

## ETT 2024 - Lexical cluster analysis of 10 Bantu A80 languages

In this presentation, I will show the results of a small-scale cluster analysis performed on a group of 10 related Bantu languages. Bantu languages are spoken from Cameroon to Kenya in the North, all the way to South Africa, covering a big part of the African continent. They are grouped together into zones, indicated with a capital letter, and subgroups, indicated with two-digit numbers (Guthrie 1971). This reference system is based on geography and does not make claims about genealogical relations between languages (updates to the original system have maintained this referential character, see e.g. Maho 2012). However, languages within a subgroup can actually be related, as is the case with the languages of our study.

The subgroup that we are concerned with here is A80, spoken at the far North-Western end of the Bantu area (Cameroon and neighboring countries); more specifically, the 10 languages in this group for which enough data is available (Cheucle 2014, Grimm 2021, Vermeir & Allasonnière-Tang, forthcoming). Even though these 10 languages (Bekwel, Gyeli, Kol, Koonzime, Kwasio, Makaa, Mpiemo, Njyem, Shiwa and Sso) are clearly related, as becomes apparent when studying the vocabulary, it is also clearly possible to add depth in the analysis by identifying subgroups of languages that are more closely related. The number of subgroups and how the 10 languages are divided amongst them optimally can be calculated using cluster analysis. Our analysis is based on a list of 556 concepts for which at least six languages of the corpus have a unique translation, providing a more in-depth study than the ones cited by Cheucle (2014, pp. 17-27). We used the Levinshtein method to calculate the phonological distance matrix between words, resulting in a hierarchical clustering. The results show a 0.8 positive correlation with the hierarchy proposed on Glottolog (Hammarström et al., 2024), with only slight differences in the lower subgroups. Subgroups clearly share distinct typological features, such as complete affrication (Gyeli, Kwasio, Shiwa, Sso), syncope (Koonzime, Njyem), and complex codas (Kol, Makaa).

Cheucle, Marion. 2014. *Étude comparative des langues makaa-njem (Bantu A80) : Phonologie, morphologie, lexique*. Université Lumière Lyon 2.

Grimm, Nadine. 2021. *A grammar of Gyeli*. Berlin: Language Science Press.

Guthrie, Malcolm. 1971. *Comparative Bantu*, Vol 2. Farnborough: Gregg Press.

Hammarström, Harald & Forkel, Robert & Haspelmath, Martin & Bank, Sebastian. *Glottolog* 4.8. Family: Narrow Bantu, <https://glottolog.org/resource/languoid/id/narr1281> (accessed 24 January 2024).

Maho, Jouni. 2012. A classification of the Bantu languages: An update of Guthrie's reference system. In: Nurse, Derek & Gérard Philippson (eds.). *The Bantu Languages*. New York: Routledge, 639-651.

Vermeir, Tessa & Marc Allasonnière-Tang. Forthcoming. Cluster analyses on the lexicon of 10 Bantu A80 languages.