How databases shape research: labial-velars distribution in Africa

(Workshop 3: Areal Phenomena in Northern Sub-Saharan Africa)

SEGERER, Guillaume
LLACAN - CNRS (segerer@vjf.cnrs.fr)

Abstract

Recent years have seen a serious increase in the number and the quality of available online linguistic databases. Yet the use of databases for research purposes is still in its infancy. I intend to show how, by using different databases, one can test scientific hypotheses and how different databases can be used for different purposes. The examples will be drawn from phonology, the domain where the most comprehensive datasets are to be found. Specifically, I will examine in some details the distribution of labial-velars in Africa, since this feature has been considered typical of the 'Macro-Sudan Belt' (Clements & Rialland 2008, Güldemann 2008), hence showing an areal distribution pattern instead of a genealogical one.

The following online databases have been explored : WALS[1] (World Atlas of Language Structures), PHOIBLE[2], LAPSyD[3] (Lyon-Albuquerque Phonological Systems Databases, Version 1.0.). All of them are freely available and provide maps for selected features. First I will show how these databases differ from each other: scope and quantity of data, their quality (i.e. reliability), presence vs absence of explicit curation, query interface, etc.

Then the specific case of labial-velars will be explored through these databases. The geographical distribution of labial-velars is known to be restricted to an area that has been labelled 'Macro-Sudan Belt' (Güldemann 2008). Languages that have one or more labial-velar consonant(s) in their inventory are all situated within this area, and no language outside the area seem to exhibit any labial-velar consonant (though there probably are a handful of exceptions in the Pacific region). The use of various online databases to assess this claim yields no big surprise, if one considers the general pattern only. In the details, though, lie a few interesting things, not every one of which are captured by online databases. The most important is the local prevalence of labial-velars, i.e. their status in each language. If the actual distribution of labial-velars is the result of contact and diffusion, one would expect that their status be more marginal at the edges of the domain. The goals of this talk will therefore be: i) to present arguments to verify (or not) the above prediction; to show that even when the use of databases is not in itself enough to answer some scientific questions, they can be most helpful in suggesting directions that could have been very difficult to take without these tools.

Keywords: databases, phonology, labial-velars

---

[1] http://wals.info/

[2] http://phoible.org/

[3] http://www.lapsyd.ddl.ish-lyon.cnrs.fr/