Natural Language Processing (NLP) for Computational Social Science

Cristian Danescu-Niculescu-Mizil and Lillian Lee

http://www.cs.cornell.edu/courses/cs6742/2015fa

Datasets:
http://www.cs.cornell.edu/home/llee/data/index.html
Why do people do what they do (when other people are involved)?

NLP: a great way to find out!
Why NLP for CSS?

Much of online human activity leaves digital traces that are recorded in natural-language format.

Exploiting these resources under a computational framework can bring a phase transition in our understanding of human social behavior and shape the future of social-media systems.
TODAY:

... Research questions
persuasion, linguistic change, framing

... Techniques
language models, Bayesian feature analysis

... Research practices
controls, feasibility, data inspection
The Social effects of linguistic Subtleties
The Social effects of linguistic Subtleties

"Motivating voter turnout" (Bryan et al., 2011)
The Social effects of linguistic subtleties

"Motivating voter turnout" (Bryan et al., 2011)

"How important it is to you to be a voter?"

"How important it is to you to vote?"
The Social effects of linguistic Subtleties

“Motivating voter turnout” (Bryan et al., 2011)

“How important it is to you to be a voter?” (identity)

“How important it is to you to vote?” (action)
The Social effects of linguistic subtleties

"Motivating voter turnout" (Bryan et al., 2011)

"How important it is to you to be a voter?" (identity)

"How important it is to you to vote?" (action)
The Social effects of linguistic subtleties

"Motivating voter turnout" (Bryan et al., 2011)

"How important it is to you to be a voter?" (identity)

"How important it is to you to vote?" (action)
The Social effects of linguistic subtleties

"Motivating voter turnout" (Bryan et al., 2011)

"How important it is to you to be a voter?" (identity)

"How important it is to you to vote?" (action)

How things are said
(vs what is said)
The Social effects of linguistic Subtleties

"The role of placebo information" (Langer et al., 1978)

"I have 5 pages. May I use the xerox machine?"
The Social effects of linguistic Subtleties

"The role of placebo information" (Langer et al., 1978)

"I have 5 pages. May I use the xerox machine?"

"I have 5 pages. May I use the xerox machine, because I need to make copies?"
The Social effects of linguistic Subtleties

"The role of placebo information" (Langer et al., 1978)

"I have 5 pages. May I use the xerox machine?"

"I have 5 pages. May I use the xerox machine, because I need to make copies?"
The Social effects of linguistic Subtleties

"The role of placebo information" (Langer et al., 1978)

"I have 5 pages. May I use the xerox machine?"  
60% agreed

"I have 5 pages. May I use the xerox machine, because I need to make copies?"  
93% agreed
The Social effects of linguistic Subtleties

"The role of placebic information" (Langer et al., 1978)

"I have 5 pages. May I use the xerox machine?" 60% agreed

"I have 5 pages. May I use the xerox machine, because I need to make copies?" 93% agreed

"I have 5 pages. May I use the xerox machine, because I am in a rush?" 94% agreed
The Social effects of linguistic Subtleties

Today's data → opportunity to discover and better understand Social effects
The Social effects of linguistic Subtleties

Today's data → opportunity to discover and better understand Social effects

A classification problem?

Example: (How) do male and female describe things differently?
The Social effects of linguistic subtleties

Today's data → opportunity to discover and better understand social effects

A classification problem?

Example: (How) do male and female describe things differently?

Gender classification
The Social effects of linguistic subtleties

Today's data → opportunity to discover and better understand Social effects

A classification problem?

Example: (How) do male and female describe things differently?

Gender classification

Issue: Gender-topic confound (Argamon et al. 2003, Sarawgi et al. 2011)

"Finance" trends male, but what about females who talk about finance?
Today's data → opportunity to discover and better understand social effects

Challenges:

* maintaining the controlled, hypothesis-driven nature of traditional studies
  > sense (and luck) to find the right data
The Social effects of linguistic Subtleties

"The role of placebo information" (Langer et al., 1978)

"I have 5 pages. May I use the xerox machine?" 60% agreed

"I have 5 pages. May I use the xerox machine, because I need to make copies?" 93% agreed

"I have 5 pages. May I use the xerox machine, because I am in a rush?" 94% agreed
The Social effects of linguistic Subtleties

"The role of placebic information" (Langer et al., 1978)

"I have 5 pages. May I use the xerox machine?" 60% agreed

"I have 20 pages. May I use the xerox machine, because I need to make copies?" 93% agreed

"I have 5 pages. May I use the xerox machine, because I am in a rush?" 94% agreed
The Social effects of linguistic Subtleties

Today’s data → opportunity to discover and better understand Social effects

Challenges:

* maintaining the controlled, hypothesis-driven nature of traditional studies

→ sense (and luck) to find the right data
The Social effects of linguistic Subtleties

Today’s data → opportunity to discover and better understand Social effects

Challenges:

* maintaining the controlled, hypothesis-driven nature of traditional studies
  > sense (and luck) to find the right data

"How to ask for a favor" (Althoff et al., 2013)

20,000 requests for … pizza
[Request] I have gotten pizza before from this subreddit, but it's Easter, and I'm stuck at school because finals for me start tomorrow, and I'm broke.

submitted 3 days ago by silentsly

comment share

[Request] I've been working on my first computer for 6 hours, only to find my GPU was DOA. Can someone hit me up with some pizza please?

submitted 3 days ago by bigbootypanda

4 comments share

[Request] Spooky podcasts go great with pizza! (California)

submitted 3 days ago by posclutelyabsotively

comment share

[REQUEST] I know this is a long shot. But I've come to the end of college and have drained my funds for it 100% I am currently waiting on an email from said college that will basically determine my future. I have never been so stressed or scared. Pizza would be a comfort. Promise to pay it forward.

submitted 3 days ago by mrsansgruber

3 comments share
[Request] I have gotten pizza before from this subreddit, but it's Easter, and I'm stuck at school because finals for me start tomorrow, and I'm broke.

submitted 3 days ago by silentsly

[Request] I've been working on my first computer for 6 hours, only to find my GPU was DOA. Can someone hit me up with some pizza please?

submitted 3 days ago by bigbootypanda

4 comments  share

[Request] Spooky podcasts go great with pizza! (California)

submitted 3 days ago by posclutelyabsotively

[REQUEST] I know this is a long shot. But I've come to the end of college and have drained my funds for it 100% I am currently waiting on an email from said college that will basically determine my future. I have never been so stressed or scared. Pizza would be a comfort.

submitted 3 days ago by mrsansgruber

Promise to pay it forward.

3 comments  share
65

[Request] I have gotten pizza before from this subreddit, but it's Easter, and I'm stuck at school because finals for me start tomorrow, and I'm broke.

submitted 3 days ago by silentsly

share

66

[Request] I've been working on my first computer for 6 hours, only to find my GPU was DOA. Can someone hit me up with some pizza please?

submitted 3 days ago by bigbootypanda

4 comments share

68

[Request] Spooky podcasts go great with pizza! (California)

submitted 3 days ago by posolutelyabsotively

share

69

[REQUEST] I know this is a long shot. But I've come to the end of college and have drained my funds for it 100% I am currently waiting on an email from said college that will basically determine my future. I have never been so stressed or scared. Pizza would be a comfort. Promise to pay it forward.

submitted 3 days ago by mrshansgruber

3 comments share
The Social effects of linguistic subtleties

Today's data → opportunity to discover and better understand such effects

Challenges:

* maintaining the controlled, hypothesis-driven nature of traditional studies
  * sense (and luck) to find the right data

"How to ask for a favor" (Althoff et al., 2013)

20,000 requests for ... pizza

Language choices can increase success rate from 9% to 57%
The Social effects of linguistic Subtleties

Today's data → opportunity to discover and better understand such effects

Challenges:

* maintaining the controlled, hypothesis-driven nature of traditional studies
  › sense (and luck) to find the right data

"Winning arguments" (Tan et al., 2016)

20,000 persuasion "contests"
CMV: the Tontine should be legalized and made a common retirement strategy.

Basically, today we have a huge problem with retirement. A tontine for retirement looks like... The yearly sum is divided evenly for all the surviving participants. The key advantages as I see it are:

* We don't need actuaries
* Management fees can be quite low
* (Another reason)

But CMV. Are there major risks I am not foreseeing? [+2 more questions]

Very interesting. I'll give a Δ because I didn't have any idea that was true and changes my idea of how the tontine should work. That said, I don't think it's unsolvable. [+44 words]

The Social Security system is basically one giant tontine. So it's already legal.

I'd imagine the tontine as a secondary system to social security though, one that is optional for people to do, not mandatory like social security. [+11 words]

Then your back to needing actuaries, to predict [+11 words]

Depends how exact you need to be [+33 words]

There are some key differences though. First, Social Security is defined by the government. [+36 words]

And a tontine would be defined by your bank. [+79 words]

A tontine is a pretty crappy retirement vehicle for most people. It pays out the least when you need the most, and the most when you need the least.

People’s income needs in retirement generally fall as they age. [...+35 words]

[URL]

The Social Security system is basically one giant Tontine. So it's already legal.

There are some key differences though. First, Social Security is defined by the government. [+36 words]

And a tontine would be defined by your bank. [+79 words]

Very interesting. I'll give a Δ because I didn't have any idea that was true and changes my idea of how the tontine should work. That said, I don't think it's unsolvable. [+44 words]

DeltaBot] Confirmed: 1 delta awarded to [red]
The Social effects of linguistic Subtleties

Today's data → opportunity to discover and better understand such effects

Challenges:

* maintaining the controlled, hypothesis-driven nature of traditional studies
  - sense (and luck) to find the right data
  - taming wild data: art to setting up the right comparisons
The Social effects of linguistic subtleties

Today's data → opportunity to discover and better understand such effects

Challenges:

* maintaining the controlled, hypothesis-driven nature of traditional studies
  > sense (and luck) to find the right data
  > taming wild data: art to setting up the right comparisons

* need to develop/adapt computational tools
Case study: catchy language

(Some) people craft (some) political and ad slogans, news items, song lyrics, etc. to achieve cultural penetration.

A depressing possibility: does content actually matter, on average?

Case study: catchy language

(Some) people craft (some) political and ad slogans, news items, song lyrics, etc. to achieve cultural penetration.

A depressing possibility: does content actually matter, on average?

Movie quotes: massively, permanently viral
Research question: Does phrasing affect memorability?

Obi-Wan: You don't need to see his identification.
Stormtrooper: [ditto]
Obi-Wan: These aren't the droids you're looking for.
Stormtrooper: [ditto]
Obi-Wan: He can go about his business.
Stormtrooper: [ditto]
Obi-Wan: Move along.
Stormtrooper: [ditto]
Bye, daddy. I hope you find the droids you’re looking for.

These aren’t the droids you’re looking for.

These aren’t the droids you’re looking for.

These aren’t the droids you’re looking for.

These aren't the droids you're looking for.
"WE DO OTHER THINGS
BESIDES LOOK FOR
DROIDS.
BUT THAT'S ALL
ANYONE EVER
REMEMBERS."

Bye, daddy. I hope you find the droids you're looking for.

These aren't the droids you're looking for.
Obi-Wan: You don't need to see his identification.
Stormtrooper: [ditto]
Obi-Wan: These aren't the droids you're looking for.
Stormtrooper: [ditto]
Obi-Wan: He can go about his business.
Stormtrooper: [ditto]
Obi-Wan: Move along.
Stormtrooper: [ditto]

Data: Movie Scripts with memorability labels (IMDB)


RQ: Does phrasing affect memorability?
RQ: Does phrasing affect memorability?

Possible prediction setting:

memorable quotes vs. all the rest

Obi-Wan: You don't need to see his identification.
Stormtrooper: [ditto]

Obi-Wan: These aren't the droids you're looking for.
Stormtrooper: [ditto]

Obi-Wan: He can go about his business.
Stormtrooper: [ditto]

Obi-Wan: Move along.
Stormtrooper: [ditto]
RQ: Does phrasing affect memorability?

memorable quotes vs. all the rest
confounds:
memorable movies (e.g., Star Wars)
memorable characters (e.g., Obi-Wan)
memorable positions (e.g., last line of a movie)
length (shorter are easier to remember)

Obi-Wan: You don't need to see his identification.
Stormtrooper: [ditto]
Obi-Wan: These aren't the droids you're looking for.
Stormtrooper: [ditto]
Obi-Wan: He can go about his business.
Stormtrooper: [ditto]
Obi-Wan: Move along.
Stormtrooper: [ditto]

Obi-Wan: You don't need to see his identification.
Stormtrooper: [ditto]
Obi-Wan: These aren't the droids you're looking for.
Stormtrooper: [ditto]
Obi-Wan: He can go about his business.
Stormtrooper: [ditto]
Obi-Wan: Move along.
Stormtrooper: [ditto]

RQ: Does phrasing affect memorability?

Controlled Setting

Match each memorable quote with a non-memorable quote.
Obi-Wan: You don't need to see his identification.
Stormtrooper: [ditto]
Obi-Wan: These aren't the droids you're looking for.
Stormtrooper: [ditto]
Obi-Wan: He can go about his business.
Stormtrooper: [ditto]
Obi-Wan: Move along.
Stormtrooper: [ditto]

RQ: Does phrasing affect memorability?

Controlled Setting
Match each memorable quote with a non-memorable quote from the same character, same place in the movie, same length... to focus on the effect of phrasing.

Research question: Does phrasing affect memorability?

Gain intuition: Look at the data

Obi-Wan: You don't need to see his identification.
Stormtrooper: [ditto]
Obi-Wan: These aren't the droids you're looking for.
Stormtrooper: [ditto]
Obi-Wan: He can go about his business.
Stormtrooper: [ditto]
Obi-Wan: Move along.
Stormtrooper: [ditto]

**Research question:** Does phrasing affect memorability?

<table>
<thead>
<tr>
<th>First quote</th>
<th>Second quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half a million dollars will always be missed</td>
<td>I know the type, trust me on this.</td>
</tr>
</tbody>
</table>

**Gain intuition:** Look at the data
Research question: Does phrasing affect memorability?

<table>
<thead>
<tr>
<th>First quote</th>
<th>Second quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half a million dollars will always be missed</td>
<td>I know the type, trust me on this.</td>
</tr>
<tr>
<td>I think it’s time to try some unsafe velocities.</td>
<td>No cold feet, or any other parts of our anatomy.</td>
</tr>
</tbody>
</table>
**Research question:** Does phrasing affect memorability?

<table>
<thead>
<tr>
<th>First quote</th>
<th>Second quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half a million dollars will always be missed</td>
<td>I know the type, trust me on this.</td>
</tr>
<tr>
<td>I think it’s time to try some unsafe velocities.</td>
<td>No cold feet, or any other parts of our anatomy.</td>
</tr>
<tr>
<td>A little advice about feelings kiddo; don’t expect it always</td>
<td>I mean there’s someone besides your mother you’ve got to forgive.</td>
</tr>
<tr>
<td>to tickle.</td>
<td></td>
</tr>
</tbody>
</table>
Research question: Does phrasing affect memorability?

<table>
<thead>
<tr>
<th>First quote</th>
<th>Second quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half a million dollars will always be missed</td>
<td>I know the type, trust me on this.</td>
</tr>
<tr>
<td>I think it’s time to try some unsafe velocities.</td>
<td>No cold feet, or any other parts of our anatomy.</td>
</tr>
<tr>
<td>A little advice about feelings kiddo; don’t expect it always to tickle.</td>
<td>I mean there’s someone besides your mother you’ve got to forgive.</td>
</tr>
</tbody>
</table>

Gain intuition: Look at the data

- **50%** Impossible (no phrasing effects, bad labels, etc.)
- **Humans: 72-78%**
- **100%** trivial
Hypothesis: Surprising combinations of words are memorable
Hypothesis: Surprising language is memorable

Technique:

- measure surprisingness using language models

Toolkits: KenLM, MIT LM Toolkit, SRILM

Creative part:

A) Where to train the language model
   i.e., "Surprising with respect to what?"

B) How to represent the language?
Hypothesis: Surprising language is memorable

Technique:

measure surprisingness using language models

Toolkits: KenLM, MIT LM Toolkit, SRILM

Creative part:

A) where to train the language model
   i.e., "Surprising with respect to what?"

B) How to represent the language?

Here:

A) Train on fiction that pre-dates the movies (to avoid contamination)
**Hypothesis:** Surprising language is memorable

**Technique:**

- measure surprisingness using language models

**Toolkits:** KenLM, MIT LM Toolkit, SRILM

**Creative part:**

A) Where to train the language model
   
   i.e., "Surprising with respect to what?"

B) How to represent the language?

**Here:**

B) represent language as sequence of words

→ surprising combinations of words are more memorable

   e.g., "I see dead people."
Hypothesis: Surprising language is memorable

Technique:

measure surprisingness using language models

Toolkits: KenLM, MIT LM Toolkit, SRILM

Creative part:

A) Where to train the language model
   i.e., "Surprising with respect to what?"

B) How to represent the language?

Here:

B) represent language as sequence of parts of speech
   → common syntax is more memorable
   e.g., "You’re gonna need a bigger boat" vs. "You’re gonna need a boat that is bigger"
Fitness and diffusion of cultural content (memes)

"Meme-tracking" Leskovec, Backstrom, Kleinberg. 2009

"Memes online" Simmons, Adamic, Adar. 2011

"What's in a name" Himabindu, McAuley, LeSkovec. 2013

"QUOTUS" Niculae, Suen, Zhang, Danescu-Niculescu-Mizil, LeSkovec. 2015
Another quick LM case study: gender bias in sports journalism [Fu et al. 2016]

Inspired by covertheathlete.com
Hypothesis: questions to female players are less about the game
Hypothesis: questions to female players are less about the game

Technique:
measure surprisingness using language models

Hypothesis (rewritten in terms of surprise):
questions to female players are more surprising wrt game language

Creative part:
Where to train the language model?
→ play-by-play game commentary
Language (models) capturing user-community dynamics

Joining
Language (models) capturing user-community dynamics

Joining  Abandoning
Main intuition: linguistic change

Language norms
  - build collective identity
  - foster individual expression

Linguistic change allows us to capture
  - relation between members and their community

"No country for old members" (Danescu-Niculescu-Mizil et al., 2013)
Longitudinal data

Complete linguistic record of three online communities:
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less surprising over time
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less **surprising** over time
Main intuition: linguistic change

Intuition check:
Norms form online: *Language* becomes less *surprising* over time
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less surprising over time

Entropy:

$$H(\hat{\theta}) = \sum_i \theta_i \log \frac{1}{\theta_i}, \quad \theta_i = P(string_i)$$
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less surprising over time

Entropy:

$$H(\hat{\theta}) = \sum_i \theta_i \log \frac{1}{\theta_i}, \quad \theta_i = P(string_i)$$

surprise to see string
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less surprising over time

Entropy:

\[ H(\hat{\theta}) = \sum_i \theta_i \log \frac{1}{\theta_i}, \quad \theta_i = P(string_i) \]

prob. [surprise] to see string
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less surprising over time

Entropy: expected surprise in a language

\[ H(\hat{\theta}) = \sum_i \theta_i \log \frac{1}{\theta_i}, \quad \theta_i = P(string_i) \]

prob. [surprise] to see string
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less surprising over time

Entropy: expected surprise in a language

Entropy according to "Snapshot" language model of January 2008
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less surprising over time

Entropy: expected surprise in a language

\[ H(\hat{\boldsymbol{\theta}}) = \sum_i \theta_i \log \frac{1}{\theta_i}, \quad \theta_i = P(string_i) \]

prob. [surprise] to see string
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less surprising over time

Entropy: expected surprise in a language
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less surprising over time

Entropy: expected surprise in a language

Alternative explanation
Main intuition: linguistic change

Intuition check:
Norms form online: Language becomes less surprising over time

Entropy: expected surprise in a language

Alternative explanation
as community size grows, LM is more informed, so harder to surprise
Main intuition: linguistic change

Intuition check:
Norms take time to learn: Newcomers start farther away
Main intuition: linguistic change

Intuition check:
Norms take time to learn: Newcomers start farther away
Main intuition: linguistic change

Intuition check:
Norms take time to learn: Newcomers start farther away

Cross-Entropy: expected surprise given a "known" language
Main intuition: linguistic change

Intuition check:
Norms take time to learn: Newcomers start farther away

Cross-Entropy: expected surprise given a "known" language

\[ H(\hat{\theta}, \varphi) = \sum_i \varphi_i \log \frac{1}{\theta_i}, \]

\[ \theta_i = P(string_i \text{ in "known" language}) \]

\[ \varphi_i = P(string_i \text{ in "new" language}) \]
Main intuition: linguistic change

Intuition check: Norms take time to learn: Newcomers start farther away

Cross-Entropy: expected surprise given a "known" language
Main intuition: linguistic change

Main results: "No country for old members" (DaneScu-NiculeScu-Mizil et al., 2013)
Main intuition: linguistic change

Main results: "No country for old members" (Danescu-Niculescu-Mizil et al., 2013)
Main intuition: linguistic change

Main results: "No country for old members" (Danescu-Niculescu-Mizil et al., 2013)

Stage 1:
user assimilates the language of the community
Main intuition: linguistic change

Main results: "No country for old members" (Danescu-Niculescu-Mizil et al., 2013)

Stage 1:
user assimilates the language of the community

Stage 2:
user's language distances itself from that of the community
Language change and Social dynamics

Other cool work (links & more on website):

"Social Dynamics of Language Change."

Regional dialects - Eisenstein. 2014

Geographic variation - Kulkarni, Perozzi, Skiena. 2016

What makes two “languages” different?

Presentation/figures follow Monroe, Colaresi and Quinn, *Political Analysis* (2008)
Persuasion: *frame* competition

Example: public discussion of GMOs in food

The *framing* of an argument emphasizes certain principles or perspectives. “One of the most important concepts in the study of public opinion”

James Druckman (2001)

Example: 106th U.S. Senate speeches on abortion

Frames we might expect from Democrats:

... women’s rights ...
... privacy ...

Frames we might expect from Republicans:

... unborn children ...
... murder ...

Assume a joint vocabulary of terms $v_i$. $p(v_i)$ and $p(v_i)$: relative frequency of $v_i$ in the blue and red samples
Ranking using $P(x|class)$

Top and bottom 20 words according to

$$p(v_i) - p(v_i)$$

important, but would be lost with stopword filtering
Aside: “stopword removal” not recommended

• Very-frequent terms have been proving “increasingly” useful, e.g., for stylistic or psychological cues

• “a” vs “the” is surprising

[for years LL assumed this was a bug, but see Language Log, Jan 3 2016]
\(P(x|\text{class})\) vs. count

\(p(v_i) - p(v_i)\) favors big counts, i.e., \(v_i\) towards the righthand side of this plot

(can’t have a large difference between two small differences)
Ranking by log odds-ratio

\[ \log \frac{p(v_i)/(1 - p(v_i))}{p(v_i)/(1 - p(v_i))} \]
Ranking by z-score of log odds-ratio, with model of variance (uniform prior)
Additional applications: Differentiating the language of ....

• **successful** vs. **unsuccessful** persuaders
• **low-status** vs. **high-status** people ...
• **males** vs **females**
• *your experimental condition A* vs. *your experimental condition B!!*

Also good for sanity-checking your data...
[The Duchess said,] `You’re thinking about something, my dear, and that makes you forget to talk. I can't tell you just now what the moral of that is, but I shall remember it in a bit.'

`Perhaps it hasn't one,' Alice ventured to remark.

`Tut, tut, child!' said the Duchess. `Everything's got a moral, if only you can find it.'
Morals you *shouldn’t* conclude (we only had two hours together...)

- “More sophisticated NLP isn’t used (or doesn’t work) for computational social science.”
  - example: topic models for differentiating language samples (Blei, Ng, Jordan 2003)
  - example: syntactic correlates of gender differences (Sarawgi, Gajulapalli and Choi 2011)
  - example: discourse modeling of conversational flow

- “We now know all the interesting problems and work there are in computational social science.”
  - not even close! (And that’s not even counting ethics, fairness, and bias questions...)
Pointers to resources

This tutorial was based on our Cornell course
“Natural Language Processing and Social Interaction”.

For links to papers, conferences, datasets, toolkits, research ideas:
http://www.cs.cornell.edu/courses/cs6742/ - most recent run (5 so far)

Add one of {2011fa,2013fa,2014fa,2015fa,2016fa} to URL to get that semester;
http://www.cs.cornell.edu/courses/cs6742/2014fa has scanned lecture notes

More datasets:
http://www.cs.cornell.edu/home/llee/data/index.html
TODAY:

... Research questions
  persuasion, linguistic change, framing

... Techniques
  language models, Bayesian feature analysis

... Research practices
  controls, feasibility, data inspection
LOOKING FORWARD:

Deeper interplay between natural language processing and how people use and are affected by language is a huge opportunity for all concerned.
I think this is the beginning of a beautiful friendship.

Thanks!