

## Language Families, Language Typology, and Creole Languages

J. Clancy Clements, Indiana University

For at least the last 25 years, an ongoing debate in Pidgin and Creole Linguistics has been whether or not stabilized pidgins and creoles (P&Cs) should be considered ‘different’ languages. There are three strands of research, divisible into two groups. One group views P&Cs as human language systems just like any other human language system, comparably complex to non-P&C languages, and they undergo change in the acquisition process just like any other language. The other group does not dispute that P&Cs are human language systems in and of themselves. The main question that drives this group of researchers is whether P&Cs are typologically different from non-P&Cs. McWhorter’s (2005, 2011, 2018) work represents one strand of this second group. He argues that P&Cs can be distinguished from non-P&Cs using a set of three features. That is, as compared to non-P&Cs, P&Cs have little or no morphological inflection, little or no grammatical or lexical tone, and little or no opaque derivational morphology. This line of research has generated considerable controversy and discussion, and has evolved and been refined over time. The other typologically motivated strand of research, which grew out of McWhorter’s work (Bakker et. al. 2011, Daval-Markussen 2015, Bakker et al. 2017, among others), uses an arguably stable set of features from well-known data bases (WALS and APiCS), along with phylogenetic trees developed for quantifying genetic relatedness of biological organisms, to measure typological distance among languages and determine the extent to which P&Cs cluster as a group. This research has been criticized for the biased selection of features and for the inappropriateness of a statistical model based on biology to examine language relatedness. Moreover, the phylogenetic tree model has not always been able to isolate the key distinguishing features responsible for the clustering of certain language groups.

Situated within this last strand of research, this paper addresses some of the criticisms leveled against the phylogenetic approach. I am part of a team who since 2016 has been working on a series of typological studies using the Random Forests algorithm with a cluster analysis. Our main research question is: do Creole languages cluster together because of their shared typological features? The algorithms selected do not have the same perceived problems the phylogenetic tree model is claimed to have. I report on three analyses so far carried out, based on Clements et al (2018). We used previously identified stable features to analyze a collection of 234 languages (34 creoles, 39 substrates, 10 lexifiers, 151 others). In the first analysis, using all 34 features, 4 features were found to be key in identifying optimal clustering of the languages into two clusters; one containing all 34 creole languages and some others and the other consisting of all 10 lexifier languages (among them French, Portuguese, and Spanish) and the majority of the other languages in the sample. We then eliminated these 4 features and ran the analysis with 30 features. Again, the optimal clustering was a 2-cluster grouping, where 31 of the 34 creoles appeared in one group. In the third analysis, with 26 features, a 5-cluster model was found to be the best fit for the features. Again, creoles languages (25/34) clustered largely within one group. In each analysis, the Creoles included in the study did not cluster with the lexifier languages, nor with substrate languages, nor primarily with the other languages included in the study. Rather, the Creoles cluster largely as a group, though not exclusively so. To account for these findings, I appeal to the processes operative in naturalistic L2 acquisition, proposing that frequency and detectability of lexical items in discourse play a large role in accounting for the clustering of creoles.

### References

- Bakker, Peter, Finn Borchsenius, Carsten Levisen & Eeva Sippola (eds.). 2017. *Creole Studies – Phylogenetic approaches*. Amsterdam & Philadelphia: Benjamins.
- Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall, and Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26(1).5-42.
- Clements, J. Clancy, Michael Frisby, Dylan Jarrett, & Fernando Melero García. 2018. Language features used in typological relatedness. Paper presented at the annual meeting of *Associação de Crioulos de Base Lexical Portuguesa e Espanhola*, Université Assane Seck, Ziguinchor, Senegal. June 11-14.
- Clements, J. Clancy, Michael Frisby, Dylan Jarrett, & Fernando Melero García. 2017. Language typology and algorithms used to determine language groupings. Paper presented at the annual meeting of *Associação de Crioulos de Base Lexical Portuguesa e Espanhola*, Stockholm University, Stockholm, Sweden, June 13-15.
- Daval-Markussen, Aymeric. 2015. Refining the notion of bi-clans in creole studies. Paper presented at the summer meeting of the Society for Pidgin and Creole Linguistics. Graz, Austria, 7 July.
- McWhorter, John. 2018. *The Creole debate*. Cambridge: Cambridge University Press.
- McWhorter, John. 2011. Tying up loose ends. A creole prototype after all. *Diachronica* 28(1).82-117.
- McWhorter, John. 2005. Defining “Creole” as a synchronic term, *Defining creole* by John McWhorter. Oxford: Oxford University Press, 9-37.