‘Definitely [sic] worth it’: engaging undergraduate corpus linguistics students in real research

John Williams
Senior Lecturer in English Language and Linguistics

john.x.williams@port.ac.uk  @lexicoj0hn
Don't let the past, make your decisions, four today
AIMS

• To share best [?] practice in the teaching of corpus linguistics

• To suggest ways of enabling students to carry out meaningful research using CL methods

• To introduce my and my students' research into 'non-standard' spellings
THE MODULE

• Researching English Vocabulary: final year optional UG module in lexicology and CL

• 'Designated Research Module' --> students are required to carry out a research-based assessment

• Students choose their own topic for the main assessment --> 3000-word research report
THE ISSUES

• Only 24 class contact hours: 12 lectures + 12 workshop classes

• Tendency for workshop classes to turn into 'how to' sessions (cf. Baker, 2009, p.75)

• Students often choose unsuitable topics (cf. Cheng, Warren & Xu, 2003, p.182)
SOME UNSUITABLE TOPICS

• "A study of conceptual metaphors and its [sic] usage in today's society"

• "I have been thinking about looking at a particular idiom such as 'every cloud has a silver lining' and then looking at different corpora and using CQL, N-grams etc to find variations and see if the use has changed over time" (slightly edited)

• The use of the word “bitch” in corpus linguistics
Some More Suitable Topics

• A comparison of the use of thin, slim, & skinny in the BNC and on 'pro-anorexia' web forums

• The prevalence of I'm good as a response to How are you?

• The lexicalization of vlog over the period 2014-2017
SOLUTIONS

• Introduce research concepts early in the module: research questions, hypothesis testing, replicability...

• Workshop activities should reinforce those concepts, eg. replicating Aitchison's (2004) analysis of disaster

• Involve students in an actual research programme ...
Speaking radicly for the moment, but no doubt academicly, logicly and statisticly soundly (though tragically for traditionalists, I know), I basicly think publically is a better spelling.
Comparing *publicly* and *publically*: AntConc Keywords

**Method** (cf. Taylor, 2010)
- Make a SketchEngine concordance for *publically*, using the enTenTen13 corpus (20548 lines)
- Download and save this as a text file (Publically13.txt)
- Make an equivalent concordance for *publicly* and use the Sample function to reduce this to 20548 lines
- Download and save this as a text file (Publicly13.txt)
- Treat these text files as corpora in AntConc; open one of them then use the other one as a reference corpus to generate keywords
COMPARING publicly AND publically: WORD LIST FOR PUBLICLY

Tokens: 324585
Types: 27751
Type /Token Ratio: 0.09
Comparing *publicly* and *publically*: **AntConc Keywords**

Using `Publicly13.txt` as reference corpus

Tokens: 327743
Types: 28717
Type /Token Ratio: 0.09
HYPOTHESIS

• Non-standard spellings of lexical words will be associated disproportionately with the most frequent lexical collocate of the corresponding standard form.

• *This is what students were asked to test in the workshop class.*
Top Ten 'Misspelled' Words in the Oxford Dictionary Corpus

1. publically
2. pharoah
3. definately
4. goverment
5. separate
6. occurred
7. untill
8. recieve
9. wich
10. accomodate

Source: https://blog.oxforddictionaries.com/2016/08/02/corpus-misspellings/?_prclt=FCuXsFJn
STUDENT FINDINGS FROM THE WORKSHOP

• 7/12 students present, 3 pairs + 1 individual: worked on *government, recieve, could of* (+ other modal verbs), *definitely*

• The keyword comparison of *government* and *goverment* does not support the hypothesis. Non-standard spelling does not feature highly in names of institutions, etc. Tends to be more associated with financial schemes, *goverment credit*, etc.

• The non-standard spelling *recieve* is more associated with letters, parcels, etc, less so with awards, grants, etc. There is a spike in the frequency of *recieve* in the post-WW2 era.
Final Assessments

- Only 3/23 students chose a non-standard spelling for their final assessment: 2 x *definitely* + 1 x *could/would/should of*, etc
- Only 1 student explicitly tested the hypothesis, and all 3 made mistakes in applying the recommended methodology.
- Nevertheless all 3 had some merit and asked some pertinent questions:

  However, in most corpora, the “of”-form still appears significantly less than the “‘ve”-form, so how come this annoys so many people so much and why does it stick out so much to us?
STUDENT FEEDBACK (n = 6)

• Post-workshop feedback sheet:

1. The most interesting/useful/enjoyable aspect of this activity was...
2. The least interesting/useful/enjoyable aspect of this activity was...

Indicate your level of agreement with the following two statements (Disagree strongly = 0; Agree Strongly = 5)

3. This activity enhanced my understanding of the research process in lexicology / corpus linguistics. (mean = 3.5)
4. This activity will help me in tackling my main assignment (or future assignments such as my dissertation). (mean = 3.17)
5. Free comment
FREE COMMENTS

• "Real research opportunities are great to practice and develop our skills, so it has been useful."

• "Doing real research that actually served a purpose. It was fun!"

• "Definately worth it!"
REFERENCES


