Sarcasm and Subtlety: Social language, human behavior and sentiment

UC Santa Cruz

Pranav Anand & Marilyn Walker
Linguistics & Computer Science
I'm not as much undecided as I am totally disgusted.
The Modern Soapbox

- 100Ks of discussions
- Millions of posts
- Any conceivable topic

<table>
<thead>
<tr>
<th>Topic</th>
<th>IAC</th>
<th>CD</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evolution</td>
<td>43,191</td>
<td>433</td>
<td>43,624</td>
</tr>
<tr>
<td>Gun Control</td>
<td>36,194</td>
<td>1,908</td>
<td>38,102</td>
</tr>
<tr>
<td>Abortion</td>
<td>37,358</td>
<td>0</td>
<td>37,358</td>
</tr>
<tr>
<td>Gay Marriage</td>
<td>19,843</td>
<td>2,582</td>
<td>22,425</td>
</tr>
<tr>
<td>Existence of God</td>
<td>7,071</td>
<td>4,304</td>
<td>11,375</td>
</tr>
<tr>
<td>Death Penalty</td>
<td>3,003</td>
<td>2,280</td>
<td>5,283</td>
</tr>
<tr>
<td>Climate Change</td>
<td>2,712</td>
<td>724</td>
<td>3,436</td>
</tr>
</tbody>
</table>

Most Popular OPEN Debates

Ninjas versus Pirates
Khabar @KhabarEU · 2h
Corruption cases involving 3 million yuan ($463,000) or more may incur
#deathpenalty
#China
Read:

China sets death penalty threshold - Khabar.eu
Corruption cases involving three million yuan ($463,000) or more may incur the death penalty in future, Chinese authorities ruled Monday.

Khabar.eu
How can we learn what someone feels from what they say?
Language of opinions, thoughts and feelings
Language of opinions, thoughts and feelings

- Rich and subtle
- Creative
- High level representation of thought
Pragmatics

I’m out of gas
Please help me get gas

It’s cold in here!
Please close the window

I’m not happy
I’m unhappy

France is a hexagon
France resembles a hexagon

If you mow the lawn, I’ll give you $5.
If you don’t mow the lawn, I won’t.

Can’t capture rich and subtle language by being pragmatically poor.
The research potential

- Language of social media has unique properties: e.g. social, dialogic, emotional, informal

- Never had this scale/variety of data for building and testing models of dialogue

- Many open research problems

- **NSF IIS-1302668**: Collaborative Research: RI: Processing Opinion Sharing Dialogue in Social Media (Walker, Anand, Fox Tree, Whittaker @UCSC)
Our research agenda

- Build classifiers for dialogue categories
  - Agreement/Disagreement
  - Factual vs. Emotional Arguments
  - Author mood/affect
  - Sarcasm

- Build corpora of dialogue and informal language ([https://nlds.soe.ucsc.edu](https://nlds.soe.ucsc.edu))

- Build theoretical insight about pragmatics of informal dialogue
Our research approach

- Hybrid of linguistic analysis and machine learning
- Want to arrive at basic psychological reasoning
  - At the level of a 7 year old
  - About daily life experiences
Roadmap

- Compositional pragmatics for sentiment
- Sarcasm via bootstrapping
Sentiment Analysis
A common approach to sentiment

Look at sentiment words to figure out emotion

Extremely **poor** made. On the top of the hammer, it has a hollow **1** hole and the manufacturer just filled it with brown **messy** glue?

**negative opinion**

I got this little hammer for jewelry stamping and it's **wonderful**! It's just the right size and has **enough** weight to it that I don't have to **slam** it down to get a **good** impression. The handle is also a **great** size and fits **well** in my hand.

**positive opinion**
A common approach to sentiment

Give me a dictionary long enough and a counter big enough and I shall know what the world thinks.
An Example

Trump beat Cruz yesterday.
An Example

Trump smoked Cruz yesterday.
An Example

My guy Trump beat Cruz yesterday.
Semantics of Benefit and Loss

Trump beat Cruz

Trump won \textit{benefit}

Cruz lost \textit{loss}
Psychology of Benefit and Loss

Trump beat Cruz

Trump won benefit
Cruz lost loss

- Author’s opinion of event: contingent on feelings towards beneficiary and loser
Psychology of Benefit and Loss

Trump beat Cruz

Trump   Cruz   Trump beat Cruz

👍   👎   👍
👎   👍   👎
Psychology of Benefit and Loss

X beat Y

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
| - | - | ?? ???
# Psychology of Benefit and Loss

<table>
<thead>
<tr>
<th>$x$</th>
<th>$E_{existing}$</th>
<th>$E_{non-existent}$</th>
<th>$E_{+property}$</th>
<th>$E_{-property}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

EXISTENCE: A participant has/lacks existence.

POSSESSION: One participant has/lacks possession of another.

AFFECTEDNESS: A participant has a positive/negative property.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$E_{have}$</th>
<th>$E_{lack}$</th>
<th>$E_{withhold}$</th>
<th>$E_{deprive}$</th>
<th>$E_{spare}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>#</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>#</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>#</td>
<td>#</td>
<td></td>
</tr>
</tbody>
</table>

VERB_CLASS: 
RESULT: 
STATE: 
FUNCTOR: 

<table>
<thead>
<tr>
<th>creation</th>
<th>destruction</th>
<th>have/lack</th>
<th>existence</th>
<th>deprived</th>
<th>result</th>
<th>state</th>
<th>functor</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>$E_{existing}$</td>
<td>$E_{non-existent}$</td>
<td>$E_{have}$</td>
<td>$E_{lack}$</td>
<td>$E_{positive}$</td>
<td>$E_{negative}$</td>
</tr>
<tr>
<td>gain</td>
<td>loss</td>
<td>benefit</td>
<td>injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Political Ads from 2008 Race

Reschke & Anand (2011)
Tweets from the NH primary

Even though I'm upset that Trump won, I'm pretty happy that Christie got walloped.

I certainly am happy that Clinton has some competition

I am worried that Cruz will eviscerate Trump

I want Trump to win the Republican nomination.
Tweets from the NH primary

I am happy that Clinton has competition
I am happy that Clinton has competition

+ event
Tweets from the NH primary

+ event

I am happy that [ Clinton has competition ]

x competition
+
-
-
+


Tweets from the NH primary

+ event

I am happy that [ Clinton has competition ]

<table>
<thead>
<tr>
<th></th>
<th>competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
Tweets from the NH primary

Even though I'm upset that *Trump* won, I'm pretty happy that *Christie* got walloped.

I certainly am happy that *Clinton* has some competition.

I am worried that *cruz* will eviscerate *trump*.

I want *Trump* to win the Republican nomination.
First person shortcut

- “I” in a comment always marks a +
- Lots of examples in first person blog posts
- Lets us learn types of events easier
  - Don’t have to worry if a + event has to do with a − participant
  - Can learn “self-helping” and “self-hurting” events

I got a promotion! I was fired
I had drinks with some friends I was stood up
First person shortcut to events (blogs)

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEADED HOME</td>
<td>LOST JOB</td>
</tr>
<tr>
<td>PLAYS</td>
<td>KILLING ME</td>
</tr>
<tr>
<td>GRILLED</td>
<td>DIAGNOSED WITH</td>
</tr>
<tr>
<td>SHARE</td>
<td>HATE MYSELF</td>
</tr>
<tr>
<td>TAKE PICTURES</td>
<td>NOT LOVE ME</td>
</tr>
<tr>
<td>THRILLED</td>
<td>LOVED YOU</td>
</tr>
<tr>
<td>PARTICIPATED IN</td>
<td>SCREAMING</td>
</tr>
<tr>
<td>DANCED</td>
<td>NOT HELP IT</td>
</tr>
<tr>
<td>SURPRISED BY</td>
<td>WANT CRY</td>
</tr>
<tr>
<td>RECOMMEND</td>
<td>CHEATER</td>
</tr>
</tbody>
</table>

- **Positive** matches contain more action-rich verbs
- **Negative** matches are more related to personal feelings
Sarcasm

TODAY'S MOOD FORECAST...

BITCHY WITH A CHANCE OF SARCASM.

STIRR OF ECHOES

NATURAL LANGUAGE AND DIALOGUE SYSTEMS
UC SANTA CRUZ
Sarcasm, new and old
Forgive me if I doubt your sincerity, but you seem like a troll to me. I suspect that you aren’t interested in learning about evolution at all. Your questions, while they do support your claim to know almost nothing, are pretty typical of creationist “prove it to me” questions.

Wrong again! You obviously can’t recognize refutation when its printed before you. I haven’t made the tag “you liberals” derogatory. You liberals have done that to yourselves! I suppose you’d rather be called a social reformist! Actually, socialist is closer to a true description.
Sarcasm is creative, but principled

- Classic case: contrast between a negative situation and a positive sentiment

I just love getting pulled over it's so amazing

Yay for not having weekends

I just love being unfollowed. It's my favorite thing!
Sarcasm is creative, but principled

- Classic case: contrast between a **negative situation** and a **positive sentiment**
- But argumentative data tends to be more complex:

**I love** how you're able to determine, not having been there, precisely what was said and precisely where it fell in relation to the law.
A more complicated example

You already knew that.
A more complicated example

You already knew that, didn’t you?
A more complicated example

But you already knew that, didn’t you?
A more complicated example

But you already knew that, didn’t you, dearie?
A more complicated example

But you (of course) already knew that, didn’t you, dearie?


| **Doubt** | Forgive me if I **doubt** your sincerity, but you seem like a **troll** to me. I **suspect** that you aren’t interested in learning about evolution at all. Your questions, while they do support your claim to **know almost nothing**, are **pretty** typical of creationist **“prove it to me”** questions. |
| **Negative Eval/Insult** | |
| **First Person Thought** | |
| **Downtoners** | |
| **Quotation** | |
| **Exclamation** | **Wrong again!** You **obviously** can’t recognize refutation when it’s printed before you. I haven’t made the tag **“you liberals”** derogatory. You liberals have done that to yourselves! I **suppose** you’d rather be called a social reformist! Actually, socialist is closer to a true description. |
| **Adverbs of Emphasis** | |
| **Quotation** | |
| **Hypothetical Marker** | |
Theoretical Models of Sarcasm

- Ironic interpretation relies on a contrast effect
- Event or Situation evokes a scale
- Event can be placed on that scale
- **Statement** about event contrasts with actual scale placement

Colston & O’Brian (2000)
A general pattern of conflict

- On scale $X$, author simultaneously commits to
  - Sentence high on $X$
  - Sentence low on $X$

But you (of course) already knew that, didn’t you, dearie?

I love how you're able to determine, not having been there, precisely what was said and precisely where it fell in relation to the law.
Bootstrapping
Bootstrapping
Bootstrapping

- Precision Error
- Recall Error
Bootstrapping

- Start with seed terms
Bootstrapping

- Start with seed terms
- Learn H precision patterns for seeds
Bootstrapping

- Start with seed terms
- Learn H precision patterns for seeds
- Extract new seeds from patterns
Bootstrapping

- Start with seed terms
- Learn H precision patterns for seeds
- Extract new seeds from patterns
- Repeat!
Linguistic Pattern Learner: AutoSlog-TS

- **SUNDANCE**: A shallow parser (Riloff, Phillips 2004)

- **AutoSlog-TS**: A weakly supervised pattern learner (Riloff, 96)

- Uses 24 syntactic templates representing types of linguistics pattern that get instantiated

<table>
<thead>
<tr>
<th>Pattern Template</th>
<th>Example Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;subj&gt; PassVP</td>
<td>&lt;subj&gt; was observed</td>
</tr>
<tr>
<td>&lt;subj&gt; ActVP</td>
<td>&lt;subj&gt; observed</td>
</tr>
<tr>
<td>&lt;subj&gt; ActVP Dobj</td>
<td>&lt;subj&gt; want explanation</td>
</tr>
<tr>
<td>&lt;subj&gt; ActInfVP</td>
<td>&lt;subj&gt; expected to find</td>
</tr>
<tr>
<td>&lt;subj&gt; PassInfVP</td>
<td>&lt;subj&gt; was used to measure</td>
</tr>
<tr>
<td>&lt;subj&gt; AuxVP Dobj</td>
<td>&lt;subj&gt; was success</td>
</tr>
<tr>
<td>&lt;subj&gt; AuxVP Adj</td>
<td>&lt;subj&gt; is religious</td>
</tr>
</tbody>
</table>

- `<subj> <dobj>`
- `<np>`
- `<subject> <preposition> <np>`
- `<subject> <preposition> <possessive> <np>`
- `<subject> <preposition> was replaced by <np>`
- `<subject> <preposition> evolved from <np>`
- `<subject> <preposition> to use as <np>`
- `<subject> <preposition> origins of <np>`
- `<subject> <preposition> question is <dobj>`
- `<subject> <preposition> like to see <dobj>`
## Sarcastic Patterns for Tweets

<table>
<thead>
<tr>
<th>F</th>
<th>P</th>
<th>Syntactic Template</th>
<th>Sample Pattern</th>
<th>Instantiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>1</td>
<td>2gram Adj Noun</td>
<td>JUST LOVE</td>
<td>I just love getting pulled over its so amazing</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>2gram Adj Noun</td>
<td>GREAT WAY</td>
<td>And now I don't feel good. Fantastic. Great way to end my great day.</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>Np Prep &lt;NP&gt;</td>
<td>YAY FOR</td>
<td>Yay for not having weekends</td>
</tr>
<tr>
<td>13</td>
<td>0.93</td>
<td>2gram Adv Adv</td>
<td>FAVORITE THING</td>
<td>I just love being unfollowed. It's my favorite thing!</td>
</tr>
<tr>
<td>9</td>
<td>0.89</td>
<td>Subj AuxVP &lt;dobj&gt;</td>
<td>LOVE GETTING</td>
<td>I love getting up at 5am...hands down, best time to start the day</td>
</tr>
</tbody>
</table>
## Sarcastic Patterns for Forums

<table>
<thead>
<tr>
<th>F</th>
<th>P</th>
<th>Syntactic Template</th>
<th>Sample Pattern</th>
<th>Instantiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>I</td>
<td>ActVp Prep &lt;NP&gt;</td>
<td>SAY ABOUT</td>
<td>You sure you weren't on a sub? You know what they say about submariners, right?</td>
</tr>
<tr>
<td>8</td>
<td>I</td>
<td>2gram Adj Noun</td>
<td>YOUR HEAD</td>
<td>Oh man Easy now I KNOW you were dropped on your head as a child....</td>
</tr>
<tr>
<td>7</td>
<td>I</td>
<td>ActVp Prep &lt;NP&gt;</td>
<td>GET OVER</td>
<td>Yea they get over it in a year or two. Haaa Haaa</td>
</tr>
<tr>
<td>7</td>
<td>I</td>
<td>2gram Adv Adv</td>
<td>AH YES</td>
<td>Ah yes, your diversionary tactics. I am sure if I provided it, you would move the goal posts, or demand evidence that you know is not yet available..</td>
</tr>
</tbody>
</table>
### Not Sarcastic Patterns for Forums

<table>
<thead>
<tr>
<th>F</th>
<th>P</th>
<th>Syntactic Template</th>
<th>Sample Pattern</th>
<th>Instantiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>1</td>
<td>2gram Adj Noun</td>
<td>NATURAL SELECTION</td>
<td>As was pointed out in some of your own quotations, survival is not an 'end' of natural selection.</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2gramAdjNoun</td>
<td>BIG BANG</td>
<td>There are certainly other theories, but the big bang is the only one supported by the evidence.</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>Np Prep &lt;NP&gt;</td>
<td>PURPOSE OF</td>
<td>Since the original purpose of the Second Amendment doesn't appear to be needed, there is a new use that is protected.</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>Np Prep &lt;NP&gt;</td>
<td>EVOLUTION OF</td>
<td>We can see the evolution of organisms in the world today.</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>Np Prep &lt;NP&gt;</td>
<td>SPECIES OF</td>
<td>There is no evidence of any interim species of apes and humans.</td>
</tr>
<tr>
<td>24</td>
<td>0.92</td>
<td>Np Prep &lt;NP&gt;</td>
<td>THEORY OF</td>
<td>Eleven years is quite swift, but it actually supports the theory of evolution through natural selection.</td>
</tr>
<tr>
<td>16</td>
<td>0.81</td>
<td>infinitiveverb &lt;dobj&gt;</td>
<td>EXPLAIN</td>
<td>Demographics plays a role, but it does not explain away the whole shift in the numbers.</td>
</tr>
</tbody>
</table>
### Learned Patterns for Factual vs. Feeling

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Text Match</th>
<th>Sentence Match</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FACT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP Prep <code>&lt;np&gt;</code></td>
<td>SPECIES OF</td>
<td>Because specific <em>species of</em> horses only exist at certain periods in the geological record.</td>
</tr>
<tr>
<td><code>&lt;subj&gt;</code> PassVP</td>
<td>EXPLANATION OF</td>
<td>So basically you are saying that evolution is the one and only true explanation of how we came to be, and that it is okay for the schools to force students to accept this fact.</td>
</tr>
<tr>
<td><code>&lt;subj&gt;</code> BE EVIDENCE</td>
<td></td>
<td>Contradiction in science <em>is not evidence</em> of doubt.</td>
</tr>
<tr>
<td><strong>FEEL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj Noun</td>
<td>MY ARGUMENT</td>
<td>My arguments are completely solid, and you haven’t even gotten close to refuting them.</td>
</tr>
<tr>
<td><code>&lt;subj&gt;</code> BE ABSURD</td>
<td></td>
<td>Funny enough there wasn’t a single scientific fact presented to support creationism, they merely tried to use science to disprove evolution. Which <em>is absurd</em>, given that all their facts were wrong.</td>
</tr>
<tr>
<td>Adv Adj</td>
<td>MORALLY WRONG</td>
<td>Since the fetus was already dead, then besides inappropriate use of a dead human body I cannot think of anything <em>morally wrong</em> with it.</td>
</tr>
</tbody>
</table>
“OF” as a preposition:
- Observed to attach to objective terminology in factual arguments

“FOR” as a preposition:
- Observed to express advocacy or refer to population groups in feeling-based arguments

<table>
<thead>
<tr>
<th>FACT “OF” Phrases</th>
<th>FEEL “FOR” Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESULT OF</td>
<td>MARRIAGE FOR</td>
</tr>
<tr>
<td>ORIGIN OF</td>
<td>STANDING FOR</td>
</tr>
<tr>
<td>THEORY OF</td>
<td>SAME FOR</td>
</tr>
<tr>
<td>EVIDENCE OF</td>
<td>TREATMENT FOR</td>
</tr>
<tr>
<td>PARTS OF</td>
<td>DEMAND FOR</td>
</tr>
<tr>
<td>EVOLUTION OF</td>
<td>ATTENTION FOR</td>
</tr>
<tr>
<td>PERCENT OF</td>
<td>ADVOCATE FOR</td>
</tr>
<tr>
<td>THOUSANDS OF</td>
<td>NO EVIDENCE FOR</td>
</tr>
<tr>
<td>EXAMPLE OF</td>
<td>JUSTIFICATION FOR</td>
</tr>
<tr>
<td>LAW OF</td>
<td>EXCUSE FOR</td>
</tr>
</tbody>
</table>

High-Probability FACT Phrases with “OF” and FEEL Phrases with “FOR”

Oraby et al. (2015)
Upshot
What tools do we need?

- Tracking arguments
  - commonsense-event outcomes
  - abstract logic of argument frames and types
  - techniques to recognize and extract arguments

- Sarcasm
  - commonsense-event outcomes
  - knowing how people talk about events
  - models robust to creative productivity

- Tracking mood in narratives
  - commonsense-event outcomes
  - knowing how people think of complex mental states
  - knowing how people construct stories
What tools do we need?

- Tracking arguments
  - commonsense-event outcomes
  - abstract logic of argument frames and types
  - techniques to recognize and extract arguments

- Sarcasm
  - commonsense-event outcomes
  - knowing how people talk about events
  - models robust to creative productivity

- Tracking mood in narratives
  - commonsense-event outcomes
  - knowing how people think of complex mental states
  - knowing how people construct stories
Some potential applications

- Higher quality and more meaningful interactions with machines
  - Eliminate and/or interpret sarcastic information
  - Mood-guided responses
  - Provide digests of online discussions

- Discover new ideas that matter

- Enable design of “proactive positive planning” systems, which curate events that help w/ mood
Common theme

- What we are looking for:
  - the affect people have to common life events
  - the “obvious” of human psychology

- Different from other cases of “world knowledge” in artificial intelligence
  - not about the physical world, but the emotional/ethical
  - varies from person to person
    - Is life sacred?
    - How does being alone make you feel?
What kind of educational training should we be giving to students?
Open questions

- Assuming Machine Learning/Data Science/Corpus Linguistics, what else is desirable?

- We feel interdisciplinary training is useful, do you?

- How important is knowledge of linguistic syntax and semantics?
  - Know the difference between a syntactic and semantic head?
  - Know different verb types, lexical relations, etc.

- How important is knowledge of pragmatic and discourse categories?
  - Know what a tag question is? What discourse markers do?
Programs completed and in progress

- Designated Emphasis in CL and NLP
  - can be added to any program in CS or Ling

- M.S. in Data Science
  - NLP Track: Core NLP, Information extraction, Information Retrieval, Data Mining (no linguistics)

- NLP M.S./M.A. ?
  - Some classes from M.S. in Data Science + some Linguistics