KUNGL.
VETENSKAPS-
AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

CrossMark

REPORT

Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications

Mario S. Di Bitetti, Julián A. Ferreras

Abstract There is a tendency for non-native English scientists to publish exclusively in English, assuming that this will make their articles more visible and cited. We tested this hypothesis by comparing the effect of language on the number of citations of articles published in six natural sciences journals from five countries that publish papers in either English or other languages. We analyzed the effect of language (English vs non-English), paper length, and year of publication on the number of citations. The articles published in English have a higher number of citations than those published in other languages, when the effect of journal, year of publication, and paper length are statistically controlled. This may result because English articles are accessible to a larger audience, but other factors need to be explored. Universities and scientific institutions should be aware of this situation and improve the teaching of English, especially in the natural sciences.

Keywords English · *Lingua franca* · Scientific evaluation · Scientific journals · Scientific literature

INTRODUCTION

Within the last few decades, English became the dominant language of science, with more than 90 % of the indexed scientific articles in the natural sciences been published in this language (Hamel 2007, 2013; Ammon 2010, 2012). Researchers that want to communicate their results to the global scientific community seem to have no other option than to write and publish their articles in English. Having a

lingua franca allows researchers from all over the world to communicate in one common language. This simplifies things because mastering one foreign language instead of three or more allows a researcher familiar with that language (currently English) to communicate and have access to information produced by researchers all over the world. However, this situation is unfair, since the outcome of writing in an unfamiliar language is usually far from desired (La Madeleine 2007) and, as a result, native English-speaking (NES) countries and researchers are at a clear advantage in science communication in a highly competitive arena (Ammon 2007).

This hinderance faced by non-NES researchers affects their scientific output. For example, Brazilian researchers with good English writing abilities outperform those with poor ones in terms of published papers, number of citations, and h-indices (Vasconcelos et al. 2008). On a global comparison, English proficiency is a good predictor of publication output in top medical journals, even more than total research spending (Man et al. 2004). Thus, at all levels, from individuals to countries, NES countries have an advantage over non-NES, especially over those with poor English training.

Despite the alleged advantages of writing in English, non-NES scientists still publish papers in their native language. This is justified, especially since some studies may be of more local or regional interest than global interest (e.g., research with local applications instead of more theoretical work). If the scientific work has national or regional importance, researchers would like to make this information available to the decision makers in their vernacular language (Bortolus 2012).

In the last decades, there has been a tendency for journals from non-NES countries to turn into English-only as their language of publication (Bordons and Gómez 2004).

\springer Springer

Even for Latin American journals that did not adopt this policy, some Editors have encouraged their potential writers to submit the manuscripts in English instead of Spanish (e.g., Agudelo 2010). However, in most non-NES countries there are journals that resist this tendency and still publish papers in their native languages, usually allowing the use of English as an alternative language. Even though at an individual level, it is known that publishing in English increases the exposure of the research work; there is no empirical evidence, as far as we are aware, that journals which accept contributions in either English or its original language are also handicapped. One possible detriment of this editorial policy is that articles published in the original language would be less cited than those published in English. If this is the case, these journals are being penalized, since indices like h, ISI (Thomson Reuters), or SJR (Elsevier) that quantify the impact of a journal in a scientific community are based on the number of citations that the average paper receives.

There are still many journals across the world that accept the submission of manuscripts and publication of articles in either English or the original language, some of them with more or less even participation of contributions in English and other languages. We used a sample of articles published in some of these journals to test the hypothesis that within a journal (an important variable to control), papers published in English have a higher probability of being cited or have a higher citation rate than those published in the original language of the publication.

MATERIALS AND METHODS

To test the hypothesis stated in the previous paragraph, we compared the probability of citation and the number of citations of papers published in six journals from five different countries. Three of them are available in the Scientific Electronic Library Online (SciELO): *Mastozoología Neotropical*, a journal published by the Argentine Society for the Study of Mammals; *Revista Mexicana de Biodiversidad*, published by the Universidad Nacional Autónoma de México; and *Revista Argentina de Microbiología*, published by Elsevier for the Argentine Association of Microbiology. These three journals have similar Scimago Journal Rank (SJR) indicators (MN = 0.26, RMB = 0.25, RAM = 0.33) and h-indices (MN = 7, RMB = 12, RAM = 17) in 2014. The other three other journals analyzed are *Acta Botanica Gallica* (France, SJR = 0.18, h index = 14), *Journal of Japanese Botany* (Japan, SJR = 0.23, h index = 10), and the *Journal of the Korean Chemical Society* (South Korea, SJR = 0.20, h index = 11).

We used Scopus to retrieve the number of citations of the papers published in these journals. For *Mastozoología Neotropical*, we looked (January 17–18 2016) for the number of citations of papers published from Vol 16 No1 year 2009 to Vol 21 No1 year 2014 (with the exception of Vol 19 No 1 of 2012 for which, for unknown reasons, we could not retrieve the number of citations with Scopus), totalizing 188 papers. For the *Revista Argentina de Microbiología*, we looked (February 9-12 2016) for the number of citations of papers published between 2009 and 2012 (Vol. 41 and 44), totalizing 197 papers. For the *Revista Mexicana de Biodiversidad,* we looked (January 19–20 2016) for the number of citations of papers published in 2010 and 2011 (Vol. 81 and 82), totalizing 241 papers. For the other three journals, we tallied (on February 19 2016) the language and the number of citations of all the articles published between 2010 and 2012 (168 articles in *Acta Botanica Gallica*, 154 in *Journal of Japanese Botany* and 380 in *Journal of the Korean Chemical Society*). We did not include in the analyses the editorials, book reviews, and obituaries. We used the language (English vs non-English), journal, the year of publication, and paper length (this variable only estimated for the three Latin American journals as last page number − first page number + 1) as predictor variables of the probability of citation and the number of citations.

We ran generalized linear models of an ANCOVA-like logistic regression type (family = binomial or quasi-binomial), using if the paper was cited at least once or not as the response variable, and language (English vs Non-English or English vs Spanish), the natural logarithm of paper length, year of publication, and journal as predictor variables. We conducted two separated analyses, one for the six journals (that did not include the natural logarithm of paper length as a predictor variable, which was only estimated for the Latin American journals) and another for the three Latin American journals, excluding in this case the nine articles published in Portuguese in *Mastozoología Neotropical*. We also conducted log-linear model tests, using as a Poisson (or quasi-Poisson) model response variable the number of citations of the articles, to assess the effect of the same predictor variables on it. Since in the binomial and Poisson models, the residual deviance was larger than the residual degrees of freedom, an indication of over-dispersion, we ran the models with a quasi-binomial and a quasi-Poisson response error structure (Crawley 2007), and we reported the results of these latter models henceforth. We ran all the analyses with R (R Core Team 2015), setting the probability level to commit a type I statistical error at 0.05.

RESULTS

Of the 1328 articles analyzed, 728 (54.8 %) were published in English. Of the papers published in English, 33.7 % had

not been cited, while 46.3 % of the articles published in other languages had not been cited when we conducted the bibliographic search (Fig. 1). This general pattern of a higher probability of citation of the articles published in English was found in the six journals. Of the 188 papers published in *Mastozoología Neotropical*, 96 were published in Spanish, 83 in English, and nine in Portuguese. Of the papers published in Spanish, 48 % had not been cited by the moment of the bibliographic search, while only 23 % of the papers in English had not been cited (Fig. 2a). Of the 197 articles of the *Revista Argentina de Microbiología*, 113 were published in Spanish and 84 in English. Twenty-five percent of the papers published in Spanish had not been cited, while only 12 % of those published in English had not been cited (Fig. 2b). Of the 143 papers published in Spanish in *Revista Mexicana de Biodiversidad*, 31 % had not been cited when we conducted this analysis, while only 18 % of the 98 published in English had not been cited (Fig. 2c). Sixty-four percent of the 64 articles published in French in *Acta Botanica Gallica* had not been cited compared to only 27 % of the 104 articles published in English (Fig. 2d). Of the 40 articles published in Japanese in *Journal of Japanese Botany*, 78 % had not been cited compared to 67 % of the 114 articles published in English (Fig. 2e). Finally, of the 135 articles published in Korean in *Journal of the Korean Chemical Society*, 64 % had not been cited compared to only 38 % of the 245 articles published in English (Fig. 2f).

An ANCOVA-like multiple logistic regression model (family quasi-binomial) indicates that the probability that a paper was cited was lower for those published in languages other than English ($t = -7.806$, df = 1320, $P < 0.001$). It decreased with the year of publication (was lower in more recently published articles, $t = -3.593$, df = 1320, $P < 0.001$) and differed among journals, being higher in the *Revista Argentina de Microbiología* ($t = 4.952$, df = 1320, $P < 0.001$) and lower in the *Journal of Japanese Botany* ($t = -5.586$, df = 1320, $P < 0.001$). When we restricted the analysis to the three Latin American journals, an ANCOVA-like multiple logistic regression model (family quasi-binomial) indicates that the probability that a paper was cited was lower for those published in Spanish than for those published in English ($t = -4.938$, df = 611, $P < 0.001$). It decreased with the year of publication (was lower in more recently published articles, $t = -3.958$, df = 611, $P < 0.001$), was higher in the *Revista Argentina de Microbiología* than in the other two journals ($t = 5.390$, df = 611, $P < 0.001$), and increased with the natural logarithm of paper length ($t = 6.896$, df = 611, $P < 0.001$). Similar results were obtained when we analyzed the effect of these variables on the number of citations with quasi-Poisson generalized linear models, being lower in articles published in languages other than English, decreasing with the year of publication, increasing with paper length, and showing statistical differences among journals. We found no evidence for statistically significant interactions between journal (as a surrogate for scientific field) and the other variables in the binomial and Poisson models.

DISCUSSION

These results suggest that there are several variables that affect the probability that an article is cited or its number of citations. These variables include the journal, paper length, year of publication, and the language. We showed that articles written in English have a higher chance of being cited and a higher number of citations than those published in other languages. So there seems to be a punishment, in terms of citation rates, for those journals accepting the publication of articles in languages other than English and for those non-NES scientists that publish their research work in their native language. It appears that for the non-NES scientists, the familiar phrase "Publish or perish" should more correctly be rephrased as "Publish in English or perish."

There may be other factors besides paper length, year of publication, and subject (journal) that may affect the probability that a paper is cited and that may confound the effect of language. For example, the origin of the researchers affects citation rates, with articles from Latin American authors being less cited than those of American or European origin (Meneghini et al. 2008). Recently, Letchford et al. (2015) found evidence that journals which publish articles with shorter titles get more citations per
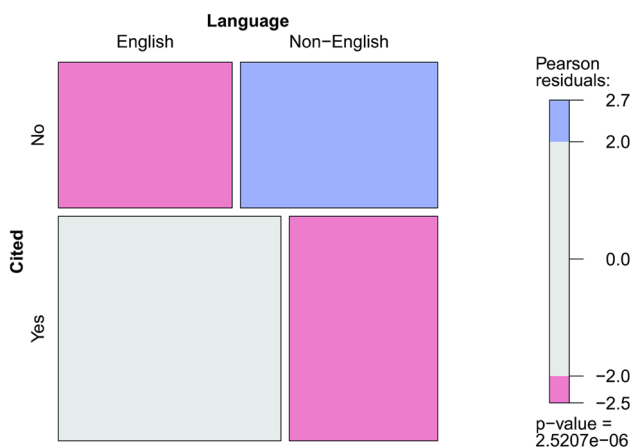

Fig. 1 Mosaic plot depicting the relative frequency of articles published in English and languages other than English plotted against the relative frequency of cited vs non-cited articles for six different journals: *Mastozoología Neotropical* (Argentina), *Revista Argentina de Microbiología* (Argentina), *Revista Mexicana de Biodiversidad* (Mexico), *Acta Botanica Gallica* (France), *Journal of Japanese Botany* (Japan), and *Journal of the Korean Chemical Association* (South Korea)
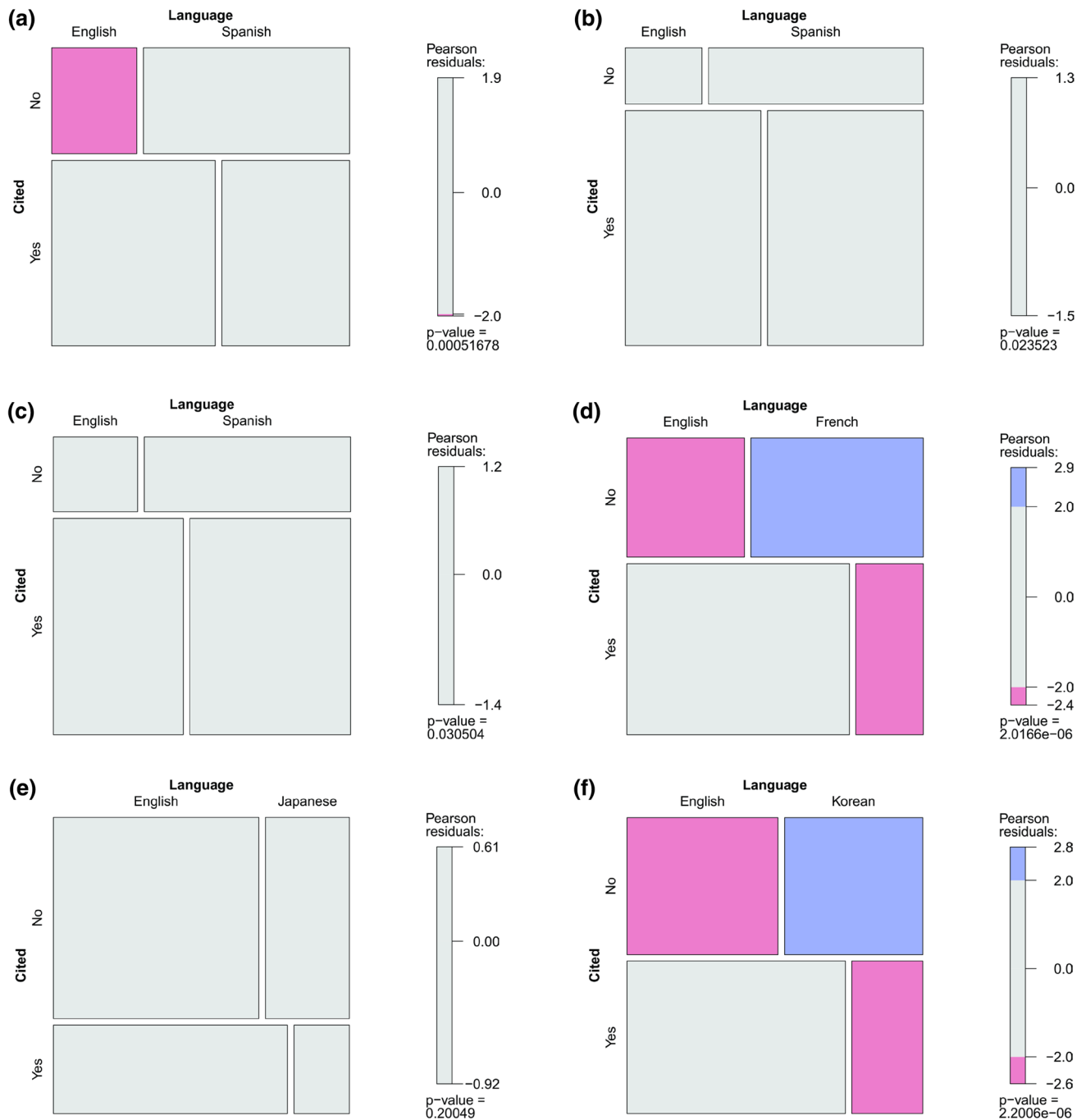
Fig. 2 Mosaic plot depicting the relative frequency of articles published in English and languages other than English plotted against the relative frequency of cited vs non-cited articles for the journals *Mastozoología Neotropical* (**a**), *Revista Argentina de Microbiología* (**b**), *Revista Mexicana de Biodiversidad* (**c**), *Acta Botanica Gallica* (France) (**d**), *Journal of Japanese Botany* (Japan) (**e**), and *Journal of the Korean Chemical Association* (South Korea) (**f**). Papers published in Portuguese in *Mastozoología Neotropical* ($N = 9$) were omitted

article than those with longer ones. There may be also other factors to consider, which could explain the strong differences in citation rates between articles written in English and those written in other languages. It is possible that authors or research groups with a higher impact on the scientific community, and thus more frequently cited, have a higher tendency to publish their work in English. Another reason may be that scientists prefer to cite articles that could be read by a broader audience (Breeze 2015). It is also possible that there is a tendency for articles of more general interest to be published in English in the journals we have analyzed and, thus, are cited more often. The idea that articles published in languages other than English are parochial or of poor quality is a widely held idea and

sustained, without empirical evidence, by both scientists and Editors (Stolerman and Stenius 2008). This may be one of the reasons why articles written in languages other than English tend to be cited less (Breeze 2015).

As a preliminary test of these last ideas, we decided to compare different attributes of papers published either in English or Spanish in the same journal. To avoid any bias from our perspective, we asked colleagues with expertise in the subjects of the journal and with experience both as authors and reviewers to perform the evaluation. They had to compare pairs of abstracts of the articles published in *Revista Argentina de Microbiología* (114 papers, 57 pairs) and *Revista Mexicana de Biodiversidad* (184 papers, 92 pairs), one member of the pair belonging to an article published in English and the other in Spanish, and to decide, based on the contents of the abstract, which one (abstract A or B) was of (a) higher general interest, (b) higher quality, (c) more original, (d) more theoretically oriented, and (e) more basic oriented (as opposed to applied). If they considered that there were no differences between articles, we asked them to choose the third option (C). The reviewers were blind in regards to the language of the article as all abstracts were presented in Spanish (independently of the language of the article all papers in these journals have an abstract in Spanish). The pairing, order of presentation (A vs B), and assignment of the abstracts to the reviewers were randomized.

The results of this analysis (Supplementary material S1) suggest that it was not easy for the reviewers to decide whether the English or Spanish article had a higher score on the selected characteristics. This is evidenced by the high frequency of option C chosen by the reviewers. It is also noted by the fact that, for *Revista Mexicana de Biodiversidad*, we sent several of the pairs ($N = 39$) to more than one reviewer in order to assess for consistency among them and, with the exception of quality, there was no statistical association between the decisions of different reviewers. This exercise had an important limitation though: the reviewers were judging the quality, general interest, etc., of the article based on the contents of the abstract, not the article itself. Even with these limitations, we found that abstracts of articles published in English were not consistently chosen as being of higher quality, general interest, novelty, etc. than those published in Spanish.

Another evidence that there is no difference in quality, general interest, etc., of articles published in English and other languages, at least in two of the journals analyzed, comes from the comparison of rejection and acceptance rates of manuscripts sent to *Revista Argentina de Microbiología* and *Mastozoología Neotropical*. It may be possible that submitted manuscripts written in English may be of better quality, general interest, or more original than those written in languages other than English. This preexisting

bias may partially explain why articles written in English are more cited than those written in other languages. However, based on data provided by the Editors, we found no difference in rejection rate of articles submitted in English and other languages in these two journals (Supplementary material S2). This further suggest that there are no objective differences in aspects, such as quality, general interest, and the like, between articles published in English and other languages in these journals. This was expected since Editors of these journals are fluent in both languages, and reviewers are asked to make recommendations based solely on the quality of the work, without consideration to other factors. We are inclined to believe that, besides the fact that the articles published in English potentially have a larger audience, there is no other objective difference between the articles published in this language and languages other than English in these journals. It remains possible that subjective aspects play an important role, and there is ample scope to explore how prejudices in relation to language affect how often an article is cited.

We hope that our analysis contributes to make better informed decisions about publishing policies. These decisions are important at different levels. First, for countries whose national governments provide support to the journals of scientific societies (Meneghini 2012), it is important to know that if these journals publish their articles in English, the impact, in terms of citations, may increase. Second, for the national or regional scientific societies and for editorial boards, it is important to know that a change of language of publishing policy of the manuscripts submitted may affect the impact of the journal. Finally, at the individual level, this issue is important for the researchers involved in this endeavor, since they are the ones that ultimately have to publish (or perish) and compete for scientific positions, grants, etc., and that are going to be evaluated, to a high degree, by the impact of their work as quantified by indices based on the number of citations (Lawrence 2008; Alberts 2013).

For several reasons, we think it is good and necessary to have a *lingua franca* and, *de facto*, this lingua is English. First, closed regional scientific communities sharing a common language not accessible to a broader audience may have a tendency to produce a very limited perspective of science, something that has been termed *institutional provincialism* (Stolerman and Stenius 2008). Second, not having a *lingua franca* may retard the dissemination of important scientific knowledge. There are many examples of scientific discoveries that took several years to be known by the global community because they were published in languages other than English (Meneghini and Packer 2007; Stolerman and Stenius 2008; Swales and Feak in Breeze 2015). Finally, not having a *lingua franca* will hamper international scientific collaborations which may help

bringing some important but neglected scientific topics to light. What we need is more solidarity among colleagues and scientific institutions worldwide to help those handicapped societies and scientists to participate in the global scientific enterprise (Clavero 2010, 2011).

As it has been already pointed out (Bortolus 2012), publishing scientific papers exclusively in English may limit the ability of non-NES scientists to communicate important results to local practitioners and decision makers (e.g., environmental managers). 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. This dilemma could be easily solved by following some of the recommendations already suggested to facilitate bi-lingual or multi-lingual publication of scientific papers (e.g., Meneghini and Packer 2007; Root-Bernstein and Ladle 2014). However, these suggestions should not be seen as an alternative to promoting the learning of scientific writing in English (Guariguata et al. 2010). Universities and scientific institutions of non-NES countries should improve the teaching and learning of scientific writing and of the English language at graduate level and for doctoral students (Vasconcelos 2007; Vasconcelos et al. 2007).

To conclude, we have shown that the language in which a scientific article is written affects its probability of citation, which may indirectly reflect the impact this paper had on the scientific community. Even though the imposition of English as the *lingua franca* poses a handicap to non-NES countries and researchers, under current circumstances promoting the use of their native languages in scientific publications will not help these countries or researchers promote their scientific agendas or careers. Even though having a *lingua franca* helps global communication and collaboration among scientists worldwide, it may negatively affect local transmission of knowledge if the researchers are pushed to publish only and exclusively in English, for which publishing systems promoting plurilingual communication of scientific results should be encouraged. Finally, we should promote a drastic change in the current evaluation system of researchers, mainly based on measuring the amount of scientific output (of articles, of citations) and turn it into a scheme that evaluates the quality of scientific work (Lawrence 2008; Sarewitz 2016) and the benefits reaped by society as a result of this work.

REFERENCES

Agudelo, J.H. 2010. Editorial: publicar en Inglés. *Revista Colombiana de Ciencias Pecuarias* 24: 1.

Alberts, B. 2013. Impact factor distortions. *Science* 340: 787.

Ammon, U. 2007. Global scientific communication: Open questions and policy suggestions. *AILA Review* 20: 123–133.

Ammon, U. 2010. The hegemony of English. In *World Social Science Report. Knowledge Divides*. Paris: UNESCO Publishing, 154 f. http://www.unesco.org/shs/wssr.

Ammon, U. 2012. Linguistic inequality and its effects on participation in scientific discourse and on global knowledge accumulation—With a closer look at the problems of the second-rank language communities. *Applied Linguistics Review* 3: 333–355.

Bordons, M., and I. Gómez. 2004. Towards a single language in science? A Spanish view. *Serials* 17: 189–195.

Bortolus, A. 2012. Running like Alice and losing good ideas: on the quasi-compulsive use of English by non-native English speaking scientists. *Ambio* 41: 769–772.

Breeze, R. 2015. Citing oustide de community? An investigation of the language of bibliography in top journals. In *English as a Scientific and Research Language*, ed. R. Plo Alastrué and C. Pérez-Llantada, 37–58. Berlin: De Gruyter Mouton.

Clavero, M. 2010. ''Awkward wording. Rephrase'': Linguistic injustice in ecological journals. *Trends in Ecology & Evolution* 25: 552–553.

Clavero, M. 2011. Unfortunately, linguistic injustice matters. *Trends in Ecology & Evolution* 26: 156–157.

Crawley, M.J. 2007. *The R book*. West Sussex: Wiley.

Guariguata, M.R., D. Sheil, and D. Murdiyarso. 2010. 'Linguistic injustice' is not black and White. *Trends in Ecology & Evolution* 26: 58–59.

Hamel, R.E. 2007. The dominance of English in the international scientific periodical literature and the future of language use in science. *AILA Review* 20: 53–71.

Hamel, R.E. 2013. El campo de las ciencias y la educación superior entre el monopolio del inglés y el plurilingüismo: Elementos para una política del leguaje en América Latina. *Trabalhos em Linguística Aplicada* 52: 321–384.

La Madeleine, B.L. 2007. Lost in translation. *Nature* 445: 454–455.

Lawrence, P.A. 2008. Lost in publication: How measurement harms science. *Ethics in Science and Environmental Politics* 8: 9–11.

Letchford, A., H.S. Moa, and T. Preis. 2015. The advantage of short paper titles. *Royal Society Open Science* 2: 150266.

Man, J.P., J.G. Weinkauf, M. Tsang, and D.D. Sin. 2004. Why do some countries publish more than others? An international comparison of research funding, English proficiency and

publication output in highly ranked general medical journals. *European Journal of Epidemiology* 19: 811–817.

Meneghini, R. 2012. Emerging journals: The benefits of and challenges for publishing scientific journals in and by emerging countries. *EMBO Reports* 13: 106–108.

Meneghini, R., and A.L. Packer. 2007. Is there science beyond English? *EMBO Reports* 8: 112–116.

Meneghini, R., A.L. Packer, and L. Nassi-Calò. 2008. Articles by Latin American authors in prestigious journals have fewer citations. *PLoS ONE* 3(11): e3804.

Root-Bernstein, M., and R.J. Ladle. 2014. Multilinguismo nas ciências ambientais: Ahora ya! [Multilingualism in Environmental Sciences: It's About Time!]. *Ambio* 43: 836–837.

R Core Team. 2015. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sarewitz, D. 2016. The pressure to publish pushes down quality. *Nature* 533: 147.

Stolerman, I.P., and K. Stenius. 2008. The language barrier and institutional provincialism in science. *Drug and Alcohol Dependence* 92: 1–2.

Vasconcelos, S.M.R. 2007. Writing up research in English: Choice or necessity? *Revista do Colégio Brasileiro de Cirurgiões* 34: 62–63.

Vasconcelos, S.M.R., M.M. Sorenson, and J. Leta. 2007. Scientist-friendly policies for non-native English-speaking authors: Timely and welcome. *Brazilian Journal of Medical and Biological Research* 40: 743–747.

Vasconcelos, S.M.R., M.M. Sorenson, and J. Leta. 2008. Researchers' writing competence: A bottleneck in the publication of Latin-American science? *EMBO Reports* 9: 700–702.

AUTHOR BIOGRAPHIES

Mario S. Di Bitetti (✉) is a Researcher and Director of the Instituto de Biología Subtropical (IBS), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) – Universidad Nacional de Misiones (UNaM), Argentina, and a Professor at the Facultad de Ciencias Forestales, UNaM. His research interests include the ecology and conservation of large mammals and tropical forests
Address: Instituto de Biología Subtropical (IBS) – nodo Iguazú, Universidad Nacional de Misiones (UNaM), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Calle Bertoni 85, 3370 Puerto Iguazú, Misiones, Argentina.
Address: Facultad de Ciencias Forestales, UNaM, Eldorado, Argentina.
Address: Asociación Civil Centro de Investigaciones del Bosque Atlántico (CeIBA), Puerto Iguazú, Argentina.
e-mail: dibitetti@yahoo.com.ar

Julián A. Ferreras is a Researcher and Vice-Director of the Instituto de Biología Subtropical (IBS), CONICET – UNaM, and a Professor at the Facultad de Ciencias Exactas Químicas y Naturales, UNaM. His research interests include the study of bacterial diseases and the diversity of bacteria.
Address: Instituto de Biología Subtropical (IBS) – nodo Posadas, Universidad Nacional de Misiones (UNaM), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Calle Jujuy 1745, 3300 Posadas, Misiones, Argentina.
Address: Facultad de Ciencias Exactas Químicas y Naturales, UNaM, Posadas, Argentina.
e-mail: juf2003@gmail.com