The development of a stance annotation scheme:

Lessons for computational linguistics and sociolinguistic theory

Scott Kiesling
Jim Fitzpatrick

Jacob Eisenstein
Umashanthi Pavalanathan

Department of Linguistics
University of Pittsburgh

School of Interactive Computing
Georgia Institute of Technology
Presentation Outline

• Stance Preliminaries
• Research Questions
• Methods: Data and Coding
• Results
• Discussion
• Q&A
STANCE?
Social Meaning in Sociolinguistics

• Brown and Gilman (1960) is an early example showing the complex interpersonal meanings associated with address terms in European languages and how such terms impact interlocutor solidarity.

• Eckert (2000) has seeded a research program investigating the ways in which sociolinguistic style affects language use.

• Silverstein (2003), Johnstone and Kiesling (2008) and Eckert (2011) have all considered important ways in which indexical meanings are created, circulated, entextualized, and resemioticized in communities and interactions.

• Kiesling (2009) and Bucholtz (2009) both show how stancetaking is central to the styles indexed by address terms in Southwestern US Spanish and US English, especially among young men.
Social Meaning in Computational Linguistics

- Lots of work on sentiment polarity and affect, originating from study of customer reviews (for an overview, see Pang and Lee 2008)
- Power dynamics (Gilbert 2012; Mizil et al 2012; Prabhakaran et al 2012)
- Politeness (Faruqui and Pado 2012; Mizil et al 2013)
What is Stance? Working definition

• Stance is the relationship of a language user to some discursive figure, and to other language users in relation to that figure, expressed through language use (and abetted in some interactions by nonverbal and other cues).
• This discursive figure can be an interlocutor, a figure represented in the discourse, the animator, ideas represented in the discourse, or other texts.
• See Kiesling (2011) and Du Bois (2007) for more!
Elements of Stance

• Building from Kiesling (2011) and Kiesling and Onuffer (2012):
  • **Affect** is the *relationship of the language user to the stance focus*: How does the author indicate they feel about the stance focus? This is a dimension of evaluation.
  • **Investment** is the *relationship of the language user to the talk itself*: How strongly does the author feel about their claim, view, etc?
  • **Alignment** is the *relationship of the language user to some other language user(s)*: For example, does the utterance show sympathy, agreement, mutual knowledge, common identity? (Note that this is not simply agreement or disagreement!)
Main Questions

• **RQ #1**: Can stance be operationalized so that it can be used in **quantitative linguistic analysis**? How? If so, can we develop a framework that works across topics and contexts?

• **RQ #2**: Can a high measure of **interrater agreement** be obtained for such a measure?
Further Question

• Should stance be broken into elements as outlined above, or should it rely on single adjectives such as *belligerent, friendly, laid back, etc.*?
METHODS
Reddit

- Reddit is an internet discussion forum site founded in 2005.
- Extremely high-traffic, according to Alexa.com’s metrics:
  - Top 25 in traffic for sites globally
  - Top 10 in traffic for US sites
  - Site is divided into thematically-organized forums, called subreddits
    - Over 626,000 as of 4/8/2015 (according to redditmetrics.com)
- Users must be registered to use many aspects of the site’s functionality, although “throwaway” accounts are common.
- Registered users may also upvote or downvote posts, with the most upvoted posts in a subreddit being displayed first.
Anatomy of a Reddit post

karma for original post

original post content: title, text, author, time...

Things for a college kid to do during the summer? (self.Atlanta)
submitted 21 hours ago by vodoman123
I will be staying in Midtown this summer and was wondering what I could do to have fun.

8 comments source share hide all child comments

all 8 comments

sorted by: best

navigate by: submitter | moderator | friend | me | admin | highlighted | gilded | IAMA | images | videos | popular | new

[-] mechukulele 2 points 20 hours ago
Shoot the hooch

[-] vodoman123 [S] 0 points 19 hours ago
I was looking for things in the city since I do not have a car
Post Collection

- Threads were collected with a Python script using the PRAW Reddit wrapper.
- We targeted posts with between 14 and 25 total comments
- 23 total threads were collected, for a total of 402 annotated stance moves, coded using the specifications covered in the following section.
- The following subreddits were targeted in two phases:
  - Phase 1: fitness, parenting, metal (music), Pittsburgh, Atlanta
  - Phase 2: ELI5 (Explain Like I’m Five)
Coding workflow

• The workflow for the stance annotation scheme implemented here consists of three primary steps:
  • Segmentation of collected data
  • Annotation of segmented data
  • Computation of agreement
Step One: Segmentation

• Each thread was randomly assigned a team of three reviewers from a pool of six available reviewers (R1, R2, R3)
• R1 divides each post into one or more stance moves.
• R2 either signs off on R1’s divisions or proposes an alternative.
• In the case of disagreement between R1 and R2, R3 breaks the tie.
• Posts to be annotated have been agreed upon by two out of three reviewers.
Segmentation Comparison

One stance move:

• A laser powerful enough to burn paper, by itself, costs upwards of $150, and that doesn't include the rest of the printer mechanics. If it became a popular enough technology, that price could come down, though. Probably not enough to make it financially viable.

• **Stance focus**: the viability of making a thermal laser printer

More than one stance move:

• "It's not necessary. I never refrigerate mustard and it doesn't spoil. However, it makes sense to keep mustard in the fridge for these reasons: it tastes better cold, it's a convenient place to keep it (near other sandwich ingredients), exercising an abundance of caution.”

• **Stance focus 1**: The idea of refrigerating mustard

• **Stance focus 2**: Hypothetical reasons to refrigerate mustard
Step Two: Annotation

- Each of the three raters assigned to each thread then rated each stance move within the thread using a 1-5 scale on the three stance elements: affect, investment, and alignment.
- Each rater also assigned a stance activity to the move (a verb ending in –ing, e.g. clarifying) and a stance adjective (e.g. helpful)
### A Look at the Annotation Environment

<table>
<thead>
<tr>
<th>User</th>
<th>ID-rep</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayoung4u</td>
<td>001</td>
<td>Why do we refrigerate mustard, if the base ingredients are all non-perishables like vinegar, water, mustard seed, salt, turmeric, and paprika?</td>
</tr>
<tr>
<td>cnash</td>
<td>002</td>
<td>Often, when you mix non-perishables with water, the result is perishable. Dehydration is one of the primary ways of making foods shelf-stable.</td>
</tr>
<tr>
<td>ThinknBoutStuff</td>
<td>003</td>
<td>From what I understand, food with more water content have higher water activity, meaning more organisms can grow within the water that isn't bonded with food particles. I do not know much, there there is actually a crap ton of science behind this.</td>
</tr>
<tr>
<td>cdb03b</td>
<td>004</td>
<td>Vinegar is perishable, it just takes a very long time. Water is perishable due to microbial growth. And all the other ingredients save salt become perishable when you add water. Mustard will spoil, we just consume it faster than it will spoil generally.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>ParentID</th>
<th>Karma</th>
<th>Stance focus</th>
<th>Affect 1=negative</th>
<th>Investment 1=Low</th>
<th>Alignment 1=Strong disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>2ypbo9</td>
<td></td>
<td>10</td>
<td>refrigeration of mustard</td>
<td>3</td>
<td>2</td>
<td>--</td>
</tr>
<tr>
<td>cpbns54</td>
<td>t3_2ypbo9</td>
<td>17</td>
<td>perishability due to water</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>cpbqwn</td>
<td>t1_cpbns54</td>
<td>5</td>
<td>water content of food</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>cpboaej</td>
<td>t3_2ypbo9</td>
<td>4</td>
<td>perishability of individual ingredients</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
Step Three: Computation

- Interrater reliability numbers were computed for each post across the three raters using Krippendorf’s alpha (Hayes & Krippendorf 2007):
  - Alpha values of zero indicate chance agreement.
  - Alpha values > 0 indicate better than chance agreement.
- Other experiments, e.g. Passonneau’s (2006) lexical semantic annotation task, have established a broad range of satisfactory alpha values between 0.3-0.7.
RESULTS
A reminder of why we’re here...

• **RQ #1**: Can stance be operationalized so that it can be used in *quantitative linguistic analysis*? How? If so, can we develop a framework that works across topics and contexts?

• **RQ #2**: Can a high measure of *interrater agreement* be obtained for such a measure?
Comparing Agreement by Phase

![Comparison of Krippendorf's Alpha by phase and type of analysis.](image)
Regression Results: Summary

• In addition to the data presented in the previous two charts, we ran a number of multiple linear regressions to determine whether different combinations of reviewers generated better agreement: linguists vs. computer scientists, professors vs. students, advisor-advisee groupings, etc.
• Phase and attribute did return significant results.
• No predictor involving who the annotator was returned significant results.
• The null results here may seem discouraging, but in actuality they lend credibility to the idea that it’s possible to train non-linguists to rate stance and have generally good interrater reliability.
DISCUSSION
Speech Activity & Subreddit “Personality”

• In Phase 1, reliability measures for investment and alignment were comparatively lower than in Phase 2. There are (at least) three possible explanations for this:
  • Over time, raters became more experienced, which would lead to more consistency in results.
  • Annotating one subreddit (i.e. ELI5) created more uniformity of discursive purpose or discursive practice.
  • ELI5 might be a subreddit where identifying investment and alignment are “easy.”
Quantitative lessons

- Stance has traditionally been analyzed (robustly and fruitfully!) using qualitative methods.
- Quantitative annotation presents challenges: interrater agreement especially.
- But if nothing else this exercise forces a sharper thinking about what stance actually is and how language users create it, because every utterance must get a stance.
- Any quantitative work begins with qualitative definitional work that overlays a theoretical perspective on the data.
Sociolinguistic lessons

- The fact that naive users can be trained relatively easily suggests that the stance dimensions are not without merit.
- While the annotation scheme may be a simplification of reality (what theory or model isn't?) it looks promising as a tool for bringing stance meanings into quantitative analyses.
- Also looks promising as a more standardized vocabulary to discuss stance differences among personae/styles/identities/communities, cf. Kiesling and Onuffer’s (2012) application of these dimensions to meetings among female academics.
Acknowledgements

• NSF Grant IIS-1110904 (Kiesling)
• NSF Grant IIS-1111142 (Eisenstein/Xing)
• Our undergraduate research assistants: Shameek Poddar and Kaylen Sanders
• Undergraduate research assistants for previous iterations of this project: Spencer Onuffer and Alex Hardware
Selected References

[θæŋk ju]

If you would like your very own personal copy of these slides, scan this QR code with your mobile device from your seat.

For additional questions and comments, email jbf24@pitt.edu or follow @jimthelinguist on Twitter.
The Stance Triangle
(adapted from original in Du Bois 2007)

Stance Taker A

aligns
positions
evaluates

Stance Focus

positions

Stance Taker B