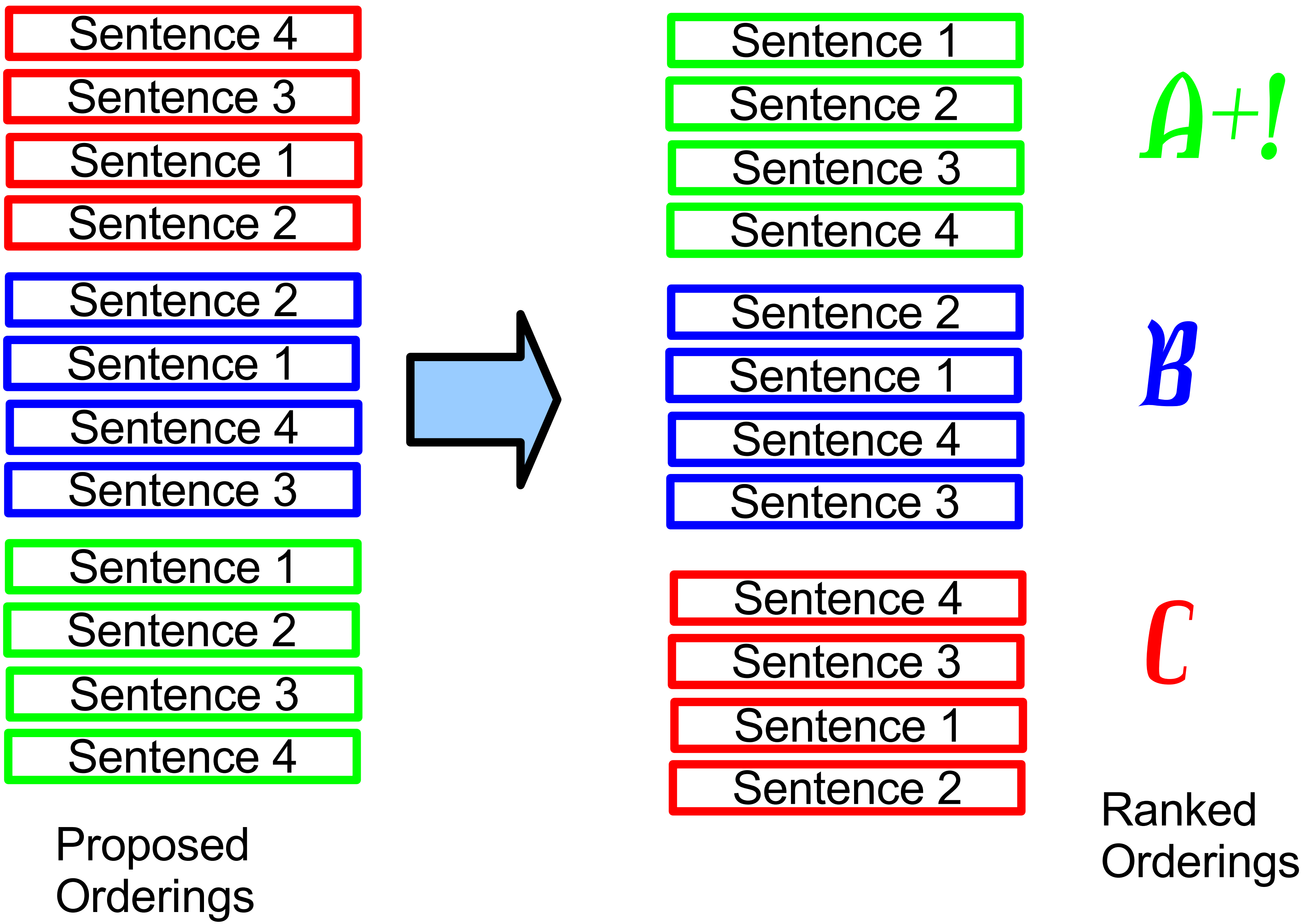
A Unified Local and Global Model for Discourse Coherence

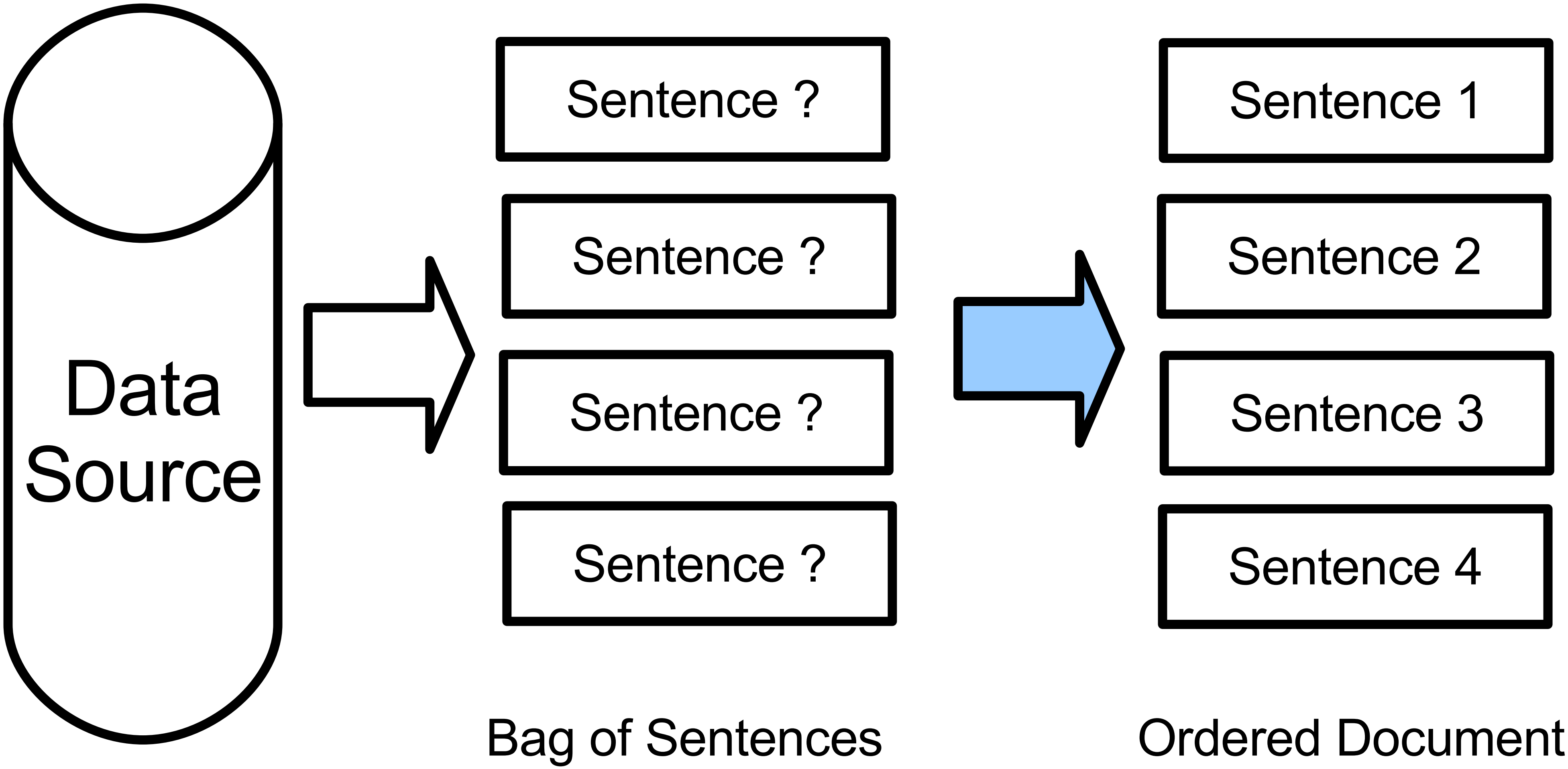
Micha Elsner, Joseph Austerweil, Eugene Charniak

Brown Laboratory for Linguistic Information Processing (BLLIP)

Coherence Ranking



Sentence Ordering



Overview

- **Previous Work: Entity Grids**
- Previous Work: Hidden Markov Model
- Relaxed Entity Grid
- Unified Hidden Markov Model
- Corpus and Experiments
- Conclusions and Future Work

An Entity Grid

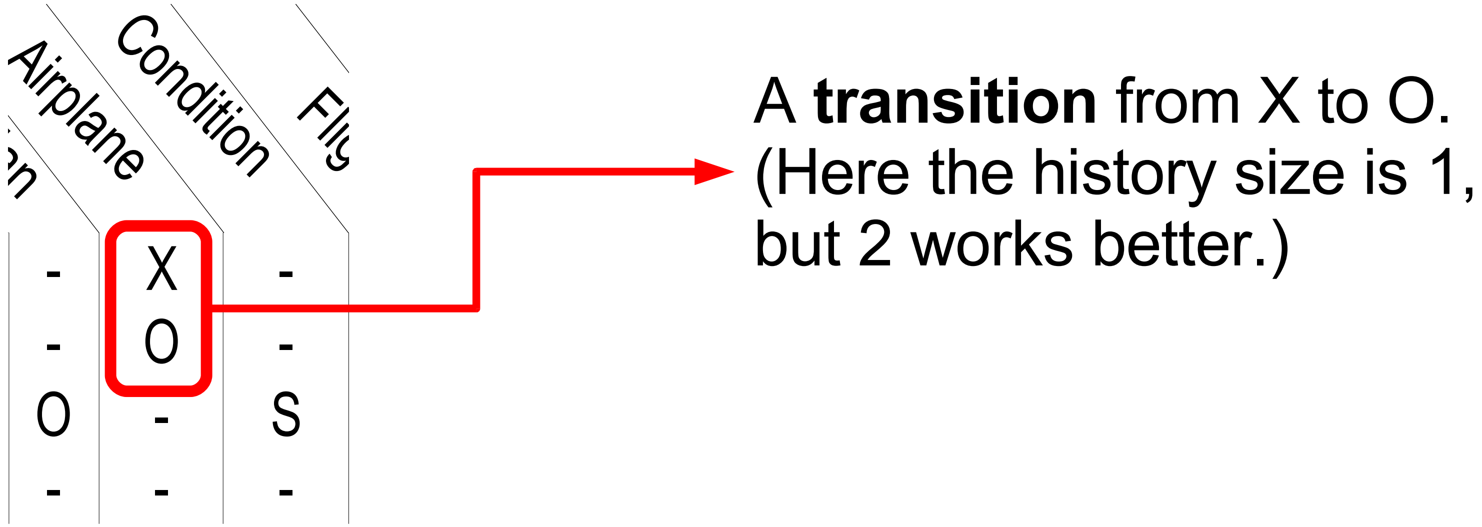
Barzilay and Lapata '05, Lapata and Barzilay '05.

The commercial **pilot**, sole **occupant** of the **airplane**, was not injured.
The **airplane** was owned and operated by a private **owner**.
Visual meteorological **conditions** prevailed for the personal cross country **flight** for which a VFR flight **plan** was filed.
The **flight** originated at Nuevo **Laredo** , Mexico , at approximately 1300.

Syntactic Role in Sentence	Plan	Airplane	Condition	Flight	Pilot	Laredo	Owner	Occupant
0	-	X	-	-	O	-	-	X
1	-	O	-	-	-	-	X	-
2	O	-	S	X	-	-	-	-
3	-	-	-	S	-	X	-	-

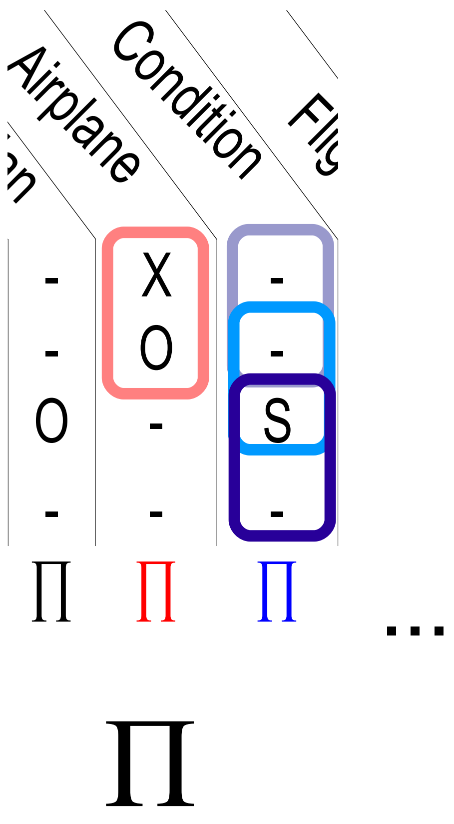
Local Coherence: Entity Grids

- Loosely based on Centering Theory.
  - Coherent texts *repeat important nouns*.
- Grid shows most prominent role of each head noun in each sentence.



Computing with Entity Grids

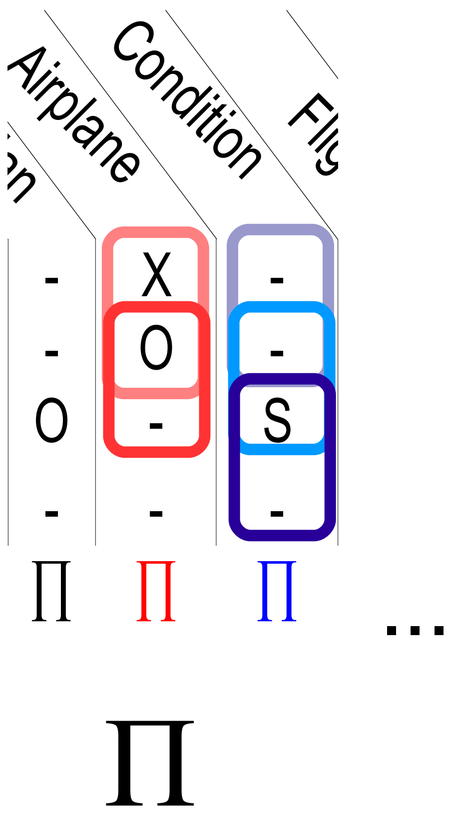
- Generatively: Lapata and Barzilay.
  - Assume independence between columns.



- This independence assumption can cause problems for the generative approach.
  - Barzilay and Lapata get better results with SVMs.

Computing with Entity Grids

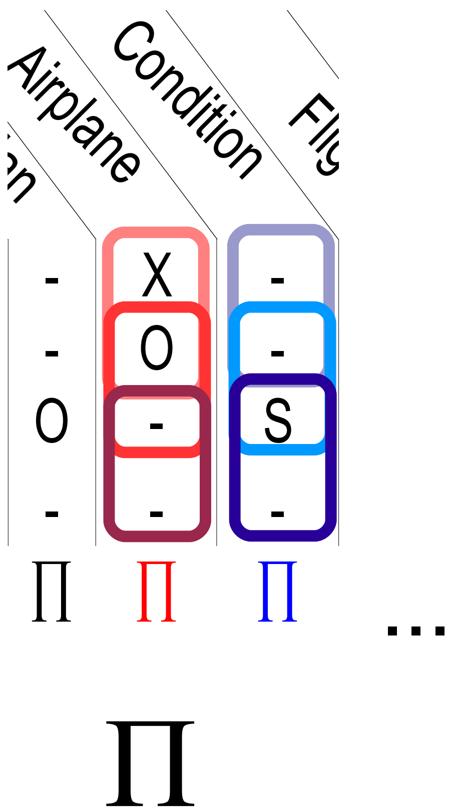
- Generatively: Lapata and Barzilay.
  - Assume independence between columns.



- This independence assumption can cause problems for the generative approach.
  - Barzilay and Lapata get better results with SVMs.

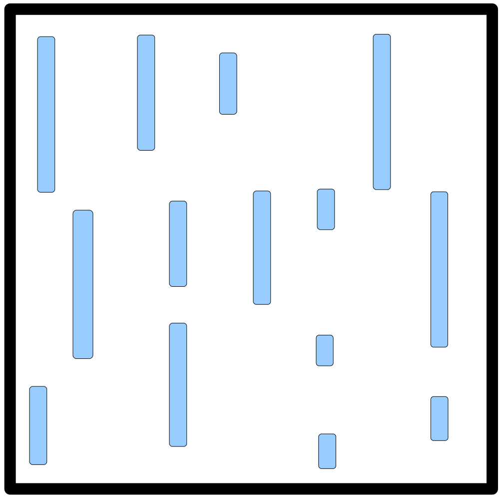
Computing with Entity Grids

- Generatively: Lapata and Barzilay.
  - Assume independence between columns.

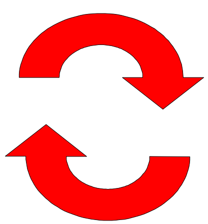


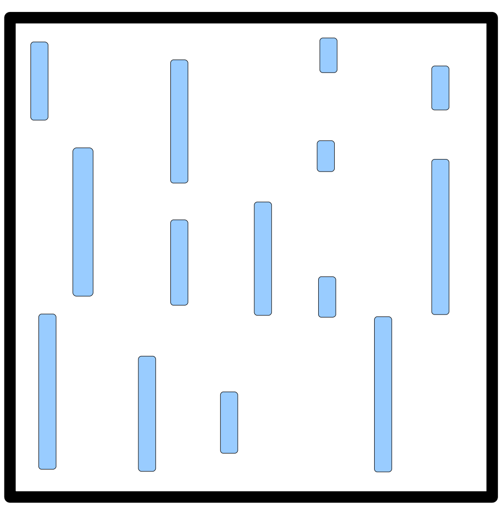
- This independence assumption can cause problems for the generative approach.
  - Barzilay and Lapata get better results with SVMs.

Entity Grids Model Local Coherence

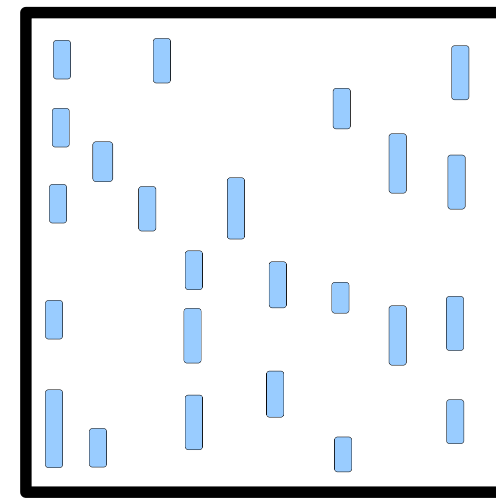


A coherent entity grid at very low zoom: entities occur in long contiguous columns.

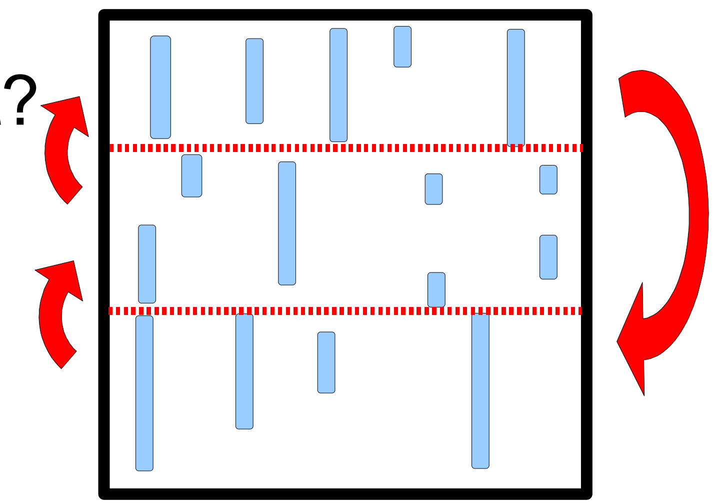




A grid for a randomly permuted document tends to look like this.



But what if we flip it? Or move around paragraphs?

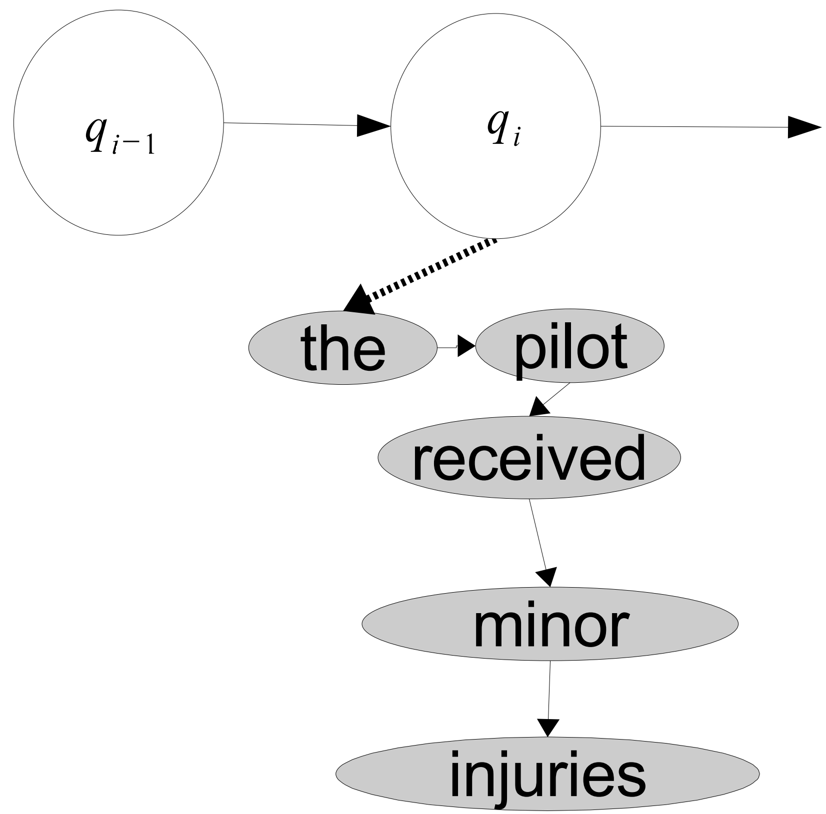


Overview

- Previous Work: Entity Grids
- **Previous Work: Hidden Markov Model**
- Relaxed Entity Grid
- Unified Hidden Markov Model
- Corpus and Experiments
- Conclusions and Future Work

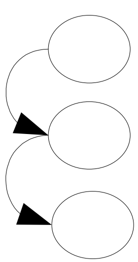
Markov Model

- Barzilay and Lee 2004, “Catching the Drift”
- Hidden Markov Model for document structure.
- Each state generates sentences from another HMM.



Global Coherence

- The HMM is good at learning overall document structure:
  - Finding the start, end and boundaries.
- But **all** local information has to be stored in the state variable.
  - Creates problems with sparsity.



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?

Creating a Unified Model

- What we want: an HMM with entity-grid features.
  - We need a quick estimator for transition probabilities in the entity grid.
  - In the past, entity grids have worked better as conditional models...

Overview

- Previous Work: Entity Grids
- Previous Work: Hidden Markov Model
- **Relaxed Entity Grid**
- Unified Hidden Markov Model
- Corpus and Experiments
- Conclusions and Future Work

Relaxing the Entity Grid

- The most common transition is from – to –.
  - The maximum likelihood document has no entities at all!
- Entities don't occur independently.
  - There may not be room for them all.
  - They 'compete' with one another.

Relaxed Entity Grid

- Assume we have already generated the set of roles we need to fill with known entities.
  - New entities come from somewhere else.

The commercial **pilot**, sole **occupant** of the **airplane**, was not injured.

The ?

new noun: owner

was owned and operated by a private ?

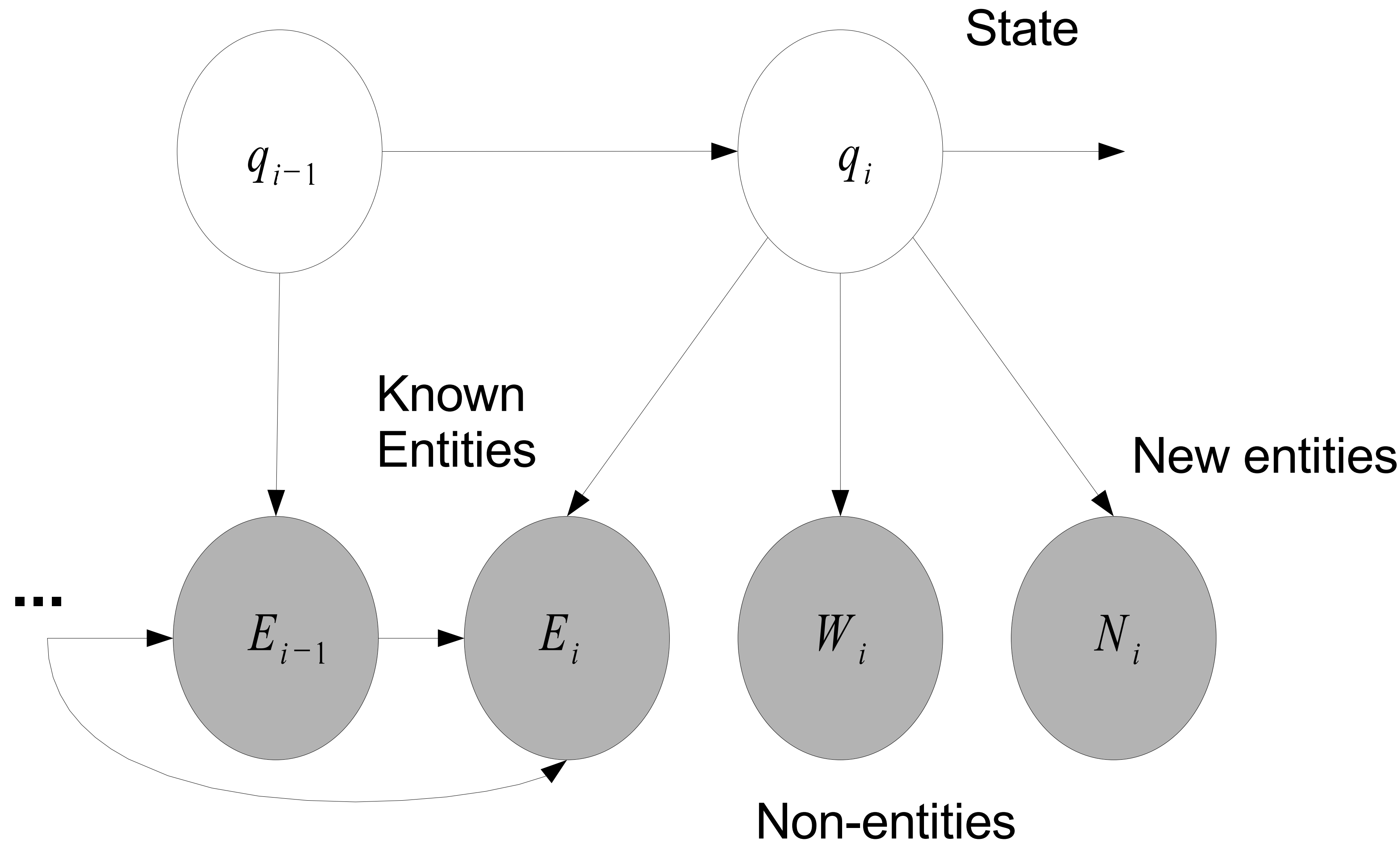
Filling Roles with Known Entities

- P(entity *e* fills role *j* | *j*, histories of known entities)
  - history: roles in previous sentences
  - known entity: has occurred before in document
- Still hard to estimate because of sparsity.
  - Too many combinations of histories.
- Normalize:
  P(entity *e* fills role *j* | *j*, history of entity *e*)
- Much easier to estimate!

Overview

- Previous Work: Entity Grids
- Previous Work: Hidden Markov Model
- Relaxed Entity Grid
- **Unified Hidden Markov Model**
- Corpus and Experiments
- Conclusions and Future Work

Graphical Model



Hidden Markov Model

- Need to lexicalize the entity grid.
  - States describe common words, not simply transitions.
- Back off to the unlexicalized version.
- Also generate the other words of the sentence (unigram language models):
  - Words that aren't entities.
  - First occurrences of entities.

Learning the HMM

- We used Gibbs sampling to fit:
  - Transition probabilities.
  - Number of states.
- Number of states heavily dependent on the backoff constants.
- We aimed for about 40-50 states.
  - As in Barzilay and Lee.

Has This Been Done Before?

- Soricut and Marcu '06:
  - Mixture model with HMM, entity grid and word-to-word (IBM) components.
  - Results are as good as ours.
- Didn't do joint learning, just fit mixture weights.
  - Less explanatory power.
- Uses more information (ngrams and IBM).
  - Might be improved by adding our model.

Overview

- Previous Work: Entity Grids
- Previous Work: Hidden Markov Model
- Relaxed Entity Grid
- Unified Hidden Markov Model
- **Corpus and Experiments**
- Conclusions and Future Work

Airplane (NTSB) Corpus

- Traditional for this task.
  - 100 test, 100 train.
- Short (avg. 11.5 sents) press releases on airplane emergencies.
- A bit artificial:
  - 40% begin: “This is preliminary information, subject to change, and may contain errors. Any errors in this report will be corrected when the final report has been completed.”

Discriminative Task

- 20 random permutations per document: 2000 tests.

Sentence 2
Sentence 1
Sentence 4
Sentence 3

vs

Sentence 1
Sentence 2
Sentence 3
Sentence 4

- Binary judgement between random permutation and original document.
- Local models do well.

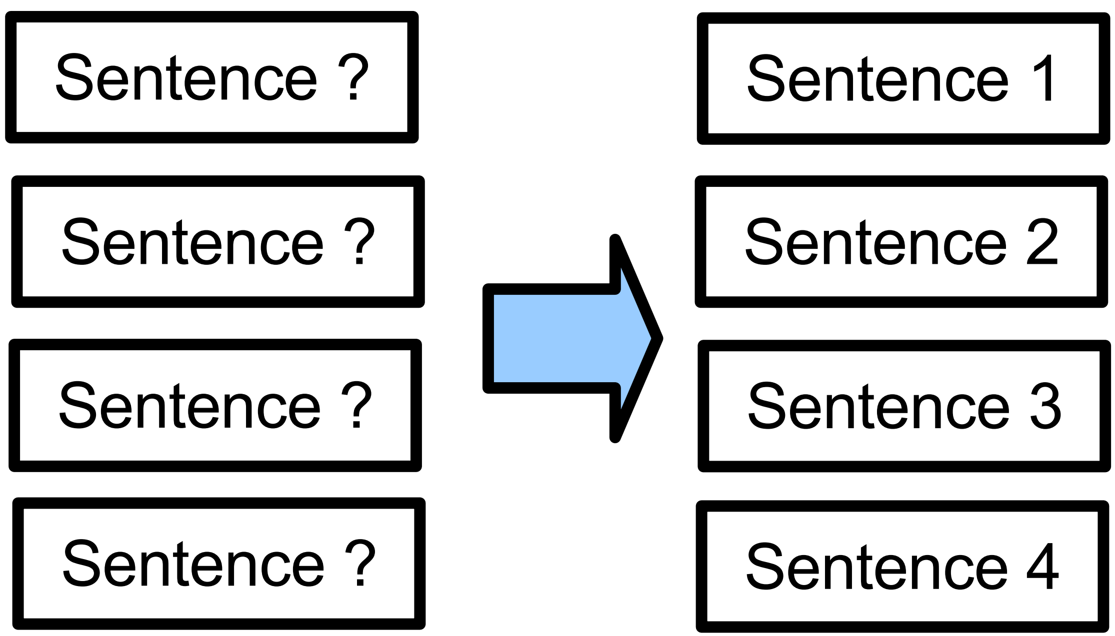
Results

Airplane Test	Discriminative (%)
Barzilay and Lapata (SVM EGrid)	90
Barzilay and Lee (HMM)	74
Soricut and Marcu (Mixture)	-
Unified (Relaxed EGrid/HMM)	**94**

Ordering Task

- Used simulated annealing to find optimal orderings.
- Score: similarity to original ordering.

$\tau = 1$



Kendall's τ metric:
-1 (worst) to 1 (best).
~ # of pairwise swaps.

Results

Airplane Test	Kendall's τ
Barzilay and Lapata (SVM EGrid)	-
Barzilay and Lee (HMM)	0.44
Soricut and Marcu (Mixture)	**0.50**
Unified (Relaxed EGrid/HMM)	**0.50**

Relaxed Entity Grid

Airplane Development	τ	Discr. (%)
Generative EGrid	0.17	81
Relaxed EGrid	0.02	87
Unified (Generative EGrid/HMM)	0.39	85
Unified (Relaxed EGrid/HMM)	**0.54**	**96**

Overview

- Previous Work: Entity Grids
- Previous Work: Hidden Markov Model
- Relaxed Entity Grid
- Unified Hidden Markov Model
- Corpus and Experiments
- **Conclusions and Future Work**

What We Did

- Explained strengths of local and global models.
- Proposed a new generative entity grid model.
- Built a unified model with joint local and global features.
  - Improves on purely local or global approaches.
  - Comparable to state-of-the-art.

What To Do Next

- Escape from the airplane corpus!
  - Too constrained and artificial.
  - Real documents have more complex syntax and lexical choices.
- Longer documents pose challenges:
  - Current algorithms aren't scalable.
  - Neither are evaluation metrics.

Acknowledgements

Couldn't have done it without:

- Regina Barzilay (code, data, advice & support)
- Mirella Lapata (code, advice)
- BLLIP (comments & criticism)
- Tom Griffiths & Sharon Goldwater (Bayes)
- DARPA GALE ($$)
- Karen T. Romer Foundation ($$)