CreoPhonPt: a collaborative database saving Portuguese creoles from digital obliteration

Sousa e Silva, Carlos Rogério

silvacarlosrogerio@gmail.com
Centro de Linguística da Universidade do Porto, Portugal

Pimentel Trigo, Luís Manuel

ltrigo@letras.up.pt
Center for Digital Humanities and Innovation - Faculdade de Letras Universidade do Porto, Portugal

While the vast majority of Portuguese-based creoles are considered endangered languages (Lewis, 2013), the linguistic literature about these languages is sparse, scarce, and, many times in closed access (Cardoso et al. 2015). To solve this issue, APiCS-Atlas of Pidgin and Creole Language Structures (Michaelis et al. 2013)- has been created. However, on the one hand, the sample of Portuguese-based creoles still leaves room for expansion. On the other hand, the phonological features of creoles are disproportionately present in comparison to syntax or semantics, which keeps phonology as the most understudied in creolistics (Smith 2008). For these reasons, we felt the need to create CreoPhon.

CreoPhon is a pilot database that, for now, only includes Portuguesebased creoles (CreoPhonPt). Its mission is to collect sound phonological data about these languages systematically, to put together a findable, accessible, interoperable, and reusable dataset, and to, by using these data, produce studies that describe so-called "creole phonology" and spread it among the scientific community and the general public.

Although the idea and its early stages of development were the responsibility of one linguist, the project expanded rapidly. First, collaboration with a data scientist was necessary to perform data transformation and management. After that, as the need for more data points increased, CreoPhonPt also became an educational opportunity. During the Phonology seminar (MA), students had practical data collection sessions and data transformation tutorials, while applying their phonological knowledge and contributing to the project.

Data collection and transformation were always performed according to a protocol written by the leading researchers. The collection of phonological inventories was discarded because it is highly dependent on the theoretical orientations of the field linguists (Kiparsky 2018). Students were asked to collect 200-words Swadesh lists (Swadesh 1955) of a given Portuguese creole of their choice with phonetic transcription, along with the cognates and translation. Then, in spreadsheets, they performed some manual annotation, like syllable division and some phonological processes. Finally, they could transform and visualize the data according to their own research interests.

Today, CreoPhonPt is comprised of 18 phonologically informed Swadesh lists representing each non-extinct Portuguese creole. It is a ready-to-use tool for advanced research either in phonology or historical linguistics. Nevertheless, it can also be used beyond Academia for the creation of orthographic norms for creoles, which are known as an important means of language conservation and revitalization (Rutkowska 2017).

Beyond linguistics, this database is also conceived as an opportunity for interdisciplinary studies and collaboration. As creoles are a consequence of unique situations of contact, their birth and development can never be fully understood without historical, geographical, ethnological, and genetic data (Hagemeijer / Rocha 2019), whose inclusion of part of the plans for CreoPhon expansion in the near future.

Bibliography

Cardoso, Hugo / **Hagemeijer, Tjerk** / **Alexandre, Nelia** (2015): "Crioulos de base lexical portuguesa". In *Manuel des anthologies, corpus et textes romans*, 670–692. Mouton deGruyter.

Hagemeijer, Tjerk / **Rocha, Jorge** (2019): Creole languages and genes: the case of sao tome and principe. *Faits de Langues* 49(1). 167 – 182.

Kiparsky, Paul (2018): *Formal and empirical issues in phonological typology* 54–106. Berlin, Boston: De Gruyter Mouton. Lewis, M. Paul (ed.). 2013. *Ethnologue: Languages of the world*. Dallas, TX, USA: SIL International seventeenth edn. https://researchr.org/ publication/ethnologue.

Michaelis, Susanne Maria / **Maurer, Philippe** / **Haspelmath, Martin** / **Huber, Magnus,** eds. (2013): *Apics online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://apics-online.info/.

Rutkowska, Hanna (2017): *Chapter 11: Orthography* 200–217. Berlin, Boston: De Gruyter Mouton.

Smith, Norval (2008): *Creole phonology* chap. 5, 98–129. John Wiley Sons, Ltd.

Swadesh, Morris (1955): Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21(2). 121–137.