LETTER

Links that speak only some languages

Ronen et al. (1) raise an intriguing question concerning how to define quantitatively "the global influence of languages." Unfortunately, the limitations and flaws in their datasets raise questions concerning the validity of their maps of global language networks.

UNESCO's Index Translationum (2) project is a wonderful database source for Ronen et al.'s map of the network of book translations. It covers more languages (1,100) than Ronen et al.'s other sources, but it relies on entries submitted by national libraries, has not been updated for many countries in 20–30 y, and is not carefully edited. The data on Arabic translations are incomplete because the data from Lebanon are from 1997. Similarly, the data on Chinese translation are inadequate because Singapore's data are from 1988. In some cases, languages are confused, e.g., a Quechua translation is grouped with Yoruba translations, and a South African poetry book is listed as including translations from Gikuyu.

The Wikipedia data are even more problematic. The authors' final dataset contained "382,884,184 edits in 238 languages by 2,562,860 editors." As of last month, there were editions of Wikipedia in 288 languages (3). Most of these have few articles. Is Winaray more important than Spanish because it has 100,000 more Wikipedia articles? Is Moldavian's connection to Hindi its most important link because some people edited Wikipedia entries in both languages?

Analysis of a billion tweets sounds impressive but seems less so when the database is cut in half because the chromium compact language detector (CLD) could not identify the language used in half of the tweets. The version of CLD that was used could only identify 73 languages. Now CLD2 (4) identifies 161 languages but the list used did not include 28 languages on the US government's list of priority languages. The latter list includes languages spoken by tens of millions of people (e.g., Hausa and Pashto, each spoken by more than 50 million people and Berber with possibly 30 million speakers), whereas the CLD list included languages spoken by relatively small groups of people (Limbu, 370,000; Icelandic, 330,000; Gaelic, perhaps 300,000).

The number of "famous people associated with each language" was estimated based on articles in Wikipedia language editions. The language Wikipedias with the most articles produced, in this case, the most famous people. However, problems with the dataset seem obvious when the results indicate that Switzerland has produced more than twice as many famous people as China and that more than a third of the famous people born between 1800 and 1950 were born in the United States or the United Kingdom. The authors also "associated each person with a language using the current language demographics for his or her country of birth." This becomes questionable with respect to highly multilingual countries. For example, for Nigeria, 30% English is the only figure given. Then they used Charles Murray's *Human Accomplishment* (5) as a "second measure of famous people," a work that many of us find inherently biased because of its basis in outdated encyclopedias.

Ann J. Biersteker[1]
African Studies Center, Michigan State University, East Lansing, MI 48824

1 Ronen S, et al. (2014) Links that speak: The global language network and its association with global fame. *Proc Natl Acad Sci USA* 111(52):E5616–E5622.
2 United Nations Education, Scientific and Cultural Organization (2009) UNESCO Index Translationum: World Bibliography of Translation. Available at www.unesco.org/xtrans/bsform.aspx. Accessed January 26, 2015.
3 Wikimedia Foundation (2015) List of Wikipedias. Available at meta.wikimedia.org/wiki/List_of_Wikipedias. Accessed March 2, 2015.
4 Compact Language Detector 2 (2013) https://code.google.com/p/cld2/wiki/CLD2FullVersion. Accessed January 27, 2015.
5 Murray C (2003) *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 B.C. to 1950* (Harper Collins, New York).

Author contributions: A.J.B. wrote the paper.
The author declares no conflict of interest.
[1]Email: bierstek@msu.edu.