Resurrecting the Dead (Languages)
Documenting, archiving, and teaching at the Linguistics Research Center of the University of Texas
Hans C. Boas and Todd B. Krause
hcbo@utexas.edu, bobtodd@math.utexas.edu

Early Indo-European Online (EIEOL)
Each Indo-European language, regardless of its relation to other languages, is a unique entity with its own distinct characteristics, historical evolution, and textual tradition. The Early Indo-European Online (EIEOL) collection seeks to highlight the diversity exhibited by the Indo-European family. EIEOL supplies structural overviews of major representatives from each branch of the linguistic family tree.

Structure and Function
Structural overviews comprise series of 5–10 lessons for each language.

Lesson Design
Each lesson conforms to a prescribed format:
- Introductions/background information on culture, language, and root
- Reading: a textual selection with word-by-word grammatical analysis and contextual gloss
- Grammar: a 5-point explanation of important grammatical features

The EIEOL Format
All language scholars maintain their own preferences as to how best to make progress in learning a language. The EIEOL format strikes a balance between the needs of the specialist and of the non-specialist. A linguist within the subdiscipline requires an introduction to the features a given language shares with its relatives. The same specialist may also need a succinct overview as to what other disciplines, such as anthropology and history, have contributed to defining the context surrounding the language and its speakers.

Non-specialists, by contrast, may wish to avoid dedicating months to a university course. They may nevertheless need more than a superficial description of the language. The EIEOL format seeks to fulfill these needs in a format derived from World War II military training and designed to allow soldiers to begin translating intercepted messages in a minimum amount of time.

- Self-standing introductions to the languages
- Written with non-specialists in mind
- Introduce technical terminology and highlight linguistic issues as needed
- Provide bibliographies pointing to sources for further study

Lesson Processing
The LRC generates webpages automatically based on author content. Perl scripts read the ASCII files, substitute the betacode transcription with Unicode characters, and insert HTML formatting. This provides three main benefits:
- Design is separated from content
- Webpages can be recreated without modifying content files
- Authors need not learn HTML

The system is lightweight
- All content files are ASCII
- All processing files are Perl scripts: i.e. ASCII files

Data can be reorganized in different fashions
- Processing takes glossed roots to create glossaries of all surface forms, as well as all headwords
- English meanings are extracted to create English-to-[Target Language] glossaries

Specialists in Language, not HTML
The content of the various EIEOL language series derives from the work of specialists in the respective languages. The LRC has designed an infrastructure that requires minimal knowledge of computers and no knowledge of computer programming languages. This frees lesson writers to focus on the content, and it allows a minimal LRC staff to integrate that content seamlessly into the greater LRC ecosystem.

- All content is written in pure ASCII files
- Computer scripts automatically convert ASCII content to linked HTML pages

Lesson Writing
To address the issue of fonts differing among computers, lesson writers create a betacode map. This is a file that describes how the lesson writer will represent a character from the indigenous script (say the ashæœ, a Unicode symbol) in terms of an ASCII sequence (perhaps a@e) on the keyboard. Whenever the author wishes an ashæœ to appear, she writes a@e in the content file. The glossed readings in each lesson require a separate file with special formatting. The LRC stipulates some guidelines:
- Betacode sequences appear in lesson content between paired Ö-signs: e.g. aö@e
- In glossed readings, each word is entered on a separate line with a rigid format:
  - surface form @ analysis of <headword> @ meaning @ [contextual translation]

References
- Black, Carl Dating A Dictionary of Selected Synonyms in the Principal Indo-European Languages (1949)
- Pokorny, Julius Indogermanisches etymologisches Wörterbuch (1959)