

The Eurasian origins of pinyin

Ulug Kuzuoglu

In 1931, Chinese and Russian revolutionaries in the Soviet Union joined hands to devise the Chinese Latin Alphabet (*Latinxua sinwenz*, CLA), the mother of contemporary *pinyin*. While the significance of the CLA in the history of Chinese language and script reforms is beyond doubt, its exact kinship to *pinyin* remains murky, for the CLA was markedly different from *pinyin* not only in its ideological make-up, but also in its letter-composition. *Yuyan* 语言 [language], for instance, was written as *yjan* in the CLA in the 1930s; or *Ladinghua* 拉丁化 [Latinization] was written as *latinxua*. The 'y' of *pinyin*, in other words, was originally written with a 'j'; 'h' was written with an 'x'; and 'yu' or 'ü' with a 'y'. Trivial as it may seem, an archaeology of these letters offers a history unlike the ones written before.

Latinization in China has largely been described either as the product of nationalism and modernization, or the outcome of Western colonialism and social Darwinism. The history of the CLA, however, compels us to question these established narratives and resituate Chinese Latinization within a greater Eurasian media history that stretches from the Ottoman Empire to the Soviet Union, and explores the technological impact of new communications technologies on non-Latin scripts, of which Chinese was one. In particular, understanding the letters of the CLA requires a novel focus on the long history of Arabic script reforms across Eurasia.

Call for separate letters

In the 1850s, a few decades prior to the invention of the CLA, Muslim and non-Muslim reformers in the Ottoman Empire and Russian Transcaucasia began rethinking the possibilities and limitations of the Arabic script. Their intellectual tribulations started almost immediately after the introduction of the telegraph into the Muslim world during the same decade, which challenged the Arabic script and imposed a new epistemology of writing. Telegraphic messages were transmitted via the Morse code, which was originally invented for the 26 letters of English. As opposed to Latin letters, the majority of Arabic letters were written differently depending on their position in a word—the letter 'ghayn' غ, for example, was غ as the initial, غ in the middle, and غ at the end of a word. The Morse code, however, demanded the dissection of each word into its letter-sound components, and designated each letter-sound as a separate entity that corresponded to a combination of dots and dashes. As such, when transmitting the word 'ghayn-ghayn-ghayn' غغغ, the telegraph operators had to send each letter separately as غ غ غ.

During the following decade, intellectuals from the Ottoman, Russian, and Iranian Empires were in conversation with each other regarding the future of the script, and for them, even more pressing than telegraphic communications (with Morse code) was the question of typography. The industrialization of the printing press in the nineteenth century, exemplified by the global dissemination of the movable metal type, imposed a similar epistemology of 'separate letters' for typesetting. Because of the number of glyphs and the various ways of combining letters, an Arabic-lettered type case had more than 500 sorts, i.e., metal types; and depending on the kind of calligraphy used for printing (e.g., *ta'liq*) the number could surpass 2000. The reformers used the exorbitant size of type cases to justify their call for separate letters, which, they argued, would increase efficiency and optimize labor not only in typesetting, but also in reading. In the eyes of the reformers who were surrounded by Latin, Cyrillic, Greek, Armenian, and Georgian alphabets, the future of a productive knowledge economy lay in separate letters.

The first proposals for a new alphabet in the Muslim world were not based on the Latin alphabet (fig.1). Some argued for a reformed Arabic script written in separate letters, while others proposed the use of Armenian letters or a combination of Cyrillic and Latin letters. It was only in the 1910s, in the aftermath of the Russian imperial reforms of 1905 and the Young Turk Revolution of 1908, that the Latin alphabet emerged as a serious contender to other script proposals. The Bolshevik Revolution in 1917 and the defeat of the Ottoman Empire in the Great War in 1918 strengthened the position of the Latin alphabet even further, as many reformers saw it as the embodiment of material and mental progress during national reconstruction. Amid incessant debates in the post-imperial Russian and Ottoman space, the Soviet Socialist Republic of Azerbaijan was the first to officially Latinize its Arabic script in 1924.

Revolutionary, Dunganese, and Chinese Latin alphabet

The modernist values of efficiency, productivity, and progress that the Latin alphabet supposedly embodied turned out to be critical for the Bolsheviks, who promised a new socialist civilization based on the same principles. Indeed, during the 1920s, socialist thinkers in the USSR were preoccupied with the so-called Scientific Organization of Labor [*nauchnaia organizatsiia truda*], an industrialist creed imported to the Soviet Union by Aleksei Gastev (1882-1939). An aficionado of American Taylorism, Gastev established dozens of institutes and utilized cutting-edge technologies to optimize bodily movements and increase labor productivity. His philosophy of efficiency extended into the realm of language and writing as well, which he claimed could also be measured and optimized in order to achieve ultimate mental productivity.

The Latinization Movement in Azerbaijan was coming to fruition in the midst of these debates. Even Lenin himself took a personal interest, allegedly claiming that "Latinization [was] the Great Revolution in the East!" After all, eyes, hands, fingers, typewriters, printing presses, and telegraph operators all functioned within a network of humans and machines, linked via the script. The Latin alphabet seemed to offer the remedy that the socialists were searching for all along—in fact, many claimed that even the Cyrillic alphabet could be Latinized for an internationalist socialist civilization!

While the Latinization of the Cyrillic alphabet never took place, the revolutionary fervor reached a peak in 1926 with the First Turcology Congress in Baku, where more than a hundred participants representing various nationalities discussed the future of the alphabet in the world. After heated discussions, the Latin alphabet was selected as the new medium of intellectual production not only in the Turco-Muslim world under the USSR, but also in the rest of the non-

Western world where socialism offered a strong alternative to extant political and economic conditions. Thus, in 1928, the Unified New Turkic Alphabet (UNTA) was devised by the socialist internationalists to Latinize the Arabic script across the Turkic world (fig.2). And starting in the same year, the UNTA was exported to non-Turkic and non-Russian lands as well, forming the material basis of not only Latinized Kurdish, Persian, and Mongolian alphabets, but also of the first Chinese Latin Alphabet.

But before the Chinese Latin Alphabet, came the 'New Dunganese Alphabet' [*novyi dungsanskii alfavit*]. The Dungans were Chinese Muslims who had emigrated to Central Asia during the turmoil in Xinjiang in the 1870s, and settled in present-day Kyrgyzstan and Kazakhstan. While it remains an unexplored subject, the Chinese Muslims had been using the Arabic alphabet to transcribe Mandarin sounds, possibly since the thirteenth century, when the Mongol rulers invited Persian and Arabic scientists to present-day Beijing. The Chinese Muslims maintained the tradition of using Arabic letters, known as *xiaoejin* 小兒錦, to write in their own tongues until the twentieth century (fig.3). And in the 1920s, those in Central Asia, i.e., the Dungans, were already in the process of reforming the Arabic alphabet itself to bring it closer to Dunganese, a dialect of Mandarin. With the wave of Latinization, however, the Dunganese revolutionaries decided to adopt the UNTA, which was invented to Latinize the Arabic alphabet in the first place, and apply it to their centuries-old *xiaoejin*. That was the New Dunganese Alphabet (fig.4).

Pinyin, an imperial lingua franca

When the Chinese and Russian revolutionaries began working on the Chinese Latin Alphabet during the same years, they already had the New Dunganese Alphabet as a template to work with, and they by and large kept its letters for the CLA. The origins of the CLA's quixotic letters that I have noted at the beginning of this essay, may thus be revealed. Why was the 'y' of *pinyin* written with a 'j' in the CLA, as in *yjan* 語言? Or why was 'h' written with an 'x'; or 'ü' written with a 'y'? The response lies in the Latinization of the Arabic alphabet itself, because the 'j' of the CLA carried a secret Arabic *ya* ي; 'x' a secret *ha* ح; and 'y' a secret *waw* و, the historical traces of which have been forgotten in scholarship. The CLA of the 1930s, in short, was born out of a Eurasian history, written in Arabic letters.

Alas, it did not live long. Stalin's increasing Russification policies culminated in a USSR-wide Cyrillicization campaign in 1938, which brought an abrupt end to the revolutionary Latinization Movement. Following the failure of internationalism in the USSR and the start of the 2nd World War, the Chinese Communist Party shelved the CLA in the 1940s. And when *pinyin* was finally invented in 1958, it neither had the revolutionary audacity to eliminate the Chinese characters, nor the internationalist vision of a former era. The new letters of *pinyin* instead stood for Mandarin as an imperial *lingua franca* that the Han imposed on their multi-ethnic frontiers—but that will be the subject of a future essay.

Ulug Kuzuoglu, Program Director and Lecturer, Dept. of East Asian Languages and Cultures, Columbia University uk2123@columbia.edu



From left to right: Fig. 1: The Qur'an transcribed in an invented separate-lettered alphabet (Tbilisi, 1879). Fig. 2: The Unified New Turkic Alphabet of the USSR, 1928. Fig. 3: An example of *xiaoejin*: Arabic script used for transcribing Mandarin. Fig. 4: The New Dunganese Alphabet (1928) and the phonetic correspondence between the UNTA, *xiaoejin*, and Cyrillic letters.