

MEASURING OPINIONATED COMMUNICATION

WHEN “IRRELEVANT” COMMENTS TURN OUT TO MATTER

THE PROBLEM

Researchers often decide whether a comment is relevant before analysing opinions, positions, or toxicity.

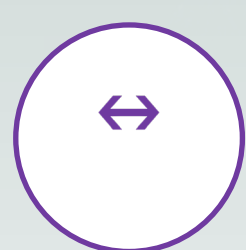
“Why is this [trans issues] talked about so much?
It’s just EVERYWHERE!!”

KEY MESSAGE

There is no universal relevance: what counts as relevant depends on what we want to find out.

Relevant for what?

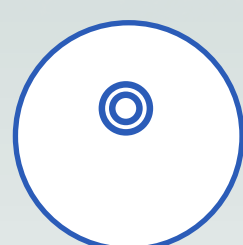
THE SAME COMMENT. DIFFERENT LENSES. DIFFERENT RELEVANCE.



PARTICIPATION

Engages with the discussion.

Relevant.



STANCE (ISSUE)

No direct opinion about trans people, but an attitude to how important the issue is.

Indirectly relevant.



AGENDA / SALIENCE

Questions the amount of attention the topic receives in public debate.

Highly informative.

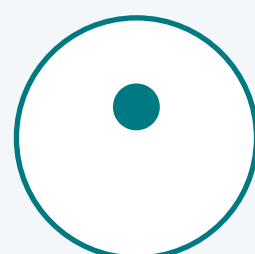


TOXICITY

May contain little or no toxicity.

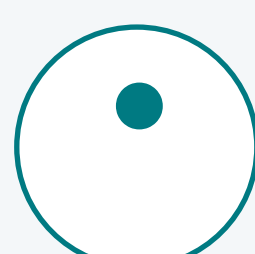
Less relevant.

CORPORA OFTEN HAVE A LIFE BEYOND THE ORIGINAL STUDY



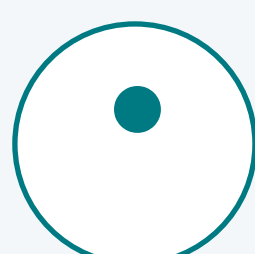
ONLINE DISCUSSION

comments



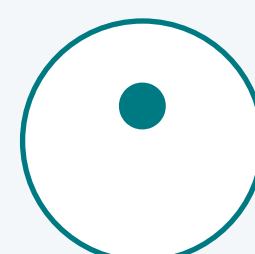
CORPUS BUILT FOR STUDY A

e.g., stance



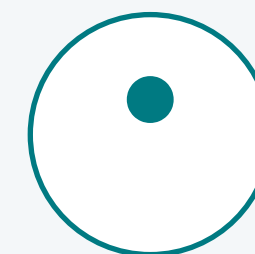
RELEVANT / “IRRELEVANT”

initial filter



CORPUS REUSED FOR STUDY B

e.g., toxicity



NEW QUESTIONS

e.g., agenda

Some comments excluded for one purpose may become useful when the corpus is reused for another.

DESIGNING FOR REUSE

Think about relevance when building the corpus, not only when analysing it.

Where possible, retaining or distinguishing different forms of relevance can make corpus data more versatile and reusable for future research.

A comment doesn’t have to mention the topic to matter. And what matters depends on what we want to find out.

BEFORE DISCARDING A COMMENT AS “IRRELEVANT”, ASK:

RELEVANT FOR WHAT?



Explore the dataset

CLARIN-LT Repository

Jūratė Ruzaitė | Vytautas Magnus University

Based on studies of (mis)gendering debates on Reddit and YouTube