A Systematic Review of Using Discipline-Specific Corpora for Lexico-Grammatical Pattern Learning: A Case Study for Computer Science Postgraduates

Shaoqun Wu, Faculty of Computing and Mathematical Sciences, University of Waikato, Hamilton, New Zealand
Liang Li, Te Hononga School of Curriculum and Pedagogy, University of Waikato, Hamilton, New Zealand
Ian Witten, Faculty of Computing and Mathematical Sciences, University of Waikato, Hamilton, New Zealand
Alex Yu, Centre for Business, Information Technology and Enterprise, Waikato Institute of Technology, Hamilton, New Zealand

ABSTRACT

This article reports on a language learning system and a program designed to help students with academic vocabulary in the New Zealand university computer science department. The system is a learner-friendly corpus-based tool that allows students to look up lexico-grammatical patterns of a given word in academic writing. The program, based on a data-driven learning approach, comprises tutorials, workshops, and follow-up exercises that help students learn useful formulaic patterns of academic words that are typical in computer science. The authors’ results capture students’ awareness of language patterns in academic text and their growing confidence in using academic words with the assistance of the tool. Not surprisingly, interpreting and transferring the corpus data into students’ own writing requires training and practice. The effectiveness and limitations of the resources and tools used in this learning program are examined, and suggestions are made for further improvement and future research.

KEYWORDS

Academic Vocabulary Learning, Corpus-Based Language Learning, Data-Driven Learning, Lexico-Grammatical Patterns

INTRODUCTION

New Zealand universities attract increasing numbers of international students into postgraduate study. Despite having fulfilled an entrance language requirement (e.g. IELTS 6.5), most students still face the challenges of producing written assignments at successively higher levels—an essential prerequisite for academic excellence. Universities can only provide limited support to help students upgrade their writing skills, and when they do, the resources tend to be generic and unrelated to the domain of students (Hyland, 2006). Many postgraduate students appear to struggle to produce written assignments that are satisfactory in terms of language proficiency and they tend to rely on their supervisors to reformulate their writing, which places an unwelcome burden on supervisors. This presents a real dilemma for both students and supervisors.

DOI: 10.4018/IJCALLT.2018010103

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
This paper reports on a learner-friendly corpus-based tool called FLAX (http://flax.nzdl.org) and a language program designed to help international postgraduate students with their academic writing by introducing lexico-grammatical patterns in academic prose. We focused on students from China who were studying computer science, although the methodology is applicable to other linguistic backgrounds and subject domains. We adopted a data-driven learning (DDL) approach advocated by (Johns, 1991), which allows students to explore lexico-grammatical patterns from authentic academic texts and has been suggested to be particularly beneficial for advanced L2 learners (Boulton & Thomas, 2012; Boulton & Pérez-Paredes, 2014; Chang, 2014; Cobb & Boulton, 2015; Vyatkina, 2016).

The language program comprises tutorials, workshops and exercises. First, the tutorials covered topics such as common word-choice errors in student writing; types of academic words (e.g. reporting verbs and shell nouns), including their functions and language patterns in writing; and different ways of starting sentences. Second, the workshops introduced the FLAX tool and how to use it to identify relevant lexico-grammatical patterns related to a particular discipline—in our case, computer science (examples are highlighted the importance of the word highlight, to solve the problem of the the word problem, and it is possible to reproduce for the word possible). Third, the follow-up exercises reinforced what had been learned in the tutorials and workshops.

In the program, students used FLAX, a free self-access language learning tool, documented in Wu, Franken and Witten (2010), and Wu and Witten (2016). FLAX automatically generates lexico-grammatical patterns, collocations, and lexical bundles from user supplied corpora, and publishes them online. It provides students with easily accessed language patterns in a simple interface, which plays a key role in the success of DDL (Łefko-Szymańska & Boulton, 2015). These features, best suited to our research purposes, distinguishing FLAX from other corpus-based tools, such as COCA, Compleat Lexical Tutor and SKELL, will be introduced in the following sections.

THE CORPORA

Three corpora were built into FLAX to support the language program. The first corpus contained twenty-five conference proceedings and journal articles, written by professors in our department that covered a wide range of topics—such as Human Computer Interaction, Machine Learning and Digital Libraries. These articles, particular to computer science as taught in our university, were chosen by the lecturer who teaches a writing for research course for international students who are likely to go on to undertake Master’s or Doctoral research in certain specialist areas. Though small, this corpus provided sample texts from which academic words and lexico-grammatical patterns salient to computer science could be derived.

The second was the much larger British Academic Written English Corpus (BAWE), which contains 2860 student assignments that scored high grades (Nesi & Gardner, 2012). This corpus was relevant to our students not only because it is academic but also because it comprises student work. It is divided into four areas: Arts and Humanities, Social Sciences, Life Sciences and Physical Sciences.

The third corpus was a learner corpus, comprising 32 literature summaries, with an average length of 200 to 500 words, written by Chinese students in the previous semester. We did not use writing samples from current students because they were not available at the beginning of the program. This learner corpus served as a complement to present learner language use (L. Flowerdew, 2001; Granger, 2009)—in our case, to identify the academic words with which the students have difficulty in the use of lexico-grammatical patterns. Though small, it probably exhibits similar language features to our current students’ writing because they all had the same educational background.

LEXICO-GRAMMATICAL PATTERNS

Language is formulaic. Biber et al. (1999), and Erman and Warren (2000) found that prefabricated language accounts for 21–52% of written text. Nation (2013) argues formulaic language, especially
Constructing a Data-Driven Learning Tool with Recycled Learner Data

The Perspectives of Iranian EFL Teachers on the Integration of ICT into their Teaching Practices
[www.igi-global.com/chapter/the-perspectives-of-iranian-efl-teachers-on-the-integration-of-ict-into-their-teaching-practices/134611?camid=4v1a](www.igi-global.com/chapter/the-perspectives-of-iranian-efl-teachers-on-the-integration-of-ict-into-their-teaching-practices/134611?camid=4v1a)