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
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
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?


