

## Language Barriers Create Knowledge Barriers in Biological Research

Eden Skopp

In a survey conducted in 2016, conservation biologists at the University of Cambridge found that 35.6% of scientific documents about conservation biodiversity available on Google Scholar in 2014 were published in a language other than English (Amano, González-Varo, & Sutherland, 2016). The conclusion of this paper, and many others agree that the standard for conducting research and publication is English. Knowledge of English opens opportunities to avenues of study and access to information, but this means extra labor for non-English speaking scientists. Research conducted in a scientist's native language can limit its success, because it does not fit the norm set by established institutions, but other researchers would not be able to access new discoveries in a language they do not understand. While using English as a set standard for research and publication allows for the establishment of some commonality, it is extremely isolating to non-native English speakers both conducting research and seeking to understand emergent scientific discoveries. This is even further complicated through the redundant stylization of scientific writing.

Papers published in English account for 86.47% of a pool of almost 20,000 papers sampled from the July-December edition of the journal, *Biological Abstracts* (Monge-Nájera & Nielsen, 2005). The next most commonly occurring language accounts for less than 2% (Chinese: 1.44%, Russian: 1.41%, Japanese: 1.01%, etc.) and at least 50% of academic web pages in Western European countries where the standard language is one other than English, are written in English. While the United States is cited as the most productive country by this study, a large portion of research from non-English speaking countries, is also published in English. It

shouldn't be assumed that all important information is available in English, but there is a lack of availability of access to research across language differences. What is the culture that encourages such an overdominance of English research, and what is the impact of this disproportionality?

In 2013, a study published in *BioScience* attempted to predict the publishing success of PhD holding biologists over the course of the next ten years (Laurance, Useche, Laurance, & Bradshaw). The factors that were considered in this study were gender, native language, prestige of the institution they received their PhD from, and their pre-PhD publication success. The Spearman rank correlations of these factors were presented and gender, and prestige did not yield significant values in terms of predicting publication success. However, when university rank and native language are considered together, they produce a significant correlation. This means that native English speakers that attend prestigious universities will achieve more publication success in their careers than non-native English speakers attending the same schools. Additionally, there are more native English speakers attending institutions of higher education than non-native English speakers, further driving this trend.

Along with language, other barriers exist that limit understanding and spread of scientific knowledge, including wealth, geographic location, and security (Amano & Sutherland, 2013). Language, location, and security within the context of an areas available wealth are significant factors to the availability of information. Countries with higher GDP's will have larger budgets available to support higher quantities and higher quality scientific research, while countries with varying priorities (threat of war, humanitarian crises) and/or less funds available will focus their resources in other areas. In another approach, even if a country is highly productive, low English skills can also intellectually isolate the country. Large volumes of research in the native language remain inaccessible to the English-operating scientific community at large (which does not only

include English speaking countries, but also countries that have high English skills and also have their research available in English), and large volumes of English research remain inaccessible to the country because the information is not available in the native language.

There must be some incentive to move away from this discriminatory trend. With the rising prevalence of online publication and information sharing, there is greater and more equalizing access to global research, and because of an increased interaction, higher demand for and availability of translated materials. A program known as Scientific Electronic Library Online (SciELO) is an open metapublisher established by several Brazilian, Latin American, and Caribbean research cohorts that should act as a role model for the success of open access to information brought to the international stage. Since its establishment in 1997, SciELO has been recognized at one of Google Scholar's top ten most accessed sites (Meneghini & Packer, 2007). In fact, several sources included in this paper were accessed through SciELO. The articles available on SciELO are reviewed by an ad hoc panel to ensure the same high level of quality and relevance to the current trends in the field expected of established journals. Finally, the founders of SciELO saw the value of this resource to be able to collect mass amounts of data in one standardized location to more easily track trends of information that would help track and understand performance indicators. At present, around 30% of the articles on SciELO are available in English ONLY. This means that a huge majority of the information can be accessed by speakers of languages, as well as English, but not just English speakers. The journals that publish on SciELO can decide the language of the content, while SciELO's only criteria is that it fits rigorous standards of quality for the use of the scientific community at large. There are more articles available in Portuguese, because the database was established by Brazilian scientific institutions for the purpose of collecting and spreading Brazilian research, but the model and

values of SciELO demonstrate a global and inclusive approach to sharing information across languages.

Biological writing prides itself in its objectivity. Research papers explain figures and interpret the meaning of results in the context of the study and previous knowledge. There is no room for personal bias when presenting evidence, and even the failure of the experiment is presented in the results. To support this standardized and formulaic approach present throughout all of scientific research, scientists learn habits to identify lexical bundles, which are formulaic groups of words that appear throughout the paper and similarly in other papers that indicate an equally standard function, to clearly understand the paper. Research oriented lexical bundles indicate aspects like date and time, procedural instruction, topic, among other things. Text oriented lexical bundles show transitions to link concepts, show the presentation of results, or indicate the limiting factor in the findings (Hyland & Tse, 2009)

*Table 1: Examples copied from Hyland & Tse, Academic Lexis and Disciplinary Practice: Corpus Evidence for Specificity*

Examples of research lexical bundles	
Location	at the start of, at the same time, in the present study
Procedure	the use of the, the role of the, the purpose of the, the operation of the
Quantifications	he magnitude of the, a wide range of, one of the most
Examples of text oriented lexical bundles	
Transition signals	“on the other hand, in addition to the, in contrast to the”
Resultative signals	“as a result of, it was found that, these results suggest that”
Framing signals	“in the case of, with respect to the, on the basis of, in the presence of, with the exception of”

Even though papers are designed to follow a formula to make them easy to understand, stylized writing known as hedging makes the text difficult to understand. While the main aim of paper writing is to remain passive and objective in tone, a practice known as “hedging” acts as a gatekeeping mechanism of overcomplications and flourishes that act in opposition to the straightforward presentation of research that a paper serves to achieve. Even in English, sometimes these papers are nearly impossible to understand. The reader wastes time wading through redundancies as the author dances around the statement, and the impact of the writer’s point is lost in unnecessary supplemental content that serves no purpose other than making the sentence longer. For a field that prides itself in the presentation of clear facts, I think it could do without dancing around the truth. In fact there are several formulas and indices available that account for average sentence length, average number of syllables per word, and percentage of “difficult” words taken from a random, representative, sample of the text (Zamanian & Heydari, 2012). If you can assume that all science articles have a large number of difficult words to account for technical terminology, papers that have unnecessarily long sentences and longer words would not score high for readability.

However, if a researcher made a bold outright claim, it would be embarrassing and discrediting if their confident statement was later disproven. The importance of hedging is to present new knowledge and possible reasons, but not accept it as finality to leave the issue open to further research and study (Hyland, 1996). It shows a proposed answer and the current version of the truth. Hedging has become engrained in the rhetoric of science, and according to Amin Brendan this is because senior researchers pass these habits on to their juniors (2013). Even though hedging serves a purpose, they write that the only reason that overcomplicated style still persists, is to show off and blow a lot of hot air to show how smart you are. When the route of

simplicity can be taken, it isn't. This harms scientific communication and makes research unintelligible for no reason other than style.

As a student, not being able to understand research written in my first language because of complicated is extremely frustrating, but the more severe issue is non-English speakers not being able to understand research written in English, or have their work understood and recognized by the English-standard community. This limits universal access to knowledge and information sharing, when it is imperative that scientists interact with different sources. SciELO is a positive and meaningful step to level the playing field and transcend language barriers, but information should be equally available across language barriers as well, without demanding unequal labor from non-English speaking scientists. This can possibly be achieved by offering incentive for making research available in multiple languages, which becomes easier as more research is available to anyone with internet access.

## References

- Amano, T., González-Varo, J. P., & Sutherland, W. J. (2016). Languages Are Still a Major Barrier to Global Science. *PLOS Biology*, *10*(1371).
- Amano, T., & Sutherland, W. J. (2013). *Four barriers to the global understanding of biodiversity conservation: wealth, language, geographical location and security*. Retrieved from
- Bredan, A. (2013). *Inheritance of poor writing habits*. Retrieved from
- Hyland, K. (1996). Talking to the academy: Forms of hedging in science research articles. *Written Communication*, *13*(2), 251.
- Hyland, K., & Tse, P. (2009). Academic Lexis and Disciplinary Practice: Corpus Evidence for Specificity. *International Journal of English Studies*, *9*(2), 111-129,168,170.
- Laurance, W. F., Useche, D. C., Laurance, S. G., & Bradshaw, C. J. A. (2013). Predicting Publication Success for Biologists. *BioScience*, *63*(10), 817-823.
- Meneghini, R., & Packer, A. L. (2007). *Is there science beyond English?* Retrieved from
- Monge-Nájera, J., & Nielsen, V. (2005). The countries and languages that dominate biological research at the beginning of the 21st century. *Revista de Biología Tropical*, *53*(1-2).
- Zamanian, M., & Heydari, P. (2012). Readability of Texts:: State of the Art. *Theory and Practice in Language Studies*, *2*(1), 42-53.