

Exploring academic literacy: preliminary insights from the Cambridge Corpus of Academic English

1. Rationale

Being literate in English is often the expectation in academia today. Our research focus is to identify **specific features of academic English** in order to explore what **academic literacy** means for native and non-native users of English and their instructors. This will enable us to identify how success in academia interacts with an individual's language background, their general level of language proficiency and their understanding and use of core and field-specific academic conventions.

Developing a robust construct of academic literacy will provide English language instructors and higher education professionals with a better understanding of the language skills required to operate successfully through the medium of English in academia. This research will inform better teaching and assessment.

3. Preliminary analysis

The documents collected were converted to plain text and categorised according to academic level. These text files were then analysed using Sketch Engine, a web-based software programme that allows the storage and interrogation of corpus data.

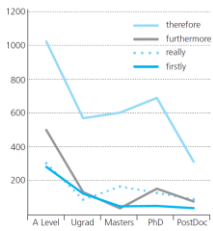
3.1 Discourse markers

We found that A Level students have a tendency to overuse certain discourse markers when compared with more experienced writers.

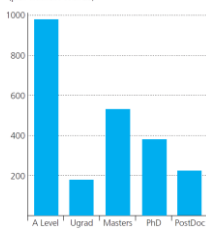
- **Pedagogical impact** may play a role: students are often provided with lists of key academic vocabulary, which they are encouraged to use.

- Experienced academics may have developed more of an ability to organise ideas and structure arguments, perhaps without using such vocabulary.

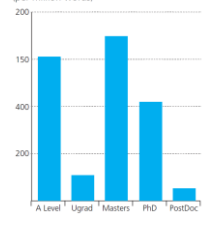
Discourse markers (per million words)



Frequency of THINK (per million words)



Frequency of I | WE THINK | THOUGHT (per million words)



3.2 The verb THINK

In our data, we found that the verb *think* is significantly more common at A Level. Interestingly, the frequency drops dramatically at undergraduate level. The trend is similar when we look at the frequency of *I/we think/thought*, although these forms are most frequent at Masters level rather than A Level.

Possible explanations include:
- Less experienced academics might use *think* as a **hedging device**. The sudden drop at undergraduate level could be a result of explicit instruction to avoid this kind of vague or cautious language.

- The relatively high frequencies at Masters and PhD levels may be a result of writers reporting on their **own research findings**, as opposed to undergraduate level, where courses are mostly taught rather than research-based. The concordance below provides typical examples of usage at Masters level:

other members of the class. Something that errors in the transcribed oral interview, negotiation of meaning in a meaningful setting, I **think** adds to Lourdes' success is that she is other-opening experience for me. I chose what I **think** it is also important to point out that I **think** the interview captures a more accurate I **thought** was an "active" approach in order to contrast I **thought** writing a letter would allow her to I **think** he understood what he was saying, I **think** several of the students would have felt

- We found that at Post Doc level, writers have a preference for the lemma *consider* rather than *think*, which could explain the low frequency of the latter at this level.

2. Pilot data collection

In 2012-13, we collected 1 million words of written academic English from a range of subject areas, academic levels and language backgrounds, as illustrated below:

Pilot data by academic subject



Pilot data by academic level



Pilot data by language background



During the pilot phase, our primary goal was to amass sufficient data to test our collection methods and carry out preliminary analyses, rather than seeking to collect a balanced corpus.

3.3 Qualitative analysis of A Level data

- Our A Level data is characterised by a significantly more **informal tone** than writing at other levels.

- We found numerous examples of **hypercorrection**, including the mis-use of *whom*, the overuse of semi colons, the over-application of features they associate with academic writing e.g. discourse markers.

- It may be that the students are aware of certain features of academic writing, but their understanding is still superficial; attempts to include these features are often stilted and clumsy:

A person whom is intelligent...

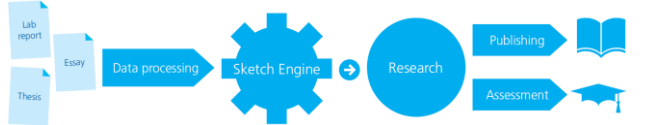
Overall therefore despite having omitting many details...

...with a view in mind as to who they were and for who they intended their works to be read by...

Women are most likely to look after their child say if the child was sick...

4. Applications

The Cambridge Corpus of Academic English will facilitate the development of EAP materials and inform assessments of academic English, for example admissions tests.



At Cambridge English, our corpus research provides us with information that we can integrate into various types of materials to create courses and assessments based on authentic English and empirical findings.

This research will provide insights into many areas, including:

- common learner errors
- frequency of vocabulary and structures
- word patterns and collocations
- differences between spoken and written English

Our research is integrated into our publications in a variety of formats:

Corpus research shows that the most common past participle that goes in the structure As in tables/figures/section/chapter is shown. What do you think are the **next** most common? The first letters are given to help you.

2. whidn	3. de	5. in	7. def	9. gl
2. di	4. e	6. se	8. dep	10. ou

3. Verb-noun collocations

Collocations are combinations of words which commonly go together. Match the verbs in the box to the nouns to make typical verb-noun collocations. These verb-noun collocations occur in the text you read in 2.3 on pages 16-17.

achieve	make (a2)	measure
satisfy	handle	take

1. academic problems	5. outcomes
2. action	6. progress
3. benefits	7. decisions
4. needs	

Research shows that in the written academic corpus the **five** most common verb-noun collocations are: 1. solve a problem 2. take a decision 3. handle a situation 4. measure 5. measure the success of something. If you have 9 words 10 deal with.

Research shows that in the spoken academic English, corpus less formal phrases are used in preference to -only adverbs. For example instead of "Culturally speaking" say "As for as culture is concerned, ...". From a cultural point of view, ... From the point of view of culture, ... In cultural terms, ... In terms of culture, ...

5. Next steps

Following the pilot project, the Cambridge Corpus of Academic English has entered a new phase of data collection. Contributors are invited to register on our website and follow three steps to take part:

1. Complete a questionnaire about yourself and your work
2. Upload your documents to the portal
3. Tick to give consent for your work to be used in research

www.cambridge.org/camcae/datacollection

Our new data portal allows us to collect data more efficiently, and facilitates contributions from students and academics around the world.

Strategies for data collection

- Emails inviting contribution sent to student lists through university department offices.
- Use of social media to generate publicity and interest.
- Workshops and talks relating to academic English and corpora provided to local institutions in exchange for participation.
- Contributors are entered into a monthly prize draw to win vouchers.

Get involved

We plan to make CAMCAE data available to researchers once the project is more developed.

If you are interested in contributing to our research, using our corpus or sharing details of related research, we would be delighted to hear from you.

www.cambridge.org/camcae