

Style and Influence in Social Text

11-27-29

Announcement

- Project reports next week
 - same drill as midterm reports
 - reverse order as midterm reports
- We know you're not done yet
 - ... but you will be by midnight Mon 12/10, right?
 - start with one slide summarizing midterm

FCE's

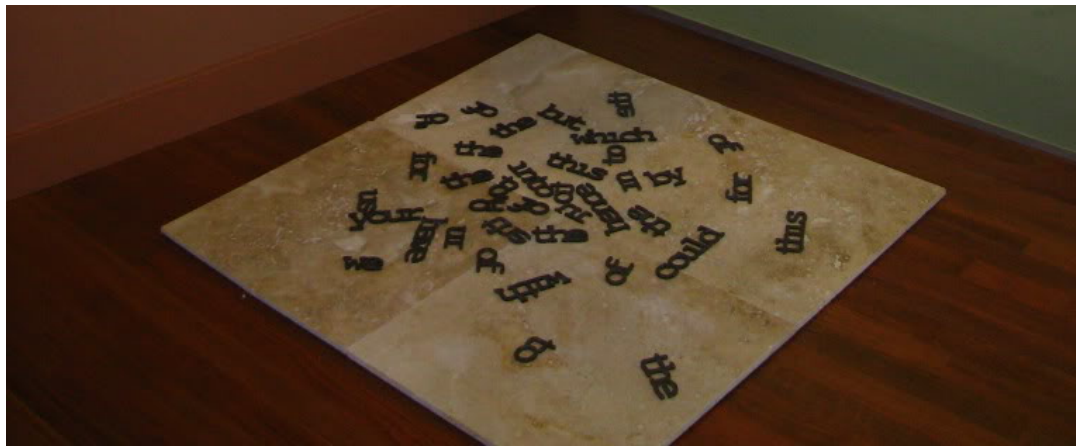
- Are now open
- We do read them...and people do care
- Especially this year
 - free-text comments on assignments/structure/layout of course *very* welcome

Puzzle time

- This sentence has 5 vowels
- i e e e a o o s o a

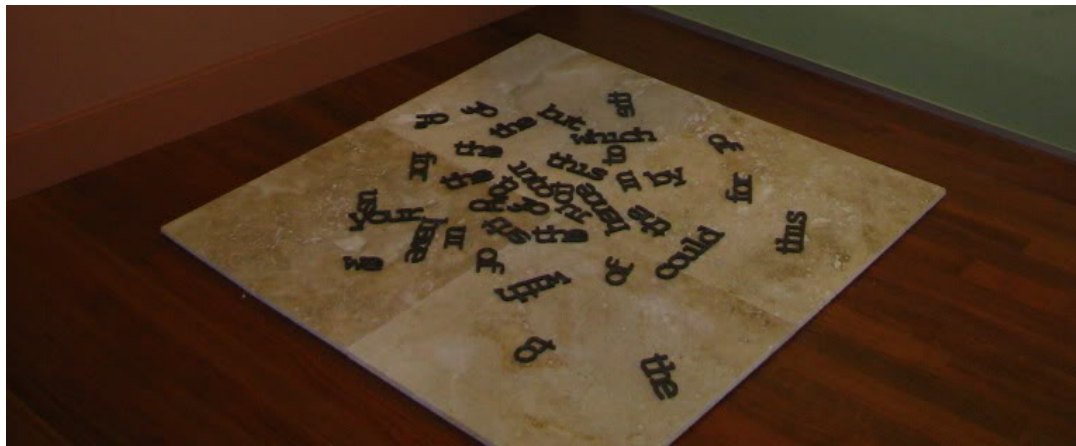
Today's topics

- Summary: there are signals in common words
 - What can you infer from how people use the most frequent words in text?



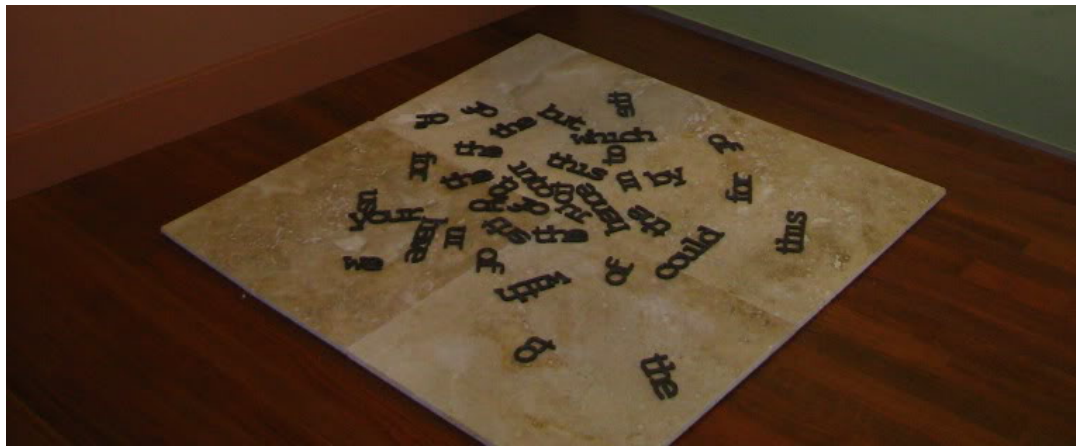
Today's topics

- **Summary:** there are signals in common words
 - What can you infer from how people use the most frequent words in text?



Today's topics

- Summary: there are signals in common words
 - What can you infer from how people use the most frequent words in text?



Today's topics

- Summary: there are signals in common words
 - What can you infer from how people use the most frequent words in text?
 - Patterns of usage → "literary style"
 - predicts: authorship, gender, ...
 - Style changes according to situation
 - and is transmitted from person to person
- Outline:
 - some background and two recent papers



Background: Authorship attribution

- Mosteller and Wallace, 1964. “Inference and Disputed Authorship”: frequency of function words can be used to classify documents by author.
 - Function words are not under conscious control
 - Function word use is independent of content
 - Histogram of function words is ok

Authorship attribution

Schlomo Argamon,
Schlomo Levitan

Author	Book	# Chapters	Avg. Words
Cather	<i>My Antonia</i>	45	1826
	<i>Song of the Lark</i>	60	2581
	<i>The Professor's House</i>	28	2172
Conrad	<i>Lord Jim</i>	45	2913
	<i>The Nigger of the Narcissus</i>	5	10592
	<i>Jude the Obscure</i>	53	2765
Hardy	<i>The Mayor of Casterbridge</i>	45	2615
	<i>Tess of the d'Urbervilles</i>	58	2605
	<i>The Europeans</i>	12	5003
James	<i>The Ambassadors</i>	36	4584
	<i>The Jungle Book</i>	13	3980
	<i>Kim</i>	15	7167
Lewis	<i>Babbitt</i>	34	3693
	<i>Main Street</i>	34	4994

Feature Set	Author	Nationality
Freq. Words	99.00%	93.50%
Freq. Pairs	91.60%	91.30%
Freq. Coll. (k=5)	88.94%	90.20%
Freq. Coll. (k=10)	84.00%	87.20%

SVM on histogram
of 200 most frequent words

Whose thumb is it anyway?
Classifying author personality from weblog text

Jon Oberlander

School of Informatics
University of Edinburgh
2 Buccleuch Place
Edinburgh, EH8 9LW

`j.oberlander@ed.ac.uk`

Scott Nowson

School of Informatics
University of Edinburgh
2 Buccleuch Place
Edinburgh, EH8 9LW

`s.nowson@ed.ac.uk`

COLING 2006

The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods

Yla R. Tausczik¹ and James W. Pennebaker¹



LIWC

- 1986: writing about emotional upheavals improved physical health (!)
- Can you refine this statement?
 - what *sort* of writings yield the *best* results?
 - but: people don't agree on ratings
 - and: “judges tend to get depressed when reading depressing stories.”
 - so: design an automatic “instrument” to rate writings (Linguistic Inquiry and Word Count) based on most frequent words

Summary Table Linking LIWC Word Categories to Published Research Studies

Category	Examples	Words in Category	Psychological Correlates	Published Articles
<i>Linguistic processes</i>				
Word count			Talkativeness, verbal fluency	2, 9, 18, 19, 20, 24, 32, 35, 36, 39, 40, 48, 53, 54, 57, 60, 66, 70, 72, 73, 74, 86, 89, 103, 115
Words/sentence			Verbal fluency, cognitive complexity	3, 7, 39, 43
Dictionary words	(Percentage of all words captured by the program)		Informal, nontechnical language	19, 42, 43, 65, 66, 85, 89
Words >6 letters	(Percentage of all words longer than 6 letters)		Education, social class	3, 19, 20, 27, 35, 36, 42, 43, 73, 74, 79, 89, 90, 93, 103, 115
Total function words		464		
Total pronouns	I, them, itself	116	Informal, personal	1, 19, 36, 43, 55, 89, 90, 119
Personal pronouns	I, them, her	70	Personal, social	58, 79
First-person singular	I, me, mine	12	Honest, <u>depressed</u> , low status, personal, emotional, informal	1, 3, 4, 5, 11, 13, 18, 27, 35, 36, 46, 55, 56, 64, 65, 66, 68, 69, 72, 73, 74, 78, 80, 81, 87, 89, 90, 92, 93, 94, 100, 101, 105, 108, 109, 112, 113, 115
First-person plural	We, us, our	12	Detached, <u>high status</u> , socially connected to group (sometimes)	1, 4, 13, 18, 35, 46, 55, 64, 65, 74, 78, 81, 87, 90, 93, 94, 97, 100, 103, 104, 105, 106, 113
Second person	You, your, thou	20	Social, elevated status	1, 18, 27, 41, 55, 90, 100, 105, 106
Third-person singular	She, her, him	17	Social interests, social support	1, 3, 14, 36, 39, 55, 64, 66, 80, 87, 88, 90, 95
Third-person plural	They, their, they'd	10	Social interests, out-group awareness (sometimes)	1, 3, 14, 39, 55, 64, 80, 87, 88, 95

LIWC words - cover about 55% of the tokens (not types) in most text
Categories are mostly designed by hand, by committee

Category	Examples	Words in Category	Psychological Correlates	Published Articles
Indefinite pronouns	It, it's, those	46		
Articles	A, an, the	3	Use of concrete nouns, interest in objects and things	19, 36, 43, 74, 79, 80, 89, 92, 115
Common verbs	Walk, went, see	383		58, 79
Auxiliary verbs	Am, will, have	144	Informal, passive voice	
Past tense	Went, ran, had	145	Focus on the past	1, 13, 37, 62, 68, 73, 79, 87, 89, 91, 93, 115
Present tense	Is, does, hear	169	Living in the here and now	13, 36, 37, 42, 62, 68, 73, 87, 89, 90, 93, 115
Future tense	Will, gonna	48	Future and goal oriented	13, 26, 37, 41, 62, 64, 76, 90, 93, 114
Adverbs	Very, really, quickly	69		58
Prepositions	To, with, above	60	Education, concern with precision	43, 79, 89, 92, 115
Conjunctions	And, but, whereas	28		
Negations	No, not, never	57	Inhibition	24, 39, 40, 48, 79, 89, 90, 114, 115
Quantifiers	Few, many, much	89		
Numbers	Second, thousand	34		19, 79
Swear words	Damn, piss, fuck	53	Informal, aggression,	58, 73, 74, 81, 98
<i>Psychological processes</i>				
Social processes	Mate, talk, they, child	455	Social concerns, social support	1, 18, 23, 27, 32, 35, 41, 55, 78, 79, 85, 88, 89, 90, 93, 95, 97, 115, 116
Family	Daughter, husband	64		18, 95
Friends	Buddy, friend, neighbor	37		18, 95
Humans	Adult, baby, boy	61		1, 11
Affective processes	Happy, cried, abandon	915	Emotionality	12, 27, 28, 32, 33, 34, 40, 44, 50, 54, 57, 58, 60, 62, 69, 77, 85, 86, 119

Category	Examples	Words in Category	Psychological Correlates	Published Articles
Positive emotion	Love, nice, sweet	406		2, 3, 4, 5, 6, 8, 10, 12, 15, 17, 21, 22, 23, 25, 28, 30, 31, 33, 36, 37, 38, 41, 45, 46, 47, 48, 49, 50, 51, 53, 54, 55, 57, 59, 60, 61, 62, 64, 66, 67, 68, 69, 70, 71, 73, 74, 75, 76, 77, 81, 82, 85, 89, 91, 93, 94, 96, 99, 107, 108, 109, 110, 113, 115, 117, 118
Negative emotion	Hurt, ugly, nasty	499		2, 3, 4, 6, 10, 12, 13, 16, 17, 20, 21, 22, 25, 28, 29, 30, 31, 33, 35, 37, 40, 44, 45, 46, 47, 48, 50, 51, 52, 53, 55, 57, 59, 61, 62, 63, 64, 66, 67, 70, 71, 72, 73, 74, 76, 79, 80, 81, 82, 84, 85, 89, 91, 92, 93, 94, 96, 99, 102, 107, 113, 115, 117, 119, 121
Anxiety	Worried, nervous	91		6, 28, 50, 66, 68, 77, 84, 85, 92
Anger	Hate, kill, annoyed	184		6, 28, 33, 50, 58, 66, 72, 74, 92
Sadness	Crying, grief, sad	101		6, 28, 33, 38, 50, 63, 66, 77, 84, 90
Cognitive processes	Cause, know, ought	730		2, 3, 5, 8, 13, 18, 21, 23, 31, 32, 34, 46, 47, 49, 55, 58, 61, 68, 69, 71, 75, 83, 84, 85, 86, 89, 92, 93, 102, 104, 119, 120
Insight	Think, know, consider	195		1, 4, 18, 19, 25, 35, 37, 45, 53, 59, 68, 73, 76, 89, 90, 91, 92, 93, 97, 99, 111, 113, 115, 118, 119, 121
Causation	Because, effect, hence	108		10, 13, 16, 20, 35, 37, 39, 45, 53, 72, 76, 89, 90, 91, 93, 97, 99, 115, 121, 122
Discrepancy	Should, would, could	76		10, 16, 18, 19, 49, 63, 74, 89, 115
Tentative	Maybe, perhaps, guess	155		18, 19, 24, 37, 38, 49, 73, 87, 89, 98, 115
Certainty	Always, never	83	Social/verbal skills, emotional stability	38
Inhibition	Block, constrain, stop	111		1, 16, 18, 19, 49, 90, 111
Inclusive	And, with, include	18		41, 60, 73, 74, 89, 115
Exclusive	But, without, exclude	17	Cognitive complexity, honesty	24, 49, 73, 80, 89, 92, 93, 115

Category	Examples	Words in Category	Psychological Correlates	Published Articles
Perceptual processes	Observing, heard, feeling	273		14, 37, 120
See	View, saw, seen	72		36
Hear	Listen, hearing	51		13, 41
Feel	Feels, touch	75		13, 88
Biological processes	Eat, blood, pain	567		36
Body	Cheek, hands, spit	180		34, 36, 37, 49, 116
Health	Clinic, flu, pill	236		
Sexual	Horny, love, incest	96		36, 94, 96, 112
Ingestion	Dish, eat, pizza	111		68, 94
Relativity	Area, bend, go	638		49, 110
Motion	Arrive, car, go	168		14, 37, 80
Space	Down, in, thin	220		14, 120
Time	End, until, season	239		1, 13, 41, 64, 93, 119, 120
<i>Personal concerns</i>				
Work	Job, majors, xerox	327		36
Achievement	Earn, hero, win	186		36, 60, 103
Leisure	Cook, chat, movie	229		
Home	Apartment, kitchen, family	93		79
Money	Audit, cash, owe	173		
Religion	Altar, church, mosque	159		41, 94
Death	Bury, coffin, kill	62		1, 2, 4, 35, 64, 68, 91, 94
<i>Spoken categories</i>				
Assent	Agree, OK, yes	30	Agreement, passivity	48, 60, 81
Nonfluencies	Er, hm, umm	8		74
Fillers	Blah, I mean, yaknow	9	Informal, Unprepared speech	9, 74

Close Relationships

Pronoun use is very important in showing the quality of a close relationship, because it shows how individuals are referring to each other. Surprisingly, first-person plural (“we”) has not been found to be related to higher relationship quality, instead use of second person (“you”) is more important in predicting lower-quality relationships. Simmons, Chambless, and Gordon (2008) found that use of second-person pronouns was negatively related to relationship quality. They found in a study of relatives of participants suffering from either obsessive–compulsive disorder or panic attacks with agoraphobia that there were differences in the use of pronouns and that these differences signaled the extent to which they had a poor relationship with the patient.

Honesty and Deception

Deceptive statements compared with truthful ones are moderately descriptive, distanced from self, and more negative. Newman, Pennebaker, Berry, and Richards (2003) investigated lying behavior in five experiments; in each experiment, lying was operationalized differently. Across the studies when participants were lying they used more negative emotion, more motion words (e.g., arrive, car, go), fewer exclusion

Status, Dominance, and Social Hierarchy

Higher-status individuals speak more often and freely make statements that involve others. Lower-status language is more self-focused and tentative. In a study of groups of three crew members, a captain, a first lieutenant, and a second lieutenant engaging in several flight simulations, the use of greater first-person plural correlated with higher rank (Sexton & Helmreich, 2000).

Another signal of rank: starting a fashion



Status, Dominance, and Social Hierarchy

Higher-status individuals speak more often and freely make statements that involve others. Lower-status language is more self-focused and tentative. In a study of groups of three crew members, a captain, a first lieutenant, and a second lieutenant engaging in several flight simulations, the use of greater first-person plural correlated with higher rank (Sexton & Helmreich, 2000).

Is literary style like fashion? Can you track literary influence? Can you find high-status, influential people by modeling literary style?

Feature Set	Author	Nationality
Freq. Words	99.00%	93.50%
Freq. Pairs	91.60%	91.30%
Freq. Coll. (k=5)	88.94%	90.20%
Freq. Coll. (k=10)	84.00%	87.20%

most frequent 200 words

Status, Dominance, and Social Hierarchy

Higher-status individuals speak more often and freely make statements that involve others. Lower-status language is more self-focused and tentative. In a study of groups of three crew members, a captain, a first lieutenant, and a second lieutenant engaging in several flight simulations, the use of greater first-person plural correlated with higher rank (Sexton & Helmreich, 2000).

Feature Set	Author	Nationality
Freq. Words	99.00%	93.50%
Freq. Pairs	91.60%	91.30%
Freq. Coll. (k=5)	88.94%	90.20%
Freq. Coll. (k=10)	84.00%	87.20%

most frequent 200 words

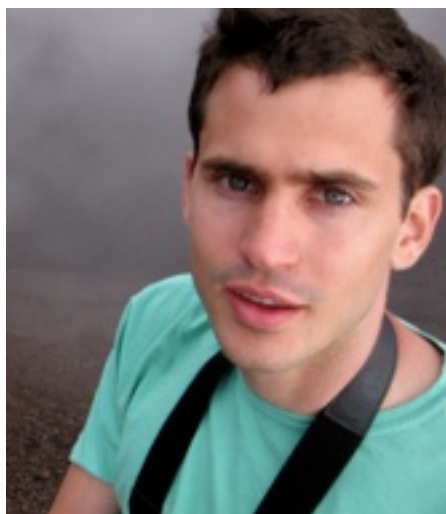
Mark My Words!

Linguistic Style Accommodation in Social Media

Cristian Danescu-Niculescu-Mizil*
Cornell University
Ithaca, NY 14853, USA
cristian@cs.cornell.edu

Michael Gamon
Microsoft Research
Redmond, WA 98053, USA
mgamon@microsoft.com

Susan Dumais
Microsoft Research
Redmond, WA 98053, USA
sdumais@microsoft.com



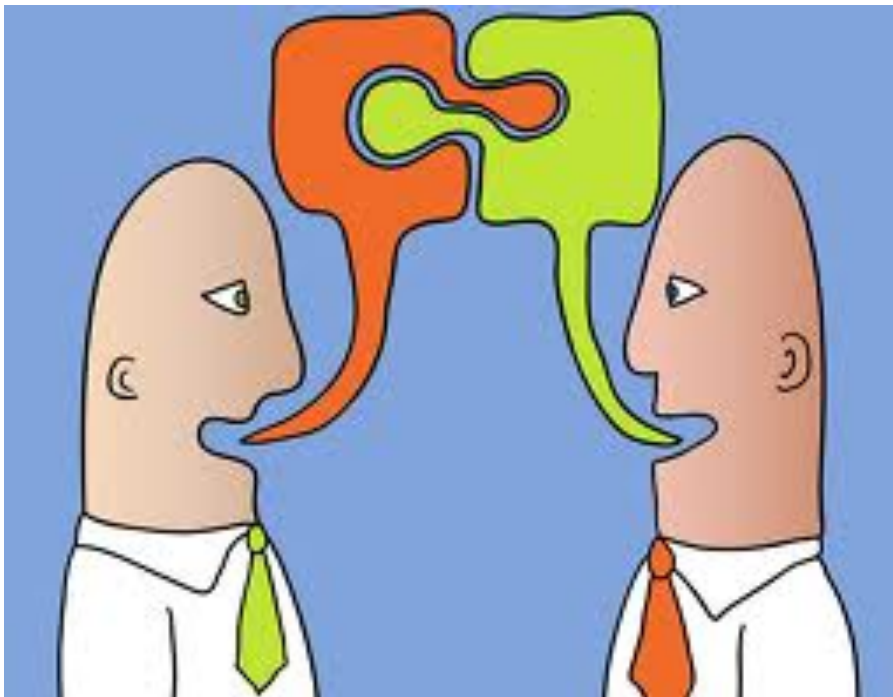
Mark My Words!

Linguistic Style Accommodation in Social Media

Cristian Danescu-Niculescu-Mizil*
Cornell University
Ithaca, NY 14853,USA
cristian@cs.cornell.edu

Michael Gamon
Microsoft Research
Redmond, WA 98053,USA
mgamon@microsoft.com

Susan Dumais
Microsoft Research
Redmond, WA 98053,USA
sdumais@microsoft.com



People adopt each other's mannerisms and style in many ways....

Feature Set	Author	Nationality
Freq. Words	99.00%	93.50%
Freq. Pairs	91.60%	91.30%
Freq. Coll. (k=5)	88.94%	90.20%
Freq. Coll. (k=10)	84.00%	87.20%

Mark My Words!

Linguistic Style Accommodation in Social Media

Cristian Danescu-Niculescu-Mizil*
Cornell University
Ithaca, NY 14853, USA
cristian@cs.cornell.edu

Michael Gamon
Microsoft Research
Redmond, WA 98053, USA
mgamon@microsoft.com

Susan Dumais
Microsoft Research
Redmond, WA 98053, USA
sdumais@microsoft.com

Table 1: Examples of dimensions for which accommodation was observed and the respective studies.

Dimension	Canonical study
Posture	Condon and Ogston, 1967
Pause length	Jaffe and Feldstein, 1970
Utterance length	Matarazzo and Wiens, 1973
Self-disclosure	Derlenga et al., 1973
Head nodding	Hale and Burgoon, 1984
Backchannels	White, 1989
Linguistic style	Niederhoffer and Pennebaker, 2002



Corpus

- Pennebaker & Niederhoffer, 2002:
 - 98 pairs in the lab + Watergate tapes
- Twitter A:
 - 1.3M “conversations” between 300k users--many are too short to analyze successfully
- Twitter B: More crawling
 - all pairs with 2+ conversations
 - *all* posts from these pairs
 - 15M tweets, 7800 users, 215k conversations, 2200 pairs

	Mean	Median	Min	Max
Number of conversations	98	60	1	1744
Average number of turns	2.7	2.6	2	16.8
Days of contact	270	257	1	886

Measuring “cohesion” for a property C

Dimension	Examples	Size
Article	an, the	3
Certainty	always, never	83
Conjunction	but, whereas	28
Discrepancy	should, would	76
Exclusive	without, exclude	17
Inclusive	with, include	18
Indefinite pronoun	it, those	46
Negation	not, never	57
Preposition	to, with	60
Quantifier	few, much	89
Tentative	maybe, perhaps	155
1st person singular pronoun	I, me	12
1st person plural pronoun	we, us	12
2nd person pronoun	you, your	20

Measuring “cohesion”

Dimension	Examples	Size
Article	an, the	3
Certainty	always, never	83
Conjunction	but, whereas	28
Discrepancy	should, would	76
Exclusive	without, exclude	17

$$Coh(C) \triangleq P(T^C \wedge R^C \mid T \leftrightarrow R) - P(T^C \wedge R^C)$$

Tweet T
contains word
from class C

Reply R
contains word
from class C

T and R are a
“turn”

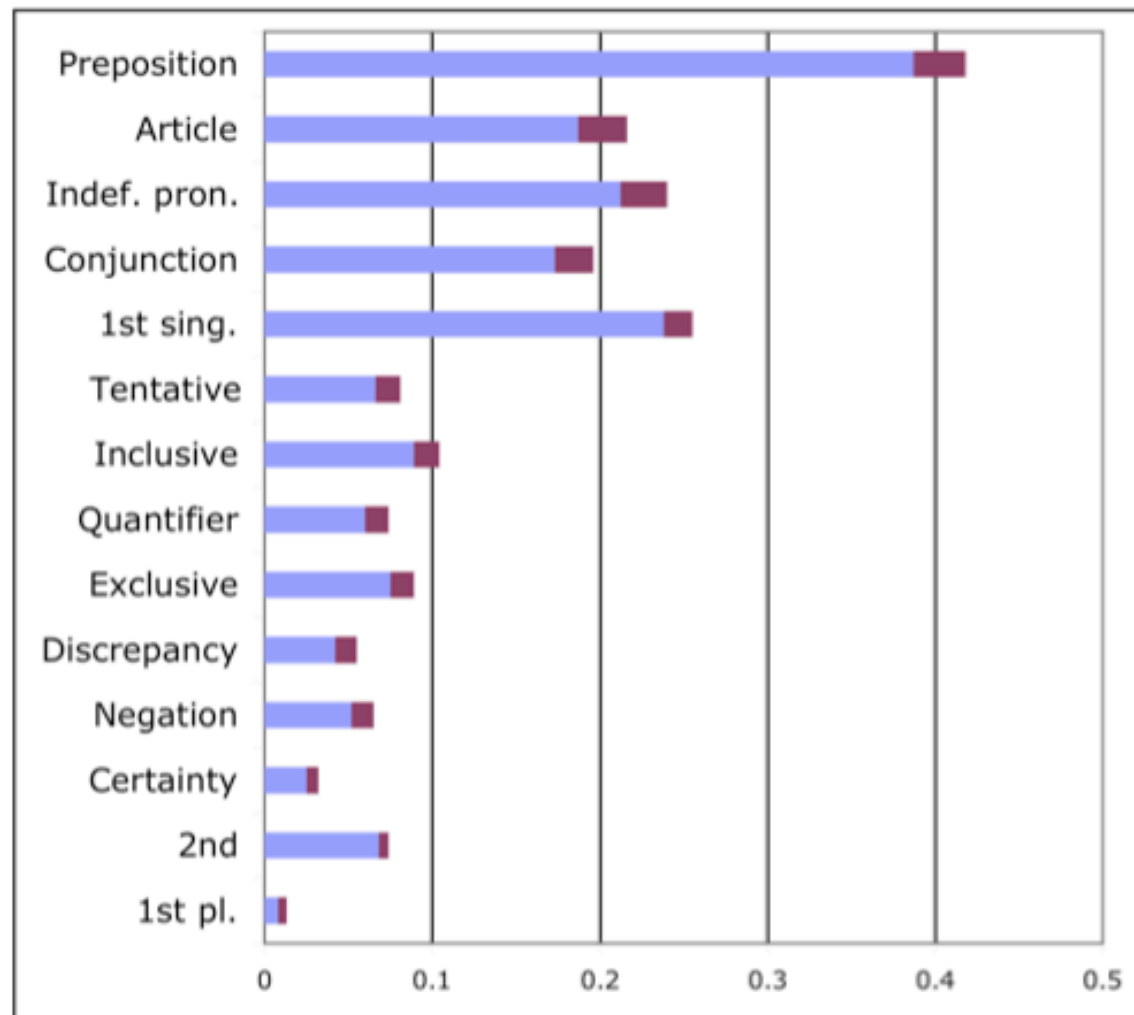
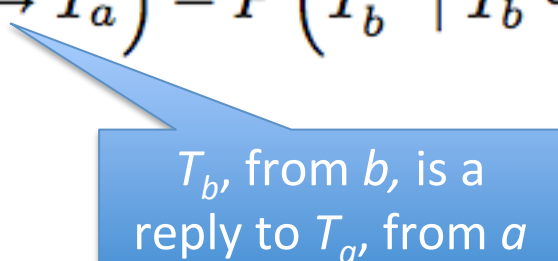


Figure 1: The effect of stylistic cohesion observed as the difference between $\hat{P}(T^C \wedge R^C \mid T \leftrightarrow R)$ (composite bars) and $\hat{P}(T^C \wedge R^C)$ (blue bars). The differences, shown in red/dark, are statistically significant ($p < 0.0001$). The dimensions are shown in decreasing order of the difference.

Measuring “accommodation” and “influence”

Formally, for a given stylistic dimension C and pair of users (a, b) , the accommodation of user b to user a is measured by how much the fact that user a exhibits C in a tweet T_a increases the probability of b to also exhibit C in a reply to T_a :

$$Acc_{(a,b)}(C) \triangleq P\left(T_b^C \mid T_a^C, T_b \hookrightarrow T_a\right) - P\left(T_b^C \mid T_b \hookrightarrow T_a\right) \quad (2)$$



T_b , from b , is a reply to T_a , from a

$$Acc(C) = E[Acc_{(a,b)}(C)]$$

$$I_{(a,b)}(C) \triangleq Acc_{(a,b)}(C) - Acc_{(b,a)}(C)$$

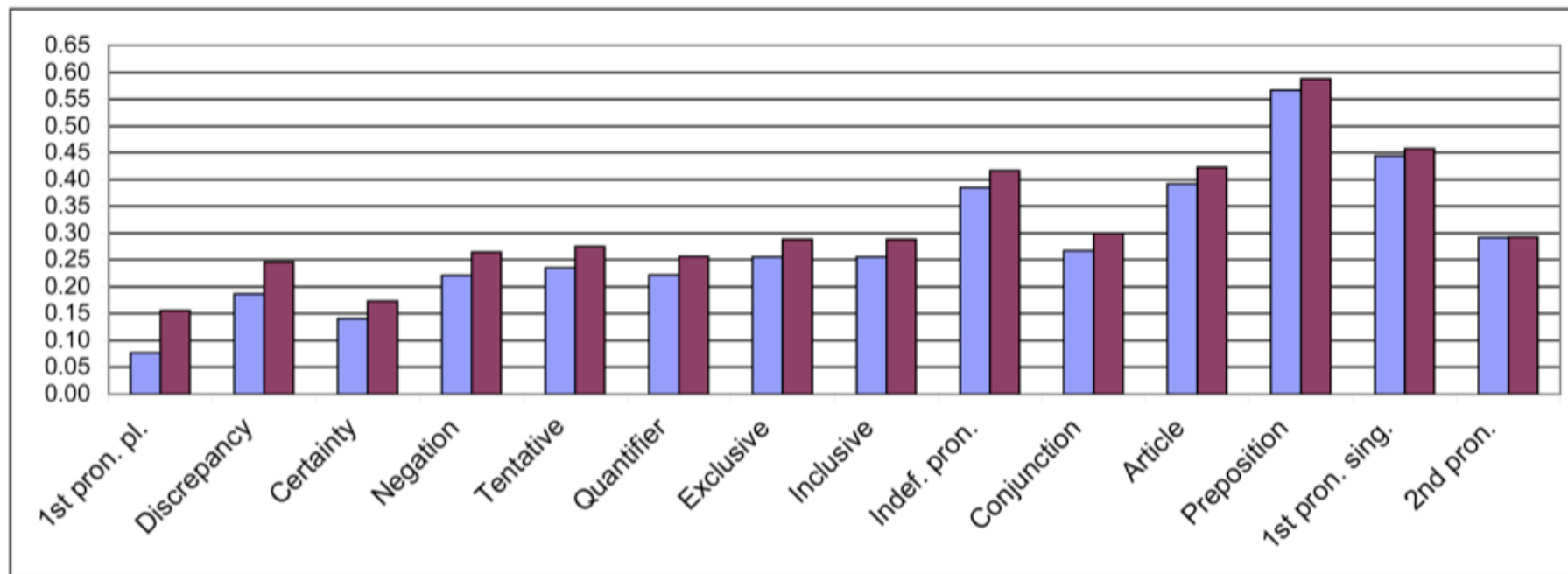
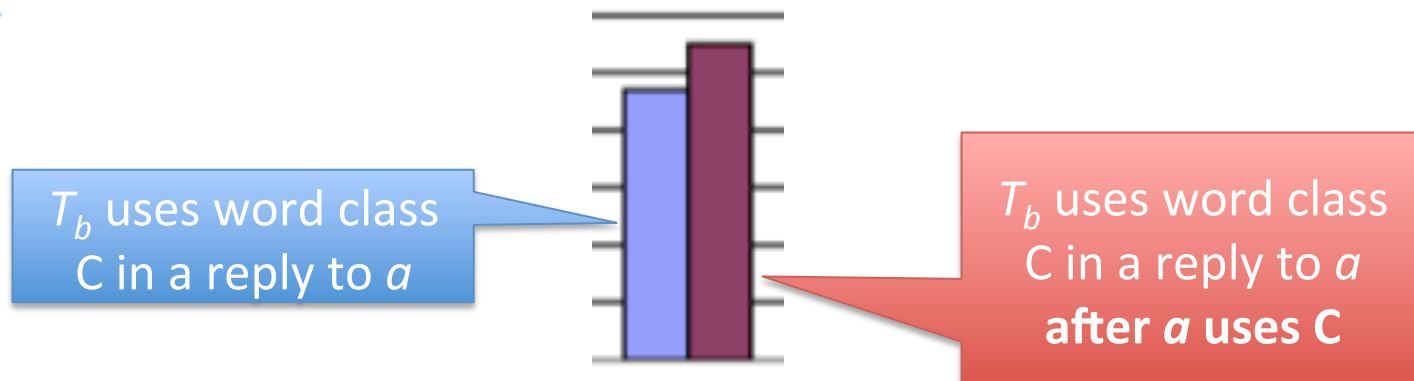
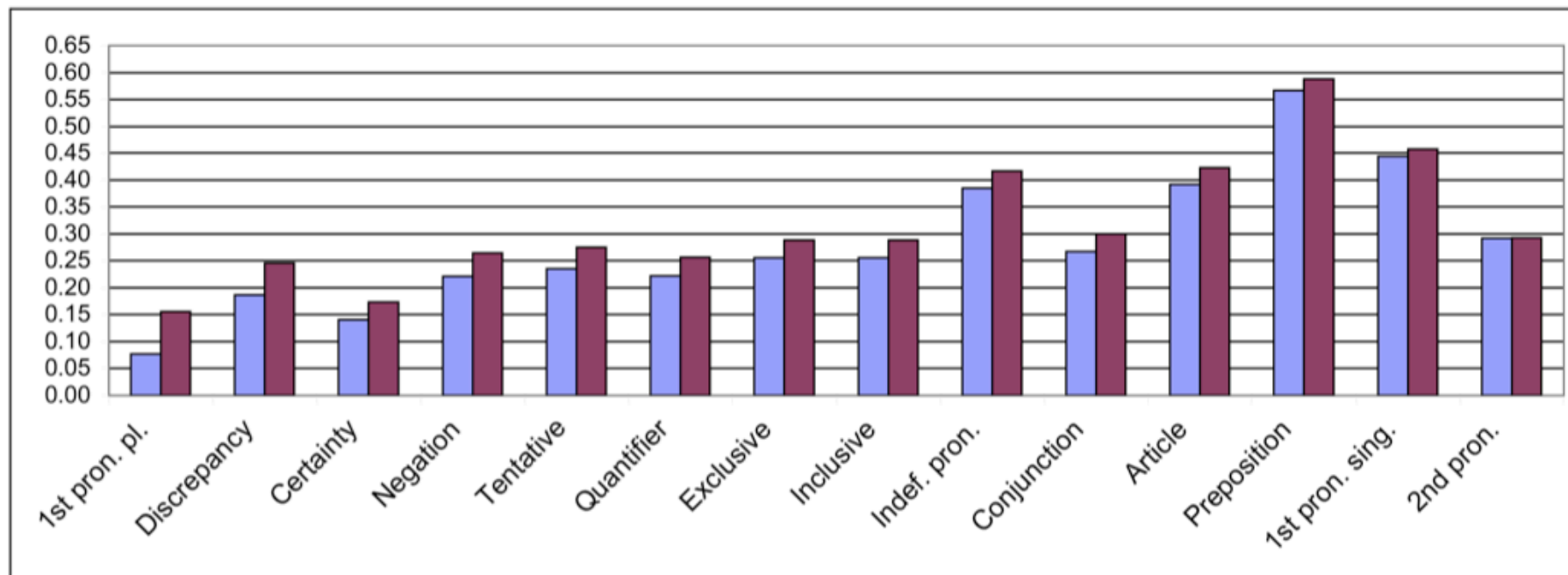


Figure 2: The effect of accommodation $\widehat{Acc}(C)$ for each strictly non-topical stylistic dimension C observed as the difference between the means of $\{\hat{P}(T_b^C | T_b \hookrightarrow, T_a) \mid (a, b) \in \text{Pairs}\}$ (blue, left) and $\{\hat{P}(T_b^C | T_a^C, T_b \hookrightarrow T_a) \mid (a, b) \in \text{Pairs}\}$ (red, right). All the differences are statistically significant ($p < 0.0001$), except for the *2nd person pronoun* category. The dimensions are ordered according to the amount of accommodation observed.

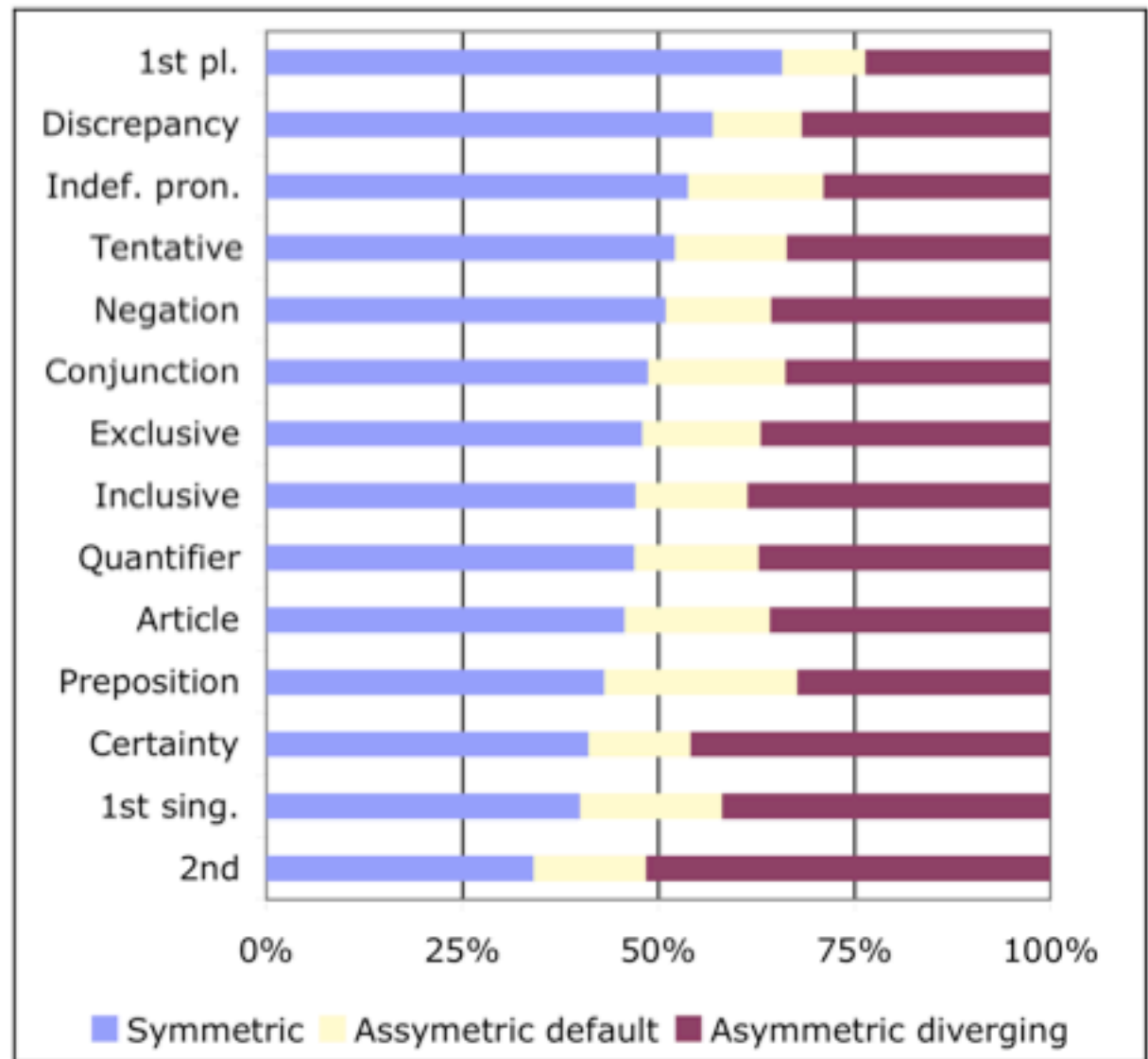




- Evidence of *fashion* in linguistic style spreading through a conversation
- Time lag suggests *influence* not *associative sorting*
- We don't have anything like direction.....

If $\text{Acc}(a,b) > 0$:

- Symmetric:
 $\text{Acc}(b,a) > 0$
- Default
asymmetric:
 $\text{Acc}(b,a) = 0$
- Divergent
asymmetric:
 $\text{Acc}(b,a) < 0$



Does one party accommodate more than the other?

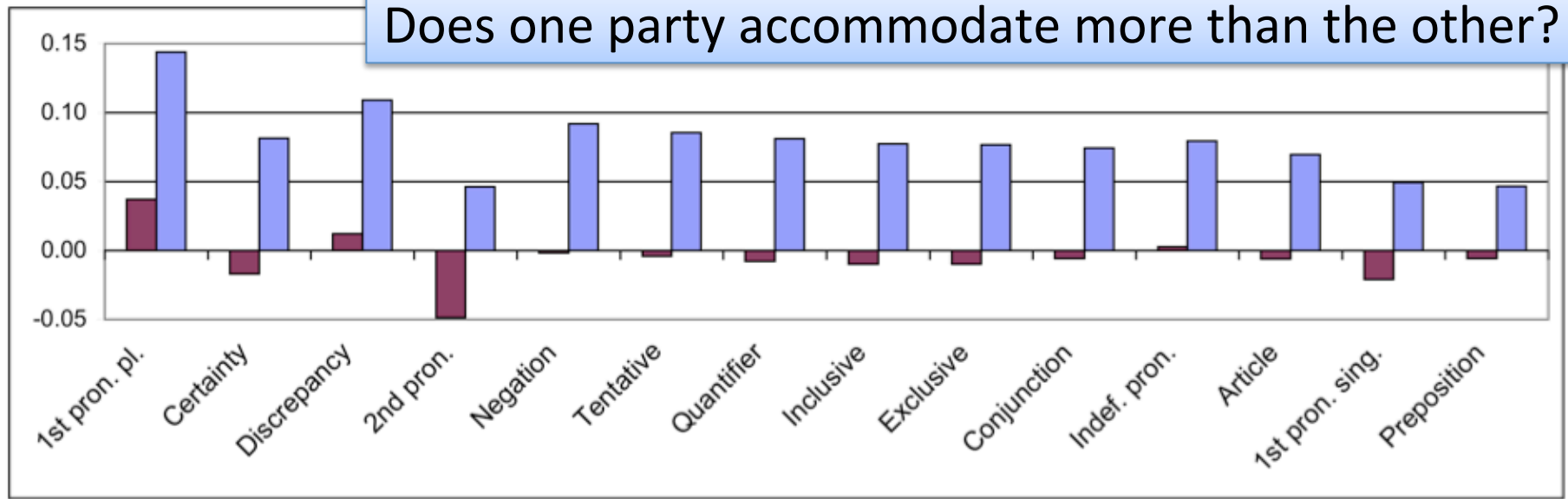


Figure 3: The effect of stylistic influence for each strictly non-topical stylistic dimension C observed as the difference between the means of $\left\{ \min \left(\widehat{Acc}_{(a,b)}(C), \widehat{Acc}_{(b,a)}(C) \right) \mid (a,b) \in \text{Pairs} \right\}$ (red, left) and $\left\{ \max \left(\widehat{Acc}_{(a,b)}(C) \right) \mid (a,b) \in \text{Pairs} \right\}$ (blue, right). All the differences are statistically significant ($p < 0.0001$). The dimensions are shown in decreasing order of the difference.

Accommodation does *not* correlate with “status” features like #followers, #days on Twitter,

Does one party accommodate more than the other?

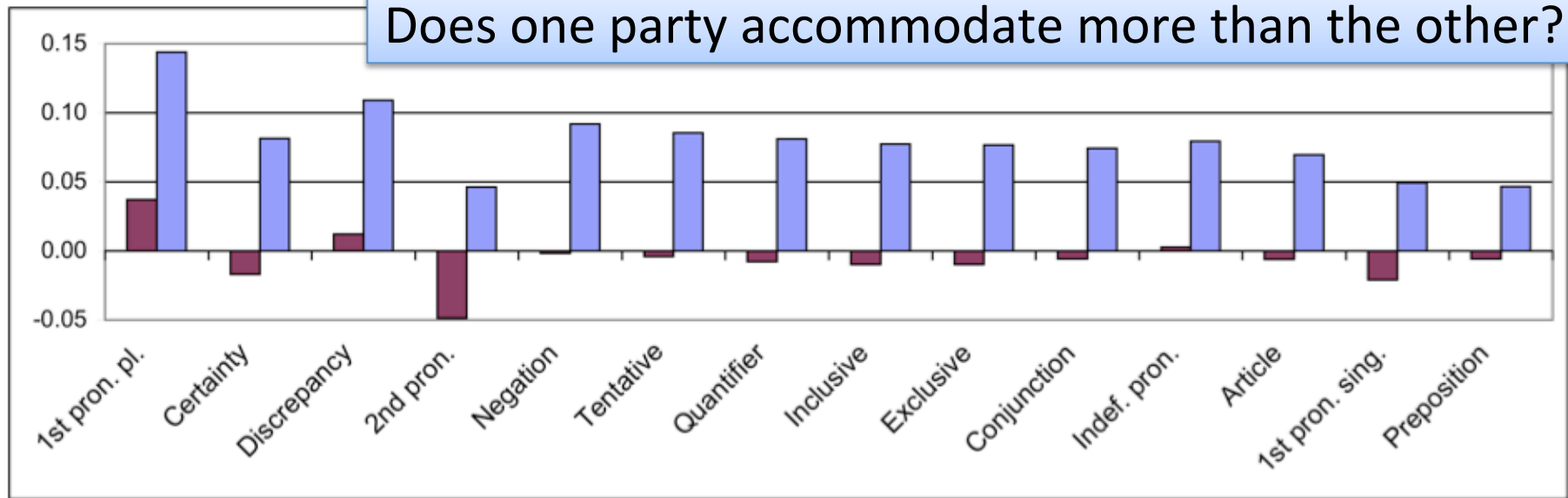


Figure 3: The effect of stylistic influence for each strictly non-topical stylistic dimension C observed as the difference between the means of $\left\{ \min \left(\widehat{Acc}_{(a,b)}(C), \widehat{Acc}_{(b,a)}(C) \right) \mid (a,b) \in \text{Pairs} \right\}$ (red, left) and $\left\{ \max \left(\widehat{Acc}_{(a,b)}(C) \right) \mid (a,b) \in \text{Pairs} \right\}$ (blue, right). All the differences are statistically significant ($p < 0.0001$). The dimensions are shown in decreasing order of the difference.

Status, Dominance, and Social Hierarchy

????

Higher-status individuals speak more often and freely make statements that involve others. Lower-status language is more self-focused and tentative. In a study of groups of three crew members, a captain, a first lieutenant, and a second lieutenant engaging in several flight simulations, the use of greater first-person plural correlated with higher rank (Sexton & Helmreich, 2000).

Echoes of Power: Language Effects and Power Differences in Social Interaction

Cristian Danescu-Niculescu-Mizil

Cornell University

cristian@cs.cornell.edu,

Lillian Lee

Cornell University

llee@cs.cornell.edu,

Bo Pang

Yahoo!

bopang@yahoo-inc.com,

Jon Kleinberg

Cornell University

kleinber@cs.cornell.edu

Datasets

- Wikipedia: wikipedia editors *talk* pages: 240k conversations; plus 32k discussions over who gets promoted to admins.
 - Status: admin vs non-admin
 - *Dependence*: learning to support/reject
- Supreme court: 50k verbal exchanges for 204 cases.
 - Status: chief justice vs justice vs lawyer
 - *Dependence*: leaning to support/learning to reject

Experiments

- Similar notion of “coordination” (=accomodation)

$$C^m(b, a) = P(\mathcal{E}_{u_2}^m \hookrightarrow u_1 \mid \mathcal{E}_{u_1}^m) - P(\mathcal{E}_{u_2}^m \hookrightarrow u_1),$$

- Hypotheses:

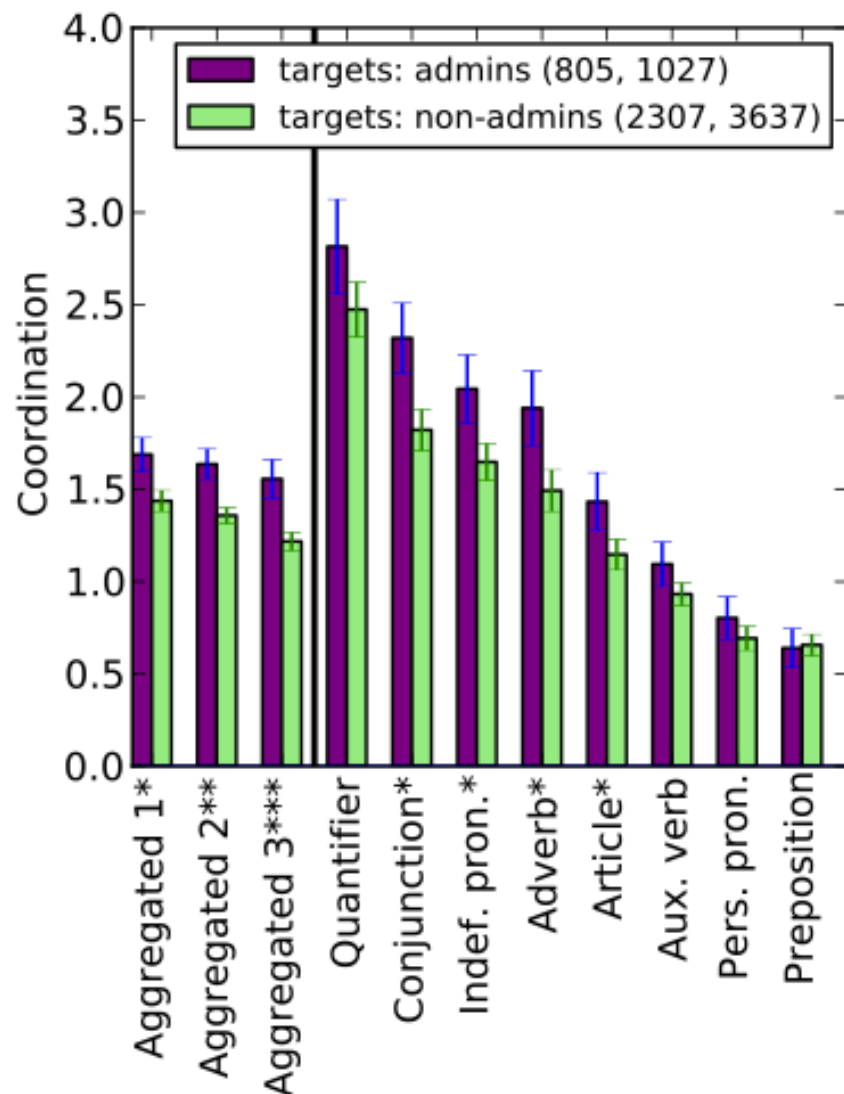
e.g., you accommodate more when speaking to a big shot

$$\mathcal{P}_{target}: C(U, G^{high}) > C(U, G^{low})$$

$$\mathcal{P}_{speaker}: C(G^{high}, U) < C(G^{low}, U)$$

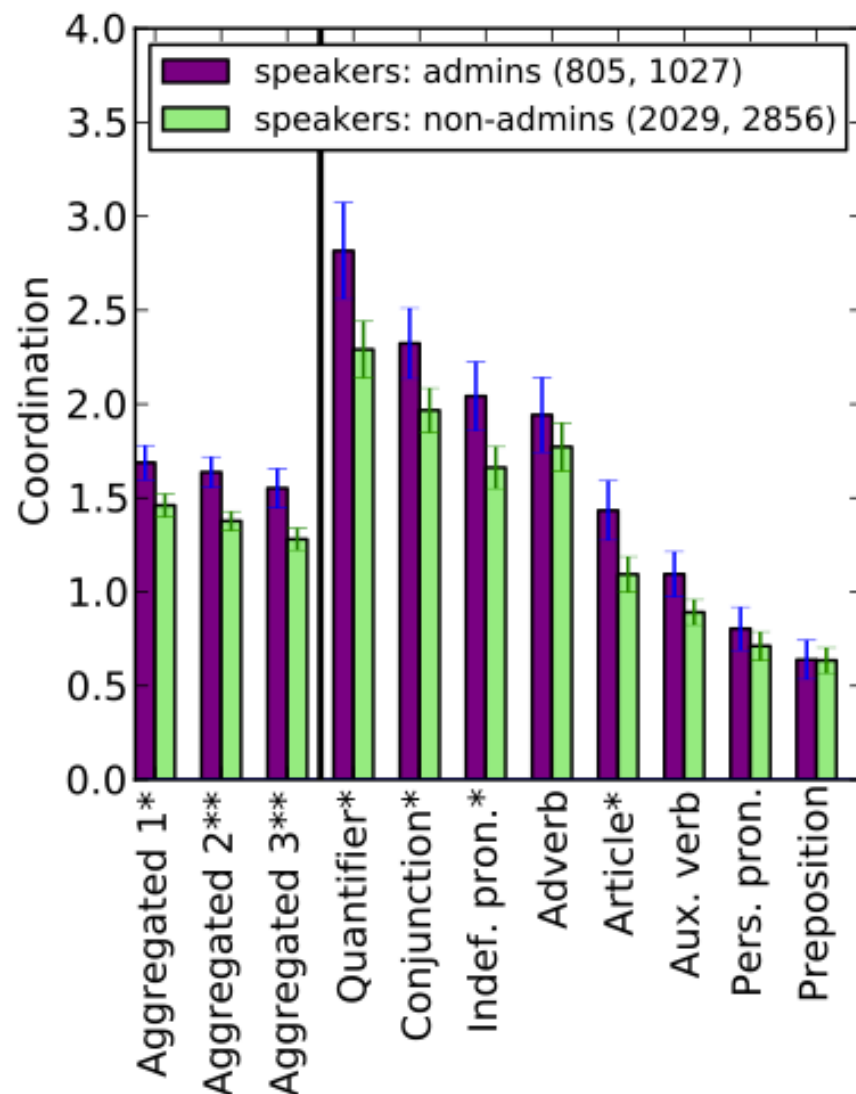
and he coordinates less with other people

	Wikipedia	
	higher power	lower power
Status	admins admins	non-admins admins-to-be (before RfAs)
Dependence	diff. vote	same vote
	Supreme Court	
	higher power	lower power
Status	Justices Chief Justices	lawyers Associate Justices
Dependence	unfavorable Justice	favorable Justice



(a) Supporting \mathcal{P}_{target}

more coordination with admins
than non-admins

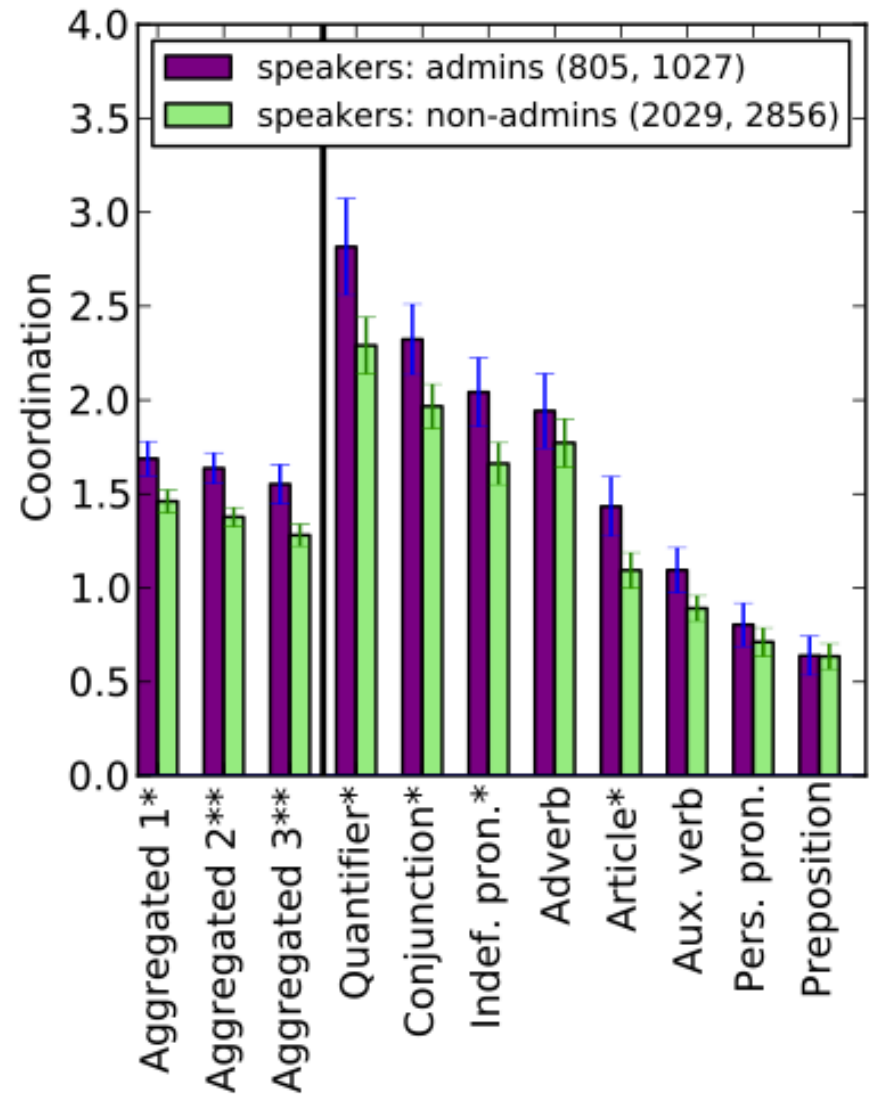


(b) Contradicting $\mathcal{P}_{speaker}$

admins coordinate **more** with
others than non-admins

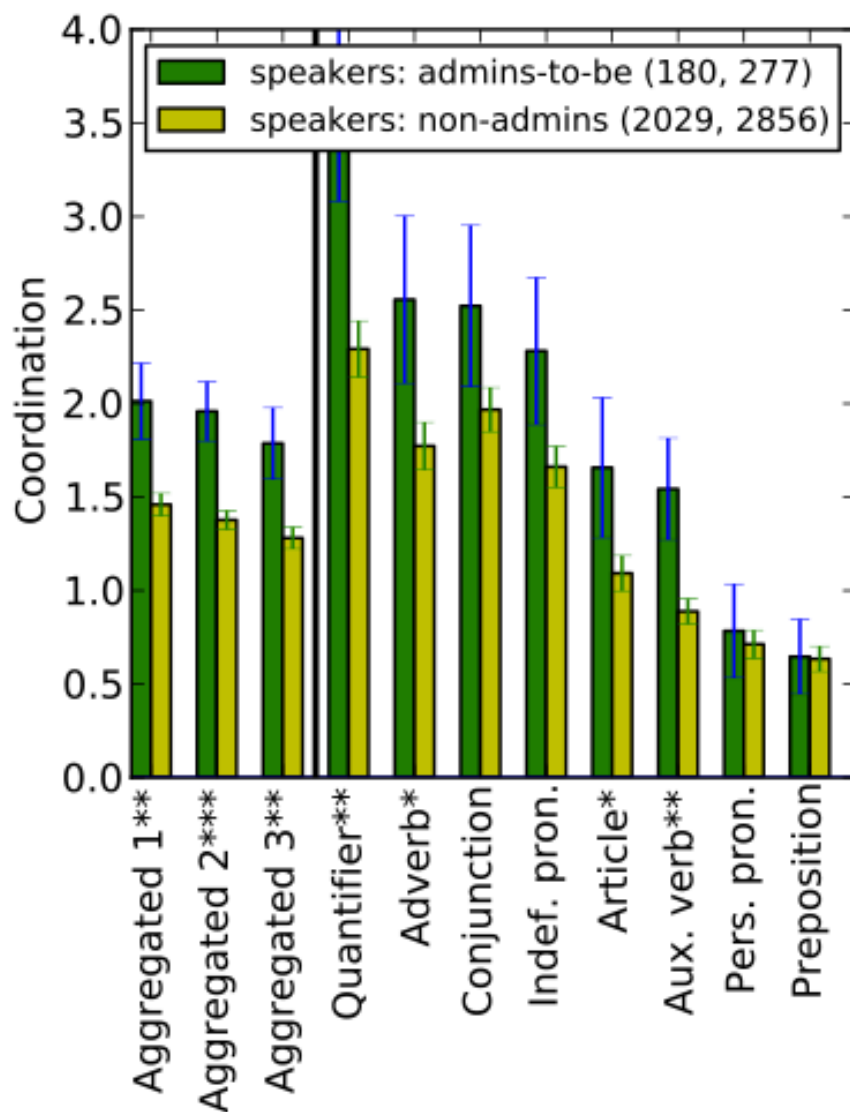
Why?

Maybe the folks that become admins are different somehow? eg more accommodating?

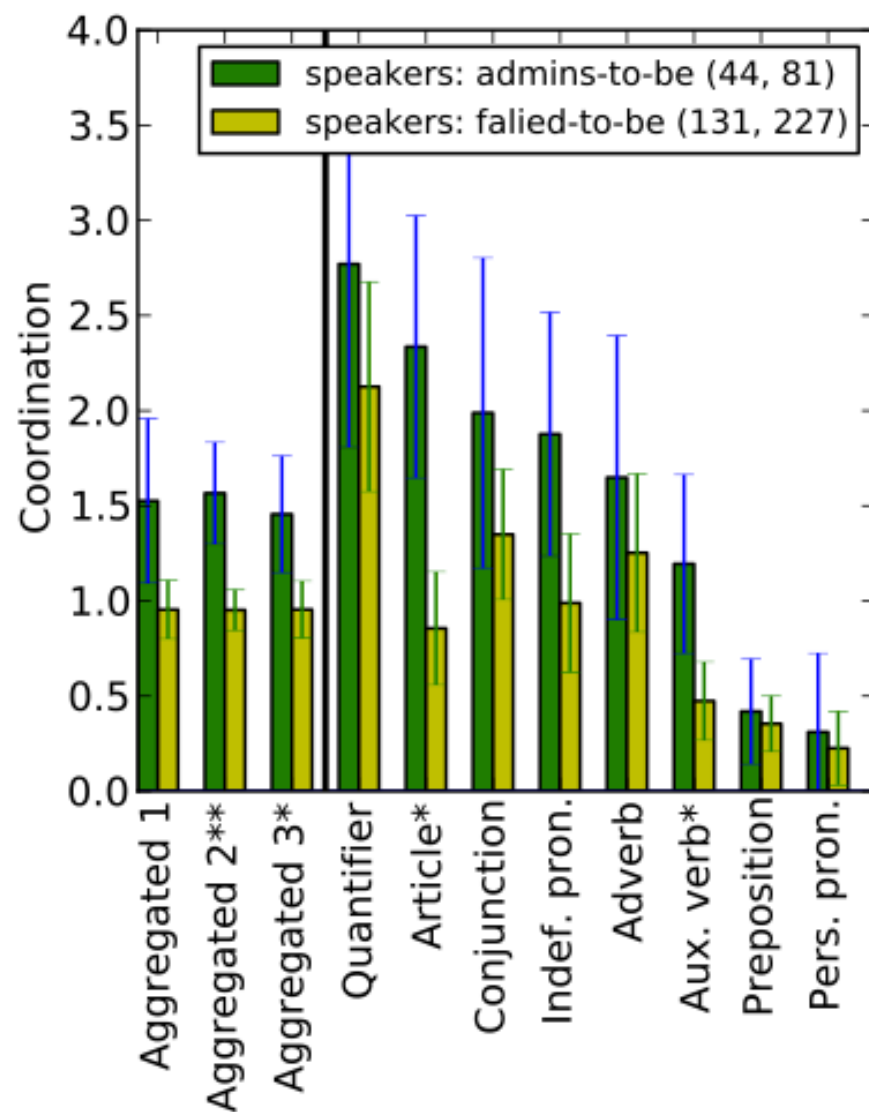


(b) Contradicting $\mathcal{P}_{speaker}$

admins coordinate **more** with others than non-admins

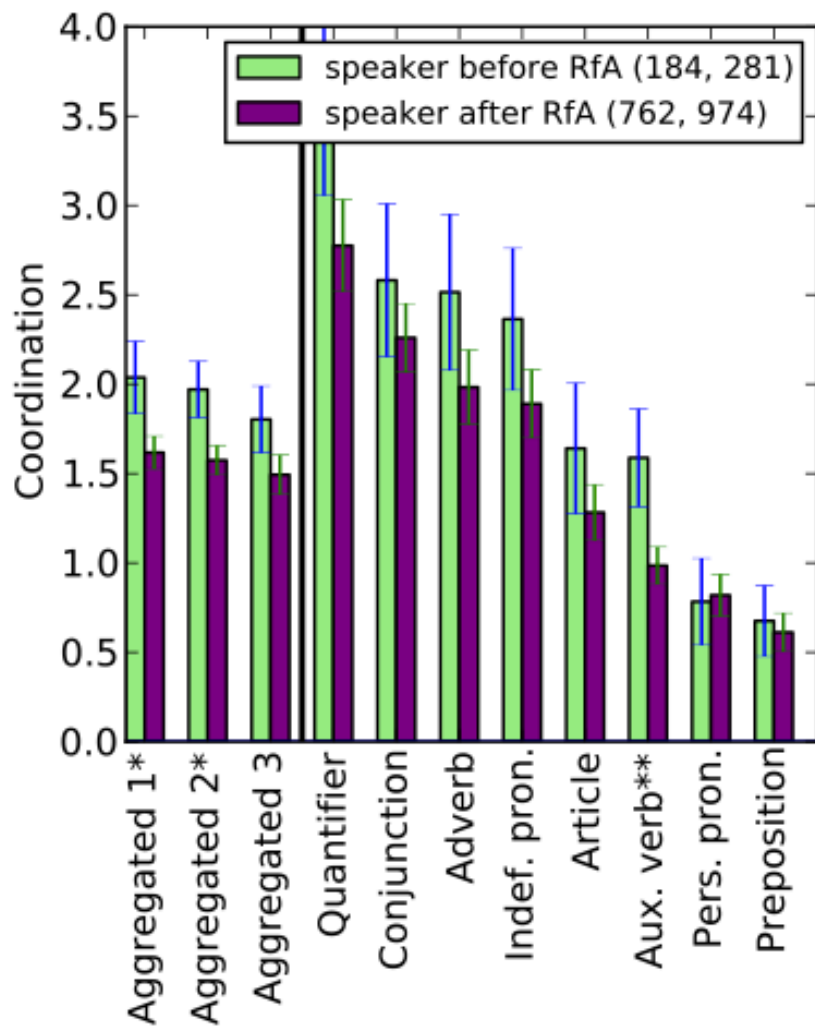


(a) Before RfA elections (\mathcal{B})

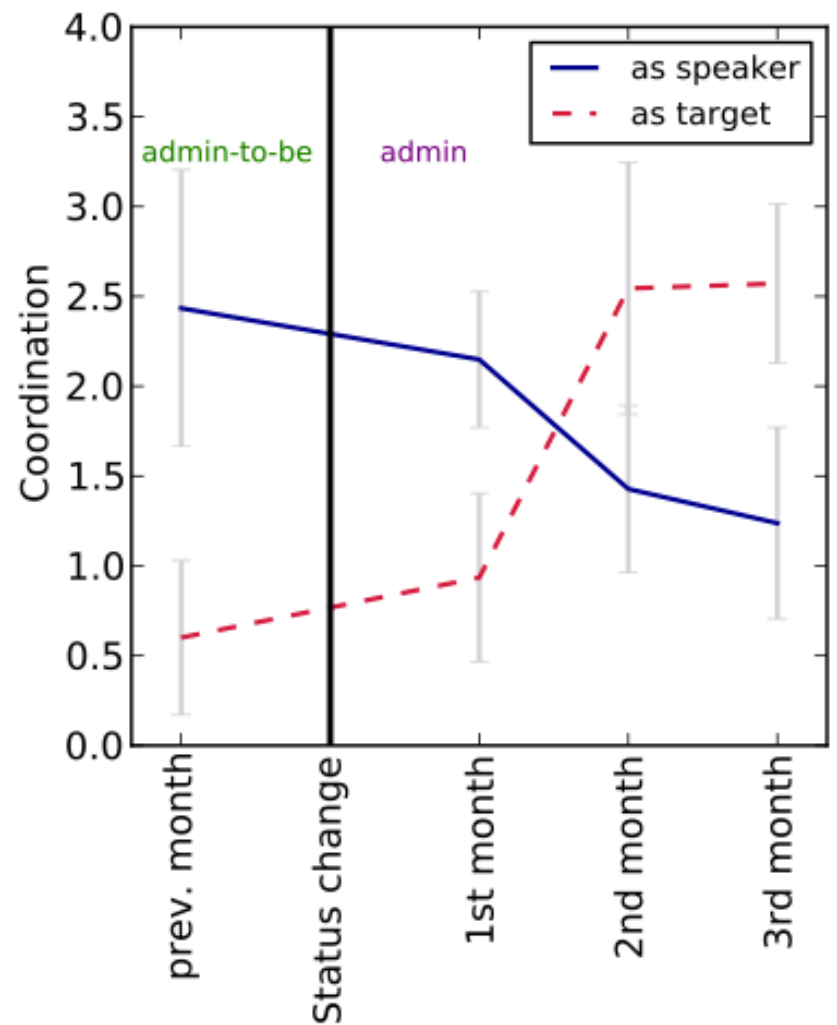


(b) In RfA discussions (\mathcal{B})

the people that **eventually become admins** coordinate more than people who **eventually fail to become admins**



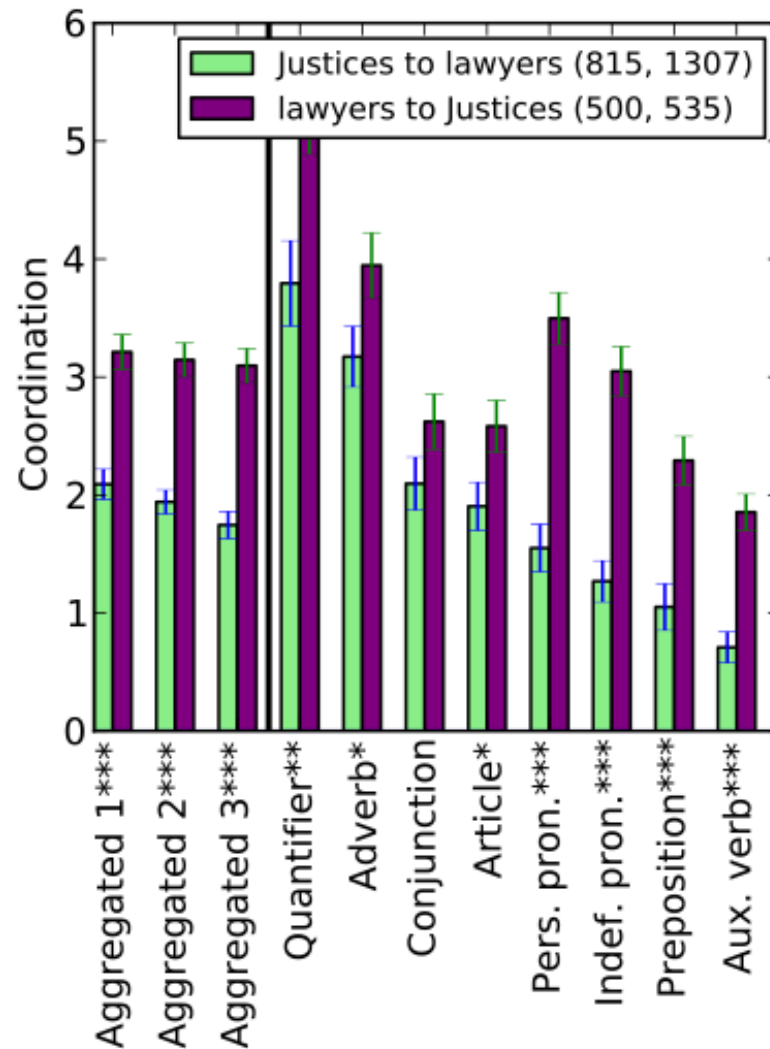
(a) Supporting $\mathcal{P}'_{speaker}$



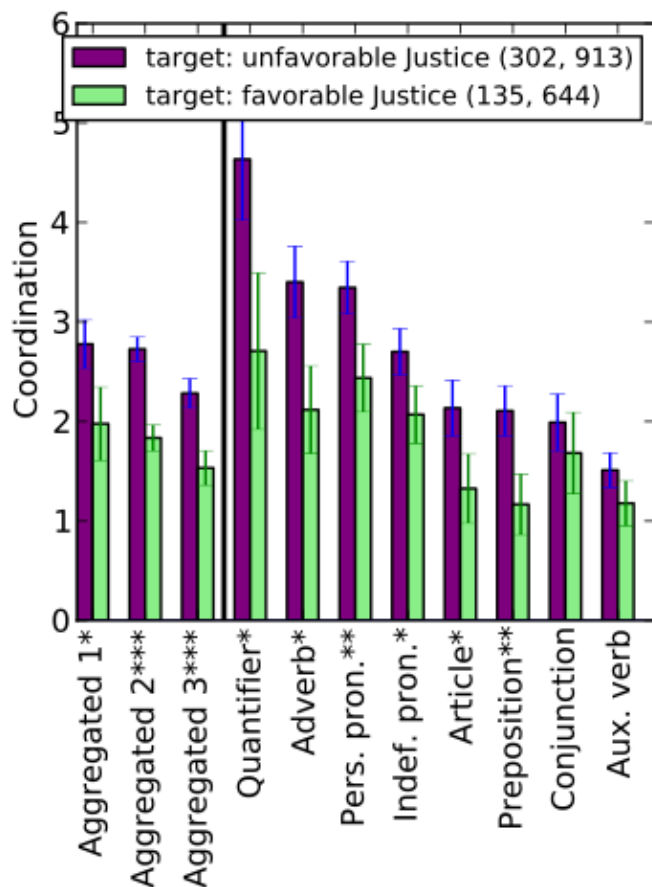
(b) Timed effect of **status change** (\mathcal{P})

revised hypothesis: **after** you become an admin you will coordinate with others **less** than you did before

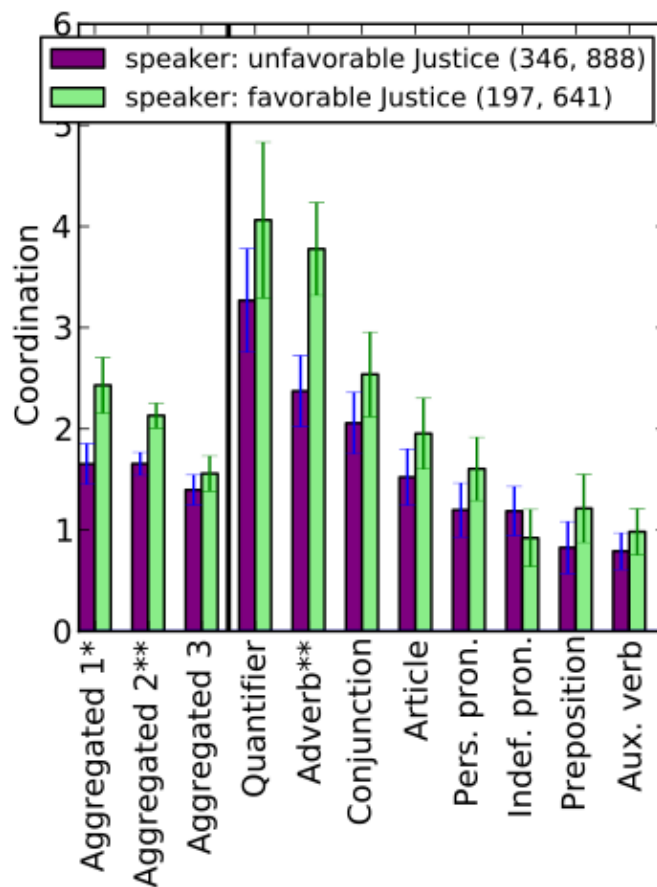
What about the court dataset?



What about the court dataset?



(a) Dependence: \mathcal{P}_{target}



(b) Dependence: $\mathcal{P}_{speaker}$

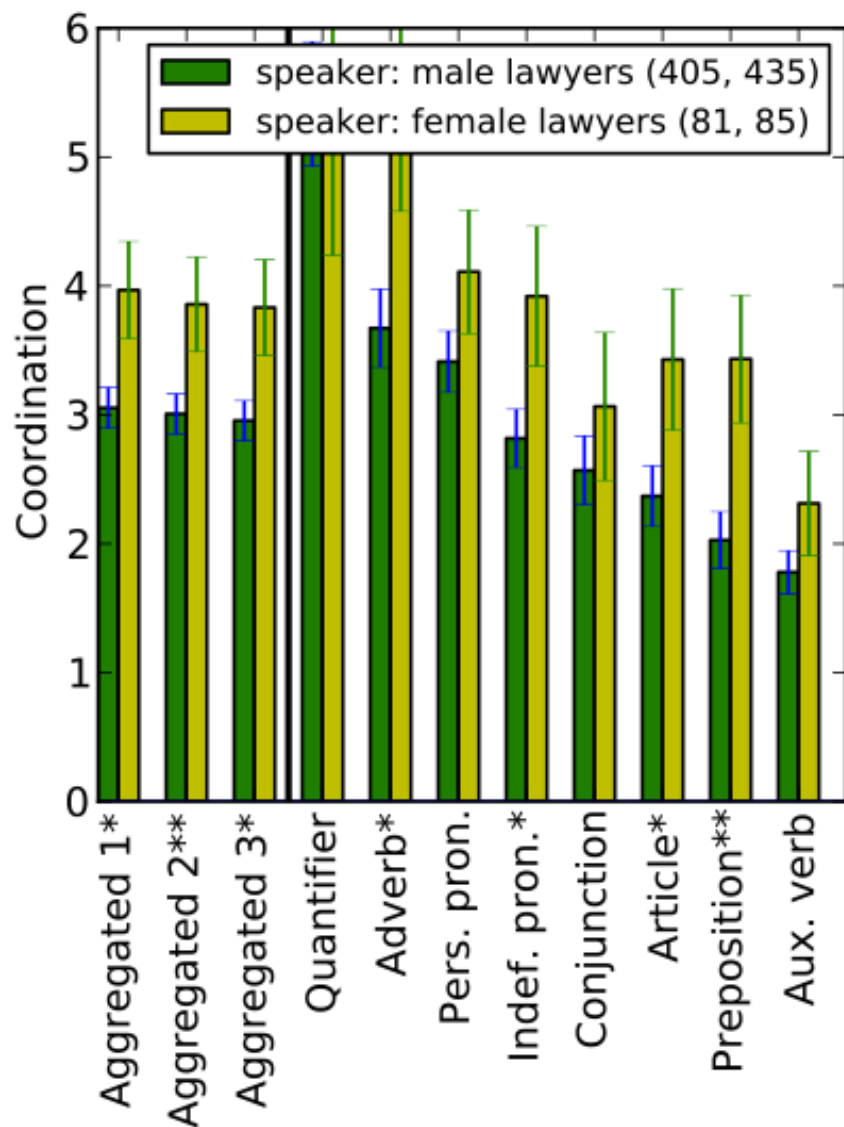
Status prediction

- Given conversation between x, y predict if $status(x) > status(y)$ or vice-versa
- Very easy to do in Supreme Court domain (“your honor,....”)
- Hard for humans in Wikipedia (inter-annotator agreement $\sim 80\%$, accuracy $\sim 70\%$)

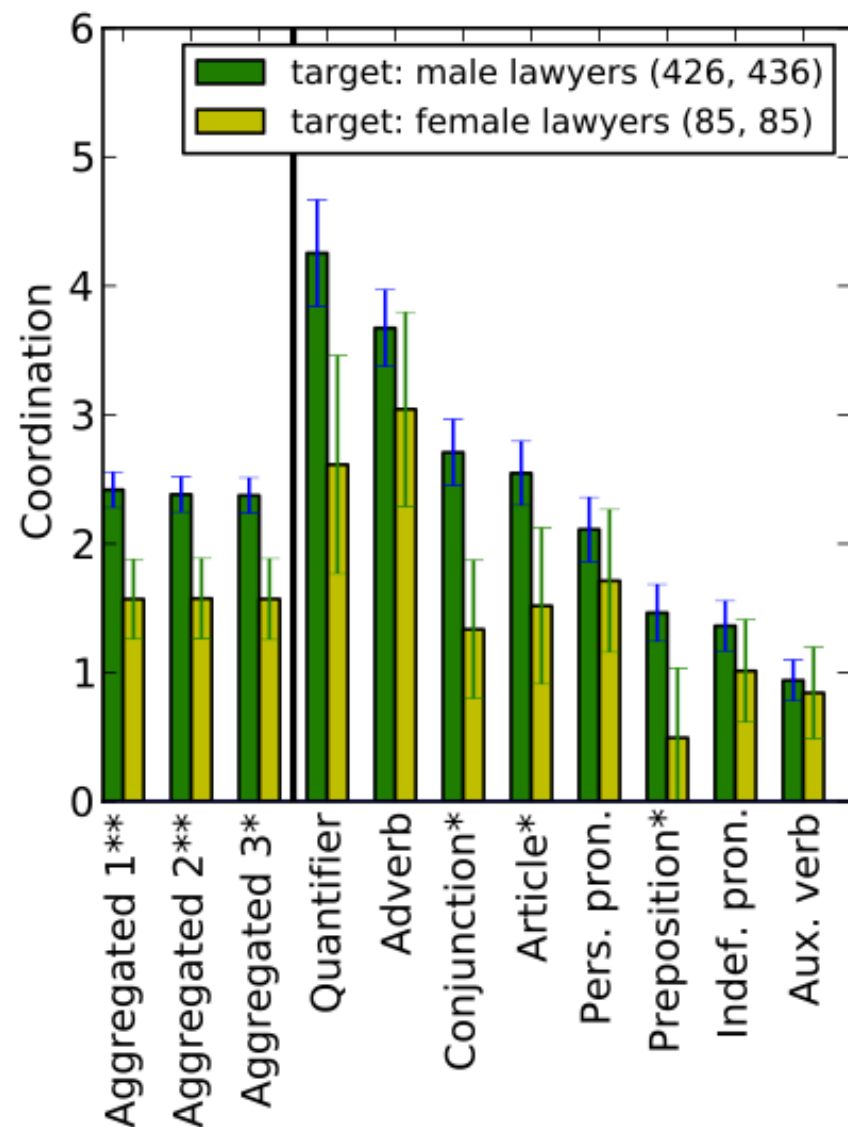
	in-domain		cross-domain	
	<i>wiki</i>	<i>court</i>	<i>court</i>	<i>wiki</i>
Training corpus	<i>wiki</i>	<i>court</i>	<i>wiki</i>	<i>court</i>
coordination features (9 altogether)	57.7	70.4	57.1	55.0
stylistic features (18 altogether)	59.2	51.4	50.0	51.9
bag of words (20,000 altogether)	51.4	99.5	45.2	40.1

Table 2: Prediction accuracy for SVM’s using various feature sets. Cross-domain results are in the right-hand two columns. Bold = results significantly better than chance.

One more observation...



(a) gender differences in speakers



(b) gender differences in targets

So to summarize...

- Summary: there are signals in common words
 - Even though we don't think about how we use them
 - Patterns of usage ➔ "literary style"
 - predicts: authorship, gender, ...
 - Style changes according to situation
 - and is transmitted from person to person
 - you can observe that transmission (accommodation, coordination) and determine its direction
 - the direction of accommodation it tells you something about the status of the speakers