

# Using Text to Predict the Real World

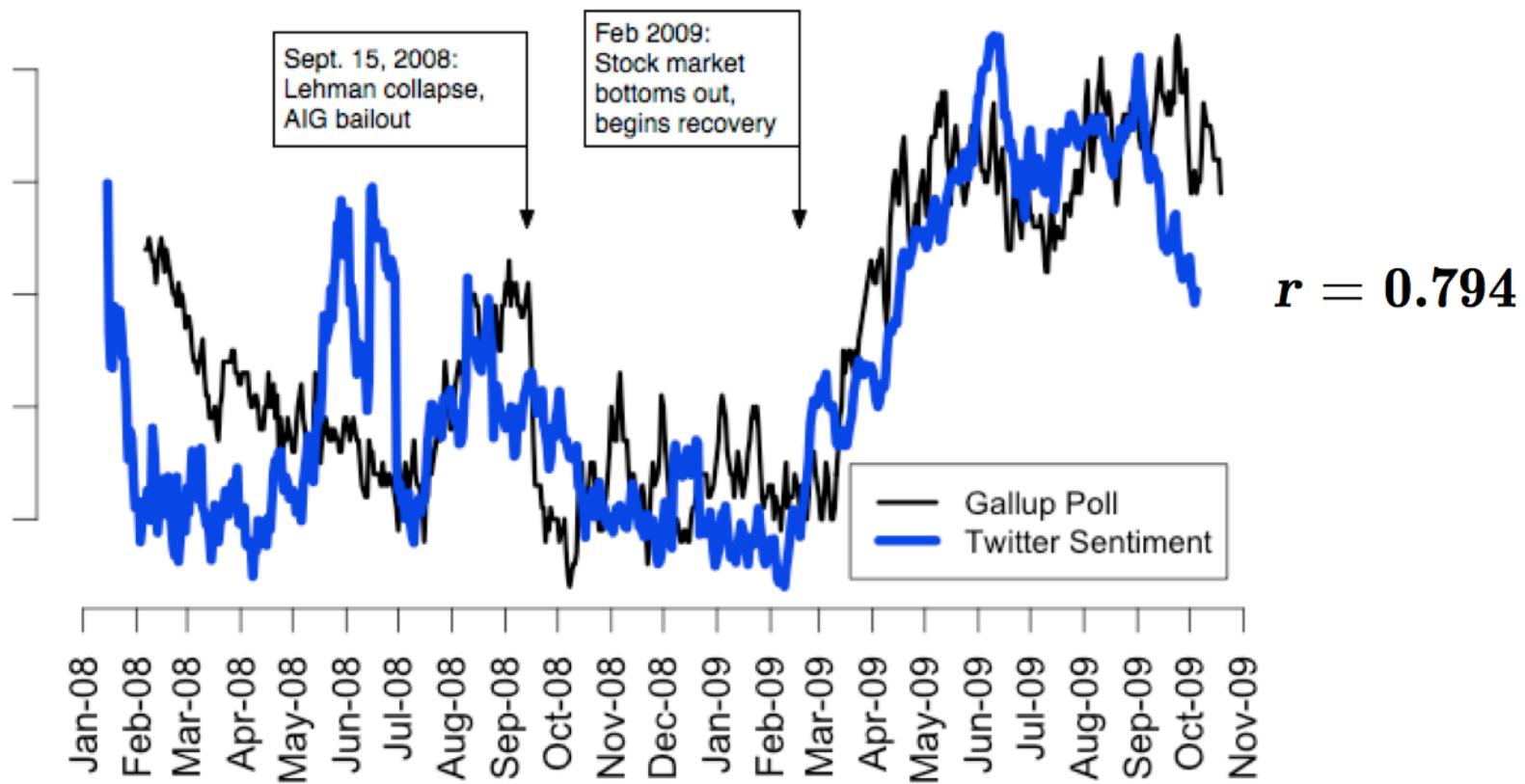
#textworld

## Part 2

Philip Resnik  
Department of Linguistics and  
Institute for Advanced Computer Studies  
University of Maryland  
[resnik@umd.edu](mailto:resnik@umd.edu)

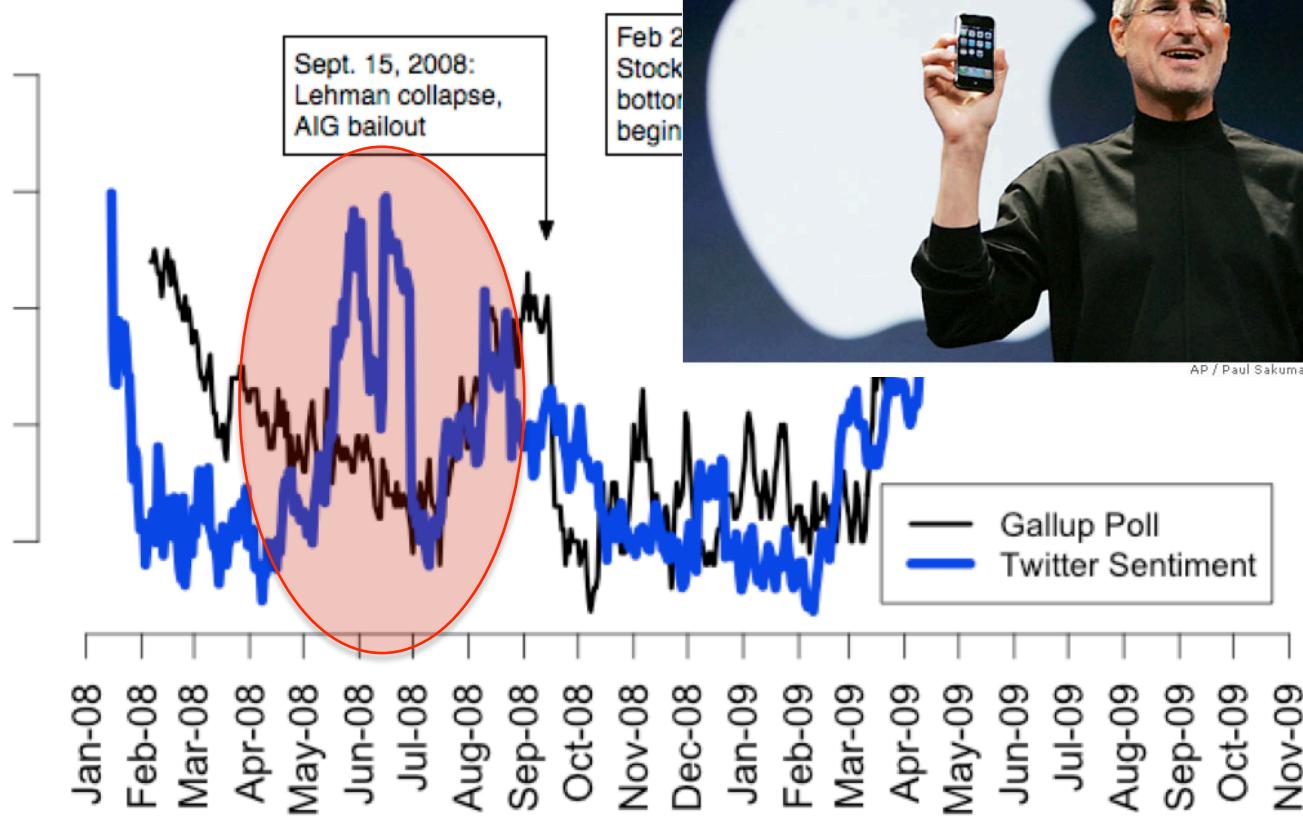
# Remember this?

## jobs on Twitter



# What's going on here?

## jobs on Twitter





AP / Paul Sakuma

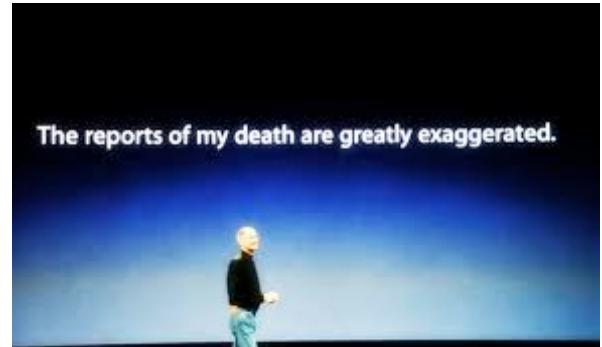
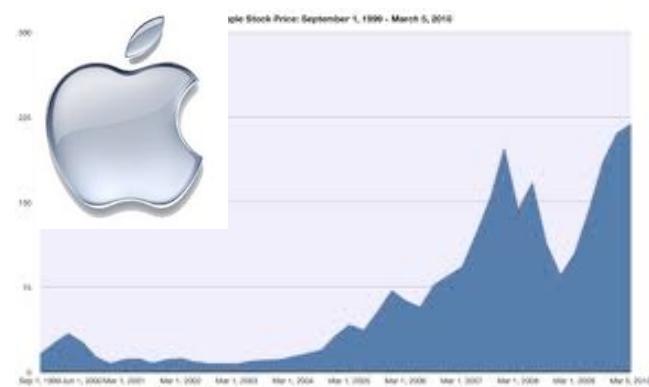
# jobs

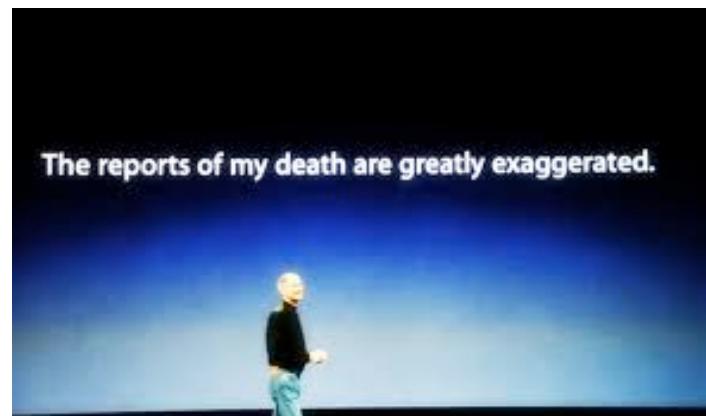
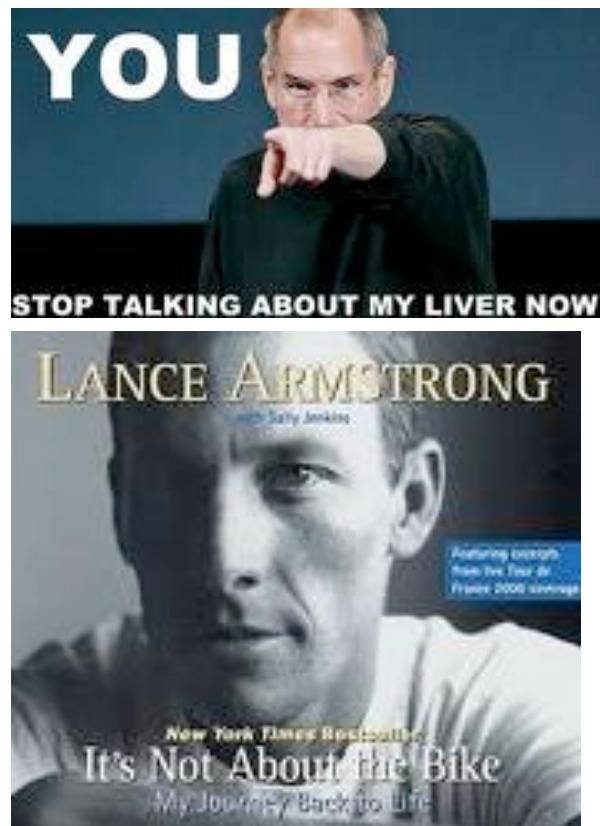
Job id	Name	User	Time	Use	S	Queue
6218104.dserv	berkeleyaligner	fture	17:22:30	R	batch	
6231169.dserv	IG_5	ynhu	08:54:44	R	batch	
6231170.dserv	IG_6	ynhu	08:54:23	R	batch	
6231171.dserv	IG_7	ynhu	08:54:46	R	batch	
6231172.dserv	IG_8	ynhu	08:54:30	R	batch	
6231173.dserv	IG_9	ynhu	08:53:51	R	batch	
6231174.dserv	IG_10	ynhu	08:53:38	R	batch	
6231175.dserv	IG_11	ynhu	08:54:17	R	batch	
6231176.dserv	IG_12	ynhu	08:53:56	R	batch	
6231177.dserv	IG_13	ynhu	08:54:04	R	batch	
6231178.dserv	IG_14	ynhu	08:54:22	R	batch	
6231179.dserv	IG_15	ynhu	08:54:18	R	batch	
6231180.dserv	IG_16	ynhu	08:54:39	R	batch	
6231181.dserv	IG_17	ynhu	08:54:14	R	batch	
6231182.dserv	IG_18	ynhu	08:53:49	R	batch	
6231184.dserv	IG_20	ynhu	08:54:11	R	batch	
6231185.dserv	IG_21	ynhu	08:54:38	R	batch	
6235366.dserv	cdec.1.0	hendra	00:58:48	R	batch	
6235367.dserv	cdec.1.1	hendra	00:58:41	R	batch	
6235368.dserv	cdec.1.2	hendra	00:59:05	R	batch	
6235369.dserv	cdec.1.3	hendra	00:58:28	R	batch	
6235370.dserv	cdec.1.4	hendra	00:58:26	R	batch	
6235371.dserv	cdec.1.5	hendra	00:59:01	R	batch	
6235372.dserv	cdec.1.6	hendra	00:59:01	R	batch	





AP / Paul Sakuma





The real world  
(the physical world,  
how we conceive the world,  
the social media conversation)  
has structure.

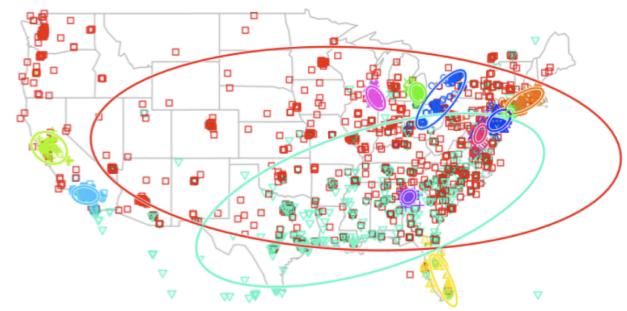
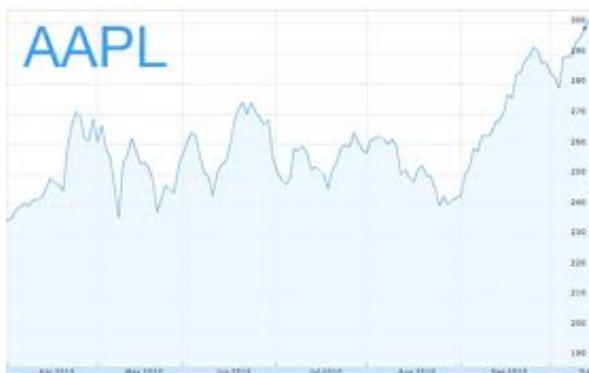
# Linking text with observable variables

pg
adult
rate r
this series
the franchise
the sequel
will smith
brittany

loss	0.025
net loss	0.017
year #	0.016
expenses	0.015
going concern	0.014
a going	0.013
administrative	0.013
personnel	0.013

- something
  - ✚ sumthin
  - ✖ suttin
- 
- lol
  - ✚ lls
- 
- lmao
  - ✚ ctfu

\$2M



# Linking text with unobservables

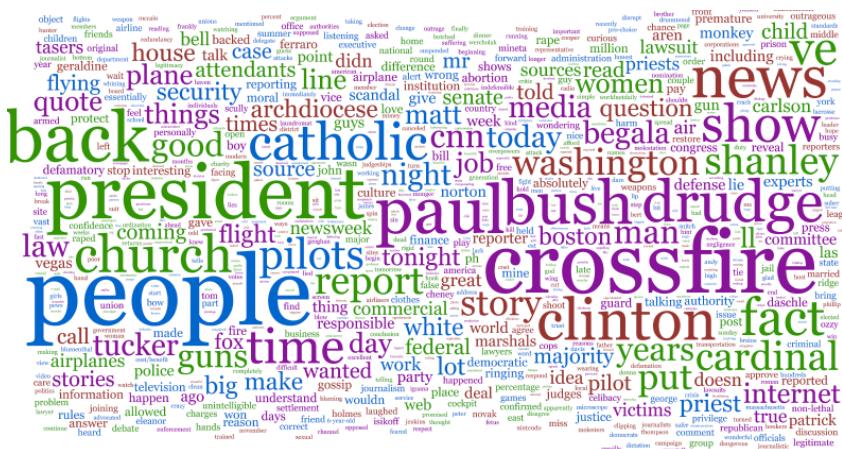
- Opinions and affect
- Topics and agendas
- Perspectives and framing
- Clinical conditions

# Linking text with unobservables

- Opinions and affect
- **Topics and agendas**
- **Perspectives and framing**
- Clinical conditions

*Work in the UMD Computational Linguistics and  
Information Processing Laboratory with  
**Jordan Boyd-Graber, Eric Hardisty, Viet-An  
Ng, and Stephan Greene.***

# Topics and agendas



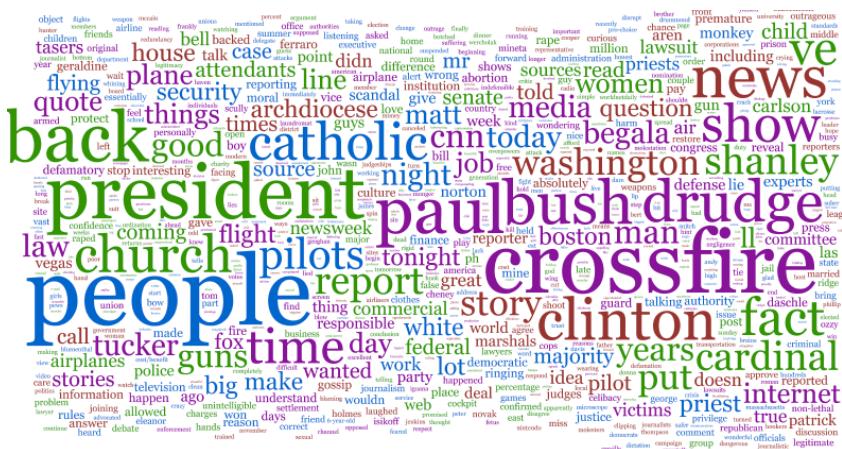
**05\_03\_02.txt.0002 BEGALA** Good evening. Welcome to CROSSFIRE, coming to you live from the George Washington University in beautiful downtown Washington, D.C. Tonight in the CROSSFIRE, the case of the Reverend Paul Shanley, the Roman Catholic priest facing child rape charges in Massachusetts. Should his superiors be held responsible? Also, Matt Drudge, founder of the Internet "Drudge Report." Is he a right-wing muckraker, an Internet gossip or a legitimate journalist? We'll ask Drudge himself when we get him in the CROSSFIRE. First, flying the not-so-friendly skies, would you feel safer if pilots were armed? One outspoken congressional critic is against having guns in the cockpit. We're going to introduce her now. Please welcome, Eleanor Holmes Norton, the Democratic delegate from the District of Columbia. Ms. Norton, thank you. Welcome back.

05\_03\_02.txt.0003 CARLSON Now, Ms. Norton, the majority, the vast majority of commercial airline pilots are strongly in favor of carrying guns in the cockpit on commercial airliners. You're against it. What do you as a delegate know about operating a commercial airliner that the majority of commercial airline pilots don't know

**05\_03\_02.txt.0004 DELEGATE** Well, I know what Transportation Secretary Norm Mineta tells me, and I know what Homeland Security Adviser Tom Ridge tells me, and they are against it. And I think the reason they are against it is you don't want the guy who's flying one of these big busters up there also with a gun in his hand trying to protect his plane. You want air marshals to do that. You want flight attendants to understand how to protect the cockpit. And you want the redundancies that we have built in, redundancy after redundancy, working for you. We are panicking the American people. They say, oh my God, I thought they had the hearings, I thought they did that. Here come the pilots saying, oh no, they haven't. We've got to have guns.

Can word frequencies tell you about the topics in a set of documents?

# Topics and agendas



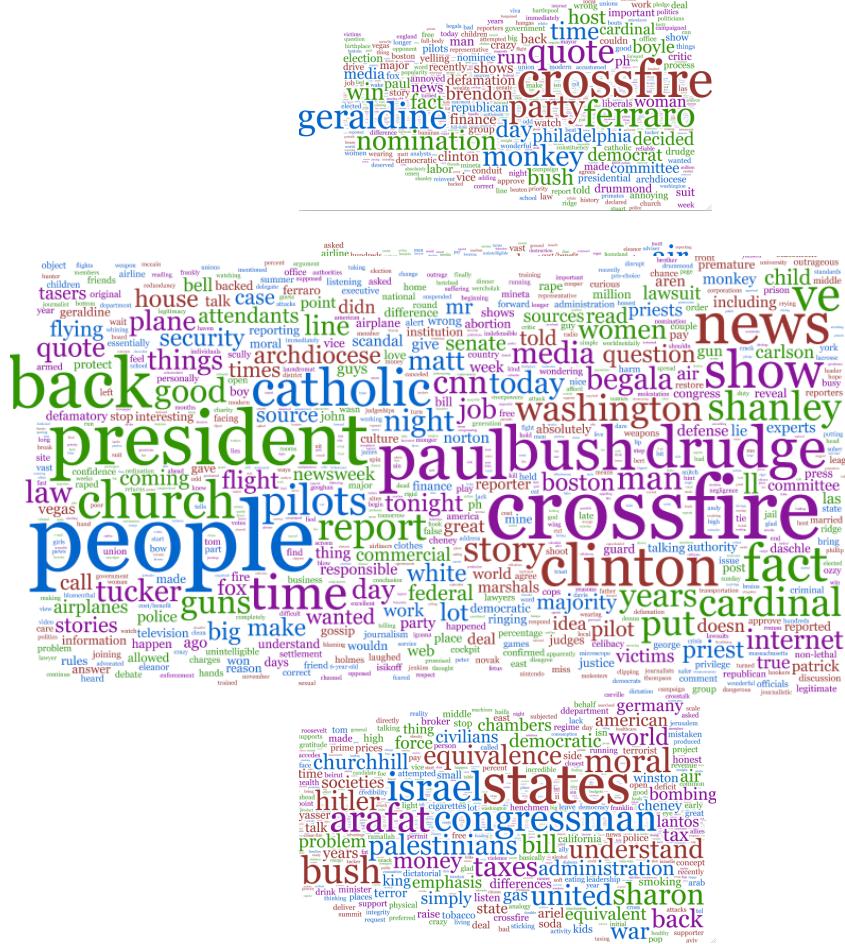
**05\_03\_02.txt.0002 BEGALA** Good evening. Welcome to CROSSFIRE, coming to you live from the George Washington University in beautiful downtown Washington, D.C. Tonight in the CROSSFIRE, the case of the Reverend Paul Shanley, the Roman Catholic priest facing child rape charges in Massachusetts. Should his superiors be held responsible? Also, Matt Drudge, founder of the Internet "Drudge Report." Is he a right-wing muckraker, an Internet gossip or a legitimate journalist? We'll ask Drudge himself when we get him in the CROSSFIRE. First, flying the not-so-friendly skies, would you feel safer if pilots were armed? One outspoken congressional critic is against having guns in the cockpit. We're going to introduce her now. Please welcome, Eleanor Holmes Norton, the Democratic delegate from the District of Columbia. Ms. Norton, thank you. Welcome back.

05\_03\_02.txt.0003 CARLSON Now, Ms. Norton, the majority, the vast majority of commercial airline pilots are strongly in favor of carrying guns in the cockpit on commercial airliners. You're against it. What do you as a delegate know about operating a commercial airliner that the majority of commercial airline pilots don't know

05\_03\_02.txt.0004    DELEGATE    Well, I know what Transportation Secretary Norm Mineta tells me, and I know what Homeland Security Adviser Tom Ridge tells me, and they are against it. And I think the reason they are against it is you don't want the guy who's flying one of these big busters up there also with a gun in his hand trying to protect his plane. You want air marshals to do that. You want flight attendants to understand how to protect the cockpit. And you want the redundancies that we have built in, redundancy after redundancy, working for you. We are panicking the American people. They say, oh my God, I thought they had the hearings, I thought they did that. Here come the pilots saying, oh no, they haven't. We've got to have guns.

Looking at just word counts often gives you a mish-mash.

# Topics and agendas



05\_03\_02.txt.0002 BEGALA Good evening. Welcome to CROSSFIRE, coming to you live from the George Washington University in beautiful downtown Washington, D.C. Tonight in the CROSSFIRE, the case of the Reverend Paul Shanley, the Roman Catholic priest facing child rape charges in Massachusetts. Should his superiors be held responsible? Also, Matt Drudge, founder of the Internet "Drudge Report." Is he a right-wing muckraker, an Internet gossip or a legitimate journalist? We'll ask Drudge himself when we get him in the CROSSFIRE. First, flying the not-so-friendly skies, would you feel safer if pilots were armed? One outspoken congressional critic is against having guns in the cockpit. We're going to introduce her now. Please welcome, Eleanor Holmes Norton, the Democratic delegate from the District of Columbia. Ms. Norton, thank you. Welcome back.

05\_03\_02.txt.0003 CARLSON Now, Ms. Norton, the majority, the vast majority of commercial airline pilots are strongly in favor of carrying guns in the cockpit on commercial airliners. You're against it. What do you as a delegate know about operating a commercial airliner that the majority of commercial airline pilots don't know

05\_03\_02.txt.0004 DELEGATE Well, I know what Transportation Secretary Norm Mineta tells me, and I know what Homeland Security Adviser Tom Ridge tells me, and they are against it. And I think the reason they are against it is you don't want the guy who's flying one of these big busters up there also with a gun in his hand trying to protect his plane. You want air marshals to do that. You want flight attendants to understand how to protect the cockpit. And you want the redundancies that we have built in, redundancy after redundancy, working for you. We are panicking the American people. They say, oh my God, I thought they had the hearings, I thought they did that. Here come the pilots saying, oh no, they haven't. We've got to have guns.

*Bayesian topic models\**  
discover the distinct topics  
interwoven in documents.

\*Wikipedia: Topic Model; Blei et al. 2003



# Topics and agendas

nooses  
long  
made  
high  
prime  
prizes  
**churchill**  
time  
bent  
small  
head  
societies  
**israel**  
**states**  
hitler  
yasser  
arafat  
congressman  
problem  
years  
bush  
palestinians  
money  
king  
emphasis  
simply  
listen  
taxes  
administration  
california  
police  
tax  
understand  
differences  
leadership  
smoking  
arab  
united  
sharon  
drink  
minister  
physical  
deliver  
psychotic  
summer  
request  
prefer  
raise  
tobacco  
deal  
crossover  
ariel  
equivalent  
adult  
sex  
adults  
kodak  
back  
supportive  
War

.03

.44

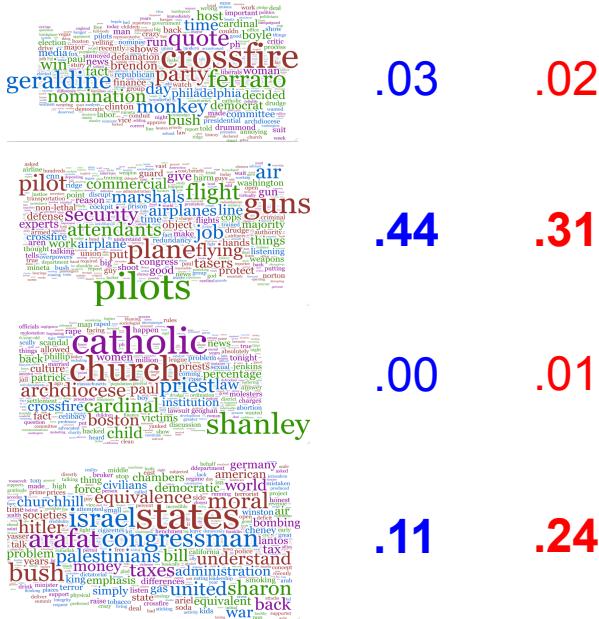
.00

.11

Well, I know what Transportation Secretary Norm Mineta tells me, and I know what Homeland Security Adviser Tom Ridge tells me, and they are against it. And I think the reason they are against it is you don't want the guy who's flying one of these big busters up there also with a gun in his hand trying to protect his plane. You want air marshals to do that. You want flight attendants to understand how to protect the cockpit. And you want the redundancies that we have built in, redundancy after redundancy, working for you. We are panicking the American people. They say, oh my God, I thought they had the hearings, I thought they did that. Here come the pilots saying, oh no, they haven't. We've got to have guns.

Any part of the conversation  
can be viewed as a  
*mixture of topics.*

# Topics and agendas

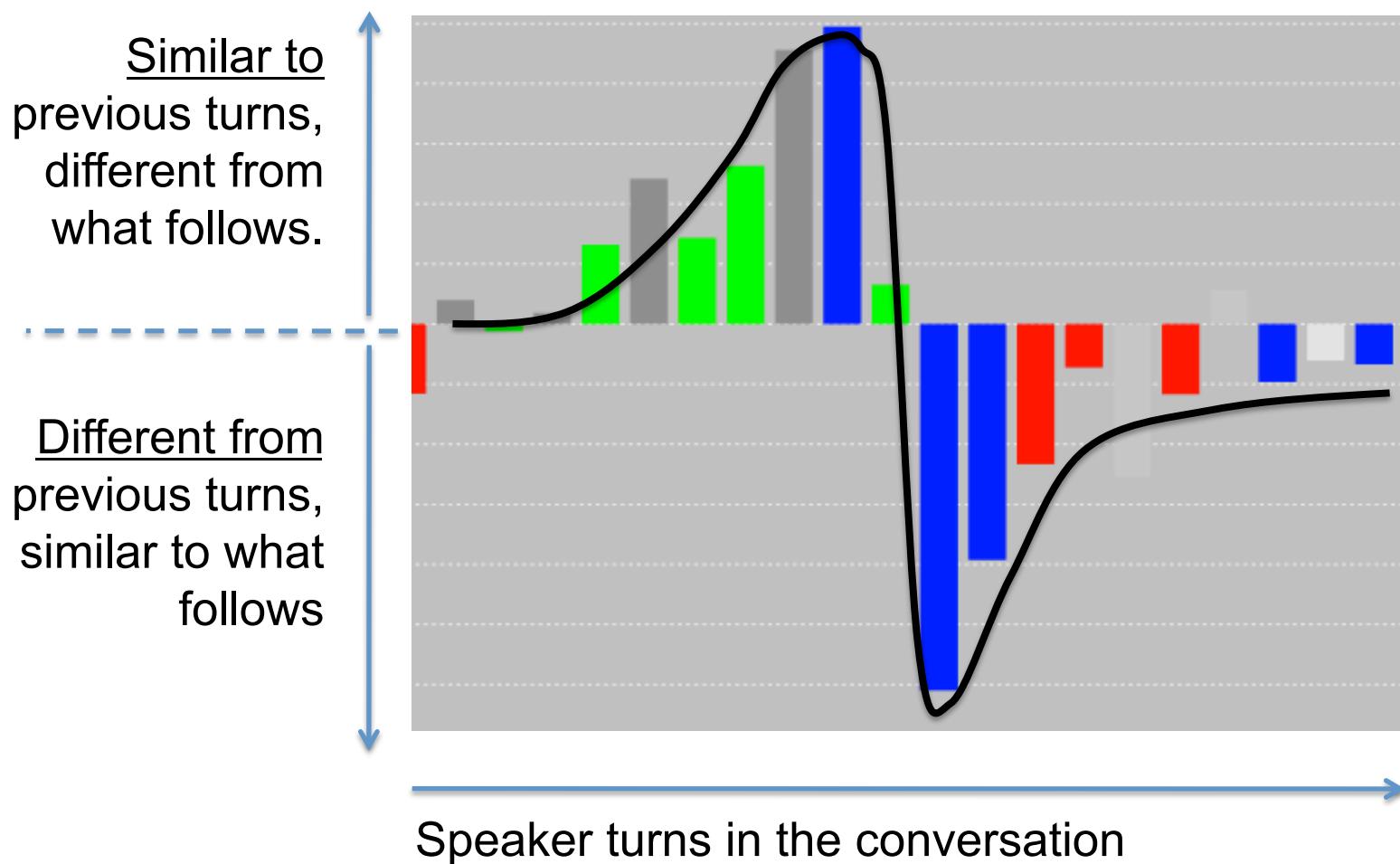


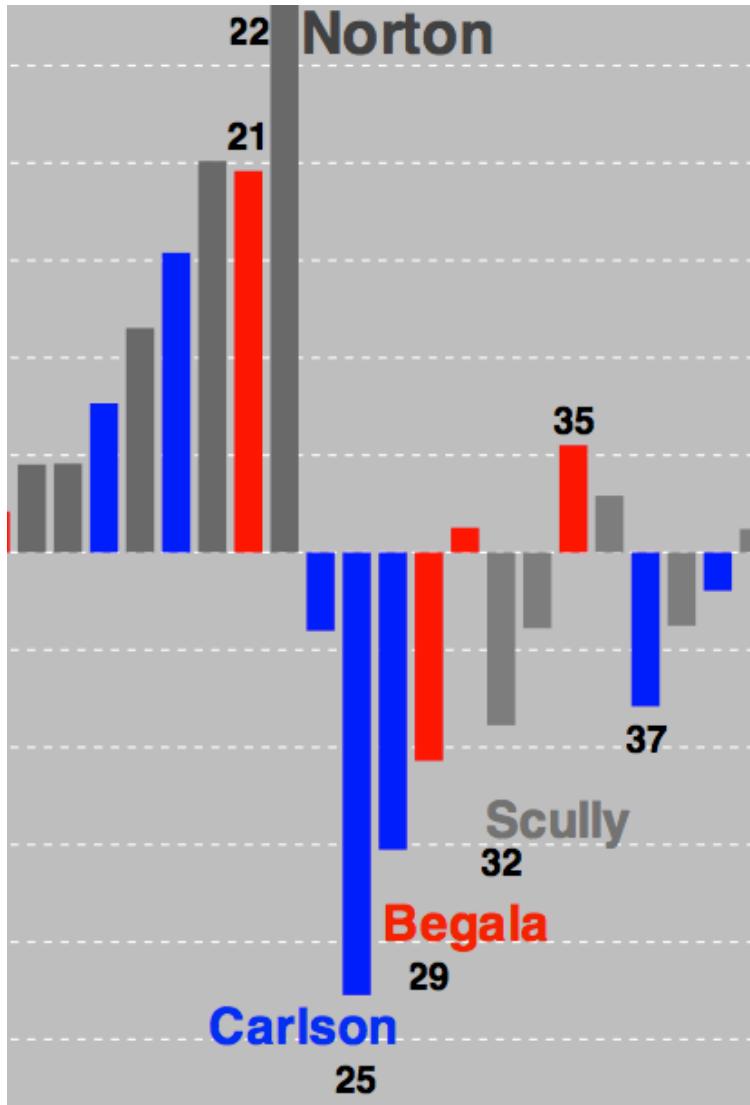
Well, I know what Transportation Secretary Norm Mineta tells me, and I know what Homeland Security Adviser Tom Ridge tells me, and they are against it. And I think the reason they are against it is you don't want the guy who's flying one of these big busters up there also with a gun in his hand trying to protect his plane. You want air marshals to do that. You want flight attendants to understand how to protect the cockpit. And you want the redundancies that we have built in, redundancy after redundancy, working for you. We are panicking the American people. They say, oh my God, I thought they had the hearings, I thought they did that. Here come the pilots saying, oh no, they haven't. We've got to have guns.

If a criminal overpowers a prison guard, the criminal has a prison guard and that is all. If a terrorist overpowers a pilot, he has an airplane he has a weapon of mass destruction. And I don't want just a front line of defense. But why not give them the same training we give our air marshals? Our -- 70 percent of our commercial pilots already have military experience, and allow them to be just another air marshal on every flight?

This means we can compare parts of the conversation to see how similar they are.

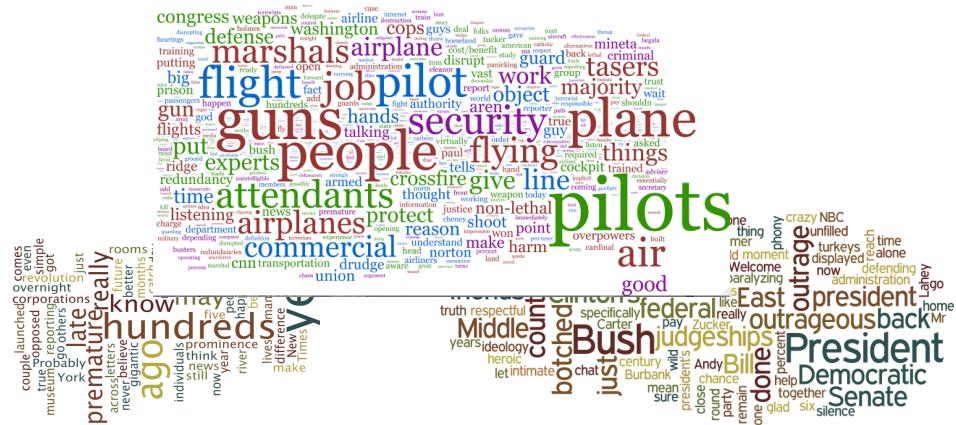
# Topics and agendas

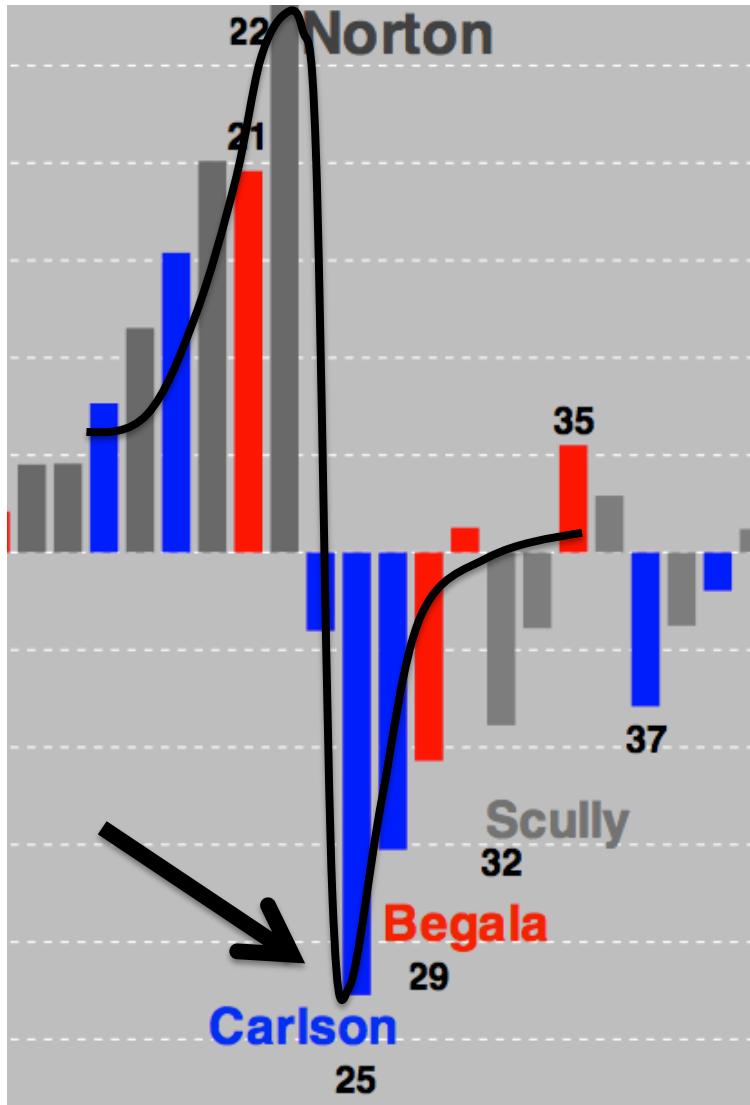




05\_03\_02.txt.0021     BEGALA    We know now what the last line of defense is. Right? We have a lot of security and it's getting better now that finally the Bush administration agreed to federalize the screeners. And as that comes online, it will improve the first line of defense. The last line of defense today is an F-16. Vice President Cheney, in his odd exercise of authority outside the chain of command, gave the order on September 11, we know now from Mr. Cheney's interview with Tim Russert that he gave the order to shoot down civilian airliners. That today is the last line of defense. I guess I'd rather move one line up and have the last line be, God forbid we might need the F-16s, why not have a pilot who is well trained with a weapon?

05\_03\_02.txt.0022 NORTON I'll make a deal with both of you guys. If we wait until the people whose job it is to tell us whether or not guns will work on airplanes, and they say, look, exactly what you said, more good than harm, considering that you might have to shoot down a plane, then I'm very open to listening. The reason I'm opening to listen is that they train these pilots, you know, they would be like cops, I don't object to cops having guns, I don't object to security guards having guns. But I object to putting guns in the hands of pilots before we know what a study would tell us, whether or not they work.

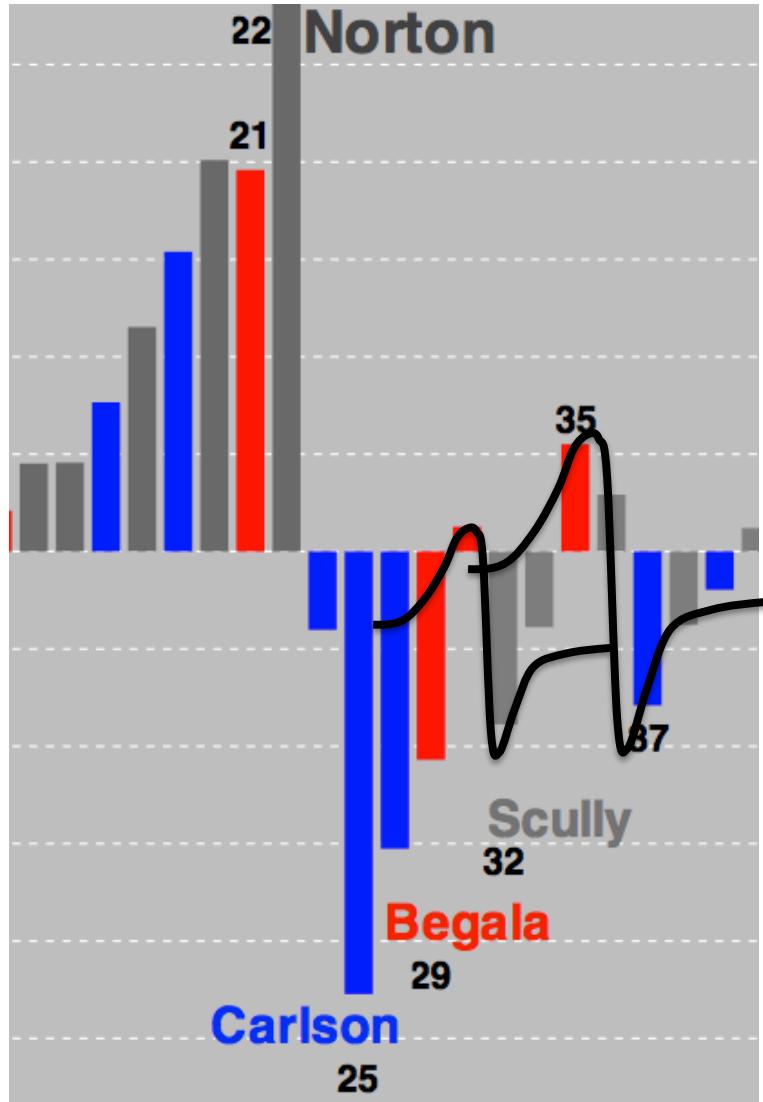




**05\_03\_02.txt.0025** CARLSON Next on CROSSFIRE, why church officials took no action against a Catholic priest now facing charges of child rape. We'll debate the case of Paul Shanley in the CROSSFIRE. Also, he's not quite human but they elected him anyway. You heard it before, details in the CROSSFIRE "News Order." We'll be right back.

**05\_03\_02.txt.0027** CARLSON Welcome back to CROSSFIRE. Former the Roman Catholic priest Paul Shanley is in jail in California tonight, facing charges of child rape. Shanley has waived his right to fight his return to Massachusetts, where he's accused of molesting dozens of boys while a priest in the Boston archdiocese. Documents released by the archdiocese show that church officials, including Cardinal Bernard Law knew about molestation complaints, but allowed Shanley to continue to serve as a priest. And in a late development tonight, the archdiocese of Boston backed out of its settlement with 86 victims of defrocked the priest, John Geoghan. Lawsuits against Geoghan, which were the catalyst for the present sexual abuse scandal would have cost the church from between \$15 million and \$40 million. For more on this scandal, let's welcome our guest, Patrick Scully, communications director of the Catholic League.



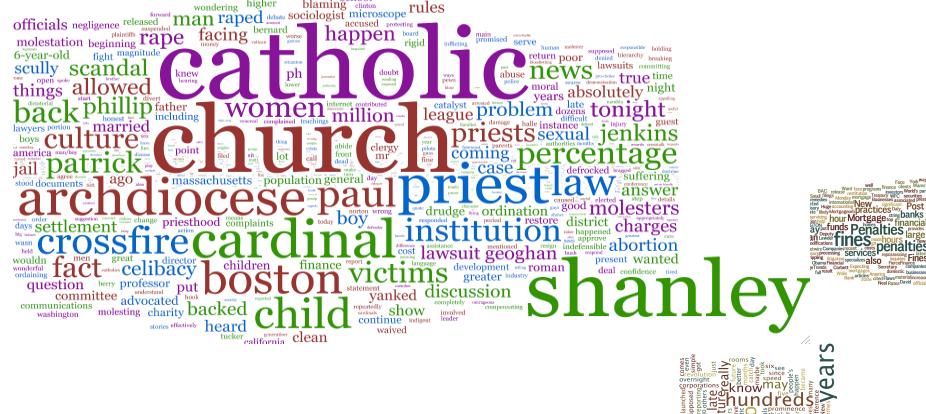


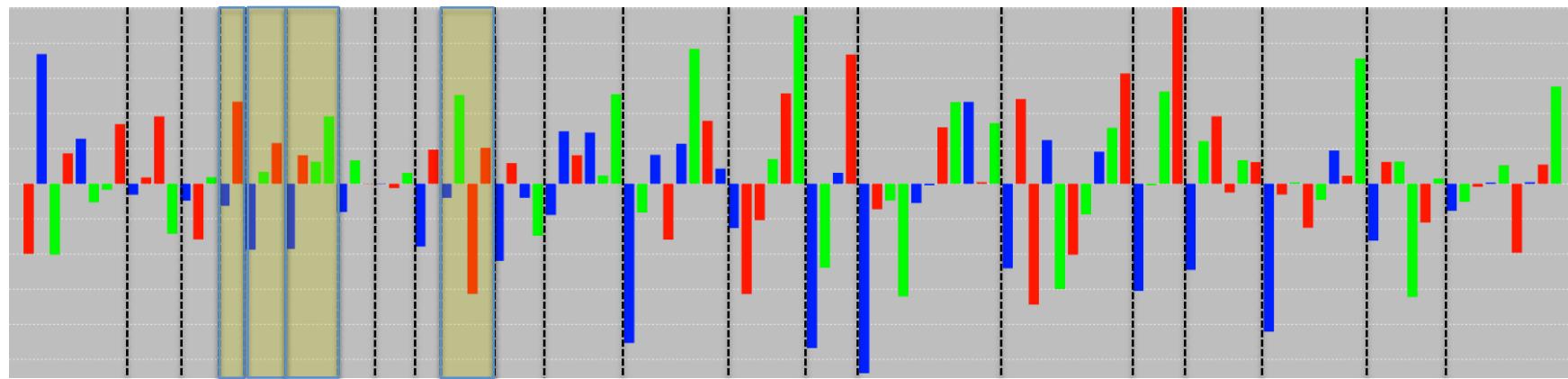
05\_03\_02.txt.0032 SCULLY Well, Cardinal Law wanted the deal to go forward. So in this case he was trying to do the right thing.

05\_03\_02.txt.0034 SCULLY The finance committee would not approve it, and that is a big difference. For this to go forward the finance committee of the archdiocese of Boston has to approve the deal [...] This is the first time, since he's become cardinal in the archdiocese that one of his requests was turned down by the finance committee

[...]

05\_03\_02.txt.0037 CARLSON Well, I have a suggestion for you then. This priest, Paul Shanley, who was arrested just the other day, was a known child molester for more than 20 years, spoke at NAMBLA conference, bragged about having venereal disease. And I'm wondering why the church didn't call the police about him 20 years ago. But how about just even a year ago? How about six months ago when the first stories of this came out?





Gwen Ifill



Joe Biden



Sarah Palin

IFILL Senator Biden, you voted for this bankruptcy bill. Senator Obama voted against it. Some people have said that mortgage-holders really paid the price.

BIDEN Well, mortgage-holders didn't pay the price. Only 10 percent of the people who are -- have been affected by this whole switch from Chapter 7 to Chapter 13 -- it gets complicated. But the point of this -- Barack Obama saw the glass as half-empty. I saw it as half-full. We disagreed on that, and 85 senators voted one way, and 15 voted the other way. But here's the deal. Barack Obama pointed out two years ago that there was a subprime mortgage [...] And there -- ways that we're offering are not being supported by -- by the Bush administration nor do I believe by John McCain and Governor Palin.

PALIN That is not so, **but because that's just a quick answer, I want to talk about, again, my record on energy versus your ticket's energy ticket, also.** I think that this is important to come back to, with that energy policy plan again that was voted for in '05. When we talk about energy, we have to consider the need to do all that we can to allow this nation to become energy independent. It's a nonsensical position that we are in when we have domestic supplies of energy all over this great land. And East Coast politicians who don't allow energy-producing states like Alaska to produce these, to tap into them, and instead we're relying on foreign countries to produce for us.

# Linking text with unobservables

- Opinions and affect
- **Topics and agendas**
- Perspectives and framing
- Clinical conditions

# Perspectives and framing

**Framing** (Entman 1993): “to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation” (Cf. “spin”)

# Perspectives and framing

- **Lexical framing**

Estate tax

Death tax

Pro-choice

Pro-life

Orcas

Killer whales

Get the government out of my

- **Two problems**

- What are the two sides?

- How do you identify the “fighting phrases”?

# Lexical framing

- Approach: non-parametric Bayesian modeling\*

November 12, 2007 Edition 41

**Is the PA beginning to resemble the SLA?**

**two Palestinian views**

**The PA cannot remain transitional much longer**

by Ghassan Khatib

The situation in Nablus can only further discredit the PA in the eyes of its own public and strengthen comparisons with the SLA.

**Unite or dissolve**

an interview with Eyad Sarraj

The Palestinians have lost the ability to govern themselves, to make war or to make peace.

**two Israeli views**

**The danger is there**

by Yossi Alpher

Any comparison between Abbas/Salam and Lahd is, to say the least, not flattering.

**The political context is totally different**

by Dani Reshef

The SLA sought to join the Lebanese army as a territorial brigade.

\*Hardisty, Boyd-Graber, Resnik 2010; in progress.

# Lexical framing

Israeli	Palestinian
zionist dream	american jew
zionist state	achieve freedom
zionist movement	palestinian freedom
american leadership	support palestinian
american victory	palestinian suffer
abandon violence	palestinian territory
freedom (of the) press	palestinian statehood
palestinian violence	palestinian refugee

# Lexical framing

- Analysis of pro- and anti-death penalty sites

79 NAME 's	176 NAME 's
46 NAME 's case	23 NAME 's car
19 NAME 's attorney	22 NAME 's home
15 NAME 's execution	22 NAME 's body
13 NAME 's lawyers	17 NAME 's murder
8 upheld NAME 's conviction	17 NAME 's house
8 fight (for) NAME 's life	16 NAME 's death
8 facts (of ) NAME 's case	15 NAME 's mother
7 NAME 's life	14 NAME 's appeal
7 NAME 's lawyer	13 NAME 's wife

# Perspectives and framing

- **Syntactic framing\***



“explosion killed eleven workers”

“eleven people were lost in an explosion and fire”

X  
            
X caused the event

verb      Y  
            
Y was affected by the event

\*Greene 2007, Greene and Resnik 2009

# Take-aways

**There is an ocean of text now available**

**We can discover the structure below the surface**

**Structure in the world**

**Structure in the social conversation about that world**

**How to relate the two**