

HUMANITIES AND THE ARTS

Musings

A new standard for cuneiform

If you are a scholar working in cuneiform and you want to send e-mail that includes this ancient writing system, for the most part you're out of luck. If you want to apply computer search capabilities to locate cuneiform in a collection of multilingual texts, again you're out of luck, unless you can make use of transliterated texts.

No standard electronic archive of the original cuneiform exists because no standard exists for reproducing on a computer cuneiform graphemes--the familiar characters originally impressed by a reed on clay tablets. Dean Snyder, senior information technology specialist at Hopkins, is working hard to remedy the situation.

Snyder is also a philologist who trained in comparative Semitics at the Oriental Institute of the University of Chicago. He has founded the Initiative for Cuneiform Encoding, an international group of cuneiformists, computer software engineers, linguists, and other specialists who want to establish a global standard for the computer encoding of Sumero-Akkadian cuneiform, the oldest known form of writing. The ICE group, which includes as faculty sponsor Jerrold S. Cooper, professor of Assyriology in Hopkins's [Department of Near Eastern Studies](#), held its first conference late last year at Homewood.

► A section of neo-Babylonian clay brick, held in Johns Hopkins's Archaeological Collection. When translated, the computer-encoded transcription (in black) means "Nebuchadnezzar, king of Babylon."



Says Cooper, "Since there are over 100,000 unread cuneiform tablets in museums and collections, and a relatively tiny number of trained cuneiformists working on them, this could be a great boon in making these unpublished tablets available to the scholarly community."

Computers store characters as numbers, according to a standard known as ASCII. But standard ASCII accommodates only 128 characters--"not even enough for the Western Romance languages," says Snyder, not to mention Russian, Greek, Hebrew, Japanese, etc. For these languages, scholars need additional proprietary encoding systems, which will work only on computer operating systems capable of recognizing them. Furthermore, separate

encoding systems frequently assign the same number to two different characters.

Enter Unicode, a new encoding standard under development by a non-profit consortium. Unicode has the capacity to assign a unique code number to every character in every language of the world. Every major computing platform has adopted Unicode as the foundation for how it handles text. "Unicode is ASCII on steroids," Snyder says.

The Unicode Consortium is in the process of assigning code numbers to the characters of various world script systems. (Someone, according to Snyder, even proposed Klingon, presumably as comic relief.) ICE hopes first of all to enlist the world's 200 cuneiformists in determining the inventory of cuneiform graphemes and assigning to them Unicode numbers. This also involves deciding how to handle things like compound signs, and signs that over a period of time merged or split. Cuneiform was used for 3,000 years by several language groups, including the Sumerians, Akkadians, Eblaites, Hittites, Elamites, and Hurrians. It changed from century to century and place to place.

Once cuneiform has been encoded, ancient tablets can be scanned, and the image processed by optical character recognition software to create a rough transliteration that scholars can fine-tune.

Researchers will be able to transmit cuneiform by e-mail and to archived digital texts. The massive Chicago Assyrian Dictionary, in progress at the University of Chicago for more than 60 years, can be put online, or on CD-ROM in an international standard format.

Since Unicode is a widely adopted standard, not a piece of proprietary technology, it is more likely to endure.

Says Snyder, "These cuneiform tables have survived 5,000 years. We don't want data systems that are gone in 10." --DK

▶ [Return to June 2001 Table of Contents](#)