

The Initiative for Cuneiform Encoding is an international group of cuneiformists, Unicode experts, software engineers, linguists, and font architects organized for the purpose of proposing a standard computer encoding for Sumero-Akkadian cuneiform, the world's oldest attested writing system.

The lack of a standard computer encoding for Sumero-Akkadian is a serious deterrent to the writing of computer software for processing and publishing cuneiform texts. ICE is an attempt to remedy this situation.

ICE is now part of the Digital Hammurabi Project, which just received \$ 1.6 million from the U.S. National Science Foundation to fund, over the next three years, basic research into 3D acquisition and rendering of cuneiform tablets. The first international ICE conference was held at Johns Hopkins University, Baltimore, Maryland, USA in November 2000; the next conference is scheduled for Spring 2003.



## INITIATIVE FOR CUNEIFORM ENCODING

### ANALYZING CUNEIFORM



There are two main phases in reading a cuneiform text - the descriptive and the interpretive.

In the descriptive phase one establish precisely what cuneiform signs, or graphemes, are actually written, i.e. one "transcodes" the text from one script system to another.

In the interpretive phase one strives to make sense of the transcoding. This is a complex, iterative process which includes grapheme sequencing, transliteration, phonemic normalization, and translation.

**DESCRIPTIVE**

	URU		NI
	KI		NI
	SU		IŠ
	NU		IŠ

**INTERPRETIVE**

NI NI IŠ URU KI SU NU  
 ḫ - lí - iš URU.KI - šu<sub>1</sub> - nu  
 iliš ḫlišunu  
 as the god of their city

Bassetki Statue - life-size, copper statue of king Naram-Sin, Old Akkadian inscription, ca. 2270 B.C. (one of the earliest explicit references to divine kingship) *Sumer* 1976, v32, p75a

This is what needs encoding

Sumero-Akkadian cuneiform, attested by hundreds of thousands of documents in many genres and several languages from various cultures spanning three millennia, is a complex syllabographic and logographic script system with perhaps a couple thousand distinct graphemes (characters). It is marked by extensive multi-valency – one grapheme can have multiple phonemic and semantic realizations.

To this day cuneiform lacks a standard computer encoding. The general practice among cuneiformists of working almost exclusively in Roman alphabetic transliteration, although adequate for its intended purposes, is a hindrance to the pervasive use of computers in cuneiform research and instruction.

The simple addition of graphemically encoded cuneiform to the current practice of transliteration would enable a dramatic increase in philological and linguistic research productivity.

For example, with cuneiform encoded:

- One could easily search for Sumero-Akkadian text in a mixed script environment, something that is practically impossible in transliteration – one could search an electronic Chicago Assyrian Dictionary, or the web, for cuneiform text.
- One could do context-free text processing of graphemically encoded texts, such as automated character recognition (cuneiform OCR) and proximity analysis of grapheme patterns.
- Plain text cuneiform would be a standard part of computer operating systems, and font architects would have much greater incentive to create the many large and complex font sets needed for rendering cuneiform usefully. And with adequate fonts the visual element of the script system would be more integral to its teaching.