Language as influence(d): Power and memorability

Lillian Lee
Cornell University
http://www.cs.cornell.edu/home/llee

Given at the Institute for Quantitative Theory and Methods, Emory University, 2012
(Not) news: “Big Data” & social science

massive amounts of data
(Probably not) news:
“Big Language Data” & A.I.

The New York Times

August 11, 2012

How Big

By STEVE LOHR

THIS has been the year of Big Data.

First, here are a few notes on the buzzword.

In March, the federal government launched a

Rick Smolan, creator of the term in headlines like the

“Dilbert” comic strip

The Big Data story is also an evocative one. The second

Big Data is a shorthand for the new wave of data

databases. The new data sources include Web-browsing data trails, social network commu

The combination of the data deluge and clever software algorithms opens the door to new

The Watson computer from I.B.M. that beat human “Jeopardy” champions last year was a

fields from business to medicine, allowing decisions to be based increasingly on data and a

“The term itself is vague, but it is getting at something that is real,” says Jon Kleinberg, a

the potential to transform everything.”

Rising piles of data have long been a challenge. In the late 19th century, census takers stru

An innovative breakthrough came in time for the 1890 census, when the population reached

invented by Herman Hollerith; these cards were the bedrock technology of the company t
Two great tastes that taste great together

Lots of language data
(documents; conversations, e.g., Facebook posts, YouTube comments, Yelp reviews) =

Lots of human-behavior samples & descriptions =

Fantastic opportunities for natural-language processing (NLP) +

the social sciences
Beyond the hype

“Big Language Data” offers the “greater resolution” needed to pick up on subtle language characteristics.
Part I: Influence → Language

Can subtle, *domain-independent* linguistic cues reveal (situational) power?

Who's in charge?
Part II: Language ➔ Influence

Does phrasing affect memorability, or is it all language-external factors?

you can put lipstick on a pig, but it’s still a pig…
you can wrap an old fish in a piece of paper called change, it’s still going to stink…

Palin’s lipstick remark

Can wording also have an effect?
Aside: on presentation style

Your goal is not to convince your audience that you are brilliant, but that your solution is trivial.

It takes a certain strength of character to take that as one's goal.

But if people think your findings are obvious, they must also believe that you are correct.

-- paraphrase of Stuart Shieber
Echoes of Power:

Language effects & power differences in social interaction

Joint with Cristian Danescu-Niculescu-Mizil, Bo Pang, & Jon Kleinberg
WWW 2012
Language reveals power: “easy” cases

Your Honor, I agree.

Thank you.

I’d love to get your thoughts on this when you are available.

Let’s discuss later.

[Gilbert 2012; Diehl et al. 2007, Scholand et al. 2010]

What about general (domain-independent) signals?
Who has the (conversational) lead?

Communicative behaviors are “patterned and coordinated, like a dance” [Niederhoffer and Pennebaker 2002]

Look for adaptation/coordination of linguistic style
Defining linguistic style coordination

Direct repetition: under speaker’s control, could just be choice of topic.

Function-class matching: unconscious & frequent
[Niederhoffer and Pennebaker 2002]

At least you were outside.

It doesn’t make much difference…

Doesn’t really matter …

It’s not important …

quantity words
blue: prepositions; red: articles; etc.
Measuring immediate influence

How much does speaker $x_1$ immediately trigger $x_2$’s use of function-word class $c$?

= how much does $x_2$ coordinate to $x_1$ on $c$?

$Pr(x_2 \text{ uses } c \mid x_1 \text{ uses } c, x_2 \text{ immediately replies})$

$- Pr(x_2 \text{ uses } c \mid x_2 \text{ immediately replies to } x_1)$

(reporting as % (multiplied by 100) in the following)

[Danescu-Niculescu-Mizil, Dumais, Gamon 2011]
Status in US Supreme Court transcripts
50,000 exchanges + metadata (download from my webpage)
Status change in Wikipedia

240,000 exchanges + metadata (download from my webpage)
Dependence in Supreme Court transcripts

Power differences can arise from dependence [Emerson 1962]

“needed to change” J’s mind

J “had your opinion”
## Classification results

<table>
<thead>
<tr>
<th>Training corpus</th>
<th>Test corpus</th>
<th>in-domain</th>
<th>cross-domain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wiki</td>
<td>court</td>
<td>court</td>
</tr>
<tr>
<td>coordination features</td>
<td>57.7</td>
<td>70.4</td>
<td>57.1</td>
</tr>
<tr>
<td>(9 altogether)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stylistic features</td>
<td>59.2</td>
<td>51.4</td>
<td>50.0</td>
</tr>
<tr>
<td>(18 altogether)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bag of words</td>
<td>51.4</td>
<td>99.5</td>
<td>45.2</td>
</tr>
<tr>
<td>(20,000 altogether)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Prediction accuracy for SVM’s using various feature sets. Cross-domain performance is in the right-hand pair of columns. Bold: results significantly better than chance.
You Had Me at Hello:
How phrasing affects memorability

Joint with Cristian Danescu-Niculescu-Mizil, Justin Cheng, & Jon Kleinberg
ACL 2012

you can put lipstick on a pig, but it’s still a pig…
you can wrap an old fish in a piece of paper…
Does phrasing affect memorability?

Implications: understanding influence, information diffusion; advertising and persuasion

- What kind of data?
- What counts as “memorable?”
- How to tease apart language from context? (the lipstick-on-a-pig problem)
Movie quotes: massively viral

"FRANKLY, MY DEAR, I DON'T GIVE A DAMN" TOPS AFI'S LIST OF 100 GREATEST MOVIE QUOTES OF ALL TIME

OTHER WINNERS INCLUDE:

THE GODFATHER, "I'M GOING TO MAKE HIM AN OFFER HE CAN'T REFUSE"

THE WIZARD OF OZ, "TOTO, I'VE GOT A FEELING WE'RE NOT IN KANSAS ANYMORE"

AND CASABLANCA, "HERE'S LOOKING AT YOU, KID"

AFI'S 100 Years...100 Movie Quotes: America's Greatest Quips, Comebacks and Catchphrases

LOS ANGELES, June 22, 2005 — The American Film Institute revealed the top movie quotes of all time in AFI's 100 Years...100 Movie Quotes, a three-hour special television event on CBS hosted by actor and action star Pierce Brosnan with commentary from many of Hollywood's most celebrated actors and filmmakers. A jury of 1,500 film artists, critics and historians selected "Frankly, my dear, I don't give a damn," spoken by Clark Gable in the celebrated Civil War epic, GONE WITH THE WIND as the most memorable movie quote of all time.
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]
"WE DO OTHER THINGS
BESIDES LOOK FOR
DROIDS.
BUT THAT'S ALL
ANYONE EVER
REMEMBERS."
Data

From ~1000 movie scripts (many lines long), pair IMDB “memorable quotes” with ~adjacent, same-length, same-speaker “non-memorable quotes”. Filter with google/bing counts: 2200 pairs.

Available off my homepage.
Pilot study

Subjects were shown 12 pairs from movies they *hadn’t seen*.  
[http://www.cs.cornell.edu/~cristian/memorability.html](http://www.cs.cornell.edu/~cristian/memorability.html)

<table>
<thead>
<tr>
<th>Movie</th>
<th>First Quote</th>
<th>Second Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jackie Brown</td>
<td>Half a million dollars will always be missed.</td>
<td>I know the type, trust me on this.</td>
</tr>
<tr>
<td>Star Trek: Nemesis</td>
<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>Ordinary People</td>
<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>

(trivial)  

50%  

impossible  

72-78%  

interesting task  

100%  

trivial  

(memo labels are bad, task too subjective, etc.)
On average, memorable quotes (significantly)…

… contain more surprising combinations of words:
    according to 1-,2-,3-gram lexical language models trained on the Brown corpus

    “…aren’t the droids…”

… are built on a more common syntactic scaffolding:
    according to 1-,2-,3-gram part-of-speech language models trained on Brown

    “You’re gonna need a bigger boat”
    [vs. “You’re gonna need a boat that is bigger”]
On average, memorable quotes (significantly)...

... are more portable/general:
and so can be used in different contexts

- fewer 3\textsuperscript{rd} person pronouns ("he" vs "we"), more indefinite articles ("a" vs. "the")
- fewer past tense verbs and more present tense verbs

"You’re gonna need a bigger boat"
[not “You’re gonna need the bigger boat"]
## Classification results

<table>
<thead>
<tr>
<th>Feature set</th>
<th># of features</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bag of words</td>
<td>962</td>
<td>59.67%</td>
</tr>
<tr>
<td>distinctiveness</td>
<td>24</td>
<td>62.05%</td>
</tr>
<tr>
<td>generality</td>
<td>4</td>
<td>56.70%</td>
</tr>
<tr>
<td>dist. + gen. features</td>
<td>28</td>
<td><strong>65.27%</strong></td>
</tr>
</tbody>
</table>
Generalizing to ad slogans

Ad slogans are closer to memorable quotes in lexical and syntactic choice:

Memorable-quote LMs predict slogans better than non-memorable-quote LMs

... and in portability:

% 3rd pers. pronoun: 0%
% indefinite articles: 5%
% past tense: 30%
Two great tastes that taste great together

Lots of language data
(documents; conversations, e.g. Facebook posts, YouTube comments, Yelp reviews) =
Lots of human-behavior samples & descriptions =
Fantastic opportunities for natural-language processing (NLP) +
the social sciences
I think this is the beginning of a beautiful friendship.

Thanks!