Transcending Concordance:
Augmenting Academic Text for L2 Writing

Shaoqun Wu, Department of Computer Science, University of Waikato, Hamilton, New Zealand

Ian Witten, University of Waikato, Hamilton, New Zealand

ABSTRACT

This paper describes an automated scheme that extracts salient linguistic features from academic text and presents them in an interface designed for L2 students who are learning academic writing. The system is guided by several common ways of utilizing corpus technology in L2 writing. The authors have developed and tested an extraction method that identifies typical lexico-grammatical features of any word or phrase in a corpus. Collocations and lexical bundles are automatically extracted; students can explore them by searching and browsing, and inspect them along with contextual information. They also present learners with common words, and academic words, hyperlinked to their usage and collocates in authentic contexts. This article uses a single running example, the British Academic Written English corpus, but the approach is fully automated and can be applied to any collection of English writing.

KEYWORDS

INTRODUCTION

This paper describes FLAX, a corpus analysis tool that makes it easy for L2 students to seek language patterns. The system is up on the web for anyone to use online;[1] and the source code is also available.[2] Like existing concordance tools, it allows students to access, analyze and discover linguistic patterns in a particular corpus, which can be chosen to match the task at hand. However, it goes far beyond simply returning concordance lines.

FLAX supports three main functions:

1. Checking vocabulary usage,
2. Seeking grammatical patterns, and
3. Looking up collocations.

First, it facilitates retrieval of typical usage of words or phrases by grouping concordance data and sorting search results to show the most common patterns first. Second, it incorporates grammar rules involving prepositions, word inflection, and articles, and makes common patterns stand out. Third, it retrieves collocations according to part-of-speech tags—for example, all adjectives associated with a particular noun—without using any special syntax.

FLAX is particularly appropriate for supporting academic writing. In this paper we illustrate its design using the British Academic Written English corpus, but it can be applied to any academic

corpus—including samples of writing collected by an individual teacher, entire textbooks (provided they are available electronically), or essays written by students.

The design incorporates the pedagogical implications of many studies of academic corpora. All academic words in a corpus are made available for users to study their usage and collocations directly. Lexical bundles, a feature of academic prose that has recently been unearthed by linguists, are shown in their original context, because they can help L2 students construct academic text. The system is also able to group words by pattern, allowing users to study word usage by showing salient lexico-grammatical patterns that involve it.

FLAX shares the advantages of existing concordancers as popular tools for supporting L2 writing. Researchers report positive responses from students using concordance data for checking grammatical errors, seeking vocabulary usage, and retrieving collocations (Gaskell & Cobb, 2004, O'Sullivan & Chambers, 2006, Yoon, & Hirvela, 2004, Varley, 2009). However, some researchers suggest that concordance data be screened before being presented to students (Varley, 2009). Others ask for commonly used linguistic patterns to be made more accessible (Coxhead & Byrd, 2007), perhaps through a simple interface for retrieving collocations (Chen, 2011). FLAX addresses these criticisms to provide a more learner-friendly interface to the underlying concordance data derived from a given corpus.

This paper is structured as follows. The next section summarizes relevant background of concordancers and how they are used for supporting L2 writing, followed by a brief review of research on academic wordlists, collocations and lexical bundles for academic purposes. Then we describe the British Academic Written English corpus, which is used as an example throughout the paper. Following that we examine and illustrate the facilities provided for searching and browsing word usage, collocations and lexical bundles, and finally draw some conclusions.

THE NATURE OF CONCORDANCERS

Concordancers are software services installed on a computer or accessed through a website that are designed to help users search, access and analyze language in a corpus; some provide a range of corpora that can be used. They are among the most frequently used tools to explore linguistic corpora, particularly when examining collocational use. They make it possible for students to obtain, organize, and study real-language data. The concordancer software mediates a corpus for the user.

Typical systems locate all occurrences of certain words or phrases that the user specifies. They show the contexts in which the item appears, and allow users to explore what is most likely to follow (or precede) it. Users specify a term, and select which of several different corpora to search. They can also associate another word or part-of-speech tag along with the search term, specifying a position— left and/or right—and a distance from the term—e.g., 4 words. The system displays chunks of text (constrained by line width) that contain the search term and associated word (if specified).

Figure 1 shows the result of a search of the Corpus of Contemporary American English website (COCA)[3] for the word *notion*, with a 4-word span on either side. Words of different syntactic types are color coded, e.g. yellow for prepositions. It is easy to see that *notion* is commonly followed by *of* and (by scrolling down) *that*, but students may not uncover typical usage patterns such as verb + *notion* + *of* / *that*, or identify verb collocates.

Retrieving collocations is another common use of concordancers, although not all provide a specific function for this. Retrieval normally requires obscure search syntax and knowledge of part-of-speech tags. For example, the modifier [v*] is used to search for verb collocates of a word in COCA; the result is a list of verbs (e.g., *is*, *was*, *have*, *are*, *had*, ... for the search term *notion*), sorted by frequency. Clicking one displays the concordance lines that contain the two words.

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/transcending-concordance/153892

Related Content

iStudyAbroad
Kathryn Murphy-Judy (2011). *Academic Podcasting and Mobile Assisted Language Learning: Applications and Outcomes (pp. 133-146).*
www.irma-international.org/chapter/istudyabroad/50674

The Impact of Four Reading Motivational Constructs on Motivating EFL Learners to Read Online Texts in English
Khalid Al Seghayer (2013). *International Journal of Computer-Assisted Language Learning and Teaching (pp. 56-81).*
www.irma-international.org/article/impact-four-reading-motivational-constructs/77627

Encouraging Communication through the Use of Educational Social Media Tools
Melissa Barnes (2017). *Multiculturalism and Technology-Enhanced Language Learning (pp. 1-12).*
www.irma-international.org/chapter/encouraging-communication-through-the-use-of-educational-social-media-tools/172859

Exploring The Impact of Integrating a MOOC-VE Blend Within English Language Pre-Service Teacher Education
Abraham Cerveró-Carrascosa (2022). *International Journal of Computer-Assisted Language Learning and Teaching (pp. 1-17).*
www.irma-international.org/article/exploring-the-impact-of-integrating-a-mooc-ve-blend-within-english-language-pre-service-teacher-education/307062

Intelligent Design of Captions in Interactive Multimedia Listening Environments
Vehbi Turel (2015). *Intelligent Design of Interactive Multimedia Listening Software (pp. 47-74).*
www.irma-international.org/chapter/intelligent-design-of-captions-in-interactive-multimedia-listening-environments/135611