New directions in DDL for EAP writers

Ana Frankenberg-Garcia¹, Robert Lew², Geraint Paul Rees¹, Jonathan C. Roberts³ & Nirwan Sharma³

¹ University of Surrey, UK
² Adam Mickiewicz University in Poznań, Poland
³ Bangor University, UK
Discovery learning (DL)

- Constructivist approach (Piaget 1936, Vygotsky 1978)
  - Discovering and understanding
  - instead of rote learning

Enter corpora

- Discovery learning through concentrated language data
  - Data driven-learning (DDL) (Johns 1991)
Two classic approaches to DDL

- The hard approach vs. the soft approach (Gabrielatos 2005)

- The hands-on approach vs. the hands-off approach (Boulton 2010)
Hands-off

1) Read the concordance lines below and try to guess what the missing word is.

at to try and sort of figure it all out in my ______ and she started trying to teach me not quite clear hadn't got that clear in my ______. Three o'clock and seven o'clock how can I Now this may just be me going off my ______ sitting in front of Newsnight he became ultimately the erm [tc text=pause] ______ of the careers service in London University. at one stage and took on MX who is now I think ______ of political science erm to do [tc text=pause] cost yeah. [M01] Yes. Yeah. [F01] The deputy ______. What kind [M01] Mm. [F01] of role does she it and FX had this terrific weight behind her ______ and she I think she did twenty er sit ups with I was a child I could not get this out of my ______. I cried at night with it going over and over up smashing the salad-cream bottle er over her ______ and [tc text=pause] I didn't want to hurt her them down my neck. I was smashed out of my ______. When I did eventually get back to the prison a lot of people er never it never entered their ______ to think about these things. [M01] Really? [F01] Mm. [F02] able to perhaps with a clearer ______ approach [F01] Mm. [F02] the more sensitive to say erm [tc text=pause] It's gone out of my ______. [F01] Right. Erm [F02] [ZF1] I th [ZF0] I [M01] MX [M02] there was MX who was then ______ of the Woodlands Unit [M01] Oh was he. [M02] Yeah. [M01] that they've g Well the deputy ______ [M02] [tc text=coughs] [M01] of curriculum erm [M01] Erm [tc text=pause] not off the top of my ______. No [F01] Okay. Well just to conclude may I [tc text=pause] I'm not putting ideas in your ______ but smoking seems to be associated with these you've got to have eyes in the back of your ______ I mean [F01] Mm [F02] having kids yourself you s [ZF1] the [ZF0] the head is he [M01] He's the ______ of the Coordination Centre but [F01] [ZGY] it on. [F01] I know. [F02] It just does my ______ in summation chronic. [F03] Yeah. [ZGY] more [F02] Eh if I do I'll be out of my ______ [M01] I'll get some more [F02]
Both have pros & cons
(Boulton 2010; Frankenberg-Garcia 2012, 2016)
Pros 😊
• Access to the combined intuitions of lots of speakers together
• Authenticity: how language is really used
• Answers to questions not documented in edited language resources
• Learners can explore whatever aspect of language they want
• Serendipity: learners can learn new things they were not even looking for
• Learner autonomy

Cons 😞
• Computers not always available in the language classroom
• Corpus software not particularly user-friendly
• Learners not always aware of what they need help with
• Learners don’t always know how to build relevant/appropriate corpus queries
• Concordances can be too difficult, non-standard or inappropriate
• Learners can misinterpret the data
• Serendipity: learners can get distracted by the data
Pros 😊

- Access to the combined intuitions of lots of speakers together
- Authenticity: how language is really used
- Answers to questions not documented in edited language resources
- No computers needed in the classroom
- Learners do not have to learn to use corpus software
- Teachers aware of learner needs, can prepare tailor-made exercises
- Teachers can edit and adapt raw corpus data if too difficult, non-standard or inappropriate

Cons 🙁

- Teachers need to learn to use corpora themselves
- Preparing own materials can be time-consuming
- Teachers need to learn how to integrate corpora to their teaching
  - Many misguided DDL materials
- What is useful to one student may not be relevant to another one
- Teachers can only cover a limited amount of language
A third approach

- Dynamic Data-Driven Learning (DDDL)

- Develop intelligent tools where
  - Learners don’t have to learn to use corpora
  - But engage with data-driven learning as and when required
  - Data is curated so that learners do not get distracted or lost in the process

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 DDDL writing assistant
  - Helps writers find the words they need
- Focus on Academic English (EAP) collocations
  - The way words combine in academic English
    - carry out + research
    - statistically + significant
    - further + studies
Existing help with EAP collocations

Textbooks
Existing help with EAP collocations

General language dictionaries
https://www.macmillandictionary.com/dictionary/british/research_1

**research** - definition and synonyms

- Noun [Uncountable]
- Pronunciation: /riˈsɜː(t)ʃən/ (research)

*the detailed study of something in order to discover new facts, especially in a university or scientific institution*

*Recent research shows that babies in the womb can be influenced by music.*

*She teaches a lot of classes and doesn’t have much time for her own research.*

*medical/historical/linguistic research*

**research into:** research into the causes of schizophrenia

**research on:** There is clearly a need for further research on this topic.

**carry out research:** Government scientists have carried out extensive research into the effects of these drugs.
Existing help with EAP collocations

Collocation dictionaries

Academic collocations list (Ackermann and Chen 2013)
Existing help with EAP collocations

EAP dictionary

ADJECTIVE + RESEARCH
future, further, previous, prior, recent, current, basic, extensive, qualitative, quantitative, comparative, empirical, scientific, academic, participatory, social, ethnographic, medical, clinical, geographic, geographical, historical

- These results are consistent both with the theory and with prior qualitative research.
- Early academic research suggested that the media more often reinforced than changed political views.

VERB + RESEARCH
conduct, undertake, carry out, do, need, require, focus on, draw on, review, describe, report, present, publish, be based on

- Many large companies employ market research agencies to conduct their research for them.
- More research is needed on how business models evolve.
Existing help with EAP collocations

Collocates of *research* in COCA
[https://corpus.byu.edu/coca/](https://corpus.byu.edu/coca/) (Davies 2008)
Learning Collocations

Family words: researched researcher researchers researches researching

<table>
<thead>
<tr>
<th>used as a noun</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>research</td>
<td>project</td>
<td>303</td>
</tr>
<tr>
<td>research</td>
<td>programme</td>
<td>237</td>
</tr>
<tr>
<td>market</td>
<td>research</td>
<td>209</td>
</tr>
<tr>
<td>research</td>
<td>findings</td>
<td>141</td>
</tr>
<tr>
<td>research</td>
<td>team</td>
<td>130</td>
</tr>
<tr>
<td>research</td>
<td>contract</td>
<td>103</td>
</tr>
<tr>
<td>research</td>
<td>work</td>
<td>102</td>
</tr>
<tr>
<td>research</td>
<td>students</td>
<td>95</td>
</tr>
<tr>
<td>research</td>
<td>interests</td>
<td>82</td>
</tr>
<tr>
<td>research</td>
<td>methods</td>
<td>80</td>
</tr>
</tbody>
</table>

Existing help with EAP collocations

<table>
<thead>
<tr>
<th>research</th>
<th>noun</th>
<th>switch to research (verb)</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>verbs with research as subject</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>suggest</td>
<td>show</td>
<td>focus</td>
<td>indicate</td>
</tr>
<tr>
<td>identify</td>
<td>prove</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>verbs with research as object</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conduct</td>
<td>fund</td>
<td>undertake</td>
<td>do</td>
</tr>
<tr>
<td>apply</td>
<td>continue</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>adjectives with research</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ongoing</td>
<td>assistant</td>
<td>underway</td>
<td>online</td>
</tr>
</tbody>
</table>

http://skell.sketchengine.co.uk/run.cgi/skell
(Baisa & Suchomel 2014)
Excellent resources available

But some limitations

• If curated (e.g. dictionaries), limited coverage
• If not curated (e.g. corpora), noise
• Learners will not look up collocations if they are not aware of their shortcomings
• Learners have to stop writing to use collocation resources (which may impair word flow)
What’s different about ColloCaid?

• Will help EAP users with collocations in real time (Frankenberg-Garcia et al. 2019; Lew et al. 2018)
• Raising awareness of collocation problems

A small increase of temperature

× Do not say increase of something
✓ Say increase in something
What’s different about ColloCaid?

- Not just correcting reactively, but helping to learn
- Raising awareness of collocations EAP users may not remember to look up in the first place

Research done as a result of...

- conduct research +
- research suggests +
- empirical research +
- research on something +
- conduct research +
- carry out research +
- undertake research +

- research carried out in the 1990s identified...
- before any research is carried out, it is essential that...
- this may mean that research can only be carried out in the dry season
What’s different about ColloCaid?

• Customizable
• Option to switch off real-time help and check text at the end

A small increase of temperature...

Research done as a result of...

• Useful for revising and proofreading
What’s different about ColloCaid?

• Customizable in more than one way

  Research

  one as a result of...

  • conduct research +
  • research suggests +
  • empirical research +
  • research on something +

Restore hidden prompts

☐ research

☐ hypothesis

☐ analyse
What’s different about ColloCaid?

- Metalanguage kept to a minimum
  - Users don’t have to know or decypher grammatical terminology

- Data for data-driven learning will be curated so that
  - Unlike corpora, users don’t get distracted from their writing with
    - irrelevant or misleading information
    - an overload of information
  - Unlike dictionaries, just the right amount
    - layered approach
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)
   - Cross-disciplinary AVL use in BAWE (Durrant 2016)
2. Academic Collocations List (Ackermann & Chen 2013)
3. Academic Keyword List (Paquot 2010)
Methdology: 513 collocation nodes
Methodology: which collocates?

**Noun bases**
- Adjectives & modifiers: e.g. a complex system, a classification system, an existing system, etc.
- Verbs (object of): e.g., design a system, develop a system, establish a system, etc.
- Verbs (subject of): e.g. the system consists of, the system functions, the system provides, etc.
- Prepositions: e.g. within a system, a system for something

**Verb bases**
- Adverbs: e.g. undoubtedly contribute, contribute significantly, contribute substantially, etc.
- Prepositions: e.g. contribute to

**Adjective bases**
- Adverbs: e.g. statistically significant, particularly significant, etc.
Methodology: which sources?

Expert academic English corpora

- Oxford Corpus of Academic English
- Pearson Corpus of Academic English
- BNC academic
- COCA academic
Methodology: which tools?

- Sketch Engine (Kilgariff et al. 2004, 2014)
- Word Sketches

Word Sketch for *research* in OCAE
Methodology: which criteria?

- logDice score of $\geq 5$
- co-occurrence $\geq 10$ for lexical collocates
- $\geq 100$ for prepositions
- Ignore collocates that are too general
- Ignore discipline-specific collocates
Methodology: which examples?

Good corpus examples

• Not too long, not too difficult
  • Save screen space, less distracting
• Three examples of each (Frankenberg-Garcia 2014)
• Collocation highlighted in examples (Dziemianko 2014)
• Colligation cues where relevant

• research carried out in the 1990s identified...
• before any research is carried out, it is essential that...
• this may mean that research can only be carried out in the dry season
Data from

- Textbooks, dictionaries
- Crowdsourced feedback (www.collocaid.uk)
- Learner corpora
  - Open Cambridge Learner Corpus
  - BAWE (Nesi et al.)
  - BrAWE (Goulart da Silva 2016)
    - Overuse: create system
    - Underuse: design system
Integration with text editors

Visualisation and graphical techniques to help writers in a maximally helpful and minimally intrusive way (Roberts et al. 2017)
Pros 😊

- Access to the combined intuitions of lots of speakers together
- Authenticity: how language is really used
- Answers to questions not documented in edited language resources
- Learners can explore whatever aspect of language they want
- Serendipity: learners can learn new things they were not even looking for
- Learner autonomy

Cons 😞

- Computers not always available in the language classroom
- Corpus software not particularly user-friendly
- Learners not always aware of what they need help with
- Learners don’t always know how to build relevant/appropriate corpus queries
- Concordances can be too difficult, non-standard or inappropriate
- Learners can misinterpret the data
- Serendipity: learners can get distracted by the data
Pros 😊
• Access to the combined intuitions of lots of speakers together
• Authenticity: how language is really used
• Answers to questions not documented in edited language resources
• No computers needed in the classroom
• Learners do not have to learn to use corpus software
• Teachers aware of learner needs, can prepare tailor-made exercises
• Teachers can edit and adapt corpus raw data if too difficult, non-standard or inappropriate

Cons 😕
• Teachers need to learn to use corpora themselves
• Preparing own materials can be time-consuming
• Teachers need to learn how to integrate corpora to their teaching
  • Many misguided DDL materials
• What is useful to one student may not be relevant to another one
• Teachers can only cover a limited amount of language
I
References


References


New directions in DDL for EAP writers: ColloCaid