Language Table and Dataset

**Overview:**

Perhaps one of the most important and challenging datasets for ethnobiological research, particularly comparative ethnobiological research, is standardizing the "Language" for which the ethnobiological data is . A particular ethnobiological datum (e.g., a plant name, the gloss of this name, plant use) is associated with the following upwardly inclusive individuals/groups:

1. the consultant

2. the community (or in some cases, group within a community such as ritual specialists)

3. the language (e.g., a number of communities that speak the same language)

4. the language group(s) (higher level nodes above the language but below the "family")

5. the language family

A screenshot of the Glottolog taxonomic structure associated with Yoloxóchitl Mixtec is given below (last page). We will try to program DEMCA/Symbiota so that searches can be "monophyletic" (i.e., a node such as "Guerrero Mixtec" is selected and the search encompasses all languages directly descended from that node; or it may be paraphyletic (e.g., all descendants of a given node except some; e.g., Yoloxóchitl and Alacatlatzala and Alcozauca Mixtec, but not Metlatonoc Mixtec) or polyphyletic (e.g., Guerrero Mixtec and Amuzgoan). Probably the most common search groupings will be monophyletic. Polyphyletic searches will be good for comparison between two datasets of languages that are not immediately related but perhaps in contiguous geographical areas (e.g., Nahuatl and Totonac). This would help in documenting contact phenemona. Paraphyletic searches might be less useful but could be employed when one wishes to exclude a certain dataset (e.g., researcher A has worked on two languages descended from a common node and wants to compare the data but not include data from researcher B who has researched another language descended from the same node).

The use of the codes established by Glottolog will therefore facilitate a wide degree of flexibility in searches by standardizing the phylogenetic relationship among languages in a wide area. It will obviously be disappointing to those who are in disagreement with either the higher level nodes of the Glottocode tree or, more importantly, with the language level codification. To accommodate this Symbiota/DEMCA will allow for textual/documentary descriptions of the language and for search capabilities for "user-defined" language internal divisions.

Example 1: A researcher on Yoloxóchitl Mixtec does not think that YM is part of the Guerrero Mixtec group but rather part of Eastern Alto Mixtec. This can be documented in a Language History tab by means of which the researcher can offer arguments to reassign YM. At this point however we can only provide this descriptive functionality but we cannot be "re-phylogenizing" the Glottocode taxonomy. Note that a user who wishes to compare YM with other Eastern Alto Mixtec languages (9 total) can do this based on the para/polyphyletic search capabilities of Symbiota/DEMCA.

Example 2: A researcher only has available a Glottocode for a language that s/he feels is too inclusive and that should be divided. For example, I study what I call Balsas Nahuatl, which is a small sub-group of Nahuatl languages within Gurrero Nahuatl (ngu; https://www.ethnologue.com/18/language/ngu/). It includes communities that speak the only tonal Nahuatl language described to date; for this and other reasons I consider it separate from Guerrero Nahuatl. One easy functionality would be to allow a project team working on any group to create a text or link a document that presents evidence for Glottocode–language internal divisions. We will also work on the possibility of adding a user-generated code that subdivides a language and allows search capabilities on this subdivision. In the above case we would allow an "annotated" Glottocode. So if Guerrero Nahuatl is guer1241 we will allow for the researcher to create a subdivision (e.g., guer1241-01). We will work with the Glottolog team to see what they might suggest. The best circumstance would be for them to include a "Balsas Nahuatl" code and incorporate Balsas Nahuatl in their phylogeny.

**Database structure**

Many potential users are more familiar with ISO 639-3 than with Glottocode. However, as noted above, Glottocode facilitates higher level and more intricate searches. Users for Symbiota/DEMCA will be able to enter either the Glottocode or ISO 639-3 codes. In most cases (we still don't know if in all cases) entering one code will, through the Glottolog API, automatically find a synonymous code in the other system. Again, we will try to establish a dialogue with the Glottolog team to look into this and other questions.

------------------------------------------------------------------------------------------------------------------------------------------

**Language table data input**

**Glottocode ISO 639-3 Click to see Glottolog phylogeny**

**Discussion of higher-level language history**  **Upload file [Note: Opens browse window]**

**Discussion of language internal divisions**  **Upload file [Note: Opens browse window]**

**Code for language subdivision**

**Upload files:**

**Orthography and phonology:** [Note: Opens browse window to search for files on local computer to upload]

**Other**

