[redacted]

Critique: “A computational approach to politeness with application to social factors”
By Danescu-Niculescu-Mizil, Sudhof, Jurafsky, Leskovec, and Potts

“A computational approach to politeness with application to social factors” by Danescu-Niculescu-Mizil et al. explores the relationship between power and politeness, investigating the social dynamics of two online communities as it develops a language classifier for politeness. The linguistic features utilized by this classifier resonated particularly with my own experiences expressing politeness on Piazza. As an instructor, I will favor less polite approaches, such as “Please do not post your code on Piazza,” whereas as a student, I will litter my posts with hedges, indirect language, and gratitude. As a student, I am constantly reframing my posts so that others will find them non-threatening, kind, and be encouraged to answer, while I spare no such thought in crafting messages as an instructor. Rather, the latter kinds of messages are more utilitarian: I can focus on content, not social cues. It was thought-provoking to see my own behaviors articulated and then identified by computational means. I think it is also interesting that I manage to operate in both modes at once, as a student and instructor in separate courses; if data allows, it would be fascinating to see if admins carry their decreased politeness over to other sites in which they are members, or if their inflated sense of importance applies only to the site in which they directly hold power.

In part due to its inclusion of linguistic features like hedges and indirection, I think the model developed in the paper also has great potential application in the realm of gendered communication. As Robin Lakoff argued in *Language and Woman’s Place,* women’s speech often contains features like “tag questions” and hedges that are used to soften the impact of statements (Lakoff, 1973). This is part of a larger phenomenon by which women are socialized to be unobtrusive and unaggressive; having too forceful an opinion can be met with social censure for being “unladylike.” Accordingly, women may overly police their language in order to be perceived as friendly. This can negatively impact their ability to be taken seriously academically and professionally, arenas where directness and assertiveness may be valued (perhaps even rewarded with promotions). Since the classifier described in this article was adept at finding the linguistic hallmarks of overly polite/dependent speech, a natural possible application of the technology could be a tool that helps women check over their speech (say, emails) and remove features like hedges. This would help them craft more assertive messages and gain better traction in the workplace.

Thus, my anecdotal experience and understanding of gendered communication convinced me of the validity of the linguistic features included in this article’s training model. However, I have to admit I remain less certain of the rationale behind (and efficacy of) using a bag of words model to classify text computationally. I wonder if a bigram model was tested and discounted, or simply never attempted—this is not discussed in the piece. It seems counter-intuitive that the frequency of unigrams would be the best possible tool when many expressions of politeness, such as “<s> Please”, “thank you”, “could you”, etc. might occur in length-2 pairs. I only wish that whatever decision heuristic the researchers employed in choosing the unigram model could have been better clarified.

Works Cited

Lakoff, R. (1973). Language and woman's place. *Language in society*, 2(01), 45-79.