**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 ethnobiological data is presented. 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 (who is a member of a **community**)
2. the community (in which one or more **languages** are spoken)
3. the language (a number of communities that speak the same language, which historically is part of a **language group**)
4. the language group(s) (these are higher level nodes above the language but below the "family"; language groups are part of a **language family**)
5. the language family

**Note** that these relationships are often one-to-many. For example, a consultant is usually a member of a single community but may speak more than one language (these multiple languages are listed under the consultant personnel record; it may be that a given consultant [e.g., a in-marrying man or woman] may not even speak the community language but communicate with community members through a lingua franca. Data from this individual may be included in a community study though the individual does not speak the "community language". Other cases will occur of bilingualism in which a consultant speaks the community language as well as another, perhaps a mother tongue). All these possibilities will be able to be databased in DEMCA.

Likewise communities may have in them more than one spoken language (again, this is documented in the **Community table**). The above 1–5 list is simply a summary presentation of the different search criteria that may be used in grouping ethnobiological data.

A screenshot of the Glottolog taxonomic structure associated with Yoloxóchitl Mixtec is given below (last page). DEMCA/Symbiota searches by default are either language specific or "monophyletic": for example, a node such as "Guerrero Mixtec" is selected and the search encompasses all languages directly descended from that node. DEMCA/Symbiota will also have the ability for paraphyletic (e.g., all descendants of a given node except some; e.g., Yoloxóchitl and Alacatlatzala and Alcozauca Mixtec, but not Metlatonoc Mixtec) and polyphyletic (e.g., Guerrero Mixtec and Amuzgoan) searches. 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 phenomena. 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. While the default use of the Glottocode taxonomy may be problematic for some researchers, we do not have the capacity to engage these debates (much as a herbarium website is not the place to engage in debates on classification). However, users will be able to search across different, including unrelated, nodes in the Glottocode hierarchy and thus create their own groupings that may represent other taxonomies or areal groupings or sprachbund. They may do this by selecting for search inclusion languages below different nodes but joined, for instance, by geographic proximity.

**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.

We had thought about including Glottocode functionalities on the DEMCA server but at this point it seems easiest to send a query to Glottolog and open up a window (e.g., of the mapping or taxonomic tree).

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

**Language table data input**

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

Upload file **Browse for file**

Topic of uploaded file:

o Orthography

o Contact languages

o Language history and phylogeny

o Language divisions

o Phonetics and phonology

o Morphology

o Syntax

o Other \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (specify)

It will often be important to add information on the language that is being "databased". For example, a project director often has a particular orthography that will need to be explained. So there is a way to upload a file on the orthography that is being used. For example, I use 1 low to 4 high for Yoloxóchitl Mixtec whereas other Mixtec orthographies use 1 high to 3 low. Some orthographies use diacritics and an underline (e.g., for mid tone). It might be that underlining is not feasible in DEMCA so perhaps the Project Director has chosen dieresis. One contributor to this list suggested noting contact languages that might have impacted the ethnobiological terminology of the language being studied. Or perhaps the nomenclature that is being databased sheds important light on language contact and division. Thus in a project on Sierra Nororiental de Puebla Nahuat I might well want to observe that there is a significant number of Totonac terms among the plant names.

Note that Glottolog also has a way to link published articles on language history/phylogeny to any given language.

