



UNIVERSITY OF  
PORTSMOUTH

TALC 2018

# ‘Definately [*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](mailto:john.x.williams@port.ac.uk)



[@lexicoj0hn](https://twitter.com/lexicoj0hn)

r vehicles it is **definately** **worth** it. T  
eing a tutor is **definately** **worth** it! T  
rilliant work is **definately** **worth** it. T  
urs but it was **definately** **worth** it. V  
s alot! But it is **definately** **worth** it if  
e end result is **definately** **worth** it. L  
melands, it is **definately** **worth** it to  
eak times, it is **definately** **worth** it as  
ly are usually **definately** **worth** it!...  
no scary bits - **definately** **worth** it or  
ob! Oh, it was **definately** **worth** it...



# 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 ...

## SPELLING VARIATIONS AND NON-STANDARD SPELLINGS



*Speaking radicly for the moment, but no doubt academicly, logicly and statisticly soundly (though tragicly 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*

AntConc 3.2.4w (Windows) 2011

File Global Settings Tool Preferences About

Corpus Files  
Publicly13.txt

Concordance Concordance Plot File View Clusters Collocates **Word List**

Hits Total No. of Word Types: 27751 Total No. of Word Tokens: 324585

Rank	Freq	Word	Lemma Word Form(s)
1	20549	doc	
2	20294	publicly	
3	12886	the	
4	8634	to	
5	7625	and	
6	6409	of	
7	4954	a	
8	4128	that	
9	3870	in	
10	3242	is	
11	2778	for	
12	2220	on	
13	2141	be	
14	1987	not	
15	1980	it	
16	1836	s	
17	1827	available	
18	1822	are	
19	1740	I	
20	1718	as	

Tokens: 324585  
Types: 27751  
Type /Token Ratio:  
0.09

**publicly** (*adverb*) Alternative PoS: noun (1,789)  
English Web 2013 (enTenTen13) freq = 340,908 (14.99 per million)

adjectives modified by "publicly"			verbs modified by "publicly"		
		13.32			87.88
accessible +	<u>6.759</u>	9.80	trade +	<u>21.320</u>	10.65
publicly accessible			publicly traded		
funded +	<u>957</u>	9.30	fund +	<u>9.556</u>	9.49
publicly funded			publicly funded		
available +	<u>27.160</u>	9.24	state +	<u>9.537</u>	8.67

# COMPARING *publicly* AND *publically*: ANTCONC KEYWORDS

Rank	Freq	Keyness	Keyword
1	20143	27729.568	publically
2	468	631.362	Publically
3	2396	98.232	I
4	47	64.702	PUBLICALLY
5	2254	40.721	available
6	1548	36.579	you
7	792	32.524	your
8	485	31.591	like
9	842	30.462	can
10	528	26.815	our
11	25	22.949	summaries
12	14	19.273	accessable
13	14	19.273	superlatively
14	21	18.111	sic
15	515	16.420	do
16	2260	15.892	it
17	215	14.787	You
18	500	14.712	data
19	2418	14.265	be
20	1232	14.107	will

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. definatly
4. goverment
5. seperate
6. occured
7. untill
8. recieve
9. wich
10. accomodate

Source: [https://blog.oxforddictionaries.com/2016/08/02/corpus-misspellings/?\\_prclt=FCuXsFJn](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 *goverment, recieve, could of* (+ other modal verbs), *definatly*
- 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 *definatly* + 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!"*
- *" Definatly worth it ! "*

# REFERENCES

- Aitchison, J. (2004). *Language Change: Progress or Decay?* (3rd ed.). Cambridge: CUP
- Baker, P. (2009). Issues arising when when teaching corpus-assisted (Critical) Discourse Analysis. In L. Lombardo (Ed.), *Using Corpora to Learn about Language and Discourse* (pp.73-96). Bern, Switzerland: Peter Lang
- Cheng, W., Warren, M. & Xu, X. (2003). The language learner as language researcher: putting corpus linguistics on the timetable. *System*, 31, 173-186
- Fligelstone, S. (1993). Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME Journal*, 17 (2), 97-109
- Oxford Dictionaries (n.d.). *Top ten misspelled words in our corpus*. Retrieved from the Oxford Dictionaries blog: [https://blog.oxforddictionaries.com/2016/08/02/corpus-misspellings/?\\_\\_prclt=FCuXsFJn](https://blog.oxforddictionaries.com/2016/08/02/corpus-misspellings/?__prclt=FCuXsFJn)
- Taylor, C. (2010). Science in the news: a diachronic perspective. *Corpora*, 5(2), 221-250. doi:10.3366/cor.2010.0106
- Williams, J. (2012a). *Is there a case for 'publically'? Part 1*. Retrieved from the Macmillan Dictionary Blog: <http://www.macmillandictionaryblog.com/is-there-a-case-for-publically-part-1>
- Williams, J. (2012b). *Is there a case for 'publically' (or 'economicly')? Part 2*. Retrieved from the Macmillan Dictionary Blog: <http://www.macmillandictionaryblog.com/is-there-a-case-for-publically-or-economicly-part-2>