

# Using a Large Computer Science Corpus to Build an Automatic Writing Suggestion System

Howard Hao-Jan Chen

Department of English, National Taiwan Normal University

hjchen@ntnu.edu.tw

## Abstract

In this paper, an automatic writing suggestion system is introduced. This system was built based on a large computer science corpus and can automatically provide suggestions of frequent academic lexical bundles for ESL/EFL learners in their English writing. To investigate the usefulness of the system, a group of EFL learners were invited to try on this system and provide feedback, and the students' overall perceptions toward the system was rather positive. For long-term development, bigger corpus data are needed for the advancement of the automatic writing suggestion system.

## Keywords

Automatic writing suggestion system, academic writing

## Introduction

There is an increasing need for ESL/EFL writers to publish their papers in various academic disciplines. It is, however, a very challenging task for many non-native English speakers to write formal academic research papers. Most ESL/EFL novice writers would need to have some help on their formal academic writing. However, few computer-assisted academic writing tools are available for English for Academic Writing (EAP) writing.

More recently, some researchers have begun to develop good tools for ESL/EFL academic writers. Phrases in academic writing have received much attention (Granger, 2017). Some researchers have used lexical bundles to develop good academic writing tools. For example, Liou, Yang, and Chang (2012) created a supportive abstract writing system. Their method involved automatically building domain-specific corpora of abstracts from the Web via domain names and related keywords as query expansions, and automatically extracting vocabulary and n-grams

from the corpora in order to offer writing suggestions. ESL/EFL learners' input in the system actively triggered a set of writing suggestions. This abstract writing assistant system facilitates interactions between learners and the system for writing abstracts in an effective and contextualized way, by providing useful suggestions such as collocations or transitional words.

Mizumoto (2017) has also collected some good quality journal articles and developed an interesting writing tool called AWSuM (Academic Word Suggestion Machine). The system was further expanded to cover all the major sections of an academic paper (Introduction-Method-Results-Discussion).

Based on Mizumoto (2017), AWSuM has the following important features: 1. it can display frequent lexical bundles within a section of a paper in a specific discipline. 2. it can suggest the most frequent lexical bundles through an auto-complete user interface. 3. checking for bundles in a specific discipline is much easier with AWSuM, because it displays the words before and after the search terms.

Although these two innovative academic writing support systems can be very useful for some learners, the content and functions of these two tools can still be improved.

## Methodology

In this study, we used a new corpus crawling tool called AntCorGen developed by Laurence Anthony at Waseda University to download 18,000 computer science articles from the PLOS ONE open-access database. These research articles were then compiled into a large academic corpus. With this large academic corpus, we were able to compile useful teaching and learning materials. About 300,000 4-6 word lexical bundles were extracted and loaded into an online database. This large lexical bundle database was

then used to support the development of an online academic writing suggestion tool. This online writing tool called Academic Phrases Automatic Suggestion System (APASS). APASS, shown below in Figure 1, allows ESL/EFL students to

input any word(s) and the writing suggestion system can automatically suggest some high-frequency lexical bundles after the search word(s).

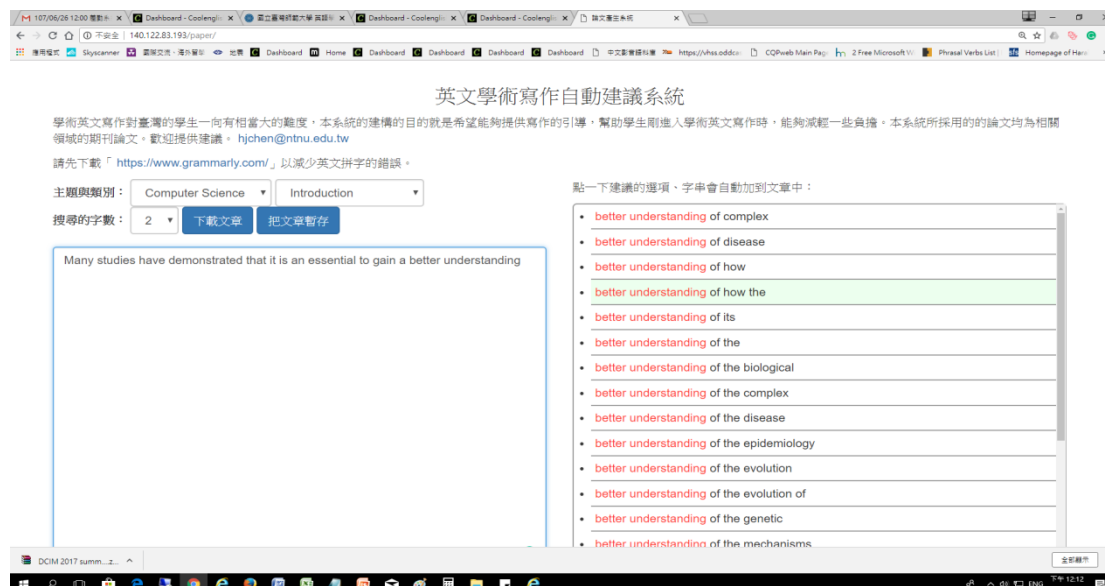


Figure 1: Academic Phrases Automatic Suggestion System (APASS)

To allow the developers to better understand the strengths and limitations of this new writing tool. The writing suggestion tool was also made available to a group of graduate students in Taiwan. About 15 students were invited to give feedback to this system.

## Results and Discussion

The feedback we received was very positive. Students found that the new tool can help them write their abstracts and papers more easily. Moreover, the new system can help them become more familiar with commonly used academic expressions. However, students also noticed that the system sometimes failed to provide good suggestions for some search words. To further enhance its performance, we plan to load more corpus data and also revise the search algorithm. For the long-term development, in addition to computer science, various corpora from the other academic disciplines will also be loaded into the writing support system. It is expected that this writing tool based on various large academic corpora can be used to help ESL/EFL students further improve their academic writing.

## References

Granger, S. (2017). Academic Phraseology: A

Key Ingredient in Successful L2 Academic Literacy. *Oslo Studies in Language*. 9,3, 9-27.

Mizumoto, A., Hamatani, S., & Imao, Y. (2017). Applying the Bundle–Move Connection Approach to the Development of an Online Writing Support Tool for Research Articles. *Language Learning*. 67,4, 885-921.

Liou, H. C., Yang, P. C., & Chang, J. S. (2012). Language supports for journal abstract writing across disciplines. *Journal of Computer Assisted Learning*. 28,4, 322-335.