Exploring Collocations with The Prime Machine

Stephen Jeaco, Xi’an Jiaotong-Liverpool University, Suzhou, China

ABSTRACT

One of the greatest impacts of corpus linguistics on language teaching has been in the recognition of the importance of collocation. A very influential guide for language teachers with regard to teaching collocation has been the Lexical Approach. Activities pointing students to rich collocational information in monolingual dictionaries, in texts and specifically in collocation dictionaries provided ways for language learners to engage with collocation information: to notice, to remember and to acquire. In recent years, there has been a growing interest in Data Driven Learning and new tools are now available to allow students to access collocation information from corpora for themselves. After introducing some pedagogic considerations, this article presents some of the features of The Prime Machine which were developed to support DDL activities focussed on collocation.

KEYWORDS

Collocation, Concordance Lines, Data Driven Learning, Language Learning

INTRODUCTION

One of the greatest impacts of corpus linguistics on language teaching has been in the recognition of the importance of collocation for effective language learning. As computer searchable databases of authentic language texts became available to linguists and lexicographers, evidence for the patterning and co-selection of word choices required greater attention (Barnbrook, Mason, & Krishnamurthy, 2013; Sinclair, 1991). A very influential guide for language teachers with regard to collocation has been the Lexical Approach (Lewis, 2008) and its activities pointing students to the rich collocational information in monolingual dictionaries, in the examples in their texts and specifically in collocation dictionaries provided ways for language learners to engage with collocation information: to notice, to remember and to acquire (Hill, Lewis, & Lewis, 2000).

While mobile phone technology and the internet have brought increased ease for checking and finding simple meanings of words, some of the pedagogic basis for spending time, thought and energy on retrieving, digesting and recalling this information has perhaps been lost. While web corpora have grown in size and scope, some of the opportunities for exploring collocations in specific contexts, perhaps for specific fields, have also to some extent been overlooked. However, one use of technology in the language learning classroom in particular – Data Driven Learning (DDL) – can offer opportunities for learners to engage directly with language data for a range of language learning pursuits (Flowerdew, 2015; Thomas, 2015; Tsui, 2004).

This paper presents an overview of some of the features of the newest version of The Prime Machine which have been developed to support the exploration of collocations as part of Data Driven Learning activities. Inspired by the theory of Lexical Priming (Hoey, 2005), The Prime Machine was developed as a learner-friendly corpus tool (Jeaco, 2015; Jeaco, 2017a). After introducing the

DOI: 10.4018/IJCALLT.2019070103

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
pedagogical and theoretical background of collocation and DDL, this paper goes on to introduce the way collocation is measured and presented in several corpus tools, before describing the ways users of The Prime Machine can view and interact with collocation data.

Defining Collocation

From a language teaching perspective, definitions of collocation can be fairly broad with the aim being to make learners more aware generally of the patterns of words around them. Developers of the Lexical Approach have introduced the concept to learners by describing the variation in the strength of relationships between words as being similar to the variations found in relationships between people (Hill, et al., 2000). Other suggestions include comparing composition to the use of a model airplane kit or drawing on expectations of their mother tongue through reflection or translation of technical phrases (Conzett, 2000; Hill, et al., 2000). With all of these explanations, the aim is to help learners start noticing collocation in the language they encounter and to encourage them to invest time in this enterprise because of the practical advantages collocations can offer. It seems that many learners find it unusual at first to examine language in front of them in units beyond individual words. Indeed, Conzett (2000) claims that explicitly making students aware of the term “collocation” will speed up class activities based around collocation. Woolard argues that definitions based on statistical information do not “guide my students’ attention to specific elements of text in a clear and directed way”; saying that for the purposes of teaching a definition focussing on expectation is more useful (Woolard, 2000, p. 29) and he also restricts the patterns covered by the term to specific combinations of word classes.

While the introduction of the term “collocation” in the classroom may be made through metaphor, reflection on translations in the mother tongue, practical applications or repeated methods of annotation, more formal definitions of collocation in linguistic theory and computational linguistics have developed over the years. The differences in the qualification and scope of collocations which mirror the different purposes that different teachers have in mind are part of a general tendency for different researchers to specify the meaning of collocation and similar phenomena in different ways. Linguistic descriptions include a wide variety of ways of limiting what should count and how it should be understood to operate, including strings of characters in raw text (Sinclair, 1991), lexical phrases (Nattinger & DeCarrico, 1992), lexical bundles (Biber, Johansson, Leech, Conrad, & Finegan, 1999), motivated or unmotivated collocations (Hunston, 2002), lexical networks across sections of a book (Phillips, 1985), within Pattern Grammar (Hunston & Francis, 2000), and as a major contribution to the identification of norms (Hanks, 2013). Each of these stipulate such aspects as whether collocation-like phenomena operate on word forms or lemma, only in specific grammatical relations or freely, and for any kind of word or only certain parts of speech. According to Hoey (2005), at times the very definition of collocation has been tied up with the methodological approach for their retrieval. He provides a definition which fuses the psychological importance of collocation with the means of detection and evaluation:

So our definition of collocation is that it is a psychological association between words (rather than lemmas) up to four words apart and is evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution (Hoey, 2005, p. 5).

Hoey introduces collocation using the example inevitable + consequence (2005, p. 2). As a piece of software purposefully designed to support the examination of the kinds of relationship between words which are introduced in Hoey’s theory of Lexical Priming, collocations are defined for The Prime Machine based on his 2005 definition. In this paper, collocation will be used as it is in the software to refer to combinations of two, three, four or five words in a four-word window either side of a node. The term “multi-word unit” will occasionally be used where combinations beyond two words in length are the focus of the discussion.
The Web-Based Intelligent English Instruction System CSIEC
(2015). Intelligent Web-Based English Instruction in Middle Schools (pp. 60-103).
www.igi-global.com/chapter/the-web-based-intelligent-english-instruction-system-csiec/119316?camid=4v1a