

A social-constructivist approach to language diversity

By Tiansheng Sun

This post was developed as part of the Columbia University course “Multilingual Technologies and Language Diversity” taught by Smaranda Muresan, PhD and Isabelle Zaugg, PhD. This cross-disciplinary course offering was a joint effort between the Institute for Comparative Literature and Society and the Department of Computer Science, developed through the generous support of the Collaboratory@Columbia.

石室诗士施氏，嗜狮，誓食十狮。氏时时适市视狮。十时，适十狮
适市。是时，施氏适市。氏视是十狮，恃矢势，使是十狮逝世。

氏拾是十狮尸，适石室。石室湿，氏使侍拭石室。石室拭，氏始试
食是十狮。食时，始识是十狮尸，实十石狮尸。试释是事。

-- 施氏食狮史 Lion-Eating Poet in the Stone Den¹

This poem written in Classical Chinese was the brainchild of celebrated Chinese linguist Yuen Ren Chao, a pioneer of the Chinese Latinization movement that was once popular. The poem, although written using multiple Chinese characters, could not be understood when spoken in modern-day Mandarin because all the words in the poem are pronounced as *shi*, with only the tones differing. Chao used this extreme example to suggest that, although problems may arise with classical Chinese texts, modern everyday spoken Chinese would have little problem and could indeed have been latinized. His idea, interestingly, is similar to what Lydia Liu (2015) suggested as a typical Eurocentric idea of the dichotomy between orality and writing.

Outside China, the Latinization of languages can be seen throughout the world. On one hand, this movement has roots in colonialism. Vietnamese, for example, a language traditionally written in Chu Nam, was converted to Latin script under the French reign. On the other hand, writing is an imperial technology of the state and has deep political roots (Liu, 2015). As Toby Lester (1997) points out, Azerbaijan converted to Cyrillic script because the Soviets wanted stronger unity among its republics and then to Latin again to detach itself from the influence of Russia. In Mongolia, the Mongolian script was similarly converted to Cyrillic by the Soviets. What I see as the biggest problem among the movement of Latinization is the potential

detachment with the literature and history of a culture. John McWhorter (2009) argues that language loss entails not the death of culture but the loss of aesthetic meaning. While I appreciate the beauty of languages and agree that language does not represent everything of a culture, McWhorter (2009) did ignore the importance of languages in documenting a nation's identity and heritage. According to Yuen Ren Chao, Chinese language could have been latinized, and Chinese would still have been Chinese, celebrating Chinese festivals and eating Chinese food. Nevertheless, Yuen Ren Chao's poem shows how Latinization may create significant confusion about classical texts. If Chinese had indeed been successfully latinized, Chinese people would then have had a much harder time learning about its classical texts and would be so detached from China's amazing history. Luckily, that did not happen, and people in China currently have a huge interest in reading classical literature and bringing it to contemporary life. At school, I can easily learn about scripts and texts written more than a thousand years ago. I am so proud when I visit a temple and am able to read what was carved into the stone about its construction history.

Despite being latinized during British colonization, Malay also managed to preserve its traditional script. Before travelling to Brunei, I knew that Malay uses the Latin script. When I arrived at Bandar Seri Begawan in Brunei, however, I saw text written in another script as well at the airport. Upon doing some research, I learned that the script is called Jawi, which is the traditional script used to write Malay language. Along with the Latin script, the Jawi script is an official script in Brunei and is more commonly used in Brunei than Malaysia. Even in Malaysia, the Jawi script is used for religious and historical texts. Indeed, it may be easy to switch to Latin script, but Jawi script carries important religious and political meanings to both countries. Therefore, even without the death of a language, the disuse of a script will cause a significant loss of cultural heritage.

How, then, do we protect the diversity of languages? One very important approach is the use of technology and the internet. Through the effort of researchers and scientists, the digitization of languages and development of multilingual technologies have made significant progress. The development of computer input methods for Chinese and Japanese, for example, definitely contributed to the thriving of their characters and the demise of the Latinization movement. The ability to type Chinese characters according to word forms, such as the Wubi

input method, made Chinese a language that could surpass English in terms of typing speed. The development of natural language processing (NLP) technologies that allowed context to play a role in determining word choice also made new pronunciation-based Chinese input methods more viable in the twenty-first century. Although current technology development has favored major languages such as English, work such as that done by Anshuman Pandey, part of a larger group at Berkeley working on developing Unicode for scripts under threat, offers hope for the preservation of language diversity.

However, developing language technologies cannot be the sole focus of language justice. We need to take more actions, especially for those languages that would be termed borderline or digitally dead according to the classifications described by András Kornai (2013). Digital technology itself cannot address the reality and underlying social problems many languages face (Bird, 2020). As in many other fields that also deal with culture, such as archaeology, research about language technology and diversity, although with good intentions, retains elements of colonialism and remains defined mostly by white western scholars with their own cultural biases. It is also not possible for scholars to know all languages, leading people to haphazardly use the same approach for all linguistic communities without understanding that different languages are unique. Therefore, we should aim for a social-constructionist, community-based approach like the one proposed by Steven Bird (2020), letting local people participate in the process of decolonizing their own languages. We should pay particular attention to the education of young children and, at the same time, allow the adult native speakers and local communities to fully explore the possibilities for their languages. The creation of N’Ko script is a good example. Realizing the inefficiency of using Arabic or Latin script for the Mande languages of West Africa, Solomana Kanter, a local Ivorian merchant, developed a script to capture the language structure of Manden languages (Rosenberg, 2011). The effort of creating a script for Manden languages and using the script for teaching in countries where the literacy rate is still low are definitely important steps for revitalizing and decolonizing these languages and an achievement for which local communities should be credited.

Although multiple efforts have been made, McWhorter (2009)’s pessimistic yet realistic claim that many languages will die in the age of globalization may actually be true. It is also important to note that languages keep evolving and any effort to keep everything as it is will not

设置格式[dolly]: 两端对齐, 缩进: 首行缩进: 12.7 毫米, 段落间距段后: 12 磅, 行距: 1.5 倍行距, 不调整中文与数字之间的空格, 不调整西文与中文之间的空格, 无孤行控制, 定义网格后不调整右缩进

follow the trend of languages evolving and changing in the future. If so, why not focus on improving educational access and developing democratic institutions that enable people to make decisions about their own languages and advocate for their use? Norwegians, for example, voted through their preferences on social media to use Bokmål rather than Nynorsk in the digital sphere (Kornai, 2013). Although Kornai (2013) expressed this event with pity, it is worth noting that Norwegian people cannot use both forms of the language when writing and need to choose one regardless. Since Nynorsk is still digitally vital and neither form has a strong colonial or imperial background, why should this necessarily be a bad thing?

设置格式[dolly]: 字体: (中文) Calibri, 图案: 清除(自动设置), 字体颜色: 自定义颜色(RGB(24,25,26))

References:

Kornai, A. (2013). Digital language death. *PLoS ONE* 8(10), 1-11.

<https://doi.org/10.1371/journal.pone.0077056>

Bird, S. (2020). Decolonising speech and language technology. *Proceedings of the 28th*

International Conference on Computational Linguistics (COLING 2020), 3504-3519.

McWhorter, J. (2009). The cosmopolitan tongue: The universality of English. *World Affairs*, 172, 2, 61-68.

Rosenberg, T. (2011, Dec. 9). Everyone speaks text message. *New York Times*.

<https://www.nytimes.com/2011/12/11/magazine/everyone-speaks-text-message.html>

Liu, L. H. (2015). Scripts in motion: writing as imperial technology, past and present. *PMLA*, 130(2), 375-383. <http://www.jstor.org/stable/44015719>

Lester, T. (1997, July). New-alphabet disease? *The Atlantic*.

<https://www.theatlantic.com/magazine/archive/1997/07/new-alphabet-disease/306207/>

删除[dolly]:

删除[dolly]:

¹ For a video recitation and translation of the poem, see:

<https://www.futilitycloset.com/2020/01/29/lion-eating-poet-in-the-stone-den/>.