

# Students as proof- readers of their own texts: Using corpus tools for writing improvement

Maggie Charles  
University of Oxford  
[maggiecharles\\_oxford@yahoo.com](mailto:maggiecharles_oxford@yahoo.com)



# Context and Rationale

- Writing a **thesis/dissertation** or preparing work for **publication** is extremely **demanding for graduate L2 writers** (Thompson 2013; Lillis & Curry 2010)
- **Corpora** have **successfully** been used to improve **lexicogrammar** (Boulton & Cobb 2017)
- **Concordancing** is useful at Burrough-Boenisch's (2013) '**language editing**' level
- Other **corpus tools** can be used at the **language, content and organization levels of proofreading/editing** (Charles 2018)

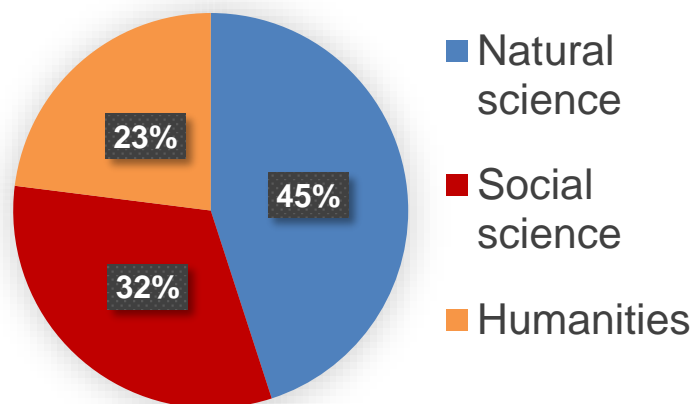
The course ***Editing your Thesis with Corpora*** was devised to help doctoral students use corpus tools to edit and proofread their texts

# *Editing your Thesis with Corpora*

## Course Details

- 90 doctoral students
  - 2-3 courses per year
  - 13 courses in total
  - one 2-hour session/week
  - 6 week course
- 
- AntConc (Anthony 2018)
  - AntFileConverter (Anthony 2017)
  - Data
    - in-class worksheets
    - final evaluation questionnaire

# 90 Course Participants 2012-2017



## Disciplinary Areas

### Natural science:

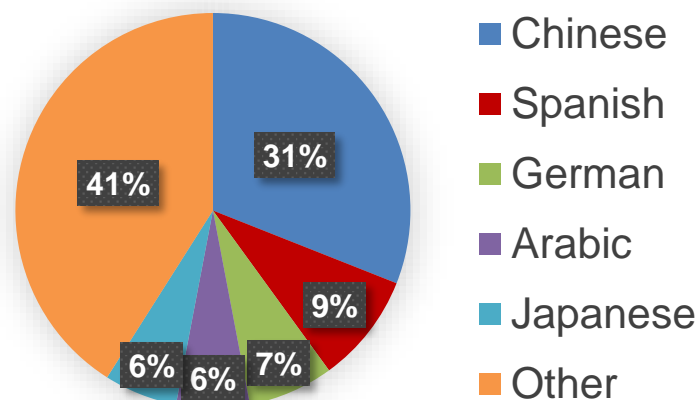
environmental science, medicine, materials science, engineering

### Social science:

sociology, education, politics

### Humanities:

Oriental studies, modern languages, archaeology



## Top L1s

Other L1s include:

4 students

Korean, Malay, Russian, Turkish

2 students

Dutch, Hindi, Italian, Lithuanian, Nepali, Portuguese, Thai

# *Editing your Thesis with Corpora*

## Course Programme

Week	Editing Topic	Corpus Tool
1.	Using concordances to answer grammar, vocabulary and usage queries	AntConc Concordance
2.	Building your corpus of research articles; answering your own editing queries	AntFileConverter Concordance
3.	Finding collocations and semi-fixed phrases; building a corpus of your own writing	Clusters Collocates
4.	Examining the words you use; checking for consistency; comparing your own writing with expert texts	Word List N-grams
5.	Tracing content, themes, terminology, citation throughout your own writing	Concordance Plot
6.	Comparing individual chapters to the rest of the thesis; comparing your own writing with expert texts	Keyword List

# Do-It-Yourself Corpora: Two Types

## **1. DIY Corpus of Research Articles in student's own field/topic area**

- based on downloaded files in own bibliography
- may include subcorpora of different topics/genres

## **2. DIY Corpus of Student's Own Writing**

- chapters of thesis as individual files
- may include subcorpora of other writing (e.g. research articles, proposals, Master's dissertation)

# Clusters in Use: Liu

Concordance		Concordance Plot	File View	Clusters/N-Grams
Total No. of Cluster Types		23	Total No. of Cluster	
Rank	Freq	Range	Cluster	
1	105	11	emergence of	
2	27	7	emerged in	
3	24	8	emergence of a	
4	16	7	emerged in the	
5	15	3	emergence of an	

- shows all the **phrases** which contain your search item
- presents them as a **list** and gives their **frequency**
- lets you choose **the length** of the cluster
- lets you select the **position** of the search item in the cluster

18	4	3	emerge through
19	4	4	emerged from the
20	3	3	emerged and
21	3	3	emerged as a
22	3	3	emergence in the
23	3	3	emerges as a

**Liu:** Chinese doctoral student in geography and the environment

**Corpora:** 31 research articles; 1,122,712 words

1 thesis chapter; 11,725 words

**Thesis:** The formation of environmental knowledge and anti-pollution mobilisation in rural China

# Clusters in Use: Liu

Concordance		Concordance Plot	File View	Clusters/N-Grams
Total No. of Cluster Types		23		
Total No. of Cluster				
Rank	Freq	Range	Cluster	
1	105	11	emergence of	
2	27	7	emerged in	
3	24	8	emergence of a	
4	16	7	emerged in the	
5	15	3	emergence of an	

- shows all the **phrases** which contain your search item
- presents them as a **list** and gives their **frequency**
- lets you choose **the length** of the cluster
- lets you select the **position** of the search item in the cluster

18	4	3	emerge through
19	4	4	emerged from the
20	3	3	emerged and
21	3	3	emerged as a
22	3	3	emergence in the
23	3	3	emerges as a

**Liu:** Chinese doctoral student in geography and the environment

**Corpora:** 31 research articles; 1,122,712 words

1 thesis chapter; 11,725 words

**Thesis:** The formation of environmental  
pollution mobilisation in rural

**Liu's Question:** ○ ○ ○

'emerge in' vs  
'emerge from'?

# Liu's Investigation: RA Corpus

Search for clusters of 2-4 words, with *emerge*\* on the left

Rank	Freq	Range	Cluster
1	105	11	emergence of
2	27	7	emerged in
3	24	8	emergence of a
4	16	7	emerged in the
5	15	3	emergence of an
6	14	6	emerge from
7	12	5	emergence and
8	10	4	emerge in
9	10	7	emerged from
10	9	5	emerge from the
11	9	3	emergence in
12	8	4	emergence of
13	8	4	emergence of a new
14	8	3	emergence of the
15	6	6	emerge as
16	6	4	emerged as
17	5	4	emerge as
18	4	3	emerge through
19	4	4	emerge from the
20	3	3	emerge and
21	3	3	emerge a
22	3	3	emerge the
23	3	3	emerge

Emerge/d in = 37  
Emerge/d from = 24

Search Term ☒ Words ☐ Case ☐ Regex ☐ N-Grams ☐ Advanced

emerge\*

Start Stop Sort

Sort by ☐ Invert Order ☐ Search Term Position ☒ On Left ☐ On Right

Cluster Size Min. 2 Max. 4

Min. Freq. 3 Min. Range 3

itute the "poli- tics" political ecology studies—emerge from a field of competing normative ideas e propositions about relationships. Once patterns emerge from dimensional profiles, more may be de in the field, theory was allowed to emerge from loosely structured, lengthy, open-en nding of the research findings: 1) Do [concepts] emerge from the data or are they imposed?...2) . Conversely, the production of new natures might emerge from the forms of consciousness made possib and streamline administration, numerous agencies emerge from the grassroots level, among which neig ppear "fragmented and episodic," its history will emerge from the "margins" and become coherent on e development and activity of intellectuals who emerge from the masses but also remain organically of politics in which subordinated social groups emerge from the periphery of society, overcome the , an increasing number of environmental movements emerge from the rural sector of China, which (firms, workers, peasants, state operatives) that emerge from the social relations of production (20 possible artists and possible works of art emerge from the "struggle for a new moral ohibited without permission. ♠When the subaltern emerge from their subordinate position and achieve the military, lest threats to the government emerge from within its ranks.10 50% of seats in

so certain that a movement will not emerge in China. Conclusion In conclusion, if th opportunities for an environmental movement to emerge in China have changed over the last that one or more movements would not emerge in China. However, there are other opportu to indicate an environmental movement would not emerge in China. However, we must take a anything but a state-sponsored social movement emerge in China? Public Participation, Political integral state, many key concepts begin to emerge in his prison notebooks, including "hegemon , etc.). Why do no writers like Flammarion8 emerge in Italy? Why has no popularized 7. , Culture and Anthropology As I hope will emerge in Part II, it is precisely this permitted new forms of everyday practice to emerge. In providing creative entry points, again a counter-hegemonic conception is likely to emerge in the course of struggle: it 'manifests figure whom I hope has begun to emerge in the course of the preceding chapters. ion in which the latter would necessarily emerge in the end as the leading class.14 allowing this form of autonomous power to emerge. In the mid 1990s, legislation was enacte idea which it codified first started to emerge in the writings of Plekhanov in 1883- by the political context within which they emerge, in turn influencing their role in social

# Clusters Outcomes

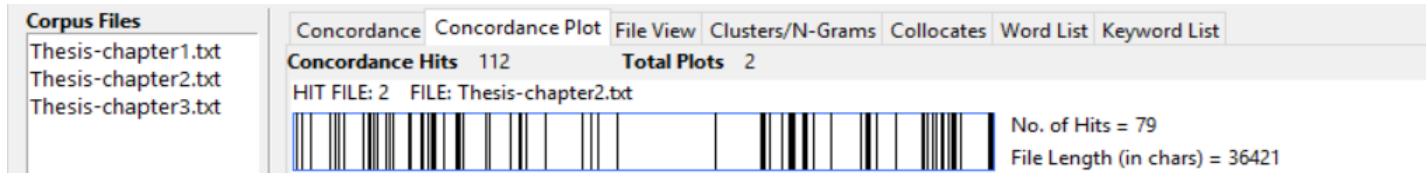
## Liu's conclusions

- *Emerge in + Place / Time*
- *Emerge from + something (a deed)*
- *Better use 'emerge from' in my context (although there are more hits for 'emerge in')*
- **Liu's Text:** *Echoing [with] the call for community level self-governance and streamline administration, numerous agencies **emerge from** the grassroots level...*

## Why use the Clusters Tool?

- check and/or discover **collocations, phraseology**
- get a **quick picture** of recurrent, disciplinary phraseology
- **cut down** on concordancing and sorting

# Concordance Plot in Use



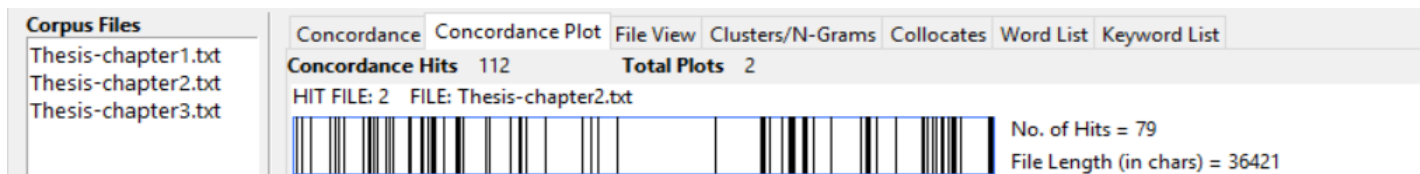
- provides a ***graphic display***
- shows where the search item occurs ***within a file***
- gives a ***simultaneous overview of all files*** in a corpus

**Naoko:** Japanese doctoral student in medicine

**Corpus:** 3 thesis chapters; 10,678 words

**Thesis:** The role of A1-MGBv **corticothalamic** feedback explored with **mistuning** of harmonic complex tones using ferrets

# Concordance Plot in Use



- provides a **graphic display**
- shows where the search item occurs **within a file**
- gives a **simultaneous overview of all files** in a corpus

**Naoko:** Japanese doctoral student in medicine

**Corpus:** 3 thesis chapters; 10,678 words

**Thesis:** The role of A1-MGBv **corticott**  
explored with **mistuning** of harmonic cor

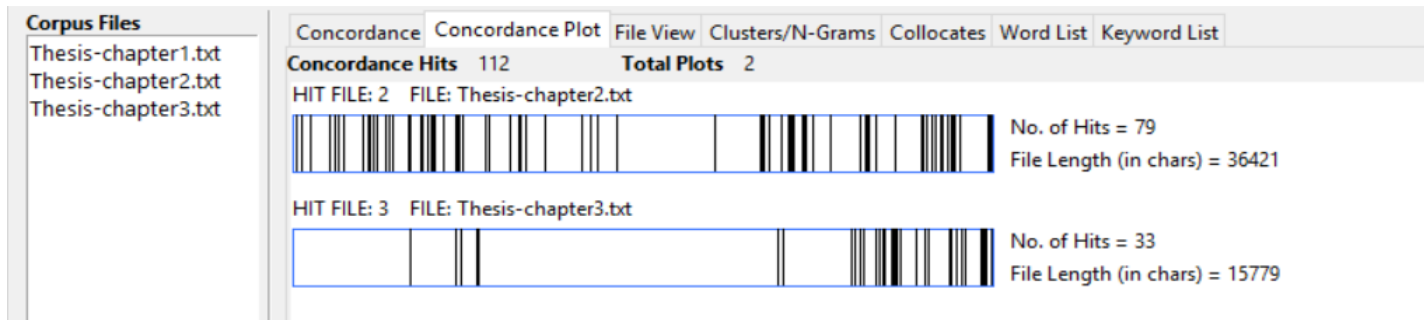
**Naoko's Question:** ○ ○ ○

Checking use of  
key terms in the  
chapters

# Naoko's Investigation 1

## Naoko's Investigation 1

Search on *mistun*\* in her thesis corpus of 3 chapters



## Naoko's Conclusion

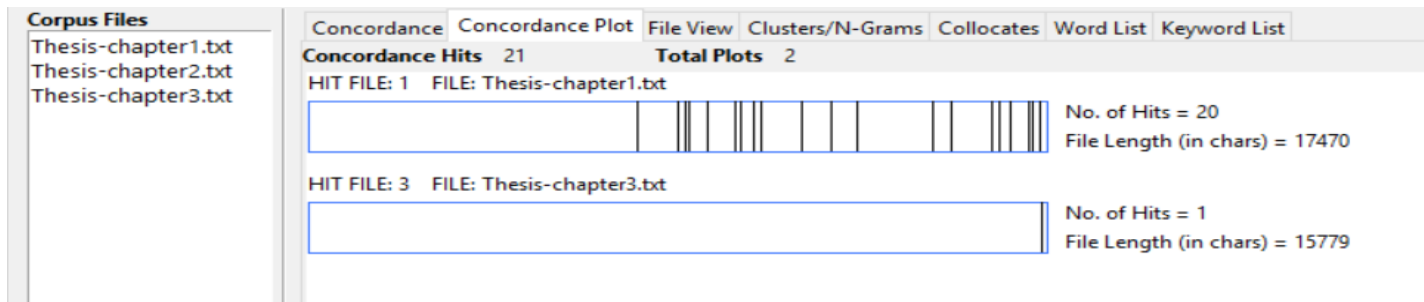
Investigation 1 *mistun*\*: 79 hits in chapter 2; 33 hits in chapter 3

I started thinking I should mention mistuning (mistuned) in chapter 1 (general introduction) because this is my main theme and keyword through my thesis.

# Naoko's Investigation 2: Plot Outcomes

## Naoko's Investigation 2

*Search on corticothalamic* in her thesis corpus of 3 chapters



## Naoko's Conclusion

*Investigation 2 corticothalamic*: 20 hits in chapter 1; 1 hit in chapter 3

This is fine because I'm going to explore the role of corticothalamic feedback system more in chapter 4 and 5.

# Naoko's Investigation 2: Plot Outcomes

## Naoko's Investigation 2

*Search on corticothalamic* in her thesis corpus of 3 chapters

### **Why use Concordance Plot?**

- **To track content, ideas, terms, references**
- **To check the balance or development of content**
- **To compare usage across multiple files**

This is fine because I'm going to explore the role of corticothalamic feedback system more in chapter 4 and 5.

# Keywords in Use: Ha-Yoon

intro.txt	Keyword Types: 727			Keyword Tokens: 2675	
	Rank	Freq	Keyness	Keyword	
	1	11	+ 11.27	thesis	
	2	30	+ 9.94	institutions	
	3	22	+ 8.55	i	

- Identifies words which are ***unusually frequent*** or ***infrequent*** in one corpus when compared to a reference corpus
- Gives insight into the ***content of individual chapters*** compared to the whole thesis

11	5	+ 5.43	korpi
12	29	+ 5.43	this

**Ha-Yoon:** Korean doctoral student in politics

**Corpus:** 3 thesis chapters; 14,397 words

**Thesis:** The political economy of inequality in developing countries: trajectories from industrialization to inequality

# Keywords in Use: Ha-Yoon

intro.txt	Keyword Types: 727			Keyword Tokens: 2675	
	Rank	Freq	Keyness	Keyword	
	1	11	+ 11.27	thesis	
	2	30	+ 9.94	institutions	
	3	22	+ 8.55	i	

- Identifies words which are ***unusually frequent*** or ***infrequent*** in one corpus when compared to a reference corpus
- Gives insight into the ***content of individual chapters*** compared to the whole thesis

11	5	+ 5.43	korpi
12	29	+ 5.43	this

**Ha-Yoon:** Korean doctoral student in politics

**Corpus:** 3 thesis chapters; 14,397 words

**Thesis:** The political economy of inequal trajectories from industrialization to

**Ha-yoon's Question:**

How does the introduction compare with the other chapters?

# Ha-Yoon's Introduction Keywords

## Positive Keywords

AntConc 3.5.6 (Windows) 2018

File Global Settings Tool Preferences Help

Corpus Files  
intro.txt

Concordance Concordance Plot File View Clusters/N-Grams

Keyword Types: 727 Keyword Tokens: 2675

Rank	Freq	Keyness	Keyword
1	11	+ 11.27	thesis
30		+ 9.94	institutions
22		+ 8.55	i
8		+ 8.35	late
9		+ 8.04	variations
9		+ 6.73	regions
6		+ 6.69	industrialized
21		+ 6.44	institutional
8		+ 6.19	my
6		+ 5.85	origins
11	5	+ 5.43	korpi
12	29	+ 5.43	this
13	16	+ 5.32	developing
14	12	+ 5.17	two

Search Term ☒ Words ☐ Case ☐ Regex

Start Stop Sort Advanced

Sort by ☐ Invert Order

Sort by Keyness

Total No. 1

Files Processed

Positive keywords:  
give aim  
and broad  
ideas

## Negative Keywords

AntConc 3.5.6 (Windows) 2018

File Global Settings Tool Preferences Help

Corpus Files  
intro.txt

Concordance Concordance Plot File View Clusters/N-Grams

Keyword Types: 727 Keyword Tokens: 2675

Rank	Freq	Keyness	Keyword
588	1	+ 0	used
589	4	+ 0	which
0	1	- 7.95	insiders
1	8	- 4.77	for
2	1	- 4.56	model
3	1	- 4.02	open
4	2	- 3.79	under
2		- 3.79	would
5	1	- 3.75	job
7	20	- 3.45	is
588	1	- 2.98	growth
599	1	- 2.98	outsiders
600	10	- 2.88	labor
601	8	- 2.74	market

Search Term ☒ Words ☐ Case ☐ Regex

Start Stop Sort Advanced

Sort by ☐ Invert Order

Sort by Keyness

Total No. 1

Files Processed

Negative keywords:  
give detail

# Keywords Outcomes

## Ha-Yoon's conclusions: Chapter 1 Introduction

Positive keywords: *thesis, institutions, I, late, variations.*

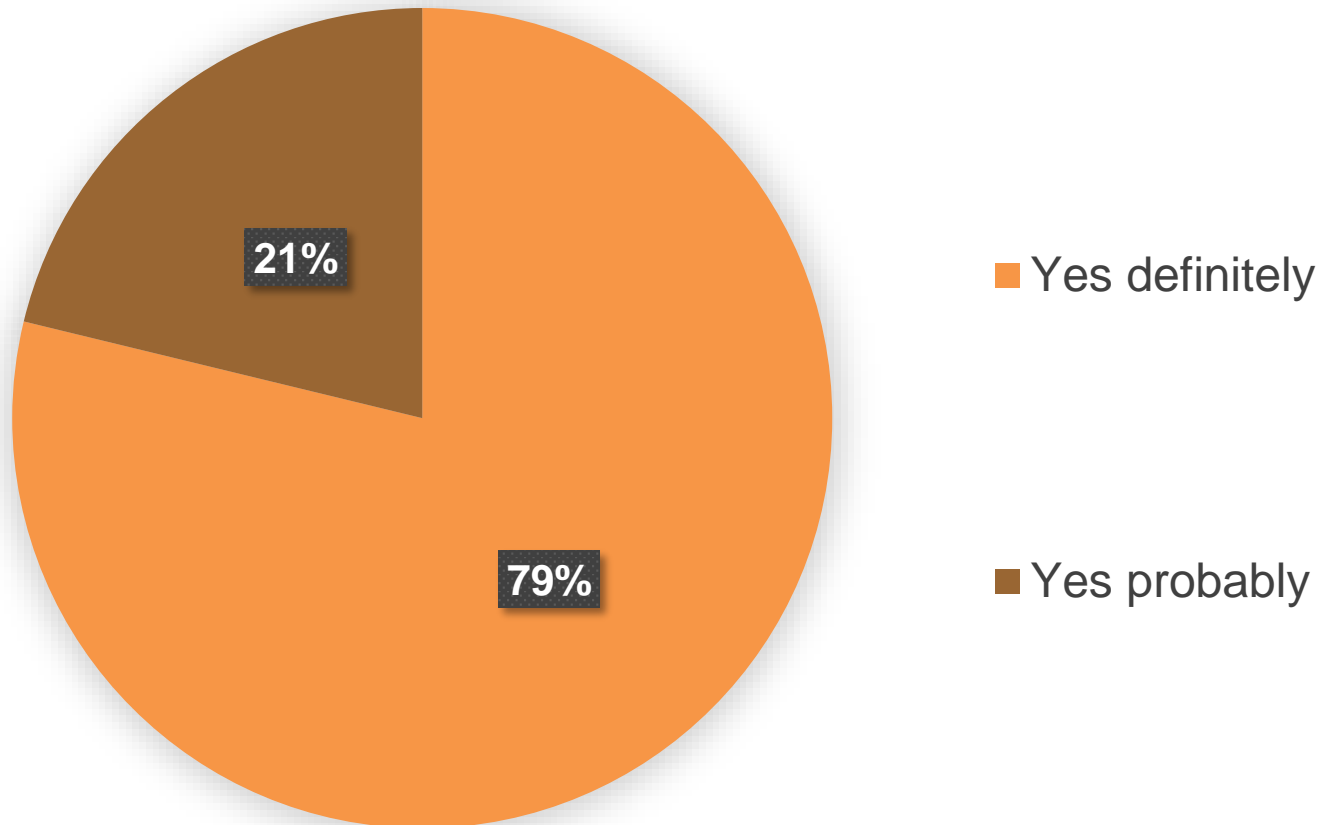
Negative keywords: *insiders, model, open, job, outsiders.*

- *'I' is positive keyword. I am considering not to use 'I' too frequently in my intro chapter.*
- *The intro chapter introduces the aim of my 'thesis' and my (I) intention'.*
- *The positive keywords show the broad ideas of my thesis, and the negative words explain detailed logic and contents. These detailed words are not explained at the beginning, and I think it's fine.*

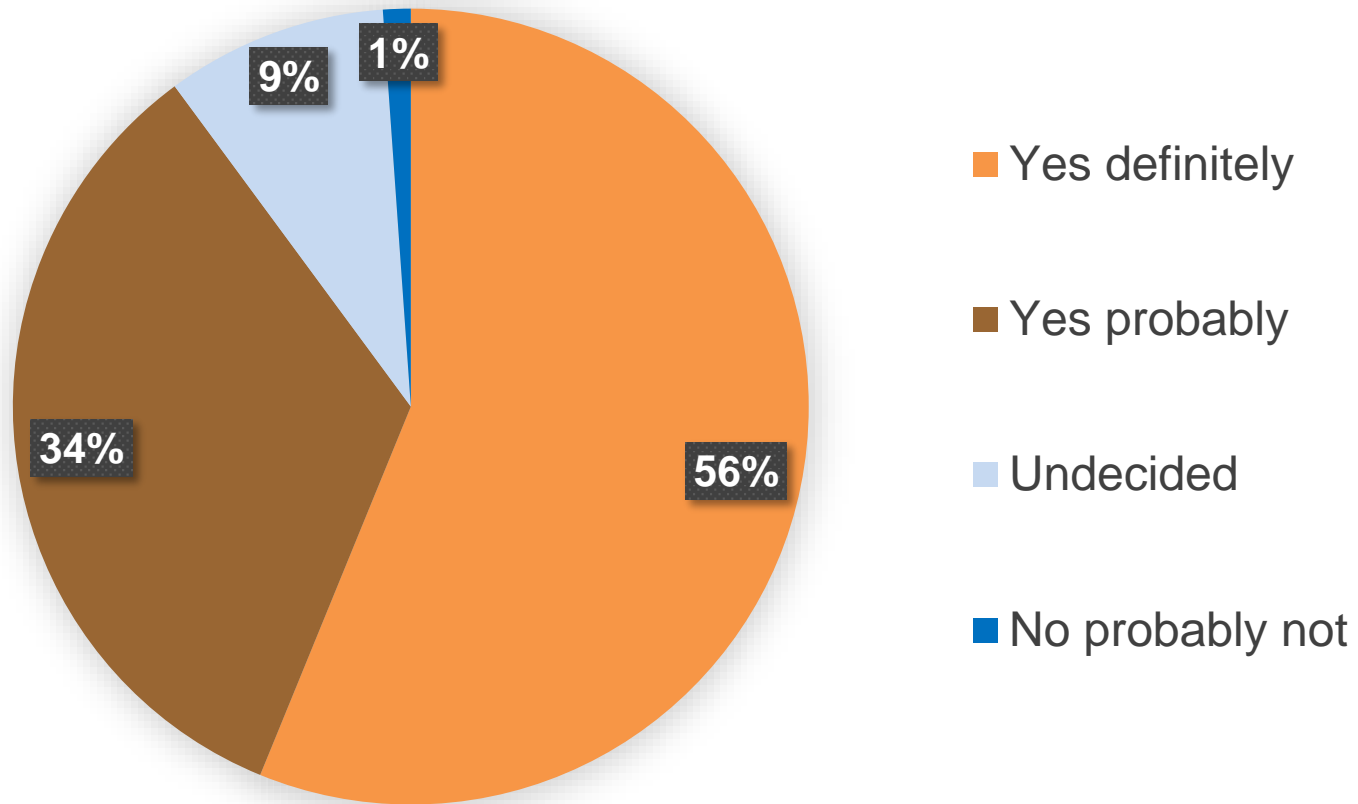
## Why use the Keywords Tool?

- to allow ***aspects of content*** to emerge
- to ***identify content issues*** the student is not aware of

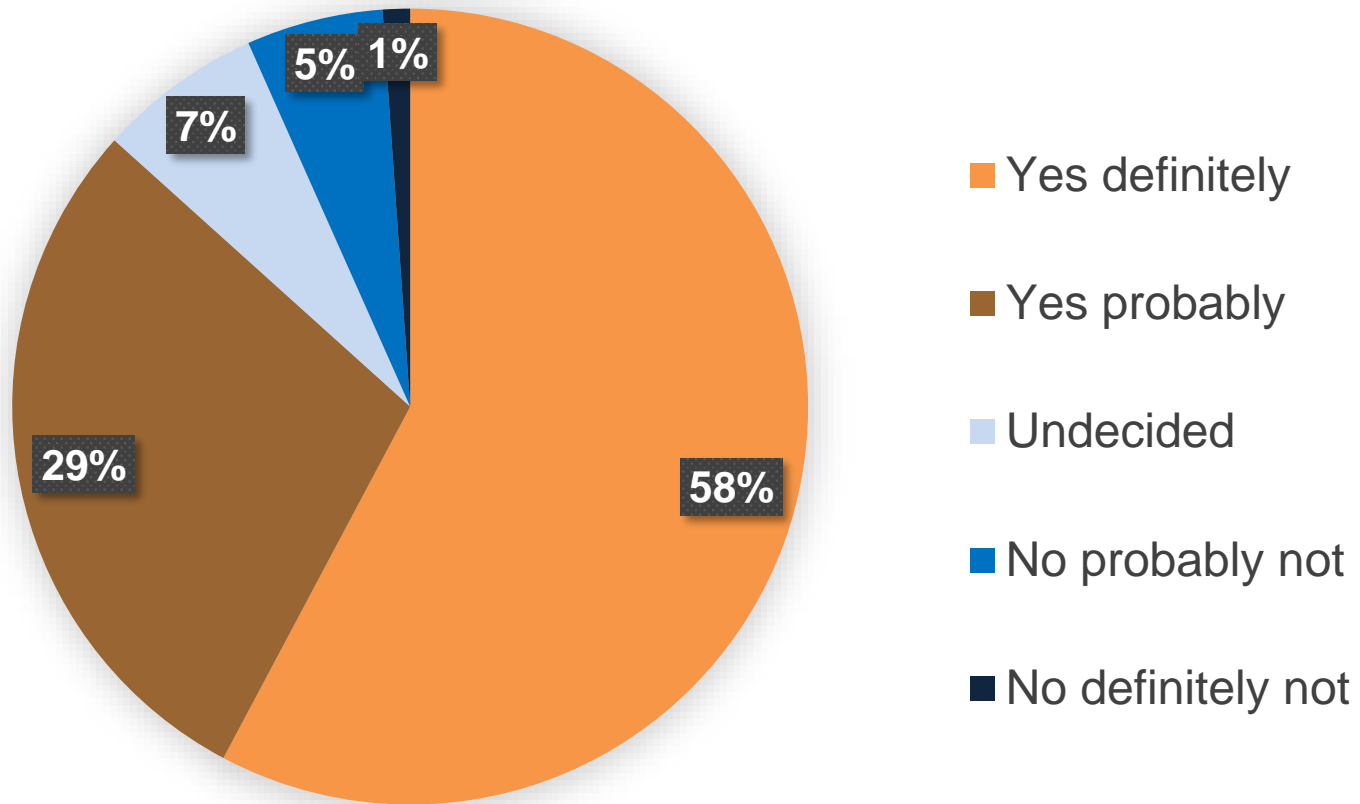
# Is it helpful to use your corpus and AntConc for editing?



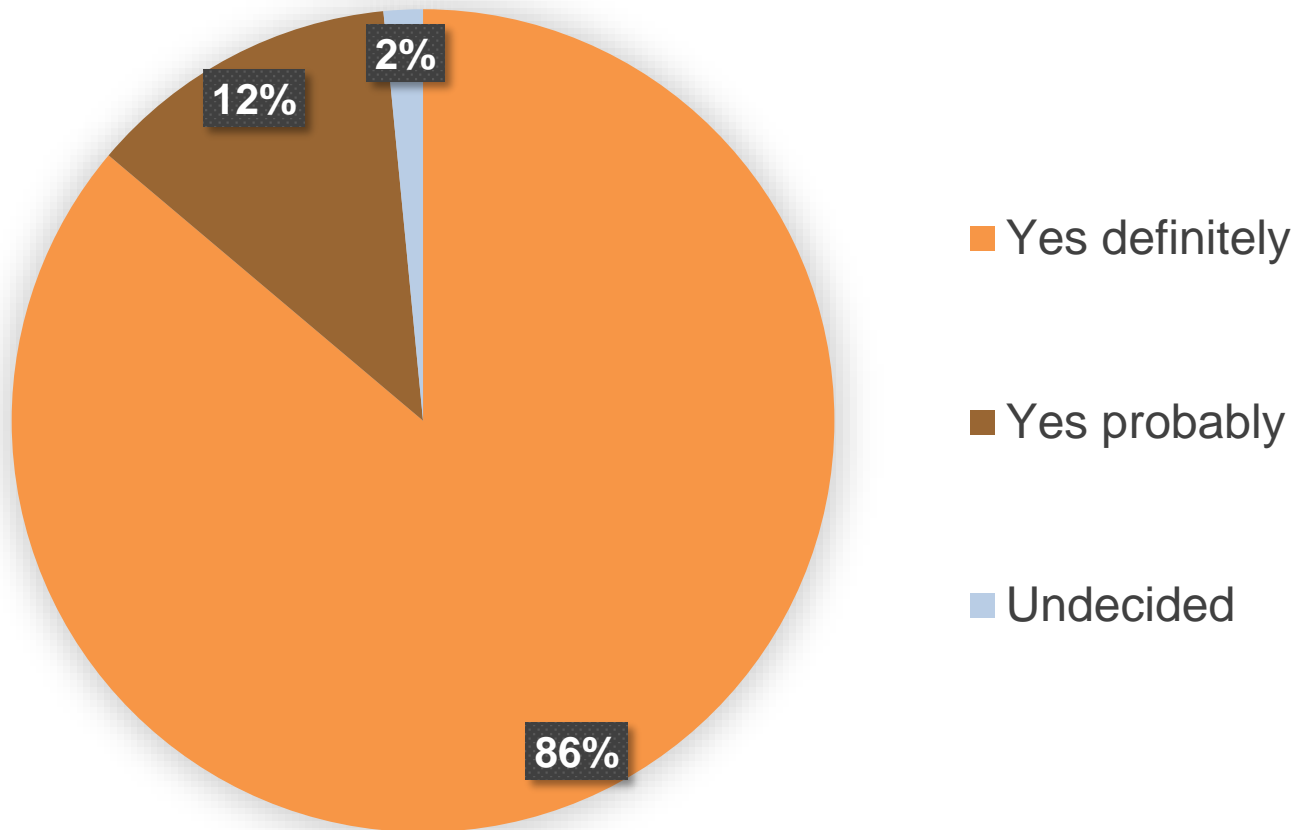
# Is it easy to use the AntConc software?



# Was it easy to build your corpus?



# Do you intend to use your corpus and AntConc for editing in the future?



# Affordances of Corpus Tools for Proofreading/Editing

- show *issues in language, content* and *organisation* that are not evident when reading linearly
- enable *comparisons* of writing e.g. between thesis chapters or between student and expert texts
- *de-familiarise* an over-familiar text

# Affordances of Corpus Tools for Proofreading/Editing

- show *issues in language, content* and
- increase writers' **autonomy** and **control**
- provide a means of **exploration** and **learning**
- help writers to **shape** their texts **themselves**
- *de-familiarise* an over-familiar text



A bird's eye  
view from  
above

A bug's eye  
view from  
below



# References

- Anthony, L. (2018). AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University.  
Available at <http://www.laurenceanthony.net/>
- Anthony, L., (2017). AntFileConverter (1.2.1). [computer program] Tokyo, Japan: Waseda University.  
Available at: <http://www.laurenceanthony.net/>
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393.
- Burrough-Boenisch, J. (2013). Defining and describing editing. In V. Matarese (Ed.), *Supporting research writing: Roles and challenges in multilingual settings* (pp. 141–155). Amsterdam: Elsevier Science.
- Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes*, 35, 30–40.
- Charles, M. (2017). Do-it-yourself corpora in the classroom: Views of students and teachers. In K. Hyland & L. Wong, (Eds.), *Faces of English education: Students, teachers and pedagogy* (pp. 107–123). Abingdon: Routledge.
- Charles, M. (2018) Corpus-assisted editing: More than just concordancing. *Journal of English for Academic Purposes*, 36, 15-25.
- Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56–75.
- Lillis, T., & Curry, M. J. (2010). *Academic Writing in a Global Context*. London: Routledge.
- Thompson, P. (2013). Thesis and dissertation writing. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (pp. 283–299). Oxford: Wiley-Blackwell.