EMNLP / NASTEA

A recap of some EMNLP talks, and my own research presented at the Second Computing News Storylines Workshop.

Dan Simonson, PhD Candidate, Georgetown University
des62@georgetown.edu / @thedansimonson

Tony Davis
tonydavis0@gmail.com

DC NLP Meetup, November 16th, 2016

Introduction

- Going to recap some EMNLP talks.
  - Selected by top tweets!
- Going to give an update on my own dissertation research on narrative schemas.

EMNLP Highlights

Introduction

- EMNLP: Empirical Methods in Natural Language Processing

Selections

- Live tweets of conference
- Top tweets from Twitter Analytics
  - Based on views.

Caveats

- Stanford NLP Group retweeted a lot of their own.
  - So their numbers are inflated by that.
  - Also, linguistically-interesting results
- I’m going to highlight the bits I found interesting.
  - ACL Anthology: Papers online for free.

7) Sluicing

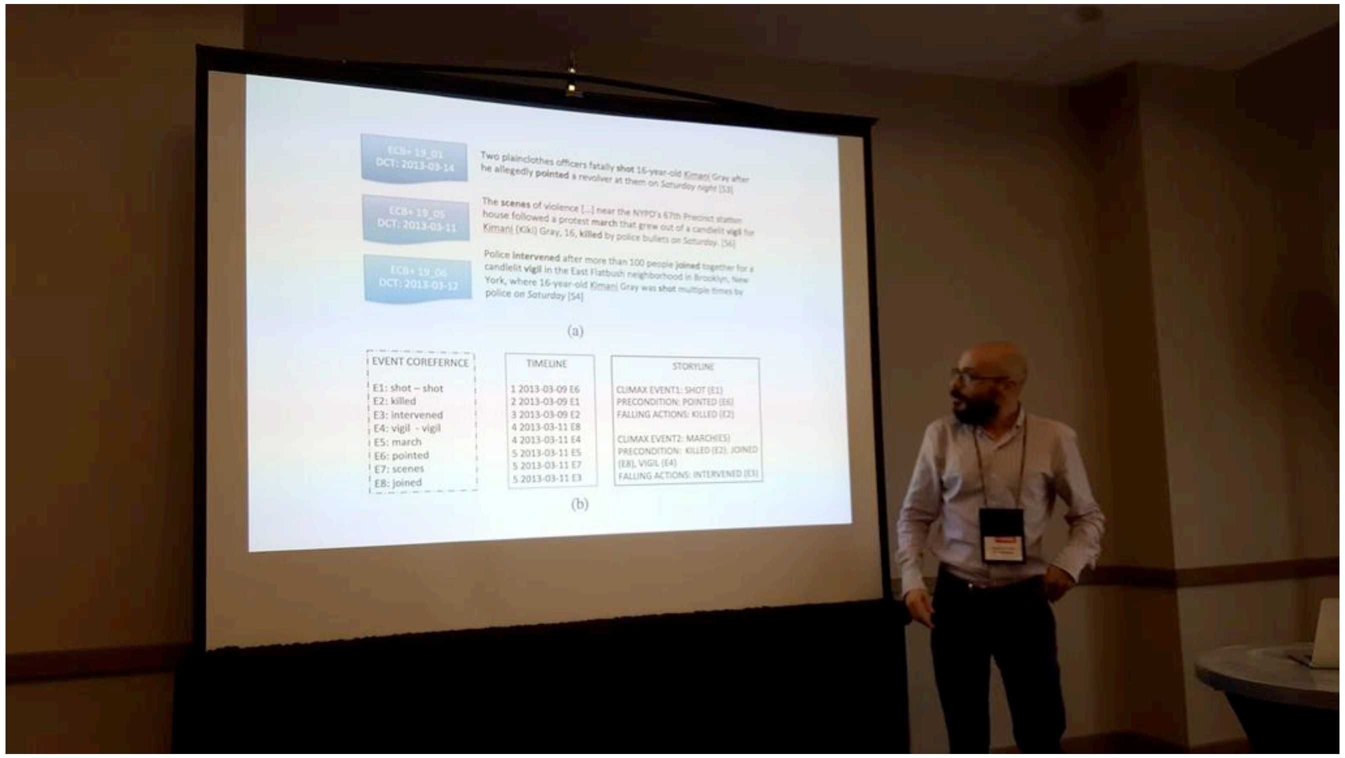
- Anand and Hardt
- Example:
    - “Harry traveled to southern Denmark to study botany. I want to know **why**.”

6) Word Problems

- Upadhyay et al.: doing algebra problems with supervision
  - e.g. Give a system an algebra word problem, and it solves it.
- In essence: system learns both
  - algebra templates
  - alignments b/w templates and problems

5) StaR

- @tommaso_caselli et al. (CNewS): Storyline representation scheme
- An exhaustive representation of the events contained in news articles: rising action, climax, falling action, time annotations, etc.



4) TweeTime

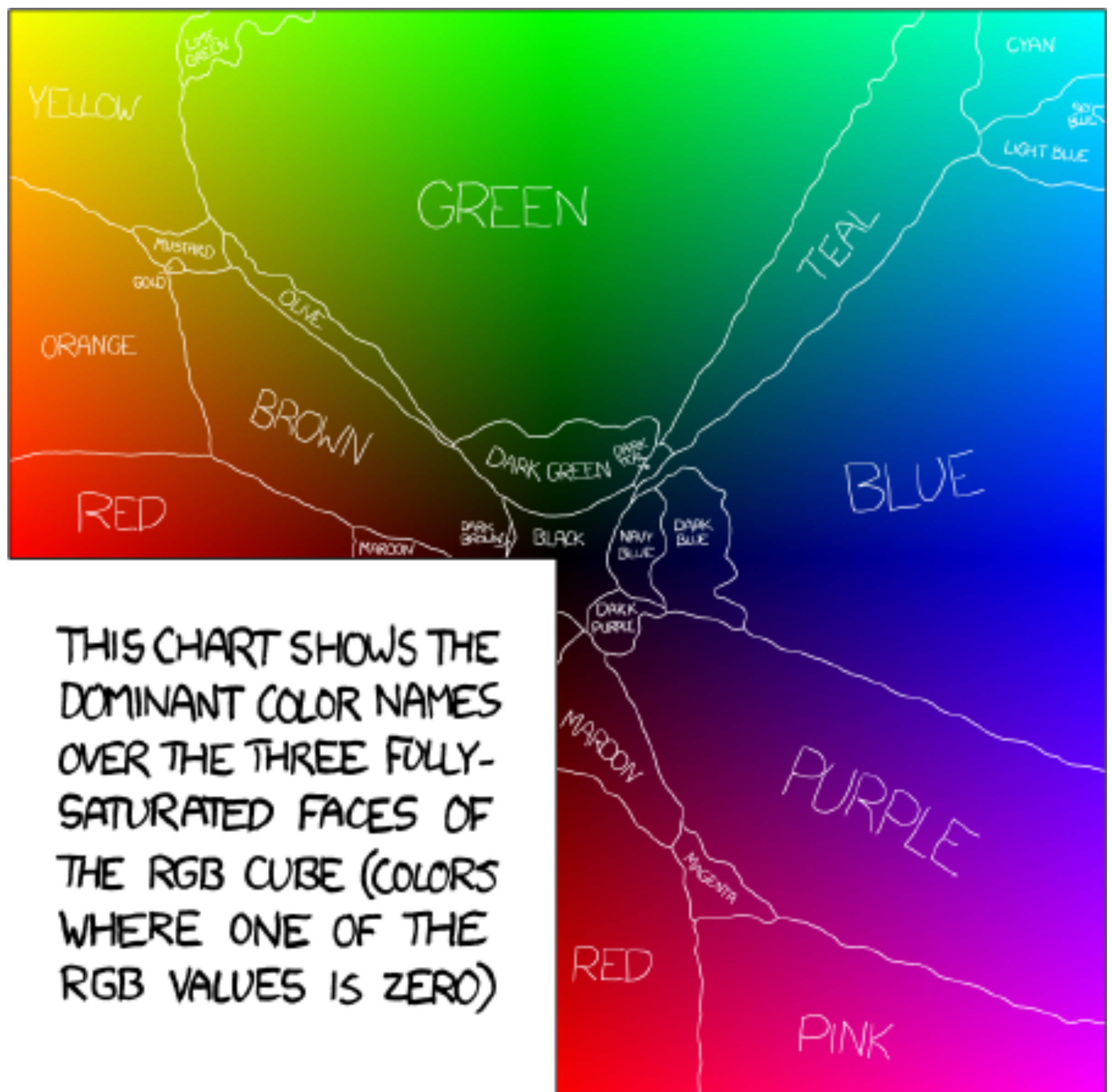
- Tabassum, Ritter, and Xu: recognizing and normalizing time expressions on Twitter
- “Distant Supervision Assumption” == Awesome

3) Color Names

- @futurulus et al.: Learning to generate color composition names
- Used Fourier analysis to get bimodally distributed color names, like "greenish"
- Generates new color names
  - "steel purple"

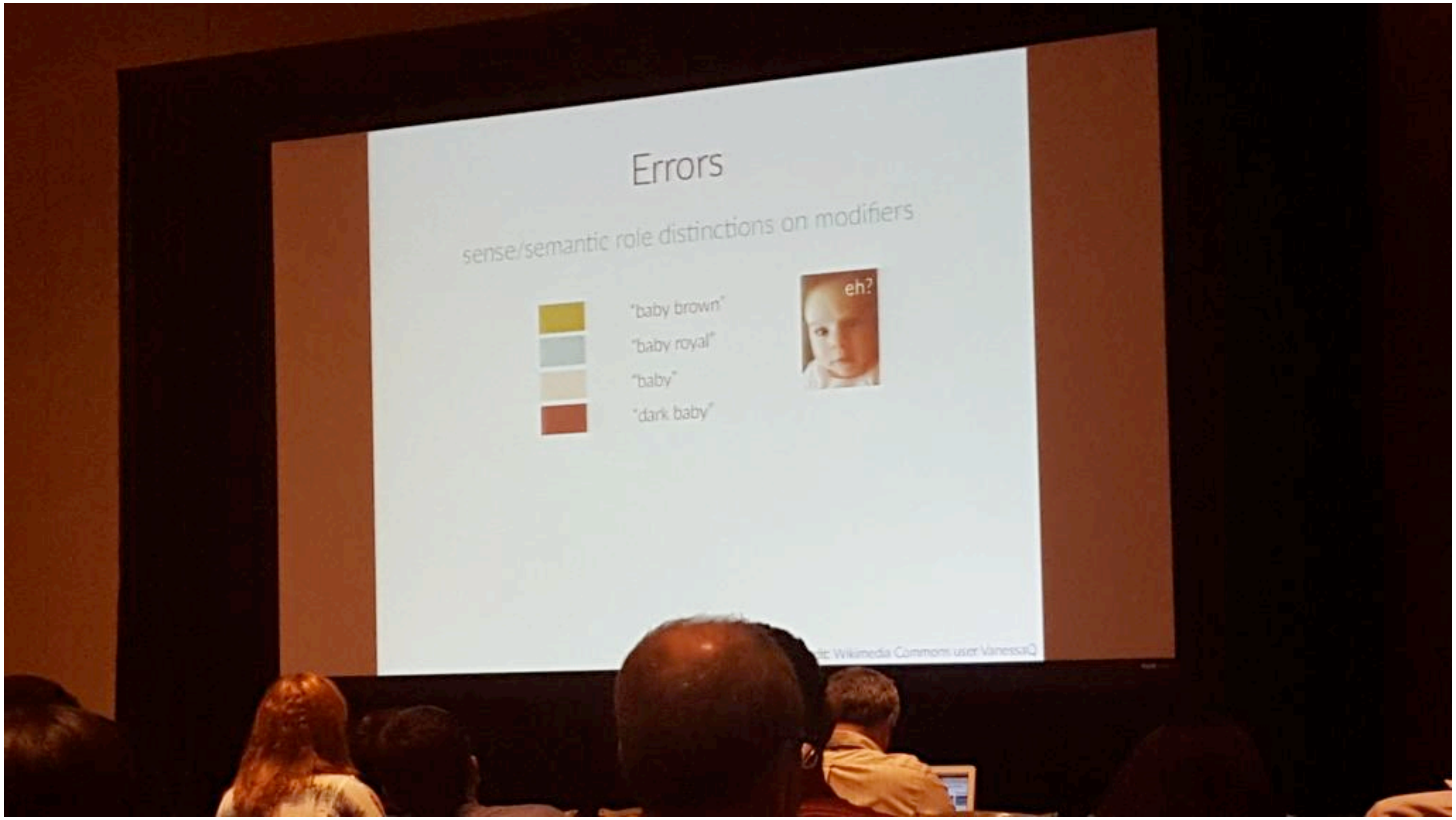
3) Color Names

- Source data: XKCD Color Survey



3) Color Names

- Example of Errors:
  - “Baby” has two senses.

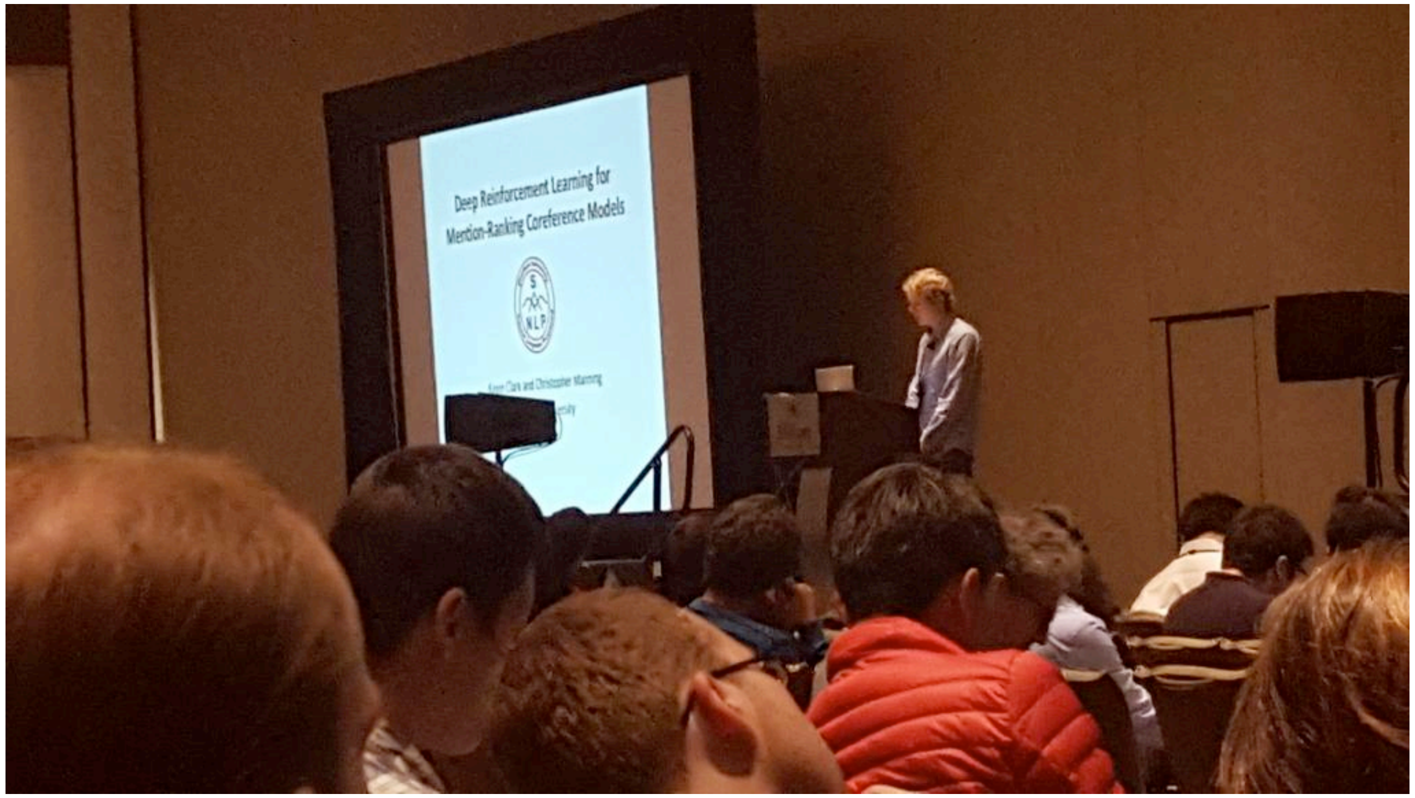


2) Coref

- Clark and @chrmanning: coreference with deep reinforcement learning.
- Makes a greater number of errors, but the errors it makes are less costly.
  - Does this by considering the global structure of a document.

2) Coref

- Code:
  - https://github.com/clarkkev/deep-coref



1) Mental NLP

- @timalthoff et al.: On counseling conversations, using #nlproc for mental health

1) Mental NLP

- Data: from a text-messaging based counseling service.
  - Texters (patients) respond to a survey after the fact.

1) Mental NLP

- Findings:
  - Good counselors spend more time solving the problem than discussing it.
  - Texters report feeling better when they talk less about self, more about future.
  - Creative, adaptable counselors performed better.

Poster Mentions

- Bouchard et al.: Generating Textual Data
  - "small data, the next big thing?"
- Augenstein et al.: Stance Detection
  - Better than just "sentiment."
- @williamlief et al.: Getting domain-specific sentiment lexicons from unlabelled data.

NASTEA

Overview

- Prior Work:
  - What are schemas?
  - Why is NASTEA needed?
- NASTEA Task
- Experiment and Data
- Results

Narrative Schemas

- Abstractions of sequences of events obtained from coreference and parses.
- Devised by Chambers and Jurafsky (2008, 2009)

Narrative Schema Examples

- We follow Chambers and Jurafsky (2009) in generating schemas, for the most part.

Making Schemas

Nonetheless, she continued working off and on... she took a job rubber-banding newspapers... She does not know exactly what will happen to her grant when she marries...

...she marries. Then, she takes time off to raise her kids. Several years hence, she seeks to re-enter the labor force... Nonetheless, she finds a job, works for 15 years or so...

Her plans to go to college to become a teacher had crumbled; in fact, she was unsure she would graduate from high school... her doctors had told her that it would be risky, to herself and the baby, to give birth while she was on dialysis... As for the future, Ms. Lorrington and Mr. Wilson said they planned to marry... And Ms. Lorrington said that while she did not know what work she would seek or be physically capable of in the future...

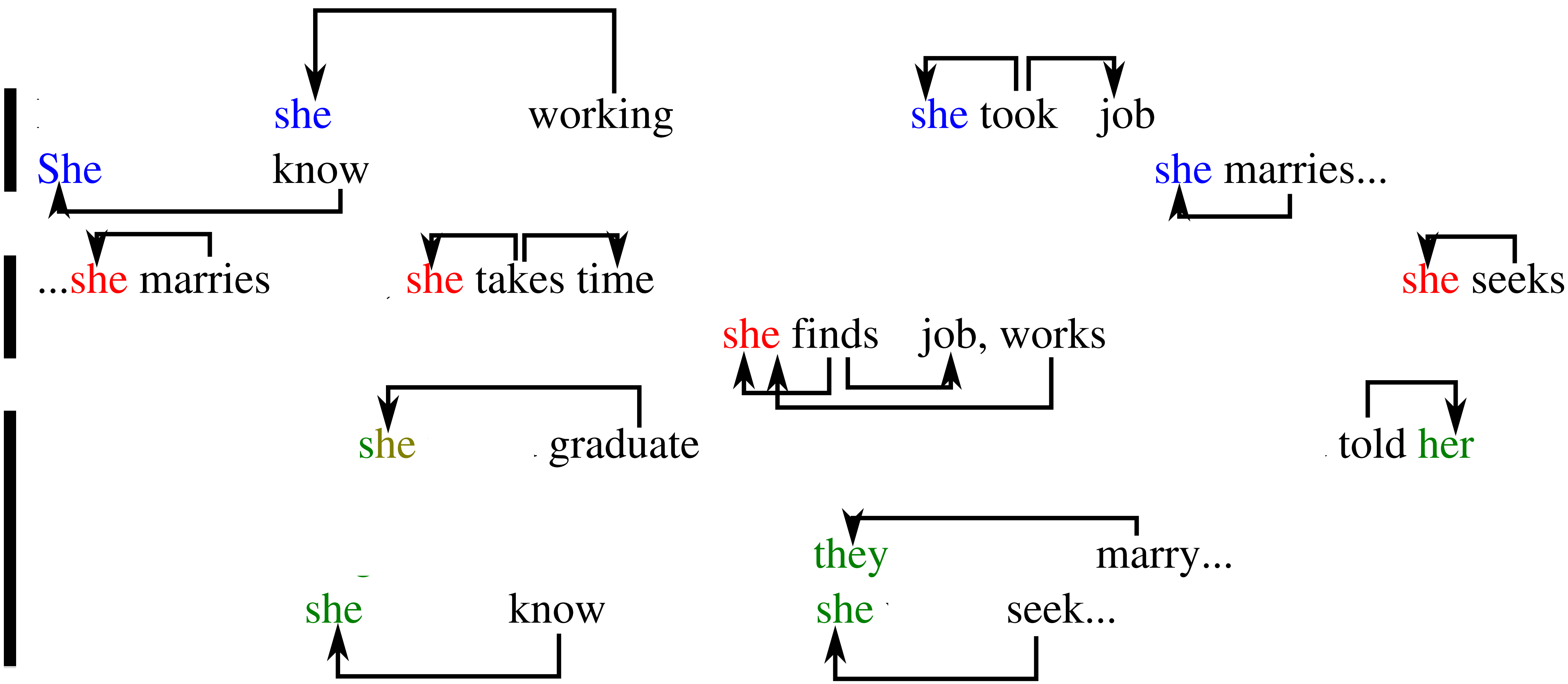
Making Schemas

Nonetheless, she continued working off and on... she took a job rubber-banding newspapers... She does not know exactly what will happen to her grant when she marries...

...she marries. Then, she takes time off to raise her kids. Several years hence, she seeks re-enter the labor force... Nonetheless, she finds a job, works for 15 years or so...

...she was unsure she would graduate from high school... her doctors had told her that it would be risky, to herself and the baby, to give birth while she was on dialysis... As for the future, Ms. Lorrington and Mr. Wilson said they planned to marry... And Ms. Lorrington said that while she did not know what work she would seek...

Making Schemas



Making Schemas

- Candidate co-referring argument pairs are scored fundamentally based on their PMI (Chambers and Jurafsky 2009).
- Schemas are generated based on this score.
  - The counter-training procedure used in Simonson and Davis (2015) was too slow for the approach to topics used here.

New Evaluation?

- Previous work does not evaluate schemas directly, we want to.
- Previous work hinted at the potential centrality of entity types in interpreting schemas (Simonson and Davis 2015).
- The NYT Corpus, our data set, has salient entity annotations: person, organization, location.

New Evaluation?

- Hypothesis: better schemas should agree with the NYT library scientists about who and what are important in an article.
- Even if we're wrong, perhaps we ought to learn something in the process.
  - Little is *known* about schemas.

NASTEA

- “Narrative Argument Salience Through Entities Annotated”
- 1) measure the “presence” of a schema in a document.
- 2) use present schemas to extract entities from a document.

Canonical Presence

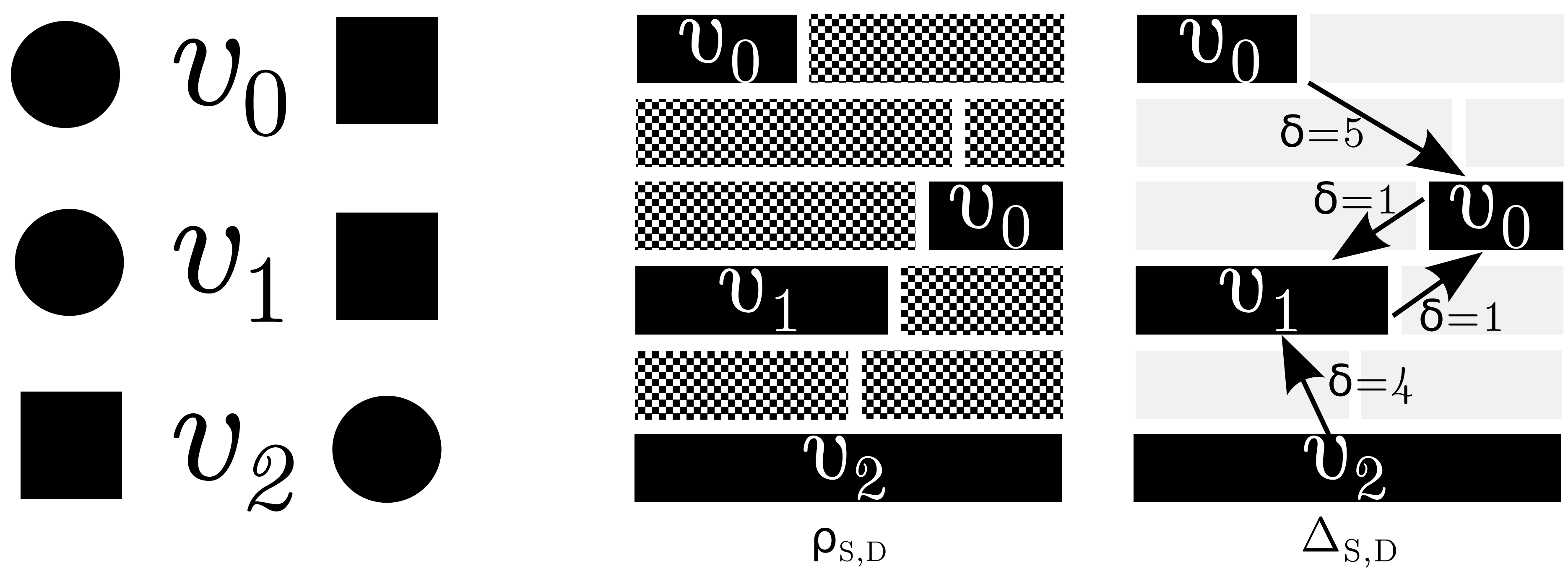
- We call the presence used in this paper "canonical presence."
  - It assumes documents are instantiations of canonical forms of a specific schema.
- We avoid *local* coreference information because it is error prone.

Canonical Presence

- We look at how the events contained in a schema are distributed inside a document.
  - Density
  - Dispersion

Canonical Presence

- Density is $\rho_{S,D}$; dispersion is $\Delta_{S,D}$.



- $p_{S,D} = \rho_{S,D} / \Delta_{S,D}$

Entity Extraction

- Use the parses from the highlighted events to grab `SUBJ`, `OBJ`, `PREP` (as relevant).
- Compare entities extracted to NYT annotations.
  - NYT annotations tokenized, normalized for case.
  - Low threshold for similarity.

Entity Extraction

- F1 scores result:
  - Precision: fraction of extracted entities contained in NYT annotations
  - Recall: fraction of NYT annotations contained in extracted entities
- Experiment with using more than one schema per document.

Data

- New York Times Corpus (Sandhaus 2008)
- Document categories chosen for being near each other in number of documents, and for variety.
  - Between 36,360 and 52,110.
  - 10% Hold-out for Evaluation
- Salient entity annotations by New York Times library scientists.

Experiment

- Q: Do topics give us better schemas?
- schemas$\nrightarrow$topic (Simonson and Davis 2015)
- But what of the converse?
  - topic$\rightarrow$schemas?
  - Do we get better schemas by conditioning them on topic?

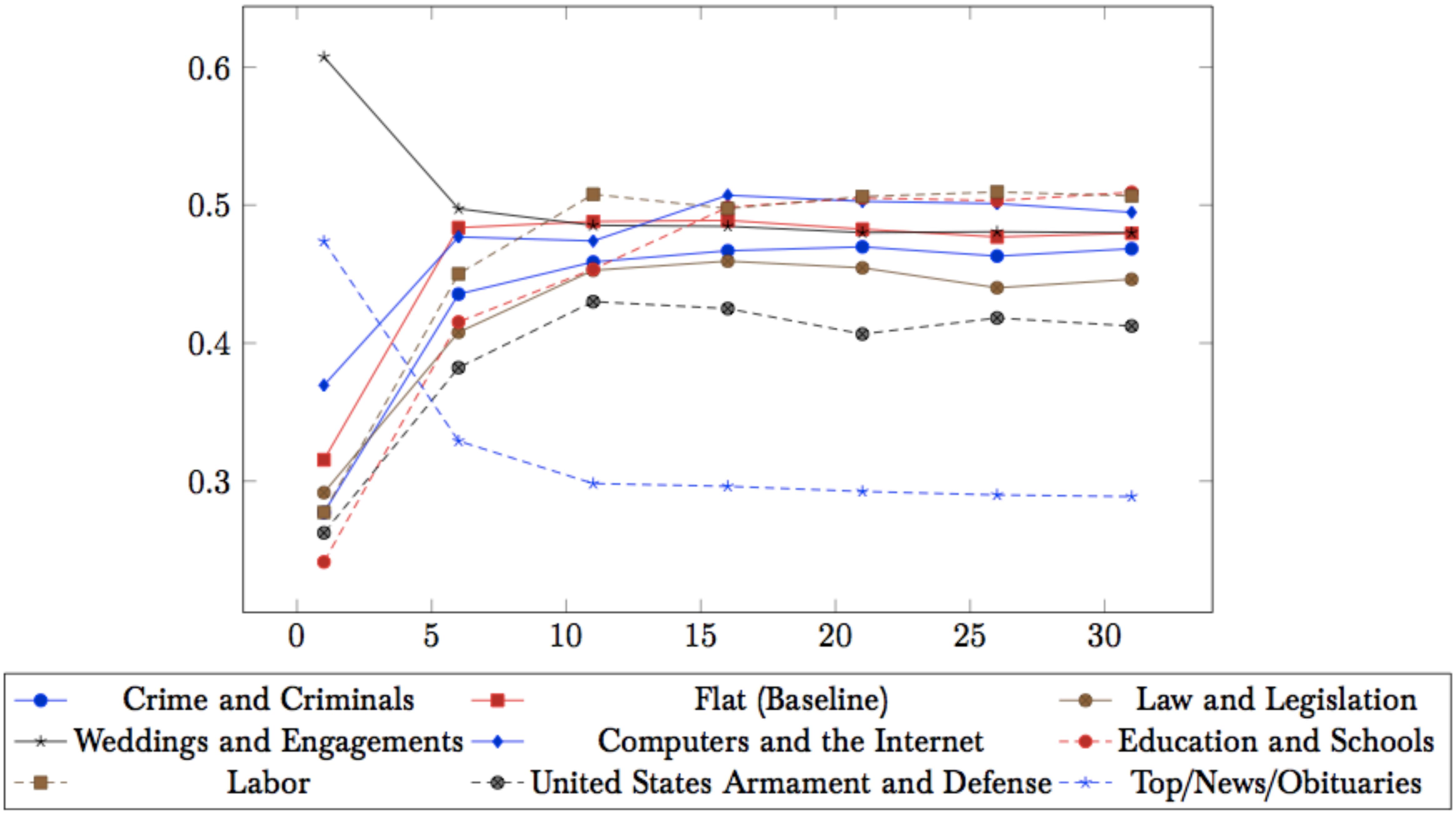
Experiment

- Generate PMI-based model for each topic, then:
  - Run narrative cloze task (Chambers and Jurafsky 2009).
  - Generate schemas for each topic, run NASTEA.
- Baseline: one large model.

Experiment

- In many cases, the most present schema fails to capture the correct entities.
  - We apply more schemas then, in increments of 5.
- We refer to the extraction using the most present schema as N_1.
  - Top 6 as N_6, Top 11 as N_{11}, etc.

Results



0.6
0.5
0.4
0.3
0
5
10
15
20
25
30
Crime and Criminals
Flat (Baseline)
Law and Legislation
Weddings and Engagements
Computers and the Internet
Education and Schools
Labor
United States Armament and Defense
Top/News/Obituaries

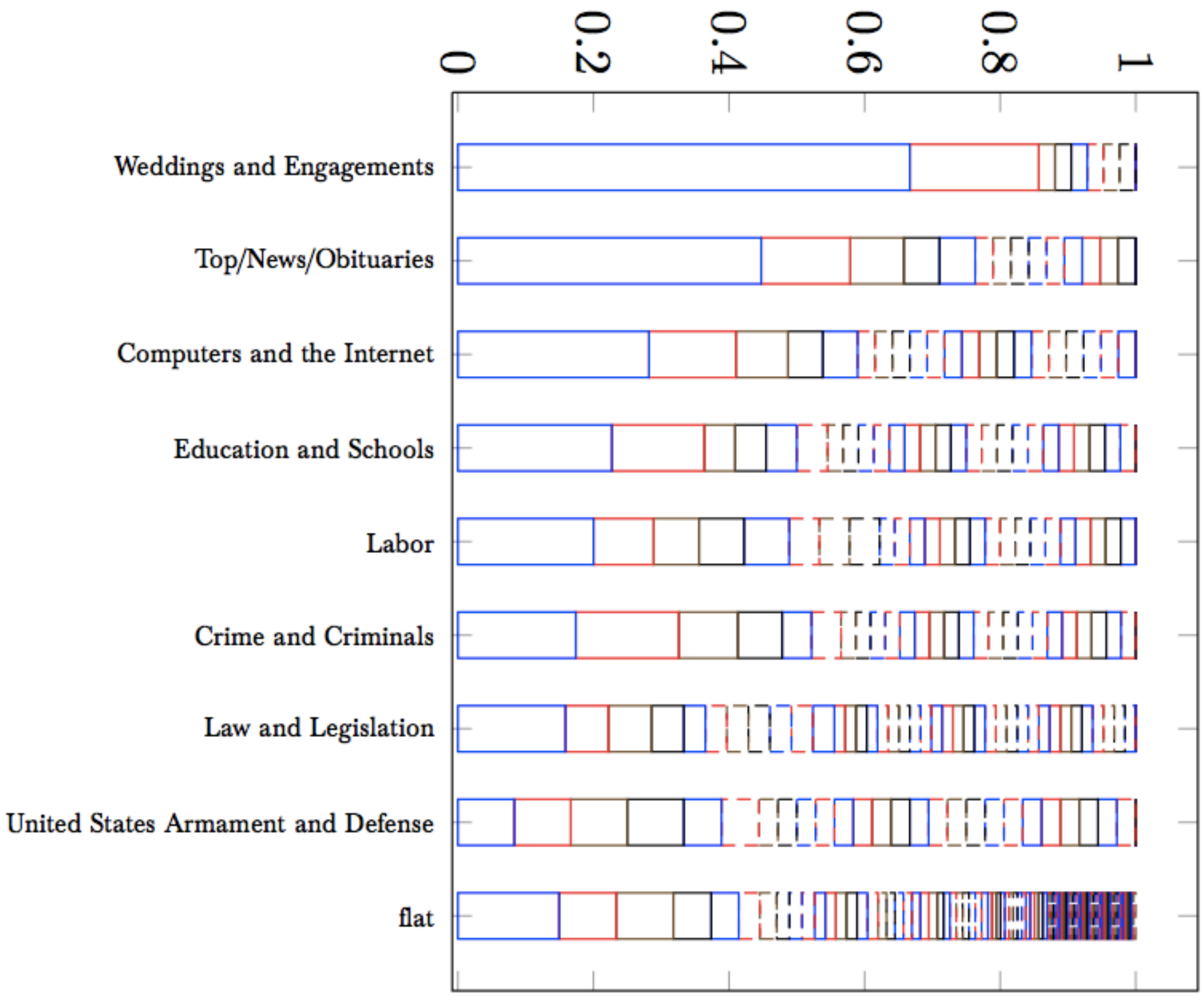
Results

Test Model	Avg. Cloze Rank	N_1
Baseline	1329	0.315
Topical (avg)	1273	0.365
Obituaries	565	0.474
Weddings and Engagements	1058	0.607
Crime and Criminals	1268	0.277
Law and Legislation	1279	0.292
Labor	1297	0.277
Computers and the Internet	1346	0.369
U.S. Armament and Defense	1805	0.262

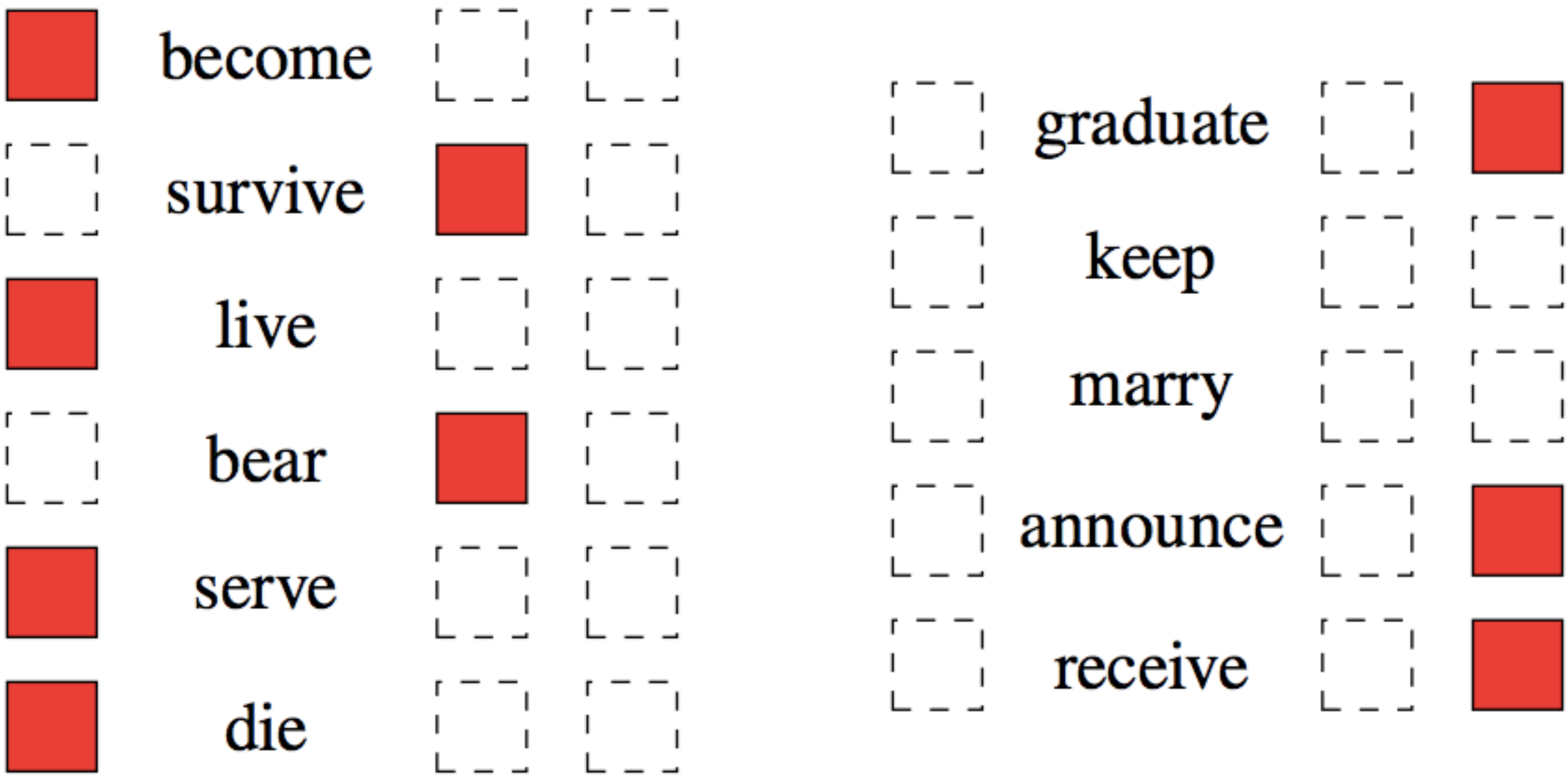
NASTEA Curves

- Some categories do better with more schemas; some do worse.
- Clear separation! But why?
- Do the N_1 high performers happen to have **a better set** of schemas, or is **a small set** of schemas really good at covering content in those topics?
- NASTEA allows us to inspect the schemas directly.

Homogeneity



Homogeneity



Homogeneity

- The ones that do better on N_1 are more homogeneous.
  - Weddings and obituaries are written from templates!
- For understanding heterogeneous documents better, we might need a better model of schemas.

Interpretations

- Within the context of our model:
  - Weddings and obituaries are more homogenous topics; news topics, more heterogeneous.
- With better N_1 as a goal:
  - A better schema model could possibly capture heterogeneous topics better.

Conclusions

- NASTEA can evaluate the quality of narrative schemas directly.
  - Trends with cloze at the large scale, local variations (to be explored).
- Some document categories are narratologically homogeneous.
  - Heterogeneity is typical of many document categories.

Thank You!

- Acknowledgements: Amir Zeldes, Nate Chambers, Georgetown University Department of Linguistics, and reviewers for helpful feedback.

References

- Bal, M. (1997) Introduction to the Theory of Narrative. University of Toronto Press, Toronto.
- Balasubramanian, N., Solderland, S., Mausam, and Etzioni, O. (2013). Generating coherent event schemas at scale. In EMNLP, pages 1721–1731. Association for Computational Linguistics.
- Chambers, N. (2013). Event schema induction with a probabilistic entity-driven model. In EMNLP, pages 1797–1807. Association for Computational Linguistics.
- Chambers, N. and Jurafsky, D. (2008). Unsupervised learning of narrative event chains. In ACL, pages 789–797. Association for Computational Linguistics.
- Chambers, N. and Jurafsky, D. (2009). Unsupervised learning of narrative schemas and their participants. In IJCNLP-AFNLP, pages 602–610. Association for Computational Linguistics.
- Chambers, N. W. (2011). Inducing Event Schemas and their Participants from Unlabeled Text. Stanford University.
- Jans, B., Bethard, S., Vulić, I., and Moens, M. (2012). Skip n-grams and ranking functions for predicting script events. In EACL, pages 336–344. Association for Computational Linguistics.
- Mostafazadeh, N., Chambers, N., He, X., Parikh, D., Batra, D., Vanderwende, L., Kholi, P. & Allen, J. (2016). A corpus and cloze evaluation for deeper understanding of commonsense stories. In NAACL HLT. Association for Computational Linguistics.
- Pichotta, K. and Mooney, R. J. (2014). Statistical script learning with multi-argument events. In EACL, volume 14, pages 220–229.
- Pichotta, K. and Mooney, R. J. (2015). Learning statistical scripts with LSTM recurrent neural networks. In Proceedings of the 30th AAAI Conference on Artificial Intelligence.
- Rudinger, R. (2015). Script Induction as Language Modeling. In EMNLP, pages 1681–1686. Association for Computational Linguistics.
- Sandhaus, E. 2008. The New York Times Annotated Cor- pus. Linguistic Data Consortium, Philadelphia.
- Simonson, D. and Davis, A. (2015). Interactions between Narrative Schemas and Document Categories. In the First CnewS Workshop, ACL 2015, Beijing, China.