

# **JALT CUE 3<sup>rd</sup> Annual ESP Symposium – Kanto 2014**

**13 September, 2014.**

**Waseda University, Tokyo.**

## **Harnessing keyness: Corpus-based approach to ESP material development**

**John Blake**

### **Summary**

Concordancers often provide an option to generate lists of keywords. Novice users tend to expect that the keyword lists produced are identical, yet there are significant differences in the lists generated. This presentation aims to show how the generation of keywords may be skewed by the choice of reference corpus and statistical formula. This knowledge enables ESP materials developers to fine-tune keyword lists to enable them to select the most appropriate vocabulary.

### **Extended abstract**

Words that commonly occur together are known as collocates; words that commonly occur in a particular type of text are called textual collocates or, more recently, keywords. The most frequent words in a text are easy to identify, but are not necessarily the most useful words to highlight in ESP materials. This is because grammatical words and high frequency general service words are likely to occupy the top positions. Words that are key, however, are likely to merit inclusion. Simply put, keyness is a measure of the frequency with which a word occurs disproportionately in a particular text type. Keyness is assessed by comparing the relative frequency of a word in one corpus to a reference corpus using a statistical formula. This way of assessing textual collocation can be harnessed by ESP material developers when evaluating which vocabulary items to focus on. This presentation aims to show how the choice of both reference corpus and statistical formula generates different lists of keywords. Materials developers can use this knowledge to make more informed choices of which vocabulary to focus on in their tailor-made materials.