Developing and Evaluating a Learner-Friendly Collocation System With User Query Data

Shaoqun Wu, University of Waikato, Hamilton, New Zealand
Alannah Fitzgerald, University of Waikato, Hamilton, New Zealand
Alex Yu, Centre for Business, Information Technology, and Enterprise, Wintec, Hamilton, New Zealand
Ian Witten, University of Waikato, Hamilton, New Zealand

ABSTRACT

Learning collocations is one of the most challenging aspects of language learning as there are literally hundreds of thousands of possibilities for combining words. Corpus consultation with concordancers has been recognized in the literature as an established way for language learners to study and explore collocations at their own pace and in their own time although not without technological and sometimes cost barriers. This paper describes the development and evaluation of a learner-friendly collocation consultation system called FlaxLC in a design departure away from the traditional concordancer interface. Two evaluation studies were conducted to assess the learner-friendliness of the system: a face-to-face user study to find out how international students in a New Zealand university used the system to collect collocations of their own interest and a user query analysis—based on an observable artefact of how online learners actually used the system over the course of one year—to examine how the system is used in real life to search and retrieve collocations.

KEYWORDS

Collocation Database, Collocation Learning, Corpus-Based Language Learning, Data-Driven Learning, User Query Data Analysis

INTRODUCTION

Collocations, recurrent word combinations, have been widely recognized as an important aspect of vocabulary knowledge (Firth, 1957; Lewis 2008; Nattinger & DeCarrico, 1992; Sinclair, 1991, Nation, 2013). Dictionaries and most recently, corpus analysis tools (i.e. concordancers and the like) are the two main resources that learners draw on while acquiring such knowledge. Printed dictionaries are the popular and traditional collocation learning resource—for example, the BBI Combinatory Dictionary of English (Benson et al., 1984), the Dictionary of Selected Collocations (Hill and Lewis 1997) and the Oxford Collocation Dictionary (2009)—dedicated to assist learners in mastering this essential knowledge. With corpus analysis tools, learners can enter a word and explore what words are most likely to occur before or after it. These tools, whether web-based (e.g. the Collins COBUILD Corpus, WebCorp, WebCollocate, Mark Davies’ Brigham Young Corpora, COCA) or stand-alone (e.g., WordSmith Tools, AntConC) are specifically designed for linguists, and come with different interfaces, search functions and the presentation of results that are mostly in the form of keyword-
in-context (KWIC) fragments and incomplete sentences. In terms of retrieving collocations, they facilitate the search for two- or three-word collocations.

Corpus-based tools have been explored by many researchers and teachers to facilitate collocation learning with promising results as demonstrated in the literature (Boulton & Cobb, 2017). They have been used in helping students find correct word combinations (e.g. “cause a problem” vs. “bring a problem”) (Yoon, 2008; Chen, 2011; Daskalovska, 2015; Vyatkina, 2016), understand the subtle meaning of certain verbs that lack direct L1 equivalents: synonyms (e.g. construct, build, and establish), hypernyms (e.g. create and compose) (Chan and Liou, 2005), and identify common word choice errors in student writing (Chambers and O’Sullivan, 2004; Wu, et al. 2009). Johns (1991) used the term “data-driven learning” (DDL) to describe this approach that centers on fostering learners’ skills in becoming a “language researcher”. Despite DDL’s great potential as presented in the literature, DDL has not been widely accepted by mainstream language educators (Leńko-Szymańska & Boulton, 2015, p. 3). Technical challenges that face both teachers and students go some way to explain the reluctance in implementing DDL in classroom. This view is supported by the results of a large-scale survey conducted by Tribble on using corpora in language teaching (Tribble, 2015).

User-friendliness and free access are reported to be two major factors in influencing the willingness of respondents to use corpora, while “don’t know how to”, “are not familiar with” are among the reasons for not using corpora. DDL researchers have also reported several factors that may hinder corpus use, including requirements of metalinguistic knowledge (e.g., part-of-speech tags) to formulate queries, unfamiliarity with complex search interfaces and functions, overwhelming results, and difficulties in locating and interpreting target language features in concordances (Yoon & Hirvela, 2004; O’Sullivan & Chambers, 2006; Yeh, Li, & Liou, 2007; Chen, 2011; Rodgers et al., 2011; Boulton, 2012a; Chang, 2014; Geluso & Yamaguchi, 2014; Daskalovska, 2015). For example, Chang (2014) asserts that the differing interfaces and functions of various corpus tools further increases the technical challenge whereby learners generally need to learn a new system in order to access a different corpus.

To make corpus tools accessible for language learners, Boulton (2012b) published a call to “simplify the technology as much as possible either the corpus itself or the associated software”. DDL researchers have also proposed improvement recommendations in corpus software development, for example, “simplified”, “controlled” or “screened” concordancer output before being presented to students (Varley, 2009; Daskalovska, 2015), and simple interfaces to return multi-word units, for example, make a presentation or give a presentation for the search query word presentation (Chan & Liou, 2005). Chen (2011) evaluated three corpus tools—the Hong Kong Polytechnic Web Concordancer, the COBUILD Concordancer, and BNC Sample Search tool—and proposed an ideal collocation retrieval tool, one that needs to include: the position of the search term (before or after the query word), functions for phrase or multiple word searches, support for using L1 in queries, syntactic information about the collocates (verb, noun, adjective, adverb), retrieval of all relevant collocations at once, and the clustering of semantically related collocations. Over the past years, it has been encouraging to see the development of dedicated tools for collocation learning (e.g. SKELL) with some of the aforementioned proposed functionalities included.

This paper describes our attempts at building and evaluating a collocation retrieval system called FlaxLC and answers the following research questions:

1. How can we design and develop a learner-friendly collocation retrieval tool?
2. How do learners use FlaxLC to look up collocations?

To answer the first research question, we proposed two design principles that aim to minimize learning and training efforts for using a corpus-based tool. The first principle is to capitalize on learners’ familiarity with traditional language resources (i.e. dictionaries); the second is to utilize learners’ existing web skills with search engines. To answer the second research question, we conducted two
Students' Perceptions About the Impact of Using Padlet on Class Engagement: An Exploratory Case Study
[www.igi-global.com/article/students-perceptions-about-the-impact-of-using-padlet-on-class-engagement/238875?camid=4v1a](www.igi-global.com/article/students-perceptions-about-the-impact-of-using-padlet-on-class-engagement/238875?camid=4v1a)

Using Blogs to Motivate English as a Foreign Language Writers
[www.igi-global.com/chapter/using-blogs-to-motivate-english-as-a-foreign-language-writers/242432?camid=4v1a](www.igi-global.com/chapter/using-blogs-to-motivate-english-as-a-foreign-language-writers/242432?camid=4v1a)