

Reviews

Technology with characters

Andrew Robinson reviews *Kingdom of Characters: a Tale of Language, Obsession and Genius in Modern China* by Jing Tsu



Shutterstock/CODEANT STUDIOS

Key issue

There are tens of thousands of Chinese characters, which creates a challenge for technologies developed based on Western alphabets.

Kingdom of Characters: a Tale of Language, Obsession, and Genius in Modern China

Jing Tsu

2022 Allen Lane
336pp £20.00hb

Chinese characters, in use by the second millennium BCE or even earlier, have functioned longer than any other script, including Egyptian hieroglyphs. Yet they have tended to isolate China, including its science, from other cultures because of their baffling complexity, especially compared to alphabets. To quote a celebrated early 20th-century Chinese writer, Lu Xun: “If the Chinese script does not go, China will certainly perish!”

This quote opens *Kingdom of Characters: a Tale of Language, Obsession and Genius in Modern China* by Jing Tsu, a sinologist at Yale University who was born in Taiwan and educated in the US. Her pioneering, fascinating, if often demanding book tells the story of China’s unique transformation over the past century or so. As its clever title suggests, Tsu discusses not only written characters but also many characterful people from the Chinese worlds of computing, librarianship, politics, science and technology behind the transformation – most of whom are relatively unknown outside specialist circles.

They include Zhi Bingyi, who earned a PhD in physics from Leipzig University, returned to China in 1946 and led a distinguished career as an engineer, before being thrown into prison during the 1960s Cultural Revolution as a supposedly “reactionary academic authority”. Living in a cowshed, deprived of even toilet paper, with only a stolen pen and the lid of a ceramic teacup as a wipeable writing surface, he invented a way of inputting Chinese characters into computers by mapping them onto an alphabetic code. This he did while contemplating eight characters on the cowshed’s wall meaning “Leniency to those who confess, severity to those who refuse.” After his eventual release, and a period as a floor-sweeper, toolmaker in a factory and warehouse guard, his breakthrough was hailed in 1978 on the front page of a Shanghai newspaper with the comment: “The Chinese script has entered the computing machine.”

The technologies discussed in the book include the typewriter, telegraph, librarian’s catalogue and digital computer, each of which receives a chapter. As Tsu sums up:

“Every technology that has ever confronted the Chinese script, or challenged it, also had to bow before it. Ideographic characters have pushed to the brink every universalist claim of Western technology, from telegraphy to Unicode [a standard international system for encoding various languages’ scripts in computers]. Having bent over backward many times to accommodate the technologies of the Western alphabet, the Chinese script, however, has not been altered in a fundamental way. Having survived, its presence has only been strengthened by those trials.”

A century ago, it was by no means clear that this triumph would occur. In 1936, in his first interview with a Western journalist, the leader of the Communist rebels in China, Mao Tse-Tung (as he was then spelt in English), said: “We believe Latinization is a good instrument with which to overcome illiteracy. Chinese characters are so difficult to learn that even the best system of rudimentary characters, or simplified teaching, does not equip the people with a really efficient and rich vocabulary. *Sooner or later, we believe, we will have to abandon the Chinese character altogether if we are to create a new social culture in which the masses fully participate* [emphasis as per original source].”

In practice, Mao – himself a lifelong calligrapher – encountered so much opposition from Chinese intellectuals that he compromised. In 1955, six years after the People’s Republic was founded, Chinese characters were officially simplified by eliminating certain variants and reducing the number of strokes in many of those remaining – a process that continues to this day. And in 1958, the government introduced romanized Chinese script, known as Pinyin (“spell sound”), as the official system for writing Chinese sounds (including tones) and for transcribing characters, so that non-Chinese

Having bent over backward many times to accommodate the technologies of the Western alphabet, the Chinese script, however, has not been altered in a fundamental way



“The Chinese Script Has Entered the Computing Machine” Wenhui Bao, 19 July 1978

Compatible code In the 1960s and 1970s physicist and engineer Zhi Bingyi mapped Chinese characters to an alphabetic code for use in computing.

speakers could roughly pronounce the language. In the Pinyin spelling of names (which does not indicate tones), Mao Tse-Tung became Mao Zedong, Peking turned into Beijing and Canton into Guangzhou.

Although Pinyin was opposed during the Cultural Revolution, when xenophobic Red Guards tore down street signs in Pinyin as evidence of China kow-towing to foreigners, it caught on. In today’s China, the world’s second largest economy, many millions of Chinese computers and smartphones, and the Chinese app WeChat – founded in 2011 and now with over one billion monthly active users – use both characters and Pinyin. As global users type in Pinyin, an array of characters appears on screen, anticipating the sentence or phrase they are composing. Indeed, Pinyin is so popular that some younger Chinese people no longer bother to learn characters the hard way; they have, to some extent, become “Latinized”, in Mao’s sense.

The issue of making Chinese script compatible with technology predates today’s computers. Consider Morse-code telegraphy, which was introduced into China by various foreign companies. In 1870, at the request of the Great Northern Telegraph Company, Danish astronomer Hans Schjellerup, who had taught himself Chinese, started work on a telegraphic code for Chinese. He compiled a proto-list of 260

characters, but then had to return to his observatory work.

A French harbour captain in Shanghai, Septime Viguier, carried on Schjellerup’s work without, apparently, much knowledge of Chinese. Soon he produced tables of characters in 20 rows and 10 columns, with each character assigned a four-digit code from 0001 to 9999, leaving empty spaces for perhaps 3000 further codes. However, the link between character and code number was completely arbitrary. Viguier’s code offered no information about the character’s shape, meaning or sound, to the irritation of native speakers.

Moreover, the Morse-code transmission of numerical digits (rather than the 26 alphabetic letters) was slow and expensive for the customer because numerals required more dots and dashes than letters. For example, you need one dot to send an “e”, the commonest English letter, but six dots to transmit the number “6”. This caused much resentment in China, not to speak of unreasonable profits for foreign companies, over the next half-century.

Only in 1929 was Chinese telegraphy made more rational by Chinese government approval of a scheme created by scientifically educated Wang Jingchun, managing director of an important railway. He devised a phonetic code for each character – three decades before Pinyin – con-

sisting of four alphabetic letters: one letter for its sound, another for its tone and two more for the spelling of the character’s component known as its “radical”. Now Chinese could be telegraphed without using numerals, more efficiently and cheaply. Nevertheless, notes Tsu, Viguier’s four-digit format “remained in use internationally and within China well into the 1980s”.

Such historic international struggles encouraged China’s newly confident 21st-century government to advance in computing without relying overmuch on foreign – especially American – technology. This determination has led to long-running international arguments about the Chinese characters in Unicode, especially with Taiwan’s computer scientists, who are generally opposed to China’s character simplification.

By 2020 there were 92856 Han Chinese characters listed in Unicode (version 13.0). This total could yet increase dramatically, given that several hundred thousand unlisted characters are being extracted and collected from old records – depending on whether China can persuade the Unicode Consortium, based in California. What an astonishing contrast with the character-doubling China of less than a century ago.

Andrew Robinson is the author of *Writing and Script: a Very Short Introduction*, published by Oxford University Press