REGULATING PLATFORM MODERATION OF HATE SPEECH

THE SHAPING OF LINGUISTIC NORMS AND SOCIO-POLITICAL VALUES

MANDY LAU, PHD IN LINGUISTICS AND APPLIED LINGUISTICS

Mandy Lau (she/her) is a PhD candidate in Linguistics and Applied Linguistics at York University. She is interested in language policy and language ideology, particularly in the contexts of digital culture, minoritized language communities, and public education. Her current research explores content moderation policies with regard to hate speech on social media as well as the relationship between voice technologies and language ideologies. She is a Graduate Associate at the Robarts Centre for Canadian Studies and holds a SSHRC Canada Graduate - Doctoral Program award.

How did you come to choose this research topic?

As a former high school teacher, curriculum consultant, and vice principal, I thought my doctoral work would be related to language learning in the K-12 public education system. During this time however, I became increasingly intrigued and alarmed by how AI systems shape how we perceive language and what counts as knowledge. While I wasn't sure how I could approach this from the field of applied linguistics, I was lucky that my supervisor, Dr. Eve Haque, encouraged me to be open to my curiosities and push disciplinary boundaries.

It took me a while and many false starts to determine how to approach research that is necessarily inter-disciplinary and outside of my immediate expertise. Fortunately, I was able to take several courses outside my program at York. Also, as many learning opportunities shifted online in response to the pandemic, I was able to participate in working groups, workshops, and conferences held around the world. Many more than I would've been able to do had this shift not occurred; the online shift made exploratory learning more feasible both

financially and practically as I juggle my studies with caring for young children. As things are now moving towards hybrid models, I can continue exploring new and wider topics online while being more intentional with what I'd like to attend in person.

Tell us about your dissertation work and how it relates to the study of Canada.

Generally, I'm interested in anything to do with language policy and language technologies. This includes language models used for generative artificial intelligence (AI) or content moderation systems and their implications for language practices.

My dissertation focuses on Canada's emerging online safety legislation as it wrestles with regulating content moderation obligations for platform companies. Specifically, I'm interested in how hate speech will be framed and how this relates to the shaping of Canadian nationhood. Coming from a language policy perspective, my analysis of the materials produced during this legislative process will examine how online linguistic norms serve to manage wider dominant social, economic, and political interests.

Tell us about the challenges you as a researcher are experiencing or had to overcome to do this work.

The biggest challenge in studying big tech policy is not being able to access meaningful policy documents. I've had to rely heavily on journalism to figure out how to access leaked documents, databases from other researchers doing adjacent work, and Access to Information and Privacy (ATIP) requests. I'm hopeful that new legislative obligations for platform transparency will be helpful for future research.

Another challenge is just having to sift through hate speech, either when reading user-generated content on platforms or internal training documents. Eventually, it can get to you. I try to set some personal boundaries and limit my "intake" and focus on the more hopeful aspects of this work.

What are some key points from your research you want others to come away with?

Whose language? When it comes to moderating platform speech, a couple of questions are always at the forefront for me: (1) whose language practices are we regulating and impacting? And (2) which languages receive investments in terms of moderation labour, training data, or the development of algorithmic tools? The first question tends to get more attention. For example, there is work on how moderation disproportionately impacts the language practices of 2SLGBTQI+, Black, and Muslim communities, or sex educators. The second question is much harder to figure out. When discussing content moderation policies, there is not a lot of discussion on which languages or how multilingualism will be

treated. The assumption is that "all languages are equal" or that multilingual practices will be dealt with in the same way when that is not the reality. For example, consider that American English is the most resourced and remains the most dominant in language technologies, often acting as the bridging language for machine translation.

Language as data: In sociolinguistics, language is understood as a social practice; it is intimately tied to social identities and power. Meanings are always in flux and negotiated. However, in order for human languages to be legible for computers, the meanings of language are processed as static and stable, devoid of politics and historicity. Natural Language Processing methods, for example, uses statistics to generate very plausible-looking strings of text. This tension between how humans employ language and how machines process and generate language will be important to follow as human-machine interaction increases. There will be more questions around how machines will shape human language practices and how humans shape languages for the machines, and what gets lost when language is valued most as data.

Classifying hate speech: For my project, I will focus on how hate speech is negotiated as a regulatory object. Legal conceptions of hate can be necessarily narrow and do not encompass the full spectrum of violent speech. So, classifying hate speech can shed some light on what speech counts as hate and the other kinds or degrees of violence deemed acceptable or legitimate. From this, we can draw out Canada's national ideals, values, and priorities.

The public and the private: The relationship between big tech and regulators is complex and iintertwined. Tech companies have a lot of sway

¹ Regulatory Capture Lab. (2023). The revolving door. https://regulatorycapturelab.ca/The-Revolving-Door.

with public regulators: for instance, their platform moderation approaches and framings often become the only possible approaches and framings, with legal and regulatory policies shaped around them. People with expertise also shuffle their labour between the regulator and the regulated, what the Regulatory Capture Lab calls a "revolving door."[1] Finally, when it comes to tracing the money, we can see that tech companies increase their lobbying dollars during times of policymaking and that there are different kinds of public funding sources available to tech companies for their research aims or their labour needs. Tracing this relationship can be informative in understanding Canadian political and economic priorities.

Tell us about what you enjoy the most about the work that you do.

I love learning within a community. While a PhD can be very isolating, I appreciate the opportunities to engage in learning and thinking, either formally in research teams or working groups or informally with a few good friends pursuing their PhDs.

What advice, lesson, or tips do you have for those starting their academic journey?

My brother and I are the first generation in our family to go to school in Canada, attend university, and now I am the first to attend graduate school. This means a lot to my family, particularly to my parents, who provide a lot of logistical support for me to participate in academic activities. The challenge is that it can sometimes be difficult figuring out how grad school works (or even what I'm supposed to be doing!). Here are some suggestions that have been helpful for me:

Have your department administrator on speed dial: literally. Our graduate program's administrator (Hi Rose!) is so knowledgeable and has helped me navigate problems of which I was not even aware.

Attend workshops organized by your program and research centres such as Robarts: I found that this was a great way to learn specific academic skills and connect with others who are figuring this out too.

Take a class outside your program: It will be more challenging to learn new disciplinary norms but it is an excellent opportunity to meet new friends with diverse disciplinary perspectives.

Stay connected to your supervisor's former doctoral students: Eve's first advice for me was to make friends and to start with Bill, her former student. Bill has become my "ghost supervisor," providing mentorship and feedback for much of my thinking

Find your people: University can be really daunting for those of us new to the scene. What has sustained me the most psychologically and emotionally is to draw from the strength of my community, friends, and family, which has helped build my confidence and sense of belonging.

What are the next steps in your research?

I can't wait for a draft of the new online safety bill to come out this year! Stay tuned!