Computational Cladistics of Afro-Asiatic

Hezekiah Akiva Bacovcin & David Wilson

Introduction: Afro-Asiatic languages are (or were) spoken throughout the African Mediterranean, the Lake Chad Basin, the Nile River Valley, the Horn of Africa and the Near East. The family is among the oldest, with Egyptian attestation beginning possibly as early as 3300 BC, and Akkadian cuneiform inscriptions dated as early as about 2600 BC. Since the work of Bender (1971), scholarly consensus has considered the Afro-Asiatic family as containing 6 major order branches; Semitic, Egyptian, Berber, Chadic, Cushitic, and Omotic. Despite this agreement on the members of the Afro-Asiatic family, there is little to no agreement on the internal subgrouping of these branches into different clades. Various proposals, such as Diakonoff (1988), Ehret (1995), or Blench (2001) differ markedly in their approach and methodology, and likewise in their results.

To address the internal cladistic subgrouping of Afro-Asiatic we propose using a computation algorithm designed to evaluate various typological properties of the Afro-Asiatic languages and, using methods of comparing different possible phylogenetic trees, to determine which tree is most congruent with the present data on the major order Afro-Asiatic branches.

Character Coding: Building on the methodology of Ringe et al. (2002), we have chosen to adopt a Character/State model. Character/State models have their origin and common use in biological taxonomic classification. In a biological Character/State model, various features of living organisms are considered as the inputs or “characters" to of the model. These characters may range from obvious morphological characters such as number of limbs, size of brain, or nature of reproduction, to the extremely minute and discrete such as cellular composition or even sequence of individual nucleotides within the genome.

Linguistic character/state models are not altogether dissimilar from their biological counterparts. Daughter languages take the place of organisms to be compared. Behavioural, anatomical and genetic characters are replaced by phonological, morphological, syntactic, and lexical characters. Alphanumeric states are likewise assigned to each language for each state, and some parsimony algorithm is applied to the states. In our study, the daughter languages chosen will be the reconstructed major order Proto-Languages wherever possible, including Proto-Semitic, Proto-Berber, Ancient Egyptian, Proto-Chadic, Proto-Cushitic and Proto-Omotic.

Since the statistical power of the tree is tied to the number of characters for which it can compare each of the daughter languages, a large number of characters was chosen representing different aspects of the lexicon, phonology, morphology and syntax of each representative branch. Wherever possible, a state value for each character was designated as the "ancestral state", which would not be utilized as possible clade defining state, to avoid the pitfall of grouping languages based on shared inheritance. Phonological characters were drawn from the correspondences of reconstructed segments in each major order branch. Morphological characters were drawn from the shapes of various morphemes such as pronominal stems, verbal actor affixes and case inflection. Lexical characters were drawn from shared lexical roots present or absent in the different branches. Examples of a phonological, morphological and lexical characters are provided below.

Computational Methods: We have used a new method of computational cladistics involving a set based metric, because it is better able to deal with polymorphism (an issue brought up in Ringe et al. (2002)) This method first identifies all of the potential clades (i.e. sets of languages that share at least one character state). The algorithm then weighs each potential clade by the number of

	Ancestral	Semitic	Egyptian	Berber	Chadic	Cushitic	Omotic
*c	*c	*s	*s	*z	*c	*c	*s
1st Sg. Ind Suffix	∅	*-aku	*-aku	*-aku	∅	∅	∅
*ḳas- "bone"	Unknown	∅	*ḳas-	*ḳas-	*ḳas-	*ḳas-	*ḳas-

characters that supports it. Using a branch-and-bound search (Felsenstein, 2004), we find the tree (or set of trees) that has the largest number of supported clades (using the previously mentioned weighting). Two clades are incompatible with each other if they have overlap and are not in a superset or subset relationship (e.g. the sets [Semitic and Egyptian] and [Egyptian and Berber]). A clade supported by many characters will make it into the tree even if there are a number of poorly supported incompatible clades.

Results: Our method concludes that there are two main branches of Afro-Asiatic (Northern and Southern). The northern branch is composed of Egyptian, Berber and Semitic, while the southern branch is composed of Chadic, Cushitic and Omotic. The northern branch is the more strongly supported of the two, being strongly identified by the shared sound changes *q̣→ *χ, *ḳ→* q̣, as well as by two constraints on root shape involving sibilants and dental consonants, and some morphological isoglosses involving verbal stem formation. The southern group is most strongly characterized by the presence of lexical tone, the absence of root constraints, a *q→ *k change and a few lexical items. Within the northern group, there is a further subgrouping of Egyptian with Semitic which is supported by a further shared change of *q→ *χ. In our talk, we will further discuss the method and the characters that underlie the various branchings in our tree, along with other clades that receive some support in our method.

References:

Bender, Marvin Lionel. 1971. The languages of ethiopia: A new lexicostatistic classification and some problems of diffusion. *Anthropological Linguistics* 165–288.

Blench, Roger. 2001. Types of language spread and their archaeological correlates: the example of berber. *Origini: Preistoria e protostoria delle civiltà antiche* 169–190.

Diakonoff, Igor M. 1988. *Afrasian languages*. Nauka.

Ehret, Christopher. 1995. *Reconstructing proto-afroasiatic (proto-afrasian): Vowels, tone, consonants, and vocabulary*, volume 126. Univ of California Press.

Felsenstein, Joseph. 2004. *Inferring phylogenies*. Sunderland, Mass: Sinauer Associates.

Ringe, Don, Tandy Warnow, and Ann Taylor. 2002. Indo-european and computational cladistics. *Transactions of the Philological Society* 100:59–129.

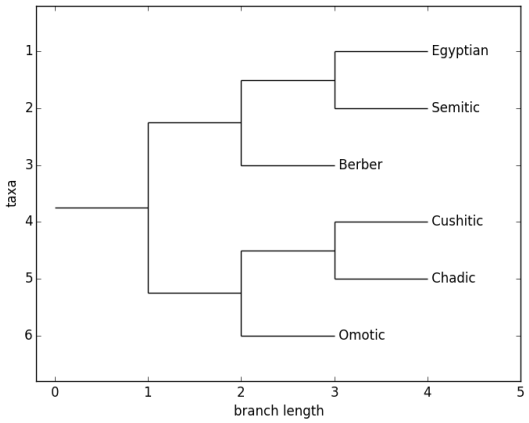

Figure 1: Predicted tree of Afro-Asiatic