

# JOHNS HOPKINS

SEPTEMBER 2003

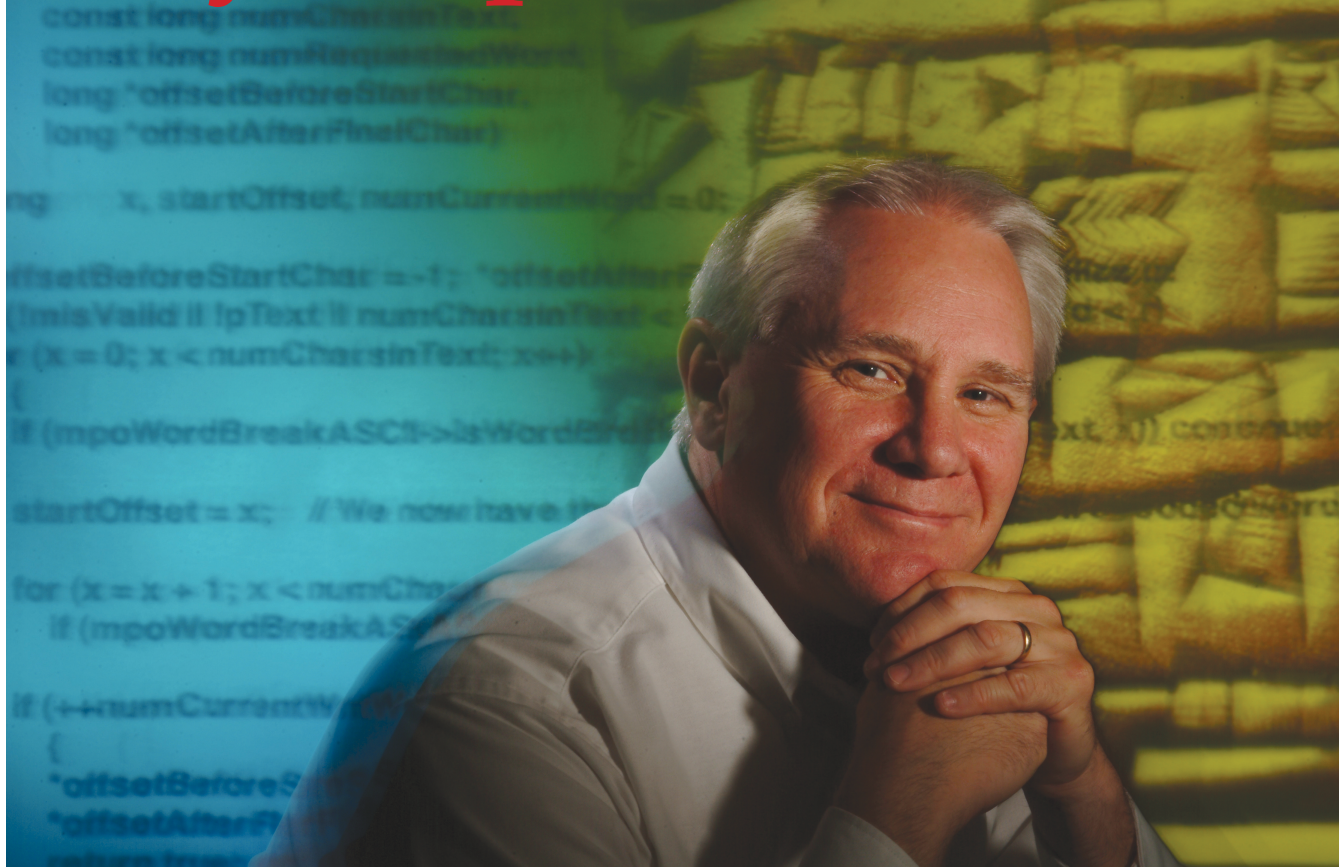
MAGAZINE



Cuneiform & Code:  
Ancient Writing Enters  
the Digital Age



# Clay, Paper, Code



It was file cards—more than a million of them—that inspired Dean Snyder's mission: to apply the newest technology to the world's oldest written language.

**BY DALE KEIGER**  
PHOTOS BY MIKE CIESIELSKI

Dean Snyder can read this:



It's cuneiform, the writing of ancient Mesopotamia. Transliterated into Babylonian syllables, it says e-nu-ma

e-lish, or the title of the Babylonian creation epic, Enuma elish, which begins, "When above heaven was not yet named..." Snyder can also read this:

```
#include <iostream.h>
int main ()
{
    cout << "That was cuneiform. ";
    cout << "This is not.";
    return 0;
}
```

It's computer code, written in a programming language called C++. Snyder is polyglot whether he's talking to people or machines. As a programmer, he can work in several languages: the relatively antique HyperTalk and Pascal, or the more current C and C++. As a scholar of ancient Semitic languages, he reads Hebrew, Aramaic, Phoenician, Moabite, Ugaritic, Syriac, Akkadian, and limited Arabic, as well as Latin and ancient Greek. As manager of the Johns Hopkins Digital Hammurabi project, he works where modern code and ancient languages meet.

Digital Hammurabi seeks to apply the newest digital technology to the world's oldest writing system. Hundreds of thousands of hardened clay cuneiform tablets rest in museums, libraries, and universities around the world.

More than half have never been studied or even read by scholars. Copying, deciphering, and publishing their contents, and checking that work, is slow and difficult because among 6.3 billion people in the world, maybe 300 read cuneiform, and their expert work is not the relatively straightforward process of, say, taking a paragraph of modern French and rendering it in English. Speakers of a half-dozen languages adopted cuneiform, and during its 3,000 years of use, it changed. Signs changed shape, changed direction, combined, in some cases split apart. Cuneiform was multivalent; that is, depending on the context, the same sign might represent one of several syllables, or it might represent a complete word, or it might be an unpronounced graphic sign classifying the name that followed as that of a god. Studying a single tablet can be a painstaking, time-consuming process. And to study a tablet, a cuneiformist has to actually go to Baghdad or Berlin or London or Chicago, wherever the tablet resides, to examine it firsthand.

Funded by \$1.65 million from the National Science Foundation, Digital Hammurabi aims to create electronic archives of detailed, three-dimensional images of cuneiform tablets. Secondly, work supported by the grant may eventually enable scholars to write cuneiform on a computer, something currently problematic because there is no standard computer encoding of cuneiform signs. If Digital Hammurabi succeeds, a cuneiformist will be able to download data files that will recreate, on a desktop computer monitor, an extraordinarily detailed three-dimensional

image of a tablet. Whatever scholarly examination the researcher needs to do — turn the tablet to examine every side, magnify a section of it, change the angle of illumination — will be done from a keyboard and mouse.

Digital archives would not only expedite decipherment of unread cuneiform by making images of the tablets accessible anytime from anywhere. The archives would also allow scholars to assemble personal collections of virtual tablets for comparative or concentrated study: an archive of legal documents



**Cuneiform tablets are difficult to scan because they often contain writing on many sides.**

from a specific city-state, for example, or the diplomatic correspondence from a designated span of time. Finally, Digital Hammurabi could create a permanent record of ancient Mesopotamian texts, a task that has gained urgency in the wake of looters ransacking Iraq's museums and, especially, archaeological sites after the recent war.

The project involves scholars, engineers, and information technology specialists throughout Hopkins. The engine behind it is Snyder, an itinerant senior IT specialist who five years ago rode into Baltimore on a motorcycle looking for work. For 20 years, he has been working at the interface of modern silicon and ancient clay.

About 10,000 years ago, settlers on the fertile alluvial land between the Tigris and Euphrates rivers began using little clay tokens, perhaps to keep track of sheep. Thousands of these tokens have been excavated. Many bear marks, simple Xs or parallel lines incised in the clay when it was still wet.

By 3300 B.C., ancient Mesopotamians at Uruk, a city in what is now southeastern Iraq, had developed complex social and political structures that required a more elaborate system of accountancy. They inscribed rectangular clay tablets with numerical systems and pictograms for commodities such as barley, oxen, and fish. During the next 500 years, people figured out that if these pictograms could be used to represent syllables of spoken language, then anything that could be said with the mouth could be recorded with the hand. They were no longer limited to merely keeping track of objects. They could express the abstract, the conceptual.

Century by century, the little pictures first used in Uruk became less complex and more abstract, until Mesopotamians had developed a system of about 1,000 signs that could be pressed into moist clay with the tip of a reed. Clay proved ideal as a medium. It preserved fine lines, so scribes could use small signs and fit a lot of text in a small space. If you had a text that you wanted to preserve, a library copy, for example, you could bake the tablet and set its contents in stone, so to speak. Most tablets were rectangular and sized to fit the palm of the hand; some still bear the lines and whorls of the scribe's palmprint. But cuneiform also appears on cylinders, monuments, stone walls, and six- or eight-sided prisms.

The content of the tablets translated to date includes diplomatic correspondence, historiography, mythology, religious rituals, legal matters, even propaganda. Yale University has a set of



tablets that appear to be recipes, calling for garlic, onions, leeks, and possibly mustard, cumin, and cypress berries. (There's some debate on the ingredients list.) A cautionary cuneiform note from 1740 B.C., displayed at a recent exhibit at New York's Metropolitan Museum of Art, reads, "You should not pass judgment when you drink beer." But roughly 80-90 percent of all deciphered tablets record transactions, inventories, that sort of thing. This has led several authors to attribute the invention of writing to the rise of commerce. Jerrold Cooper, Hopkins professor of Near Eastern studies and a participant in Digital Hammurabi, cites the rise of bureaucracy as a major impetus.

"If you have thousands of people whom you want to organize to dig a canal or build a huge temple," he says, "you have to feed them and keep track of them. So [the writing on many tablets] is about organizing and controlling and feeding large numbers of people. It wasn't private trading that led to the complexity that necessitated the invention of writing. Merchants probably kept their accounts in their heads, as they do in a lot of parts of the world. You can have illiterate commerce. You can't have illiterate bureaucracy."

People used cuneiform for a long time. Says Cooper, "That's the cool thing. It started around 3300 B.C., and the latest dated text we have is from around 75 A.D. They continued to write Sumerian when no one spoke it" — it was superseded by Aramaic — "and that's interesting. It was a prestige language." The tablets have survived the last scribes who could read them by two millennia. Clay can chip and shatter, of course, but unlike leather or papyrus it doesn't rot and it doesn't burn. When an invading army sacked a city, not an uncommon occurrence in ancient Mesopotamia, the fires set by invaders actually helped to preserve libraries and archives by hardening the tablets, much as a ceramicist fires a pot in a kiln. It is from these libraries that scholars have pieced together the epic of Gilgamesh and Enamum elish, tracked the diplomatic relations between city-states, and extracted from all that bureaucratic re-

cord-keeping the structure of ancient Mesopotamian commerce, details of diet and agriculture, and information about the construction of temples and public works.

**B**etween the eras of inscribed clay and the etched silicon computer chip, there was paper, and it was paper — specifically file cards — that motivated Dean Snyder to learn how to program a computer.

The Oriental Institute of the University of Chicago has been compiling the Chicago Assyrian Dictionary for 82 years. The heart of the project is more than 1.3 million file cards that bear words, definitions, and citations,



**Jerrold Cooper estimates that half of the world's trove of cuneiform tablets have never been read.**

and constitute the dictionary's database. For decades, scholars working on ancient Semitic languages at the institute have created their own personal banks of cards specific to their research. These cards eventually drove Snyder right out of the institute, where he was doing doctoral research.

His path to Chicago and ancient languages covered several states and was, he is convinced, divinely inspired. In 1969, he was serving in the Air Force when his mother became disabled. The military granted him a hardship



discharge, and he moved her to Tucson, Arizona, hoping it would help her arthritic condition. He recalls, "I was working in Kmart, in the camera department, and I didn't know what I wanted to do with my life." Snyder is a devout Christian, and he says, "I actually fasted and prayed for a week about what to do. And I heard four words, and I believe they were from God: 'Study Hebrew and Greek.'" He's a bit self-conscious talking about this now, but he took the words to heart and plotted a new life course. "If you want to really study ancient Hebrew," he says, "you have to study the other related languages — Akkadian, Arabic, Syriac, Ugaritic, Aramaic, Phoenician. So that's what I did."

First he returned to the University of Oklahoma, where he had done some undergraduate work prior to the Air Force, and started a new undergraduate degree in Hebrew and Greek. He finished those studies at Southern Connecticut State University, while doing independent linguistic research at Yale's Sterling Memorial Library. Finally, he entered the doctoral program at the Oriental Institute to work on comparative Semitics.

One of his interests at the institute was linguistic evidence of early Hebrew. Usually early Hebrew was written without vowels, but the Hexapla, a 3rd-century polyglot edition of the Hebrew Bible that is arranged in six columns of parallel texts in Hebrew and Greek, includes vowels. So Snyder embarked on a study of the Hexapla, looking, for example, at how all i-vowels following a guttural consonant are transcribed into Greek. That required starting his own collection of file cards, indexing the words found in the Hexapla's multiple texts. "In the late fall of 1982, I was working through a 9th-century manuscript of the Hexapla and had amassed 1,076 cards in three weeks. I sat back and realized that I had thousands of cards to go, and every time I wanted to look up something it would take hours, even days, of manually flipping through all the cards, one by one. That's when

I decided that this should be done on computers."

Snyder went to the University of Chicago's information technology department, where several specialists told him that no software existed to do the sort of multilingual work he needed. "They said, 'If you're going to do anything with this, you'll have to do it on your own.'" Suddenly, Snyder had a new mission, inspired not by the voice of God, but the counsel of some IT advisers. He left the doctoral program in 1983. "I just quit," he says. "I fell off the face of the graduate earth."

His mother became seriously ill again, requiring his help, so he worked in construction for a few years before buying the computer equipment he needed to begin programming the tools he envisioned. Snyder spent \$25,000 on all the best gear he could find, including an Apple Mac II. He learned a programming language, HyperTalk, and wrote some code to create concordances, indexes that note every appearance of a given word in a text. That program, which could concord 10,000 words per second, proved too slow, so he taught himself another language, Pascal, and wrote new code that, after some optimization, would concord 600,000 words per second. He learned yet another programming language, C, founded his own

chusetts to North Carolina (with a few airline trips to Texas and California), from one job interview to another. After 26,000 miles, he accepted a position at Hopkins, because he wanted to be back at a university and Hopkins was the southernmost institution with a library adequate for his research in Hebrew and Greek — Snyder and his wife were tired of Chicago winters. In early 1999 he began technology support for the departments of Romance languages, history of science, and philosophy in the Krieger School of Arts and Sciences. On his own time, he kept thinking about computers and cuneiform. Silicon and clay.

**P**ick up four books about ancient Mesopotamia and you're likely to encounter four different estimates of how many tablets are available for study. Some go as high as 500,000. Cooper thinks that estimate's too big: "The problem is whether you're counting tablets or fragments. At Ebla, when they said they'd found 14,000 tablets, once they glued all the pieces together it was down to around 6,000 or 8,000." Though more keep turning up. "In the British Museum about 12 years ago, they were dismantling a case in which Sumerian objects were displayed, and in the base they found built into it crates

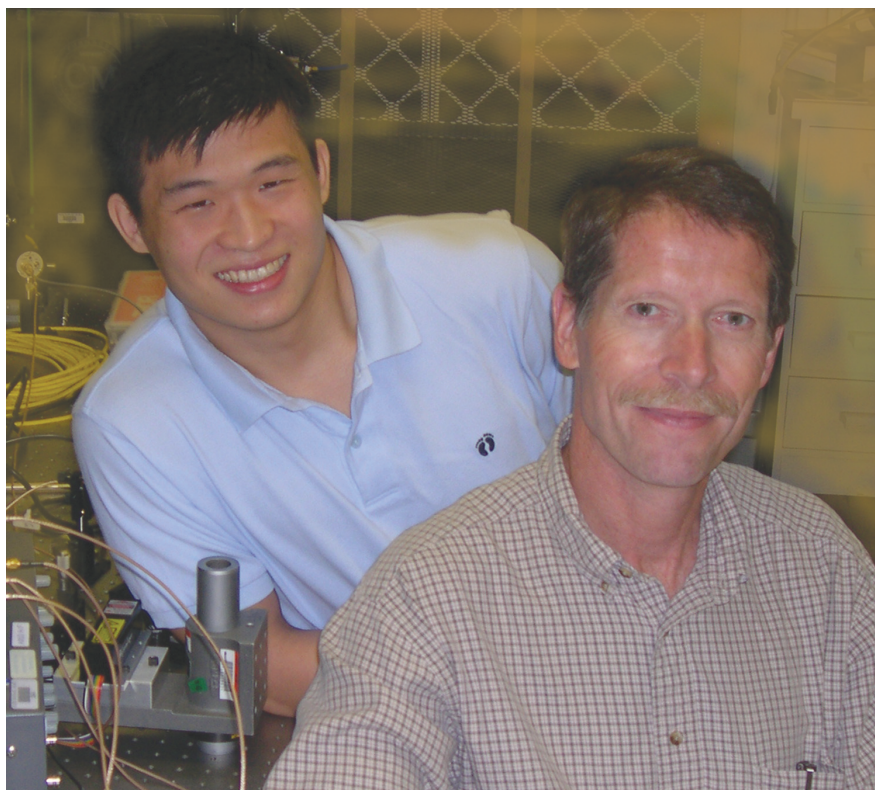
**"I had thousands of cards to go, and every time I wanted to look up something, it would take hours, even days," says Snyder. That's when he turned to computers.**

business, and began writing commercial software. After he got married in 1988, he took a job with a software company because he could make more money and would be, he thought, more secure.

In 1998, that company downsized, a word that seems inadequate: out of 250 employees, the company laid off 225, one of them Snyder. By now, though, he'd taught himself to be a skilled coder, and several universities and software firms vied for his services. For about five months, Snyder rode a Suzuki 850 motorcycle from Massa-

of tablets. Then they found more unopened crates in the basement." Cooper estimates that half of the world's trove has never been read.

Reading tablets requires expertise, patience, and the ability to minutely examine them. Almost none are flat like a slate or a piece of paper; instead, most are convex on at least one side. Even the flatter tablets usually have curved edges, and scribes routinely extended lines of cuneiform around these curves. This makes a two-dimensional representation from photographs or flat-bed scans



**APL's Donald Duncan (right) and Jason Liang '05 use laser light to create detailed, three-dimensional scans of cuneiform tablets.**

problematic, because you need multiple images of all the surfaces. Furthermore, a scholar cannot hold a two-dimensional image of a tablet at different angles to the light, to better discern faint or damaged inscriptions in the clay.

Snyder knew that for a digital archive of tablets to have much use, scans of those tablets would have to be detailed, three-dimensional, and in a form that could be manipulated on a computer screen in the same way a cuneiformist might turn a tablet in the light. The technology that makes those scans would have to be portable, so it could be taken to wherever tablets were stored or discovered, and fast, because there are so many to scan. Also, some way would have to be found to hold the tablets safely and manipulate them so that every surface was exposed to the scanner.

In late 1999, at this point still pursuing the project in his spare time because the National Science Foundation hadn't yet made its grant, Snyder approached the Hopkins Applied Physics Laboratory to see what sort of technology might be usable. He was referred to Donald Duncan, a member of APL's

principal research staff. For many years Duncan has been interested in non-destructive evaluation of materials. "This seemed like a natural fit," Duncan says. "In NDE, you're interested in measuring deformations of objects, for instance, using optical techniques." A cuneiform sign pressed into clay can be considered a deformation of the clay's smooth surface. Measure it accurately enough and you have the data you need to create a detailed, three-dimensional image. "With conventional imagery like a photograph," Duncan says, "you get information about the x and y axes. We need the z dimension, the range dimension, as well."

Intrigued, Duncan, like Snyder, began fooling around with the problem in his spare time, working with a high school student, Jason Liang, who made it his senior project. Liang, now a junior mathematics major at Hopkins, helped Duncan explore a technique called structured light. "Imagine that you project a

laser line on an object," Duncan says. "If the object is a plane and you look at it obliquely, that light forms a straight line. But if the object has any departure from a plane, look at it obliquely and you'll see displacement of that line." Accurately measure that displacement and you've begun to assemble the data you need to form a three-dimensional model.

Duncan, now funded by part of the NSF grant, continues to test various means of using laser light to scan cuneiform tablets. He has been evaluating existing technology that might be modified to create the scanner they need. "It's looking more and more like we'll have to adopt a number of different technologies and integrate them in a system that gives us what we need."

Whatever they use will surely produce massive data files. Jonathan Cohen and Subodh Kumar, Hopkins assistant professors of computer science at the Whiting School of Engineering, are working on ways to handle so many 0's and 1's. Says Cohen, "The initial data could easily be 100 gigabytes or so for one palm-sized tablet." For perspective, a 25-micro-meter laser scan of a cuneiform tablet no bigger than your hand — that is, a scan in which each dot represents 25/1000 of a millimeter of the tablet's surface — could produce a single file twice the capacity of the hard drives that come with many personal computers. "But there are lots of opportunities to compress that kind of data," Cohen adds. "There's a lot of redundancy." For example, a tablet may need to be scanned several times, each time with the laser set at a different angle, to create the three-dimensional image. Each pass of the scanner covers much of the same relatively flat surface, creating a large amount of redundant data. Cohen and Kumar are developing algorithms that can strip out that redundant data without losing any vital information. They also note that not every part of a tablet may need the same fine resolution. Damaged portions, or portions with faint markings, may require a scan with high resolution, while clearly marked or smooth sections of the same tablet may need only a low-res scan.



The software engineers don't have any 25-micrometer scans to play with yet, but they have been experimenting with lower resolution files. Last June, they demonstrated how a digital tablet could be displayed and manipulated on a computer monitor. They turned the tablet every which way, simulated changes in the angle of light to better see various details, and magnified a portion of it. Several cuneiformists, who were in town for a conference on encoding cuneiform, watched the demonstration, and they were impressed by the technology's potential.

Some of them seemed less convinced of the potential of the part of Digital Hammurabi that seems closest to Snyder's heart — creation of a standard cuneiform encoding that will allow someone to compose cuneiform text on a computer. Modern cuneiformists move smoothly from 5,000-year-old clay tablets to e-mail and Internet Web sites. But when they exchange messages about, say, ancient Akkadian or Ugaritic texts, they do so not in cuneiform but in sign-to-syllable transliteration: e-nu-ma e-lish la na-bu-ú shá-ma-mu. There are no standard means for them to tap a key on a computer and have the correct cuneiform sign appear on the monitor, in a document, or in e-mail.

Snyder thinks there needs to be. Computers don't know "d" from "K" in the way that any 6-year-old does. The computer simply catalogs letters and other characters by assigning numbers to them: 65 for "A," 97 for "a," 38 for "&." Strike the "a" key on your keyboard and you've told your computer to display what it regards as character 97. The list of these coded numbers has to be part of the computer's operating system for it to correctly display whatever you type. Years ago programmers adopted an encoding system called ASCII that could accommodate 256 letters, numbers, punctuation marks, and symbols like "\$" and "@." This worked for English, French, Spanish, German, and other languages that use the Roman al-



**Cuneiformists will be able to manipulate virtual tablets to simulate changes in the angle of the light.**

phabet. But what about Chinese, with its tens of thousands of characters? What about Arabic or Hebrew, which not only have their own characters but read from right to left?

The answer was Unicode, a numbering convention that can accommodate millions of different characters and has been adopted as standard by all the world's major computer operating systems and programming languages. A non-profit organization, the Unicode Consortium, governs the standard, approving new entries for additional scripts. As more scripts enter the Unicode system, writers can be more polyglot all the time, writing notes and scholarly papers in Urdu or Tamil, creating concordances and textual commentaries in Khmer, Mongolian, even Old Norse runes.

But not cuneiform, because it's not yet been prepared for Unicode. For a script to become part of the Unicode standard, someone has to create the definitive table of all that script's characters so that each can be matched with a Unicode number. English uses 26 letters, and no one lobbies for a list that includes three more. But eminent cuneiformists like Miguel Civil and Rykle Borger have spent years compiling differing sign lists for cuneiform, and those lists are still being revised. There's a corpus of characters that everyone accepts — without it, there'd be no translation of Gilgamesh (and there are now several). But what to do about the word dividers employed by Old Assyrian? Or punctuation? Or line breaks that are syntactically significant? Scholars who study lesser-known cuneiform

languages like Hurrian and Eblaite continue to find new values in those languages for existing signs. All of this must be decided because once elements enter the Unicode standard, they can't be removed. You have to get it right the first time.

In 2000, Snyder, again on his own time, organized the Initiative for Cuneiform Encoding (ICE) — a collaboration of cuneiformists, Unicode experts, software engineers, and font architects — to make cuneiform part of the Unicode standard. ICE held its first conference at Hopkins in November 2000, to discuss the theoretical and practical issues of encoding cuneiform. Since then, ICE has become part of the Digital Hammurabi project, and last May a second conference, ICE2, assembled at Shriver Hall. The conferees made substantial progress on a Unicode encoding, despite a fundamental divide on its ultimate usefulness.

Snyder believes that the Unicode standard will prove to be of great significance for cuneiform scholarship. He says, "I think there will be software tools based on this encoding that will dramatically increase productivity in the research into cuneiform." He envisions, for example, an online edition of the Chicago Assyrian Dictionary, which a scholar or student could search using cuneiform text. He imagines automated cuneiform optical character recognition, for computer-generated transliteration of cuneiform text. "You could automatically generate an index of all the words written in Sumero-Akkadian cuneiform in multilingual texts if cuneiform is encoded," he says. "You could do automated parsing, concordance generation, proximity searching for linguistic features. These tools do not exist for cu-



**Subodh Kumar (left) and Jonathon Cohen impressed cuneiformists with their digitized tablets.**

neiform now. They are difficult to write, and there's no incentive to write them when the base technology, the Unicode encoding, isn't there."

Jerrold Cooper has a different perspective: "The only reason I'm doing this Unicode thing is because if it's not done by cuneiformists, it's going to be done by enthusiasts, amateurs, and I didn't want that to happen. If they're re-

ally going to establish a standard that's going to be built into software and operating systems, it should be something that cuneiformists will respect rather than laugh at. But no cuneiformist that I know can actually envisage a use for the Unicode standard in our work."

Once a tablet has been copied and deciphered, scholars working with its text usually don't bother with the cunei-

form anymore; they simply use transliteration. That, says Cooper, is adequate. He doesn't need to be able to write cuneiform e-mail. "The computer people all assure us that there are things that we can't imagine now that we'll be able to do," he says. "I think it boils down to people who aren't working cuneiformists who think it will be wonderful to do a whole series of things that actual cuneiformists aren't eager to do." He believes that some of Snyder's ideas, like cuneiform OCR, would involve such complexity it's doubtful computers could effectively perform the necessary tasks. "It's not hard to teach a computer to recognize signs. But because of the multivalence of cuneiform, you need more than recognition. Once the sign is identified, the proper reading has to be chosen. The algorithms for making such choices would be very hard to write."

But Cooper admits he came out of ICE2 with a more positive attitude toward the Unicode project. "I saw how it might be done, and because the younger cuneiformists in our group are enthusiastic, I feel less skeptical about the whole venture. Dean's much more a visionary than I am."

Snyder's vision — paper file cards and clay tablets become electronic 0's and 1's, the oldest surviving texts living on as bits and bytes on hard drives — is becoming manifest. And he's looking well past what he imagined 20 years ago: "Scholars, at their desks, will be able to instantly navigate between two-dimensional images of Aramaic papyri and Hebrew ostraca and three-dimensional images of neo-Babylonian cuneiform tablets. Scholars will print out three-dimensional plastic replicas of pertinent tablets for their students to work on. More students will study ancient languages, and we will see greater productivity in research and publication. And we will throw away our index cards, once and for all." ■