

Ground Writing v3

Ground Writing

-

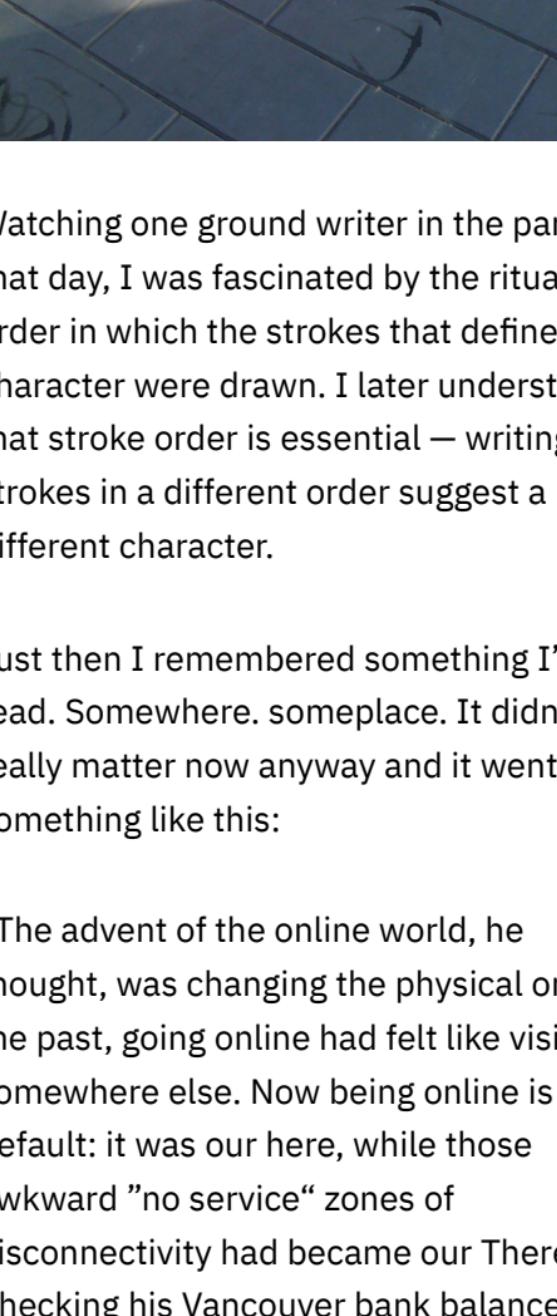
I tapped the first three letters into the omnipresent search bar on my Team 6AC:

"w-a-t..."

Software cook over and finished my thought, soon searching the Internet for "water":

Water, a substance composed of the chemical elements hydrogen and oxygen and existing in gaseous, liquid, and solid states. It is one of the most plentiful and essential of compounds. A tasteless and odorless liquid at room temperature, it has the important ability to dissolve many other substances. Density: 997 kg/m³ Boiling point: 212°F (100°C) Molar mass: 18.01528 g/mol Melting point: 32°F (0°C) Formula: H₂O

水



I remembered the water writers I'd seen about eight years ago in Jingshan Park right next to the Forbidden City in the middle of Beijing. They were mostly older men but not exclusively. One woman I remembered in particular because she held brushes in both hands and wrote two columns of mirrored calligraphy at the same time.

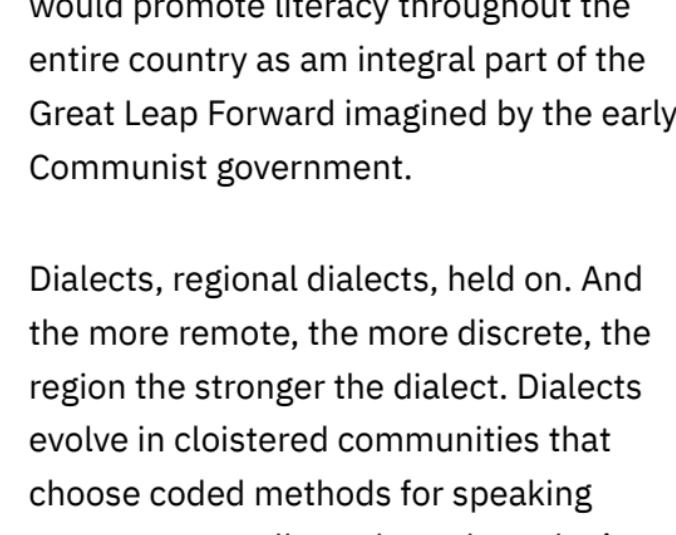
Water calligraphy, better known as Ground Writing or 地书 (Pinyin "Dishu") in Chinese, appeared at the start of the 1980s in Beijing. This was around the time of the "Reform and opening up" policy and since, ground writing has spread to public parks all over China. Most of the writers are retired. One 79-year-old described, "This is just for fun! You have to do something when you get old right?"

The calligraphers write poetry, mostly short traditional verse. They use homemade brushes fashion from broomsticks or umbrella sticks and hand-cut brushes made from sponges, rags, mops, whatever. They write directly on the ground, in public spaces. Since it is written in water not ink, it is explicitly ephemeral.

In Jingshan Park, there were about five writers position with a graceful distance between each other and in close proximity to where newspapers hung in public for reading. The proximity, even overlap, of reading and writing in this public space made a strong impression and that arrived back in full force at the front of my brain just now. I recalled the story of British Romantic poet John Keats who died very young and asked to be buried in an unidentified grave marked only by the epitaph:

Here lies One Whose Name was Writ in Water.

All writing should be done like that, I thought. Carefully. Temporarily. In public.



Watching one ground writer in the park that day, I was fascinated by the ritualized order in which the strokes that define any character were drawn. I later understood that stroke order is essential — writing strokes in a different order suggest a different character.

Just then I remembered something I'd read. Somewhere. someplace. It didn't really matter now anyway and it went something like this:

"The advent of the online world, he thought, was changing the physical one. In the past, going online had felt like visiting somewhere else. Now being online is the default: it was our here, while those awkward "no service" zones of disconnectivity had became our There. Checking his Vancouver bank balance from an ATM in Los Angeles struck him suddenly as spooky. It didn't matter where you were in the landscape; you were in the same place in the datascape. It was as though cyberspace were turning inside out, or 'everting' — consuming the world that had once surrounded it."

I focused on my phone and the Telegram app as I scrutinized the characters trying to perform a reverse lookup in my brain, to find a connection, a tendril line between shape and meaning.

I moved up six years ago from Los Angeles. Now I stood somewhere south of 24th St., just at the edge of what was used to be called the Northeast mission and now is most certainly Noe Valley. Who cares. I returned my full attention to the interface. The first character was easy enough but there were five of them and I recognized the second as the sign for "water." If I didn't know that much by now I may as well give up on my little self-improvement project.

The message came from about halfway around the world, from some friends I met last summer, who were now busy organizing a large group of people. They use, almost exclusively, secure messaging. Telegram specifically. Myself, I was standing in a fairly tight perimeter of our corporate offices in downtown San Francisco.

Telegram was introduced as a simple messaging service and as an alternative to the standard iOS or android messaging apps. It was not owned by a social network although it was directly in competition with WhatsApp. Telegram is encrypted. Every message along its entire route is unreadable by both the servers that touch it, and also by any of the intermediate software that routes a message from one place to the other.

Public / private secure key pairs at either side of the transmission between trusted users encode and then decode the message only at the sender and receiver. There's no data capture even possible en route.

But alternate character text input on the app was awkward, well at least for Chinese traditional or simplified characters. All of this was brand new for me. Nonetheless I was compelled to learn how to write Chinese, at least some subset of characters. Tackling the full collection of characters which make up traditional Chinese writing systems was completely impossible. I was less interested in learning how to speak (a much more common approach from a distance). The characters are logograms where each one stands for a syllable in a compound word and that works very differently from an alphabet of sounds produce words. Pronunciations and vocabulary differ between languages, of course. To me, it seemed that maybe writing in logograms was simpler and more direct, something more like traversing the world in images, somehow just less linear and literal.

So on New Year's Eve five years ago at a party in Santa Clara, I made a resolution. I was always good at both making New Year's resolutions and following through with what I planned. So this year when I lay down the challenge of grappling with some tens of thousands of discrete characters, I knew it would be a busy year.

I started by visiting the Wikipedia page "Chinese text Input by Computer." First the simple bits regarding the language itself, not exactly quoted more like a gloss:

Traditional Chinese consists of 70,000 characters. Learning to read and write it is complicated. Simplified Chinese was introduced around the cultural revolution in China and reduced the character set to somewhere around 20,000 (8,105) unique logograms. The not-entirely-unique base shapes are combined to make words which might refer to things or concepts. Simplified Chinese was intended to be much easier to learn to read and write and would promote literacy throughout the entire country as an integral part of the Great Leap Forward imagined by the early Communist government.

Dialects, regional dialects, held on. And the more remote, the more discrete, the region the stronger the dialect. Dialects evolve in cloistered communities that choose coded methods for speaking among some cells so that others don't understand is said. As a result, connections become stronger, but also

more fluid. The speed of dialect creation in dataspace is much faster and in every way more fluid.

I just remembered something I'd heard before:

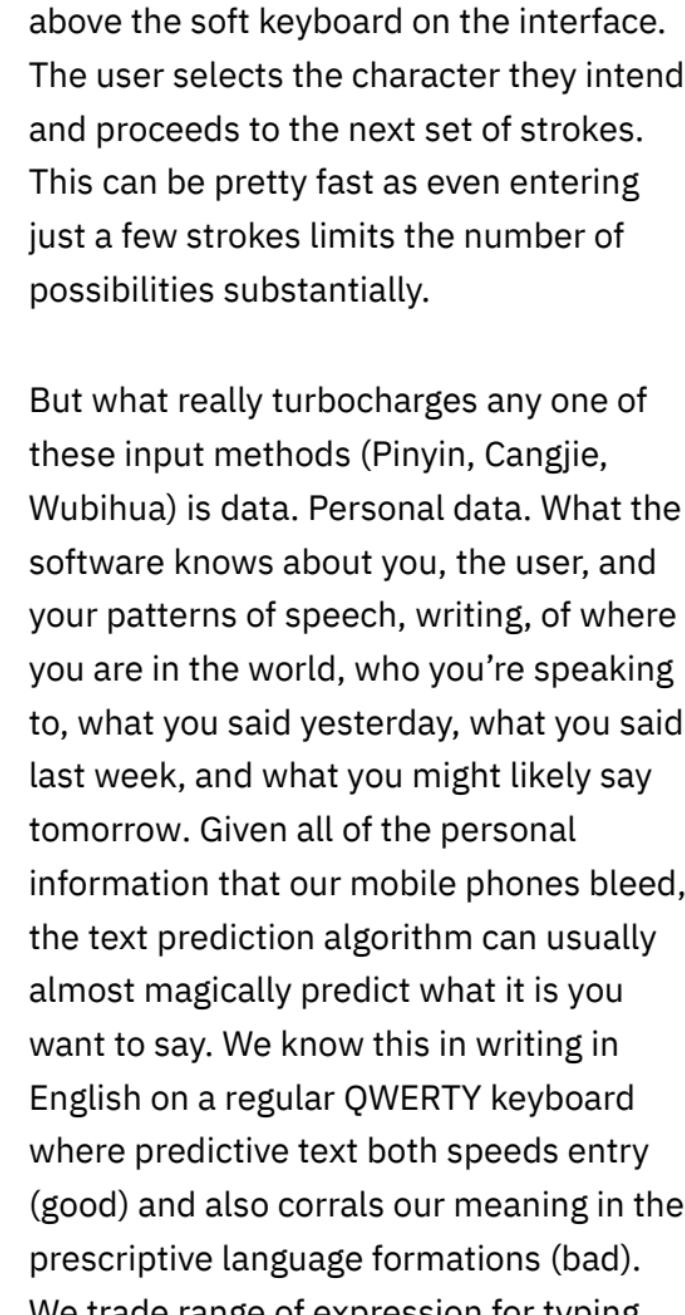
"Be like water."

Cheesy! A quote by Bruce Lee from a 1953 movie. Why do I keep these things in my brain. And still, the rest follows:

"Don't get set into one form, adapt it and build your own, and let it grow, be like water. Empty your mind, be formless, shapeless — like water. Now you put water in a cup, it becomes the cup; you put water into a bottle it becomes the bottle; you put it in a teapot it becomes the teapot."

In Hong Kong, I understand that students learn to write traditional Chinese in school but learn to speak Cantonese. Cantonese is a distinct spoken dialect. Cantonese emerged in southern China during the Song Dynasty about 1000 years ago. Mandarin is another dialect which evolved in the north under the influence of neighboring cultures, and remains the official state spoken language of China. The two dialects are distinct and mutually unintelligible. I understand that poetry in Mandarin doesn't rhyme. In Cantonese, it does. Both are written, however, using the same traditional Chinese characters. Writing for me, like anyone my age, is mostly on my phone. Mine's brand new, flexible, and waterproof — pretty slick. The text entry interface which shipped with the phone perhaps hadn't caught up to its external form with its grapheme coating to keep the device cool and unbreakable diamond glass on the front. Absolutely edge-to-edge, back and front interface. And still, referencing these characters was awkward. Writing in characters was a chore (it was also new to me) and so I tried a few of the different methods commonly available. I understood from talking to friends that if you ask 10 different Chinese speakers to type the same message in traditional characters on their phone, each of the 10 would do it a different way. Some of the basics to try first:

The most popular is Pinyin. It's based on phonetics, although perhaps surprisingly is typed in Roman characters with English pronunciation. Using a standard QWERTY keyboard, the user types in a phonetic approximation of the sound of the character they're searching for. After entering perhaps three English letters then a constellation of possible Chinese characters which begin with that sound appear above the keyboard to choose from. This method of text input relies on being able to speak the language, but there's no need to be able to write it, exactly. Recognizing the symbol is enough, rather than recalling it from memory and knowing how to write it. (I don't like this about it.)



It is, however, the most popular way to enter text but it comes at the expense of the persistence of written Chinese language. The more people know how to speak it and say it and approximate it with English characters, the less essential it is to know how to write Chinese in the first place.

Other methods use the shapes of letters. The most common of these methods is Cangjie. Cangjie is a set of 24 root characters, or radicals, which are used in combination to specify a more complex Chinese character. The system was developed around 1976 when it became obvious that text entry on computers would soon dominate written communication. Computers use the keyboard already, so having a Chinese "alphabet" of sorts was a handy solution to the unwieldiness of the written language. 24 was also a handy number seeing as it neatly mapped on to the existing hardware.

The first commercially successful typewriter was developed from around 1868 by Americans Christopher Latham Sholes, Frank Haven Hall, Carlos Glidden, and Samuel W. Soule in Milwaukee, Wisconsin. An early market for the writing machine, was the telegraph industry. Telegrams were messages sent between geographically distant points on electric wires, allowing long-distance communication, a frighteningly new thing in the middle of the 19th century. Telegraph transmissions were sent using Morse code, a combination of short and long electric pulses which stand for each of the letters of the English alphabet. The typewriter might then be used to transcribe a telegraph transmission, and produce a clear and human-readable document. Sholes, Hall, Glodde, and Soule's first client for their writing machine was nearby Porter's Telegraph College in Chicago.

In 1868, their first machine used a keyboard of 28 keys, laid out like a piano with A-N left-to-right on the black keys, and O-Z from right-to-left on the white keys. It looked like this:

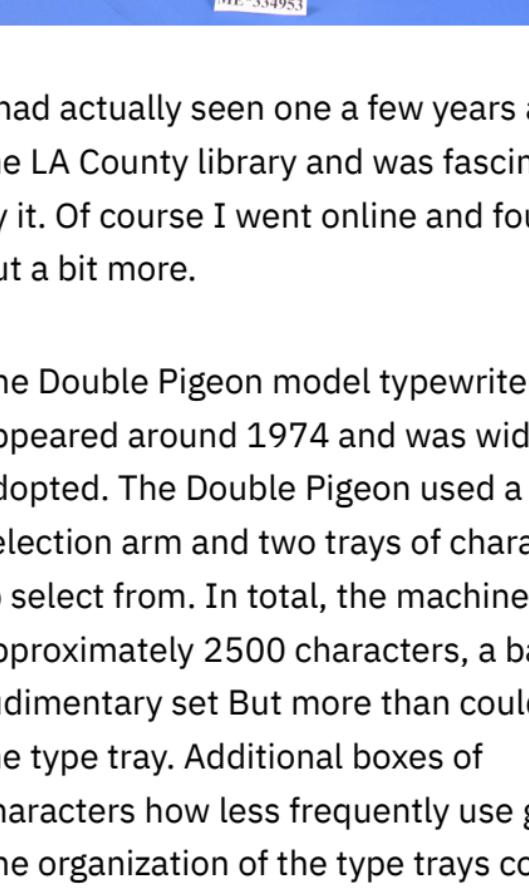


Fig. 1. Keyboard of Hughes-Phelps Printing Telegraph [4]

You'll notice there are no number keys. Porter's Telegraph College noticed this also and insisted that a more compact layout be instituted. A couple of years later the keyboard had been significantly re-organized into four rows, with the bottom two rows made up of consonants in alphabetical order, the third row with vowels and punctuation, while the top row had the digits 2-9 plus a hyphen. There was neither a "1" nor a "0" key:

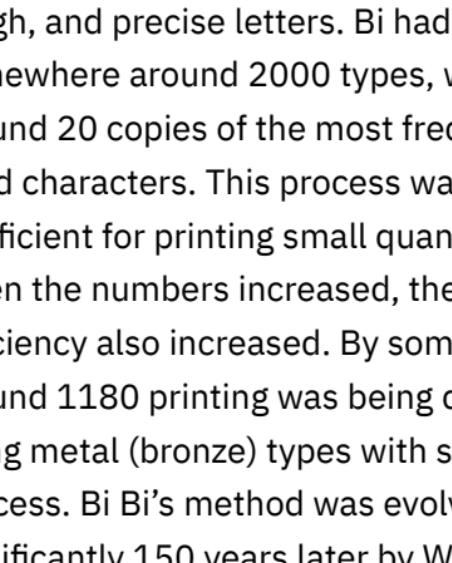
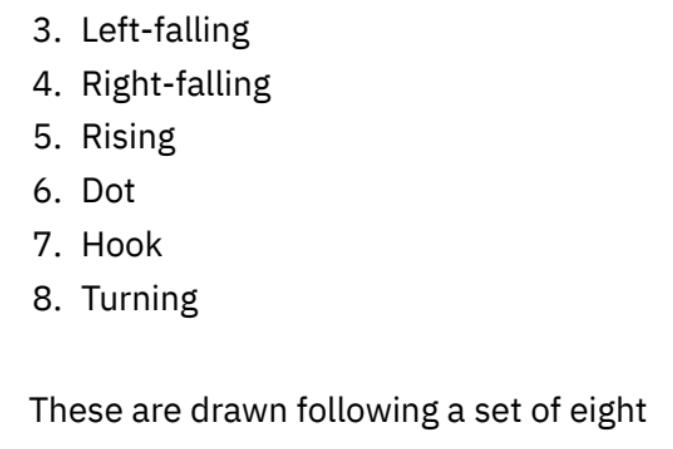


Fig. 2. Presumable keyboard arrangement in April, 1870

For cost efficiency and manufacturing simplicity, "1" was typed as the letter "I" and the letter "O" was used for "0." When the machine was offered to the American Telegraph Company, additional refinements were made with the frequently typed letter "T" moved to the top center of the keyboard, the W moved to the top row with the vowels, and the less-used "Q" moved to the left-edge of the keyboard.

By 1872, when the machine was presented to the Western Union Telegraph Company, additional keyboard layouts were suggested by the telegraph operators. Since telegrams used numbers frequently in the short messages, "I" and "O" were moved to the third row far right, close by and just underneath the rest of the digits. The hyphen key was placed at the top right of the keyboard, as "—" was used to identify paragraph breaks in a telegram. Looking at a QWERTY keyboard today you can see the vestiges of these decisions with the I, the O, and the hyphen remaining in these positions.

Chinese typewriters were something else. To mechanically arrange the sequence of characters possible in as many as even a bare minimum of 2500, distinct metal pieces had to be arranged in some kind of accessible and indexable manner. This required more than one tray of characters and a selection mechanism for choosing the required character. Chinese typewriter designs often employed more than one deck of metal characters arranged in so many trays. The most successful of these typewriters was called the Double Pigeon.

I had actually seen one a few years ago at the LA County library and was fascinated by it. Of course I went online and found out a bit more.

The Double Pigeon model typewriter appeared around 1974 and was widely adopted. The Double Pigeon used a selection arm and two trays of characters to select from. In total, the machine had approximately 2500 characters, a barely rudimentary set but more than could fit in the type tray. Additional boxes of characters how less frequently use glyphs. The organization of the type trays could be changed to suit the user. Individual characters are selected by a movable arm along a grid to select a character, which is picked up and used to print on a paper roll. The machines had no keyboard. Using the a Double Pigeon was complex operation with the operator hunting and searching and selecting and printing and returning and hunting and searching and selecting and printing and returning in an intricate dance of body and memory. (MC Hammer even named the signature dance move from his Grammy-winning hit "You Can't Touch This" (1990), "The Chinese Typewriter.")

The machines reminded me of my classes in metal typesetting while a student at Stanford. Typewriters brought the crisp and repeatable printing of movable type letterpress setting to an individual.

Movable type was invented in China during the Northern Song dynasty around 1040 AD by inventor Bi Sheng. He first used wooden individual characters set together in an iron frame to make multiple copies of a text. However the woodgrain was a problem, making the character less legible and the letter also less durable. Instead he invented a working system using ceramic characters, specifically porcelain, which resulted in extraordinarily durable, tough, and precise letters. Bi had somewhere around 2000 types, with around 20 copies of the most frequently used characters. This process was inefficient for printing small quantities, but when the numbers increased, the efficiency also increased. By somewhere around 1180 printing was being done using metal (bronze) types with some success. Bi Bi's method was evolved most significantly 150 years later by Wang Zhen who created 30,000 characters (this time of wood) and also invented a sorting method to find a specific character when needed, as needed. As well, he published a book describing his process called *A method of making movable wooden types for printing books*. (It wasn't until the middle of the 15th century that printing with movable type appeared in western Europe.)

With movable type as well as for Chinese typewriters, the most perplexing problem was always how to arrange the characters so that they could easily be found when needed. This organization scheme, protocol, or lookup table, is absolutely essential to make typesetting or typewriting (or later computer text entry) in Chinese characters practical.

Approximately 3,000 to 4,000 characters are needed to read and write Chinese fluently. More like 40,000 characters are often used to set complicated texts.

Around 1959 in Cambridge, Massachusetts at MIT, Dr. Samuel Caldwell approached this problem of Chinese text input. Caldwell was an electrical engineer and a computer scientist before there was really either computer science or much in the way of practical computers. However, he became interested in how to design a computer which would understand and produce written Chinese. This was at the height of the Cold War, and the United States "needed a win," according to Stanford historian Tom Mullaney. The Soviet Union had recently launched the Sputnik satellite into space and in 1958 Mao had instituted China's Great Leap Forward. If the United States could develop a computer that operated in Chinese, the possibility of producing text and therefore propaganda was enormous. Faced with this challenge, Dr. Caldwell consulted Chinese graduate students over casual group dinners and learned something surprising to him. Written Chinese has an orthography. A Chinese character is always written the same way with strokes drawn in a prescribed order which follows a set of discrete rules. Further, there are a limited number of stroke types, eight main categories are agreed on today.

These are drawn following a set of eight rules which determine the stroke order:

1. Horizontal
2. Vertical
3. Left-falling
4. Right-falling
5. Rising
6. Dot
7. Hook
8. Turning

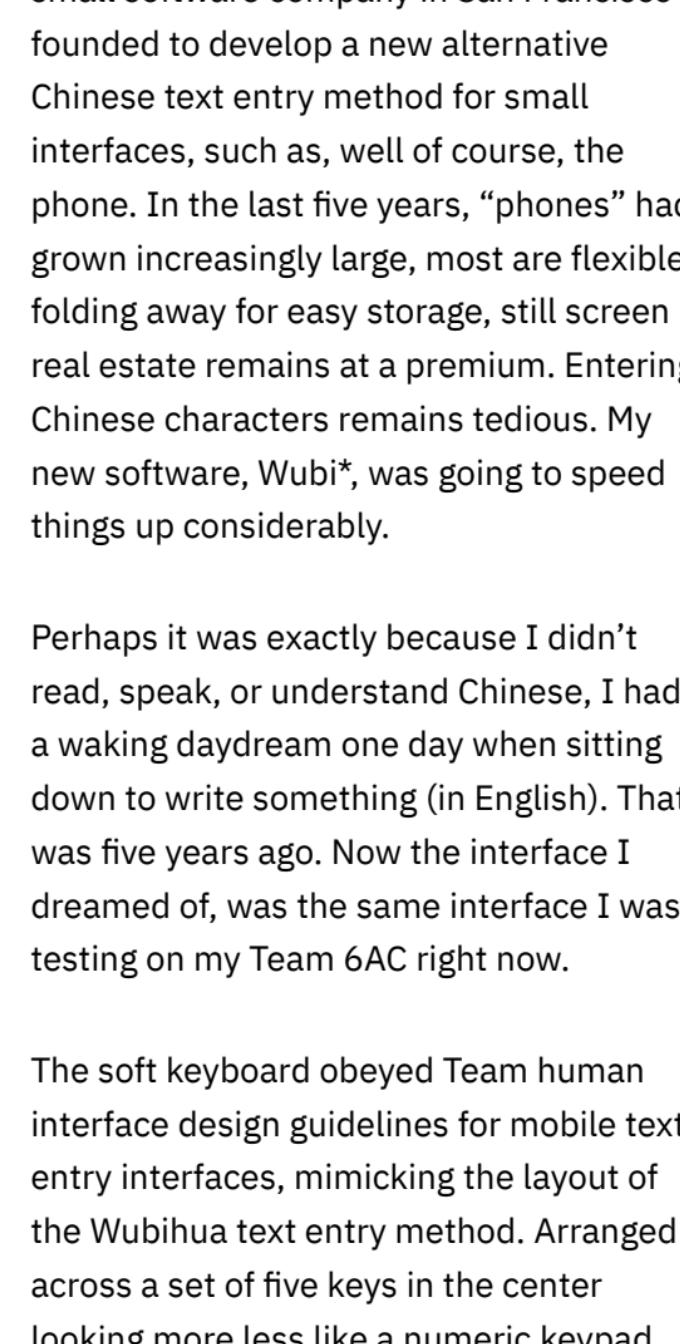
These are drawn following a set of eight rules which determine the stroke order:

1. Horizontal then vertical
2. Left-falling then right-falling
3. Top to bottom
4. Left to right
5. If framed from above, frame first
6. If framed from below, frame last
7. Frames closed last
8. If symmetrical, middle first then sides

These rules determine a definite order, which when combined with a limited collection of strokes add up to a working orthography for writing Chinese. This was a revelation for Dr. Caldwell and his team. If Chinese characters had a definite construction order and they are made up from a basic set of strokes, then a particular character could be looked up by writing the strokes in the correct order. He soon also realized something even more remarkable which was that Chinese stroke order lookup is highly redundant. And so a character maybe made up of as many as 13 strokes or more however entering just the first four or maybe five strokes was enough to definitively identify the character in a lookup table. This logic worked beautifully on a computer which is designed to process large amounts of data and can index a large database from a particular query. Given the first four or five strokes as a query the computer could quickly find the matching character from a database of 40–70,000 possibilities. This lookup query and matching made Chinese possible and practical to enter on a computer. Typing was actually more like searching.

Given this collection of strokes and the set of rules, Dr. Caldwell landed on a 22-character “alphabet” of sorts. Working with a team including a professor of near Eastern languages at Harvard, Caldwell identified a small set of root glyphs, combinations of strokes written in a certain order, which frequently appear in written Chinese characters. Of course this wasn’t really an alphabet. In an alphabet, individual characters stand for specific sounds in a spoken language, and in this case the glyphs instead are only keys used to lookup the desired character in an indexing system. A 22-glyph set was convenient for mapping Chinese text entry onto the existing English QWERTY keyboard. A typist at this keyboard need only the first few “letters” of a Chinese word in order to identify the correct character. Rather than typing or spelling a word completely, this was more of a case of describing the intended word, searching a database and selecting the correct result. This team of engineers and linguists had unwittingly invented auto-completion.

Chinese characters stand for single syllables. This derived from the fact that most ancient Chinese words were only one syllable. (This remains the case today.) As the language expanded and forked into the 70,000-character set of classical Chinese, words also became more complex. Characters adapted and it often required more than one character to represent a word. More often, characters are combined to make composite characters. Chinese writing evolved a clever graphic technique of combining multiple root glyphs into one character. Characters must always fit in a square. This is the case for single characters. And this is also the case when combining one or more characters into a compound character. in a square. By simply horizontally compressing the constituent characters, the resulting compound glyph still fits neatly in a square. An asterisk-shaped gridded paper is used when learning Chinese calligraphy for this purpose.



Stroke order in Chinese writing is fundamental. It can be thought of, as Dr. Samuel Caldwell at MIT realized at the end of the 1950s, as a kind of “spelling.” Spelling in English determines the order that you write the letters that make up a word. In Chinese, rules determine the order that you write the strokes to make a word (character).

Writing “water” in Chinese (following the rules of stroke order), the first mark to draw is the long stroke in the center from top to bottom. The next stroke is a falling left stroke which connects to it. The next is the rising right stroke in the top-right of the character, and the final stroke is a downward right following stroke. The sequence looks like this:

|
| →
| → J
or
水

English also relies on written order of the component letters that make up a word. So for example, writing “water” looks like this:

W
W-A
W-A-T
W-A-T-E
W-A-T-E-R

WATER Making the marks in a different order (left-to-right) produces a different word, order produces meaning:

TAWER (a leather worker)

Or several: A WET “R”

You *could* write the letters in a different temporal sequence, first write the “W” then write the “E” then write the “A” then right the “R” then write the “T” but in their regular linear positions. However nobody writes English like that. The orthography demands that it is always written from left to right. In English, you can read the order of writing, left to right, in the spelling of the word.

It’s not like that in Chinese. As Chinese characters always fit neatly into a square, then the stroke order or the order in which the marks were made is illegible in the final character. If an English word is a linear sequence of the order of writing, a Chinese character is not nearly so forthcoming. The final mark is a temporal whole where the stroke order of its writing is invisible.

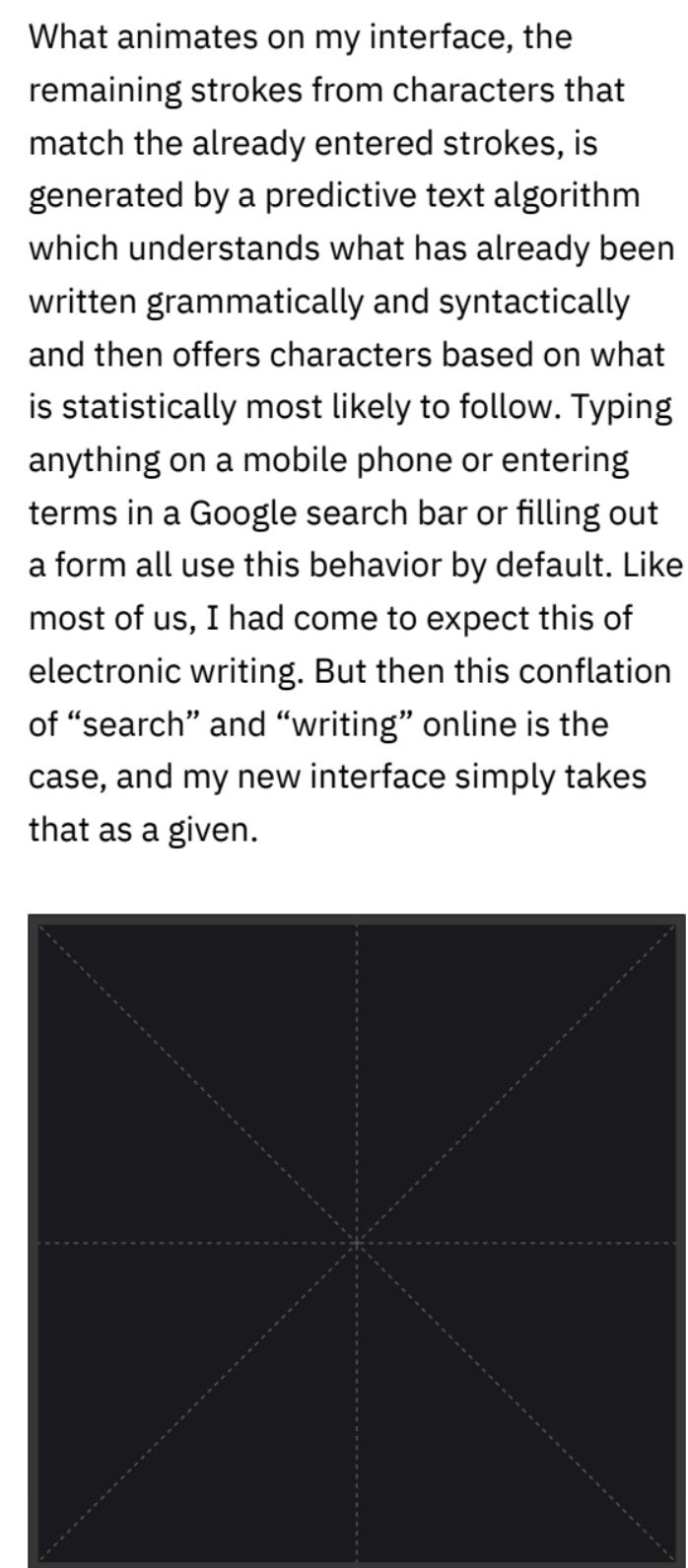
Years, ago seeing the old man in Jingshan Park carefully perform his ground writing was a revelation. And after bit of time marinating in my brain, it emerged as a really good, distinctly new idea for Chinese text input and formed the basis for my current work (five years after that fortuitous New Year’s resolution) with the small software company in San Francisco I founded to develop a new alternative Chinese text entry method for small interfaces, such as, well of course, the phone. In the last five years, “phones” had grown increasingly large, most are flexible, folding away for easy storage, still screen real estate remains at a premium. Entering Chinese characters remains tedious. My new software, Wubi*, was going to speed things up considerably.

Perhaps it was exactly because I didn’t read, speak, or understand Chinese, I had a waking daydream one day when sitting down to write something (in English). That was five years ago. Now the interface I was testing on Team 6AC right now.

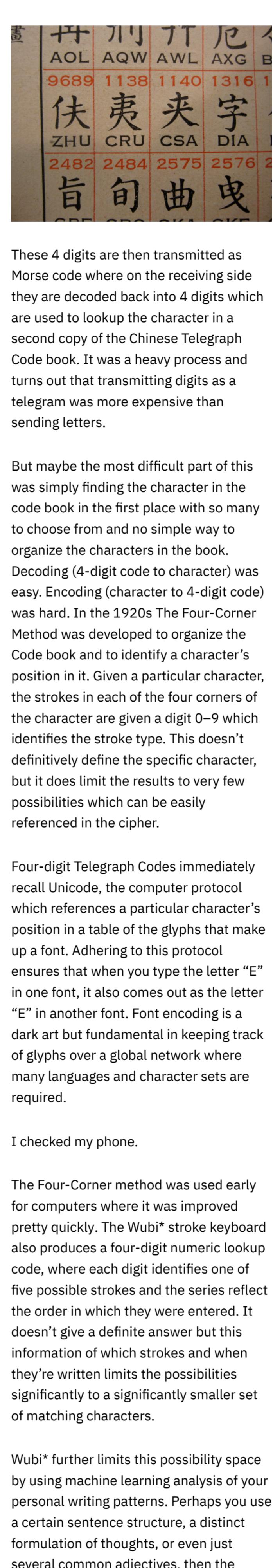
The soft keyboard obeyed Team human interface design guidelines for mobile text entry interfaces, mimicking the layout of the Wubihuā text entry method. Arranged across a set of five keys in the center looking more less like a numeric keypad from top left to bottom right appeared the same five constituent strokes. When typed in a specific order these also produce a series of possible Chinese characters. This worked essentially the same as Wubihuā.

What I changed was the display of the possible matches. Typically as a user enters the stroke radicals on a Wubihuā keyboard, possible Chinese characters appear above which match the order of the strokes entered so far. Each stroke typed limits the number of characters that match those strokes in that order.

Instead of a static table of possibilities organized on an expandable row, my display showed a flickering compound display of possible matches. Organized against the scaffold of an 8-pointed asterisk-shaped grid, the strokes I entered appeared in the order that they were entered surrounded by the ambient flicker of an animation of all possible characters which fit this stroke sequence. It reminded me of that scene from The Matrix, a 90s film starring Keanu Reeves, where the matrix is shattered into a sparkling green curtain of data, each character changing into every possible other character.



What animates on my interface, the remaining strokes from characters that match the already entered strokes, is generated by a predictive text algorithm which understands what has already been written grammatically and syntactically and then offers characters based on what is statistically most likely to follow. Typing anything on a mobile phone or entering terms in a Google search bar or filling out a form all use this behavior by default. Like most of us, I had come to expect this of electronic writing. But then this conflation of “search” and “writing” online is the case, and my new interface simply takes that as a given.



Entering Chinese characters in particular into a text interface is an exercise in “describe and retrieve.” Given the smallest amount of descriptive information (search terms), a database (the complete Chinese character set) is queried and possible matches (specific characters) are returned as results. Entering Chinese text into a machine has always relied on a lookup process like this.

The challenge of sending Chinese language over a telegraph line required a similar system. Telegrams are transmitted in Morse code, where strings of dots and dashes correspond to the letters of a Latin alphabet and digits. This would never work for 70,000 characters, so a secondary encoding was necessary. Chinese Telegraph Code was developed to address the issue and it maps a subset of approximately 7000 characters to discrete 4-digit numbers which map the characters lookup position in a code book. The code book is 100 pages, each organized as a 10 row by 10 column table of characters. The first two digits in the code identify the page number, the next identifies the row, and the last digit identifies the column.

These 4 digits are then transmitted as Morse code where on the receiving side they are decoded back into 4 digits which are used to lookup the character in a second copy of the Chinese Telegraph Code book. It was a heavy process and turns out that transmitting digits as a telegram was more expensive than sending letters.

But maybe the most difficult part of this was simply finding the character in the code book in the first place with so many to choose from and no simple way to organize the characters in the book. Decoding (4-digit code to character) was easy. Encoding (character to 4-digit code) was hard. In the 1920s The Four-Corner Method was developed to organize the Code book and to identify a character’s position in it. Given a particular character, the strokes in each of the four corners of the character are given a digit 0–9 which identifies the stroke type. This doesn’t definitively define the specific character, but it does limit the results to very few possibilities which can be easily referenced in the cipher.

Four-digit Telegraph Codes immediately recall Unicode, the computer protocol which references a particular character’s position in a table of the glyphs that make up a font. Adhering to this protocol ensures that when you type the letter “E” in one font, it also comes out as the letter “E” in another font. Font encoding is a dark art but fundamental in keeping track of glyphs over a global network where many languages and character sets are required.

I checked my phone.

The Four-Corner method was used early for computers where it was improved pretty quickly. The Wubi* stroke keyboard also produces a four-digit numeric lookup code, where each digit identifies one of five possible strokes and the series reflect the order in which they were entered. It doesn’t give a definite answer but this information of which strokes and when they’re written limits the possibilities significantly to a significantly smaller set of matching characters.

Wubi* further limits this possibility space by using machine learning analysis of your personal writing patterns. Perhaps you use a certain sentence structure, a distinct formulation of thoughts, or even just several common adjectives, then the system learns this over time. Wubi* observes all of your writing, mails, texts, telegrams, and harvests all of this extra-linguistic information which, together, provides a composite picture of your writing style. With this as a predictive template, Wubi* more easily and more accurately guesses what character you are looking for next. This is something like machine-assisted search or auto-completion (neither of which is exactly brand new tech), but there is also more.

Wubi* flashes each possible next stroke at a rate faster than your conscious eyes can keep up. Still, the more the user learns to let go and trust her intuition, engaging the full capabilities of the human body to identify one in a large sequence of possible answers, the more fluid the writing becomes. Like writing in water, this gets better both with practice and by thinking about it less. Writing is then freed from the tiny screen and a pair of clumsy thumbs and returned to the body, as it was meant to be. Faster, sure – but more importantly, writing becomes more liquid, less forced. The consequences remain to be seen.

I checked my phone again. I entered a vertical stroke first. My display hummed, blitz-animating any of 4000 potential matching characters over the asterisk grid.

2

With each stroke entered on a Wubihua keyboard, the number of matches reduces significantly. It’s like entering more and more precise google search terms.

I entered the next stroke, →, a hook. The display slowed a bit, barely noticeable as the search space narrowed to hundred results and became a measure less frenetic. Cues offered by my interface indicate the most likely matches. I’d recruited one of the early search algorithm wizards from Alphabet (Google) to streamline the character search. Latency was not acceptable when you’re writing. Latency was hardly even a thing these days.

2-5-3

“For us, of course, things can change so abruptly, so violently, so profoundly, that futures like our grandparents’ have insufficient ‘now’ to stand on. We have no future’s because our present is too volatile.”

I clicked my display again, entering one final stroke, right-falling (↖).

2-5-3-4

“But intuition is real, and texts and artworks take on lives of their own, and sometimes it feels like technology does, too. It can seem like you’re saying the future. Really you’re just participating in history.”

My display came to a definitive halt on one matching character waiting to be accepted. I tapped, selected, and pressed “send.” The Telegram left my immediate concern arriving on the opposite side of Earth, where the character landed on someone’s device, was correlated to its

5

Unicode position, drawn in a simplified Chinese font, displayed, read, understood, and furtively distributed. Then something happened.

+