

Phylogenetic magic in Creolistics: Language acquisition, feature values and family values

Michel DeGraff, Trevor Bass & Robert Berwick



Formal approaches to Creole studies III — November 15–16, 2012 – Universidade de Lisboa

Goals and outline

A three-step argument against recent instances of Creole Exceptionalism:

1. Creole languages do not, and cannot, reflect “early interlanguages” as defined in Plag (2008, 2009)
2. So far, there’s no “irrefutable” or “staggering” evidence that “Creoles are typologically distinct from non-Creoles
(e.g., the data and computational phylogenetic methods in Bakker *et al* 2011 and Daval-Markussen & Bakker 2012 do not reconstruct Creole formation).
3. Once the Comparative Method is adequately applied, Creole formation appears indistinguishable from language change.

Family values?

1. Atlantic Creoles as descendants of their European superstrate languages (e.g., Greenfield 1830, Hall 1958; Mufwene 2008; DeGraff 2009; etc.).
2. Creoles as “orphans”? (Taylor 1956; Thomason & Kaufman 1988; cf. Adam 1883).
3. Creoles as “simple” languages due to “pidgin” origins (Baissac 1880, Adam 1883, Taylor 1956, Hall 1958, McWhorter 1998ff, Parkvall 2008, **Plag 2008ff, Bakker *et al* 2011, Daval-Markussen & Bakker 2012**, etc.)

Goals and outline

A three-step argument against recent instances of Creole Exceptionalism:

- 1. Creole languages do not, and cannot, reflect “early interlanguage” as defined in Plag (2008, 2009)**
2. So far, there’s no “irrefutable” or “staggering” evidence that “Creoles are typologically distinct from non-Creoles
(e.g., the computational phylogenetic methods in Bakker *et al* 2011 and Daval-Markussen & Bakker 2012 do not reconstruct Creole formation).
3. Once the Comparative Method is adequately applied, Creole formation appears indistinguishable from language change.

Are Creoles “conventionalized interlanguages of an early stage”?

1. **The Early Interlanguage Hypothesis (EIH):** Creole formation mostly depends on processing procedures that are attested in the *earliest* stages of L2A—and not on those that are attested in the advanced stages (Plag 2008a,b, 2009a,b).
2. **Corollary:** Creoles are exceptional, with origins from early L2A interlanguages in a fossilized state of ‘arrested development’ (cf. Hall 1962:152).

**Two processing procedures
that are absent in *early* interlanguages
and, thus, in Creoles (Plag 2008a,b)**

1. Intra- and inter-phrasal information exchange: at t_3 and t_5 , respectively
2. Clausal embedding: the “S’-procedure”, at t_5

L2A sequence: $t_1 < t_2 < t_3 < t_4 < t_5$

Our claim: *These “advanced” processing procedures exist in all natural languages, including Creoles—necessarily so (i.e., as part of C_{HL}).*

Lack of information exchange in early L2A (Pienemann's 2005 Processability Theory)

1. "... the beginning learner is unable to produce any structures which rely on the exchange of specific L2 grammatical information using syntactic procedures, or *in LFG terms the 'unification' of lexical features...*"
2. "[In] the NP *a child* ... [t]he lemma CHILD is marked 'singular', and the value of this diacritic feature has to match that of the determiner."
3. "agreement between heads of different phrases as in subject-verb agreement ... involves the exchange of grammatical information between phrases"

Pienemann's (2008) exchange levels

Information Exchange				
	Stage	Loci of Exchange	Example	Illustration
↑	Sentence	Within sentence	Peter sees a dog	<p>Diagram illustrating the hierarchical structure of the sentence "Peter sees a dog". The root node is S, which branches into NPs and V. NPs branches into N (Peter) and V branches into V (sees) and NP (a dog). Arrows indicate the flow of information exchange from the lower-level nodes up to the higher-level nodes.</p> <p>[3rd pers sg] [3rd pers sg, pres, non-cont.]</p>
	Phrase	Within phrase only	Two kids	<p>Diagram illustrating the hierarchical structure of the phrase "Two kids". The root node is NP, which branches into Det (Two) and N (kids). Arrows indicate the flow of information exchange from the lower-level nodes up to the higher-level node.</p> <p>Det [pl] N [pl]</p>
	Category	No exchange	Talk-ed	<p>V [past]</p>

Information exchange vs. inflectional morphology

1. Information exchange is a ***necessary condition*** for the appearance of syntactically-determined inflectional morphology (= “contextual inflection”).
2. But information exchange is **not a *sufficient condition*** for contextually-determined inflection.
3. Booij 2005 includes examples of **information exchange in syntax *without* morphology.**
 - a) ***het soepvlees*** “the soup meat” (neuter) vs.
 - b) ***de vleessoep*** “the meat soup” (non-neuter)

Syntax vs. morphology in Pienemann's (2005) Processability Theory [PT]

1. “the PT hierarchy is based on information exchange ***to the exclusion of phonological, derivational or specific morphophonological processes.***”
2. “LFG is able to represent abstract grammatical information such as ‘SUBJECT of’ or ‘subject PERSON’ at the level of f-structure ***regardless of whether that information is expressed through morphology or syntax.***”

Agreement *sans* inflection in Haitian Creole (HC) and Saramaccan

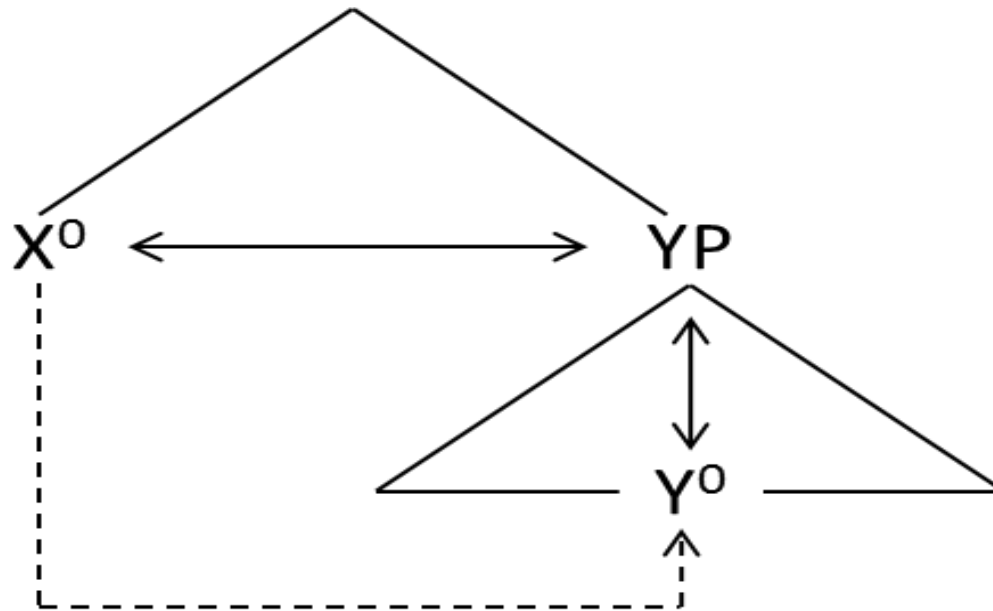
1. *senk liv yo /*la* (HC)
five books DET.PL / DET.SG
'the **five** book*(s)'

2. a) *di man kodo womi* (Saramaccan)
the.SG one single man
'the single/unique man'

- b) *dee dii womi*
the.PL three man
'the/these three men'

(Rountree & Glock 1992; Aboh, to appear)

Selectional restriction as information exchange



--> : selectional restriction
↔ : information exchange

Selectional restriction in Haitian Creole: V^0 selects P^0

1. *Mwen* [_{VP} ***konte*** [_{PP} *(*sou*) *wou*]]
I count on you
'I rely on you'

2. * *Mwen konte de / anba / anlè wou*
I count of / below / above you
'I rely on you'

Agreement *plus* clausal embeddings in Saramaccan (Aboh, to appear)

1. ***Di*** *fisi* [_{CP} ***di*** *mi tata kisi*] *bigi*
the.SG fish that.SG my father caught big
'the fish that my father caught is big'

2. ***Dee*** *fisi* [_{CP} ***dee*** *mi tata kisi*] *bigi*
the.PL fish that.PL my father caught big
'the fish that my father caught are big'

(Rountree & Glock 1992; Aboh, to appear)

Selectional restriction plus clausal embeddings in HC = *long-distance* information exchange

1. *Jan kwè [CP Mari te di [CP li te konte sou Woje]]*
 John believe Mary ANT say 3sg ANT count on Roger
 ‘John believes that Mary had said that (s)he counted on Roger’

2. *Sou ki moun [CP Jan kwè [CP Mari te di [CP li te konte t]]]?*
 On which person John believe Mary ANT say 3sg ANT count
 ‘On whom does John believe that Mary had said that (s)he counted?’

3. **Sou ki moun [CP Jan kwè [CP Mari te chichote [CP li te konte t]]]?*
 ‘On whom did John believe that Mary had whispered that (s)he counted?’

Clausal embeddings in Creoles across time and across space

1671 Mermaid text from Martinique (Hazaël-Massieux 2008)

mouchié faire [_{TP} *yon autre negro* *courir après li* [_{CP} *pour prendre li avec ligne*]]
man make a other negro run after 3sg for take 3sg with line
'The man made another negro run after it in order to catch it with a line'

Also consider: *Wh*-movement, focus constructions and predicate clefts across unbounded domains in Saramaccan (Aboh 2006), Jamaican Creole (Durrelman 2008), Capeverdean (Alexandre 2012, *Abels & Alexandre 11/16/2012, Duarte et al 11/16/2012*) and various other Creoles (Byrne & Winford 1993, Holm & Patrick 2007, etc.)

<http://www.clul.ul.pt/en/component/content/article/458-facs3>

Displacement and other “non-canonical” mappings entail information exchange

1. “In PT ...the linearisation problem [i.e., the pronunciation of syntactic phrases in non-canonical position] is modelled using ***feature unification that permits the exchange of lexical feature information within and across constituents...***” (Pienemann 2005)

2. Information exchange in non-canonical mappings:

- a) *wh*-questions (***Obata et al 11/15/2012***)
- b) clefts (***Duarte et al 11/16/2012***)
- c) passives (***Rendall et al 11/15/2012***)
- d) causatives (***Luís 11/15/2012***)
- e) ...

Creoles are ... *natural* languages, really!

1. Plag's "Early Interlanguages" lack two *fundamental* properties of natural languages:
 - a) information exchange
 - b) clausal embedding

2. ***Information exchange is one essential ingredient*** for implementing:
 - a) feature unification in LFG, HPSG, etc. (à la Pienemann);
 - b) Probe-Goal relationships in Minimalism (Agree, Feature Valuation, ...).

3. ***Merge is one essential computational building block of natural languages. Merge is recursive.*** (Chomsky 2008)

4. Applications of merge include *wh*-movement as Internal Merge, thus cross-clausal information exchange of unbounded depth (via, e.g., copies of the displaced *wh*-phrase).

5. ***In Creoles as in all natural languages, Merge is recursive (as in, e.g., clausal embeddings with *wh*-movement).***

Goals and outline

A three-step argument against recent instances of Creole Exceptionalism:

1. Creole languages do not, and cannot, reflect “early interlanguage” as defined in Plag (2008, 2009)
- 2. So far, there’s no “irrefutable” or “staggering” evidence that “Creoles are typologically distinct from non-Creoles” (e.g., the computational phylogenetic methods in Bakker *et al* 2011 and Daval-Markussen & Bakker 2012 do not reconstruct Creole formation).**
3. Once the Comparative Method is adequately applied, Creole formation appears indistinguishable from language change.

Are Creoles *exceptional*?

The Exceptional Creole Typology Hypothesis (“CTH”):

1. Parkvall 2008: “Creole simplicity”
2. Bakker, Daval-Markussen, Parkvall & Plag 2011:
“Creoles are typologically distinct from non-Creoles”
3. Daval-Markussen & Bakker 2012:
“Explorations in creole research with phylogenetic tools.”

Is Creole Exceptionalism *exceptionally* convincing?

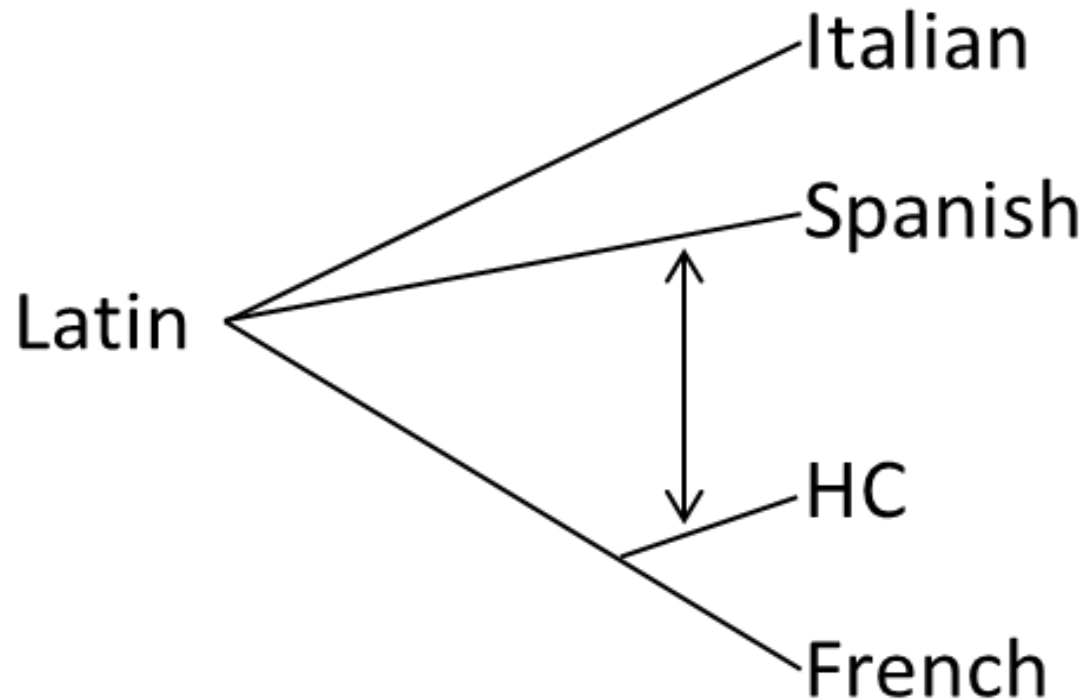
1. “The results of the application of the software to the dataset are quite *staggering*.” (B&al:33)
2. “[Our] results *invariably* cluster all the creoles and pidgins, quite separately from the non-creole languages of the world” (B&al:33).
3. “Creoles as a group *stand out as being less complex* than non-creoles” (B&al:8)
4. “[T]he conclusion that creoles (and pidgins, for that matter) are typologically distinct from the languages of the world is *inescapable and robust*” (B&al:35).
5. Parkvall 2008 and B&al are cited in McWhorter’s (2011:10) search for “a *litmus test for creole status*” in synchronic structural terms.
6. B&al is “the *strongest piece of evidence*” for a Creole typology, evidence that is improved by D-M&B’s “*irrefutable evidence*” (D-M&B:94).

***B&aI* on SplitsTree**

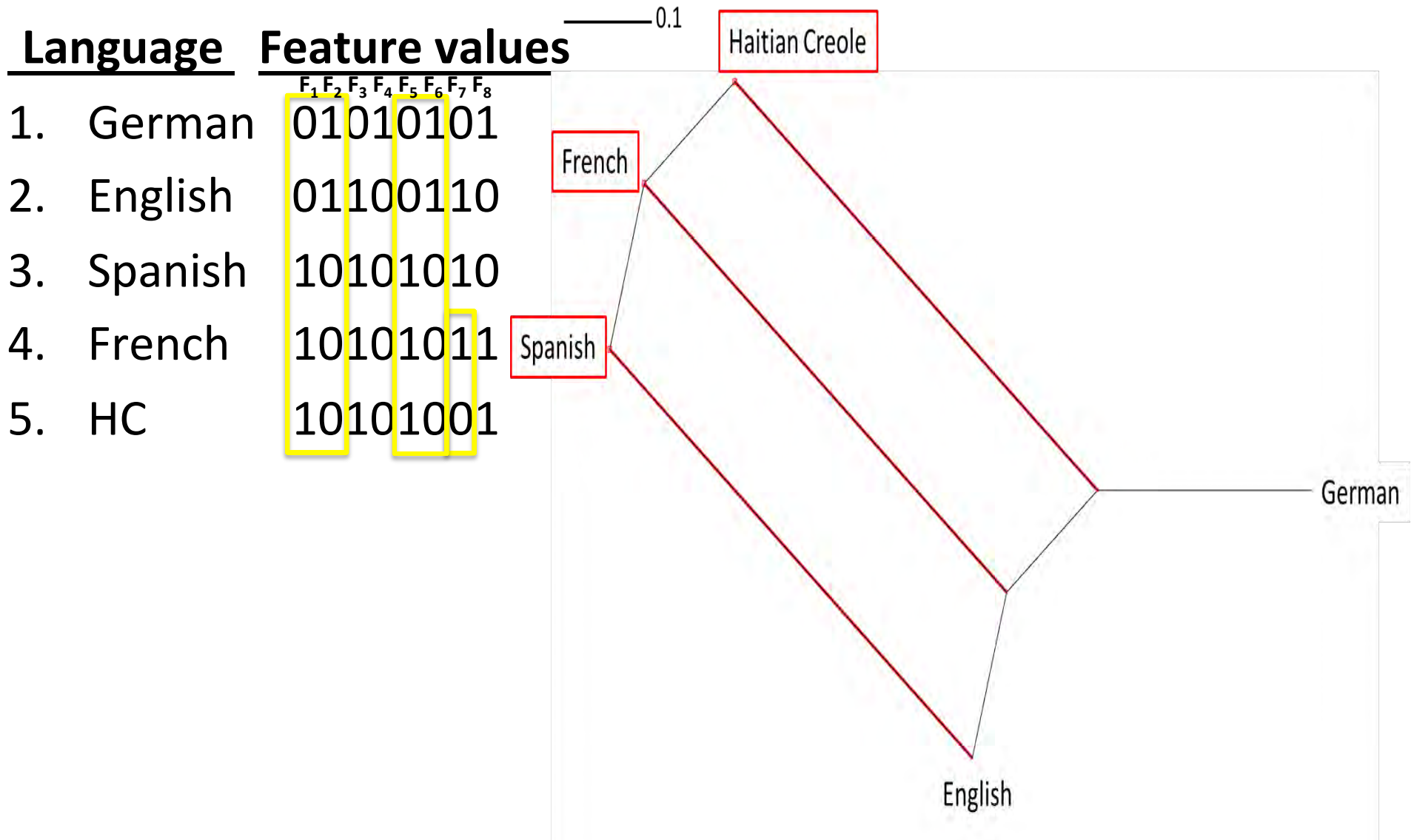
1. “In recent years, a number of algorithms [e.g., SplitsTree; Huson & Bryant 2006] have been developed by bioinformaticians to help visualize biological evolution ...The resulting phylogenetic networks have a number of advantages over the old evolutionary trees. ***First, they can account for horizontal relationships, i.e. contact phenomena***”
2. “[These] algorithms now make it possible to draw trees that show not only inheritance, but also ***horizontal influence (contact, borrowing)***”

Explicit phylogenetic networks

A phylogenetic network for a toy example—
horizontal transfer from Spanish into Haitian Creole:



SplitsTree: An *implicit* phylogenetic networks



The semantics of SplitsTree networks

“[SplitsTree] represents graphically how the input data ... do not fit a tree exactly. **Thus the graph represents a combination of tree-like signal and noise in the data. In particular, the internal nodes of this graph do not represent ancestors of the given languages ...**” (Nichols & Warnow 2008:764f)

Limitations of SplitsTree

B&al: “[E]ven a cursory look at the clustering of non-creoles shows that *the non-creoles do not classify along genetic or areal lines*. For example, Basque (isolate, Western Europe), Hindi (Indo-European, India/South Asia), Burushaski (isolate, North Pakistan), and Hunzib (East-Caucasian, Caucasus) cluster ...”

The basic logic in *B&al* and D-M&B

1. Select a “reasonable” set $C_0 \subset C$ of languages in C (in *B&al*, $C = \text{“Creoles”}$).
2. Select a “reasonable” set $N_0 \subset N(C)$ of languages $N(C)$ not in C .
3. Select a “reasonable” set $F_0 \subset F$ of available features F .
4. Use *SplitsTree* to show that the languages in C_0 are, as a set, closer to each other by normalized Hamming distance than to the languages in N_0 .
This causes SplitsTree to separate them via splits (“Split decomposition”).

Showing “exceptionality”?

1. A class of languages C is argued to be “exceptional” if there exist “reasonable” subsets $F_0 \subset F$, $C_0 \subset C$, and $N_0 \subset N(C)$ such that the languages in C_0 are closer to each other than to the languages in N_0 using normalized Hamming distance with respect to F_0 .
2. NB: B&al *generally* defines “reasonable” for the selected sets of languages as “diverse” rather than as “representative” or “sufficiently large.”

A “reasonable and balanced” sample?

B&al: “The creoles selected had seven different lexifiers: Arabic (Nubi), Assamese (Nagamese), Dutch (Negerhollands, Berbice Dutch), English (Jamaican, Krio, Ndyuka, Tok Pisin), French (Dominican, Haitian, Seychellois), Portuguese (Angolar, Cape Verdean, Guinea Bissau Creole, Korlai Creole), Spanish (Palenquero, Zamboangueño). Two of them show noteworthy admixture from African languages (Angolar: Portuguese/Bantu; Berbice Dutch: Dutch/Ijo) and one has been considered both or either Portuguese and/or Spanish (Papiamentu).

This can be considered a reasonable and balanced spread over lexifiers.”

The basic flaw in B&al's logic (i.e., bias)

Are the members of the set C of 15-year-old males significantly taller, as a class, than males of other ages? (F_0 is height)

1. **Select a set $C_0 \subset C$ of 15-year-old males.** Make sure that this set is “reasonable and balanced” in the sense that it is diverse with respect to many factors: ethnicity, nationality, hair color, eyesight, age of parents at birth...
2. **Select a set $N_0 \subset N(C)$ of males who are not 15 years old.** Make sure that this set is similarly diverse, adding age as a diversity factor.
3. ***If the 15-year-olds we select happen to all be young basketball stars and the factors we use to justify diversity do not allow us to identify this very important fact, the answer is likely to be ‘yes’ (erroneously so)*** ³⁰

How to avoid bias?

Dunn *et al*'s (2008) guidelines:

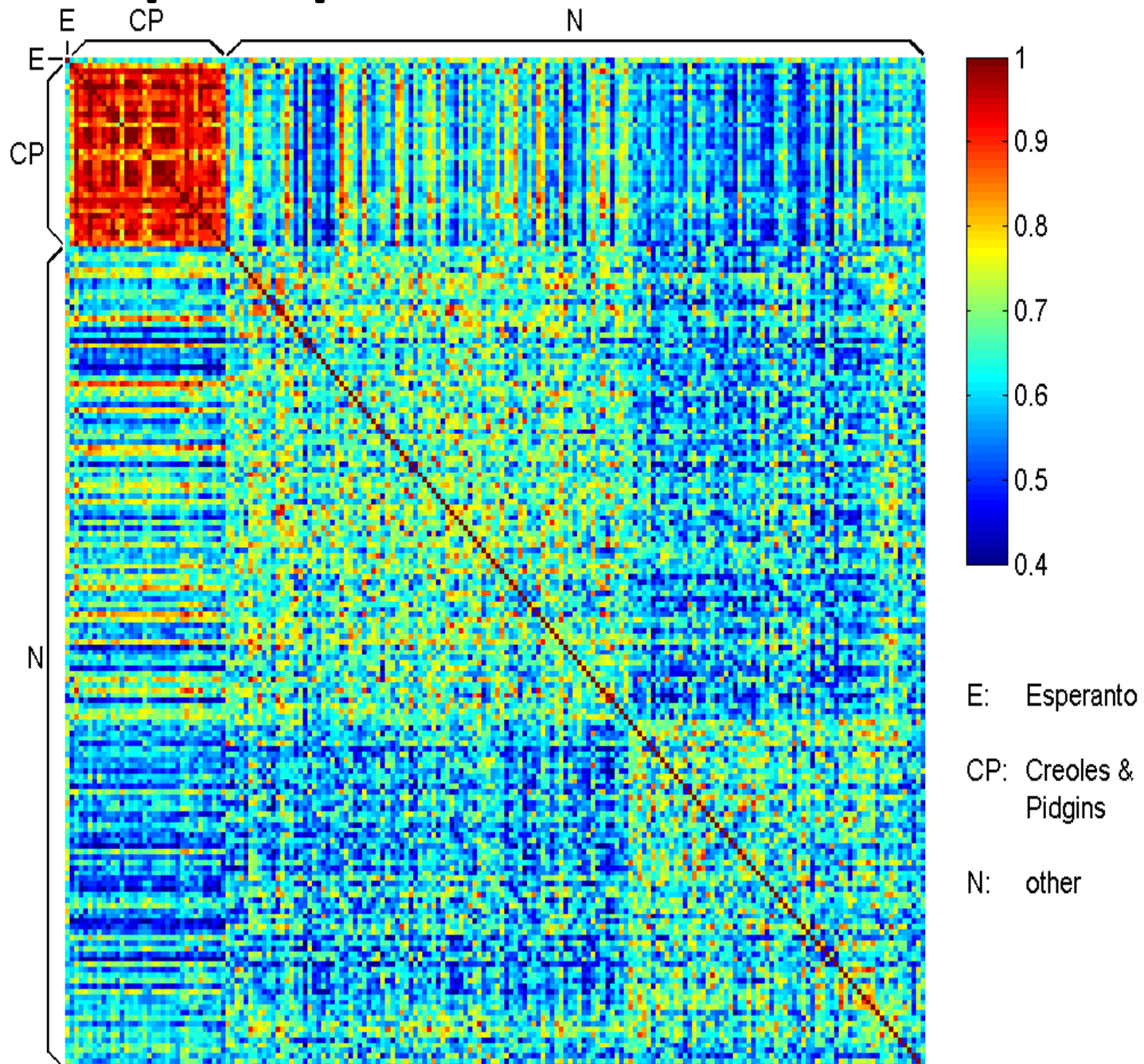
- 1. Investigate “as many abstract structural features from as many parts of the grammar as possible”**
- 2. Provide “a large body of basic features for each language, which together give a broad typological profile, regardless of whether any given feature seems typologically significant.**
- 3. “[A]void the charge of ‘hand-picking’ features by including in [the] sample the widest feasible range of noninterdependent typological phenomena”.**

Compare with Bakker & *al* 2011 (“B&*al*”)

Take the study that is based on Parkvall 2008:

1. **43 features almost exclusively from morpho-syntax**, for the comparison of 188 languages, 34 of which are Pidgins and Creoles.
2. Of the 8,084 possible feature values defined by this language/feature matrix, 1,743 are marked “?” for **missing data, nearly 22% of the total**
3. Contrast: Dunn *et al* 2008 used 115 features for their sample of 22 Papuan languages, with features ranging over phonology, morphology and syntax

Heatmap for pairwise similarities in B&al



Logically inter-dependent features in B&al

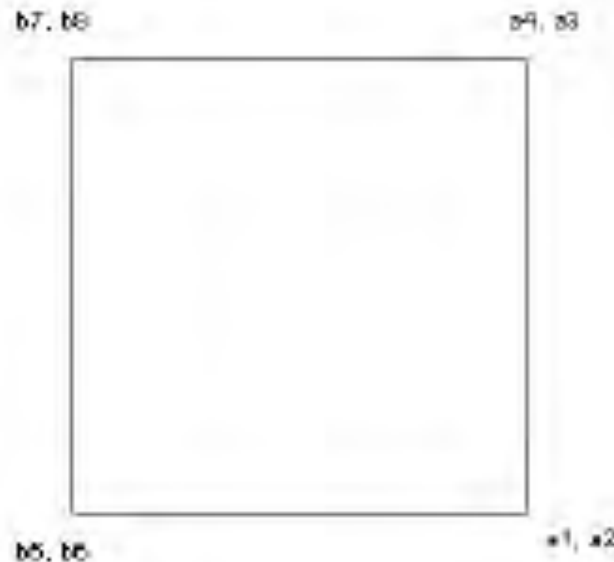
1. F06/F07 (“overt marking of direct object”/“double marking of direct object”) for WALS 23 (“locus of marking the clause”). F07 → F06
2. F08/F09 (“Possession by double marking”/“Overt possession marking”) for WALS 24 “Locus of Marking in Possessive Noun Phrases”). F08 → F09
3. F11/F13 (“Gender”/“Non-semantic gender assignment”) for WALS 30–32 (“Number of Genders”/“Sex-based and Non-sex-based Gender Systems”/“Systems of Gender Assignment”). F13 → F11
4. F23/F24 (“Ordinals exist as a separate class beyond ‘first’”/“Suppletive ordinals beyond ‘first’”) for WALS 53 (“Ordinal Numerals”). F24 → F23
5. F30/F31 (“Grammaticalized past/non-past” /“Remoteness distinctions of past”) for WALS 66 (“The Past Tense”). F31 → F30
6. F36/F37 (“Evidentiality (grammatical)”/“Both indirect and direct evidentials”) for WALS 78/77 (“Coding of Evidentiality”/“Semantic Distinctions of Evidentiality”). F37 → F36

Logical contradictions in B&aI

1. Consider the interdependent F06 and F07: It is logically impossible for any language to simultaneously exhibit “double marking of direct object” and no “overt marking of direct object.” $F07 \rightarrow F06$ means that $F07 = 1$ and $F06 = 0$ is a logical contradiction.
2. Yet, this is what we find in B&aI for seven languages: Amele, Chukchi, Ewe, Georgia, Goonyandi, Diola-Fogny and Slave.
3. Similar contradictions obtain almost every other combination of inter-dependent features. Out of the 188 languages in the Parkvall-based dataset used in B&aI, 152 have at least one inconsistent pair, for a total of 215 inconsistent pairs

The problem of missing data: SplitsTree's interpretation of “?”

Data	Missing	SplitsTree4 output	Clusters
a1: 11111111			
a2: 11111111			
a3: 11110000			{a1, a2}
a4: 11110000			{a3, a4}
b5: 00001111	None		{b5, b6}
b6: 00001111			{b7, b8}
b7: 00000000			
b8: 00000000			



Data	Missing	SplitsTree4 output	Clusters
a1: 1111??11	Scenario 1		{a1, a2, a3, a4}
a2: 111111??			{b5, b6, b7, b8}
a3: 1111?0?0			
a4: 11110?0?			
b5: 0000?11?			
b6: 00001?1?			
b7: 00000???			
b8: 0000?0?0			

Data	Missing	SplitsTree4 output	Clusters
a1: ??111111	Scenario 2		{a1, a2, b5, b6}
a2: ?1?11111			{a3, a4, b7, b8}
a3: 1?1?0000			
a4: 1??10000			
b5: ?00?1111			
b6: ?0?01111			
b7: 0??00000			
b8: ?00?0000			

More problems with missing data

- 1. Any biases in missing data may artificially create or magnify splits.** For example, if there is little overlap between the missing features in Creoles vs. the missing features in non-Creoles, then this could result in spurious splits or in splits of greater magnitude between Creoles and non-Creoles.
- 2. The dataset from Parkvall 2008 has a highly non-uniform distribution of missing data loci.** For example, features F35 (Morphological Optative) and F36 (Evidentiality) are given values for all Creoles and Pidgins but are missing (“?”) in approximately 74% of languages other than Creoles, Pidgins, and Esperanto.

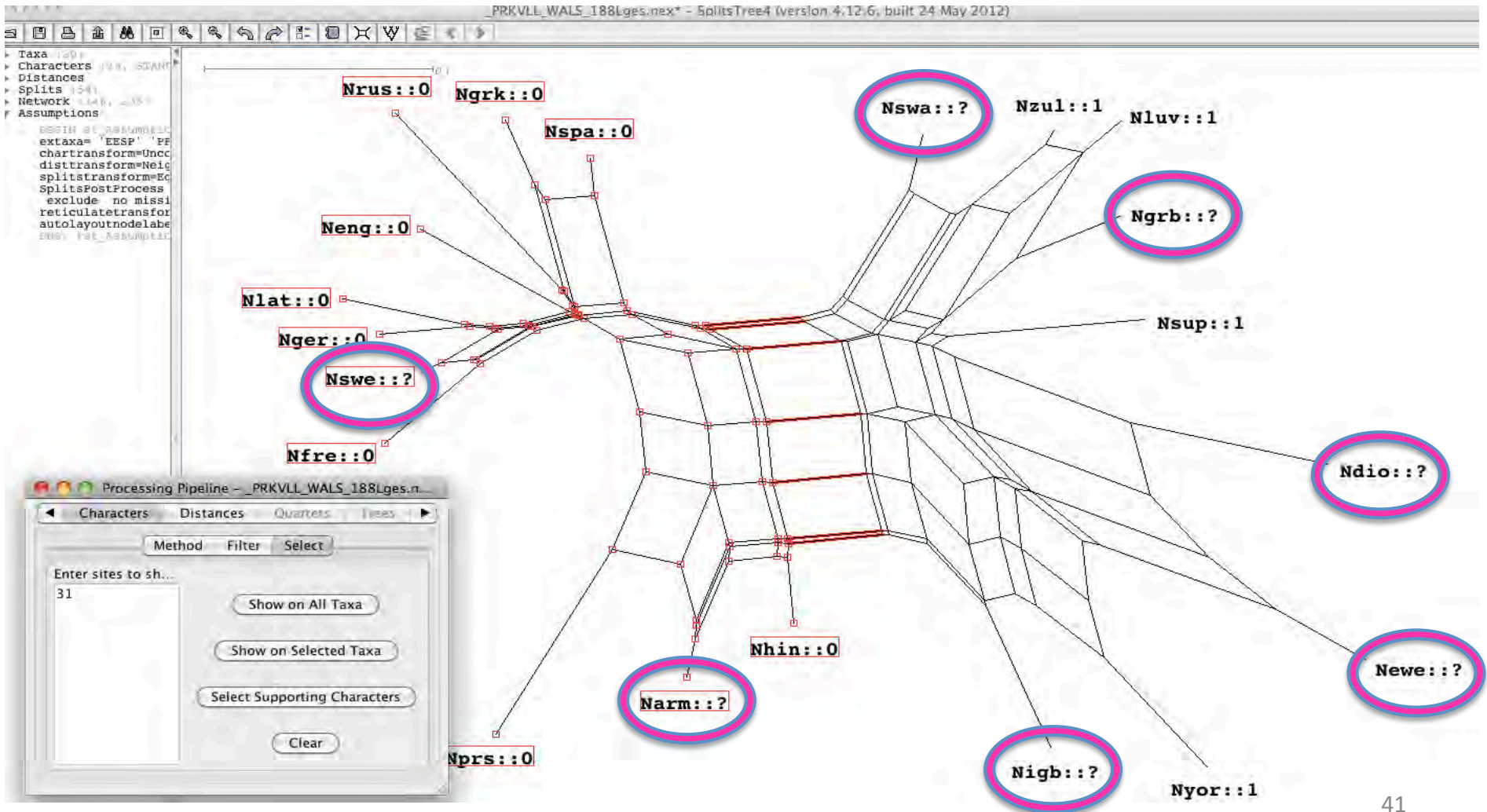
One troubling example of “missing” data

Reduplication (“F10”):

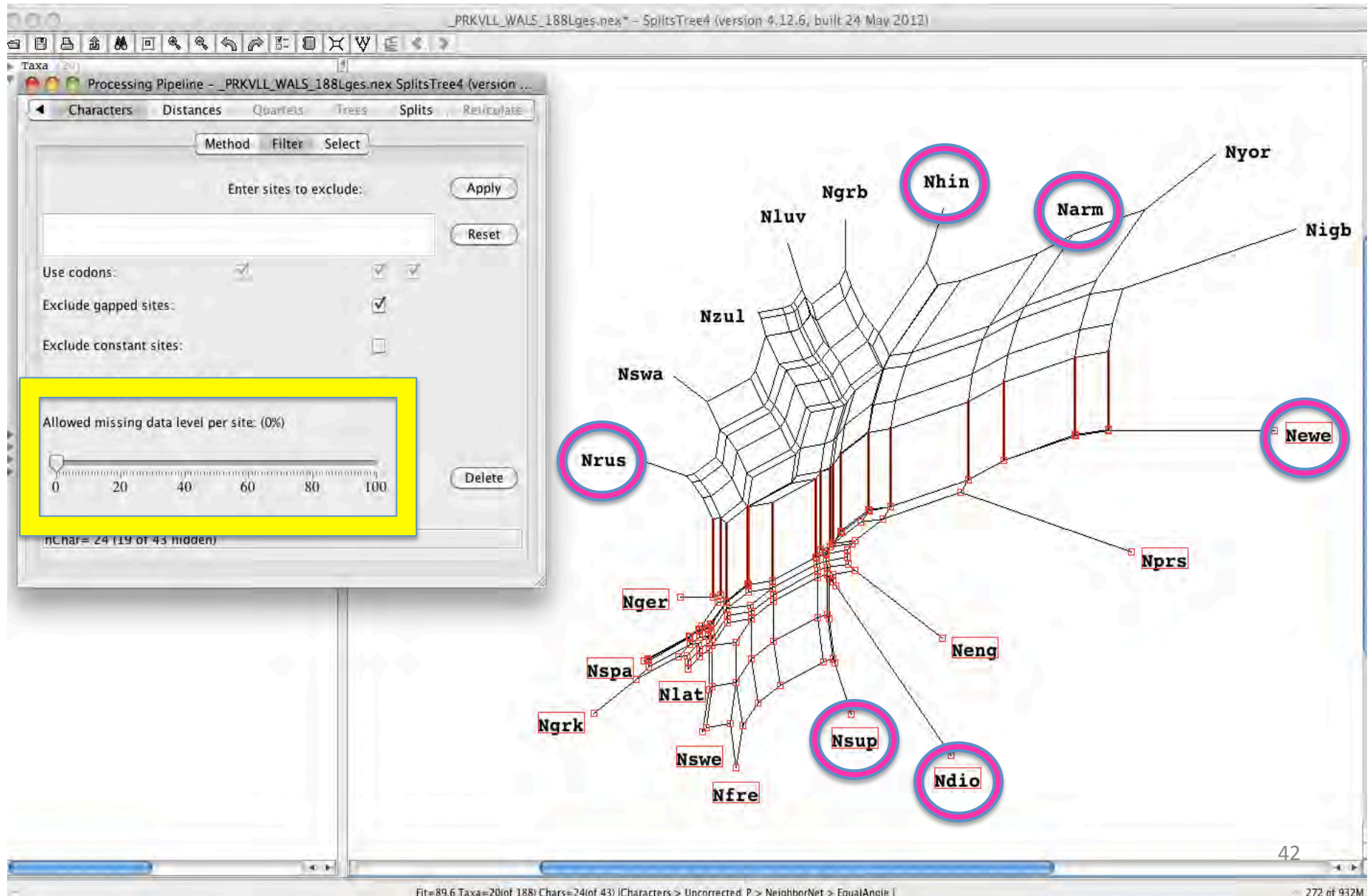
1. F10 = “?” (unknown) for all but two Pidgins and Creoles
2. Yet reduplication is well documented in Creole languages (Kouwenberg 2003; cf. *Luís 11/15/2012* and *Alleesaib 11/15/2012* here at FACS III)
<http://www.clul.ul.pt/en/component/content/article/458-facs3>
3. Bakker & Parkvall 2005: “reduplication [is] *fairly common in Creoles, possibly more so than in languages in general*”
4. Plag 2009: “The pertinent studies show *a wide range of reduplication processes across a wide range of creoles.*”
5. F10 = 1 *for English, French, German and Spanish* even though WALS indicates that such languages have no productive reduplication (ditto in Plag 2009:358)

Splits based on missing data

Split Decomposition network of Indo-European vs. Niger-Congo based on the relevant subset of B&aI's data:



Split Decomposition network of Indo-European and Niger-Congo with “Allowed missing data level per site” = 0%



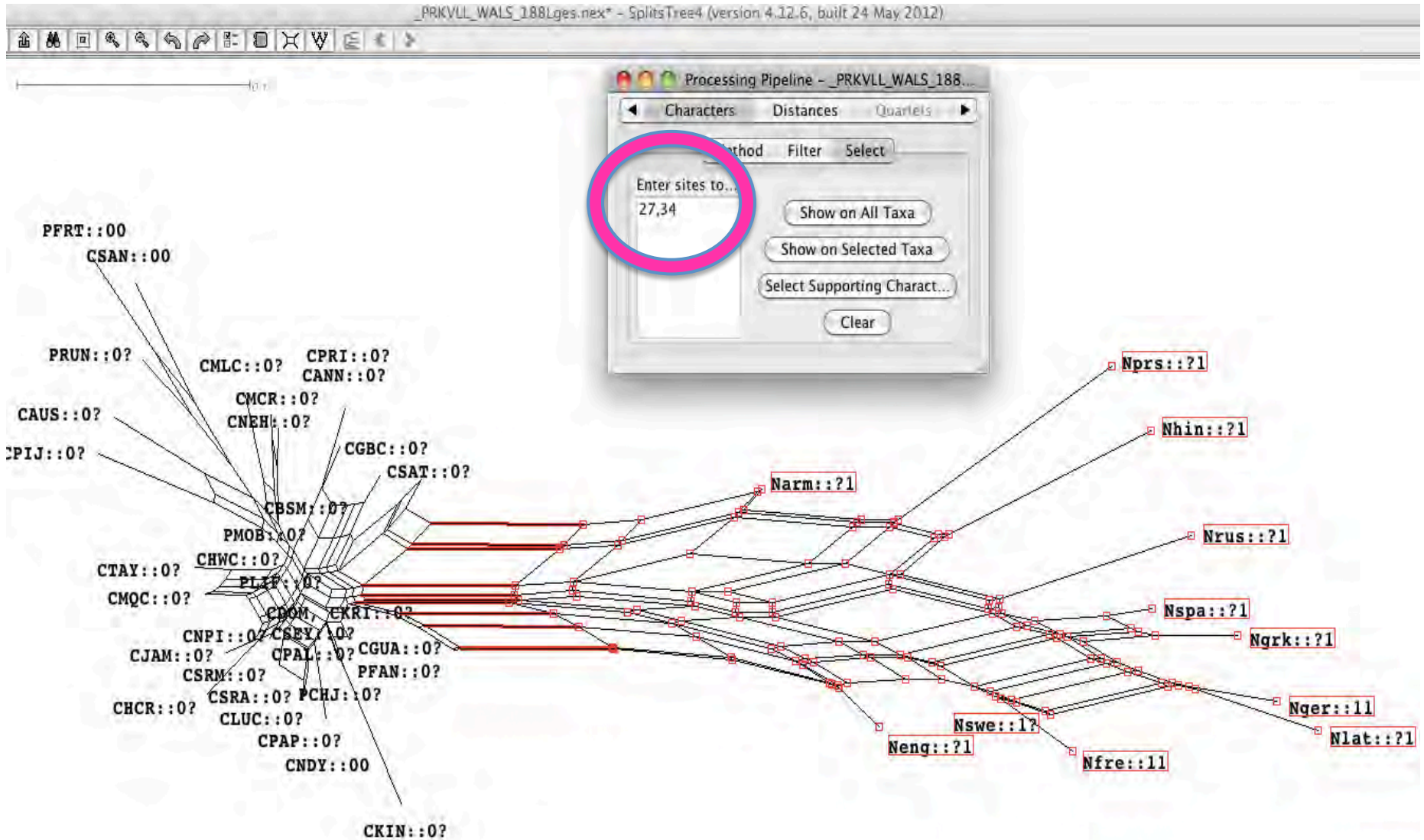
What are the facts about F39 “non-neutral marking of pronouns”?

1. **B&al's claim in their appendix:** English, French, German, Greek, Hindi, Latvian, Persian, Russian and Spanish have no case-marking on pronouns.
2. **But:** English: I vs. *me*; French: *je* vs. *me*, German: *ich* vs. *mich*...
3. The F39 values for the Indo-European languages in B&al are **all contradicted by WALS**.
4. **WALS also contradict all the Niger-Congo values for F39 in B&al, except for Yoruba.**

B&al's online appendix:

<http://aal.au.dk/fileadmin/www.aal.au.dk/lingvistik/afdelingen/bakkeretal2011JPCLappendix.pdf>

Split Decomposition network of Pidgins/Creoles vs. IE based on the relevant subset of B&a's data



Morphological optatives and antipassives as “supporting characters” for the split

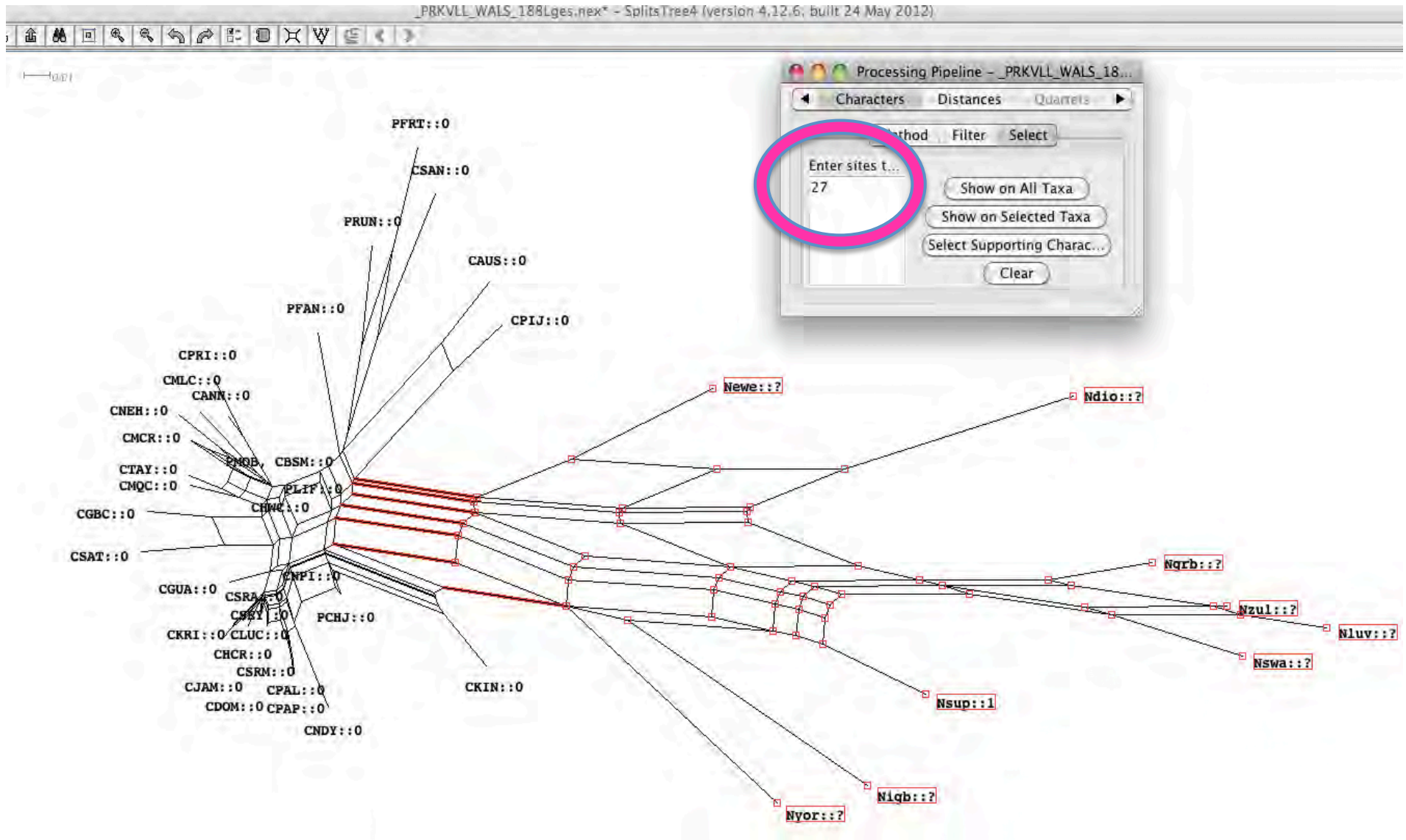
1. “The [WALS] map documents 319 languages. Worldwide, so it seems, **morphological optatives are a fairly infrequent phenomenon** (48 languages in our sample). Our sample does not contain any languages with inflectional optatives from Australia or from Africa; we have just two cases in Europe (Albanian and Karaim), one in New Guinea, and only a few in the Americas and Asia outside of the Caucasus and the area of northern India and Nepal.” (Dobrushina *et al* 2011)
2. **Antipassive** as well is a cross-linguistically rare feature: according to WALS, it is **totally absent in Europe (except for Basque), like it is among Pidgins and Creoles.**

The devil is in the feature matrices of *B&al's appendices*

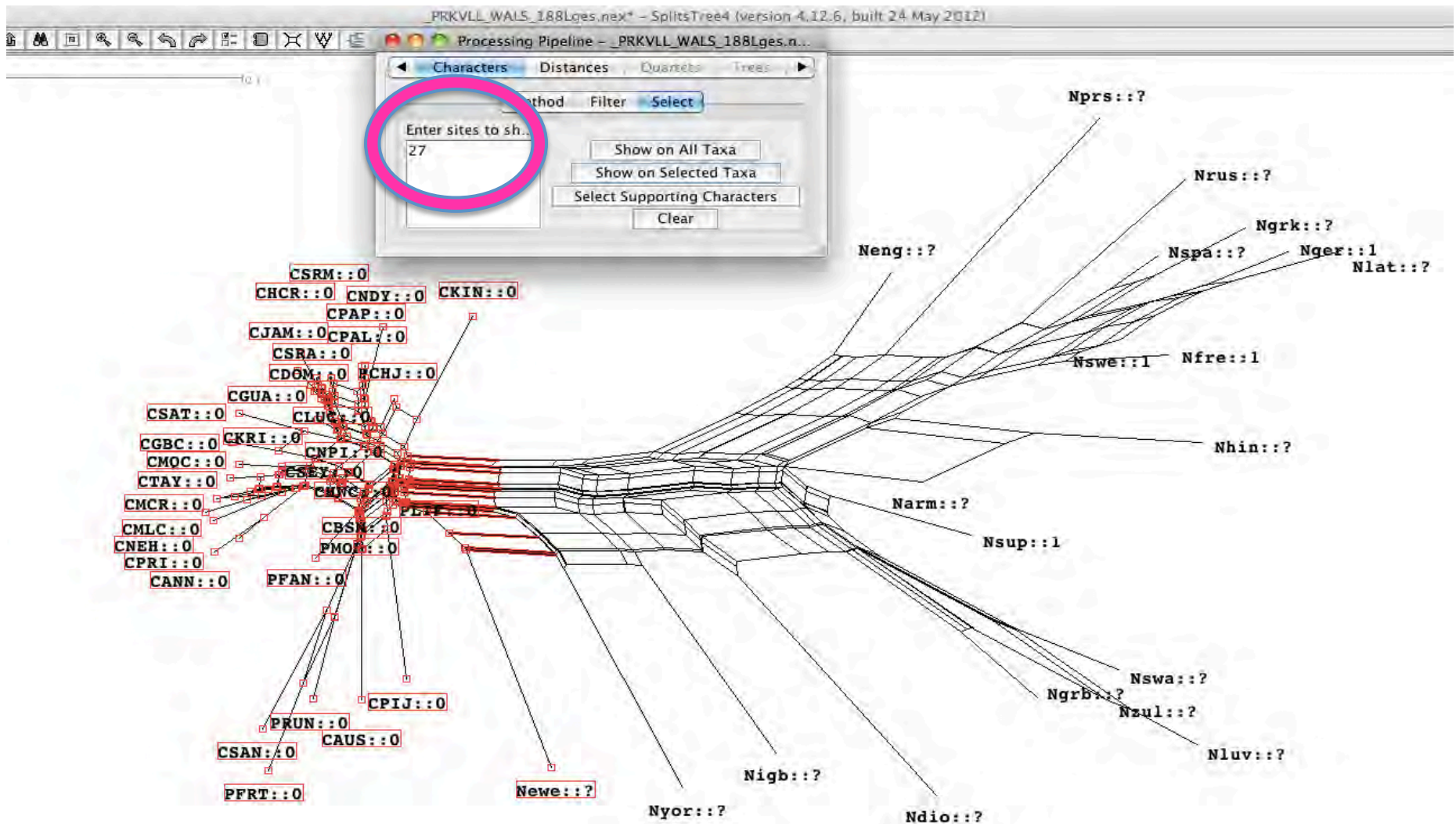
<http://aal.au.dk/fileadmin/www.aal.au.dk/lingvistik/afdelingen/bakkeretal2011JPCLappendix.pdf>

1. *B&al's* distribution of features such as “morphological optative” and “antipassive” contradicts not only the descriptions in *WALS*, but even basic knowledge about well-studied languages.
2. French and German are described as having both morphological optatives and antipassive.
3. Antipassive is claimed to even exist in English, Spanish and Greek.
4. In *WALS*, both features are shown to be largely absent in Europe—except for optatives in Albanian and Karaim.

Split Decomposition network of Pidgins/Creoles vs. Niger-Congo based on the relevant subset of B&a's data



Split Decomposition network of Pidgins/Creoles vs. N-C vs. I-E based on the relevant subset of B&a's data



Missing data as *wild* “wild cards”

Consider the marking “?” for Ewe’s F35 “morphological optatives” (in column 27) in the preceding networks:

1. **Ewe’s F35 is treated as a “0” in the network with Pidgins/Creoles, Niger-Congo and Indo-European, which brings Ewe closer to Pidgins/Creoles in this network**
2. **Ewe’s F35 is treated is as a “1” in the network for Pidgins/Creoles vs. Niger-Congo, which brings Ewe closer to Niger-Congo in this network.**
3. Given the actual distribution of morphological optatives in the relevant languages, all these networks are based on dubious empirical grounds.
4. **Missing data (i.e., about 22% of the data in B&al) introduce similarly substantial noise in the networks produced by SplitsTree.**

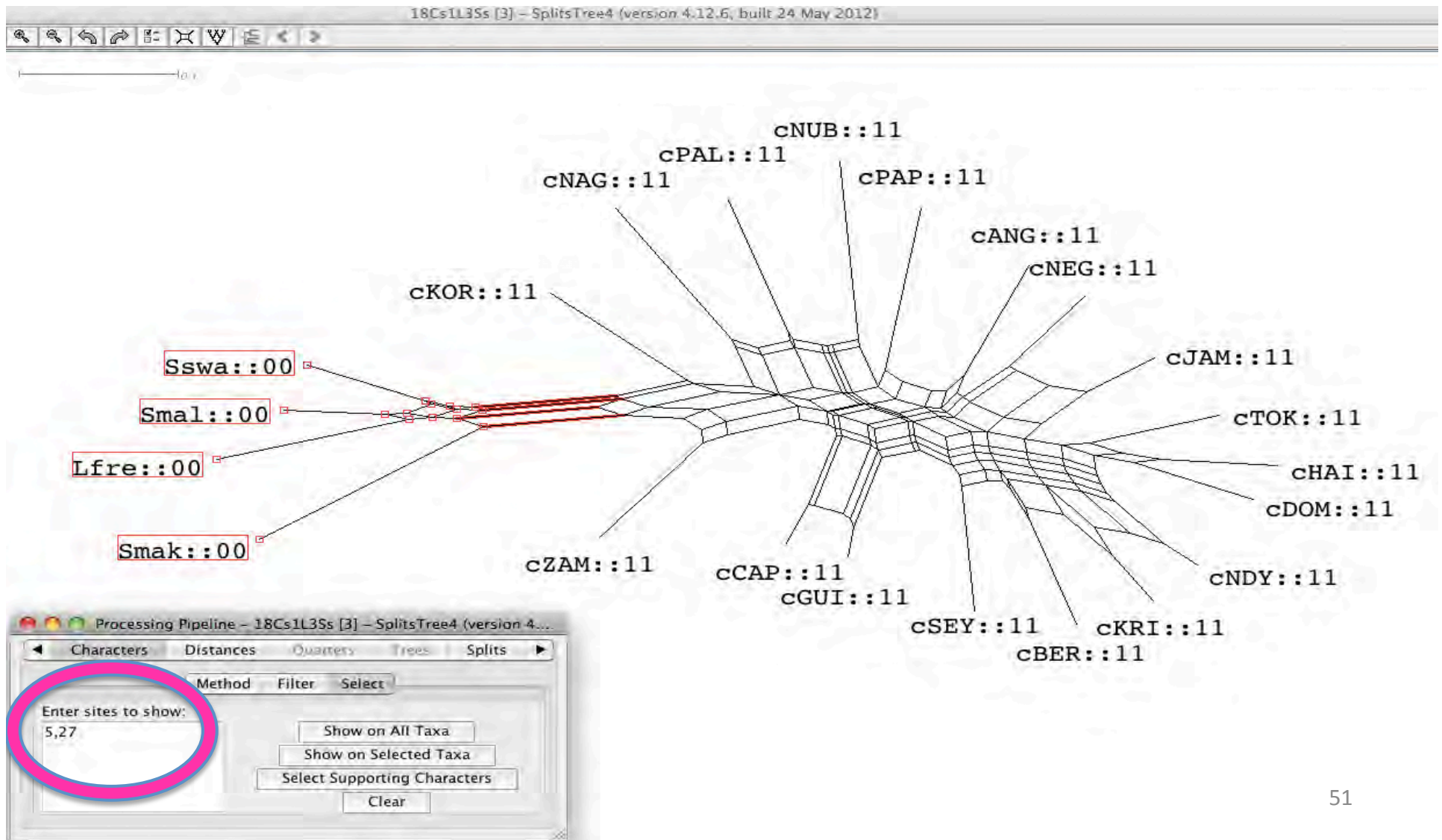
Daval-Markussen & Bakker 2012 (“D-M&B”):

More interdependence, biases contradictions, etc.

1. **Interdependence:** Inherited from Holm & Patrick 2007 and Parkvall 2008
2. **Bias:** “[T]he software was able to detect a clear phylogenetic signal in only a few cases, which in itself is not surprising, since **the features were originally selected as representative of the Atlantic creoles ...**” (D-M&B)
3. **Contradictions in D-M&B: 4 WALS features shared by 80% of Creoles:**
 - a) **Indefinite article same as numeral *one* (18 out of 18)**
 - b) **No tense-aspect inflection (18 out of 18)**
 - c) double negation for Angolar and Palenquero; negative particle for the other 16 Creoles
 - d) predicate possession is locational for Berbice Dutch and Krio; predicate possession is *have* for the other 16 Creoles

Interdependence

D-M&B's network of 18 Creoles (including Seychellois), 3 Substrates (Malagasy, Makuwa and Swahili) and 1 Lexifier (French)



Interdependence at work

- 1. D-M&B's Claim:** “[This] network ... shows that even though only languages involved in the creation of Seychellois were included, the creoles cluster together, which goes against the predictions of a feature pool view, according to which Seychellois would be expected to appear close to the languages that were involved in its formation.”
- 2. Challenge:** The “supporting characters” in columns 5 and 27 are **CCS “2.1 Statives with past reference [indicated with anterior (or past) tense]”** and **CCS “7.3 Other auxiliary-like elements including modals.”** Both features are part of the same CCS complex of 40 inter-related features that are *defined based on the profiles of Atlantic Creoles’ TMA and verbal systems.*

Contradictions

1. French “statives with past reference [indicated with anterior (or past) tense]”:
J'aimais Marie ‘I loved Mary.’
cf. HC: *Mwen te renmen Mari*
2. French “auxiliary-like elements including modals”:
Je dois partir ‘I must leave’ (cf. HC *Mwen dwe pati*)
3. French periphrastic verbal constructions fed into the emergence of French-based Creoles “auxiliary-like elements,” including those of Seychellois (Michaelis 1993).

More contradictions:

D-M&B's claim: for 18 out of 18 Creoles, indefinite article = *one*

1. Hancock 1987 and Holm & Patrick 2007 (cited in Bakker *et al* 2011) describes **indefinite articles in *some* Creoles.**
2. Bakker *et al*'s (2011) study based on Parkvall 2008: F16 “**Indefinite article**” = “**0**” (i.e., *absent*) ***for every single Pidgin or Creole***
3. Daval-Markussen & Bakker 2011: ***In all their 18 Creoles, definite article is same as numeral ‘one’***

**Abrupt formation of indefinite articles as *one*—
between 2011 and 2012?**

What are the facts?

1. WALs: The use of the numeral for *one* as indefinite article is the most popular pattern among languages that have indefinite articles, and it is found among most of the Atlantic Creoles' lexifier languages: French, Portuguese and Spanish.
2. The grammaticalization of the cardinal *one* into indefinite articles is a well-known case of grammaticalization as in the history of Germanic and Romance (Lehmann 2002).
3. Indefinite article = *one* does not require any Pidgin phase.

Another contradiction:

D-M&B's claim: No tense-aspect inflection in 18 out of 18 Creoles, including Portuguese-based Creoles such as Capeverdean Creole and Korlai

1. Tense-aspect inflection is documented in Holm & Patrick 2007, Holm 2008, Luís 2008, **Cardoso & Luís 11/16/2012** (one hour ago)
<http://www.clul.ul.pt/en/component/content/article/458-facs3>
2. Luís 2008: tense/aspect suffixes in Korlai are *not* post-creolization developments, but are derived from the specific languages in contact
3. Holm 2008: Most Atlantic Creoles show few tense-aspect affixes since their source languages are “partially inflected superstrates and largely non-inflected, isolating substrate languages”

“Exceptional” vs. “Distinguishable”

- 1. Statistical fact:** With enough languages and features from which to draw, distinguishability can be demonstrated for an astoundingly large number of *arbitrary* sets of languages.
- 2. Statistical flaw:** A procedure that compares too few features across too many languages inevitably creates spurious “typologies.”
For example, D-M&B’s 4 features put Albanian, Dutch, English, Lakota and Yaqui in the same equivalence class.
- 3.** In WALS, there are some 10^{22} such equivalence classes based on 4 features.

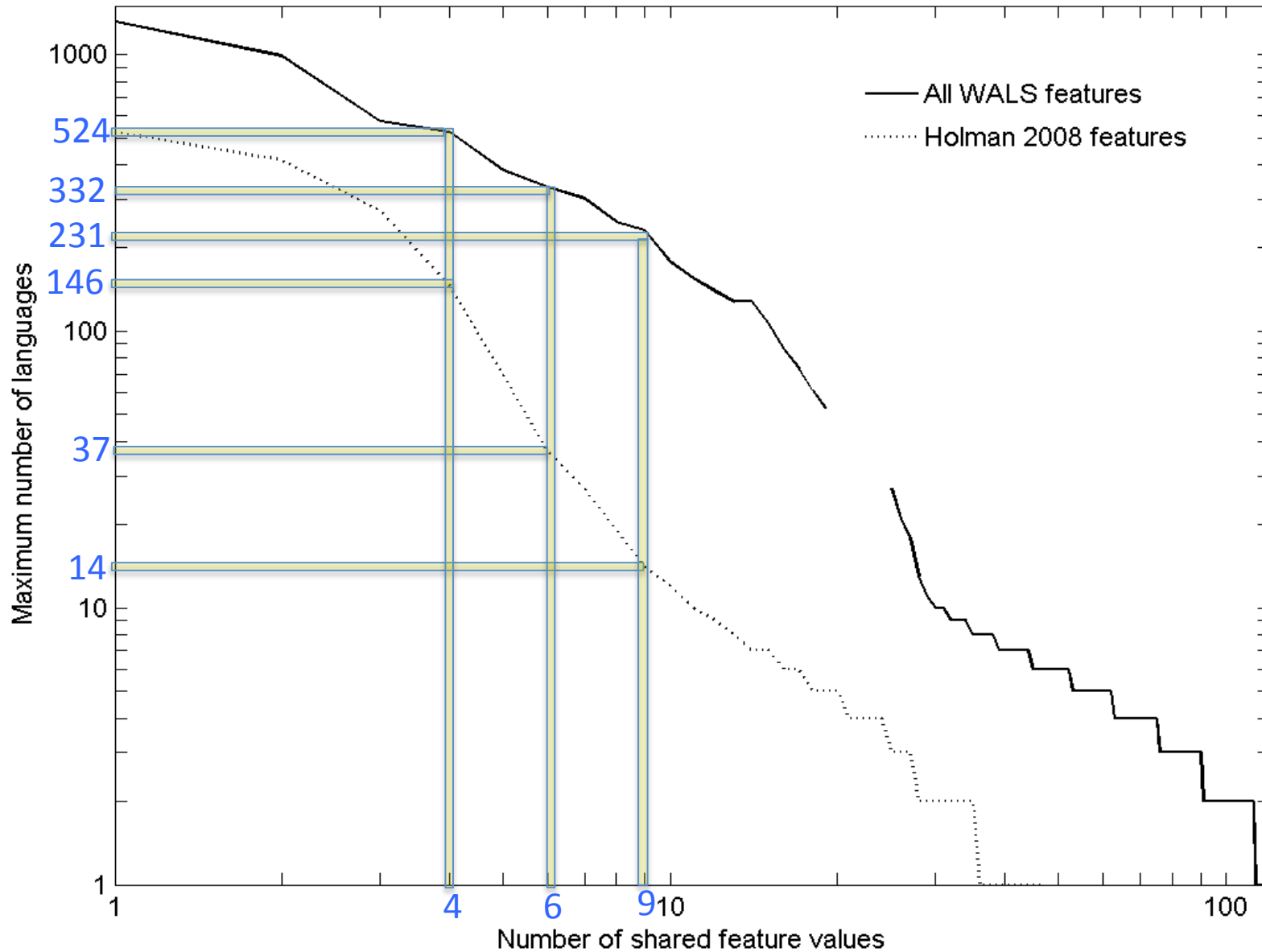
Distinguishable languages in WALS

Number of shared features	Maximum number of languages	
	All WALS features	Holman 2008 features
1	1,316	525
2	988	415
3	575	272
4	524	146
5	382	69
6	332	37
7	302	27
8	248	19
9	231	14
10	178	12
11	155	10
12	140	9
13	128	8
14	128	7
15	107	7
16	86	6
17	74	6
18	61	5
19	53	5

*Holman (2008)
eliminates inter-
dependent features
from WALS and
retains **47**
“approximately
independent”
features*

Distinguishable languages in WALS

Maximum number of WALS languages that share feature values



Distinguishable sets based WALS features

Common features	Count	Likelihood of language sets sharing at least this many features	Expected number of language sets sharing at least this many features
0	17,958,358	1.0	1.1×10^{25}
1	64,680,158	0.82	0.9×10^{25}
2	15,375,145	0.17	1.9×10^{24}
3	1,895,367	0.020	2.2×10^{23}
4	86,784	0.00091	1.0×10^{22}
5	3,986	0.000042	4.6×10^{20}
6	191	0.0000020	2.2×10^{19}
7	11	0.00000011	1.2×10^{18}

Distinguishable based on “approximately independent” features

Common features	Count	Likelihood of language sets sharing at least this many features	Expected number of language sets sharing at least this many features
0	82,995,585	1.0	1.1×10^{25}
1	16,534,833	0.17	1.9×10^{24}
2	466,220	0.0047	5.2×10^{22}
3	3,355	0.000034	3.7×10^{20}
4	7	0.000000070	7.7×10^{17}

Goals and outline

A three-step argument against recent instances of Creole Exceptionalism:

1. Creole languages do not, and cannot, reflect “early interlanguage” as defined in Plag (2008, 2009)
2. So far, there’s no “irrefutable” or “staggering” evidence that “Creoles are typologically distinct from non-Creoles
(e.g., the computational phylogenetic methods in Bakker *et al* 2011 and Daval-Markussen & Bakker 2012 do not reconstruct Creole formation).
- 3. Once the Comparative Method is adequately applied, Creole formation appears indistinguishable from language change.**

Back to Family Values

1. The most robust genealogical classifications to date have relied on **lexical and phonological data** (Paul 1890, Bloomfield 1933, Meillet 1958, etc.)
2. **Typological data are best when they complement lexical data** (Nahkleh *et al* 2005, Wichmann & Saunders 2007, Holman *et al* 2008)
3. Dunn *et al* 2008 use typological data only because the lexical data are no longer reliable because of the time depth of his data (some 3,000 years!)
4. This time-depth limitation (10,000 years according to Dunn *et al* 2008: 710–712; cf. Wichmann & Saunders 2007:378) certainly does not apply to Creole languages.

Trees vs. Networks

1. In biology, there is a growing consensus that phylogenetic trees alone cannot account for all of biological evolution. Witness reticulate events such as hybridization and horizontal gene transfer (Huson & Bryant 2006:254).
2. But reticulate events do not eliminate the fact that **mutation and speciation are among the most important mechanisms for biological evolution.**
3. **Mutation and speciation are still best captured by phylogenetic trees** of the traditional sort—similar to the *Stammbaumtheorie* tree in the Comparative Method.
4. **Phylogenetic networks** that explicitly represent reticulate events “provide an ‘explicit’ representation of evolutionary history, generally depicted as a **phylogenetic tree with additional edges**” (Huson & Bryant 2006)

HC and the “Gold Standard of historical linguistics” ...

... to be applied whenever *cognate sets can be reasonably established within a time depth of some 10,000 years* (Dunn et al 2008; Wichmann & Saunders 2007)

1. Cardinal numbers from French: *en* ‘1’, *de* ‘2’, *twa* ‘3’, *kat* ‘4’ ... *san* ‘100’ ... *mil* ‘1,000’ ... from French *un, deux, trois, quatre ... cent ... mille ...*
2. Ordinal numbers, including the suffix /-jɛm/ and its morphophonology (sandhi, suppletion, etc.), from French: *premye* ‘1st’, *dezyèm* ‘2nd’, *twazyèm* ‘3rd’, *katryèm* ‘4th’, ... *santyèm* ‘100th’ ... *milyèm* ‘1,000th’ ... from French *premier, deuxième, troisième, quatrième, ... centième, ...*
3. Kinship terms from French *frè* ‘brother’, *sè* ‘sister’, *kouzen* ‘cousin’, *kouzin* ‘cousin (feminine)’ ... from French *frère, soeur, cousin, cousine ...*
4. Color terms from French: *blan* ‘white’, *nwa* ‘black’, *rouj* ‘red’ ... from French *blanc, noir, rouge ...*
5. Body-part terms from French: *cheve* ‘hair’, *zòrèy* ‘ear’, *je* ‘eye’, *nen* ‘nose’, *bouch* ‘mouth’, *dan* ‘tooth’, *lang* ‘tongue’ ... from French *cheveux, oreille, yeux, nez, bouche, dent, langue ...*

More “Gold standard” for HC

6. TMA markers from French: *te* ‘ANT’, *ap* ‘PROG, FUT’, *ava* ‘IRREALIS’, *fini* ‘COMPLETIVE’ ... from French *étais/était/été* (imperfect and participle of ‘to be’), *après* ‘after’, *va(s)* ‘go +3sg/2sg+PRES’, *finir/fini(s)* ‘to finish’ ...
7. Prepositions from French: *nan* ‘in’, *pou* ‘for’, *apre* ‘after’, *anvan* ‘before’, *devan* ‘in front of’, ... from French *dans, pour, après, avant, devant, ...*
8. Determiners, demonstratives, etc., from French: *yon* ‘a’, *la* ‘the’, *sa* ‘this/that’ from French *un, la/là, ça*.
9. Pronouns from French: *m(wen)* ‘1sg’, *ou* ‘2sg’, *li* ‘3sg’, *nou* ‘1pl, 2pl’, *yo* ‘3pl’ ... from French *moi, vous, lui, nous, eux ...*
10. Complementizers from French: *ke* ‘that’, *si* ‘if’, *pou* ‘for’ ... from French *que, si, pour ...*

HC meets the Gold Standard!

11. Almost all derivational morphemes from French: for example, HC *de-* as in *deboutonnen* ‘to unbutton’ and *dezose* ‘to debone’ from French *de-* which, like HC *de-*, has inversive and privative uses.
12. Morphophonological phenomena with French roots: liaison phenomena as in *an Bèljik* ‘in Belgium’ vs. *ann Ayiti* ‘in Haiti’; *de zan* /de zã/ ‘two years’, *twa zan* /twa zã/ ‘three years’, *san tan* /sa tã/ ‘one hundred years’ ... (cf. the pronunciation of the HC and French ordinal and cardinal numbers above; Cadely (2002, 2003) provides further examples of HC-French correspondences in phonology)

These HC-French correspondences meet Nichols’ (2006) “individual-identifying” threshold: these correspondences contain a fair amount of “faits particuliers” (or “language-particular idiosyncratic properties” in Meillet’s terminology) that reliably rule out chance correspondences.

Cf. **Henri & Bonami 11/15/2012** and **Araújo et al 11/16/2012** on Mauritian Creole, Gulf of Guinea Creoles and Surinamese Creoles:

<http://www.clul.ul.pt/en/component/content/article/458-facs3>

Envoi

1. Great care must be taken in the use of language-acquisition data and phylogenetic tools in Creole studies.
2. The (apparent?) conviction that Creole languages must be “exceptional” leads to *exceptionally* gross errors.
3. The Comparative Method applies straightforwardly to Atlantic Creoles: they descend from their European ancestors, with modification—modification via L1A and L2A, and partly influenced by Niger-Congo substrates.
4. Our results go counter to one long-standing orthodoxy in linguistics—Creole Exceptionalism (Saint-Quentin 1872, Adam 1883, Taylor 1956, Bickerton 1984, Thomason & Kaufman 1988, Ringe *et al* 2002, Nahkleh *et al* 2005b, Labov 2007, etc, etc; see overview in DeGraff 2005, 2009).