Global Trends in Democratisation Practices in Online Discourse

Martin Schweinberger

m.schweinberger@uq.edu.au

IAAK / RESEARCH PROJECTS / LANDPOSTS IN (DE-)DEMOCRATIZATION

LANDPOSTS IN (DE-)DEMOCRATIZATION

Interdisciplinary Workshop, 25 – 26 June 2025, University of Bonn

Martin Schweinberger

- Lecturer in Applied Linguistics at the University of Queensland (at UQ since 2018)
- PhD in English Linguistics at the University of Hamburg (2014)
- Quantitative Corpus Linguistics – combining traditional language-based research with modern computational methods

Underlying interest

- Linguistically, I want to improve our understanding language use, linguistic variability, and cultural shifts reflected in language
- Methodologically, I'm intersted in statistical modelling of lamnguage data, reproducibility and transparency in language-based research, and in Big Data (not massive data)

Underlying theme

- I have an interest in asepcts of language that are not thought in school, yet fundamentally improtant for communication

Martin Schweinberger

Director *Language Technology and Data Analysis Laboratory* (LADAL, www.ladal.edu.au)
which is financed through the *Language Data Commons of Australia* (LDaCA, www.ldaca.edu.au)
LDaCA is the largest research infrastructure project in the humanities and social sciences
in Australia (role: CI & steering commitee)

Relevant positions in Centres of Excellence

- 2018-2023: Affiliate Australian Research Foundation (ARC) *Centre of Excellence for the Dynamics of Language* (CoEDL, Australian CoE)

- 2021-2023: Principal Data Science Advisor & Asscoiate Professor II at the AcqVA Aurora Lab in the *Aurora Center for Language Acquisition, Variation & Attrition* (Norwegian CoE)

- 2025: International collaborator Research Council of Finland's Centre of Excellence in Climate Action (CLIMACTION, Finnish CoE, application submitted)

- 2025: CI ARC Centre of Excellence for the Future of Language (Australian CoE, not successful)

# Background to what I want to present on today

Background

In 2021, together with colleagues from UQ, The Queensland University of Technology, and the University of Eastern Finland, we applied for funding for a project called **F%$# Twitter** from the UQ Centre for Digital Cultures and Societies.

The results of this project have led to several presentations at international conferences and publications including Special Issue of *Lingua* entitled ***Bad Language and Vulgarity Online and in Public Discourse*** (Eds. Martin Schweinberger, Paula Rautionaho, & Kate Burridge)

Schweinberger & Burridge (2025) Vulgarity in online discourse around the English-speaking world went viral…

- Features in the national afternoon and evening news in Australia

- Over 70 newspaper articles including *The Guardian*, *CNN*, *Der Speigel*,

- *The Conversation*, *Deutsche Welle*, *The Washington Times*, etc.

- Several live radio interviews (I alone appeared on 5 ABC radio shows)
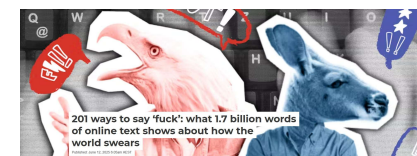
Structur of today's talk

What I want to talk to you about today are results of that research project that may be relevant in the context of (De)Democratisation

Prior Research (on Vulgarity/Swearing)

Methodology

Case Studies

• Vulgarity in Online Discourse around the English-Speaking World (Schweinberger & Burridge, 2025)

• F%$# Twitter. A corpus-based analysis of vulgar language on Twitter (Schweinberger et al., fc.a)

Key Findings and Limitations

Discussion & Future Directions

Previous Research on Vulgarity

Previous Research

Swearing in Online Discourse

Vulgarity and swearing figure prominently in studies focusing on hatespeech and online bullying. However, ther is more to these phenomena!

Online communication shows distinctive patterns of vulgarity
(Cheng et al., 2017; Dynel, 2012; Jay & Janschewitz, 2008; McEnery & Xiao, 2004; Seargeant & Tagg, 2014)

- Higher frequency of swearing in anonymized or semi-anonymous platforms (e.g., Reddit, Twitter, 4chan)
- Vulgarity used both for hostile purposes (e.g., trolling, flaming) and for affiliative functions (e.g., memes, banter)
- Strong ties to stance-taking, political identity, and in-group signalling
- Swearing constitutes ~0.5–0.7% of all spoken tokens in English corpora
- A small set of taboo words (e.g., *fuck*, *shit*, *damn*) accounts for the majority of instances

Norms and Taboo Dynamics

- Swearing norms are historically and culturally variable
  words that were taboo 50 years ago may now be mild or even mainstream (Allan & Burridge, 2006; Hughes, 2006)
- Increasing desensitization
  Especially in media and music (cf. corpus studies of lyrics and TV scripts) (Fägersten & Bednarek, 2022)
- Taboo often shifts from religious to sexual to racial/identity-based profanity (Burridge, 2017; Jay, 2009)

Previous Research

Functions of Swearing/Vulgarity

Vulgarity and swearing a highly functional feautures of human interaction and communication

Social functions

- Indexing Identity/Group Membership: can signal in-group solidarity, shared norms, or cultural/regional identity (Beers Fägersten, 2012)
- Establishing|Challenging Social Hierarchies: can assert dominance, express resistance, or subvert social norms (Dynel, 2012; Jay, 2009)
- Expressing Emotion & Intensity: used as an emotional intensifier and stress relief (Stephens, Atkins, & Kingston, 2009; Wajnryb, 2005).
- Managing Social Distance & Politeness: may reduce social distance or enhance authenticity in informal settings (Culpeper, 2011; Kádár & Haugh, 2013)
- Resistance and rebellion: Used to challenge norms, institutions, or power structures (e.g., punk lyrics, protest slogans).

Linguistic functions

- Lexical Creativity & Innovation: often involves creative word formation, euphemisms, and taboo-avoidance strategies (Allan & Burridge, 2006; Ljung, 2011).
- Pragmatic & Discourse Marking: serves as a discourse strategy to mark stance, topic shifts, or emphasis (Dewaele, 2016; Holmes, 2013)
- Sociolinguistic Variation: varies by age, gender, class, and region, making it a rich domain for sociolinguistic study (McEnery & Xiao, 2004; Schweinberger, 2018; Stapleton, 2003)
- Language Norms and Taboo: reveals how societies construct and negotiate linguistic taboos (Burridge, 2006; Pinker, 2007).

Previous Research

Functions of Swearing/Vulgarity

Vulgarity and swearing figure prominently in studies focusing on hatespeech and online bullying. However, ther is more to these phenomena!

Psychological Functions

- Emotional regulation: helps people manage pain, fear, or frustration (e.g., Stephens et al., 2009).

- Catharsis: Offers emotional release or venting in high-stress or conflict situations.

- Identity assertion: Reinforces a sense of self or group belonging, particularly in adolescence or marginalised communities.

Cultural Functions

- Indexing modernity or authenticity: In some contexts, swearing signals being "real," "honest," or even progressive, challenging linguistic conservatism.

- Taboo reinforcement and negotiation: defines what is unacceptable and thereby also delineates what a culture values or suppresses.

- Cultural memory and intertextuality: can reference historical moments, media (e.g., Tarantino-style dialogue), or pop-cultural icons.

Aesthetic and Rhetorical Functions

- Stylistic resource: Authors, artists, or speakers may use swearing to craft voice, realism, or intensity.

- Humour and irony: Vulgarity is frequently used for comedic effect or to subvert expectations (especially in stand-up or satire).

- Provocation and attention: Swearing can deliberately shock or disrupt, functioning as a rhetorical or performative tool.

Previous Research

What is the connection between Swearing/Vulgarity and(De)Democratisation? (based on Putensen, 2025)

The use of vulgarity is tied to **Colloquialization** (Mair, 2024)

- General trend towards informality in writing
- Increase in spoken(-like) features in written and public discourse
- Linked to wider societal trend linked to cultural and political changes in the 60s and 70s (Hobsbawm, 1994)

Today, the use of "easy" language with short texts and spoken-like, informal features is a common feature of populist discourse (Trump) (see Schilling & Fuchs, 2025)

Link between colloquialisation to a broader democratisation of discourse (Fairclough, 1992)

- **Democratisation**: phasing out of overt markers of power asymmetry
- **Commodification**: increasing marketisation of language, where informal, engaging styles are used for persuasion
- **Technologisation**: influence of digital communication in reshaping linguistic expectations

With the rise of digital media, the boundaries between speech and writing, formal and informal, public and private discourse have become increasingly blurred - social media platforms shift traditional genre boundaries which, in turn, accelerates colloquialisation on a global scale

Study 1
Vulgarity in Online Discourse around the English-Speaking World

Vulgarity in Online Discourse around the English-Speaking World

Motivation

Substantive body of research on vulgarity and social factors that correlate with its use, but no large-scale comparative studies that analyze the use of vulgarity across English-speaking regions

RQs

- How frequent is vulgar language in online discourse across English-speaking regions?

- What vulgar elements are used in the data and are there region-specific differences or preferences?

- Is there are difference in the variability (creativity) of vulgar elements that are used?

Data

Global Web-Based English Corpus (GloWbE) – large dataset consisting of scraped web texts designed to allow comparisons across English-speaking regions (Davies & Fuchs, 2015)

App. 1.7 billion words across 20 English-speaking regions

Methodology

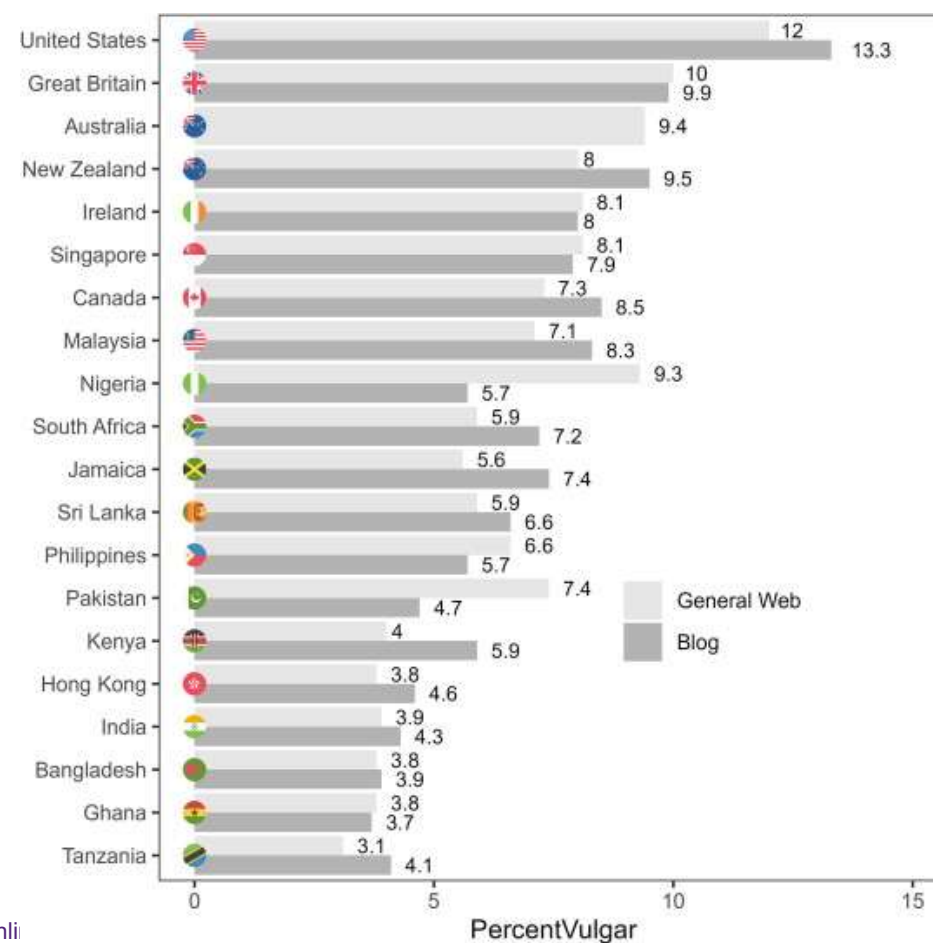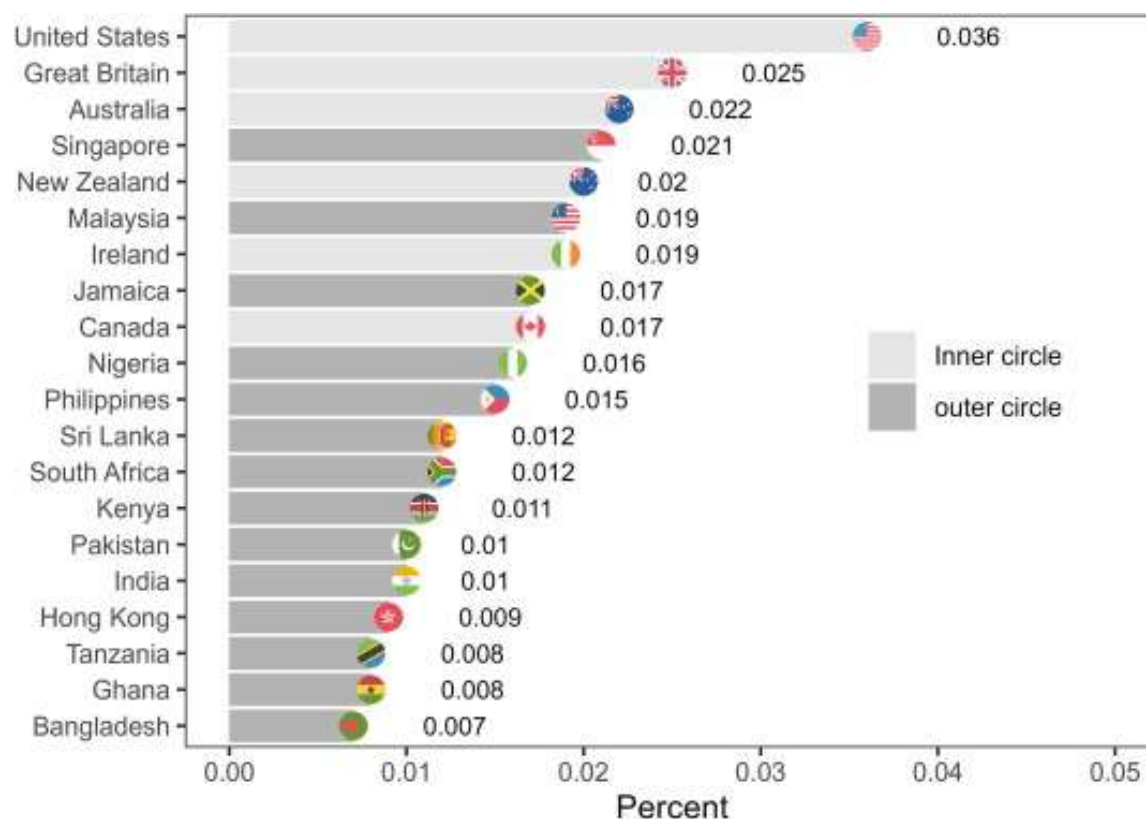Extraction of vulgar elements using regular expressions plus manual cross-checking.
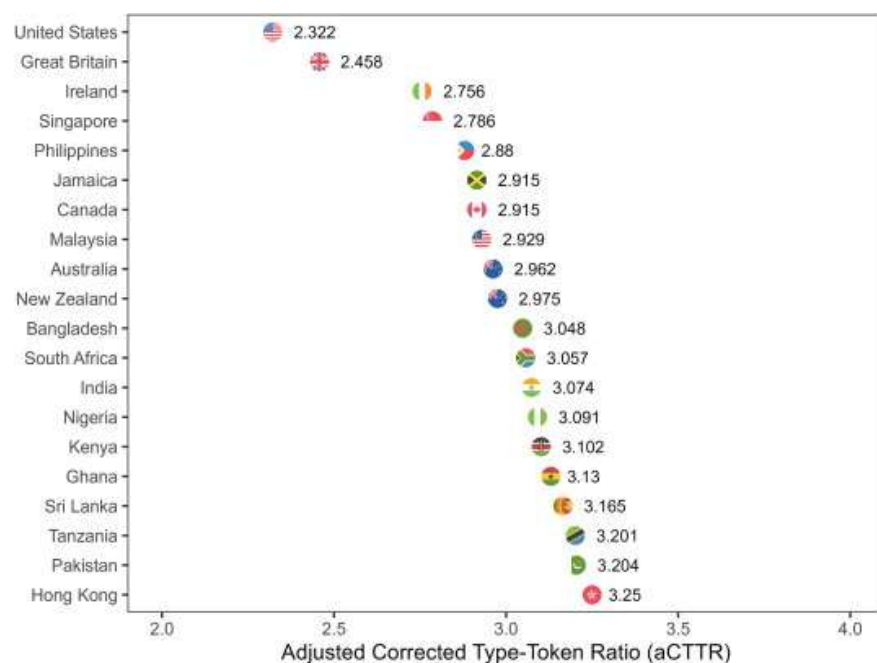Basic statistics (Configural Frequency Analysis, adjusted Corrected Type-Token Ratios)

Regex used to identify variants of the lemma <fuck>

```
"\\b(cluster|head|mother|motha|mutha|m
ada|cock|mom|mum|daddy|father|sister|
brother)*[f=ƒ][uµüúû|@a4äáå|e3€ëéê|o0
øöóõ|*]*[c¢©(]*[k|<{(][e3€ëéê]*[r]*(head|f
ace|wit|ing|er|a|ed|ers|az)*\\b",

"\\b[f=ƒ][c¢©(*k<{(uµüúû*]+(ing|er|a|ed|e
rs|az)*\\b",
```
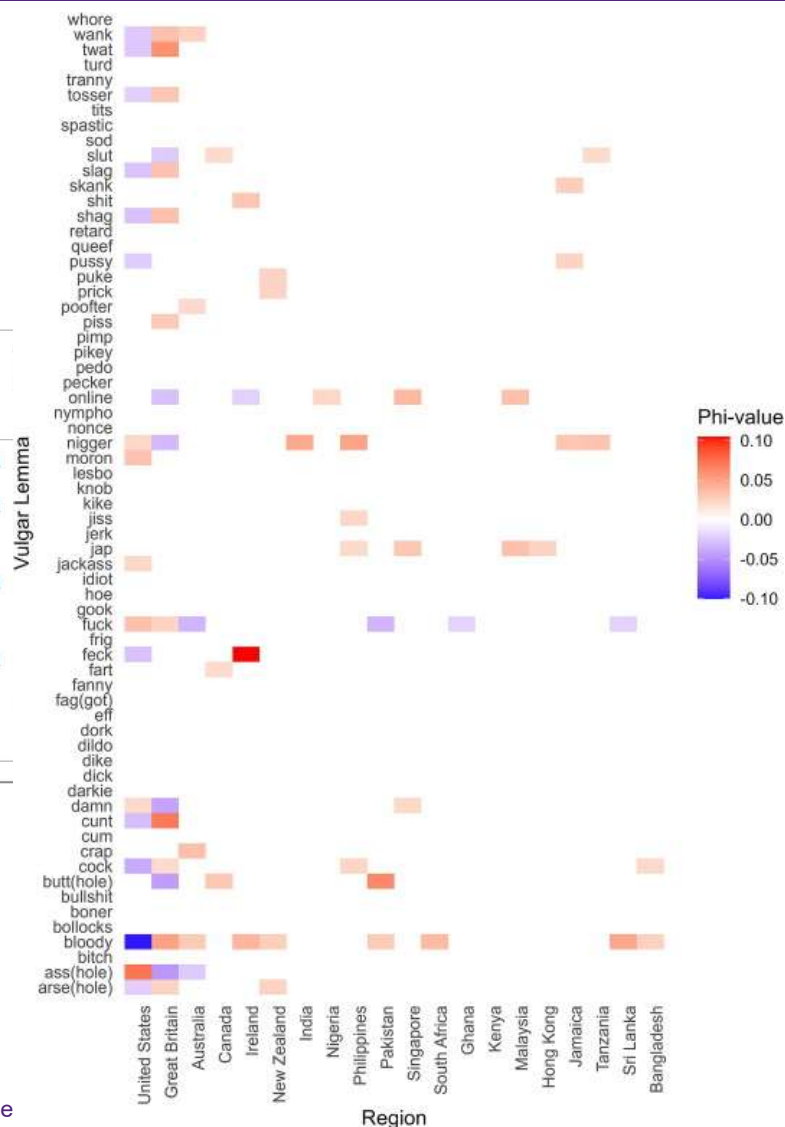
Findings

Schweinberger: Global Trends in Democratisation Practices in Onli

Schweinberger: Global Trends in Democratisation Practices in Online Discourse

Region	Vulgar lemma	Type
Ireland	feck	Preferred
United States	ass(hole)	Preferred
Great Britain	cunt	Preferred
Pakistan	butt(hole)	Preferred
United States	bloody	Avoided

Vulgarity in Online Discourse around the English-Speaking World

Highlights

First large-scale empirical comparative analysis of vulgarity in online discourse across several varieties of English

- Inner circle varieties exhibit more vulgarity compared with outer circle varieties.

- Speakers in the United States use vulgarity most.

- Inner circle variety speakers make use of a wider variety of vulgar forms.

- Speakers from different regions show preferences for specific vulgar items.

Study 2
F%$# Twitter. A corpus-based analysis of vulgar language on Twitter

Vulgarity in Online Discourse around the English-Speaking World

Motivation

Substantive body of research on vulgarity and social factors that correlate with its use, but traditional social factors cannot be analyzed reliably on social media.

Twitter data samples are often not representative or balanced rendering previous analyses problematic.

RQs

- What forms of vulgarity are used on Twitter in the United States, the United Kingdom, and Australia?

- What topics are associated with vulgarity on Twitter in the United States, the United Kingdom, and Australia?

- What factors correlate with vulgarity on Twitter?

Vulgarity in Online Discourse around the English-Speaking World

Data

Twitter data accessed using the Twitter API v2 via academic access and Twarc software.

June 21, 2022
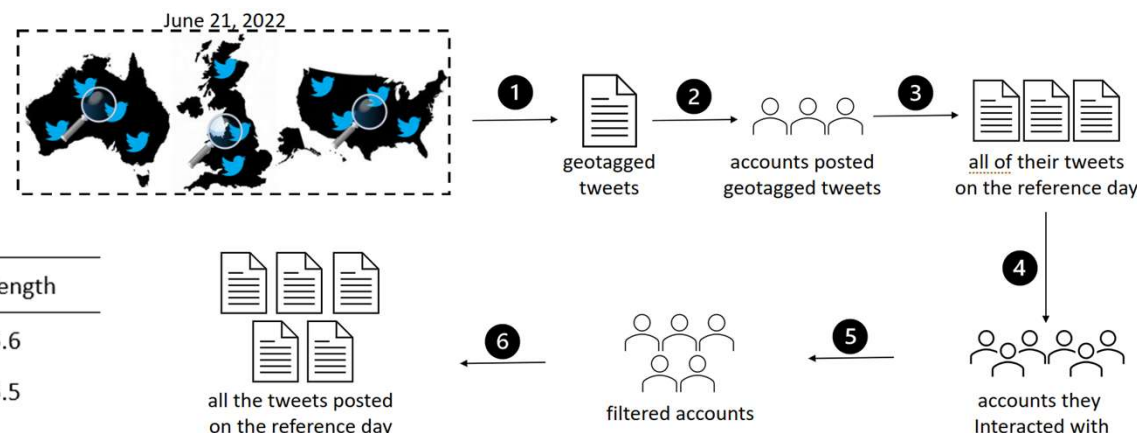
	Accounts	Tweets	Tokens	Avg. length
AU	7,527	87,000	1,357,599	15.6
UK	50,780	475,355	7,370,433	15.5
US	180,893	1,811,029	25,479,355	14.1
Total	239,200	2,373,384	34,207,387	14.4

Necessary: extensive data cleaning and vetting (removal of non-English tweets, etc.)

Advantage: this way we could create a comparable dataset and enrich it with metadata about the speakers (number of followers, when did the account start, how active is the account, etc.)

Vulgarity in Online Discourse around the English-Speaking World

Methodology

Extraction of vulgar elements using regular expressions.

What counts as vulgar?

- Based on pre-existing lists used in online moderation
- Previous research
- Dictionaries
- Test: exists a more appropriate alternative? (fuck included as less vulgar forms, such as make love with, are available).
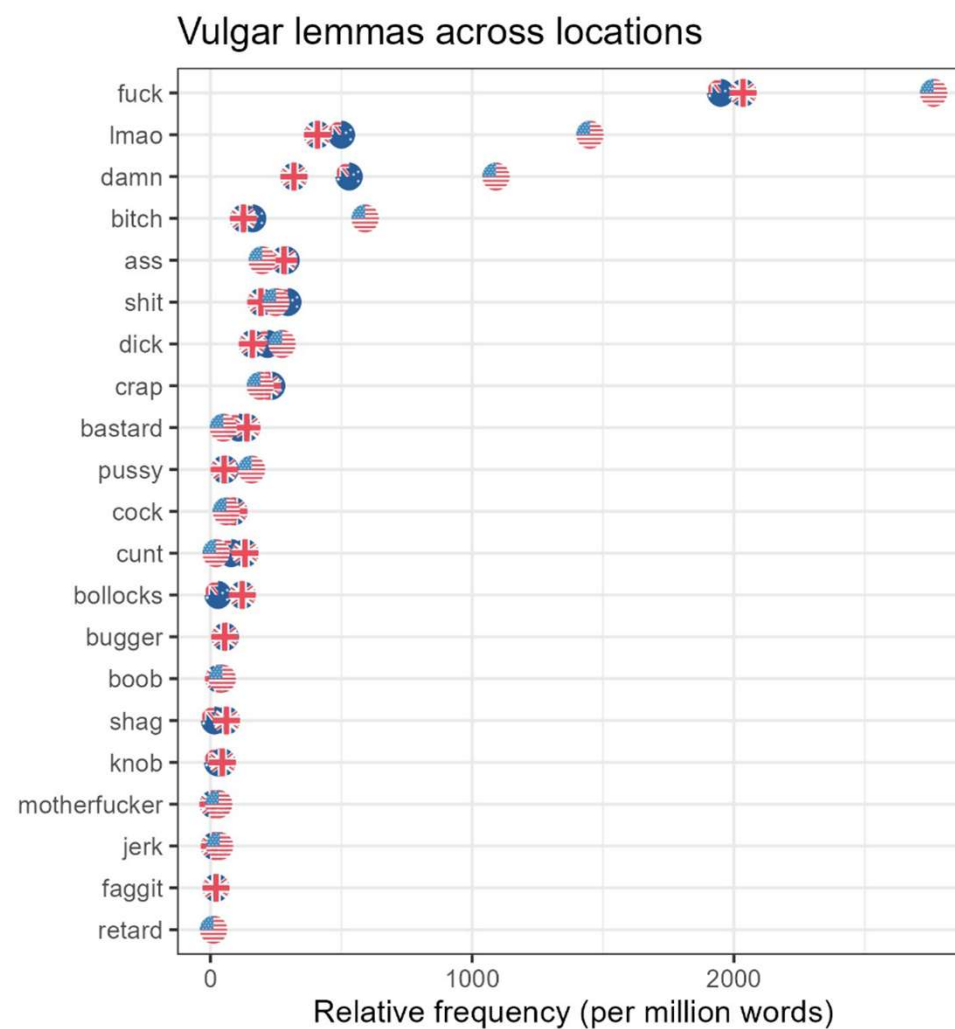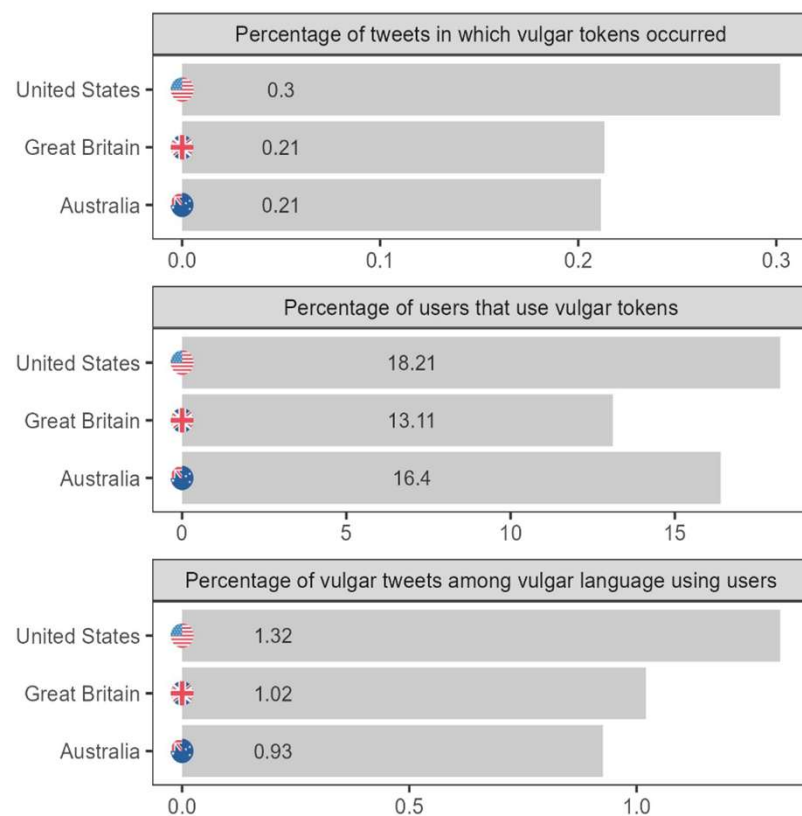
Topic modelling

Allows to group documents by "topic", i.e. based on the words that appear in them
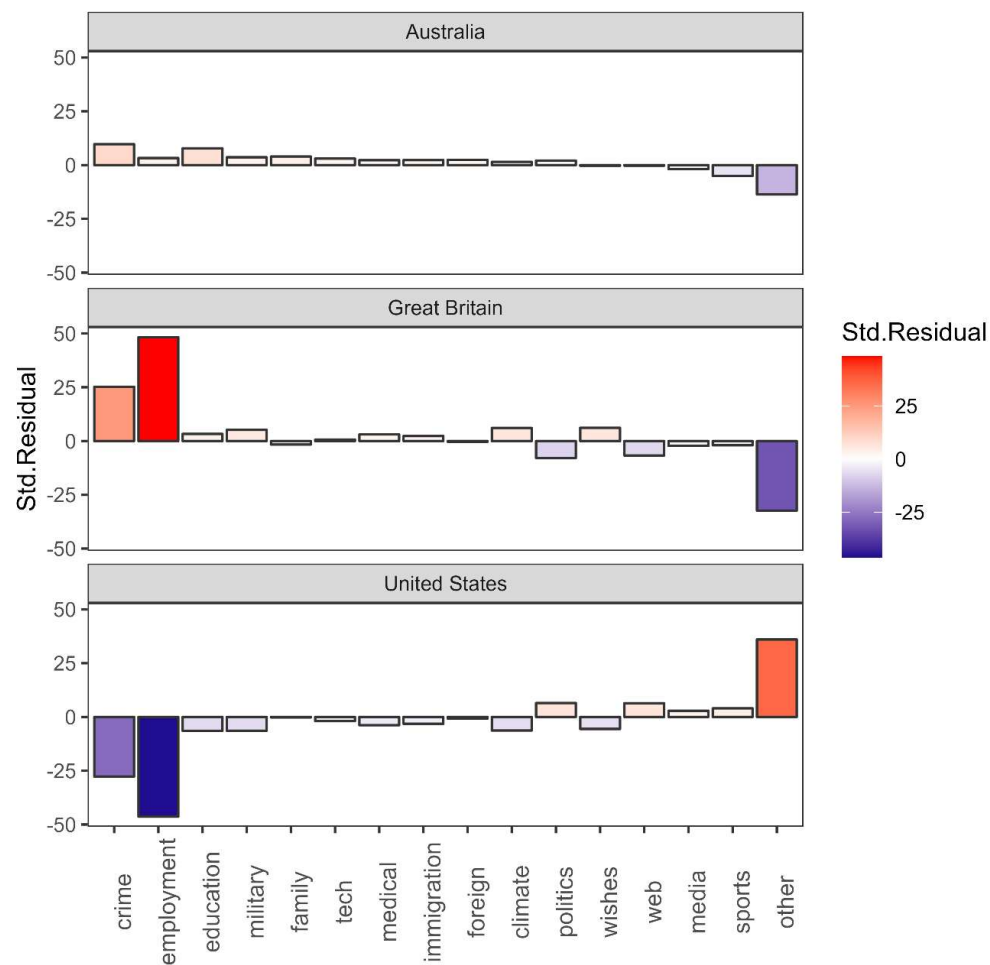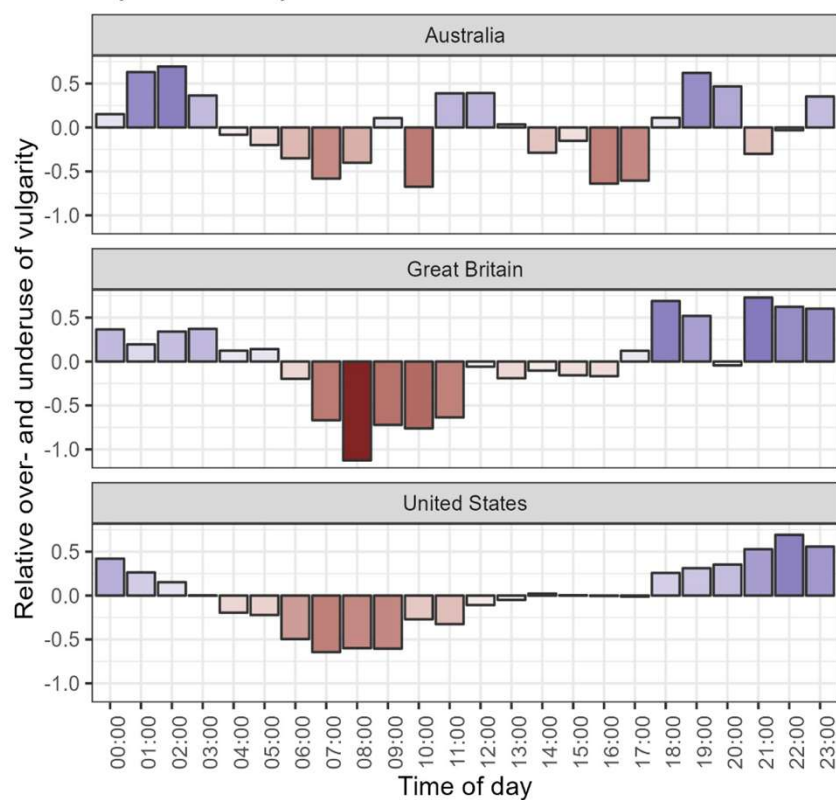
Boruta

- Extension of Random Forests (tree-based bootstrapped partitioning method developed in machine learning for classification)
- Advantage over random forests allows to reliably differentiate between relevant and non-relevant factors
- Advantages over regression analysis: not affected by multicollinearity (correlation between predictors) and very stable.

Findings

Percentage of tweets in which vulgar tokens occurred

Location	Value
United States	0.3
Great Britain	0.21
Australia	0.21

Percentage of users that use vulgar tokens

Location	Value
United States	18.21
Great Britain	13.11
Australia	16.4

Percentage of vulgar tweets among vulgar language using users

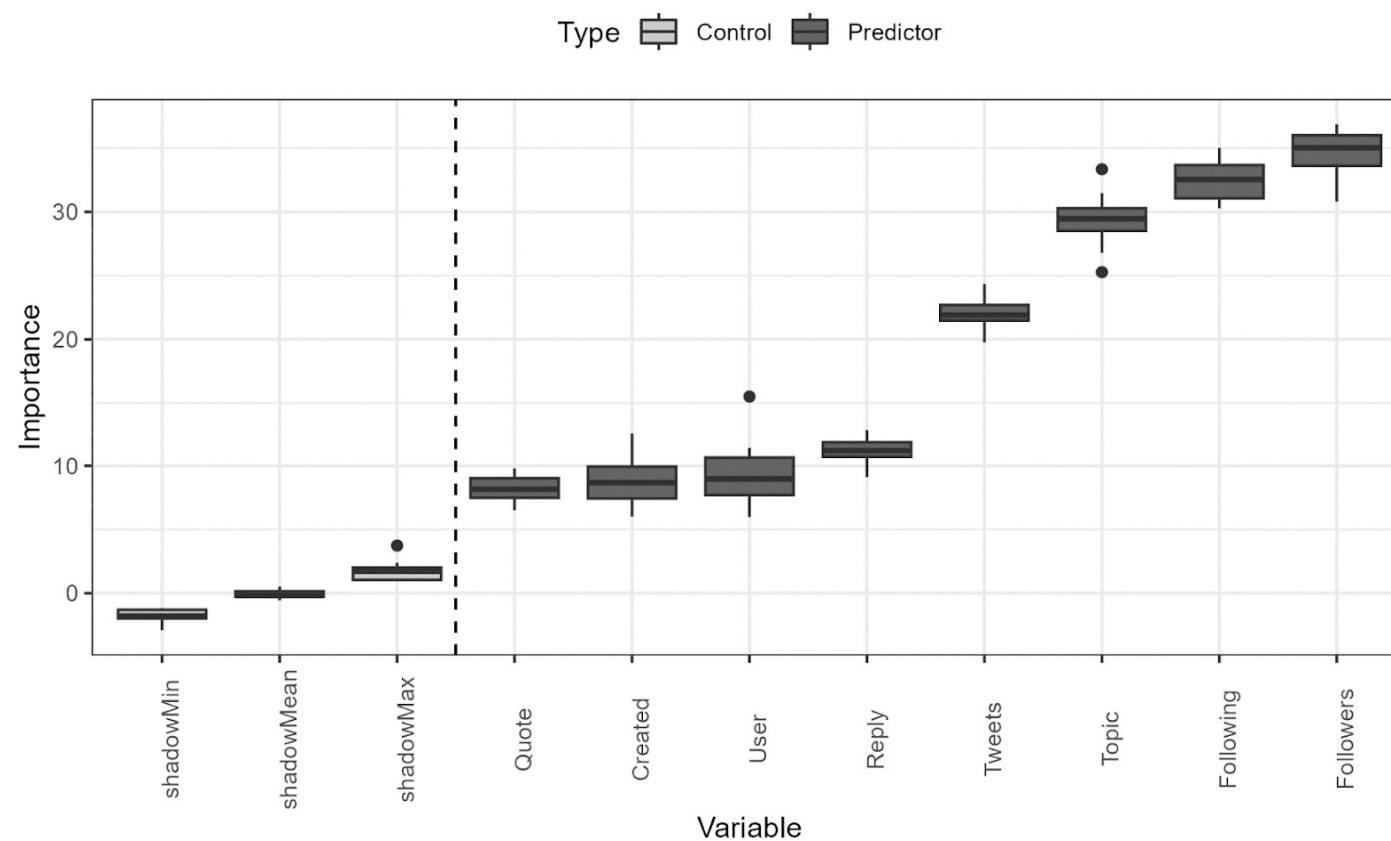| Location	Value
United States	1.32
Great Britain	1.02
Australia	0.93

Vulgar lemmas across locations

Findings

Percentage of observed against expected vulgarity by time of day and location

Findings

F%$# Twitter. A corpus-based analysis of vulgar language on Twitter

Highlights

One of the first large-scale comparative analysis of vulgarity on Twitter

- Users from the United States use most vulgarity

- Variants of <fuck> are the most common form of vulgarity (followed by forms of abbreviations typically associated with online discourse, e.g., *wtf*, *lmao*, *asf*, etc.)

- Vulgarity is underused during working hours

- Users from the United States avoid vulgarity in tweets related to work/employment and crime while users from Great Britain relatively overuse vulgarity in tweets around these topics.

- Factors associated with users' online status affects the use of vulgarity (numbers of followers, numbers of accounts followed, topic, activity, etc.) confirming that enriching online data has explanatory power

Key Findings & Problems

Key Findings

Cross-Platform and Cross-Regional Usage

- Vulgarity is widespread but concentrated in inner-circle English varieties, especially the US and UK.

- In both studies, users from the US used most vulgarity

Lexical and Functional Variation

- Most vulgarity comes from a limited core lexicon (e.g., fuck, shit), yet regional preferences are evident (e.g., preference for *feck* by Irish users, *ass(hole)* by US users, or *cunt* by UK users).

Time and Context Sensitivity

- Temporal variation observed: use decreases during working hours.

- Vulgarity correlates with social context (e.g., less in tweets about jobs/crime in the US, more in such tweets from the UK).

Sociolinguistic Insights

- Users' online status (followers, activity level, account age) correlates with likelihood of using vulgar language.

Problems

There are too many issues and limitations to be discussed in detail.

Among the most obvious issues are…

Data-related issues

- Due to missing information, confounds are unknown (data skewed towards younger people rather than older users – they is likely to vary across regions)

- Genuine-ness of the data hard to validate (is data from India really produced by a speaker of Indian English?) – "true" behavior hard to capture

- In study 1, non-English vulgarity was not considered.

Statistical Method

- Boruta only informs about the existence of an effect – not its direction. Further analysis required…

Discussion & Future Directions

Discussion – Vulgarity and (De)Democratisation

Colloquialisation as Democratization?

- The (increased) use of vulgarity reflects colloquialization, i.e., a shift towards informality in public discourse.

- Results presented here support Fairclough's (1992) hypothesis that democratization reduces linguistic markers of power asymmetry.

Informality ≠ Anti-Democratic

- Informal, vulgar language can increase engagement, signal authenticity, and enhance solidarity.

- Especially in online contexts, vulgarity often functions as a tool of inclusion, not exclusion.

Populism and Linguistic Authenticity

- Populist discourse often embraces vulgarity to reject elite norms and appeal to "ordinary people".

- Vulgarity becomes a rhetorical device to claim truth-telling and anti-establishment ethos (Trump: "They don't know what the fuck they are doing…" (23/4 June, 2025)

Blurring of Public/Private Boundaries

- Social media collapses traditional genre boundaries which contributes to the legitimation of informal, emotionally charged, and provocative discourse in public arenas.

Future Directions

- Vulgarity is not merely offensive language—it performs complex social, emotional, and rhetorical functions

- Its presence in online discourse reflects broader sociolinguistic trends (colloquialisation, personalization, and polarization)

- Patterns of use differ across regions, platforms, and topics, shaped by sociotechnical affordances and cultural norms

- Vulgarity in online discourse can act as a democratizing force, but it also challenges civility norms and tests the limits of tolerance

Future Directions

- Multimodal vulgarity: Explore how swearing interacts with memes/emojis (first results very promising)

- Beyond English: Comparative studies across languages and cultural contexts (e.g., German)

- Discourse ecology: Situate vulgarity within broader patterns of (de)legitimation, resistance, and identity performance

References

Allan, K., & Burridge, K. (2006). Forbidden Words: Taboo and the Censoring of Language. Cambridge University Press.

Beers Fägersten, K. (2012). *Who's Swearing Now? The Social Aspects of Conversational Swearing*. Cambridge Scholars Publishing.

Burridge, K. (2006). Taboo words and the polite language. *English Today, 22*(1), 3–9.

Culpeper, J. (2011). *Impoliteness: Using Language to Cause Offence*. Cambridge University Press.

Dewaele, J.-M. (2016). Self-reported use and perception of swear words across languages by multilinguals. *Intercultural Pragmatics, 13*(3), 343–369.

Dynel, M. (2012). Swearing methodologically: The (im)politeness of expletives in anonymous commentaries on YouTube. *Journal of English Studies, 10*, 25–50.

Fägersten, K. B., & Bednarek, M. (2022). The evolution of swearing in television catchphrases. *Language and Literature, 31*(2), 196-226.

Gries, S. Th. (2020). On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory, 16*(3), 617-647.

Holmes, J. (2013). "Discourse markers." In R. F. Simonsen & P. Mæhlum (Eds.), *Pragmatics of Society* (pp. 343–364). De Gruyter Mouton.

Jay, T. (2009). The utility and ubiquity of taboo words. *Perspectives on Psychological Science, 4*(2), 153–161. https://doi.org/10.1111/j.1745-6924.2009.01115.x

Kádár, D. Z., & Haugh, M. (2013). *Understanding Politeness*. Cambridge University Press.

Laitinen, M., Rautionaho, P., Fatemi, M., & Halonen, M. (2025). Do we swear more with friends or with acquaintances? F#ck in social networks. *Lingua*, 320, 103931. https://doi.org/10.1016/j.lingua.2025.103931

Ljung, M. (2011). *Swearing: A Cross-Cultural Linguistic Study*. Palgrave Macmillan.

McEnery, T., & Xiao, R. (2004). Swearing in Modern British English: The Case of Fuck in the BNC. *Language and Literature, 13*(3), 235–268.

Pinker, S. (2007). *The Stuff of Thought: Language as a Window into Human Nature*. Viking.

Posit team (2025). *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA. url: http://www.posit.co/.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing,   Vienna, Austria. url: https://www.R-project.org.

Schweinberger, Martin (2018). Swearing in Irish English – A corpus-based quantitative analysis of the sociolinguistics of swearing. *Lingua 209*: 1-20. https://doi.org/10.1016/j.lingua.2018.03.008

Schweinberger, Martin and Kate Burridge. 2025. Vulgarity in Online Discourse around the English-Speaking World. *Lingua*, 321: 103946. https://doi.org/10.1016/j.lingua.2025.103946

Schweinberger, M., Laitinen, M., Haugh, M., Rautionaho, P., Fatemi, M., Hames, S., & Takahashi, M. (n.d.). F%$# Twitter: A corpus-based analysis of vulgar language on Twitter [Manuscript submitted for publication]. *Corpora*.

Schweinberger, M., Haugh, M., Laitinen, M., Fatemi, M., Hames, S., Rautionaho, P., & Takahashi, M. (n.d.). Swearing online: Exploring dimensions of variability in born-digital data [Manuscript submitted for publication]. *Digital Scholarship in the Humanities*.

Stapleton, K. (2003). Gender and swearing: A community practice. *Women and Language, 26*(2), 22–33.

Stephens, R., Atkins, J., & Kingston, A. (2009). Swearing as a response to pain. *NeuroReport, 20*(12), 1056–1060. https://doi.org/10.1097/WNR.0b013e32832e64b1

Wajnryb, R. (2005). *Expletive Deleted: A Good Look at Bad Language*. Free Press.

Thank you very f§%$ much

Global Trends in Democratisation Practices in Online Discourse

Martin Schweinberger

m.schweinberger@uq.edu.au

IAAK / RESEARCH PROJECTS / LANDPOSTS IN (DE-)DEMOCRATIZATION

LANDPOSTS IN (DE-)DEMOCRATIZATION

Interdisciplinary Workshop, 25 – 26 June 2025, University of Bonn