

**College University Educators ESP Symposium**  
**7 September 2013**  
**Japan Advanced Institute of Science and Technology**

**Statistics for scientists:**  
**Incorporating data-driven decision making in the publishing process**

**Abstract**

Feedback on draft publications written by non-native English speakers tends to be rather subjective. The impetus for this project stemmed from the desire to enable writers to make decisions based on objective statistical evidence rather than subjective advice. The aim is for the writers to better understand which aspects of their articles in terms of genre, grammar and lexical choice need to be addressed. A small corpus of research articles of the target publication is compiled for each writer. The generic integrity and lexical repertoire of drafts received are then analyzed with respect to the specially-created target corpus. Various concordancing tools are harnessed to create comparative data on keyness, readability and word types. In addition, standard lexicogrammatical errors and marked usage are identified manually. A large scientific corpus is used to identify the statistical probability of the occurrence of any marked forms. Each writer is given a statistical summary of their draft manuscript and an explanatory guidance sheet showing how to interpret the statistics and providing general advice on how to improve the generic integrity. Specific advice is given on how to address lexical or grammatical errors.