Racism in online interactions

Francesco Bailo (with Gerard Goggin)

Department of Media and Communications, The University of Sydney
Gale Digital Scholar Lab

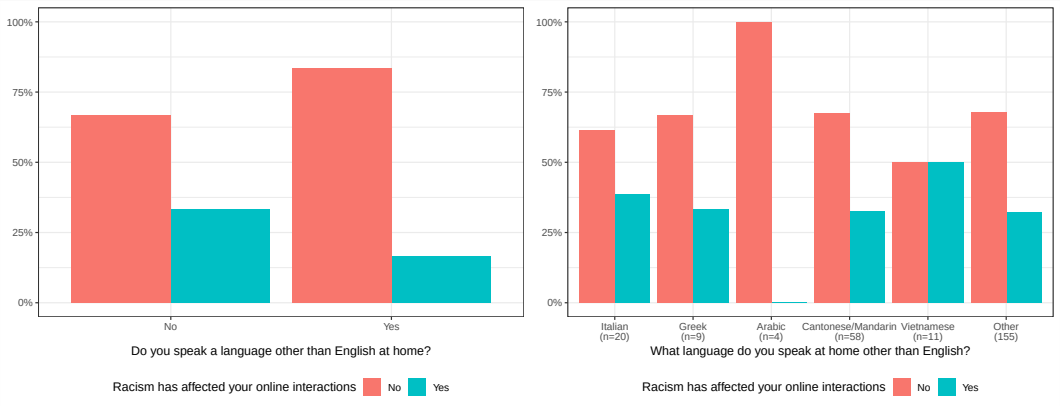
3 October 2018

Justification and research question

Research design

Justification and research question

According to a survey we conducted within the Digital Rights and Governance project among 1,600 Australians, about one-third of respondents who declare to speak a language other than English at home have been subjected to racism in their online interactions. (Goggin et al., 2017)



Research question?

Based on empirical observation, how likely an Australian Twitter user with a non-Anglo-Saxon background is subjected to racism?

Research design

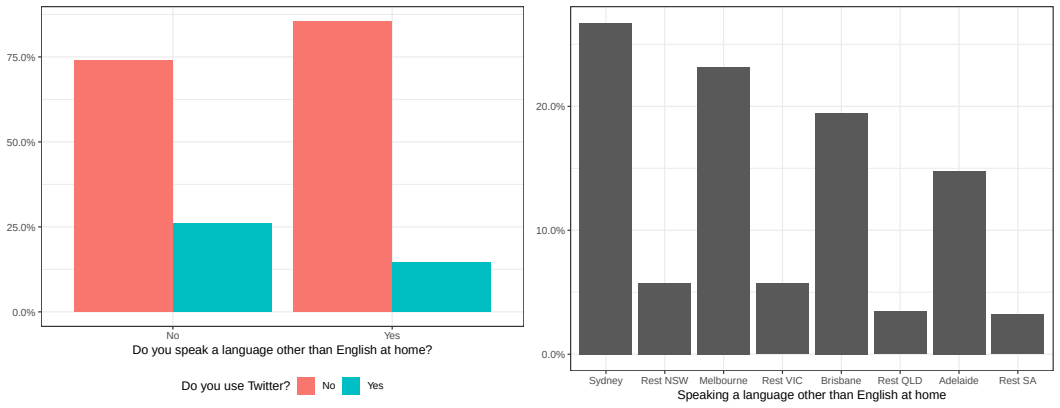
Main challenges

In order to answer to the research question, I need to

1. Create a *representative sample* of Twitter users and tweets;
2. Identify the *ethnic background* of Twitter users;
3. Assess if tweets are *racist* or not.

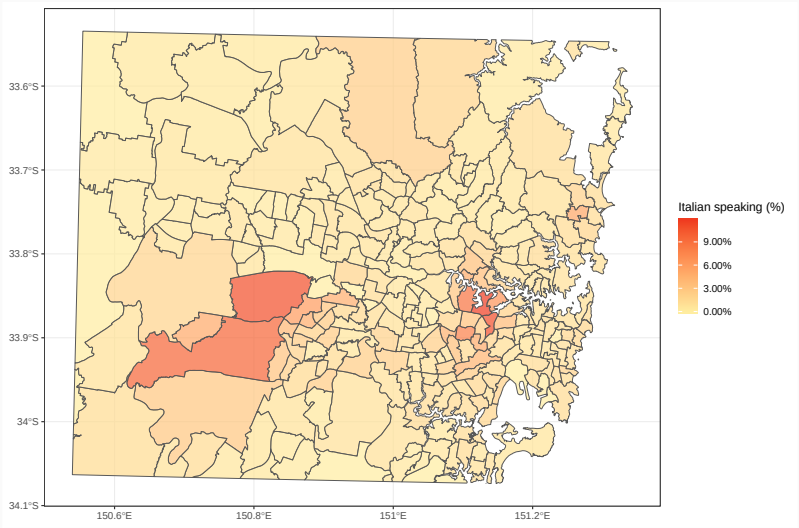
Note: The research project is under revision by the research and ethics committee of the University.

Creating a representative sample of Twitter users and tweets

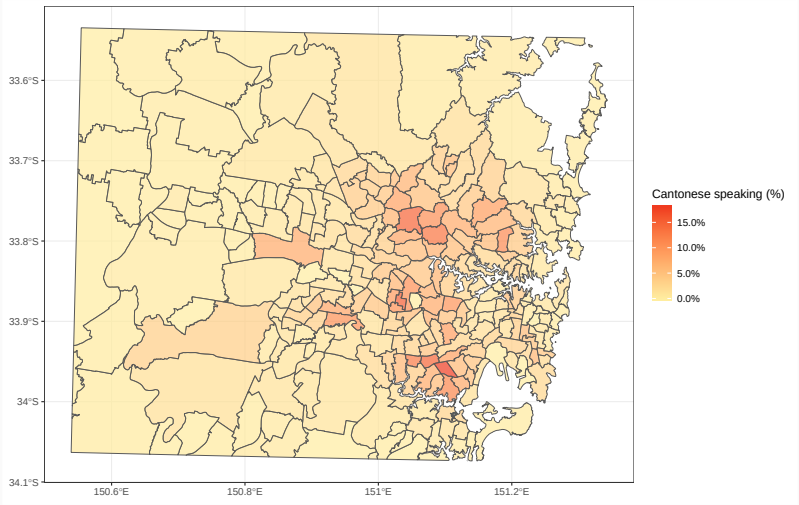


Twitter is used by less than 15% of Internet users who speaks a language other than English but they are highly concentrated in cities.

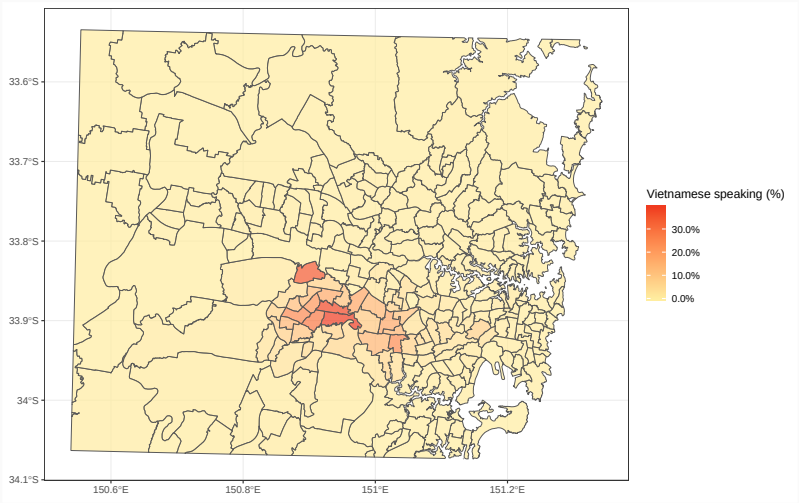
Distribution of Italian speaking population (Sydney, 2016 census)



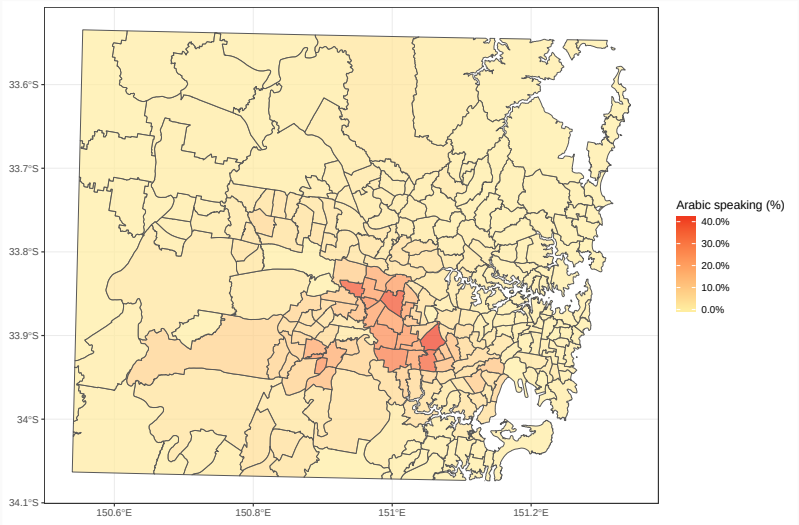
Distribution of Cantonese speaking population (Sydney, 2016 census)



Distribution of Vietnamese speaking population (Sydney, 2016 census)



Distribution of Arabic speaking population (Sydney, 2016 census)



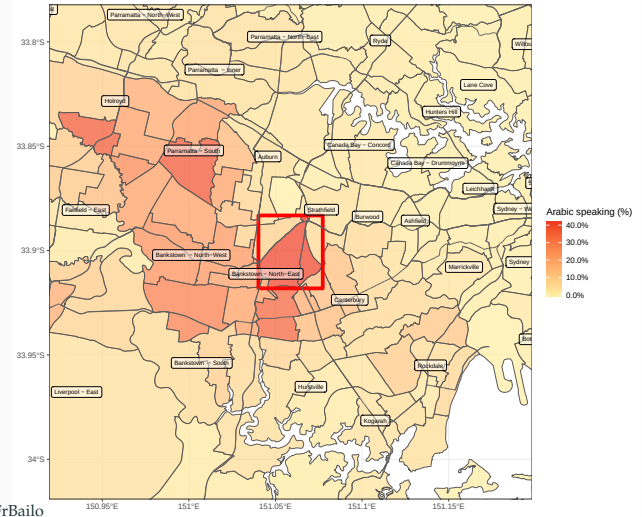
Creating a representative sample of Twitter users and tweets

- 23,401,892 (100%) people living in Australia on census night 2016;
- 6,388,717 (27%) people speaking a language other than English (PLOE);
- 4,400,000 (70%, estim.) PLOE on the Internet;
- 650,000 (14%) PLOE on Twitter.

Residents speaking a language other than English tend to cluster in specific areas.

I can use the filtering functions of the Twitter API to collect tweets published in those areas.

Distribution of Arabic speaking population (Sydney, 2016 census)



Limits to filter realtime Tweets via API for *Greenacre*

`statuses/filter &`

`locations=151.04014,`

`-33.91810, 151.07729,`

`-33.88318`

Testing Greenacre API query

Over 4 days (long weekend):

- 35,584 tweets;
- 41% of tweets are a reply;
- 6,338 Twitter users.

Creating a representative sample of Twitter users and tweets

Defining 30 bounding boxes in Sydney, Melbourne, Brisbane and Adelaide and filtering results from the Streaming API based on the diffusion of languages.

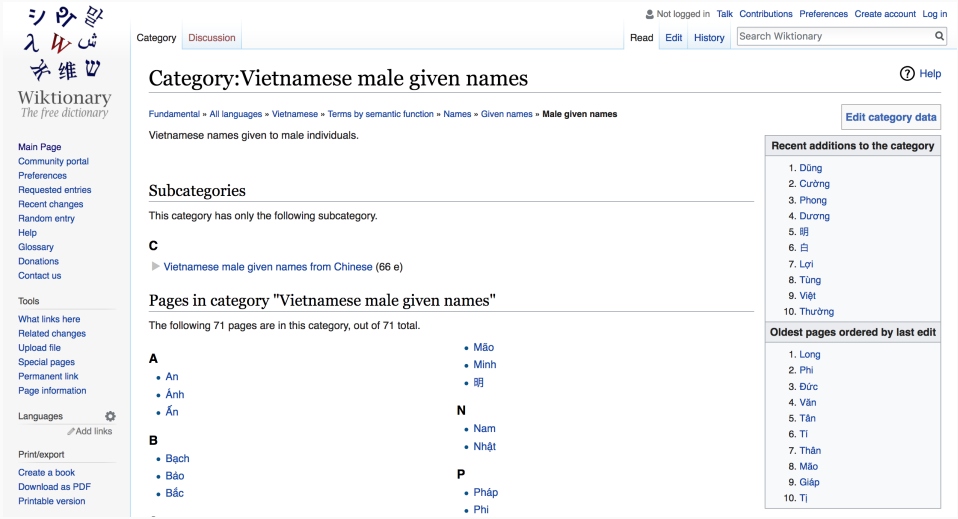
- Italian;
- Greek;
- Arabic;
- Cantonese;
- Mandarin;
- Vietnamese.

Identifying the ethnic background of Twitter users

The identification of the ethnic background of Twitter users is conducted by comparing user names to a dictionary of common first names derived from en.wiktionary.org.

The dictionary is compiled by parsing all given names/surnames and their variations in the categories:

- Italian male given names;
- Italian female given names;
- Greek male given names;
- Greek female given names;
- ...
- English surnames from Chinese;
- Vietnamese male given names.
- Vietnamese female given names.



Wiktionary
The free dictionary

Main Page
Community portal
Preferences
Requested entries
Recent changes
Random entry
Help
Glossary
Donations
Contact us

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information

Languages
Add links

Print/export
Create a book
Download as PDF
Printable version

Not logged in Talk Contributions Preferences Create account Log in

Category Discussion

Read Edit History Search Wiktionary

Category:Vietnamese male given names

Help

Fundamental » All languages » Vietnamese » Terms by semantic function » Names » Given names » **Male given names**

Edit category data

Vietnamese names given to male individuals.

Subcategories

This category has only the following subcategory.

C

▶ Vietnamese male given names from Chinese (66 e)

Pages in category "Vietnamese male given names"

The following 71 pages are in this category, out of 71 total.

A
- An
- Ánh
- Ấn

B
- Bạch
- Bảo
- Bắc
- Mão
- Minh
- 明

N
- Nam
- Nhật

P
- Pháp
- Phi

Recent additions to the category
1. Dũng
2. Cường
3. Phong
4. Dương
5. 明
6. 白
7. Lợi
8. Tùng
9. Việt
10. Thường

Oldest pages ordered by last edit
1. Long
2. Phi
3. Đức
4. Văn
5. Tân
6. Tí
7. Thân
8. Mão
9. Giáp
10. Tị

Identifying the ethnic background of Twitter users

Labelling approaches:

- Simple match (given name/surname is present in Twitter user name);
- Supervised learning: training an algorithm by manually coding a selected sample.

Issues:

- False positives (e.g. the Arabic name `Ali` will match `Ali` short for *Alice*);
- Supervised learning requires significant time investment for manual labelling and assessment of predictions.

Creating a representative sample of Twitter users and tweets

When user names are identified as belonging to PLOE, replies directed to them will be pulled from TrISMA (trisma.org).

TrISMA have been tracking and storing tweets of the 'most active Australian Twitter accounts'.

TrISMA data will allow to expand the dataset longitudinally.

Assessing if tweets are racist or not

Labelling approaches:

Dictionary-based approach : A dictionary was created from the Wiktionary category English ethnic slurs (containing 332 terms).

Supervised learning approach based on three categories *Toxicity*, *Aggression* and *Personal Attacks* based on the manually coded dataset released by the 'Wikipedia Detox project'.

Assessing if tweets are racist or not

The **Dictionary-based approach** will return a binary answer to the question if a tweet contains racist terms.

But the **Supervised learning approach** will return a more nuanced answer based (racism is a subjective experience!) on three dimensions:

Toxicity How strongly the message is 'perceived as likely to make [the recipient] want to leave the discussion'.

Aggression 'How strongly the message is perceived as aggressive' towards the recipient.

Personal Attack '[W]hether it contains a personal attack' towards the recipient.

References

Goggin, G., Vromen, A., Weatherall, K., Martin, F., Adele, W., Sunman, L. & Bailo, F. (2017). *Digital Rights in Australia*. Sydney: The University of Sydney. Retrieved September 6, 2018, from https://ses.library.usyd.edu.au/handle/2123/17587