Explorations in creole research with phylogenetic tools

Aymeric Daval-Markussen
Research Centre for Grammar and Language
Use, Aarhus University
aymemeric@gmail.com

Peter Bakker
Research Centre for Grammar and Language
Use, Aarhus University
linpb@hum.au.dk

Abstract

The principal goal of this paper is to illustrate various ways in which phylogenetic tools can advantageously be put to use in investigating and visualizing the relationships of creole languages to other languages, both creoles and non-creoles. After introducing a test study on the English-based creoles, the major theories seeking to explain the emergence and development of creoles will be reviewed and assessed. The final part of the paper is concerned with the typological status of creoles, where various samples will be used to show that creoles form a typologically coherent group among the world's languages.

1. Introduction

Although the linguistic processes underlying creolization remain far from being fully understood, computational methods offer today the opportunity of uncovering complex mechanisms of language evolution on the basis of quantitative investigations. More sophisticated and powerful algorithms now available enable the visualization of patterns in a straightforward manner that was not possible before.

Creole languages emerged in situations of intense contact between several languages, more often than not in the context of massive forced population displacements as were typical of European slave-trading ventures. A diglossic situation with a high-prestige variety (the superstrate) and several low-prestige languages (the substrates) characterized the settings in which creoles developed. Therefore, creole languages can be said to have several parents, and possess as well many often recurring features that appear *ex nihilo*. Thus, the problem of determining relationships between creole languages and other creoles or unrelated languages has long haunted creolists and been recognized as one of the challenges in the field.

Following recent developments in creolistics, where phylogenetic networks were used to investigate questions inherent to the field (Bakker et al. 2011, Daval-Markussen 2011, Daval-Markussen and Bakker 2011), the aim of this paper is to argue that creole languages offer an unparalleled venue for exploratory research in language evolution, and that available computational tools now permit to graphically represent the relationships between the languages considered. In our demonstration, we will exemplify various ways in which phylogenetic networks may advantageously be used to visualize the results.

Following the argumentation in Daval-Markussen (2011: 6-13), only structural features will be taken into account in the present study, since the lexical stock of a creole is mainly derived from a single source language, and this would likely be reflected in the resulting graphs.

In the first part of the paper, phylogenetic tools are used to represent the relationships between 33 English-based creoles, for which 62 typological features were selected and encoded binarily[1]. The second part examines the various scenarios proposed to account for the emergence of creole languages in the light of phylogenetic networks. To this end, samples of various sizes and including creoles as well as non-creole languages (mostly languages involved in the emergence of creoles) were used in order to visualize the impact of the various languages present in the contact situation on the new vernaculars. The final part deals with the typological status of creoles, a topic hotly debated in creolistics (e.g. DeGraff 2003; McWhorter 1998, 2011). Basing our analysis on samples of languages selected from the *World Atlas of Linguistic Structures* (Dryer and

---

[1] These correspond to the structural features described in Daval-Markussen and Bakker (2011).

Haspelmath 2011, hereafter WALS), we will show that phylogenetic methods represent a unique tool for exploring the relationships between creole languages and other languages of the world and can provide invaluable insights into questions on language birth and evolution.

2. Classification

A number of studies have sought to classify English-based creoles on the basis of shared similarities (Hancock 1969, 1987; McWhorter 1995; Baker 1999 to mention just a few). The results presented in these studies are in accordance on several higher-level groupings (the West African, the Suriname, and the Eastern/Western Caribbean groups). We expect therefore to find similar groupings in our results, even though we are using different features and a different method.

The languages selected for this investigation are presented in Hancock's (1987) seminal study on the relationships between 33 Atlantic English-based creoles[2]. On the basis of these data, Hancock (1987: 324-325) attempted to construct a historical scenario explaining their distribution. Although he summarized the results in a tree structure, it is worth mentioning that Hancock indicated the influences between varieties with dotted lines, thus foreshadowing the approach advocated here, i.e. when investigating relationships between (especially creole) languages, lateral influences must be taken into account and somehow be graphically depicted.

The 33 languages were analyzed and 62 typological features attesting the presence vs. absence of a particular phenomenon were selected for binary encoding (see also the Supplementary Materials).

The data were used as input for the software SplitsTree v. 4.12.3 (Huson and Bryant 2006) and returned the network presented in Fig. 1. The geographic location of each variety is indicated with the following colors: blue = Leeward Is., brown = Pacific; green = Caribbean; grey = Suriname; pink = Windward Is.; red = West Africa; yellow = mainland US.

Several clusters are immediately evident in the network reproduced in Fig. 1. While the color codes help visualize the geographic distribution of the included languages, the varieties which appear closest to one another also correspond to
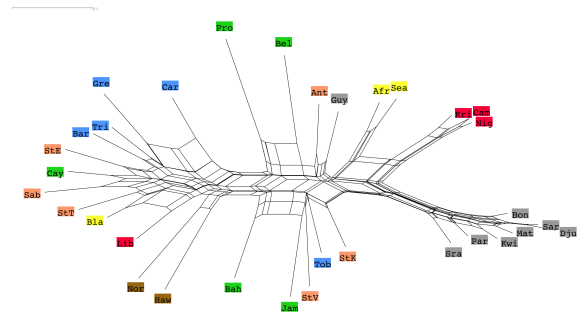

Figure 1: Phylogenetic network of 33 English-based creoles with 62 typological features

groupings identified by other authors (such as the Suriname creoles in grey in the lower right side of the graph). In several cases, we can observe disrupted groups, such as Afr/Sea and Bla (in yellow), or Kri/Nig/Cam and Lib (in red) in clusters which correspond to genealogical groupings (e.g. the Suriname creoles and the varieties of West Africa). The reason for these discrepancies is to be found in the histories of these vernaculars, which have developed apart from one another (see Daval-Markussen and Bakker 2011 for an overview). Besides, the results in Fig. 1 go against the conclusions of Donohue et al. (2011), who claim that the various clusterings observable in phylogenetic networks are due to the effects of areality and geography rather than to genealogy.

This indicates that phylogenetic networks can confidently be used to shed light on the relationships between creoles by presenting the results in such a visually appealing manner.

3. The challenges of creolistics

The present section focuses on the various explanations proposed to account for the similarities observed between creoles, an issue which has been central to creolistics ever since its beginnings. How creoles came about is still a matter of controversy, and in practice, most creolists agree on a working definition encompassing both the linguistic and sociohistorical aspects of creoles.

The main theories seeking to explain how creoles came into being claim that the languages that have played a major role in the creation of the nascent vernaculars were: i) the superstrate, or lexifier (the superstratist school); ii) the substrate languages spoken by the displaced populations (the substratist account); iii) only the superstrate and substrates provide the features available for competition and selection and

---

[2] The full list of languages and features used throughout the essay with the corresponding abbreviations is found in the Supplementary Materials.

nothing else (the Feature Pool hypothesis); and finally, iv) no language in particular and creole similarities are to be explained by restructuring universals, where similar solutions were found in order to optimize communication (the universalist approach). These approaches are not necessarily mutually exclusive (e.g. Mufwene 1986).

In order to determine whether the predictions made by the various theories accounting for the emergence and development of creoles are borne out by the facts, a sample of creoles and non-creole languages was carefully selected and binary oppositions were encoded according to the 97 morphosyntactic features presented in Holm and Patrick's *Comparative Creole Syntax* (Holm and Patrick 2007, henceforth CCS). The 18 creoles originally described in the Holm and Patrick volume were included, as well as languages known to be involved in the creation of the creoles. Apart from the 18 CCS creoles, we included 19 substrates, 7 lexifiers, as well as 8 non-creoles, selected because of their analytic character and relative low complexity so as to match the character of creoles (see Bakker et al. 2011).

In order to facilitate the interpretation of the figures, it should be kept in mind that the abbreviations used in each network provide the following information: a three-letter code was attributed to each language and written in upper-case for creoles and in lower-case for all non-creoles. A further distinction is made in all the abbreviations, where an initial capital letter (L, S or X) indicates whether the language is a lexifier, a substrate or a non-creole respectively, while a lower-case 'c' precedes all the abbreviations for the creoles.

In the following, with the help of phylogenetic networks, we will test the predictions made by each of the four major proposals seeking to account for the emergence and development of creoles.

3.1 The superstratist view

The main idea within the superstratist framework is that the structural similarities between creoles are due to the role played by the superstrate (or lexifier) language in the period of creole formation. Thus, in this view, creoles are mere continuations of the European languages which provided the bulk of their lexicon and are thus genetically related to their lexifiers. Moreover, according to Mufwene (e.g. 2000, 2008) and Chaudenson (1992, 2003), creolization results

from normal language change under particular sociohistorical circumstances and is more a sociological process rather than a linguistic one, and therefore creoles are in this view indistinguishable from non-creoles structurally and typologically.

In order to test the validity of the superstratist approach, we produced a network
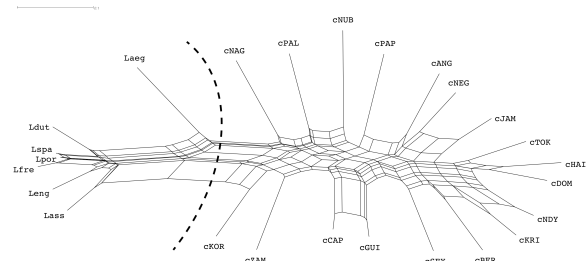

Figure 2: A network of 18 creoles and 7 lexifiers

including the 18 creoles of the CCS sample with the seven lexifiers involved in their creation. Fig. 2 shows the resulting network.

The seven lexifiers all cluster together on the left hand of the network, separated from the creoles by a curved line. All the Indo-European languages cluster neatly together, while Arabic (Laeg) shows up further removed, and the creoles all appear away from the lexifiers. Obviously, the visual interpretation of the network in Fig. 2 does not support the superstratist view, since the creoles do not group with their respective lexifiers, as otherwise expected. This strongly suggests that the superstrates have had a rather limited influence on the grammatical makeup of the incipient creoles at the time of restructuring.

3.2 The substratist position

The substratist school of thinking emphasizes the role of the substrate languages involved in the creation of a creole, which, in this view, was highly influenced by the languages of the enslaved populations (e.g. Holm 1989). Obvious influences are found in the lexicons of individual creoles in the form of borrowings and syntactic structures such as serial verb constructions have also been claimed to be inherited from substrate languages (e.g. Sebba 1987).

In order to assess the extent to which the substratist approach is able to account for the resemblances between creoles, we will examine a network including the 18 CCS creoles with a set of 19 languages which have been claimed to be substrate languages of the various creoles.
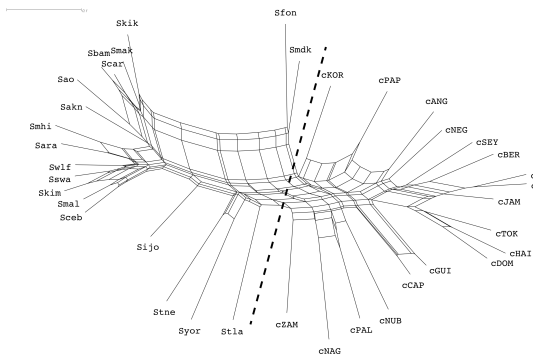
Figure 3: Phylogenetic network of 18 creoles and 19 substrates

The network in Fig. 3 shows an obvious clustering of all the creoles to the right of the dotted line, while all the substrates appear on the left side. Several West African languages often mentioned in the context of creoles are found in the vicinity of the creole cluster and form a transition zone between creoles and non-creoles. Fon (Sfon), Ijo (Sijo), Mandinka (Smdk), Temne (Stne) and Yoruba (Syor) are known to share some structures with creoles (they all use preverbal TMA markers for instance), and this is clearly reflected in the network, where they appear between the creole and non-creole clusters.

Similarly, Tolai (Stla) appears relatively close to the creoles without affecting its relative position to Tok Pisin (cTOK), hence suggesting that the substratist approach fails to fully account for the facts.

3.3 The feature pool approach

The main proponent of the feature pool approach is Mufwene (e.g. 2001, 2008 - see also Aboh and Ansaldo 2006), who advocates a view inspired by genetics and applied to language change. In this context, languages are conceived of as biological species, and processes of language change are explained through the lens of population genetics and Darwinian evolution. Hence, in this view, the roles of the dominant substrates and non-standard varieties of the lexifier are critical, since they provide the feature pool where particular items compete for selection, first in individual idiolects, and then in the wider linguistic community (Mufwene 2008).

In order to test the validity of the feature pool approach, two languages which are known substrates of Seychellois (cSEY), Makhuwa (Smak) and Malagasy (Smal), were specifically included, as well as a third substrate language,

Swahili (Sswa), which was encoded for being a suggested substrate for another creole, Nubi Arabic (cNUB). In this sense, we voluntarily tipped the balance in favor of the languages that we know were present at the time when Seychellois emerged and must therefore have provided the feature pool from which the various items were available options for competition and selection in this framework.
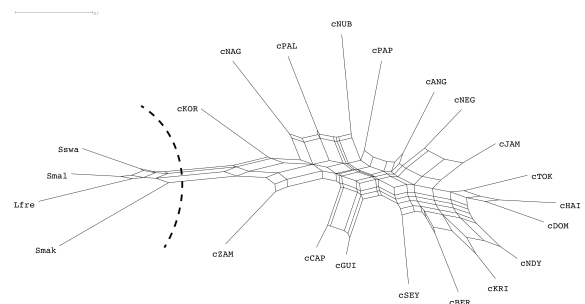

Figure 4: A network of 18 creoles, 3 substrates and one lexifier

The network in Fig. 4 shows that even though only languages involved in the creation of Seychellois were included, the creoles cluster together, which goes against the predictions of a feature pool view, according to which Seychellois would be expected to appear close to the languages that were involved in its formation. The topology of the creole cluster in this and previous networks remains strikingly similar, without affecting the position of Seychellois in the lower right side of the graph. This strongly suggests that the role of the languages involved in the formation of creoles is overstated in a feature pool approach.

3.4 The universalist view

Another hypothesis on creole formation posits that structural similarities between creoles are due to an innate biological propensity, to cognitive constraints or to universals of language restructuring. Thus, language creators drew on the same cognitive resources and universal linguistic processes when trying to solve the communicative problems they encountered, and which in turn resulted in the observed similarities between creolized vernaculars. Bickerton (1981, 1984) further claims that a language-acquisition device in the human brain is the source of creole similarities and regulates the outcome of imperfect language acquisition with the default (or unmarked) settings due to the limited input.

In order to assess the validity of the last hypothesis, that of a universalist account of creole formation, a sample of 52 languages including both substrates and lexifiers was used. Fig. 5 presents a network illustrating the relationships between these languages.
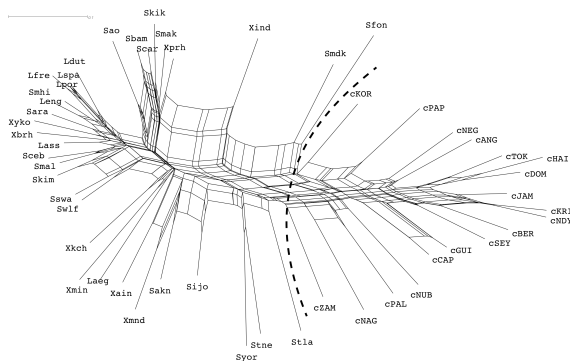

Figure 5: A network of 18 creoles, 19 substrates, 7 lexifiers and 8 non-creoles

In the graph in Fig. 5, the creole cluster is again clearly identifiable to the right of the dotted line. On closer inspection, the network reveals on the upper left a cluster including all the Indo-European languages, Assamese (Lass), Dutch (Ldut), English (Leng), French (Lfre), Marathi (Smhi), Portuguese (Lpor) and Spanish (Lspa). However, this cluster is disrupted by the presence of Arawak (Sara), Kolyma Yukaghir (Xyko) and Brahui (Xbrh). Another genealogical cluster comprising the Afro-Asiatic languages Egyptian Arabic (Laeg) and Mina (Xmin) is found in the lower left side of the graph.

The four Bantu languages that were included, Kikongo (Skik), Kimbundu (Skim), Makhuwa (Smak) and Swahili (Swa), appear in different clusters. Only Kimbundu and Swahili do show up in a cluster on the left in spite of their belonging to widely different branches of the Bantu family (respectively Bantu P and Bantu G in the Guthrie 1948 classification), whereas Kikongo and Kimbundu, which are both Bantu H, appear in opposite ends of the graph. Similarly, the other Niger-Congo languages of West Africa, Akan (Sakn), Bambara (Sbam), Fon (Sfon), Ijo (Sijo), Mandinka (Smdk), Temne (Stne), Wolof (Swlf) and Yoruba (Syor), all appear in different clusters, with the exception of Temne and Yoruba. As for the three Austronesian languages of the sample, Cebuano (Sceb) and Malagasy (Smal) appear in a cluster on the left, far from Tolai (Stla).

Thus, the software was able to detect a clear phylogenetic signal in only a few cases, which in itself is not surprising, since the features were originally selected as representative of the Atlantic creoles (Holm and Patrick 2007: vi). Hence, the results lend support to the universalist position. Besides, all the graphs presented so far also support the idea that creoles form a relatively homogeneous group of languages, in that the creoles are clearly visible and easily distinguishable from the other languages.

4. Creole typology

In this section, we will deal with the typological status of creoles. The issue has long been controversial in creolistics, and the debate has severely suffered from a paucity of systematic cross-linguistic empirical studies directly addressing the question. However, recent studies have shed new light on the matter with the help of phylogenetic tools (Cysouw 2009, Bakker et al. 2011, Daval-Markussen 2011). In the following section, we will show in a similar spirit that creoles pattern similarly in phylogenetic analyses, thus providing further support to the claim that creoles form a synchronically distinguishable sub-group among the world's languages.

4.1 Sampling and method

Traditionally, comparative work in creolistics has focused on subdomains of syntax and/or lexicon for individual languages and/or a restricted number of languages to compare with. In spite of an increasing awareness of the lack of comparative studies encompassing creoles other than the ones lexified by Indo-European languages (partly remedied for with the publication of Holm and Patrick's long-awaited *Comparative Creole Syntax* in 2007), a majority of investigations on creoles still focus on varieties derived from Indo-European superstrates, partly because of the rarity of creole with a non-European lexical base.

The question whether creoles are structurally distinguishable as a group against other languages of the world can be visualized with Neighbor-Joining trees (Saitou and Nei 1987), which have the advantage of quickly returning clear-cut groupings. Since in this part of the study, we are less concerned with pinpointing the reticulation events that shaped creoles than in establishing typological relationships, we have opted for using NJ trees in the remainder of this article.

The study conducted by Parkvall (2008) was the first to make use of quantitative cross-linguistic data for creoles, and the results presented strongly suggested that creoles are structurally distinguishable from non-creoles from a simplicity point of view. In his investigation of the complexity of creoles, Parkvall selected 43 features, 37 of which were taken directly from the WALS (Dryer and Haspelmath 2011), with a further six added by that author. These features were used 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). He then added data on 30 creoles and pidgins with diverse lexifiers (2008: 278), some of which are also found in the CCS sample: Dominican, Guinea Bissau, Haitian, Jamaican, Nubi, Negerhollands, Palenquero and Tok Pisin, i.e. almost half of the Holm and Patrick sample. Parkvall interpreted his results as evidence that creoles are structurally less complex than other natural languages. Consequently, another conclusion reached by the author was that creoles form a typological group characterized by a relatively low level of structural complexity.

The dataset used by Parkvall (2008) also served to provide additional evidence that creoles do indeed form a typological group (Bakker et al. 2011) on the basis of a quantitative empirical analysis. The strongest piece of evidence Bakker et al.'s (2011) large-scale investigation presents relies on Parkvall's (2008) data: 34 creoles and pidgins distinctly cluster in a network including 155 non-creole languages of the world. However, the validity of the results is somewhat undermined by the fact that the data which allowed the authors to reach this conclusion were specifically selected on the basis of creole properties, in that Parkvall selected the features in WALS that could be quantified in terms of complexity (e.g. presence versus absence of a grammatical distinction, or less versus more of a particular phenomenon). Thus, one could object that the results do not really reflect what the authors claim them to (see also Kouwenberg 2010 for a critical assessment).

In order to provide additional, and this time irrefutable evidence, these results must be replicated with different samples and different features. This is what will be attempted in this section, first with the CCS languages and features used in the previous section, then with

various samples of WALS languages and features.

4.2 Using the CCS features and languages

The morphosyntactic features described in Holm and Patrick (2007) are divided up into 20 overarching categories covering various areas (such as TMA systems, NPs or relativization strategies), which were reduced to 18 binary features by selecting for each category the feature(s) that were shared by most creoles. The major linguistic families are indicated with colors in order to facilitate the interpretation of
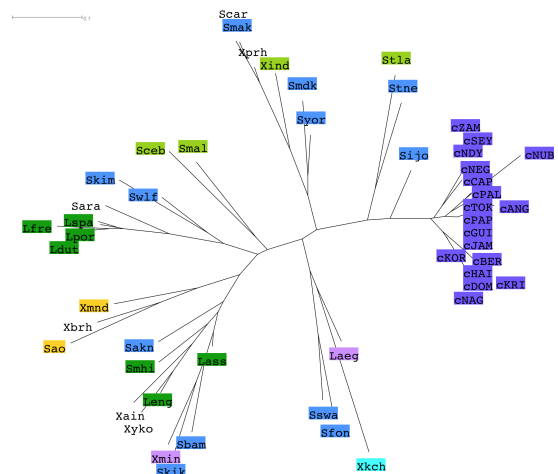

Figure 6: NJ tree of 50 languages with 18 binary features

the graphs (cobalt blue = creoles; cyan = Nilo-Saharan; dodger blue = Afro-Asiatic; green = Indo-European; light blue = Niger-Congo; olive = Austronesian; orange = Austro-Asiatic; pink = Altaic; purple = Uto-Aztecan; red = Australian; tangerine = Trans New Guinea; yellow = Sino-Tibetan).

A cluster consisting of all the creoles is immediately visible in the right end of the tree. The analysis producing these results is based on binary characters only attesting the presence vs. absence of a feature. Therefore, we will introduce in the next section new samples selected from a different database that allows the inclusion of finer-grained distinctions.

4.3 Using WALS to settle the matter

The database provided by the *World Atlas of Linguistic Structures* (Dryer and Haspelmath 2011) consists of descriptions of 144 typological features in a wide variety of languages of the world (2678 as of March 2012 in the constantly updated online version). In the following, different datasets will be extracted from the

WALS in order to further explore the relationships of creoles in the context of the world's linguistic diversity.

The features that are shared by at least 60% of the CCS languages were retained and used to produce several trees based following the multiple-state encoding of the WALS. In the following figures, the same color codes were applied to identify typological clusters.
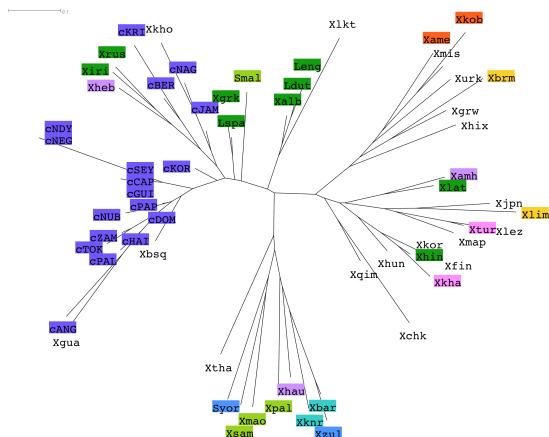

Figure 7: NJ tree of 61 languages with 9 multi-state features

The creole cluster is immediately evident on the left side of the graph in Fig. 7, but a closer look reveals several anomalies: two unrelated languages, Basque (Xbsq) and Guarani (Xgua), appear in the periphery of the core creole cluster, while further up the tree on the initial branch of the creole cluster, four non-creoles show up: the Indo-European Irish and Russian (Xiri and Xrus), the Afro-Asiatic language Hebrew (Xheb) and the Khoisan language Khoekhoe (Xkho).

In order to test the robustness of these results, a larger, more representative sample of the world's languages is required. A logical result of the operation of reducing the number of features increases the number of languages, therefore we gathered another sample of 76 languages based on 6 multi-character features, which returned the tree presented in Fig. 8.

The graph reveals a much denser creole cluster compared to the previous graphs, thus emphasizing the relative homogeneity of creoles as a group. However, several non-creoles appear within the creole cluster: Basque (Xbsq) and Khoekhoe (Xkho) in the core cluster, and four Indo-European languages, Dutch (Ldut), English (Leng), Greek (Xgrk) and Spanish (Lspa), the Uto-Aztecan language O'odham (Xood) and Hungarian (Xhun, Finno-Ugric) in its periphery.
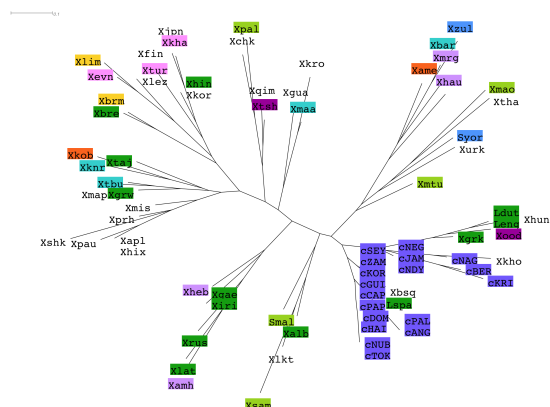

Figure 8: NJ tree of 76 languages with 6 multi-state features

In order to further increase the number of languages included, we have kept the 4 features which were shared by at least 80% of the CCS creoles. The resulting tree is presented in Fig. 9.
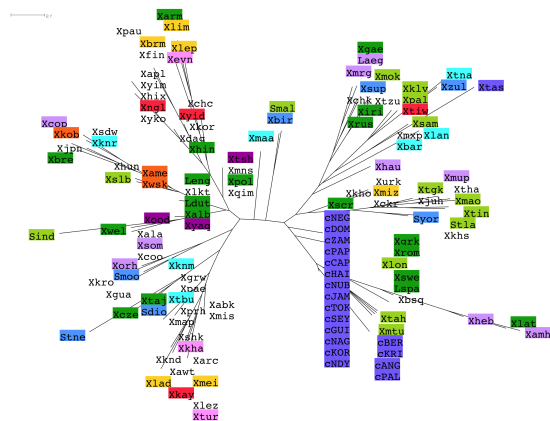

Figure 9: NJ tree of 134 languages with 4 multi-state features

In this final graph, the creole cluster is once again unequivocally identifiable. This time, a majority of the creoles (16 out of 18) are present on a single branch, while two creoles, Krio (cKRI) and Berbice Dutch (cBER) appear on an adjacent branch together with three Austronesian languages, Loniu (Xlon), Motu (Xmtu) and Tahitian (Xtah). The graph in Fig. 9 thus provides conclusive evidence as to the status of creoles: they do form a coherent group of languages that can be distinguished solely on synchronic grounds, as is clearly visible in this, and in previous graphs.

5. Conclusion

We have shown that the application of phylogenetic tools can help shed new light on the typological relationships between languages in general on the one hand, and, on the other hand,

more specifically on the relationships between creoles and other languages, both creoles and non-creoles. The problem of the classification of creoles was shown to be a manageable task with the help of phylogenetic networks. The various theories seeking to account for the similarities between creoles were investigated with the help of phylogenetic networks, and it was found that the chosen analysis was advantageous in that it allowed to graphically represent the relationships between the various languages involved in the emergence of creoles. Finally, the controversial question of whether creoles form a distinguishable subgroup among the world's languages was similarly satisfactorily answered using phylogenetic trees. Moreover, we introduced different ways of depicting the results, where color codes were used so as to instantly identify linguistic patterns.

The availability of freely accessible online databases is constantly increasing, and the prospects for future research are many in the perspective of the interplay between computational methods and linguistics, and, more specifically, in the context of creole languages, as their emegence raise questions on the very nature of language evolution as well.

References

Enoch O. Aboh and Umberto Ansaldo. 2006. The role of typology in language creation. In Umberto Ansaldo, Stephen Matthews and Lisa Lim (eds.), *Deconstructing creole*, 39–66. Amsterdam/Philadelphia: Benjamins.

Philip Baker. 1999. "Investigating the origin and diffusion of shared features among the Atlantic English Creoles". In Philip Baker and Adrienne Bruyn (eds.), *St Kitts and the Atlantic Creoles. The texts of Samuel Augustus Mathews in perspective,* 315-364. London: Battlebridge.

Peter Bakker, Aymeric Daval-Markussen, Mikael Parkvall and Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. In Parth Bhatt and Tonjes Veenstra (eds.), *Journal of Pidgin and Creole Languages.* 26(1): 5-42.

Derek Bickerton. 1981. *Roots of Language.* Ann Arbor: Karoma.

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

Robert Chaudenson. 1992. *Des îles, des hommes, des langues.* Paris: L'Harmattan.

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

Michael Cysouw. 2009. APiCS, WALS, and the creole typological profile (if any). Paper presented at the 1st APiCS conference, Leipzig, 5-8 November 2009.

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

Aymeric Daval-Markussen. 2011. Of networks and trees in contact linguistics: new light on the typology of creoles. MA thesis, Aarhus University.

Michel DeGraff. 2003. Against Creole Exceptionalism. *Language.* 79(2): 391-410.

Mark Donohue, Simon Musgrave, Bronwen Whitting, and Søren Wichmann. 2011. Typological feature analysis models linguistic geography. *Language.* 87(2): 369-383.

Matthew Dryer and Martin Haspelmath (eds.). 2011. *The World Atlas of Linguistic Structures Online.* Munich: Max Planck Digital Library. Available online at http://wals.info/. Accessed on 2012-01-15.

Malcolm Guthrie. 1948. *The classification of the Bantu languages.* London: Oxford University Press.

Ian F. Hancock. 1969. A provisional comparison of the English-based Atlantic Creoles. *African Language Review.* 8: 7-72.

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

John Holm. 1989. *Pidgins and Creoles. Vol. 2. Reference survey.* Cambrige: Cambridge University Press.

John Holm and Peter L. Patrick (eds.). 2007. *Comparative Creole Syntax.* London: Battlebridge.

Daniel H. Huson and David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution.* 23: 254-267.

Silvia Kouwenberg. 2010. Creole studies and linguistic typology: Part 2. *Journal of Pidgin and Creole Languages.* 25(2): 359-380.

John H. McWhorter. 1995. Sisters under the skin: a case for genetic relationship between the Atlantic English-based Creoles. *Journal of Pidgin and Creole Languages.* 10(2): 289-333.

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

John H. McWhorter. 2011. *Linguistic simplicity and complexity: Why do languages undress?* Berlin: Mouton de Gruyter.

Salikoko S Mufwene. 1986. The Universalist and Substrate Hypotheses Complement One Another. In Pieter Muysken and Norval Smith (eds.), *Substrata versus universals in creole*

genesis, 129-162. Amsterdam/Philadelphia: John Benjamins.

Salikoko S. Mufwene. 2000. Creolization is a social, not a structural, process. In Ingrid Neumann-Holzschuh and Edgar Schneider (eds.), *Degrees of restructuring in creole languages*, 65-84. Amsterdam/Philadelphia: John Benjamins.

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

Salikoko S. Mufwene. 2008. *Language evolution: Contact, competition, and change.* London/New York: Continuum Press.

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

Naruya Saitou and Masatoshi Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution.* 4(4): 406-425.

Mark Sebba. 1987. *The Syntax of Serial Verbs.* Amsterdam/Philadelphia: John Benjamins.