Putting corpora to work:

applications of corpus tools in the EAP classroom

Aims

1. **Introduce** widely-available tools to enable students and teachers to use corpora in the EAP classroom.

1. **Practise** using these tools for:
    a. Using formulaic language (e.g. thesis statements).
    b. Providing constructive feedback on student writing.
    c. Preparing for presentations
    d. Micro-writing practice focusing on improving word choice (e.g. collocations)

1. **Reflect** on how these tools might be used in the classroom.

Which quote is more true for you?

"Now corpora, large and small, are seen by many teachers as useful tools, and are being put to use more and more every day. Access has become fairly easy on standard small computers, user-friendly software is available for most normal tasks, websites are accumulating fast, and **corpora are** almost **part of the pedagogical landscape**" (Sinclair, 2004, p.2).

"While the contribution of computer corpora to language description is by now an uncontested fact, with corpus linguistics being firmly established as a branch of linguistic inquiry in its own right, the **application of computer corpora for language teaching** is still a rather **neglected area**"(Kaltenböck & Mehlmauer-Larcher, 2005, p.65).

How we ended up here...

1. We want learners to engage with language as it is actually used.
2. We want learners to be given tools to grow in their skills and autonomy.
3. We want learners to be empowered to make informed decisions about their language use, rather than being provided with a set of acceptable answers by a teacher.
4. We want to encourage learners to be critical users of online and digital tools.

Rationale/principles for choice of activities

We want to put corpora to work in way which:

1. **Is easy to use: for both teachers and students, corpora need to be straightforward to use, and not a distraction from other priorities.**
2. **Develops and promotes learner autonomy.**
3. **Helps with differentiation: putting corpora to work provides good opportunities for differentiation.**
4. **Is adaptable: not requiring the rewriting of curricula or generating complex materials (nothing we present here took us more than a couple of hours to create).**

Corpora we have used in teaching

1. **The British Academic Written English (BAWE) corpus** contains approx 3000 good-standard student assignments (**6,968,089 words**) distributed across **four broad disciplinary areas** (Arts and Humanities, Social Sciences, Life Sciences and Physical Sciences) and across **four levels of study** (undergraduate and taught masters level).

2. **The British Academic Spoken English (BASE) corpus** consists of **160 lectures** and **38 seminars** recorded in a variety of university departments and contains **1,477,281 words**. Holdings are distributed across the same four disciplinary groups as BAWE.

3. **(The British National Corpus (BNC)** has an academic component.)

Using formulaic language to develop thesis statements

- **Take advantage of formulaic language to develop parts of students' writing.**
- **Example: Thesis statements.**
  - **"This essay will…"**

Using formulaic language to develop thesis statements

Students guess words likely to follow 'This essay will...'

Use corpora to provide data to check guesses.

Follow up with observation questions:

- *E.g. write 5 words which follow the phrase more than once.*

Students then developed their thesis statements using this data.

Which do you think are the top 3 verbs which follow 'this essay will...'?

Using formulaic language to develop thesis statements

Which do you think are the top 3 verbs which follow 'this essay will…'?

Using formulaic language to develop thesis statements
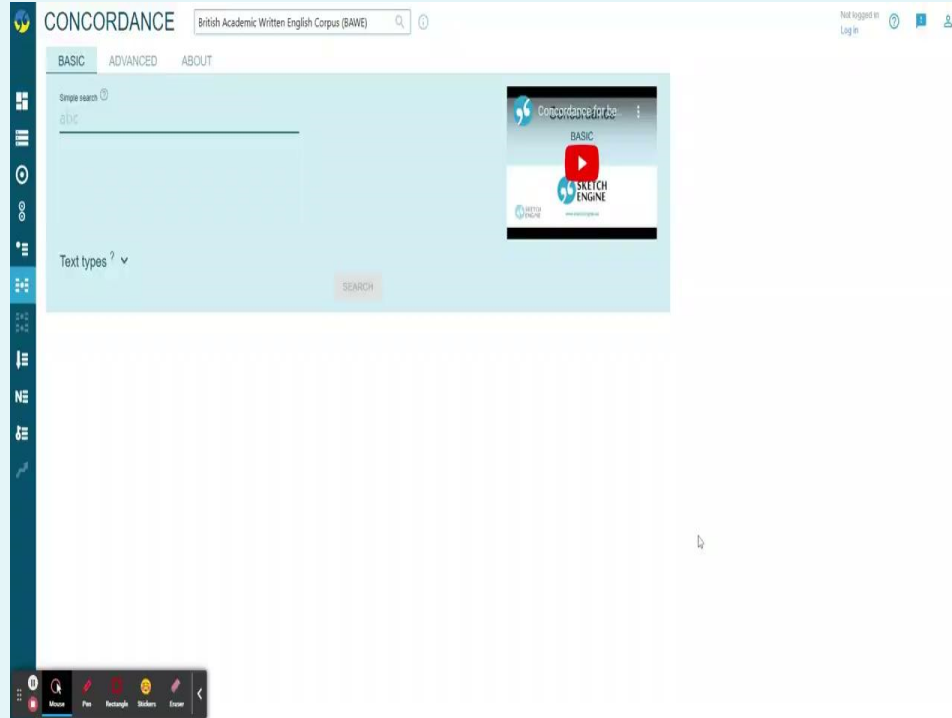
Using formulaic language to develop thesis statements: for lower levels

	Word	Word	Frequency	Relative ?
1	This essay will	discuss	14	1.68 •••
2	This essay will	examine	14	1.68 •••
3	This essay will	attempt	14	1.68 •••
4	This essay will	focus	13	1.56 •••
5	This essay will	argue	11	1.32 •••
6	This essay will	look	11	1.32 •••
7	This essay will	explore	10	1.20 •••
8	This essay will	first	10	1.20 •••
9	This essay will	analyse	8	0.96 •••
10	This essay will	consider	7	0.84 •••

Using formulaic language to develop thesis statements

Students guess words likely to follow 'This essay will...'

Use corpora to provide data to check guesses, ideal for higher levels.
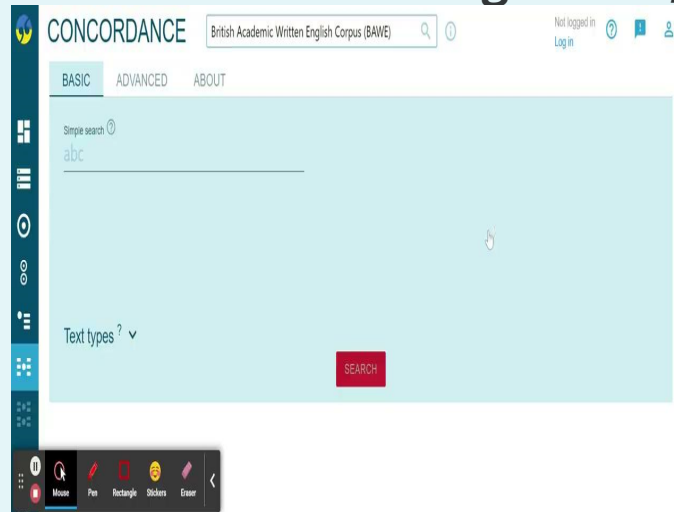
Activity: Using formulaic language to develop thesis statements

Can you use the corpus to start investigating so you can create something you could potentially use with a class of students?

Use Sketch Engine (concordance)

- Type the formulaic language and see what you get.
- Can you think of any other applications of this tool?

[BAWE QUICKLINKS](#) and OWN LINKS

- based at Coventry University

- a database of hyperlinks to concordance lines, collocations lists or word sketches from BAWE in response to some common student errors

- can be accessed by anyone

- our experience: making own links allows for differentiation and specificity.

Making your own links for error correction: Some principles

Differentiation:

- Important to note that there is a range of searches you can perform and provide to the students. Some students will benefit from a more open-ended search, maximising scope for autonomy, analysis and decision-making.

- Some students will need something which provides them with a limited range of possibilities from which to choose.

Ease of use:

- Some searches require more skill and expertise from the teacher and the learner.

- In general, the lower level the student, the greater the expertise needed by the teacher.

Making your own links for error correction: Some tools

1. Filter by part-of-speech

1. Corpus Query Language *(handout available, but we will not cover this today)*

Tool 1: Part of Speech

a predetermined set of mathematical theories. AI is now a crucial element in this new modern life style , as it is adopted in every field , for example according to survey made by Steffen Kinkel , Marco Baumgartner and Enrica Cherubini,(2022), 23% of the 3000 company surveyed had adopted AI applications in some processes or services, and 23% at least one AI pilot project. Yet, there is still some hesitation and worry that relevant for taking part in business due to its negative impact as 22% of 3000

Language

Review this word choice.

Filtering by part-of-speech is often enough

Errors 1-3: Filter by part-of-speech

Access Handout 2 on Sched to try this now.

Making your own quicklinks for error correction: Some tools

1. Filter by part-of-speech ✔

1. Using Corpus Query Language
   a. Enables more specific, tailored examples.
      i. More granular grammar (e.g. superlative adjectives, modal verbs).
      ii. Either/or choices.
      iii. Numerous tools to make exemplars more specific (and thus helpful to lower levels).

Using BASE to prepare for presentations

Context:

- **Students preparing for presentations.**
- **Presenting issue of scripts.**

Aim:

- **Understanding differences between written and spoken academic language.**
- **Evaluate progress on presentations and make improvements.**

NB: This also reinforces some of the learning about distinct conventions of academic written language.
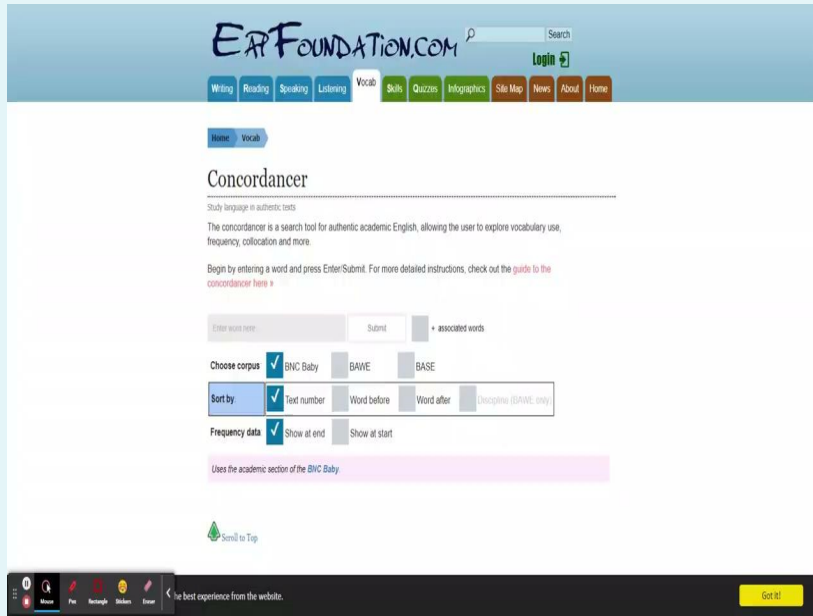
Using BASE to prepare for presentations
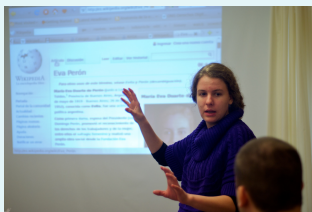
'And so today, I'm going to explain the findings of my research into…'

We considered this phrase, and asked students to note features of the language. Students guessed at relative frequency of the following in written and academic English.

- 'So today'
- 'I'm going to'

What do *you* think?

Using BASE to prepare for presentations

Using BASE to prepare for presentations

Lots of examples were then provided of language, some of which were much more common in BASE (e.g. we're going).

Compare the following examples using EAP Foundations concordancer (eapfoundation.com/vocab/concordancer)

a. **we're going**
b. **move on**
c. **consider**
d. **think about**
e. **consideration**

1. What information does this give a learner?
2. How could they use this information?
3. What features of language could this tool help you teach?

Developing writing skills using [ColloCaid](#).

- **University of Surrey, Bangor University, Adam Mickievicz University in Poznan, Poland**

- **suggests collocations and provides examples in real time**

Developing writing skills using [ColloCaid](#).

We asked students to report the information below first without ColloCaid, then relying on collocations suggestions offered by ColloCaid.

The British Social Attitudes Survey 2021.

Information:
- 40% are very concerned about the environment. 22% were very concerned in 2011.
- 45% view climate change as being the most important environmental issue (up from 19% in 2011).
- 76% of those who are 'very concerned' about the environment are willing to pay to protect it.10% of people say they will not pay to protect the environment, despite their concern.

Source: British Social Attitudes Survey (2021)

Developing writing skills using ColloCaid.
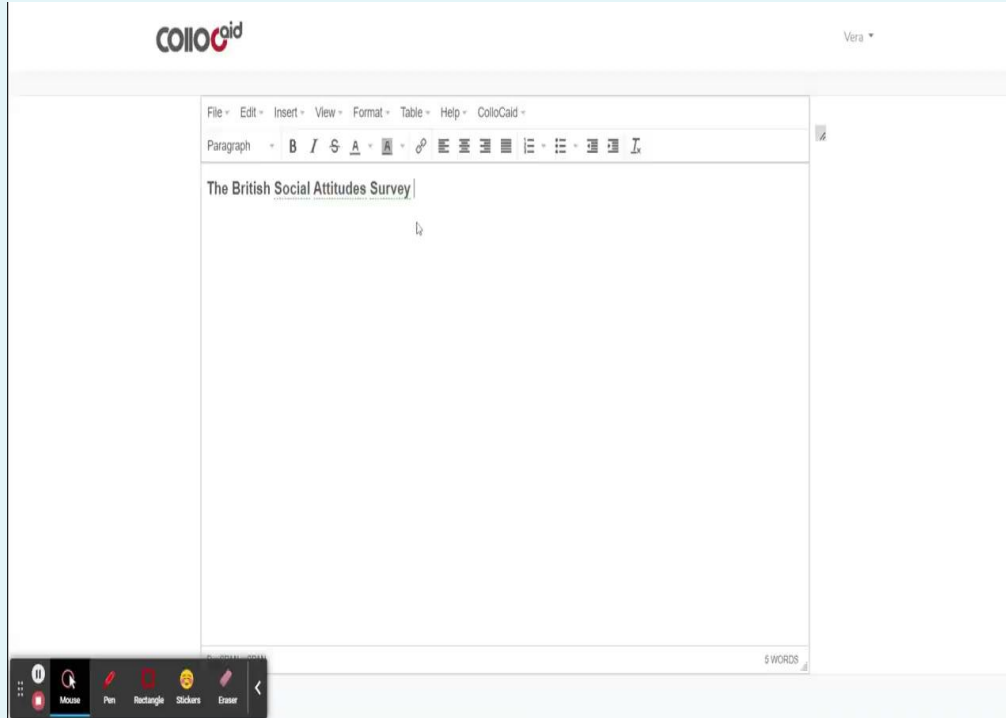
Some examples of student writing

Sentences without ColloCaid	Sentences with ColloCaid
According to the British Social Attitudes Survey (2021), 40% of British are extremely concerned about the environment, which has doubled from 2011.	*The British Social Attitudes survey reveals 40% of British view climate change as the most important environmental issue, which is up from 19% in 2011.*
There was a dramatic rise in the number of people considering climate change to be a critical issue between 2011-2021, which increased from 19% to 45%.	*According to the survey, the numbers of people concerned about the environment doubled between 2011 and 2021…*

Activity: Developing writing skills using [ColloCaid](http://www.collocaid.uk).

Try using ColloCaid www.collocaid.uk - Try the prototype - Register

**Use your own ideas or report the same information from the British Social Attitudes Survey 2021:**

Information:
- 40% are very concerned about the environment. 22% were very concerned in 2011.
- 45% view climate change as being the most important environmental issue (up from 19% in 2011).
- 76% of those who are 'very concerned' about the environment are willing to pay to protect it.10% of people say they will not pay to protect the environment, despite their concern.

Source: British Social Attitudes Survey (2021)

Activity: Developing writing skills using [ColloCaid](#).

Reflection (suggested questions)

- **What did you think? Could you use it?!**

- **What kind of learning activities could it support?**

- **How might it be used in a differentiated way?**

[ColloCaid](): initial reflections

- **works well with higher level students, but can be confusing for lower level students (complex grammar)**
- **good potential to develop autonomy**
- **requires repeated exposure**
- **works better for writing in real time, not for proofreading/introducing improvements/corrections**

Corpora and the tools to access them

BAWE
The British Academic Written English

↓

Sketch Engine http://www.sketchengine.eu/

EAP foundations concordancer
https://www.eapfoundation.com/guides/concordancer/

ColloCaid
https://collocaid.uk/prototype/editor/public/

BAWE Quicklinks - lexico-grammatical patterns common in academic writing
https://bawequicklinks.coventry.domains/

BASE
The British Academic Spoken English

↓

Sketch Engine
http://www.sketchengine.eu/

EAP foundations concordancer
https://www.eapfoundation.com/guides/concordancer/

BNC
The British National Corpus

↓

BNC lab
http://corpora.lancs.ac.uk/bnclab/search

BNC https://www.english-corpora.org/bnc/

References

Kaltenböck, G., & Mehlmauer-Larcher, B. (2005). Computer corpora and the language classroom: on the potential and limitations of computer corpora in language teaching. *ReCALL*, *17*(1), 65–84. https://doi.org/10.1017/S0958344005000613

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: ten years on. Lexicography, 1, 7-36, . http://www.sketchengine.eu/

Lew, R., Frankenberg-Garcia, A., Rees, G., Roberts, J. C., & Sharma, N. (2018). ColloCaid: A real-time tool to help academic writers with English collocations. Paper presented at Proceedings of the XVIII EURALEX International Congress, Ljubljana, Slovenia. https://www.collocaid.uk/

Sinclair, J. (2004). Introduction. In: Sinclair J. M. (ed.) *How to use corpora in language teaching* (pp.1-10). J. Benjamins.

Smith, C.H.S. (2013). *Concordancer.* EAP Foundation. https://www.eapfoundation.com/vocab/concordancer/

Vincent, B. & Nesi, H. (2018) The BAWE Quicklinks project: a new DDL resource for university students. *Revue de Linguistique et de Didactique des Langues (Lidil),* 58: http://journals.openedition.org/lidil/5306 https://bawequicklinks.coventry.domains/