Research Update: The linguistic production of learning opportunities

Chris Proctor
Amir Lopatin Fellowship
April 1, 2018

Background

The Amir Lopatin Fellowship funded my research project on how identity and ideologies shape the way adolescents learn computer science. We know there exists a substantial participation gap in computer science participation across categories of race and gender. Margolis & Fisher (2003) document factors such as differing motivations and interests, opportunities to gain experience and confidence, and the local culture of computing, which can either support women in developing as computer scientists or serve as barriers. Margolis et al (2010) document comparable factors promoting racial marginalization or inclusion in computing. Barron (2006) analyzes how students' interest in computation can be sustained across multiple sites in a learning ecology, and the roles parents can play in supporting this learning (Barron, 2009). In her (2016) doctoral dissertation, Michelle Friend (who originally recruited me to teach at the Girls' Middle School, and then encouraged me to apply to graduate school) conducted longitudinal research on graduates of the Girls' Middle School and interviewed women who ended up working in computational fields.

What we don't yet understand is the fine-grained processes by which social roles and identities, which afford students learning opportunities (Brown, Collins, & Duguid, 1989; Hull & Greeno, 2006), are shaped by categories such as race, gender, and ethnicity, which in turn are produced through discourse. A sociolinguistic approach, studying how people present themselves and interact through talk, is particularly promising because it treats students as potentially actively engaged in reshaping the meanings of the categories that mediate their ability to access learning opportunities. Therefore, my research project funded by the Amir Lopatin Fellowship is focused on answering two questions:

1. What language ideologies dominate the discourse of Computer Science, and how do they shape the learning identities available to Computer Science learners?

2. What strategies are employed by women, people of color, and others who are marginalized by the dominant computing culture to challenge, subvert, or redefine the categories on which they are marginalized?

Methods

I proposed to answer these questions by learning about the experiences of graduates from GMS as they went through high school. What happens when early adolescents go from a supportive culture in which they learn how to be computer scientists, into high schools which might be much less intentional in
resisting racialized and gendered ideas about who can be a legitimate computer scientist? I started with individual interviews of GMS alumnae (currently in 10th grade).

with plans to then convene small group discussions synthesizing and sharing experiences. I plan to synthesize students' experiences through a process of open coding within grounded theory affinity analysis (Bernard & Ryan, 1998), with a research team composed of alumnae. Finally, I will organize a "critical computation summer camp" focused on helping participants identify and respond to the challenges previously identified.

One aspect of the research project which has emerged since the grant proposal is the use of natural language processing techniques to compare participant interviews with discourse characteristic of a dominant-culture computer science community. Using a decade's worth of posts from Hacker News, an online community dedicated to CS, startups, and the tech industry, I created a model of how linguistic meaning within Hacker News is different from everyday word meanings. This model captures jargon and an increased interest in technical topics, but it also subtly encodes attitudes about the kinds of identities that belong and do not belong within the space. I plan to contrast language use within interviews with language use within Hacker News.

Progress

Interviews

I met with several stakeholders at the Girls' Middle School, including the director of alumni relations, the academic dean, and both computer science teachers, to discuss the project and participant recruitment. I launched my recruitment effort by putting up a website and sharing it at GMS Alumni Day in June 2017, and the school followed up with an email to all the alumni with whom they are in contact. I recruited a high school senior to help lead the student research team and to conduct some of the interviews. Aware that my position as their former CS teacher (and possibly, my familiarity with their backgrounds and families) might affect the interview, I asked participants whether they would prefer to be interviewed by me or by another member of the research team. Participants unanimously said they wanted to have the interview with me.

Of the 67 students in the cohort, 8 replied, and so far I have been able to conduct and transcribe 7 interviews, ranging from 40 minutes to over an hour. The content of the interviews (discussed in the next paragraph) has been fascinating, but I wish I had been able to recruit more participants. In addition to direct outreach, I relied on "snowball recruitment," in which participants would encourage their friends to sign up. This worked to some extent; most of the members of a social group that stays in close contact have signed up. However, they do not maintain social relations with many of their alumnae peers. In my interviews, some students told me they hesitated to sign up because they are not currently taking Computer Science, so they thought their experience would not be relevant to my research. This is not the case, and in a follow-up email to the cohort, I corrected this impression. Still, I was their computer science teacher, so they may be reluctant to participate because they think they will personally disappoint me if they tell me they are not continuing to study Computer Science.
As I am still in the process of coding the interviews, I can only offer general impressions at this point. I have noticed a stark contrast between the participants who participate in some kind of computer science and those who do not. Those who do not participate in computer science appear to see themselves as mainstream, normal girls. Those who do participate in some form of computing self-identify as outsiders, or being part of a niche social group. These individuals are much more aware of the way cultural pressure shapes expectations on their identities, and in several cases they reject gender-based expectations by wearing their hair or clothing in particular ways, or by identifying as gender non-binary or transsexual. These interviews have generally included quite a lot of detail about the spaces in their schools where computer science, robotics, engineering, and related activities are practiced. Generally they are male-dominated and range from unwelcoming to hostile. As is consistent with prior research, there has often been a mentor who has played a pivotal role in making the space available (Barron et al, 2009). It has been particularly interesting to hear the terminology participants use to discuss these challenges: 'cis males,' 'toxic masculinity.'

**Analyzing Discourse**

At the same time, I have been developing methodology for using natural language processing to interpret the interviews within the larger culture of computing. I want to understand how insiders understand themselves to be doing computer science, so that I can ground the ideas and identities I hear from interviewees. To this end, I have begun working with Hacker News, a discussion forum affiliated with the Bay Area startup incubator Y Combinator. Many of the participants on Hacker News have high social capital within the broader CS community, and participants talk freely to a perceived internal audience, so this is an excellent site for studying CS discourse practices. I downloaded the ~12 million comments which have been posted to Hacker News over the last decade by ~300,000 users (~10% of whom have over 50 posts).

I used an algorithm called Word2Vec (Mihkolov et al, 2013) to create a model of linguistic meaning based on the distribution of words in the comments, and took a snapshot of how word meanings changed every month over the course of a decade. (This was a novel application of Word2Vec; a paper for a technical audience is in preparation.) By comparing the positions of words in each monthly language model, we can observe semantic change over time. The charts below each illustrate how word meanings shift over time with respect to a particular relational axis. (The model of everyday language, from which these models depart, would be just to the left of each graph.) The words selected were all among the twenty words which moved the most over time. In the first chart below, some of the often-discussed stereotypes about women in technology are clearly visible. The second chart uses [black, dark] and [white, pale] as anchors. Even though these have non-racialized meanings, the words whose meanings shift the most over time are heavily laden with racialized innuendo.

I can also use each model to estimate the likelihood of observing particular sequences of words, which allows me to observe how individual users' alignment with the discourse community changes over time. This strategy is an extension of one used by Danescu-Niculescu-Mizil et al (2013), who explore the relationship between change in online linguistic communities and relative change in
individual users’ language with respect to community norms. I am currently analyzing the extent to which users’ alignment with community linguistic norms predicts their future participation. From there, I plan to look for alignment between evidence from the interviews and this model.
Next steps

I am currently actively engaged in recruiting more participants for one-on-one interviews through one additional round of emails, and an attempt to recruit through prior participants. In particular, I have recruited very little socioeconomic and racial diversity. I have reached out to the teacher who coordinates support for students on scholarship at GMS for help connecting with students who were on scholarship at GMS. Though the timeline has been delayed by my difficulty in recruiting participants, I still intend to pursue a participatory process for interpreting the interviews, and to organize a summer camp which combines computing with reading and discussing critical computational literacy.

Conclusions

Through this research, my goal is to recast the conversation about equitable access to computer science so that the potential agency of youth is included. Without discounting the seriousness of social and cultural forces, a focus on how adolescents make meaning within their local contexts can help us better understand the mechanisms behind inequitable participation as well as the ways in which adolescents boldly and creatively push back against and subvert beliefs and practices which exclude them. This research has turned out to be more central to my PhD work than I originally anticipated. Here are some of the ways it ties in with other ongoing research projects:

- I have been studying this cohort's CS learning during middle school. Using all of their CS work and reflective writing through middle school, and a summative task designed to measure their
computational thinking skills, I was able to conduct the most substantial longitudinal study to date on early computer science learning (Proctor & Blikstein, 2018). The present work further extends my study of this cohort into high school.

- I have been exploring interactive storytelling as a medium which combines writing and programming. Through a series of design-based research studies, I found that it is a particularly effective medium for using computational tools to investigate social issues. In particular, some workshops have helped participants to develop critical computational literacy (Proctor, 2018). This is exactly the kind of development I hope to support through the planned critical computation summer camp. I plan to use what I learned, as well as Unfold Studio, the web application I developed, in the camp.

I am deeply grateful for the support of the Amir Lopatin Fellowship, which has made it possible for me to carry on this research. I look forward to the completion and publication of this research.
References


