



John G. Dvorak

## Kiss Your ASCII Goodbye

**B**usiness and communications are going international, but the PC hesitates to cross the language barrier. Though you really don't want to type in any character set except your native language, what if you want output in Icelandic or Arabic? For

intercultural communications, ASCII just doesn't cut it. The solution is Unicode, and it's long overdue.

It was my experimenting with translation software that brought the problem to my attention. I write for the French and German counterparts of *PC Magazine* and decided to send them translated documents. Because of character set differences, I couldn't send the documents electronically. Even printing and faxing were problems. Too many characters were seen as control codes instead of Latin-1 characters (an ASCII superset that includes various symbols plus European letters). Sure, you can struggle with binary files and graphics instead of character sets. But what a hassle!

An easier and more universal solution is the Unicode standard. Within the next year or two you can expect worldwide adoption of this 16-bit character standard to replace 7- and 8-bit ASCII. While ASCII has 128 characters and Latin-1 has 256 characters, Unicode will open the door to 65,536.

It took an international battle to get Unicode accepted. Back in the heyday of Xerox PARC in the early 1980s, Xerox was promoting an international character set, but some members of the computing community decided they could do better. So an ISO (International Standards Organization) committee was formed to develop a 32-bit character set of all known characters and symbols. The committee, ISO 10646, instantly got bogged down in politics.

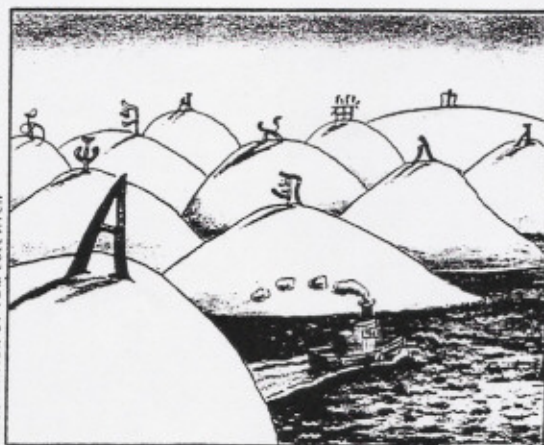
Meanwhile, a group of multilingual system implementers at Xerox,

including Joe Becker, Lee Collins (now at Taligent), and Eric Mader and Dave Opstad (both now at Apple), were already thinking about Unicode. (Becker coined the term.) They began with an ambitious project to unify the commonalities of Asian character sets, ultimately persuading the Chinese, Japanese, and Koreans to take joint ownership of this effort. Inspired by Mark Davis (from Apple's international division, and now at Taligent), they broadened participation to leading industry representatives, including Bill English (Sun Microsystems), Asmus Freytag (Microsoft), Mike Kernigan (Metaphor Computer Systems), Rick McGowan (NeXT), Isai Scheinberg (IBM), Karen Smith-Yoshimura (Research Libraries Group), Ken Whistler (a UC Berkeley linguist now at Metaphor), and many others.

Finally, a breakthrough compromise allowed Unicode to be a subset of ISO 10646, and ASCII and Latin-1 to be subsets of Unicode. ISO 10646 was passed by the member countries in June.

I don't know how ISO 10646 will ever fill up its over 4 trillion character slots, but until it does, Unicode will supply all the characters needed for the languages of the modern world—28,706 characters so far.

Unicode is published as *The Unicode Standard, Worldwide Character Encoding* (ISBN: 0-201-56788-1), \$32.95 from Addison-Wesley, 800-447-2226. Unicode Inc. provides tapes and seminars. You can contact Unicode at 1965 Charleston Rd., Mountain View, CA 94043; 415-961-4189.



*Unicode supports all the languages of the modern world. It may finally break the PC's language block.*