

# **Statistics for Scientists:** harnessing textual profile tools

by John Blake and Xavier Blake

Background

Aim: provide statistical evidence on draft publications for non-native English speaking authors.

To enable authors to make decisions based on objective statistics, comparative data is provided on the vocabulary fit, readability and word type profile in relation to the target journal. Feedback on style & usage is provided in the form of frequency ratios. Lexicogrammatical errors are also identified.

## **Summary statistics**

Aim: provide a simple statistical overview of the generic integrity<sup>1</sup> and lexicogrammatical accuracy of

### **Vocabulary fit**

Peltokorpi Peltokorpi among team-based size data memory a often bow show and the based below the based bel

leader & Mavekawa

informal projects

<sup>\*</sup>people

social DOPPOS

teams develop organization

Aim: compare key vocabulary. Keyness is identified to assess the likelihood of paper-journal fit<sup>2</sup> based on vocabulary. Argote Organization WORK network role WORK Network role

team leaders

#### the draft publication.



<sup>1</sup>Bhatia, V. K. (1993). Analysing genre: Language use in professional settings. London: Longman.

## Readability

#### Aim: compare the readability in terms of word & sentence length.





<sup>2</sup> Hyland, K. (2011). Welcome to the Machine: Thoughts on writing for scholarly publication. Journal of Second Language Teaching and Research, 1 (1), 58-68.

# Word types

Aim: compare the ratio of academic words<sup>3</sup> and first **2000 words using a** vocabulary profiler.<sup>4</sup>



their draft to fit the style of the target publication.

|  |  | 4 |  |  |  |  |
|--|--|---|--|--|--|--|
|  |  | 5 |  |  |  |  |
|  |  |   |  |  |  |  |

See the following page for detailed guidance on how to make the best use of this analysis.

## **Style & usage**

Aim: show occurrence ratios calculated using a variety of corpora.

Errors of style or usage are manually identified and then occurrence ratios calculated using an appropriate corpus.

| severe disabilities . I              | first provide | <b>background</b> material on the p     |
|--------------------------------------|---------------|---|
| t nations , it needs to              | first provide | <b>basic</b> human rights for its citiz |
| net . In this article we             | first provide | demographic information as              |
| he <mark>intervening</mark> state to | first provide | evidence for its actions . The          |
| unity <mark>of scholars</mark> , we  | first provide | some historical background . <i>i</i>   |
| recorded messages to                 | first provide | <b>you</b> with the information you r   |

| Marked usage         | Ratio             | Suggestion                  |
|----------------------|-------------------|-----------------------------|
| People provide first | 0:9 <b>(COCA)</b> | People <b>first</b> provide |

<sup>3</sup> Coxhead, A. (2000). A new academic word list. TESOL Quarterly, 34, 213-238. <sup>4</sup> Cobb , T. (2013). *Web Vocabprofile* www.lextutor.ca/vp/

## Lexicogrammatical errors

### Aim: understand and correct the error.

Errors are manually identified, corrections suggested and comments given to enable authors to understand the underlying rules.

a former employee to participate to in this study for to being responsible



