



UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
ARTS AND LAW

Data and Humanities: love it or loathe it?

Petar Milin

<p.milin@bham.ac.uk>



Coping with Public Sexual Violence in the Digital Sphere:



Spain #LaManada vs #SarahEverard UK

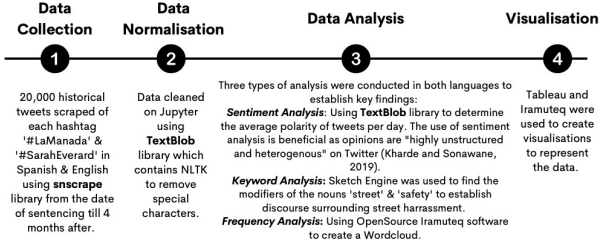
INTRODUCTION & PURPOSE

This exploratory corpus linguistic study comparatively investigates the response to two public sexual violence cases in the UK and Spain on Twitter, following a call by Bisiada (2021) for more cross-linguistic studies in this field.

The cases of Sarah Everard and La Manada provide a strong cross-linguistic standpoint as they:

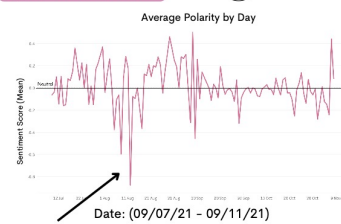
- are similar in criminal nature, and can be used to establish discourse surrounding 'public sexual harassment' on the street, where there is a paucity of research.
- have gained momentum within the domain of feminist hashtag activism, being utilised along other key hashtags that have been largely researched including #MeToo, #YoSÍTeCreo, #Cuéntalo (Bisiada, 2021).
- similar in their stance towards the hashtags, with both countries having reported positive language and discourse (ibid), something this novel study will seek to confirm.

METHODOLOGY



RESULTS

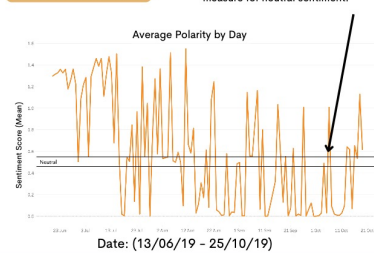
Figure 1 #saraheverard



Sharp negative minimum (10/08 - 12/08 - reported same dates as murder of another British woman Sarah Hussein.

A different library was used to obtain the Spanish sentiment score, therefore it cannot be measured in the same way as in English, and the researcher decided the region of 0.45 - 0.55 to be measure for neutral sentiment.

Figure 2 #lamanada



DISCUSSION

Spanish

Sentiment is higher to start with perhaps due to overturn of previous trial where perpetrators did not receive full sentence (Figure 2)

Uses more action verbs for the collocates of 'street' such as 'explotar', 'protestar', 'meter', 'gritar' in their view of the street, seeing it as a place of change and power (Figure 3)

The concept of seeking justice is central to the discourse (Figure 4), with the use of terms like 'justicia', 'sentencia', 'violación'.

English

Largely positive sentiment (F1 & 2)

'Safety' is seen as a right for women, with both using verbs such as 'garantizar' and 'ensure' (F3)

Figure 4: Twitter as a safe community for women to express their views, with established hashtags appearing frequently in both corpora including: #yoSiTeCreo, #Inocencia and #sheWasJustWalkingHome #enoughisEnough (Bisiada, 2021; Clark, 2016)

Negative spike in Figure 1 which could be linked to death of Sarah Hussein who was largely underreported in media

Figure 3 shows 'street', with the collocates representing the concept of freedom, such as 'room', 'light', 'desert' and 'reclaim', which evokes the movement #ReclaimTheStreets

The discourse focuses on sexual violence at the institutional level, with the use of 'trust', 'metpoliceuk', 'power' (Figure 4)

FIGURE 3: Most frequent verb collocates with the nouns 'street' and 'safety', in both Spanish and English in both respective corpora.

Verbs which take the following 'nouns' as objects:			
CALLE	SEGURIDAD	STREET	SAFETY
valer	garantizar	walk	ensure
explotar	orientar	patrol	deserve
protestar	reforzar	make	improve
aplaudir	aumentar	reclaim	threaten
gritar	dar	light	provide
meter		room	address
parecer		keep	regard
servir		desert	put
volver		police	

FIGURE 4: Word Cloud highlighting the most frequent words in both respective corpora.

FIGURE 4: Word Cloud highlighting the most frequent words in both respective corpora.

REFERENCES

Bisiada, M., (2021). *Movement or debate? How#MeToo is framed differently in English, Spanish and German Twitter discourse.*

Clark, R. (2016). "Hope in a hashtag": The discursive activism of #WhyStayed. *Feminist media studies*, 16(5), pp.788-804.

Iramuteq (2008). Available at: <http://www.iramuteq.org> (Accessed: 10th May 2022)

Kharde, V. and Sonawane, P. (2016). *Sentiment analysis of twitter data: a survey of techniques.*

Tableau (2003). Available at: <https://www.tableau.com> (Accessed: 12th May 2022)

Sketch Engine (2008). Available at: <https://www.sketchengine.eu> (Accessed: 12th May 2022)

Limitations & Future Research:

- Due to the number of historical tweets, the researcher could not obtain results from across the whole period of the movement (#), which would be more reflective of the discourse, therefore future research could be extended to include this.
- A literature review pointed the value in Twitter as better representation of feminist discourse than traditional media (i.e. newspapers), therefore this study could be conducted again using newspapers as a corpus which would provide a novel comparative standpoint.







Epistemological abyss?



UNIVERSITY OF
BIRMINGHAM

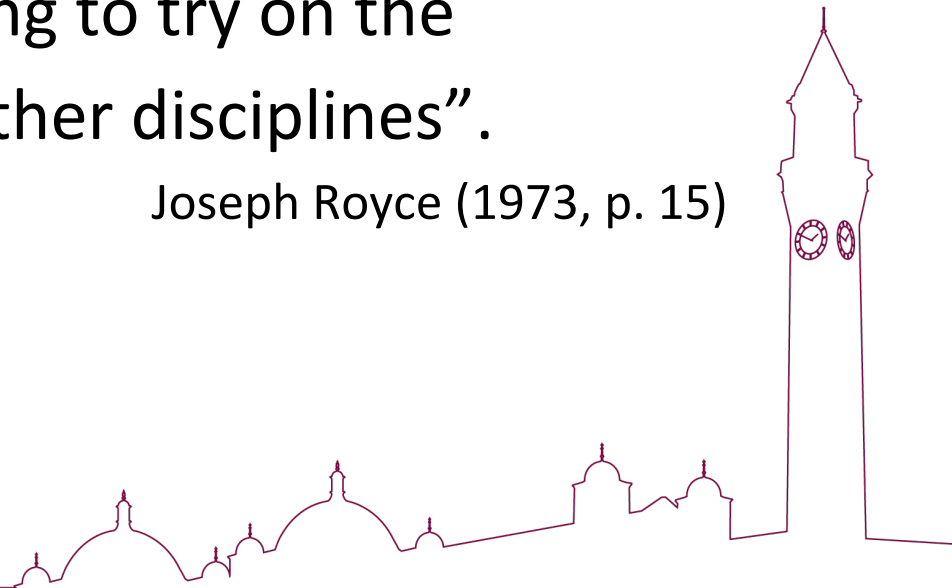
COLLEGE OF
ARTS AND LAW

“Psychology [is] unique in the extent to which its institutionalization preceded its content and its methods preceded its problems.”

Sigmund Koch (1959, p. 783)

“[Thus, it] has been too willing to try on the theoretical orientations of other disciplines”.

Joseph Royce (1973, p. 15)

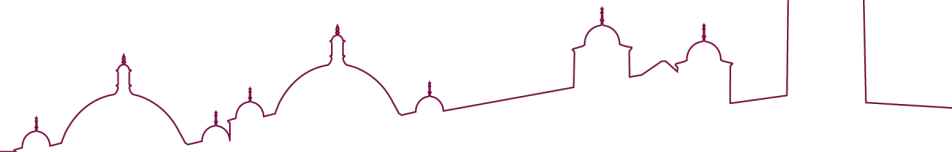


~~Epistemological abyss?~~

How to bridge the divide?

- Humanities like to keep themselves number-pasteurised
 - lack of tradition
 - lack of specific skillset
 - distrust in ability to make a difference

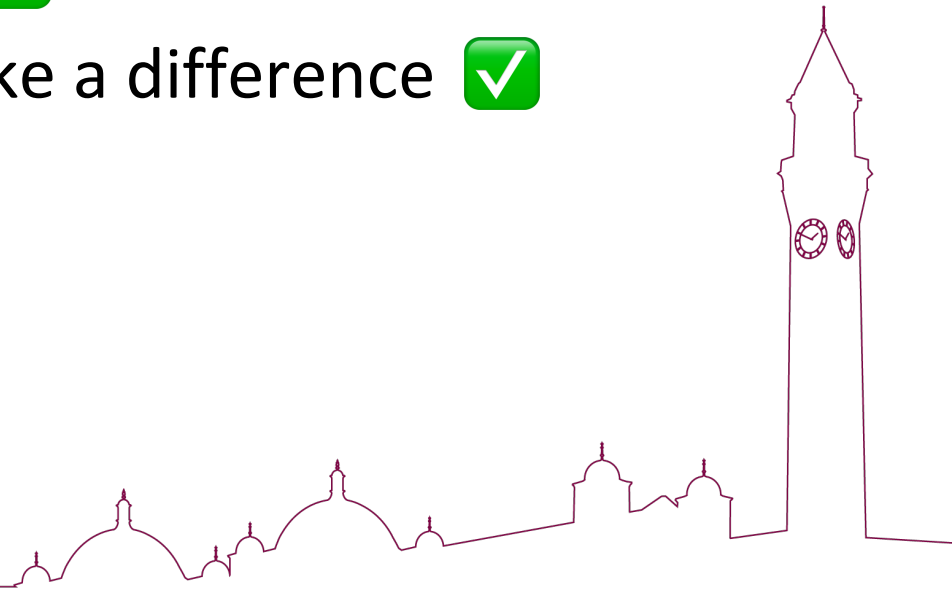
= selling short **Both sides!**



~~Epistemological abyss?~~

How to bridge the divide?

- lack of tradition
- lack of specific skillset ✓
- distrust in ability to make a difference ✓



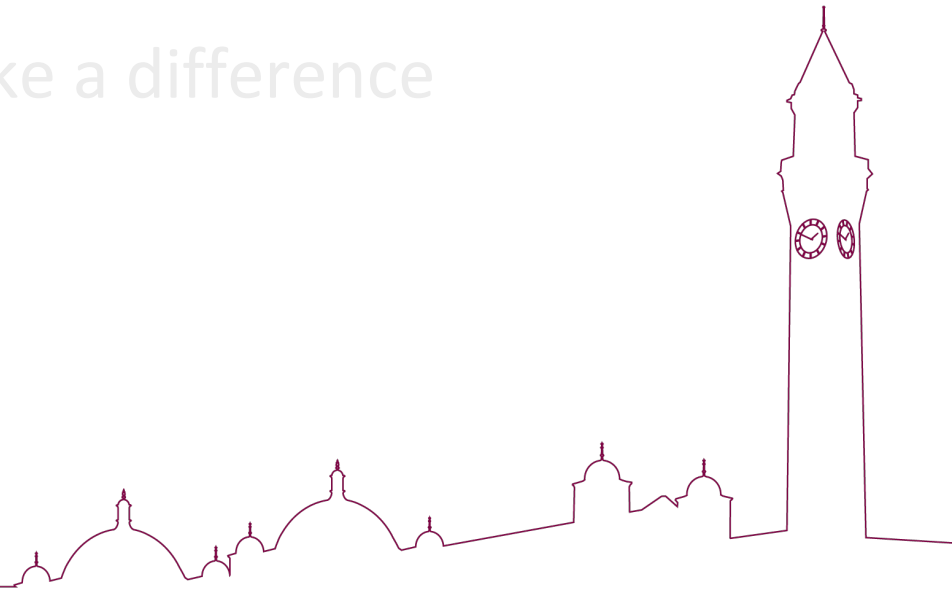
~~Epistemological abyss?~~

How to bridge the divide?

→ **lack of tradition ?**

lack of specific skillset

distrust in ability to make a difference



Can/Should we change the tradition?



Can/Should we change the traditions?

- Current focus on the *more embedded enterprise and employability provision*
→ that's easy... anything more?







Here's a heresy...

Data Science needs Humanities!

Obstacles to pass

Challenges to meet



UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
ARTS AND LAW

Here's a heresy...

Obstacles to pass

Challenges to meet

- The more interesting the data, the more difficult (humane)
- Humanities (“soft” disciplines) data is
 - more diverse and challenging/interesting
 - in next 5-10 years, large-scale processing of such data will be a principal, global must







✘ The photo you want to upload does not meet our criteria because:

- Subject eyes are closed

Please refer to the technical requirements.
You have 9 attempts left.

Check the photo [requirements](#).

Read more about [common photo problems and how to resolve them](#).

After your tenth attempt you will need to start again and re-enter the CAPTCHA security check.

Reference number: 20161206-81

Filename: Untitled.jpg

If you wish to [contact us](#) about the photo, you must provide us with the reference number given above.

Please print this information for your records.



✗ The photo you want to
our criteria because:

- Subject eyes are c

Please refer to the tech
You have 9 attempts left

Check the photo [require](#)

Read more about [commo](#)
[how to resolve them.](#)

After your tenth attempt
start again and re-enter
check.

Reference number: 201

Filename: Untitled.jpg

If you wish to [contact us](#)
must provide us with the
given above.

Please print this informa

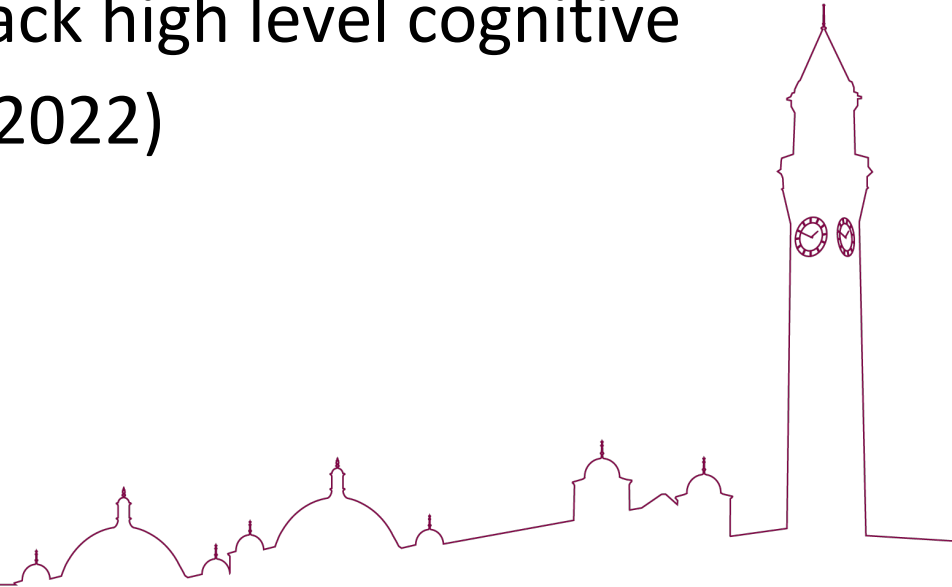


I hope they accept this one



Language sciences

- Sophisticated (Deep) Convolutional Networks suggest only rather shallow associations at the level of inter-word relationships (Budts, 2020)
- BERToids fundamentally lack high level cognitive capabilities (Anonymous, 2022)



Language sciences

<https://outofourminds.bham.ac.uk/>

- BERT did inhumanly good on a simple language task
- Simple but tractable network made an epic fail...
 - ... but achieved human-like performance once fed with carefully curated information



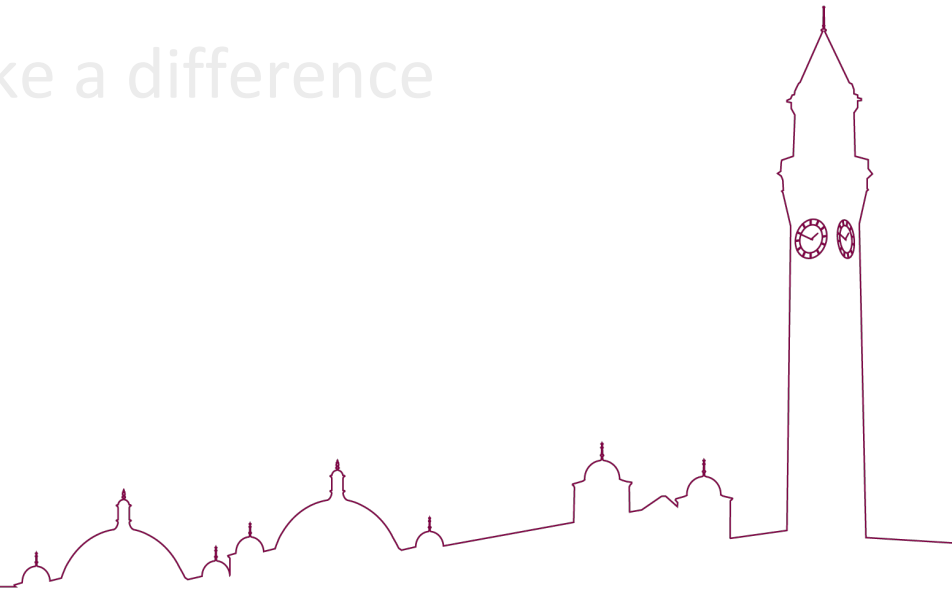
~~Epistemological abyss?~~ How to bridge the divide?

number-pasteurised

→ **lack of tradition ?**

lack of specific skillset

distrust in ability to make a difference



How?!?

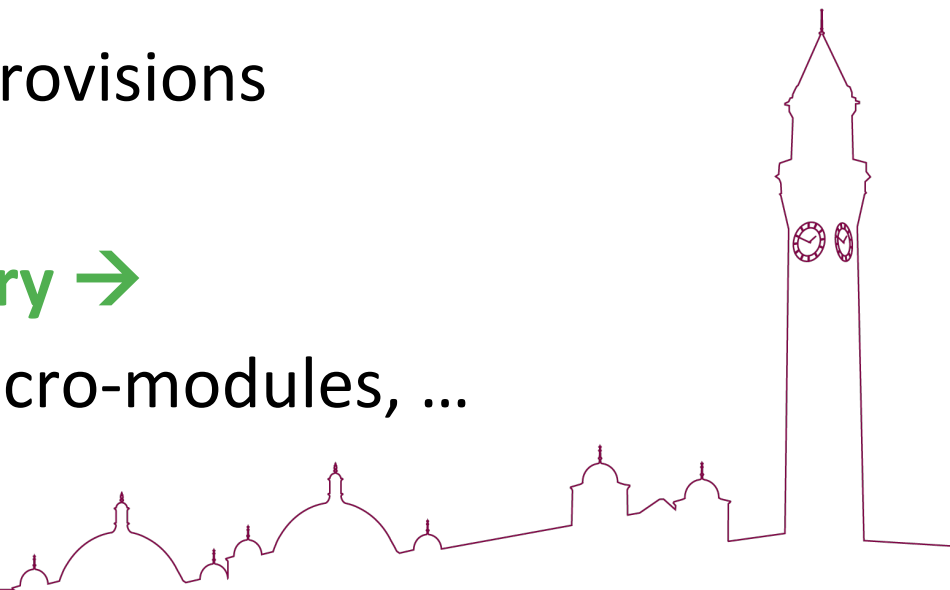


UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
ARTS AND LAW

Guiding principles

- **Proactivity** →
in setting the data science needs
- **Integration & Re-Integration** →
with main curricula provisions
- **Focused & Flexible delivery** →
e.g., targeted bits, micro-modules, ...



General focus

- Data-Awareness

- Data-Literacy

- ? Humanities as the *interface*

Data Science ↔ Public

→ Data Translators



The Data Science Components

- In the beginning... rather coarsely...
 - Statistics
 - Machine Learning



The Data Science Components

- However, there is more to it...
 - ... it's emerging, dynamically developing field...
 - Data Preprocessing
 - Software Engineering
 - Programming
 - Mathematics
 - **Statistics**
 - **Machine Learning**
 - Data Visualisation
 - Soft Skills



The Data Science Components

- However, there is more to it...
 - ... it's emerging, dynamically developing field...
 - **Data Preprocessing**
 - Software Engineering
 - Programming
 - Mathematics
 - Statistics
 - Machine Learning
 - **Data Visualisation**
 - **Soft Skills**



The Data Science Components

- However, there is more to it...
 - ... it's emerging, dynamically developing field...
 - Data Preprocessing
 - e.g., obtaining data, selecting & defining features...
 - Data Visualisation
 - e.g., composition, comparison, relation...
 - Soft Skills
 - e.g., critical thinking, presentation & storytelling, communication...



UoB/CAL emerging strategy

- General
- Beyond



UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
ARTS AND LAW

UoB/CAL emerging strategy

■ General

- Many existing modules with data elements (Law, EDACS, LCAHM)
- CPDs for other existing modules to raise data awareness



UoB/CAL emerging strategy

■ Beyond

● Summer Schools

- Statistics for Linguistics
- Corpus Linguistics
- our teaching-power contributes globally to many other summer schools, masterclasses, workshops
- ...



UoB/CAL emerging strategy

■ Beyond

- Summer Schools
- EU Twinning Project
 - ~60 PGRs over 3 years; 15 from the UoB/CAL
- West Midlands Combined Authority

Data Skills for Creative Industries

- 30 students with degrees in Humanities looking for job



Thank you very much!

