Dialogue Structure in Microtext

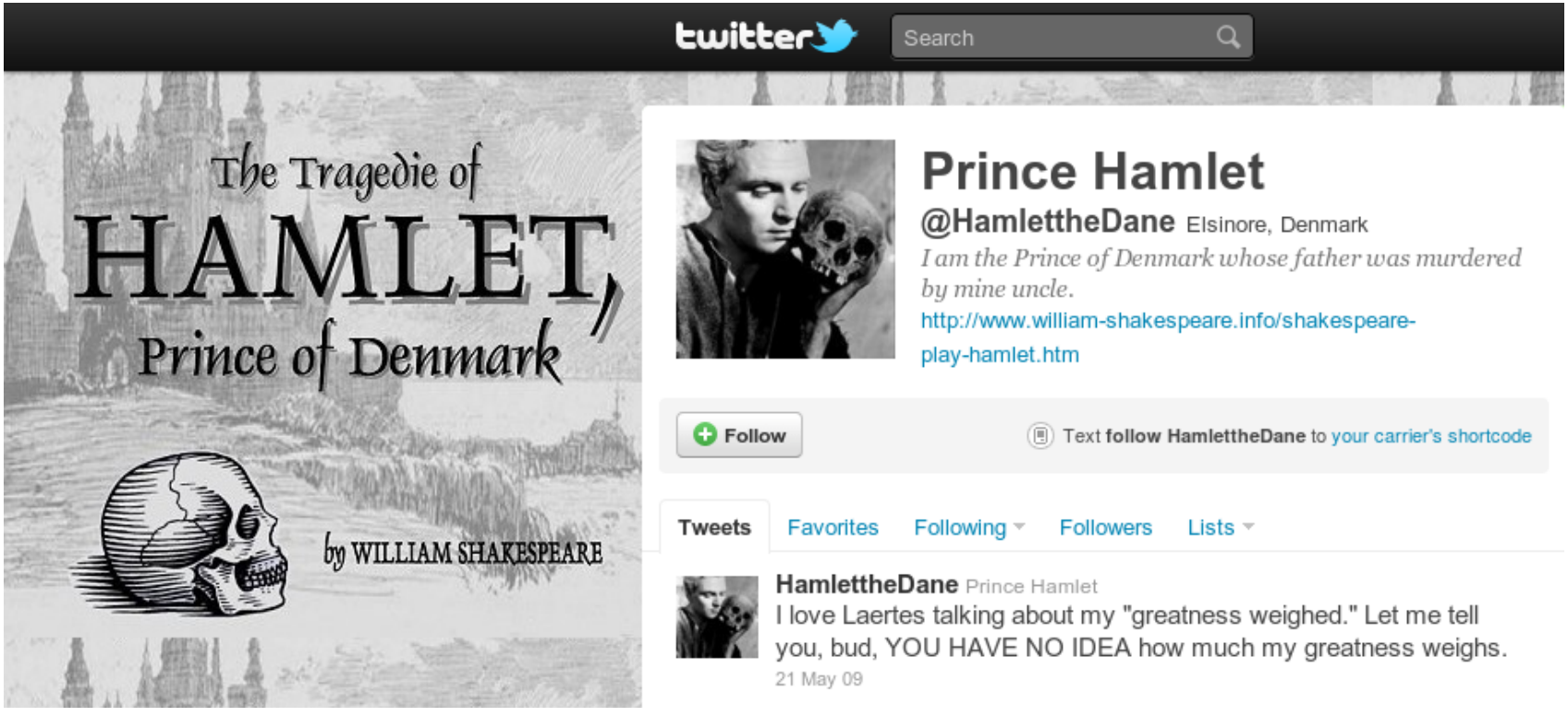
AAAI-11 Workshop on Analyzing Microtext

Micha Elsner

School of Informatics
University of Edinburgh

8 August, 2011

Unconventional interactions



Where do we fit in?

Speech

- Prosody, tone of voice
- No latency
- Short utterances
- Strictly sequential

Text

- Just words
- Ultra-high latency
- Long documents
- Hierarchical / searchable

Interfaces matter!

- *Mostly* text, some multimedia
- Low latency (IM) vs high latency (web forum)
- Short turns (twitter) vs long turns (email)
- Sequential (Huffington comments) vs structured (Slashdot comments)

These differences affect the language we see!

Overview

Introduction

Case study: IRC chat

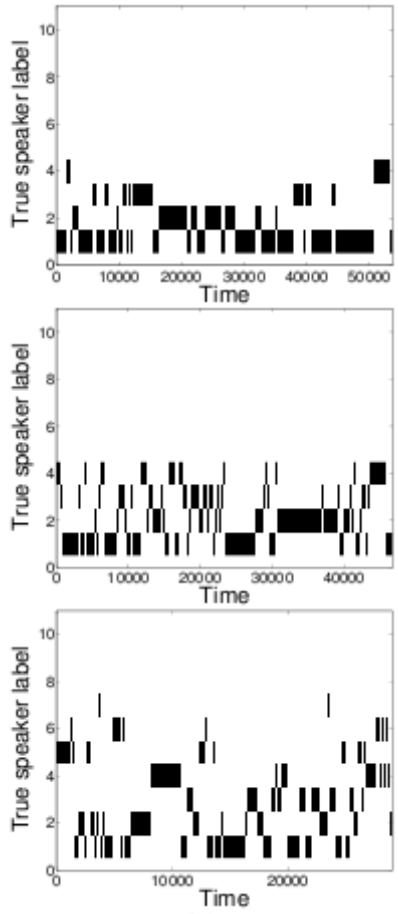
IRC vs speech

Text and speech models for disentangling IRC

Conclusions

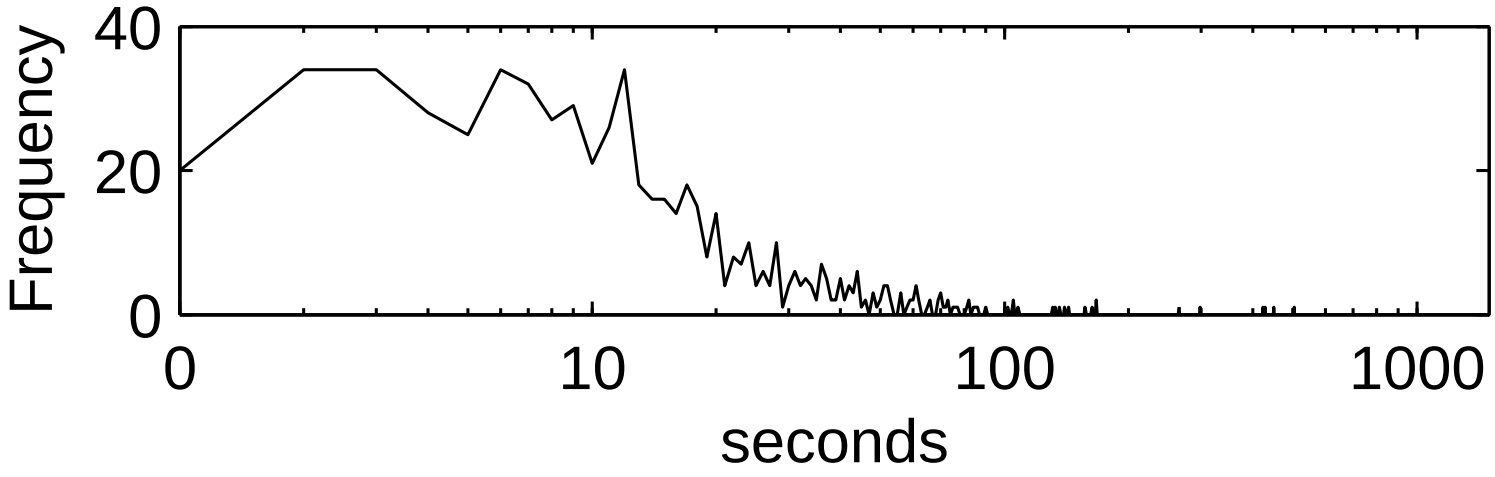
Speech: conversational structure

- One speaker at a time
  - Has *the floor* (Sacks et al)
- Speaker signals intent to keep talking or finish
- Coordination via short utterances:
  - Filled pauses “uh”, backchannels “yeah”



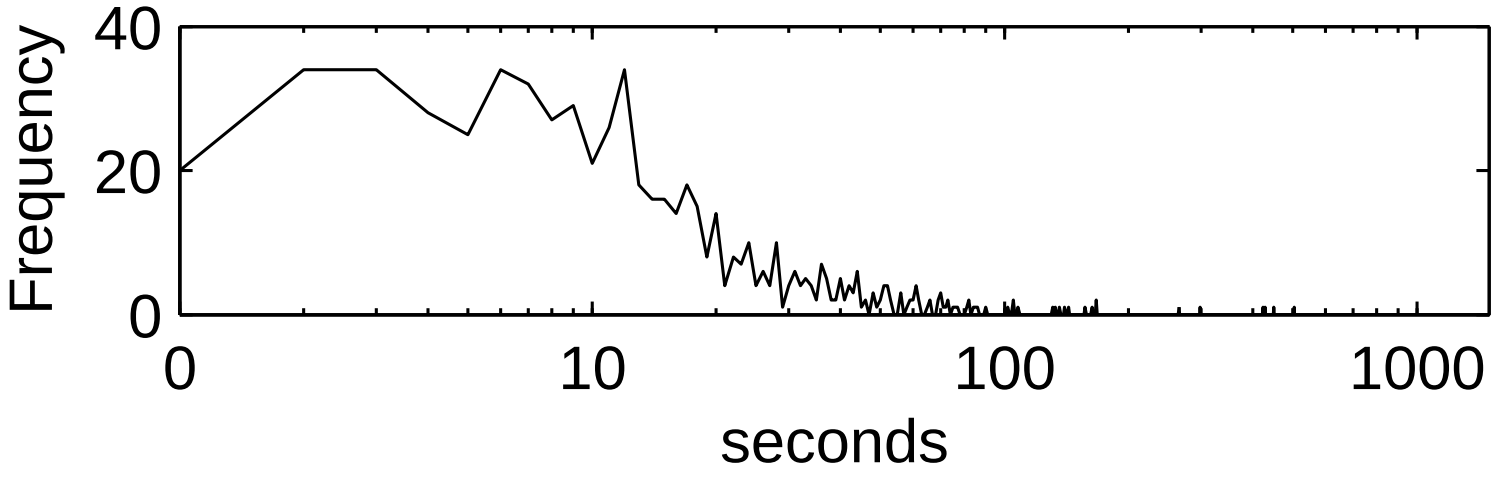
(Fox, Sudderth et al: A Sticky HDP-HMM)

Turn-taking in IRC chat



- More tolerant of long pauses
- And possibly of “interruption”

Turn-taking in IRC chat



- More tolerant of long pauses
- And possibly of “interruption”

- Some backchannel-like utterances:
  - $\sim$ 10% one-word comments: “lol”, “ok”
  - Switchboard: $\sim$ 17% backchannels: “yeah”, “uh-huh”

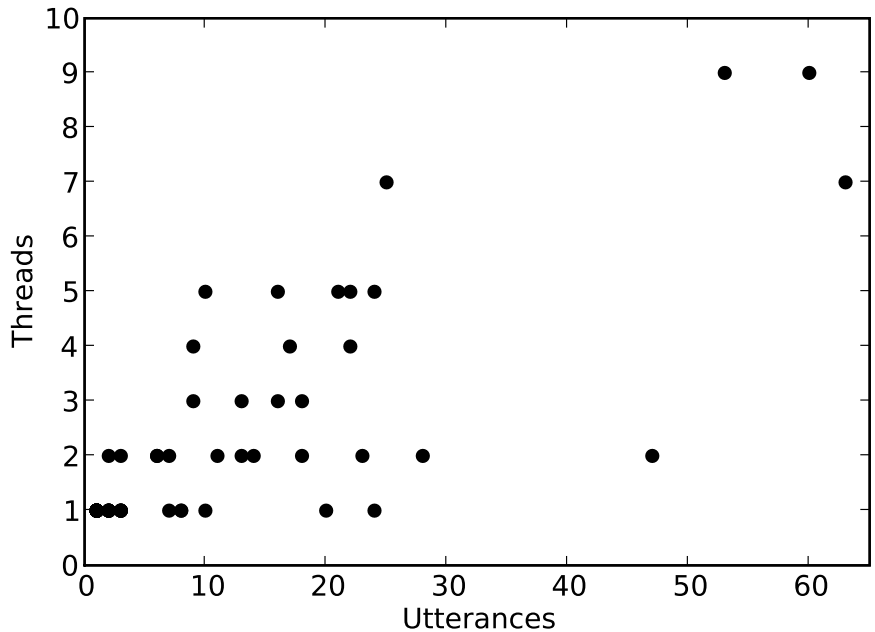
Multiple floors

Usually several conversations at a time

- Between 2 and 3 active during each utterance

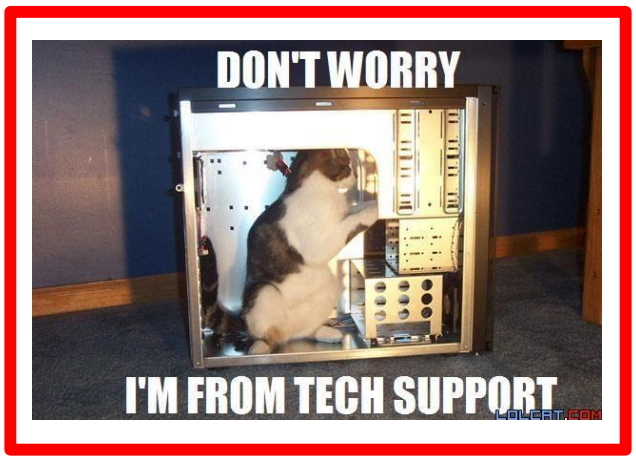
Chatters participate in many conversations

- The more one speaks, the more threads they speak in



Questions

What information is useful?
How well do text/speech models adapt?



Disentanglement (threading)

Google solved my problem.

You guys have never worked in a factory before, have you?

There's some real unethical stuff that goes on

Of course, that's how they make money!

You deserve a trophy!

People lose limbs, or get killed.

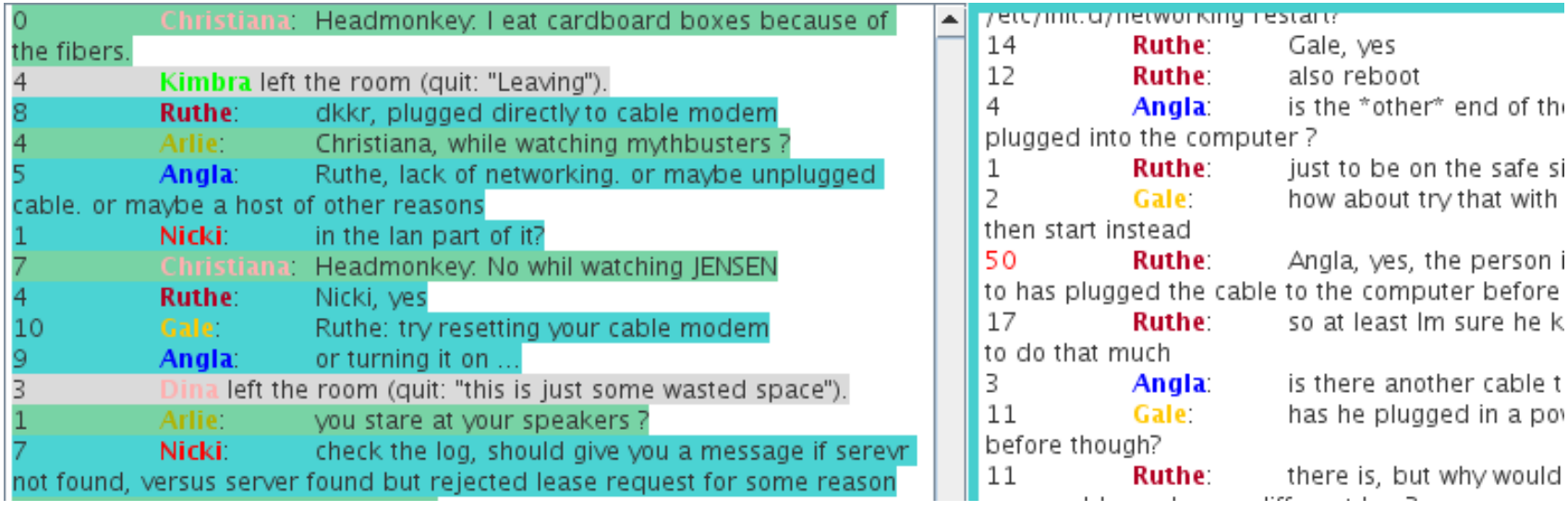
Excellent!

(Elsner+Charniak ACL 08, CL 10, ACL 11)

Preliminaries

Six annotators marked 800 lines of chat

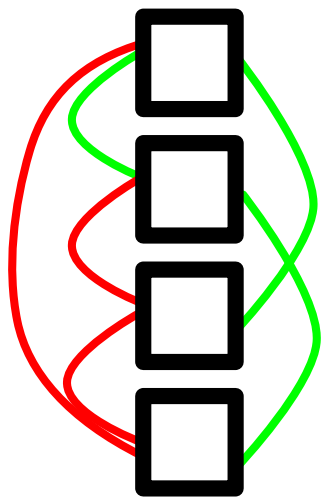
- From a Linux tech support forum on IRC



An initial model

Correlation clustering framework:

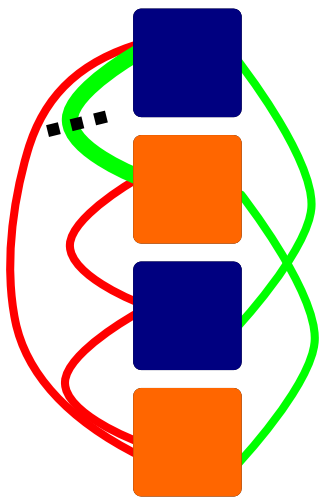
- Classify each pair of utterances as “same thread” or “different thread”
- Partition the transcript to keep “same” utterances together and split “different” ones apart
  - NP-hard, so we use heuristics



An initial model

Correlation clustering framework:

- Classify each pair of utterances as “same thread” or “different thread”
- Partition the transcript to keep “same” utterances together and split “different” ones apart
  - NP-hard, so we use heuristics



Classifying utterances

Pair of utterances: same conversation or different?

Chat-based features (F 66%)

- Time between utterances
- Same speaker
- Speaker's name mentioned

Classifying utterances

Pair of utterances: same conversation or different?

Chat-based features (F 66%)

Discourse features (F 58%)

- Questions, answers, greetings, etc.

Classifying utterances

Pair of utterances: same conversation or different?

Chat-based features (F 66%)

Discourse features (F 58%)

Word overlap (F 56%)

- Weighted by word probability in corpus
- Simplistic coherence feature

Classifying utterances

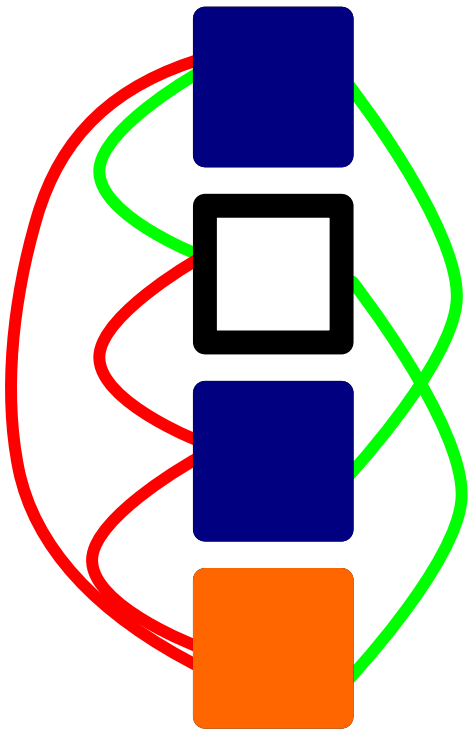
Pair of utterances: same conversation or different?

- Chat-based features (F 66%)
- Discourse features (F 58%)
- Word overlap (F 56%)
- Combined model (F 71%)

Assigning a single sentence

It's easy to maximize the objective locally...

- Even though the global problem is hard



	Accuracy
Same as previous	56
Corr. Clustering	**76**

Models from text and speech

Models which may apply here...

- Initially designed for putting sentences in order
- Distinguish coherent sequence of utterances from randomness
- Many different aspects of language
- *Not* all our own work.

Entity grid

Model of transitions from sentence to sentence
(Lapata+Barzilay '05,Barzilay+Lapata '05):

Text	Syntactic role
Suddenly **a White Rabbit** ran by her.	**subject**
Alice heard **the Rabbit** say "I shall be late!"	**object**
The Rabbit took a watch out of its pocket.	**subject**
Alice started to her feet.	**missing**

Topical entity grid

Relationships between *different* words

"a crow infected with **West Nile**..."
"**the outbreak** was the first..."

Our own work.

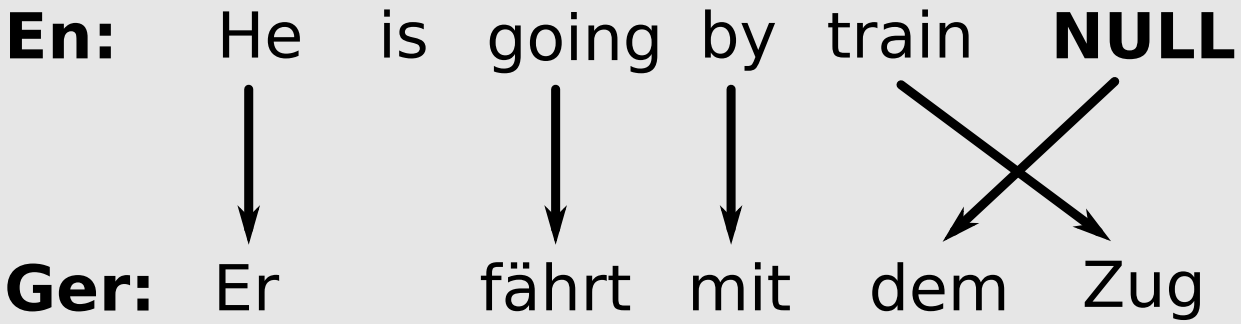
- Represents words in a "semantic space": LDA (Blei+al '01)
- Entity-grid-like model of transitions
- "Semantics" can be noisy...
  - More sensitive than the Entity Grid, but easy to fool!

IBM Model 1

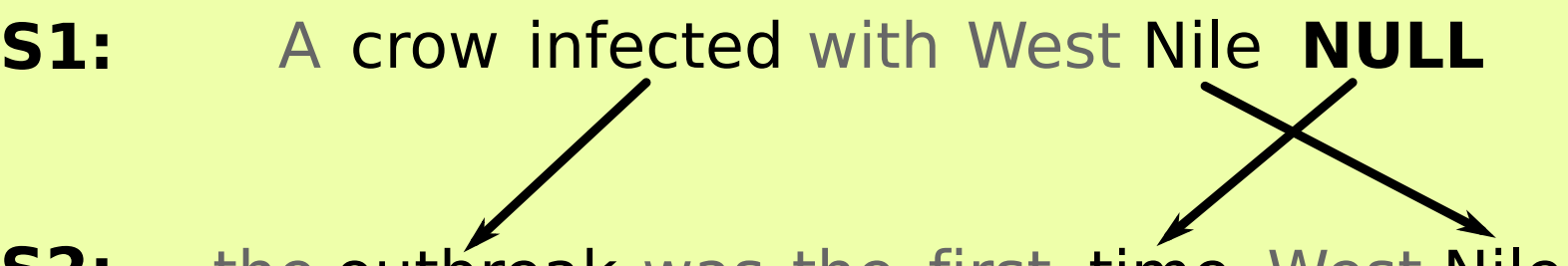
Single sentence of context

Learns word-to-word relationships directly

Brown+al `90



Soricut+Marcu `06



Pronouns

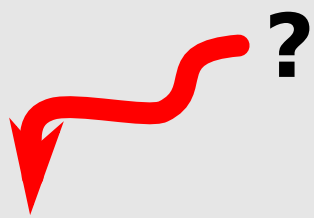
Detect passages with stranded pronouns:
(Charniak+Elsner '09), (Elsner+Charniak '08)

Good

Marlow sat cross-legged.

He had sunken cheeks.

Bad



The day was ending.

?

He had sunken cheeks.

Old vs new information

New information needs complex packaging

"Secretary of State Hillary Clinton"

Old information doesn't

"Clinton"

Soft constraints: put the "new"-looking phrase first

(Elsner+Charniak '08) following (Poesio+al '05)

Old vs new information

New information needs complex packaging

"Secretary of State Hillary Clinton"

Old information doesn't

"Clinton"

Soft constraints: put the "new"-looking phrase first

(Elsner+Charniak '08) following (Poesio+al '05)

Works well for news, poorly on speech and chat

- Entities introduced in different ways

Synthetic speech transcripts

Have you ever been called by a computer?

And you have to stop everything to run to the phone?

It's just a computer voice on the line.

IThose are the ones I really, really hate.

What would you serve at a dinner party?

I try to keep to things I can prepare ahead of time.

I have one recipe for a really good type of meatball.

If it was informal, my first choice would be crawfish.

Have you ever been called by a computer?

What would you serve at a dinner party?

I try to keep to things I can prepare ahead of time.

And you have to stop everything to run to the phone?

I have one recipe for a really good type of meatball.

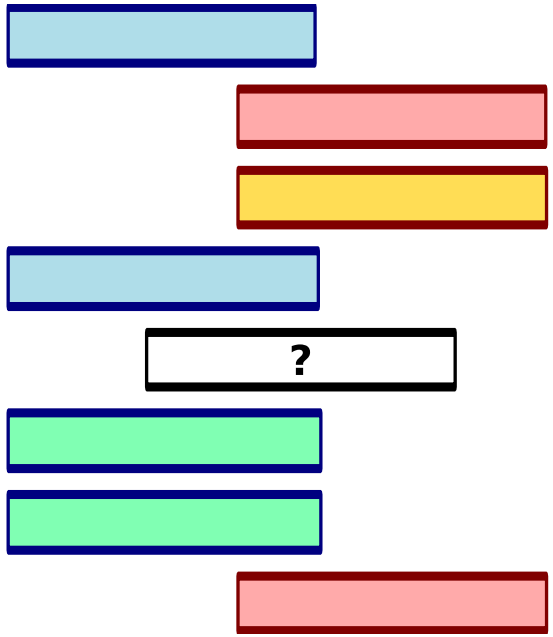
It's just a computer voice on the line.

IThose are the ones I really, really hate.

If it was informal, my first choice would be crawfish.

Speech results

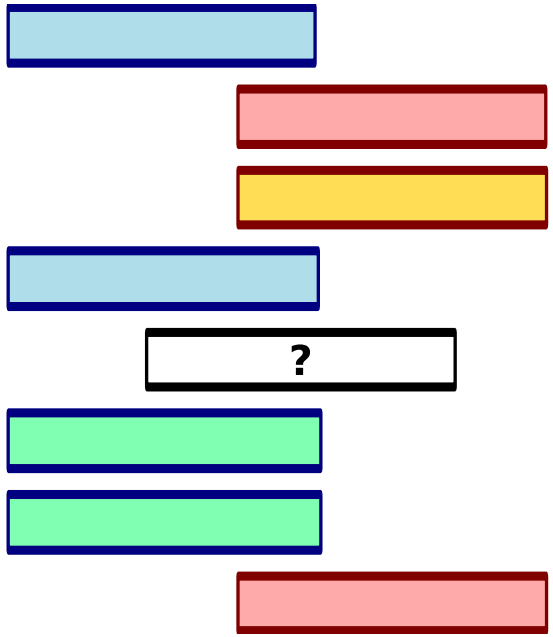
Chance	50
Corr. Clustering	**76**
EGrid	77
Topical EGrid	78
IBM-1	69
Pronouns	52
Time	58
Combined	**83**



- Coherence approach outperforms previous

Speech results

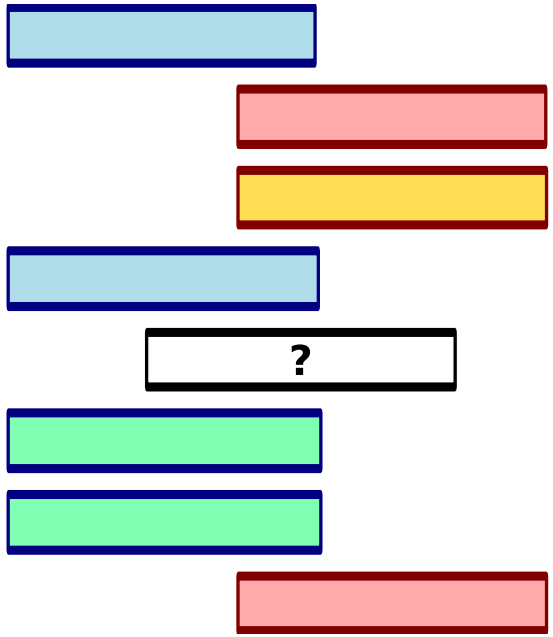
Chance	50
Corr. Clustering	**76**
EGrid	77
Topical EGrid	**78**
IBM-1	69
Pronouns	52
Time	58
Combined	**83**



- Coherence approach outperforms previous
- Topical model is useful

Speech results

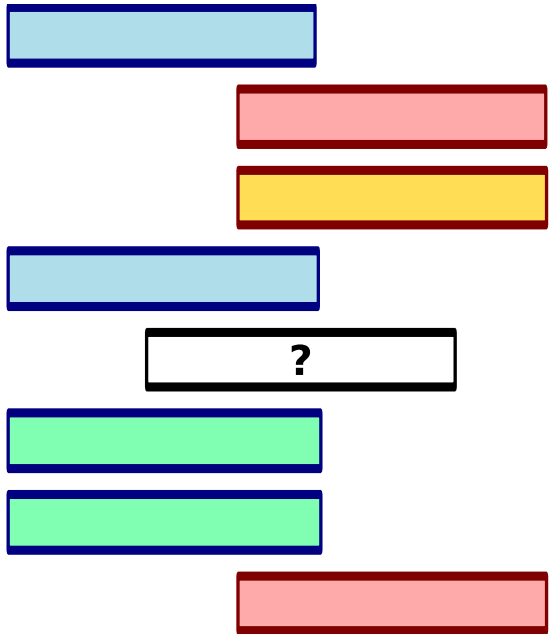
Chance	50
Corr. Clustering	**76**
EGrid	77
Topical EGrid	78
IBM-1	69
Pronouns	**52**
Time	58
Combined	**83**



- Coherence approach outperforms previous
- Topical model is useful
- Pronouns very bad

Speech results

Chance	50
Corr. Clustering	**76**
EGrid	77
Topical EGrid	78
IBM-1	69
Pronouns	52
Time	58
Combined	**83**



- Coherence approach outperforms previous
- Topical model is useful
- Pronouns very bad

Best models: sensitive, many-sentence context

Pronominals in speech and text

Different usage patterns...

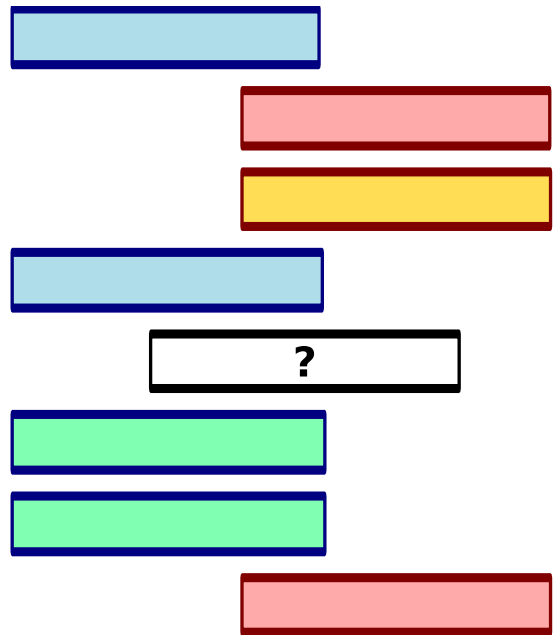
Corpus	Deictics	Pronouns	3rd person pronouns
WSJ	.04	0.64	0.52
Switchboard	.12	1.18	0.39
IRC	.09	0.92	0.31

News models totally inadequate here...

- Microtext also differs from speech pattern

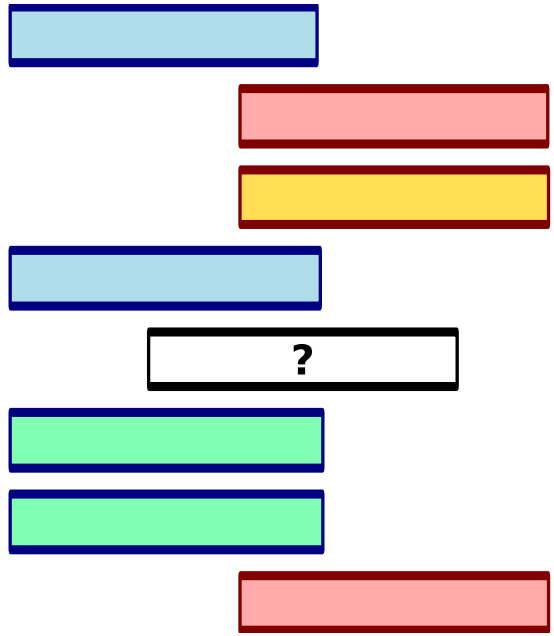
IRC results

Chat-specific	74



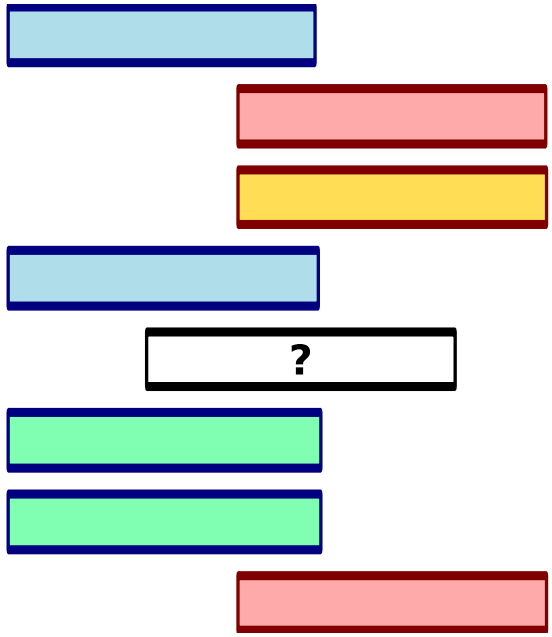
IRC results

Chat-specific	74
Corr. Clustering	76



IRC results

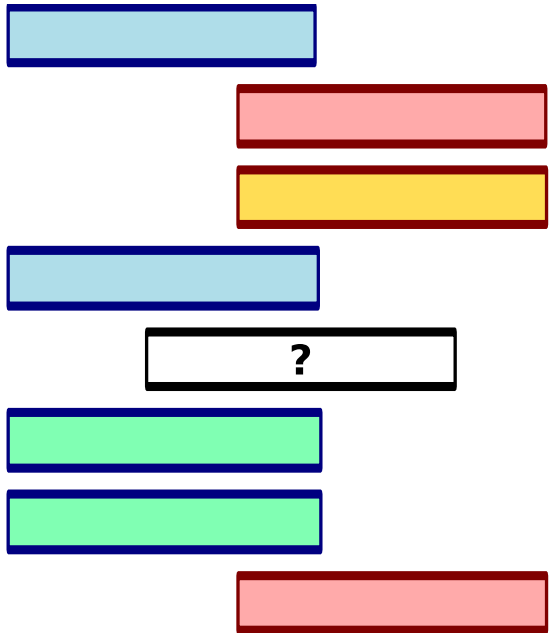
Chat-specific	74
Corr. Clustering	76
Chat+EGrid	**79**



- Coherence still outperform previous

IRC results

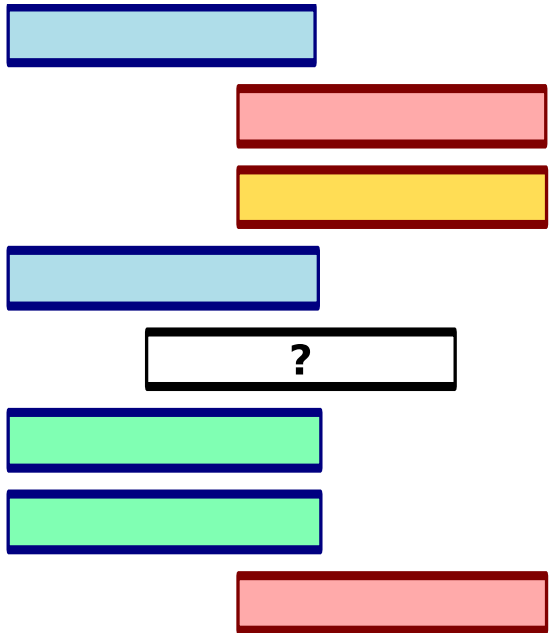
Chat-specific	74
Corr. Clustering	76
Chat+EGrid	**79**
Chat+Topical EGrid	77
Chat+IBM-1	76



- Coherence still outperform previous
- Lexical models not as good
  - Lack of data: trained on phone conversations

IRC results

Chat-specific	74
Corr. Clustering	76
Chat+EGrid	**79**
Chat+Topical EGrid	77
Chat+IBM-1	76
Chat+Pronouns	74



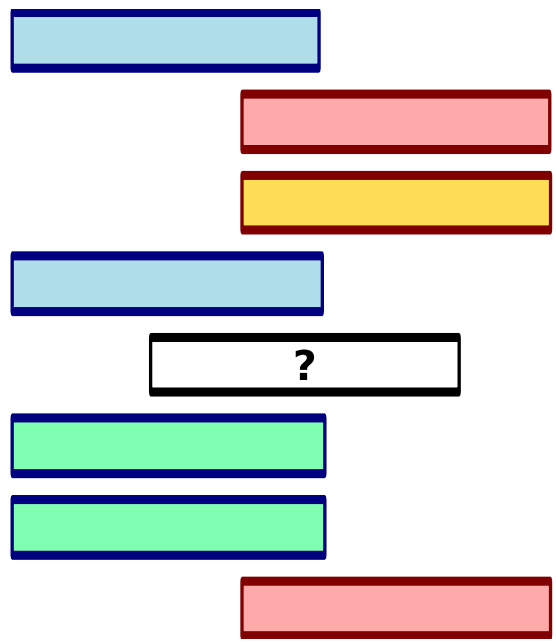
- Coherence still outperform previous
- Lexical models not as good
  - Lack of data: trained on phone conversations
- Pronouns same as before

More data

800 utterances not enough for you?

- Much larger corpora from (Martell+Adams '08)
  - Using our annotation software and protocol
  - $\sim$ 20000 total utterances from three newsgroups

	M+A corpora
Corr. Clustering	89
EGrid	**93**



Full-scale disentanglement

Unfortunately, scalability problems with advanced models...

- So results only for simple model

	Annotators		
Agreement	53		

Full-scale disentanglement

Unfortunately, scalability problems with advanced models...

- So results only for simple model

	Annotators	Best Baseline	
Agreement	53	35 (Pause 35)	

Full-scale disentanglement

Unfortunately, scalability problems with advanced models...

- So results only for simple model

	Annotators	Best Baseline	**Corr. Clustering**
Agreement	53	35 (Pause 35)	**41**

Full-scale disentanglement

Unfortunately, scalability problems with advanced models...

- So results only for simple model

	Annotators	Best Baseline	**Corr. Clustering**
Agreement	53	35 (Pause 35)	**41**

Best overall result: (Wang+Oard '09) 47

What we've learned

IRC chat is like speech

- Turn-taking and floor control
- Models based on lexical/entity coherence
  - But resource-poor (should be surmountable)

Real differences exist

- Floors more fluid
- Referring behavior
  - Full NPs
  - Pronominals

Conclusions

Chat disentanglement

- Sophisticated models can help!
- Still technical problems
  - Scaling inference, building topic models...
- Some real differences from speech
  - Coreference is a new challenge

Conclusions

Chat disentanglement

- Sophisticated models can help!
- Still technical problems
  - Scaling inference, building topic models...
- Some real differences from speech
  - Coreference is a new challenge

Microtext

- Interface determines communication behavior
- May vary from *any* previous mode of communication
  - Important to consider before applying off-the-shelf models

Thanks!

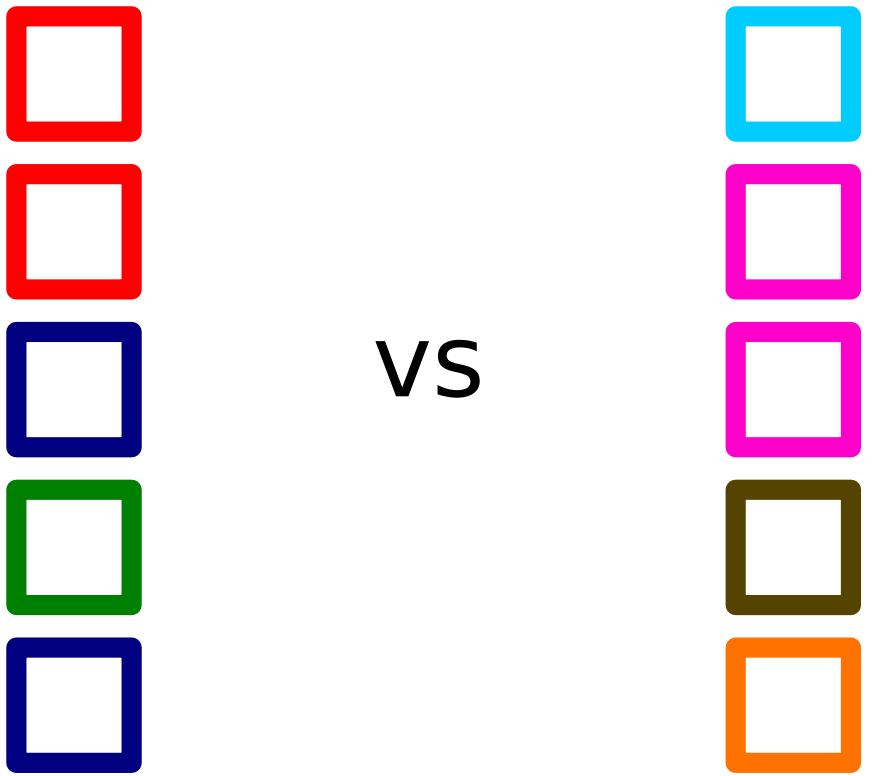
Thanks to...

- Eugene Charniak, Mark Johnson, Regina Barzilay
- Former labmates at Brown University
- Google Fellowship for NLP
- Craig Martell for NPS dataset

Corpus and software available

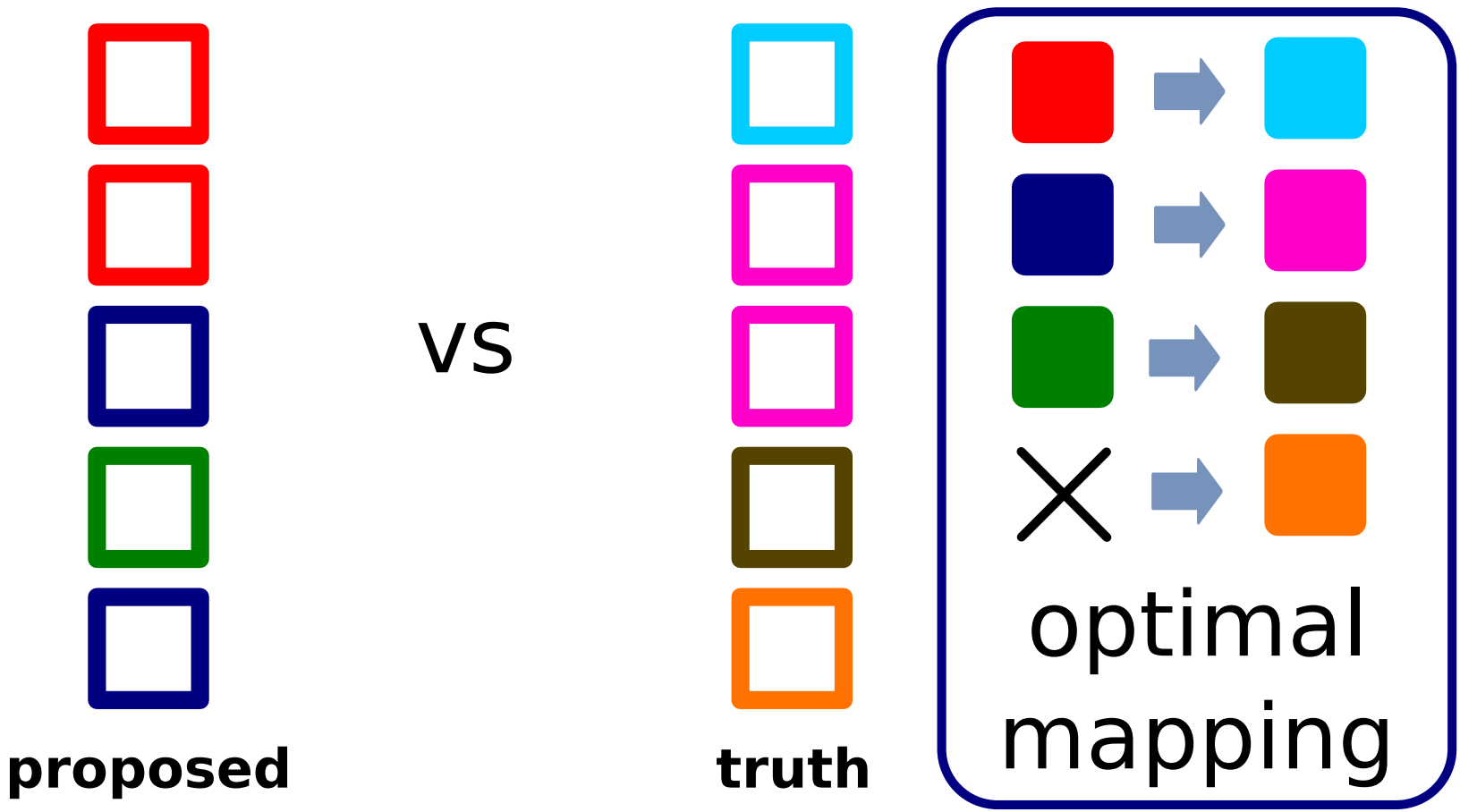
cs.brown.edu/∼ melsner
bitbucket.org/melsner/browncoherence

One-to-one overlap

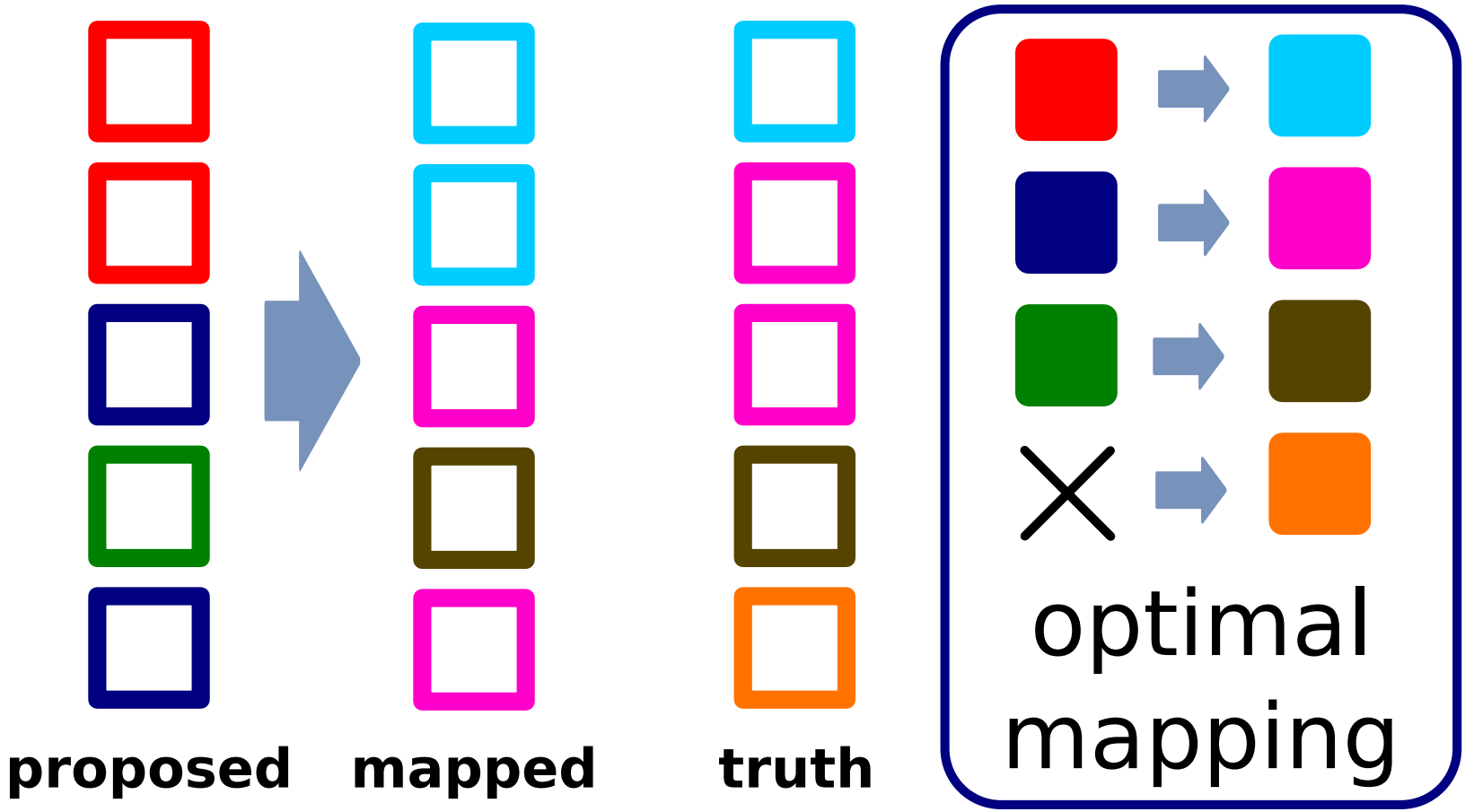


proposed **truth**

One-to-one overlap



One-to-one overlap



One-to-one overlap