

In the format provided by the authors and unedited.

Grammars are robustly transmitted even during the emergence of creole languages

Damián E. Blasi^{1,2*}, Susanne Maria Michaelis^{2,3}  and Martin Haspelmath^{2,3} 

¹Department of Comparative Linguistics, University of Zürich, Plattenstrasse 54, 8032 Zürich, Switzerland. ²Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Kahlaische Str. 10, 07745 Jena, Germany. ³Department of English Studies, Leipzig University, Nikolaistrasse 8-10, 04109 Leipzig, Germany. *e-mail: damian.blasi@uzh.ch

Supplementary materials for
**Grammars are robustly transmitted even during the emergence of
creole languages**

Damián E. Blasi^{1,2}, Susanne M. Michaelis^{3,2} and Martin Haspelmath^{2,3}

1: Department of Comparative Linguistics, University of Zürich, Plattenstrasse 54, 8032 Zürich, Switzerland.

2: Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Kahlaische Str. 10, 07745 Jena, Germany.

3: Department of English Studies, Leipzig University, Nikolaistrasse 8-10, 04109 Leipzig, Germany

Element	Page
Supplementary table 1.....	2
Supplementary table 2.....	6
Supplementary table 3.....	8
Supplementary table 4.....	12
Supplementary table 5.....	14
Supplementary table 6.....	15
Supplementary figure 1.....	16
Supplementary figure 2.....	17
Supplementary figure 3.....	18

Supplementary tables

WALS (non-creole) languages

WALS code	Language	Genus	Family
abk	Abkhaz	Northwest Caucasian	Northwest Caucasian
aco	Acoma	Keresan	Keresan
aeg	Arabic (Egyptian)	Semitic	Afro-Asiatic
ain	Ainu	Ainu	Ainu
ala	Alamblak	Sepik Hill	Sepik
ame	Amele	Madang	Trans-New Guinea
apu	Apurinã	Purus	Arawakan
arm	Armenian (Eastern)	Armenian	Indo-European
arp	Arapesh (Mountain)	Kombio-Arapesh	Torricelli
awp	Awa Pit	Barbacoan	Barbacoan
bag	Bagirmi	Bongo-Bagirmi	Central Sudanic
bma	Berber (Middle Atlas)	Berber	Afro-Asiatic
brm	Burmese	Burmese-Lolo	Sino-Tibetan
brs	Barasano	Tucanoan	Tucanoan
bsq	Basque	Basque	Basque
bur	Burushaski	Burushaski	Burushaski
cha	Chamorro	Chamorro	Austronesian
chk	Chukchi	Northern Chukotko-Kamchatkan	Chukotko-Kamchatkan
ckr	Canela-Krahô	Ge-Kaingang	Macro-Ge
cmn	Comanche	Numic	Uto-Aztecan
coo	Coos (Hanis)	Coosan	Oregon Coast
eng	English	Germanic	Indo-European
epe	Epena Pedee	Choco	Choco
eve	Evenki	Tungusic	Altaic
ewe	Ewe	Kwa	Niger-Congo
fin	Finnish	Finnic	Uralic
gar	Garo	Bodo-Garo	Sino-Tibetan
geo	Georgian	Kartvelian	Kartvelian

goo	Gooniyandi	Bunuban	Australian
grb	Grebo	Kru	Niger-Congo
grk	Greek (Modern)	Greek	Indo-European
grw	Greenlandic (West)	Eskimo	Eskimo-Aleut
gua	Guaraní	Tupi-Guaraní	Tupian
hau	Hausa	West Chadic	Afro-Asiatic
hin	Hindi	Indic	Indo-European
hix	Hixkaryana	Cariban	Cariban
hmo	Hmong Njua	Hmong-Mien	Hmong-Mien
hun	Hungarian	Ugric	Uralic
hzb	Hunzib	Avar-Andic-Tsezic	Nakh-Daghestanian
imo	Imonda	Border	Border
ind	Indonesian	Malayo-Sumbawan	Austronesian
iri	Irish	Celtic	Indo-European
irq	Iraqw	Southern Cushitic	Afro-Asiatic
jak	Jakaltek	Mayan	Mayan
jpn	Japanese	Japanese	Japanese
juh	Jul'hoan	Ju-Kung	Kxa
kay	Kayardild	Tangkic	Australian
ket	Ket	Yeniseian	Yeniseian
kha	Khalkha	Mongolic	Altaic
khm	Khmer	Khmer	Austro-Asiatic
kho	Khoekhoe	Khoe-Kwadi	Khoe-Kwadi
khs	Khasi	Khasian	Austro-Asiatic
kio	Kiowa	Kiowa-Tanoan	Kiowa-Tanoan
knd	Kannada	Southern Dravidian	Dravidian
knm	Kunama	Kunama	Kunama
knr	Kanuri	Western Saharan	Saharan
koa	Koasati	Muskogean	Muskogean
kor	Korean	Korean	Korean
krk	Karok	Karok	Karok
kro	Krongo	Kadugli	Kadu
kse	Koyraboro Senni	Songhay	Songhay

kut	Kutenai	Kutenai	Kutenai
lan	Lango	Nilotic	Eastern Sudanic
lat	Latvian	Baltic	Indo-European
lav	Lavukaleve	Lavukaleve	Solomons East Papuan
lez	Lezgian	Lezgic	Nakh-Daghestanian
lkt	Lakhota	Siouan	Siouan
mal	Malagasy	Barito	Austronesian
mao	Maori	Oceanic	Austronesian
map	Mapudungun	Araucanian	Araucanian
mar	Maricopa	Yuman	Hokan
mau	Maung	Iwaidjan	Australian
may	Maybrat	North-Central Bird's Head	West Papuan
mei	Meithei	Kuki-Chin	Sino-Tibetan
mnd	Mandarin	Chinese	Sino-Tibetan
mun	Mundari	Munda	Austro-Asiatic
mxc	Mixtec (Chalcatongo)	Mixtecan	Oto-Manguean
myi	Mangarrayi	Mangarrayi	Australian
ngi	Ngiyambaa	Pama-Nyungan	Australian
nht	Nahuatl (Tetelcingo)	Aztecán	Uto-Aztecán
niv	Nivkh	Nivkh	Nivkh
nug	Nunggubuyu	Nunggubuyu	Australian
orh	Oromo (Harar)	Lowland East Cushitic	Afro-Asiatic
pau	Paumari	Arauan	Arauan
prh	Pirahã	Mura	Mura
prs	Persian	Iranian	Indo-European
qim	Quechua (Imbabura)	Quechuan	Quechuan
ram	Rama	Rama	Chibchan
rus	Russian	Slavic	Indo-European
shk	Shipibo-Konibo	Panoan	Panoan
sla	Slave	Athapaskan	Na-Dene
snm	Sanuma	Yanomam	Yanomam
spa	Spanish	Romance	Indo-European

sup	Supyire	Gur	Niger-Congo
swa	Swahili	Bantoid	Niger-Congo
tab	Taba	South Halmahera - West New Guinea	Austronesian
tag	Tagalog	Greater Central Philippine	Austronesian
tha	Thai	Kam-Tai	Tai-Kadai
tiw	Tiwi	Tiwan	Australian
tru	Trumai	Trumai	Trumai
tuk	Tukang Besi	Celebic	Austronesian
tur	Turkish	Turkic	Altaic
vie	Vietnamese	Viet-Muong	Austro-Asiatic
war	Wari'	Chapacura-Wanham	Chapacura-Wanham
wch	Wichí	Matacoan	Matacoan
wra	Warao	Warao	Warao
yag	Yagua	Peba-Yaguan	Peba-Yaguan
yaq	Yaqui	Cahita	Uto-Aztecan
yim	Yimas	Lower Sepik	Lower Sepik-Ramu
yko	Yukaghir (Kolyma)	Yukaghir	Yukaghir
yor	Yoruba	Defoid	Niger-Congo

Supplementary Table 1: non-creole languages extracted from WALS

APiCS (creole) languages

Language	Lexifier group	Substrate group
Ambon Malay	Other	Austronesian
Angolar	Romance	Macro-Sudan
Bahamian Creole	Germanic	Macro-Sudan
Belizean Creole	Germanic	Macro-Sudan
Berbice Dutch	Germanic	Other
Bislama	Germanic	Austronesian
Cameroon Pidgin English	Germanic	Macro-Sudan
Casamancese Creole	Romance	Macro-Sudan
Creolese	Germanic	Macro-Sudan
Diu Indo-Portuguese	Romance	Other
Early Sranan	Germanic	Macro-Sudan
Fa d'Ambo	Romance	Macro-Sudan
Ghanaian Pidgin English	Germanic	Macro-Sudan
Guinea-Bissau Kriyol	Romance	Macro-Sudan
Gullah	Germanic	Macro-Sudan
Guyanais	Romance	Macro-Sudan
Haitian Creole	Romance	Macro-Sudan
Hawaii Creole	Germanic	Other
Jamaican	Germanic	Macro-Sudan
Juba Arabic	Other	Other
Kinubi	Other	Other
Korlai	Romance	Other
Krio	Germanic	Macro-Sudan
Kriol	Germanic	Other
Lingala	Other	Other
Louisiana Creole	Romance	Macro-Sudan
Martinican Creole	Romance	Macro-Sudan
Negerhollands	Germanic	Macro-Sudan

Nengee	Germanic	Macro-Sudan
Nicaraguan Creole English	Germanic	Macro-Sudan
Nigerian Pidgin	Germanic	Macro-Sudan
Norfolk	Germanic	Austronesian
Palenquero	Romance	Other
Papia Kristang	Romance	Other
Papiamentu	Romance	Macro-Sudan
Pichi	Germanic	Macro-Sudan
Principense	Romance	Macro-Sudan
Reunion Creole	Romance	Other
Sango	Other	Other
Saramaccan	Germanic	Macro-Sudan
Seychelles Creole	Romance	Other
Sranan	Germanic	Macro-Sudan
Sri Lanka Portuguese	Romance	Other
Tayo	Romance	Austronesian
Ternate Chabacano	Romance	Austronesian
Tok Pisin	Germanic	Other
Trinidad English Creole	Germanic	Macro-Sudan
Vincentian Creole	Germanic	Macro-Sudan

Supplementary Table 2: creole languages extracted from APiCS and their lexifier and ancestry groups.

	Lexifier			Substrate			Base rate
	lfdr	Rate	Cramer's V	lfdr	Rate	Cramer's V	
Order of subject, object and verb	0.207	0	0.22	0.149	0	0.36	0.938
Order of genitive and noun	0	0.209	0.48	0.612	0.063	0.25	0.458
Order of adjective and noun	0	0.083	0.58	1	0	0.19	0.625
Order of adposition and NP	0.224	0	0.22	1	0	0.21	0.917
Order of demonstrative and noun	0.066	0.083	0.32	1	0.02	0.17	0.542
Order of numeral and noun	0.004	0.041	0.73	1	0	0.2	0.917
Order of RC and noun	0.299	0	0.2	0.665	0	0.25	0.938
Order of degree word and adjective	0.036	0.063	0.36	0.422	0	0.25	0.562
Pos. of interrogative phrases	0.002	0.062	0.44	0.142	0.02	0.38	0.688
Gender in indep. pers. pronouns	0.066	0	0.35	0.665	0	0.25	0.771
Inclusive/exclusive dist. in indep. pers. pron.	0.207	0	0.26	0.142	0.021	0.5	0.875
Politeness in pronouns	0.224	0	0.21	0.665	0	0.22	0.732
Indefinite pronouns	1	0	0.16	0.612	0	0.32	0.812
Occurrence of nominal plurality	0.036	0	0.36	0.612	0	0.26	0.771
Coding of nominal plurality	0.002	0.02	0.54	0.009	0.062	0.52	0.542
Definite articles	0.002	0.062	0.69	0.001	0.125	0.5	0.542
Indefinite articles	0.004	0.042	0.64	0.328	0	0.26	0.625
Pronominal and adnominal demonstratives	0.14	0.063	0.24	0.422	0.084	0.28	0.479
Distance contrasts in demonstratives	0.207	0	0.27	1	0	0.2	0.667
Ordinal numerals	0.078	0.048	0.43	0.665	0	0.38	0.357
Numeral classifiers	0.207	0	0.23	0.665	0	0.2	0.938
Locus of marking in pos. NP	0.002	0.104	0.48	0.001	0.229	0.61	0.417
Suppletion acc. to tense and aspect	1	0.021	0.22	1	0.021	0.19	0.521
Prohibitive	0.036	0.043	0.37	0.001	0	0.47	0.702
Alignment of case marking and full NP	0.036	0	0.4	0.149	0	0.37	0.896
Alignment of case marking of pronouns	0.101	0.042	0.3	1	0	0.09	0.708
Ditransitive. constructions, give	0.066	0.062	0.29	0.043	0.104	0.38	0.542
Expression of pronominal subjects	0.015	0	0.48	0.018	0	0.44	0.771
Comitatives and instrumentals	0.14	0	0.26	0.149	0	0.31	0.766

Nominal and verbal conjunction	0.015	0.083	0.44	0.328	0.021	0.33	0.646
Zero copula for predicative nominals	0.101	0.083	0.31	0.665	0.083	0.23	0.5
Nominal and locational predication	0.078	0.084	0.3	1	0	0.16	0.562
Predicative possession	0.071	0	0.32	0.665	0	0.34	0.909
Intensifiers and reflexive pronouns	0.015	0.093	0.48	0.149	0.07	0.44	0.628
Reciprocal constructions	0.207	0	0.31	1	0	0.26	0.674
Applicative constructions	0.071	0	0.48	1	0	0.2	0.979
Want complement subjects	1	0	0.12	0.665	0	0.24	0.771
Negative morphemes	0.299	0	0.25	1	0	0.22	0.95
Negative indefinite pronouns and predicate negation	0.066	0	0.33	0.328	0	0.26	0.792
Tone	0.14	0	0.26	0.032	0.063	0.5	0.66
Vowel nasalization	0.036	0.041	0.44	0.665	0	0.23	0.688
Position of indefinite article in the noun phrase	0	0.069	0.66	0.61	0	0.25	0.886
Dual in independent personal pronouns	0.224	0	0.22	0.149	0	0.41	0.886
Person syncretism in independent personal pronouns	0.207	0	0.23	0.328	0	0.28	0.864
Special dependent person forms for subject and object	1	0.023	0.19	0.399	0.068	0.31	0.409
Interrogative.pronouns	0.036	0.091	0.42	0.179	0.046	0.41	0.295
The associative plural	0.224	0.023	0.26	0.011	0.046	0.53	0.545
Nominal plural marker and third person plural pronoun	0.015	0.159	0.45	0.005	0.204	0.52	0.341
Functions of reduplication	0.066	0	0.43	0.328	0	0.37	0.5
Pronominal and adnominal demonstratives	0.207	0.046	0.23	0.179	0.114	0.33	0.477
Adnominal distributive numerals	0.299	0	0.21	1	0	0.14	0.634
Sortal numeral classifiers	0.224	0	0.23	0.665	0	0.23	0.932
Gender agreement of adnominal adjectives	0.015	0	0.4	0.665	0	0.19	0.818
Tightness of the link between the past marker and the verb	0.207	0.045	0.32	0.005	0.136	0.71	0.364
Tightness of the link between the progressive marker and the verb	0.078	0.07	0.36	0.328	0	0.36	0.465
Tense-aspect systems	0.224	0	0.18	0.001	0.069	0.63	0.886
Negation and tense, aspect and mood marking	0.299	0	0.23	0.665	0	0.26	0.69

Present reference of stative verbs and past perfective reference of dynamic verbs	0.071	0	0.34	0.001	0.068	0.55	0.682
Aspect markers and inchoative meaning	0.224	0.07	0.28	0.149	0.139	0.37	0.349
Aspect change in verb chains	0.101	0.128	0.32	0.298	0.102	0.35	0.385
Ability verb and epistemic possibility	0.224	0.091	0.22	0.328	0.114	0.34	0.5
Marking of patient noun phrases	0.066	0	0.38	0.149	0	0.37	0.909
Noun phrase conjunction and comitative	1	0.093	0.17	1	0.046	0.17	0.349
Predicative noun phrases	0.066	0.045	0.29	0.043	0.068	0.41	0.523
Predicative adjectives	0.042	0	0.32	0.043	0.113	0.42	0.432
Predicative locative phrases	0.14	0	0.24	0.179	0	0.31	0.591
Existential verb and transitive possession verb	0.066	0.046	0.33	1	0	0.25	0.545
Motion to and motion from	0.14	0	0.26	1	0	0.21	0.568
Directional serial verb constructions with come and go	0.015	0.091	0.55	0.005	0.159	0.58	0.636
'Give' serial verb constructions	0.066	0.023	0.39	0.005	0.137	0.65	0.477
Complements of 'think' and 'want'	0.101	0.114	0.34	1	0.046	0.24	0.295
Verb doubling in temporal clauses	0.299	0	0.22	1	0	0.13	0.909
Verb doubling and focus	0.071	0.091	0.34	0.001	0.25	0.65	0.568
Vocative markers	0.224	0.024	0.39	1	0.047	0.28	0.395
Paralinguistic usages of clicks	0.14	0.052	0.34	0.665	0.026	0.28	0.474
Pequenino	0	0.181	0.64	0.399	0.045	0.3	0.614
Savvy	0.036	0.091	0.42	1	0	0.12	0.636
Hand and arm	0.015	0.113	0.39	1	0	0.11	0.614
Finger and toe	0	0	0.53	0.149	0	0.36	0.643
Body hair and feather	0.071	0.047	0.43	1	0	0.15	0.581
Hear and smell	0.101	0	0.31	1	0	0.23	0.841
Green and blue	0.207	0	0.37	1	0	0.19	0.886
Syllable onsets	0.101	0.023	0.3	1	0	0.12	0.568
Syllable codas	0.078	0.07	0.34	1	0.047	0.21	0.488
Vowel height distinctions	0.036	0.091	0.35	0.142	0.091	0.36	0.591
The schwa vowel	1	0	0.18	0.665	0	0.27	0.682
Labiodental fricatives	0.224	0.023	0.28	0.018	0.046	0.53	0.636
Palato-alveolar sibilants	0.224	0.046	0.24	0.298	0.069	0.38	0.386

The voiced sibilant z	1	0	0.22	0.149	0.045	0.4	0.682
Interdental fricatives	0.071	0	0.38	1	0	0.26	0.795
The palatal nasal	0.14	0.023	0.33	1	0	0.24	0.432
Prenasalized consonants	0.224	0	0.18	1	0	0.16	0.705

Supplementary Table 3: APiCS features analyzed. Features in bold have a counterpart in WALs, and features not in bold are APiCS-specific features that do not have mixed values. The lfdr column indicates local false discovery rates. The lexifier and substrate rates are the increase in the fraction of correctly predicted language features when information about the ancestry is considered, in relation to the baseline classification rate (the classification that results from assigning every language to the most frequent feature value). Bias-corrected Cramer's V association values are reported for both lexifier and substrate. Variables in yellow are the ones for which ancestry dependency could not be established using an anticonservative test (45/94), and variables in red (37/94) are the ones for which the conservative test yield an association.

Rule length	Dataset	Values	Criterion	Mean recall	SD recall	Mean precision	SD precision
1	Complete	Empirical	$F_{0.5}$	0.77	0.0038	0.69	0.0157
			F_1	0.78	0.0426	0.69	0.0319
			F_2	0.94	0	0.58	0.0095
		Permutations	$F_{0.5}$	0.52	0.1374	0.41	0.0425
			F_1	0.76	0.0901	0.36	0.0314
			F_2	0.83	0.0454	0.33	0.0212
	Reduced	Empirical	$F_{0.5}$	0.86	0.043	0.58	0.0197
			F_1	0.89	0.0325	0.57	0.0232
			F_2	0.91	0.0094	0.56	0.0239
		Permutations	$F_{0.5}$	0.52	0.1374	0.41	0.0425
			F_1	0.76	0.0901	0.36	0.0314
			F_2	0.83	0.0454	0.33	0.0212
2	Complete	Empirical	$F_{0.5}$	0.73	0.0971	0.9	0.0597
			F_1	0.86	0.0243	0.83	0.0248
			F_2	0.88	0.02	0.8	0.0585
		Permutations	$F_{0.5}$	0.45	0.0931	0.47	0.0532
			F_1	0.61	0.0735	0.41	0.0448
			F_2	0.67	0.0453	0.37	0.0309
	Reduced	Empirical	$F_{0.5}$	0.83	0.0456	0.8	0.0266
			F_1	0.85	0.0214	0.79	0.0316
			F_2	0.9	0.0236	0.7	0.0557
		Permutations	$F_{0.5}$	0.45	0.0931	0.47	0.0532
			F_1	0.61	0.0735	0.41	0.0448
			F_2	0.67	0.0453	0.37	0.0309
3	Complete	Empirical	$F_{0.5}$	0.79	0.0756	0.93	0.0286
			F_1	0.85	0.0241	0.9	0.0271
			F_2	0.87	0.01	0.85	0.0452
		Permutations	$F_{0.5}$	0.42	0.0686	0.5	0.0607

	Reduced	Empirical	F_1	0.51	0.0587	0.45	0.0555
			F_2	0.54	0.0411	0.4	0.0409
			$F_{0.5}$	0.8	0.0677	0.89	0.0347
		Permutations	F_1	0.86	0.0183	0.86	0.0238
			F_2	0.87	0.0138	0.84	0.0456
			$F_{0.5}$	0.42	0.0686	0.5	0.0607
			F_1	0.51	0.0587	0.45	0.0555
			F_2	0.54	0.0411	0.4	0.0409
4	Complete	Empirical	$F_{0.5}$	0.81	0.0592	0.94	0.0203
			F_1	0.84	0.021	0.93	0.0232
			F_2	0.85	0.0109	0.91	0.0441
		Permutations	$F_{0.5}$	0.39	0.0479	0.51	0.0699
			F_1	0.43	0.0488	0.47	0.0673
			F_2	0.45	0.0391	0.43	0.0543
	Reduced	Empirical	$F_{0.5}$	0.78	0.053	0.93	0.0312
			F_1	0.84	0.0237	0.9	0.02
			F_2	0.84	0.0171	0.88	0.0418
		Permutations	$F_{0.5}$	0.39	0.0479	0.51	0.0699
			F_1	0.43	0.0488	0.47	0.0673
			F_2	0.45	0.0391	0.43	0.0543

Supplementary Table 4: Precision and recall values for the rule-based creole profile model for both datasets (full or reduced), divided between the estimates for the actual data and the mean and standard deviation obtained in the permuted datasets, for each rule length and criterion ($F_{0.5}$, F_1 , F_2).

	Forest	Rule 1	Rule 2	Rule 3	Rule 4
Creoles misclassified as non-creoles	Sri Lanka Portuguese, Korlai, Lingala, Ghanaian Pidgin English, Berbice Dutch, Angolar, Tayo	Korlai (.75), Lingala (.62), Sri Lanka Portuguese (.66)	Angolar (.93), Korlai (.75), Lingala (1), Sri Lanka Portuguese (.78), Tayo (.69)	Angolar (.92), Fa d'Ambo (.69), Korlai (1), Lingala (1), Sri Lanka Portuguese (.95), Tayo (.92),	Angolar (.95), Fa d'Ambo (.86), Korlai (1), Kriol (.71), Lingala (1), Sri Lanka Portuguese (1), Tayo (.95)
Non-creoles misclassified as creoles	Arapesh, Bagirmi, Chamorro, English, Hmong Njua, Iraqw, Malagasy, Maung, Persian, Spanish, Tukang Besi, Wari, Yoruba	Arabic (.62), Bagirmi (.71), Berber (.64), Basque (.62), English (1), Greek (.85), Guarani (.62), Hmong Njua (.55), Iraqw (.74), Kunama (.55), Krongo (.58), Malagasy (.75), Mapudungun (.64), Maung (.54), Maybrat (.52), Mixtec (.59), Oromo (.77), Piraha (.62), Persian (1), Quechua (.62), Spanish (.85), Wari (.56), Wichi (.59), Yoruba (1)	Arabic (.56), Bagirmi (.51), English (1), Greek (.95), Malagasy (.5), Persian (.75), Spanish (.95), Yoruba (.75)	Basque (.52), English (1), Greek (.89), Quechua (.52), Spanish (.9),	English (1), Greek (.85), Spanish (.87)

Supplementary Table 5. Misclassified languages for the rule-based and probabilistic profiles based on the reduced dataset. Numbers between brackets indicate the fraction of all instances in which the language is misclassified. For the rule-based profile, only languages misclassified on the majority of imputations were included.

Feature	Creoles	Lexifiers	Substrates
Order of subject, verb and object	Overwhelmingly SVO	Germanic and Romance SVO	Macro-Sudan diverse, Austronesian diverse
Order of adposition and NP	Overwhelmingly prepositional	Germanic and Romance prepositional	Macro-Sudan diverse, Austronesian prepositional
Order of RC and noun	Overwhelmingly NRC	Germanic and Romance NRC	Macro-Sudan and Austronesian NRC
Gender distinction in indep. pronouns	Overwhelmingly absent	Germanic and Romance mostly present in 3SG	Macro-Sudan mostly absent, Austronesian overwhelmingly absent
Incl./excl. distinction in indep. personal pronouns	Overwhelmingly absent	Germanic and Romance absent	Macro-Sudan overwhelmingly absent, Austronesian present
Zero copula for predicative nominals	Diverse	Germanic and Romance impossible	Macro-Sudan and Austronesian diverse
Ordinal numbers	Diverse	Germanic and Romance diverse	Macro-Sudan first, two-th, three-th, Austronesian mixed
Numeral classifiers	Overwhelmingly no classifiers	Germanic and Romance no classifiers	Macro-Sudan no classifiers, Austronesian optional or obligatory
Comitatives and instrumentals	Overwhelmingly identity	Germanic and Romance identity	Macro-Sudan diverse, Austronesian different
Reciprocal constructions	Diverse	Germanic diverse, Romance mixed	Macro-Sudan diverse, Austronesian distinct from reflexive
Want complement subjects	Overwhelmingly subject is left implicit	Germanic and Romance subject is left implicit	Macro-Sudan and Austronesian diverse
Negative morphemes	Overwhelmingly negative particle	Germanic and Romance negative particle	Macro-Sudan and Austronesian diverse
Predicative possession	Overwhelmingly have-possessive	Germanic and Romance have-possessive	Macro-Sudan and Austronesian diverse

Supplementary Table 6: Properties of the distribution of feature values for some features for which the data do not allow to conclude there is an observable dependency between them and the ancestry a creole is related to. As it can be observed, in many of these cases the feature value in creoles is either very frequent and reflected in their ancestry, or diverse reflecting diversity in the ancestry as well.

Supplementary figures

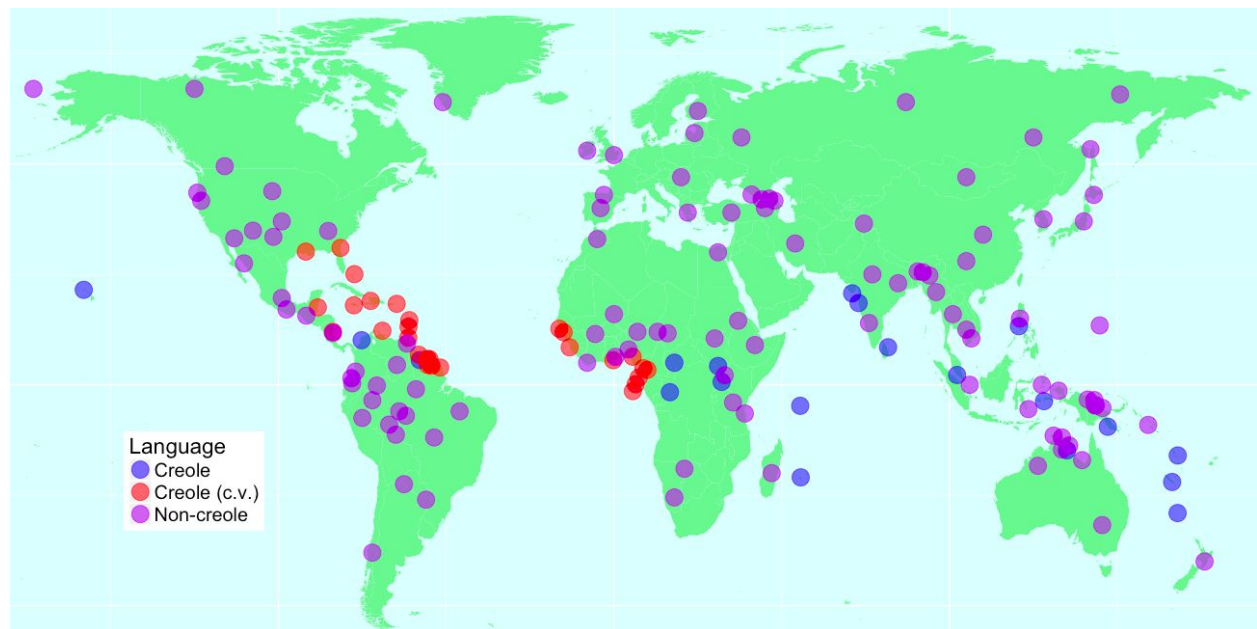
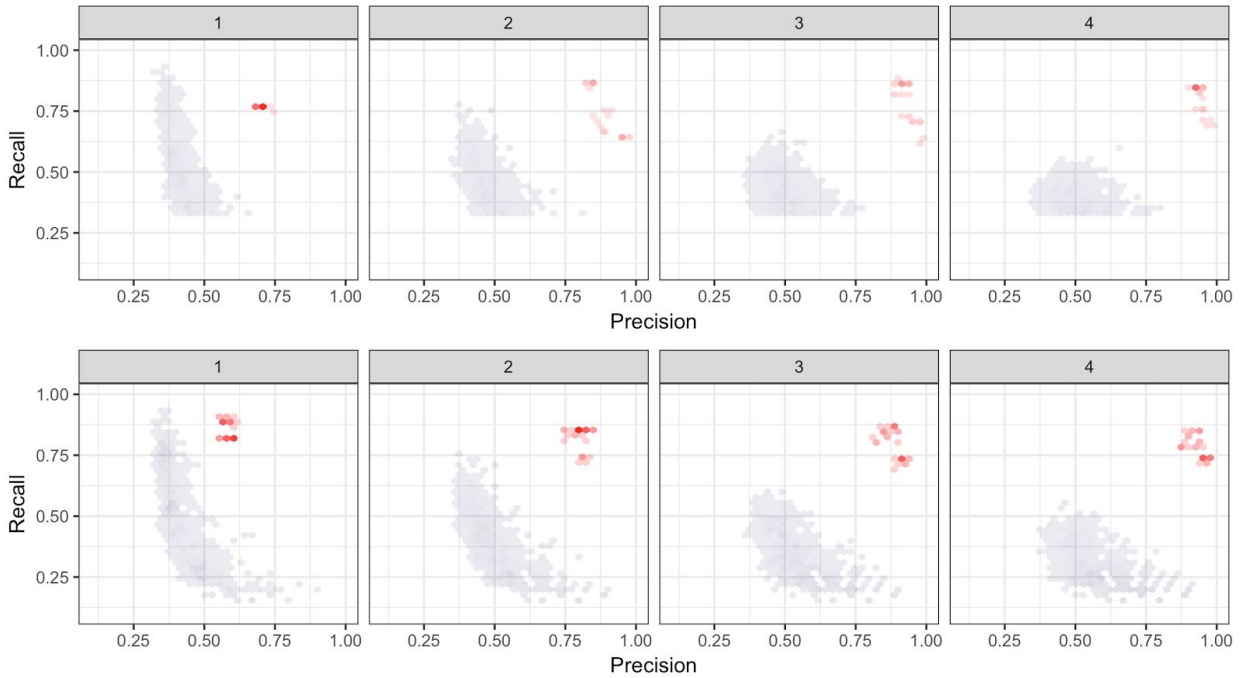
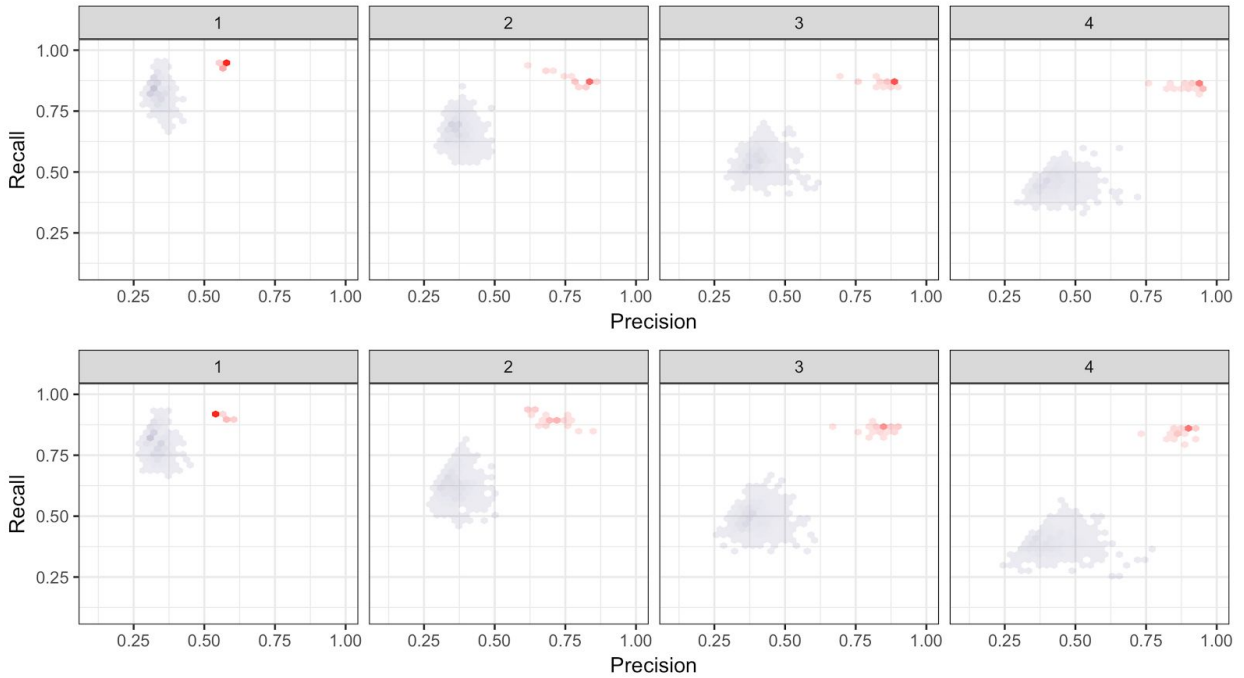


Figure 1: Geographic distribution of the languages used in the study. Creoles of the common variety (“c.v.”) are those with Western European lexifiers and Macro-Sudan substrates.



Supplementary Figure 1: Classification under the rule-based creole profile for the full (above) and reduced (below) datasets, with rules chosen by best $F_{0.5}$ score. Distributions of the empirical (red) and randomized (blue) precision and recall for rule lengths involving from 1 to 4 variables. For rules of length larger than 1, the majority of the rules obtained in the empirical data have better precision and recall values than the randomized controls, which supports the notion that creoles can be distinguished from non-creole languages.



Supplementary Figure 2: Classification under the rule-based creole profile for the full (above) and reduced (below) datasets, with rules chosen by best F_2 score. Distributions of the empirical (red) and randomized (blue) precision and recall for rule lengths involving from 1 to 4 variables. For rules of length larger than 1, the majority of the rules obtained in the empirical data have better precision and recall values than the randomized controls, which supports the notion that creoles can be distinguished from non-creole languages.