

ColloCaid: A Corpus-Based Writing Assistant for Academic English Collocations

Keywords: lexicography; writing assistant; collocation; academic English

Dictionaries as we have traditionally known them (that is, as self-contained products) are slowly but surely becoming a thing of the past. In their place, lexical and lexicographic products increasingly take the form of specialized tools, and are often designed with specific tasks in mind. One such prominent task with high demand for lexicographic assistance is text writing.

The ColloCaid project aims to produce an online tool that assists in the process of writing in English by suggesting collocations appropriate for Academic English. It is primarily aimed at scholars writing in English and university students studying in English. Due to the global role of English in the research and educational academic contexts, this makes for a very substantial population of potential users (though ColloCaid may also be useful to other users, e.g. science journalists, politicians, etc.). Research indicates that, contrary to naïve belief, there are no native users of academic language (Hyland, 2006; Kosem, 2010; Frankenberg-Garcia, 2018), and so a collocational tool for Academic English should be of value, not just to non-native speakers of English, but also to native-English-speaking writers who do not have sufficient command of academic English collocations.

Currently, in order to find appropriate collocations, writers may seek help in: English dictionaries (general or collocation-specific), corpora, or existing writing assistants. However, none of these tools serve the specific needs of Academic English writers with respect to collocational choices.

ColloCaid is conceived of as a tool that will be able to suggest collocational choices during the process of writing, with a focus on academic English. The lexicographic database underlying ColloCaid includes at this time: 557 lemmas in 702 senses; 9,401 core non-discipline-specific collocations extracted from corpora of expert academic writing; 30,203 curated corpus examples of core collocations in context; and 31,927 additional interdisciplinary collocations (see also Frankenberg-Garcia, 2015; Frankenberg-Garcia et al., 2019).

Our database draws on up-to-date academic vocabulary lists (Gardner and Davies, 2014) to identify the relevant set of collocational bases (nodes). In turn, several authoritative corpora of academic English have been used as sources of the salient collocates of these collocational bases. The largest part of the collocational bases are nouns, for which we include: typical pre-modifiers, verbs taking those nouns as subjects and objects, respectively, as well as any characteristic prepositions. Verb bases form the second most numerous subclass, and for these we supply modifying adverbs and prepositions would be added. Finally, a number of adjectives are included as bases, with their salient pre-modifying adverbs.

The philosophy of ColloCaid is to provide carefully curated content based on relevant and extensive resources focusing on General Academic English, as opposed to automatically extracted collocates. In this way, writers are less likely to be distracted by noisy data or an overload of information. Currently, ColloCaid has a working, publicly accessible implementation (<https://collocaid.uk/prototype/editor/public/>), which employs the [TinyMCE](#) environment, compatible with multiple devices and operating systems, without the need to download additional software. We intend to give a live demonstration of how the tool works during the presentation.

Word count: 498

Acknowledgement

This research is funded by the UK Arts and Humanities Research Council (AHRC), ref. AH/P003508/1.

References

- Frankenberg-Garcia, Ana. 2015. Dictionaries and Encoding Examples to Support Language Production. *International Journal of Lexicography* 28(4): 490-512.
- Frankenberg-Garcia, Ana. 2018. Investigating the Collocations Available to EAP Writers. *Journal of English for Academic Purposes* 35: 93-104.
- Frankenberg-Garcia, Ana, Robert Lew, Jonathan C. Roberts, Geraint Paul Rees and Nirwan Sharma. 2019. Developing a Writing Assistant to Help EAP Writers with Collocations in Real Time. *ReCALL* 31(1): 23-39.
- Gardner, Dee and Mark Davies. 2014. A New Academic Vocabulary List. *Applied Linguistics* 35(3): 305-327.
- Hyland, Ken. 2006. *English for Academic Purposes: An Advanced Resource Book*. London/New York: Routledge.
- Kosem, Iztok. 2010. *Designing a Model for a Corpus-Driven Dictionary of Academic English*. Ph.D., Aston University, Birmingham.