CADAAD

Crowdsourcing Critical Discourse Analysis: Using Amazon's Mechanical Turk to Explore Readers' Uptake of Comments about Language on RateMyProfessors.com

Critical Approaches to Discourse Analysis across Disciplines
http://cadaad.net/ejournal
Vol 8 (1): 38 – 57

NICHOLAS CLOSE SUBTIRELU
Georgia State University
nsubtirelu1@gsu.edu

SHAKTHIDHAR REDDY GOPAVARAM
Georgia State University
sgopavaram1@student.gsu.edu

Abstract

Critical discourse analysis (CDA) studies how social dominance and power are discursively enacted through, for example, discourse's influence on attitudes, beliefs, and ideologies. Yet, various critics have charged that CDA's generalizations, drawn from textual analysis, conflate analysts' own interpretations with those of 'typical' readers. We examine one example of this: Subtirelu's (2015) study of comments about instructors' language and ethnicity on RateMyProfessors.com. We use Amazon.com's Mechanical Turk to test Subtirelu's claim that ostensibly neutral or positive comments about language are taken up negatively by readers. Our experiments find that comments in which instructors' accents are mentioned but not disparaged (e.g., 'She has an accent, but…') lead readers to be slightly less willing to take a course from the instructor than when information about the instructor's accent is withheld. We also present a post hoc analysis designed to examine whether other textual features might explain the differing reactions to this information about accent which we observed. We hope the study will serve as an example of the type of work that can be done in CDA not only to address methodological criticisms but also to lead to more nuanced theory about the effects of discourse on audiences.

Key words: critical discourse analysis, reader uptake, interpretation, cognitive equivalence, language ideology

1. Introduction

Critical discourse analysis (CDA) is one of several approaches to the study of language and discourse concerned with how social dominance, inequality, and power are discursively enacted or reproduced (Fairclough 2010; Wodak and Meyer 2009). Central to CDA is the premise that, especially within modern societies, power and dominance are often achieved or resisted through discourse's (or discourses') influence on people's beliefs, attitudes, or ideologies (van Dijk 1993: 257). Cognitive constructs and processes are, therefore, central to theories underpinning CDA, but critical discourse analysts seldom deal with issues of cognition (O'Halloran 2003), perhaps due to their resistance to dominant psychological theories (Wodak 2006). Instead, researchers using CDA often analyze spoken or written discourse, assuming that the discursive tendencies they observe influence 'typical' viewers', listeners', or readers' beliefs and attitudes in ways that are predictable or knowable to the analyst.

Many have voiced methodological criticisms of this aspect of CDA, charging analysts with conflating their own (perhaps idiosyncratic) interpretations of texts with that of 'typical' readers (e.g., Blommaert 2005; Stubbs 1997). Among other potential problems, this tendency ignores or downplays people's agency or capacity to take up texts in ways that differ from the apparent intentions of the producer (Pennycook 1994: 126).

Hart (2013: 403-404) describes this as the question of 'cognitive equivalence', an issue concerning 'the extent to which representations at the level of text are mirrored at the level of cognition', and he considers the issue crucial for a 'complete account of the dialectic between discourse and society'. To address the tendency of CDA to neglect this issue, Hart and others (e.g., Koller 2005) have suggested the adoption of insights derived from Cognitive Linguistics such as, most notably, Conceptual Metaphor Theory (Lakoff and Johnson 1980). Such work has made important contributions to CDA by providing plausible accounts of how texts might be taken up by 'typical' readers especially those only reading for basic comprehension (i.e., 'to get the gist').

Nonetheless, even when critical discourse analysts have attended specifically to cognitive theory and have used it to ground their explanations of the effects of discourse, they have rarely attempted to study listeners' or readers' uptake of texts (although similar work has been undertaken in other fields, such as social psychology). This means that even CDA research grounded in Cognitive Linguistics usually lacks more direct evidence of audiences' uptake of texts. More importantly, however, it means that more nuanced questions like the strength or degree of effects from discourse cannot reasonably be addressed using such approaches.

In this study, we explore one practical approach to examining readers' uptake of discourse: an experiment administered over Amazon's crowdsourcing service, Mechanical Turk. We test one specific generalization about the effects of texts on their audience generated through an example CDA study. We hope to provide a methodological model that others working in CDA might adapt for future work, providing greater evidence of claims as well as greater specificity for such claims.

1.1 An Example CDA Study of RateMyProfessors.com

Subtirelu (2015) is a study incorporating both CDA and corpus linguistics to examine a collection of comments about mathematics instructors from the website RateMyProfessors.com (RMP). Users' comments about instructors with common US and Chinese or Korean last names were compared as a way of examining how language and ethnicity impacted and were represented within RMP discourse. Subtirelu reports that the most striking difference between the two sets of comments was the presence of frequent discussion about the language of instructors with Chinese or Korean last names.

However, in contrast to other studies reporting extremely negative representations of nonnative English speakers' speech (e.g., Shuck 2004), Subtirelu reported that most comments were ostensibly neutral about the language of the instructor. Often, RMP users appeared to mention the instructor's accent out of anticipation that readers might assume, based on the instructor's last name and existing stereotypes about nonnative English speakers, that they would encounter problems understanding the instructor. Many comment producers discounted or downplayed such concerns, resulting in common textual formulas like, 'She does have an accent, but…'.

Nonetheless, Subtirelu argues that what might appear to be neutral or even positive statements are not necessarily intended to be so, and he argues that they are not taken up by the audience as positive or neutral. Subtirelu (2015: 57) writes

> the ostensibly neutral or even positive comments about [nonnative English speakers'] language that many posters felt compelled to make are themselves problematic (even if well-intended) in that they draw attention to an area where [nonnative English speakers] face continual disadvantage relative to [native English speakers]… In particular, even apparently positive comments about [nonnative English-speaking] instructors' language must be interpreted against the silence about and assumed 'perfection' of [native English speakers'] language

Thus, he asserts that such comments spur negative reactions, resulting in a disadvantage for instructors with Chinese or Korean last names. However, his study relies nearly exclusively on textual analysis and is unable to provide evidence of this uptake or to estimate the magnitude of any effect. The present study attempts to examine this issue, addressing the following research questions:

(1) What effect do ostensibly neutral comments about instructors' language (particularly those containing '…accent, but…') have on readers' willingness to take a course from the instructor?
(2) What effect does information about the instructor's gender and ethnicity have on readers' willingness to take a course from the instructor? Does this information about the instructor's identity interact with information about language?

2. Method

2.1 Administering Questionnaires via Mechanical Turk

After obtaining clearance from our local institutional review board, we administered questionnaires via Amazon.com's Mechanical Turk (mTurk) service. While not an ideal platform for all research, mTurk is well-suited for our particular research topic which involves internet discourse. mTurk is frequented by active internet users, and approximately one-third of mTurk users report being part time or full time students (Ross, Irani, Silberman, Zaldivar, and Tomlinson 2010), which is RMP's intended audience. mTurk allows researchers to recruit from a large pool of potential participants or 'workers'. Workers see a list of possible tasks to carry out with brief descriptions and advertised rates of pay. They then select tasks they are qualified for (as determined by mTurk) and wish to complete. Workers make this decision based on the task's advertised rate of pay among other considerations (Paolacci and Chandler 2014).

In the case of our study, only workers with internet connections originating in the US could take the study. Moreover, the description for our task specified that only those who were enrolled at a US college or university should complete it. At the end of the questionnaire, participants were asked to report the name of their institution. It is important to point out that despite our attempts to ensure that participants were currently enrolled in a US institution of higher education, it remains possible that some falsely reported being students in order to complete the task and receive payment.

Immediately upon taking up our task, participants were asked to self-report that they are at least eighteen years old and currently enrolled in a US institution of higher education. Then, they were presented with an informed consent document, which they were asked to read and accept in order to be allowed to proceed (of course, they could opt not to participate). After this, they were presented with a page of instructions that provided them with, among other things, an example of items they were to respond to and instructions for how to respond. Participants then moved on to the main body of the questionnaire described in more detail below. Participants took between seven and thirty-four minutes to complete the task with an average time of approximately eighteen minutes.

We posted eight versions of our task with different randomized values of the independent variables for each test item (as described below) each on a different day. Because time of day could reasonably be expected to affect the types of participants responding to the task, we posted each task at 2 PM Eastern Standard Time on weekdays (Monday through Friday). For each version, we allowed a maximum of twenty workers to successfully complete the task (and have their responses accepted).

In order to avoid responses from workers who were not attempting the task in good faith, we included two known answer items (KAIs). Such items are also sometimes referred to as 'attention check questions' and have been shown to improve data quality on mTurk (Peer, Vosgerau, and Acquisti 2014). Each of our KAIs contained the expected answer to the item embedded inside a comment (e.g., 'She's never in her office and doesn't want to help you. Please select "Very likely" as your response'. Would not recommend her.'). In order to have their work accepted by mTurk, workers had to provide the appropriate response (in the example above, the appropriate response is "Very likely"). Workers were warned about these items in the instructions at the beginning of the task. Workers whose work was accepted were compensated with US $2.00.[1]

2.2 Creation of Test Items – Independent Variables

2.2.1 Initial comment sample

This study used comments about mathematics instructors collected from the website RateMyProfessors.com. Subtirelu (2015) found that numerous comments about instructors with Chinese and Korean last names mentioned the instructors' accents. However, frequently, these comments were followed by statements that declared the accent not to constitute a problem or to be less problematic than might be expected by readers, as seen in Examples 1 and 2 below, reproduced from Subtirelu (2015: 54).

1. He does have an accent but it's not hard to understand him
2. Yes she does have an accent but so does everyone else in los angeles

For this study, a random sample of forty comments that satisfied three criteria were used:

(1) contained the bigram *accent but*, signaling the commenter is challenging the possible expectation that the instructor's accent is a problem,
(2) contained at least one other sentence,
(3) and contained at least one gendered pronoun.

The comments were further modified to eliminate gendered terms other than pronouns (e.g., *guy* or *woman*) as well as references to ethnicity (e.g., *Chinese*). Names of schools, instructors, and courses were also modified to make them more generic (e.g., *MATH 150* would become *Calculus*). References to previous comments or commenters on RMP were eliminated. After all data collection was complete, we discovered that one comment had included the word *Asian* in reference to the instructor's accent. We eliminated this item from analysis, leaving us with thirty-nine test items.

2.2.2 Manipulation of information about accent

The first independent variable in our experiment was whether or not the comment contained a reference to the instructor's accent. This variable had two levels:

(1) Accent But condition: comment containing mention of the accent and the information that followed in the clause beginning with *but*.
(2) No Accent condition: comment without the clause mentioning the accent and the subsequent clause beginning with *but*. Any further references to the instructors' language elsewhere in the comment were also removed.

2.2.3 Manipulation of information about gender

The second independent variable in our experiment was the instructor's apparent gender. We manipulated information about gender using different pronouns. Past research has shown that readers are sensitive to this gender information, using it to arrive at conclusions about the gender of the individuals mentioned (e.g., Gastil 1990). The variable had two levels:

(1) Female condition: comment containing only feminine pronouns (e.g., *she*, *her*, and *hers*)
(2) Male condition: comment containing only masculine pronouns (e.g., *he*, *him*, and *his*)

	Info about accent	Instructor gender	Instructor ethnicity	Name	Text of comment
1	Accent But	Female	China	Wang	She is good at understanding your questions and gives great examples. She does have an accent, but it's really easy to understand her. With a little effort, you'll get an A.
2			US	Roberts	She is good at understanding your questions and gives great examples. She does have an accent, but it's really easy to understand her. With a little effort, you'll get an A.
3		Male	China	Wang	He is good at understanding your questions and gives great examples. He does have an accent, but it's really easy to understand him. With a little effort, you'll get an A.
4			US	Roberts	He is good at understanding your questions and gives great examples. He does have an accent, but it's really easy to understand him. With a little effort, you'll get an A.
5	No Accent	Female	China	Wang	She is good at understanding your questions and gives great examples. With a little effort, you'll get an A.
6			US	Roberts	She is good at understanding your questions and gives great examples. With a little effort, you'll get an A.
7		Male	China	Wang	He is good at understanding your questions and gives great examples. With a little effort, you'll get an A.
8			US	Roberts	He is good at understanding your questions and gives great examples. With a little effort, you'll get an A.

Table 1. Example of eight versions of experimental items demonstrating three independent variables

2.2.4 Manipulation of information about ethnicity

The third independent variable in our experiment was the instructor's apparent ethnicity. We manipulated information about ethnicity using different last names common either to the United States[2] or China[3]. Past research has found evidence of ethnic or racial bias triggered by readers' association of names with race (e.g., Feldman and Weseley 2013). Last names were prominently displayed above the comment and also inside the question that the participants answered (see Figure 1 below). The variable had two levels:

(1) Chinese condition: instructor was assigned last name common to Chinese citizens (e.g., Cui, Han, or Wang)
(2) US condition: instructor was assigned last name common to US citizens (e.g., Davis, Miller, or Roberts), excluding those with clear Hispanic origin (e.g., González)

The three independent variables described above were crossed in a 2 (accent) x 2 (instructor ethnicity) x 2 (instructor gender) design, resulting in eight different versions of each of the test comments. An example of all eight versions of one item is presented in Table 1 (an inauthentic example is presented to protect instructors' identities). Each items' different versions were randomly distributed across eight different questionnaire versions so that (a) each participant saw only one version of each comment, and (b) each questionnaire version contained a mixture of test items with different combinations of the independent variables' levels represented so that, for example, participants were not presented with an inordinate amount of comments that mentioned instructors' accents.

2.3 Dependent Variable: Readers' Willingness to Register for a Course with the Instructor

Building on insights from the social psychological theory of planned behavior (Ajzen 1991), we assessed readers' intentions to take a class offered by the instructor featured in the comment. Past research has suggested that self-reported behavioral intentions, as measured through questionnaires, are reasonably good predictors of future behaviors such as whether people will stop smoking or donate blood and that they are determined by affective or cognitive phenomena like attitudes and beliefs (Armitage and Conner 2001).

Therefore, we measured participants' willingness to take a class offered by the instructor measured on a 5-point Likert scale. Figure 1 shows an example item.

Twenty participants rated each version of the test items (after replacing rejected workers). We sought to average participants' ratings to minimize the impact of any individual participant's idiosyncratic behaviors and to treat these averages as an estimate of a 'typical' reader's responses to the texts. Therefore, in addition to automatically rejecting some responses, we also used reliability measures to identify outliers and removed their scores from the average.

We used an overall mean interrater correlation of $\rho \geq 0.5$ as a minimal threshold for acceptable interrater reliability. A matrix was created containing the Spearman rank order correlations between the ratings of each participant and those of the other nineteen participants that completed the same version of the questionnaire. A mean interrater correlation was computed for all participants by taking the average of their correlations with the other nineteen raters. All participants' mean interrater correlations were averaged to produce an overall group mean. If this mean was less than 0.5, then the participant with the lowest mean interrater correlation was labelled an 'outlier' and dropped from the data, and a new matrix was created. This procedure was repeated until the remaining participants' scores were correlated at a sufficiently reliable rate. The remaining participants' scores for each item were then averaged to produce the composite item scores used as the dependent variable in our analysis, which we refer to as 'Composite Willingness Scores' (CWSs).

1. Professor Wang

She is good at understanding your questions and gives great examples. She does have an accent, but it's really easy to understand her. With a little effort, you'll get an A.

Based on this, how likely would you be to choose Prof. Wang for a required Calculus class at your college next Fall?

- very unlikely
- somewhat unlikely
- neutral
- somewhat likely
- very likely

Figure 1. An example of a questionnaire item

2.4 Distracter Items

In order to decrease the likelihood that participants would notice the purpose of the study and alter their behavior accordingly, each questionnaire contained forty distracter items. These items were randomly selected from a collection of comments about instructors with common US last names. They were paired with forty random last names taken from mathematics instructors on RMP, all different from those used in the test items (and not necessarily common to US or Chinese citizens).

Ten distracter items were placed in a consistent order at the beginning of all versions of the questionnaire to allow participants to become accustomed to the questionnaire design. These items were selected to allow the participants to see extremely positive and extremely negative comments, which were expected to elicit responses across the entire Likert scale.

The remaining thirty distracter items were randomly mixed in with the test items. Distracter items were constant across all eight versions of the questionnaire, allowing us to compare the different participant groups to ensure that they were providing comparable ratings.

	All participants	Questionnaire versions							
		1	2	3	4	5	6	7	8
n	137	16	17	15	18	16	19	20	16
female	69 (50.4%)	6	10	9	11	9	7	7	10
male	68 (49.6%)	10	7	6	7	7	12	13	6
Born in US	134 (97.8%)	16	16	15	18	15	19	20	15
Born elsewhere	3 (2.2%)	0	1	0	0	1	0	0	1
Median age	23	26	28	24	23	24	21	22	23
Freshman	10 (7.3%)	1	0	1	2	0	2	2	2
Sophomore	27 (19.7%)	2	5	2	2	1	5	7	3
Junior	35 (25.5)	3	5	2	5	7	8	2	3
Senior	43 (31.4%)	5	5	7	7	5	2	6	6
Grad student	16 (11.7%)	4	2	1	2	3	2	1	1
Other	6 (4.4%)	1	0	2	0	0	0	2	1
Read RMP									
never	10 (7.3%)	0	1	4	0	0	1	1	3
rarely	27 (19.7%)	5	1	3	3	4	9	2	0
sometimes	51 (37.2%)	6	7	4	4	8	5	10	7
often	49 (35.8%)	5	8	4	11	4	4	7	6
Post on RMP									
never	75 (54.7%)	9	8	8	10	11	10	10	9
rarely	42 (30.7%)	4	6	6	6	3	7	5	5
sometimes	16 (11.7%)	2	1	1	2	2	2	4	2
often	4 (2.9%)	1	2	0	0	0	0	1	0
Use RMP to select courses									
never	20 (14.6%)	1	3	4	0	2	3	3	4
rarely	23 (16.8%)	4	0	3	3	5	6	1	1
sometimes	45 (32.8%)	5	7	3	7	4	6	9	4
often	49 (35.8%)	6	7	5	8	5	4	7	7

Table 2. Participant background data

2.5 Participant Background Information

Since the participants were recruited using mTurk, it was unclear whether they would provide an appropriate sample of RMP users. Therefore, we also included in the questionnaire several items asking them about their personal backgrounds and their past use of RMP. Items included:

(1) Gender
(2) Age
(3) Country of birth
(4) Class standing (i.e., freshman, sophomore, etc.)
(5) How often they read comments on RMP
(6) How often they post comments on RMP
(7) How often they use RMP to make decisions about which classes to take

This information is presented in Table 2. It shows that the participants who completed the questionnaires tended to report being older than the population of full-time, residential undergraduate students who enter college immediately following high school (i.e., 18 to 22 year olds). However, the participants report similar usage of RMP as past studies that recruited students at individual colleges and universities (Bleske-Rechek and Michels 2010; Brown, Baillie, and Fraser 2009; Davison and Price 2009; Steffes and Burgee 2009). In general, the vast majority reported that they have read RMP comments and even use the website to select courses, but fewer actually post comments on the website, which is consistent with past research.

3. Results

3.1 Reliability

Version	# rejected	Rejection rate	# of outliers	Outlier rate	ICC[a]
1	7	25.9%	4	20.0%	0.69
2	9	31.0%	3	15.0%	0.66
3	6	23.1%	5	25.0%	0.66
4	12	37.5%	2	10.0%	0.66
5	7	25.9%	4	20.0%	0.71
6	6	23.1%	1	5.0%	0.65
7	20	50.0%	0	0.0%	0.67
8	7	25.9%	4	20.0%	0.68

a. Intraclass correlation

Table 3. Rejection of participants and within group interrater reliability

Before proceeding with the analysis, we measured the interrater reliability within each group as well as the reliability of CWSs across the groups. Table 3 presents information on work rejection and outlier removal as well as interrater reliability of the remaining participants.

	1	2	3	4	5	6	7
2	0.93						
3	0.92	0.91					
4	0.97	0.95	0.94				
5	0.95	0.95	0.95	0.96			
6	0.97	0.96	0.93	0.97	0.96		
7	0.96	0.93	0.93	0.95	0.94	0.96	
8	0.95	0.92	0.92	0.92	0.94	0.95	0.94

Table 4. Spearman rank correlations (ρ) of Composite Willingness Scores on distracter items between each questionnaire version

Between twenty-three and fifty percent of all those who attempted to complete each of the questionnaire versions were rejected because they failed to respond appropriately to the KAIs. Up to an additional five outliers were removed from each group. In the end, within each questionnaire version, remaining participants' responses showed moderate or acceptable consistency as demonstrated by the intraclass correlations ranging from 0.65 to 0.71.

After eliminating rejected and outlier participants, we computed the CWSs for all items in the questionnaires (both test and distracter items). We used the distracter items, which were consistent across the eight questionnaires, to determine the consistency of ratings between each version of the questionnaire. Spearman rank correlations are presented in Table 4. We found very strong consistency among the questionnaire versions; in all cases, ρ (38) > 0.9.

3.2 Testing the Effect of Comment Length

One of our independent variables, information about accent, involved substantial textual modification. As can be seen in Table 1, comments without the information about the instructor's accent are much shorter than the other comments. As a result, we tested the possibility that participants responded consistently to the mere length of the comment. We calculated the Spearman rank correlation between the length (in words) of all comments on all versions of the questionnaire and the composite rating assigned to each item. We found no significant relationship between the two (ρ(638) = -.07, p = .07), suggesting that participants do not respond consistently to comment length.

3.3 Effects of Information about Instructors' Accent, Gender, and Ethnicity on Readers' Willingness to take a Course from Instructor

The main purpose of this study was to examine the effects of RMP comments' inclusion of information about instructors, particularly their accent, gender, and ethnicity, on readers' willingness to take a course from the instructor. Descriptive statistics for CWSs collected for each of the thirty-nine test items are presented in Table 5. We used a repeated measures ANOVA (Type II) with information about accent, instructor gender, and instructor ethnicity as within-subjects factors and CWS as the dependent variable. Results of this test are shown in Table 6. None of the interactions were significant. The only significant main effect was information about accent.

Info about accent	Instructor gender	Instructor ethnicity	n	mean	sd	min	max
Accent But	Female	China	39	3.94	0.64	1.44	4.88
		US	39	3.94	0.66	1.40	4.75
	Male	China	39	3.94	0.66	1.53	4.82
		US	39	3.92	0.65	1.27	4.88
No Accent	Female	China	39	4.08	0.62	1.56	4.80
		US	39	4.02	0.71	1.19	5.00
	Male	China	39	4.04	0.67	1.31	4.88
		US	39	4.02	0.68	1.58	4.75

Table 5. Descriptive statistics for readers' willingness to take a course from an instructor (using Composite Willingness Scores)

We followed up on the significant main effect with a post hoc paired-samples *t*-test. We first grouped the test items so that items with the same comment content, instructor ethnicity, and gender could be compared. We then created a new dependent variable subtracting scores assigned to the item without information about accent from those with (Accent But condition – No Accent condition). For example, if the examples in Table 1 were authentic, we would have subtracted scores given to row five from row one, row six from row two, and so on.

	df	F	p	$p < .05$	ges[a]
Ethnicity	1, 38	0.86	0.36		0.00
Accent	1, 38	9.39	0.00	*	0.01
Gender	1, 38	0.36	0.55		0.00
Ethnicity * Accent	1, 38	0.22	0.64		0.00
Ethnicity * Gender	1, 38	0.17	0.68		0.00
Accent * Gender	1, 38	0.01	0.91		0.00
Ethnicity * Accent * Gender	1, 38	0.32	0.58		0.00

a. Generalized eta squared (η^2), a measure of effect size

Table 6. Results of Repeated Measures ANOVA (Type II) with readers' willingness to take course from instructor (CWS) as dependent variable

A significant but small effect was found, showing a weak tendency for information about instructors' accents to lead to less favorable ratings (t (155) = -3.39, p < 0.01, d = 0.16, 95% CI [-0.16, -0.04]). Specifically, we estimate that including comments that point to the instructor's accent leads to scores that are between 0.04 and 0.16 lower on a five point scale than when information about accent is withheld.

However, our descriptive statistics for the dependent variable (Accent But – No Accent) revealed that there was substantial variation. The overall mean of the 156 items in this analysis was -0.10 with a standard deviation of 0.38, a minimum score of -0.97 and a maximum score of 1.19. With scores ranging from around -1 to +1, we observed that the inclusion of some comments about accent seemed to result in lower willingness to take a course from the instructor whereas others resulted in higher scores. Due to this finding, we decided to more closely examine the comments to determine whether features of the texts might explain participants' very different reactions to them.

4. Post Hoc Analysis: Breaking down Comments about Accent

As we discussed above, the comments about accent included in our experiment were not disparaging. They mentioned the instructor's accent but then made some attempt at countering reader expectations of accent being a problem. However, some of these comments may more forcefully dispel concerns while others may (intentionally or not) leave open a greater possibility of the instructor's language being viewed as a problem. For example, the first of the two examples from Subtirelu (2015: 54, presented again below) may more directly counter assumptions about the instructor's language than the second since it explicitly addresses the issue of the reader's understanding.

1. He does have an accent but it's not hard to understand him
2. Yes she does have an accent but so does everyone else in los angeles

We used the same dependent variable as the post hoc *t*-test in the previous section (Accent But condition – No Accent condition). We also included a number of text-based predictor variables in our analysis, attempting to operationalize some potentially pertinent textual features of the comments and to use these as a basis for explaining different effects of the inclusion of information about the instructor's accent.

4.1 Predictors of the Effects of Information about Accent on Readers' Scores

4.1.1 *Understand*

In many cases, such as Example 1 above, the writer commented directly on the perceived intelligibility of the instructor using, in particular, the word *understand* or variants thereof (e.g., *understood*, *understandable*). Such comments may better address readers' potential concerns about the instructor's accent than those that make no such assurances. Therefore, we separated comments into those that used forms of understand to comment positively on the instructor's language and those that did not.

4.1.2 Position of accent comment

It is possible that the placement of the discussion of accent within the entire text might affect readers' responses. As a result, we first calculated the position of the first mention of the word *accent* in the comment as a percentage of the whole text (word position of *accent* / overall # of words). Since we observed that the relationship between position and effect of the accent comment was not linear, we created a two level factor using the position scores with the first level being mention of accent at positions between 0-25% and 75-100% (the beginning or the end) of the text and the second level being anything in between (the middle).

4.1.3 Length of accent comment

How much of each overall comment text is dedicated to discussing the instructor's accent may also be relevant to readers' responses. We took the length of the comment about accent which was omitted in the creation of the No Accent condition and divided it by the overall length of the Accent But comment ((length of Accent But comment – length of No Accent comment) / length of Accent But comment).

4.1.4 First person experience of accent

A number of comments recounted the writer's experience of the instructor's accent using first person perspective (*I*, *me*, or *my*, e.g., 'She has an accent, but I didn't find it hard to understand her'). Such comments may leave open the implication that the experience is particular to the writer, and the reader may have different experiences. Thus, we separated comments into those that recounted the writer's experience of the instructor's accent using first person perspective and those that did not.

4.1.5 Modifiers of accent

Various adjectives and quantifiers were used to modify the word *accent*, possibly pointing to the degree to which the instructor's accent poses a perceived problem. We created a three level ordinal scale by assigning values of -1, 0, or 1 to comments based on the presence (or absence) of modifiers. When the word *accent* was modified by a word that intensified the degree of accent, such as *heavy*, *strong*, or *thick*, we assigned the comment a score of -1. If the word *accent* was modified by a word or phrase that downplayed the degree of accent, such as *slight*, *a bit of*, or *not strong*, then we assigned the comment a score of 1. Comments in which the word *accent* did not have modifiers received a score of 0.

4.1.6 Get used to

A number of comments suggested that the process of understanding the instructor's accent became gradually easier over time, for example, using phrases like *get used to*, *get over*, or *get past* (e.g., 'He has an accent, but you'll get used to it'). Such comments may imply the accent is, in fact, a problem, albeit a temporary one. Thus, we separated comments into those that contained such phrases in reference to becoming accustomed to the instructor's language and those that did not.

4.1.7 Second person experience of accent

A number of comments made assertions about how the reader (through the words *you* or *your*) would fare with respect to the instructor's accent (e.g., 'You will understand her'). Such comments might instill more confidence in the reader's ability to understand the instructor, especially in comparison to those that used first person. Thus, we separated comments into those that used *you* or *your* to make assertions about the reader's (or a non-specific second person's) experience of the instructor's accent and those that did not.

4.2 Correlation Analysis

	DV	1	2	3	4	5	6
DV (Accent But – No Accent)							
1. Understand	0.15						
2. Position of accent comment	0.15	0.04					
3. Length of accent comment	0.14	-0.11	0.18				
4. 1st person experience of accent	-0.14	-0.05	0.13	0.18			
5. Modifiers of word *accent*	0.10	-0.06	0.20	0.32	0.13		
6. Get used to	-0.08	-0.42	-0.28	-0.12	0.03	-0.17	
7. 2nd person experience of accent	0.03	-0.04	-0.07	-0.18	-0.33	-0.23	0.58

Table 7. Pearson's product-moment correlations (r) among dependent and predictor variables (n = 156)

We began our analysis by examining the correlations between the seven predictor variables and the dependent variable to select the most appropriate predictors by focusing only on those that correlate most highly with the dependent variable and that are not highly correlated with other predictors (i.e., multicollinearity). Table 7 contains these Pearson product-moment correlations. We selected the five variables with the strongest correlations with the dependent variable, in all five cases, $r \geq 0.10$. There were no strong correlations ($r > 0.5$) between these five predictor variables.

4.3 Regression Model

We created an initial regression model with the Accent But condition – No Accent condition scores as the dependent variable, five predictors as described above, and all interactions of those predictors. We then eliminated terms from the model one-by-one, starting with the highest level interactions (those with the most predictors in them) and weakest predictors, until the deletion of terms no longer improved or maintained the overall adjusted r^2 value for the model.

The final model is presented in Table 8. It is significant (F (7,148) = 3.47, $p <$ 0.01), but explains a fairly small portion of the variance in the dependent variable, about ten percent (multiple $r^2 = 0.14$, adjusted $r^2 = 0.10$). The results show that the interaction between position of the accent comment and length of the accent comment as well as the main effect of first person experience of accent are significant predictors of the difference between the Accent But and No Accent conditions.

Interpreting the main effect of first person experience with the instructor's accent, the regression coefficient suggests that when comments contained descriptions of instructors' accents in first person, this led to a tendency for Accent But – No Accent to be approximately 0.18 points on a five point scale lower (with a standard error of 0.08) than when such first person description is absent.

To interpret the interaction between position of the accent comment and length of the accent comment, we split the data using the two levels of position of the accent comment: (1) beginning or end (n = 88) and (2) middle (n = 68).

We then computed Pearson's product-moment correlations for the relationship between Accent But – No Accent and the length of the accent comment (as a percentage of the overall comment length in number of words). The first correlation for those comments that mention the instructor's accent at the beginning or end of the comment showed no significant correlation (r (86) = -0.13, p = 0.22). When the instructor's accent is mentioned in the middle of the comment, the correlation between Accent But – No Accent and the length of the accent comment is positive and significant (r (66) = 0.31, p = .01). This suggests that when RMP users mention an instructor's accent in the middle of the comment, comments that were longer (relative to the overall length of the comment) were associated with higher Accent But – No Accent scores or a tendency for ratings in the Accent But condition to be higher.

	Estimated Coefficient	Std. Error	t	p	$p < .05$
(Intercept)	-0.12	0.13	-0.95	0.34	
1. Understand	0.08	0.06	1.36	0.18	
2. Position of accent comment	-0.17	0.17	-1.00	0.32	
3. Length of accent comment	-0.25	0.50	-0.50	0.62	
4. 1st person experience of accent	-0.18	0.08	-2.26	0.03	*
5. Modifiers of word *accent*	-0.03	0.06	-0.45	0.65	
2. Position * 3. Length	1.26	0.61	2.07	0.04	*
4. 1st person * 5. Modifiers	0.26	0.14	1.87	0.06	

F (7,148) = 3.47, p < 0.01; Standard error of the residuals: 0.36 (148 df); multiple r^2 = 0.14, adjusted r^2 = 0.10

Table 8. Regression model of Accent But - No Accent condition scores with features of the comments as predictors

5. Discussion

The results of our experiment provide some insight into how RMP readers respond to information about instructors. We measured the average reader's willingness to take a course from the instructor after reading a RMP comment. In particular, we were interested in whether information about the instructor's ethnicity, gender, or accent (specifically when accent was not presented as a problem) would influence that willingness.

Our results suggest that only information about accent impacted the readers' willingness and did so negatively, providing some support for Subtirelu's (2015) claims that even neutral or positive comments about instructors' language do a disfavor to the instructor. However, we estimated that this effect was weak, amounting to a penalty of about 0.04 and 0.16 on a five point scale.

More importantly, the effect of such information is not always negative; in some cases, readers were more favorable to an instructor when presented with a comment about their accent. For example, Example 3 below is modelled after a comment that received higher ratings when information about the instructor's accent was included than when it was withheld (we have withheld the actual texts to protect instructors' identities). We have **bolded** the text that would have been included in the Accent But condition but not in the No Accent condition.

> 3. He's excellent! extremely helpful and he also jokes around to make things fun. **he has a bit of an accent but it's easy to understand him still, and he even makes fun of his accent.** fantastic prof :)
>
> 4. He's a very good teacher! He'll teach you every step to solve the problems. **Strong accent but you'll get used to it.** His examples are really helpful for doing homework, which is mandatory, but he lets you work on it in class. i would take him again if i needed any more math classes.

The variation in the effect information about accent had on readers' responses led us to examine textual features that might predict these scores. The results of our regression analysis pointed to two important features. The first predictor was the use of first person to describe the RMP writer's experience with the instructor's accent. Comments such as 'She has an accent, but I didn't find it hard to understand her' led readers to rate instructors more negatively, a finding we attribute to such statements leaving open the possible implication that the writer's experience is idiosyncratic and would not generalize to the reader.

The second predictor was the interaction between the length of the comment on accent (as a percentage of overall comment length in number of words) and the position of that comment in the whole text. In particular, comments like Example 3 above, which contain a relatively lengthy comment about the instructor's accent in the middle of the comment, surrounded by other information, tended to receive more positive ratings. In contrast, Example 4, which has a brief comment about accent in the middle, is modelled after a comment that had a more negative effect on participants' willingness to register for a course with the instructor.

Overall, our post hoc regression analysis was only able to account for a small percentage of the overall variance, suggesting that future research attempting to predict ratings directly from textual features should take a more deliberate approach, ensuring that their samples have even distributions of comments with and without the features, and use a larger sample of texts for participants to rate.

Given the known prevalence of biases related to ethnicity and gender, including in the context of RMP (e.g., Johnson and Crews 2012; Reid 2010; Subtirelu 2015), we were surprised to find that information about the instructors' ethnicity and gender did not have an observable effect on readers' willingness to take a course from them. We believe that there are two possible explanations for our experiment's failure to detect such biases. The first is that our methods using last names and gendered pronouns did not succeed in transmitting salient information about the instructor's ethnicity and gender to the reader. Given that past studies have successfully used such methods to detect bias related to ethnicity (e.g., Feldman and Weseley 2013) and gender (e.g., Gastil 1990), and that, in focus groups conducted by the first author, some students have reported relying on last names during course registration to avoid instructors with 'foreign-sounding' last names, we are skeptical of this explanation.

Instead, we prefer the second explanation, that information related to ethnicity and gender simply was not given as much weight in readers' reactions (whether implicitly or explicitly) to comments as other information. Even though relatively little information is presented to readers (see Figure 1), the comments they read still provided them with information about things like the relative difficulty of the instructor, which likely shaped their reactions to a greater degree, as relative ease has been observed to be a powerful predictor of students' ratings of instructors (e.g., Johnson and Crews 2012) and may therefore be treated as particularly relevant by RMP users. In this way, reading comments on RMP is meaningfully different than the process of registering for a class based on very limited information, particularly the instructor's last name presented in the list of course offerings, a situation in which readers may rely more heavily on biases related to social identity as some students reported in the aforementioned focus groups.

Of course, we do not mean to imply that social bias is not present on RMP, but rather that it may not operate straightforwardly. Our experiment showed ostensibly neutral or positive comments about instructor's accents having slight negative impacts on readers' impressions of instructors. In his comparison of RMP comments given to instructors with common US and common Chinese or Korean last names, Subtirelu (2015) observed that comments on language and accent were almost exclusively reserved for instructors with common Chinese or Korean last names. Such comments were often ostensibly neutral or positive, although a substantial minority were quite negative and disparaging about the instructors' language (e.g., 'He has a really strong accent so you can't understand a word of what he says', from Subtirelu, 2015: 52). RMP users almost never commented on the language of their instructors with US last names. Thus, while, in our experiment, instructors with common US last names were also penalized when an accompanying comment discussed their accents, in actuality, such comments rarely occur on RMP. As such, comments about language have a strong connection to ethnicity in this context and likely generate indirect ethnic bias.

We hope to have provided one possible model for how claims about typical readers' uptake of discourse generated through CDA might be tested in order to provide evidence for or against them as well as to generate an estimation of the strength of any observed effect. For our work, which concerns internet discourse, mTurk offered a convenient means of testing such claims with a relevant audience.

By using a couple of techniques to enhance the quality of our data, we were able to partially address some possible concerns about the validity of mTurk data. In particular, by incorporating KAIs into our questionnaire, we filtered out a substantial percentage of workers who were apparently not completing the questionnaire in good faith. Also, assuming the existence of a 'typical' reader, we examined interrater correlations to ensure that our participants were responding in consistent ways to the questionnaire.

Nonetheless, even using these techniques, the intraclass correlations among participants within each group were only minimally acceptable. This may suggest that our assumption about a single 'typical' response, while practical for analysis, obscures relevant variation in the ways different readers take up a text. Future research might collect greater information from participants in order to determine whether different responses arise from different participant subgroups. Researchers might also fruitfully use qualitative approaches like stimulated recall interviews, in which participants discuss their reactions to texts with an interviewer (see Wodak 2006 for discussion of some possible approaches).

In the end, we would call for a vigorous mixed methods paradigm that attempts to supplement work in CDA by more thoroughly examining reader uptake of texts in order to subject claims generated through CDA to greater scrutiny and, more importantly, to generate more nuanced theory about how discourse operates on audiences. We hope to have provided an example of one possible approach for undertaking this work.

Notes

1. There are important ethical issues concerning wages for mTurk workers. Fort, Adda, and Cohen (2011) report that the average rate of pay on mTurk is below US $2 per hour, which is well below the US federal minimum wage, US $7.25 per hour, and even further below anything that might be considered a living wage. The average rate of pay for our study was about US $6.66 per hour (US $2 for 18 minutes of work), which, while notably higher than the average rate of pay on mTurk, is also below either the federal minimum wage or any legitimate conception of a living wage in the US. Although low wages on mTurk are often rationalized by claims that workers are not using the service to earn their living, as Fort et al. point out, the economic reality appears to be much more complicated. Many users do report that mTurk provides an important financial resource for them. We believe that all researchers should compensate mTurk workers more fairly for their labor than has been the practice up until now.
2. List of common surnames in North America. Online: http://en.wikipedia.org/wiki/List_of_most_common_surnames_in_North_America; accessed January 30, 2013.
3. List of common Chinese surnames. Online: http://en.wikipedia.org/wiki/List_of_common_Chinese_surnames; accessed January 30, 2013.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2): 179-211.

Armitage, C. J., and M. Conner (2001). Efficacy of the Theory of Planned Behaviour: A meta-analytic review. *British Journal of Social Psychology, 40*(4): 471-499.

Bleske-Rechek, A., and K. Michels (2010). RateMyProfessors.com: Testing assumptions about student use and misuse. *Practical Assessment, Research and Evaluation, 15*(5). http://pareonline.net/pdf/v15n5.pdf

Blommaert, J. (2005). *Discourse*. Cambridge: Cambridge University Press.

Brown, M. J., Baillie, M., and S. Fraser (2009). Rating RateMyProfessors.com: A comparison of online and official student evaluations of teaching. *College Teaching, 57*(2): 89-92.

Davison, E., and J. Price (2009). How do we rate? An evaluation of online student evaluations. *Assessment and Evaluation in Higher Education, 34*(1): 51-65.

Fairclough, N. (2010). *Critical discourse analysis: The critical study of language* (2nd ed.). Harlow, UK: Longman.

Feldman, M. E., and A. J. Weseley (2013). Which name unlocks the door? The effect of tenant race/ethnicity on landlord response. *Journal of Applied Social Psychology, 43*: E416-E425.

Fort, K., Adda, G., and K. B. Cohen (2011). Amazon Mechanical Turk: Gold mine or coal mine? *Computational Linguistics, 37*(2): 413-420.

Gastil, J. (1990). Generic pronouns and sexist language: The oxymoronic character of masculine generics. *Sex Roles, 23*(11-12): 629-643.

Hart, C. (2013). Event-construal in press reports of violence in two recent political protests: A cognitive linguistic approach to CDA. *Journal of Language and Politics, 12*(3): 400-423.

Johnson, R. R., and A. D. Crews (2012). My professor is hot! Correlates of RateMyProfessors.com ratings for criminal justice and criminology faculty members. *American Journal of Criminal Justice*: 1-18.

Koller, V. (2005). Critical discourse analysis and social cognition: evidence from business media discourse. *Discourse and Society, 16*(2): 199-224.

Lakoff, G., and M. Johnson (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

O'Halloran, K. (2003). *Critical discourse analysis and language cognition*. Edinburgh: Edinburgh University Press.

Paolacci, G., and J. Chandler (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*(3): 184-188.

Peer, E., Vosgerau, J., and A. Acquisti (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46*(4): 1023-1031.

Pennycook, A. (1994). Incommensurable discourses? *Applied Linguistics, 15*(2): 115-138.

Reid, L. D. (2010). The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.Com. *Journal of Diversity in Higher Education, 3*(3): 137-152.

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., and B. Tomlinson (2010). *Who are the crowdworkers?: shifting demographics in Mechanical Turk*. Paper presented at the CHI '10 Extended Abstracts on Human Factors in Computing Systems, Atlanta, Georgia, USA.

Shuck, G. (2004). Conversational performance and the poetic construction of an ideology. *Language in Society, 33*(2): 195-222.

Steffes, E. M., and L. E. Burgee (2009). Social ties and online word of mouth. *Internet Research, 19*(1): 42-59.

Stubbs, M. (1997). Whorf's children: Critical comments on critical discourse analysis. In A. Wray and A. Ryan (eds.), *Evolving models of language*. Clevedon: Multilingual Matters. pp. 100-116

Subtirelu, N. (2015). "She does have an accent but…": Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com. *Language in Society, 44*(01): 35-62.

van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse and Society, 4*(2): 249-283.

Wodak, R. (2006). Mediation between discourse and society: assessing cognitive approaches in CDA. *Discourse Studies, 8*(1): 179-190.

Wodak, R., and M. Meyer (2009). Critical discourse analysis: History, agenda, theory, and methodology. In R. Wodak and M. Meyer (eds.), *Methods of critical discourse analysis*. London: Sage. pp. 1-33