

~~Is it all~~ *What's* in the phrasing?  
**Computational explorations in  
how we say what we say,  
and why it matters**

**Lillian Lee**

Dept. of Computer Science, Cornell University

<http://www.cs.cornell.edu/home/llee>

**For persuasion, influence, and diffusion,  
(how much) does wording matter?**

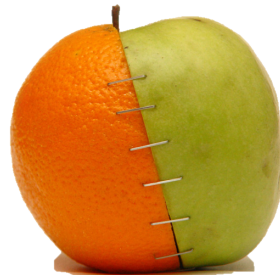
# One aspect: framing

Amber Boydstun's talk

Philip Resnik's talk

The *framing* of arguments emphasizes certain principles or perspectives.

"green revolution"



"Frankenfood"

[Eunsol Choi, Chenhao Tan, Lillian Lee, Cristian Danescu-Niculescu-Mizil, Jennifer Spindel 2012]

It's "one of the most important concepts in the study of public opinion" [James Druckman, 2001]

## But perhaps phrasing doesn't help at all?

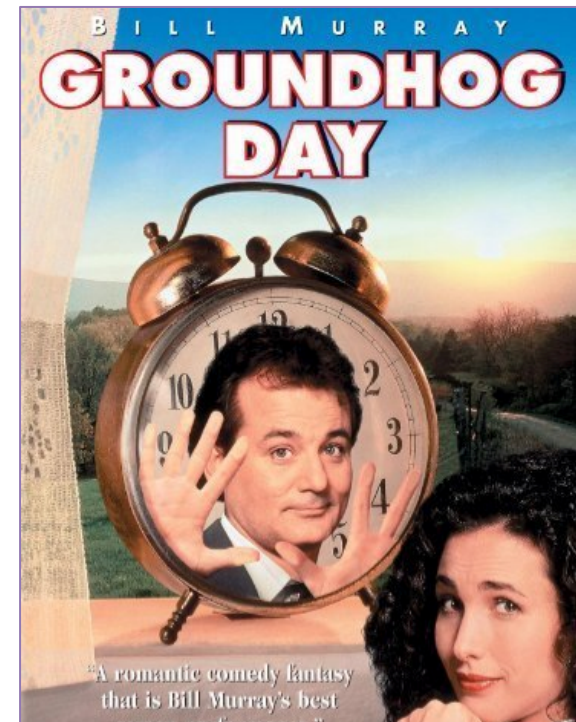
1. Daniel Hopkins, SSRN 2013: " there is no evidence that groups targeted by specific frames [such as "death panels" in the health care debates] respond accordingly."
2. **Justin Grimmer**, Solomon Messing, Sean Westwood, APSR '12: *total number of messages* mattered more than amount the messages described.
3. One of Sasa Petrovic, Miles Osborne, and Victor Lavrenko, slashdot comment, '14: "It is all about the followers [in] predicting when a tweet would be retweeted ... Basically, a famous person can write anything and it will be retweeted. An unknown person can write the same tweet and it will be ignored."

# Controlling to focus on effect of phrasing

## *Natural experiment* paradigm:

Exploit situations with *many* instances of:

- the same speaker
  - in the same situation, or conveying the same info...
  - varies their wording (beyond a fixed set of lexical choices)
- and see the effects.



<http://www.imdb.com/media/rm2963188736/tt0107048/>

# Case study: information diffusion

Surprisingly many Tweeters post about the *same* URL but varying their "sales pitch", with significantly different retweet results



cactus\_music

I know at some point you've have been saved from hunger by our rolling food trucks friends. Let's help support them! [bit.ly/P6GYCq](http://bit.ly/P6GYCq)



cactus\_music

Food trucks are the epitome of small independently owned LOCAL businesses! Help keep them going! Sign the petition [bit.ly/P6GYCq](http://bit.ly/P6GYCq)



GeorgeMonbiot

read @ameliagentleman's report today, then tell me Tories are no longer the nasty party: [guardian.co.uk/society/2012/o...](http://guardian.co.uk/society/2012/o...)



GeorgeMonbiot

Work capability tests: designed by bastards, performed by idiots. [guardian.co.uk/society/2012/o...](http://guardian.co.uk/society/2012/o...)

Try it! <http://chenhaot.com/retweetedmore/quiz>

# Example classification results

## Estimate of human accuracy:

(sample of 100 pairs; 106 judges; 39 judgments/pair)

61.3% per-human average,

62.4% per-response average

Our classifier on (much larger) truly\* held-out data: 65.6%

\*We ran only one experiment on it, at submission time

## Example feature results

Don't be too different from the community  
(a general Twitter language model).

*On average*, the more-retweeted version had a significantly higher score than its twin.

But also be true to yourself  
(a user-specific language model).

*On average*, the more-retweeted version had a significantly higher score than its twin.



## Case study: longer-term cultural penetration (application: slogans)



*Lines from the same scene.*

**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]

<http://www.blu-ray.com/movies/screenshot.php?movieid=14903&position=6>

# Example classification results

Movie	First Quote	Second Quote
Jackie Brown	Half a million dollars will always be missed.	I know the type, trust me on this.
Star Trek: Nemesis	I think it's time to try some unsafe velocities.	No cold feet, or any other parts of our anatomy.
Ordinary People	A little advice about feelings kiddo; don't expect it always to tickle.	I mean there's someone besides your mother you've got to forgive.

Estimate of human accuracy:

(6 judges or 11K judges)

72-78% per-human average

Our classifier (10-fold xval): 64.27%

***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”]

[Cristian Danescu-Niculescu-Mizil, Justin Cheng, Lillian Lee, Jon Kleinberg  
ACL 2012]

## Study of claim strength: motivating example

The US embassy initially referred to [attacks at Kunming] as: “**terrible and senseless act of violence in Kunming**”.

A weibo user: “If you say that the Kunming attack is a ‘terrible and senseless act of violence’, then **the 9/11 attack can be called a ‘regrettable traffic incident’**”

**After Prodding, U.S. State Department Labels Kunming Attack ‘Terrorism’**

By DIDI KIRSTEN TATLOW   MARCH 4, 2014 1:05 AM   19 Comments

# Another motivating example:

## An actual review

(post-talk comments: from 2007; not written by Graeme)


"The authors claim that they are addressing a document classification problem without using any prior linguistic knowledge - to which I am tempted to ask, what is this paper doing being submitted to the A C \*L\* conference?



Nonetheless I think the paper is a reasonable fit, especially since the technique actually does make use of several facts about language that are different from other sorts of data (photographs, etc.):


- Written English can be tokenized into meaningful 'words' at whitespace ..."

[much more follows]

# Aside: Facebook group "Reviewer 2 Must Be Stopped!"





**Reviewer 2 Must Be Stopped!**[Members](#)[Photos](#)[Join Group](#)[+ Create Group](#)



**Jian Raymond Rui**  
June 23 at 6:24pm · Buffalo, NY



Reviewer 2 said yes you found this interaction effect but I don't believe it. What can I say? Yes sir, I confess that I'm a liar.



[Like](#) · [Share](#)


 20 people like this.


 [View 4 more comments](#)


**Catherina Becker** could be ma'm too  
Yesterday at 6:36am · [Like](#) ·  2


**Anne Dickmeis** Jian Raymond Rui ... You seem to be in the exact domain I sent my first paper to at ICA (for others: this is NOT the one I talked about in this forum, which was in my second domain).  
It had a reviewer 2 who claimed his statistical skills were "well a... [See More](#)  
19 hours ago · Edited · [Like](#) ·  1


**Jian Raymond Rui** Anne Dickmeis ICA and NCA need to improve the quality of reviewers. I can't stand biased, irresponsible, even ridiculous comments on my studies. I'm just whining.  
14 hours ago · [Like](#) ·  1

**ABOUT**  
 **Open Group**  
This group is for those who believe Reviewer 2 must be stopped. In the name of science and all t... [See More](#)

**SUGGESTED GROUPS**  
**World Transhumanist Association**  
3,498 members  
[+ Join](#)

**Scientific Transhumanism**  
8,357 members  
[+ Join](#)

**Just Science**  
276 members  
[+ Join](#)

**Association for Contextual Behavior...**  
4,000 members



## Case study: strength revisions

On the arxiv e-print archive, authors post LaTeX source for different versions of the *same* paper, allowing for sentence alignment.

In order to overcome this inconsistency, an additional constraint due to the requirement of extensivity is needed in the maximization procedure.

Therefore, an additional constraint due to the requirement of extensivity is needed in the maximization procedure, leading to a novel generalized maximization procedure.

## Case study: strength revisions

On the arxiv e-print archive, authors post LaTeX source for different versions of the *same* paper, allowing for sentence alignment.

Circadian pattern and burstiness in  
human communication activity

Circadian pattern and burstiness in  
mobile phone communication.



## Example finding from human subjects: People can be swayed by specificity

S1: ... using data from numerics and experiments.

S2: ... using data sets from numerics in the point particle limit and one experimental data set.

Subject comment: “S2 is more specific in its description which seems stronger.”

# Looking forward

Deeper analysis of phrasing is an important and fascinating broader direction.

Datasets are available from my homepage.

Deeper interplay between

*NLP* 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!

# Additional refs from post-talk discussions

Bryan, Christopher J, Gregory M Walton, Todd Rogers, and Carol S Dweck. 2 August 2011. [Motivating voter turnout by invoking the self](#). *Proceedings of the National Academy of Sciences* 108 (31): 12653-12656.

Bailey, Michael, Daniel J. Hopkins, and Todd Rogers, SSRN, May 26, 2014. [Unresponsive and Unpersuaded: The Unintended Consequences of Voter Persuasion Efforts](#)

Hsee, Christopher K., Loewenstein, George F., Blount, Sally, Bazerman, Max H. [Preference Reversals Between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis](#)." 1999. *Psychological Bulletin* 125 (5): 576-590,

Rogers, Todd and Michael I Norton. June 2011. [The artful dodger: Answering the wrong question the right way](#). *Journal of Experimental Psychology: Applied* 17 (2): 139