Saying Mean Words to Strangers: A Comparative Study of Politeness Features in Internet Arguments

Connor Milligan

May 3, 2024

Abstract

This study seeks to delve into argumentative discourse in online forums. Online argument is a fairly common occurrence in internet subculture. The main point of study is to take a closer look at cases of argument and look into how mood indication might vary across different platforms. It also seeks to answer if a more goal-oriented community exhibits more politeness in argumentative discourse than that of a more traditional forum-focused community.

This study uses Cornell's *Conversations Gone Awry* Dataset, which compiles curated data from both Wikipedia and Reddit. The dataset contains conversations which derail into personal attacks from both the ChangeMyView Subreddit, a sub-community where Reddit users present their viewpoints for other users to challenge, and the talk pages of Wikipedia articles, where contributors discuss changes and revisions to articles. This will be used in conjunction with the Politeness Strategies function based on the work of Cristian Danescu-Niculescu-Mizil (2013) to gauge politeness of utterances. Politeness is then measured against datasets of non-specific Reddit and Wikipedia talk page data to observe both the difference in politeness between argumentative and general discourse, and argumentative discourse between the two platforms.

1 Introduction

Humans have only been interacting over the internet for less than 50 years. Launching in the late 80s and becoming more widely accessible into the mid to late 90s, the World Wide Web took only a little over 20 years to evolve into the modern, reactive internet known today. Today, social media has become one of the most defining features of the internet. That is to say, interaction is one of the driving forces of internet usage. Interacting over computers actually precedes the internet itself, this being the BBS, or bulletin board system, which saw use throughout the 80s. These would be accessed through the phone line with a dial-up connection. These boards facilitated conversation through message boards, which were often divided up into topics such as hobbies and interests[1]. With the dawn of the internet, and gradual shutdown of traditional BBS servers, this discourse moved from BBS to traditional forums and social news websites, which aggregate links to other pages on the web and include a comment section for links.[2] For the purposes of this study, discourse associated primarily with said sites will be defined as Opinion-Oriented Discourse, that is, where discussion is primarily motivated by opinion.

While most internet discussion was in transition to adopting forum suites and aggregators, a new type of discourse was finding its home in various spots around the internet. The late 90s to early 2000s saw the jump in popularity of large collaborative projects. One of the largest being Wikipedia, an online collaborative encyclopedia, springing up in 2001. Things too, such as open-source software projects[3] grew very popular with the releases of popular software licenses in the early 90s, one of the most notable projects being the Linux kernel. Both of these projects saw use the use of various communications platforms to facilitate discussion about development, collaboration, and problem-solving related to their respective projects. The Linux kernel project, as with many other software

---

[1]Bulletin board systems also often had other functions, such as news, file exchange and even games. Usenet systems held similar functions and existed around the same time.

[2]See sites such as *Digg*, *Reddit*, *Hacker News*.

[3]Software projects where the source code is freely available and worked on collaboratively, often by volunteers. See projects such as *GIMP*, *VLC*, *Firefox*

projects, accommodated communications with mailing lists[4], which became more and more popular during this time period. Wikipedia used its own on-site discussion forums called Talk pages, each of which was tied to an article, where uses could discuss changes and debate over edits. Projects and sites such as these saw the emergence of a new type of discourse, which will be referred to in this study as Task-Oriented Discourse, that is, discussion focusing more on problem-solving and commonality within a single goal.

1.1 Online Argument

With human discourse comes argumentation, and the online world is particularly notorious for many a heated exchange. From the ability of a user to formulate responses in their own time, to simple anonymity, there are many reasons this could be the case. Another consideration is how heated conversation is by no means limited to casual, often Opinion-Oriented Discourse, but is found still in more professional or scientific settings. Mailing lists have been quite notorious still for scathing remarks, being surrounded by a culture of "Flame Wars"[5].

The flame war has been the subject of study for some researchers looking into online discourse. It is pretty well accepted that the proliferation of flame wars is intrinsically linked to the internet. A study on a computational chemistry mailing list concluded that the function of the list as an open platform for members of all backgrounds, experience and academic accomplishments could engage in discussions, was one of the major contributing factors of tension (Hocquet & Wieber, 2018, p. 53). To turn back to the Linux kernel mailing list, flaming by the head of the project, Linus Torvalds, has become infamous, due to harshness. "Reverting commit b22d127a39dd and just having a "ok, if we need to allocate, then drop the lock, allocate, re-get the lock, and see if we still need the new allocation" is *beautiful* code compared to the diseased abortion you just posted."[6]. This is all to say, even more professional, Task-Oriented discourse sees argumentation similar to that of Opinion-Oriented discourse.

1.2 Research Question

With it defined that both Opinion-Oriented and Task-Oriented Discourse see argument present, it leads directly into the main question of this study: Does Task-Oriented Discourse tend to see more politeness features in argument than that of Opinion-Oriented Discourse? The hypothesis here is that Opinion-Oriented Discourse is likely to see less politeness markers in argumentative discussion than that of Task-Based. This touches into a term coined by the aforementioned computational chemistry mailing list study, referred to as "fruitful trolling" (Hocquet & Wieber, 2018, p. 53). The idea is that the argument and debate will lead to productive outcome. The Linux kernel mailing list may see heated arguments, but new version releases come out all the same. The thought is that with Opinion-Oriented Discourse, the lack of a wider unifying goal leads to what is likely more personal and unproductive debate. The two chosen platforms to make this comparison are Reddit and Wikipedia Talk pages. Reddit primarily facilitating Opinion-Oriented Discussion and Wikipedia Task-Oriented. These two are chosen as "closed networks"[7] that is, each with community moderation, defined and agreed upon etiquette and extremely contextual discussion environments (Forestal, 2023). This isolates the variable to the nature of the discussion rather than the structure of the site. There exist also platforms with little or no moderation at all, though this exists primarily as the exception to the rule. It is also difficult to collect data from such sites, as the most prominent typically do not archive content[8].

2 Politeness Background

Politeness as a metric is important to look at, particularly with the growth of large online communities. In general terms, computational analysis of pragmatic features to isolate instances of things like

---

[4]Linux kernel mailing list: https://subspace.kernel.org/vger.kernel.org.html

[5]The usage of "flaming" and "flame war" appear in jargon files as far back as the 80s referencing Usenet subcultures. (http://www.catb.org/~esr/jargon/html/F/flame.html)

[6]https://lkml.iu.edu/hypermail/linux/kernel/1212.2/02623.html

[7]Forestal defines the difference between open and closed networks as follows: "In closed networks, individuals have more shared connections with one another. In open networks, by contrast, there are fewer connections between individuals." (Forestal, 2023)

[8]*4chan* and other anonymous image boards are the most popular of this category and will automatically delete inactive threads. The tendency of extreme content makes usage undesirable.

Table 1: Isolated politeness strategies paired with human annotated scores. Sourced directly from (Danescu-Niculescu-Mizil et al., 2013, p. 5)

Strategy	Politeness	Example
Gratitude	0.87	**I** really **appreciate** that you've done them.
Deference	0.78	**Nice work** so far on your rewrite.
Greeting	0.43	**Hey**, I just tried to . . .
Positive lexicon	0.12	**Wow**! / This is a **great** way to deal. . .
Negative lexicon	−0.13	If you're going to **accuse** me . . .
Apologizing	0.36	**Sorry** to bother you . . .
Please	0.49	Could you **please** say more. . .
Please start	−0.30	**Please** do not remove warnings . . .
Indirect (btw)	0.63	**By the way**, where did you find . . .
Direct question	−0.27	**What** is your native language?
Direct start	−0.43	**So** can you retrieve it or not?
Counterfactual modal	0.47	**Could**/**Would** you . . .
Indicative modal	0.09	**Can**/**Will** you . . .
1st person start	0.12	**I** have just put the article . . .
1st person pl.	0.08	Could **we** find a less complex name . . .
1st person	0.08	It is **my** view that ...
2nd person	0.05	But what's the good source **you** have in mind?
2nd person start	−0.30	**You**'ve reverted yourself . . .
Hedges	0.14	I **suggest** we start with . . .
Factuality	−0.38	**In fact** you did link, . . .

toxicity in messages are important tools for online moderation. Large companies look to automatic moderation tools to help identify problematic and terms of service breaking behavior. This extends as well to pre-review of things such as job applications or peer-reviewed studies (Nobarany & Booth, 2015). Questioning the general attitudes of things such as online platforms and their discourse could prove useful to advertisers who do not wish for their products to appear next to potentially hateful content. Another case is of an online community startup, seeking to identify moderation features and community structures that minimize hostile interaction between users.

2.1 Defining Politeness

Politeness as a concept is primarily pragmatic, and can be difficult to quantify or even identify at times. As much is seen in a preliminary analysis done in a study by Cristian Danescu-Niculescu-Mizil et al. (2013). The findings of this study serve as the majority foundation for research of this paper. Their study describes a case in which various utterances had to be ranked using a slider from "very polite" to "very impolite". While the results showed standard deviations, the reviewers noted the difficulty in defining what the boundary between polite and impolite speech was (Danescu-Niculescu-Mizil, Sudhof, Jurafsky, Leskovec, & Potts, 2013). For a study seeking to define an algorithmic approach, the subjectivity by human annotators is not sufficient to base an algorithm upon. That is to say, it is not possible to isolate the subjectivity which is taken into account when a human annotator must give a binary categorization to an utterance with ambiguous politeness. The next step to seek to make such data more quantifiable in their research, and by extension this study, was to employ Politeness theory.

Defined in their book, *Politeness: Some universals in language usage*[9], Penelope Brown and Stephen Levinson propose the idea of Politeness theory, which surrounds nearly all sociolinguistic

---

[9]Brown, Penelope and Stephen C. Levinson. 1987.

Table 2: Utterance counts per corpora, General Reddit and Wikipedia Talk corpora and respective Conversations Gone Awry (CGA) variants as of March 2024.

Corpus	General	CGA
Reddit	297,132	42,964
Wikipedia	391,294	30,021

analysis of politeness today. The principal idea of politeness theory concerns theory of face, defined as the "public self that individuals enact in social interaction" (Henningsen, 2017). Across an interaction, face is negotiated, with actions promoting positive, or negative face. It is major features of such politeness strategies that were isolated for use in categorizing politeness computationally (Table 1). These strategies are divided up into a few categories, though the main division comes down to both general politeness markers (i.e. apologies, gratitude, positive lexicon), and indirect speech (i.e. modals, and 1 person pronouns). While these features can be isolated, it's important to put it in the context of an algorithm.

2.2 Data and Algorithms

The primary tool of the study is Stanford University's ConvoKit (Chang et al., 2020), a suite of tools in the form of a Python 3 library to "extract conversational features and analyze social phenomena in conversations"[10]. The primary feature of analysis to be used is the Politeness Strategies function, the politeness analysis algorithm from the results of the aforementioned study on computing politeness (Danescu-Niculescu-Mizil et al., 2013). When run on an utterance, the function will output the counts of various politeness strategies as defined in the study (Table 1). These values are what will be weighted to determine general politeness of an utterance by features.

This study will make use of four datasets across two sources. All corpora are sourced from the same ConvoKit tool resources[11]. These datasets include metadata designed to be parsable by ConvoKit functions. The first two datasets used are general corpora of Wikipedia Talk and Reddit. Wikipedia Talk Pages Corpus[12] will be used to source the Wikipedia-side of the data to use in this study. This data was made available after being collected and used for a study on social media interaction (Danescu-Niculescu-Mizil, Lee, Pang, & Kleinberg, 2012). The Reddit Corpus[13] used will be its small variant to reduce computational intensity and better match the size of the corresponding Wikipedia corpus. It is composed of data from the top 100 most active Subreddits and should provide a representative sample of most users of the site.

The next two corpora are the Reddit and Wikipedia versions of **Conversations Gone Awry** (CGA), used by a 2018 study on predicting indicators of conversations which fall into personal attacks (Zhang et al., 2018). These datasets derive from the same datasets mentioned, but with the argumentative instances isolated with Perspective API, a machine learning API used to identify toxic utterances for the purposes of online moderation[14]. The Reddit variant of the CGA dataset is focused specifically on the **ChangeMyView** (CMV) Subreddit, a community that invites other users to challenge the stated opinion of the poster. This is chosen for the high frequency of conversation derailment.

These datasets provide a large sample (Table 2) that allows for argumentative discourse to be compared against a control of politeness indicators in each community as a whole.

2.3 Controversies

The outline of this study is not without some potential issues that may impact the results. One of the first potential issues is connected into how reliable indirect speech acts are in determining politeness. This is shown to not always be the case, including in a cross-cultural context (Blum-Kulka, 1987). This is an issue directly addressed in the research surrounding the used article, which concludes that, despite

---

[10]https://convokit.cornell.edu/documentation/index.html

[11]https://convokit.cornell.edu/documentation/datasets.html

[12]Wikipedia Talk Pages Corpus collected 2012 (https://convokit.cornell.edu/documentation/wiki.html)

[13]Reddit Corpus collected October 2018 (https://convokit.cornell.edu/documentation/subreddit.html)

[14]https://www.perspectiveapi.com/

this, indirect speech is still sufficiently reliable as a means to determine politeness (Danescu-Niculescu-Mizil et al., 2013, p. 4). This is ideally counterbalanced by the size of the datasets, as well as users matching the etiquette of the online platform over specific cultural characteristics. Essentially, the majority of users will interact in a way that is more synergetic to the platform and its users, speech that doesn't match this can be considered, for the most part, outliers.

Another potential point of note is the accuracy of an algorithmic analysis of both Reddit and Wikipedia. In a study comparing language models trained to classify hate speech across YouTube, Reddit, Twitter and Wikipedia. Wikipedia, and Reddit in particular, saw some of the lowest accuracies compared to their counterparts (Salminen et al., 2020, p. 21-22). While this study relies on language models, which classify speech differently from the ConvoKit algorithm, it's important to note the potential difficulty to classify utterances algorithmically. In the case of Reddit, users make use of a lot of site specific speech, which can get even more specific in specialized Subreddits, which have the potential to disrupt the weights of the politeness strategies collected. Despite this, the numbers in the cases of the language models still falls in a range of acceptability where meaningful conclusions can be made.

One other point that could skew data is the cutoff point of the Wikipedia Talk Corpus. While the other data sources cutoff data at around 2018, the Wikipedia Talk Corpus cuts off data in 2012. This does leave a potential gap in data which could alter some of the generalizable findings. While not ideal, the majority of the timespan of the datasets overlap, also no major global events occurred during this time that may have altered moods of general discourse. This will not scupper the utility of the dataset, it may indeed introduce some potential variance that should be noted moving forward.

3 Methodology

As mentioned previously, ConvoKit serves as the primary source of data and tooling. To accommodate its use, a short program was written to easily process the data[15]. The program was written in the Python programming language, which allows for easy interfacing with its features. While easy to use, the primary disadvantage of python is its speed, which can prove particularly painful for some of the larger spats of data. Breaking down the actual control flow of the program, after running and specifying the corpus to be used, the program will automatically download and load the corpus into the program. The corpus will then be parsed and loaded into a form readable by the program. Once parsed and readable by the program, an algorithm previously mentioned that is capable of isolating the politeness features (Table 1) is run on the individual utterances in the loaded corpus. This will attach a count of each politeness feature to its corresponding corpus entry. Once finished, each utterance will be assigned a politeness value based on the feature occurrence (Function 1), which will then be written out into a CSV[16] file to be parsed and analyzed with graphing software. Utterances in each corpus are contained in *dictionary*[17] data structures, meaning no metadata is embedded in the text of the utterance itself, but the data structure.

3.1 Identifying Utterance Politeness

As mentioned, once the program parses the corpus and isolates the politeness features of each utterance, the program is able to determine the politeness of the whole utterance. At a high level, the program is able to compose the counts with the weighted perceived politeness of each feature into a single score. This is accomplished with this function (Function 1). In the function, the counts of the number of instances of each feature is multiplied by the perceived politeness of each feature (Table 1). This is then normalized to fit into the same range of values the politeness features are classified with, -1.0 as most impolite and 1.0 as most polite.

This can also be parsed more mathematically. If we let $f_{\text{politeness}}$ represent the set of politeness features observed in an utterance, where each feature i is associated with a weight w_i. The politeness score S for the utterance is calculated by summing the product of the count of each feature and its corresponding weight:

---

[15]Source: https://github.com/ConnorMilligan/politeness

[16](Comma Separated Values), a common file format that can be read by spreadsheet software.

[17]A dictionary is a collection of data in which each value can be fetched with a *key*. This mimics a real dictionary in which one must look up a word (key) to find a definition (value).

Figure 1: Code snippet of the strategy for determining a politeness score of an entire utterance based on counts of contained politeness strategies.

```python
def get_politeness_score(utt):
    # Take each politeness feature and multiply it by its weight, then sum the values
    score = sum([utt.meta["politeness_strategies"].get(politeness_feature, 0) * weight
     for politeness_feature, weight in politeness_features.items()])

    # Sum of absolute weights
    abs_weights_sum = sum(abs(weight) for weight in politeness_features.values())

    # Scale the score based on the sum of absolute weights
    # Avoid division by zero
    if abs_weights_sum == 0:
        normalized_score = 0
    else:
        normalized_score = score / abs_weights_sum

    return normalized_score
```

$$S = \sum_{i \in f_{\text{politeness}}} (\text{count of } i) \times w_i$$

Next, we compute the sum of the absolute values of all weights, denoted as W_{abs}:

$$W_{\text{abs}} = \sum_{i \in f_{\text{politeness}}} |w_i|$$

To ensure a well-scaled politeness score and avoid division by zero, we normalize S by W_{abs}. If W_{abs} is zero, indicating no observed features, the normalized score N is set to zero. Otherwise, S is divided by W_{abs} to obtain N:

$$N = \begin{cases} 0, & \text{if } W_{\text{abs}} = 0 \\ \frac{S}{W_{\text{abs}}}, & \text{otherwise} \end{cases}$$

The resulting normalized score N represents the politeness level of the utterance, scaled to fall within the range of -1 to 1. This normalization process ensures that the score is robustly adjusted to the distribution of weights, providing a reliable measure of politeness.

Here is an example of these calculations done by hand with the following utterance: "Could you restore Category:Second wave synthpop acts (or at least the bands formerly listed therein) to my userspace? I was checking out some of the bands and bookmarked it to check out some others later, but found that it was deleted. Thanks"[18]. This utterance, when parsed, finds five instances of politeness markers: One of **Gratitude** with "*thanks*", one of **1st Person** with "*my*", one of **1st Person Start** with "*I*", one of **2nd Person** with "*you*", and one of **Subjunctive** with "*could*". With the counts of all of these, we will multiply the counts by their politeness weights (Table 1) and add them together.

$$\text{Score} = (1 \times 0.87) + (1 \times 0.08) + (1 \times 0.12) + (1 \times 0.05) + (1 \times 0.47) = 1.59$$

This will provide us with the raw score, but the next step is to normalize the data. We will take the sum of the absolute values of all the weights and divide it by our score to normalize. The sum of the absolute values of the weights will always be the same.

$$W_{\text{abs}} = 6.439999999999998$$

$$\text{Normalized Score} = \frac{1.59}{W_{\text{abs}}} = 0.2468944099378883$$

Rounding up to 0.25 we have our final normalized politeness score of the utterance.

---

[18] Wikipedia Talk Corpus, ID: 262154.

4 Results

Due to the size of the dataset, the time to run the data was not insignificant, with all the algorithms run in about 1 hour and 20 minutes on a powerful computer[19]. The collected data was organized in a count of the number of utterances per range groups of 0.02, emanating from 0. For example, on the positive axis, groups range from 0.0, 0.2, 0.4, 0.6, ... and so on. A value fits into one of these ranges, defined by the lower value. For example, A value of 0.23 would fit into group 0.22 and a value of -0.03 would fit into -0.2. Also collected was the percentage use of each politeness feature for each corpus.

Figure 2: A percent distribution of the occurrence counts of utterances in ranges of 0.02 from 0 on a politeness scale across Reddit and Wikipedia Talk.

(a) Comparison of general datasets for Reddit and Wikipedia.

(b) Comparison of the CGA dataset for Reddit and Wikipedia.
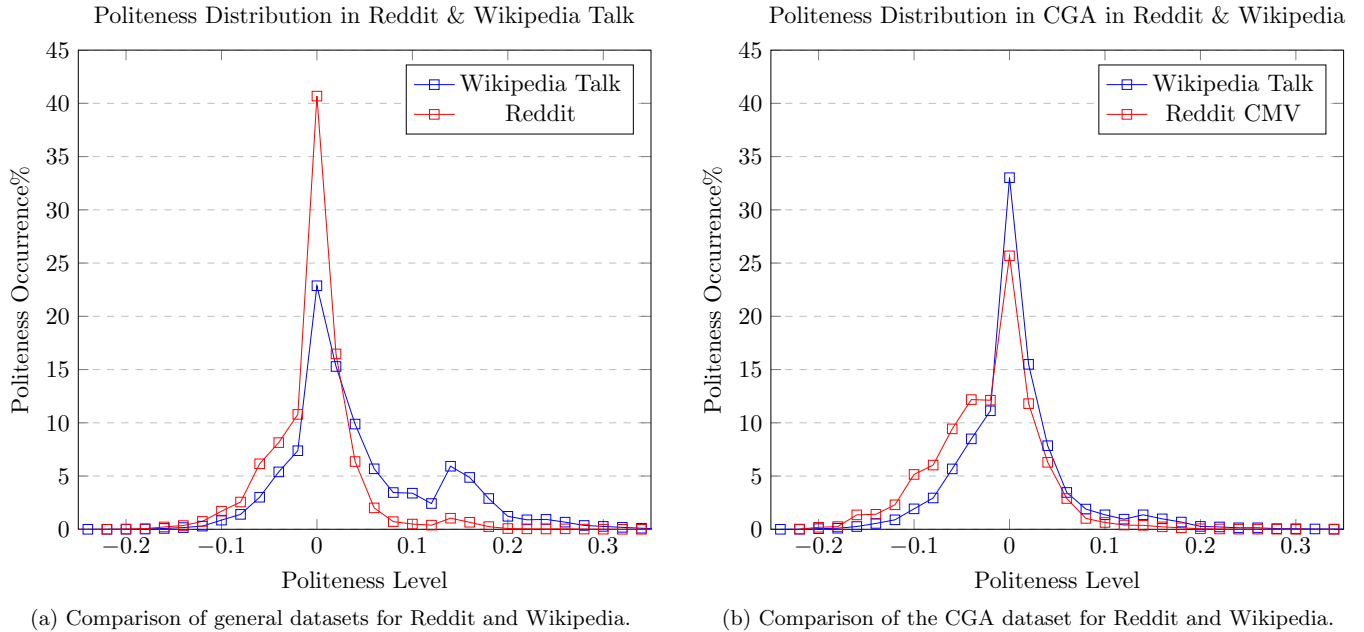
4.1 General Reddit & Wikipedia

The general corpora for Reddit and Wikipedia showed similar patterns in general, but with differing amplitude (Figure 2a). The largest percentage is 0, with utterances typically too small to have any identifiable indicators. Moving towards the positive axis, percentage of politeness occurrence dramatically falls off for both Reddit and Wikipedia before spiking again between 0.1 and 0.2 after which it slowly tapers off, with utterances not found with politeness levels of around 0.35 and higher. Throughout the positive axis, Wikipedia maintains a higher level of politeness than that of its counterpart, with the observed 0.1 to 0.2 spike being far more substantial, with Wikipedia trending 4.88% higher than Reddit at its apex. As for the negative side of the axis, politeness falls much more dramatically than on the positive side. It is also worth noting that there is no equivalent spike in −0.1 to −0.2. While Reddit trended less polite in the positive side, it trended more impolite than Wikipedia as both eventually tapered off, with few utterances extending past −0.24.

As for the actual frequency of strategies, Wikipedia utterances trended more polite (Figure 3). Vocabulary-wise, there are many more occurrences of positive connotated words in Wikipedia than Reddit. In terms of negatives, they are about at parity, with Reddit trending just slightly above. Pronouns are found more on Wikipedia, with the more negatively perceived *2nd person start* falling to close to parity, but with Wikipedia still leading. In terms of other negative features, Reddit surpass Wikipedia in both features of direct speech. In terms of positivity, more obvious features of politeness

---

Figure 3: Comparison of general datasets for Reddit and Wikipedia.

Frequency of Politeness Strategies in Reddit & Wikipedia Talk
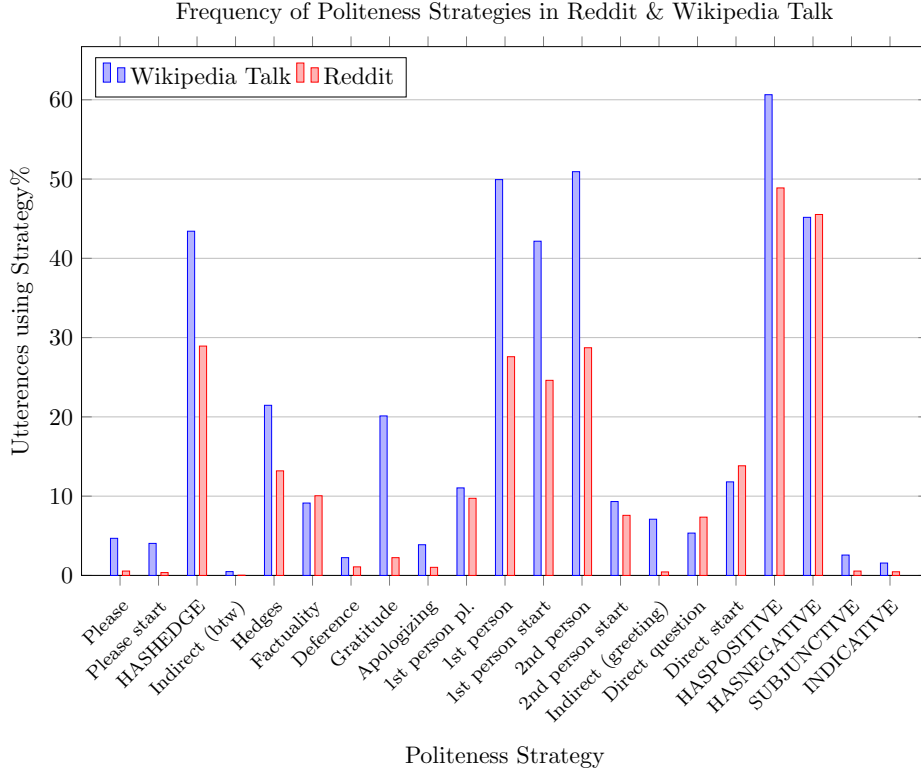
trend towards Wikipedia, with significant variance in use of *please*, hedges, and most notably gratitude, leading nearly 17% over Reddit. Of note as well is the disparity between negative and positive vocabulary, with negative tokens appearing at similar distributions, with them only slightly higher on Reddit. With positive occurrences, the gap is far wider, with Wikipedia seeing occurrences of positive tokens at more than a 10% greater frequency.

4.2 CGA Reddit & Wikipedia

When it comes to the more specialized CGA datasets, the politeness distribution is generally at closer parity (Figure 2b). Both platforms contain the same apex at 0 in their politeness levels, though while Reddit's falls from a bit over 40% in the general to just about 25%, Wikipedia's 0-politeness utterances climbs by nearly 10%. This indicates that Reddit falls over a wider distribution and Wikipedia smaller in comparison to the general corpora. Beginning again with the polite side of the chart, results remain close to parity, with Wikipedia holding a slight lead in politeness. What is of note additionally it the lack of the same spike in politeness between 0.1 and 0.2. While it is still present, it is of a dramatically lower amplitude. Occurrences of higher politeness utterances fall off gradually and appear less frequently than the general corpora. In terms of the negative side of the chart, Wikipedia demonstrates a similar trend as its general results, albeit with a moderately higher amplitude. Reddit on the other hand sees instances of lower politeness decline less gradually than before, all while trending higher than its Wikipedia counterpart.

In terms of the actual frequency of occurrences of detected politeness strategies in CGA (Figure 4) the changes are quite significant. Reddit usurps Wikipedia in all but usage of *please*, indirect markers, gratitude and apologizing. These are, incidentally, most of the highest rated strategies for politeness (Table 1). Save from the change in strategy occurrence, the general trends match the frequencies of the general corpora fairly closely, with most polite strategies that Wikipedia leads with decreasing across both datasets.

Figure 4: Comparison of CGA datasets for Reddit and Wikipedia.

Frequency of Politeness Strategies in CGA on Reddit & Wikipedia Talk
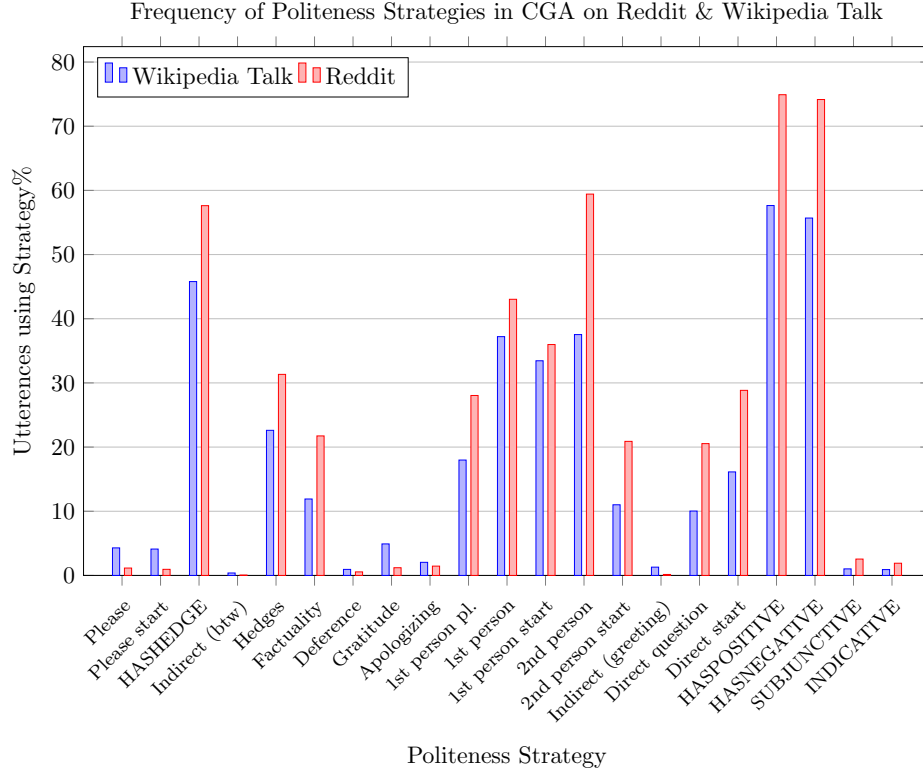
5   Discussion

In explaining some of the trends present in the graph, there are a few points to make note of. The significant apex in utterances tagged with 0 politeness is most certainly those with smaller utterances without enough tokens to have any features to detect. The other significant trend is the previously-mentioned jump in politeness between 0.1 and 0.2. This is likely in cases in which a user is making an active effort to produce a polite utterance. Aside from some of these trends, the trends resemble a normal distribution, as would be expected for the parameters of this study.

Overall, the results taken show that both general and argumentative discourse on Wikipedia demonstrates higher clemency than that of its counterpart with Reddit. This is indicative of Task-Oriented Discourse showing more politeness characteristics than Opinion-Oriented Discourse. It is difficult to make a concrete conclusion as to if this is a trend that is observable in most cases due to the limited scope of the size of this study.

This study is far from perfect, and would likely see some value in preforming it again with an altered methodology. The main issue concerns the accuracy of a lexical analysis of politeness. While it results in scoring that is generally accurate, some finer points are not classified. This also has the issue of missing some surrounding context, which is an issue of the dataset and its organization. Including surrounding context is possible, but requires a specialty dataset and accompanying algorithm.

If this study were performed again, it would likely use a specially trained neural network to identify politeness. This has the benefit of being able to classify aspects that are not as explicit as simple lexical features, as well as surrounding context. The main drawback would be consistency, as well as a lack of concrete metrics for classification.

Moving forward, the results of this study to show some meaningful trends in the theory of Task-Oriented Discourse and Opinion-Oriented Discourse politeness. It is worth continuing the study by looking at data from more sources that fit into these two categories of discourse to see if the current hypothesis can be generalized at all.

Results from this study could prove useful in an age where online hostility and misinformation are both major issues in the digital landscape. Those with aspirations to create new and unique online social platforms could be influenced by results such as these when designing its structure and function.

References

Blum-Kulka, S. (1987). Indirectness and politeness in requests: Same or different? *Journal of pragmatics*, *11*(2), 131-146.

Chang, J. P., Chiam, C., Fu, L., Wang, A. Z., Zhang, J., & Danescu-Niculescu-Mizil, C. (2020). Convokit: A toolkit for the analysis of conversations. *arXiv.org*.

Danescu-Niculescu-Mizil, C., Lee, L., Pang, B., & Kleinberg, J. (2012). Echoes of power: language effects and power differences in social interaction. In *Proceedings of the 21st international conference on world wide web* (p. 699-708). ACM.

Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors. In *arxiv.org*. Ithaca: Cornell University Library, arXiv.org.

Forestal, J. (2023). Social media, social control, and the politics of public shaming. *The American political science review*, 1-15.

Henningsen, M. (2017). *Encyclopedia of communication research methods*.

Hocquet, A., & Wieber, F. (2018). Mailing list archives as useful primary sources for historians. *Internet histories*, *2*(1-2), 38 - 54. Retrieved from https://shs.hal.science/halshs-01916970 doi: 10.1080/24701475.2018.1456741

Nobarany, S., & Booth, K. S. (2015). Use of politeness strategies in signed open peer review. *Journal of the Association for Information Science and Technology*, *66*(5), 1048-1064.

Salminen, J., Hopf, M., Chowdhury, S. A., Jung, S.-g., Almerekhi, H., & Jansen, B. J. (2020). Developing an online hate classifier for multiple social media platforms. *Human-centric computing and information sciences*, *10*(1), 1-34.

Zhang, J., Chang, J. P., Danescu-Niculescu-Mizil, C., Dixon, L., Hua, Y., Thain, N., & Taraborelli, D. (2018). Conversations gone awry: Detecting early signs of conversational failure. *arXiv.org*.