

Developing a Corpus Concordancer as a Learning Tool for EST Students

Shoichi Tsuchida
Shi Laboratory
Department of Communication Engineering and Informatics
The National University of Electro-Communications (UEC Tokyo)

Jie Shi
Graduate School of Informatics and Engineering
The National University of Electro-Communications (UEC Tokyo)

Abstract

Corpus tools are essentially tools for linguistic research or treating texts for linguistic analysis, e.g. lexical features, instead of being educational or learning tools for ESP practitioners and students. EFL learners or users who must handle high level academic written or spoken communication in English independently or with little professional assistance tend not to choose a corpus tool as their first choice to solve language or linguistic problems. This research demonstrates the process of developing a new concordancer that is not only an analytical tool but also a learning tool especially for EFL learners to understand professional and genre-based English. Specific stages of the development of the new concordancer will be reported, i.e. the choice of a computer language, selection of functions, interface design, programming, database preparation, user-friendliness features, testing, and piloting. In addition, the result of a survey to find out the preferences of concordancer functions of graduate students at a Japanese university of science and engineering will be presented.

Introduction

Definition of Corpus by Cheng (2012):

A corpus is a collection of texts that has been compiled for a particular reason. In other words, a corpus is not a collection of texts regardless of the types of texts collected or, if a variety of text types (i.e., genres) are in the corpus, the relative weightings assigned to each text type. A corpus, then, is a collection of texts based on a set of design criteria, one of which is that the corpus aims to be representative. (p. 3)

Classification of Corpora

- Specialised corpus
- General corpus
- Comparable corpora
- Parallel corpora
- Learner corpus
- Pedagogic corpus
- Historical or diachronic corpus
- Monitor corpus

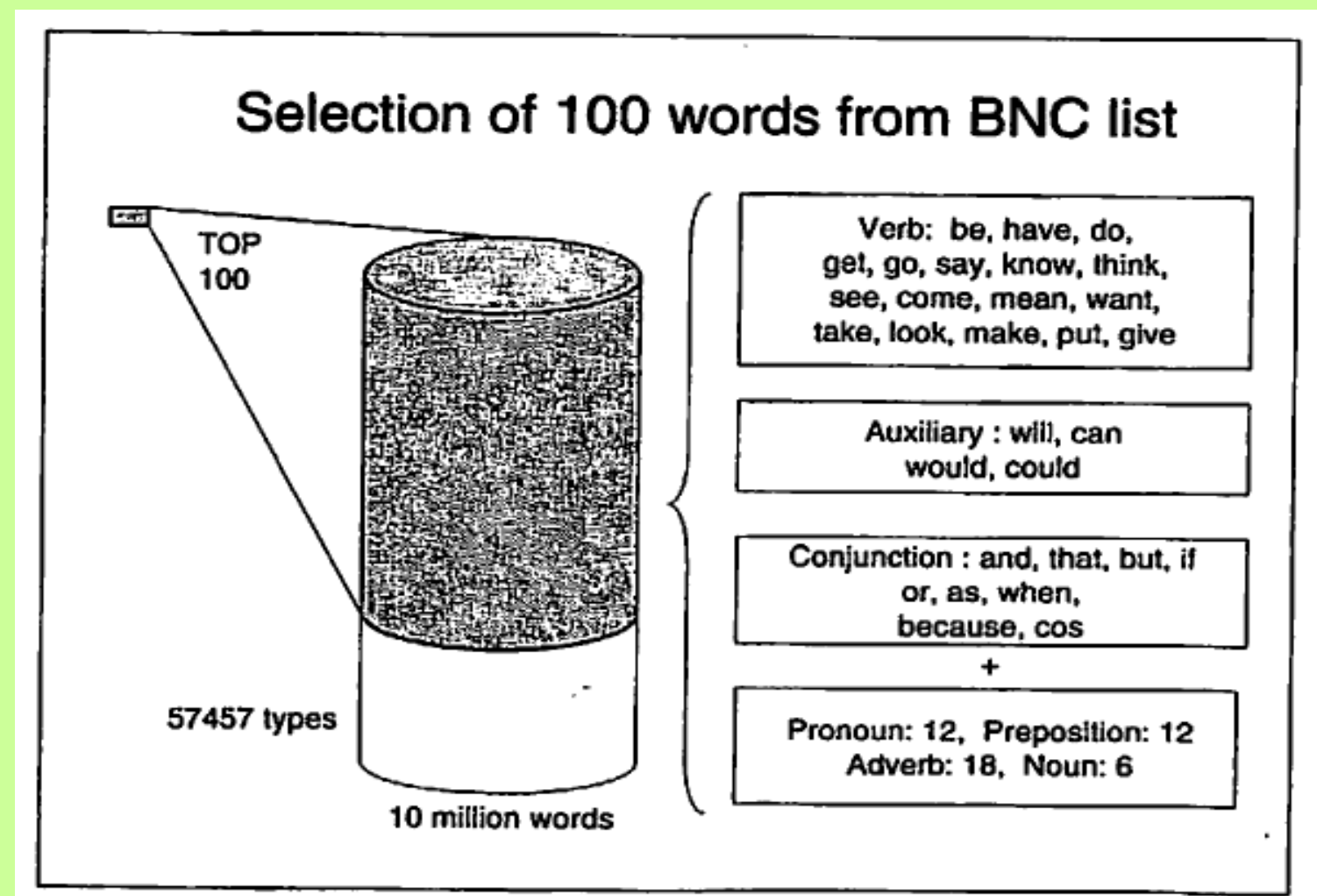


Figure 1.1 Selection of 100 words from the BNC-spoken list (Tono, 2011c, p. 7)

Research Purpose

To develop a corpus linguistic tool which is user-friendly for researchers, teachers and learners especially of EST.

Methodology & Design

A. Programming Language: Python

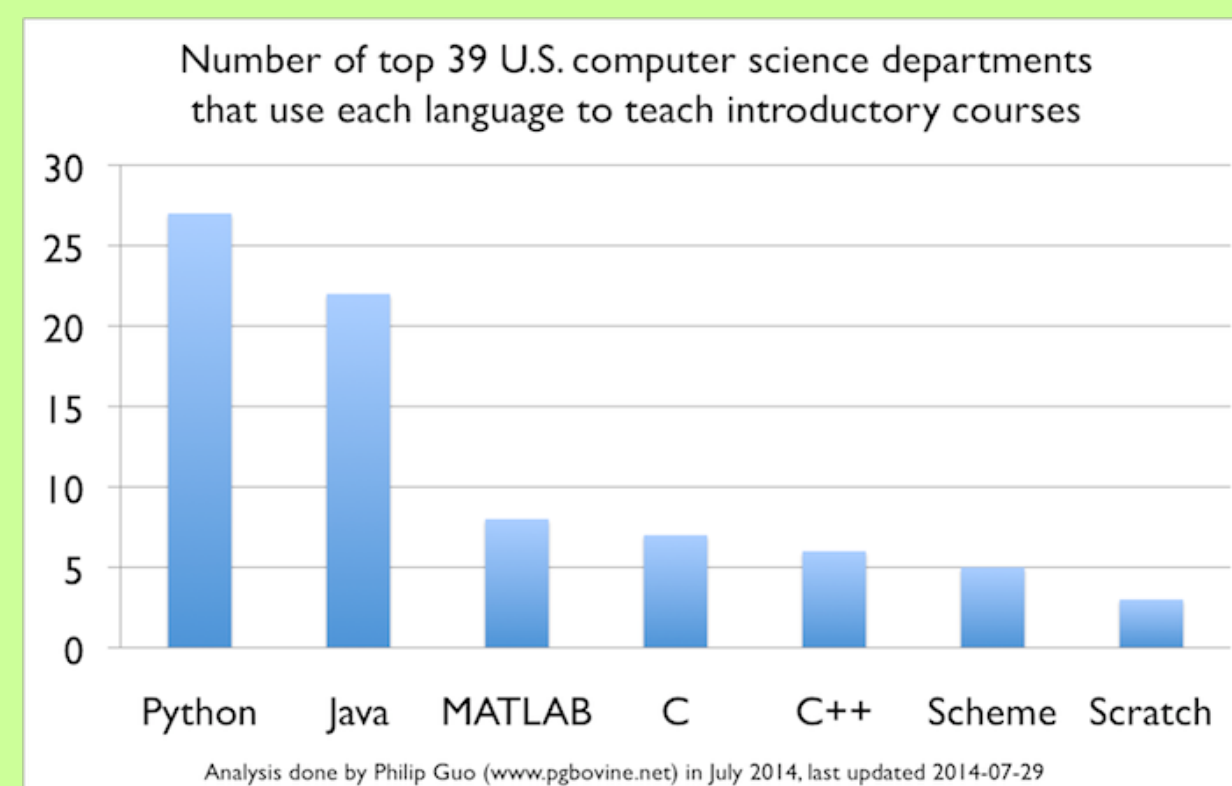


Figure 4.1 Number of top 39 U.S. computer science departments that use each language to teach introductory courses

B. Aiming at User-Friendliness

- GUI = Interface
- Levels or Fields
- Display of Functions
- Bilingual instructions
- Participatory features

C. Selection of Functions

- KWIC (Key Word In Context)
- Word List
- N-gram
- Cluster
- Collocate
- File View
- Charts
- Exercises
- Participatory

Results

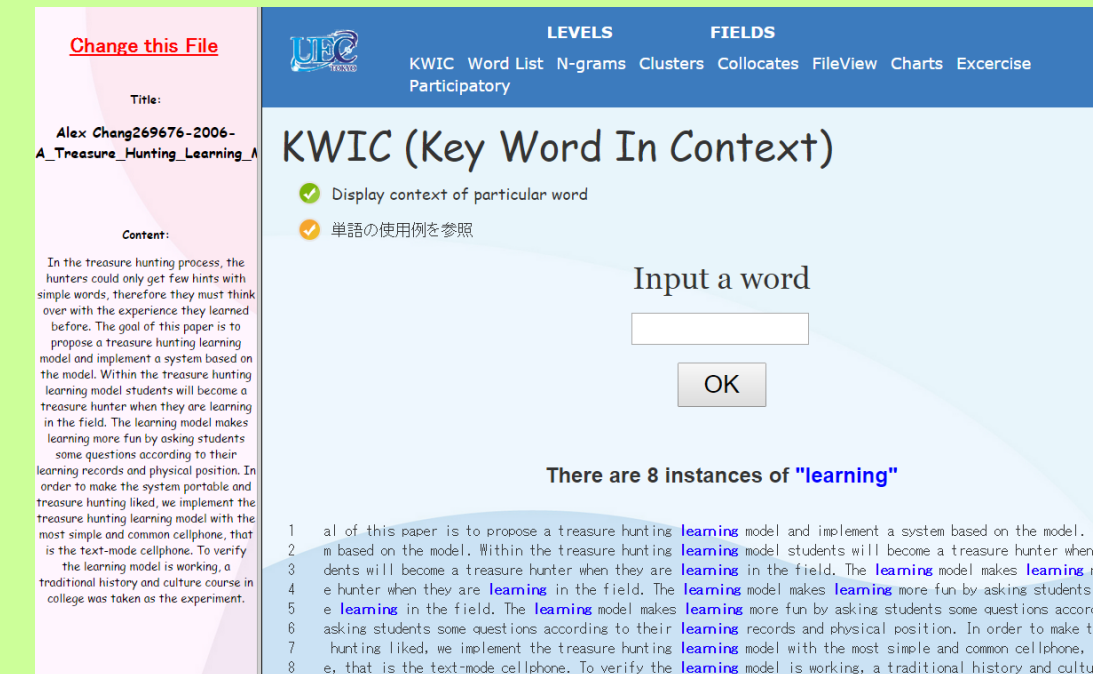
Top Menu



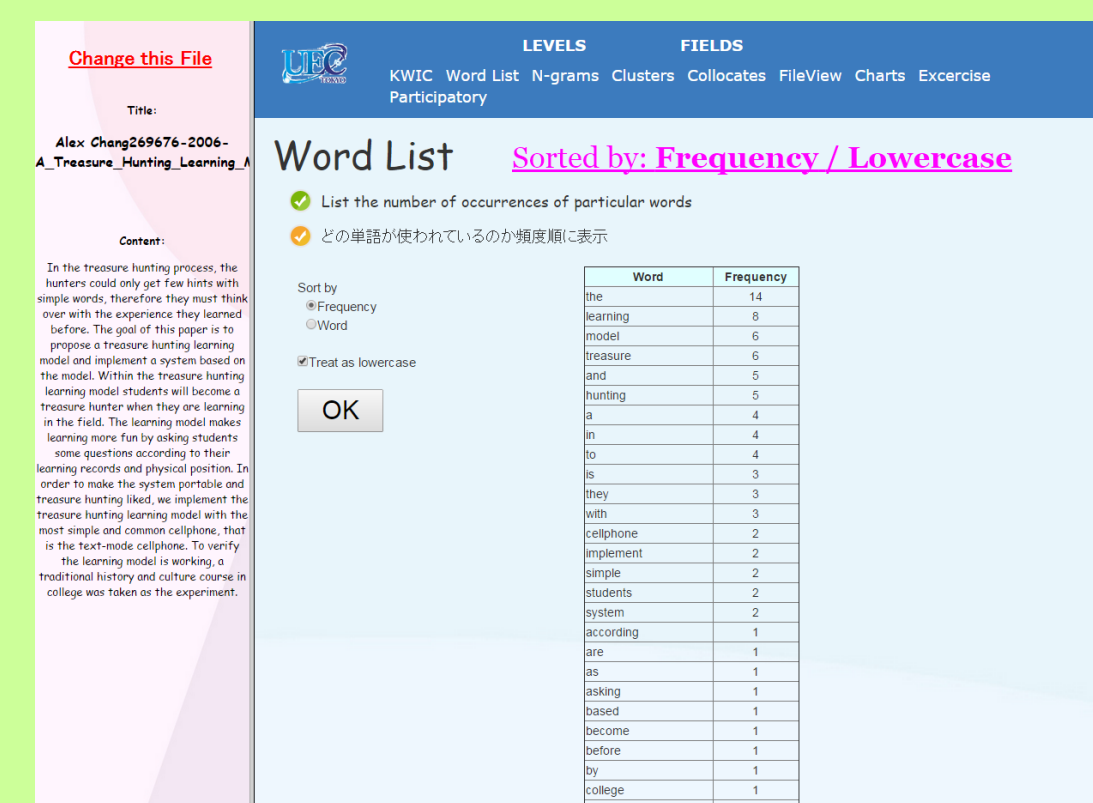
Select a File



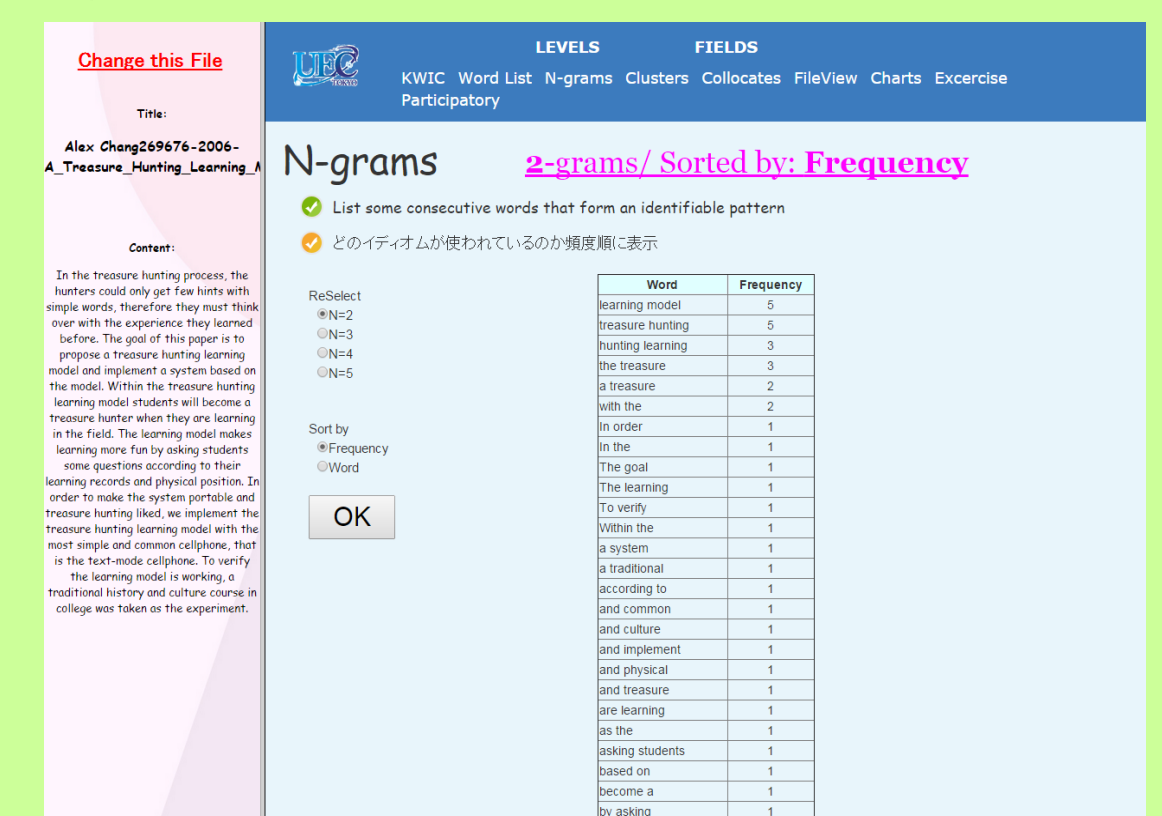
KWIC



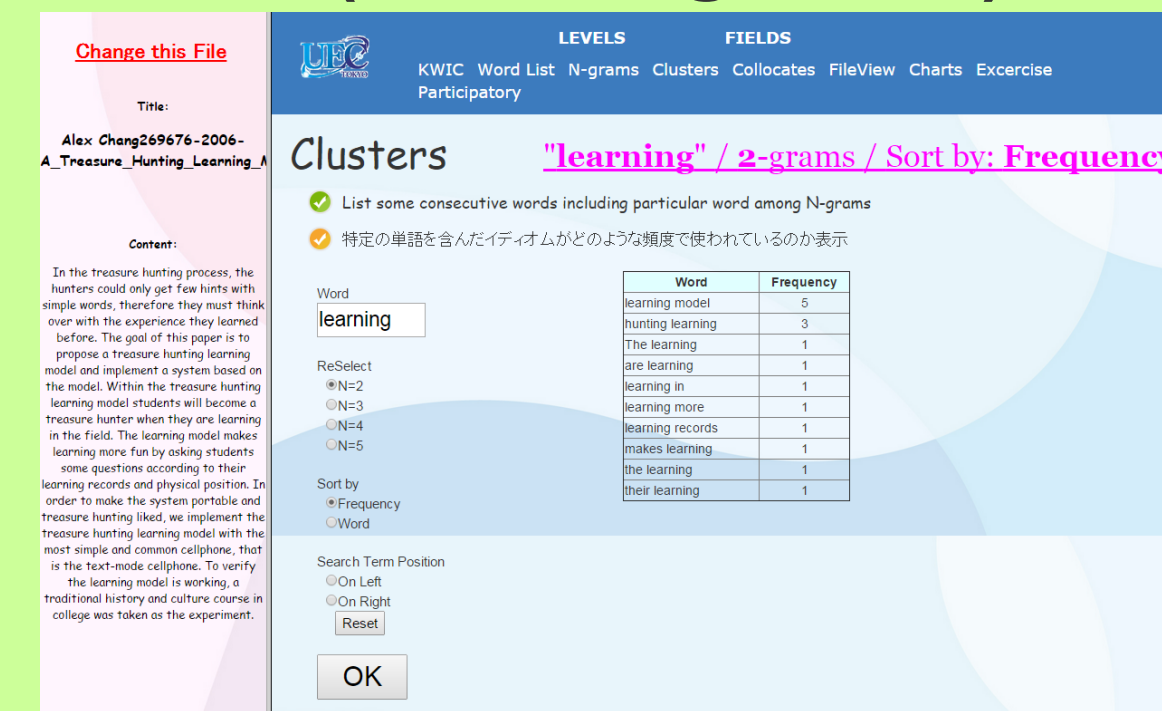
Word List



N-grams (N=2)



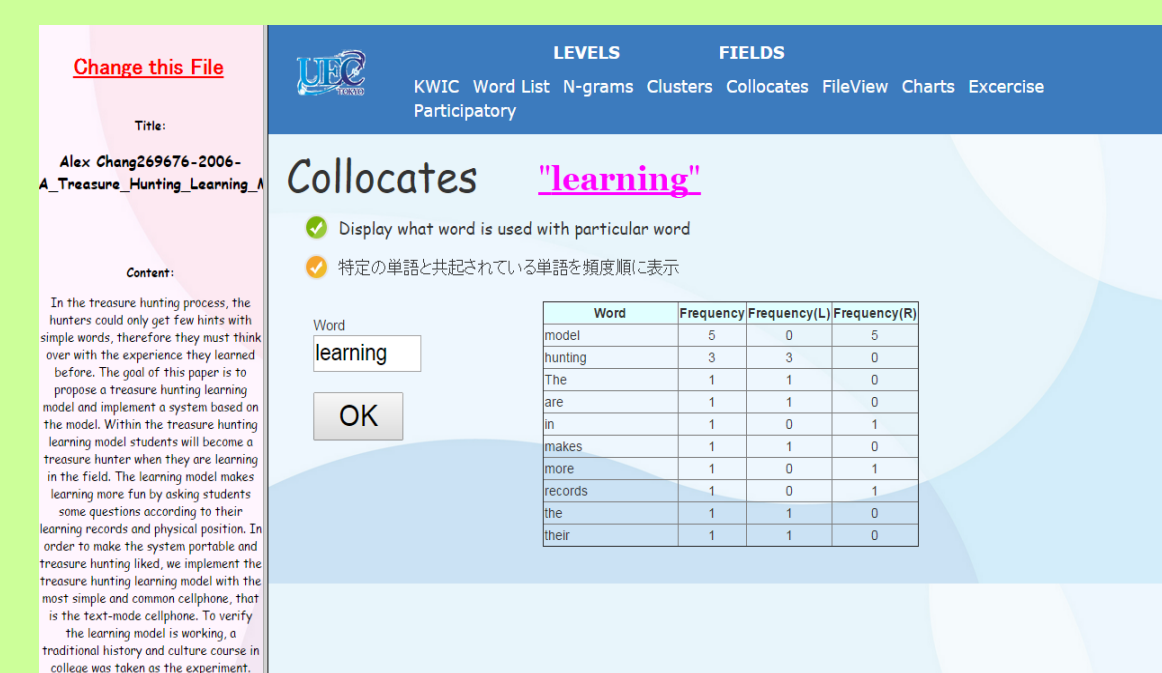
Clusters ("learning", N=2)



File View



Collocates



Conclusion

- Have developed a concordancer for students or researchers of EST
- Python: A suitable language for interface and low-maintenance
- Positive feedback from graduate students
- 6 functions completed
- Web-browser style of interface

References

- Cheng, W. (2012). *Exploring Corpus Linguistics*. New York, United States of America: Routledge.
- Guo, P. (2014, 7 7). *Python is Now the Most Popular Introductory Teaching Language at Top U.S. Universities*. Retrieved from Communications of the ACM: <http://cacm.acm.org/blogs/blog-cacm/176450-python-is-now-the-most-popular-introductory-teaching-language-at-top-us-universities/fulltext>
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge, United Kingdom: Cambridge University Press.
- McCarthy, M., & O'Keeffe, A. (2010). *The Routledge Handbook of Corpus Linguistics*. Routledge.
- Nesi, H. (2013). ESP and Corpus Studies. In B. Paltridge, & S. Starfield, *The Handbook of English for Specific Purposes* (p. 407). Wiley-Blackwell.
- Tono, Y. (2011). TaLc in action: recent innovations in corpus-based English language teaching in Japan. In A. G. Frankenberg, G. Aston, & L. Flowerdew, *New Trends in Corpora and Language Learning* (pp. 7-8). Bloomsbury.