a step further. We use the model also with exclusively structural-typological features, but this time for a historical group of languages, creoles, where we know in most cases that there are no historical connections between them: the creoles selected are widespread geographically, they have different lexifiers and came into being under quite different circumstances (viz. fort, plantation, maroon). Even though the trees and networks have been designed for mapping evolution, we use them for finding similarities in languages that came into being independently (in most cases) from one another, and that are not in areal contact.

These networks that account for both historical relationships and borrowed items are ideal for application to creole languages, as both inheritance and contact played an important role in the formation and development of creoles. Creoles take over features from their lexifiers (historical), from substrate languages (historical and contact), from neighboring languages (contact). These relations as established by the programs can represent loans, structural borrowings, shared inheritance, substratal influences or independent developments, some of them called 'adstrate influence' by creolists. It is interesting to know, of course, how these languages would group together on the basis of the number of shared features.

4.3   Creole language trees and networks

Previous Stammbaum trees with creole languages tended to be of two types. In trees based on inherited vocabulary, creoles unsurprisingly cluster with the lexifier languages. For instance, in McMahon's tree of Indo-European languages, Sranan (the only creole included) clusters with English (McMahon & McMahon 2003: 34, there called *takitaki*). Traditionally, however, both lexical and structural (especially morphological) features are deemed necessary for establishing a genetic connection between languages. Creoles typically show lexical continuity with their lexifiers, but only limited continuity in their structural make-up, making it strictly seen impossible to consider a creole language a genetic descendant of its lexifier.

The other types of trees are based on creole languages of the same lexifier, where creolists try to come up with a subclassification among these. Trees drawn on this basis tend to form geographical clusters (see e.g. Hancock 1987, Smith 1987, Daval-Markussen & Bakker, 2011, for the English Atlantic creoles).

We can formulate a number of hypotheses with regard to the connections between different creole languages.

1. If one emphasizes the role of the lexifier languages, then one would expect the creoles with the same lexifier to cluster together.
2. If one emphasizes the role of the substrate languages, one would expect the creoles to cluster more or less by region (e.g. Caribbean creoles, with roots in West Africa and the set of substrate languages spoken there).
3. If one assumes that creoles get *less* creole-like in time, one would expect the oldest creoles to be more deviant from the lexifier than the ones that came into being later.
4. If one assumes that creoles get *more* creole-like in time, for instance because successive generations produce approximations of the language of preceding generations, one could expect older creoles to be less deviant from the lexifiers.
5. If one believes that universal patterns play a role in creole genesis, then one expects no clustering on the basis of lexifier, geography, substrate or historical connections (but it is likely that creoles themselves form a group).

These hypotheses can be tested on the basis of a sample of creoles. By relying only on structural-typological features, one can abstract away from inheritance of forms, and compare creoles across lexifiers. This will be done in Section 6.

5.   Samples and data

The aim of this paper is to study the similarities and dissimilarities of creole languages and non-creole languages. In order to do so, we carried out five different studies. Most of them are based on samples selected by other scholars, and for different purposes. We used other scholars' pre-existing samples in order to avoid any potential bias by the present authors with regard to the selection of languages or features. The use of software also guarantees that all features have the same weight, thus minimizing the bias. In this section we discuss in detail those samples and data that we employ in more than one study. The other samples and data are only briefly introduced in this section, and more detailed discussion can be found in the pertinent subsections of Section 6.

In the first study (study 1), we establish that structural-typological features are highly suitable for establishing the kinds of relationships among languages that we are interested in. This is important since any cross-linguistic study of language similarities across language families needs to rely on structrural rather than lexical data. We study this methodological question on the basis of Hancock's (1987) sample of a particular subset of creole languages, i.e. the English-lexifier creoles of the Atlantic. Based on this sample (which is discussed in more detail in

Section 6.1), it is established that the structural-typological features can indeed be used for our purposes (Section 6.1).

Studies 2 through 5 are then devoted to questions of typological similarities among creoles, and of similarities between creoles on the one hand and non-creoles on the other. In studies 2 through 4 (see Sections 6.2 through 6.4) we investigate the degree of creoleness of these languages (study 2), the typological clustering of creoles (study 3) and the typological clustering of creoles and non-creoles (study 4) on the basis of structural properties of 18 creoles sampled in Holm & Patrick's *Comparative Creole Syntax* (2007, henceforth CCS).

CCS uses a set of 97 structural features, all somehow assumed to be typical of creole languages. Specialists in 18 different creole languages were asked to describe the languages with regard to these features. These were neatly summarized in tables, in which <+> meant presence, <−> meant absence, <?> meant unknown and <R> meant 'rare'. In order to reduce the distinctions to binary oppositions, we merged <R> with the category of <+>.[5] Question marks were maintained.

The creoles selected had 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), Spanish (Palenquero, Zamboangueño). Two of them show noteworthy admixture from African languages (Angolar: Portuguese/Bantu; Berbice Dutch: Dutch/Ijo) and one has been considered both or either Portuguese and/or Spanish (Papiamentu). This can be considered a reasonable and balanced spread over lexifiers.

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 (Zamboangueño).

Also, the age of the different creoles could be a factor. If one takes a Bickertonian (e.g. Bickerton 1981, 1984) view of sudden genesis of creoles, combined with the sociolinguistic phenomenon of decreolization, one would assume that those creoles that came about longest ago, will display fewer creole traits (unless the creole has not undergone decreolization). On the other hand, if one assumes that creoles developed gradually over many generations (see the contributions in Selbach et al.

---

5. The authors also merged the <R> category with <−>. This caused no major differences in the overall results, only certain details concerning the topology of the resulting networks and the rankings of languages (see 6.2 to 6.5 below) were slightly affected.

2009, or, for a different view of gradualism, Chaudenson 2001, 2003), one would assume that the creoles that have developed earliest, have become more creole-like in time, e.g. through 'approximations of approximations' in successive generations.

Finally, the creoles in the sample emerged in different circumstances. Some developed among escaped slaves (Ndyuka, Angolar, Palenquero), others emerged around trading places as 'fort creoles' (Cape Verdean, Guinea Bissau, Zamboangueño, Korlai, Nubi), Nagamese as a means of interethnic communication in trade and the rest came about on multilingual plantations.

Only a few of these creoles are presumed to be historically connected (see e.g. Baker 1988, 1993, 1999, Baker & Huber 2001, Hancock 1987, Smith 1987, McWhorter 1995): Krio, Ndyuka and Jamaican are likely to derive (at least in part) from a common ancestor. Guinea-Bissau and Cape Verdean are closely connected historically (they are in fact treated together in the Holm & Patrick book; see also Jacobs 2009). For all other creoles it is certain that they emerged completely independently from the others in this sample: Berbice Dutch, Korlai, Nagamese, Negerhollands, Nubi, Palenquero, Seychellois, Tok Pisin and Zamboangueño.

In short, this is a balanced sample, with a fair distribution across lexifiers, geography and circumstances of genesis, including at least one that is fairly deviant structurally and not always classified as a creole (Nagamese).

The degree of creoleness of the languages in the CCS sample (study 2, Section 6.2) is gauged by a rather simple measure, namely the number of typical creole features present in each of the 18 languages in the sample. In a regression analysis we test hypotheses (1) through (5) by modeling whether lexifier, area, type and age can predict the degree of creoleness for a given creole. It is shown that the degree of creoleness for a given creole is not dependent on superstrate, area of origin, the type of sociolinguistic setting (plantation, fort or trade) or the creole's age.

Study 3 uses the CCS features and languages to investigate the typological similarities among creoles based on phylogenetic networks and trees. The results are very similar to the ones of study 2: creoles are quite similar to one another, and the similarities are not primarily based on lexifier, area or substrate.

Study 4 also uses the CCS sample of features, but now includes not only the 18 creoles of the CCS sample, but also a sample of 12 non-creole languages. It appears that the creoles and non-creoles cluster separately (6.4).

Study 5 finally uses a different sample of languages and a different sample of features. We investigate 43 features that are taken from Parkvall (2008), who in turn took them from WALS, the *World Atlas of Language Structures* (Haspelmath et al. 2005). This sample will be discussed in detail in Section 6.4. His sample of languages comprised 155 languages sampled from WALS, to which he added 30 pidgins and creole languages. From WALS he took all languages for which at least 30 of the selected features were known. We use Parkvall's sample of languages for

study 5. The aim of study 5 is the same as that of study 4, only with a different set of features and a different set of languages: we compare creoles and non-creoles and model their typological clustering using again phylogenetic networks (6.5). Here again, creoles cluster separately.

The results of studies 2 through 5 all point in the same direction: creoles form a typologically distinct group of languages, in favor of hypothesis 5.

6.   Results

6.1 English-lexifier creoles: Lexicon and structure

The first part of our study to be discussed is a test in evolution, for which the SplitsTree software was developed originally (Huson & Bryant 2006). First applied to lexical evolution and later to structural evolution in linguistics, the biological models appeared to work for both. Holman et al. (2008) show that in fact the most successful method of language development and classification combines lexical with grammatical (typological) data.

In his seminal study, Hancock specifies that his sample '(…) provides a body of synchronic data for others to use in whatever productive way they may' (1987: 268). His sample consists of a set of 50 sentences elicited from speakers of 33 distinct varieties of English and English creoles. The varieties under scrutiny are the following (the abbreviations recur in the figures further below):

(1)   Afro-Seminole (Afr), American Black English (Bla), Antigua (Ant), Bahamas (Bah), Barbados (Bar), Belize (Bel), Boni (Bon), Cameroon (Cam), Carriacou (Car), Caymans (Cay), Grenada (Gre), Guyana (Guy), Hawaii (Haw), Jamaica (Jam), Krio (Kri), Kwinti (Kwi), Liberia (Lib), Matawai (Mat), Ndyuka (Dju), Nigeria (Nig), Norfolk (Nor), Paramaccan (Par), Providencia (Pro), Saba (Sab), Saramaccan (Sar), Sea-Islands Creole English (aka Gullah) (Sea), Sranan (Sra), St. Eustatius (StE), St. Kitts (StK), St. Thomas (StT), St. Vincent (StV), Tobago (Tob), Trinidad (Tri)

All but the creoles of Norfolk Island and Hawaii are spoken in the Atlantic area. To date, this sample represents the most comprehensive database of directly comparable material for the Atlantic English-based creoles that provides both lexical and grammatical data. These data were therefore chosen as a testing ground for the methodology proposed here. We decided to make a list of all the differences observed between the sentences, and divided them into formal and structural features. 'Formal' includes lexical and phonological differences, whereas under the 'structural' heading, the more abstract typological features such as the presence

or absence of certain types of copula, or the order of genitive and noun, etc. are subsumed. All in all, 122 lexical and structural-typological features were noted and coded into binary oppositions, i.e. presence vs. absence of features (<1> or <0>, and <?> for unknown and <−> for inapplicable features) as required by the software. Roughly half of them were related to form, half of them to structure. For more details, including the chosen features and their values, see Daval-Markussen & Bakker (2011). At present it is not possible to specify which individual features are responsible for the clusterings, but as the results are so robust, we can be quite confident in their validity and reliability.

One interesting result is that the phylogenetic networks drawn by the program are, except for some minor differences, highly similar in their topologies. Figures 1 and 2 show two of these networks. The first in Figure 1 is based on forms, the second in Figure 2 based on structures.
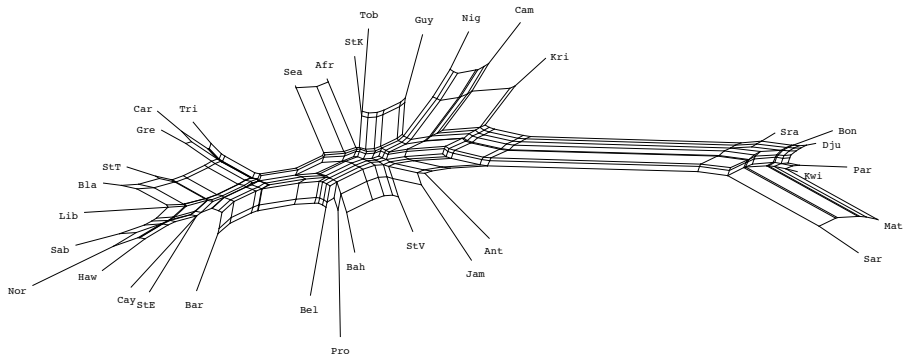

Figure 1. Phylogenetic network for Atlantic Englishes based on forms only.
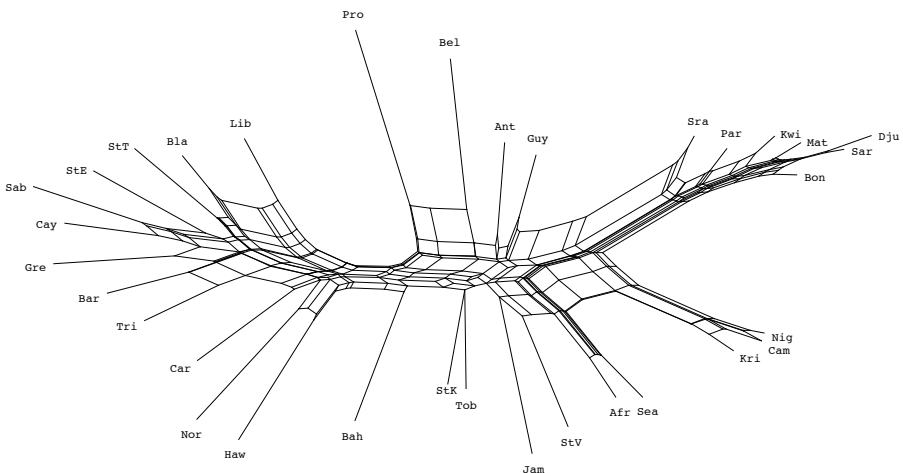

Figure 2. Phylogenetic network for Atlantic Englishes based on structure only.

In both graphs, three major groups appear, distributed as follows: a first group (to the right) made up of all the creoles of Suriname (Bon/Dju/Kwi/Mat/Par/Sar/Sra), a second group comprising the restructured vernaculars of West Africa (Cam/Kri/Nig), the North American varieties (Afr/Sea), the clusters StK/Tob, Ant/Guy and the Western Caribbean cluster Bel/Jam/Pro, and a third group (to the left) made up of five clusters (the two Pacific vernaculars (Haw/Nor), two Eastern Caribbean clusters (Bar/Tri and Car/Gre), and two other, seemingly unrelated clusters (Cay/Sab/StE and Bla/Lib/StT)).

It is common practice to benchmark one's results against a well-understood model of evolution in order to test the validity of the results (Nichols and Warnow 2008: 777), but in our case, since no classification of English-based creoles is yet widely accepted, we are unable to compare the results with a particular model (see Daval-Markussen & Bakker 2011 for a more detailed study). In genetics, an outgroup (a taxon which is known to have separated from the other taxa a long time ago) is usually chosen in order to root the network. Therefore we have included the putative ancestor from which all nodes descend, directly or indirectly, in order to root the tree. To this end, we included English (Eng) in the dataset, for which binary values were encoded in the software SplitsTree, and created a rooted tree (Figure 3). We expected the most acrolectal varieties to appear near the root of the tree, and this is exactly what we observed, as Figure 3 shows, where English is the leftmost language, closest to the root.

Further evidence for the adequacy of this method lies in the fact that the data have been adjusted and corrected several times, and that the resulting graphs consistently showed the same clusters. Moreover, we also tried to skew the data in order to observe and assess the effects of these changes on the graphs (for example by inverting all the values, by using only half of the features in each category, or by assigning Sar only <+>s or Eng only <−>s),[6] and this also resulted in consistent clusters. Moreover, we have ran several bootstrap analyses[7] of the data using SplitsTree's built-in function with 1000 iterations, and this also resulted in strong support for our results.

---

6. Saramaccan is considered by most creolists to be the most 'radical' creole (see e.g. Byrne 1987) and therefore we expected that by assigning it only <+>s by reverting the values accordingly, it would reinforce its position in the rightmost end of the graphs, in the same way that we expected that assigning only <−>s to English would reinforce its position in the leftmost end of the graph. However, this did not affect the graphs in any significant way.

7. A bootstrap analysis is a statistical resampling method used to estimate support values for a dataset. We ran bootstrap analyses using 10,000 samples. The resulting values for the final branches in Figure 1 (form features) ranged between .51 and 1.0, with only 7 of the final nodes having values below .95. For Figure 2 the bootstrapping support was even higher, with only four final nodes having values below .95, and an overall very narrow range of values between .83 and 1.0.
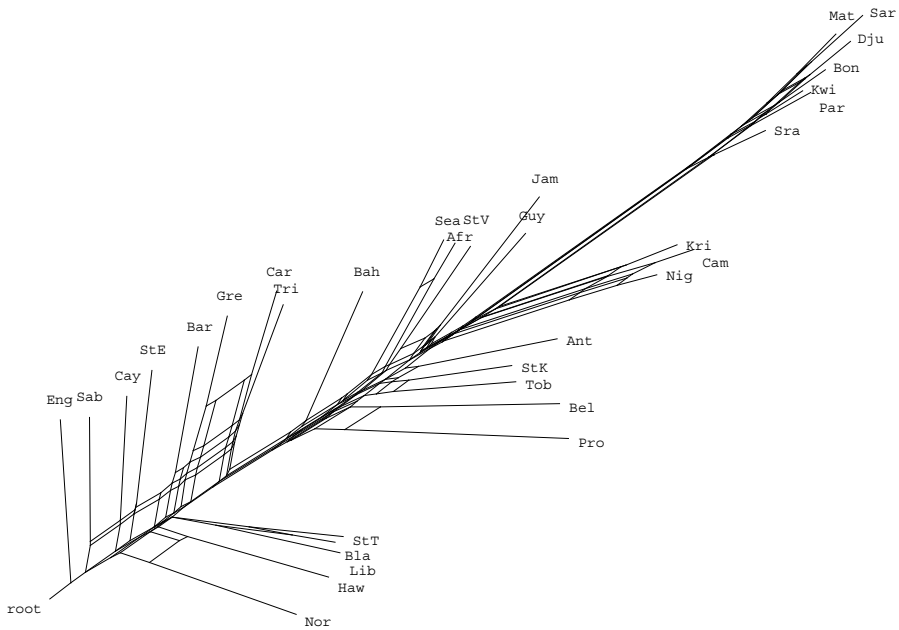
Figure 3. Rooted split network for form and structure combined.

The networks based on structure and on form appeared quite similar. We can conclude that structural features may be safely used for evolutionary studies (cf. Dunn et al. 2005), even though often only lexical-formal data have been used in most classifications.

In the next section we will see that a network based on purely structural features for creoles of different lexifiers shows the typological similarities among creoles.

6.2 How creole-like are the languages in the Holm & Patrick sample?

Holm & Patrick selected 97 features assumed to be typical of creole languages to be included in their survey of creole structures on the basis of previous work on common structures of creole languages. These features are discussed in the chapter texts on the different languages in the book, and are summarized in tables with a <+>, a <−>, a <?> or <R>, meaning 'rare' (see the appendix).

We have counted the number of plusses and minuses for each of the 18 creoles, and the results of this count can be seen in the column 'CCS features' in Table 1, where the languages are presented in alphabetical order. The basic idea now is that the number of features could be taken as a proxy for the degree of creoleness. We have also classified each language according to the socio-historical circumstances,

Table 1. Classification of creole languages and the number of CCS and creole features

Creole	Lexifier	Area	Age	Type	CCS features	Creole features
Angolar (ang)	Portuguese	Atlantic	1550	maroon	65	63
Berbice Dutch (ber)	Dutch	Atlantic	1600	plantation	72	57
Cape Verdean (cap)	Portuguese	Atlantic	1500	trade	71	62
Dominican (dom)	French	Atlantic	1700	plantation	67	55
Guinea Bissau (gui)	Portuguese	Atlantic	1500	fort	73	63
Haitian (hai)	French	Atlantic	1600	plantation	73	55
Jamaican (jam)	English	Atlantic	1650	plantation	73	63
Korlai (kor)	Portuguese	Indian Ocean	1500	fort	54	48
Krio (kri)	English	Atlantic	1780	plantation	81	59
Nagamese (nag)	Assamese	Indian Ocean	1800	trade	57	48
Ndyuka (ndy)	English	Atlantic	1700	maroon	77	60
Negerhollands (neg)	Dutch	Atlantic	1650	plantation	66	59
Nubi (nub)	Arabic	Africa	1850	trade	54	55
Palenquero (pal)	Spanish	Atlantic	1600	maroon	53	55
Papiamentu (pap)	Portuguese	Atlantic	1600	trade	63	58
Seychellois (sey)	French	Indian Ocean	1770	plantation	71	60
Tok Pisin (tok)	English	Pacific	1850	trade	67	58
Zamboangueño (zam)	Spanish	Pacific	1800	trade	53	51

the source language of the lexicon, the area where it is spoken and its age (indicated by the approximate date of its emergence).[8] This classification allows one to investigate whether the typological properties of these creoles, as gauged by the number of features, are independent of these factors. These factors and the respective values for each creole are also listed in Table 1 (the column 'creole features' will be discussed further below). Some of the classifications given in Table 1 are controversially discussed in the literature. In such cases we have settled for what appeared to be the most common view. It will become clear, however, that different codings in controversial cases would not lead to dramatically different results.

---

8. Philip Baker (p.c.) points out that 'several of the creoles in Holm & Patrick begin with input from pre-existing P/Cs': Jamaican, Krio, Tok Pisin, Dominican, Haitian, Seychellois are in all likelihood continuations of earlier pidgins or creoles spoken elsewhere, which would make them older than indicated here. We follow the dates given by the contributors to CCS for reasons of consistency.

Before we turn to a proper statistical analysis[9] let us first look at Table 1. Krio (81), Ndyuka (77), Guinea-Bissau (73), Haitian (73) and Jamaican (73) are the five creoles with the highest scores. This subset includes, impressionistically at least, creoles that are structurally far removed from their lexifiers (Ndyuka) and creoles that are fairly mesolectal, like Jamaican. They cover three lexifiers and three regions. The five languages with the lowest scores are Palenquero (53), Zamboangueño (53), Korlai (54), Nubi (54) and Nagamese (57) — incidentally, all languages that came into being independently from all the others. Here again we find a fair spread with five lexifiers and four regions. An impressionistic look at Table 1 thus seems to support a claim that neither lexifier, type of creole nor region are relevant for the structural properties of creoles (against hypotheses 1 and 2, in favor of 5). Note also that the creoles occupy a rather narrow range of values on the CCS scale, which theoretically ranges from zero to 97. Creole values range only between 53 and 81 (median = 67, SD = 8.7), which means that they scatter only across about one fifth of the whole scale, hence they appear to be a rather homogeneous group of languages.

In order to properly investigate the relation between the degree of creoleness and the other factors, we carried out a multiple linear regression analysis with the number of features as the dependent variable and LEXIFIER, AREA, TYPE and AGE as predictor variables. We use multiple regression as a statistical technique because it is especially well suited to test the influence of many variables at a time (as in this case), namely by calculating the effect of one variable while holding all other variables constant (see, for example, Baayen 2008 for an introduction to multiple regression in linguistics).

The regression model shows that none of the factors has a significant influence on the number of features present in the creole, providing evidence for hypothesis 5 and against 1–4. The model itself does not reach significance either (F-statistic: 1.805 on 11 and 5 DF, p-value: 0.267). In non-sequential analyses of variance (Type II as well as Type III) of the regression model, none of the predictors reaches significance.[10] The result of the anova is documented in Table 2.[11]

---

9. For the statistical analysis, we used the statistical package R (R Development Core Team, 2007).

10. Type III has traditionally been the recommended method in unbalanced designs (cf. Kirk 1968; Roberts and Russo 1999), and thus it is the default used by statistical software such as SPSS and SAS (cf. Langsrud 2003). However, more recent studies have argued that Type II anovas are often to be preferred, with the actual differences being often rather small. In our case, we performed both types of analysis, with basically the same results. We document the Type III results.

11. Note that the multiple regression analyses had to be carried out under exclusion of Nubi, since this was the only language with the value afr. For mathematical reasons, in such situations (called 'singularities') the computation of a linear regression leads to unsatisfactory results.

Table 2. Anova Table (Type III tests) for CCS feature

	Sum Sq	Df	F value	*p* (>F)
(Intercept)	9.17	1	0.1994	0.6739
AGE	9.46	1	0.2057	0.6692
LEX	337.93	5	1.4704	0.3413
AREA	86.97	2	0.9460	0.4483
TYPE	11.37	3	0.0824	0.9667
Residuals	229.83	5		
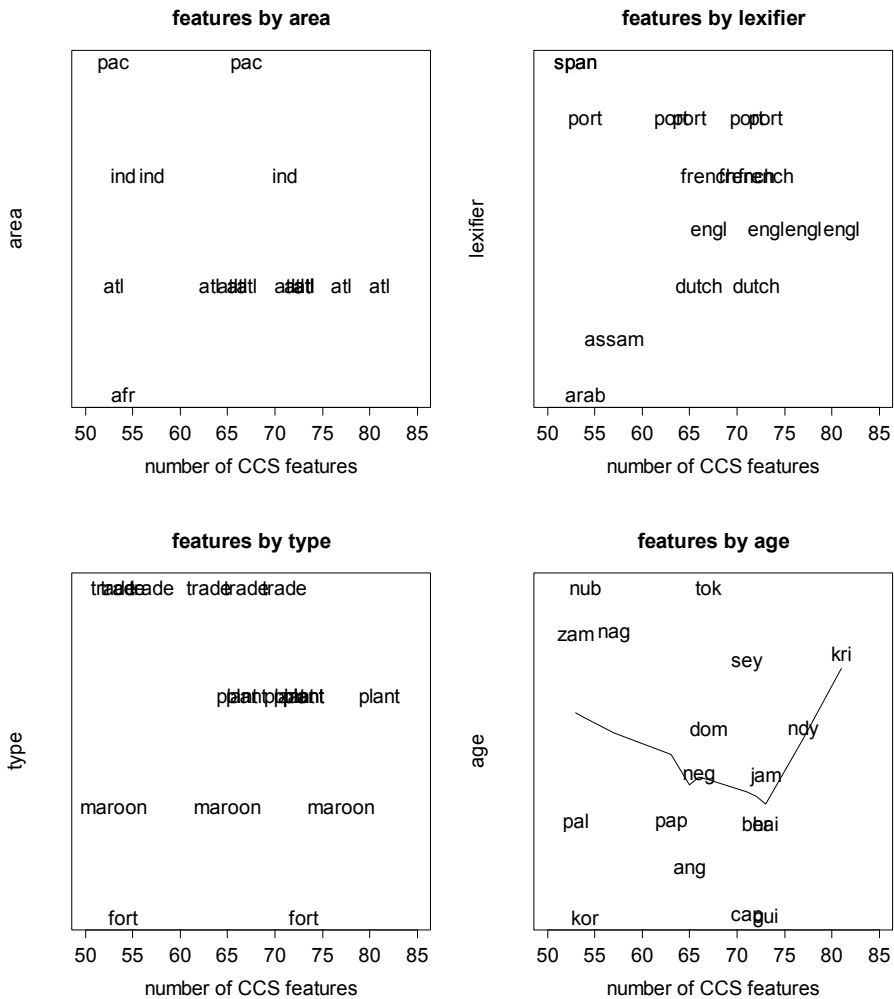
Figure 4. The effect of area, lexifier, type and age on the number of CCS features.

The relations between the predictor variables and the number of CCS features are illustrated in the graphs shown in Figure 4.

On the x-axis we see the number of CCS features, and on the y-axes the four respective predictors. We can see that for those areas for which a number of creoles are in the CCS sample (i.e. Indian Ocean/India and Atlantic) we find a nice spread across the scale of features, suggesting partial evidence against hypothesis 2. As for the lexifiers, we see that Portuguese-based creoles are spread across the lower three quarters of the scale, while the English-based creoles are spread more across the upper half of the scale, and French- and Dutch-based in the middle, which speaks against hypothesis 1. Overall, however, these different tendencies do not reach significance due to the large overlaps. The same holds for type and for age (against hypotheses 3 and 4), with none of these variables clustering significantly with certain values in a particular range of the feature scale. The right plot in the lower panel includes a line reflecting a non-parametric scatterplot smoother fit through the data (Cleveland 1979), showing the main trend in the data, and further illustrates the pertinent result of the statistical analysis, namely that there is no significant relationship between the two variables (cf. also the separate correlation coefficient for these two variables: $rho = -0.11$, $p = 0.66$, Spearman test). Creoles with a rather low number of features can be of different ages and the same is true for creoles with larger numbers of features. This provides evidence against hypotheses 3 and 4.

In sum, there is no significant relation between the degree of creoleness (as measured by the presence of CCS features) and the kind of lexifier, the kind of socio-historical situation, the area or the age of a given creole. This lends independent support to the hypothesis that similarities between creoles do not arise from these factors (against hypotheses 1–4, in favor of hypothesis 5).

One could, however, raise the objection that the more or less mechanical application of plusses and minuses does not really indicate the creoleness of these languages. In fact, it is true that one cannot automatically interpret a <+> as being more creole-like than a <−>. For instance, items 12.6 and 12.7 constitute 'comparison with pass' and 'comparison as in [lexifier]', where a <+> in 12.6 would be a more creole-like structure, and a <+> in 12.7 a non-creole structure. Obviously, none of the languages scored 97 plusses, but it may be interesting to compare the languages with regard to those features that are present in the majority of the languages in the CCS sample.

Therefore we decided to make a new creole hierarchy based only on features that are actually present in most creoles, be they expressed by minuses or pluses. As our criterion for 'typically creole', we used the condition that a feature must be present in at least two thirds of the languages of the CCS sample. This meant that we had to drop 28 features, and we were left with 69 of the original 97 features (be
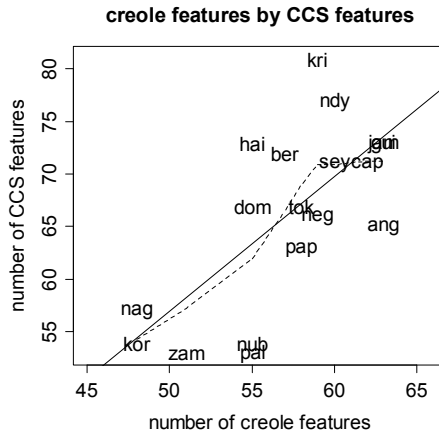
Figure 5. Correlation between creole features in the narrower sense and CCS features.

Table 3. Anova Table (Type III tests) for creole features in the narrower sense.

	Sum Sq	Df	F value	p (>F)
(Intercept)	8.886	1	0.3399	0.5852
AGE	2.572	1	0.0984	0.7664
LEX	98.469	5	0.7533	0.6182
AREA	17.424	2	0.3332	0.7314
TYPE	12.068	3	0.1539	0.9228
Residuals	130.714	5		

they marked by a plus or a minus in CCS).[12] For lack of a better term, we label these features 'creole features in the narrower sense'. For a new regression analysis we now created a new variable CREOLE FEATURE which holds the number of 'creole features in the narrower sense' for each creole. The respective value for each creole is listed in the rightmost column of Table 1. We can see that the creoles occupy only a small proportion of the overall range of zero to 69, they range between 48 and 63, which covers about 24 % of the whole scale (median = 58, SD = 4.7). Again this is evidence for the idea that creoles form a typologically rather homogeneous group of languages.

Let us first compare the CCS FEATURE values with the CREOLE FEATURE values. It turns out that the two variables are highly positively correlated (rho = 0.61,

---

12. The values of the following features were given by minuses: 6.4 irrealis, 8.3 'for' as modal, 10.2 negation, 12.2 and 12.6 adjectives/ verbs, 15.8 and 15.11 NP, 16.1 and 16.3 possession, 17.2, 17.4, 17.5 and 17.6 pronouns, case. An anonymous reviewer pointed out that some of these features, interestingly enough, have been considered central in creole studies.

$p = 0.007$, Spearman). Thus, an increase in the number of CCS features goes together with a higher value for the number of creole features in the narrower sense. The correlation is illustrated in Figure 5, with a non-parametric scatterplot smoother line (broken line) and a straight linear regression line (unbroken line) showing the main trend of the data.

A regression model with CREOLE FEATURE as dependent variable and the four predictors AGE, AREA, TYPE and LEXIFIER[13] shows basically the same result as the model with the CCS FEATURE as dependent variable. The model does not reach
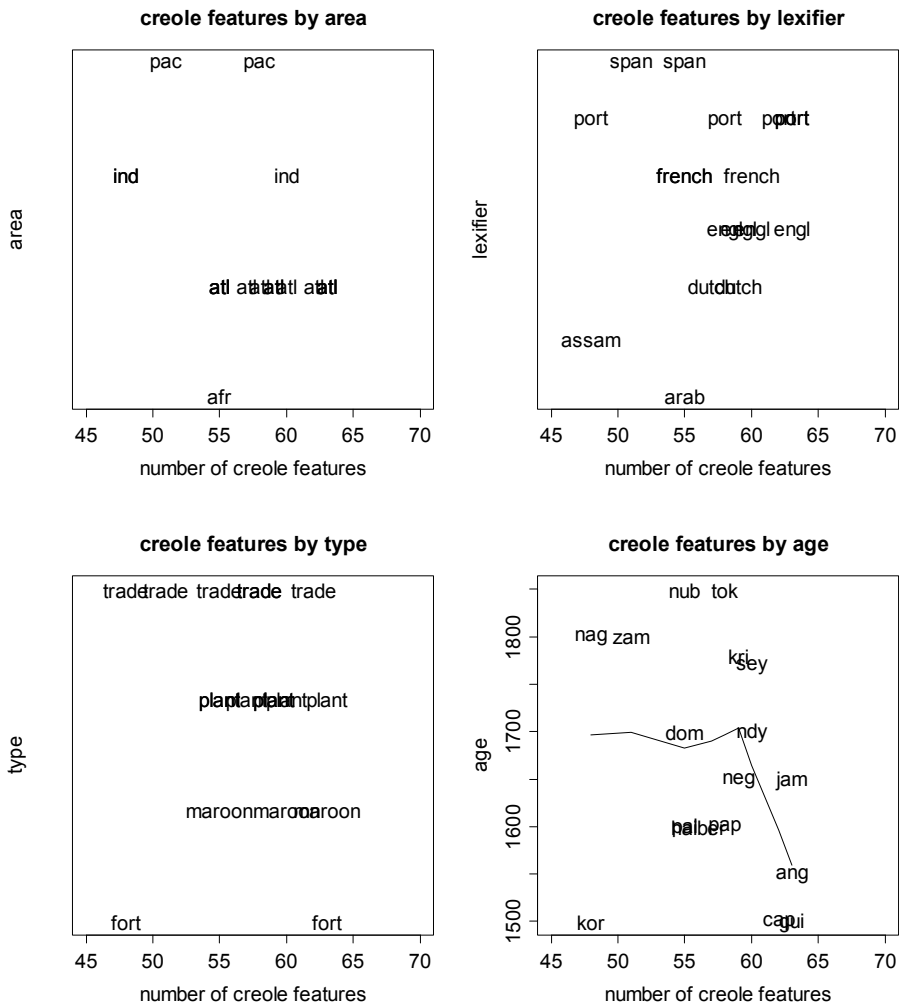

Figure 6. The effect of area, lexifier, type and age on the number of creole features.

---

13. Again, Nubi had to be excluded from the multiple regression analysis.

significance (F-statistic: 0.8305 on 11 and 5 DF, p-value: 0.6312), nor does any of the predictors (AREA, TYPE, AGE or LEXIFIER). The results of the type III anova are documented in Table 3 (again, a type II anova yields very similar results).

The plots in Figure 6 show the relations between the predictors and the target variable.

The results for the creole features in the narrower sense are very similar to the results obtained for the whole of the CCS features.

To summarize this section, it was shown that, no matter whether we use all of the CCS features or only the subset of 'creole features in the narrower sense', we find no statistically significant effects of age, area, lexifier or type on the number of features present in a given creole. Furthermore, it was shown that creoles are very similar to each other with regard to the number of creole features they instantiate. The statistical analysis thus adds fuel to the idea that creoles form a typologically rather coherent group of languages.
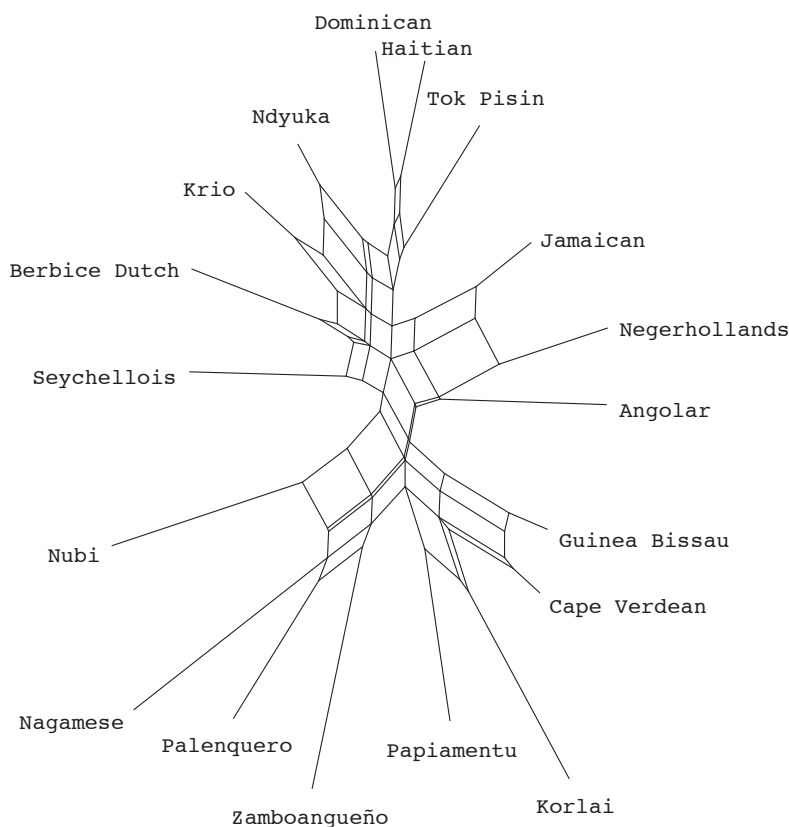

Figure 7. A network of 18 creoles based on the CCS features

6.3 A CCS network

Figure 8 shows a network of the 18 creole languages in the CCS sample in Holm & Patrick (2007) based on 97 features. The network shows that most languages are approximately equidistant from one another (perhaps not surprisingly, as features were selected as typical of creoles), but, except for the historically connected Guinea Bissau Creole and Cape Verdean, no creoles seem especially closely connected. It is also striking that there are no clusters based on type of creole (maroon, fort, plantation), nor clusters based on geographical area (Caribbean, West Africa, East Africa and islands, India), nor clusters based on the lexifiers (see Figure 8a). These results strongly suggest that hypothesis 5 is correct. No precise data on substrates have been used here, but it is well known that creoles spoken in the same region, say the Caribbean, often have similar substrates. Figures 8a and 8b show that lexifiers and substrates play a limited role.

There is indeed **some** clustering according to both lexifier and substrate in general. The Ibero-Romance creoles cluster together (8a), and some Caribbean creoles cluster as well (8b), but generally the relative lack of clustering is equally striking. The Dutch creoles (Negerhollands and Berbice Dutch) are nowhere near one another, and neither are the French-lexicon varieties of the Seychelles or the Caribbean, the latter placing themselves between English-lexicon varieties. The same holds with regard to substrate languages — Tok Pisin and Zamboangueño (with partly and entirely Austronesian substrates, respectively) appear at opposite
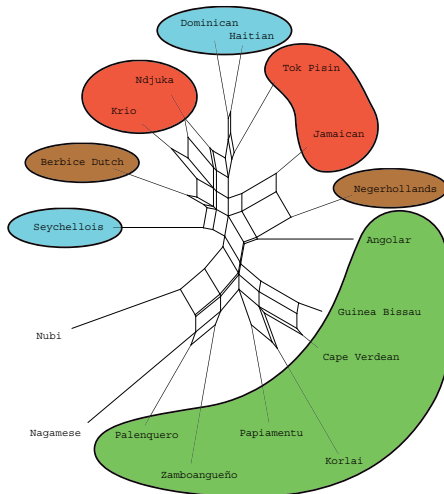
Figure 8a. Partial clustering according to lexifiers.
Green = Ibero-Romance, red = English, blue = French, orange = Dutch
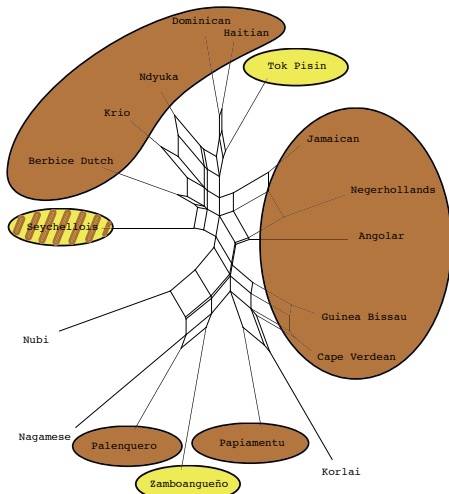
Figure 8b. Partial clustering according to substrates.
Brown = Niger-Congo, yellow = Austronesian (or partly Austronesian), no color: others

ends of the diagram, both occuring between creoles with a Niger-Congo substrate. Korlai Creole similarly splits the Niger-Congo group in two. Even though creoles with a non-Niger-Congo substrate mostly cluster together, the creoles as a group remain a unit. In addition, geographically proximate languages are known to influence one another, and therefore some of the partial clusterings could also reflect areal effects. In other words, creoles are quite similar to one another, and the similarities are not primarily based on lexifier, area or substrate. This result is very similar to that of the regression analysis in Section 6.2. This supports hypothesis 5, and is problematic for hypotheses 1 and 2.

6.4 The typology of creoles and non-creoles: CCS features

Holm & Patrick have identified 97 features that they, or other creolists, have assumed to be more or less typical of creoles. What if we look for these same features in non-creole languages? If creoles form a typological subgroup, we would expect that non-creoles do not display the same features in the same way as creoles do. In order to investigate this, we used a sample of twelve non-creoles (nine unrelated languages and three Niger-Congo languages, one Bantu, one Kwa and one Mande[14]). The non-creoles were chosen on the basis of Ruhlen's (1987) superfamilies of the world, and on the most creole-like profile of the languages. We are aware of the fact that this classification is highly controversial, but that is irrelevant here, as long as one agrees that the chosen languages are genetically unrelated, and that is what all classifications agree on.

The non-creole languages included here are chosen on the basis of one or both of the following two criteria: a relatively isolating structure within the phylum (i.e. similar to the typological profile usually associated with creole languages), or a complexity score around that established for creoles in Parkvall (2008). In his study, Parkvall selected 155 languages on the basis of 46 quantifiable features found in the WALS database (Haspelmath et al. 2005) covering all areas of grammar, from phonology to syntax. The complexity scores thus obtained range from 0.62 (for the most complex language in his sample, Burushaski) to 0.15 (for the least complex language, Sango). Languages in the range 0.38–0.39 can be considered of relative average complexity, while creoles fall within the 0.13–0.33 range. Choosing non-creole languages with a complexity score around that established for creole languages biases the sample of non-creoles in such a way that the non-creoles are rather similar to the creoles. If, in spite of this similarity, the analysis still shows that creoles and non-creoles cluster differently, this would constitute strong evidence that creoles indeed form a typological class distinct from other

---

14. Some Africanists consider the Mande languages as unrelated to Niger-Congo.

Table 4. Non-creole languages selected
Relatively Isolating/ Low complexity score (creoles range from 0.13 to 0.33)

Ainu	Japan; Isolate; Refsing 1986 (complexity score: 0.39; rank 87 out of 155 in Parkvall 2008)
Akan	Ghana; Niger-Congo, Kwa; Christaller 1875
Bambara	Mali; Niger-Congo, Mande; Binger 1886
Brahui	South & Central Asia; Dravidian; Bray 1909 (complexity score: 0.35, rank 120 out of 155)
English	UK, etc..; Indo-European ; (own knowledge); (complexity score: 0.42; rank 78 out of 155 in Parkvall 2008)
Indonesian (standard)	Indonesia; Austronesian; Sneddon 1996 (complexity score: 0.26; rank 146 out of 155 in Parkvall 2008)
Kimbundu	Angola; Niger-Kongo, Bantu; Maia 1964
Kolyma Yukaghir	Russia; Yukaghir, isolate; Maslova 2003; (complexity score: 0.36; rank 112 out of 155 in Parkvall 2008)
Koyra Chiini	Mali; Songhay, Nilo-Saharan; Heath 1998 (complexity score: 0.32; rank 129 out of 155 in Parkvall 2008)
Mina	Cameroon; West Chadic, Afro-Asiatic; Frajzyngier & Edwards 2005
Mandarin	China; Sinitic, Sino-Tibetan; Li & Thompson 1981 (complexity score: 0.39; rank 89 out of 155 in Parkvall 2008)
Pirahã	Brazil; Mura, Amerind; Everett 1986 (complexity score: 0.18, rank 154 out of 155)

languages, even from those non-creoles that are at first sight very similar to them. Languages from other families like Eskimo-Aleut, Turkic or Athabaskan would be less expected to follow the creole profile. The twelve languages that are isolating and/or low in complexity are listed in Table 4. Information about location and classification is added, both Ruhlen's macro-groupings and more conservative groupings, and our information source is indicated.

Figure 9 shows the 18 creole languages of the CCS and as presented in Figure 8, but with the addition of the twelve non-creole languages structurally most similar to creoles. Please note that the internal relations of the creoles are different from Figure 8, after addition of the non-creoles, but here again no subgroupings appear on the basis of social history, age, area or lexifier, against hypotheses 1 to 4. Note that the three Niger-Congo languages do not appear anywhere close to the creole cluster.

Figure 9 shows clearly that morphologically challenged languages and languages within the same range of complexity as creoles do cluster separately from non-cre-
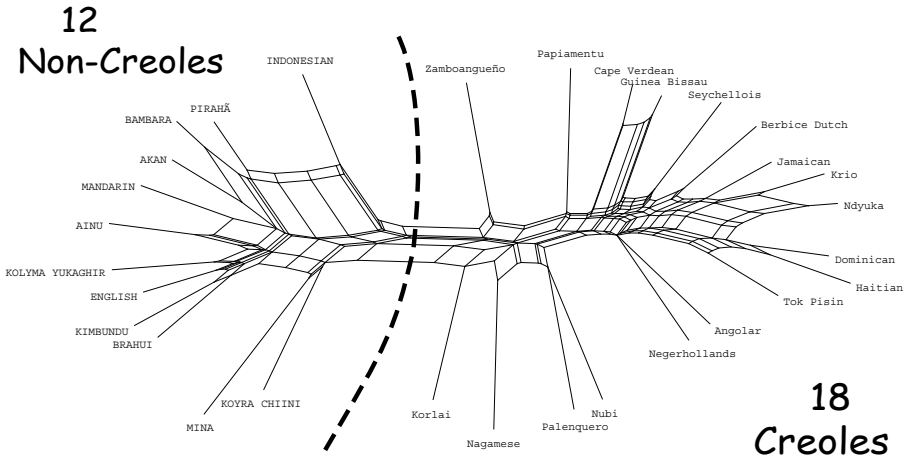
Figure 9. 18 creoles and 12 non-creole languages in a network.

oles, indicating that creoles have more in common than a relative scarce morphology or lower complexity. Again this shows that creoles form a typological class.

The inclusion of Pirahã and Brahui show that complexity is not connected to a lack of morphology. Both of these languages display a considerable amount of morphology, especially in the verb, but overall their complexity score turned out low.

It goes without saying that the addition of, say, Zulu, Georgian or Apache would make the creoles stand out even more. The comparison here involves only languages which stand a fair chance of siding with the creoles — and yet, even those fail the test. This again supports hypothesis 5.

6.5 The typology of creoles and non-creoles: WALS features

In 6.4 we investigated a small subset of the world's language families from a creole perspective, i.e. we started out with features known or thought to be characteristic of creoles. In this section we shall do it the other way around, and take a set of structural-typological features of the languages of the world as our point of departure, in order to see how creoles fit in. If creoles are not a specific typologically distinguishable subset of languages, then we would — again — not expect them to cluster.

In order to investigate this, we used the typological datasets as collected for the *World Atlas of Language Structures* (WALS) (Haspelmath et al. 2005). We did not use all the WALS features, but only the subset selected by Parkvall (2008). Of the 53 features chosen by that author, we discarded the 10 features which were not readily quantifiable (Parkvall's features F01, F02, F04, F05, F12, F18, F20, F26,

F51, F53) in order to take them into account in our own investigation.[15] Parkvall used them to calculate the relative complexity of the 155 languages included in the WALS for which at least 30 features were known (see Parkvall 2008 for the selection criteria). For our investigation we converted the datasets into binary data sets that could be used with the SplitsTree software. Parkvall (2008: 278) added data on 30 creoles and pidgins with diverse lexifiers. Some of these are also found in the CCS sample: Dominican, Guinea Bissau, Haitian, Jamaica, Nubi, Negerhollands, Palenquero and Tok Pisin, i.e. almost half of the Holm & Patrick sample.

The results of the application of the software to the dataset are quite staggering. Whether one uses the software to draw trees or networks, the results invariably cluster all the creoles and pidgins, quite separately from the non-creole languages of the world. In *some* trees/networks, Hmong (South East Asia) clusters with the creoles, but this applies *only* to Hmong and never to *any* other language. This is the case also in Figure 10.

Each language was given a three letter code. All non-creole languages start with N with a three letter code in lowercase, and all pidgins start with P and all creoles start with C and the codes for both are typed in uppercase. Figure 10 shows a network of all 155 languages from WALS (which included two pidgins/creoles, Sango and Ndyuka) plus 32 pidgins and creoles from all over the world and one constructed language. The region where all of the 34 pidgins and creoles (and Hmong) are clustered is visible in Figure 10, marked with a dotted line. The dotted line clearly shows that there is *not a single* pidgin or creole which clusters with any non-creole. Several lexifiers (Arabic, English, French, Spanish, circled in Figure 10) of the creoles of the CCS sample are found in the network as well, but quite far removed from the creoles: 40, 53, 54, and 48 steps out of 76, when the languages are ranked according to their relative distance to the creole cluster. For convenience, we also marked known substrates with stars and the non-creoles used in 6.4 with squares, and both of those are found fairly evenly spread among the non-creoles.

The validity of the results is confirmed by a range of observations.

First, in both tree-drawing and network-drawing, the pidgins and creoles cluster, but within their cluster they are grouped together in distinct ways with the two techniques.

This indicates that, despite the different criteria used in the two drawing methods, the creoles are identified as a group, albeit on different grounds.

Second, even a cursory look at the clustering of non-creoles shows that the non-creoles do not classify along genetic or areal lines. For example, Basque (isolate, Western Europe), Hindi (Indo-European, India/South Asia), Burushaski

---

15. The authors wish to thank Michel DeGraff for pointing out some shortcomings in a previous version of this paper..
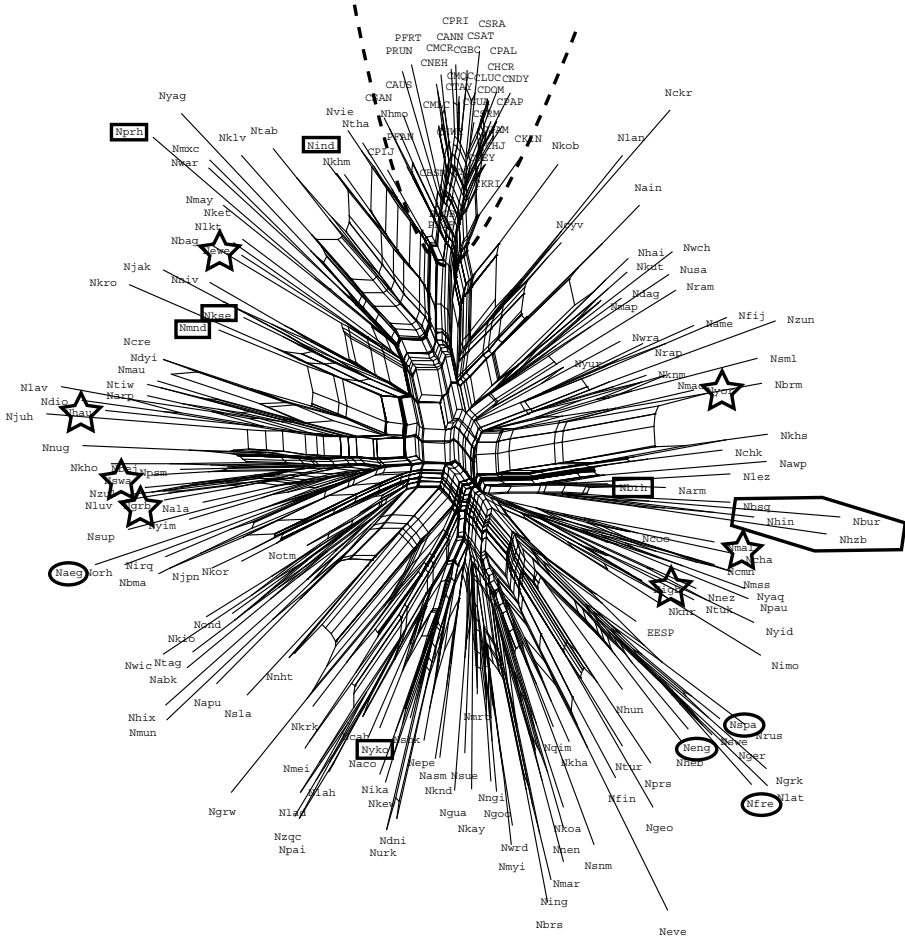
Figure 10. Network of 153 non-creole languages, 34 pidgins and creoles, and Esperanto (EESP).

(isolate, North Pakistan), and Hunzib (East-Caucasian, Caucasus) cluster (boxed on the right side of the network in Figure 10).

One might object that Parkvall's choice of features was biased. Even if they were picked from a representative selection of features established by linguists with no specific interest in creoles, they were included in his study because they might shed light on the levels of complexity in various languages. Still, if it were the case that the label 'creole' had no typological correlate, it ought to be impossible to have selected features that result in such a spectacular clustering.

Moreover, a third observation is highly relevant here: the classifications suggested by the program do *not* follow the complexity scale for creoles (or non-

creoles): the languages do not cluster from lowest to highest complexity,[16] showing that it is not simply complexity that is measured. So, while the original selection of features by Parkvall (2008) had complexity in mind, complexity has no bearing on this case. Furthermore, only 31 non-creole languages (out of 155 of the sample) fall within the complexity range of the 32 creoles included, which shows that there are just as many non-creoles as creoles at the lowest end of his complexity scale.[17] Nevertheless, not all low-complexity languages cluster, only the pidgins and creoles do. The conclusion that creoles (and pidgins, for that matter) are typologically distinct from the languages of the world is inescapable and robust, also from the perspective of the structures of the languages of the world. Again, this supports hypothesis 5.

7.   Conclusion

Parkvall (2008) showed that creoles are a distinct group from non-creoles with regard to complexity. In this paper it is demonstrated for the first time on the basis of a large dataset that creoles constitute a typologically distinct group and on the basis of structural-typological features that were not differentiable with regard to complexity. Whether one takes creole properties and looks for those in non-creole languages, or whether one takes a set of typological properties used for typological research, and then looks at a sample with creoles and non-creoles, the results are the same: *creoles stand out*. Creoles are languages that can be identified on the basis of structural features not found in the same constellation in non-creoles.

Ansaldo & Matthews (2007: 14) wrote: 'creole exceptionalism is a set of sociohistorically-rooted dogmas, with foundations in (neo)colonial power relations, *not a scientific conclusion based on robust empirical evidence*' (our emphasis). We have shown that creole exceptionalism, or rather distinctiveness, is not a dogma, not an ideology, but a fact. The robust empirical evidence demanded has now been delivered. Those who defend the view that creoles are not typologically distinct from non-creole languages, rarely use typological arguments or language data to prove their point. One of the fiercest opponents of creole exceptionalism stated:

---

16. Not surprisingly, most languages close to the creole cluster have a low complexity score, but not all. The ten languages closest to the creoles are Thai (ranking 132 of 152, complexity 0.30), Vietnamese (147; 0.26), Khmer (130; 0.33), Indonesian (146; 0.26), Taba (133; 0.30), Kobon (152; 0.20), Lango (121; 0.34), Canela-Crahô (125; 0.33), Cayuvava (139; 0.29), Ainu (87; 0.39). Hmong Njua scores 153; 0.20.

17. For example, Wichi is fairly close to the creole cluster, but with a score 0.43 it ranks in the more complex half (74).

'There are plenty of linguistic reasons to hope that creolistics can make a contribution to general linguistics, in the same way that one can select any *arbitrary set of languages* and hope to make a contribution to, for instance, linguistic typology (…).' (our emphasis) (Mufwene 2003: 284–285). We think we have done just that — and proven that creoles are not an arbitrary set of languages.

Creoles can be distinguished from other languages of the world on structural grounds. Creoles are a typologically coherent class of languages. Why would that be? The obvious explanation that suggests itself lies in the shared aspects of their sociohistorical origin. Probably all students of creole languages, including those who as yet do not recognize them as a synchronically valid distinct grouping, share the opinion that creoles have particular comparable historical events in common (although opinions vary regarding how extraordinary those events were).

In all cases, our interpretation is that groups or individuals without a common language had to create a new means of communication, and this process seems to involve getting rid of irregularities, of inflections, and, more generally, of features unnecessary for (elementary) communication. In the long run, such simplified forms of interethnic makeshift languages were insufficient for communication, both for direct communication and the more indirect forms of communication commonly expressed by accents, styles, dialectal features and the like. The users of the reduced forms of communication used their creativity to fill in gaps and create novel means of expressing necessary or just handy grammatical distinctions, for example by introducing them from other languages, by grammaticalizing lexical words and by regularizing some of the variation.

Despite the obvious merits of the list of creole properties in Holm & Patrick (2007), it is not perfect — some of the features, for instance those for the tense-mood-aspect systems, are dependent on one another, some are more relevant for Atlantic creoles, etc. It could thus be that we would benefit from similar studies being carried out with a different, improved list of creole properties. A study involving more languages would also further bolster our claims. The results for creoles as a typological group are very robust. In the meantime, there are a number of additional studies we are working on that should strengthen the conclusions, e.g. on substrates and on Hmong. Some creolists might propose that many of the properties common to creoles are due to them having the same (e.g. West African), or typologically partially similar (e.g. West African vis-à-vis Austronesian) substrates. This can be falsified by using a sample of substrate languages and compare them with creoles. Substratists would predict the individual creoles to side with their substrates, but we hypothesize that the creoles would again form their own grouping, with the substrates in one or more other groupings. As we saw above, the same argument can be applied to superstratists, i.e. those who see the lexifier as the source for most creole structures.

The study based on WALS (in 6.5) was only based on the quantifiable features. This is justified in a study focusing on complexity, but more representative data would be preferable. An expansion to all the WALS features would therefore be interesting.

Finally, the *Atlas of Pidgin and Creole Language Structures*, to be published in the near future, will supply new databases that can be used in additional tests of the structural uniqueness of creoles. We doubt, however, that any empirical study using sufficiently large samples of creoles and non-creoles, will reach opposite conclusions.

Acknowledgements

References

Aboh, Enoch & Umberto Ansaldo. 2007. The role of typology in language creation: a descriptive take. In Umberto Ansaldo, Stephen Matthews, & Lisa Lim (eds.), *Deconstructing Creole*, 39–66. Amsterdam: John Benjamins Publishing Company.

Ansaldo, Umberto. 2004. Contact, typology and the speaker: the essentials of language. *Language Sciences* 26. 485–494.

Ansaldo, Umberto, Stephen Matthews, & Lisa Lim (eds.). 2007. *Deconstructing Creole*. Amsterdam: John Benjamins Publishing Company.

Ansaldo, Umberto & Stephen Matthews. 2007. Deconstructing creole: the rationale. In Umberto Ansaldo, Stephen Matthews & Lisa Lim (eds.), *Deconstructing Creole*, 1–18. Amsterdam: John Benjamins Publishing Company.

Atkinson, Quentin D. & Russell D. Gray. 2005. Curious parallels and curious connections — Phylogenetic thinking in biology and historical linguistics. *Systematic Biology* 54 (4) 513–526.

Baayen, R. Harald. 2008. *Analyzing linguistic data. A practical introduction to statistics.* Cambridge: Cambridge University Press.

Baker, Philip. 1988. Combien y a-t-il eu de genèses créoles à base lexicale française? *Études Créoles* 10 (2). 60–76.

Baker, Philip. 1993. Australian influence on Melanesian Pidgin English. *Te Reo* 36. 3–67.

Baker, Philip. 1999. Investigating the origin and diffusion of shared features among the Atlantic English Creoles. In Philip Baker & Adrienne Bruyn (eds.), *St-Kitts and the Atlantic Creoles,* 315–365. London: Battlebridge.

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

Bakker, Peter. 2004. Phoneme inventories, language contact and grammatical complexity: a critique of Trudgill. *Linguistic Typology* 8 (3). 368–375.

Bakker, Peter. 2008. Pidgins versus Creoles and Pidgincreoles. In Silvia Kouwenberg & John Singler (eds.), *The Handbook of Pidgin and Creole Studies*, 130–157. Oxford: Wiley-Blackwell.

Bakker, Peter. 2009. Phonological complexity in pidgins. In Nicholas Faraclas & Thomas Klein (eds.), *Simplicity and Complexity in Pidgin and Creole Languages*. London: Battlebridge.

Bandelt, H.-J. & A. Dress. 1992. Split-decomposition: A new and useful approach to phylogenetic analysis of distance data. *Molecular Phylogenetics and Evolution* 1. 242–252.

Bickerton, Derek. 1981. *Roots of Language*. New York: Karoma Publishers.

Bickerton, Derek. 1984. The language bioprogram hypothesis. *The Behavioral and Brain Sciences* 7. 173–188.

Binger, Louis-Gustave. 1886. *Essai sur la langue Bambara, parlée dans le Kaarta et dans le Bélédougou, suivi d'un vocabulaire*. Paris: Maisonneuve Frères et Ch. Leclerc.

Bobyleva, Ekaterina. 2009. *Substrate vs. superstrate competition in creole genesis: what determines the winner? Evidence from creole DP*. Paper presented at the Eighth Creolistics Workshop: *Pidgins and creoles in a comparative perspective*, *University of Giessen, Germany*, 2–4 April 2009.

Bray, Denys S. 1909. *The Brahui Language. Vol. 1: Introduction and Grammar*. Calcutta. Reprinted Delhi 1986: Gian.

Byrne, Francis. 1987. *Grammatical Relations in a radical Creole. Verb complementation in Saramaccan*. Amsterdam: John Benjamins Publishing Company.

Chaudenson, Robert. 2001. *Creolization of language and culture*. London and New York: Routledge.

Chaudenson, Robert. 2003. *La créolisation: théorie, applications, implications*. Paris: L'Harmattan.

Christaller, Johann Gottlieb. 1875. *A Grammar of the Asante and Fante Language Called Tshi [Chwee, Twi]: Based on the Akuapem Dialect with Reference to the Other (Akan and Fante) Dialects*. Basel: Basel Evangelical Missionary Society.

Cleveland, W. S. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74. 829–836.

Coelho, Adolpho. 1880. Os Dialectos Romanicos ou Neo-Latinos na Africa, Asia e América. *Lisboa* 2. 129–196 (1880–1881); 3. 451–478 (1882); 6. 705–755 (1886). Reprinted in Jorge Morais-Barbosa (ed.), *Estudos linguísticos crioulos.* Lisbon: Academia Internacional de Cultura Portuguesa.

Daval-Markussen, Aymeric & Peter Bakker. (2011). A phylogenetic networks approach to the classification of English-based Atlantic creoles. *English World-Wide* 32 (2), 115–146.

DeGraff, Michel. 2001. On the origin of Creoles: A Cartesian critique of 'neo'-Darwinian linguistics. *Linguistic Typology* 203. 213–310.

DeGraff, Michel. 2003. Against Creole Exceptionalism. *Language* 79 (2). 391– 410.

DeGraff, Michel. 2005a. Linguists' most dangerous myth: The fallacy of Creole exceptionalism. *Language in Society* 34 (4). 533–591.

DeGraff, Michel. 2005b. Morphology and word order in 'creolization' and beyond. In Guglielmo Cinque & Richard Kayne (eds.), *Oxford Handbook of Comparative Syntax*, 293–372. Oxford: Oxford University Press.

Donohue, Mark, Søren Wichmann, & Mihai Albu. 2008. Typology, areality and diffusion. *Oceanic Linguistics* 47 (1). 223–232.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley, & Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309. 2072–2075.

Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink, & Angela Terrill. 2008. Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language* 84 (4). 710–759.

Everett, Dan. 1986. Pirahã. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages*, Vol. 1, 200–325. Berlin: Mouton de Gruyter.

Farquharson, Joseph. 2007. Creole morphology revisited. In Umberto Ansaldo, Stephen Matthews, & Lisa Lim (eds.), *Deconstructing Creole*, 21–37. Amsterdam: John Benjamins Publishing Company.

Frajzyngier, Zygmunt & Adrian C. Edwards. 2005. *A Grammar of Mina*. Berlin: Mouton de Gruyter.

Goury, Laurence & Isabelle Léglise. 2005. Introduction. *Études Créoles* 28 (1). 9–19.

Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426. 435–439.

Gray, Russell D., A. J. Drummond, Simon J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323. 479–483.

Hagège, Claude. 2001. Creoles and the notion of simplicity in human languages. *Linguistic Typology* 5 (2/3). 167–175.

Hall, Robert. 1966. *Pidgin and creole languages*. Ithaca: Cornell University Press.

Hammarström, Harald. 2008. Complexity in numeral systems with an investigation into pidgins and creoles. In Matti Miestamo, Kaius Sinnemäki, & Fred Karlsson (eds.), *Language Complexity. Typology, Contact, Change*, 287–304. Amsterdam: John Benjamins Publishing Company.

Hancock, Ian. 1987. A preliminary classification of the anglophone Atlantic creoles with syntactic data from thirty-three representative dialects. In Glenn Gilbert (ed.), *Pidgin and Creole Languages. Essays in Memory of John E. Reinecke,* 264–333. Honolulu: University of Hawaii Press.

Haspelmath, Martin, Matthew Dryer, David Gil, & Bernard Comrie (eds.). 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Heath, Jeffrey. 1998. *A Grammar of Koyra Chiini: The Songhay of Timbuktu*. Berlin: Mouton de Gruyter.

Holm, John & Peter L. Patrick (eds.). 2007. *Comparative Creole Syntax. Parallel outlines of 18 creole grammars*. London: Battlebridge.

Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, Andre Müller, & Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42 (2). 331–354.

Huson, Daniel H. & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23 (2). 254–267.

Jacobs, Bart. 2009. The Upper Guinea origins of Papiamentu. Linguistic and historical evidence. *Diachronica* 26 (3). 319–379.

Kirk, Roger E. 1968. *Experimental design procedures for the behavioral sciences.* Belmont: Brooks & Cole.

Klein, Thomas. 2005. *Phonological typology of phonemes in Portuguese- and Spanish-lexified creole languages.* Paper presented at the Annual Meeting of the Association for Portuguese and Spanish lexically based creoles. Orléans, France, June 2005.

Klein, Thomas. 2006a. Creole phonology typology: phoneme inventory size, vowel quality distinctions and stop consonant series. In Parth Bhatt & Ingo Plag (eds.), *The structure of creole words: segmental, syllabic and morphological aspects,* 3–21. Tübingen: Niemeyer.

Klein, Thomas. 2006b. Segmental typology of African Creole languages. In Olaoba Frank Arasanyin & Michael Pemberton (eds.), *Selected Proceedings of 36th Annual Conference on African Linguistics (ACAL* 36), 42–50. Somerville, Ma: Cascadilla.

Klein, Thomas. This volume. Diversity and complexity in the typology of syllables in Creole languages.

Langsrud, O. 2003. ANOVA for unbalanced data: use Type II instead of Type III sums of squares. *Statistics and Computing* 13 (2). 163–167.

Lefebvre, Claire. 2000. What do creole studies have to offer to mainstream linguistics? *Journal of Pidgin and Creole Languages* 15 (1). 127–153.

Li, Charles & Sandra A. Thompson. 1981. *Mandarin Chinese. A Functional Reference Grammar.* Berkeley: University of California Press.

McMahon, April & Robert McMahon. 2003. Finding families: Quantitative methods in language classification. *Transactions of the Philological Society* 101. 7–55.

McMahon, April & Robert McMahon. 2006. Keeping contact in the family: Approaches to language classification and contact-indiced change. In Yaron Matras, April McMahon, & Nigel Vincent (eds.), *Linguistic areas. Convergence in Historical and Typological Perspective,* 51–74. Houndmills Basingstoke: Palgrave MacMillan.

McMahon, April, Paul Heggarty, Robert McMahon, & Warren Maguire. 2007. The sound patterns of English: representing phonetic similarity. *English Language and Linguistics* 11. 113–143.

McWhorter, John H. 1995. Sisters under the skin: a case for genetic relationship between the Atlantic English-based Creoles. *Journal of Pidgin and Creole Languages* 10 (1). 289–333.

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

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5 (2/3). 125–166. With peer commentaries, 167–387.

McWhorter, John H. 2005. *Defining Creole.* New York: Oxford University Press.

Maia, António da Silva. 1964. *Lições de gramática de quimbundo: português e banto, dialecto omumbuim.* Cucujães: Escola Tipográfica.

Maslova, Elena. 2003. *A grammar of Kolyma Yukaghir.* Berlin: Mouton de Gruyter.

Mufwene, Salikoko S. 1986. Les langues créoles peuvent-elles être définies sans allusion à leur histoire? *Etudes Créoles* 9 (1). 135–150.

Mufwene, Salikoko. 1999. Accountability in descriptions of creoles. In John Rickford & Suzanne Romaine (eds.), *Creole genesis, attitudes and discourse,* 157–185. Amsterdam: John Benjamins Publishing Company.

Mufwene, Salikoko S. 2001. *The Ecology of Language Evolution.* Cambridge: Cambridge University Press.

Mufwene, Salikoko S. 2003. Genetic linguistics and genetic creolistics: A response to Sarah G. Thomason's 'Creoles and genetic relationships'. *Journal of Pidgin and Creole Languages* 18 (2). 273–288.

Mufwene, Salikoko. 2007a. Population movements and contacts in language evolution. *Journal of language contact* 1. 63–91.

Mufwene, Salikoko S. 2007b. 'Les créoles : de nouvelles variétés indo-européennes désavouées ?' *Actes du colloque 'Créolisation linguistique et Sciences Humaines*,' ed. by Marie-Paul Ansie, 59–70. Paris: Presses Universitaires Haïtiano-Antillaises.

Muysken, Pieter. 1988. Are creoles a special type of language? In Frederik Newmeyer (ed.), *Linguistics: the Cambridge Survey II*, 285–302. Cambridge: Cambridge University Press.

Nichols, Johanna & Tandy Warnow. 2008. Tutorial on Computational Linguistic Phylogeny. *Language and Linguistics Compass* 2/5. 760–820.

Parkvall, Mikael. 2008. The simplicity of creoles in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki, & Fred Karlsson (eds.), *Language Complexity. Typology, Contact, Change*, 265–285. Amsterdam: John Benjamins Publishing Company.

Plag, Ingo. This volume. Creolization and admixture: Typology, feature pools, and second language acquisition.

R Development Core Team. 2007. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Refsing, Kirsten. 1986. *The Ainu Language: The Morphology and Syntax of the Shizunai Dialect*. Aarhus: Aarhus University Press.

Rijkhoff, Jan & Dik Bakker. 1998. Language sampling. *Linguistic Typology* 2. 263–314.

Roberts, Maxwell J. and Russo, Ricardo. 1999. *A student's guide to analysis of variance*. London: Routledge.

Ruhlen, Merritt. 1987 [1991]. *A Guide to the World's Languages: Volume 1: Classification; with a postscript on recent Developments*. London: Edward Arnold.

Schuchardt, Hugo. 1914. Die Sprache der Saramakkaneger in Surinam. *Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam. Afdeeling Letterkunde*, n.s. 14(6). Amsterdam: Johannes Müller.1914 [1979: 73f., 77]). Partly reprinted In Glenn Gilbert (ed., 1980). *Pidgin and Creole languages: Selected essays by Hugo Schuchardt*, 89–126. Cambridge: Cambridge University Press.

Selbach, Rachel, Hugo Cardoso, & Margot Van den Berg (eds.). 2009. *Gradual Creolization. Studies celebrating Jacques Arends*. Amsterdam: J. Benjamins.

Smith, Norval. 1987. *The Genesis of the Creole Languages of Surinam*. Amsterdam, Netherlands: University of Amsterdam dissertation.

Sneddon, James N. 1996. *Indonesian. A Comprehensive Grammar*. London: Routledge.

Winford, Donald. 2005: Contact-induced changes. *Diachronica* 22 (2). 373–427.

Winford, Donald. 2008: Atlantic Creole Syntax. In Silvia Kouwenberg & John Singler (eds.), *Handbook of Pidgin and Creole Studies*, 19–47. Oxford: Blackwell.