Title	**A corpus-based study of vocabulary distribution in the finance register**
Author(s)	**Ha, YH**
Citation	**The 2013 International Conference on Register Revisited: New Perspectives on Functional Text Variety in English, Vechta, Germany, 27-29 June 2013.**
Issued Date	**2013**
URL	**http://hdl.handle.net/10722/184000**
Rights	**Creative Commons: Attribution 3.0 Hong Kong License**

A corpus-based study of vocabulary distribution in the finance register

Ying Ho Ha (The University of Hong Kong)
althea.ha@hku.hk

As far as is known to date, no attempts have been made to comprehensively describe the vocabulary in the finance register in terms of its lexical density, range, and frequency by financial sectors, text types, modes of communication, and geographical locations around the world.

This corpus-based study aims at investigating vocabulary distribution in the finance register. Authentic texts of the finance register from 20 companies in sectors of banks, financial services, insurance, and real estate (Industry Classification Benchmark, 2011) were collected to build a 2,030,324-token corpus. Text samples are from five geographical locations which are English-speaking developed financial markets. The majority of the data are texts or transcripts produced between 2010 and 2011. Three of the text types are in written mode, namely annual reports, news releases, and product and service descriptions. The other four are in spoken mode, namely earnings calls, interviews, speeches, and presentations plus a question-and-answer session. The average numbers of tokens per file vary enormously among the seven text types. Texts in the spoken mode cover 16.43% of the whole corpus.

Results showed that the combined General Service List of English Words (West, 1953) and Academic Word List (Coxhead, 2000) covers from 83% to 85% for the three text types in written mode and over 90% for the three text types except speeches in the spoken mode. The most striking and expected difference between written and spoken mode lies in lexical density with a gap of almost 10%. The spontaneity of speeches may contribute to the differences in coverage between text types in spoken mode. Results for scripted speech are between those for the written mode and those for the other text types in spoken mode. Lexicalisations of different types can be found in the finance register and the most prominent ones are affixation and compounding from the corpus evidence.

References

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213-238.

Industry Classification Benchmark. (2011). *Product Specification ICB Universe Data Services.* Retrieved from http://www.icbenchmark.com/ICBDocs/ICB%20Product%20Spec%20-%20Oct2011.pdf

West, M. (1953). *A general service list of English words*. London: Longman, Green, & Company.