British English swearing in corpora of casual conversation

Dr Robbie Love
Lecturer in English Language
Aston University

r.love@aston.ac.uk
@lovermob

Spoken British National Corpora

- Transcriptions of recorded conversations: 1990s and 2010s

- **Spoken BNC1994** (BNC Consortium, 2007)
  - c. 5 million words casual conversation

- **Spoken BNC2014** (Love et al., 2017)
  - c. 11 million words casual conversation

McEnery et al. (1999, 2000); McEnery & Xiao (2004); McEnery (2006)

Love (2021)

- Comparing swear word frequency among all speakers in conversational components of the BNC1994 and BNC2014

- Focus specifically on 'pure' swear words (and not also terms of abuse)

  – Comparing the ranking of individual swear word frequency in both corpora
  – Quantifying the usage of swear words across age groups (apparent time) and between corpora (real time)

Stenström & Love (in review)

- Comparing functions of *fuck* among London teenage speakers in BNC1994 and southeast England teenage speakers in BNC2014
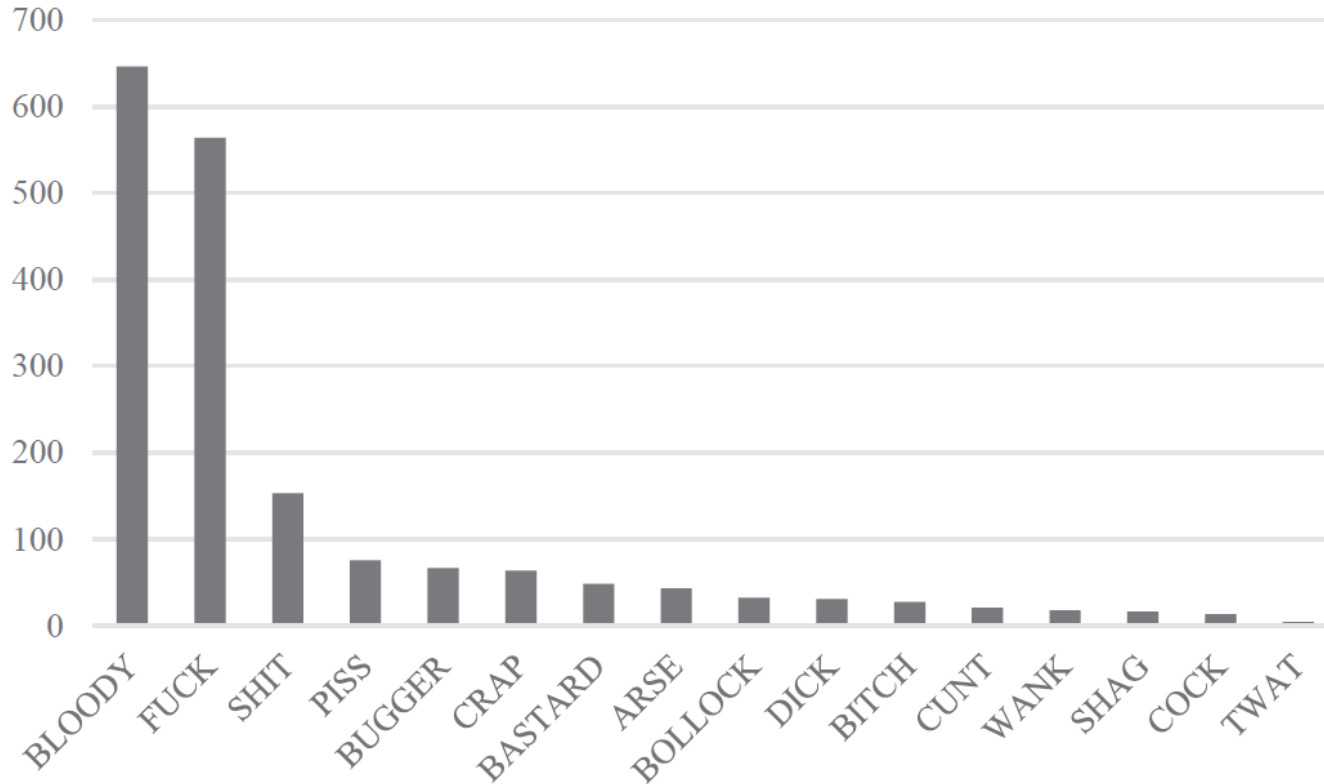
ARSE
BASTARD
BITCH
BLOODY
BOLLOCK
BUGGER
COCK
CRAP
CUNT
DICK
FUCK
PISS
SHAG
SHIT
TWAT
WANK

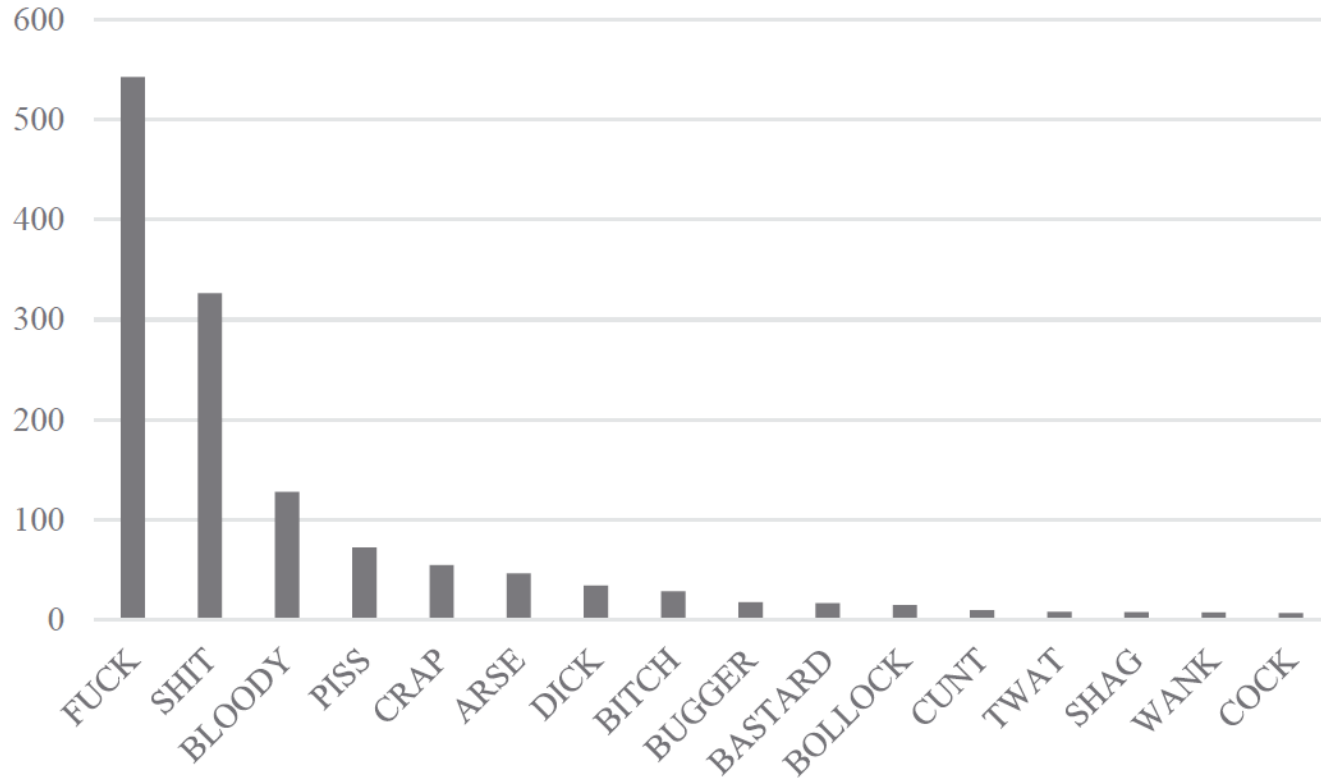**ARSE, BASTARD, BITCH, BLOODY, BOLLOCK, BUGGER, COCK, CRAP, CUNT, DICK, FUCK, PISS, SHAG, SHIT, TWAT, WANK**

- Frequency

  – BNC1994**:** 2,265 per million (**0.23%**)

  – BNC2014: 1,428 per million (**0.14%**)

- Most British English conversational swearing is performed using three words:
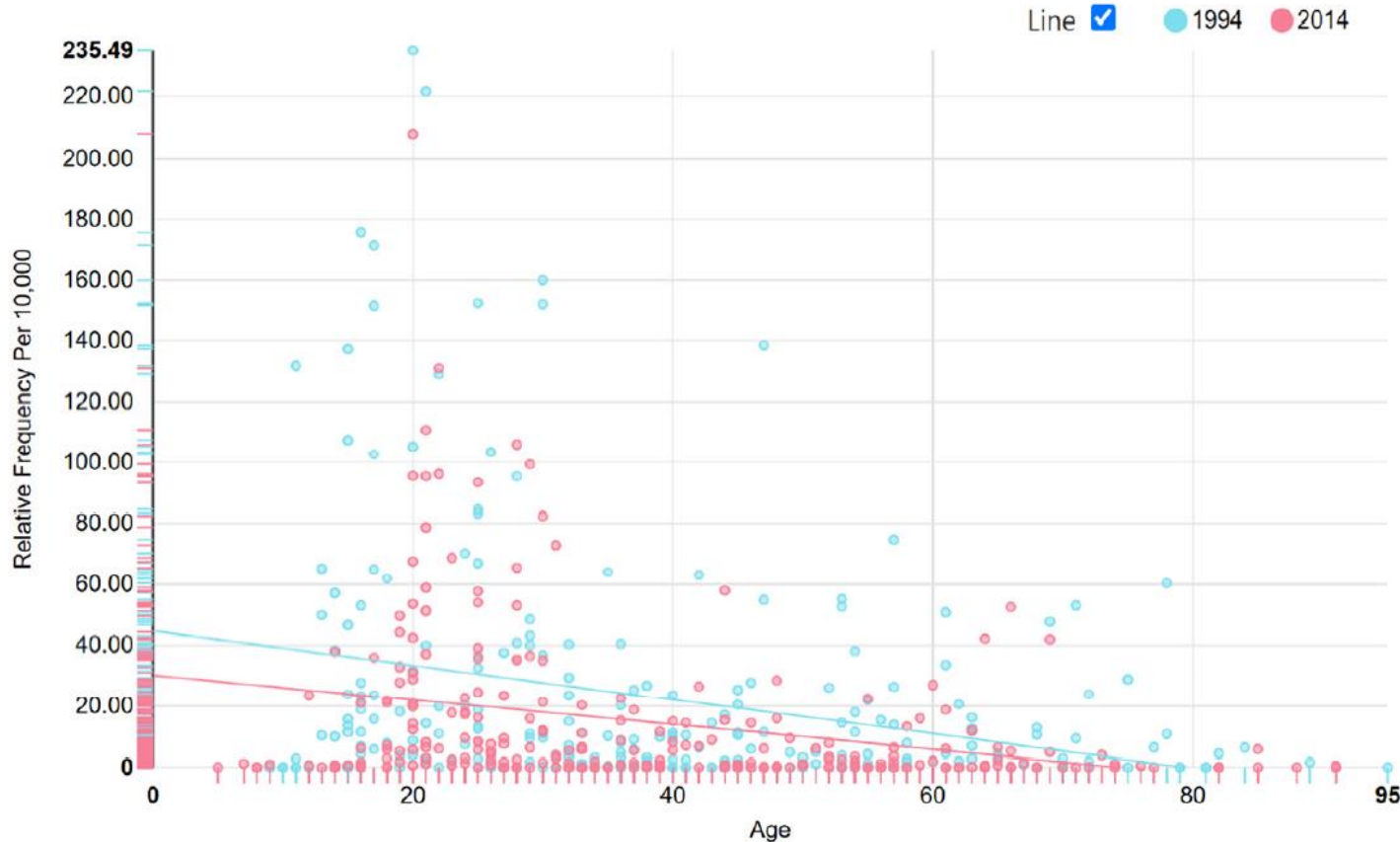
FUCK, SHIT and BLOODY

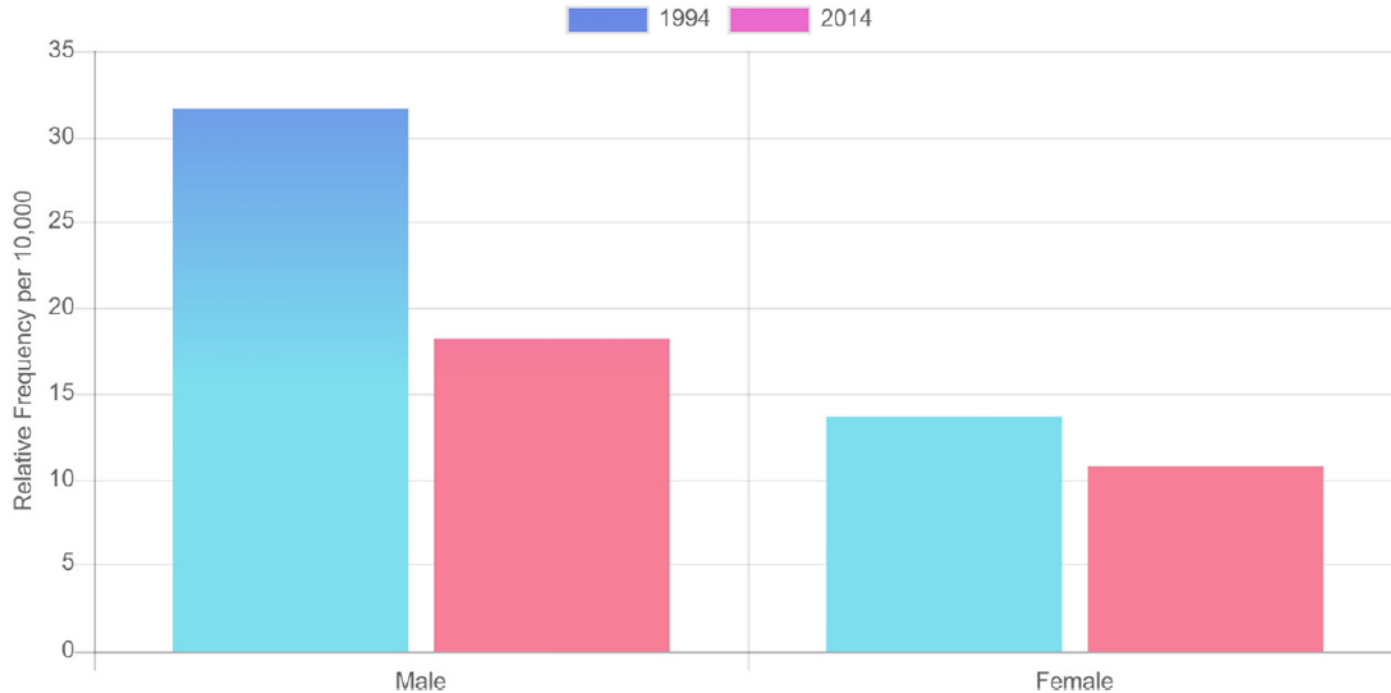# Ranking swear words (1990s)

Ranking swear words (2010s)

Swearing and age

ARSE
BASTARD
BITCH
BLOODY
BOLLOCK
BUGGER
COCK
CRAP
CUNT
DICK
FUCK
PISS
SHAG
SHIT
TWAT
WANK

Swearing and gender

ARSE
BASTARD
BITCH
BLOODY
BOLLOCK
BUGGER
COCK
CRAP
CUNT
DICK
FUCK
PISS
SHAG
SHIT
TWAT
WANK

- The most common swear word is **FUCK**

- Stenström & Love (in review) – investigating *fuck* among London teenagers in the BNC1994 and BNC2014

  - COLT: a corpus comprising London teenage speakers (and part of BNC1994)

    - 83 speakers, c. 600,000 tokens

  - Sub-corpus comprising southeast teenage speakers from BNC2014

    - 15 speakers, c. 300,000 tokens

- Word class

- Swearing function (McEnery & Xiao, 2004)

Functions of *fuck* (based on McEnery & Xiao, 2004)

Category	Criteria	Example(s)
G	General expletive	*(oh) fuck*
P	Personal insult referring to defined entity	*you fuck / that fuck*
C	Cursing expletive	*fuck you / me / him / it*
D	Destinational usage	*fuck off / he fucked off*
L	Literal usage denoting taboo referent	*he fucked her*
E	Emphatic intensifier	*fucking marvellous / in the fucking car*
O	'Pronominal' form	*like fuck / fat as fuck*
I	Idiomatic 'set phrase'	*fuck all / give a fuck / thank fuck*
A	Predicative negative adjective	*this game is fucked*
X	Metalinguistic or unclassifiable due to insufficient context	The use of the word "fuck" / you never fucking

Frequency and form

- COLT: 727 instances of FUCK (652 pmw)

- BNC2014 sub-corpus: 171 instances of FUCK (547 pmw)

- Most swearing by FUCK is performed by three forms: *fucking*, *fuck* and *fucked*

  – This aligns with the findings of McEnery & Xiao (2004: 258) for all BNC1994 speakers

Word form	COLT		Spoken BNC2014 sub-corpus	
	Frequency	%	Frequency	%
fucking	395	54.33	60	35.09
fuck	239	32.87	88	51.46
fucked	62	8.53	20	11.70

Word class

- Significant difference in distribution of word class categories between corpora

- Biggest difference: FUCK as noun (9.6% → 24%)
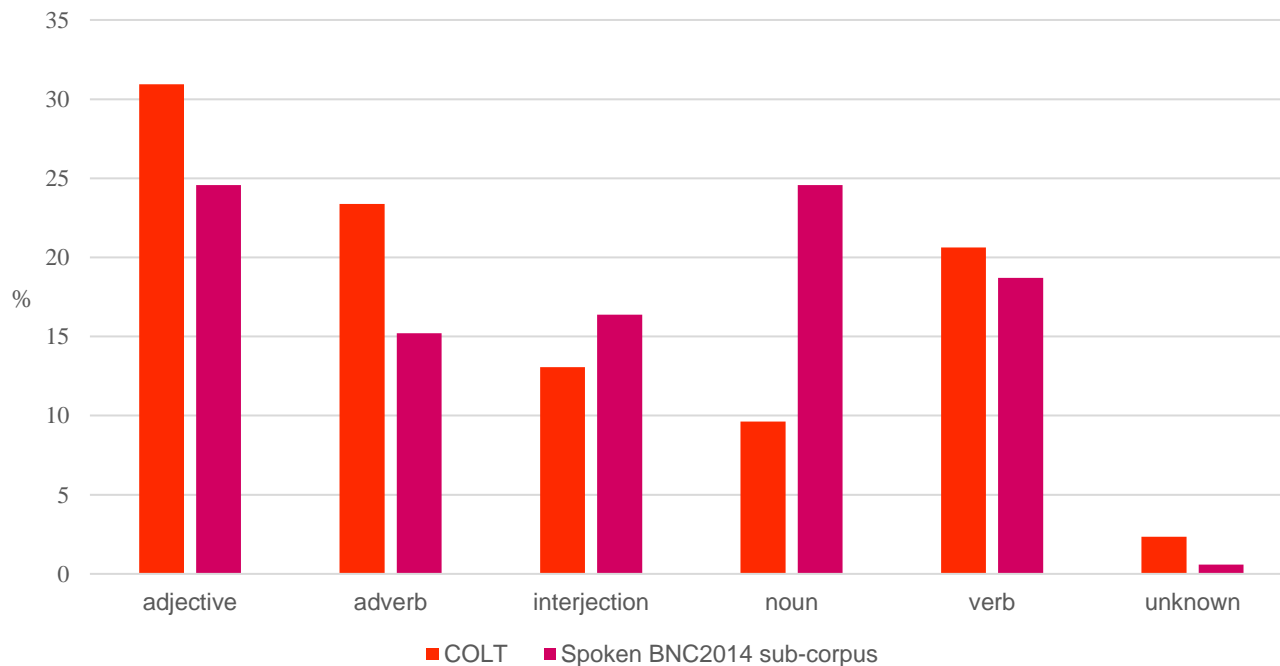
Swearing function

Code	Description	COLT		Spoken BNC2014 teenage sub-corpus		% diff
		Frequency	%	Frequency	%	
A	Predicative negative adjective	29	3.99	12	7.02	3.03
C	Cursing expletive	17	2.34	6	3.51	1.17
D	Destinational usage	59	8.12	5	2.92	-5.19
E	Emphatic intensifier	359	49.38	54	31.58	-17.80
G	General expletive	90	12.38	28	16.37	3.99
I	Idiomatic 'set phrase'	86	11.83	59	34.50	22.67
L	Literal usage denoting taboo referent	40	5.50	4	2.34	-3.16
O	'Pronominal' form	1	0.14	1	0.58	0.45
P	Personal insult referring to defined entity	23	3.16	1	0.58	-2.58
X	Metalinguistic or unclassifiable due to insufficient context	23	3.16	1	0.58	-2.58
	Total	727	100	171	100	

- **A common category in both: emphatic intensifiers**

  – *I can't be **fucking** bothered anymore* (BNC1994)

  – *my teacher was a weird **fucking** psycho who fed us conspiracy theories* (BNC2014)

- **Strong negative semantic prosody**, but less explicit in BNC2014, e.g. as an adjective, pre-modifying other taboo words:

  – COLT: *arse** (4)*, bastard* (4), bitch* (7)*, bollocks* (1)*, cunt* (3)*, dick* (2)*, gay boy* (1)*, lesy* (1)*, penis* (1)*, poofter* (1)*, shit* (4)*, slag* (1)*, tart* (1)*, wanker* (4)*, whore* (1)

  – BNC2014: *bitch* (1)*, idiot* (1)*, prick* (1)*, pussy* (1)*, queer* (1)

Swearing function

- **More idiomatic usage**

    – *did I fuck something up?* (BNC2014)

    – *Jesus that's like two Jager Bombs each which yeah will fuck you over* (BNC2014)

- In BNC2014 sub-corpus, 64% of idiomatic cases used FUCK as noun (up from 47%)

    – *we were just sitting here thinking okay what the fuck is going on?* (BNC2014)

    – *oh for fuck's sake when was that like four AM or something?* (BNC2014)

- **More general expletive usage**

    – *I mean that's the reason you should want to go to uni oh yeah fuck yeah for the course if you if you end up it's not for the drinking* (BNC2014)

- **Even less literal usage**

    – *You know I'm the one, who fucked your mum* (BNC1994)

    – *the beginning bit when she's fucking a man she sticks an axe through him* (BNC1994)

- Biggest development of FUCK is **increased usage in idiomatic (fixed)** expressions

  - The root form *fuck* "is most frequently used idiomatically, as in *what the fuck, for fuck's sake, give a fuck* and *fuck up/around/about*" (McEnery & Xiao, 2004: 258-9)

  - **Idiomatization** (as derived from **lexicalization**) (Brinton & Traugott, 2003)**:** "a semantic process which occurs within lexicalization and/or grammaticalization and causes the obscuration of meanings of combinations" (Rodriguez-Puente, 2012)

- High proportion of **expletive and intensifying functions**

  - **Subjectification:** "Meanings tend to become increasingly based in the speaker's subjective belief state/attitude toward the proposition" (Traugott, 1989)

  - **Semantic bleaching:** "the development from lexical to grammatical meaning as a "process of generalization or weakening of semantic content," whereby "meanings are emptied of their specificities" (Traugott, 1989)

Perception of the strength of *fuck* is declining in public opinion, although it remains relatively high

Millwood-Hargrave (2000)

- 71% rated *fuck* "very severe" + 22% "fairly severe"

Ofcom (2021)

- "**There was greater diversity in views towards *fuck*.** Although *fuck* was categorised variously as strong, moderate and mild by different groups of qualitative participants, they largely agreed, on a precautionary basis, that they did not want children to be exposed to this word."

- "**Older participants from the general groups were more likely to rate *fuck* as strong, while middle-aged participants consistently saw it as moderate. Younger participants held more mixed views, with different groups rating it from mild to strong.**"

- "Those who felt *fuck* was not strongly offensive suggested that it tended to be **used in a more general way** rather than targeting an individual or group"

- FUCK is highly frequent but still considered overall to be 'strong' (Ofcom, 2021) – an example of the **'swearing paradox':**

  – "how this highly offensive behaviour (according to ratings studies) can also enjoy such a high rate of occurrence (according to frequency studies)"

  (Beers-Fägersten, 2007: 16)

- Beers-Fägersten (2007) posits that the swearing paradox may be caused by deficiencies in the methods used to elicit public opinion on the strength of swearing

- Swearing strength is known to be highly context-dependent, even within a specific register like casual conversation "there is a strong and selective interaction between swearwords and macrostructures in discourse" (McEnery et al., 2023: 46)

Operationalising the functional coding scheme

- It was challenging to establish mutual exclusivity of function categories, as revealed by inter-rater reliability testing

- This was helped by noting **correspondence between word class and function categories**, as most functions were limited to specific word classes, e.g. emphatic intensifier = adjective or adverb

    - "Parts of speech are clearly important to the categorisation scheme, but the scheme itself is not simply a relabelling of parts of speech. Also, and interestingly, just because a particular word covered in the LCA has a part of speech connected with a category does not mean to say that the word will appear in that category." (McEnery, 2006: 28)

- For a few cases of one function category – idiomatic 'set phrase' – we could not establish full exclusivity from other categories, e.g.

    - *fucking hell* = idiomatic (I) **and** general expletive (G)?

    - *fuck off* = destinational (D) **and** idiomatic?

Sample size and comparability

- The sub-corpora are not a perfect match in terms of speaker region – more contemporary data from London teenagers would allow testing of observations

Thank you

@lovermob
https://robbielove.org/
r.love@aston.ac.uk