

An Exploration Into the Placement of Creole Languages in Language Evolution Phylogeny Estimation

Sarah Schieferstein

Creole languages

- Not single-parent evolution
- One parent: lexifier
 - Gives language lexicon
- Other parents: substrates
 - Possibly gives morphosyntax
- Other creoles:
 - Most similar syntactically
 - Because they undergo same process

mo ki sou lang mwen
cheri
pa nan lang mwen

se mo ki anba lang
ki fè lanm anndan m'
avan yo pwente tèt yo
sou pwent bouch mwen
epi yo jete kò yo nan lank
yo pran fè lanmou
tankou lanp

sa m' gen pou m' di ou
cheri
pi plis pase sa m' panse

men panse sa w' vle
pa panse m' ka panse
kè m' k'ap senven

cheri

annou kite mo yo gonfle
venn yo devan tout devenn

les mots sur ma langue
chérie
ne sont pas dans ma langue

ce sont des mots sous la langue
faisant lames en mon âme
avant de pointer leur tête
sur le bout de mes lèvres
puis se jettent dans l'encre
se mettent à faire l'amour
comme des lampes

ce que j'ai à te dire
chérie
est plus qu'un surplus de pensée

mais pense ce que tu veux
pense pas que je peux panser
mon cœur saignant

chérie

laissons les mots gonfler leurs
veines face à toutes déveines

Language evolution phylogeny with creoles

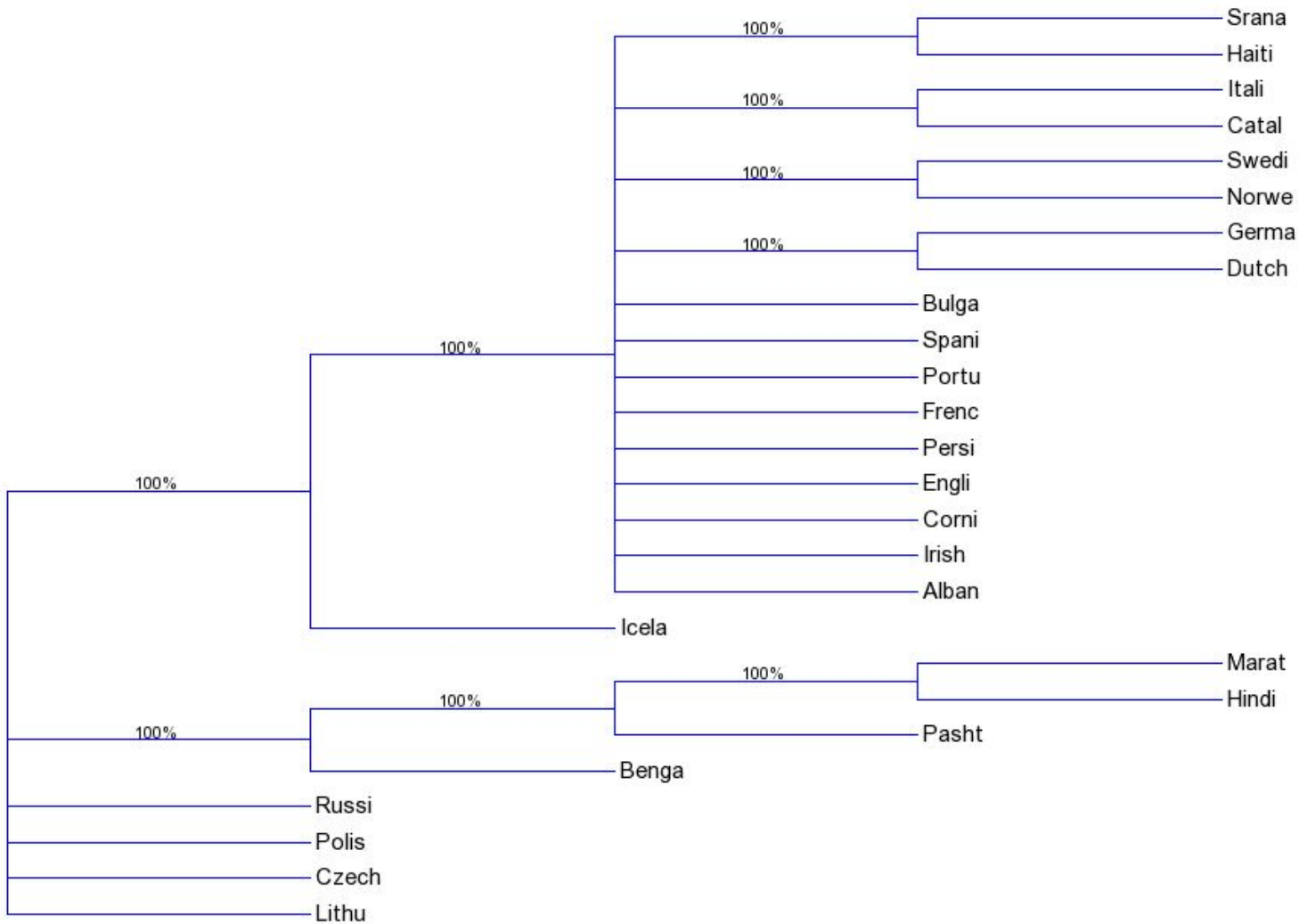
- Studied before with NeighborNet and SplitsTree
 - But they assume single parent evolution
- Since they have multiple parents and are ‘hybrids’... use existent hybridization methods
 - Ideally will find ‘reticulations’ to its multiple parents
- Compare with single parent evolution estimation like ML and MP

Input Data

- Compare syntactic features, **not** cognates
- Examine trees
 - With no creoles
 - With 1 / 2+ creoles
 - With substrates
- Each site is a syntax feature from APiCS/WALS, each row is a language

Alban	2	1	1	0	1	2	1	0	1	0	2	2	1	0	1	0	1	1	1	1	1	0
Lithu	0	1	4	0	2	2	2	0	3	1	2	2	3	0	1	0	2	0	1	1	4	0
Irish	3	1	3	1	1	2	1	1	1	3	2	1	1	0	1	1	1	0	1	0	1	0
Corni	0	1	3	0	0	2	1	0	1	0	0	0	2	0	1	0	0	0	1	0	0	0
Enqli	3	1	1	1	1	3	1	1	2	1	2	2	1	1	1	2	1	1	1	1	1	4

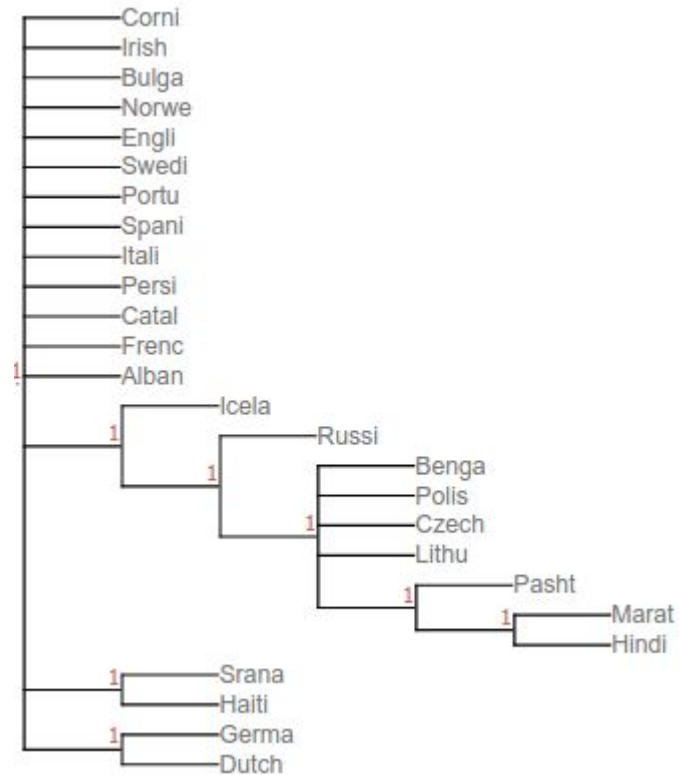
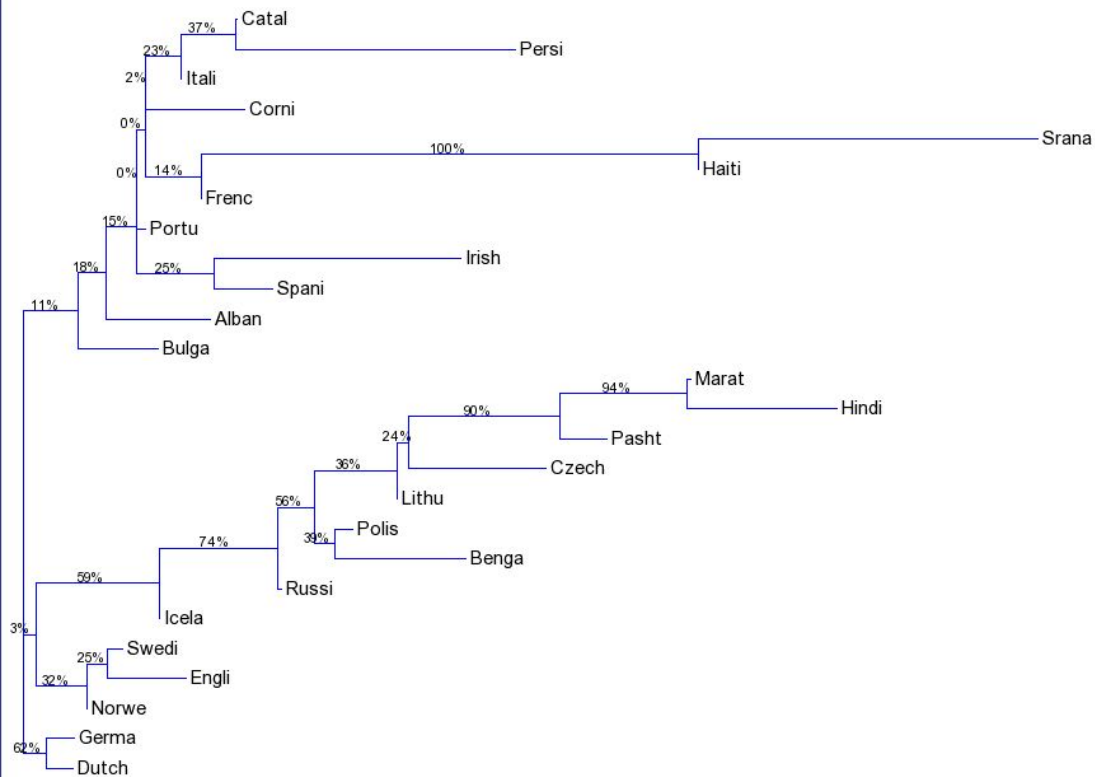
MP



MP evaluation

- **Groups creoles together**
- Confused about some Romantic / Celtic groupings
- The two Iranian languages aren't close at all

ML (best ML tree + consensus)



ML evaluation

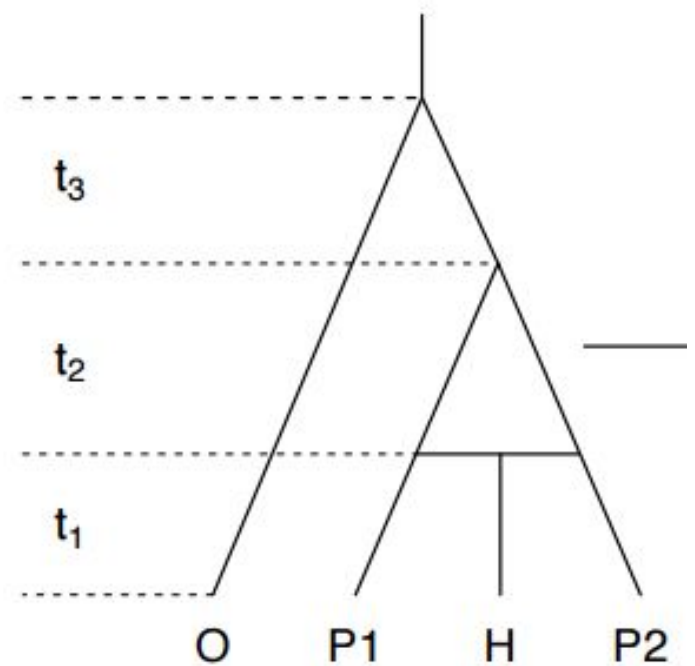
- Same as MP but it's more unsure about Romance language groupings
- Can see from the best ML tree that it places the creoles next to one of their lexifiers, French

Summary of single-parent evolution

- These trees aren't great (vs. the ground truth for Indo-European)
 - The usual input for these trees is cognate data
 - Making language evolution trees with syntax is new, I have to do it less naively
 - Maybe only using certain types of syntax features will help increase the phylogenetic signal
- Creoles are grouped together even though they are not related at all by evolution!!!
 - This is because their syntax is similar
 - Maybe using lexical features as well could help group them less
- Creoles are grouped near their lexifiers, an actual parent

Hybridization

- Network, not tree
- H has DNA from P1 and P2
- Analogy: Creoles get syntax features from
 - Its lexifier
 - Its substrates
- So a phylogeny with a creole should be a network and denote this hybridization process with extra edges



Future work

- Include substrates in the trees (they are the other parents of creoles)
- Estimate a phylogeny that assumes hybridization can occur
- Explore different inputs in order to make the trees better
- In-depth evaluation: is hybridization good enough? Should a new model be made that considers creole languages explicitly? What should this model consider? What type of input should it have? Can the historically correct parents of creole languages even be estimated in a phylogeny under a custom model?

References

<https://www.quora.com/How-do-the-French-perceive-Canadian-French> for slide 2's picture

Gerard, David, H. Lisle Gibbs, and Laura Kubatko. "Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling." *BMC evolutionary biology* 11.1 (2011): 291. (For slide 10's picture)

Campbell, Lyle. *Historical linguistics*. Edinburgh University Press, 2013.

Dyen, Isidore, Joseph B. Kruskal, and Paul Black. "An Indoeuropean classification: A lexicostatistical experiment." *Transactions of the American Philosophical society* 82.5 (1992): iii-132.

Daval-Markussen, Aymeric, and Peter Bakker. "Explorations in creole research with phylogenetic tools." *Proceedings of the EAACL 2012 Joint Workshop of LINGVIS & UNCLH*. Association for Computational Linguistics, 2012.

Nakhleh, Luay, et al. "A comparison of phylogenetic reconstruction methods on an Indo-European dataset." *Transactions of the Philological Society* 103.2 (2005): 171-192.

Shu, Kevin, et al. "Syntactic phylogenetic trees." *Foundations of Mathematics and Physics One Century After Hilbert*. Springer, Cham, 2018. 417-441.