Editor's Column

Languages in the Time of Corona

AROUND MIDNIGHT ON 30 DECEMBER 2019, NINE DAYS BEFORE the World Health Organization (WHO) issued warnings about the coronavirus outbreak that would be known as COVID-19, a Canadian start-up, BlueDot, noticed a cluster of unusual pneumonia cases in Wuhan and flagged it. Unlike the WHO, or the United States Centers for Disease Control and Prevention (CDC), BlueDot doesn't just accept official statistics. "We know that governments may not always be relied upon to provide information in a timely fashion," said Kamran Khan, its founder and CEO (qtd. in Nitler). Instead, a machine-learning algorithm using natural language processing combs through a vast database every day, looking for syndromes of over 150 diseases in sixty-five languages. Local news stories, medical bulletins, even livestock and animal disease data are all on the algorithm's reading lists. When an outbreak is identified, its location is linked to global air-travel data to pinpoint other cities at risk. It takes the algorithm just a few seconds to process all this information, providing updates every fifteen minutes, twenty-four hours a day. A group of epidemiologists and physicians regularly reviews and interprets the data. In the case of the pneumonia cluster in Wuhan, everything from outbreak detection to information dissemination occurred in under twelve hours. Alerts went out at 10:00 a.m. on 31 December to BlueDot's clients: health agencies, hospitals, and commercial airlines around the world.

Khan, a practicing infectious-disease physician and professor of medicine at the University of Toronto, founded BlueDot after his experience at St. Michael's Hospital during the 2003 SARS epidemic. SARS originated in Guangdong province in China and spread to Hong Kong and then Toronto, where it killed forty-four people, including several frontline health-care workers. "In 2003, I watched

https://doi.org/10.1632/pmla.2020.135.5.841 Published online by Cambridge University Press

the virus overwhelm the city and cripple the hospital," Khan recalled. "There was an enormous amount of mental and physical fatigue, and I thought, 'Let's not do this again'" (qtd. in Nitler).

The name BlueDot—inspired by Carl Sagan's observation that the earth looked like "a pale blue dot" in a picture taken by *Voyager 1*—reminds us just how vulnerable we are (Vendeville). We aren't immune from viruses, and we can't fight them singlehandedly without wearing ourselves out. BlueDot proposes to ease that burden through automation, harnessing the power of artificial intelligence (AI) to sound the alarm about pandemics before threats become apparent to human beings. "We don't use artificial intelligence to replace human intelligence," Khan says. "We basically use it to find the needles in the haystack and present them to our team" (qtd. in Stieg). "When you're dealing with an outbreak, time and timing is everything," he said emphatically (qtd. in "This Canadian Startup").

An algorithm able to sift through local news stories in Chinese would save time at a critical juncture. The sixty-five-language capability eventually developed by BlueDot is one of the key lessons learned from SARS. Languages matter. Life and death depend on them. Picking up keywords in Thai, Korean, and Japanese, as well as Chinese, the algorithm was able to predict that the coronavirus would spread from Wuhan to Bangkok to Seoul to Taipei and Tokyo. It predicted twelve of the first twenty cities to be hit. On 23 March 2020 Gavin Newsom, the governor of California, singled out BlueDot in his daily coronavirus update and announced that his state was partnering with the Canadian company to fine-tune its pandemic response ("When").

Yet, in spite of this endorsement, and Khan's subsequent interview on *60 Minutes*, BlueDot remains largely unknown to the general public in the United States and around the world. This muted presence says something about the rarefied operations of AI in general. It also says something about the linguistic inaccessibility of this particular algorithm: the topic modeling underwriting its data analytics, its emphasis on prediction, its primarily extractive relation to what it reads, and the kinds of knowledge it is able—or not—to generate under that protocol. The language competency given the algorithm is a text-mining competency, designed to turn a vast, unstructured database into a meaningfully structured one, yielding predictive patterns with direct social, economic, and health-care consequences.

Elimination of low-relevance data is key. Using machine learning, the algorithm is taught not just to spot certain keywords but to classify them according to their value in risk assessment. Does a given keyword relate directly to the emergence and spread of an infectious disease? or to a secondary issue, such as new therapeutic drugs in development? or to something entirely unrelated, like the heavy-metal band Anthrax (Khan)? Winnowing down the data by eliminating irrelevant terms, as well as retellings of the same story, the algorithm can extract epidemiological information with maximum speed: where and when the disease originated, how many people have contracted it, and how many have died. This information is in turn correlated with human-mobility data to predict the degrees of disruption likely to result as the epidemic spreads from one locale to another.

Massive databases, processing speed, and efficiency in prediction—these are the strengths of AI as an early-warning system. They also set the boundaries of what it's able to do. Since the goal here is eliminative, extractive, and predictive, what's needed is just a small numbers of keywords. In the case of the Wuhan outbreak, these were *pneumonia* and *unknown cause* (O'Brien). In fact, it's not clear how much more of the Chinese language the algorithm has learned, since only semantic tags indicative of infectious dis-

https://doi.org/10.1632/pmla.2020.135.5.841 Published online by Cambridge University Press

eases are relevant here. A deeper engagement with language might be counterproductive, diminishing the speed of zeroing in on one bit of data and extracting from it information structured enough for human beings to process. Mining text with ease in sixty-five languages, BlueDot's algorithm wouldn't be able to understand the nuances of a single news story. Close reading of that sort would require a different linguistic capability, a different set of goals, and greater human input. Less spectacular than large-scale data analysis, this in-depth focus is nonetheless able to do some things that text mining cannot.

The Web-based *COVID-19 Dashboard* points to just that alternative. Developed at Johns Hopkins University by the Center for Systems Science and Engineering (CSSE), this interactive dashboard takes the form of a global map, displaying in real time the number of confirmed cases and deaths in every affected country. It is a quantitative rather than linguistic exercise. Yet because its operations began when COVID-19 was still centered in China, it has been necessarily bilingual from the outset, suggesting that knowledge of more than one language is a prerequisite for more disciplines than we realize. In this instance, with the help of two languages, the CSSE dashboard has emerged as the go-to pandemic resource for millions of people around the world, enjoying a public presence matched by few other science platforms. News media rely on it, as do health agencies and policy makers. For the rest of us, just hearing the words "Johns Hopkins University" every day in the news is oddly reassuring. The site gets 1.2 billion hits a day (Raymond). But it started out as a DIY project, put together in under eight hours, expanded and eventually automated only through popular demand.

On 21 January 2020 Lauren Gardner, an associate professor of civil and systems engineering at Johns Hopkins, was having coffee with two Chinese students, Ensheng Dong and Hongru Du, her first-year PhD advisees. Gardner, who works with spatial modeling for infectious diseases, had been following the Wuhan outbreak since it was first reported. Dong and Du had been following the outbreak as well, worried about family and friends back home. When Gardner came up with the idea of a tracking map, the two jumped at the opportunity. The dashboard was ready that very evening.

Knowledge of Chinese was critical, and for the first ten days all the information was assembled by Dong and Du. Dong entered the data manually. As the disease spread around the world, this human-only approach was no longer feasible. "Initially, I was trying to update two to three times a day, at 12 p.m. and 12 a.m.," Dong said. "But people were so anxious to see the dashboard, so I had to update every three to four hours. But sometimes by the time I had collected it all, the data had updated in the original data source, so I would have to tear down everything I'd done and collect it again" (qtd. in Swenson). The dashboard switched to a semiautomated live data-stream format on 1 February 2020, with various cross-checks in place and an algorithm now doing the hourly updating, but with room left for manual input (Dong et al.).

The dashboard is a "volunteer-based public service effort," Gardner explains (qtd. in Kaiser). The data collected are freely available to all, initially through Google Sheets and now through a GitHub repository, which allows viewers to check all the sources ("COVID-19 Data Repository"). Because the site operates twenty-four hours a day, seven days a week, it is staffed by international volunteers from different time zones. And people from all over send e-mail messages to report new cases (Swenson). They also weigh in on how it is run, some making irate requests for correction, known as pull requests. Even though the site now has an automated anomaly detector, a "million eyeballs" trained on the data guarantee a level of error detection surpassing even AI (Kaiser).

https://doi.org/10.1632/pmla.2020.135.5.841 Published online by Cambridge University Press

The bilingualism of the *COVID-19 Dashboard*, though less impressive than BlueDot's sixty-five-language capability, is adequate for its goal: collaboration. Back in January, the site's primary source was the *DXY*, a Chinese-language platform run by members of the medical community in China (www.dxy.cn). These days, it works on a globalized but similarly granular scale, getting its information from a large number of state, county, and city agencies, extending from Maryland to the Northern Mariana Islands, and not shying away from contested territories such as Hong Kong, Kosovo, Palestine (the West Bank and Gaza), Scotland, and Taiwan. It has also partnered with nongovernmental entities like the *Berliner Morgenpost* (Germany), covid19br (Brazil), Dati COVID-19 Italia, OpenCOVID-19 France, and rtve (Spain). Avoiding aggregated national statistics whenever possible, the dashboard offers data on the province level in China, the city level in the United States, and subregionally in many other countries (Dong et al.). When there's a discrepancy between the national and local statistics, the latter is more accurate and up-to-date.

Speaking in Chinese on *YouTube* about the dashboard they helped put together, Dong and Du emphasized the importance of global cooperation in a pandemic ("Chinese Students"). Knowing Chinese is not always a plus, however, nor is cooperation always possible. In fact, Chinese is nothing less than a minefield amid fractious global politics, and never more so than when viruses are involved. The politicization of the coronavirus as the "Chinese virus" is only the most glaring example. For the CSSE dashboard, the biggest pushback has been from some Chinese speakers, furious with the site for "calling Taiwan Taiwan" (Kaiser). The nuances of that impossibly vexed designation would be lost on algorithms, of course, and on any human observer not following the linguistic acrobatics performed by the People's Republic of China in naming the island that it considers part of its territory. The *COVID-19 Dashboard* might see itself as no more than an infectious-disease platform; those political fine points will plague it every step of the way.

Political fine points, though, are what humanize this dashboard, what make it emotionally resonant for so many people. Two other Johns Hopkins entities now join forces with CSSE: the Center for Health Security and the Centers for Civic Impact. These three have teamed up to read the small print. To highlight the health disparities emerging so starkly during the pandemic, with infection rates disproportionately high among African Americans, Latinx Americans, and Native Americans, a map of the United States was added to the global tracker, offering a county-by-county breakdown of case loads, correlating race, income levels, Medicaid usage, and health-insurance rates with testing capacity, hospital capacity, and rates of infection (Cruickshank). Questions about civil rights, health inequity, and the ethics of pandemic rationing are now part of the coronavirus conversation.

Collaboration among a large and diverse team makes this outcome almost inevitable. The dashboard now has to speak not just a quantitative language documenting the present spread of the virus but also a historical language capturing the legacy of the systemic racism that makes infectious disease such a profoundly unequal playing field. The six thousand physicists and cosmologists who stopped work on 10 June 2020 in support of #Strike4BlackLives, #ShutDownSTEM, and #ShutDownAcademia were speaking this historical language, facing up to the inequities embedded in our political, cultural, and intellectual environments (Overbye). The *COVID-19 Dashboard* had already learned that language, building a new platform to give it expression.

The current objectives of the dashboard couldn't have been imagined in its early days. And the site will continue to evolve in the months and perhaps years ahead, not because

https://doi.org/10.1632/pmla.2020.135.5.841 Published online by Cambridge University Press

machine-learning algorithms are doing this on their own but because human needs are driving that evolution. There's no better example of the human-nonhuman cyborg envisioned by Donna Haraway (149–82). AI here provides the tools for sharable knowledge at variable scales. That scalability and the in-depth health data it offers in turn put AI on one particular evolutionary path, aligning it with the well-being of the most vulnerable and pointing to crowdsourced science as a mitigating force.[1]

As the dashboard develops these granular levels of analysis, it also becomes mostly monolingual. It could be that monolingualism is the default position even for those who can speak more than one language. But it could also be that the kinds of analysis closest to the lived experience of ordinary people need to be conducted in the language that poses the least impediment to that population. Monolingualism is not, in any case, always a shortcoming or drawback. It is one side of a double formation. For global predictions, knowledge of more than one language is crucial, and algorithms designed to that end would certainly help. But to recognize preexisting conditions in one particular locale and the ways epidemiology is tangled up with social injustice, a language local to the population that lives there, one that can never be fully automated, is equally crucial. A humanly meaningful pandemic response requires both.

Narrative Medicine International—reading and storytelling workshops linking healthcare workers around the world—illustrates just that dynamic. Created by Rita Charon and others at Columbia University, these workshops are the latest additions to a twenty-year effort to bring the humanities to bear on the practice of medicine. Within ten days of the COVID-19 outbreak, international Zoom sessions were being held so that participants could share reading experiences and tell stories across national borders. These literary networks are especially important for clinicians facing social isolation and existential threat, connecting them with colleagues worldwide.

Even for these polyglot groups speaking English with no difficulty, it's important to switch off that global language from time to time and talk about books and share healthcare ordeals in the local tongue they are most comfortable with. The workshops soon started proliferating in many languages, including Greek, Italian, Polish, and Spanish (Charon). Close reading and close telling make all the difference in a pandemic, and for most of us these can be done only monolingually—a low bar, to be sure, and robust and democratic for that reason. But monolingualism cannot and should not exist in one language alone. There are, after all, as many different forms of it as there are languages. Anything less would be an unwarranted loss of data, an inaccurate tally of the world.

Wai Chee Dimock

Note

1. This evolutionary path is by no means a given. Facial-recognition technology, for instance—widely used by law enforcement and with an error rate five to ten times higher for African Americans than for whites (Magid)—seemed the dominant path before it began to receive critical scrutiny amid the nationwide protests against police brutality.

Works Cited

Charon, Rita. E-mail to the author. 18 May 2020.

"Chinese Students behind Globally-Shared COVID-19 Data Map." *YouTube*, 10 Apr. 2020, www.youtube.com/watch?v=fJQFYG4wrto.

"COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University." *GitHub*, github.com/CSSEGISandData/COVID-19.

Cruickshank, Saralyn. "Johns Hopkins Launches New U.S.-Focused COVID-19 Tracking Map." *Johns Hopkins University*, 14 Apr. 2020, hub.jhu.edu/2020/04/14/johns-hopkins-launches-us-coronavirus-tracking-map/.

https://doi.org/10.1632/pmla.2020.135.5.841 Published online by Cambridge University Press

Dong, Ensheng, et al. "An Interactive Web-Based Dashboard to Track COVID-19 in Real Time." *The Lancet*, 19 Feb. 2020, www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30120-1/fulltext.

Haraway, Donna J. *Simians, Cyborgs, and Women: The Reinvention of Nature*. Routledge, 1991.

Kaiser, Jocelyn. "'Every Day Is a New Surprise': Inside the Effort to Produce the World's Most Popular Coronavirus Tracker." *Science*, 6 Apr. 2020, www.sciencemag.org/news/2020/04/every-day-new-surprise-inside-effort-produce-world-s-most-popular-coronavirus-tracker.

Khan, Kamran. E-mail to the author. 14 June 2020.

Magid, Larry. "IBM, Microsoft, and Amazon Not Letting Police Use Their Facial Recognition Technology." *Forbes*, 12 June 2020, www.forbes.com/sites/larrymagid/2020/06/12/ibm-microsoft-and-amazon-not-letting-police-use-their-facial-recognition-technology/#7b4660441887.

Nitler, Eric. "An AI Epidemiologist Sent the First Warnings of the Wuhan Virus." *Wired*, 25 Jan. 2020, www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/.

O'Brien, Cillian. "How a Canadian Company Spotted the Coronavirus Outbreak before Health Officials." *CTV News*, 7 Feb. 2020, www.ctvnews.ca/health/how-a-canadian-company-spotted-the-coronavirus-outbreak-before-health-officials-1.4802942.

Overbye, Dennis. "For a Day, Scientists Pause Science to Confront Racism." *The New York Times*, 10 June 2020, www.nytimes.com/2020/06/10/science/science-diversity-racism-protests.html.

Raymond, Adam K. "JHU's Coronavirus Site Gets 1.2 Billion Interactions a Day." *New York*, 10 Apr. 2020, nymag.com/intelligencer/2020/04/jhus-coronavirus-site-gets-1-2-billion-interactions-a-day.html.

Stieg, Cory. "How This Canadian Start-up Spotted Coronavirus before Anyone Else Knew about It." *CNBC Make It*, 6 Mar. 2020, www.cnbc.com/2020/03/03/bluedot-used-artificial-intelligence-to-predict-coronavirus-spread.html.

Swenson, Kyle. "Millions Track the Pandemic on Johns Hopkins's Dashboard: Those Who Built It Said Some Missed the Real Story." *The Washington Post*, 29 June 2020, www.washingtonpost.com/local/johns-hopkins-tracker/2020/06/29/daea7eea-a03f-11ea-9590-1858a893bd59_story.html.

"This Canadian Startup Used AI to Track Coronavirus and Raised Alarm Days before the Outbreak." *ETtech.com*, 19 Feb. 2020, tech.economictimes.indiatimes.com/news/technology/this-canadian-startup-used-ai-to-track-coronavirus-and-raised-alarm-days-before-the-outbreak/74206120.

Vendeville, Geoffrey. "U of T's Infectious Disease Expert's AI Firm Now Part of Canada's COVID-19 Arsenal." *U of T News*, 27 Mar. 2020, www.utoronto.ca/news/u-t-infectious-disease-expert-s-ai-firm-now-part-canada-s-covid-19-arsenal.

"When the Coronavirus Hit, California Turned to Artificial Intelligence to Help Map the Spread." *60 Minutes Overtime*, 23 Apr. 2020, www.cbsnews.com/news/coronavirus-mapping-artificial-intelligence-california-governor-gavin-newsom-2020-04-23/?trk=organization-update-content_share-video-embed_share-article_title.

https://doi.org/10.1632/pmla.2020.135.5.841 Published online by Cambridge University Press