

基于科学推文视角的非正式科学交流语言分布研究^{*}

余厚强 董 克 王曰芬 章成志

摘 要 语言是科学交流的载体,科学交流的语言分布反映了全球科技话语权的格局,本文基于科学推文视角揭示了非正式科学交流中的语言分布,并与基于科技文献的正式科学交流语言分布做了对比分析。研究发现:①科学推文语言集中分布在英语(91%)、日语(2.4%)和西班牙语(1.7%),而科技文献语言集中分布在英语(94.2%)、中文(4.3%)和土耳其语(0.4%);②科学推文和科技文献的语言分布均存在学科差异,体现出不同语言作者对特定学科关注度的不同;③除了沙特阿拉伯之外,世界各国即使母语不是英语,均以英语科学推文占主导地位,母语科学推文屈居第二;④英语、德语、日语、法语等语种的科技文献获得科学推文最多,而科技文献量和质量名列前茅的中文和土耳其语文献获得的科学推文很少。结果表明,英语俨然已经成为非正式科学交流中的通用语,提高中文在科学交流中的国际影响力任重道远。图3。表3。参考文献15。

关键词 替代计量学 科学推文 语言分布 非正式科学交流 推特

分类号 G250.252

Study of Language Distribution in Informal Scientific Communication from the Perspective of Scientific Tweets

YU Houqiang, DONG Ke, WANG Yuefen & ZHANG Chengzhi

ABSTRACT

Language is a media of scientific communication. Language distribution of scientific communication reflects the status of global scientific power. The study, based on scientific tweets, has revealed the language distribution in informal scientific communication, and comparative analysis is done with language distribution of scientific literature, in order to help understand the function and influence of major languages in informal scientific communication in major countries around the world.

Firstly, bibliographic data were collected from Scopus for scientific publications of all languages published in June 2015, including country, discipline, citation count, title and DOI. Secondly, Python program was used to match records in Altmetric.com dataset (October 2011 to June 2016) by matching DOI and title. Scientific tweets of these publications were collected, including geographic coordinates and full text of tweet content. Thirdly, descriptive statistics were used for analyzing the data and comparison was conducted.

^{*} 本文系国家社会科学基金重大项目“面向知识创新服务的数据科学理论与方法研究”(编号:16ZDA224)的研究成果之一。(This paper is an outcome of the key project “Study of Data Science Theory and Methods Oriented to Knowledge Innovative Service” (No. 16ZDA224) supported by National Social Science Foundation of China.)

通信作者:余厚强,Email:yuhouq@yeah.net, ORCID: 0000-0002-9241-6630 (Correspondence should be addressed to YU Houqiang, Email:yuhouq@yeah.net, ORCID:0000-0002-9241-6630)

According to Scopus, 183 699 scientific publications of 25 languages were published in June 2015. These publications obtained 451 982 scientific tweets. 1) From the perspective of general distribution of scientific tweets, English scientific tweets have the highest percentage. Languages of scientific tweets are concentrated in English (91%), Japanese (2.4%) and Spanish (1.7%). In comparison, languages of scientific publications are concentrated in English (94.2%), Chinese (4.3%) and Turkish (0.4%), and the concentration is getting more skewed than before. 2) From the perspective of language distribution in disciplinary level, there exists disciplinary difference in informal scientific communication, reflecting the different level of attention they have for different disciplines. In comparison, language distribution of scientific publications in disciplinary level reflects that certain countries have advantage in certain disciplines. 3) From the perspective of language distribution in country level, English has undoubtedly become lingua franca for informal scientific communication. Countries around the world, regardless of whether the native language is English or not, have English scientific tweets in the dominant position while their native language scientific tweets rank the second, except for Saudi Arabia. For all countries, the top three languages have occupied over 95% of scientific tweets. The distribution also shows some scientific and cultural difference, for example, in Japan, Japanese scientific tweets have similar percentage as English scientific tweets. 4) From the perspective of language distribution for publications of different languages, scientific tweets are concentrated for publications of English, Germany, Japanese, French, Portuguese and Spanish, while Chinese and Turkish publications of high percentage receive no scientific tweets, indicating publications of these two languages obtain low attention in informal scientific communication.

Although scientific tweet is an important form of informal scientific communication on Twitter, it cannot represent informal scientific communication from all scientists, and there are also other forms of scientific communication on Twitter. China has restricted the use of Twitter so China is not in the scope of analysis of language distribution in country level.

The study quantitatively measures the language distribution of informal scientific communication. With detailed statistical analysis from country level, disciplinary level and language level, the author points out that English has also become the lingua franca in informal scientific communication. 3 figs. 3 tabs. 15 refs.

KEY WORDS

Altmetrics. Scientific Tweets. Language distribution. Informal scientific communication. Twitter.

0 引言

语言既是文化和生活的象征符号,也是科技论文的载体^[1],在科学交流中扮演了非常关键的角色。在不同的历史时期,某一种或若干种语言成为全球科学交流中频繁和普遍使用的媒介语言,称之为科学的通用语(Lingua Franca)。科学中心在世界范围内几度发生迁移,全球科学的通用语同样经历着变迁,从拉丁

文到德语,再到英语^[2],日语也一度成为重要的中介语言,世界各国重要的研究成果被翻译成日语,再交流到中国或传递到其他国家^[3]。毋庸置疑,科技语言的地位客观上反映了其母国的综合国力,对了解世界科技格局有重要启示。

当今时代,正式科学交流的通用语无疑是英语。对科技论文语言的研究尤其具有说服力,期刊是正式科学交流的标准输出渠道,随着全球学术交流和国际合作的兴起,学者被迫在国际期刊上发表文章,而通常所用语言正是英

语^[4]。超过 85% 的 WoS (Web of Science) 索引期刊是英语期刊, 2015 年 Scopus 索引的科技文献中 92% 是用英语撰写。守门人的角色诸如编委会和审稿人, 通常由英语为母语的人担任^[5], 没有英文参考文献的稿件被拒稿的概率大大增加^[6]。世界各国学者为了提高学术可见度和学术影响力, 更是为了迎合本国的科研考评制度, 竞相在英文期刊上发表高质量论文; 非英语母语的学者为了提高本国期刊的竞争力, 也竞相开办英文版期刊, 以此争取早日被 Web of Science 等顶级数据库收录。英语同时也是国际组织和顶级会议的主导性语言^[7]。英语的垄断地位形成于多种原因^[8], 包括历史环境、信息获取和经济发展等。英语作为垄断语言的优缺点已经有过较成熟的研究, 举例来说, 垄断地位的英语可能作为交流科学的通用语言, 促进不同母语科学家的学术交流^[9], 但是对非盎格鲁语系的学者会造成偏见, 贬低其他语言的地位^[10],

不同维度的证据指出了英语作为国际语言带来的问题, 亦称为 EILS (English as an International Language of Science) 问题。

与正式科学交流相比, 非正式科学交流中学者拥有更多的自主性。对于掌握两种或多种语言的学者而言, 他们有独立的裁量权决定在何种场合使用何种语言。迄今为止, 我们还缺乏有效的手段去衡量非正式科学交流中的语言分布情况。推特上的科学交流是典型的非正式学术交流, 在推特上作者可以用母语发推, 也可以用外语发推。Rowlands 等人^[11]于 2011 年的测算结果表明, 约 10% 的研究人员将推特用于专业研究用途, Adie 等人^[12]则发现围绕学术成果发生的社交媒体活动每个月以 5% 到 10% 的速度增长。科学推文是指包含学术成果痕迹的推文, 这种痕迹可以是 DOI (Digital Object Identifier) 或链接, 也可以是其他任意可被追踪的形式。科学推文的示例如图 1 所示。

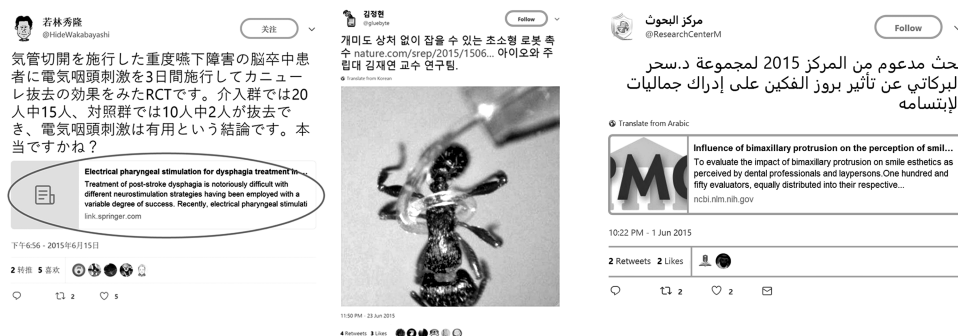


图 1 科学推文示例图

科学推文是学术成果在推特上被传播和讨论的一种表现形式, 属于典型的非正式科学交流, 为我们分析非正式交流情况提供了契机。科学推文形式的科学交流非常普遍, 据笔者统计, Altmetric.com 数据集 (2011 年 10 月到 2016 年 6 月) 中 390 余万个学术成果具有科学推文, 涉及不同的用户 263 余万名, 科学推文总量达 2 069 万条。与此同时, 推特被世界范围内 75 个

国家^①广泛使用, 并且是这些国家的主流微博服务, 为作者提供多语言版本, 作者可自主选择, 所以我们可以透过这个窗口去窥视非正式科学交流中语言分布的现状。

通常情况下, 作者会选择用本国母语来发推文, 但是也存在其他情形, 所以可以分成四种情形: ①作者以母语在母国发推; ②作者以非母

① 见 <https://transparency.twitter.com/information-requests/2015/jul-dec>

语在母国发推;③作者以母语在非母国发推;④作者以非母语在非母国发推。在上述四种情形中,第一种情形是最正常不过的情形,第三种和第四种情形都可能仅出于朋友圈受众考虑而为,第二情形具有特别的意义,因为这种情形最能说明语言在非正式科学交流中的地位,那么问题就变得很有趣,非英语母语的学者是否会选择在推特上的非正式科学交流中使用外语呢?推特这种典型的非正式学术交流环境中,是否也存在“统治性”的语言。

本文通过研究科学推文的语言分布,来检验非正式学术交流中是否存在通用语现象。科学推文与科学出版物的一大区别是,科学推文作者不限于学者,其推文语言偏好反映了他们的交流模式。如果作者选择用非母语的語言来发推,例如德国人用英语发推,他们可能是试图获得面向全球范围更广的受众群。使用不止一种语言发推的作者通常是活跃的知识传递者。传播的及时性是考虑选择什么语言发推的另一个重要因素。我们集中比较英语母语国家和非英语母语国家的情形,便于与正式科学交流中英语的垄断情形做对比。所以,本文主要回答以下研究问题。

(1)科学推文总体的语言分布是怎样的。具体来说,目前是否存在主导性的语言,在全球范围内举足轻重的各大语言表现如何,这种分布与科技文献的相比有什么特点。

(2)科学推文分学科的语言分布是怎样的。科学推文的语言分布是否存在学科差异,与科技文献相比具有什么特点。

(3)不同国家的科学推文语言分布是怎样的。不同国家有着自己固有的文化背景和施行的科技政策,反映在科学推文的语言分布上,会体现出怎样的分布特点。

(4)不同语种科技文献的科学推文语言分布是怎样的。不同语种科技文献的科学推文语言分布,反映了正式交流与非正式交流中各语言作为科技语言的强弱关系。

1 数据获取与处理

为了回答上述研究问题,笔者首先从 Scopus 数据库采集不同语种科技文献的元数据,然后从 Altmetric.com 数据集中采集这些科技文献的科学推文数据。Altmetric.com 的数据集更新到 2016 年 6 月,根据之前的研究结论^[13],绝大多数科学推文发生在科技论文正式发表后 360 天内,所以选取 Scopus 中 2015 年 6 月出版的科技文献作为研究对象,以收集到相对及时并且完整的科学推文数据。

经检索,Scopus 数据库中 2015 年 6 月发表各类科技文献一共 183 699 篇,但是 Scopus 数据库对可下载的题录数据数量做了限制,符合检索条件的检索结果中,完整的题录和引文数据仅允许下载两千条(其中含有语种等本研究所需字段),单纯的引文数据仅允许下载两万条(其中不包含语种等字段),所以笔者制定以下分解检索下载策略:①依据 Scopus 中的时间分学科进行检索,示例检索式是“PUBDATETXT(JUNE 2015) AND SUBJAREA(AGRI)”;②若某学科科技文献数量仍然大于两千条,继续按语言划分,示例检索式是“PUBDATETXT(JUNE 2015) AND SUBJAREA(AGRI) AND LANGUAGE(ENGLISH)”;③若按学科和语言划分后,科技文献数量仍然超出两千条,则继续按国家划分,检索式是“BDATETXT(JUNE 2015) AND LANGUAGE(ENGLISH) AND AFFILCOUNTRY(“UNITED STATES”) AND SUBJAREA(AGRI)”。数据下载时间是 2016 年 12 月 18 日。

获得 Scopus 数据集后,根据每条记录的 DOI 和标题利用 Python 程序从 Altmetric.com 数据集中检索这些科技文献获得的科学推文,共收集到 451 982 条科学推文及其元数据,具有科学推文的科技文献比例为 5%。

本研究所需核心数据项,例如科技文献语种、科学推文作者所在国、科技文献所属学科,

均已获取,尚未获取的是科学推文的语种。笔者经过比较,采用Python包langdetect 1.0.7来识别科学推文语言,该程序包后台采用谷歌公司的语言识别API,识别准确率为99.1%^①。实际测试表明,由于科学推文中存在链接,正文内容有时偏少,导致识别准确率发生些许下降,但不影响整体结果。科技文献中,存在少量双语(通常为英语和母语)科技文献,均计作母语文献。

2 研究结果

2.1 科学推文与科技文献总体语言分布比较分析

从表1看到,2015年6月Scopus收录的论

文中,以英语为主,约占94.2%。中文论文居第二,但是仅占到约4.3%。土耳其语论文居第三,仅占约0.4%。这充分说明了英语在正式科学交流中的垄断地位。值得注意的是,中文和土耳其语文献的比重,已经超越了西班牙语和德语。从篇均被引量角度来看,中文论文的质量也高于其他语种的文献,说明我国的科技论文不论是数量还是质量均取得了显著进步。但是,还没有哪一种语言能挑战英语的地位。英语文献的总量和篇均被引量分别是中文文献的24倍和7倍。

从表1还可以看到,这些科技文献的科学推文中,英语推文占91.3%,可见英语已成为科学推文的主体语言。与科技文献语言分布不同的是,日语科学推文量居第二位,约占2.4%。西

表1 科学推文与科技文献总体语言分布比较

排名	科学推文语言	推文比例(%)	科技文献语言	论文比例(%)	篇均被引(次)
1	英语	91.30	英语	94.16	3.11
2	日语	2.35	中文	4.25	0.47
3	西班牙语	1.66	土耳其语	0.39	0.17
4	法语	1.41	西班牙语	0.22	0.29
5	阿拉伯语	0.61	德语	0.20	0.4
6	罗马尼亚语	0.44	法语	0.20	0.26
7	荷兰语	0.34	葡萄牙语	0.14	0.2
8	加泰罗尼亚语	0.32	日语	0.13	0.19
9	意大利语	0.28	韩语	0.11	0.4
10	葡萄牙语	0.23	意大利语	0.05	0.14
11	德语	0.19	克罗地亚语	0.04	0.04
12	印尼语	0.11	俄语	0.03	0.47
13	挪威语	0.11	荷兰语	0.02	0.12
14	瑞典语	0.08	匈牙利语	0.02	0.18
15	俄语	0.07	捷克语	0.01	0.51
16	丹麦语	0.06	波斯尼亚语	0.01	0.14
17	韩语	0.06	加泰罗尼亚语	0.01	0
18	土耳其语	0.06	斯洛文尼亚语	0.01	0.41
19	南非荷兰语	0.05	波斯语	0.01	0.07
20	波兰语	0.04	南非荷兰语	0.01	0.7
21	爱沙尼亚语	0.04	波兰语	0.00	0.13
22	芬兰语	0.03	希伯来语	0.00	0.14
23	索马里语	0.02	斯洛伐克语	0.00	1
24	威尔士语	0.02	乌克兰语	0.00	0
25	斯洛文尼亚语	0.01	阿拉伯语	0.00	0

① 见 <https://shuyo.wordpress.com/2012/02/21/language-detection-for-twitter-with-99-1-accuracy>

班牙语推文量也较高,约占 1.7%。此外,法语和阿拉伯语在科学推文中也有较高的显示度。相较之下,阿拉伯语在科学推文中的可见度要显著高于在科技文献中的表现。

2.2 科学推文与科技文献分学科语言分布比较分析

由于英语在各个学科都占到最大比例(90%左右),为了更清晰地展示其他语种的分布,将英语刨去,绘制分布图,如图 2 和图 3 所示,科学推文和科技文献的语言分布都存在显著

的学科差异。图 2 中看到,中文论文在大部分学科都占了较大比例,例如农学与生物科学(AR-
GI)、化学工程(CHNG)、化学(CHEM)、数学(MATH)、医学(MEDI)等,但是也有例外,例如在兽医学(VETE),没有中文文献,德语文献最多。在保健科学(HEAL),也是德语文献最多。在牙医学(DENT),则是法语文献最多。土耳其语文献集中在综合性科学(COMP)、工程学(ENGI)和神经科学(NEUR)中。科技文献语言分布的学科差异体现了各语种在相应学科研究中实力的强弱关系。

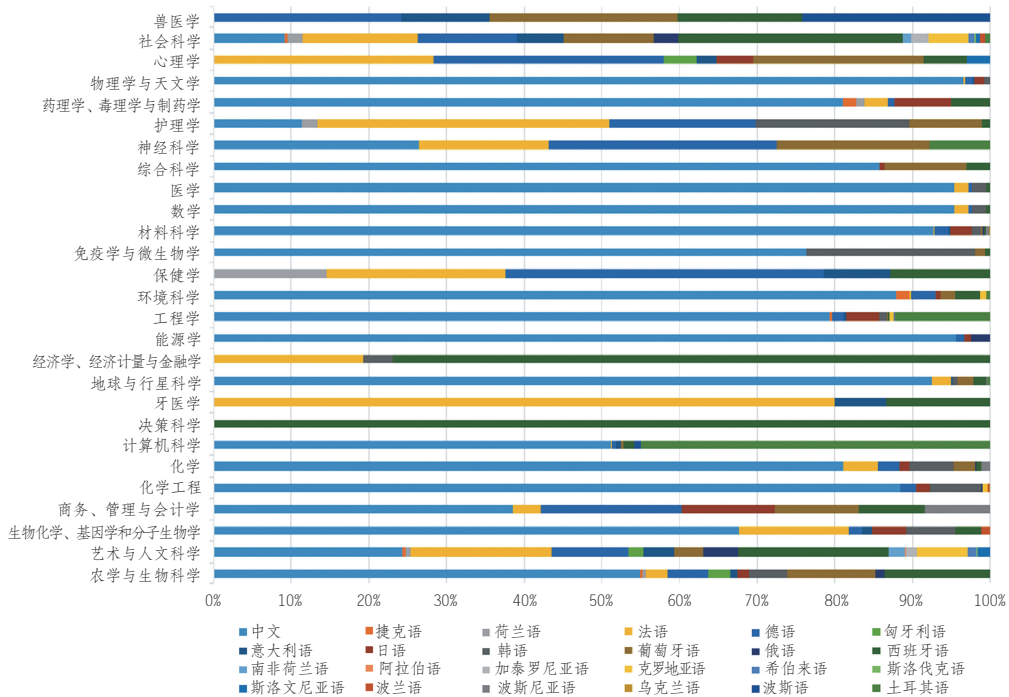


图 2 各学科科技文献语言分布图(英语除外)

图 3 展示了各个学科的科学推文语言分布,部分语言在特定学科产生了较多的科学推文,例如地球与行星科学(EART)的罗马尼亚语科学推文较多,经济学(ECON)、艺术学(ARTS)、社会学(SOCI)等法语科学推文最多,物理学(PHSY)、农学与生物科学(AR-
GI)、免疫学(IMMU)等日语科学推文最多,牙医学(DENT)、保健科学(HEAL)和护理学(NURS)

西班牙语科学推文最多。科学推文语言分布的学科差异反映了各语种对相应学科关注度的高低程度。

2.3 世界范围内不同国家科学推文语言分布

表 2 展示了世界各大区域主要国家的科学推文语言分布。从表 2 可以看到,除了沙特阿拉伯以外,各国均以英语科学推文比例最高(占 50%

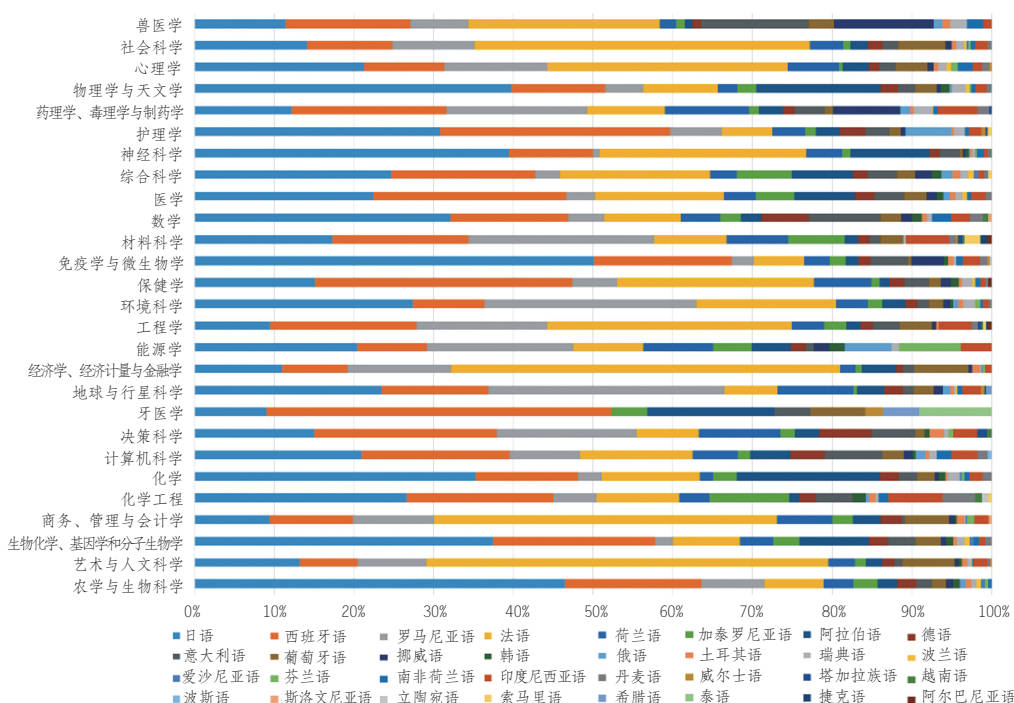


图3 各学科科学推文语言分布图(英语除外)

以上),母语科学推文屈居第二,例如西班牙科学推文的第二语言是西班牙语,法国科学推文的第二语言是法语。德语曾经在世界科技交流中扮演重要角色,但是即使在德国本国,英语推文量竟高达95%,德语科学推文自身只占到约2%。还有南美诸国以及北欧诸国均是如此。以英语为母语的国家的英文推文比例均高达90%以上,例如北美和大洋洲的国家。非英语母语的国家的国家中,有少数国家的母语科学推文比例与英语科学推文比例相近,例如日本的英语科学推文占约53%,而日语科学推文占约44%。不论哪个国家,占主导性的前三种语言的科学推文均占到所有推文的95%以上。

这些结果表明:①英语已成为非正式科学交流中的通用语。即使没有外界的显性压力,为了扩大交流圈子和提高影响力,各国作者在推特上传播和讨论科学研究成果的时候,也会选择使用英语作为媒介。②英语科学推文的高

比例,客观上反映了科学研究全球化带来的影响。大科学时代,各国学者展开的科研合作达到了前所未有的深度和广度。不同国家的科学家在交流时,学习和使用范围最广的英语成了主流选择,因此不管英语母语国家还是非英语母语国家的国际交流,各国学者都在使用英语进行沟通和讨论,科学推文也可视作这种讨论的延伸。③科学推文的语言分布,在一定程度上反映了各国科学文化上的差异。沙特阿拉伯是G20集团唯一的阿拉伯国家,也是相对保守的伊斯兰教国家,在科学交流中倾向于使用阿拉伯语,所以成为唯一一个英语科学推文不占主导地位的国家。日本是世界第三大经济体,科技极为发达,拥有发达的科技翻译产业,科技工作者即使不懂英语,也能及时阅读到最新的科技成果,这种科技自信和便利的科技文献交流体系,使得日语科学推文与英文科学推文相当。

表 2 世界主要国家科学推文语言分布

国家	区域	第一语言	比例 (%)	第二语言	比例 (%)	第三语言	比例 (%)	其他语言 比例 (%)
西班牙	西欧	英语	75.2	西班牙语	20.4	加泰罗尼亚语	1.6	2.9
法国	西欧	英语	78.9	法语	17.7	葡萄牙语	2.4	1.1
德国	西欧	英语	95.3	德语	2.2	意大利语	0.8	1.8
荷兰	西欧	英语	85.3	荷兰语	13.2	法语	0.4	1.1
俄国	西欧	英语	85.6	俄语	8.7	德语	1.5	4.2
日本	东亚	英语	52.9	日语	43.6	韩语	0.7	2.8
韩国	东亚	英语	91.1	韩语	3.7	日语	3.5	1.7
泰国	东亚	英语	95.1	泰语	2.9	罗马尼亚语	1.4	0.6
墨西哥	南美	英语	90.2	西班牙语	7.2	法语	1.1	1.5
巴西	南美	英语	88.3	葡萄牙语	6.9	加泰罗尼亚语	1.3	3.4
瑞典	北欧	英语	89.8	瑞典语	6.2	法语	1.6	2.5
芬兰	北欧	英语	94.0	芬兰语	4.0	索马里语	0.4	1.7
挪威	北欧	英语	86.6	挪威语	9.6	意大利语	1.4	2.5
南非	非洲	英语	98.5	罗马尼亚语	0.3	波兰语	0.2	1.0
肯尼亚	非洲	英语	97.6	索马里语	1.0	法语	0.7	0.7
埃及	非洲	英语	88.9	阿拉伯语	7.4	法语	2.5	1.2
沙特阿拉伯	西亚	阿拉伯语	68.1	英语	30.7	索马里语	0.5	0.7
阿联酋	西亚	英语	88.6	阿拉伯语	7.6	加泰罗尼亚语	3.0	0.8
希腊	东欧	英语	93.1	希腊语	5.2	法语	0.8	1.0
土耳其	东欧	英语	80.7	土耳其语	17.1	挪威语	0.9	1.3
波兰	东欧	英语	91.8	波兰语	4.9	法语	1.6	1.6

2.4 不同语种科技文献的科学推文语言分布

表 3 展示了不同语种科技文献的科学推文语言分布情况,本数据集中 Scopus 共收录了 25 种语言的文献,但是仅有英语、德语、日语、法语、葡萄牙语、西班牙语六种语种文献有科学推文,这些语种在 Scopus 中被收录的数量都在前八位,体现了推特这类社交媒体与文献题录数

据库对这些语种科技文献关注度的一致性。但是我们知道,中文文献和土耳其语文献是仅次于英语文献被 Scopus 收录最多的文献,可是却没有获得一条科学推文。说明尽管从引文角度来说,这两个语种的文献论文质量都较高,但是其在社交媒体上被关注的程度较低。

表 3 不同语种科技文献科学推文的语言分布

文献语种	第一推文语种 比例 (%)	第二推文语种 比例 (%)	第三推文语种 比例 (%)	第四推文语种 比例 (%)
英语文献	英语 (93.2)	西班牙语 (1.6)	日语 (1.2)	法语 (1.2)
德语文献	英语 (93.6)	罗马尼亚语 (3.7)	印尼语 (1.8)	德语 (0.9)
日语文献	日语 (90.9)	英语 (9.1)		
法语文献	英语 (49.4)	法语 (46.3)	葡萄牙语 (4.2)	
葡萄牙语文献	英语 (94.0)	罗马尼亚语 (3.8)	印尼语 (1.9)	葡萄牙语 (0.3)
西班牙语文献	英语 (93.0)	罗马尼亚语 (3.7)	印尼语 (1.9)	西班牙语 (1.4)

3 结论与讨论

3.1 主要结论

本文的主要结论有以下几点。

(1)从科学推文总体数量分布来看,英语在非正式科学交流中的比例最高。科学推文语言分布集中在英语(约占91%)、日语(约占2.4%)和西班牙语(约占1.7%)。相较之下,科技文献语言分布集中在英语(约占94.2%)、中文(约占4.3%)和土耳其语(约占0.4%),并且这种集中趋势与过去^[14]相比还在加强。

(2)从学科层次科学推文语言分布来看,非正式科学交流中的语言分布存在学科差异。例如法语作者关注商业、管理与会计学较多,罗马尼亚语作者关注材料科学较多,这种学科差异体现了不同语言作者对特定学科的关注程度不同。相较之下,学科层次科技文献语言分布反映了个别国家在特定学科的优势地位,例如德国在兽医学和保健科学很强,法国在牙医学方面表现很好。

(3)从国家层次科学推文语言分布来看,英语无疑成为非正式科学交流的通用语。世界各国不论母语是不是英语,均以英语科学推文居

主导地位,母语科学推文通常屈居第二,沙特阿拉伯例外。不论哪个国家,科学推文主导性的前三种语言均几乎占到所有推文的95%以上。各国科学推文语言分布在一定程度上也体现了科技文化差异,例如日本的日语科学推文与英语科学推文比例相当。

(4)从不同语种科技文献语种层次科学推文的语言分布来看,科学推文集中在英语、德语、日语、法语、葡萄牙语和西班牙语六个语种科技文献上,科技文献比例较高的中文和土耳其文没有获得任何科学推文,说明这两种语言科技文献在非正式科学交流语境下受关注度不高。

3.2 讨论与思考

值得注意的是,科学推文的语言分布与推文的语言分布是不同的。根据 Mocanu 等人^[15]的大规模研究,推文中使用最多的语言依次是英语、西班牙语、印尼语、马来语和日语,而本研究指出科学推文中使用最多的语言依次是英语、日语、西班牙语、法语和阿拉伯语。不仅如此,Mocanu 等人^[15]的研究也测定了各国推文语言分布情况,结果均以母语为主,与该国各语言使用者数量分布保持一致。对比之下,各个国家的科学推文语言分布与该国的推文语言分布

也是不同的。这些差异说明,科学推文的语言分布确实反映了作者在科学交流中的语言偏好,而与语言使用人数没有直接关系,从而切实揭示了非正式科学交流中的语言分布特征。

中文文献和土耳其文献在 Scopus 中收录的比例较高,却没有获得任何相应的科学推文。这一方面可能是因为中国和土耳其均对推特的使用做了限制,使得本国作者很难发表科学推文,另一方面也反映了其他国家作者对中文和土耳其语科技文献没有表现出很大的兴趣,语言障碍的影响可能是重要原因。

综上所述,英语不仅是正式科学交流中的通用语,占据垄断地位,而且在非正式科学交流中,也已经成了通用语。英语科学推文在世界各国的主导地位,客观反映了大科学时代科研全球化带来的影响。当各国科学家走出国门参与国际合作交流的时候,英语这个业已被众多国家作为第二语言学习的语言成了不二之选。所以,不管是中国学者到了日本,还是德国学者到了瑞典,不论是学术会议还是颁奖典礼,英语都成为沟通和连接不同母语者的工具。体现在科学推文上,非英语母语国家作者选择使用英语去推广研究成果,事实上,他们中的许多人已经完全使用英语去发推。英语的这种媒介作用

毋庸置疑具有积极意义,但是对其他语言产生的消极作用也很明显,使得原本在正式科学交流中已经突显的弊端更加严重,非英语母语国家的学者为了学习和使用英语付出了巨大成本,却在国际会议、国际研讨等重要方面很难进入核心角色。但是也有例外,沙特阿拉伯和日本刚好是两个典型,不管是因为文化保守,还是科技自信和制度保障,这两个国家作者坚持用母语来发科学推文。

本研究的主要局限是数据来源可能本身存在不足,虽然 Scopus 宣称大约 21% 的文献是非英语文献,包涵 40 余种本地语言^①,但是从 2015 年 6 月的收录情况来看,非英语文献仅占 6%,且只有 25 种语言的文献。与此同时,本文从科学推文视角出发,并不意味着科学推文代表了所有科学家的非正式科学交流,推特上的非正式科学交流形式也不只科学推文一种,例如,推文内容与科学相关,却不涉及学术成果的讨论,是另一种形式的科学交流,只不过尚难以捕捉。中国对推特平台的使用做了限制,所以中国不在国家层次科学推文语言分布的分析范围。

致谢:感谢 Altmetric.com 公司为本研究提供的数据支持和武汉大学邱均平教授的指导。

参考文献

- [1] Williams C H. Language in geographic context[J]. *Modern Language Journal*, 1988, (73):13-25.
- [2] Atlantic T. The hidden bias of science's universal language[EB/OL]. [2017-01-25]. <http://www.university-worldnews.com/article.php?story=20150828105145992>.
- [3] 石国进,龚耘. 科学是平的——“汤浅现象”新解[J]. *科学学研究*, 2012, (12): 1769-1780. (Shi G J, Gong Y. Science is flat: new explanation of Yuasa phenomenon[J]. *Studies in Science of Science*, 2012, (12): 1769-1780.)
- [4] Flowerdew J. The non-anglophone scholar on the periphery of scholarly publication[J]. *AILA Review*, 2007, 20(1): 14-27.
- [5] Canagarajah A S. *A Geopolitics of academic writing*[M]. University of Pittsburgh Press, 2002.
- [6] Canagarajah A S. *Non-discursive requirements in academic publishing, material resources of periphery scholars*,

① 见 https://www.elsevier.com/_data/assets/pdf_file/0007/69451/scopus_content_coverage_guide.pdf.

- and the politics of knowledge production[J]. Written Communication, 1996, 13(13): 435-472.
- [7] Brambrink A M, Ehrler D, Dick W F. Publications on paediatric anaesthesia: a quantitative analysis of publication activity and international recognition[J]. British Journal of Anaesthesia, 2000, 85(4): 62-556.
- [8] Ammon U. The dominance of English as a language of science[M]. Walter de Gruyter, 2001: 3-26.
- [9] Butler D. French scientists turn to journals in English[J]. Nature, 2000, 405(6786): 500.
- [10] Dalen H P V, Henkens K. What makes a scientific article influential? the case of demographers[J]. Scientometrics, 2001, 50(3): 455-482.
- [11] Rowlands I, Nicholas D, Russell B, et al. Social media use in the research workflow[J]. Learned Publishing, 2011, 24(3): 183-195.
- [12] Adie E, Roe W. Altmetric: enriching scholarly content with article-level discussion and metrics[J]. Learned Publishing, 2013, 26(1): 11-17.
- [13] Yu H Q, Xu S M, Xiao T T, et al. Global science discussed in local altmetrics: Weibo and its comparison with Twitter[J]. Journal of Informetrics, 2017, 11(2): 466-482.
- [14] Albarillo F. Language in Social Science databases: English versus Non-English articles in JSTOR and Scopus[J]. Behavioral & Social Sciences Librarian, 2014, 33(2): 77-90.
- [15] Mocanu D, Baronchelli A, Perra N, et al. The twitter of babel: mapping world languages through microblogging platforms[J]. PLoS ONE, 2012, 8(4): e61981.

余厚强 南京理工大学经济管理学院信息管理系副教授。江苏 南京 210094

董 克 武汉大学信息管理学院讲师。湖北 武汉 230074

王曰芬 南京理工大学经济管理学院信息管理系教授。江苏 南京 210094

章成志 南京理工大学经济管理学院信息管理系教授。江苏 南京 210094

(收稿日期:2017-08-19;修回日期:2017-11-09)