Avoiding bias in comparative creole studies: Stratification by lexifier and substrate

Susanne Maria Michaelis

Leipzig University & Max Planck Institute for Evolutionary Anthropology (Leipzig)
susanne.michaelis@uni-leipzig.de

How to cite: Michaelis, Susanne Maria. 2020. Avoiding bias in comparative creole studies: Stratification by lexifier and substrate. *Isogloss. Open Journal of Romance Linguistics* 6/ 8. DOI: https://doi.org/10.5565/rev/isogloss.100

Abstract

One major research question in creole studies has been whether the social/diachronic circumstances of the creolizaton processes are unique, and if so, whether this uniqueness of the evolution of creoles also leads to unique structural changes, which are reflected in a unique structural profile. Some creolists have claimed that indeed the answer to both questions is yes, e.g. Bickerton (1981), McWhorter (2001), and more recently Peter Bakker and Ayméric Daval-Markussen. But these authors have generally overlooked that cross-creole generalizations require representative sampling, especially when working quantitatively. Sampling for genealogical and areal control has been a much discussed topic within world-wide typology, but not yet in comparative creolistics. In all available comparative creoles studies, European-based Atlantic creoles are strongly overrepresented, so that typical features of these languages are taken as "pan-creole" features, e.g. serial verbs, double-object constructions, or obligatory use of overt pronominal subjects. But many of these Atlantic creoles have the same genealogical/areal profile, i.e. European (lexifier) +

Macro-Sudan (substrate). I therefore propose a new sampling method that controls for genealogical/areal relatedness of both the substrate and the lexifier, which I call "bi-clan" control (where "clan" is a cover term for linguistic families and convergence areas).

Keywords: creole languages; creole universals; sampling; genealogical and areal bias; grammaticalization.

Table of Contents

1. Introduction[1]

One major research question in creole studies has been whether the social/ diachronic circumstances of the creolization processes are unique, and if so, whether this uniqueness of the evolution of creoles also leads to unique structural changes, which are reflected in a unique structural profile. Some creolists have claimed that indeed the answer to both questions is yes, e.g. Bickerton (1981), McWhorter (2001), Bakker et al. (2011), Daval-Markussen (2014), and most recently Daval-Markussen & Bakker (2017) and Daval-Markussen (2018).

To demonstrate the unique structural profile, these creolists have proposed that creoles share a set of "pan-creole" features. A unique structural profile implies that creoles are internally uniform through their pan-creole features and that they are externally distinctive with respect to non-creoles world-wide. But to show that creoles uniformly share pan-creole features, one does not only need to find a set of such features, but one also needs to examine a *representative* sample of creoles, i.e. a sample with historically and areally maximally independent languages. Sampling has been a very much discussed topic within typology (e.g. Dryer 1989, 1992; Rijkhoff & Bakker 1998; Perkins 2001; Bickel 2008; D. Bakker 2011), where it has been widely recognized that a biased picture can result if language samples contain languages that are not independent from each other, either because they are descended from a common ancestor or are neighbouring languages that may have influenced each other. Thus, one needs to control for

genealogical and areal bias when looking for universal features in languages world-wide. However, in studies that look for creole universals, we still lack such a discussion of what a representative sample of creoles should look like.

The present paper is an attempt to fill that gap by suggesting a new method for sampling in contact linguistics, which involves introducing the notion of BI-CLAN (a set of languages that share the same lexifier clan and substrate clan, where clan is a cover term for families and convergence areas).

The paper is organized as follows: In §2, I briefly look at existing samples in comparative creole studies and show that they are biased towards one areal group, Atlantic creoles. In §3, I give an overview of the *Atlas of Pidgin and Creole Language Structures* (*APiCS*, Michaelis et al. 2013a), before I introduce the notion of bi-clan as a way of stratifying samples of creoles in §4. In §5, I look at the implications of bi-clan sampling for pan-creole features and creole universals. In §§6-7, various structural features in (pidgin)creoles will be checked against the bi-clan distribution, and §8 concludes the paper.

The main claim of this article is the following: For far too long creolists have concentrated on the analysis of a specific areal type of (pidgin)creole languages, namely Atlantic (pidgin)creoles, and have extrapolated from this narrow profile to creole languages in general (e.g. Bickerton 1981). But with the publication of *APiCS*, the most comprehensive source of systematic comparable data of a large number of creole languages, we are now in a position to assess the impact of the oversampling of the Atlantic creoles and to introduce a new sampling method via bi-clans.

2. Sample bias in comparative creole studies

Compared to typological studies, in comparative creole studies we are dealing with rather small sample sizes between 10 and maximally 80 languages (e.g. Taylor 1971, Goodman 1964, Hancock 1987, Holm & Patrick 2007, *World Atlas of Varieties of English*, (*WAVE*; Kortmann & Lunkenheimer 2011, *APiCS*, Daval-Markussen 2018). One major reason for this is that the great majority of languages called "creoles" have been restricted to certain contact varieties, one major type being contact languages that have emerged and evolved between the 15th and 19th centuries in European colonial settings. Out of these creole languages, only a subset has been described for a fair amount of grammatical phenomena. But there are more and more studies on non-European pidgins and creoles (see e.g. Buchstaller, Holmberg & Almoaily 2013). At present, there must exist many more creole languages in other parts of the world, for instance in China. But either these varieties have not been discovered and described yet, and/or scholars do not see themselves as being part of the creolist community and therefore do not want to classify their contact varieties as pidgins, pidgincreoles or creoles. But future research in contact languages will certainly enlarge existing databases of pidgins and creoles.

In the present article, I use the term "creole" to embrace classical creoles like Saramaccan, Mauritian Creole, and Tayo, but also pidgincreoles, i.e. expanded pidgins that are used in a wider set of linguistic functions even though

they are not the mother tongues of all their speakers, (Bakker 2008: 131ff) such as Cameroon Pidgin English or Tok Pisin.

One characteristic of all existing samples of creoles is that they are strongly biased. In all available studies, Atlantic creoles (i.e. languages of West Africa and the Caribbean) – whether English-, Dutch-, French-, or Ibero-Romance-based – are heavily overrepresented. Thus, Nigerian Pidgin, Jamaican, Saramaccan, Sranan, Haitian Creole, Principense, Guinea-Bissau Kriyol and Cape Verdean Creole and the like have been at the center of interest, whereas Papiá Kristang, Kriol, and Tayo have been much less discussed in the literature. For accidental historical reasons, the latter languages do not have as many sister creoles to be studied as, for instance, the Caribbean creoles have. This fact may have given rise to the view in creole studies that somehow the Atlantic creoles can be taken as the prototypical creoles, only because they are more numerous and better studied than less well-known creole languages in other parts of the world (see, e.g., Szmrecsanyi & Kortmann 2009: 34[2] , Daval-Markussen 2018).

Many of these overrepresented Atlantic creole languages happen to have the same contributor profile: European lexifiers and Sub-Saharan West African substrate(s). Table 1 shows some of the most recent comparative creole studies, the number of pidgins and creoles analyzed, and the percentage of Atlantic pidgins and creoles out of these languages.

Table 1. Overrepresentation of Atlantic pidgins and creoles in comparative studies*

comparative creole studies	pidgin/creoles/ mixed lgs	Atlantic creoles	
Holm & Patrick 2007	18	12	**67%**
Parkvall 2008	31	18	**58%**
eWAVE 2011	26	17	**65%**
APiCS 2013	76	33	**43%**
Daval-Markussen 2018	64[3]	36	**56%**

*Bakker et al. (2011) used the samples by Holm & Patrick 2007 and Parkvall 2008, and thereby they inherited the overrepresentation of Atlantic creoles of these data sets.

As can be clearly seen from Table 1, the Atlantic creoles are overrepresented in all cited works. Even the most comprehensive comparative work, the *APiCS* (Michaelis et al. 2013) with 76 contact languages world-wide, shows more than 40% of all languages featuring this narrow profile.

---

[2]   "This first-ever comprehensive survey of nonstandard English morphosyntax worked in a very simple way: We compiled a catalogue of 76 features – *essentially, the 'usual suspects' in previous dialectological, variationist, and creolist research* – and sent out this catalogue to the authors of the chapters in the morphosyntax volume of the Handbook" [emphasis is mine], the Handbook referring to Kortmann et al. 2004, which is the basis for the *Electronic World Atlas of Varieties of English*, Kortmann & Lunkenheimer (2013). In this atlas, 17 out of 26 creoles are Atlantic-based (see Table 1).

[3]   Daval-Markussen (2018) contains various creoles with non-European lexifiers/non-Macro-Sudan substrates which are not included in *APiCS*, such as Yilan Creole (Japanese-based), Nheengatu Creole (Tupi-based), Hiri Motu Creole (Motu-based).

Why should this be a problem? If we want to study creoles in order to find universal mechanisms of creolization or unique feature profiles of these languages, we had better try to sample the languages maximally independently from each other to control for genealogical and areal relatedness. But this is not what has happened so far: All existing theories of creolization and potentially unique typological profiles of creoles have been based on genealogically, and areally, biased data, namely mainly on Atlantic creoles. Not only do these creoles have closely related lexifiers (French, Portuguese, English and Dutch), but also their substrates share many linguistic features: they belong to a convergence area (cf. the notion of Macro-Sudan belt, see below). Thus, all potential generalizations over creoles which have been proposed so far may manifest only genealogical or areal features of a specific group of creoles, but not the claimed universal features of creoles in general. For a universal claim about creoles to hold, one would have to control for genealogical and areal relatedness. The task then would be to sample creoles in such a way that the genealogical and areal predominance of Atlantic creoles gets reduced in a principled way to account for the world-wide diversity of creoles. Interestingly, creolists have not seen any problem with the bias of Atlantic creoles in their respective sample (e.g. Kortmann et al. 2004). On the contrary, Bakker et al. (2011: 5), for instance, say:

> 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*" (emphasis is mine).

Note that Bakker et al.'s data come from Holm & Patrick (2007) and Parkvall (2008), both strongly Atlantic-biased samples of creole languages. But one should also stress that some scholars have been very clear about their specific narrow perspective on Atlantic creoles. Holm & Patrick (2007: vi) for instance themselves clarify that:

> [t]he syntactic features chosen for examination are generally those which distinguish the *Atlantic Creoles (those of the Caribbean and West Africa) from their source languages*." (emphasis is mine)

As the seminal work by Holm & Patrick (2007) is the first systematic comparative study based on 18 creole languages with different lexifiers, some creolists have used the comparative data, but they have overlooked the important restriction which the authors themselves stated very clearly in their introduction. Thus, Bakker et al. (2011: 16) claim that the 97 structural features chosen by Holm & Patrick 2007 are "all somehow assumed to be *typical of creole languages*" (emphasis is mine). They pursue their Atlantic-biased method even further to test for creoleness in other contact languages only by taking typically Atlantic features into account:

> The basic idea now is that the number of features (in Holm & Patrick 2007) could be taken as a proxy for the degree of creoleness. (Bakker et al. 2011: 21)

Unsurprisingly, most non-Atlantic creoles in their sample (which is the sample in Holm & Patrick 2007) show a lower score of shared linguistic features: Palenquero, Chabacano of Zamboanga, Nubi, and Korlai. This is exactly what one would expect if viewing these non-Atlantic creoles with a feature set that has been deliberately conceived of as one that distinguishes Atlantic creoles from their lexifiers. This observation is important, not only with respect to comparing creole languages, but to comparing languages in general, because there is no such thing as a neutral set of features or feature values to which one can compare different languages, but every feature questionnaire and its framing has consequences for the obtained results.

    The view of creolists that the most widely known and studied Atlantic creoles somehow constitute the default case of creole languages may have been influenced by the fact that non-typologists seem both to overestimate the linguistic distance between the European lexifier languages, and to underestimate consistent areal clusters/patterns in the substrate/adstrate languages, e.g. in sub-Saharan Africa. From the perspective of grammar, the linguistic distance between the European lexifiers such as French, Portuguese, and Spanish, all members of the Romance family, is quite small. But even the Germanic languages, like English and Dutch, are very closely related to the Romance languages compared, for instance, with other branches of the same Indo-European language family, such as Indo-Aryan or Iranian. But most importantly, the West African substrate languages which are at the basis of most of the studied Atlantic creoles also show areal patterns of convergence, which led Güldemann 2010 to speak of the so-called "Macro-Sudan belt" (see Güldemann 2010, zone III in Figure 1). Güldemann shows a converging feature profile which cuts across various language families. These converging features allow him to propose a core zone (brown colour) and peripheral zones of this large Sub-Saharan macro area (orange and yellow colours).

Map 1. Macro areas in Africa
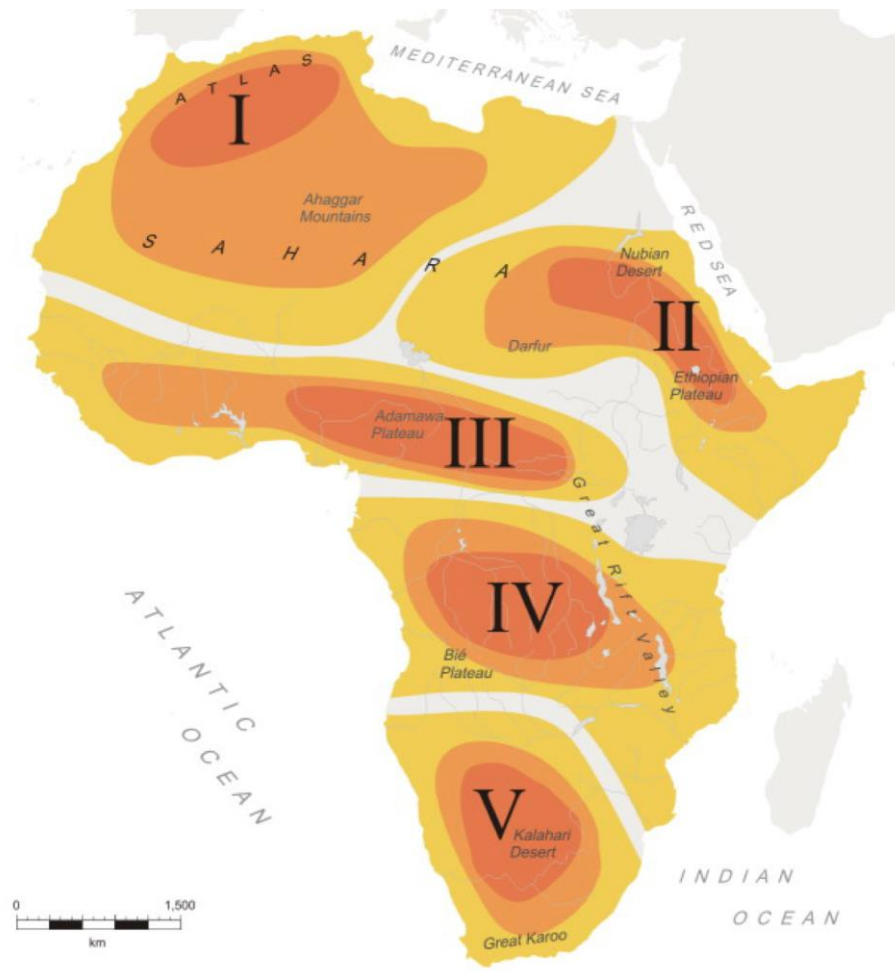
Source: Güldemann (2010); III = Macro-Sudan belt

Güldemann takes phonological and syntactic features into account, as for instance, labial-velar consonants, ATR vowel harmony, logophoricity and S-(AUX)-O-V-X. In this view the Macro-Sudan belt embraces the following families. In the core area we find: Mande, Kru, Gur, Kwa, Benue-Congo (except Narrow Bantu), Adamawa-Ubangi, Bongo-Bagirmi, and Moru-Mangbetu. In the periphery of this macro area are the following families: Atlantic[4], Dogon, Songhai, Chadic, Ijoid, Narrow Bantu, and Nilotic. Many of the different substrate languages of the Atlantic creoles are either core languages of the Macro-Sudan belt, such as Fongbe, Akan, or Yoruba, or peripheral languages, such as Wolof, Mandinka etc. (see, e.g., Alleyne 1980, Boretzky 1983, Holm 1988, and Parkvall 2000 on African substrate languages in Atlantic creoles).

---

[4]   The term "Atlantic" in this context relates to a sub-family of African languages mainly spoken in Senegal, Gambia, Guinea, Guinea-Bissau (cf. also North-Central Atlantic language family in glottolog.org) and should not be confused with the term "Atlantic" in "Atlantic creoles", which as a cover-term refers to creole languages in West Africa and the Caribbean.

The idea is now that the languages of these different families share certain linguistic features, presumably due to long-standing contact. This view on West and Central African languages implies that even though potential substrates of Atlantic creoles belong to different substrate families, they may still show convergent structures. Therefore, speakers of these different African languages may have initiated similar linguistic changes in the different language contact situations with European speaking colonists during creolization processes in Africa and the Caribbean leading to similar outcomes in the resulting creoles.

Once one is aware of the Atlantic bias in all available samples of contact languages, can one claim something about the typological profile of creole languages in general on the basis of this sample? My answer is clearly no. If one wants to generalize over the class of creoles as such, the first step is to better balance one's sample, and this is in my view possible even without collecting new data.

3. The *APiCS* database

The present paper is based on the large-scale comparative database of pidgins and creoles, the *APiCS* (Michaelis et al. 2013a). In *APiCS*, 76 contact languages world-wide are investigated with respect to 130 structural features, some 330 segmental features, and 28 sociolinguistic features. As Map 2 illustrates, *APiCS* covers all major world regions. It especially embraces information on non-Atlantic contact languages in South Asia, Southeast Asia, Melanesia, and Australia. It also contains contact languages which have non-European base languages, like Arabic, Bantu, and Malay.

Map 2. 76 contact languages
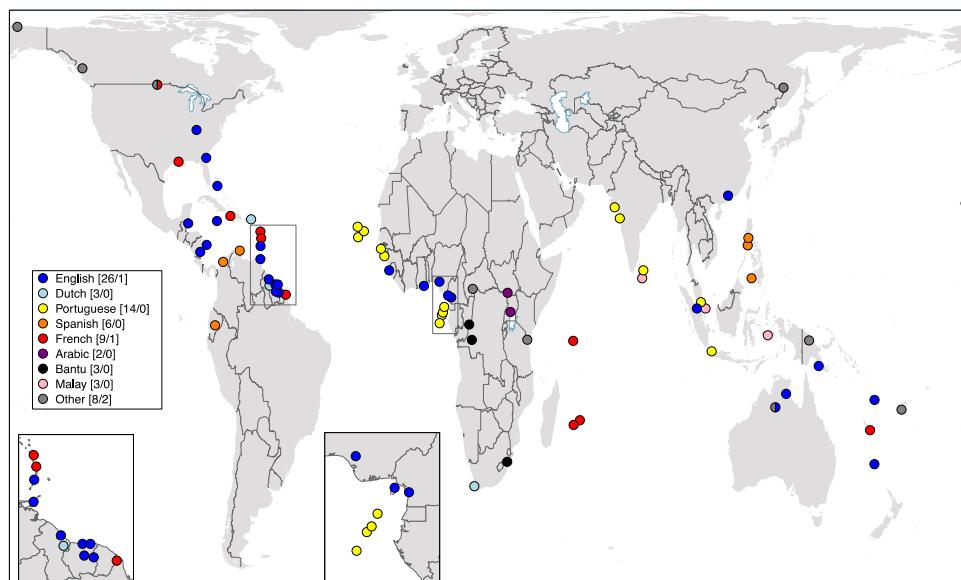
Source: APiCS[5]

---

[5]   For space reasons, the labels of the languages cannot be given in this map and the subsequent maps. For a list of the languages, see *APiCS Online* http://apics-

Each language is the responsibility of a language expert or team of experts. *APiCS* has been published as a four volume print work: the first volume displays chapters written by the four editors (Michaelis, Maurer, Haspelmath, and Huber) on each of the 130 structural features with the corresponding map where each coloured dot represents the given feature value for a given language, and the three other volumes constitute the *Survey of pidgin and creole languages* (Michaelis et al. 2013b) where each expert gives a concise overview of the social history, sociolinguistic situation and of the grammar of the contact language. Besides the paper publication, the underlying database *APiCS Online* (apics-online.info) has been published electronically as part of the *CLLD* databases (clld.org) with more than 15,000 fully glossed and translated examples as well as many references[6]. 48 structural features of *APiCS* have been taken over from the *World Atlas of Language Structures* (Haspelmath et al. 2005). This gives us the unique possibility to see the *APiCS* languages against the background of languages world-wide. *APiCS Online* provides special views for this comparison.

4. Sample stratification

4.1. The problem of bias in language typology

Typologists have for some time been aware that sample bias may be a serious problem (e.g. Bell 1978; Dryer 1989; Rijkhoff & Bakker 1998; Perkins 2001). If one considers data from a range of languages that are not historically independent of each other, then one may get a skewed picture, even if one looks at a large number of languages (also known as *Galton's problem* in cross-cultural studies in general). For example, if one's sample has many languages from Eurasia, one may wrongly conclude that the order of adjective and noun correlates with the order of possessor and noun, whereas this is in fact not the case (Dryer 1989). According to Rijkhoff & Bakker (1998: 264-265), there are basically two kinds of samples, variety samples (which display the greatest possible variety) and probability samples (which are designed to be quantitatively representative of the entire population). Variety samples are most suitable for exploratory research, when little is known about the phenomenon. By contrast, when one is interested in any kind of quantitative evaluation, one needs a probability sample (cf. also Bickel 2008: 222). This also applies to comparative creole studies that make quantitative statements with universal scope.

   In language typology, genealogical bias and areal bias are the best-known kinds of sample bias, i.e. too many languages from a well-described family (e.g. Indo-European) are chosen, or too many languages from a well-described area (e.g. Europe). Such biases can be avoided by stratification, i.e. by creating mutually exclusive subgroups of languages (families or areas) which have equal status and are the basis for the selection of languages. Since the great majority of

---

online.info/contributions#2/30.3/10.0. All *APiCS* maps in this paper were designed by Hans-Jörg Bibiko, Max Planck Institute for Evolutionary Anthropology (Leipzig).

[6] Since 2017, the Atlas and the survey chapters have been freely available as part of *APiCS Online* as well.

universals are statistical trends rather than exceptionless generalizations, a stratified world-wide sample is a necessary ingredient of any large-scale study that makes universal claims. There are of course many practical problems (such as determining the right families, and determining areas within which contact-induced convergence has taken place, cf. Song 2001: §1.5.3), but there is no doubt that stratified sampling is the least that one needs to support universal claims.[7]

4.2. Stratification of creole samples through bi-clans
If one is interested in universal features of creole languages, one needs a stratified sampling method, too, but there are two possible sources of bias: from the substrate and the lexifier. Therefore, I would like to propose a sampling method that controls for genealogical and areal relatedness of both the substrate(s) and the lexifier, what I call BI-CLAN SAMPLING. A CLAN[8] is a language or a family or a linguistic area, and a BI-CLAN is a combination of a lexifier clan and a substrate/adstrate clan[9]. For example, the lexifier clan "English" combined with the substrate clan "Macro-Sudan" gives rise to the bi-clan "English/Macro-Sudan". Nigerian Pidgin, Jamaican and Saramaccan are for instance members of this bi-clan. The lexifier clan "Portuguese" combined with the substrate/adstrate clan "Indic" constitutes the bi-clan "Portuguese/Indic". Languages that belong to this bi-clan are Korlai, Diu Indo-Portuguese and Sri Lanka Portuguese. While we often know very well which lexifier is at the base of a given creole, the identification of the relevant substrates is a much more difficult matter. Therefore, we have the option of lumping different entities into a clan: A clan can either be a single language (e.g. English or French), a family (e.g. Indic, Malay) or a linguistic area (e.g. Macro-Sudan). The important issue here is that we try to keep potentially historically related creoles in the same bi-clan, whereas historically unrelated creoles should be in different bi-clans.

I interpreted Güldemann's Macro-Sudan belt (see above §2) narrowly and only took the core families of the Macro-Sudan belt to be part of the clan "Macro-Sudan", most importantly Mande, Kru, Gur, Kwa, Benue-Congo (except Narrow Bantu), whereas the families in the periphery (Atlantic, Ijoid, Narrow Bantu, and Nilotic) each make up their own clan, giving us bi-clans such as Dutch/Ijoid (with its member Berbice Dutch), English/Bantu (with its members Pichi and Cameroon Pidgin English), or Portuguese/Atlantic (with its members Cape Verdean creole varieties, Casamancese Creole and Guinea-Bissau Kriyol). Note that the term "Atlantic" here refers to a specific language family of West Africa (with e.g. Wolof and Balanta, see also footnote 4).

---

[7]   Dryer (1989) argues that not even this will always work, because in the case of word order, there are continent-sized linguistic areas, and if just one language per area is chosen, the sample is too small. I leave this problem aside and assume that it does not arise frequently.

[8]   The term "clan" was suggested to me by Bernard Comrie.

[9]   A similar approach was adopted by Dryer (1989, 1992) for world-wide samples in the study of language universals. Dryer suggests the unit *genus* which is a level between the individual language and the larger family. A typical example of genera are the subfamilies of Indo-European, e.g. Germanic, Slavic, Celtic, Romance.

The 76 *APiCS* languages fall into 34 bi-clans, out of which 20 are represented by only one language. Many pidgins and all mixed languages in the sample happen to constitute a bi-clan of their own, as their areal/genealogical profile is unique. For example, Chinese Pidgin Russian is the only member of the bi-clan Russian/Sinitic, the mixed language Gurindji Kriol belongs to the bi-clan Gurindji (a language of Northern Australia)+Kriol (a creole language which arose from the contact between English and languages of Northern Australia), and Media Lengua is in the bi-clan Spanish+Quechua[10].

In the present paper, I will concentrate on the 59 creoles in *APiCS*, whose bi-clan distribution is shown in Table 2.

Table 2. 20 creole bi-clans

bi-clan	number of lgs	creoles belonging to the bi-clan
English/Macro-Sudan	16	Early Sranan, Sranan, Saramaccan, Nengee, Creolese, Trinidad English Creole, Vincentian Creole, Jamaican, Belizean Creole, San Andres Creole English, Nicaraguan Creole English, Bahamian Creole, Gullah, Krio, Ghanaian Pidgin English, Nigerian Pidgin
French/Macro-Sudan	5	Haitian Creole, Martinican Creole, Guadeloupean Creole, Guyanais
Ibero-Romance/Macro-Sudan	5	Papiamentu, Angolar, Santome, Principense, Fa d'Ambô
Dutch/Macro-Sudan	1	Negerhollands
Portuguese/Atlantic[11]	5	Cape Verdean Creole of Brava, Santiago and São Vicente, Guinea-Bissau Kriyol, Casamancese Creole
Dutch/Ijoid	1	Berbice Dutch
English/Bantu	2	Pichi, Cameroon Pidgin English
Spanish/Bantu	1	Palenquero
Spanish/Philippinic	3	Chabacano of Ternate, Cavite and Zamboanga
French/Bantu	3	Reunion Creole, Mauritian Creole, Seychelles Creole
Portuguese/Indic	3	Diu-Indo Portuguese, Korlai, Sri Lanka Portuguese
Portuguese/Malay	2	Papiá Kristang, Batavia Creole
English/Oceanic	2	Tok Pisin, Bislama
French/Oceanic	1	Tayo
Arabic/Southern	2	Juba Arabic, Kinubi

---

[10]   The parts of bi-clans of pidgins, (pidgin)creoles and restructured varieties are separated by a slash, whereas the two parts of mixed languages are connected by a „+".

[11]   For the term "Atlantic", see footnote 4.

Sudanese		
Bantu/Bantu	2	Lingala, Kituba-Kikongo
English/Australian	1	Kriol
Malay/Central Malayo-Polynesian	1	Ambon Malay
English/various	2	Norf'k, Hawai'i Creole
Ngbandi/Central African	1	Sango

Source: *APiCS*

Table 2 clearly shows (i) that English/Macro-Sudan is the bi-clan with by far the greatest membership (16), and (ii) that the creoles in a bi-clan that features a West European lexifier language (English, Dutch, French, Portuguese, Spanish) and Macro-Sudan as their substrate area number as many as 27 languages, which is 46% of all creoles in *APiCS*. After English/Macro-Sudan, the following bi-clans are the next best represented with 5 languages each: French/Macro-Sudan, Ibero-Romance/Macro-Sudan, and Portuguese/Atlantic. Other bi-clans have two or three members, e.g. Arabic/Southern Sudanese with Juba Arabic and Kinubi. Some creoles make up their own bi-clan, for instance Tayo (French/Oceanic), Palenquero (Spanish/Bantu), and Berbice Dutch (Dutch/Ijoid).

The granularity and the classification of the proposed bi-clans is open to discussion[12]. But the present approach should be taken as a first attempt to do justice to the different genealogical/areal linguistic profiles of creoles and at the same time to reduce the impact of typologically uniform languages of the same bi-clan, in order to achieve the ultimate goal, namely to assess potential universals in creole languages.

I will now turn to the discussion of various structural features in the context of the bi-clan distributions.

5. Implications of bi-clan sampling for pan-creole features and creole universals

In the next section (§6), I will examine various grammatical features and I will discuss their cross-creole distribution in *APiCS*. One of the leading questions will

---

[12]   I am very much aware of the fact that the splitting of the Western European lexifiers into clans, such as French, Portuguese, English etc., is not on a par with the splitting of the substrates. But lumping these lexifiers together as "Western European" would drastically reduce the number of bi-clans, as the great majority of well-described creoles has a Western European lexifier. The sample of bi-clans would then be too small for quantitative evaluation. Furthermore, the substrate assignments may seem too coarse-grained. It is true that the social histories of most creole languages are much more complex than a single bi-clan assignment may suggest. Different slave groups were deported at different points in time from different geographical sites. Often the linguistic backgrounds were as different as Malagasy and Bantu languages from East Africa, as in the case of Reunion and Mauritian Creole. I opted for a single substrate clan given the pivotal role of this substrate clan in the formation of the creole.

be whether a given feature is wide-spread enough among the different creole languages so that we can call it a pan-creole feature. The bi-clan sampling will help us to address this question.

For a feature to qualify as a pan-creole feature, it should be

- widespread in an unbiased sample of creoles, i.e. in a maximal number of bi-clans, not just in the majority of creoles surveyed.

If this feature additionally is

- more likely to be found in creoles than in non-creoles, and
- not found in the contributing lexifier/substrates of a given creole

then we will have good reasons to classify this feature as a creole universal, i.e. a feature that has arisen through special cognitive and/or social conditions of the creolization process.

In section §6.1, I will look at features that seem wide-spread in creoles and therefore at first glance look like good candidates for pan-creole features, but on closer inspection turn out to have a clear areal distribution. In section §6.2, I will examine features that occur rarely in creoles.

In §7, I will consider pan-creole features and ask whether the additional criteria are fulfilled so that they can be regarded as universal creole features.

6. Areally-restricted features

As the focus of creole studies has long been on the major Atlantic creoles such as Jamaican, Haitian Creole, Santome and Krio, it does not come as a surprise that the grammatical features used for creole comparison have often been those which are typical of Atlantic creoles.

6.1 Features that seem widespread in creoles

The serial verb construction is a prominent type of construction which is widespread in Atlantic creoles, but which also belongs to a set of features that has been claimed by some authors to belong to the core features of creole languages in general (e.g. Taylor 1971, Bickerton 1989, 1996, Byrne 1987, Kortmann & Szmrecsanyi 2004).

Let us first look at directional serial verb constructions with 'come' and 'go', where the motion verb as second verb specifies the direction of the action of the first verb. (For a more detailed discussion of this construction, see Maurer, Michaelis & *APiCS* Consortium 2013). The first verb of such a construction may be either intransitive (1-2) or transitive (3).

(1) Santome (Portuguese-based, Gulf of Guinea; Hagemeijer 2013)
    *Nansê ka   subli **ba**   ôbô   ê!*
    2PL   IPFV go.up go   forest PCL
    'You go up to the forest!'

(2) Papiamentu (Spanish-based, Caribbean; Kouwenberg 2013)
   *E   bebi   a   gatia **bai**   den   kushina.*
   DEF   baby   PFV   creep go   in   kitchen
   'The baby crept into the kitchen.'

(3) Haitian Creole (French-based, Caribbean; Fattier 2013)
   *Li   voye   bòn   nan   **ale**.*
   3SG   send   maid   ART   go
   'She dismissed the maid.'

Map 3 shows the world-wide distribution of this construction in the 59 creole languages in *APiCS*. The red dots show the presence of such a construction, whereas white dots represent languages where such a construction does not exist.

Map 3. Directional serial verb constructions with 'come' and 'go' in 59 creoles of *APiCS*

- ● 'Come' and 'go' directionals exist [34]
- ○ 'Come' and 'go' directionals do not exist [24]
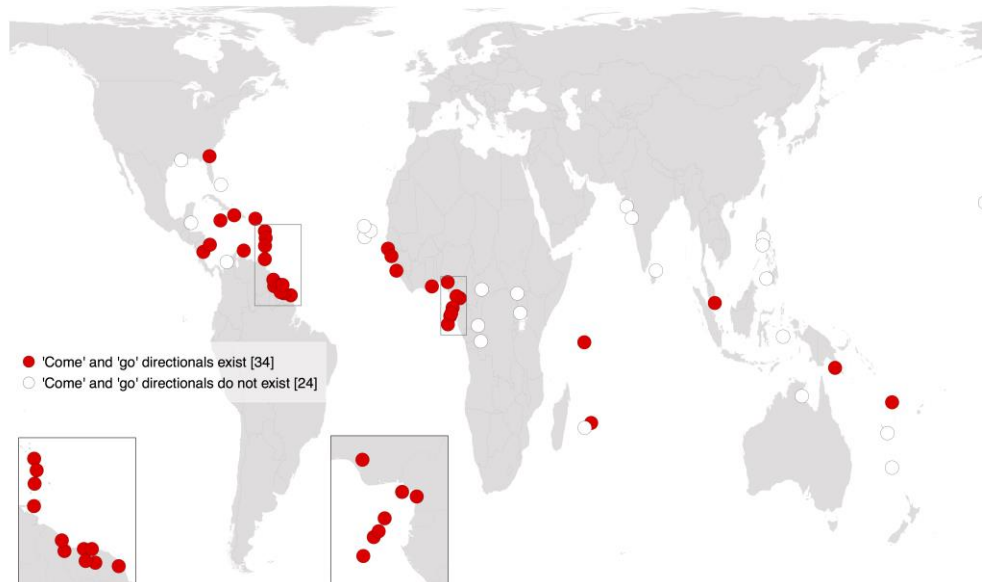
Source: Maurer & Michaelis & *APiCS* Consortium (2013)

At first glance, the expectation that the majority of the creoles in *APiCS* show directional serial verb constructions seems to be fulfilled (see Table 3): 34 (59%) out of 58 creoles with data for this feature show a type of directional serial verb construction, whereas 24 creoles (41%) lack this construction. Already from eye-balling one can see that the large Macro-Sudan bi-clans (English-, French-, Ibero-Romance/Macro-Sudan) all feature this type of directional serial verb construction.

Table 3. Directional serial verb constructions with 'come' and 'go'

	APiCS creoles		APiCS bi-clans of creoles	
'come' and 'go' directionals exist	34	59%	10	**42%**
'come' and 'go' directionals do not exist	24	41%	14	**58%**

Source: *APiCS*

And indeed, when we consider bi-clans instead of languages, the results change: Now only 10 of the 24 bi-clans (42%) have a directional serial verb construction, with the other 14 bi-clans (58%) lacking such a construction.

How do we determine the bi-clan numbers? Languages of the same bi-clan are genealogically (and/or areally) closely related, so they show similar typological profiles. This means that languages of the same bi-clan often show the same feature value for a given feature. In the present feature, for instance, the 16 creoles of the English/Macro-Sudan bi-clan all have directional serial verb constructions (red dots on Map 3). But that is what we expect from genealogically/areally related languages, namely that they share many linguistic features, regardless of whether they are creoles or not. These 16 instances of uniform marking should therefore not be given the same weight as other languages with no close relatives. Thus, in the bi-clan distribution, a bi-clan is counted only once if its members show uniform behavior. The English/Macro-Sudan languages thus contribute 16 points to the language count, but only one point to the bi-clan count. Of the five languages that belong to the French/Macro-Sudan bi-clan, four have directional 'come' and 'go', whereas one language (Louisiana Creole) lacks this construction. Therefore, this bi-clan is counted twice, once for the existence of this construction and once for its absence. In this way, we capture the linguistic diversity within and across bi-clans. The advantage of this method is straightforward: Bi-clans are treated alike independently of their size – bi-clans with few languages have the same impact as bi-clans with many more languages. The crucial criterion is whether the languages of one and the same bi-clan show the same or different feature values.

This methodological step is crucial if we want to find pan-creole and ultimately universal features in creoles. Just counting creoles with and without directional serial verb marking may blur the picture: Thanks to the large group of historically related creoles with Macro-Sudan substrates, the overall majority of the *APiCS* creoles shows this serial construction. But in the bi-clan distribution, the majority relation is flipped around: Now it is the bi-clans that lack a directional serial verb construction that makes up the majority of cases.

This feature is instructive in many ways: The bi-clan distribution of directional serial verb constructions clearly shows that this feature is not a pan-creole feature, but at the same time the distribution is not just random. We see clear areal patterns: Atlantic creoles with a few exceptions (Cape Verdean varieties, Palenquero, Bahamian, Belizean and Louisiana Creole) all show this construction as well as some of the Indian Ocean creoles, plus Bislama and Tok Pisin. However, the non-serial-verb areas are also clearly detectable: creoles in South Asia, the Philippines, Australia and some in Oceania (Tayo and Norf'k).

Such a patterning indicates that the construction is likely to originate in the lexifier or in the substrates, and is not due to the cognitive or social conditions of creolization. And indeed, it has long been noted that this type of serial construction is found in a wide area of sub-Saharan Africa (see, e.g., Boretzky 1983, Parkvall 2000:71ff.). Interestingly, the Cape Verdean creoles lack the directional serial-verb construction, as does Wolof, its main substrate language, whereas the Upper Guinea varieties spoken on the African mainland, Guinea-Bissau Kriyol and Casamancese Creole, show directional serial verb constructions, as does Balanta, one of the main sub-/adstrates of these two Portuguese-based creole languages. The same is true for other substrate/adstrate languages in other parts of the world (see Maurer, Michaelis & *APiCS* Consortium 2013: 342)[13].

Another feature that is relevant here is ditransitive constructions with the verb 'give', as in examples (4)-(6). All examples show a double-object construction, with no preposition marking the recipient (or the theme).

(4) Palenquero (Spanish-based, Caribbean; Schwegler 2013)
     El   tan   nda   ele   un   regalo muy   epesial.
     he/she FUT   give   him/her a   gift   very   special
     'He/she is going to give him/her a very special gift.'

(5) Krio (English-based, West Africa; Finney 2013)
     di   uman   gi   di   titi   sɔm   mɔni
     the   woman   give   the   girl   some   money
     'The woman gave the girl some money.'

(6) Seychelles Creole (French-based, Indian Ocean; Michaelis & Rosalie 2013)
     Mon   'n   donn   Marcel en   mang.
     1SG   PRF   give   Marcel a   mango
     'I gave Marcel a mango.'

It has been claimed that creoles typically show double-object constructions (Bickerton 1995, Bruyn et al. 1999) even if their lexifiers, for instance the Romance languages, have an indirect-object construction (with a preposition marking the recipient, e.g. French *J'ai donné une mangue à Marcel*). But if we look at the world-wide distribution of creoles in *APiCS*, the picture is not uniform at all (Map 4).

---

[13]   McWhorter (1997: 35-39) similarly argues for substrate influence in a wide range of serial verb constructions for Atlantic and non-Atlantic creoles.

Map 4. Ditransitive constructions with 'give' in 59 creoles of *APiCS*
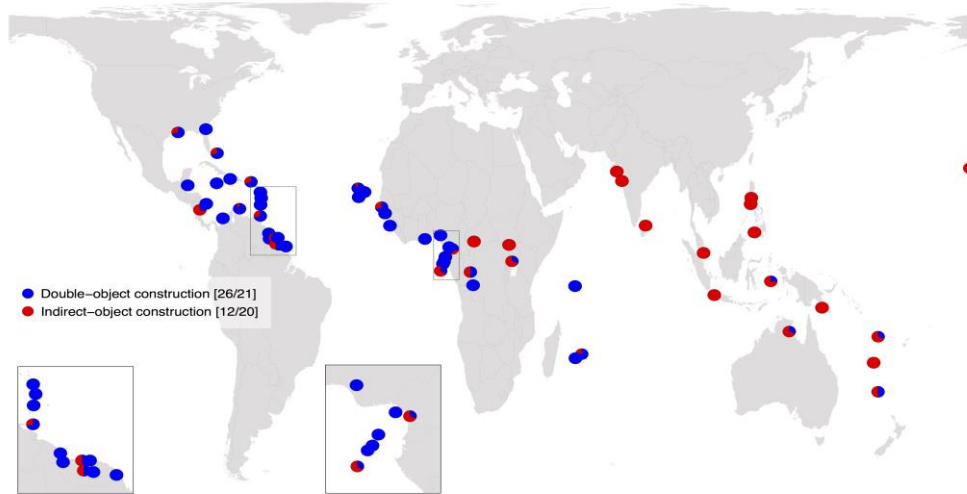
Source: Haspelmath, Michaelis & *APiCS* Consortium (2013a)

First of all, languages can have both construction types, double-object constructions (blue dots) and indirect-object constructions (red dots)[14], which is shown in the pie-charts on Map 4. But if for simplicity we restrict ourselves to creoles with exclusive double-object constructions and exclusive indirect-object constructions, that is representing single-coloured dots, the figures are as follows:

Table 4. Ditransitive constructions in creoles (exclusive marking only)

	APiCS creoles		*APiCS* bi-clans of creoles	
Double-object constructions	26	69%	9	**56%**
Indirect-object constructions	12	31%	7	**44%**

Source: *APiCS*

Indeed a clear majority of creoles (69%) feature the double-object construction, but again if we apply the bi-clan distribution, the majority shrinks and we nearly have an equal split between languages with exclusive double-object constructions (56%) and those with exclusive indirect-object constructions (44%). Here the bi-clan subdivision helps us to realize that the indirect-object construction in the non-Atlantic creoles, mainly in South and Southeast Asia and the Pacific, also constitutes a widely represented construction type of the world's creoles. In ditransitive constructions, creoles also clearly reflect their substrate/adstrate pattern against possibly conflicting patterns in their lexifiers. This can be detected from a comparison with the corresponding *WALS* map and the information on areal patterning of the constructions in question (for a detailed discussion see Michaelis & Haspelmath 2003 and Haspelmath, Michaelis &

---

14   We do not investigate the secondary-object construction due to its marginal status in the contact languages that are studied here.

APiCS Consortium 2013a). So here again, the narrow perspective on Atlantic creoles has considerably blurred the picture on creoles world-wide.

Another putative typical creole feature in this context is the obligatory use of overt pronominal subjects (see Lipski 1996, Bartens & Sippola 2013), as illustrated in examples (7)-(9):

(7) Louisiana Creole (French-based, North America; Neumann-Holzschuh & Klingler 2013)

> ***Li***   *va*   *kote*   *vye*   *mile-la.*
> 3SG   go   to   old   mule-ART.DEF.SG
> 'He goes over to the old mule.'

(8) Negerhollands (Dutch-based, Caribbean; van Sluijs 2013)

> *Wa*   ***ju***   *lō*   *du*?
> what   2SG   PROG   do
> 'What are you doing?'

(9) Casamancese Creole (Portuguese-based, West Africa; Biagui & Quint 2013)

> ***N***   *kudá* *baŋ* *kumá* ***bu***   *sebé kumá* ***i***   *beŋ kasa.*
> 1SG.SBJ think   PST COMP 2SG.SBJ know COMP 3SG.SBJ come house
> 'I thought that you knew that he had come home.'

On the corresponding *APiCS* map "Expression of pronominal subjects" with all 76 contact languages (Haspelmath & *APiCS* Consortium 2013b; Map 5 below), obligatory pronoun words/affixes are by far the most prominent pattern (obligatory pronoun words/affixes 53, optional pronouns 18). In addition, a striking areal pattern arises: All *APiCS* languages in Africa, the Atlantic and the Americas show obligatory pronoun words/affixes, as well as Australian and Pacific languages, whereas the languages of the Indian Ocean, Southeast Asia and New Guinea allow for optional pronoun words, as in (10) and (11), where there is no pronoun expressed:

(10) Diu Indo-Portuguese (Portuguese-based, South Asia; Cardoso 2013)

> *Kwɔn* *kõpr-o*?
> when buy-PST
> 'When did [you] buy [it]?'

(11) Ternate Chabacano (Spanish-based, South East Asia; Sippola 2013)

> *Ya*   *camina y*   *camina, ta*   *pasa*   *nah*   *monti.*
> PFV   walk and   walk IPFV pass   LOC   mountain
> '[They] walked and walked, while [they] passed the mountain.'

This areal distribution is impressive, as the different marking patterns form very well circumscribed, coherent linguistic areas.

Map 5. Expression of pronominal subjects in 76 *APiCS* contact languages

Obligatory pronoun words [49]
Pronoun affixes [4]
Optional pronoun words [18]
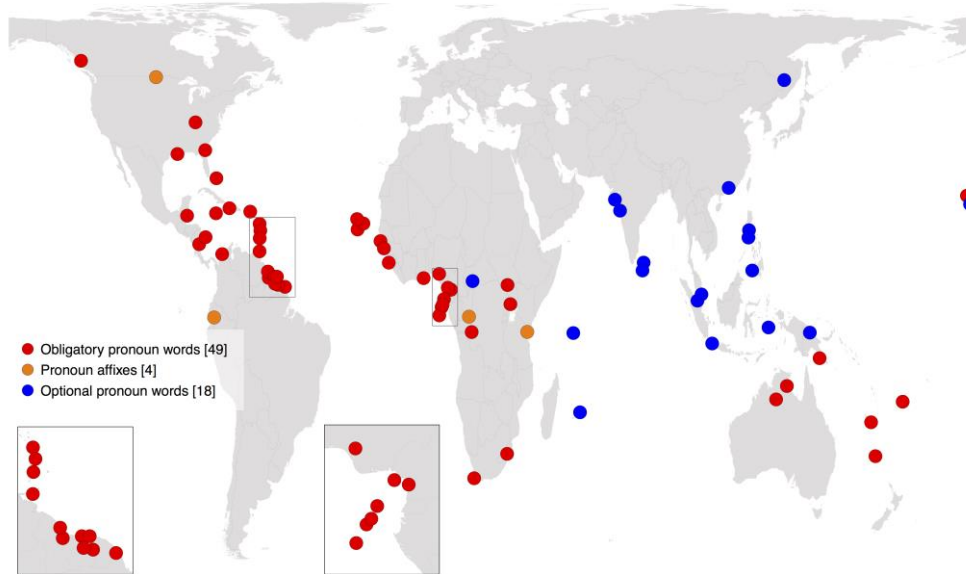
Source: Haspelmath & *APiCS* Consortium (2013b)

Note that this areally compact distribution of obligatory and optional marking holds over different types of languages: pidgins, (pidgin)creoles, mixed languages and restructured varieties[15], strengthens the idea of large areal patterns irrespective of the type of the contact language[16]. When one restricts the view to the group of creoles in *APiCS* (see Table 5), 79% of the languages show obligatory pronominal subjects and 21% have optional pronoun words.

Table 5. Expression of pronominal subjects[17]

	APiCS creoles		*APiCS* bi-clans of creoles	
Obligatory pronoun words/affixes	44	79%	15	**71%**
Optional pronoun words	12	21%	6	**29%**

Source: *APiCS*

In the bi-clan distribution, the figures shift towards 29% of creoles featuring optional marking against 71% obligatory marking. Even if the figures do not change dramatically, the bi-clan perspective again reduces the weight of uniformally-marked large bi-clans (here again European/Macro-Sudan) and enhances at the same time the weight of bi-clans which are represented by fewer

---

15 For more information on the different types of contact languages in *APiCS* see the *Survey of pidgins and creole languages*, ed. by Michaelis et al. (2013b), volumes I-III.

16 See also Bisang (2013, 2015) for a discussion of pronominals and pro-drop in terms of hidden complexity.

17 Some minor values are omitted; for the full picture see Haspelmath & *APiCS* Consortium (2013b).

languages (e.g. Portuguese/Indic, Spanish/Philippinic). This method thus gives a much more realistic picture of the diversity in creoles world-wide. Obligatory pronoun words are just one strategy of creoles world-wide. It so happens that Atlantic creoles overwhelmingly show this feature, but as we have seen, it does not imply that this feature is therefore a pan-creole feature.

As with the other areal features, we also suspect substrate/adstrate influence as the driving force for this clear-cut areal distribution. When we compare the corresponding *WALS* map (Dryer 2005a), the facts are striking: West African substrate languages show a very strong tendency to have obligatory subject pronoun words or affixes (see also Creissels 2005), and even the Portuguese-based creoles of the Atlantic consistently show obligatory subject words whereas their lexifier Portuguese has no such strategy. For the corresponding data of South Asian and Asian substrate languages, see Haspelmath & *APiCS* Consortium (2013b).

6.2. Features that seem rare in creoles
Finally, I will discuss another type of features, namely those features which seem to be rare in creoles world-wide and therefore apparently negligible for the discussion of typical creole or pan-creole features (see also Bakker et al.'s (2011) method of eliminating rarer features from their "typical creole feature list"). The two features are (i) dual in independent personal pronouns and (ii) inclusive/exclusive distinction in independent pronouns. They constitute the mirror image of the areal features which we discussed earlier in that they cluster in non-Atlantic regions. Likewise, as there are much fewer creoles in non-Atlantic regions of the world, these features seem quite marginal at first glance. But again, once we count bi-clans rather than languages, the picture changes significantly.
Dual forms in independent personal pronouns are lacking in the overwhelming majority of *APiCS* creoles, as Map 6 shows (white dots). It is only in the Pacific and Australian areas that languages with pronominal duals are common (red dots), e.g.

(12a) Tayo (French-based, Pacific; Ehrhart & Revis 2013)
   nu ekri a $^{\eta}$gra let
   1PL write INDF.ART long letter
   'We are writing a long letter.'

(12b) Tayo (French-based, Pacific; Ehrhart & Revis 2013)
   nu^{n}de *vote pu USTKE*
   1DU.INCL vote PREP USTKE
   'The two of us vote for the syndicate USTKE.'

(13) Bislama (English-based, Pacific; Meyerhoff 2013)
   yufala   'you all' vs.
   yutufala 'you two'

Map 6. Dual in independent personal pronouns in 59 creoles of *APiCS*
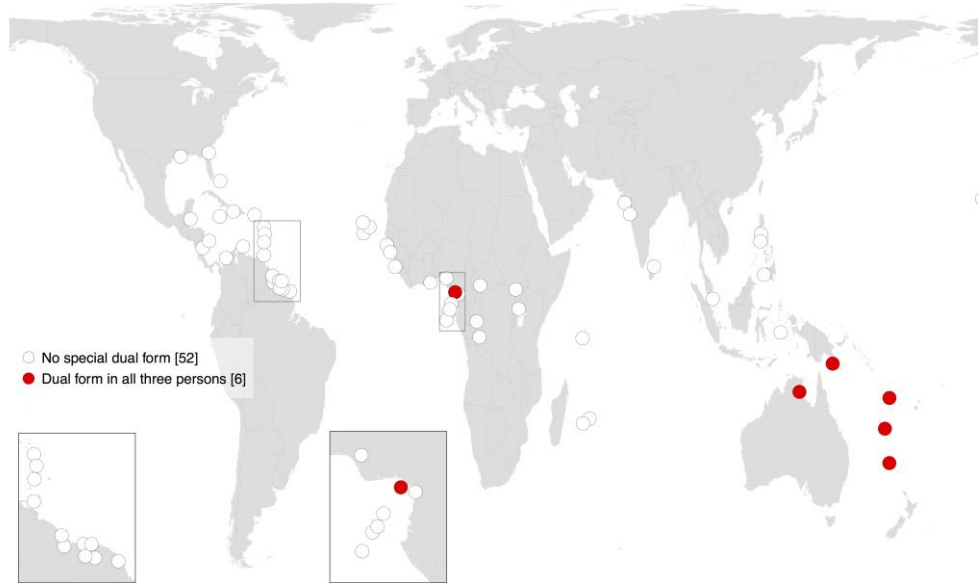
Source: Haspelmath & *APiCS* Consortium (2013a)

Table 6 summarizes the figures: 90% of the creoles lack duals whereas only 10% of them show this marking.

Table 6. Duals in independent pronouns

	APiCS creoles		*APiCS* bi-clans of creoles	
No dual forms	52	90%	17	**78%**
Dual forms	6	10%	5	**23%**

Source: *APiCS*

However, in the bi-clan distribution, the percentage of languages with dual forms more than doubles from 10% to 23%. This means that nearly a quarter of the creole bi-clans in *APiCS* do have dual forms in independent personal pronouns. The presence of dual pronouns is thus a feature that is well represented in creoles, but only in a restricted area of the world. As this area is a non-Atlantic area and comprises relatively few languages, this grammatical phenomenon has not found its way into other cross-creole comparisons (not present in Holm & Patrick 2007 nor in *eWAVE*). We see once again that the bi-clan perspective better represents the (areal) diversity in creoles than mere counting of single languages.

A somewhat similar feature is the inclusive/exclusive distinction in independent personal pronouns (Haspelmath, Michaelis & *APiCS* Consortium 2013b). An inclusive pronoun means 'we including the hearer, i.e. you and me', and an exclusive pronoun means 'we excluding the hearer, i.e. me excluding you', as in:

(14) Tok Pisin (English-based, Pacific; Smith & Siegel 2013)
    yumi   1PL.INCL   'we' = 'you and me'
    vs.
    mipela/mipla 1PL.EXCL   'we' = 'me (excluding you) and he/she/they'

This *APiCS* feature, which was inspired by *WALS* (Cysouw 2005), shows a similar distribution in the *APiCS* creoles as does the preceding feature on dual pronouns.

Map 7. Inclusive/exclusive distinction in independent personal pronouns in 59 creoles of *APiCS*
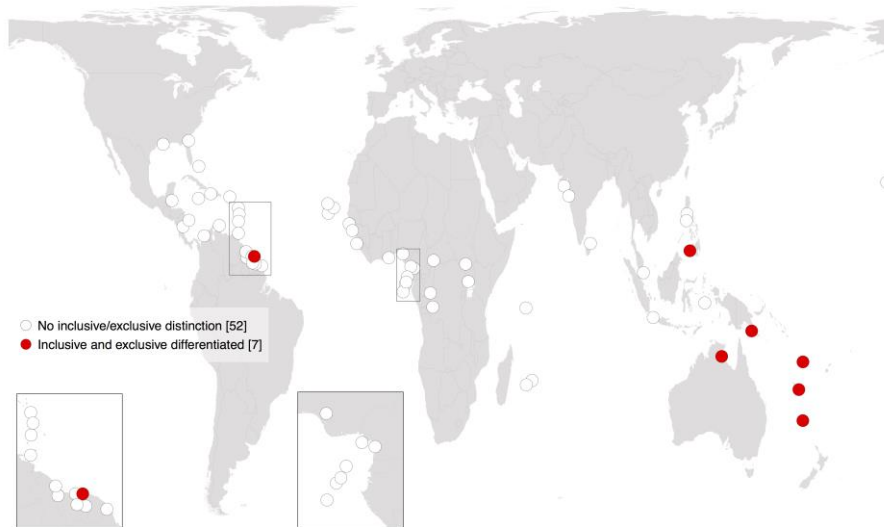
Source: Haspelmath, Michaelis & *APiCS* Consortium (2013b)

Here again, the overwhelming majority of creoles (88%) does not make the inclusive/exclusive distinction, whereas 12% of the creoles make it.

Table 7. Inclusive/exclusive distinction in independent personal pronouns

	APiCS creoles		*APiCS* bi-clans of creoles	
No inclusive/exclusive distinction	52	88%	17	**74%**
Inclusive and exclusive differentiated	7	12%	6	**26%**

Source: *APiCS*

In the bi-clan distribution, the inclusive/exclusive distinction again more than doubles to 26%, i.e. more than a quarter of the creole bi-clans worldwide have this distinction. Thus, this feature cannot be said to be rare in creoles in general.

Both these features, dual and inclusive/exclusive pronouns, are clearly areally-restricted features worldwide. But this areal restriction is in principle of the same nature as directional serial verbs or double-object constructions in Atlantic creoles. As can be seen from the *WALS* map on inclusive/exclusive

distinction in independent personal pronouns (Cysouw 2005), areas where such a distinction is widespread are the Philippines, Australia, and Melanesia. Thus, it is clear that the presence of these features in the creoles of Australia and Melanesia is due to similar patterns in the substrates/adstrates of these contact languages (see Keesing 1988 and subsequent scholars).

For arbitrary historical reasons, these two features have never made it onto any list of pan-creole features. They are prevalent in a region of the world that has not led to a large number of well-established and well-described creole languages. But I showed earlier that directional serial verbs, double-object constructions and obligatory subject pronoun words show the same areal restrictedness, even though in a different area of the world, the Atlantic. Again, for arbitrary historical reasons the Atlantic features have made it on several lists of pan-creole features even though they, too, are just areal features, but present in bi-clans with the largest number of creoles. Qualitatively, they must be treated in the same way as duals and inclusive/exclusive pronouns. Thus, none of the features discussed in this section can be considered a pan-creole feature.

## 7. Candidates for creole universals

As mentioned earlier (§5), candidates for creole universals should fulfill three requirements. They should be

> (i) pan-creole features
> (ii) more likely to be found in creoles than in non-creoles, and
> (iii) not found in the contributing lexifier or substrates of a given creole.

All of the features presented in §6.1 do not even meet the first requirement, as they turn out to be areally-restricted features. So which features are widespread enough over most bi-clans and could thus satisfy the first and potentially also the two other conditions for creole universals? I will consider four *APiCS* features here: comitative/instrumental identity, SVO order, prepositions, and occurrence of nominal plurality. We will see that only one of them is a possible creole universal.

7.1. Pan-creole features which are not creole universals
The first feature, comitatives and instrumentals, was studied by Maurer & *APiCS* Consortium (2013). Languages can express the concepts of comitative and instrumental in the same way (identity; yellow dots), as in English *with* or French *avec* 'together with' or 'by means of', or by different markers (differentiation; red dots), as in Korlai (Portuguese-based creole of South Asia), where *kosid* has comitative meaning, whereas *ku* has instrumental meaning (Clements 2013). Some creoles have two markers, one of which refers both to comitative and instrumental. This pattern is classified as overlap (orange dots); an example is French-based Reunion Creole *ek* 'together with, by means of', and *ansanm* 'together with' (Bollée 2013). Map 8 reflects the overwhelming pattern of identical and overlap marking of comitatives and instrumentals in the creoles of *APiCS*.

Map 8. Comitatives and instrumentals in 59 creoles of *APiCS*
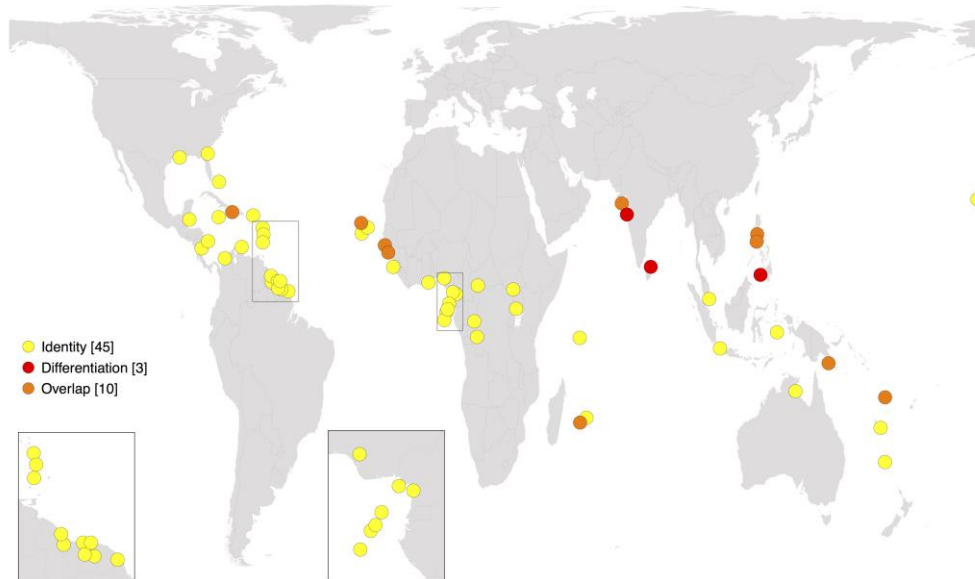
Source: Maurer & *APiCS* Consortium (2013)

The figures in Table 8 illustrate the cross-creole pattern. Not only does the vast majority of creoles show identity or overlap of the two functions in question (95%), but also the bi-clan distribution speaks in favor of a pan-creole feature: 92% of the creole bi-clans in *APiCS* can mark comitative and instrumental in the same way.

Table 8. Comitatives and instrumentals

	APiCS creoles		*APiCS* bi-clans of creoles	
Identity	45	78%	17	**68%**
Overlap	10	17%	6	**24%**
Differentiation	3	5%	2	**8%**

Source: *APiCS*

When we compare these data with the corresponding *WALS* map (Stolz et al. 2005), the second condition cited above also seems to be fulfilled: twice as many languages and genera world-wide have different words to refer to comitative and instrumental, whereas creoles seem to prefer identical expression of both concepts.

But is the third condition also fulfilled? When we examine the lexifiers and substrates of the creoles, we see that it is clearly not fulfilled: All European lexifiers and some important African substrates, too, show the identity or overlap pattern. Therefore, it is quite possible that the creoles have simply retained this polysemous marking from either lexifier or substrate languages, which weakens the idea of a creole universal that has arisen through the special cognitive and socio-cultural conditions of creolization.

The same is true for two other features which are widespread in the creoles of *APiCS*, even in the overwhelming majority of creole bi-clans, SVO word order and prepositions (rather than postpositions). But here too, I would say that in both features the creoles just reflect their contributing languages. The often cited test cases here are in my view only apparent (see Bakker 2008: 140f.).[18]

•   Berbice Dutch, which has SVO word order, despite the SOV order of the Ijo substrate and Dutch's non-dominant word order (as reflected in *WALS*, Dryer 2005b). But Dutch certainly has enough contexts with SVO word order that could have been the model in creolization;
•   Bakker cites Juba Arabic, which has SVO word order, and compares it to Classical Arabic, which had VSO word order. But Juba Arabic must of course go back to *spoken* Arabic varieties, which all have SVO as their preferred value to begin with.

As for prepositions, the situation is similar: The great majority of creoles show prepositions. But there are no cases where a bi-clan with two non-prepositional clans has given rise to a prepositional-marking creole. Therefore, here too, creoles replicate the patterns of their contributing languages. Why word order features often follow the European/lexifier patterns against the African substrate patterns is still an open question[19].

7.2. A possible creole universal feature

The last feature to be discussed in this section is occurrence of plural markers in creole languages. As Map 9 illustrates, the great majority of creole languages in *APiCS* have variable plural marking, i.e. notionally plural noun phrases are sometimes, but not always, plural-marked.

---

[18]   Some South Asian and Philippine creoles are said to have had SVO in earlier stages, but changed since then to SOV word order because of adstrate pressure (see Bakker 2008: 140f.).
[19]   See Blasi et al. 2017 for a quantitative analysis confirming the fact that word order patterns in creoles overwhelmingly mirror word order patterns of their lexifiers.

Map 9. Occurrence of nominal plural markers in 59 creoles of *APiCS*
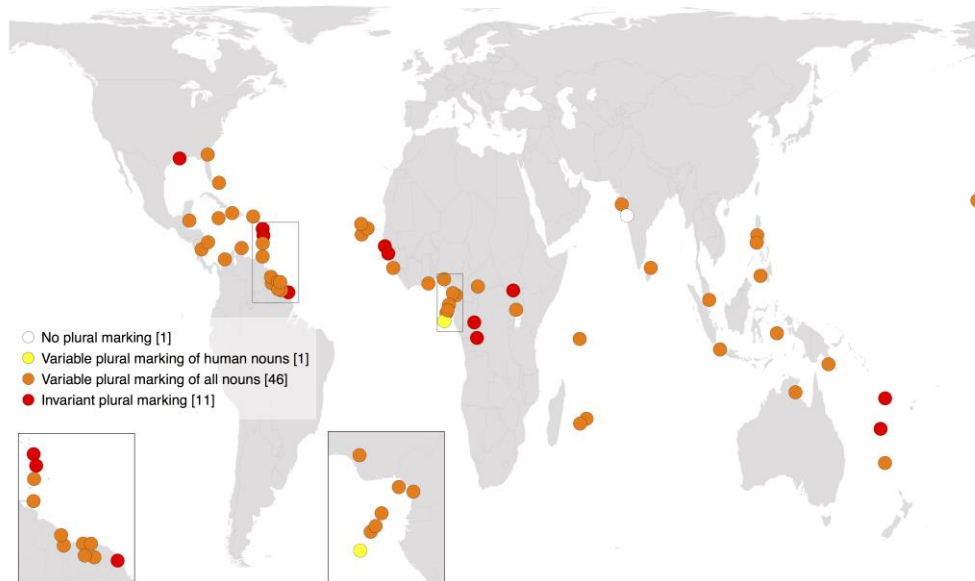
Source: Haspelmath & *APiCS* Consortium (2013c)

The variable occurrence can be conditioned by different factors, often involving animacy and definiteness (see Haspelmath & *APiCS* Consortium 2013c). Some areally-restricted bi-clans (e.g. European/Macro-Sudan, English/Oceanic) have innovated plural markers that derive from third-person-plural pronouns[20], e.g.

(15) Nigerian Pidgin (English-based, West Africa; Faraclas 2013)
     got   dèm,   ston   dèm
     goat   3PL   stone   3PL
     'goats, stones'

Often these plural markers can only occur with definite noun phrases, a restriction which derives from the semantics of personal pronouns, which are definite by nature. In these languages the plural marker thus has a two-fold function, marking definiteness and plurality.

(16) Papiamentu (Spanish-based, Caribbean; Kouwenberg 2013)
     *baka-**nan***
     cow-(3)PL
     'the cows' (only definite)

(17) Nengee (English-based, Caribbean; Migge 2013)
     ***den**   pikin*
     (3)PL   child
     'the children' (only definite)

---

20   This is another areally-restricted feature that was proposed as a creole universal
        (e.g. by Taylor 1971 and Markey 1982).

Other creoles with variable plural marking have innovated their plural markers from words meaning 'all' (Diu Indo-Portuguese *tud* from Portuguese *tudo* 'all') or words deriving from a noun meaning 'group' (e.g. Seychelles and Mauritian Creole *bann*, from French *bande*; see Bollée 2000, Michaelis & Haspelmath 2020).

(18) Diu Indo-Portuguese (Portuguese-based, South Asia; Cardoso 2013)
     *Es   tud   ɛ   kaz   də **tud** pad.*
     this   PL   COP.NPST house   of PL priest
     'These are the houses of the priests.'

(19) Seychelles Creole (French-based, Indian Ocean; Michaelis & Rosalie 2013)
     bann   *zanfan*
     pl   child
     'the children'

82% of the creoles in *APiCS* have variable plural marking. When we count the languages by bi-clan 76% have variable plural marking, whereas 24% have invariant plural marking.

Table 9. Occurrence of nominal plural markers

	APiCS creoles		*APiCS* bi-clans of creoles	
Variable marking of human or inanimate nouns	46	80%	18	**72%**
Variable marking of human nouns	1	2%	1	**4%**
Invariant plural marking	11	19%	6	**24%**

Source: *APiCS*

     Variable marking is not present in 6 bi-clans. This may seem to disqualify the occurrence of plural marking as a pan-creole feature in the first place. But the argument goes as follows: For some of the creoles with variable plural marking, we can again safely invoke lexifier or substrate patterns as we did for features in §7.1. Many Southeast Asian, Philippinic, and Australian substrate/adstrate languages have variable plural marking. Thus, it may be that creoles such as Papiá Kristang (Portuguese-based, Southeast Asia), the Chabacano varieties (Spanish-based, Southeast Asia), and Kriol (English-based, Australia) mirror the pattern of their substrates/adstrates. But this is the first feature where we also find creoles which go against their contributing languages. We find clear cases where both lexifier and substrate show invariant marking, but the corresponding creole has variable marking, e.g. many English-based Atlantic creoles. Fongbe and Ewe, but also English, have invariant plural marking, whereas Nigerian Pidgin, Saramaccan, and Jamaican have variable marking (cf. *WALS* chapter by Haspelmath 2005).
     Why is this feature important in the discussion of creole universals? Variable plural marking in Nigerian Pidgin or Seychelles Creole points to

diachronic processes by which new grammatical categories are on their way to being grammaticalized. Much of the old plural-marking morphology of the lexifiers got lost during the creolization process. Therefore, new strategies have been created and gradually grammaticalized. Variability is one of the key properties of new plural – and other grammatical – markers during the grammaticalization process, where constructions have been fixed to a certain degree, but have not reached invariance in each plural context[21]. Therefore, the behavior of plural markers in creoles is one salient feature which points in the direction where we should systematically look for universal creole features: features which reflect diachronic processes in creolization[22]. Many of the grammatical features which I have discussed in this paper result from language change processes where essentially the lexifier's and/or substrate's structural pattern prevails in the new creole language. But the features unique to creoles, i.e. creole universals, are really *diachronic universals* (cf. Bybee 2006). Thus, variability in grammatical marking, here specifically: variable marking of nominal plural seems to be one of the most promising creole universals as it relates to the nature of freshly grammaticalized markers in creoles. But this is a topic of another paper (see Michaelis & Haspelmath 2020).

8. Conclusion

If we want to generalize over creole languages, we need to avoid bias and consider cases that are as independent of each other as possible. Just counting creole languages in a large database (such as *APiCS*) irrespective of their genealogical and areal relatedness is not enough. Thus, I suggest that groups of creoles which are historically closely connected and share both the lexifier and the substrate type should be counted only once. In other words, rather than counting languages, one should count bi-clans.

  Furthermore, I showed that features which seem wide-spread in creoles may turn out not to have a pan-creole status once the bi-clan distribution is considered. Likewise, features which seem rare in creoles world-wide turn out to be not rare, but just areally-restricted, where areal restriction often points to substrate/adstrate influence.

  Finally I suggested that creole universals are really diachronic universals: The loss of much grammatical marking (not only inflectional marking) and the subsequent restructuring and renewal processes in creole languages have left their unique footprints: the unusual amount of newly-grammaticalized structures often entails variable marking, which then is one good diagnostic of creole grammars.

---

[21] Cf. McWhorter (2011: 91ff.)

[22] See also Daval-Markussen & Bakker (2017), Daval-Markussen (2018: Ch. 8).

References

Alleyne, Mervyn C. 1980. *Comparative Afro-American: An historical-comparative study of English-based Afro-American dialects*. Ann Arbor: Karoma.

Bakker, Dik. 2011. Language sampling. In Jae Jung Song (ed.). *The Oxford handbook of linguistic typology*, 100-127. Oxford: Oxford University Press.

Bakker, Peter. 2008. Pidgins versus creoles and pidgincreoles. In Silvia Kouwenberg & John Victor Singler (eds.), *Handbook of Pidgin and Creole Studies*, Oxford: Wiley-Blackwell, 130-57.

Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall & Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26(1). 5–42. doi:10.1075/jpcl.26.1.02bak.

Bartens, Angela & Sippola, Eeva. 2013. Subject null arguments in creole languages. Paper read at the conference "Subject: cognitive, typological and functional approaches", University of Helsinki.

Bell, Alan. 1978. Language samples. In Joseph Greenberg & Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of Human Language.* Cambridge, Mass.: MIT Press, 123-156.

Biagui, Noël Bernard & Nicolas Quint. 2013. Casamancese Creole structure dataset. In Susanne Maria Michaelis & Philippe Maurer & Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/34.

Bickel, B. 2008. A refined sampling procedure for genealogical control. *Language Typology and Universals (STUF)* 61. 221–233.

Bickerton, Derek. 1981. *Roots of language*. Ann Arbor: Karoma.

Bickerton, Derek. 1989. Seselwa serialization and its significance. *Journal of Pidgin and Creole Languages* 4. 155–83.

Bickerton, Derek. 1995. The syntax of creole languages. In Jacobs, Joachim & von Stechow, Arnim & Sternefeld, Wolfgang & Vennemann, Theo (eds.) *Syntax: An international handbook of contemporary research.* 2 vols. Berlin: de Gruyter, 1451-62.

Bickerton, Derek. 1996. Why serial verb constructions in Isle de France Creole can have subjects: A reply to Corne, Coleman & Curnow. In Philip Baker & Anand Syea (eds.), *Changing meanings, changing functions.* London: University of Westminster Press, 155–69.

Bisang, Walter. 2013. Language contact between geographic and mental space. In Peter Auer, Martin Hilpert, Anja Stukenbrock & Benedikt Szmrecsanyi, *Linguistic Perspectives on Space: Geography, Interaction, and Cognition*. Berlin: Mouton de Gruyter, 61-100.

Bisang, Walter. 2015. The neglected side of complexity and its consequences. *Linguistic Vanguard* 1:1.

Blasi, Damián E. & Michaelis, Susanne Maria & Haspelmath, Martin (2017). Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour*, 1, 723–729.

Bollée, Annegret. 2000. La restructuration du pluriel nominal dans les créoles de l'Océan Indien. *Études Créoles* 23. 25-39

Bollée, Annegret. 2013. Reunion Creole structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/54

Boretzky, Norbert. 1983. *Kreolsprachen, Substrate und Sprachwandel*. Wiesbaden: Otto Harrassowitz.

Bruyn, Adrienne, Pieter Muysken & Maaike Verrips. 1999. Double-object constructions in the creole languages: Development and acquisition. In Michel DeGraff (ed.), *Language Creation and Language Change*, 329–373. Cambridge, MA: MIT Press.

Buchstaller, Isabelle, Anders Holmberg & Mohammad Almoaily. 2014. *Pidgins and creoles beyond Africa-Europe encounters*. Amsterdam: Benjamins.

Bybee, Joan. 2006. Language change and universals. In Ricardo Mairal & Juana Gil (eds.), *Linguistic universals*, 179–194. Cambridge: Cambridge University Press.

Byrne, Francis. 1987. *Grammatical relations in a radical creole: Verb complementation in Saramaccan*. Amsterdam: John Benjamins.

Cardoso, Hugo C. 2013. Diu Indo-Portuguese structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/39.

Clements, J. Clancy. 2013. Korlai structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/40

Colot, Serge & Ludwig, Ralph. 2013. Martinican Creole structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://apics-online.info/contributions/51

Creissels, Denis. 2005. A typology of subject and object markers in Africanlanguages. In Voeltz, F. K. Erhard (ed.), *Studies in African linguistic typology*, 43–70. Amsterdam/Philadelphia: John Benjamins.

Cysouw, Michael. 2005. Inclusive/exclusive forms for "we." In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 162–169. Oxford: Oxford University Press.

Daval-Markussen, Aymeric. 2014. First steps towards a typological profile of creoles. *Acta Linguistica Hafniensia* 45(2), 141–50. doi:10.1080/03740463.2014.880606

Daval-Markussen, Aymeric. 2018. *Reconstructing creole*. Doctoral thesis, University of Aarhus.

Daval-Markussen, Aymeric & Bakker, Peter. 2017. Creole typology II: Typological features of creoles: from early proposals to phylogenetic approaches and comparisons with non-creoles. In Peter Bakker, Finn Borchsenius, Carsten Levisen & Eeva Sippola (eds.), *Creole Studies – Phylogenetic Approaches*. Amsterdam/Philadelphia: John Benjamins, 103-140.

Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257 – 292.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138.

Dryer, Matthew S. 2005a. Expression of pronominal subjects. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 410–13. Oxford: Oxford University Press.

Dryer, Matthew S. 2005b. Order of subject, object and verb. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 330–3. Oxford: Oxford University Press

Ehrhart, Sabine & Revis, Melanie. 2013. Tayo structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://apics-online.info/contributions/57

Faraclas, Nicholas. 2013. Nigerian Pidgin structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/17.

Fattier, Dominique. 2013. Haitian Creole structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/49.

Finney, Malcolm Awadajin. 2013. Krio structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/15

Goodman, Morris F. 1964 . *A comparative study of creole French dialects*. Den Haag: Mouton.

Güldemann, Tom. 2010. Sprachraum and geography: Linguistic macro-areas in Africa. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation*, vol. Vol. 2: Language mapping, 561–585, Maps 2901–2914. (HSK 30.2). Berlin, New York: De Gruyter Mouton.

Hagemeijer, Tjerk. 2013. Santome structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of*

Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/35.

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

Haspelmath, Martin. 2005. Occurrence of nominal plurality. In Martin Haspelmath & Matthew S. Dryer & David Gil & Bernard Comrie (eds.) *The World Atlas of Language Structures*. Oxford: Oxford University Press, 142-145.

Haspelmath, Martin & Matthew S. Dryer & David Gil & Bernard Comrie (eds.). 2005. *The world atlas of language structures*, Oxford: Oxford University Press.

Haspelmath, Martin & The *APiCS* Consortium. 2013a. Dual in independent personal pronouns. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The Atlas of Pidgin and Creole Language Structures*, Oxford: Oxford University Press, 54–57. https://apics-online.info/parameters/14.chapter.html

Haspelmath, Martin & The *APiCS* Consortium. 2013b. Expression of pronominal subjects. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The Atlas of Pidgin and Creole Language Structures*, Oxford: Oxford University Press, 244–247. https://apics-online.info/parameters/62.chapter.html

Haspelmath, Martin & The *APiCS* Consortium. 2013c. Occurrence of nominal plural markers. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The Atlas of Pidgin and Creole Language Structures*, 84-87. Oxford: Oxford University Press. https://apics-online.info/parameters/22.chapter.html

Haspelmath, Martin, Susanne Maria Michaelis & The *APiCS* Consortium. 2013a. Ditransitive constructions with "give." In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The Atlas of Pidgin and Creole Language Structures*, Oxford: Oxford University Press, 236–239. https://apics-online.info/parameters/60.chapter.html

Haspelmath, Martin, Susanne Maria Michaelis & The *APiCS* Consortium. 2013b. Inclusive/exclusive distinction in independent personal pronouns. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures*, Oxford: Oxford University Press, 58–59. https://apics-online.info/parameters/15.chapter.html

Haspelmath, Martin & Michaelis, Susanne Maria. 2017. Analytic and synthetic: Typological change in varieties of European languages. In Isabelle Buchstaller & Beat Siebenhaar (eds.), *Language Variation – European Perspectives VI: Selected papers from the 8th International Conference on Language Variation in Europe* (*ICLaVE 8*), *Leipzig 2015*, Amsterdam: Benjamins, 3-22.

Holm, John A. 1988. *Pidgins and creoles. Volume 1: Theory and structure*. Cambridge: Cambridge University Press.

Holm, John A. & Peter L. Patrick. 2007. *Comparative creole syntax: Parallel outlines of 18 creole grammars*. London: Battlebridge.

Keesing, Roger M. 1988. Melanesian pidgin and Oceanic substrate. Stanford: Stanford University Press.

Kortmann, Bernd & Kerstin Lunkenheimer (eds.). 2013. *Electronic World Atlas of Varieties of English (eWAVE).* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://ewave-atlas.org/.

Kortmann, Bernd, Kate Burridge, Rajend Mesthrie, Edgar W. Schneider & Clive Upton (eds.). 2004. *Varieties of English*. Berlin: Mouton de Gruyter.

Kortmann, Bernd & Benedikt Szmrecsanyi. 2004. Morphological and syntactic variation in English. In Bernd Kortmann, Kate Burridge, Rajend Mesthrie, Edgar W. Schneider & Clive Upton (eds.), *Varieties of English*. Berlin: Mouton de Gruyter, 1142-1202.

Kouwenberg, Silvia. 2013. Papiamentu structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/47

Lipski, John.1996. The evolution of null subjects in Philippine Creole Spanish. *1994 Mid-American Linguistics Conference Papers Volume II*, Lawrence Kansas: University of Kansas Linguistics Department, 387-401.

Markey, Thomas L. 1982. Afrikaans: Creole or Non-Creole? *Zeitschrift für Dialektologie und Linguistik*. 49 (2). 169–207.

Maurer, Philippe & The *APiCS* Consortium. 2013. Comitatives and instrumentals. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The Atlas of Pidgin and Creole Language Structures*, Oxford: Oxford University Press, 276-279. https://apics-online.info/parameters/70.chapter.html

Maurer, Philippe, Susanne Maria Michaelis & The *APiCS* Consortium. 2013. Directional serial verb constructions with "come" and "go." In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The Atlas of Pidgin and Creole Language Structures*, Oxford: Oxford University Press, 334–337. https://apics-online.info/parameters/84.chapter.html

McWhorter, John H. 1997. *Towards a new model of creole genesis*. New York: Peter Lang.

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2-3). 125–166.

McWhorter, John H. 2011. *Linguistic Simplicity and Complexity: Why do languages undress?*, Berlin: Mouton de Gruyter.

Meyerhoff, Miriam. 2013. Bislama structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/23.

Michaelis, Susanne & Martin Haspelmath. 2003. Ditransitive constructions: Creole languages in a cross-linguistic perspective. *Creolica*. http://www.creolica.net/article-28.html

Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.). 2013a. *The Atlas of Pidgin and Creole Language Structures*. Oxford: Oxford University Press.

Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.). 2013b. *The Survey of Pidgin and Creole Languages*, 3 vols. Oxford: Oxford University Press.

Michaelis, Susanne Maria & Marcel Rosalie. 2013. Seychelles Creole structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/56.

Michaelis, Susanne Maria & Haspelmath, Martin. 2020. Grammaticalization in creole languages: Accelerated functionalization and semantic imitation. In Walter Bisang & Andrej Malchukov (eds.), *Grammaticalization scenarios: Cross-linguistic variation and universal tendencies*, vol. 2, Berlin: De Gruyter Mouton, 1109-1128.

Migge, Bettina. 2013. Nengee structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://apics-online.info/contributions/4

Neumann-Holzschuh, Ingrid & Klingler, Tom. 2013. Louisiana Creole structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://apics-online.info/contributions/53

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*, Amsterdam: Benjamins, 265–285.

Perkins, Revere D. 2001. Sampling Procedures and Statistical Methods. In: Haspelmath, Martin & Ekkehard König & Wulf Oesterreicher & Wolfgang Raible (eds.), *Language Typology and Linguistic Universals – an International Handbook*, Vol. 1, Berlin/New York: Walter de Gruyter, 419-434.

Rijkhoff, Jan & Bakker, Dik. 1998. Language sampling. *Linguistic Typology* 2-3: 263-314.

Schwegler, Armin. 2013. Palenquero structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://apics-online.info/contributions/48.

Sippola., Eeva. 2013. Ternate Chabacano structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

https://apics-online.info/contributions/44

Sluijs, Robbert van. 2013. Negerhollands structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/27.

Song, Jae Jung 2001. *Linguistic Typology: Morphology and Syntax*. Harlow and London: Pearson Education (Longman).

Smith, Geoff P. & Jeff Siegel. 2013. Tok Pisin structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/22.

Stolz, Thomas, Cornelia Stroh & Aina Urdze. 2005. Comitatives and Instrumentals. In: Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures,* 214-17. Oxford: Oxford University Press.

Szmrecsanyi, Benedikt & Bernd Kortmann. 2009. Vernacular universals and angloversals in a typological perspective. In: Filppula, Markku, Juhani Klemola & Heli Paulasto (eds), *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. London/New York: Routledge, 33-53.

Taylor, Douglas. 1971. Grammatical and Lexical Affinities of Creoles. In Dell Hymes (ed.), *Pidginization and Creolization of Languages*. Cambridge: Cambridge University Press, 293–296.