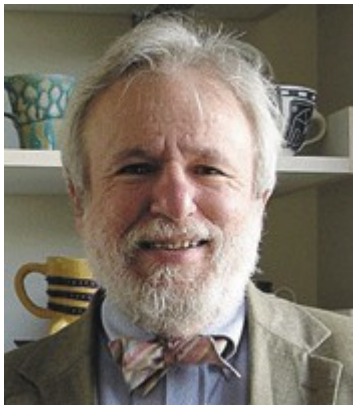
Entity-based Coherence: Going Off the Grid

Micha Elsner

Elsner, Austerweil, Charniak: NAACL '07
(Unified Model of Local and Global Coherence)

Elsner, Charniak: ACL '08
(Coreference-inspired Coherence Modeling)



BLIP BROWN

Text!

- Sir Walter Elliot, of Kellynch Hall, in Somersetshire, was a man who never took up any book but the Baronetage.
- Sir Walter had improved it by adding the day he had lost his wife.
- There followed the history of the ancient family.
- Vanity was the beginning and end of Sir Walter Elliot's character.
- He had been remarkably handsome in his youth.

Coherence

- Consistent topic.
- Earlier sentences provide context for later ones.

Sir Walter Elliot, of Kellynch Hall, in Somersetshire, was a man who never took up any book but **the Baronetage**.
Sir Walter had improved **it** by adding the day he had lost his wife.

Coherence

- Consistent topic.
- Earlier sentences provide context for later ones.

Sir Walter Elliot, of Kellynch Hall, in Somersetshire, was a man who never took up any book but **the Baronetage**.
Sir Walter had improved **it** by adding the day he had lost his wife.

- Not:

Sir Walter had improved **it** by adding the day he had lost his wife.
He had been remarkably handsome in his youth.

Applications

- **Create** coherent text:
  - Summarize
  - Add new facts
- **Evaluate** texts:
  - Essay scoring
- **Understand** text pragmatics:
  - Coreference
  - Topicality

Our Approach: Entities

- Objects in the world:

- Referred to in language:
  Sir Walter Elliot
  Sir Walter
  He

- Coherence: *what* gets mentioned, and *how*.
  - Other approaches: lexical, rhetorical.

Overview

- Evaluation Tasks
- Previous: Entity Grids
- Topics
- Referring Expressions
- Open Problems...

Overview

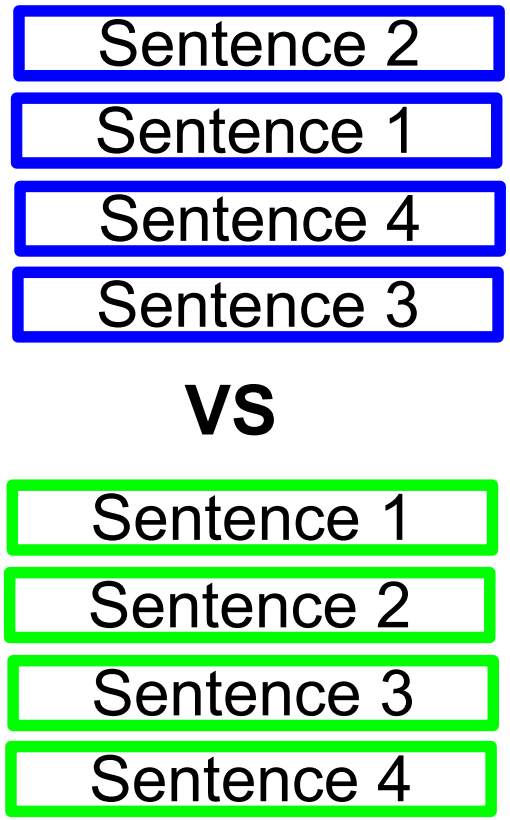
- **Evaluation Tasks**
- Previous: Entity Grids
- Topics
- Referring Expressions
- Open Problems...

Corpora

- Airplane
  - Reports of plane crashes
  - Short (11 ss)
  - Stereotyped: 40% begin “This is preliminary information”
- WSJ
  - Standard news corpus
  - Longer (25 ss)
  - More natural syntax

Discriminative task

- Binary judgement between random permutation and original document.

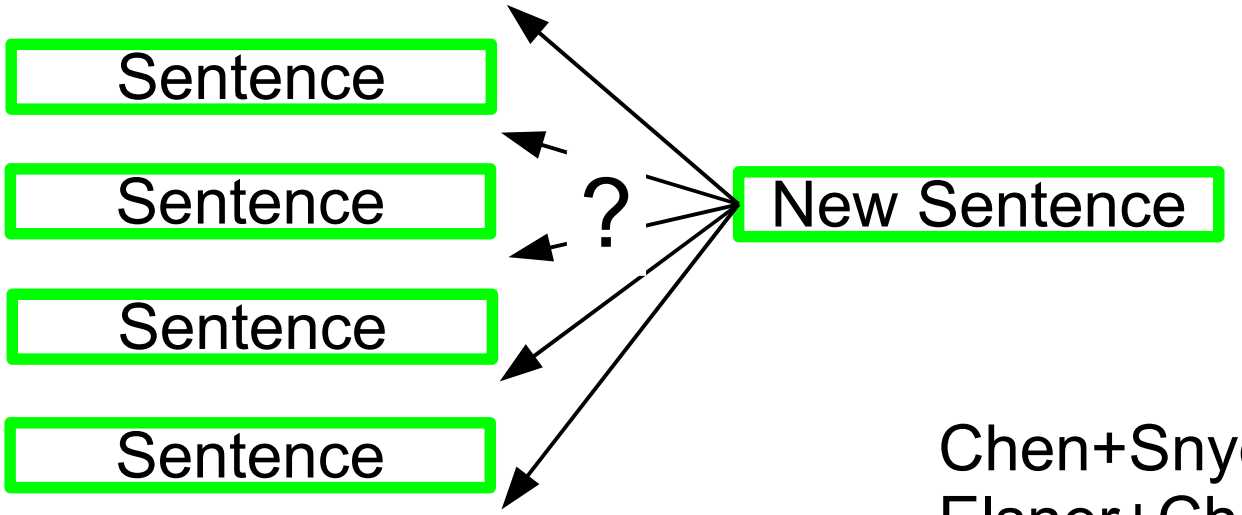


- Fast, convenient test.
- Longer documents are *much* easier!
- F-score (classifier can abstain).

Barzilay+Lapata '05

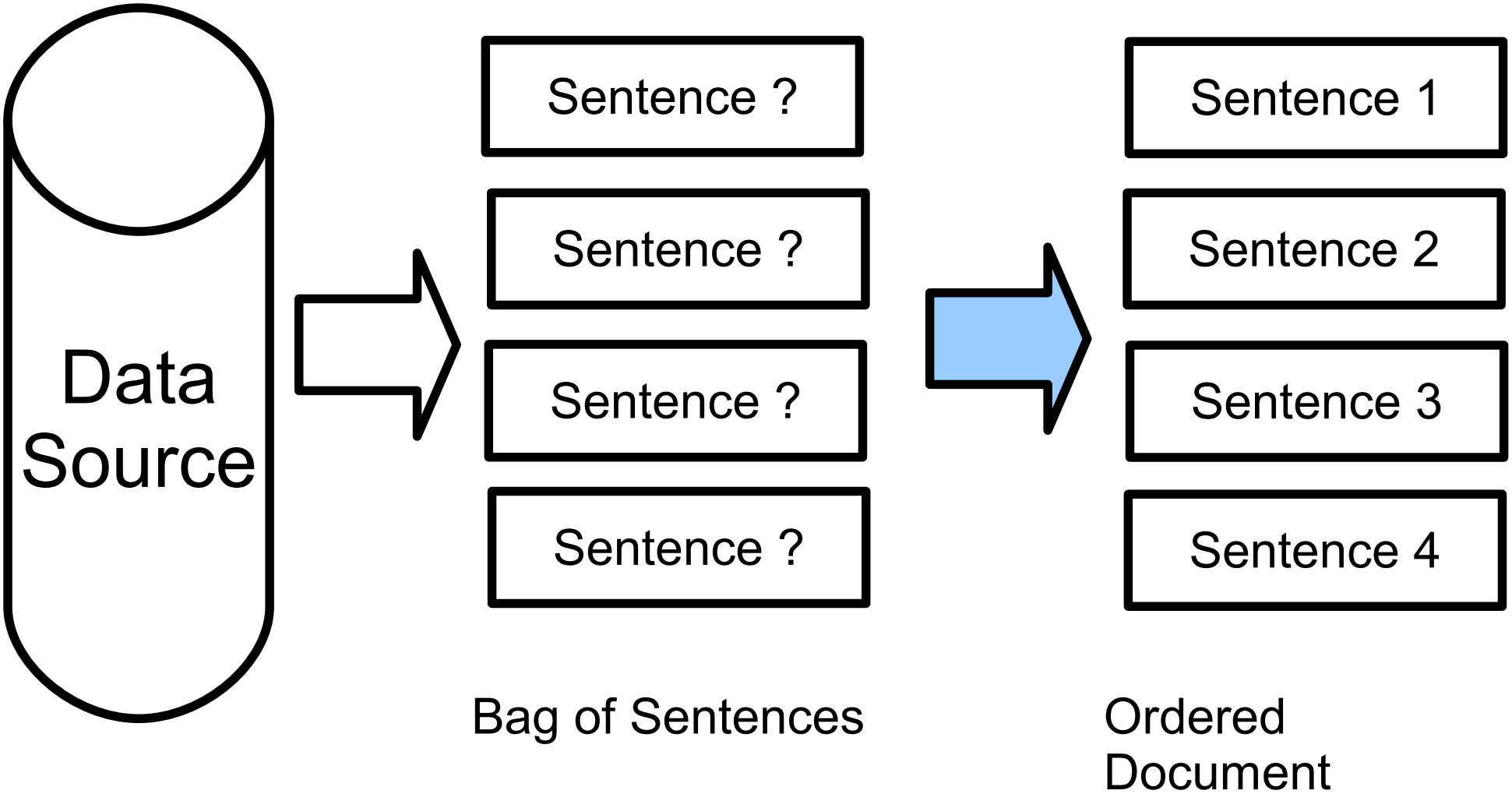
Insertion task

- Remove and re-insert one sentence at a time.
- Examines permutations closer to the original ordering.
  - Hard even for long documents.
  - Report percent exactly correct.



Chen+Snyder+Barzilay '07
Elsner+Charniak '07

Sentence Ordering



Ordering metric

- Kendall's Tau (rank ordering distance)
- Counts pairwise swaps
- No concept of structure
  - Moving a paragraph vs. moving sentences
  - Good for short documents (Lapata 2006)

Overview

- Evaluation Tasks
- **Previous: Entity Grids**
- Topics
- Referring Expressions
- Conclusion
- Open Problems...

Entity Grid

Lapata+Barzilay '05

Entities in text (NPs)

Walter	Hall	Somerset	Bt.age	day	wife	history	family

Entity Grid

Lapata+Barzilay '05

Entities in text (NPs)

	Walter	Hall	Somerset	Bt.age	day	wife	history	family
Sentence #: 1	S	X	X	O				

Sir Walter Elliot, of Kellynch Hall, in Somersetshire, was a man who never took up any book but the Baronetage.

Entity Grid

Lapata+Barzilay '05

Entities in text (NPs)

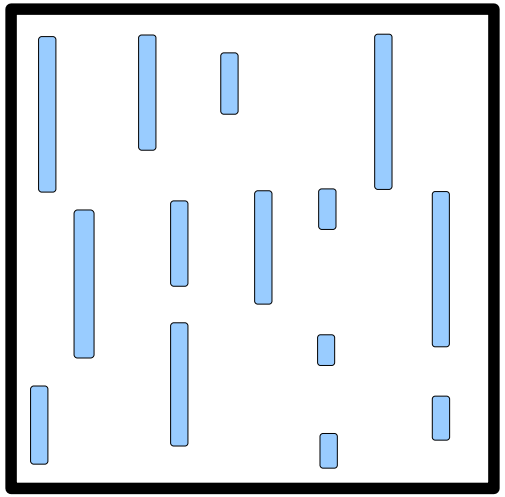
Sentence #:	Walter	Hall	Somerset	Bt.age	day	wife	history	family
1	S	X	X	O				
2	S			O	X	O		
3							X	X

Sir Walter Elliot, of Kellynch Hall, in Somersetshire, was a man who never took up any book but the Baronetage.

Sir Walter had improved it by adding the day he had lost his wife.

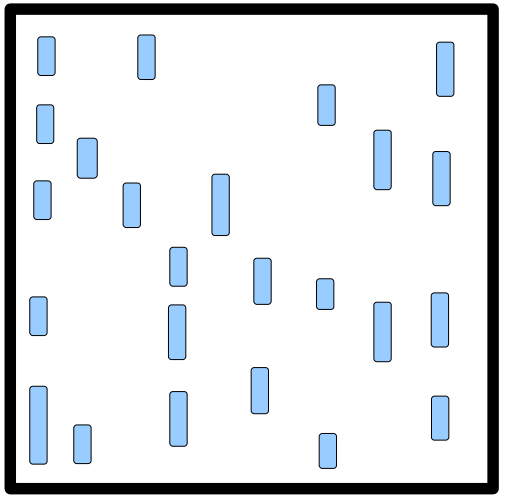
There followed the history of the ancient family.

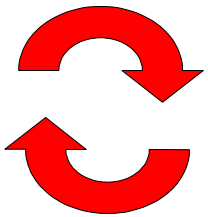
Local coherence

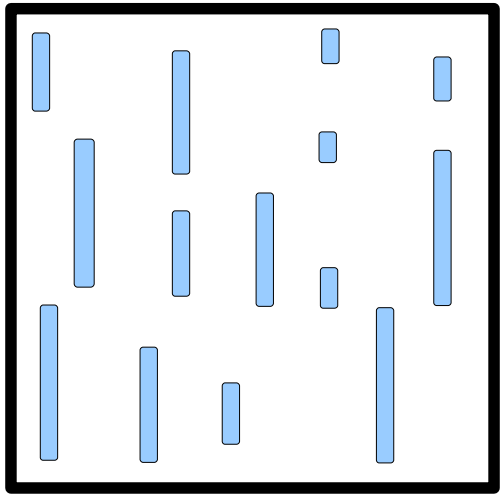


Very low zoom:
entities in long contiguous columns.

A randomly
permuted document:

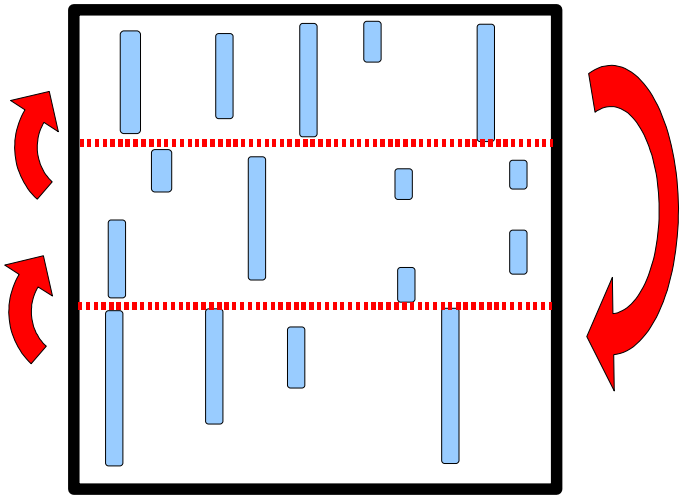






Backwards?

Move the
paragraphs?



Independence assumptions

- Real entities:
  topically related.

- Grid entities:
  independent!

	Walter	wife	family
	S		
	S	O	
			X

Referring expressions

- NPs treated as transparent:
  Sir Walter Elliot
  Elliot
  both handled the same.
- 'Same head' heuristic to fake coreference.
  - About 2/3 accurate (Poesio+Vieira).

Results

Airplane	Disc (%)	Ordering (τ)
Barzilay+Lapata (EGrid)	90	
Generative EGrid	81	0.17

WSJ	Disc (F)	Ins (prec)
Generative EGrid	73	18.1

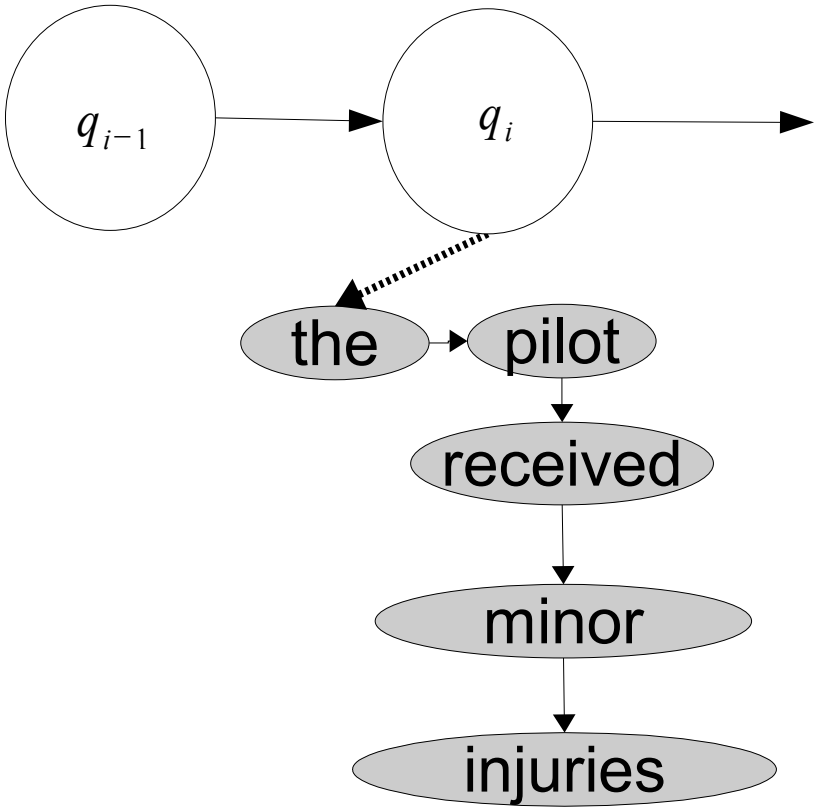
- % vs. F: roughly equivalent here
- Good discrimination, poor ordering.

Overview

- Evaluation Tasks
- Previous: Entity Grids
- **Topics**
- Referring Expressions
- Open Problems...

Markov Model

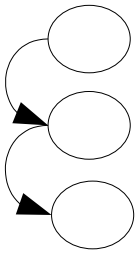
- Hidden Markov Model for document structure.
- Language model for each state.



Barzilay and Lee 2004

Global Coherence

- HMM learns overall document structure:
  - Start, end, topic shift.
- **All** local information stored in the state.
  - Sparsity issues.



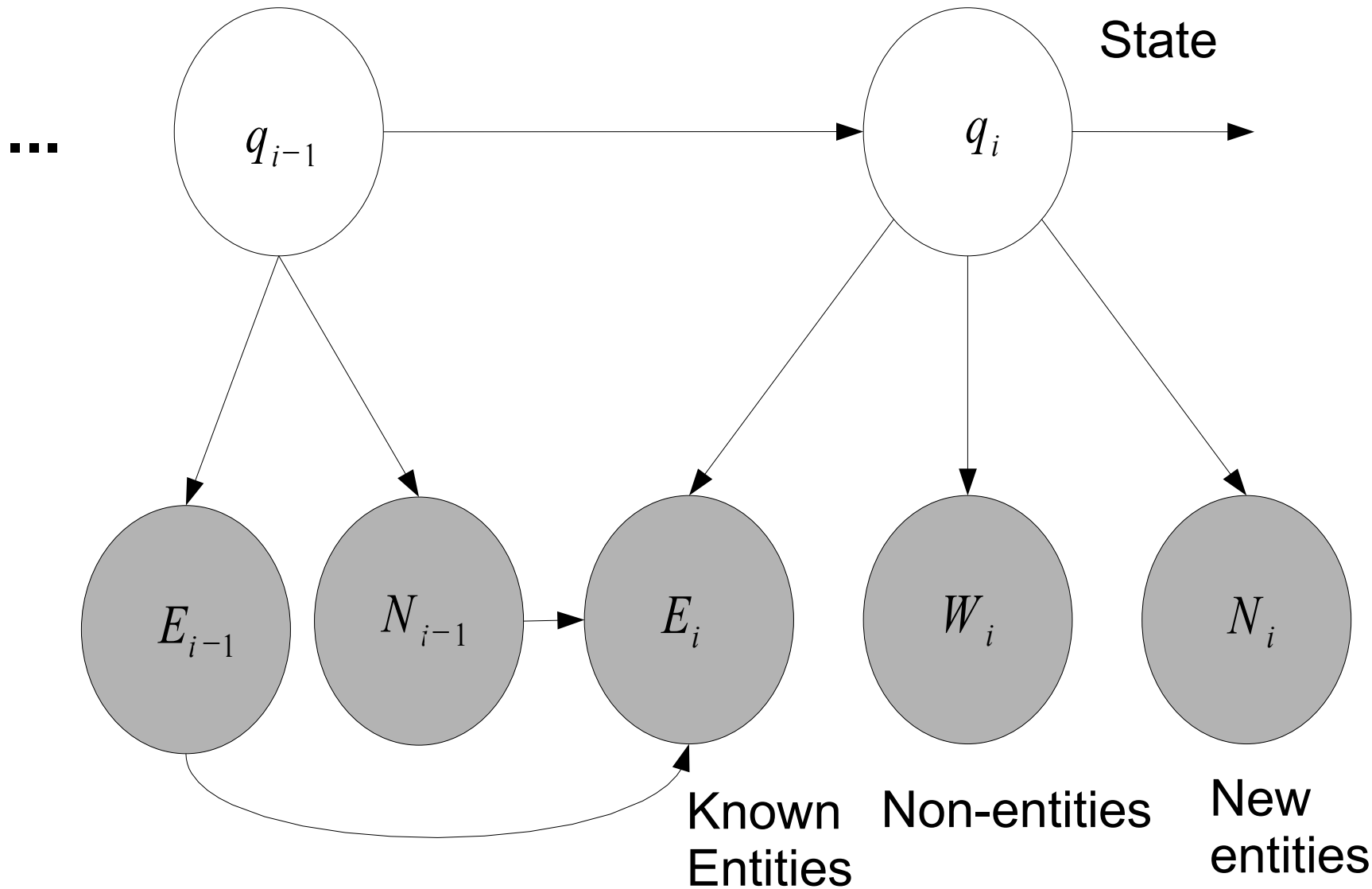
A **wombat** escaped from the cargo bay.
Finally the **wombat** was captured.
The last major **wombat** incident was in 1987.

- Is there a state q-wombat?

Unified Model

- HMM structure:
  - States generate entities.
  - Back off to Entity Grid.
  - Also generate other words.
- Entity Grid prior:
  - Repeat entities regardless of state.
  - (New estimator for the entity grid; mistake in original results.)

Graphical Model



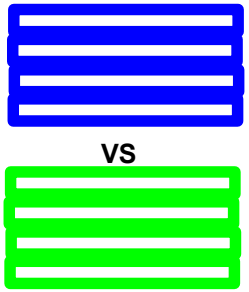
Soricut + Marcu '06

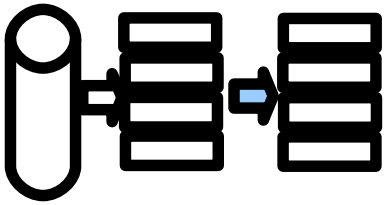
- Mixture model:
  - HMM, entity grid and word-to-word (IBM) components.
  - Results are as good as ours.
- No joint learning.
  - No relationship between topic and grid.
- Uses more information (ngrams and IBM).
  - Might be improved by adding our model.

Results

Airplane Corpus: short documents

Airplane Test	Disc (%)	Ordering (τ)
Barzilay+Lapata (EGrid)	90	-
Generative EGrid	81	0.17
Barzilay+Lee (HMM)	74	0.44
Soricut+Marcu (Mixture)	-	**0.50**
Unified (Egrid/HMM)	**94**	**0.50**





Overview

- Evaluation Tasks
- Previous: Entity Grids
- Topics
- **Referring Expressions**
- Open Problems...

Anatomy of an unfamiliar NP

full name and title

Sir Walter Elliot, of Kellynch Hall, in Somersetshire,

was a man who...

- Lots of linguistic markers to introduce this guy...
  - because you don't know who he is.

Anatomy of an unfamiliar NP

full name and title

long phrasal modifier

Sir Walter Elliot, of Kellynch Hall, in Somersetshire,

was a man who...

- Lots of linguistic markers to introduce this guy...
  - because you don't know who he is.

Anatomy of an unfamiliar NP

full name and title

long phrasal modifier

Sir Walter Elliot, of Kellynch Hall, in Somersetshire,

was a man who...

copular verb

- Lots of linguistic markers to introduce this guy...
  - because you don't know who he is.

Terminology

- First mention of entity: ***discourse-new***
  - Usually unfamiliar: *hearer-new*
  - Hearer-new NPs typically marked.
- Subsequent: *given*, ***discourse-old***
- Discourse-new isn't always hearer-new.
  - **Unique** entities (the FBI)

Prince '81

Lots of features!

- **Appositives**: Mr. Shepherd, *a civil, cautious lawyer...*
- **Restrictive relative clauses**: *the first man to...*
- **Syntactic position**: subject, object &c
- **Determiner / quantifier**: *a* (new), *the* (complicated!)
- **Titles and abbreviated titles**:
  - *Sir*, *Professor* (usually new); *Prof.*, *Inc.* (usually old)
- **How many modifiers?:** More implies newer.

- Most important feature: **same head occurred before?**

Vieira+Poesio '00
Ng+Cardie '02
Uryupina '03 ...

Previous work (classifiers)

- Used for coreference resolution:
  - Don't resolve the **new NPs**.
  - Do resolve the **old ones**.

Joint decisions:
Denis+Baldridge '07

Sequential:
Poesio+al '05
Ng+Cardie '02

- Almost any machine learning algorithm available...
- All score about 85%.
- (Relies on document being in order.)

Modeling coherence

Sir Walter Elliot, of Kellynch Hall, in Somersetshire

he

his

Walter Elliot

Sir Walter

himself

Sir Walter

Sir Walter Elliot

vs

Sir Walter

he

his

Sir Walter

Walter Elliot

Sir Walter Elliot, of Kellynch Hall, in Somersetshire

himself

Sir Walter Elliot

Now some computation...

P(new | Sir Walter Elliot, of Kellynch Hall, in Somersetshire)

P(old | he)
P(old | his)
P(old | Walter Elliot)
P(old | Sir Walter)
P(old | himself)
P(old | Sir Walter)
P(old | Sir Walter Elliot)

Where do the labels come from?
Full coreference!

$$P(\text{chain}) = \Pi\ P(\text{np})$$

$$P(\text{doc}) = \Pi\ P(\text{chain})$$

More realistic computation...

P(new|Sir Walter Elliot, of Kellynch Hall, in Somersetshire)
P(old|Walter Elliot)
P(old|Sir Walter Elliot)

One coreferential chain turns into two.
(Bad, but survivable.)

P(new|Sir Walter)
P(old|Sir Walter)

P(old|he)
P(old|his)
P(old|himself)

And what about the pronouns?
We'll come back to them later.

What else can go wrong?

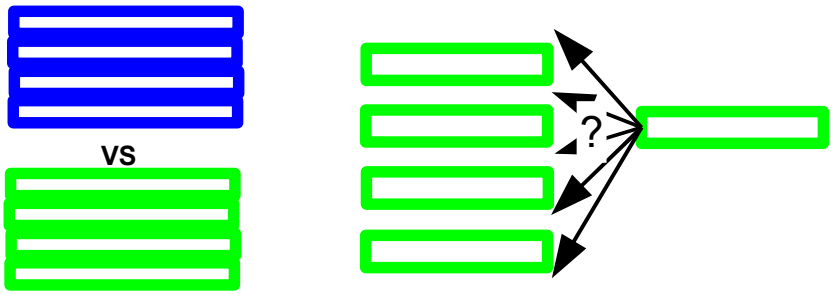
- Not all new NPs are unfamiliar.
  - Unique referents: **The FBI, the Golden Gate Bridge, Thursday**
  - Our technique will mislabel these.
- We can reduce error by distinguishing three classes: **new**, **old**, **singleton**
  - **singleton**: no subsequent coreferent NPs
  - often look more like old than new

corpus study: Fraurud '90
classifiers: Bean+Riloff '91
Uryupina '03

Results

WSJ corpus: longer documents

WSJ	disc (F)	ins (prec)
Entity Grid	73.2	18.1
NP syntax	**72.7**	**16.7**
Grid, NP syntax	**77.6**	**21.5**



Pronoun coreference

- Pronouns occur close after their antecedent nouns.

1 **Marlow** sat cross-legged right aft, leaning against the mizzen-mast.
2 **He** had sunken cheeks, a yellow complexion, a straight back, an ascetic aspect, and... resembled an idol.
3 The **director**, satisfied the anchor had good hold, made his way aft and sat down amongst us.
We exchanged a few words lazily. Afterwards there was silence on b
4 did not begi
and fit for nothing but placid staring. The day was ending in a serenity of still and exquisite brilliance.

No possible antecedents here!

Violations cause incoherence

1 **Marlow** sat cross-legged right aft, leaning against the mizzen-mast.

3 The **director**, satisfied the anchor had good hold, made his way aft and sat down amongst us.

We exchanged a few words lazily. Afterwards there was silence on board the yacht. For some

4 reason or

dominoes

but placid staring. The day was ending in a serenity of still and exquisite brilliance.

2 **He** had sunken cheeks, a yellow complexion, a straight back, an ascetic aspect, and... resembled an idol.

No possible antecedents here!

What sort of a model?

- Typical coreference models are conditional: P(antecedent | text)

Marlow sat ...

P(Marlow | he) = .99

He had sunken cheeks...

- Probability of linking the pronoun to each available referent.
- High for unambiguous texts...

What sort of a model?

- Typical coreference models are conditional: P(antecedent | text)

Marlow sat ...

P(Marlow | he) = .99 (still!)

We exchanged a few **words** lazily.

P(words | he) ≈ 0

There was **silence** on board the **yacht**.

P(yacht | he) ≈ 0

He had sunken cheeks...

Generative coreference

- Not only tell good *coreference assignments* from bad ones...
- But good *texts* from bad ones.
  - So we need P(text | antecedent)

- Luckily we can do that (sort of)...
  - Ge+Hale+Charniak '98
  - Charniak+Elsner '09 (talk Thursday!)

Results

- Improvements continue...
  - On its own, this model is not as strong as the syntactic one.

WSJ	disc (F)	ins (prec)
Entity Grid	73.2	18.1
Pronoun	**63.1**	**13.9**
EG, NP	77.6	21.5
EG, NP, Prn	**78.2**	**22.7**

Overview

- Evaluation Tasks
- Previous: Entity Grids
- Topics
- Referring Expressions
- **Open Problems...**

Topic model revisited

- Longer documents, larger domains:
  - Still use one state per topic?
  - How fine-grained are topics?
  - How predictable are transitions?
- Ordering task:
  - Doesn't scale!
  - Hierarchical ordering?
  - How do we score?

Referring expressions

- Full coreference?
  - Generative models now exist (Haghighi+Klein '07)
- Reference rewriting:
  - Related task.
  - Nenkova+McKeown '03; now a shared task: Belz+Gatt, Generation Challenge '08,'09
- Inferrables:
  - Mention of one entity (**Sir Walter**) allows definite description for another (**the family**)

Ordering paradigm

- Coherence: not just ordering!
- Relationship to general readability?
  - Sentence structure
  - Word choice

- Realistic source of *incoherence*...
  - Better than permuted documents.

Conclusion

- Started with Entity Grid (prev. work)
- Added:
  - Topic (HMM)
  - Disc-new NP detection
  - Pronoun coreference
- Much still to do...

Other research: chat

Does anyone here shave their head?

How do I limit the speed of my internet connection?

I shave part of my head.

A tonsure?

Use dialup!

Nope, I only shave the chin.

How do you cluster the different conversations?

Elsner+Charniak ACL '08

Other research: named entities

Unsupervised learning:

Which references go together?

What is their structure?

President Bill Clinton
President Clinton
Bill
Mr. Clinton

Secretary Hillary Clinton
Hillary Rodham Clinton
Hillary Clinton
Ms. Clinton

Clinton

Elsner+Charniak+Johnson, NAACL '09

Other research: copyediting

California voters will ~~get a chance in a November vote~~ **have the chance to vote in November** **on whether** to end gay marriage...

- How does editing change a document?
- Do our models have the same preferences as editors?

Thanks!

- Regina Barzilay, Erdong Chen
- Mirella Lapata
- Olga Uryupina
- all of BLLIP (and Tom Griffiths)
- DARPA GALE, Karen T. Romer Foundation
- Everyone here!

Code is available:
http://www.cs.brown.edu/people/melsner