Journal of Informetrics 12 (2018) 605–617

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

Is there Lingua Franca in informal scientific communication? Evidence from language distribution of scientific tweets

Houqiang Yu [a,*], Shenmeng Xu [b], Tingting Xiao [c]

[a] School of Economics & Management, Nanjing University of Science & Technology, Nanjing, China
[b] School of Information and Library Science, UNC Chapel Hill, USA
[c] Jiangsu Institute of Quality and Standardization, Nanjing, China

ARTICLE INFO

Article history:
Received 18 December 2017
Received in revised form 27 February 2018
Accepted 7 June 2018

Keywords:
Altmetrics
Scientific tweets
Language distribution
Informal scientific communication
Twitter

ABSTRACT

Language distribution in scientific communication reflects the influence of different languages on science in global perspective. The study, based on over 450 thousand scientific tweets of all publications indexed by Scopus in June 2015, reveals the language distribution in informal scientific communication. Moreover, this result is compared with the language distribution in formal scientific communication reflected in scientific publications. Results show: (1) The language of scientific tweets is concentrated in English (91%), Japanese (2.4%) and Spanish (1.7%), while the language of scientific publications is concentrated in English (90.6%), Chinese (5%) and German (1.1%). (2) Both scientific tweets and scientific publications present disciplinary differences in language distribution, reflecting the different amount of attention that authors of different languages have on certain disciplines. (3) Except Saudi Arabia, investigated countries all over the world, regardless of whether their native language is English or not, all have English scientific tweets in the dominant position. For the vast majority of these countries, the native language scientific tweets only rank the second position. (4) Overall, 26% of tweeters use more than one language to tweet scientific products, while 49% of scientific tweeters tweet everything in English only. The results indicate that English has undoubtedly become the lingua franca in informal scientific communication.

1. Introduction

Language is a symbol of both culture and life. As the media of scientific publications (Williams, 1988), language plays a fundamental role in scientific communication. Across different historical stages, different languages were widely and frequently used for global scientific communication and became so-called *Lingua Franca* of scientific communication. Lingua franca is a commonly used term in language studies referring to a common language used by speakers of different languages. The scientific centers in the world had shifted several times, so did the lingua franca of global scientific communication: from Latin to German, then to English (Atlantic, 2017). Japanese was once an influential intermediate language when important research from all over the world were translated into Japanese and then exported to China or other countries (Shi & Geng, 2012).

* Corresponding author.
E-mail address: yuhouq@yeah.net (H. Yu).

https://doi.org/10.1016/j.joi.2018.06.003
1751-1577/

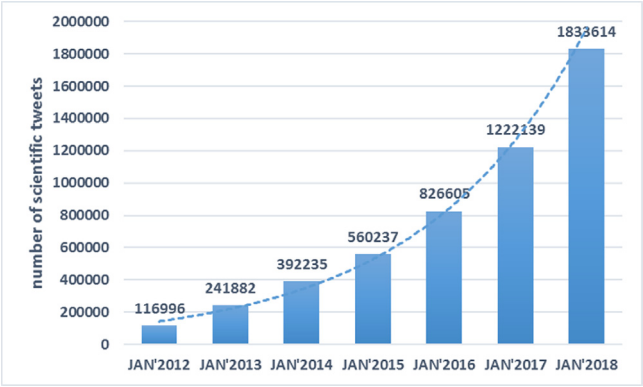

Fig. 1. Number of scientific tweets recorded by Altmetric.com company.

Nowadays, the lingua franca of formal scientific communication is with no doubt English. Language studies of scientific publications show convincing evidence. Scholarly journal is one of the main output channels of formal scientific communication. With the growth of global academic collaboration, scholars are motivated to publish in international journals of which the prevailingly adopted language is English (Flowerdew, 2007). Over 85% of journals indexed in Web of Science are in English. In 2015, 92% of Scopus indexed scientific publications are in English. Gatekeeper roles like editorial board member and reviewer are usually possessed by native English speakers (Canagarajah, 2002). Rejection rate of articles without English references is much higher (Canagarajah, 1996). English is simultaneously the dominant language of international organizations and top conferences (Brambrink, Ehrler, & Dick, 2000). Scholars around the world prefer to publish high quality articles in English journals in order to improve their visibility and academic impact. Their publishing preference is severely shaped by the scientific evaluation system. For example, publications in prestigious journals like *Nature* or *Science* are highly rewarded in some countries (Quan, Chen, & Shu, 2017). Non-anglophone countries have established English journals and endeavored to get them indexed by top bibliographic database like Web of Science, so as to improve their international competitiveness.

There are a number of reasons for English to be dominant (Ammon, 2001), including but not limited to historical reason, scientific competitiveness and monopolistic publishers. The pros and cons of English as the dominant scientific language are well studied. For example, the dominance promotes scientific communication between scholars from different countries when they are unable to understand each other without an intermediate language (Butler, 2000). However, the dominance causes biases to non-anglophone scholars and devalue other languages (Dalen & Henkens, 2001). The various biases caused by the dominance of English are well known as the EILS (English as international language of science) problem (Tardy, 2004).

Compared with formal scientific communication, scholars are less influenced by official factors in informal scientific communication. For scholars who speak two or more languages, they are free to choose a language to use in different occasions. For example, on Twitter they can tweet in their mother tongue, or they can tweet in English. Informal scientific communication on Twitter is becoming increasingly popular, reflected in the fast-growing number of scientific tweets (see Fig. 1, the average annual growth rate is 160%). A *scientific tweet* is a tweet that contains links, oftentimes with DOIs, or other trackable traces to a scientific product. There are plenty of studies of Twitter usage in scholarly communication, as shown in detail in Section 2. However, this study focuses on the language of scholarly communication on Twitter from a macro perspective. The driving question of this study is, in informal circumstances, will scholars use English when they can choose not to? Is there also a dominant language in informal scientific communication?

From previous studies on tweeting language, we know English is also the dominant tweeting language of which the percentage is 6 times that of Spanish, the second most popular tweeting language. However, tweeting language is far more diverse than scientific publishing language. In addition to popular languages like English, languages such as Indonesian, Malay, Japanese, Korean and Portuguese all have good visibility on Twitter (Mocanu, Baronchelli, & Perra, 2012). Considering that most users tweet in their native language, the language distribution of general tweets to some extent reflects the distribution of population using different languages. By contrasting the language distribution of scientific tweets with that of general tweets, we can determine whether language distribution of scientific tweets shows a different pattern. Here, we use *general tweets* to refer to tweets other than scientific tweets.

Based on different tweeting locations and tweeting languages, four scenarios are presented in Fig. 2. If users tweet both scientific products and general content in their native language, we define this as scenario A. Scenario B is of particular interest because it demonstrates the influence of certain foreign languages in domestic informal scientific communication in some non-anglophone countries. Scenario C and scenario D are difficult to capture because the geographic data provided by Twitter usually reflects the registration location which is usually the home country rather than the real-time location of Twitter users. The geo-coordinates of tweets that reflect real-time locations are only available when users explicitly choose to enable the “Share precise location” feature. In other words, if a Twitter user travels and posts tweets abroad, these tweets