Developing a tool to help EAP writers with collocations in real time

Ana Frankenberg-Garcia1, Robert Lew2, Geraint Paul Rees3, Jonathan C. Roberts3, & Nirwan Sharma3

1 University of Surrey, UK
2 Adam Mickiewicz University in Poznan, Poland
3 Bangor University, UK

If you came to my talk yesterday...

1. Current research has _______ that...
2. The objective is to _____________ a system that...
3. The data ____________ during the process...
4. The information _________ in Table 3...
5. They attempted to _______________ the effect of...
6. Another ___________ change observed...
7. These decisions play a/an ___________ role in...
8. The analysis was __________________ in two stages...
9. An additional factor that ________

Novice EAP writers have difficulties remembering core academic collocations
Even when their L1 is English

Existing help with core academic collocations

Corpus-based textbooks

https://www.macmillandictionary.com/dictionary/british/research_1

Corpus-based collocation dictionaries

Academic Collocations List (Ackermann and Chen 2013)

Corpus-based EAP dictionary
Existing help with core academic collocations

Direct corpus use

Corpus of Contemporary American English

Collocates of research in COCA

https://corpus.byu.edu/coca/ (Davies 2008)

Existing help with core academic collocations

Direct corpus use

SkELL

Collocates of research in SkELL

http://skell.sketchengine.co.uk/run.cgi/skell

(Baisa & Suchomel 2014)

Excellent help available, but some limitations

EAP textbooks
- Okay in the EAP classroom
- Not so practical when you are actually writing a paper, essay, etc.

Dictionaries
- May not have the exact information you need
- May not have enough information to help you (e.g. examples)
- The information you need might be hidden away in entry

Corpora
- Software generally not very intuitive
- Learners don't always know how to build relevant or appropriate corpus queries
- Concordances can be too difficult, non-standard or inappropriate
- Learners can get distracted by the data

More limitations

Dictionaries and corpora
- Learners have to stop writing to consult references, which can interrupt their flow of words
  (Yoon 2016; Tarp et al. 2017)
- Learners will not look up collocations if they are not aware of their collocation problems
  - Evidence that learners tend to overestimate knowledge of collocations
    (Frankenberg-Garcia 1999, 2005; Laufer 2011)

An alternative solution

Develop intelligent tools that attempt to address these limitations

www.collocaid.uk

- AHRC AH/P003508/1 (Surrey, Bangor and Poznan)
- Ana Frankenberg-Garcia (PI), Jonathan Roberts (CI), Robert Lew (CI)
  Geraint Rees (PDRA) & Nirwan Sharma (PDRA)

What is ColloCaid?

- A text editor that helps writers find the words they need
- Focus on Academic English (EAP) collocations
  - CARRY OUT + research
  - STATISTICAL + significant
  - FURTHER + studies
What’s different about ColloCaid?

- Will help EAP users in real time
- Raising awareness of collocation problems
  - A small increase of temperature
  - the reasons for this increase in temperature are not fully understood
  - a temporary increase in temperature that can result in loss of...
  - symptoms can deteriorate with an increase in temperature
- No auto-correction, helping to learn

Friday, 20 July 2018

What’s different about ColloCaid?

- Customizable
- Option to switch off real-time help and check text at the end
- Useful for revising and proofreading

Friday, 20 July 2018

What’s different about ColloCaid?

- Works in real time, writers do not have to stop writing to consult external resources
- Is not just about feedback on miscollocations...
- Helps writers to expand their collocation repertoire autonomously
- Lexicographic data is curated (scaffolding) so that users don’t get distracted from their writing with
  - irrelevant or misleading information
  - information that is difficult to find
  - an overload of information
  - a flexible, layered approach, with as much (or as little) information as needed
- Metalanguage is kept to a minimum
  - Focus on DDL
  - Users don’t have to know or decipher grammatical terminology

Friday, 20 July 2018

Methodology: which collocations?

Starting point are individual words, not collocations
- Learners ask “What verb can I use with research?”
- Not “Where in my text can I fit in conduct research?”

Circa 500 maximally useful interdisciplinary academic words
1. Academic Vocabulary List (AVL) (Gardner & Davies 2014)
2. Cross-disciplinary AVL lemmas in BAWE (Durrant 2016)
3. Academic Collocations List (Ackermann & Chen 2013)
4. Academic Keyword List (Paquot 2010)
Methodology: 513 collocation nodes

Expert academic English corpora
- Oxford Corpus of Academic English
- Pearson Corpus of Academic English
- BNC academic
- COCA academic
- Cambridge Academic?

Methodology: which collocates?

Strong, frequent, interdisciplinary collocations

Methodology: which sources?

- Expert academic English corpora
  - Oxford Corpus of Academic English
  - Pearson Corpus of Academic English
  - BNC academic
  - COCA academic
  - Cambridge Academic?

Methodology: which tools and criteria?

- Sketch Engine
- Word Sketches
- logDice score ≥ 5
- co-occurrence ≥ 10 for lexical collocates
  100 for prepositions
- Ignore collocates that are too general
- Ignore discipline-specific collocates

Methodology: which examples?

Good corpus examples (Kilgarriff et al. 2008)
- Not too long, not too difficult
  - Save screen space, less distracting
  - Fair use, no plagiarism
- Three examples of each (Frankenberg-Garcia 2014)
- Collocation highlighted in examples (Dziemianko 2014)
- Colligation cues where relevant

Methodology: which errors?

Data for from
- Textbooks, dictionaries
- Crowdsourced feedback (www.collocaid.uk)
- Learner corpora
Visualisation and graphical techniques to help writers in a maximally helpful and minimally intrusive way (Roberts et al. 2017)

Methodology: how to present the data?

ColloCaid

More details  www.collocaid.uk