QWERTY in China: Chinese Computing and the Radical Alphabet

Thomas S. Mullaney

Technology and Culture, Volume 59, Number 4 Supplement, October 2018,
pp. S34-S65 (Article)

Published by Johns Hopkins University Press
DOI: https://doi.org/10.1353/tech.2018.0149

QWERTY in China

Chinese Computing and the Radical Alphabet

THOMAS S. MULLANEY

ABSTRACT: Since the late 1980s, computers throughout the Sinophone world have featured QWERTY keyboards, employing input techniques that rely upon the Latin alphabet. In this article, I argue that historians of modern China and modern information technology alike have profoundly misunderstood China's QWERTY keyboard and oversimplified the history of China's engagement with the Latin alphabet during the nineteenth, twentieth, and twenty-first centuries. Scholars have been too quick to fixate on the narrower issue of phoneticization: that is, of attempts to re-inscribe Chinese by creating Latin alphabet-based writing systems that rewire the circuitry of Chinese linguistic signification with the goal of bypassing (and ultimately abolishing) Chinese characters altogether. The historical record alerts us to a much broader history of "Chinese alphabets," however. Based upon three cases, this article explores some of the many schemes in which the goal was to alphabetize Chinese, while also leaving character-based Chinese writing intact.

Ever since the mass manufacture of keyboard typewriters began in the United States at the turn of the century, engineers and entrepreneurs imagined a day when this new device would conquer Chinese, just as it had practically every other writing system in the world. It never did, with Remington, Underwood, Hammond, and IBM failing to enter the Chinese market.[1] This dream was rekindled in the age of computing and, beginning in the 1990s, seemed at last to be coming true: computers throughout the Sinophone world began to look "just like the real thing," even including the

Thomas S. Mullaney is associate professor of Chinese history at Stanford University. The author wishes to thank Suzanne Moon, Chiara Vernari, and anonymous reviewers for their insightful comments on early versions of this article. Gratitude is also extended to all participants and supporters of the 2017 conference "Shift CTRL: New Perspectives on Computing and New Media," which took place at Stanford University.

0040-165X/18/5904-0003/S34–S65

1. For more on the history of the Chinese typewriter, see Thomas S. Mullaney, *The Chinese Typewriter*.

familiar QWERTY keyboard, which by today is ubiquitous across all East Asian IT markets. During this same decade, moreover, Chinese keyboarding methods came to rely more and more heavily on Hanyu Pinyin, a Latin alphabet–based phoneticization system for Chinese characters first promulgated decades earlier. Designed by Chinese linguists shortly after the Communist revolution of 1949, pinyin steadily became ubiquitous in mainland China, functioning as a paratextual technology that runs alongside and supports character-based Chinese writing. When Chinese toddlers first learn to read and write Chinese characters, for example, they learn pinyin at the outset in order to assist them with the memorization of standard, nondialect pronunciation. More centrally for the present discussion, when computer users in mainland China sit down at their laptops, the keyboards they use are of the standard QWERTY variety, and the input methods they employ are in large part premised on phoneticized Hanyu Pinyin input. Although Chinese characters are still with us, then, it would seem that the computer—and perhaps more specifically, the QWERTY keyboard—has finally begun its conquest of Chinese, the last major nonalphabetic holdout in a world otherwise dominated by alphabet-centered information technologies.

When we place the Chinese QWERTY keyboard in broader historical perspective, this seeming conquest assumes even greater importance. Chinese is one of the most widely spoken languages in the world. It is also the world's only major non-alphabetic writing system, employing neither the letters of the Latin alphabet nor any comparable system to represent phonemes, as in Hebrew, Greek, Arabic, or Cyrillic. Beginning in the nineteenth century, however, dozens of "Chinese alphabets" were proposed, first by foreign Christian missionaries for the purposes of proselytization and conversion, and later by Chinese reformist elites for the purposes of state-led modernization.[2] Regarding Chinese characters as inhibitors to mass literacy and thus spiritual engagement, missionaries often aligned their phoneticization proposals with broader critiques of character-based Chinese script, with phonetic script offering up the promise of widespread non-elite literacy, along with the ability to produce textual materials for many millions of Chinese dialect speakers for whom no dedicated, dialect-specific scripts existed. Later, as Elizabeth Kaske, W. K. Cheng, and others have shown, a vocal subset of modernizing Chinese elites took up the banner of phoneticization in the late nineteenth and early twentieth centuries, with at least twenty-nine proposals for Chinese phoneticization advanced between the 1890s and 1911—many of which advocated the use of Latin alphabetic letters as a replacement for Chinese characters.[3] "If the roman script enabled the missionaries to reach the people not only efficiently but

2. T. H. Darlow, Horace Frederick Moule, and A. G. Jayne, *Historical Catalogue*.

3. Elizabeth Kaske, *The Politics of Language in Chinese Education, 1895–1919*; W. K. Cheng, "Enlightenment and Unity"; Jing Tsu, *Sound and Script*.

so economically," these elites reasoned, "what would happen if phoneticization served the purpose of popular enlightenment instead?"[4]

This question remained unanswered, as it were, insofar as each of these dozens of turn-of-the-century phoneticization efforts either failed or was never implemented. Character-based Chinese script was not displaced by an alphabetic orthography—much to the disappointment of missionaries, Chinese reformist elites, and corporations such as Remington, Underwood, Mergenthaler Linotype, Olivetti, and IBM. When viewed from this historical perspective, then, QWERTY's present-day conquest of Chinese is significant, not only within the confines of the history of computing, but also in its fulfillment of a dream over one century deferred. The phoneticization of Chinese—and thus the displacement of character-based script—has finally begun, thanks to the computer.

Closer examination reveals a more complex history, however. In this article, I argue that historians of modern China and modern information technology alike have profoundly misunderstood China's QWERTY keyboard and oversimplified the history of China's engagement with the Latin alphabet during the nineteenth, twentieth, and twenty-first centuries. When broaching the issue of Chinese alphabetization and romanization, scholars have been too quick to fixate on the much narrower issue of *phoneticization*: that is, of attempts to re-inscribe Chinese by creating Latin alphabet-based writing systems that rewire the circuitry of Chinese linguistic signification with the goal bypassing Chinese characters altogether. No doubt influenced by the history of abortive Chinese phoneticization schemes outlined above, there has emerged a deep-seated assumption, whether explicit or implicit, that alphabetization and phoneticization are effectively one and the same thing—and by extension that alphabetization schemes are by their very nature allied with efforts to replace and abolish character-based Chinese writing. To alphabetize or romanize Chinese, and yet to leave character-based Chinese writing intact, is regarded as a contradiction in terms.

There is nothing oxymoronic about the coexistence and coordination of Romanization and character-based Chinese script, however, for the simple reason that there is much more to the history of Chinese alphabetization than merely Chinese *phoneticization*. Based upon three cases—an experimental Chinese telegraph code developed in the 1920s by Wang Jingchun; the first Chinese computer, developed at MIT in the 1950s by Samuel Caldwell; and a Chinese character encoding system for computers and word processors developed in the 1970s by Chinese-American psychologist H. C. Tien—this article examines a variety of largely unexplored Chinese alphabetization proposals whose semiotic strategies involved min-

4. Cheng, "Enlightenment and Unity," 479.

ing and exploiting the capacities of the Latin alphabet (and alphabets in general) far beyond simply its utility as a technology of phonetic capture or transliteration. These alternate strategies were premised on the recognition that the Latin alphabet is "good for" much more than simply its ability to "sound out" Chinese words using the letters A through Z.

One strategy we will examine exploited the Latin alphabet as a repository of *variables*, wherein the letters "a" through "z" were used, not to represent the phonemic values of /a/ through /z/, but as arbitrary stand-ins for separate semiotic values altogether. Were we to imagine a hypothetical English-language example, this might be like rendering the word "nation" as "nax" wherein the first two letters "n-a" are used in the conventional phonetic mode as a representation of the ˈneɪ in ˈneɪʃən, but where the final letter "x" is used as a variable *placeholder* set equal to the sound-value conventionally rendered as "tion." In such as a system, the appearance of "x" within a sequence might serve either in its conventional phonetic capacity (as the letter "x" in "taxation," for example), or as a variable whose value is arbitrarily set equal to an altogether different phonetic value. While the use of letters as variables is, of course, a quotidian affair in mathematics, chemistry, physics, computer programming, and other domains, such strategies remain uncommon in the contexts of everyday writing. In such a system, the English language would continue to be rendered in the Latin alphabet, and yet it would exploit a broader set of semiotic possibilities than those offered up by the alphabet when we restrict our understanding of the alphabet as simply a phoneticization or transliterative technology.

Other Chinese alphabets we will encounter used the Latin alphabet in ways that will be less familiar to readers: the use of letters, neither as means to encode phonetic values nor as mathematical-like variables, but purely as graphical shapes exploited for their structural similarities to glyphs in other writing systems (in our case, Chinese). To draw an example from contemporary new media practice, and a somewhat insidious one at that, trans-orthographic isomorphisms have been to carry out so-called "homograph attacks" on the Internet, wherein users are duped into navigating to a URL that appears on their device as, perhaps, "HM.com" (i.e., a well-known clothing retailer). In such a homograph attack, however, the "H" in the website url is not, in fact, the Latin alphabetic letter "H" (/h/) but rather the Cyrillic letter "H" (/n/), thereby routing the user to an entirely separate destination. Other homographic pairs include 0 (the numeral) and O (the capital letter), B (Cyrillic letter /v/) and B (Latin letter /b/), P (Cyrillic letter /r/) and P (Latin letter /p/), among dozens of other examples.

As we will see, more than one experimental Chinese alphabet has exploited Latin-Chinese isomorphisms in precisely the same way, although on a far more extensive scale, and for non-malicious purposes. Rather than merely the occasional usage of such homographic relationships, a number of Chinese alphabets fashioned in the late twentieth century—particularly for the purposes of computational text input—have been developed through the systematic creation of dozens if not hundreds of homographic equivalences between alphabetic letters and Chinese characters. Once again, then, we are confronted with fully "alphabetic" modes of encoding Chinese words, yet modes in which the Latin alphabet is being used to achieve semiotic strategies that extend far beyond mere phonetic transliteration. Phrased more broadly, we encounter the Latin alphabet being used in ways that depart dramatically from the way in which it is used in English, French, Spanish, and other language contexts.

In narrating a more expansive history of Chinese alphabetization, the goal of this article is to lay the groundwork for a history of Chinese computing, in which the assumed synonymy of alphabetization and phoneticization is particularly prevalent. China's QWERTY keyboard, this article seeks to show, is not what most have assumed it to be.

The Letter as Radical: Wang Jingchun and Pinyin Telegraphy

In order to understand the history of China's engagement with the Latin alphabet within the context of computing, we must first delve into the history of Chinese alphabetization during the age of telegraphy. During the nineteenth century, the technology of electrical telegraphy expanded beyond the Latin alphabetic world.[5] As cables were laid from Suez to Aden and Bombay, and from Madras to Penang, Singapore, and Batavia, Morse Code came into contact with languages, scripts, alphabets, and syllabaries that it was not designed to handle. In 1871 in particular, telegraphic communication reached the shores of the Qing empire, with a single line opened between Shanghai and Hong Kong. Installed by two foreign companies—the Great Northern Telegraph Company of Denmark and the Eastern Extension A&C Telegraph Company of the United Kingdom—this line marked the first step in the construction of an empire-wide communications web, woven one filament at a time. With these modest beginnings, Morse Code came into direct contact with the world's one major non-alphabetic script: Chinese, which contains upwards of seventy thousand characters. How was one to design a variant of Morse Code without using letters?

These encounters raised profound questions for the trajectory of modern information technology. Would the inclusion of new scripts and writing systems prompt a radical reconceptualization of the global telegraphic system, or would these scripts be absorbed and subordinated within Morse

5. Daniel Headrick. *The Tentacles of Progress*, chap. 11; see also W. Bull. "A Short History of the Shanghai Station," Shanghai: NP, 1893 [Handwritten Manuscript], DOC/EEACTC/12/10, 4, in CAWA.

Code and its attendant, alphabet-centered semiotic logic and syntax (some more comfortably than others)? In the case of Chinese, the answer to this question came quickly.[6] In 1871, the first internationally recognized Chinese telegraph code was developed by a Danish professor of astronomy named H. C. F. C. Schjellerup, and formalized by Septime Auguste Viguier, a French harbormaster in Shanghai.[7] In the code, approximately 6,800 common usage Chinese characters were placed in standard dictionary order and then assigned distinct, four-digit numerical codes running from 0001 to 9999.[8] To transmit the character *shuo* (說 *to speak*) using this system, for instance, one would first translate it into the cipher "5356"—the number assigned to this character in the code book—and then use standard Morse Code sequences to transmit this numeric code. Upon receipt of the transmission, the recipient would transcribe a sequence of four-digit codes, and then use a corresponding code book to translate these numeric codes back to the intended Chinese-character message.[9] Phrased differently, Chinese became perhaps the only major world language that could not participate in Morse Code the same way that other languages did, namely by translating messages directly into dots and dashes. Instead, the 1871 code relegated Chinese telegraphy to a condition of double-encryption and double-translation: an initial translation from Chinese characters to numerals, and only then from numerals to Morse Code pulse patterns.

Because of its reliance on numerical notation, the Chinese telegraph code of 1871 had profound implications for China's starting point within the emerging global telegraphic infrastructure. First, numerals were by far

6. A few fleeting-yet-fascinating ruminations cropped up during this period which explored how the semiotic structure of telegraphy might be entirely reimagined. These ruminations were in large part inspired by the puzzle and challenge of creating a telegraphic code for Chinese. For an extended exploration of the most radical of these, put forth by Pierre-Henri Stanislas Escayrac de Lauture (1826–68), see Mullaney, *The Chinese Typewriter*, chap. 2.

7. For an in-depth discussion of the 1871 Chinese telegraph code, see Thomas S. Mullaney, "Semiotic Sovereignty"; and Erik Baark, *Lightning Wires*.

8. Approximately 3,000 blank spaces were left at the end of the code book, and a few blank spaces left within each radical-class, so that individual operators could include characters that were essential for their work but infrequently used by most other people. Septime Auguste Viguier (Weijiye [威基謁]), *Dianbao xinshu* [電報新書] (Guangxu 18), "Extension Selskabet—Kinesisk Telegrafordbog," 1871, Arkiv nr. 10.619, "Love og vedtægter med anordninger," GN Store Nord A/S SN China and Japan Extension Telegraf, in DNA.

9. Viguier's original list of characters was later slightly adjusted by De Mingzai (德明在), attaché to the Qing mission to France in 1871, who detected flaws, he explained, in the Frenchman's stroke-count sequence. Following the 1949 revolution, we witness the further creation of two versions of the code, one in the mainland and one in Taiwan, which both made use of four-digit codes but with different code assignments. Even when accounting for these changes, we find that the basic model of Viguier's system became the industry standard in China for over a century.

the most expensive kinds of symbols within Morse Code, with the shortest code sequence being five short "dots" (to transmit the numeral 5), and the longest being five long "dashes" (for the numeral 0). Insofar as the Chinese telegraph code of 1871 relied exclusively on numerals, all Chinese transmissions were thus much costlier than their alphabetic counterparts (to give some sense of the cost differential, the shortest code unit for Latin alphabetic letters was for the transmission of the letter "e," which required only one short "dot" pulse). Second, from an early moment within the history of telegraphy, a pressing issue facing cable companies and governments was the rapid spread of coded languages and ciphers.[10] These systems of encryption were geared toward protecting the content of the message, and more so toward reducing the cost of transmission through the usage of abbreviations and ciphers that stood in for longer words, passages, and even entire sentences. While a boon for individuals and companies, these codes and ciphers threatened to erode the profits of telegraph companies, and were quickly subjected to higher tariffs and certain limitations on usage.[11]

The moment that the Chinese language entered into the international telegraph system, it did so under a rubric of a ciphered or "numbered language"—that is, it was understood by the prevailing tariff and legal regimes of international telegraphy as a kind of *inherently secret code*. If languages such as English, French, Russian, German, Hebrew, Arabic, and others could be transmitted either "in the clear" or "encrypted," Chinese technically speaking could not: "plaintext Chinese" legally did not exist. As such, even for messages that were not subject to any cost-saving or cryptographic processes beyond the initial 4-digit encipherment, nevertheless such messages were subject to a large and growing body of regulations that applied to secret or cost-saving transmissions. From 1871 onward, that is, all Chinese telegrams were treated a if they were "secret" or "encrypted" transmissions.[12]

10. Tom Standage, *The Victorian Internet*, 105–26.

11. Mullaney, *The Chinese Typewriter*, chap. 2.

12. Telegraph companies with vested financial interests in the Qing—particularly the Danish company Great Northern and British company Eastern Extension Australia and China Telegraph Company—focused on compensating for the disadvantaged position of Chinese script by establishing preferential transmission rates. They also advocated special status for the Chinese number code, securing agreements as early as 1893 such that the price of Chinese transmissions would be assessed differently than conventional "numbered language" transmissions. Domestically, various Chinese regimes also promulgated preferential pricing systems. Circa 1933, for example, a regulation was passed that exempted customers from paying the cost of encipherment and decipherment, deferring it instead to the telegraph offices themselves. Even with such makeshift exemptions, however, the Chinese telegraph code was not on equal footing with codes for other languages. Bull, "A Short History of the Shanghai Station," Shanghai: NP, 1893 [Handwritten Manuscript], DOC/EEACTC/12/10, 4, in CAWA; Zhu Jiahua, *China's Postal and Other Communications Services*, 166.

Confronted with these challenges, some in the closing years of the nineteenth century began to explore alternatives to the numeric code of 1871—particularly codes based upon the letters of the Latin alphabet. By employing the Latin alphabet instead of numerals, it was reasoned, it would not only be possible to develop a more rapid and cost-effective transmission protocol for Chinese messages (since letters were cheaper to send than numerals), but also to advance broader efforts advocated by some to replace character-based Chinese script with a fully alphabetic, phoneticized alternative. Prominent examples included the "Phonetic Shorthand" (*chuanyin kuaizi*) system invented by Cai Xiyong (1847–97), a revised version of the same system proposed by Tang Jinming, and an exploration of pinyin-based transmission by Wang Bingyao in his "Pinyin Character Table" (*Pinyin zipu*).[13]

The most influential early proponent of pinyin-based transmission was Wang Jingchun, China's representative at the 1925 Paris International Telegraph Conference and one-time Managing Director of the Beijing-Hankou Railway. Wang Jingchun was a vocal advocate of Chinese phoneticization, arguing at length in his writings that character-based Chinese script should be displaced by a Latin alphabet-based writing system. The pinyin telegraphy system he proposed reflected this conviction and commitment.

Before proceeding further, it is necessary to pause for a moment to contemplate one of the major linguistic and technical challenges that confronted Wang, along with others who sought to transform Chinese into a fully alphabetic script capable of functioning independently of a character-based writing system: the preponderance of homophones in the Chinese language, or characters that have identical pronunciations, but distinct meanings. Whereas in the English language the phonemic value /rīt/ forms the pronunciation of four semantically different words—*right*, *rite*, *wright*, and *write*—homophony in the Chinese language is considerably more common and complex. In one of the more extreme examples, no fewer than eighty semantically different Chinese characters all have the pronunciation *li*—among them *li* (禮 ceremony/ritual), *li* (李 a common surname), *li* (力 strength), and *li* (荔 litchi), to outline only a handful.

In his development of a new Chinese telegraph code, Wang succeeded in creating one based entirely on the letters of the Latin alphabet, but not in fashioning one that was independent of character-based Chinese writing. While he did use the letters of the Latin alphabet to transliterate the phonetic properties and tonal values of Chinese words, he also used them in ingenious ways to encode the non-phonetic *structural* properties of Chinese characters as well—a strategy he needed in order to overcome the challenges associated with homophony. To introduce the operations of the code, let us consider two characters with identical phonetic values, but with

13. Wang Kaijie, *Wo guo dianxin fazhan jianshi.*

different tonal values: 汪 (*wāng*, "an expanse of water," with the pronunciation of "wang" in the first tone) and 枉 (*wǎng*, "crooked," with the pronunciation of "wang" in the third tone). We will focus here on how these two characters were encoded according to Wang Jingchun's proposed scheme, and then return to consider the implications of his strategies.

汪 = wangbux

枉 = wangxum

For each of these two characters, the first segment of the tripartite code was identical: a phonetic rendering of the character's pronunciation (*wang*) using a revised version of the Wade-Giles romanization scheme, widely used during this period.[14] Prima facie, then, Wang's pinyin telegraphy code might strike us as a straightforward case of phonetic transliteration, insofar as all vestiges of the original Chinese characters are nowhere to be found in the resulting code sequence—the characters have, it would seem, been fully purged from the chain of signification. Delving more deeply into the code, however, two further semiotic strategies and layers are revealed. Following this initial phonetic component, the second and third components of Wang's code exploited the Latin alphabetic *non-phonetically*, with the letters of the alphabet serving not as means of spelling out the phonemic values of Chinese words, but as variables with which to encode first the tonal values of characters, and second their key structural properties.

Specifically, the second segment of the code—the segment "b" in the case of the first example "wangbux," and the segment "x" in the second "wangxum"—represented the tone of each character. The tonal codes in Wang's system were as follows:

b = pingxeng (平聲) first tone

p = yangping (陽平) second tone

x = xangxeng (上聲) third tone

c = ciuxeng (去聲) fourth tone

r = ruxeng (入聲) fifth tone[15]

It is worth noting that Wang and his colleagues could in theory have used Arabic numerals to signify tones (e.g., "wang1ux" and "wang3um"),

14. Wang Jingchun, "The New Phonetic System," 151; Jiaotongbu pinyin dianbao yanjiuhui yiding [交通部拼音電報研究會議定]. Jiaotongbu guiding guoyin dianbao fashi [交通部規定國音電報法式], 1928, 5.

15. Modern standard Chinese has four tones, making Wang Jingchun's inclusion of five markers surprising, perhaps, to some readers. In developing his system, Wang Jingchun appears to have drawn upon typologies of Chinese tones more commonly encountered in early modern and late imperials treatises, in which the inclusion of a fifth tone is not uncommon. For more on early modern and premodern taxonomies of Chinese tones, see Endymion Wilkinson, *Chinese History*.

yet this would have proven financially disadvantageous. As noted above, numerals in Morse Code have much longer dot-dash code sequences than do letters, thereby incurring significantly higher costs in their transmission (a problem already well known by Chinese telegraphers, thanks to their experience with the 1871 numerical code and its descendants). What is more, the interpellation of numbers into an otherwise alphabetic code sequence ran the risk of being flagged once again as a "secret" or "cost-cutting" transmission—the same fate that befell the code of 1871 within the international telegraphic tariff regime. To avoid this, Wang Jingchun made sure to restrict his code exclusively to the letters of the Latin alphabet.

How Wang Jingchun settled upon these fives letters illustrates the complex semiotic strategies at play, strategies which steadily moved Wang away from a simple case of alphabetization-as-phoneticization. In his explanation of "b" as a suitable code for representing the first tone, for example, the inventor provided the following rationale: "B whose sound approaches P, the initial letter of *pingxeng* (平聲), is adopted for *yinping* (陰平), the first tone."[16] Within this one short passage, a daisy-chain of no fewer than four layers of mediation were at play: "B" representing "P" (for no other reason that Wang Jingchun required the letter "P" elsewhere in his system, and thus could not reuse it here); "P" as representing *pingxeng*, the phoneticized representation of the Chinese character term 平聲 (signifying "level tone"), and the Chinese term 平聲 as the synonym of an earlier Chinese character term *yinping* 陰平, often considered the premodern counterpart of the contemporary concept of "first tone"). Similarly elaborate chains of mediation connected the remaining code—p, x, c, and r—to their corresponding tonal values. In this second part of the code, then, we have clearly begun to depart from a purely phonetic scheme, even as the system remained wholly alphabetic.

The key for our discussion is the third and final part of the code, however—the sequence "ux" in "wangbux" and "um" in "wangxum." This was the moment when Wang employed the Latin alphabet to smuggle into his ostensibly "phoneticized/romanized" system the very Chinese characters that his system sought to replace (and help abolish). Comprising a two-letter sequence, this portion of Wang's code used letters to represent not the pronunciation of Chinese words, but rather the *physical shapes of Chinese characters*—or more specifically, the components out of which the characters in written Chinese are composed, referred to in English as *radicals* or *classifiers*. By doing so, Wang's code was able to differentiate between otherwise homophonic Chinese words (as in our example above of two characters that shared the pronunciation *wang*).

Like their tone-code counterparts, this segment within Wang Jing-

16. Wang, "The New Phonetic System," 151.

chun's telegraph code was based upon complex forms of mediation that used the letters of the Latin alphabet to achieve much more than mere phoneticization. Specifically, this two-letter code represented the *orthographic inversion of the Latin alphabetic letters used in the romanized pronunciations of the names of Chinese radicals*. To clarify this admittedly opaque description, these final code units were designed using a three-step process: First, the phonetic value of the relevant Chinese radical was transcribed using the letters of the Latin alphabet. In the case of 枉, the relevant radical was the tree-radical 木 or "mu," which appeared on the left side of the character. In the case of 汪, the relevant radical was the water-radical 氵 (rendered "xueh" in Wang's romanization scheme, as opposed to the now common pinyin romanization "shui"). Second, for any phoneticized values containing more than two letters (as in the case of Wang's "xueh," which contains four letters), these were cropped down to leave only the two initial letters thereof (transforming the four-letter "xueh" into the two-letter "xu," but leaving the two-letter "mu" untouched). Finally, these two-letter renderings were inverted—from "mu" to "um" and from "xu" to "ux," respectively—so as to prevent them from being misrecognized as phonetic values that encoded pronunciation. Each of the radicals in Wang's code was thereby assigned a unique, two-letter code (fig. 1).

Taken together, these three components were concatenated to form a fully alphabetic and yet predominantly *non-phonetic* code sequence, which would thereafter be transmittable using standard Morse pulse-patterns. The character *wang* (汪) was thus encoded as "wangbux" while the character *wang* (枉) was encoded as "wangxum." To decode the code sequence "wangbux," then, the recipient would need to deduce the three properties of the Chinese term being transmitted: that it was pronounced "wang," that it was pronounced in the first tone, and importantly for us, that the underlying Chinese character contained the "mu" or "tree"-radical. While the first two of these code sequences were thus premised on the phonetic value of the Chinese word in question, the third and arguably critical code unit was premised upon the structural properties of the Chinese character itself—a structural property unrelated to the phonetic value of the word.

Wang's experimental telegraph code came closer than any in history to replacing the numeric code of 1871. His proposal received the formal support of a network of esteemed Chinese engineers, telegraphers, linguists, and policymakers.[17] Approved by the central government in 1928, the new system was poised to replace the four-digit telegraph code with a semi-phonetic/semi-structural alphabetic code, scheduled to take effect beginning on 1 January 1929. Political crises delayed and ultimately short-circuited the plan, however, and the code was never put into use. Pinyin-based transmission continued to attract attention, however, and would be

17. Ibid., 149.

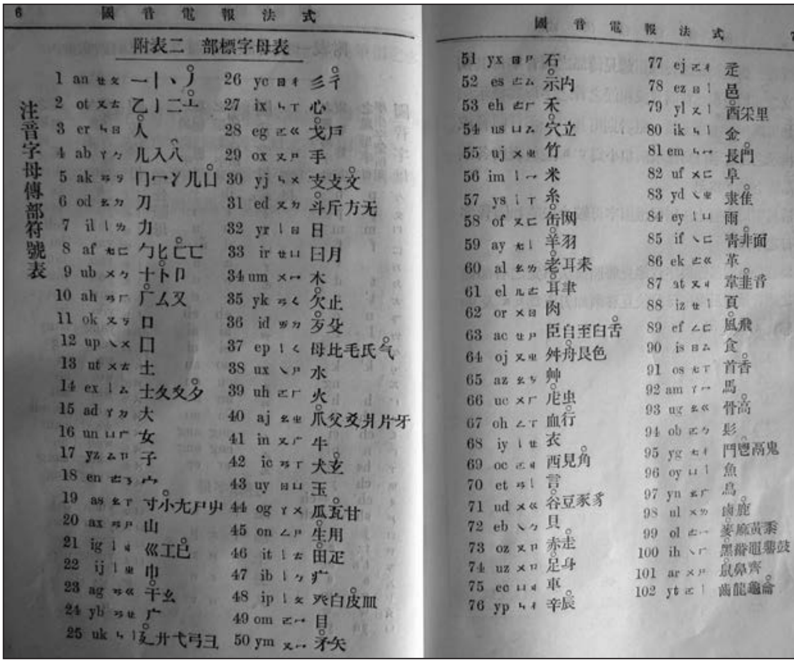


FIG. 1 Codes for the Transmission of Radicals, by Wang Jingchun (1928). (Source: Chinchun Wang, "The New Phonetic System of Writing Chinese Characters." *Chinese Social and Political Science Review* 13 [1929]: 158–59.)

the focus of experimentation well into the 1970s. At the same time, four-digit transmission was never unseated.

For our purposes, the key dimension of Wang Jingchun's abortive experiment with alphabetization was the way in which even this vocal advocate of full-scale phoneticization, when given the opportunity to create an alphabet-based semiotic system for Chinese telegraphic transmission, nevertheless smuggled into his system the very things that he and a number of his contemporaries were attempting to abolish: namely, Chinese characters themselves. Whether "ux," "um," or any one of the other dozens of codes by which Wang encoded Chinese radicals, the non-phonetic, structural properties of characters remained not only present in the code, but critical for it to function. Without "um" (that is, 木) and "ux" (that is, 氵), there would be little besides linguistic context and probabilistic reasoning to enable a message recipient to understand with any certainty which of the dozens of Chinese words pronounced "wang" were being sent across the wire. Were Wang truly to have abandoned Chinese characters—operating exclusively within a framework of alphabetization-as-phoneticization—his code would have fallen prey to the challenges of homophones and disambiguation, rendering it useless for the purposes of telegraphic transmission. This most "alphabetized" of Chinese telegraph codes, then, quietly depended upon the continued presence of the very character-based writing it sought to replace.

The Letter as Stroke: Samuel Caldwell and the Sinotype

A new "Chinese alphabet" emerged in the 1950s, this time in the contexts of computing and phototypesetting. Had everything gone precisely to plan, United States President Dwight D. Eisenhower would have delivered a speech in the summer of 1959, unveiling to the world the first Chinese computer in history: the "Ideographic Composing Machine." Not only could such technology be presented as a "gift" from the capitalist world to the Chinese people, scoring a technological and cultural victory for the U.S. Cold War camp, but so too did it offer up the possibility of a powerful new infrastructure for the global dissemination and translation of Chinese-language material. Whoever possessed such a device could flood the world with Chinese texts at a rate never before witnessed in human history.

Such was the optimistic vision of the "Ad Hoc Working Group on Ideographic Composing Machine," at least, which in its memo of May 1959 urged the "earliest public announcement of this machine by the President as a major breakthrough by the United States in the long and continuing struggle to improve mutual understanding among peoples of the world by better communication."[18] Alongside this somewhat romantic goal, a more prominent objective was likely found in an accompanying justification, namely that "the urgency for earliest public announcement stems from our inability to withhold knowledge of the machine and its operation from the Communists, and thereby, the risk of an announcement first by the Communists that they have developed such a machine." There was, the committee explained, a "risk of being 'scooped' by the Communists from whom knowledge of the invention cannot be withheld."[19]

The machine at the center of this plan was developed by MIT engineer Samuel Hawks Caldwell (1904–60), a pioneer in the field of logical circuits who, despite neither speaking nor reading a word of Chinese himself, became fascinated by the puzzle of Chinese computing halfway into his career. Caldwell was born in Massachusetts in 1904, going on to study at MIT under renowned analog computer designer Vannevar Bush (1890–1974). In 1933, Caldwell received his doctoral degree with a dissertation

18. "Briefing Memo re Agenda Item 2, 'Interim Report of Ad Hoc Working Group on Ideographic Composing Machine', Board Assistants' Meeting, May 8, 1959," 7 May 1959, in OCB Secretariat Series, Box 3: Ideographic Composing Machine, in DEL.

19. "Memorandum for the Executive Office: Chinese Ideographic Composing Machine—Briefing Memo on Deferral of Board Consideration," 20 May 1959, in OCB Secretariat Series, Box 3: Ideographic Composing Machine, in DEL.

entitled "Switching Circuits and Logical Design."[20] Caldwell later joined the faculty of MIT as a professor of Electrical Engineering, advising such luminaries as pioneering information theorist David Huffman (1925–99). In his spare time, he played the organ, making guest appearances with the Boston Pops, and played Christmas carols for the neighborhood during the holiday season. He was a serious man, as his granddaughter Ann Welch recalls, but Christmastime was a uniquely joyous and relaxing season.[21]

Thanks to its global prowess within fields of science and engineering, MIT was a prime destination for international students from the early twentieth century onward, including Chinese students eager to contribute to the modernization efforts of their home country. Caldwell had never traveled to Asia until later in life, but clearly took some pleasure in his interactions with his Chinese students. Betty Caldwell, Samuel Caldwell's wife, often invited his Chinese students over to the house, venturing to make them feel more at home by preparing Chinese dishes—dishes that she learned from a family friend who happened to run a cooking school in Boston: Joyce Chen, doyen of Chinese-American cuisine and author of the classic *Joyce Chen Chinese Cook Book*.[22]

These interactions evidently made a strong impression on Caldwell, particularly those with MIT student Francis Fan Lee. As Caldwell would later explain, it was Lee who first alerted him to the "basic strokes" used in the composition of Chinese, and that—with certain exceptions—the stroke-orders of Chinese characters were invariant.[23] In Caldwell's writings, the sudden and remarkable appearance of China and the Chinese language cut a striking silhouette against a body of work otherwise dominated by deeply technical studies on circuit design and electrical engineering. "A Chinese character is not a haphazard structure," Caldwell explained, a phrasing that suggests either his own once-held assumptions about the language, or perhaps his expectation about the assumptions of his readers.[24] "Chinese has a 'spelling,'" Caldwell explained in another publication, "in the sense that the sequence of strokes used in writing a given character is invariant. . . . if the strokes are regarded as the 'letters' of an 'alphabet,' the Chinese always 'spell' a word the same way each time it is written."[25]

20. Samuel H. Caldwell, *Switching Circuits and Logical Design*, 1958.

21. Interview with Ann Welch (daughter-in-law of Samuel Hawks Caldwell), 5 September 2013.

22. Ibid.

23. Caldwell, "Machine Seen as Possible 'Breakthrough' in Chinese Printing," FEFile no. 147, 22 June 1959, accession no. 98055-16.370/376, box no. 276, 1–6, in PLP.

24. Caldwell, "Progress on the Chinese Studies," in "Second Interim Report on Studies Leading to Specifications for Equipment for the Economical Composition of Chinese and Devanagari," by the Graphic Arts Research Foundation, Inc., addressed to the Trustees and Officers of the Carnegie Corporation of New York, accession no. 98055-16.370/376, box no. 276, p. 2, in PLP.

25. Ibid., 1.

Having once perhaps thought that Chinese calligraphy was subject to no orthographic laws, Caldwell soon discovered something to the contrary:

> What kind of writing skill does the Chinese student acquire in the course of his elementary education? Strangely enough, it turns out that he learns to write ideographic characters very much as his alphabetic brother learns to write words. . . . There is nothing whimsical about the process of setting down the strokes that comprise a Chinese character. Every Chinese learns to write a character by using exactly the same strokes in exactly the same sequence. . . . It is found that Chinese writing can be analyzed into 21 distinct elemental strokes, disregarding the position and the scale of each stroke within the body of a character.

OCTOBER
2018
VOL. 59

Caldwell's intellectual curiosity had been whetted and reoriented in a direction entirely distinct from his career thus far: he set out to build a computer for Chinese. As part of this shift, Caldwell took on a new role as Director of Research at the Graphic Arts Research Foundation in Cambridge, Massachusetts. "Nearly all of the advancements of mechanization in the type composing art," he would soon write, "have been made to facilitate composition in languages having relatively few distinct typographical characters as compared with the numbers of comparable characters in ideographic languages, such as Chinese."[26]

Consulting Lien-Sheng Yang, esteemed professor and first full-time historian of China in the department of Department of Far Eastern Languages at Harvard University, Caldwell enlisted Yang's help in conducting a thorough analysis of the structural make-up of Chinese characters. In particular, he relied on Yang to oversee four critical components of the study: the determination of the number and types of fundamental Chinese strokes, the selection of vocabulary to use in the analysis, the recruitment of Chinese analysts, and the final checking of the analysis.[27]

26. Caldwell, "Ideographic Type Composing Machine," patented 30 August 1960.

27. Without delving too deeply into Caldwell's research process, there were three design challenges he faced: ambiguity of strokes, characters that shared identical "spellings," and the problem of "spelling" characters that themselves were also Chinese radicals which appeared in other characters. With regard to the twenty-one fundamental strokes, Caldwell and Yang soon discovered that "in many characters it is difficult for even well-educated Chinese to be sure whether a given stroke is long or short." The second problem involved what Caldwell termed "spell-alike-look-different": characters who had the same "spelling" according to his system, such as 已己巳 (yi, ji, si). "Fortunately," Caldwell explained in his findings, "there are relatively few of these 'spell-alike-look-different' characters in the vocabulary." Within 2,178 characters used, only ten duplications of spelling were discovered. To resolve this issue, Caldwell outfitted his keyboard with three Chinese numerals—one (yi), two (er), and three (san)—with which the user could choose between characters in the rare instance that more than one exhibited the same stroke-spellings. The third challenge was the radical-character problem.

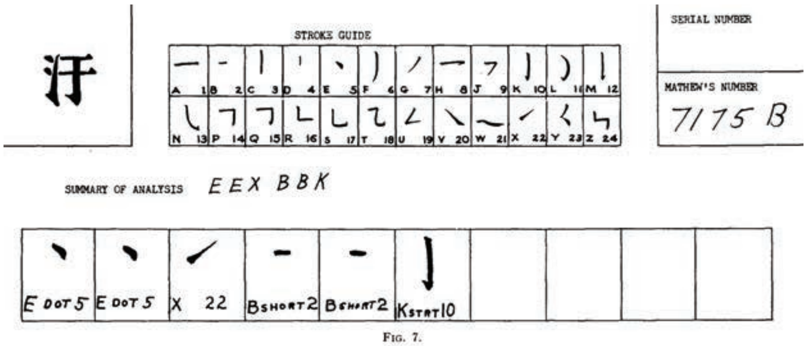


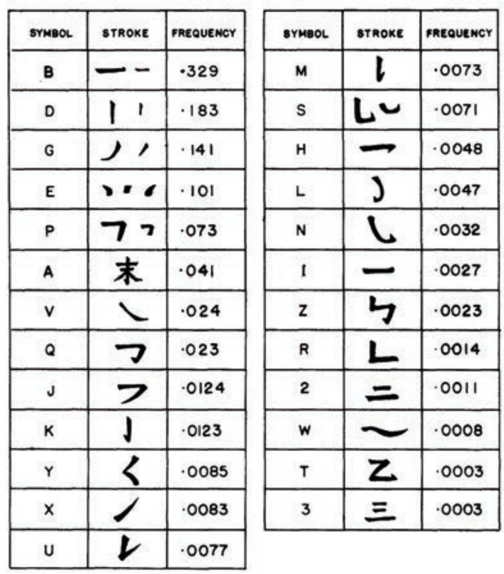
FIG. 2 Data Collection Card for Caldwell-Yang Chinese Stroke Analysis. (Source: Samuel H. Caldwell, "The Sinotype A Machine for the Composition of Chinese from a Keyboard." *Journal of The Franklin Institute* 267, no. 6 [June 1959]: 471–502, 479.)

Caldwell's use of the word "spelling," both in his early exploratory writings, and in his patent document and project report, is at once revealing and misleading. The term alerts us to the ways in which Caldwell was attempting to translate a set of concepts familiar to him within his native, English-language context to the novel domain of Chinese information processing. By referring to the "spelling" of Chinese, and by focusing on Chinese strokes, Caldwell was attempting to draw a line of equivalence between Chinese "strokes" and alphabetic "letters"—to imagine, as it were, that an English "word" bore roughly the same relationship to "letters" as a Chinese "character" did to calligraphic "strokes." This intellectual move in turn raised the question as to how many stroke-letters there were in this "Chinese alphabet"—relatively few, as in the case of writing systems like English, Russian, Arabic, or German; or perhaps many, as in the case of Burmese or Thai? Insofar as Caldwell intended to use the Western alphabetic typewriter as the input device for his system, however, his and Yang's research into Chinese strokes was bounded at the outset: ideally, there could be no more Chinese stroke-letters than there were letters of the Latin alphabet, since this would make the Western typewriter keyboard unusable; but also not too few, insofar as this would have disallowed Caldwell from making the fullest use of the Western typewriter keyboard in its en-

---

Caldwell quickly discovered that some stand-along characters, like the word "mouth" (*kou* 口), also appeared in the "spelling" of other characters, such as "to eat" (*chi* 吃), thereby raising the question of how to differentiate for the system which is desired. To resolve this, Caldwell developed a special Chinese stroke-key labeled "*mo* (末)," the Chinese term for "ending." By "spelling" the character mouth/*kou* and then depressing the "end" key, the system was thereby instructed to produce the standalone character "mouth," rather than any number of the other characters that contain the mouth/*kou*-component.

OCTOBER
2018
VOL. 59



SYMBOL	STROKE	FREQUENCY	SYMBOL	STROKE	FREQUENCY
B		·329	M		·0073
D		·183	S		·0071
G		·141	H		·0048
E		·101	L		·0047
P		·073	N		·0032
A		·041	I		·0027
V		·024	Z		·0023
Q		·023	R		·0014
J		·0124	2		·0011
K		·0123	W		·0008
Y		·0085	T		·0003
X		·0083	3		·0003
U		·0077			

FIG. 3 Chinese Stroke Frequency Analysis by Samuel Caldwell and Lien-Sheng Yang. (Source: Samuel H. Caldwell, "The Sinotype A Machine for the Composition of Chinese from a Keyboard." *Journal of The Franklin Institute* 267, no. 6 [June 1959]: 471–502, 475.)

tirety. By consequence, Caldwell and Yang ultimately decided upon a set of twenty-one "fundamental" Chinese strokes (later expanded to twenty-four), each of which was assigned a single typewriter key and Latin alphabetic counterpart (fig. 3).

At first glance, then, it would once again appear that in Caldwell's system we are dealing with the wholesale "alphabetization" of Chinese, albeit not by phonetic means. Although Caldwell's system was not premised upon the "sounding out" of Chinese words using Latin alphabetic letters, it was premised on "shaping out" Chinese words using Caldwell's hybridized concept of stroke-letters. Although it was not a phoneticized alphabet, Caldwell had managed, it seemed, to fully subordinate the Chinese writing system under the alphabetic logic of "spelling"—in this case, using the letters of the Latin alphabet, each paired with a "fundamental stroke" to type out the user's desired Chinese characters piece-by-piece. While this was not alphabeticization-as-phoneticization in a literal sense, then, it upheld the spirit of alphabetic spelling through a rather ingenious form of analogy.

In practice, however, the way that Caldwell's Sinotype actually "spelled" Chinese characters was nothing like the way that a Western typewriter spelled out alphabetic words. To "spell" the word pronounced /kəmˈpjutɚ/ using a typewriter, word processor, or computer is to depress a sequence of keys to "sound out" the word in accordance with the spelling conventions that govern the language in question: i.e., C-O-M-P-U-T-E-R. The "spelling" of this word is said to be complete only when all eight letters are present, in the correct sequence, and without error. Such was not the case with the Sinotype, however. Instead of using the stroke-keys to "sound out"—or even to "shape out"—each character in full, the core objective on the Sinotype was, as Caldwell himself explained, "to furnish the input and output data required for the switching circuit, which converts a character's spelling to the location coordinates of that character in the photographic storage matrix."[28] In other words, the Sinotype operator would use the keys of the keyboard to provide the system with one or another "address" with which the system would then retrieve the operator's desired character from internal memory, and present it to the user for confirmation or cancellation. Sinotype was not primarily an inscription device in this regard, but rather a *retrieval* device. Inscription took place only after retrieval was successfully completed.[29]

While this distinction might at first seem minor, the implications are profound—for Caldwell, and for our understanding of Chinese technolinguistic frameworks and the functioning of the Latin alphabet therein. In the course of his research, Caldwell made a startling discovery. Not only did Chinese characters have a "spelling" but, as he wrote, "the spelling of Chinese characters is highly redundant."[30] It was redundant, Caldwell explained, because it was almost never necessary for Caldwell to enter every stroke within a character in order for the Sinotype to retrieve it unambiguously from memory. "Far fewer strokes are required to select a particular Chinese character than are required to write it," Caldwell explained. An English-language analogue might be that of retrieving, rather than spelling, the word "xylophone" or "crocodile": the first five letters are sufficient to form an unambiguous (or nearly unambiguous) match with the complete word. Since there exist few, if any, words in English that start with these five letters, it becomes unnecessary (in the context of retrieval) for the user to continue the selection process beyond the establishment of a match. In spelling, however, neither of these two words can be said to be complete until each of the component letters are inscribed on the page or

28. "Final Report on Studies Leading to Chinese and Devanagari," 14, in GARF.

29. Caldwell's Sinotype was not the first Chinese system to employ this mode of human-machine interaction. Rather, an experimental Chinese typewriter known as the MingKwai machine, invented in the 1940s by Lin Yutang, was the first. For an in-depth examination of MingKwai and Chinese retrieval-inscription machines, see Mullaney, *The Chinese Typewriter*.

30. Caldwell, "Progress on the Chinese Studies," in "Second Interim Report on Studies Leading to Specifications for Equipment for the Economical Composition of Chinese and Devanagari," by the Graphic Arts Research Foundation, Inc., addressed to the Trustees and Officers of the Carnegie Corporation of New York, accession no. 98055-16. 370/376, box no. 276, p. 2, in PLP.

TABLE 1
COMPARISON OF FULL AND "MINIMUM" SPELLINGS FOR SAMPLE CHARACTERS

Mathew's number	*Spelling in full*	*Strokes required to write characters*	*Minimum spelling*	*Strokes required to "find" or select characters*
7592	AAK	3	AAK	3
1752	ABB CPB BBG ENE	12	ABB	3
4278	ABC AAF C	7	ABC	3
6748	ABD ACP BBB GE	11	ABD AC	5
5187	ABD AGM GV	8	ABD AG	5
4678	ABD BDA GDP BR	11	ABD BD	5
5817	ABD BNE	6	ABD BN	5

screen. What takes nine letters to "spell" in the conventional alphabetic sense might therefore take only five letters to "retrieve."

In many cases, the difference between "spelling in full" and "minimum spelling" (Caldwell's terms) was dramatic. For one character containing fifteen strokes, for example, it was necessary for the operator to enter only the first five or six strokes before the Sinotype arrived at a positive match. In other cases, a character with twenty strokes could be unambiguously matched with only four keystrokes[31] (table 1).

Caldwell pushed these observations further. He plotted the total versus minimum spelling of some 2,121 Chinese characters and, in doing so, determined that the median "minimum spelling" of Chinese characters fell between five and six strokes, whereas the median total spelling equaled ten strokes. At the far end of the spectrum, moreover, no Chinese character exhibited a minimum spelling of more than nineteen strokes, despite the fact that a number of Chinese characters required twenty or more strokes to compose. Thus, in addition to developing the first Chinese computer in history, Caldwell had also unwittingly invented what we now know as "autocompletion."

Caldwell's system quickly garnered attention, including financial backing from the Carnegie Foundation, the United States Army, and the United States Air Force.[32] By May 1959, momentum for Caldwell's project grew further, with a delegation from the United States Information Agency (USIA) planning to visit Cambridge "to cover the trial run of a newly-invented Chinese language photocomposition machine." "This invention has been described as the greatest step forward in printing of ideographic languages

31. Ibid.

32. Caldwell, "Machine Seen as Possible 'Breakthrough' in Chinese Printing," FEFile no. 147, 22 June 1959, accession no. 98055-16.370/376, box no. 276, 1–6, in PLP.

since the Chinese invented movable type in the 11th century," the team reported in advance, "and is believed to be as important to the Chinese language as the inventing of the linotype machine was to Romance language printers."[33] There was also growing concern among the team that PRC scientists might be nearing their own breakthrough, which would severely undercut the psychological impact of Caldwell's invention. In particular, the team referred to an "unconfirmed NCNA [New China News Agency] report on 14 November 1958 of Chinese Communist scientists developing automatic typesetting machine capable setting 20,000 characters per hour."[34]

Despite the optimism of the Ad Hoc committee, summer 1959 passed by without remark, at least when it came to the history of Chinese information technology. Eisenhower did not unveil a Chinese computer, and the Ideographic Composing Machine was never debuted to the world. Doubts persisted as to the readiness of the device, and whether it would withstand scrutiny by the international community and military analysts. Would it prove viable for Chinese users? Was it, indeed, as potentially field-changing as the designers had come to believe? Ultimately, it was decided that the risk of premature announcement was too great, and the project was postponed.

The following year, the project was dealt a heavy blow. In 1960, Samuel Caldwell passed away. Without this pioneering leader to oversee the project, enthusiasm in military circles necessarily diminished, if only as they waited for an eventual successor. The life of the machine continued, though, persisting for decades along a tortuous chain of custody that counted among its members a veritable who's who of the military-industrial-academic complex: the Air Force, Army, Pentagon, Itek, MIT, RCA, and many others. The machine would be re-christened along the way, first as the *Sinowriter*, then as the *Chicoder*, thereafter as the *Ideographic Encoder*, and later as the *Sinotype II* and *Sinotype III*. More importantly, the conceptual and technical framework Caldwell and his team had laid down would remain foundational for Chinese computing well into the 1980s.

Letter as Isomorph: H. C. Tien and the Chinese Transalphabet

The foregoing analysis has focused exclusively on alphabetization schemes developed prior to the advent and popularization of Hanyu Pinyin—the now-official and widespread system of romanization in the People's Republic of China. As the focus of our analysis ventures into the age of pinyin, a reasonable assumption might be that our story begin to close the gap between alphabetization and phoneticization. With the wide-

33. "Chinese Language Photocomposition Machine," 4 May 1959, accession no. 98055-16.370/376, box no. 193, in PLP.

34. Ibid.

spread standardization and popularity of pinyin, should not "alphabetization" and "phoneticization" at long last become equivalent terms? Phrased differently, might not the experiments of Wang Jingchun and Samuel Caldwell be explained rather simply: namely, that during their times, the stage had simply not been adequately set for the full-scale phoneticization of the Chinese script to take place?

In the balance of this article, we will examine an aspiring technologist who, as part of his work in the 1970s, was not only fully aware of pinyin, but also a vocal advocate of its usage—a man who, like Wang Jingchun in the 1920s, was a vocal proponent of replacing Chinese characters with a fully phoneticized orthography. Even in this context and with these commitments, however, we will see that the system he developed continued to occupy the liminal position between the phonetic and the structural.

H. C. Tien (Tian Xincheng 田心澂) was born in China in 1929. As the son of Chinese diplomat Tian Fangcheng (田方城), he spent a portion of his childhood in Portugal during the Second World War, where he studied English among a variety of other subjects. Tien went on to attend Adrian College in Michigan, supported by the Boxer Indemnity Fund, and later received two master's degrees: one in neurology from the University of Michigan, and one in electrical engineering from Michigan State University. Tien went on to receive a medical degree, attending the University of Michigan, and completing his psychiatric residency at Ypsilanti State Hospital in Michigan.[35]

Throughout his professional career, Tien was ambitious, not a little bit eccentric, and characterized by self-reliance. In the late 1960s he launched his own journal, the *World Journal of Psychosynthesis*, in which he and his co-contributors explored an array of subjects. Tien was deeply enthralled by the promise of cybernetics, for example, particularly when combined with telecommunicative technologies. Throughout the 1960s he attended summer classes at MIT on the subject of cybernetics, and in 1978 wrote and self-published *Videology: Theory and Techniques*, wherein he outlined a program for the psychiatric applications of video technology.

Tien was also deeply disturbed by the perils and social fragmentation of the post-World War II world and the emotional turmoil they imparted upon the modern family and individual. Tien's psychiatric practice was oriented strongly towards the community and the group, with Tien arguing that the true foundation of psychological treatment was not the individual in isolation, but the individual nested within a family or social setting.

As a member of the East Lansing Human Rights Commission, and as a supporter in the anti-Vietnam War and Civil Rights movements, Tien was also deeply influenced by Chinese Communist ideology and discourse. Reading the *World Journal of Psychosynthesis*, one is immediately struck by his passionate tone. "World Journal of Psychosynthesis heralds the Third

35. Interview with Aled Tien (son of H. C. Tien), Singapore, 16 March 2015.

Psychiatric Revolution!" one issue announced. "Life is struggle," another declared. And perhaps most evocatively: "Mankind needs a cosmic identity. Cosman has already disclosed his emerging identity, when his hand first touched the moon in 1969. . . . Who is Cosman? We, people and earth are one. We are Cosman."

Tien was never more deeply impressed than upon learning of Henry Kissinger's clandestine conversations with delegates of the People's Republic of China, and with Richard Nixon's announced visit to the PRC. In short order, *World Journal* began to dedicate increasing space to China, Chinese acupuncture, and other subjects—with most of these articles authored by Tien himself. In the wake of Nixon's visit, Tien began to dedicate a section of each issue of *World Journal* to Chinese questions—and to the Chinese language in particular.

Citing Mao Zedong's endorsement of the romanization of Chinese writing, Tien began to contemplate his own system to "speed up the transphonation of Chinese calligraphy into a more universal Latin alphabetization of the Chinese language."[36] Not only would this system facilitate language acquisition by Chinese children in the People's Republic, he argued, but it would also make possible the development of Chinese language proficiency for foreigners. Once in possession of Chinese language facility, the world would have opened up to it a fount of Chinese wisdom—most notably the "medical miracle" of Chinese acupuncture. "Discovery of anesthesia," Tien pondered rhetorically, "is almost like a cultural bombshell on the world: How and why does a tiny needle create such great miracles in surgery and medicine?"[37] "Having acquired the Chinese language, cultural treasures hidden there for hundreds of years beckon us."[38] Beyond this lied the archives of Mao Zedong Thought and the Chinese regime's success in the "abolition of drug abuse, prostitution, personality disorders, and diseases of poverty." "Western mental health workers should study the thoughts of Mao Tsetung in Chinese." All this wisdom and perhaps more awaited the wider world, if only the Chinese writing system could be rendered in the Latin alphabetic form, and thus a more approachable framework.

Major challenges confronted the romanization of Chinese, Tien explained: Chinese cultural attachments to character-based writing, regional differences between the panoply of Chinese dialects, technical challenge of representing tones in the Latin alphabet, and above all, the preponderance of homophones in the Chinese language, as we examined earlier in the case of Wang Jingchun.[39] To confront the steep challenges posed by homophony, Tien introduced the central concepts of his Romanization program in

36. H. C. Tien, "On Chinese Acupuncture and Chinese Language," 4–5.
37. Ibid.
38. Ibid.
39. Tien, "On Learning Chinese," 4–5.

July 1972: *transalphabetization* and the "Chinese transalphabet." In developing his transalphabetization program over the course of six articles, Tien settled upon two core components: a system of "letter-doubling" by which Roman letters would be used to indicate the tones of standard Chinese, and what Tien called a "radical transalphabet," in which Latin alphabetic letters would be used to represent not the sound of Chinese radicals (as in Wang Jingchun's model from the 1920s), but the *graphical* properties of these radical components. As such, not only would letters serve their expected role of helping spell or sound out the phonetic value of Chinese words (spelling out the phonetic value of "li," for example), but also as variables (to indicate tone), and most intriguingly of all, as *isomorphs*: that is, as *shapes* whose value would be derived from their structural similarity to components of writing in the Chinese script.[40]

To begin with "letter-doubling," Tien distilled each of the tones of standard Chinese in letter form, avoiding the use of diacritical markings that he regarded as problematic for a variety of reasons. For "li" characters in the first tone, these would be rendered directly as "li." For "li" in the second tone, the initial letter would be doubled, producing "lli." For the third tone, the final would be doubled, producing "lii." Finally, for the fourth tone, both the initial and final would be doubled, producing "llii."[41] In this first part of the code, then, letters served both as phonetic elements—sounding out the phoneme *li*—but also as protocological variables governed by an artificial system devised by Wang himself. When encountering the doubled "ii" formation, this did not prompt a reader to elongate the vowel sound, nor did the doubled "ll" formation indicate a distinct phonological property. Instead, Tien was exploiting other dimensions—other semiotic resources beyond phonemic description—of the Latin alphabet in service of his system.

By itself, the letter-doubling technique was insufficient to disambiguate between all homophonic Chinese characters. In the case of "li," for example, the distinction between four tonal values alone was not enough to describe fully which "li" was being referenced out of the many dozens of possibilities. A further component would be needed within the code to identify characters uniquely.

Tien's answer to this was his "radical transalphabet": Latin alphabetic letters that would be used, neither for their phonemic value, nor as protocological variables, but as isomorphic counterparts to Chinese radicals. That is, Tien sought to exploit Latin alphabetic letters for their perceived structural similarities to certain structural shapes within character-based

40. Ibid.; Tien, "The Transalphabetization of Chinese Characters," 35–38.

41. Tien based his opposition to diacritical tone marks on reasons of unintelligibility and technical challenge. Diacritical marks, he argued, rendered the act of reading more challenging, and also complicated technological processes of typing, transmitting, and printing.

Chinese writing—to exploit Latin alphabetic letters as "characters" in and of themselves, comparable to the examples given at the outset of this article in which the Cyrillic letter "H" (/n/) is treated as a homographic counterpart of the Latin alphabetic letter "H" (/h/).

To grasp how Tien's encoding scheme worked, we can use the example of 電 (*dian*), a character signifying "electricity." Based on the pinyin romanization of the character's phonemic value, the first component in Tien's code was "d-i-a-n." As with our example of the phonemic value of li, however, *dian* corresponded to a wide variety of Chinese characters, such as 點 (*dian*, "point, dot"), 典 (*dian*, "canon" or "classic"), 滇 (*dian*, toponym signifying "Yunnan"), and many dozens more. As with the case of *li*, then, Tien's code would also need to disambiguate between these various homophones. In the case of *dian*, Tien forwent usage of his letter-doubling/tonal technique, and focused instead fully of his "transalphabet" in order to achieve unique identification. Focusing upon the graphical qualities of the upper part of the character—the "rain-radical" 雨 located at the top of the character *dian* 電—Tien encoded this radical using a series of Latin alphabetic letters that, within Tien's "Chinese transalphabet," corresponded to the rain-radical *isomorphically*.

The classes of isomorphic equivalence were fashioned in the following three steps (fig. 4). First, Tien isolated the horizontal and vertical strokes located in the top-center of the rain-radical—two strokes that, when treated alone, bore resemblance to the upper-case letter "T"—and thus encoded them as "T." The second alphabetic isomorph, the uppercase letter "M," was set equal to a cluster of strokes located in the middle section of the rain-radical. Finally, a pair of lowercase "v" letters were set equal to the smaller strokes located within this M-shaped structure. Taken together, then, the full "transalphabetic" code for the rain-radical 雨 was "T-M-V-V." When combined with the first part of Tien's code—the "dian" derived from the pinyin romanization of the overall character—the full code for *dian* (electricity) was thus D-I-A-N-T-M-V-V. In Tien's proposal, a keyboard operator would enter this alphabetic sequence into the system using a standard QWERTY keyboard, whereupon the system would retrieve and display the character 電 from memory. No other character with the pronunciation "dian" would be presented, because none of those characters would contain the rain-radical.

Tien created transalphabetic equivalences for every radical in Chinese, using one or more Latin alphabetic letters in the same way as the rain-radical example above. To represent the "grass radical" (艹), Tien's system used "tt"—relying upon the graphic resemblance between two, lower-case "T's" and the Chinese character component. Still others included the isomorphic pairings of "b" and 白, "p" and 尸, "pq" and 門, "pttq" and 鬥, "bb" and 比, "mm" and 马, "bv" and 身, and many others (fig. 5).

Over the next two decades, Tien's engagement with mainland China

OCTOBER
2018
VOL. 59

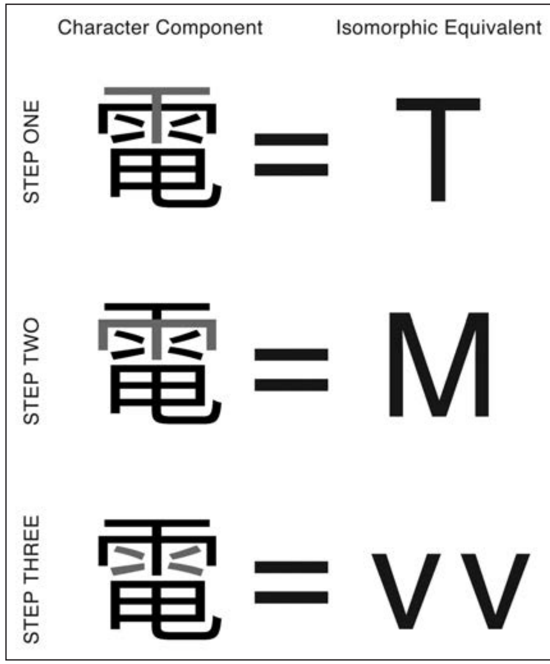


FIG. 4 Encoding "dian" (電 electricity) using H. C. Tien's "radical transalphabet" method. (Source: Image by author.)

grew steadily, as did his commitment to the project of "transalphabetization." When it became apparent that voyages from the United States to mainland China would once again become possible, Tien was able to realize a lifelong dream.[42] In 1973 he travelled with his family back to the country of his birth, proceeding first to Hong Kong, and then walking on foot over the bridge to present-day Shenzhen. From there, the family boarded a train to Guangzhou, before continuing on to Beijing. In the capital, the family stayed in the Wangfujing neighborhood in the "Overseas Chinese Mansion," one of two hotels available (the other being the Peking Hotel). In the city, they reunited with a number of Tien's aunts and uncles, including one great-uncle who worked in the Chinese Academy of Sciences in the Institute of Mathematics. It was perhaps thanks to this relative that Tien was successful in arranging a meeting during his visit with members of the Chinese Language Reform Commission.

No doubt inspired by his travels, Tien's writings on China multiplied and his commitment redoubled to the project of transalphabetization. In the early 1980s, H. C. Tien took a decisive step in forming a new company to advance his transalphabetic project into the realm of information technology, venturing headlong in the wide-open terrain of early Chinese personal

42. Interview with Aled Tien, Singapore, 16 March 2015.

FIG. 5 "The First Radical Transalphabet Table" by H. C. Tien (samples). (Source: H.C. Tien. "The Transalphabetization of Chinese Characters. VI: The Radical Transalphabet." *World Journal of Psychosynthesis* 4, no. 12 [December 1972]: 30–31, 30.)

computing. Tien formed Chinese Computer Communications, Inc., hiring his son Aled in 1983 to help program, design fonts, and develop user manuals. By 1985, Aled recalled, the company had a saleable software package.

The platform was based on Compaq and IBM PCs. The family-run business pursued contact with these companies, along with HP, Wang Laboratories, Apple, Microsoft, Multitech (predecessor of Acer), and others, eager to attract their interest in his Chinese text-processing program.

Such conversations did not materialize, however, and Tien's venture ultimately did not yield lucrative results. In the wide-open field of early Chinese personal computing and word processing, however, experimentation with the semi-phonetic, semi-structural possibilities of the Latin alphabet would continue into the present day.

Hiding in Plain Sight: The Continued Presence of Chinese Characters in Pinyin Input

By way of conclusion, we return to the present day to consider Chinese text processing in the contemporary period, which has witnessed some of the most successful and lucrative engagements between Chinese and the Latin alphabet. As noted at the outset of this article, the QWERTY keyboard is by now ubiquitous within Chinese computing, with the vast majority of computer users employing it in concert with one or another phonetic, Hanyu Pinyin-based Input Method Editor (IME). An IME is a "middleware" software program that operates in the background on one's desktop, laptop, tablet, or mobile system, intercepting the user's keystrokes, and then presenting Chinese character candidates to the user for confirmation on the basis of a protocol specific to the particular IME one is using (e.g., Google Pinyin, QQ Pinyin, Sougou Pinyin, or otherwise). From the moment the first key is depressed, the IME sets off on a dynamic, iterative process, searching in computer memory for those Chinese characters intended by the user. With the first key depression—the letter marked "C," perhaps—the IME begins to present the user with options in a pop-up menu that follows along at the margins of the screen, typically along the bottom—in this case, Chinese characters whose phonetic value begins with "C," such as "chi" (吃 to eat), "cai" (才 "only then," among many other meanings), or dozens of other possibilities.

When the user depresses a second button—let's say "H"—the IME rapidly adjusts its list of recommended Chinese characters, now offering up only those characters whose pronunciations begin with "ch" (thereby excluding the example of "cai" above). Proceeding letter by letter, then, Input Method Editors enable one to use the letters of the Latin alphabet, and with them the keys of the QWERTY keyboard, to produce Chinese character output. This process also makes available a wide variety of shortcuts. If the user were to enter the sequence "CX"—a sequence that does not correspond to any recognized phoneme in Hanyu Pinyin—the IME interprets this as a request to retrieve and present all meaningful, two-character Chinese compounds in which the pronunciation of the first character begins with the value "C" and the pronunciation of the second character begins with the value "X"—such as "chuangxin" (创新 innovation) and "chaoxi" (抄袭 plagiarism). Meanwhile, to enter the sequence "YL" yields still other two-character offerings, such as "Yelu" (耶鲁 Yale) and "yanlei" (眼泪 tears). Such abbreviated input sequences can even extend to three- and four-character sequences, and sometimes even longer ones. To enter the sequence "ZMKH," for example, retrieves the four-character idiomatic expression "to look at flowers while riding horseback" (走马看花 *zouma kanhua*, taken to mean a superficial understanding of something due to cursory analysis). Whether the user is composing a document in Microsoft Word, surfing the Web, or otherwise, he or she is constantly engaged in this iterative process of criteria, candidacy, and confirmation, using one IME or another. Using these pinyin-based input systems, along with a wider number of lesser-used non-phonetic Chinese input method editors, millions of Chinese computer and new media users have transformed China from a backwater of the global information infrastructure to one of its engines, cores, and most lucrative marketplaces.

When confronted with the technolinguistic success story of Input Method Editors and Chinese keyboarding techniques—rather than the long string of prototypes, theories, and failures examined above—what becomes of our central question: namely, the question of where and how the Chinese character is present within present-day systems of Chinese alphabetization, even when it appears to be absent? Within Hanyu Pinyin-based Input Method Editor systems, there is no evidence to suggest that Chinese characters are being smuggled in by means of the keystroke sequence itself, as in Wang's telegraphic transmission and its use of letter-based codes for radicals and tones; nor as stand-ins for Chinese strokes, as in Caldwell's system; and certainly not as structural isomorphs meant to correspond to the shapes of Chinese characters. Does this mean, then, that we have at long last arrived full-circle, fulfilling the goals of many nineteenth- and twentieth-century reformers, missionaries, and entrepreneurs of a fully phoneticized script that finally departs wholesale from character-based orthography?

The answer to this question is hiding in plain sight, as it were, not in the form of keystroke sequences, variables, or isomorphs, but rather in the now ubiquitous feature that accompanies all Chinese Input Method Editors: the pop-up menu. It is here, in the iteratively repopulated list of Chinese character candidates, constantly presented to the user as he or she depresses the keys of the QWERTY keyboard, where the structural and non-phonetic properties of Chinese characters continue to play an indispensable role in computational and informational processes. Now, rather than using Chinese radicals to help disambiguate between homophones, as in Wang's work, it is Chinese characters themselves that serve this purpose. Unlike Wang, who, in order to disambiguate and transmit the homophones 汪 (*wang*) and 枉 (*wang*) by telegram, used the Latin alphabetic suffixes "bux" and "xum" to encode both Chinese tones and radicals; or Tien, whose code

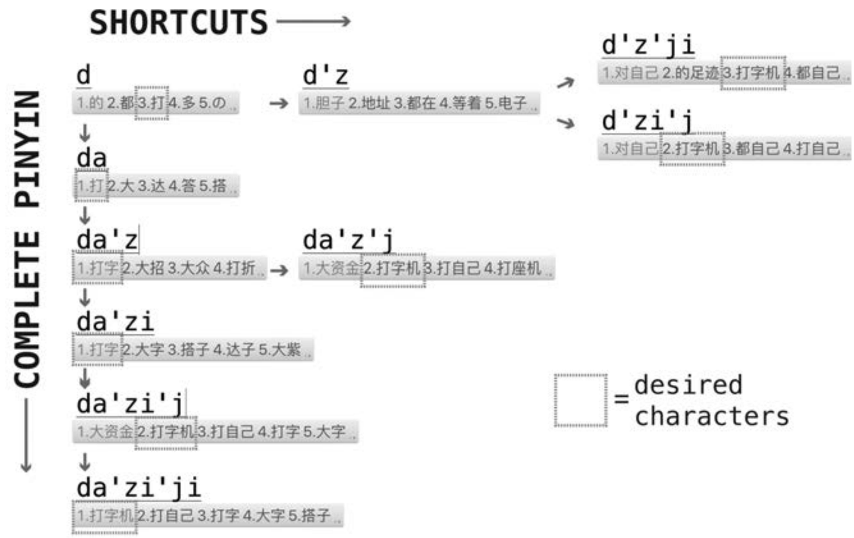


FIG. 6 Present-day Pinyin Chinese Input (sample of inputting "daziji/typewriter"). (Source: Image by author.)

employed such non-phonetic Latin alphabetic sequences as "tt" and "pq" to smuggle in the Chinese "grass radical" (艹) and "door radical" (門); contemporary pinyin-based Input Method Editors make use of the pop-up menu to perform precisely the same kind of technolinguistic work (fig. 6). The pop-up menu is now the standard "location" in which Chinese characters continue to occupy a vital "absent presence" in all contemporary engagements between Chinese script and the Latin alphabet. What the cases of Wang, Caldwell, Tien, and present-day pinyin-based IMEs tell us to consider—indeed, urge us—is quite simple: when it comes to understanding the full and complex history of China's engagement with the Latin alphabet, the theoretical and historiographic approaches we take to such a phenomenon need to be as complex and nuanced as the phenomenon itself.

Character Glossary

Cai Xiyong 蔡錫勇
chi 吃
Chuanyin kuaizi 傳音快字
ciuxeng 去聲
De Mingzai 德明在
Dianbao xinshu 電報新書
grass radical 艹
ji 己
kou 口

ma 馬
mo 末
mu 木
niu 牛
pingxeng 平聲
Pinyin zipu 拼音字譜
ruxeng 入聲
shuo 說
si 巳
Tang Jinming 湯金銘
Tian Fangcheng 田方城
Tian Xincheng 田心澂
Wang Bingyao 王炳耀
Wang Jingchun 王景春
Wang Kaijie 王開節
wǎng 枉
wāng 汪
Weijiye 威基謁
xan 山
xangxeng 上聲
yangping 陽平
yi 已

Bibliography

Archival and Oral Sources

Museum of Printing, Haverhill, MA
Graphic Arts Research Foundation papers (GARF)
Hoover Institution Library and Archives, Stanford University, Palo Alto, CA
Pardee Lowe Papers (PLP)
Dwight D. Eisenhower Presidential Library, Abilene, KS (DEPL)
Tien, Aled (son of H. C. Tien), telephone interview with author, Singapore, 16 March 2015.
Welch, Ann (daughter-in-law of Samuel Hawks Caldwell), telephone interview with author, 5 September 2013.
Cable and Wireless Archive, Porthcurno, UK (CAWA)

Published Sources

Baark, Erik. *Lightning Wires: The Telegraph and China's Technological Modernization, 1860–1890*. Westport, CT: Greenwood Press, 1997.
Byrne, David. *Playing the Building* (2005). Available at www.davidbyrne.com/archive/art/art_projects/playing_the_building/ (accessed 4 September 2018).

Caldwell, Samuel H. “Ideographic Type Composing Machine.” United States No. 2950800. Filed 24 October 1956. Patented 30 August 1960.

_____. “The Sinotype—A Machine for the Composition of Chinese from a Keyboard.” *Journal of The Franklin Institute* 267, no. 6 (June 1959): 471–502.

_____. *Switching Circuits and Logical Design*. New York: John Wiley & Sons, 1958.

Cheng, W. K. “Enlightenment and Unity: Language Reformism in Late Qing China.” *Modern Asian Studies* 35, no. 2 (May, 2001): 469–93.

Darlow, T. H., Horace Frederick Moule, and A. G. Jayne. *Historical Catalogue of the Printed Editions of Holy Scripture in the Library of the British and Foreign Bible Society*. New York: Kraus Reprint Corp, 1963.

Headrick, Daniel. *The Tentacles of Progress: Technology Transfer in the Age of Imperialism, 1850–1940*. Oxford: Oxford University Press, 1988.

Kaske, Elisabeth. *The Politics of Language in Chinese Education, 1895–1919*. Leiden: Brill, 2008.

Mullaney, Thomas S. *The Chinese Typewriter: A History*. Cambridge, MA: MIT Press, 2017.

_____. “Semiotic Sovereignty: The 1871 Chinese Telegraph Code in Global Historical Perspective.” In *Science and Republican China*, edited by Jing Tsu and Benjamin Elman, 153–84. Leiden: Brill, 2014.

Standage, Tom. *The Victorian Internet: The Remarkable Story of the Telegraph and the Nineteenth Century’s On-line Pioneers*. New York: Berkeley Books, 1999.

Tien, H. C. “On Chinese Acupuncture and Chinese Language.” *World Journal of Psychosynthesis* 4, no. 6 (June 1972): 4–5.

_____. “On Learning Chinese.” *World Journal of Psychosynthesis* 4, no. 7 (July 1972): 4–5.

_____. “The Transalphabetization of Chinese Characters. I: The Letter-Doubling Technique.” *World Journal of Psychosynthesis* 4, no. 7 (July 1972): 35–38.

_____. “The Transalphabetization of Chinese Characters. IV: The Isomorphic Principle of the Silent Letters (V, W, X, Y, and T).” *World Journal of Psychosynthesis* 4, no. 10 (October 1972): 32–35.

_____. “The Transalphabetization of Chinese Characters. VI: The Radical Transalphabet.” *World Journal of Psychosynthesis* 4, no. 12 (December 1972): 30–31.

Tsu, Jing. *Sound and Script in Chinese Diaspora*. Cambridge, MA: Harvard University Press, 2011.

Wang, Jingchun. “The New Phonetic System of Writing Chinese Characters.” Jiaotongbu pinyin dianbao yanjiuhui yiding [交通部拼音電報研究會議定]. Jiaotongbu guiding guoyin dianbao fashi [交通部規定國音電報法式]. 1928.

Wang, Kaijie 王開節, *Wo guo dianxin fazhan jianshi* 我國電信發展簡史 [The Development of Telegraphy in China—A Brief History]. Taibei: Zhongguo jiaotong jianshe xuehui 中國交通建設學會, 1954.
Wilkinson, Endymion. *Chinese History: A New Manual*. Cambridge: Harvard University Asia Center, 2012.
Zhu, Jiahua [Chu Chia-hua]. *China's Postal and Other Communications Services*. Shanghai: China United Press, 1937.