

(Asymmetric) language adaptation

Quick notes about the readings for next time, ~~and~~ ; today's lecture:
pedagogy

the goal now in this part of the course is to find inspiration for course projects

You are not reviewing these papers!

For your "proposal" assignments, this should be ~~the~~ focus your

Think about:

what are interesting related problems?

- or -

what ~~are interesting~~ could ~~be~~ else could be done w/ the datasets?

- or -

what would some interesting related datasets be?

~~if you have a good grasp of the techniques used~~ what else could be done w/ the techniques?

- or -

what else could be done to the techniques?

I don't want you to get hung up on what you don't know.

budget about ~1.5 hr 1st pass.

Don't forget ^{for} you actually need to come up with write a proposal.

~~Blah blah~~

~~If you want to then go back on the paper.~~

~~then go~~

~~what did you like about how they analyzed~~

This

Q: Pick two people for ~~next time~~ next Th., exempt them from next ~~two~~ days "proposals".
Jean, Mathre.

- need to schedule sign-ups.

Asymmetric language synchronization in social interaction

Removed: image licensed for powerpoint presentation only; review can be viewed here: [http://www.slideshare.net/chartrand/2009-01-27-chartrand-and-bargh-on-social-interaction](#)

Conversational synchrony

People tend to adopt the behaviors of their interlocutors.
[Giles et al. '91, Chartrand & Bargh '99, inter alia]

Non-verbal	Posture [Condon and Ogston '67] Nodding [Hale and Burgoon '84]
"Non-semantic"	Pause length [Jaffe and Feldstein '70] Backchannels [White '84]
Language content	Words, esp. referring expressions [Brennan and Clark '96, Nenkova et al. '08, Stoyanchev and Stent '09,] Word classes [Niederhoffer and Pennebaker '02] Degree of self-disclosure [Derienga et al. '73]

→ "lock-in" on certain "vocab" choices

• note the dates - lots of prior work has been in small-scale lab settings.

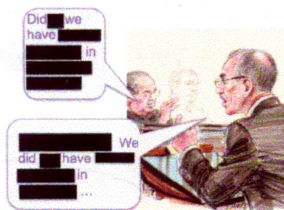
People tend to adopt the behaviors of their interlocutors...

...but participants can entrain to different degrees.

Asymmetric conversational synchrony can tell us a lot about user relationships.

Preview of Part I: Pairwise adaptation and power

Who's in charge?



Preview of Part II: Adaptation to a group and long-term engagement

Removed: image licensed for powerpoint presentation; review can be viewed here: [http://www.slideshare.net/chartrand/2009-01-27-chartrand-and-bargh-on-social-interaction](#)

Isn't all this obvious?

Paraphrasing Stuart Shieber:

[http://www.hawaii.edu/comp/chartrand/2009-01-27-chartrand-and-bargh-on-social-interaction](#)

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.

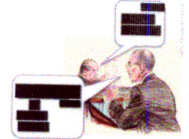
- discussion of how to present things as "clear" w/out making them seem "trivial"
• make people aware of the counterhypothesis
• ask a q, split the crowd's answer
=> answer can't have been obvious.

In class, we talked about strategies for presenting things in a way to make them "obvious" without having your audience think the work is *trivial*.

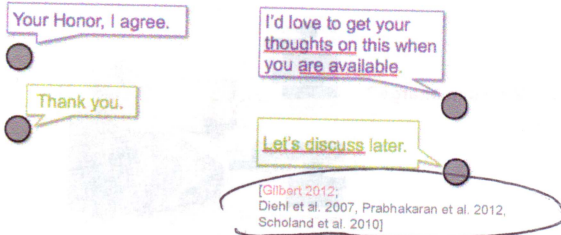
- make your audience aware of the counter-hypotheses
- ask your audience what they think the outcome will be; for an interesting question, the audience should be divided, which "proves" that the eventual answer can't be obvious.
- admit that there are easy/obvious cases, but also show harder ones.

Echoes of Power: Language effects & power differences in social interaction

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, & Jon Kleinberg
WWW 2012



Language reveals power: "easy" cases



What about general (domain-independent) signals?

here we talked about why so much work on Enron - the allure of annotated data. (here, the org chart)

Why word classes instead of words?

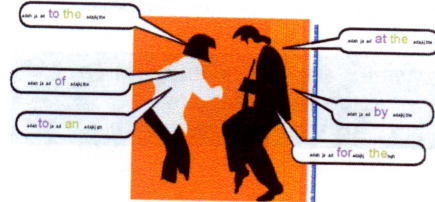
Function-class matching: unconscious & frequent
[Niederhoffer and Pennebaker 2002]



In contrast, **direct repetition** is under the speaker's control, and could just be choice of topic. ☹

Who has the (conversational) lead?

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



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

3.3 Harvard General Inquirer

The Harvard General Inquirer is a lexical coding system, semantic and syntactic classification tool of research papers written in English. It is based on the work of the Harvard General Inquirer. The system is available at <http://www.cis.upenn.edu/~jduffy/gi/>. The system is available at <http://www.cis.upenn.edu/~jduffy/gi/>. The system is available at <http://www.cis.upenn.edu/~jduffy/gi/>.

Table 3. A list of the Harvard General Inquirer categories.

Category	Number	Meaning
1	ADJECTIVE	ADJECTIVE
2	ADVERB	ADVERB
3	CONJUNCTION	CONJUNCTION
4	PREPOSITION	PREPOSITION
5	PRONOUN	PRONOUN
6	VERB	VERB
7	NUMERAL	NUMERAL
8	PROPER NOUN	PROPER NOUN
9	COMMON NOUN	COMMON NOUN
10	ARTICLE	ARTICLE
11	OTHER	OTHER

Table 4. A list of the Harvard General Inquirer categories.

Category	Number	Meaning
12	POSITIVE	POSITIVE
13	NEGATIVE	NEGATIVE
14	EMOTIONAL	EMOTIONAL
15	POWER	POWER
16	CONFIDENCE	CONFIDENCE
17	COMPLEXITY	COMPLEXITY
18	CONCRETE	CONCRETE
19	ABSTRACT	ABSTRACT
20	PERSONAL	PERSONAL
21	IMPERSONAL	IMPERSONAL
22	FORMAL	FORMAL
23	INFORMAL	INFORMAL
24	ACADEMIC	ACADEMIC
25	POPULAR	POPULAR
26	SCIENTIFIC	SCIENTIFIC
27	NON-SCIENTIFIC	NON-SCIENTIFIC
28	TECHNICAL	TECHNICAL
29	NON-TECHNICAL	NON-TECHNICAL
30	LEGAL	LEGAL
31	NON-LEGAL	NON-LEGAL
32	MEDICAL	MEDICAL
33	NON-MEDICAL	NON-MEDICAL
34	BUSINESS	BUSINESS
35	NON-BUSINESS	NON-BUSINESS
36	RELIGIOUS	RELIGIOUS
37	NON-RELIGIOUS	NON-RELIGIOUS
38	POLITICAL	POLITICAL
39	NON-POLITICAL	NON-POLITICAL
40	SPORTS	SPORTS
41	NON-SPORTS	NON-SPORTS
42	ENTERTAINMENT	ENTERTAINMENT
43	NON-ENTERTAINMENT	NON-ENTERTAINMENT
44	EDUCATION	EDUCATION
45	NON-EDUCATION	NON-EDUCATION
46	HEALTH	HEALTH
47	NON-HEALTH	NON-HEALTH
48	ENVIRONMENTAL	ENVIRONMENTAL
49	NON-ENVIRONMENTAL	NON-ENVIRONMENTAL
50	OTHER	OTHER

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 ?

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

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

[Danescu-Niculescu-Mizil, Dumais, Gamon WWW 2011]

not always available for small-data settings

Reported as % (multiplied by 100) in the following.

Getting a more immediate measure will be cool.

But correlation itself isn't good enough to show influence.

Supreme Court Oral Args: samples

Interruptions:

MR. MILLER: Justice Stevens, and may it please the Court: These two contract cases concern whether the Government is liable in money damages under the Contract Disputes Act and section 110 of the Indian Self-Determination Act when the Secretary asks to fully pay a contract price for the - JUSTICE O'CONNOR: Would you mind explaining to us ...

Short interchanges with disfluencies, direct repetitions:

MR. MILLER: It - it is precisely, Justice Kennedy, what - what you stated. In fact, we can turn directly to page 149 -

JUSTICE KENNEDY: I - I stated an or. I - what I - I stated at the outset?

MR. MILLER: The - the Government took money for its own Federal functions

instead of - including contract monitoring -

JUSTICE KENNEDY: Including contract monitoring.

MR. MILLER: Including contract monitoring. Including contract monitoring. In fact, they say so in their brief and - and defend that action. The - the sections on page 126 and 128 articulate in quite a bit of detail the nature of the contract support costs that must be added.

JUSTICE SCALIA: These are strange contracts, of course, aren't they? ...

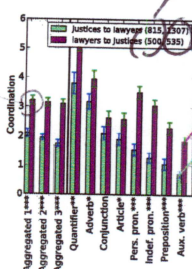
Joking around:

MR. SRINIVASAN: It does, Justice Scalia - Justice Souter. Excuse me.

JUSTICE SOUTER: Thank you, but apologize to him. (Laughter.)

Status in US Supreme Court transcripts

50,000 exchanges + metadata (download from my webpage)



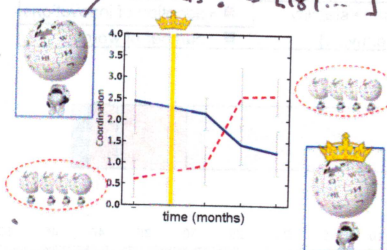
actually show bars since s.d. results marked in x.

of users for all markers
total # of users for "It markers"
low status to high status
high status to low status

Nice to indicate amount of data

Status change in Wikipedia

240,000 exchanges + metadata (download from my webpage)

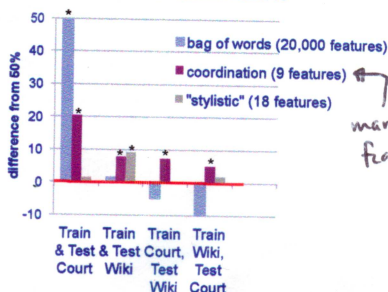


- the time lag is of interest

- discussion of swing votes, the importance of oral arguments vs. the internal discussion afterwards, Clarence Thomas.
- posted the private analysis to Piazza.

Evidence of domain independence

SVM classification with various features



many fewer features!

(could do better w/ more sophisticated alg., of course)

No country for old members:

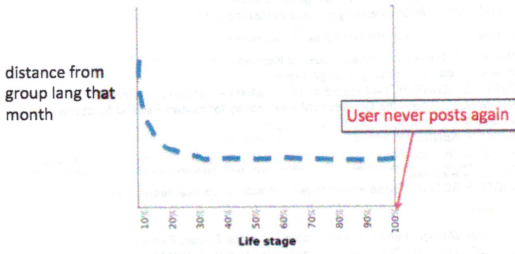
User lifecycle & linguistic change in online communities

C. Danescu-Niculescu-Mizil, R. West, D. Jurafsky, J. Leskovec, & C. Potts
Best paper award, WWW 2013

how

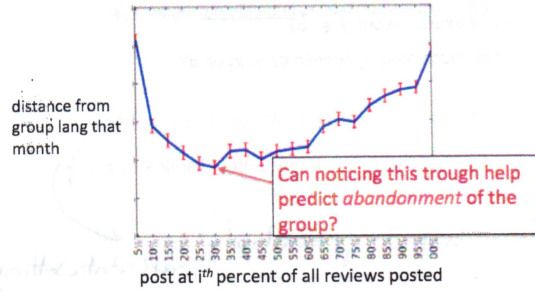
Discussion of fact that this is a long-lasting community w/ very active members, and probably more of a 'true' community compared to, e.g., amazon reviewers.

Hypothesis: a user starts out of sync, then synchronizes



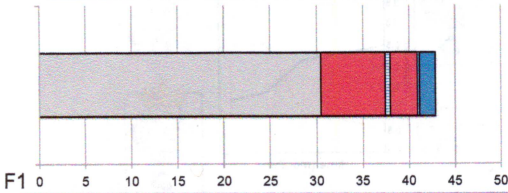
post at i^{th} percent of all reviews posted

Actual lifecycle pattern



Predicting imminent exit

- Activity baseline
- +language stability
- + {I, me, my,....}
- +similarity to community
- +adoption of innovations
- +number of words



Language gives 12 point absolute (40% relative) improvement; synchronization is the lions' share.