

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Pragmatics

journal homepage: [www.elsevier.com/locate/pragma](http://www.elsevier.com/locate/pragma)

# “I appreciate u not being a total prick ...”: Oppositional stancetaking, impoliteness and relational work in adversarial Twitter interactions

Camilla Vásquez

University of South Florida, USA



## ARTICLE INFO

### Article history:

### Keywords:

Online aggression  
 Conversational toxicity  
 Face-work  
 Relational work  
 Stancetaking  
 Twitter interaction

## ABSTRACT

By examining a set of Twitter interactions following a 2018 media event, this article contributes to a growing body of research addressing the fluid and shifting nature of relational work in digital contexts. The Twitter interactions were posted in response to a tweet by documentarian Hari Kondabolu, in which he thanked actor Hank Azaria for Azaria's apology about voicing the controversial character of “Apu” on the long-running animated television program, *The Simpsons*. Kondabolu's tweet prompted responses from over 500 Twitter users, some of whom contributed to larger discussions about racism and structural inequality on one hand, to extreme political correctness on the other. As they participated in acts of online stancetaking to express their often-opposing positions, Twitter users simultaneously engaged in face-work to manage the interpersonal dimensions of their communication. Providing a micro-analysis of two case study examples, I highlight some of the relationally-oriented discourse-pragmatic strategies employed by Twitter users in interaction with others. One instance shows how a dismissal of the other user's perspectives leads to increased aggression, while the other instance offers insights into how users might signal their willingness to engage in a civil exchange, despite opposing stances.

© 2021 Elsevier B.V. All rights reserved.

## 1. Introduction

In contemporary flows of information, mass media outlets and digital media platforms operate in a recursive cycle. As controversial news events circulate via mass media sources, comments on these news events proliferate on social media platforms, such as Twitter. In turn, some mass media outlets then feature selected tweets in their subsequent reporting on the same topic. The present study highlights exactly this phenomenon, by examining a set of Twitter interactions following a 2018 media event: a public statement made on a late-night television program by actor, Hank Azaria. The statement from Azaria – who voices the character of “Apu” on the long-running animated television program, *The Simpsons* – addressed explicit criticism raised by documentarian Hari Kondabolu about ethnic stereotyping and offensive representations of South Asians in popular media. Subsequently, numerous mainstream news stories reporting on this media event featured tweets that were posted in response to a tweet by Hari Kondabolu, in which Kondabolu thanked Hank Azaria for his public apology and for his

E-mail address: [cvasquez@usf.edu](mailto:cvasquez@usf.edu).

willingness to “step aside” and stop voicing the character of Apu. Kondabolu's tweet expressing gratitude was met with over 500 replies, several of which would evolve into longer networked conversations among Twitter users.

On the surface, online discussions about a television cartoon character may seem trivial. However, this media event inspired numerous Twitter users to participate in larger discussions about racism and structural inequality on one hand, to extreme political correctness on the other. The preponderance of tweets ostensibly arguing over the portrayal of a cartoon character, but in reality engaging in larger debates about more serious social and political issues, supports sociologist [Murthy's \(2013\)](#) claim that profound and banal discourses often exist side-by-side on Twitter.

In this article, I focus exclusively on networked dialogues, or conversation-like interactions involving two Twitter users, which resulted from over 500 comments reacting to Kondabolu's message of thanks to Azaria on Twitter. Drawing on the interrelated notions of face ([Goffman, 1967](#); [Brown and Levinson, 1987](#); [Haugh, 2013](#)), relational work ([Locher, 2008](#); [Locher and Watts, 2005, 2008](#)) and linguistic impoliteness ([Bousfield 2008](#); [Culpeper, 2005](#)), I examine a range of discourse-pragmatic strategies employed by Twitter users to attend to the interpersonal dimensions of communication: in some cases, to signal their dismissal of perspectives which differ from their own, and in other – much more rare – instances, to signal their willingness to engage in a civil interactional exchange, despite opposing stances.

This study explores how online participants, who take part in controversial interactions on Twitter engage in relational work while, at the same time, they take up stances in which they actively *disalign* with their interlocutors. Specifically, I aim to show different ways in which social media users manage the face wants of others with whom they are actively disagreeing online. This study contributes to a growing domain of research investigating impoliteness in digital contexts (e.g., [Angouri and Tseliga, 2010](#); [Graham and Hardaker, 2017](#); [Lorenzo-Dus et al., 2011](#)) – and lends support to those scholars suggesting that new pragmatic norms may be evolving on different social media platforms ([Matley, 2018b](#)). Additionally, as I will elaborate below, this study also engages with recent questions raised by researchers in computer science about what kinds of language may be helpful in shifting “toxic” online interactions back to a more civil mode of discourse (e.g., [Zhang et al., 2018](#)). As such, this work draws attention to an important area for further inquiry, by raising the issue of how to promote constructive online communication among individuals who may hold radically different beliefs and opinions – a matter of no small consequence in an era of increased political division and polarization.

## 2. Media event

While the specifics of the particular media event that is the focus of this study are unique, it is hardly an exceptional occurrence. Indeed, as [Márquez-Reiter and Haugh \(2019\)](#) have observed, much contemporary discourse in both mass media and social media is characterized by public denunciations and the discursive negotiations of morality which tend to follow. As noted earlier, the specific media event that sparked the Twitter interactions under investigation in this study was a public apology, made by actor Hank Azaria on a late-night television talk show in April 2018. Among Azaria's many roles, he is well-known for voicing the character of Apu on *The Simpsons*. Azaria's character, Apu Nahasapeemapetilon, is an Indian immigrant who owns a convenience store and is the married father of eight children. Azaria has been voicing Apu – as well as dozens of other characters on *The Simpsons* – since the program's first season aired in 1990. However, around 2007, critical responses reacting to Apu as an offensive cultural stereotype began to appear in the U.S. popular media. One of the most outspoken critics of the character has been Hari Kondabolu, a North American comedian of Indian descent, who directed a documentary film called *The Problem with Apu* ([Kondabolu, 2017](#)).

Perhaps unexpectedly, the writers of *The Simpsons* decided to respond to Kondabolu's and others' critiques about Apu within an actual episode of the program, which aired on April 8, 2018. In this episode ([Westbrook, 2018](#)), the character of Lisa Simpson is shown in her bedroom speaking to her mother, and on her nightstand rests a framed photo of Apu. Referring indirectly to Apu, Lisa turns to the audience, breaking the fourth wall, and says: “Something that started decades ago and was applauded and inoffensive is now politically incorrect. What can you do?” Immediately after the episode aired, several media outlets critiqued this response from *The Simpsons* writers as “tone-deaf,” “awkward” and “dismissive.”

However, two weeks following the airing of the episode, Hank Azaria apologized on *The Late Show with Stephen Colbert*. Azaria not only apologized to anyone who had been offended or hurt by his portrayal of Apu, but he also disaligned with the stance that had been taken by the show's writers – and he offered to stop voicing the character of Apu altogether. In this way, Azaria accepted some of the blame for accusations of cultural insensitivity related to the Apu issue, and in [Goffman's \(1971\)](#) terms, his apology served as remedial work functioning to repair the interaction order.

Azaria's public apology prompted Kondabolu to tweet: *Thank you, @Hank Azaria. I appreciate what you said & how you said it.* As shown in [Fig. 1](#), Kondabolu's expression of gratitude was incorporated into a retweet from *The Late Show*, and responded intertextually to an embedded video clip of Azaria's apology on the program. The data for this study consist of some of the 500+ replies to this particular tweet, specifically those which resulted in dialogic interactions among other Twitter users.

Initially, I collected the comments responding to Kondabolu's tweet expecting to find other metapragmatic evaluations of Azaria's apology. However, there were almost none. Instead, the majority of users' comments involved oppositional stancetaking around issues of representation in the media. In other words, when a Twitter user expressed agreement with Kondabolu's message – essentially sharing the perspective that the portrayal of the Apu character reinforced culturally-offensive stereotypes – if any reactions followed, then they typically came from Twitter users who expressed the opposite perspective. Consequently, rather than exploring metapragmatic evaluations of apology, I became interested in how Twitter



Fig. 1. Kondabolu's tweet.

users managed the interpersonal dimensions of their communication while engaging in dialogues that were most often oppositional and frequently adversarial.

As is characteristic of so many contemporary social media discourses, the texts that are the focus of my analysis are heavily steeped in popular culture – centering on a character who has appeared on a popular television program for over three decades – as well as in current identity politics, at a moment when issues of race, representation and inequality have emerged as major topics of public debate in the United States and elsewhere. At the same time, these same texts are also densely intertextual in that they reference multiple layers of prior textual activity: a television cartoon, a documentary, and a late-night talk show, among others. Because such references are often made indirectly, it can be difficult to determine which specific prior text/s Twitter users are referring to in their comments.

Many of the tweets analyzed here are also highly metapragmatic, in the sense that this particular issue began with a critique of how a specific individual literally gives voice to a certain type of character and continued with public evaluations on social media regarding the appropriateness, or necessity, of Azaria's speech act of apology that followed (One of the many critiques raised was the issue of why Apu was portrayed by Azaria, a non-Asian, rather than by a South Asian actor.) It is worth pointing out that while there is nothing antagonistic about Kondabolu's tweet of thanks (Fig. 1), inasmuch as it functioned to align Azaria's actions with Kondabolu's prior public denunciations of the Apu character, it sparked a number of antagonistic comments and reactions from other Twitter users who maintained an opposing stance on the issue. Beyond discussions of whether or not the character of Apu is offensive, in responding to Kondabolu's tweet other Twitter users engaged in debates related to a much wider range of controversial social issues – extending even beyond stereotypes, racism and inequality into issues such as gun control, healthcare, and so on.

I have offered the detailed background information above to provide macro- and meso-level context (Garcés-Conejos Blitvich and Sifianou, 2019) for the interactions I will analyze below. To address how Twitter users attended to the interpersonal dimensions of communication while engaging in often adversarial dialogues, in the next section I discuss the key concepts of face, relational work and stancetaking in online interactions. In the section that follows, I focus more specifically on concepts associated with adversarial or antagonistic discourse, which are highly relevant to the analysis: linguistic impoliteness, face attack and – a related construct from computational studies – conversational toxicity.

### 3. Face, relational work and stancetaking in digital contexts

The notion of face as an individual's public self-image first theorized by Goffman (1967:5) as “the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact” has been central in pragmatics. The role of face in accounting for variation in the expressions of different speech acts – in particular, face-threatening acts and the strategies associated with their mitigation – became the main focus of Brown and Levinson's Politeness theory. Yet, in spite of its vast influence on scholars working across a wide range of disciplines, Brown and Levinson's work came under critique in the early 2000s as a result of several perceived limitations. Locher and Watts (2005, 2008), for instance, argued that the management of social relations through language (i.e., face-work) entailed a broader range of linguistic phenomena than “politeness” alone – including what they call more unmarked forms of “politic” behavior, as well as impoliteness. In addition, they maintained that all of these are discursive phenomena that must be studied in interaction, not solely at the level of the utterance. Therefore, instead of focusing on politeness, Locher (2008) and others have

advocated that in the field of pragmatics, analytic attention should instead be given to “relational work,” which refers more broadly to the interpersonal dimensions of language used in interaction.

In terms of relational work in networked communication, [Bou-Franch and Garcés-Conejos Blitvich \(2018: 138\)](#) note that while there is a growing body of scholarship on im/politeness research focusing on interactions in various digital contexts, “one particular shortcoming arises from empirical research that focuses on [...] one specific interactional behavior without taking into account how other behaviors may be brought into the same interaction.” Indeed, in their detailed analysis of a networked polylogue comprising 54 comments posted by 12 users on Facebook, the authors highlight the interplay between different kinds of relational work, as they demonstrate how relational practices fluctuate within a specific interaction – and sometimes even within a single comment.

Also relevant to understanding relational work is the dimension of interactional stancetaking known as “alignment.” Alignment concerns the extent to which a language user’s expressed stance is convergent in perspective to that of some other language user, real or imagined. Building on [Du Bois’ \(2007\)](#) model of stance, [Kiesling et al. \(2018: 685\)](#) explain that “alignments and disalignments are created between (canonically two) speakers as they display similarity and difference” in their evaluation of some object or idea. [Kiesling \(2009\)](#) conceptualized stance “as a person’s expression of their relationship to their talk *and to their interlocutors*” ([Kiesling et al., 2018: 707](#), emphasis mine); in this way, verbal alignment – or disalignment – with one’s interlocutor is clearly a key aspect of relational work. Social media spaces, such as Twitter, which include affordances that encourage the sharing of opinions can be considered “stance-rich” online environments ([Barton and Lee, 2013: 92](#)).

#### 4. Face, impoliteness, and “conversational toxicity”

Although definitions of impoliteness are abundant ([Garcés-Conejos Blitvich and Sifianou, 2019](#)), impoliteness is generally associated with “face-aggravating behavior” ([Locher and Bousfield, 2008: 3](#)) and can be described as any behavior that causes or is “presumed to cause offence” ([Culpeper, 2010: 3233](#)). Forms of impoliteness may vary in their intensity and not all impoliteness is necessarily intentional, which adds to the complexity in defining the construct. Recent years have seen a growing recognition of, and scholarly interest in, different forms of online impoliteness in digital contexts, often discussed in terms of online aggression, incivility, or face-attack.

Social media scholars addressing impoliteness have found that, due to “factors such as anonymity, invisibility and minimisation of authority online” ([Matley, 2018a, p. 37](#)), the pragmatic norms that would be expected in face-to-face interaction may be quite different in computer-mediated interactions. Specifically, the disembodied nature of much online communication has been linked to online aggression. In accounting for online aggression and hostility, many scholars draw on deindividuation theory, which posits that the disembodied – and (in some cases) anonymous – nature of online communication leads participants to not treat others as individuals. Consequently, deindividuation is associated with increased polarization among groups as well as in-group cohesiveness ([Khazarie and Talehbzadeh, 2020](#)). Most notably, trends of linguistic impoliteness have been documented in several recent studies, which have examined, among others: aggressive comments on Facebook ([Parvaresh and Tayebi, 2018](#)), politically controversial messages posted by a single user on Twitter ([Terkourafi et al., 2018](#)), as well as playful face-attacks (or “ludic impoliteness”) in a corpus of tweets targeting a politician ([Vladimirou and House, 2018](#)). As these and other studies have shown, certain topical contexts, such as politics and religion, are susceptible to antagonistic behavior online.

Referring specifically to impoliteness in online contexts, conversational toxicity is a closely-related construct that comes from computational studies. Just as multiple definitions of impoliteness abound in pragmatics, conversational toxicity has also been variably conceptualized by computational scholars, ranging from mild forms of linguistic impoliteness to extreme face-attack: in other words “content that can be anywhere from slightly abusive to hate inducing” ([Gunasekara and Nejadgholi, 2018: 21](#)). Conversational toxicity is often treated as a commonsense category ([Wulczyn et al., 2017](#)), with little attention paid to disambiguating possible differences in intensity, or other characteristics, among different forms. Regardless of the definition adopted, computational studies often have as their goal the development of automated toxicity detectors. While these systems generally involve identifications of toxicity at the level of lexis (i.e., “bad words lists”) or single utterance, more recent studies have begun examining the extent to which accurately identifying toxic content involves context-sensitivity, or taking into account the surrounding discourse ([Pavlopoulos et al., 2020](#)).

One computational study ([Zhang et al., 2018](#)) which included a focus on pragmalinguistics, involved the identification of specific language features at early stages of online interaction that would predict whether or not an interaction would eventually turn toxic. Examining a set of *Wikipedia Talk* discussions, Zhang et al. found that the presence or absence of just a few pragmatic features in the opening moves of an interaction could predict whether a conversation would remain positive and “on-track,” or whether it would devolve into a more toxic interaction. The opening moves associated with toxicity included more second-person pronouns and more direct, challenging questions, whereas those that remained “on-track” included more first-person pronouns, especially when appearing in subjectivizers such as *I mean*, or *I understand*. Acknowledging that the specific context they explored (i.e., *Wikipedia Talk* pages) likely played a significant role in their findings, the researchers called for further exploration of other, perhaps less collaborative, online public discussion venues, such as Reddit or Twitter. The researchers also called for future work to investigate “the complementary problem of identifying pragmatic strategies that can help bring uncivil conversations back on track” (9). Motivated to explore the latter issue, I was curious to discover if there were any instances of antagonistic Twitter interactions in which participants ended up being

more linguistically polite than when they first started. And, if so, what kinds of pragmalinguistic strategies, or interactional resources, characterized such interactions?

Adopting a discursive approach to pragmatics, I offer a microanalysis of two textual interactions on Twitter, which considers the larger social context(s) in which they are embedded. Where impoliteness is concerned, a discursive pragmatics approach focuses on interactional issues taking stretches of discourse as a unit of analysis. Blitvich and Sifanou (2019: 93) note that discursive approaches to im/politeness also “devote more attention to social function and context” than did earlier approaches which tended instead to focus more narrowly on utterance-level meaning. I agree with these authors that ultimately, impoliteness is an interactional accomplishment and that it is “a feature of context and interpretation rather than an intrinsic property of certain linguistic forms” (Blitvich and Sifanou, 2019: 94). Nevertheless, a well-established set of conventionalized impoliteness formulae have been proposed by Culpeper (2010) and others, which include insults, use of taboo words, pointed criticisms, challenging questions, and condescension, scorn and ridicule. When analyzing a stretch of discourse, the presence of one (or more) of these features may be indicators of linguistic impoliteness; however, any assessment of an utterance, turn, or interaction as impolite requires further, multi-level, contextualized analysis.

## 5. User interactions on Twitter

As the most popular microblogging platform, Twitter is very much “event-driven,” with many tweets addressing breaking news and other trending topics. Yet it is not a neutral platform; as Papacharissi and de Fatima Oliveira (2012) have pointed out, Twitter “combines news, opinion and emotion” (p. 266). Tagg et al. (2017) suggest that political discussions and oppositional stancetaking may be more common on Twitter than on Facebook, for example. They speculate that Twitter's affordances, specifically, the “non-reciprocal participation structure” may “play some role in the potential for Twitter to act as a space for dispute and conflict” (50).

In terms of linguistic analyses, Twitter has served primarily as a source for big data or corpus studies (Webb et al., 2017), owing to its open access and the availability of tools that enable researchers to capture large quantities of data from the platform. Conversely, analyses of smaller interpersonal interactions taking place on Twitter are more the exception rather than the norm. As one set of authors points out “to date little work has been carried out on the social organisational characteristics of Twitter interaction” (Housely et al., 2017: 568). Although their own analysis centers primarily on the textual content of individual tweets, Housely et al. (2017) argue for the importance of analyzing threaded interactions on social media. A recent example of such a study is Dynel and Poppi (2019), whose analysis of Stormy Daniels' witty retorts to insults directed toward her on Twitter illustrate the potential of examining sequential interactions in the form of adjacency pairs occurring on the platform. Additionally, Bou-Franch and Garcés-Conejos Blitvich's (2018) close analysis of relational work being carried out in one threaded social media interaction serves as model for the present study, which is concerned with understanding the sequential nature and the interpersonal dimensions of networked dialogues on Twitter.

## 6. Methods

Focusing on the responses to Kondabolu's tweet (Fig. 1), my sampling involved capturing those publicly-available data that any other user of Twitter would have access to. At the time of data collection, Twitter's API displayed only 334 of the 500+ replies to Kondabolu's tweet. After downloading the 334 visible replies, I further divided the replies into those that were followed by comments from other users from those replies that led to no further interaction. As shown in Table 1, almost half

**Table 1**  
Replies to Kondabolu's tweet with subsequent comments from other Twitter users.

Reply + 0 comments	Reply + 1 comment	Reply + 2 comments	Reply + 3–9 comments	Reply + 10+ comments
151	134	21	15	8

(151/334) of the replies to Kondabolu's tweet of thanks received no response from other users, supporting claims that much of the talk on Twitter is actually “monologic” and never responded to (Murthy, 2013: 69). This also corroborates prior observations made by Housely et al. (2017), who in a dataset of 12,000+ tweets, found that a significant proportion of tweets (30%) were followed by no comments.

Table 1 shows that the next largest category (134/334) were replies that were followed by only 1 comment from another user, with no further interaction resulting beyond that. In other words, the majority (285/334, or 85%) of replies to Kondabolu's tweet of did not develop into extended conversation-like dialogues. Conversely, only 15% (44/334) of the replies resulted in threaded interactions, involving 3 or more turns. My analysis focused on these 44 interactions. After reading through the data multiple times, I found that while a few of these extended interactions involved participants who agreed with one another, the vast majority of these interactions involved participants who took opposing stances on the issue(s) being discussed. In nearly every case, these oppositional interactions included some – and frequently several –

pragmalinguistic features conventionally associated with impoliteness, face-attack, and aggression: including challenging questions, taboo language, sarcasm, and insults (Culpeper, 1996, 2008).

As mentioned earlier, my primary aim was to learn more about what discourse strategies appeared in antagonistic Twitter interactions where participants ended up being more civil, or linguistically polite, than when they first started. In the 44 interactions, I found only one such case. In the analysis below, I focus on two interactions, because of the contrast they provide. Interaction 1 is typical and thus representative of the majority of the interactions in these data, in that participants took opposing stances and their interaction concluded with (some form of) face attack. In contrast, Interaction 2 was anomalous, as the only example involving two participants with opposing stances where an initially antagonistic interaction concluded on a more civil note. Following these analyses, I highlight possible discourse-pragmatic strategies that may be helpful for contributing to more harmonious, or at least less overtly antagonistic, communication online.

Where public online discourse is concerned, research ethics remain thorny. Specifically, there is a lack of consensus on the ethical dimensions of reporting research about Twitter (Webb et al., 2017). This is because, on the one hand, Twitter is a public space and no institutional permissions are required for researchers to use data from Twitter. On the other hand, as Spilioti and Tagg (forthcoming) explain, where digital discourse analytic research is concerned, additional, contextualized, micro-level considerations may apply beyond regulatory ethics. In this particular case, due to the emotionally-charged and political nature of the topics discussed in my data – and following the general research ethical principle of doing no harm – I have chosen to use pseudonyms to refer to individual participants rather than users' actual Twitter IDs, as well as to anonymize all other identifying data. Nevertheless, Twitter is an open, publicly-searchable platform, which means that it is possible for readers to retrieve the content of the tweets posted here. However, I have no reason to believe that my reporting on these public tweets in a scholarly venue poses any greater risk to their authors than their original posts did. I refer to all participants using the gender pronouns that correspond to the gender identities displayed in their Twitter profiles.

## 7. Adversarial interaction ending with aggression: interaction 1

The first interaction analyzed is typical of those interactions in this dataset involving Twitter users with opposing stances: it begins with face-attack and ends with further aggression – in this case, with participants insulting one another. Interaction 1 is an 10-turn exchange taking place between two Twitter users, who I refer to as “A” and “B.”

As seen in Fig. 2, A responds to Kondabolu's tweet and begins this interaction with a claim to ethos, with which he positions himself as having some authority on which to speak to the Apu issue. A's claim may be paraphrased as: “I've known

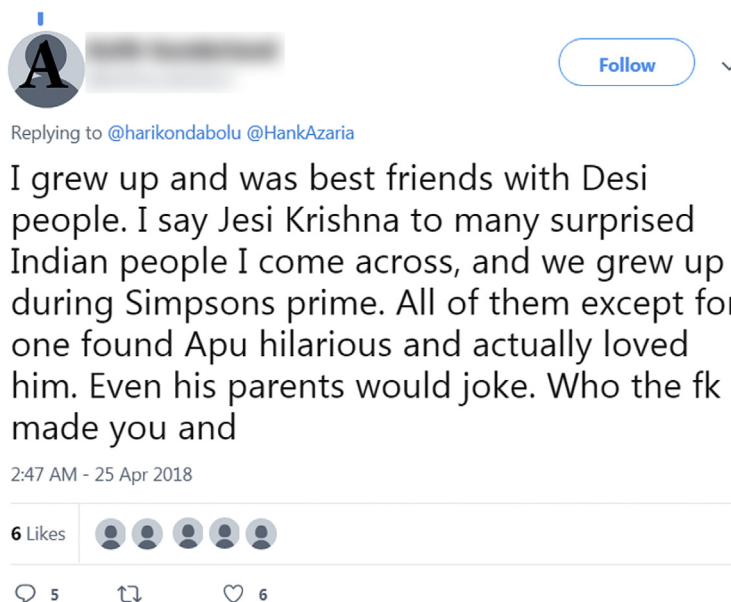


Fig. 2. Interaction 1: Turn 1, Part 1.

lots of Indian people in my life – and all but one of them thought Apu was funny.” Constrained by Twitter's character limit and thus unable to complete his message in a single tweet, A then continues his message by carrying it over into a second tweet (Fig. 3), in which he clearly disaligns with Kondabolu's stance and argues that Kondabolu's critiques of the character of Apu are unjustified. He further complains that Kondabolu is among those who are “ruin[ing] comedy for the rest of us.”

Although A is ostensibly responding to Kondabolu's tweet, it is evident that in A's tweet he is, in fact, addressing some other prior text(s). In other words, the stance focus of A's message is Kondabolu's position about Apu as a stereotype – and not Kondabolu's expression of gratitude addressed to Azaria. Thus A's message (Fig. 2) clearly extends intertextually beyond Kondabolu's simple tweet of *Thank you, @Hank Azaria*, and instead references either Kondabolu's documentary and/or some previous statements made by Kondabolu elsewhere. This is evident from A's attribution of Kondabolu's stance not solely to Kondabolu, but also to fellow actor, Kal Penn (@Kalpenn), as seen in the continuation of A's tweet in Fig. 3. In Fig. 3, A's comment (*couldn't get a multimillion dollar roles [sic] ...*) seems to be addressing some of Kondabolu's prior complaints about ethnic stereotyping in the entertainment industry, showing that debates on Twitter often extend well-beyond the current interaction and build on prior social/cultural discourses, which may be referenced only indirectly.

In the next turn of this interaction (Fig. 4), another Twitter user, B, replies to A's prior two-tweet message and disaligns with A's stance by challenging his position. B does this by repeating the very same syntactic structure used by A in *Who made [X] the majority of Indians in this country?* but replacing the names in the question with the phrase *your friends*. Without adding any further information, B's tweet simply turns A's logic back on himself. Veale et al. (2006) use the term “trumping” to refer to this type of structural parallelism, which may be used to subvert the argument of an interlocutor. In this instance, B intentionally “molds his attack” (315) to mimic the syntax of A's challenging question, engaging in a type of “verbal duel” (331).

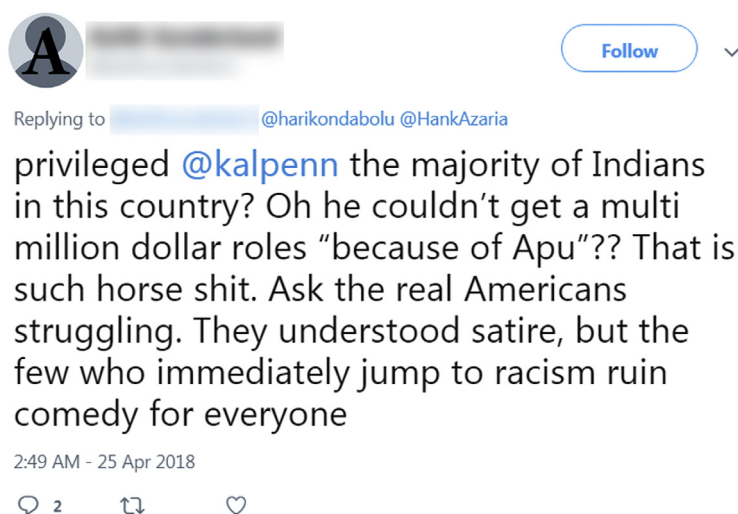


Fig. 3. Interaction 1: Turn 1, Part 2.



Fig. 4. Interaction 1: Turn 2.

In this sequence A's opening texts include features of linguistic impoliteness (Culpeper, 1996) such as name-calling (*privileged*), expletives (*fk, horseshit*) as well as a hyperbolic accusation (i.e. that Kondabolu and his colleague are responsible for *ruining comedy for everyone*). In addition, the pragmalinguistic features of both A and B's turns in the opening sequence of this Twitter interaction (Figs. 2–4), include both direct questions as well as second person pronouns, two of the features that Zhang et al.'s (2018) computational study identified as predicting toxicity. These features (i.e., “unpalatable” questions and personalization of insults) have also been previously identified by Culpeper (2011) as expressions of linguistic impoliteness. Here, in both cases, the rhetorical questions also function as a challenge (*Who made X the majority of Indians [...]?*), calling into question different individuals' rights to speak on behalf of Indian people in the United States. This opening sequence depicted in Figs. 2–4 is then followed by a series of 6 other turns (not shown here due to space limitations) in which these two Twitter users continue their debate in a similarly oppositional and antagonistic fashion. The interaction concludes with the two users recapitulating their initial positions in their final turns, shown here as Figs. 5 and 6.

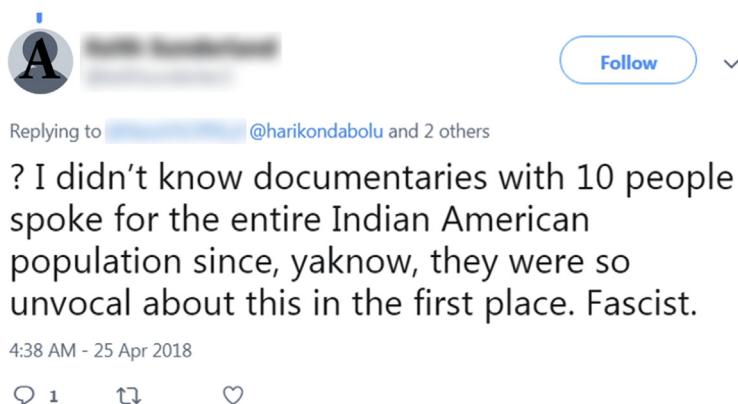


Fig. 5. Interaction 1: Turn 9.

In his closing turn (Fig. 5), A returns to his original line of argumentation, but this time making a more direct intertextual reference to Kondabolu's documentary (i.e., *documentaries with 10 people*). He ends his tweet with a single-word insult, *fascist*. It is unclear if this insult is directed at B, his interlocutor on Twitter, or if instead it is directed at Kondabolu, the primary target of his indignation. Regardless of A's intent, in his reply B (Fig. 6) again mirrors and repeats the same linguistic structures used by A: starting with the turn-initial sarcastic *I didn't know* followed by a proposition, and then ending with a face-attack in the form of a single-word insult, which is structurally similar (*racist/fascist*). Even the unconventional message-initial question mark used by A is repeated by B. B's response in Fig. 6 can again be considered “trumping” (Veale et al., 2006), wherein B fights back with a response that mimics the structure of A's preceding turn.

This Twitter interaction – like the majority of interactions in this particular data set – involves users with divergent perspectives engaging in oppositional stancetaking. Because the opening turns of this interaction included several features associated with linguistic aggression (insults, challenging questions), it comes as no surprise that the interaction concludes in as “toxic” a fashion as it began, with interlocutors trading one-word insults. That impoliteness is often followed by retaliation in kind has been discussed previously by scholars of linguistic impoliteness (Culpeper et al., 2003), as well as those working in other disciplines (Andersson and Pearson, 1999). In both offline and online contexts, aggression often begets more aggression.

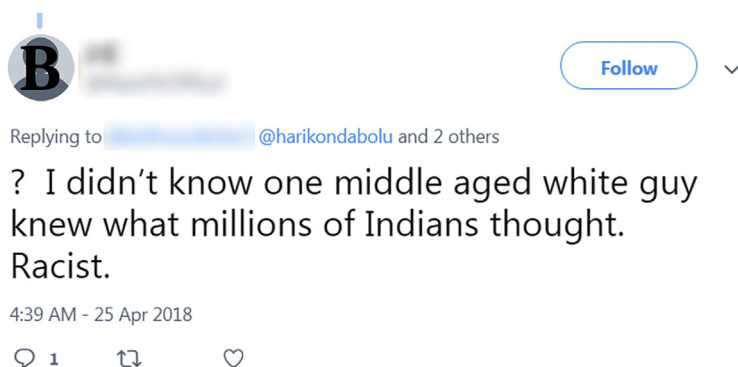


Fig. 6. Interaction 1: Turn 10.

## 8. Adversarial interaction ending with civility: interaction 2

The next example provides an interesting contrast to previous interaction. As I will show, Interaction 2, starts out in a similarly oppositional fashion, but ends up taking a different turn at the end. In this respect, whereas Interaction 1 represents the dominant trend in the dataset, Interaction 2 is anomalous.

As in the previous interaction, Interaction 2 is a multi-turn exchange taking place between two Twitter users, who I refer to as “C” and “D.” Unlike Interaction 1, the Twitter user who posts the first turn in this interaction, C, aligns with Kondabolu’s general stance on the Apu issue, and in doing so, he challenges previous replies to Kondabolu’s tweet by those Twitter users who, through their stancetaking, placed themselves on the opposite side of the issue. Once again, it is worth noting although C’s tweet is technically replying to Kondabolu’s tweet in which he expressed his gratitude to Azaria, the content of C’s tweet extends intertextually well beyond that individual tweet, in its indirect referencing of a number of prior texts and discourses. The first turn (Fig. 7) includes many of the same features predictive of toxicity that appeared Interaction 1, including rhetorical questions, challenging questions, second-person pronouns, expletives and insults (*Why do so many assholes feel like [...] ? Why are you here?*).

However, in his response to C’s tweet, D takes a different approach (Fig. 8). Although his reply to C’s comment also includes a question, it is important to note that D’s question is formulated from a first-person perspective (*So my question is ...*), and in fact, there is no second-person reference anywhere in his tweet. As Zhang et al. (2018) have suggested, the first person perspective adopted here may contribute to diffusing the potential conflict here, rather than to escalating it. D’s question also presents the larger issue in much more impersonal terms (*... at what point does it not matter ...*), which seems to move the discussion to a different level, in shifting the focus more to the issue itself rather than to the person who is articulating an opposing position. Although this too could be considered a rhetorical question, it does not function as a direct challenge, as was the case with the questions in the previous interaction, specifically in the tweets shown above in Figs. 2–4.

Taking up this question posed by D, C in his response (Fig. 9) uses analogic reasoning to extend the discussion further. Specifically, C compares the line of argumentation introduced by D to one that is used in the (U.S.) gun debate. He does so by producing an instance of hypothetical reported speech from an imagined speaker, who is portrayed as voicing a pro-gun stance. Notably, C’s language here begins to match D’s more impersonal framing of the issue, by referring to generic *people* who would use such an argument, thus avoiding the use of first- and second-person pronouns altogether. While C’s response bears some structural similarities to D’s prior turn (impersonal pronoun + question), this is very different from the parallelism resulting in verbal dueling that characterized Interaction 1.

In his next turn responding to C (Fig. 10), D extends his interlocutor’s gun debate analogy to the broader topic of *laws against killing people*. He ends his tweet by reformulating the opening move that he used in his previous turn, shown in Fig. 8 (i.e., *someone is going to always be offended by something*). This time, however, D impersonalizes this argument to an even greater extent by using an existential construction combined with passive voice, which is both agentless and patientless: *There will always be something to be offended by*. Beyond these impersonalization strategies, another key difference between this interaction and the preceding one is that here, C and D attempt to engage with the content of what the other is communicating. In contrast, in Interaction 1, B attended more to the form of A’s utterances, turning A’s words against him. Here, C and D do appear to be working toward reaching at least some degree of intersubjective understanding.

Not willing to give up just yet, in his next turn (Fig. 11) C then tries using a different issue as an analogy: the futility of having laws in place for preventing drinking and driving. However this time, C’s question is phrased as a direct challenge to his interlocutor, by attributing a particular position to him by means of a second-person pronoun used as direct address (*So [...] you’re opposed to [...]?*). Although C’s question is genuine in the sense of seeking a response from D, having an unstated position attributed to D by his interlocutor represents an imposition, which could be construed face-threatening.

In his tweet that follows (Fig. 12), D responds to C’s questioning challenge with a simple negative *No*, followed by a reformulation (*what I mean is [...]*), which he uses to clarify his position. This construction (*What I mean is*) serves as an intention-signaling device, functioning both to clarify the previous misconstrual of his stance and to metapragmatically show that he is trying to explain himself.

Having tried several different approaches, in his fourth turn (Fig. 13) C finally gives up and decides to end the interaction. Yet instead of simply dismissing his interlocutor – as many Twitter users do, especially when they are involved in disagreement with others – C does something different. He begins his tweet by offering a conciliatory statement (*I didn’t mean to lead us so far off track with the analogy [...]*). Here, the negated *I mean* construction is another instance of meta-commentary designed to clarify C’s intention about his previous contributions to their interaction. In an online context, a move such as this one may also serve as a form of reassurance to C’s interlocutor that he has not wasted his time interacting with a troll (Hardaker, 2013), who had been deliberately trying to provoke a reaction from him.

Following this clarification, C expresses appreciation for his interlocutor’s behavior in an unexpected and highly contextually-specific example of relational work: *I appreciate u not being a total prick*. Here, this instance of a “verbal formula mismatch” (Culpeper, 2011: 174) – where a politeness formula precedes an impolite utterance (in this case, an insult) – functions as an expression of appreciation of their exchange, while still maintaining and acknowledging their differences of opinion. Furthermore, it is possible that, in this interactional context, the hedged insult (*not [...] a total prick*) actually serves as an indication of C’s sincerity. Several examples in the dataset included instances of exaggerated mock politeness that were clearly sarcastic: for instance, at the end of an antagonistic exchange, one Twitter user ended her part of the discussion by posting the ambiguous, though, in this context, most likely condescending *bless your heart honey*, to which her interlocutor



Fig. 7. Interaction 2: Turn 1.

responded *same to you sweetheart*. The sudden appearance of politeness markers – which are not qualified in any manner – in the midst of otherwise antagonistic discourse may be construed as mock politeness. If that is the case, then the jocular mitigated insult (*I appreciate u not being a total prick*) serves as a means for C to indicate that his appreciation is genuine. C concludes his turn by conceding that there is no point in further attempts at communication and even wishes his interlocutor a good day.

In his final turn (Fig. 14) D replies by expressing his alignment with C – not about the larger issue, but about C's assertion that it makes sense to end their discussion at this point. D also responds in kind, by wishing his interlocutor, C, a good day, and engages in further relational work by using the casual and friendly homosocial masculine vocative, *buddy*.

Interestingly, Interaction 2 began in a similar way to Interaction 1, with Twitter users not only taking an oppositional stance toward prior statements, but also by using features associated with linguistic aggression and conversational toxicity. However, what differed between the two interactions was the type of discourse that followed. In Interaction 1, the second

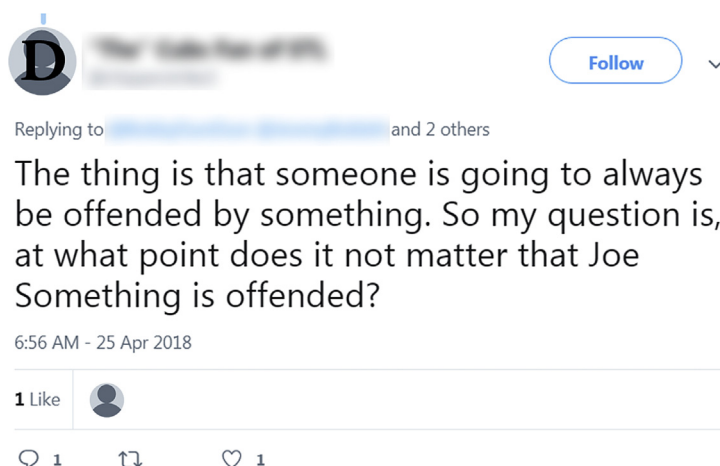


Fig. 8. Interaction 2: Turn 2.

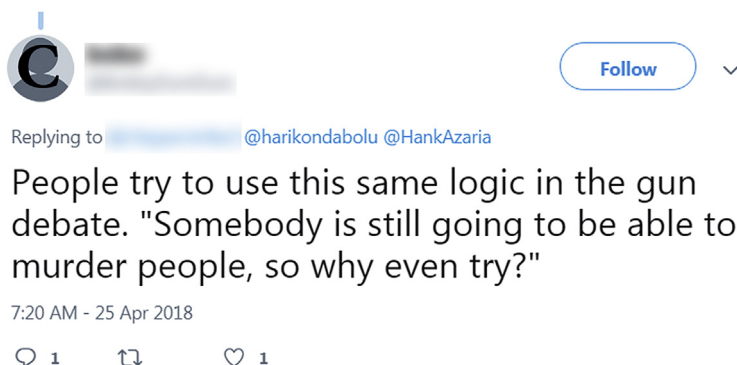


Fig. 9. Interaction 2: Turn 3.

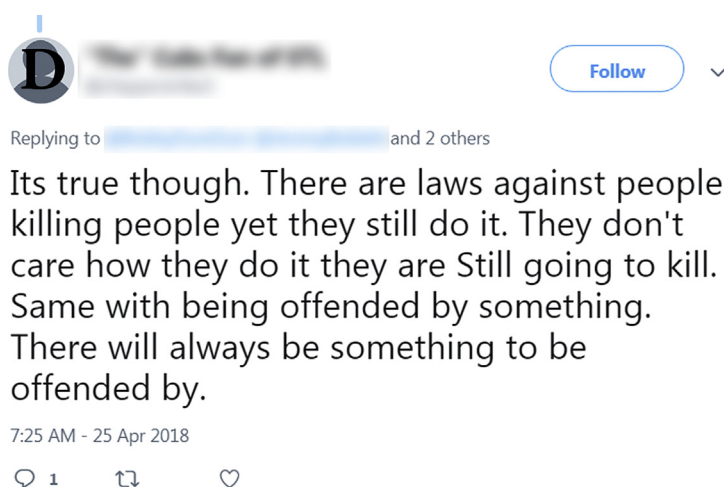


Fig. 10. Interaction 2: Turn 4.

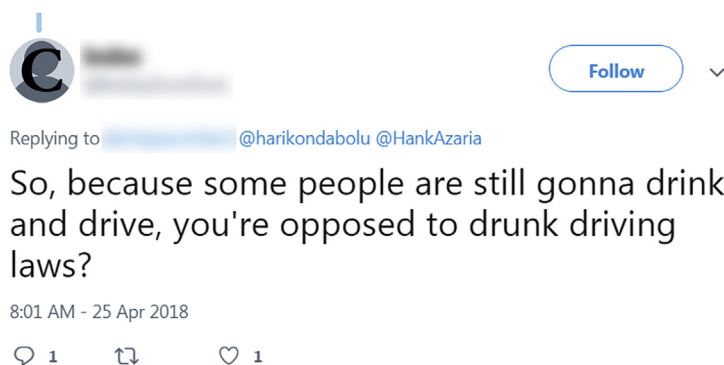


Fig. 11. Interaction 2: Turn 5.

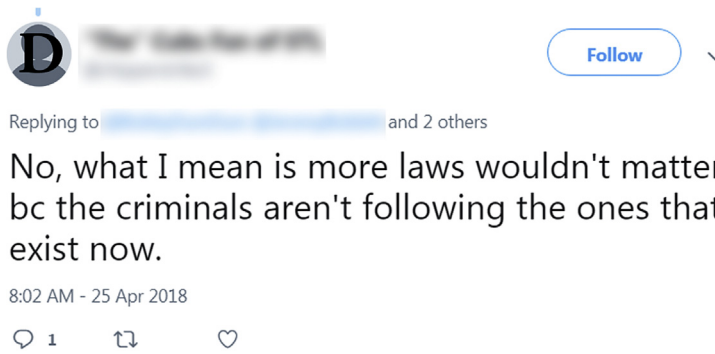


Fig. 12. Interaction 2: Turn 6.

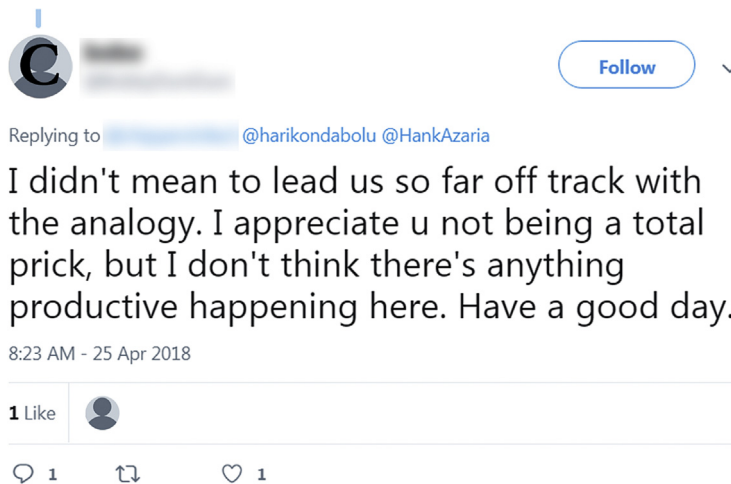


Fig. 13. Interaction 2: Turn 7.

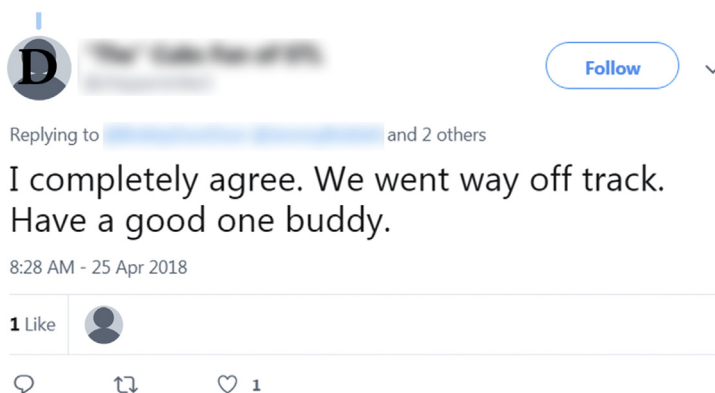


Fig. 14. Interaction 2: Turn 8.

party who engaged in a dialogue continued using features associated with toxicity (i.e., challenging questions, second-person pronouns). In contrast, in Interaction 2, the second party who engaged in dialogue managed to re-frame the discourse, by emphasizing the larger issue, rather than by leveling an attack at the person on the opposite side of the issue. Linguistically, this included avoiding second-person pronouns and using more impersonal language instead. Although it represents the exception and not the norm, Interaction 2 nevertheless indicates that it is possible to de-escalate an interaction that begins with aggression. These interactions demonstrate that just as impoliteness is an interactional achievement, so too, is turning an adversarial dialogue into a more civil mode of discourse.<sup>1</sup>

## 9. Conclusions

The previous analysis of two interactions with similar openings but different closings revealed some promising patterns in terms of their pragmalinguistic features. Corroborating observations from computational research (Zhang et al., 2018), second-person pronouns and direct questions (especially challenging questions) did appear in interactions that eventually concluded with some type of face-threat or face-attack. However, as my analysis of the second interaction showed, one way to reverse this tendency is for users to frame the controversial issue in a more impersonal manner. Moreover, first person pronouns – especially when they appear as part of subjectivizers (such as *I mean* or *I understand*) – seemed to help signal, or clarify, a users' communicative intent. The metapragmatic signaling of communicative intent may be an especially important discursive resource in text-based digital contexts, which present users with different non-verbal resources than those typically relied upon in face-to-face interactions.

Certainly the usual limitations associated with a case study approach apply to what has been presented here. That is, rather than serving as a basis for broader generalizations, the value of the type of analysis and the findings presented above is to show exactly how situated and contextual both meanings and pragmatic norms are. Furthermore, the larger question addressed here regarding what types of relational work are helpful, or necessary, to shift an interaction which begins as “toxic” to a more civil tone is one which clearly requires further research. Additional work on this topic is needed with different datasets, from different platforms, and focusing on interactions orienting to different types of social and political issues. Such research is not only bound to be richly generative, in terms of expanding our understanding of contextually-specific forms of relational work in a variety of online sites, but it is also extremely important, as we continue to explore productive ways to engage in dialogues with people whose beliefs and values differ radically from our own. In addition, this study suggests that future collaboration between computational studies and pragmatics research may potentially enrich our understandings of both impoliteness online and conversational toxicity.

As Tagg et al. (2017) observe, in spite of several highly-publicized cases of hostility and aggression taking place on Twitter, there has been a rather surprising paucity of scholarship on this topic. Many interactions about controversial issues on Twitter involve not only oppositional stancetaking, but also feature various forms of antagonistic discourse (Housely et al., 2017), including extreme forms of linguistic impoliteness, often in the form of ad hominem attacks and other forms of verbal aggression. But some exceptions to this larger trend can be found – and these exceptions are worthy of our analytic attention for what they can tell us about the potential for conducting civil interactions about controversial issues. One of the greatest challenges facing society in an era marked by increasing political polarization is the need for individuals to productively converse or dialogue with others who hold opposing points of view. All too often, such encounters quickly devolve into face-attacks. Further research in pragmatics is needed to shed light on how individuals can instead engage in oppositional stancetaking online in ways that are more respectful and productive.

## Declaration of competing interest

There is no conflict of interest.

## References

- Andersson, Lynne, Pearson, Christine, 1999. Tit for tat? The spiraling effect of incivility in the workplace. *Acad. Manag. Rev.* 24 (3), 452–471.
- Angouri, Jo, Tseliga, Theodora, 2010. “You have no idea what you are talking about!” from e-disagreement to e-impoliteness in two online fora. *J. Politeness Res.* 6 (1), 57–82. <https://doi.org/10.1515/jplr.2010.004>.
- Barton, David, Lee, Carmen, 2013. *Language Online: Investigating Digital Texts and Practices*. Routledge, London.
- Bousfield, Derek, 2008. *Impoliteness in Interaction*. Amsterdam, Benjamins.
- Brown, Penelope, Levinson, Stephen, 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press, Cambridge.
- Culpeper, Jonathan, 1996. Toward an anatomy of impoliteness. *J. Pragmat.* 25 (3), 349–367.
- Culpeper, Jonathan, 2005. Impoliteness and entertainment in the television quiz show: the weakest link. *J. Politeness Res.* 1 (1), 35–72.
- Culpeper, Jonathan, 2008. Reflections on impoliteness, relational work and power. In: Bousfield, D., Locher, M. (Eds.), *Impoliteness in Language*. Mouton de Gruyter, Berlin, pp. 22–53.
- Culpeper, Jonathan, 2010. Conventionalised impoliteness formulae. *J. Pragmat.* 42 (12), 3232–3245.
- Culpeper, Jonathan, 2011. *Impoliteness: Using Language to Cause Offence*. Cambridge University Press, Cambridge.

<sup>1</sup> As one reviewer pointed out, it unlikely that C and D actually learned much from each other, came around to what the other was saying, or reached any compromise – instead they “agreed to disagree.” Indeed, when participants have intractable differences of opinion, “agreeing to disagree” may be the most positive outcome possible from such an interaction.

- Culpeper, Jonathan, Bousfield, Derek, Wichmann, Anne, 2003. Impoliteness revisited: with special reference to dynamic and prosodic aspects. *J. Pragmat.* 35, 1545–1579.
- Du, Bois, John, W., 2007. The stance triangle. In: Englebretson, R. (Ed.), *Stancetaking in Discourse*. Benjamins, Amsterdam, pp. 139–182.
- Dynel, Marta, Poppi, Fabio, 2019. Arcana imperii: the power of humorous retorts to insults on Twitter. *J. Lang. Aggress. Confl.* 1–31. <https://doi.org/10.1075/jlac.00031.dyn>.
- Garcés-Conejos Blitvich, Pilar, Sifanou, Maria, 2019. Im/politeness and discursive pragmatics. *J. Pragmat.* 145, 91–101.
- Garcés-Conejos Blitvich, Pilar, 2018. Relational work in multimodal networked interactions on Facebook. *Inter. Pragmat.* 1 (1), 134–160. <https://doi.org/10.1075/ip.00007.bou>.
- Goffman, Erving, 1967. *Interaction Ritual: Essays on Face to Face Behavior*. Doubleday, New York.
- Goffman, Erving, 1971. *Relations in Public: Micro Relations of the Public Order*. Penguin, Harmondsworth.
- Graham, Sage Lambert, Hardaker, Claire, 2017. (Im)politeness in digital communication. In: Culpeper, J., Haugh, M., Kádár, D. (Eds.), *The Palgrave handbook of linguistic (im)politeness*. Palgrave MacMillan, London, pp. 785–814.
- Gunasekara, Isuru, Nejadgholi, Isar, 2018. A review of standard text-classification practices for multi-label toxicity identification of online content. In: *Proceedings of the Second Workshop on Abusive Language Online*. Association for Computational Linguistics, Brussels, pp. 21–25.
- Hardaker, Claire, 2013. “Uh..... not to be nitpicky, but... the past tense of drag is dragged, not drug”: an overview of trolling strategies. *J. Lang. Aggress. Confl.* 1 (1), 57–85. <https://doi.org/10.1075/jlac.1.1.04har>.
- Haugh, Michael, 2013. Disentangling face, facework and im/politeness. *Sociocult. Pragmat.* 1 (1), 46–73. <https://doi.org/10.1515/soprag-2012-0005>.
- Housely, William, Webb, Helena, Edwards, Adam, Procter, Rob, Jirotko, Marina, 2017. Membership categorization and antagonistic Twitter formulations. *Discourse Commun.* 11 (6), 567–590. <https://doi.org/10.1177/1750481317726932>.
- Khazraie, Marzieh, Talebzadeh, Hossein, 2020. “Wikipedia does NOT tolerate your babbling!”: impoliteness-induced conflict (resolution) in a polylogal collaborative online community of practice. *J. Pragmat.* 163, 46–65.
- Kiesling, Scott F., 2009. *Style as stance*. In: Jaffe, A. (Ed.), *Stance: Sociolinguistic Perspectives*. Oxford University Press, Oxford, pp. 171–194.
- Kiesling, Scott F., Pavalanathan, Umashanthi, Fitzpatrick, Jim, Han, Xiaochuang, Eisenstein, Jacob, 2018. Interactional stancetaking in online forums. *Comput. Ling.* 44 (4), 683–718. <https://doi.org/10.1162/colia00334>.
- Kondabolu, Hari, 2017. *The Problem with Apu* [documentary film].
- Locher, Miriam, 2008. Relational work, politeness and identity construction. In: Antos, G., Ventola, E. (Eds.), *Handbook of Interpersonal Communication*. De Gruyter, Berlin, pp. 509–540.
- Locher, Miriam, Bousfield, Derek, 2008. Introduction: impoliteness and power in language. In: Bousfield, Derek, Locher, Miriam (Eds.), *Impoliteness in Language: Studies on its Interplay with Power in Theory and Practice*. Mouton de Gruyter, Berlin, pp. 1–13.
- Locher, Miriam, Watts, Richard, 2005. Politeness theory and relational work. *J. Politeness Res.* 1 (1), 9–33. <https://doi.org/10.1515/jplr.2005.1.1.9>.
- Locher, Miriam, Watts, Richard, 2008. Relational work and impoliteness: negotiating norms of linguistic behavior. In: Bousfield, D., Locher, M. (Eds.), *Impoliteness in Language: Studies on its Interplay with Power in Theory and Practice*. Mouton de Gruyter, Berlin, pp. 77–99.
- Lorenzo-Dus, Nuria, Blitvich, Garcés-Conejos Pilar, Bou-Franch, Patricia, 2011. Online polylogues and impoliteness: the case of postings sent in response to the Obama Reggaeton YouTube video. *J. Pragmat.* 43 (10), 2578–2593. <https://doi.org/10.1016/j.pragma.2011.03.005>.
- Márquez-Reiter, Rosina, Haugh, Michael, 2019. Denunciation, blame and the moral turn in public life. *Discourse Context Media* 28, 35–42. <https://doi.org/10.1016/j.dcm.2018.09.001>.
- Matley, David, 2018a. “This is NOT a #humblebrag, this is just a #brag”: the pragmatics of self-praise, hashtags and politeness in Instagram posts. *Discourse Context Media* 22, 30–38. <https://doi.org/10.1016/j.dcm.2017.07.007>.
- Matley, David, 2018b. “Let’s see how many of you mother fuckers unfollow me for this”: the pragmatic function of the hashtag #sorrynotsorry in non-apologetic Instagram posts. *J. Pragmat.* 133, 66–78. <https://doi.org/10.1016/j.pragma.2018.06.003>.
- Murthy, Diraj, 2013. *Twitter: Social Communication in the Twitter Age*. Polity, Cambridge.
- Papacharissi, Zizi, de Fatima Oliveira, Maria, 2012. Affective news and networked publics: the rhythms of news storytelling on #Egypt. *J. Commun.* 62, 266–282. <https://doi.org/10.1111/j.1460-2466.2012.01630.x>.
- Parvareh, Vahid, Tayebi, Tamineh, 2018. Impoliteness, aggression, and the moral order. *J. Pragmat.* 132, 91–107. <https://doi.org/10.1016/j.pragma.2018.05.010>.
- Pavlopoulos, John, Sorenson, Jeffrey, Dixon, Lucas, Thain, Nithum, Androutopoulos, Ion, 2020. Toxicity detection: does context really matter?. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 4296–4305. ArXiv: 2006.0098v1.
- Spilioti, Tereza, Tagg, Caroline, 2021. *Research ethics*. In: Vásquez, C. (Ed.), *Research Methods for Digital Discourse Analysis*. Bloomsbury, London (forthcoming).
- Tagg, Caroline, Seargeant, Philip, Brown, Amy A., 2017. *Taking Offence on Social Media: Conviviality and Communication on Facebook*. Springer, Cham.
- Terkourafi, Marina, Catedral, Lydia, Haider, Iftikhar, Karimzad, Farzad, Melgares, Jeriel, Mostacero-Pinilla, Cristina, Nelson, Julie, Weissman, Benjamin, 2018. Uncivil Twitter: a sociopragmatic analysis. *J. Lang. Aggress. Confl.* 6 (1), 26–57. <https://doi.org/10.1075/jlac.00002.ter>.
- Veale, Tony, Feyaerts, Kurt, Bröne, Geert, 2006. The cognitive mechanisms of adversarial humor. *Humor* 19 (3), 305–339.
- Vladimirou, Dimitra, House, Julianne, 2018. Ludic impoliteness and globalization on Twitter: ‘I speak England very best’ #agglia\_Tsipra, #Tsiptas #Clinton. *J. Pragmat.* 134, 149–162. <https://doi.org/10.1016/j.pragma.2018.05.005>.
- Webb, Helena, Jirotko, Marina, Stahl, Bernd Carsten, Housely, William, Edwards, Adam, Williams, Matthew, Procter, Rob, Rana, Omer, Burnap, Pete, 2017. The ethical challenges of publishing Twitter data for research dissemination. In: *WebSci 17, Proceedings of the 2017 ACM on Web Science Conference*, pp. 339–348. <https://doi.org/10.1145/3091478.3091489>.
- Westbrook, Jeff, April 8, 2018. *No Good Deed Goes Unpunished. The Simpsons. Season 29, Episode 15*.
- Wulczyn, Ellery, Thain, Nithum, Dixon, Lucas, 2017. Ex machina: personal attacks seen at scale. In: *International World Wide Web Conference Committee (IW3C2)*, April 3–7, 2017, Perth, Australia. <https://doi.org/10.1145/3038912.3052591>. ArXiv 1610.08194.
- Zhang, Justine, Chang, Jonathan, Danescu-Niculescu-Mizil, Cristian, Dixon, Lucas, Hua, Yiqing, Taraborelli, Dario, Thain, Nithum, 2018. *Conversations Gone Awry: Detecting Early Signs of Conversational Failure*. <https://doi.org/10.18653/v1/P18-1125> arXiv: <https://arxiv.org/pdf/1805.05345.pdf>.

**Camilla Vásquez** is a Professor of Applied Linguistics at the University of South Florida, where she teaches courses in Pragmatics, Discourse Analysis and Sociolinguistics. Her publications include *The Discourse of Online Consumer Reviews* (Bloomsbury, 2014), *Language Creativity and Humor Online* (Routledge, 2019), as well as recent articles in *Discourse Context & Media*, *Intercultural Pragmatics*, *Journal of Sociolinguistics*, *Narrative Inquiry*, among others.