**SpellingBee: an entomological dictionary for word processor spellcheckers**

Adam Haberski1, Michael S. Caterino1

1 Department of Plant and Environmental Sciences, 277 Poole Agricultural Center, Clemson University, Clemson, SC 29634-0310, USA

**Abstract**

Word processor spellcheckers do not recognize many entomological terms and incorrectly flag them as spelling errors. These “false positives” waste time and can obscure actual errors. To alleviate this problem, we created SpellingBee, a dictionary of 15,224 entomological terms and taxon names that can be installed into common word processors to improve the spellchecking of entomological manuscripts. We tested its usefulness by spellchecking published literature across several entomology subdisciplines. ­­SpellingBee significantly reduced the number of false positive errors compared to the standard Microsoft Word dictionary. It was most effective for systematics literature (13% reduction) but less effective for medical entomology and pest management. Proper nouns and taxa below subfamily remain substantial sources of false positive errors. SpellingBee can be downloaded for free and will be updated periodically with more terms to improve its usefulness.

**Introduction**

Writing is a fundamental part of science, and publication quantity and quality are considered for job placement and performance evaluations (Brennan 2020). An important tool in the modern writing workflow is the Word processor spellcheck function. Spellcheckers flag words that are spelled incorrectly and suggest alternative spellings, however, they often fail to recognize entomological terms, resulting in large numbers of “false positive” flags. These are not only distracting but create a real burden, because each must be manually verified least they obscure a genuine error. Publishing typos is not only embarrassing, but in taxonomy, spelling counts and in rare cases typos could confuse nomenclature [ICZN art. 33].

This perceived bug is actually a feature. Spellcheckers work by comparing the words in a document against a reference dictionary and flagging words that do not match. Words omitted from the reference are flagged, regardless of spelling. Why then are standard dictionaries (those built into word-processing applications) not comprehensive? Word processors are designed to appeal to a large user base and including proper nouns, slang, and technical jargon can increase the occurrence of homophones (like *their* vs *there*) and real-word errors (*from* vs *form*) and decrease their efficacy for typical users.

Entomology is a jargon-rich field, and few entomological terms are included in standard spellcheck dictionaries, leading to a high rate of false positives during spellcheck. Missing words can be added to a list of known exceptions that will not be flagged in the future (referred to as a “custom dictionary” in most applications), but this must be done manually for each word by every user, which is inefficient. Additionally, an author could add a misspelled word to their custom dictionary perpetuating their mistake rather than fixing it.

The problem is pervasive enough that numerous commercial scientific spellchecking applications are for sale (e.g., Spellex, https://www.spellex.com; Inductel, https://inductel.com). These products claim to reduce false positives for scientific terms, but they can be expensive, and none are advertised as entomology specific.

Here, we introduce SpellingBee, a free spellcheck dictionary of entomological terms to improve the spellchecking of entomology manuscripts. It can be imported into most word processors to supplement standard dictionaries and reduce the occurrence of false positives. We evaluate its usefulness by testing it on existing literature.

**Methods**

We searched the literature for words that were absent from at least one of three common word processor standard dictionaries: Microsoft Word, Open Office, and Apple Pages. The largest source of terms was The Torre-Bueno Glossary of Entomology (Nichols 1989). We next searched the literature and bugguide.net (VanDyk 2022; https://bugguide.net) for the names of extant and extinct arthropod taxa down to the subfamily level. Below subfamily, the number of names increases exponentially, and are impractical to include. Latin binomina were only included for the most common model organisms.

We tested SpellingBee’s efficacy on articles from five journals published by the Entomological Society of America with open-source articles: Environmental Entomology, Insect Systematics and Diversity, Journal of Insect Science, Journal of Integrated Pest Management, and Journal of Medical Entomology. Combined, these journals encompass a broad scope of entomological subdisciplines. Our search was limited to “open-access” or “free” research articles, which yielded 11,183 results. Ten were chosen at random from each journal (*n* = 50) and the principal text (introduction, methods, results, conclusions, and captions) was copied and pasted into Microsoft Word v16.65 as plain text. We excluded reference sections because they contain numerous proper nouns and abbreviations that would inflate the number of false positives and slow down the test.

We then ran the spellcheck tool with the standard dictionary, and again with the standard dictionary plus SpellingBee to compare the number of errors detected. The proofing language was set to American English. We assumed that published manuscripts contain relatively few genuine spelling errors, and the majority of detections would be false positives. The difference in the mean number of reported errors was compared with a paired T-test. Flagged words were then extracted using a custom macro script and categorized by error type.

**Results/Discussion**

We compiled 15,020 entomological terms and taxon names absent from common spellcheck dictionaries. Our evaluation on published journal articles found SpellingBee significantly reduced the number of false positives detected during spellcheck (*t* = 3.77, *df* = 49, *p* = 0.0004; Table 1). It was most effective in the journal Insect Systematics and Diversity (13% reduction) and least effective in the Journal of Integrated Pest Management (4%). Through this exercise we identified an additional 204 terms and taxon names, which were added to SpellingBee to reach a final total of 15,224 terms. As expected, the number of genuine spelling errors averaged fewer than 1 per article.

**Table 1:** Comparison of errors reported during spellcheck of 10 articles each from 5 journals (*n* = 50) published by the Entomological Society of America, using SpellingBee versus the standard Microsoft Word dictionary. Table values are means ± standard errors.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Standard dictionary | | SpellingBee |
| All journals combined | | | 233.8 ± 36 | 211.5 ± 32 \* |
| Environmental Entomology | | 149.7 ± 42 | 141.0 ± 40 \* |
| Insect Systematics and Diversity | | 573.0 ± 21 | 500.3 ± 20 \* |
| Journal of Insect Science | | 127.4 ± 125 | 114.3 ± 111 |
| Journal of Integrated pest Management | | 174.3 ± 34 | 167.2 ± 33 \* |
| Journal of Medical Entomology | | 144.7 ± 22 | 134.8 ± 20 |
| \* indicates significantly different means at α = 0.05. | | | | |

The number of false positives remained high when using SpellingBee. The majority were taxonomic names below subfamily (Fig. 1). Resolving this problem would require adding all available taxonomic names, an un-ending process as new species are described each year. A comprehensive name checker for plants exists as a stand-alone web application, The Taxonomic Name Resolution Service (https://tnrs.biendata.org/). It attempts to match submitted names against a database of all published names to correct spelling errors (Boyle et al. 2013). A similar tool would be useful for entomology, but the number of described insect species is 2.5 times greater than plants (Stork 2018) and creating such a system is far more complicated than compiling a dictionary. Most manuscripts from our test contained relatively few names repeated many times. We therefore suggest authors carefully add names to their custom dictionaries at the outset of writing, after checking their spelling against a trusted source such as BugGuide or the Global Biodiversity Information Facility database (gbif.org).



**Figure 1**: Relative proportion of error types detected while using SpellingBee to spell check 10 articles each from 5 journals (*n* = 50) published by the Entomological Society of America. Errors were categorized as taxonomic names (below subfamily), other proper nouns, abbreviations, non-American English words, genuine spelling errors, or “other”.

Most non-taxonomic proper nouns reported as errors were author names from in-text citations, but also place names, businesses, and products. Author names are best managed through a citation manager, like the open-source Zotero (www.zotero.org), that automatically inserts accurate author names into text. Although this will not reduce false positives, it removes the burden of manually checking them.

SpellingBee can be downloaded for free through GitHub (https://github.com/ahaberski/SpellingBee.git). Installation instructions vary by application and operating system but are generally simple. Instructions for Microsoft Word can be found in the GitHub readme file. Instructions for other applications can be found in the applications’ “help” menus. Installing it will not overwrite users’ existing custom dictionaries.

SpellingBee is not comprehensive or infallible, so it will be updated periodically to increase its usefulness. Those who wish to contribute terms or corrections should contact the first author.

**Acknowledgements**

We thank Peter Adler for his feedback which improved this manuscript, and Curt Harden who contributed terms.

**References**

Brennan, J., 2020. Good Work If You Can Get It: How to Succeed in Academia. Johns Hopkins University Press. Baltimore, MD.

Boyle, B., Hopkins, N., Lu, Z., Garay, J. A. R., Mozzherin, D., Rees, T., Matasci, N., Narro, M. L., Piel, W. H., McKay, S. J., Lowry, S., Freeland, C., Peet, R. K., & Enquist, B. J. (2013). The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics, 14(1), 16. https://doi.org/10.1186/1471-2105-14-16

Nichols, S.W., 1989. The Torre-Bueno Glossary of Entomology. New York Entomological Society.

Stork, N.E., 2018. How many species of insects and other terrestrial arthropods are there on Earth? Annual Review of Entomology, 63(1), pp.31-45.

VanDyk, J., ed. 2021. BugGuide.Net: Identification, Images, & Information For Insects, Spiders & Their Kin For the United States & Canada. Iowa State University. Accessed January 30 2022. https://bugguide.net/.

**Appendix: Installation instructions for Microsoft Word**

For Windows:

1. Download SpellingBee.dic and save it to the desired folder.
2. Open the **Custom Dictionaries** dialog box: From the toolbar go to File > Options > Proofing.
3. Select **Add**.
4. Locate the folder containing SpellingBee and then double-click the dictionary file.

For MacOS:

1. Download SpellingBee.dic and save it to the desired folder.
2. Open the **Spelling & Grammar** dialog box: From the toolbar go to Word > Preferences > Spelling & Grammar.
3. Select **Dictionaries**.
4. Select **Add**.
5. Locate the folder containing SpellingBee, click the dictionary file and select **Open**.