Entity-based Models of Discourse Structure

Micha Elsner

School of Informatics
University of Edinburgh

March 16, 2012

Coherence

Structure of information in a discourse–
Gives readers **context** they need...
to understand new information (Halliday+Hasan '76)

Coherent text

Alice was sitting by her sister.
Suddenly a White Rabbit ran by her.
Alice heard the Rabbit say "I shall be late!"

Incoherent text

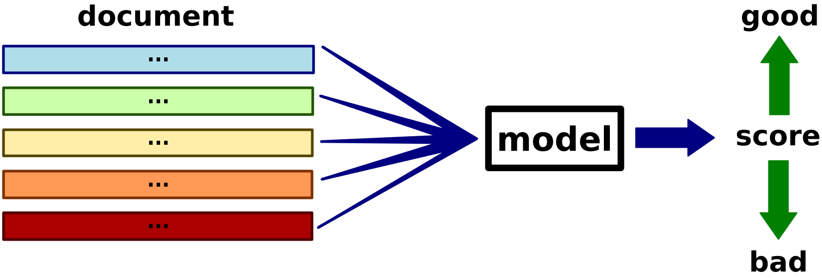
Alice heard the Rabbit say "I shall be late!"
The Mouse did not notice this question.
"It isn't", said the Caterpillar.

Why coherence?

Fundamental linguistic concept

Helps predict which utterances are pragmatically appropriate
Tells us which utterances are closely related

Computational work: essay scoring



- Predict human assessments of quality
- Mostly in education (Miltsakaki,Higgins,Burstein,others)

Multidocument summarization

- Sentence selection

Now the CDC is testing blood from patients who had symptoms.

The outbreak was the first time West Nile had been seen in this hemisphere.

Officials announced they had found a crow infected with West Nile.

...

...

Multidocument summarization

- Sentence selection
- “Patients who had symptoms”– of what?

Now the CDC is testing blood from patients who had symptoms.

The outbreak was the first time West Nile had been seen in this hemisphere.

Officials announced they had found a crow infected with West Nile.

...

...

Multidocument summarization

- Sentence selection
- “Patients who had symptoms”– of what?
- Reordering posed as search for most coherent order
  - Can also do some rewriting (Nenkova+McKeown ‘03)

1 Now the CDC is testing blood from patients who had symptoms.

2 The outbreak was the first time West Nile had been seen in this hemisphere.

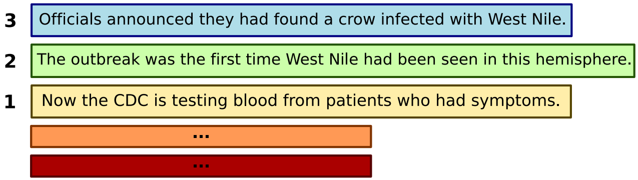
3 Officials announced they had found a crow infected with West Nile.

...

...

Multidocument summarization

- Sentence selection
- “Patients who had symptoms”– of what?
- Reordering posed as search for most coherent order
  - Can also do some rewriting (Nenkova+McKeown ‘03)



Modeling coherence

Coherence created by large-scale *(global)* sequence of topics:

- HMMs, trees (Eisenstein), Mallows model (Chen+al '09)

and *local* surface properties of text:

- Lexical: consistent topic and vocabulary
  - Tf-idf, topic models, IBM model 1
- Rhetorical: related propositions
  - Temporal (Bollegala+al '09)
  - Discourse parsing (rare) (Lin+al '11)
- Entity-based: focus on set of objects
  - This talk
  - Also (Karamanis) and others

Intuitions

A text is about **entities**: things in the world

Suddenly **a White Rabbit** ran by **her**.
Alice heard **the Rabbit** say “I shall be late!”
The Rabbit took a watch out of its pocket.
Alice started to her feet.

Coherence created by repeated entity mentions

More specific theories, eg Centering (Grosz+Sidner)

Overview

Baseline: the entity grid

Barzilay and Lapata CL ‘08

Referring expressions: form and content

Elsner and Charniak ACL ‘08

Different types of entity behave differently

Elsner and Charniak ACL ‘11, ACL ‘10

New evaluation: phone dialogues

Elsner and Charniak ACL ‘11b

Implications

High-level structure: plot

Elsner EACL ‘12

Representing a text

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**

Representing a text

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**

Grid

White Rabbit	subj	obj	subj	–

Representing a text

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**

Grid

White Rabbit	subj	obj	subj	–
Alice	other	subj	–	subj

Representing a text

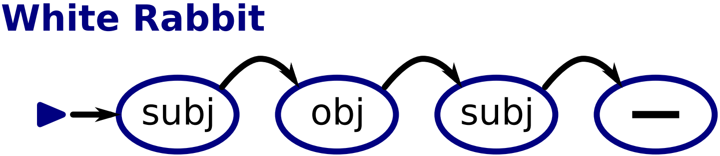
Text	Syntactic role
Suddenly **a White Rabbit** ran by her.	**subject**
Alice heard **the Rabbit** say “I shall be late!”	**object**
The Rabbit took a watch out of its pocket.	**subject**
Alice started to her feet.	**missing**

Grid

White Rabbit	subj	obj	subj	–
Alice	other	subj	–	subj
watch	–	–	obj	–
feet	–	–	–	other

Modeling (simplified)

Entities treated independently...
Modeled via Markov chain:



Generative and discriminative grids
both use these features

Just what is an entity?

Coreference?

We don't use it!

- Only sometimes improves results (Barzilay+Lapata '05)...
- Input documents must be fairly coherent
- Instead: link mentions with same head noun

Mention detection?

Use all nouns as mentions.

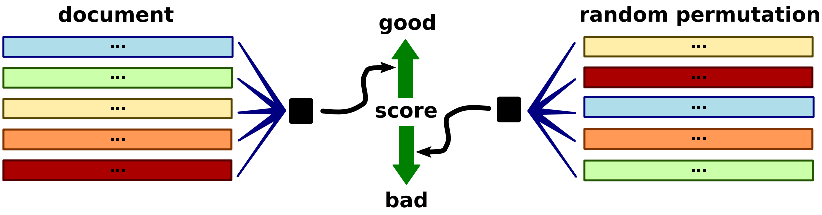
- Pick up premodifiers like "a **Bush** spokesman"
- Maximize coreference recall
- Improves over NPs as mentions by 4%

Standard ordering benchmarks

Discrimination

(Barzilay+Lapata '05) following (Karamanis+al '04)

- Proxy for summarization reordering
- No human judgements required
- Too easy on long documents

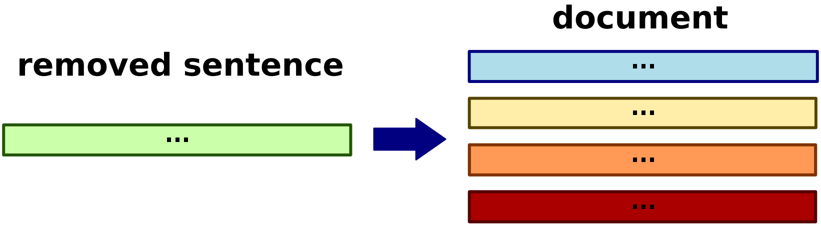


Standard ordering benchmarks

Insertion

(Chen+al '07) and (Elsner+Charniak '07)

- Proxy for updating an article
- Also no human judgements
- Harder for longer documents



Results on WSJ test

1004 documents in test set

	Discrimination	Insertion
Random	50	13
Grid	**80**	**21**

Overview

Baseline: the entity grid
Barzilay and Lapata CL '08

Referring expressions: form and content
Elsner and Charniak ACL '08

Different types of entity behave differently
Elsner and Charniak ACL '11, ACL '10

New evaluation: phone dialogues
Elsner and Charniak ACL '11b

Implications

High-level structure: plot
Elsner EACL '12

Referring expressions

- The grid sees entities via syntactic roles
- Doesn’t care what the mention looks like

Expressions referring to entities constrained by **information structure** (Prince ‘81 and subsq.)

- New information has to be grounded
- Old information gets shorter mentions
- Very salient entities get pronouns or demonstratives

Old vs new NPs

New information needs complex packaging

"Secretary of State Hillary Clinton"

Old information doesn't

"Clinton"

Relatively easy to build an old/new classifier:

(Poesio+al '05 and others)

- Linear classifier with syntactic features
- Trained on coref corpus or using same-head coref
- Accuracies usually 80-90% with document order
  - $\sim$ 60% without

Ordering model based on same-head coref

Pronouns

Pronouns refer to very salient entities
Possible to find references automatically with $\sim$ 70% accuracy

- Number and gender constraints
- Syntactic tree distance within sentence (Hobbs)
- Nearly all antecedents within 2 previous sentences
- Have to handle pleonastic/expletive pronouns (Bergsma '11)
  - "It appears that..."
- Unsupervised algorithm (Charniak+Elsner '09)

Using pronouns for coherence

First idea:

- Resolve pronouns, add to grid (Barzilay+Lapata '05)
  - Doesn't work
  - Coreference too inaccurate on disordered documents
  - Pronouns can usually find *some* potential referent

But these referents are often poor!

- New idea: use $p(pronoun|antecedent)$!
- (Requires generative model)

Pronouns

Detect passages with stranded pronouns:

Good

Marlow sat cross-legged.

He had sunken cheeks.

Bad

The day was ending.

?

He had sunken cheeks.

Results

	Discrimination	Insertion
Random	50	13
Grid	80	21
Disc-new	**70**	**16**
Pronouns	**65**	**16**

- Both models reasonably good
- Combined results improve over grid
  - (later in talk)

Overview

Baseline: the entity grid
Barzilay and Lapata CL ‘08

Referring expressions: form and content
Elsner and Charniak ACL ‘08

Different types of entity behave differently
Elsner and Charniak ACL ‘11, ACL ‘10

New evaluation: phone dialogues
Elsner and Charniak ACL ‘11b

Implications

High-level structure: plot
Elsner EACL ‘12

Entity types

We've seen how different referring expressions belong earlier or later...

Can also use the referring expressions to learn about the entity itself:

- Semantics affect likelihood of importance
- News focuses on people and companies...
- Not so much on dates
- Many references don't act like names
    - "Half of them", "members of the union", "my problem with that" (Elsner+Charniak '10)

Modeling redux

Features from previous work

White Rabbit = subj	obj of previous
	subj of prev-1
	occurs 3x total

These features aren't enough...

White Rabbit vs **watch**

Features of important entities

White Rabbit = subj	obj of previous
	subj of prev-1
	occurs 3x total
	is a proper NP
	is named entity class PERSON
	has some modifiers
	is singular

Features separate **White Rabbit** from **watch**

Similar features useful in coref/summary tasks

Coreference features

Spurious entities

Formed around nouns like “care”, “increase”, “percent” Don’t throw away, but should distinguish

an increase = subj	obj of previous ... **in MUC6, but never coreferent** **rarely has coreferent pronouns**

- Automatic pronoun coreference on large dataset

What we learn

Baseline

P(**May 25/President Bush** = subj | missing in previous
other in prev-1
occurs 3x total)

= .045

What we learn

Baseline

P(**May 25/President Bush** = subj | missing in previous
other in prev-1
occurs 3x total)

= .045

Our model

P(**May 25** = subj | missing in previous
...
NE type DATE
never corefers in MUC6)

= .001

What we learn

Baseline

P(**May 25/President Bush** = subj | missing in previous
other in prev-1
occurs 3x total)

= .045

Our model

P(**President Bush** = subj | missing in previous
...
NE type PERSON
proper NP
corefers in MUC6
modifiers)

= .133

Results on WSJ test

Alone

	Disc.	Ins.
Random	50	13
Grid	80	21
Extended Grid	**84**	**24**

Combined

Grid+other models	83	24
ExtEGrid+other models	**86**	**27**

Overview

Baseline: the entity grid
  Barzilay and Lapata CL ‘08

Referring expressions: form and content
  Elsner and Charniak ACL ‘08

Different types of entity behave differently
  Elsner and Charniak ACL ‘11, ACL ‘10

New evaluation: phone dialogues
  Elsner and Charniak ACL ‘11b

Implications

High-level structure: plot
  Elsner EACL ‘12

Are we really measuring coherence?

Experiments on simplified summarization tasks

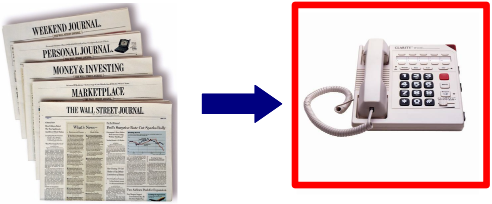
- Incoherent documents aren't realistic
- Standard corpora are newswire and short reports

Essay grading

- Annotations subjective and somewhat unreliable
- Data is expensive (and mostly proprietary)
- Coherence is just one factor in quality

Important to try other tasks and domains!

From news...
to phone conversations!

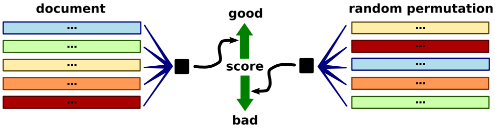


Phone dialogues on selected topics

- Manually transcribed/parsed
- Switchboard corpus

Results in this section include a lexical model, IBM-1 (Soricut+Marcu '06) and a lexical-entity model, Topical Entity Grid (TGrid) (Elsner+Charniak '11b)

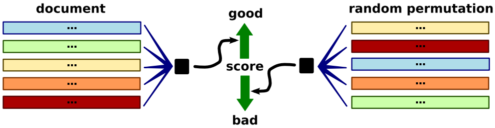
Discrimination on dialogue



Chance	50
Grid	86
TGrid	71
IBM-1	85
Pronouns	72
New info	55
Combined	88

Main effect of document length (uninteresting)

Discrimination on dialogue



Chance	50
Grid	86
TGrid	71
IBM-1	85
Pronouns	72
New info	**55**
Combined	88

Main effect of document length (uninteresting)

New-information model restricted to news

Disentangling conversations

Many simultaneous conversations

- Crowded rooms
- Push-to-talk radio
- Internet chat

Chanel:	How do I limit my internet connection speed?
Felicia:	Use the keyword “throttling” in google.
Chanel:	Felicia, google solved my problem.
Gale:	You guys have never worked in a factory, have you?
Gale:	There’s some real unethical stuff that goes on.
Arlie:	Of course, that’s how they make money.
Chanel:	You deserve a trophy!
Gale:	People lose limbs, or get killed.
Felicia:	Excellent!

How do participants cope?

Individual conversations must be coherent...

Participants know the structure is this:

Chanel:		How do I limit my internet connection speed?
Felicia:		Use the keyword “throttling” in google.
Chanel:		Felicia, google solved my problem.
	Gale:	You guys have never worked in a factory...
	Gale:	There’s some real unethical stuff that goes on.
	Arlie:	Of course, that’s how they make money.
Chanel:		You deserve a trophy!
	Gale:	People lose limbs, or get killed.
Felicia:		Excellent!

How do participants cope?

Individual conversations must be coherent...

Not this:

Chanel:		How do I limit my internet connection speed?
Felicia:		Use the keyword “throttling” in google.
Chanel:		Felicia, google solved my problem.
	Gale:	You guys have never worked in a factory...
	Gale:	There’s some real unethical stuff that goes on.
	Arlie:	Of course, that’s how they make money.
Chanel:		You deserve a trophy!
	Gale:	People lose limbs, or get killed.
	Felicia:	Excellent!

How do participants cope?

Individual conversations must be coherent...

Especially not this!

Chanel:		How do I limit my internet connection speed?
Felicia:		Use the keyword “throttling” in google.
	Chanel:	Felicia, google solved my problem.
	Gale:	You guys have never worked in a factory...
Gale:		There’s some real unethical stuff that goes on.
	Arlie:	Of course, that’s how they make money.
Chanel:		You deserve a trophy!
Gale:		People lose limbs, or get killed.
Felicia:		Excellent!

Synthetic 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.

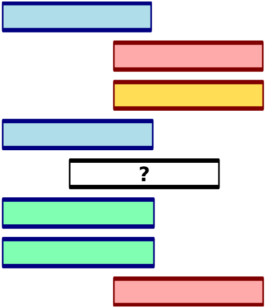
It's just a computer voice on the line.

IThose are the ones I really, really hate.

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

Assigning a single utterance

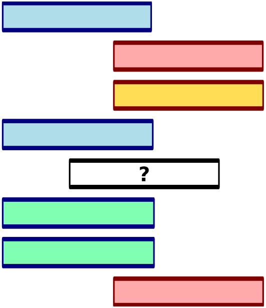
Chance	50
Grid	77
TGrid	78
IBM-1	69
Pronouns	52
Time	58
Combined	**83**



- Coherence approach is effective

Assigning a single utterance

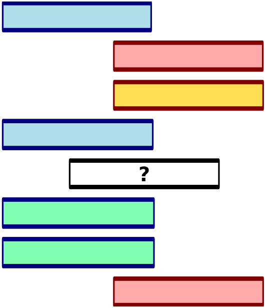
Chance	50
Grid	77
TGrid	78
IBM-1	69
Pronouns	**52**
Time	58
Combined	**83**



- Coherence approach is effective
- Pronouns very bad here

Assigning a single utterance

Chance	50
Grid	77
TGrid	78
IBM-1	69
Pronouns	52
Time	58
Combined	**83**

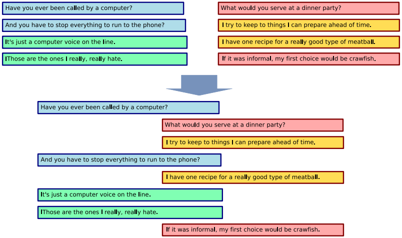


- Coherence approach is effective
- Pronouns very bad here

Best models: sensitive, many-sentence context

Recovering the original conversations

Chance	50
Grid	59
TGrid	60
IBM-1	56
Pronouns	54
Time	55
Combined	**64**



- Results predictable from single utterance

Overview

Baseline: the entity grid
Barzilay and Lapata CL ‘08

Referring expressions: form and content
Elsner and Charniak ACL ‘08

Different types of entity behave differently
Elsner and Charniak ACL ‘11, ACL ‘10

New evaluation: phone dialogues
Elsner and Charniak ACL ‘11b

Implications

High-level structure: plot
Elsner EACL ‘12

What we've seen

- Coherence is improved by appropriate patterns of **entity mentions**
- The syntax of a mention depends on its **information status**
- Important entities distinguished by **form**, **number** and **syntactic roles** of mentions

- Coherence models can be **automatically validated** via reordering tasks
- ...and **disentanglement**
- Models don't always work on **all domains**

Challenges

- Can we use relationships between entities?
  - Topical entity grid tries to do this, but...
- Better models for discourse-new NPs and pronouns on conversation
  - See (Rahman+Ng '11)
- An account of other coref. phenomena?
  - Bridging/mediated reference, demonstratives, etc.
- Integration of local and global models?
  - See (Elsner+Austerweil+Charniak '07) but doesn't scale

- Evaluate models against human judgements...
- Stylistic acceptability, not just coherence
  - How "journalistic"/"formal"/etc.?

Overview

Baseline: the entity grid
- Barzilay and Lapata CL '08

Referring expressions: form and content
- Elsner and Charniak ACL '08

Different types of entity behave differently
- Elsner and Charniak ACL '11, ACL '10

New evaluation: phone dialogues
- Elsner and Charniak ACL '11b

Implications

High-level structure: plot
- Elsner EACL '12

...follows the main character Elizabeth Bennet as she deals with issues of manners, upbringing, morality, education and marriage... (Wikipedia)

The story turns on the marriage prospects of the five daughters of Mr. and Mrs. Bennet... (Amazon.com)

...follows the main character Elizabeth Bennet as she deals with issues of manners, upbringing, morality, education and marriage... (Wikipedia)

The story turns on the marriage prospects of the five daughters of Mr. and Mrs. Bennet... (Amazon.com)

“Bingley.” Elizabeth felt Jane’s pleasure. “Miss Elizabeth Bennet.” Elizabeth looked surprised. “FITZWILLIAM DARCY” Elizabeth was delighted. Elizabeth read on: Elizabeth smiled. “If! “Dearest Jane! (Jason Huff: Microsoft Word ‘08)

First steps:

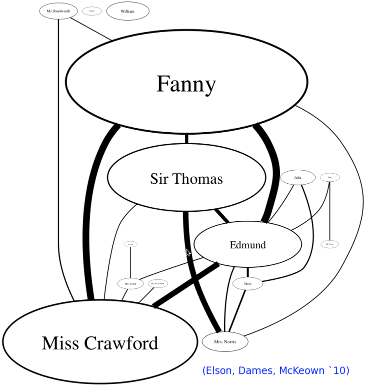
- We're not going to solve this all at once...

Similarity between novels

- Helpful for information retrieval:
  - Find another novel like "Pride and Prejudice".
- Clustering and organization:
  - Are there "plot type" clusters?
- Project knowledge about training novels to unknown:
  - This novel is *like* "Pride and Prejudice"; maybe it's a romance.

Plot is *high-level*...

Two basic insights:



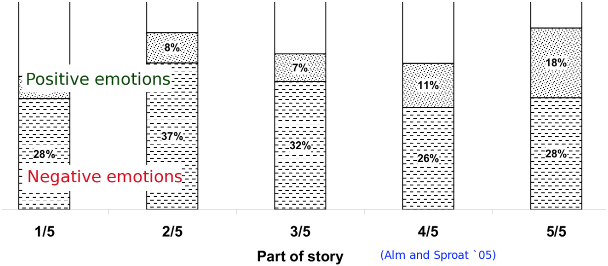
(Elson, Dames, McKeown `10)

Characters...
forming a *social network*

(Elson+al ‘10)

Plot is *high-level*...

Two basic insights:

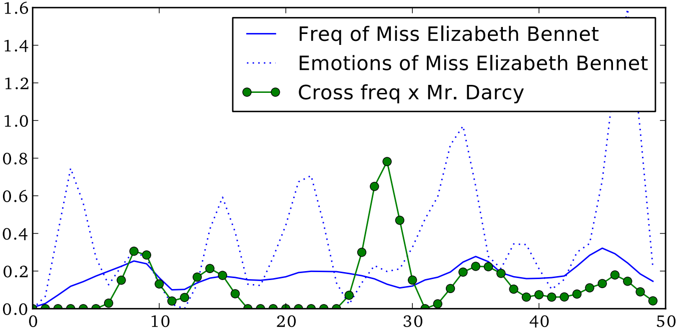


Story has an *emotional trajectory*

(Alm+Sproat '05)

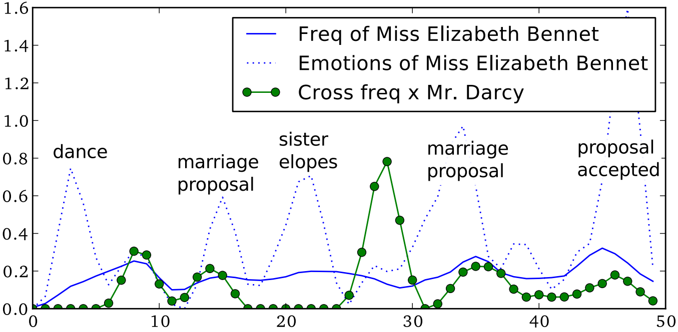
Combine the two:

- Compute a trajectory for each character
- Observe social relationships through time



Combine the two:

- Compute a trajectory for each character
- Observe social relationships through time



Preprocessing

- Chop the novel into paragraphs
- Parse everything and retrieve proper NPs
- Simple coreference on the NPs to find characters
- Emotion: “strong sentiment” cues from (Wilson+al ‘05)

Coreference

Similar to cross-document coreference:

- Shared name elements
- Presence in same documents
- List of gendered names and titles

"Miss Elizabeth Bennet" (f)	Elizabeth Bennet Elizabeth Miss Elizabeth Bennet Miss Bennet
"Miss Eliza" (f)	Miss Eliza Eliza
"Miss Elizabeth" (f)	Miss Elizabeth
"Lizzy" (?)	Lizzy

Use this representation to measure similarity...

Kernel function

$k(x, y)$: similarity between x and y

0: no similarity; $>$ 0: more similar

basic ML building block

Use this representation to measure similarity...

Kernel function

$k(x, y)$: similarity between x and y
0: no similarity; > 0: more similar
basic ML building block

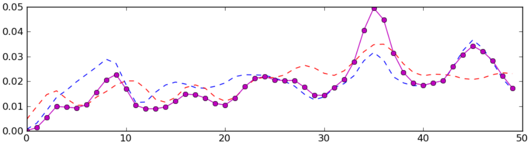
Use *convolution theorem* (Haussler '99) to build a complex kernel out of simpler ones:

$$k(x, y) = \sum_{ch_1 \in X} \sum_{ch_2 \in Y} \underbrace{c(ch_1, ch_2)}_{\text{kernel over characters}}$$

Similarity between characters

$e(ch_1, ch_2)$:

- Similarity for trajectory curves
- Normalized integral of the product
- Used for frequency and emotion



$d(ch_1, ch_2)$

- Nearby words

replied Elizabeth	17
Elizabeth felt	14
Elizabeth looked	10
Elizabeth's mind	7

...

First-order character kernel

$$c_1(ch_1, ch_2) = d(ch_1, ch_2)e(ch_1, ch_2)$$

Adding social network features

Characters are more similar if:

- They each have close friends...
  - (Measured by co-occurrence frequency)
- ...who are also similar

Second-order character kernel

$$c_2(ch_1, ch_2) = c_1(ch_1, ch_2) \sum_{u' \in X} \sum_{v' \in Y} \underbrace{e(\widehat{u, u'}, \widehat{v, v'})}_{\text{relationship strength}} c_1(u', v')$$

Testing similarity

- First, simple proof of concept
- Independent of particular critical theory
- Difficult for very naive models

Testing similarity

- First, simple proof of concept
- Independent of particular critical theory
- Difficult for very naive models

Order discrimination

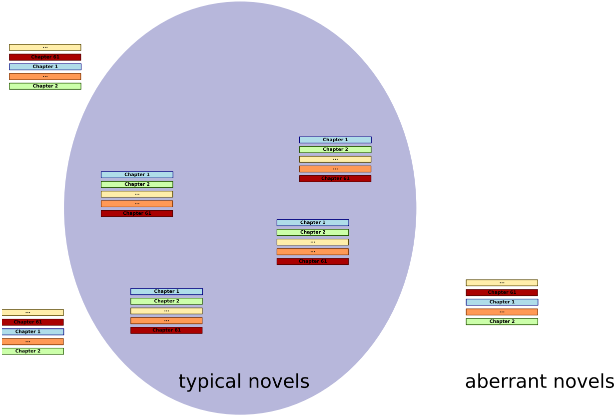
(Karamanis+al ‘04) (Barzilay+Lapata ‘05)

Chapter 1		**...**
Chapter 2		**Chapter 61**
...	**vs**	**Chapter 1**
...		**...**
Chapter 61		**Chapter 2**

Weighted nearest-neighbor

For training set T, is:

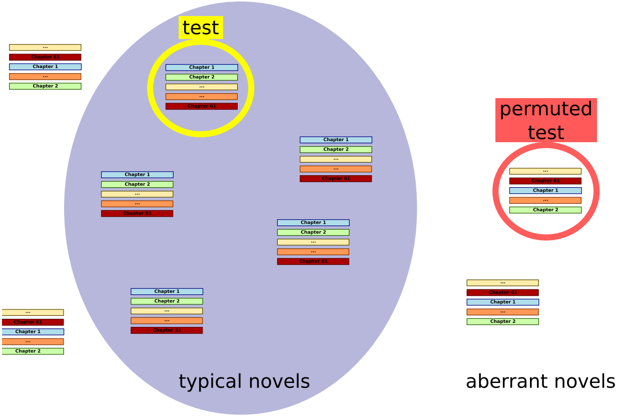
$$\sum_{t \in T} k(t, y) > \sum_{t \in T} k(t, y_{perm})?$$



Weighted nearest-neighbor

For training set T, is:

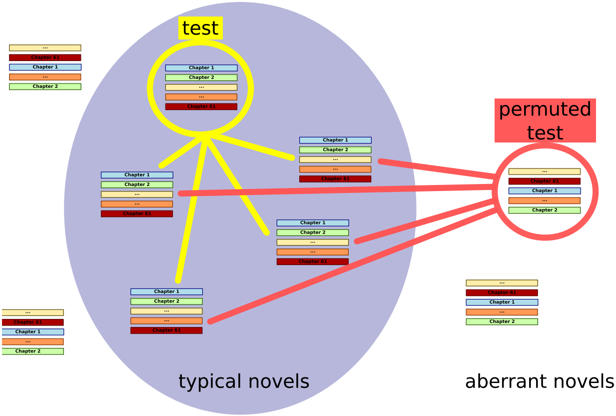
$$\sum_{t \in T} k(t, y) > \sum_{t \in T} k(t, y_{perm})?$$



Weighted nearest-neighbor

For training set T, is:

$$\sum_{t \in T} k(t, y) > \sum_{t \in T} k(t, y_{perm})?$$



Results (30 novels from Project Gutenberg)

Binary classifications

Chance accuracy 50%
Significance via kernel-based non-parametric test (Gretton+al '07)

	Random perm	Reversed

Results (30 novels from Project Gutenberg)

Binary classifications

Chance accuracy 50%
Significance via kernel-based non-parametric test (Gretton+al '07)

	Random perm	Reversed
Whole-novel traj.	50	53

Results (30 novels from Project Gutenberg)

Binary classifications

Chance accuracy 50%
Significance via kernel-based non-parametric test (Gretton+al '07)

	Random perm	Reversed
Whole-novel traj.	50	53
First-order k_1	77	63

Results (30 novels from Project Gutenberg)

Binary classifications

Chance accuracy 50%
Significance via kernel-based non-parametric test (Gretton+al '07)

	Random perm	Reversed
Whole-novel traj.	50	53
First-order k_1	77	63
Second-order k_2	**90**	**67**

We've seen:

- Plot structure: based on *character* and *emotion* over *time*
- Simple ordering test as proof of concept

We've seen:

- Plot structure: based on *character* and *emotion* over *time*
- Simple ordering test as proof of concept

Future work:

- Eventually: search and summarize stories
- Topic modeling: match emotions to events
  - With Dae-Il Kim and Victoria Adams
- Interface for writers to visualize their work
  - With Jon Oberlander and Victoria Adams

Connections

- Very high-level models can still be **entity (character)-based**
- Still possible to validate **automatically** by **disentanglement**
  - Analogues of other tests? Insertion? Human scoring?
- Can we learn more from **sentence-level** models?
  - Do novels have “local transitions”?

Conclusion

- Entities shape discourse at many levels
- Coherence affects referring expressions
- ...ideas from coref help

- Automatic tests are great, but use several to avoid getting fooled

- Fiction: an exciting new area!

Software available

bitbucket.org/melsner/browncoherence

Topical entity grid

(Elsner+Charniak ‘11b)

Relationships between *different* words

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

- 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