



A Computational Approach to Politeness

Azure Zhou, Muhammad Khattak, Clare Chua



Overview

- How do we operationalize domain-indep. concepts of politeness?
- What is the relationship between politeness and social power dynamics?



An exercise...



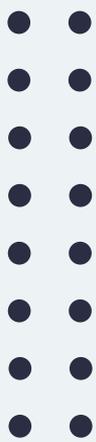
<< Left (most impolite) ...

... Right (most polite)>>

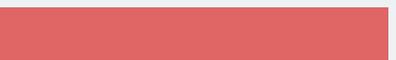
Answer from Stack Overflow: “You can't parse [X]HTML with regex. Because HTML can't be parsed by regex. Regex is not a tool that can be used to correctly parse HTML. As I have answered in HTML-and-regex questions here so many times before, the use of regex will not allow you to consume HTML.”

Edit Request from Wikipedia: “Most of the information in here seems fine, except for 1 thing: the June record low of -11F at Anaktuvuk Pass in Alaska is dubious because it seems like that year that the Anaktuvuk Pass weather station was having data errors. Source: <https://xmacis.rcc-acis.org/>”

Question from Stack Overflow: “How does Python's slice notation work: when I write code like `a[x:y:z]`, `a[:]`, `a[::2]` etc., how can I understand which elements end up in the slice? Please include references where appropriate.”



Data Collection



“I would consider politeness to be more of a scale rather than a binary variable.”

– Simrin Kalkat

“I think the quantiles 2 and 3 should have been merged as a neutral category.”

– Roberto Lopez

“The annotation process is... questionable.”

– Xiaoyuan Ni

“I would like to hear from other classmates about whether they have any criticisms on the division of quartiles for politeness.”

– Kaiyu Ren



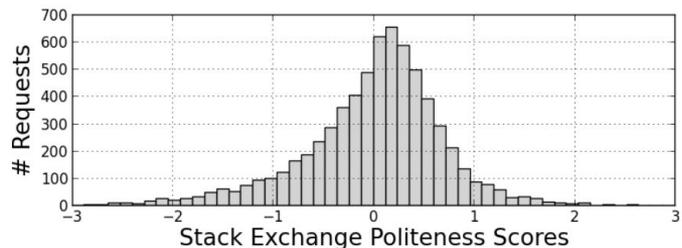
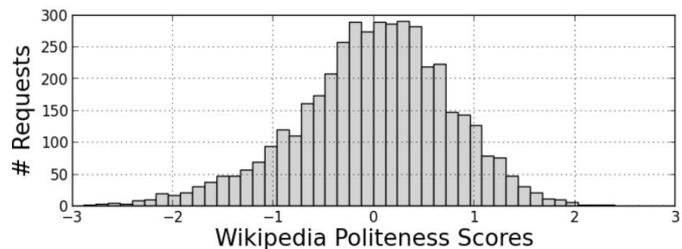


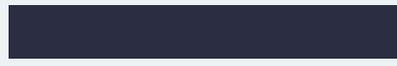
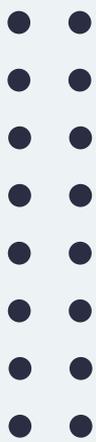
Figure 1: Distribution of politeness scores. Positive scores indicate requests perceived as polite.

How reliable is this scale?

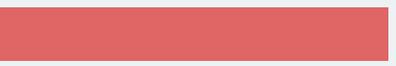
Would you have designed the data collection process in the same way?

Quartile:	1 st	2 nd	3 rd	4 th
Wiki	62%	8%	3%	51%
SE	37%	4%	6%	46%

Table 2: The percentage of requests for which all five annotators agree on binary politeness. The 4th quartile contains the requests with the top 25% politeness scores in the data. (For reference, randomized scoring yields agreement percentages of <20% for all quartiles.)



Classification Model



Strategy	Politeness	In top quartile	Example
1. Gratitude	0.87***	78%***	I really appreciate that you've done them.
2. Deference	0.78***	70%***	Nice work so far on your rewrite.
3. Greeting	0.43***	45%***	Hey , I just tried to ...
4. Positive lexicon	0.12***	32%***	Wow! / This is a great way to deal ...
5. Negative lexicon	-0.13***	22%**	If you're going to accuse me ...
6. Apologizing	0.36***	53%***	Sorry to bother you ...
7. Please	0.49***	57%***	Could you please say more ...
8. Please start	-0.30*	22%	Please do not remove warnings ...
9. Indirect (btw)	0.63***	58%**	By the way , where did you find ...
10. Direct question	-0.27***	15%***	What is your native language?
11. Direct start	-0.43***	9%***	So can you retrieve it or not?
12. Counterfactual modal	0.47***	52%***	Could/Would you ...
13. Indicative modal	0.09	27%	Can/Will you ...
14. 1st person start	0.12***	29%**	I have just put the article ...
15. 1st person pl.	0.08*	27%	Could we find a less complex name ...
16. 1st person	0.08***	28%***	It is my view that ...
17. 2nd person	0.05***	30%***	But what's the good source you have in mind?
18. 2nd person start	-0.30***	17%**	You've reverted yourself ...
19. Hedges	0.14***	28%	I suggest we start with ...
20. Factuality	-0.38***	13%***	In fact you did link, ...

Negative politeness
–Brown and Levinson

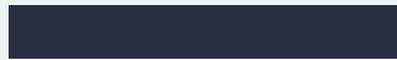
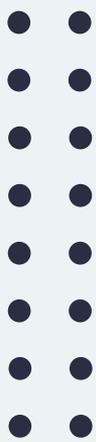


Classifier Results

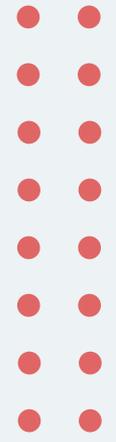
Train	In-domain		Cross-domain	
	Wiki	SE	Wiki	SE
Test	Wiki	SE	SE	Wiki
BOW	79.84%	74.47%	64.23%	72.17%
Ling.	83.79%	78.19%	67.53%	75.43%
Human	86.72%	80.89%	80.89%	86.72%

How comprehensive do you find the politeness markers? Would you have included other features in the model (e.g. honorifics)?

Does it make sense to look at politeness without regard to context?

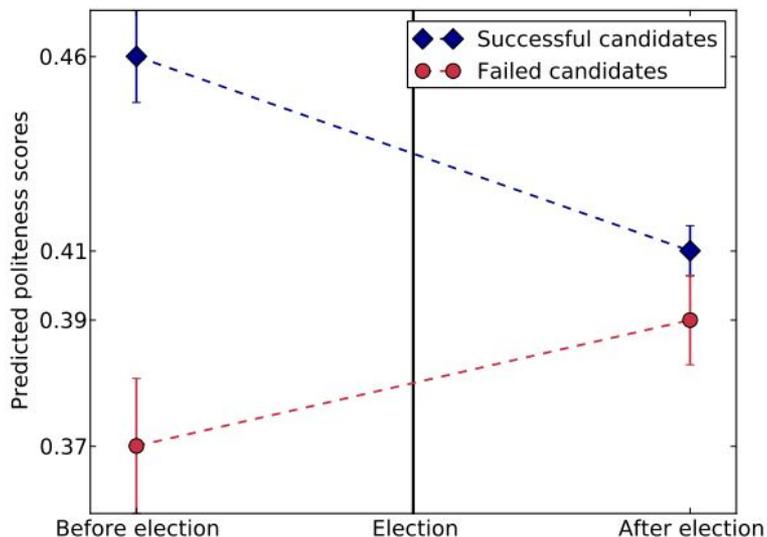


Results



Eventual status	Politeness	Top quart.
Admins	0.46**	30%***
Non-admins	0.39***	25%
Failed	0.37**	22%

Politeness and social outcome



“Are the ones who succeed in being voted admins for example have forcefully been more polite before? What’s the relation here, is it causal?”

- Axel Peytavin

“Why would people who were impolite during an election suddenly become more polite when there is nothing to win?”

- Adhitya Venkatraman

Politeness on Stack Exchange

Role	Politeness	Top quart.
Question-asker	0.65 ^{***}	32% ^{***}
Answer-givers	0.52 ^{***}	20% ^{***}

How could these results extend?

What other relationships would you expect to see between politeness and social power (on other platforms and in real life)?

Reputation level	Politeness	Top quart.
Low reputation	0.68 ^{***}	27% ^{***}
Middle reputation	0.66 ^{***}	25%
High reputation	0.64 ^{***}	23% ^{***}

Table 7: Politeness and Stack Exchange reputation (texts by question-askers only). High-reputation users are less polite. Analysis conducted on 25k requests (4.5k low, 12.5k middle, 8.4k high).

PL name	Politeness	Top quartile
Python	0.47 ^{***}	23%
Perl	0.49	24%
PHP	0.51	24%
Javascript	0.53 ^{**}	26% ^{**}
Ruby	0.59 ^{***}	28% [*]

Table 8: Politeness of requests from different language communities on Stack Exchange.

“not only do people interact differently online than offline, they also interact differently in other online communities, and in other cultures, spaces, etc.”
– Nourya Cohen

How generalizable are these results? Do you expect them to transfer to other contexts and situations?

On what platforms might we see the classifier being useful today (a decade later)? Are the results from 2013 applicable to current digital interaction platforms?

What are some potential applications of this work?





**How can we consider
“politeness” as a reflection
of English ability, education
level, socioeconomic status,
regional norms, cultural
background, in addition to
intent and power?**



What if a classifier classifies someone's speech as “impolite” but actually it could be polite (given its subjective)?

Would this type of model work when some users use dialect, or in informal settings?

“A lot of this does depend on our culturally specific consensus on a sense of politeness.”

– Joel Johnson

“Preliminary results”



**Midwestern
Politeness**

**Female
Wikipedians**

**Mixed-signal
Stack Exchange**

How do broader social patterns affect how we should interpret the results?

How can we factor compounding effects of identity?

Is there future work to investigate the causes of these underlying trends?

Thank you!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

