



Al Lim

June 14th, 2021

Methodological Pivoting in COVID-19: Experimenting with Critical Corpus-based Analysis

Questions about methods and data continue to animate my thinking about research. COVID-19 has been resurging in Southeast Asia at the time of writing this in the middle of 2021. The restrictions around in-person ethnographic work on research call for continued thinking about alternative research methodologies such as digital ethnography and remote forms of research. What methodological intersections or adaptations can generate in-depth social science research, and how can one gain proficiency in these forms of research methodologies?, *writes Al Lim*

My methods seminars never included “What to do during a pandemic” before COVID-19. Like many others, COVID-19 was a major disruption to research plans. The amount of time for research allocated to master’s students for their dissertation is not long. By the time the proposal is written, ethics approved, and one gets to their field site (especially for ethnographic work), it is a matter of weeks to analyze, write up, and submit the dissertation. Producing a substantive piece of work in such a short amount of time is quite a challenge, and even more so with the pandemic.

This reflection explains what I ended up doing for my master’s research, pivoting my research methods towards critical corpus-based analysis to study smart urban surveillance. I hope that this provides some resources for master’s researchers who might be interested in the intersections of qualitative and quantitative methods, or who might also be in the midst of experimenting and adapting various forms of social science research.

The Pivot

Initially, I had planned to investigate the smart surveillance “Eagle Eyes” project in Phuket with ethnographic fieldwork. This was part of the ASEAN Smart Cities Network project that the city had proposed, and Kong and Woods (2018; 2019) have masterfully set a regional research agenda that called for greater empirical work. I was trying to draw from literature on smart cities and surveillance from security studies, critical geography, and urban studies, which help to unpack the (mis)alignment between existing practices of community policing and the newly introduced CCTVs. The most important part to this ethnographic plan was getting there.

Cue late March 2020 in London—there was a particularly anxiety-inducing Thursday when seemingly every international student that I knew had decided to book flights home. Colleagues that had possibly contracted COVID were not admitted to hospitals and told to self-isolate in London, amplifying stress levels. At this point, any plans of going to Phuket was moot, as cases in Southeast Asia had been rising even before the situation escalated in the UK.

So, this forced me to pivot. Digital ethnography was not a plausible alternative for my study, as I did not have much ethnographic access to pursue this form of research. Many of the other methodological options: archives, visual analysis, interviews, focus groups, surveys, and econometrics either did not help me answer my research questions or I did not have the capacity to gather requisite data in time.

At this point, I turned to a critical corpus-based analysis (see Paterson and Gregory 2019). This method takes up discourse as its main object of study, combining the quantitative corpus linguistics method and qualitative critical discourse analysis (CDA). It treats text as data, while also connecting it to its entangled socio-political agents and values. I analyzed newspapers as a key locus for data and meaning-making, given its importance as part of national discourse regimes during critical junctures like the pandemic. The pivot also involves changing my topical focus. I had trouble finding a huge amount of discourse and empirical data online for Phuket itself, which is why I turned to a comparative study between Singapore and Phuket, while narrowing my focus to a key time period of the pandemic.

The Study

I explored 423 articles from *The Straits Times*, *TODAY*, *Thai Rath*, and *The Thaiger* and 8 speeches by Prime Minister Lee Hsien Loong during 14 April to 13 May 2020. My main research question took up the discursive role of “smart”, “surveillance”, and privacy in the Singapore and Phuket during the pandemic, and the analysis includes how often (frequency) and how (mode of deployment) these terms were used in newspapers. In doing so, the research presented how various configurations of the state legitimized surveillance and co-opted various civil liberties in the two cities. Instead of taking a normative stance on this, I focused on what questions should continue to be asked in light of pandemic surveillance.

Practically speaking, there were four steps in my approach: (1) acquiring documents for the corpus, (2) pre-processing the data, (3) computing frequency and collocations, and (4) analyzing these in tandem with its wider socio-political processes.

First, I gathered articles in Singapore and Phuket related to COVID-19 from mid-April to mid-May, a period that saw a rapid rise in COVID-cases for both cities. There were over 3,000 articles across the four sources, so I had to figure out how to account for duplicates and include ones that were pertinent to my study.

Following this selection, the second step was to pre-process the data. I converted these into TXT format and uploaded it onto Antconc, a processing software (see Joss et al. 2019). Third, the selected features of “smart city”, “surveillance”, and “privacy” were then computed to measure their absolute frequency and collocations. Lemma words did not appear often in the corpus, but the equivalence class of surveillance was expanded to include metonyms such as “contact tracing.”

At this point, most quantitative text analyses are followed up by scaling or validation (see Grimmer and Stewart 2013; Jurafsky and Martin 2019; Lowe and Benoit 2013). The analytical outcomes of these methods did not align my research questions, which is why I adopted CDA. CDA refers to investigating discourse through its power relations, as well as the social, historical, and material conditions of possibility for their formation and constraint (see Fairclough 2001; Hook 2001; Reisigl and Wodak 2009).

What I found interesting from the data analysis were the agents associated with pandemic smart surveillance. This helped frame two chapters in my dissertation about the construction of a smart virus for a smart solution, as well as the involvement of state spatial projects. The former explores how the pandemic was treated as a sentient, non-human entity that could “act” in a smart way. The latter investigates state spatial projects (see Brenner 2004), and the multi-scalar actors driving the smart surveillance project. In this manner, this methodology was generative in exploring key power relations, their agents, and how these impact civil liberties such as privacy.

The Afterlife

One year after this study, questions about methods and data continue to animate my thinking about research. COVID-19 has been resurging in Southeast Asia at the time of writing this in the middle of 2021. The restrictions around in-person ethnographic work on research call for continued thinking about alternative research methodologies such as digital ethnography and remote forms of research. What methodological intersections or adaptations can generate in-depth social science research, and how can one gain proficiency in these forms of research methodologies?

A key part of research is data. Reiterating Boellstorff et al.’s (2015) call that “data is too big to be left to the data analysts,” how can social scientists expand our remit to investigate emergent forms of

data and big data? Data is not a homogenous ground, and its multiplicity and ethnographic openness charts future trajectories for researching smartness, urbanity, surveillance, and social science broadly speaking (Douglas-Jones et al. 2021).

Thinking about data also necessarily implicates its object and model. As Kockelman (2020: 332) explains, an object causes data, data informs a model, and model represents the object; at the same time, the object constrains the model, the model conditions the data, and data indexes the object. Here, the text-as-data in my research was mediated by the model of analysis (Antconc) and its object (smart pandemic surveillance), and each of these entities can be expanded in many ways. For instance, increasing the parameters of the data and object by including social media sites, additional news sources over a longer period of time, and a range of other ASCN cities who have been heavily implementing pandemic surveillance techniques would have increased external validity.

Though simply increasing the amount of data through the corpus also has its drawbacks. The initial plan for having the two cities in comparison was due to their investment in CCTV cameras. My examiner's point that the comparative endeavor could have been re-evaluated is a point well taken. Drawing out different kinds of relations between these two cities or diving deeper into each of their contexts could have produced far more insight.

This model of data analysis and its afterlife entail further modeling. Having spent time working with corpus analysis, I am now thinking of the ways different software like NVivo or alternative modes of analysis can enable different interpretive grounds. Moreover, models of COVID cases and their presuppositions continue to play vital roles in the way governments and civil society respondents address the ongoing pandemic, calling for sustained scrutiny to extant and widening inequities. Thus, exploring data, models, and objects in their radical multiplicities/particularities open not just more research about smartness per se, but also its performance and limits.

References

Boellstorff, Tom, Bill Maurer, Genevieve Bell, Melissa Gregg, and Nick Seaver. 2015. *Data, Now Bigger and Better!* Prickly Paradigm Press.

Brenner, Neil. 2004. *New State Spaces: Urban Governance and the Rescaling of Statehood*. Oxford, UK: Oxford University Press.

Douglas-Jones, Rachel, Antonia Walford, and Nick Seaver. 2021. "Introduction: Towards An anthropology of Data." *Journal Of The Royal Anthropological Institute* 27 (S1): 9–25.

Fairclough, Norman. 2001. "Critical Discourse Analysis as a Method in Social Scientific Research." In *Methods of Critical Discourse Analysis*, edited by Ruth Wodak and Michael Meyer, 121–38. London: SAGE Publications.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97.

Hook, Derek. 2001. "Discourse, Knowledge, Materiality, History: Foucault and Discourse Analysis." *Theory & Psychology* 11 (4): 521–47. <https://doi.org/10.1177/0959354301114006>.

Joss, Simon, Frans Sengers, Daan Schraven, Federico Caprotti, and Youri Dayot. 2019. "The Smart City as Global Discourse: Storylines and Critical Junctures across 27 Cities." *Journal of Urban Technology* 26 (1): 3–34.

Jurafsky, Daniel, and James H. Martin. 2019. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." 3rd Edition Draft. <https://web.stanford.edu/~jurafsky/slp3/>.

Kockelman, Paul. 2020. "The Epistemic and Performative Dynamics of Machine Learning Praxis." *Signs and Society* 8 (2): 319–55.

Kong, Lily, and Orlando Woods. 2018. "The Ideological Alignment of Smart Urbanism in Singapore: Critical Reflections on a Political Paradox." *Urban Studies* 0 (0): 1–23.

———. 2019. "Scaling Smartness, (De)Provincializing the City, The ASEAN Smart Cities Network and the Predictable Politics of Technocratic Regionalism." LSE Southeast Asia Forum, London School of Economics and Political Science.

Lowe, Will, and Kenneth Benoit. 2013. "Validating Estimates of Latent Traits from Textual Data Using Human Judgment as a Benchmark." *Political Analysis* 21 (3): 298–313. <https://doi.org/10.1093/pan/mpt002>.

Paterson, Laura L., and Ian N. Gregory. 2019. "Corpus Linguistics, Critical Discourse Analysis, and Poverty." In *Representations of Poverty and Place: Using Geographical Text Analysis to Understand Discourse*, edited by Laura L Paterson and Ian N Gregory, 19–39. Cham: Springer International Publishing.

Reisigl, Martin, and Ruth Wodak. 2009. "The Discourse-Historical Approach (DHA)." In *Methods for Critical Discourse Analysis*, edited by Ruth Wodak and Michael Meyer, 2nd ed., 87–121. London: SAGE Publications.

About the research

This blog post discusses the methodological pivoting from ethnography to critical corpus-based analysis for a master's dissertation on smart urban surveillance and reflects on continuing methodological questions relating data, objects, and models.

For citation: Lim, A. (2021). Methodological Pivoting in COVID-19: experimenting with critical corpus-based analysis. Field Research Methods Lab at LSE (14 June) Blog entry. URL: <https://blogs.lse.ac.uk/fieldresearch/2021/06/14/methodological-pivoting-in-covid-19-experimenting-with-critical-corpus-based-analysis>

*The cover image is by Nick Morrison on Unsplash

About the author



Al Lim

Al Lim is a PhD student in the joint Anthropology and Environmental Studies program at Yale University. He is a member of the LSE-Southeast Asia Early Career Researcher Network. He completed his MSc in Urbanisation and Development at LSE and a B.A. (Hons) in Urban Studies at Yale-NUS.

Posted In: Digital research | Remote fieldwork | Research design



© LSE 2024