Text analysis for when you can't be there

Rich Nielsen, MIT

Prepared for APSA "MENA Virtual Research Methods: TEXT ANALYSIS", 5/18/2021

Many of us want to be elsewhere right now. Personally, I was supposed to get a long-awaited research leave and travel to Morocco with my family to work on a book about Islamic authority in the digital age. There was even supposed to be some surfing involved. Instead, I've spent my research leave working from my home office that is also my bedroom, in my small apartment in Boston where I help keep my kids on task with their online schoolwork at the same time. I would love to be elsewhere right now, and I suspect each of you can resonate with my disappointment and add to it 1000 more disappointing stories as well.

When the pandemic hit, I started thinking a lot about how early-career researchers should respond. I'm lucky -- my book is in complete shambles and may be delayed for years because I couldn't do the fieldwork I planned, but my job is secure. My PhD students are not so lucky...losing a year of research can prematurely end a promising academic career. I came across a paper by an Economist, Ryan Hill, who studies the economics of academic careers; particularly astronomers whose careers are literally dependent on whether there are clear skies on the nights they are scheduled to use scarce observatory time with the world's most highly powered telescopes. His conclusion is that researchers can survive and even thrive after bad luck at the observatory if they diversify their research portfolio and shift effort between research streams to circumvent their challenges.

This resonated with me because it echoes how I got started with text analysis. I entered graduate school in 2007 with the plan to do field research in the Middle East, but then my wife and I had a child in 2008 and we did not have the financial ability to spend a year abroad. We were broke, and research fellowships don't pay for families. I actually never got to do the year of fieldwork I'd hoped for, though I was able to take a few shorter trips to Cairo.

At the same time, I was getting increasingly curious about online fatwas I was seeing as I tried to keep my Arabic fresh. And I got lucky. My grad school mentors and peers were just getting interested in text analysis, and I found a match between the tools we were developing and the texts I was finding on the internet. Instead of writing a dissertation where I felt apologetic about my curtailed fieldwork, I did something that was genuinely new: the first Arabic-language automated text analysis in the political science discipline. That in turn saved my academic career.

The rest of my talk is about how to pivot your own work if you face similar challenges.

Why might text analysis -- and image and audio analysis -- be a pivot you might want to explore?

1) It engages deeply with the sources that make many of us curious about politics.
2) It relies on many of the same skills as fieldwork: language, attention to nuance, and ethnographic sensibility.
3) Yet its accessibility is not hampered by a virus, in fact more text, audio, and images have moved online. And it allows research in settings where fieldwork is traditionally hard: violent organizations, autocratic governments.
4) It opens new questions and new answers. People worry about text analysis being method-driven, but the best way to do *problem*-driven research is to have many tools.
5) Is compatible with both positivist and interpretivist approaches. [My thoughts on combining positivism and interpretivism]
6) Results in research that is satisfyingly rich but also satisfyingly comprehensive. Close reading and far reading combined.

What does a pivot look like? With permission, I'll share how one of my PhD students, Aidan Milliff, has used text and image analysis to pivot when his fieldwork got interrupted by COVID. He wrote a reflection in PS.

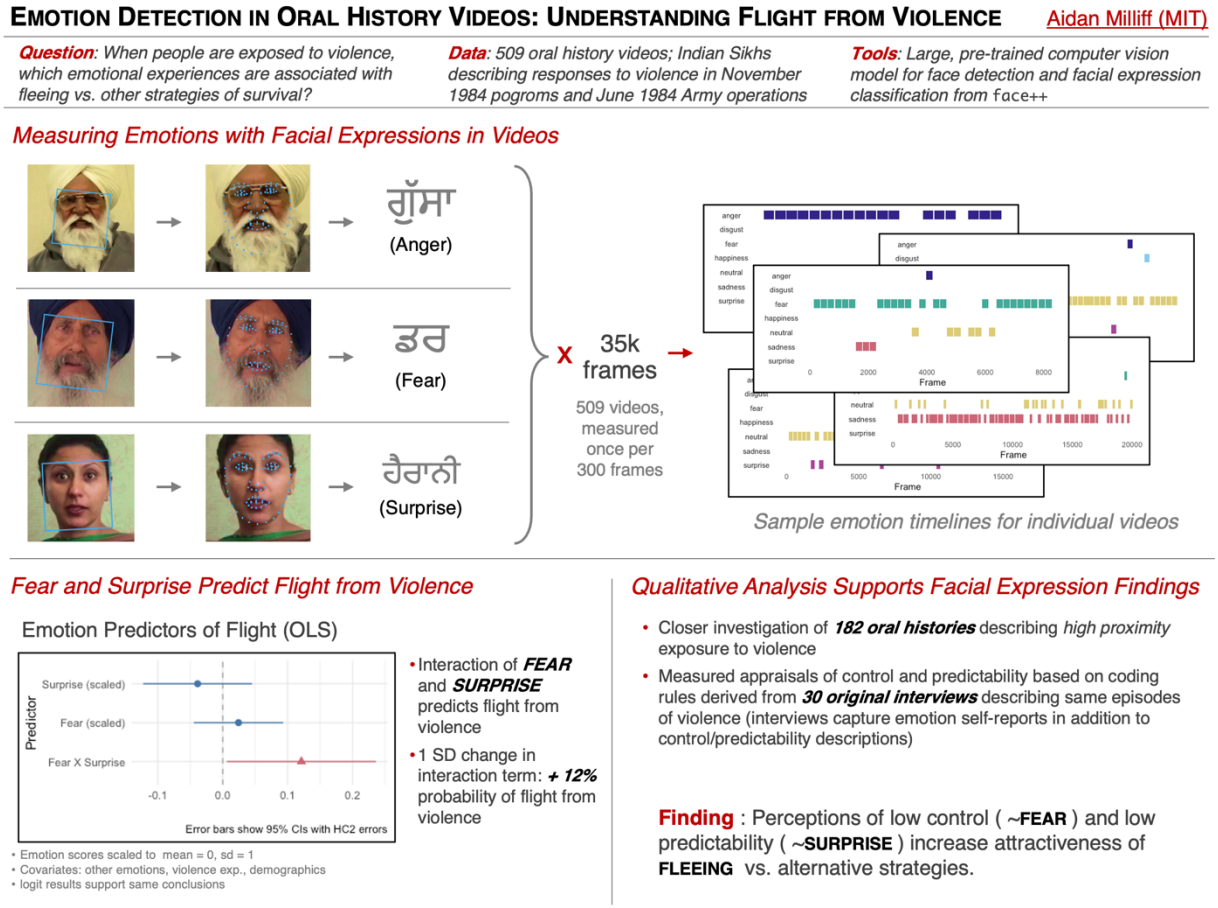

Figure 1: Aidan Milliff's dissertation in one slide.

Two routes into a strong text analysis project.

1) Look at examples of other text analysis projects and have your imagination expanded. Then figure out how you could adapt that approach to your own texts.
    a) Justin Grimmer's Dissertation/book -> My dissertation/book -> Michael Freedman's JCR.
2) Identify a set of texts that you find puzzling.
    a) Develop a clear question (biggest pitfall!)
    b) How you would answer that question from the texts if you had infinite time to read?
    c) Find an automated method that best approximates (b). This will almost always be some version of supervised or unsupervised learning.

Routes into a weak text analysis project

- Not having a clear question to which your text analysis is the answer.
- Not reading or understanding your texts.
- High-tech models are NOT essential. I've published word searches in the best journals.

The best way to get started is to get started! I taught myself and you can too. I've made my workflow open source, and it is linked here along with other resources: http://www.mit.edu/~rnielsen/helpful.htm