

Methodology

During this research, I used online interviews and Twitter searches to supplement reports and articles. The participants of the interviews were international scholars that have worked on Sri Lanka's civil war and its aftermath, members of the Tamil diaspora, participants in the 2002 peace negotiation, and members of non-governmental organizations (NGOs) working in Sri Lanka. All interviews were in English and took place via video conferencing. The Twitter searches only analyzed posts using the Latin alphabet.

Twitter analysis

During this research, I used public Twitter posts between 31 December 2008 and 31 August 2020. To access to the tweets, I created a public account, @SLCWdissert, and was approved to use Twitter's Application Programming Interface (API) at the premium level as an academic researcher. I ran two searches, and the Python code I wrote is in Appendix 4. The first search pulled tweets using keywords relating to the Rajapaksas between January 2009 and August 2020. The second showed tweets using the word "Eelam" between May and August 2020.

The use of Twitter in research has grown from 2007, the year after the social media platform was created. Studies using Twitter cross a broad range of subjects such as business, economics, medicine, and political science (Weller, 2014; Zimmer & Proferes, 2014). There is an advantage to using Twitter because of the public access to tweets, Twitter's use of timestamps, hashtags,¹ and location, and the popularity and use of Twitter by people from a wide range of backgrounds (Weller, 2014). There are multiple methods to use Twitter in research such as experiments and quantitative, network, linguistic, or content analysis. Data sets that researchers retrieve include tweets, networks of users, and results of experiments or surveys (Ibid.). I chose to use Twitter in my study because of its ability to reflect "offline events [that] can trigger 'online tension'" (Burnap, et al., 2013, p.96).

At the beginning of the project I planned to use a programmed sentiment analysis for the first search, however, due to time constraints, I could not clean the raw data and feed it into a machine learning program (Gokulakrishnan et al., 2012). Additionally, as English may not be the first language for many of the users, I elected to manually interpret and code the tweets as positive, neutral, or negative. I was able to pick up on context clues and tone that an artificial intelligence program would have read literally.

Data collection

Through the API, I had access to tweets containing keywords posted since 2006 (Twitter, 2020a). I discuss the findings in Chapter 5 and analyze them in the context of militarization and reconciliation in Chapters 6 and 7, respectively. Additional data from the searches are in Appendices 2 and 3.

The code that I wrote uses a Python library called *SearchTweets* to access the tweets. Through this method, I used keywords and dates to search through the database. I searched for "Gotabaya", "Mahinda", "Rajapaksa", and "white van". I searched for the brothers individually and their family to pull the attitudes that people had towards the president and defense secretary during the war, the opinions towards the family during the regime, and reactions towards the

¹ A hashtag (#) is a keyword or phrase that places the tweet in the context of the wider Twitter conversation about that topic (Twitter, 2020d). A hashtag cannot have spaces between the words.

brothers at the time of the Easter bombings and the first months of Gotabaya's presidency. I selected "white van" because it was a defining characteristic of Sri Lanka's censorship and citizen oversight.

Twitter is set to Greenwich Mean Time, five and a half hours behind Sri Lanka. To accommodate this in the search the "from" date is the day before the period I want to examine. I was unable to specify the language or location of the tweets, so when I cleaned the data to prepare it for analysis, I discarded tweets that did not use the Latin Alphabet or did not meet my location criteria. In the first search, I removed 2425 tweets out of the total 3219 that I collected, leaving me with 794 to analyze. The breakdown of tweets per keyword and date are in Table 1. I manually coded the tweets as positive (1), neutral (0), or negative (-1). I also coded for bots, verified accounts,² and hashtags; these are in Appendix 2. I designated the tweets as positive, neutral, or negative depending on the text available; in some cases, I did not have the entire tweet, and I could not follow links to articles that expired.

The second search that I ran pulled tweets with the word "Eelam" posted each month between May and August 2020. I searched for the opinions of the Tamil diaspora, so I discarded tweets that originated from Sri Lanka or did not specify their location. I kept 122 of the 400 tweets that were pulled and coded the tweets and hashtags that used the Latin alphabet.

There are ethical considerations one must account for when using Twitter in research. As the contemporary use of social media evolves faster than good practices for research are updated, there are concerns regarding the protection of users' anonymity (Williams, Burnap, & Sloan, 2017; Zimmer & Proferes, 2014). I set the location of my account to Sri Lanka and followed verified accounts so I could observe the trending hashtags and events leading up to the parliamentary election, but I did not retweet or like any of the posts and I did not direct message accounts. A few accounts reciprocated my follow by following me back, but I did not otherwise interact with them.

Limitations

When I collected the Twitter data, my inability to read Sinhala and Tamil also limited the tweets that I kept. As a result, I could not use tweets that did not use the Latin alphabet. Additionally, the code did not have the capacity to specify the geographic area from which I pulled the tweets, so I discarded tweets that did not fit my criteria. For example, tweets could have originated from Sri Lanka but if the location was set as "None" I could not keep them.

Online, there is not an even distribution of Twitter users across the population. Due to the nature of the tweets and the information that users share I was not able to assess the demographics of the posters. It is additionally difficult to identify the location of users at a level that would enable further analysis (Weller, 2014) as not all profiles or tweets provide specific geographic locations (Twitter, 2020b). Due to these limitations, the information I gathered consists of individuals' observations and is a snapshot of the social and political climate in Sri Lanka.

² A bot is an account that posts and interacts with other tweets automatically by liking or retweeting them (Wojcik, et al., 2018) and Twitter (2020c) defines a verified account as one that is of public interest and is determined to be authentic.

Table 1. Tweets by keyword and date.

Keyword	Date Range	Tweets kept	Score
Mahinda	31 December 2008-31 May 2009	48	0.354
Gotabhaya	31 December 2008-31 May 2009	18	0.056
Rajapaksa	1 June 2009 - 31 December 2014	409	0.054
white van	31 December 2008-31 December 2014	0	NA
Rajapaksa	31 December 2014 -31 December 2015	93	-0.43
Rajapaksa	20 April 2019-30 April 2019	31	0.484
Gotabhaya	20 April 2019-18 August 2020	144	0.125
Mahinda	31 December 2019-18 August 2020	51	0.02

Appendix 2. First Twitter search categories

Category	Tweets Kept	Score
Bot	122	0.082
Verified	38	-0.237
Hashtag	193	-0.238

Appendix 3. Second Twitter search hashtags

Hashtag	Count
AntiTamilPogrom	2
BlackJuly1983	2
eelam	7
Google	3
Justice4DisappearedTamils	3
Justice4EnforcedDisappeared Tamils	1
Justiceforatamils	1
lka	9
non-English	9
prabakaran	5
Prabhakaran	2
soltng	1

srilanka	7
sundarpichai	3
tamil	16
Tamileelam	3
TamilNadu	1
Thalaivar	1

Appendix 4. Twitter API search code³

```
import json
import pandas as pd
from searchtweets import load_credentials, gen_rule_payload, ResultStream
import yaml

# save credentials to yaml file (Rastogi, 2020)
config = dict(
    search_tweets_api=dict(
        account_type='premium',

endpoint='https://api.twitter.com/1.1/tweets/search/fullarchive/NAME.json',
        consumer_key='CONSUMER KEY',
        consumer_secret='CONSUMER SECRET'
    )
)
with open('twitterkeys.yaml', 'w') as config_file:
    yaml.dump(config, config_file, default_flow_style=False)

# load credentials and create bearer token (Rastogi, 2020; Hammar, 2019)
premium_search_args = load_credentials("twitterkeys.yaml",
                                       yaml_key="search_tweets_api",
                                       env_overwrite=False)

print(premium_search_args)

# search rule (Rastogi, 2020; GitHub, 2020; Hammar, 2019)
rule = gen_rule_payload("KEYWORD",
                       results_per_call = 100,
                       from_date = 'YYYY-MM-DD',
                       to_date = 'YYYY-MM-DD'
                       )

rs = ResultStream(rule_payload=rule,
                  max_results=100,
                  **premium_search_args)

print(rule)
print(rs)

# save results to json file (Rastogi, 2020; Hammar, 2019)
with open('FILE.json', 'a', encoding='utf-8') as f:
    for tweet in rs.stream():
        json.dump(tweet, f)
```

³ I wrote this code with guidance from Gregory (2020), GitHub (2020), Rastogi (2020), and Hammar (2019).

```
f.write('\n')

# convert json file to csv (Gregory, 2020)
df = pd.read_json('FILE.json', lines=True)
df.to_csv('FILE.csv')
print('done')
```

References

- Burnap, P., Rana, O. F., Avis, N., Williams, M., Housley, W., Edwards, A., Morgan, J., Sloan, L. (2013). Detecting tension in online communities with computational Twitter analysis. *Technological Forecasting & Social Change*, 95. [Online]. Available at: <http://dx.doi.org/10.1016/j.techfore.2013.04.013>. [Accessed 5 June 2020].
- GitHub. (2020). *twitterdev/search-tweets-python*. [Online]. GitHub, Inc. Available at: <https://twitterdev.github.io/search-tweets-python/>. [Accessed 14 July 2020].
- Gokulakrishnan, B., Priyanthan, P., Ragavan, T., Prasath, N., and Perera, A. (2012). Opinion mining and sentiment analysis on a Twitter data stream. *The International Conference on Advances in ICT for Emerging Regions*. [Online]. Available at: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6423033>. [Accessed 31 July 2020].
- Gregory, M. (2020). JSON to CSV code. (Python). (see Appendix 3).
- Hammar, L. (2019). *Collecting old Tweets with the Twitter Premium API and Python*. [Online]. Luca Hammar. Last updated: 4 June 2020. Available at: <https://lucahammer.com/2019/11/05/collecting-old-tweets-with-the-twitter-premium-api-and-python/>. [Accessed 14 July 2020].
- Rastogi, P. (2020). *Extracting Tweets using Twitter Premium Search API and Python*. [Online]. Medium. Last updated: 1 May 2020. Available at: <https://medium.com/swlh/extracting-tweets-using-twitter-premium-search-api-and-python-2d025144e8a4>. [Accessed 14 July 2020].
- Twitter. (2020a). *About Twitter's APIs*. [Online]. Twitter. Available at: <https://help.twitter.com/en/rules-and-policies/twitter-api>. [Accessed 5 June 2020].
- Twitter. (2020b). *Docs*. [Online]. Twitter. Available at: <https://developer.twitter.com/en/docs>. [Accessed 30 June 2020].
- Weller, K. (2014). What do we get from Twitter – and what not? A close look at Twitter research in the social sciences. *Knowledge Organization*, 41. [Online]. Available at: <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-47768-2>. [Accessed 4 June 2020].
- Williams, M. L., Burnap, P., and Sloan, L. (2017). Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and

algorithmic estimation. *Sociology*, 51. [Online]. Available at:
<https://doi.org/10.1177/0038038517708140>. [Accessed 4 June 2020].

Zimmer, M. and Proferes, N. J. (2014). A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66. [Online]. Available at:
<https://doi.org/10.1108/AJIM-09-2013-0083>. [Accessed 4 June 2020].