Natural Language Understanding

Lecture 16: Entity-based Coherence

Mirella Lapata

School of Informatics
University of Edinburgh
mlap@inf.ed.ac.uk

March 28, 2017

Reading: Barzilay and Lapata (2008).

Coherence in Text

Coherence:

- is a property of well-written texts;
- makes them easier to read and understand;
- ensures that sentences are meaningfully related;
- and that the reader can work out what expressions mean;
- the text is thematically organized;
- temporally organized;
- rather than a random concatenation of sentences.

In this lecture, we will discuss Barzilay and Lapata's (2008) entity-based model of coherence.

Coherence in Text

Summary A

Britain said he did not have diplomatic immunity. The Spanish authorities contend that Pinochet may have committed crimes against Spanish citizens in Chile. Baltasar Garzon filed a request on Wednesday. Chile said, President Fidel Castro said Sunday he disagreed with the arrest in London.

Summary B

Former Chilean dictator Augusto Pinochet, was arrested in London on 14 October 1998. Pinochet, 82, was recovering from surgery. The arrest was in response to an extradition warrant served by a Spanish judge. Pinochet was charged with murdering thousands, including many Spaniards. Pinochet is awaiting a hearing, his fate in the balance. American scholars applauded the arrest.

Entity-based Coherence

- The way entities are introduced and discussed influences coherence (Grosz et al., 1995).
- *Entities* in an *utterance* are ranked according to salience.
  - Is an entity pronominalized or not?
  - Is an entity in a prominent syntactic position?
- Each utterance has one *center* ($\approx$ topic or focus).
- Coherent discourses have utterances with common centers.
- Entity transitions capture degrees of coherence (e.g., in Centering theory CONTINUE $>$ SHIFT).

Notions of salience, utterance, ranking are left unspecified.

Entity-based Local Coherence

John went to his favorite music store to buy a piano.

He had frequented the store for many years.

He was excited that he could finally buy a piano.

He arrived just as the store was closing for the day.

John went to his favorite music store to buy a piano.

It was a store John had frequented for many years.

He was excited that he could finally buy a piano.

It was closing just as John arrived.

Entity-based Local Coherence

John went to his favorite music store to buy a piano.

He had frequented the store for many years.

He was excited that he could finally buy a piano.

He arrived just as the store was closing for the day.

John went to his favorite music store to buy a piano.

It was a store John had frequented for many years.

He was excited that he could finally buy a piano.

It was closing just as John arrived.

The Entity Grid

Can we compute a discourse representation automatically?

- Does it capture coherence characteristics?
- What linguistic information matters for coherence?
- Is it robust across domains and genres?

What is an appropriate coherence model?

- View coherence rating as a machine learning problem.
- Learn a ranking function without manual involvement.
- Apply to text-to-text generation tasks.

Inspired from entity-based theories, not a direct implementation of any theory in particular.

The Entity Grid

1 Former Chilean dictator Augusto Pinochet, was arrested in London on 14 October 1998.
2 Pinochet, 82, was recovering from surgery.
3 The arrest was in response to an extradition warrant served by a Spanish judge.
4 Pinochet was charged with murdering thousands, including many Spaniards.
5 He is awaiting a hearing, his fate in the balance.
6 American scholars applauded the arrest.

The Entity Grid

1 [Former Chilean dictator Augusto Pinochet], was arrested in [London] on [14 October] 1998.

2 [Pinochet], 82, was recovering from [surgery].

3 [The arrest] was in [response] to [an extradition warrant] served by [a Spanish judge].

4 [Pinochet] was charged with murdering [thousands], including [many Spaniards].

5 [He] is awaiting [a hearing], [Pinochet's fate] in [the balance].

6 [American scholars] applauded the [arrest].

The Entity Grid

1 [Former Chilean dictator Augusto Pinochet]$_S$, was arrested in [London]$_X$ on [14 October]$_X$ 1998.

2 [Pinochet]$_S$, 82, was recovering from [surgery]$_X$.

3 [The arrest]$_S$ was in [response]$_X$ to [an extradition warrant]$_X$ served by [a Spanish judge]$_S$.

4 [Pinochet]$_O$ was charged with murdering [thousands]$_O$, including many [Spaniards]$_O$.

5 [Pinochet]$_S$ is awaiting [a hearing]$_O$, [his fate]$_X$ in [the balance]$_X$.

6 [American scholars]$_S$ applauded the [arrest]$_O$.

The Entity Grid

1	Pinochet$_S$	London$_X$	October$_X$	
2	Pinochet$_S$	surgery$_X$		
3	arrest$_S$	response $_X$	warrant$_X$	judge$_O$
4	Pinochet$_O$	thousands$_O$	Spaniards$_O$	
5	Pinochet$_S$	hearing$_O$	Pinochet$_X$	fate$_X$	balance$_X$
6	scholars$_S$	arrest$_O$		

The Entity Grid

	Pinochet	London	October	Surgery	Arrest	Warrant	Judge	Thousands	Spaniards	Hearing	Fate	Balance	Scholars
1													
2													
3													
4													
5													
6													

The Entity Grid

	Pinochet	London	October	Surgery	Arrest	Extradition	Warrant	Judge	Thousands	Spaniards	Hearing	Fate	Balance	Scholars
1	**S**													
2	**S**	–												
3	–													
4	**O**													
5	**S**													
6	–													

The Entity Grid

	Pinochet	London	October	Surgery	Arrest	Extradition	Warrant	Judge	Thousands	Spaniards	Hearing	Fate	Balance	Scholars
1	S	X	X	–	–	–	–	–	–	–	–	–	–	–
2	S	–	–	X	–	–	–	–	–	–	–	–	–	–
3	–	–	–	–	S	X	X	O	–	–	–	–	–	–
4	O	–	–	–	–	–	–	–	O	O	–	–	–	–
5	S	–	–	–	–	–	–	–	–	–	O	X	X	–
6	–	–	–	–	O	–	–	–	–	–	–	–	–	S

The Entity Grid

S	X	X	–	–	–	–	–	–	–	–	–	–	–
S	–	–	X	–	–	–	–	–	–	–	–	–	–
–	–	–	–	S	X	X	O	–	–	–	–	–	–
O	–	–	–	–	–	–	–	O	O	–	–	–	–
S	–	–	–	–	–	–	–	–	–	O	X	X	–
–	–	–	–	O	–	–	–	–	–	–	–	–	S

The Entity Grid

S	X	X	–	–	–	–	–	–	–	–	–	–	–
S	–	–	X	–	–	–	–	–	–	–	–	–	–
–	–	–	–	S	X	X	O	–	–	–	–	–	–
O	–	–	–	–	–	–	–	O	O	–	–	–	–
S	–	–	–	–	–	–	–	–	–	O	X	X	–
–	–	–	–	O	–	–	–	–	–	–	–	–	S

S	S	–	X	X	–	–	–	–	–	–	–	–	–	–
–	–	X	–	–	X	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	X	X	O	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	O	O	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–	O	X	X
–	–	–	–	–	–	O	–	–	–	–	–	–	–	–

The Entity Grid

S	X	X	–	–	–	–	–	–	–	–	–	–	–	
S	–	–	X	–	–	–	–	–	–	–	–	–	–	
–	–	–	–	S	X	X	O	–	–	–	–	–	–	
O	–	–	–	–	–	–	–	O	O	–	–	–	–	
S	–	–	–	–	–	–	–	–	–	O	X	X	–	
–	–	–	–	O	–	–	–	–	–	–	–	–	S	

S	S	–	X	X	–	–	–	–	–	–	–	–	–	–
–	–	X	–	–	X	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	X	X	O	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	O	O	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–	O	X	X
–	–	–	–	–	–	O	–	–	–	–	–	–	–	–

Entity Transitions

Definition

A local entity transition is a sequence $\{\mathbf{S}, \mathbf{O}, \mathbf{X}, -\}^n$ that represents entity occurrences and their syntactic roles in n adjacent sentences.

Feature Vector Notation

Each grid x_{ij} for document d_i is represented by a feature vector:

$$\Phi(x_{ij}) = (p_1(x_{ij}), p_2(x_{ij}), \ldots, p_m(x_{ij}))$$

m is the number of predefined entity transitions
$p_t(x_{ij})$ the probability of transition t in grid x_{ij}

Entity Transitions

Example (transitions of length 2)

	SS	SO	SX	S–	OS	OO	OX	O–	XS	XO	XX	X–	–S	–O	–X	––
d_1	0	0	0	.03	0	0	0	.02	.07	0	0	.12	.02	.02	.05	.25
d_2	0	0	0	.02	0	.07	0	.02	0	0	.06	.04	0	0	0	.36
d_3	.02	0	0	.03	0	0	0	.06	0	0	0	.05	.03	.07	.07	.29

Entity Transitions

Example (transitions of length 2)

	SS	SO	SX	S–	OS	OO	OX	O–	XS	XO	XX	X–	–S	–O	–X	––
d_1	0	0	0	.03	0	0	0	.02	.07	0	0	.12	.02	.02	.05	.25
d_2	0	0	0	.02	0	.07	0	.02	0	0	.06	.04	0	0	0	.36
d_3	.02	0	0	.03	0	0	0	.06	0	0	0	.05	.03	.07	.07	.29

Linguistic Dimensions

Salience: Are some entities more important than others?

- Discriminate between salient (frequent) entities and the rest.
- Collect statistics separately for each group.

Coreference: What is its contribution?

- Entities are coreferent if they have the same surface form.
- Apply a coreference resolution system.

Syntax: Does syntactic knowledge matter?

- Use four categories $\{\mathbf{S}, \mathbf{O}, \mathbf{X}, -\}$.
- Reduce categories to $\{\mathbf{X}, -\}$.

Learning a Ranking Function

Training Set

Ordered pairs (x_{ij}, x_{ik}), where x_{ij} and x_{ik} represent the same document d_i, and x_{ij} is more coherent than x_{ik} (assume $j > k$).

Goal

Find a parameter vector $\vec{w}$ such that:

$$\vec{w} \cdot (\Phi(x_{ij}) - \Phi(x_{ik})) > 0 \ \forall j, i, k \text{ such that } j > k$$

Support Vector Machines

Constraint optimization problem can be solved using the search technique described in Joachims (2002).

Text Ordering

Motivation

- Determine a sequence in which to present a set of items.
- Essential step in generation applications.

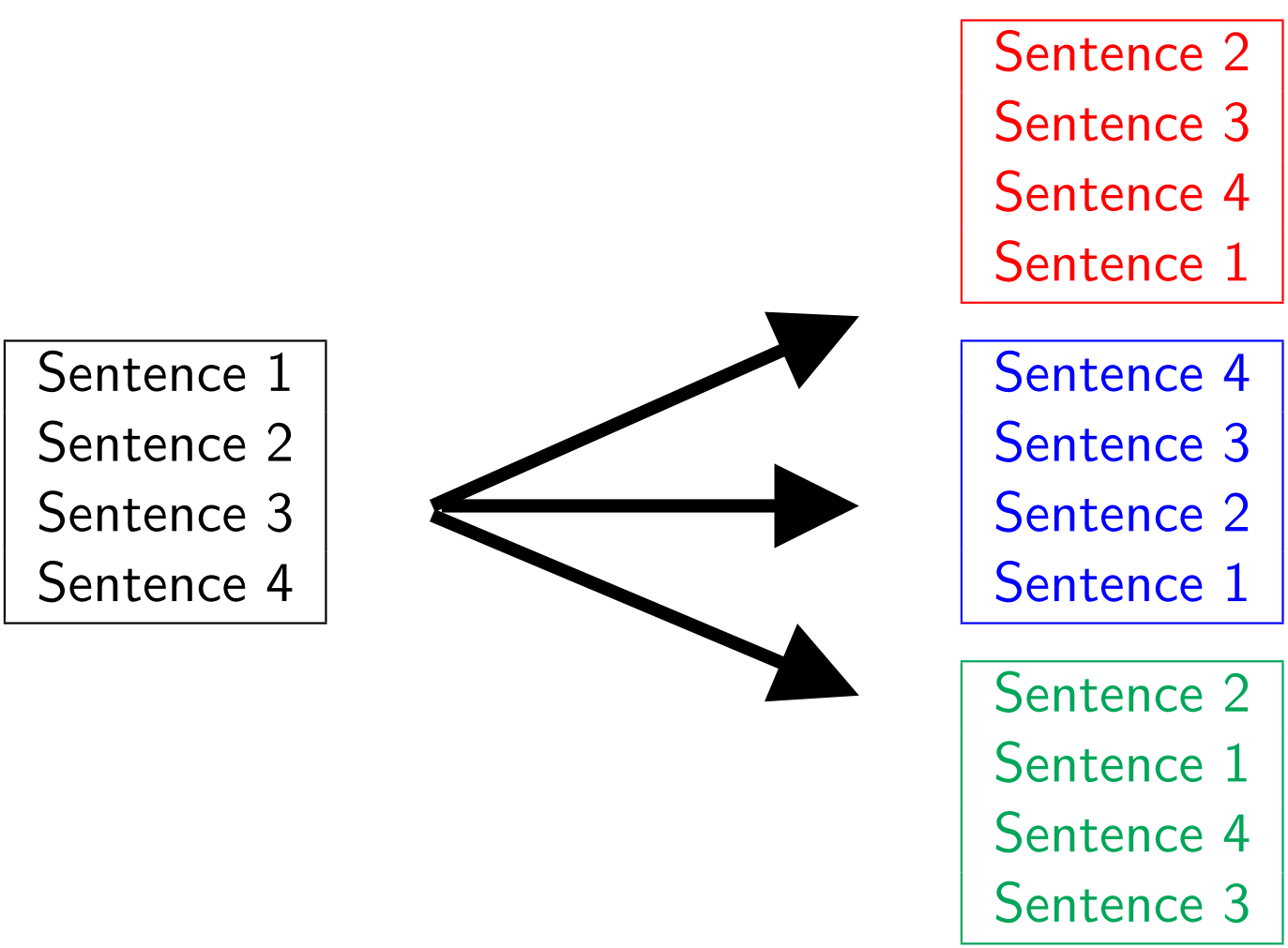
Data

- Source document and permutations of its sentences.
- Original order *assumed coherent.*
- Given k documents, with n permutations, obtain $k \cdot n$ pairwise rankings for training and testing.
- Two corpora, Earthquakes and Accidents, 100 texts each.

Text Ordering

Sentence 1
Sentence 2
Sentence 3
Sentence 4

Text Ordering

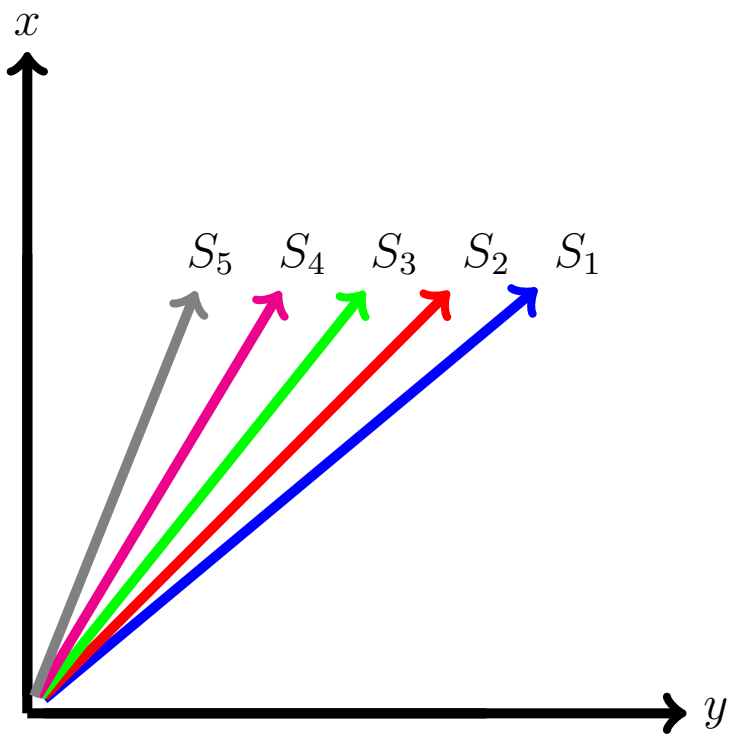


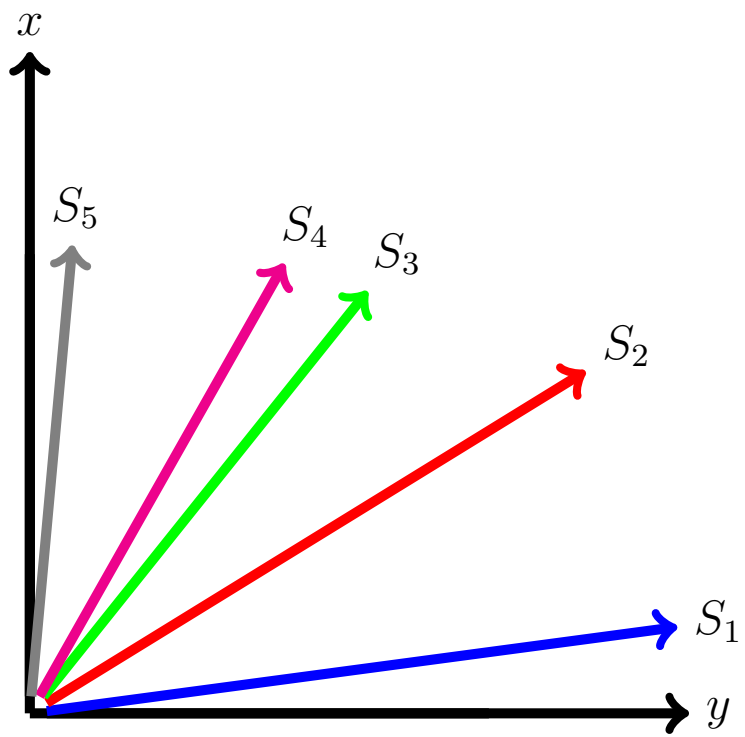
Comparison with State of the Art

Vector-based Model (LSA, Foltz et al., 1998):

- Meaning of individual words is represented in vector space.
- Sentence meaning is the mean of the vectors of its words.
- Average distance of adjacent sentences.
- *Unsupervised, local, lexicalized, domain independent.*

Comparison with State of the Art



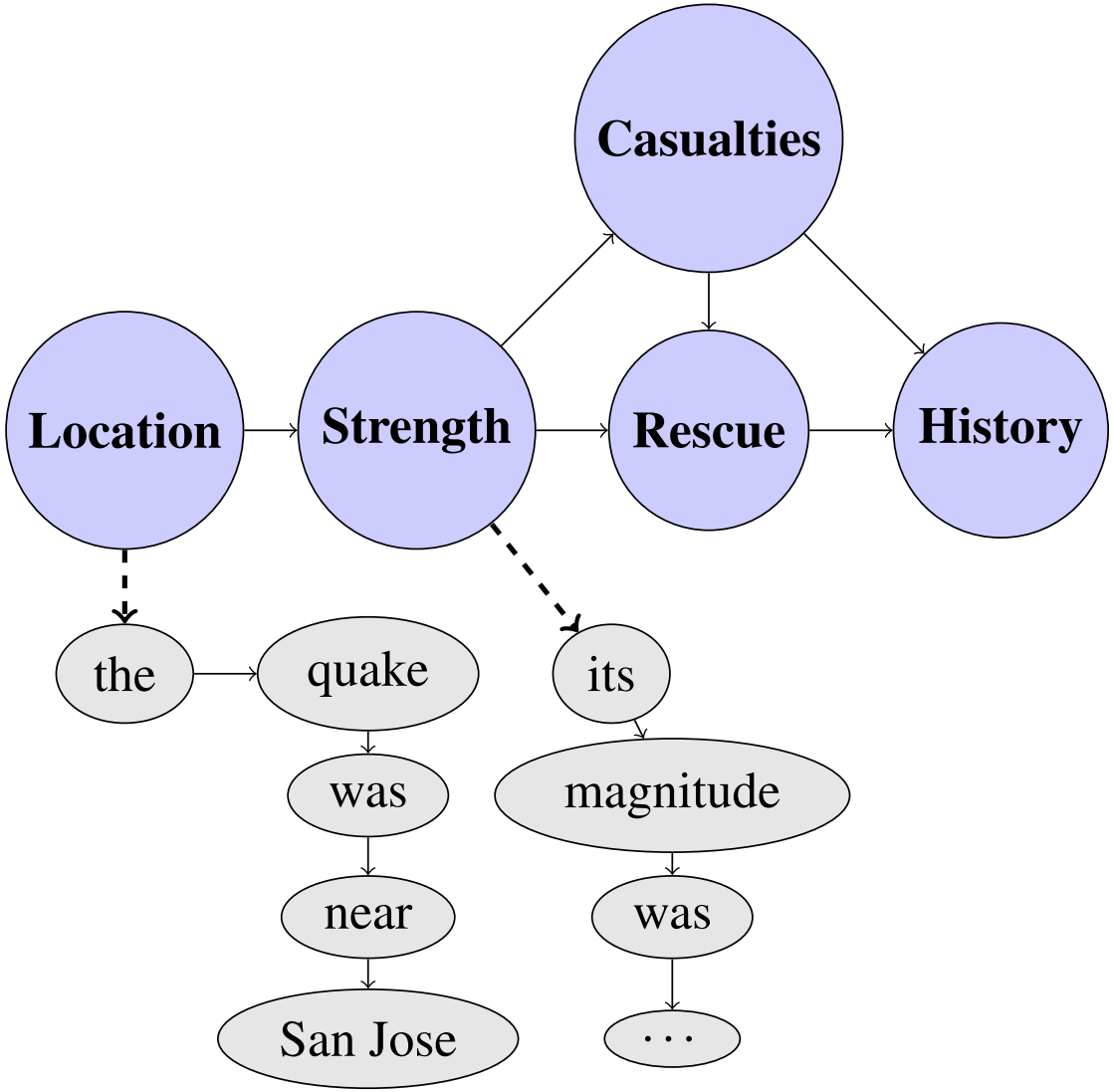


Comparison with State of the Art

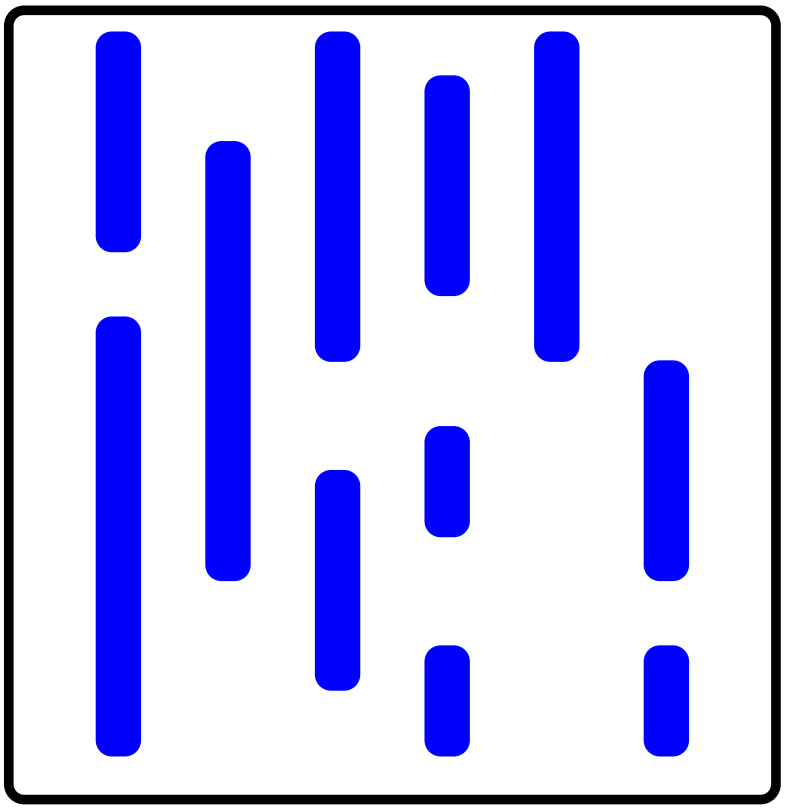
HMM-based Content Models (Barzilay and Lee, 2004):

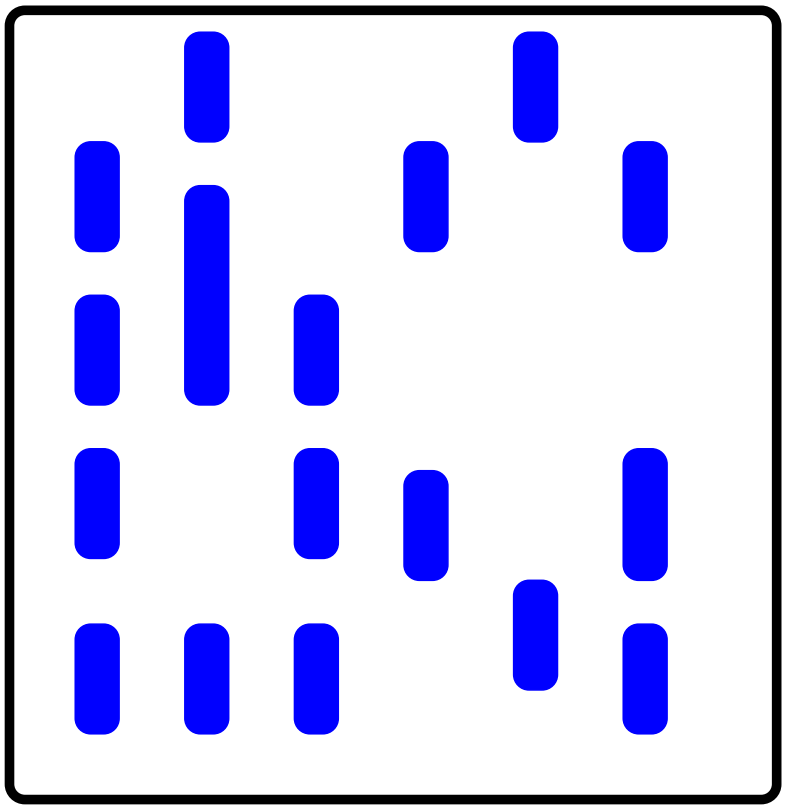
- Model topics and their order in texts.
- Model is an HMM: states correspond to topics ($\approx$ sentences).
- Model selects sentence order with highest probability.
- *Supervised, global, lexicalized, domain dependent.*

Comparison with State of the Art

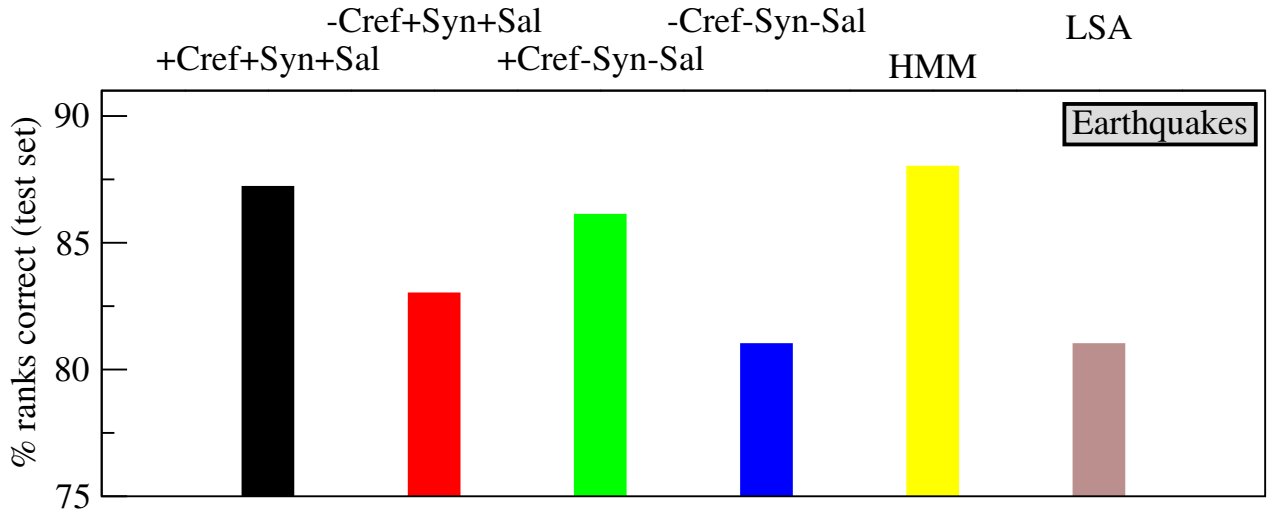


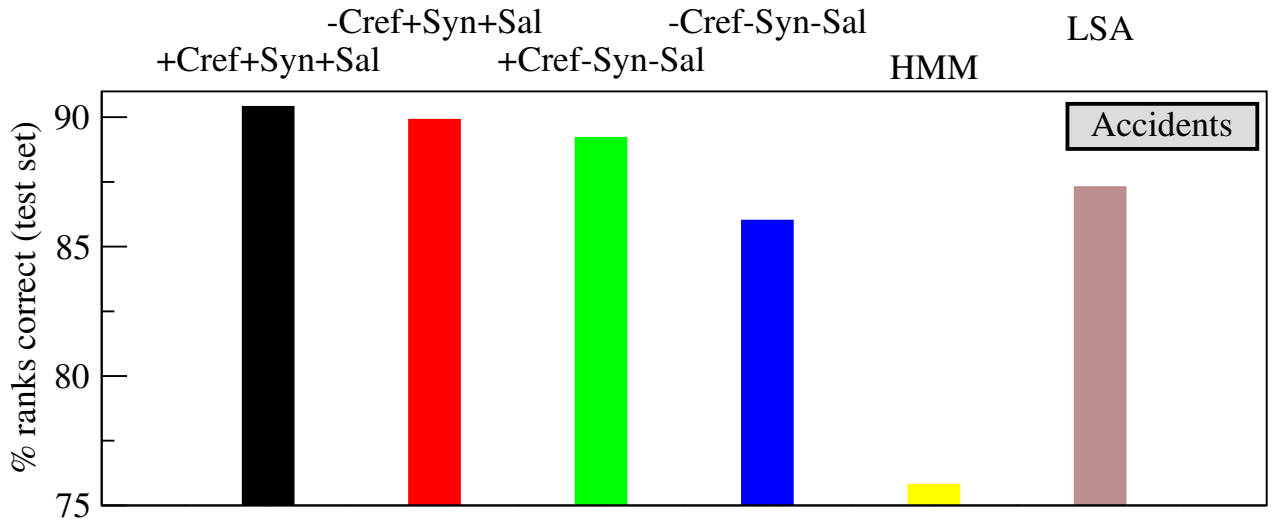
The Entity Grid





Results: Ordering





Discussion

- Omission of coreference causes performance drop.
- Syntax and Salience have more effect on Accidents corpus.
- Linguistically poor model generally worse.
- Entity model is better than LSA.
- HMM-based content models exhibit high variability.
- Models seem to be complementary.

Summarization

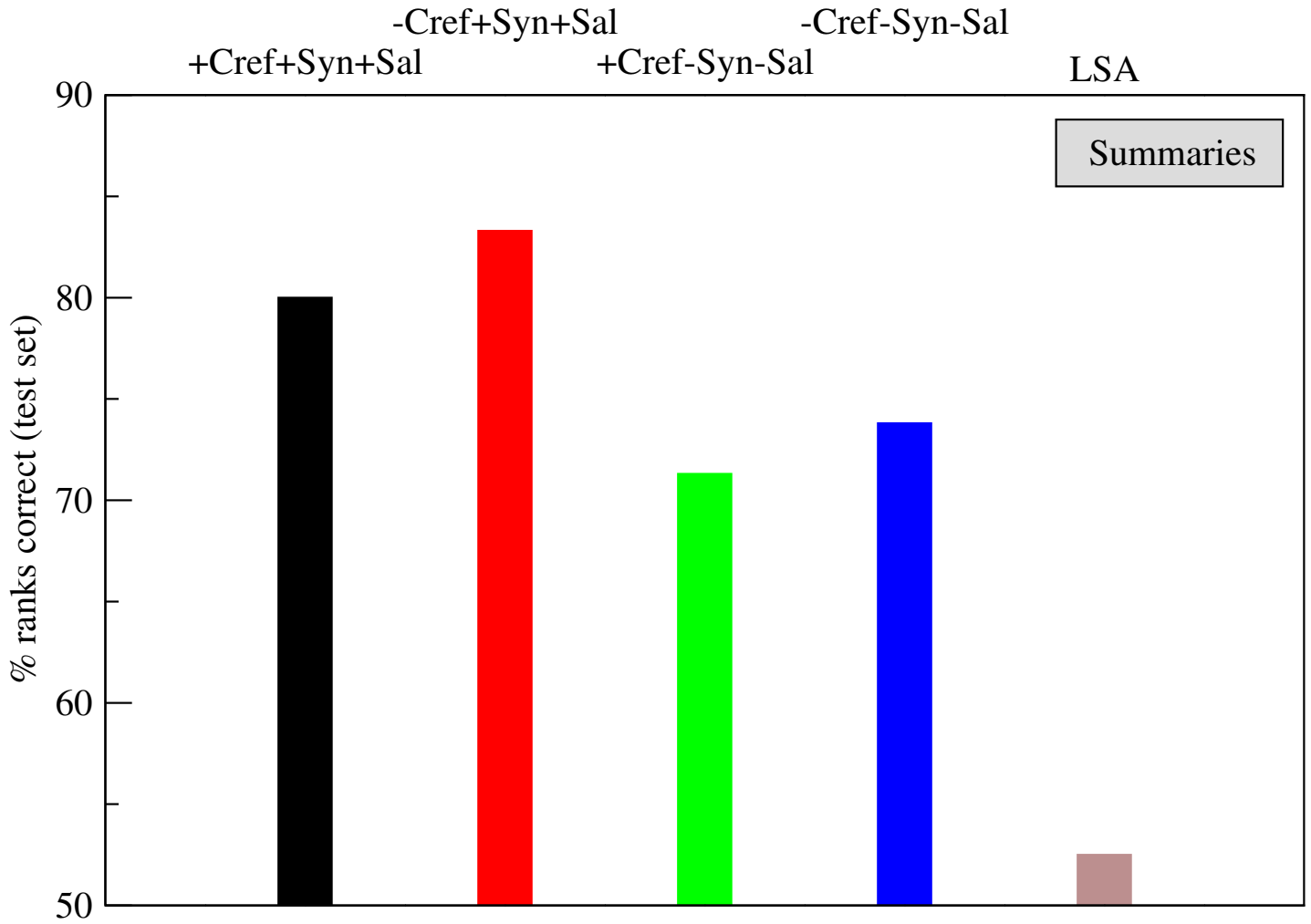
Motivation

- Summaries naturally exhibit coherence violations.
- Compare model against rankings elicited by human judges.
- Useful for automatic evaluation of machine generated text.

Data

- Outputs of 5 multi-document summarization systems and corresponding human authored summaries (DUC 2003).
- Participants assign readability score on a seven point scale.
- 144 summaries, 177 participants (23 per summary).

Results: Summarization



Results

- Coreference decreases accuracy (machine generated texts).
- Salience seems to have more of an impact here.
- Linguistically poor model is generally worse.
- Entity model performs better than LSA.
- LSA is unsupervised and exposed only to human texts.
- Training corpus is unsuitable for HMM-based content models.

Summary

Strengths:

- Novel framework for representing and measuring coherence.
- Entity grid and cross-sentential transitions.
- Suited for learning appropriate ranking function.
- Fully automatic and robust, useful for system development.

Weaknesses:

- Entity grid doesn't contain lexical information.
- Doesn't contain a notion of global coherence.
- Can't model multi-paragraph text.