

Conclusion

Chin Man KWAN

January 16, 2022

Contents

Chapter 1

**examiner: Aragorn II Elessar # only
required if PhD; otherwise delete or
comment out**

Placeholder

1.1 Metadata

1.2 Front Matter

1.3 Bibliography

1.4 PDF Formatting

1.5 Output Options

Chapter 2

Introduction

Placeholder

Chapter 3

Literature Review

Placeholder

3.1 Theories

3.1.1 What is national identity?

3.1.2 Impact of conceptions of national identity on attitudes towards immigrants

3.1.3 Chineseness in Hong Kong

3.2 Empirics

3.2.1 How do different political camps in Hong Kong perceive national identity?

The pro-Beijing camp

The pro-Democracy camp (or pan-democrats)

The localist camp

3.2.2 Asylum seekers in Hong Kong

Hong Kong's policy regime

Responses by political camps

Chapter 3

Methods

After reviewing both the theories of the impact of national identity on xenophobia as well as how political camps in Hong Kong imagine the city's national identity, how could we know the possible associations between each camp's definition of the city's national identity and its attitude towards asylum seekers who are mainly non-ethnic Chinese? This section will discuss the strategy of answering this question.

3.1 Hypotheses

Based on the aforementioned literature in the previous section, I will formulate hypotheses to be tested and explain why I expect them to go in certain directions. The main hypothesis to be tested will be:

H₁: Pro-Beijing media outlets are more likely to negatively portray asylum seekers than media outlets from other political camps.

I expect the pro-Beijing camp news media to frame asylum seekers more negatively based on the literature on the relationship between national identity and xenophobia (e.g. Hjerme (1998) and Heath and Tilley (2005)) which states that a more ethnic-based conception of national identity is generally associated with a higher degree of xenophobia towards immigrants, of which asylum seekers is one of the (most vulnerable) subgroups of immigrants. Accordingly, since pro-Beijing politicians and officials (including the Chinese President *Xi Jinping* himself) (Lin and Jackson 2021) resort heavily to ethnic-based discourses for justifying Chineseness of the Hong Kong identity, pro-Beijing media outlets can be expected to hold more negative attitudes towards the mostly non-ethnic Chinese non-refoulement claimants. In particular, the pro-Beijing camp's emphasis on how Hong Kong citizens belong to the

3. Methods

greater Chinese national identity based on shared ethnicity with mainland Chinese might lead to its affiliated media outlets to portray asylum seekers as a threat to society since their ethnic distinctiveness marks them as ever-being strangers of the family of the majority Chinese population as Erni (2012) describes.

By contrast, since the pro-democracy camp focuses more on civic values (e.g. civil liberties, the rule of law) while defining the Hong Kong identity, the non-Chineseness of asylum seekers should appear less threatening to the Hong Kong identity from the pro-democracy camp's point of view. Lastly, the localist camp's primary focus on the mainland-Hong-Kong division regarding the city's identity again prevents a clear direction for H_1 from being formulated beforehand.

3.2 Data collection

I will study news media reports on asylum seekers and non-refoulement claimants in Hong Kong by local mass media during the year of 2019. Specifically, media publishing in **Chinese** in printed format will be included in the data¹. For the retrieval of the relevant newspaper articles, I will rely on the *WiseNews* database which can fetch news articles with search keywords published in Hong Kong². While it would have been more ideal to also include online news media into the dataset as Lee (2018) mentions that they have become quite prominent in the 2010s for providing counter-narratives to the mainstream media which are often co-opted by the Chinese Government³, *WiseNews* does not directly provide the texts of online media news article and thus requires additional steps of web scrapping which are, unfortunately, beyond the author's capacity by the time of writing. That being said, including only the printed news media during 2019 would still cover 16 of the well-known news media with associations to different political camps according to the survey on media credibility by the Chinese University of Hong Kong (CUHK 2020). In other words, just searching for printed news media outlets still allows me to discover whether the attitudes of reporting on asylum seekers by some of the most well-known and widely circulated media outlets in Hong Kong might be related to their political associations.

The reason for choosing to collect media reports published in 2019 is because it was when the Anti-Extradition Law Amendment protest first broke out in June and then lasted for several months, a

¹The full list of the included newspaper outlets is provided in *table 3.1*.

²I would like to thank my friend in Hong Kong helping me gain access to the database, which would otherwise have been extremely difficult.

³It should be noted, however, that by the very beginning of 2022, several online media outlets such as *Stand News* has defuncted after the flagship pro-Democracy newspaper outlet *Apple Daily* had dissolved in June 2021. The implication of this phenomenon to the media's narrative on asylum seekers in Hong Kong will be discussed further in the conclusion section.

3. Methods

period when one can argue the manifestation of political ideologies should be the most salient for all the political camps. This should indeed be the case for both the pro-democracy camp whose supporters participated in several large-scale protests and sometimes even engaged in numerous violent clashes with the Hong Kong police to oppose a law that would have permitted extradition of fugitives to mainland China, and the pro-Beijing camp whose authority was once again immensely challenged after the 2014 Umbrella Movement. Accordingly, it would be interesting to investigate whether the issue of asylum seekers still remained its salience and was reported in similar manners given the co-occurrence of another arguably more salient political event (i.e. the anti-Extradition Law protests), as compared to when the issue first had become prominent back in 2016. Specifically, analysing newspaper articles in 2019 provides an opportunity to observe whether and how asylum seekers in Hong Kong might be represented by the media in tandem with major political events.

It is worth discussing some of the limitations of using media reports as the data for this thesis. For starters, media reports do not necessarily reflect how individuals identifying with each political camp might perceive asylum seekers directly and might not be fit for investigating the causal mechanism of how definition of Hong Kong's national identity may cause a political camp to frame asylum seekers in particular ways. Despite these shortcomings, media reports can still adequately help prompt the hypothesis of whether depictions of asylum seekers in Hong Kong might be associated with their political camp's definition of the city's national identity. In fact, previous research about the framing of asylum seekers by mass media (Cooper et al. 2017; McKay, Thomas, and Warwick Blood 2011; I. Ng, Choi, and Chan 2019; O'Doherty and Lecouteur 2007) point to its potential impact on affecting how the public may perceive the issue via being a major provider of related information, and sometimes these media reports may even serve to justify policymaking or reflect the government's position on the issue. Drawing from Fong (2017) and Lee (2018), the media industry in Hong Kong is quite remarked for their political stances, and quite a number of outlets have their owners being affiliated with the Chinese and/or HKSAR Governments. Therefore, reports published by media outlets should be able to reflect political camps' attitudes towards asylum seekers and, perhaps more importantly, the version of framing that they would like to persuade their readers to accept, thereby making them a suitable data source to test the hypotheses proposed above.

Language-wise, including only Chinese-language news media omits how asylum seekers may be represented by English-language media in Hong Kong (e.g. the South China Morning Post) and thus

3. Methods

cannot investigate whether the representation of asylum seekers by media outlets to potential non-Chinese audience may also vary based on each outlet’s political affiliation. Moreover, solely focusing on textual content of the media reports means discarding graphical and (if also published online) audio-visual materials accompanied in a given news report, which may contain additional information that are not observed in the texts of news articles. Nevertheless, analysing the textual content of the media reports should still yield a considerable amount of information about the representation of asylum seekers in Hong Kong by media outlets linked to different political camps, since after all a large proportion of the content in news reports are represented textually.

To search for relevant media reports on WiseNews, I will refer to the keywords used by I. Ng, Choi, and Chan (2019) which refer to the group of population-in-interest. These terms can either be directly describing this group of population (e.g. refugees, illegal immigrants) or policies related to asylum seeking (e.g. non-refoulement claim, immigration paper). *Table 3.1* lists all the Chinese terms used for searching the relevant news articles on WiseNews and their meanings in English. Additional notes are used for indicating whether some articles were omitted and why. The author also omitted articles only mentioning the keywords briefly throughout the main text which discusses issues not directly related to asylum seekers. As for the keywords “South Asian”, “illegal immigration” and “human snake”, the author only included news entries fetched from these keywords if at least one of the other keywords directly referring to asylum seekers (i.e. non-refoulement claim(ant), torture claim(ant), asylum seeking, refugee and immigration paper) co-appeared in either the title or main text. This is to ensure that the news articles in the dataset are obviously describing non-refoulement claimants in Hong Kong.

Table 3.1: Translation of the keywords used to search for media reports on WiseNews

Original terms in Chinese pinyin	Literal meaning in English	Notes
mian3 qian3fan3 sheng1qing3	Non-refoulement claim(ant)	
ku4xing2 sheng1qing3	Torture claim(ant)	
feilfa3 ru4jing4, ren2she2	Illegal immigration, human snake (alias of illegal immigrants)	Excluded illegal immigration from mainland China
xun2qiu2 bi4hu4	Asylum seeking	Excluded asylum seekers <i>outside of</i> Hong Kong
nan4min2	Refugee	Excluded refugees <i>outside of</i> Hong Kong and Hong Kong activists seeking refuge overseas
nan2ya4	South Asian	Articles about the South Asian region are excluded.

3. Methods

Original terms in Chinese pinyin	Literal meaning in English	Notes
xing2 jiel zhi3	Immigration paper	Documents issued by the Immigration Department to non-refoulement claimants.

3.3 Modelling

Since the period under study spans for the year 2019, the dataset contains almost 700 articles in total which would be extremely time- and resource-consuming to qualitatively analyse each article line by line, whereas only focusing on a small subset of the articles may risk losing generalisability. On the other hand, while it is true that using human coding (whether it be done by trained experts or crowd coding platforms) for sentiment analysis will be likely to yield the most accurate results as van Atteveltdt, van der Velden, and Boukes (2021) have noted, doing so will be too resource-intensive for a master-level thesis, and this method may also fall short of reliability, i.e. how other researchers may replicate the study if wanted. Despite some of the caveats listed by van Atteveltdt, van der Velden, and Boukes (2021) (e.g. accidentally identifying spurious relationships between variables, problem of overfitting) which can be remedied with corresponding solutions, I will use supervised machine learning to test the hypothesis. This means H_1 will be treated as a sentiment analysis problem which classifies the *polarity* of the news articles (i.e. whether they reported on asylum seekers *positively*, *neutrally* or *negatively*). How the independent (both the media outlet’s political camp and other controls) and dependent variables are operationalised and coded will be discussed in the next section, and for now I will talk about what kinds of tools and models will be used for the analysis.

Starting with the models that will be used for the analysis, I will rely on some commonly used classification models. Candidate models include logistic regression, support vector machine (SVM), random forest and gradient boosting. Although van Atteveltdt, van der Velden, and Boukes (2021) also mention that deep learning is another popular choice of machine learning model for sentiment analysis tasks, I did not use such a model for this paper due to the relatively small size of the dataset at hand as well as the difficulty of interpreting the features’ impact on the dependent variable directly from neural network models. I will first run some baseline models with the default parameter settings of the models, and then select one for further tuning based on their performance on their *f1 scores* (which measures the harmonic mean of *precision* and *recall*)⁴. Using f1 score as the evaluation metric can better measure

⁴Just a reminder, the formula of calculating a model’s F1 score is: $2 \cdot \frac{(\text{precision} \cdot \text{recall})}{(\text{precision} + \text{recall})}$ (Pedregosa et al. 2011).

3. Methods

a model's performance when there is class imbalance in the dependent variable compared to metrics like accuracy and the ROC-AUC score.

The model which obtains the overall highest fl score will then be used for calculating the *SHAP values* of the features (Lundberg 2022) which evaluate how much impact each feature has on the model prediction when the features is at certain values versus at its baseline value. In essence, the higher the SHAP value of a feature, the higher its impact of the model's prediction.

In order to pre-process the articles into suitable formats as machine learning model inputs, I will also transform the news articles into a term-document matrix which records the frequency of each word (i.e. token) appearing in each article (i.e. document). Instead of using a simple bag-of-word approach, I will use the term-frequency inverse-document-frequency (TF-IDF) matrix transformation which also considers in how many documents a word appears out of all the documents in the corpus. This is because words that appear too frequently across all the documents in the corpus may not convey much meaning of each document and thus should be treated as stop words. Due to the relatively small size of the dataset (with only around 550 articles), I will also adopt non-negative matrix factorisation (NMF) to reduce the dimensionality of the dataset to prevent overfitting. According to Stevens et al. (2012) (p.953), the matrix denoted as H which captures the weight of each topic (as columns) in each document (as rows) of the corpus can help summarise the information of the articles in terms of which topic(s) they primarily focus on. Thus, I will use NMF to transform the TF-IDF matrix into another which summarises the news articles into a certain number of pre-defined latent topics (the exact number will be figured out in the *preprocessing* section later), and then use this matrix as part of the inputs for the models.

The majority of the implementation from data preprocessing, modelling to validation will be done via the `scikit-learn` package in Python (Pedregosa et al. 2011), whereas the tokenisation of the Chinese-language news articles will be done by the `jieba` package which is designed for natural language processing in Chinese (Junyi 2021). For better tokenisation of the articles, I will also remove stop words, punctuations and hyper-links of the texts and add words related to Hong Kong politics (downloaded from elgarteo (2021)) and asylum seekers into the dictionary. As for the gradient boosting model, the `XGBoost` package (Chen et al. 2015) will be used since it is one of the most renowned machine learning models in data science, and it also offers a lot of options for tuning the model for better performance.

3. Methods

3.4 Operationalisation

3.4.1 Main independent variable

Starting from the main independent variable which classifies the associated political camp of the media outlets, I will mainly refer to the literature exploring how Hong Kong's media industry is intertwined with politics and perform additional research when needed. I consulted the articles by Fong (2017) and Lee (2018) on the development of post-1997 Hong Kong's media industry, of which they have documented the co-optation of Hong Kong media outlets by the Chinese Government (e.g. the major shareholders holding positions in such as Chinese People's Political Consultative Conference). Building on the concept of "allocative control" (p.11) by Lee (2018) about media outlet owners having the ultimate decision on the personnel and resources at the newsroom's disposal for reporting news which may lead to self-censorship of the frontline journalists, I will code media outlets whose owners are affiliated with the official Chinese and/or Hong Kong Government organisations as belonging to the pro-Beijing camp. As for those media owners with less apparent links with the Chinese or Hong Kong Governments, I will decide the classification according to whether the owners are known to be affiliated with a certain political camp and how a given media outlet is perceived for its political stance. Media outlets without apparent political stances are labelled as "neutral". Table 3.2 is the resulting classification scheme. Note that due to the medium of publication of localist news media being almost exclusively online, none of the outlets in that camp was included in the dataset. Therefore, it would not be possible to examine how the localist media outlets might report on asylum seekers and non-refoulement claimants in news articles.

Table 3.2: Classification of the associated political camp of media outlets in Hong Kong included into the sample, source: Fong (2017); Lee (2018); author's own research

Associated political camp	Media outlets
Pro-Beijing camp (directly owned by the Liaison Office or the HKSAR Government)	Wen Wei Po, Hong Kong Government News, Hong Kong Commercial Daily, Ta Kung Pao
Pro-Beijing Camp (owners affiliated with political institutions on mainland China/ HKSAR Government)	Oriental Daily News, Sing Tao Daily, Sing Pao, Headline Daily, Hong Kong Economic Journal, Hong Kong Economic Times, am730, Sky Post, HK01 Newspaper
Pro-Democracy camp	Apple Daily, Kung Kao Po
Localist camp	N/A
Neutral	Metro Daily, Ming Pao Daily News

3. Methods

3.4.2 Dependent variable

As mentioned before, the dependent variable of H_1 will be the sentiment of the news articles towards asylum seekers in the polarities of *positive*, *neutral* and *negative*. But how exactly should these three polarities mean in this context? Again referring to the literature can help devise a sound coding scheme. Drawing McKay, Thomas, and Warwick Blood (2011), Cooper et al. (2017) and I. Ng, Choi, and Chan (2019), *positive* articles usually acclaim the person or event in concern, describe personal stories, report expert opinions without derogatory terms or depict refugees favourably, *neutral* articles simply report an incident without an overt sentiment or opinion towards particular stakeholders and with little interpretation by the journalists, and *negative* articles frame asylum seekers either as bringing problems to the host society or use inaccurate terms to delegitimise this group of population. I will follow these coding criteria for the polarities of the news articles in the dataset since these definitions suit the purpose of this study. The numerical representation of the polarities in the dataset will be as follows, namely, positive articles will be coded as 2, neutral articles will be coded as 1, and negative articles will be coded as 0.

For the context of Hong Kong about the “incorrect” terms and information used for describing asylum seekers, I will refer to expert opinions about whether descriptions of asylum seekers in Hong Kong are founded⁵. For example, I. Ng, Choi, and Chan (2019) point out that the term “fake/bogus refugees” (*jia3 nan4min2*) should be considered as both derogatory and inaccurate because those applying for non-refoulement claims are not automatically qualified as refugees while awaiting the final decisions of their applications. *Therefore, the polarity of news articles in which the news media directly refers non-refoulement claimants as “fake refugees” in its reportage will be coded as negative, but the report should be coded at least as neutral instead if the news outlet is merely quoting someone who describes non-refoulement claimants as “fake refugees”.* Furthermore, I. Ng (2020) mentions that non-refoulement claimants are forced to be illegal immigrants by the policy design of the HKSAR Government so that this group of population can be excluded from social rights that legal residents enjoy. News articles equating non-refoulement claimants as illegal immigrants will thus also be coded as negative since such reportage is justifying the exclusion of non-refoulement claimants by resorting to their illegality that had been forced upon by the HKSAR Government.

⁵A detailed discussion of the validity of the depiction of asylum seekers in Hong Kong can be found in I. Ng (2020) and I. Ng, Choi, and Chan (2019).

3. Methods

3.4.3 Control variables

I will also incorporate some control variables to see if the association between political camps and polarities of the news articles still remains significant afterwards. For starters, the time when the articles were published will be included as a control, though I will bin the twelve months into four evenly split quarters (*Quarter*) with the first quarter beginning in January and ending in March. Since 2019 was also the year when one of the most large-scale and long-lasting pro-democracy protests broke out, It would be interesting to see how asylum seekers may be portrayed during the peak of the protest. Moreover, the presence of racial labels will be included as a control, since it is possible that more overtly racialised discourses may be associated with more negative sentiments towards asylum seekers in general according to the literature. This variable will be named *Racial_label* in the dataset and will be binary, with 1 meaning there is presence of racial labels in the main text and 0 otherwise. Racial labels referring to both regional (e.g. South Asia, Africa) and national (e.g. the nationalities included in figure 2.2) will be counted. Lastly, the metadata of the news articles, namely the character counts of titles and main texts, will also be added to see if the lengths of the articles may be associated with the sentiment towards asylum seekers.

Chapter 4

Results

Placeholder

4.1 Exploratory data analysis (EDA)

4.1.1 How does the number of news articles vary by political camps and month?

4.1.2 Polarities of the news articles

4.1.3 Presence of racial labels

4.1.4 Character lengths of news articles and titles

4.2 Sentiment analysis

4.2.1 Preprocessing

4.2.2 Training the model

4.3 Is the pro-Beijing camp more likely to portray asylum seekers in 2019 more negatively than other outlets?

4.3.1 Overall magnitudes of SHAP values for each feature

4.3.2 SHAP values of predicting negative news articles

4.3.3 SHAP values of predicting neutral news articles

4.3.4 SHAP values of predicting positive news articles

4.4 Discussion

Chapter 5

Conclusion

In conclusion, it is found that at least in 2019, the political camp of media outlets was associated with their attitudes towards asylum seekers in Hong Kong. Specifically,

- I will sum up the findings after editing the main sections

5.1 How might the instigation of the National Security Law affect the public discourse on asylum seekers in Hong Kong?

Just a year after the anti-extradition law protest had started and once again mobilised a huge section of Hong Kong's society against the authority, the HKSAR Government promulgated the National Security Law in July 2020 which aims to tighten the control over the political landscape of Hong Kong. As a result, non-pro-Beijing forces have witnessed their political influence being severely restricted, both in official institutions and the society. Even though the Legislative Council had been one of the few avenues where pro-democracy activists and politicians could run for office (despite lacking substantive authority to participate in policymaking by proposing bills), with the conclusion of the recent 2021 Legislative Council election after an overhaul of the electoral system which essentially permits candidacy only to the "patriots" loyal to the Beijing and HKSAR Governments (Lau and Yam 2021), the door for pro-democracy camp to advocate alternative political discourses has been shut. This means when pro-Beijing lawmakers propose non-binding bills to combat the so-called "fake refugee" problems as Holden Chow did back in 2016 (Cheng 2016), such bills will very unlikely be opposed by others, and the voting results on the however unbinding bills may then be used by the Government to justify more stringent measures against non-refoulement claimants.

5. Conclusion

Even the presence of alternative media which strive to provide an alternative narrative of news contrary to the state-coopted mainstream media as Lee (2018) describes may not be the case anymore. In June 2021, the flagship pro-democracy newspaper *Apple Daily* was forced to shut down after the HKSAR Government had frozen the media outlet's asset under the pretext of violating the National Security Law (Tan 2021). Just days before the year 2021 ended, the popular pro-democracy online media *Stand News* had to shut down as well after the Hong Kong police raided its office and arrested senior staff (E. Ng and Pomfret 2021), followed by the closure of another online pro-democracy news outlets days after ("Hong Kong: Pro-democracy Website Citizen News Is Latest to Close" 2022). It is likely that more alternative media may be pressured to defunct in the future as well due to the threat from the National Security Law, and this implies that alternative discourses will exist in the media industry of Hong Kong. Consequently, since the pro-Beijing camp media in general are more hostile to non-refoulement claimants, it is likely that Hong Kong's media landscape will be dominated by more negative rhetoric towards asylum seekers in the city given that pro-democracy outlets are now finding increasing difficulties to operate without running into legal repercussion.

With two major channels of formulating and propagating political discourses being shut off, it can be expected that the pro-Beijing camp will face much less counter-discourse from the oppositions while attempting to frame certain issues, including on the status and legitimacy of non-refoulement claimants. As the study by Boomgaarden and Vliegenthart (2009) shows, the tone of how asylum seekers are reported in news media is negatively associated with the public attitude towards asylum seekers in later periods, meaning that more pejorative coverage of refugees by the media may cause the public to be more averse towards this group of vulnerable population, and the effect of negative news coverage may become even more salient when the levels of immigration and asylum seeking applications are perceived to be high.

In short, the gradual decrease of