

" Catastrophe! The Looting and Destruction of Iraq' s
Past

April 10— December 31, 2008

Home > Research > Projects > The Afroasiatic Index Project

Print this Page

Research Projects

Archaeology

Philology

Anatolia

Egypt/Nubia

Levant

Mesopotamia/Syria

Persia

Yemen

Individual Scholarship

Publications

Symposia

Research Archives - Library

CAMEL

Computer Laboratory

Academic/Degree Programs

Volunteer Program

OI History

Website Information

THE AFROASIATIC INDEX PROJECT

- INTRODUCTION
- THE AFROASIATIC LANGUAGES
- HISTORY OF THE PROJECT
 - THE CUSHITIC LEXICON PROJECT
 - THE AFROASIATIC INDEX PROJECT
- PRESENT STATE OF THE PROJECT
- EXCURSUS: WHAT IS ETYMOLOGY?
- ARTICLES
- ANNUAL REPORTS



INTRODUCTION

The Afroasiatic Index Project is a scholarly initiative that aims at creating an etymological database of Afroasiatic languages.

THE AFROASIATIC LANGUAGES

Afroasiatic languages are a group of related languages spoken by various communities from a large area in West African centered around Lake Chad (*Chadic*), all the way across North Africa (*Berber*) into Egypt (*Egyptian*), Ethiopia, and Somalia, and down the Great Rift Valley to the foot of Mt. Kilimanjaro (*Cushitic / Omotic*). Crossing over into Western Asia (*Semitic*), they are also spoken in the Middle East through Palestine and Syria, down around the Arabian Peninsula into Yemen and Oman, and up into Iraq. This huge area encompasses and enormous variety of people who have shared over many thousands of years a common linguistic bond, the nature of which can be discovered only by careful investigation and comparison of the respective languages.

It was to facilitate this process of bringing together as many Afroasiatic cognate sets, and as much pertinent literature, as possible that the Afroasiatic Index Project was conceived.

HISTORY OF THE PROJECT

THE CUSHITIC LEXICON PROJECT

The predecessor of the Afroasiatic Index Project was the Cushitic Lexicon Project, or Cushlex. The purpose of the Cushitic Lexicon Project was to provide interested investigators with access to the comparative lexical information contained in cognate sets existing within the 80 odd members of the Cushitic and

The Cushitic Lexicon, whose infrastructure is now largely complete, was done using hardware and software resources (e.g., desktop computer systems and off-the-shelf database management systems) readily available to any historical linguist. The idea was that, in addition to contributing to comparative Afroasiatic studies, the Project would provide faster and more accurate searches of the literature, and would offer query, display, and report facilities that would put old-fashioned 3x5 card files, notes, and even printed dictionaries to shame.

The data for the Cushitic Lexicon was designed to be independent of the programs that operate on it. The programs themselves were relatively straightforward, and can be used as written, or can provide a springboard for reformulation into as many distinct etymological and lexical tools as there are distinct sets of query and report facilities that can be used on data existing in this format (which is a lot).

The file format and programs were not specific to the language base being investigated here (i.e., Cushitic and Omotic), and could be used as-is to encode etymological information for sets of related words in other language groups. Presumably, if they were used on another language family, differences in goals and methods would lead to changes in the structure and presentation of the data. For example, certain aspects of how cognates display in the Cushitic Lexicon set-up stem from the fact that while cognacy within one major sub-family of Cushitic is fairly straightforward, it can be somewhat problematic within the context of Cushitic as a whole. One therefore needs two different kinds of presentation, one for each level. It is quite possible that in other language areas, one might not need to make this distinction. Or, conceivably, one might need to make more such distinctions.

Originally, data and query mechanisms for the Cushitic Lexicon were developed using dBASE IV, which functioned, during the 1980s, as a kind of *lingua franca* for off-the-shelf database management systems. Because the Cushitic Lexicon was written in dBASE IV, the data could be used by any interested linguist who had access to dBase IV, or to some compatible database management package, such as FoxPro, Clipper, or Paradox. Those few who used packages that did not utilize the dBASE IV file format were typically equipped with data and program conversion utilities for turning dBASE files into something they could use.

In order to improve the look and performance of the dictionary, considerable effort was spent in the late 80s and early 90s on developing a new interface program in FoxPro for Windows. This mouse-and-menu-driven program permitted a user to find words in the corpus according to various criteria of etymology, shape, or meaning, to see what words it is cognate with in its immediate sub-family (e.g., East Cushitic, Agaw[= Central Cushitic]), and to find and compare any claims about a word's wider cognacy relations (e.g., with words in other branches of Cushitic, or in other branches of Afroasiatic, such as Egyptian, Semitic, Berber, etc.) that had been proposed in the literature. It was also possible for a user to formulate new cognate sets and add them to the database on a permanent or temporary basis, to edit database entries, to add new entries, and to incorporate upgrades. Again, for portability and compatibility's sake, the interface program retained the dBASE IV format.

For more details, see the following documents:

- 1992-93 Annual Report
- 1991-92 Annual Report

From about 1994 two serious shortcomings in the Cushitic Lexicon approach became apparent. One was that the database programs and data formats were becoming obsolete. Another, more important, one was that programming and interface design were becoming white elephants. These also threatened to become even more of a burden as the Project, as a whole, began to look outward, beyond Cushitic and Omotic, toward the more general Afroasiatic language superfamily.

Fortunately, just when the project threatened to stall, a new approach to interface design emerged, based on HTML and the World-Wide Web. On the one hand, HTML offered a simple language for designing interfaces. On the other hand, the World-Wide Web provided a way of making those interfaces available to the public at large. After a short period of experimentation and study in early 1995, development exploded. To start with, the old materials in dBASE IV format were finalized, consolidated, and stored. Then a prototype Web interface was developed for etymological work on Semitic languages. An initial attempt was also made at translating the older Cushitic Lexicon data into a format suitable for use on the Web.

It was during this period that the project, which had moved well outside the Cushitic-Omotic realm, was re-named the Afroasiatic Index Project.

PRESENT STATE OF THE PROJECT

At present we hope to make once more available a prototype Semitic Index giving etymological relations within Semitic, but not yet to any branches of Afroasiatic outside Semitic, and follow up with a Cushitic Index. Since, as indicated, the Cushitic index already notes cognates in other branches, a rudimentary, Cushitic-oriented Egyptian, Berber and Chadic index will begin to take shape by default. It is our hope that linking to or embedding from data provided by competent outside sources will provide more adequate coverage in these areas.

In addition to the lexical and grammatical information provided by these modules of the project, we plan to include sample text browsing capabilities. A preview of these text-browsing capabilities can be seen in the approach developed in a sister site the Achaemenid Royal Inscriptions project.

EXCURSUS: WHAT IS ETYMOLOGY?

Etymology is the study of word histories. For example, our word for pig meat, *pork*, is actually an Old French word, *porc*--borrowed after the Norman conquest brought French culture to the British Isles. In turn, Old French *porc* itself developed out of an earlier Latin word, *porcus*. Our English word *pork*, therefore, comes down to us, through the intermediary of Old French, all the way from classical Latin.

The story does not stop here, however, for Latin *porcus* itself is related to another English word, *farrow* 'a litter of pigs.' But in this case the relationship is not that of a lender to a borrower (as with Old French and English *porc*) or of a parent to a child (as with Latin *porcus* and Old French *porc*). In reality, their relationship is more one of sibling to sibling. Latin *porcus* and English *farrow*, that is, stem from a common ancestor, conventionally designated as *Indo-European* (IE), a speech community which existed many millennia ago, and whose exact geographic location is still the subject of considerable debate. In IE, the root form was something like *porko-*. The original /p/ sound in IE, which remained /p/ in Latin, became /f/ in Germanic languages like English (compare Latin *pater* and English *father*). Likewise, the final /k/ sound, mutated the same way it did in, say, *arrow* (cf. Latin *arcus*). Through a series of such mutations,

called "sound changes" Indo-European *porco*- gradually evolved into *porcus* in Latin and *farrow* in English. As was mentioned above, Latin *porcus* later evolved into Old French *porc*, which turned into our modern English *pork*. *Farrow* and *pork*, then, are really just different historical derivatives of the same Indo-European form!

What we see in the history of *pork*, therefore, is a kind of recapitulation of how Latin turned into French, and of how French language and culture changed the development of English in medieval Britain. In the combined histories of *pork* and *farrow* we also see an example of how Latin, French, and English all evolved from a common parent, Indo-European. Indo-European itself no one has ever actually heard or seen written down. We can only infer its existence and nature by careful comparison of words in the various Indo-European languages--that is, from words like *farrow*, which differ in regular, often totally predictable, ways from words in other Indo-European languages (e.g., Latin *porcus*).

What makes etymology such an interminably fascinating pursuit is that *farrow*, *pork*, etc. are nothing unusual: every word has its own individual history, and the large-scale evolution of the language as a whole is the outcome of the interaction of countless such individual mutations. And this is, of course, true not only for words in Indo-European languages, but also for words in other language families, such as *Afroasiatic*.

Articles

- [1996 Article: ETYMOLOGY AND ELECTRONICS: THE AFROASIATIC INDEX](#)

Annual Reports

- [1993-94 Annual Report](#)
- [1992-93 Annual Report](#)
- [1991-92 Annual Report](#)

Revised: July 30, 2007

[Home](#) > [Research](#) > [Projects](#) > [The Afroasiatic Index Project](#)

[Visit Us](#) | [Contact Us](#) | [Get Involved](#) | [Calendar of Events](#) | [Order Online](#) | [What's New](#)

© 2008 The Oriental Institute | 1155 East 58th Street Chicago, IL 60637

Ph: (773) 702-9514 |