

# Rhetorical Patterns in Legislative Speech

Jacob Eisenstein  
@jacobeisenstein

Georgia Institute of Technology

October 17, 2015



# Models of discourse structure

**Discourse** refers to the broad range of linguistic phenomena beyond the sentence level.

- ▶ Well-structured texts are said to be **coherent**.
- ▶ Coherence is typically (not always!) characterized in terms of **discourse relations**.
- ▶ Discourse relations elucidate structural connections between segments of text.

(1) Well, first of all, Rand Paul shouldn't even be on this stage. He's number 11, he's got 1 percent in the polls...

As far as temperament, I think I have a great temperament. I built a phenomenal business with incredible, iconic assets, one of the really truly great real-estate businesses.

And I may be an entertainer, because I've had tremendous success with number-one bestsellers all over the place, with "The Apprentice" and everything else I've done.

But I will tell you this: What I am far and away greater than an entertainer is a businessman.

(2) **Well, first of all**, Rand Paul shouldn't even be on this stage. He's number 11, he's got 1 percent in the polls...

**As far as** temperament, I think I have a great temperament. I built a phenomenal business with incredible, iconic assets, one of the really truly great real-estate businesses.

**And** I may be an entertainer, **because** I've had tremendous success with number-one bestsellers all over the place, with "The Apprentice" and everything else I've done.

**But** I will tell you this: What I am far and away greater than an entertainer is a businessman.

- Discourse relations may be marked with explicit connectives.

- (3) Well, first of all, **Rand Paul shouldn't even be on this stage. He's number 11, he's got 1 percent in the polls...**

As far as temperament, I think I have a great temperament. I built a phenomenal business with incredible, iconic assets, one of the really truly great real-estate businesses.

And I may be an entertainer, because I've had tremendous success with number-one bestsellers all over the place, with "The Apprentice" and everything else I've done.

But I will tell you this: What I am far and away greater than an entertainer is a businessman.

- ▶ Other relations are implicit, particularly when the relation is semantically "obvious" (Asr & Demberg, 2013)

- (4) Well, first of all, Rand Paul shouldn't even be on this stage. He's number 11, he's got 1 percent in the polls...

As far as temperament, I think **I have a great temperament. I built a phenomenal business with incredible, iconic assets**, one of the really truly great real-estate businesses.

And I may be an entertainer, because I've had tremendous success with number-one bestsellers all over the place, with "The Apprentice" and everything else I've done.

But I will tell you this: What I am far and away greater than an entertainer is a businessman.

- ▶ Other relations are implicit, particularly when the relation is semantically "obvious" (Asr & Demberg, 2013)

**Research question:** Is the frequency of discourse relations affected by political covariates?



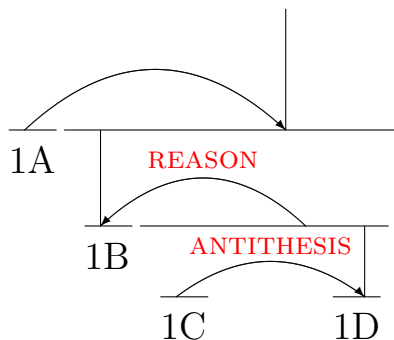
**Research question:** Is the frequency of discourse relations affected by political covariates?

**Rhetorical Structure Theory** Discourse units are built compositionally from smaller elements, into a tree structure.

**Penn Discourse Treebank** Discourse relations hold between arbitrary spans of text; no global document structure.

Applied to Convote transcripts of 3000 congressional floor speeches (Thomas et al., 2006).

# Rhetorical Structure Theory (RST)



[It is quite obvious that]<sup>1A</sup> [not enacting these reforms]<sup>1B</sup> [does not come from a lack of ability,]<sup>1C</sup> [but from a lack of will.]<sup>1D</sup>

- ▶ Discourse relations build combine “elementary” discourse units into larger structures.
- ▶ The **nucleus** is the more important element in each relation.
- ▶ Interrater agreement on relations is 65%  
Best systems get 60% accuracy (Ji & Eisenstein, 2014).

# RST: Condition

Realization of the nucleus clause depends on the satellite clause.

- (5) **If** this were to pass, 60 percent of those employers would either freeze or terminate their pension plans. (Richard Neal, D-MA)

# RST: Contrast

A multinuclear relation in which the two nuclei are similar in some respects, but differ in a few key ways.

- (6) It is quite obvious that not enacting these reforms does not come from a lack of ability, **but** from a lack of will. (Thomas Davis, R-VA)

# RST: Purpose

The satellite clause is to be realized through the action described in the nucleus.

- (7) ... a rules package [that actually lowers our ethics standards,] [**so that** they may promote their own agenda].... (Joe Baca, D-CA)

# RST: Reason

The action described in the nucleus clause was carried out by a volitional agent; the satellite provides the explanation for why the action was taken.

- (8) I voted for this energy bill **because** I thought we needed a national plan.  
(Dennis Cardoza, D-CA)

# RST relations in Convote

RST relations are detected using state-of-the-art DPLP parser (Ji & Eisenstein, 2014).

	total	per document
CONDITION	2532	0.912
CONTRAST	3750	1.35
PURPOSE	12,482	4.49
REASON	1426	0.513
All discourse units	349,528	125.9

The hierarchical structure is ignored.

# The Penn Discourse Treebank

The PDTB focuses on local discourse relations (Prasad et al., 2008).

- ▶ All relations that are explicitly marked by a connective.
- ▶ Implicit relations between all adjacent sentences.
- ▶ Relation inventory is hierarchical, and “lexically-anchored” in specific connectives.
- ▶ 94% agreement on four-way relation senses; 80% agreement on fine-grain senses.



# PDTB relations

“Unambiguous” explicit markers of top-level PDTB relations:

**COMPARE** although, but, however, though, yet,  
instead

**CONTINGENCY** because, if, thus, as a result

**EXPANSION** also, and, for example, moreover, for  
instance, in fact, indeed, or, unless

**TEMPORAL** after, before, until, when, later, once,  
previously

# PDTB relations in Convote

PDTB relations are detected by simple string matching.

	total	per document
COMPARE	6097	2.20
CONTINGENCY	5974	2.15
EXPANSION	44,500	16.02
TEMPORAL	4898	1.76
All words	1,507,348	542.8

# Regression model

$$y_{r,i} \sim \text{Binomial}(\mu_{r,i}, N_i) \quad (1)$$

$$\mu_{r,i} = \sigma(\theta_r^\top \mathbf{x}_i + b_r). \quad (2)$$

Predictors in  $\mathbf{x}_i$ :

- ▶ republican
- ▶ yes vote
- ▶ cross-party vote
- ▶ idp1 (Carroll et al., 2009)
- ▶ abs(idp1)

# RST results

	repub.	yes vote	x-party	idp1	abs idp1
CONDITION	-0.168	-0.015	0.166	0.276	-0.386
CONTRAST	-0.069	0.072	0.118	0.029	0.166
PURPOSE	0.104	-0.034	-0.059	-0.100	0.128
REASON	0.187	0.018	0.189	-0.133	-0.029

# RST results

	repub.	yes vote	x-party	idp1	abs idp1
CONDITION	-0.168	-0.015	0.166	0.276	-0.386
CONTRAST	-0.069	0.072	0.118	0.029	0.166
PURPOSE	0.104	-0.034	-0.059	-0.100	0.128
REASON	0.187	0.018	0.189	-0.133	-0.029

No statistically significant coefficients after False  
Discovery Rate correction 😞

# PDTB results

	repub.	yes vote	x-party	idp1	abs idp1
COMPARE	0.017	<b>-0.210**</b>	0.097	0.051	-0.030
CONTINGENCY	-0.025	-0.076	0.042	0.080	<b>0.360*</b>
EXPANSION	0.028	0.031	0.004	-0.000	-0.060
TEMPORAL	<b>-0.279*</b>	-0.070	-0.005	<b>0.264*</b>	0.017

# PDTB: Compare

The COMPARE relation is used more often in speeches explaining “no” votes ( $p < .01$ ).

- (9) The federal budget should be an honest blueprint for the spending priorities of the government. **However**, this budget is not honest. (David Scott, D-GA)

# PDTB: Compare

The COMPARE relation is used more often in speeches explaining “no” votes ( $p < .01$ ).

- (10) **Instead** of requiring the EPA to change testing procedures, the Rogers amendment requires the EPA to change the adjustment factors that the EPA currently uses to make the fuel economy label accurate. (John “Joe” Schwartz, R-MI)



# PDTB: Compare

The COMPARE relation is used more often in speeches explaining “no” votes ( $p < .01$ ).

- (11) **Instead** of requiring the EPA to change testing procedures, the Rogers amendment requires the EPA to change the adjustment factors that the EPA currently uses to make the fuel economy label accurate. (John “Joe” Schwartz, R-MI)

(No improvement on vote prediction, sadly)

# PDTB: Contingency

The CONTINGENCY relation is associated with abs idp1, indicating partisanship ( $p < .05$ ).

- (12) So **if** you are a believer in big government and world government and you believe in giving up the prerogatives of the congress and not assuming our responsibility, I would say, go with the WTO. But **if** you believe in freedom, **if** you believe in the constitution and **if** you really believe in free trade, i would say we should vote to get out of the WTO. (Ron Paul, R-TX, *idp1* = .974)

# PDTB: Contingency

The CONTINGENCY relation is associated with abs  
idp1, indicating partisanship ( $p < .05$ ).

- (13) Our people can not be desperate **if**, in  
fact, we want a strong America. (Barbara  
Lee, D-CA, *idp1* =  $-.710$ )

# PDTB: Contingency

The CONTINGENCY relation is associated with abs idp1, indicating partisanship ( $p < .05$ ).

- (14) But **if** this legislation was only about principle, why would the majority party be distributing talking points in the other body declaring that “this is a great political issue” and that by passing this bill “the pro-life base will be excited”? (John Conyers, D-MI, *idp1* =  $-.687$ )

# PDTB: Temporal

The TEMPORAL relation is negatively associated with Republican legislators ( $p < .05$ ).

- (15) Mr. Speaker, in 1995 **when** the WTO was established, our trade deficit was \$96 billion. (Bernie Sanders, I-VT,  $idp1 = -.523$ )

# PDTB: Temporal

The TEMPORAL relation is negatively associated with Republican legislators ( $p < .05$ ).

- (16) These prices were well on their way to hitting record highs long **before** Hurricane Katrina. (Dennis Kucinich, D-OH,  $idp1 = -.604$ )

# PDTB: Temporal

The TEMPORAL relation is negatively associated with Republican legislators ( $p < .05$ ).

- (17) The bill **before** us today proposes to gut the clean air act in order to promote construction of more refineries. (Edward Markey, D-MA,  $idp1 = -.512$ )

# PDTB: Temporal

The TEMPORAL relation is negatively associated with Republican legislators ( $p < .05$ ).

- (18) The bill **before** us today proposes to gut the clean air act in order to promote construction of more refineries. (Edward Markey, D-MA,  $idp1 = -.512$ )

...But **within** each party, temporal relations are used by more conservative legislators.



# Summary

- ▶ Discourse structural analysis can help to distinguish content from rhetorical style.
- ▶ PDTB relations are easier to detect than RST, better correlated with political covariates.

# Summary

- ▶ Discourse structural analysis can help to distinguish content from rhetorical style.
- ▶ PDTB relations are easier to detect than RST, better correlated with political covariates.
- ▶ Next steps:
  - ▶ 2007-2010 data would help with collinearity between votes, party, and ideology.
  - ▶ Different discourses for different audiences?
  - ▶ Discourse metrics for linguistic complexity (Spirling, 2015).
  - ▶ Better modeling: legislator random effects, structural motifs, discourse + topics, ...

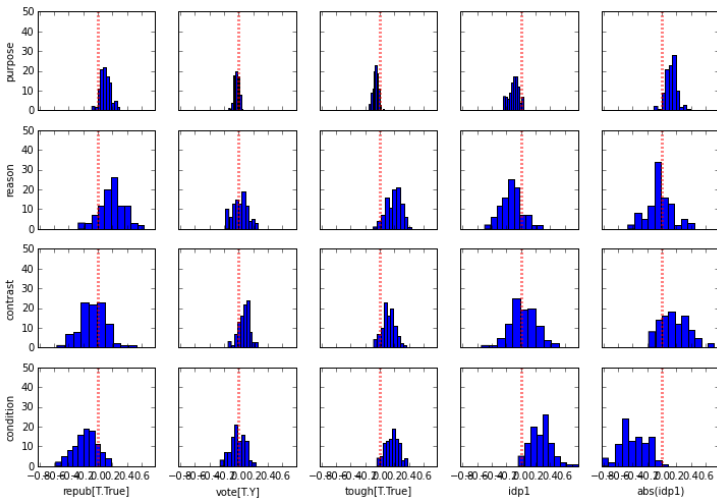
# References I

- Asr, F. T. & Demberg, V. (2013). On the information conveyed by discourse markers. In *Proceedings of the Fourth Annual Workshop on Cognitive Modeling and Computational Linguistics*, (pp. 84–93).
- Carroll, R., Lewis, J. B., Lo, J., Poole, K. T., & Rosenthal, H. (2009). Measuring bias and uncertainty in dw-nominate ideal point estimates via the parametric bootstrap. *Political Analysis*, 17(3), 261–275.
- Ji, Y. & Eisenstein, J. (2014). Representation learning for text-level discourse parsing. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse Treebank 2.0. In *Proceedings of LREC*.

# References II

- Spirling, A. (2015). Democratization and linguistic complexity: The effect of franchise extension on parliamentary discourse, 1832–1915. *Journal of Politics*, forthcoming.
- Thomas, M., Pang, B., & Lee, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, (pp. 327–335).

# RST bootstrap coefficients



# PDTB bootstrap coefficients

