

**Is it all in the phrasing?**  
**Computational explorations in**  
**how we say what we say**

**Lillian Lee**

Dept. of Computer Science, Cornell University

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

# Language Effects

The rise of user-generated content and social media offers:

*new opportunities to examine influence, information diffusion, and so on...*

*...and the role that language (phrasing) plays.*

(These are *long-standing* issues in rhetoric, communication, the sociology of language, marketing, etc.)

# Part 1 of 2: Memorability

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...

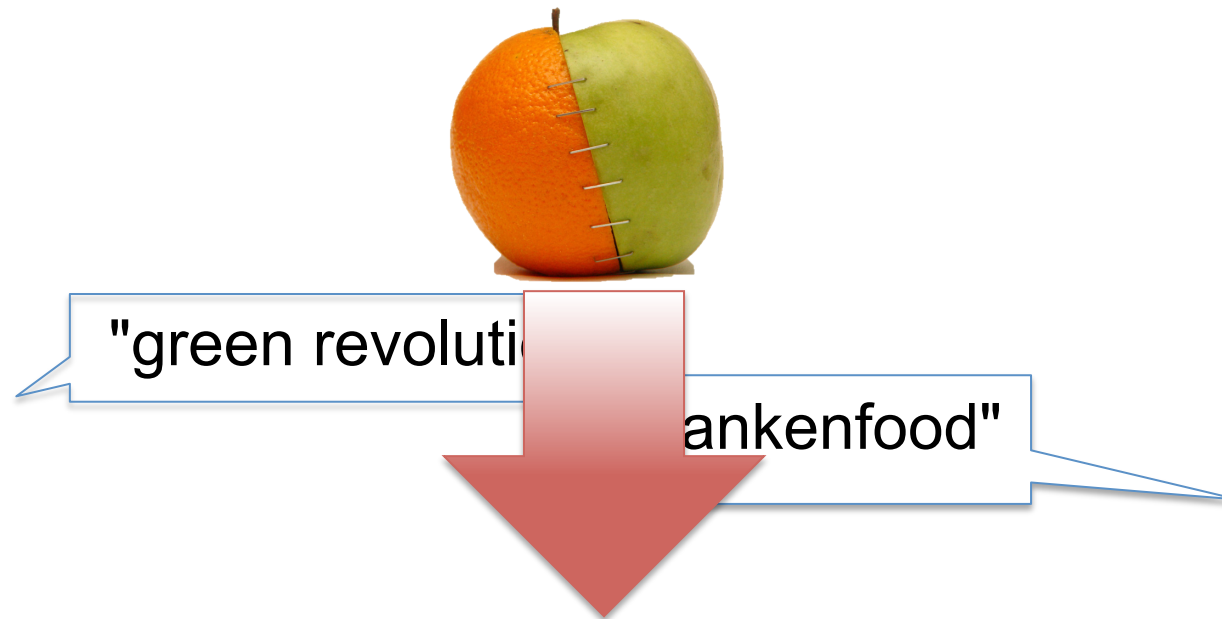
*Palin's lipstick remark*

*Can wording also  
have an effect?*



## Part 2 of 2: Framing

Can we characterize how different sides use phrasing to promote different principles?



New kind of question:

~~Do the two sides use distinctive vocabularies?~~

Does one side adopt a different *style* of discourse?

# 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

# 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 and information diffusion; advertising and persuasion.

[Much related work in many fields; see paper for refs. Our direct inspiration: Leskovec, Backstrom, & Kleinberg '09.]

- 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



<http://www.afi.com/100years/quotes.aspx>

**"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.



# Controlling for speaker and context



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]

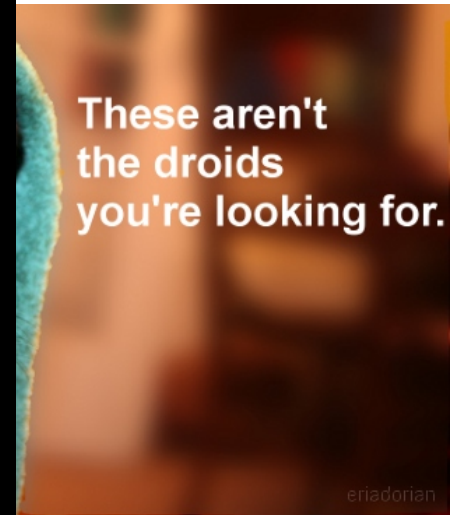
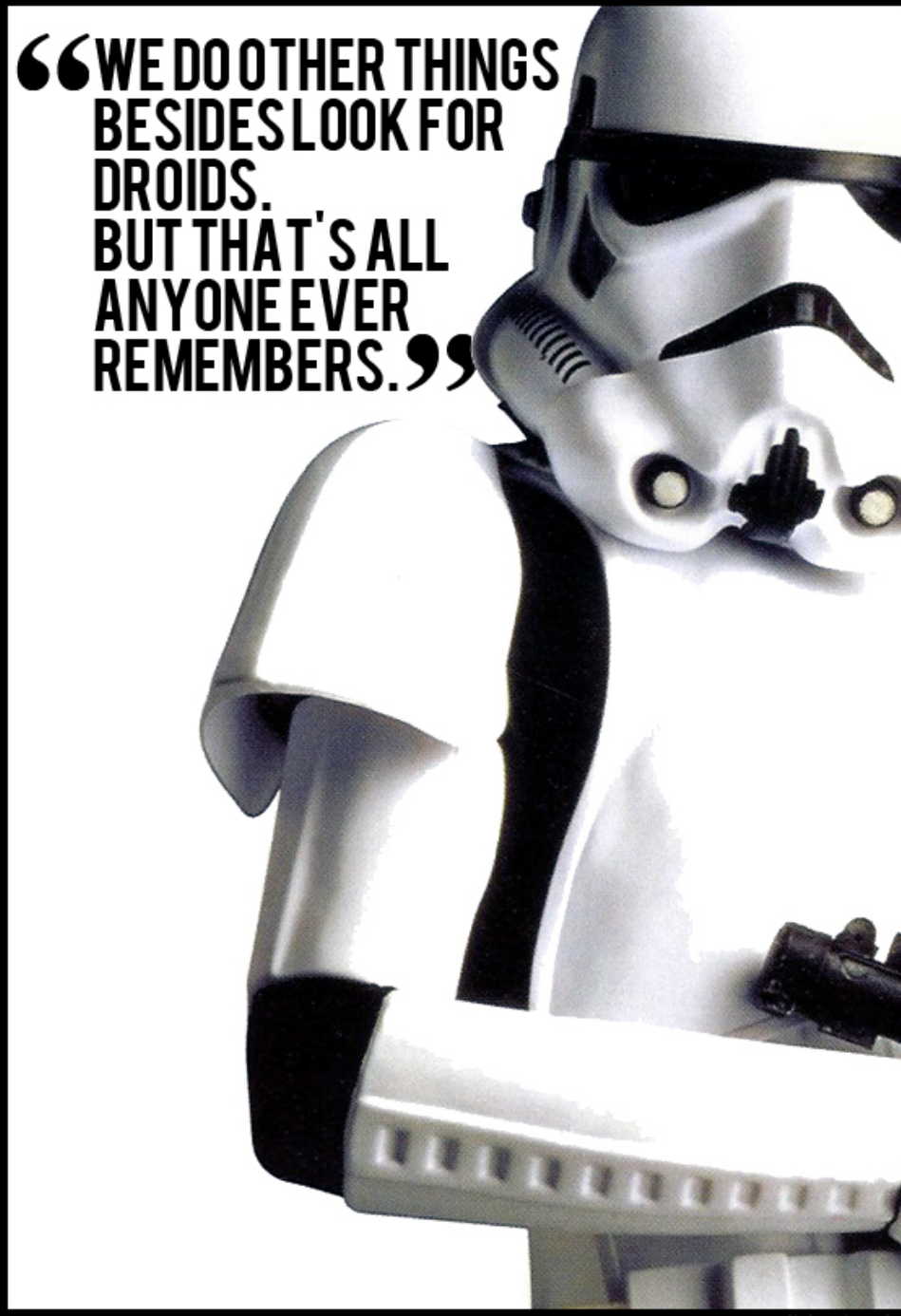
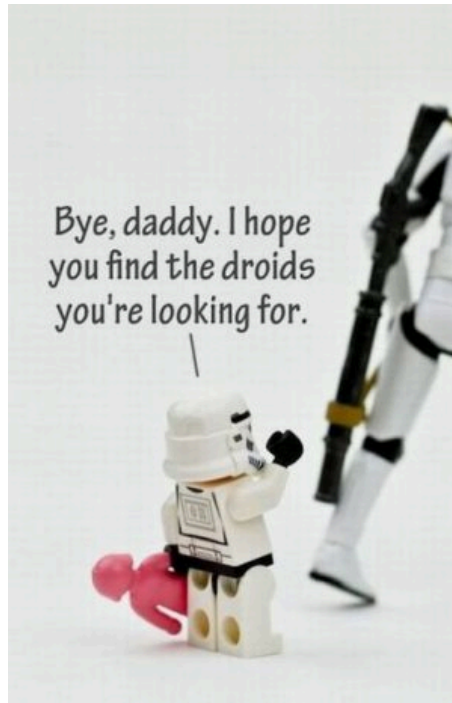
Obi-Wan: He can go about his business.

Stormtrooper: [ditto]

Obi-Wan: Move along.

Stormtrooper: [ditto]

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



<http://mikiedaniel.files.wordpress.com/2011/09/noopers.jpg>  
[http://ic.ics.livjournal.com/rev/\\_speak\\_47171608842/245440/245440\\_original.jpg](http://ic.ics.livjournal.com/rev/_speak_47171608842/245440/245440_original.jpg)  
<http://bloodyot.com/wp-content/uploads/2009/11/droids-we-were-looking-for1.jpg>

# 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 at

<http://www.mpi-sws.org/~cristian/memorability.html>

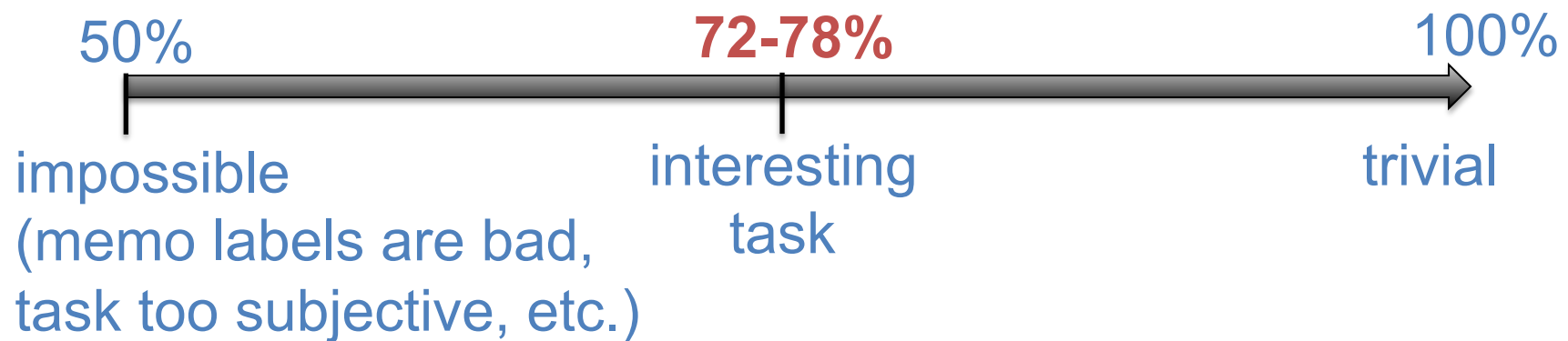


# Pilot study

Subjects were shown 12 pairs from movies they *hadn't* seen.

<http://www.mpi-sws.org/~cristian/memorability.html>

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.



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**:

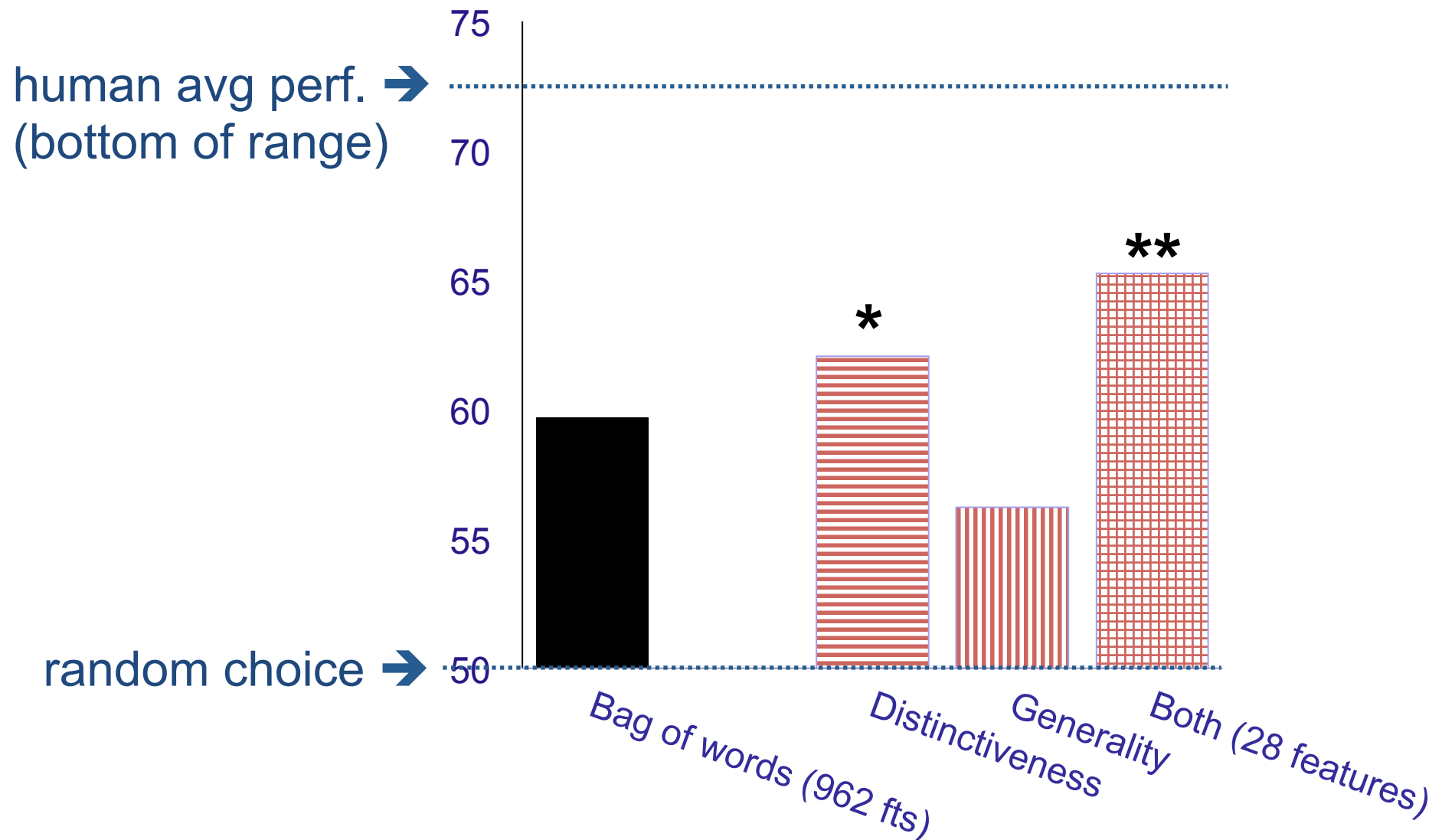
- more indefinite articles

*“You’re gonna need a bigger boat”*

[not *“You’re gonna need **the** bigger boat”*]

- fewer 3<sup>rd</sup> person pronouns (“he” vs “we”),
- fewer past tense verbs and more present tense verbs

# Improved classification accuracy (SVMs, 10-fold cross-validation)



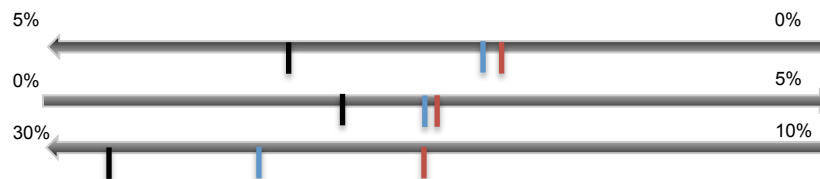
# Generalizing to ad slogans

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 generality:

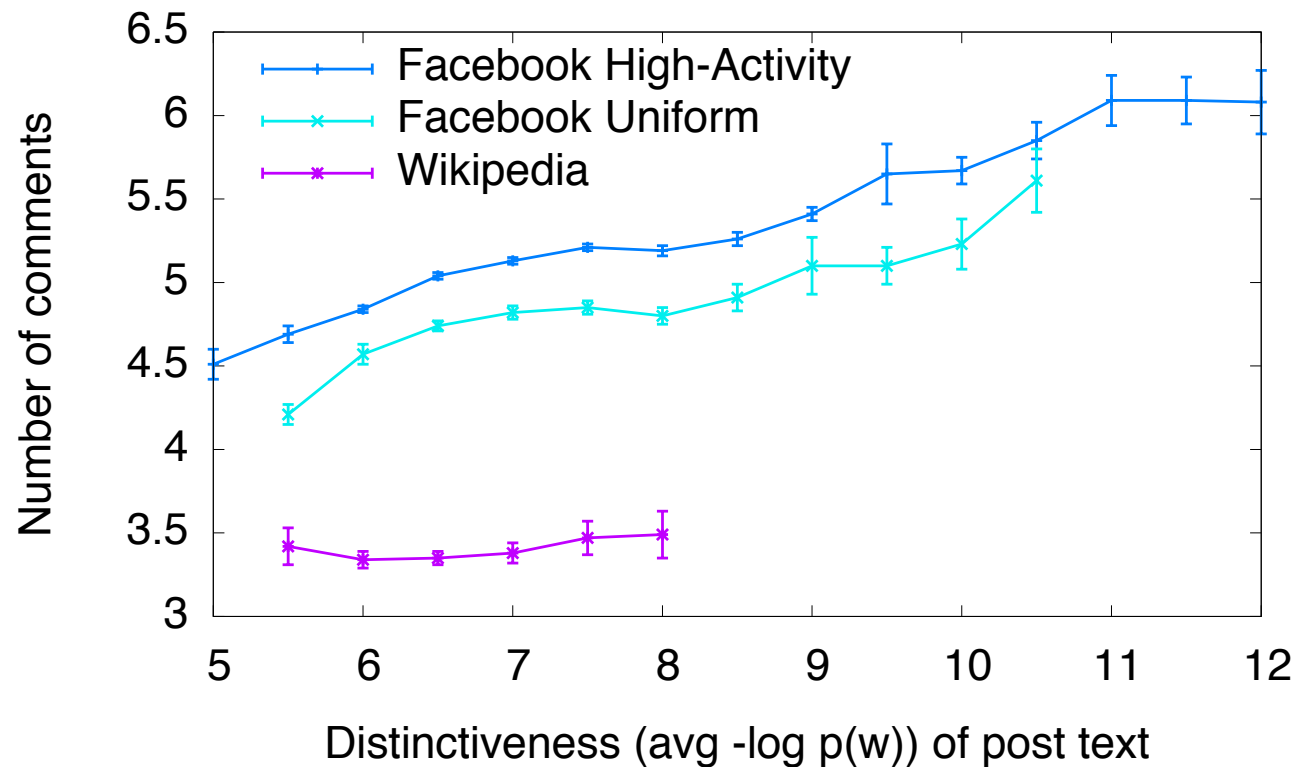
% 3rd pers. pronoun:  
% indefinite articles:  
% past tense:





# Applications to social-media UI

joint with L. Backstrom, J. Kleinberg, C. Danescu-Niculescu-Mizil, WSDM 2013



More-unusual Facebook posts get more comments  
(under certain circumstances), but not so with Wikipedia.

# Aside: re-entry prediction: A new "interestingness" characterization

## (1) Length/volume

## (2) Re-entry

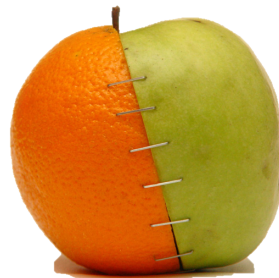


Image: Chase Mitchell, College Humor, 2010

# *Hedge detection as a lens on framing in the GMO debates*

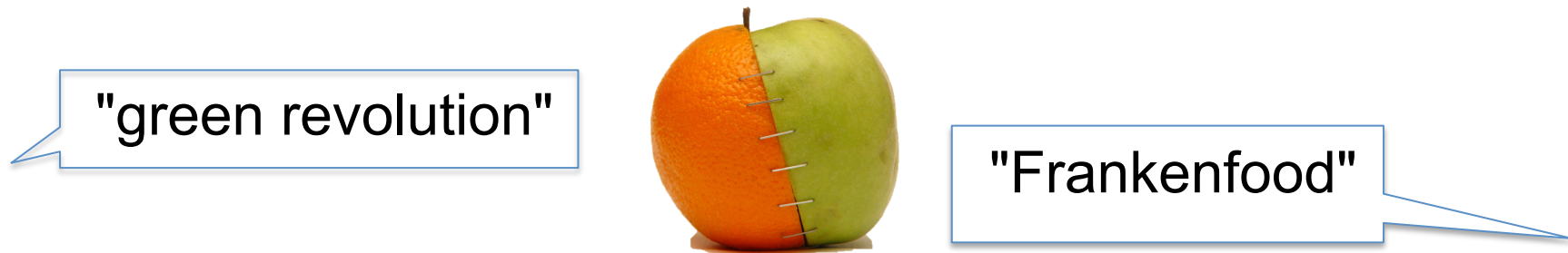
A pilot study and proposal

Joint with Eunsol Choi, Chenhao Tan, Cristian Danescu-Niculescu-Mizil, &  
Jennifer Spindel  
EXPROM 2012



# Framing effects

The *framing* of arguments to emphasize certain principles or perspectives is "one of the most important concepts in the study of public opinion" [Druckman 2001]



Prior work: Distinctive vocab to distinguish different perspectives [e.g., Lin, Wilson, Wiebe, Hauptmann '06]; syntactic packaging [Greene & Resnik '09]

## A challenge: framing via style

"...experiment ... on ... media coverage of transgenic ("GMO") crops... I have a hypothesis that the positive articles and negative articles not only use different words ... but **employ different discourses**, and that this has a subsequent effect on the reader... **I think the positive articles, even those from lay newspapers, employ a more scientific discourse (meaning *more uncertainty*, fewer emotionally-laden words, more jargon)"**

Jen Spindel, Ph.D. student in plant breeding



# Hedging as "scientific"

- *"It seems that* this group plays a critical role in orienting the carboxyl function"
- *"We wish to suggest* a structure for the salt deoxyribose nucleic acid (D.N.A.)"

But scholars of scientific rhetoric disagree on the *relative* degree of hedging in pro-science vs. pop-science. [see paper for refs]

Opportunity to settle debate in another field!

# Data and baselines available

- CoNLL 2010 shared task (a competition with shared annotated training and test data) was on ... hedge detection!

[We've also distributed in-domain data.]

- Preliminary annotation results: more uncertainty in pop-science GMO articles!
- Our re-implementation of a top CoNLL-2010 system did not successfully transfer to this new domain.
- This remains a nice opportunity (for someone else?) to do novel, deeper computational work on framing.

# Conclusions (I)

I had not yet come to the conviction:  
It is not what you say that matters  
but the manner in which you say it;  
there lies the secret of the ages."

--- William Carlos Williams,  
preface to *Selected Essays*

Phrasing *does* have some effect on  
memorability and engagement.

Further progress on the detection of hedging-  
type phrasing can inform current debate on  
framing (and other questions).



## Conclusions (II)

Deeper analysis of phrasing (broadly construed to reach as far as framing) is an important broader direction in NLP.

I view deeper interplay between *NLP* and *how people use and are affected by language* — such as in questions of *influence* — as a huge opportunity for all concerned.



**I think this is the beginning  
of a beautiful friendship.**

Thanks!