Title: Facilitating academic words learning: a data-driven approach using a collocation consultation system built from open access research papers

Keywords: academic words, collocation learning, data-driven learning, corpus consultation

Abstract: It is essential and beneficial for ESP students to master collocations of a set of core academic words. Corpus analysis tools (e.g. concordancers) have been widely used in facilitating collocation learning, and with promising results having been demonstrated in the literature. This paper presents a learner friendly collocation consultation system built from 50,000 open access research papers made available by CORE (https://core.ac.uk/). The research papers are grouped into four disciplines: Arts and Humanities, Physical Sciences, Life Sciences and Social Sciences. From these articles, useful syntactic-based word combinations (e.g., verb+noun, noun+noun, adjective+noun) are extracted, organized by syntactic patterns, sorted by frequency, and linked to their context sentences. Learners can look up the collocations and look up the usage of an academic word in any of these four disciplines by simply entering the word or selecting it from one of pre-compiled academic word lists. The paper will also show how the system was used in an initial study carried out with 15 international students at the Computer Science Department at University of Waikato, New Zealand.