AEBMedicine

Letter

Linguistic underrepresentation in medical publishing: perspectives around the COVID-19 pandemic

Jose Manuel Gonzalez-Rayas[1,*] & Marcela Mendoza-Sigala[1]

[1] Tecnologico de Monterrey, School of Medicine and Health Sciences Monterrey, Mexico
* Corresponding author: contact.jmgr@gmail.com

Keywords: COVID-19; languages; underrepresentation; medical publishing

> "…publishing scientific papers exclusively in English may limit the ability of non-NES—native English-speaking—scientists to communicate important results to local practitioners and decision makers…This creates the moral dilemma of deciding whether to publish in English and making the results accessible to a broader audience or transferring knowledge to local experts."
>
> Di Bitetti & Ferreras, 2017.[1]

COVID-19 and the "new normal" have come here to stay. What initially emerged as a local outbreak transitioned into a global pandemic, and has caused more than 6 million deaths.[2] The sudden onset of the SARS-CoV-2 crisis challenged health care systems in terms of both patient overflow and information paucity. As the new infectious agent was previously unknown, evidence was scarce and highly controversial. Nonetheless, after the novel coronavirus continued spreading, medical literature became available to the point that more than 72,000 articles were published in PubMed in 2020 (Figure 1). This number is around 80 times greater than the previous year.

After analyzing the set of languages in which COVID-19 research was written, it is easy to note that English arose as the dominant language, with a 96% representation (Table 1). Spanish, French, German, Chinese and Portuguese had a much lower participation of 1.120%, 0.585%, 0.578%, 0.365% and 0.361% respectively. These figures are in stark contrast to the number of total language users each language integrates. In the case of Mandarin Chinese, the disparity is extreme, since this language is used by more than 1.3 billion speakers (~21.1% of the world's population).[3]

The reason behind the dominance of English in science is a matter of debate. This phenomenon was shaped by both the need for a lingua franca and by a complex series of historical events. However, to understand the role of English in scientific communication, it is important to briefly comment on the hierarchical model of De Swaan.[4–6] On top of the scheme, English is termed the "hypercentral language". On a second sphere, "supercentral languages", such as French, Spanish, Russian, Mandarin Chinese, Japanese, Arabic, Hindi, German and Portuguese, extend to more than one country as they were the once official languages of colonial powers. "Central languages", mostly of national and regional use, are part of the third level, and have low international diffusion. Finally, "peripheral" or vernacular languages represent the fourth hierarchy, but make up 98% of the world's linguistic heritage. They represent the mother tongue of ethnic groups but lack the official recognition in their home countries.[4–6] Unfortunately, as Hamel, 2007 highlights: "vernacular languages almost never appear in the debates about languages in science, since their status and corpus are considered unfit to express scientific thought and research findings".[4] As a result, speakers of vernacular languages are frequently excluded from the scientific endeavor, since to do so, they are required to learn both a central or supercentral language (with official status in their country), followed by English. This is the case of indigenous scientists in Mexico, who must overcome the multiple socio-cultural and linguistic barriers they encounter.

Citation: Gonzalez-Rayas, JM & Mendoza-Sigala M. Linguistic underrepresentation in medical publishing: perspectives around the COVID-19 pandemic. *AEBMedicine*, 1(1), 1-5.

Received: 20/03/2023
Revised: 21/04/2023
Accepted: 24/04/2023
Published: 28/04/2023

Evidence level: opinion.

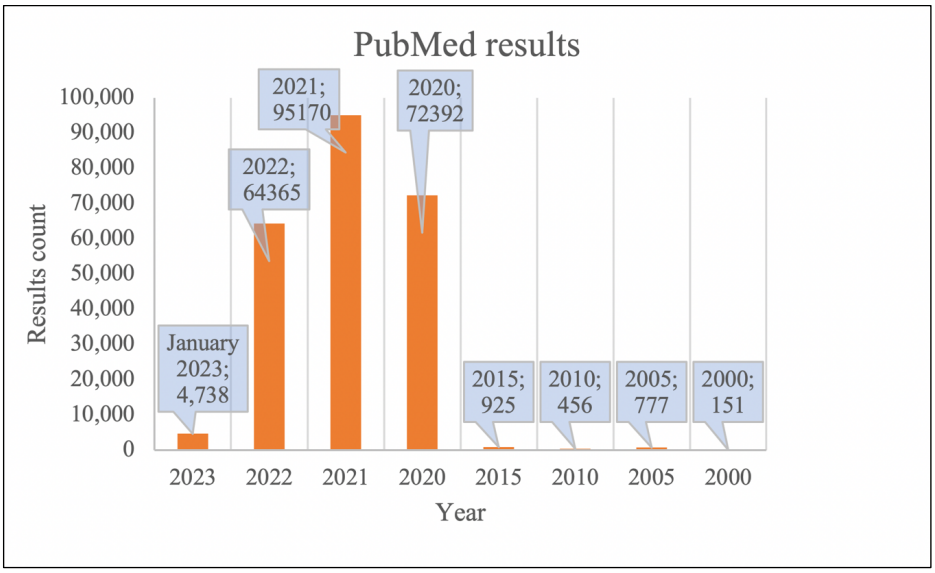

Figure 1. Number of articles regarding "coronavirus" published in PubMed according to the year.

Currently, some researchers sustain that through our history, and long before English became the hypercentral language, there has always existed a lingua franca for science in Occident. Some examples are Sumerian, Greek, Arabic and Latin. Nevertheless, amid the 15th century, plurilingualism arose with the more frequent use of French, English, German, Italian and Russian for the publication of scientific discoveries.[4] Actually, during the first part of the 20th century, three languages (English, French and German) played a "balanced" role in science, although in different fields.[4] English was mainly used for economy and geology, French for law and political sciences and German for medicine, biology and chemistry. However, modern multilingualism was interrupted by the two World Wars, which accelerated the economic and political rise of the United States. The latter finally led to the emergence of English as the dominant language for international and scientific communication. In fact, according to Hamel, 2007, English is the language of 75% of the publications in social sciences and humanities, while this figure increases to 90-98% in the case of the natural sciences.[4,7]

In the case of Mexico and Spanish speaking-countries, scientific interaction is frequently restricted to Spanish. Portuguese-speaking countries (Brazil and Portugal) show a similar behavior and have been integrated into the Hispano-American scientific sphere. The latter is due to the similarity between the two languages and to the rich scientific output of Brazilian researchers (who author more than 40% of the total articles produced in Latin America). Databases such as Latindex and SciELO (Scientific Electronic Library Online, created in 1997 in Brazil) have further encouraged the union between Luso-Hispanic authors.[4,8] As a consequence, the Spanish/Portuguese-speaking cluster has developed a certain degree of academic independence. This is reflected in the considerable number of scientific meetings and congresses within Latin America, which are almost universally held in Spanish or Portuguese.[4]

Compared to the local scientific communication undertaken in Latin America, the case of academic publishing in Spanish and Portuguese is less favorable. For instance, more than 90% of the research output generated by Colombian researchers is published in English.7 Furthermore, of the 5,986 scientific journals edited in Brazil, only 17 are part of the Science Citation Index or SCI (one of the most reputed databases for research).[4] A critical reason behind the low rate of inclusion of Latin American journals on international indexes (such as SCI, Scopus or PubMed) is the small number of citations they receive. This phenomenon was studied by Di Bitetti and Ferreras in 2017 and, after taking into account the journal, the year of publication and the paper's length, they found that articles written in English receive more citations than their counterparts in other languages.[1] This citation bias directly affects Latin American journals, since they are mostly edited in Spanish and/or Portuguese. As a consequence, few articles are published in languages other than English, as happened during the COVID-19 crises.

Table 1. PubMed results filtered by language

Language	PubMed results	Percentage of total results	Total language users in all countries	Percentage of total speakers
Afrikaans	0	0.000	17,537,980	0.279
Albanian	0	0.000	5,901,310	0.094
Arabic	1	0.000	335,176,770	5.335
Armenian	0	0.000	3,847,000	0.061
Azerbaijani	0	0.000	23,140,520	0.368
Bosnian	0	0.000	2,690,810	0.043
Bulgarian	0	0.000	8,143,260	0.130
Catalan	0	0.000	9,177,310	0.146
Mandarin Chinese	1,242	0.365	1,323,796,770	21.071
Croatian	0	0.000	6,716,350	0.107
Czech	10	0.003	13,389,650	0.213
Danish	76	0.022	5,616,830	0.089
Dutch	219	0.064	24,393,350	0.388
English	328,156	96.390	1,268,100,190	20.185
Esperanto	6	0.002	2,000,000	0.032
Estonian	0	0.000	1,249,990	0.020
Finnish	0	0.000	5,825,800	0.093
French	1,990	0.585	276,570,840	4.402
Georgian	0	0.000	3,879,020	0.062
German	1,968	0.578	131,631,870	2.095
Greek, Modern	11	0.003	13,192,550	0.210
Hebrew	58	0.017	9,328,950	0.148
Hindi	0	0.000	637,271,710	10.144
Hungarian	144	0.042	12,532,990	0.199
Icelandic	19	0.006	322,620	0.005
Indonesian	0	0.000	198,984,560	3.167
Italian	357	0.105	67,687,060	1.077
Japanese	223	0.066	126,358,970	2.011
Kinyarwanda	0	0.000	13,125,250	0.209
Korean	22	0.006	79,368,420	1.263
Latin	0	0.000	-	-
Latvian	0	0.000	1,718,980	0.027
Lithuanian	0	0.000	2,994,890	0.048
Macedonian	0	0.000	1,656,090	0.026
Malay	0	0.000	81,578,326	1.299
Malayalam	0	0.000	37,919,870	0.604
Maori	0	0.000	159,700	0.003

Multiple Languages	0	0.000	-	-
Norwegian	180	0.053	5,307,610	0.084
Persian	0	0.000	65,113,060	1.036
Polish	65	0.019	40,631,460	0.647
Portuguese	1,228	0.361	252,152,760	4.014
Pushto	0	0.000	53,131,900	0.846
Romanian	0	0.000	24,536,480	0.391
Russian	541	0.159	257,962,060	4.106
Sanskritt	0	0.000	5,007,800	0.080
Scottish gaelic	0	0.000	60,130	0.001
Serbian	0	0.000	8,831,666	0.141
Slovak	1	0.000	7,227,280	0.115
Slovenian	0	0.000	2,228,650	0.035
Spanish	3,813	1,120	537,905,240	8.562
Swedish	48	0.014	12,804,400	0.204
Thai	0	0.000	60,683,370	0.966
Turkish	66	0.019	85,197,130	1.356
Ukrainian	3	0.001	33,144,480	0.528
Undetermined	0	0.000	-	-
Vietnamese	0	0.000	76,972,360	1.225
Welsh	0	0.000	573,050	0.009
Total	**340447**	**100**	**6,282,457,442**	**100.000**

References

1. Di Bitetti MS, Ferreras JA. Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications. Ambio 2017;46(1):121–7. Doi: 10.1007/s13280-016-0820-7.
2. Johns Hopkins University. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). COVID-19 Map - Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html. 28/04/2023.
3. Ethnologue : Languages of the World. Dallas, Texas :SIL International, 2023.
4. Hamel RE. The dominance of English in the international scientific periodical literature and the future of language use in science. AILA 2007;20:53–71. Doi: 10.1075/aila.20.06ham.
5. De Swaan, A. 1993. The emergent world language system: An introduction. International Politi-
6. cal Science Review 14(3): 219–26. De Swaan, A. 2001. Words of the World. The Global Language System. Cambridge: Polity Press.
7. Ramírez-Castañeda V. Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science: The case of Colombian researchers in biological sciences. PLoS ONE 2020;15(9):1–15. Doi: 10.1371/journal.pone.0238372.
8. Bojo Canales C, Fraga Medín C, Hernández Villegas S, Primo Peña E. SciELO: un proyecto cooperativo para la difusión de la ciencia. Rev esp sanid penit 2009;11(2). Doi: 10.4321/S1575-06202009000200004.