CS 4650/7650
Discourse Structure

Jacob Eisenstein

November 4, 2014

Outline

Overview

Coherence

Segmentation

Zoning and ordering

Centering and Reference

Relational models of discourse structure
- Rhetorical structure theory
- The Penn Discourse Treebank

Applications

Discourse

Discourse is concerned with language processing beyond the sentence level:

- resolving references to **entities** and **events**
- identifying the characteristics that make text **coherent**
- describing the rhetorical and narrative **relationships** between units of text.

We have already talked about reference resolution.
Today we'll talk about coherence and discourse relations.

Outline

Overview

Coherence

Segmentation

Zoning and ordering

Centering and Reference

Relational models of discourse structure
- Rhetorical structure theory
- The Penn Discourse Treebank

Applications

Coherence

- In syntax, we are concerned with **grammaticality**.
- In discourse, we are concerned with **coherence**.

1. Sue hid Lakshmi's keys. She was drunk.
2. ?? Sue hid Lakshmi's keys. She likes spinach.

Coherence

- In syntax, we are concerned with **grammaticality**.
- In discourse, we are concerned with **coherence**.

1. Sue hid Lakshmi's keys. She was drunk.
2. ?? Sue hid Lakshmi's keys. She likes spinach.
3. Uma offered spinach to anyone who could keep Lakshmi from driving. Sue hid Lakshmi's keys. She likes spinach.

Coherence

Which version of the story seems more coherent?

1. Serena went to her favorite music store to buy a piano.
2. It was a store Serena had frequented for many years.
3. She was excited that she could finally buy a piano.
4. It was closing just as Serena arrived.

1. Serena went to her favorite music store to buy a piano.
2. She had frequented the store for many years.
3. She was excited that she could finally buy a piano.
4. She arrived just as the store was closing for the day.

Coherence

Which version of the story seems more coherent?

1. **Serena** went to **her** favorite *music store* to buy a piano.
2. *It* was a store **Serena** had frequented for many years.
3. **She** was excited that **she** could finally buy a piano.
4. *It* was closing just as **Serena** arrived.

1. **Serena** went to **her** favorite *music store* to buy a piano.
2. **She** had frequented *the store* for many years.
3. **She** was excited that **she** could finally buy a piano.
4. **She** arrived just as *the store* was closing for the day.

Discourse and semantics

We have seen that a meaningless sentence can be **grammatical**:

$Colorless\ green\ ideas\ sleep\ furiously$

The discourse analogue of **grammaticality** is **coherence**.
Can a text be coherent without meaning?

An example discourse

In today's society, college is ambiguous. We need it to live, but we also need it to love. Moreover, without college most of the world's learning would be egregious. College, however, has myriad costs. One of the most important issues facing the world is how to reduce college costs. Some have argued that college costs are due to the luxuries students now expect. Others have argued that the costs are a result of athletics. In reality, high college costs are the result of excessive pay for teaching assistants.[1]

[1] From Lee Perelman, `http://www.cbc.ca/spark/wp-content/uploads/2012/05/Essays-for-Robo-Reader.pdf`

An example discourse

I live in a luxury dorm. In reality, it costs no more than rat infested rooms at a Motel Six. The best minds of my generation were destroyed by madness, starving hysterical naked, and publishing obscene odes on the windows of the skull. Luxury dorms pay for themselves because they generate thousand and thousands of dollars of revenue. In the Middle Ages, the University of Paris grew because it provided comfortable accommodations for each of its students, large rooms with servants and legs of mutton. Although they are expensive, these rooms are necessary to learning.

An example discourse

The second reason for the five-paragraph theme is that it makes you focus on a single topic. Some people start writing on the usual topic, like TV commercials, and they wind up all over the place, talking about where TV came from or capitalism or health foods or whatever. But with only five paragraphs and one topic youre not tempted to get beyond your original idea, like commercials are a good source of information about products. You give your three examples, and zap! you're done. This is another way the five-paragraph theme keeps you from thinking too much.

Automated essay scoring

This essay receives a perfect score according to ETS's E-rater software.

- The essay contains many elements of well-structured discourse, even if it doesn't make sense.
- "E-rater is not designed to be a fact-checker" – Paul Deane, principal research scientist.
- Why is it coherent?

Educational Testing Service

(http://www.nytimes.com/2012/04/23/education/robo-readers-used-to-grade-test-essays.html)

Discourse connectors

In today's society, college is ambiguous. We need it to live, but we also need it to love. Moreover, without college most of the world's learning would be egregious. College, however, has myriad costs. One of the most important issues facing the world is how to reduce college costs. Some have argued that college costs are due to the luxuries students now expect. Others have argued that the costs are a result of athletics. In reality, high college costs are the result of excessive pay for teaching assistants.

Discourse connectors

In today's society, college is ambiguous. We need it to live, but we also need it to love. Moreover, without college most of the world's learning would be egregious. College, however, has myriad costs. One of the most important issues facing the world is how to reduce college costs. Some have argued that college costs are due to the luxuries students now expect. Others have argued that the costs are a result of athletics. In reality, high college costs are the result of excessive pay for teaching assistants.

Lexical chains

In today's society, college is ambiguous. We need it to live, but we also need it to love. Moreover, without college most of the world's learning would be egregious. College, however, has myriad costs. One of the most important issues facing the world is how to reduce college costs. Some have argued that college costs are due to the luxuries students now expect. Others have argued that the costs are a result of athletics. In reality, high college costs are the result of excessive pay for teaching assistants.

Lexical chains

In today's society, college is ambiguous. We need it to live, but we also need it to love. Moreover, without college most of the world's learning would be egregious. College, however, has myriad costs. One of the most important issues facing the world is how to reduce college costs. Some have argued that college costs are due to the luxuries students now expect. Others have argued that the costs are a result of athletics. In reality, high college costs are the result of excessive pay for teaching assistants.

Lexical chains

In today's society, college is ambiguous. We need it to live, but we also need it to love. Moreover, without college most of the world's learning would be egregious. College, however, has myriad costs. One of the most important issues facing the world is how to reduce college costs. Some have argued that college costs are due to the luxuries students now expect. Others have argued that the costs are a result of athletics. In reality, high college costs are the result of excessive pay for teaching assistants.

Discourse relations

1. In today's society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world's learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.

Discourse relations

1. In today's society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world's learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.

Discourse relations

1. In today's society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world's learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.

Discourse relations

1. In today's society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world's learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.

Discourse relations

1. In today's society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world's learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.

Discourse relations

1. In today's society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world's learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.

Overview of discourse structure

What kind of structures can represent coherence?

- linear segmentation
- zoning and ordering
- relations between adjacent spans
- relations between arbitrary spans
- graphs

Outline

Overview

Coherence

Segmentation

Zoning and ordering

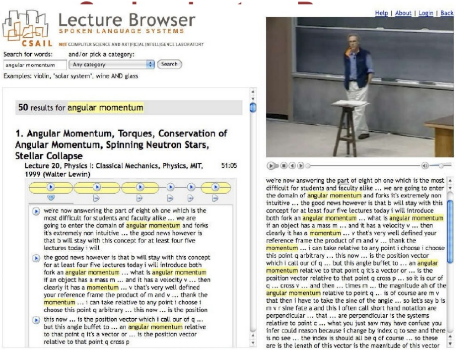
Centering and Reference

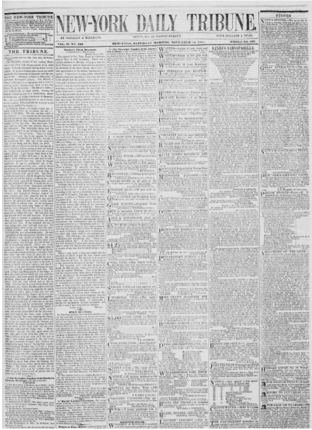
Relational models of discourse structure

- Rhetorical structure theory
- The Penn Discourse Treebank

Applications

Segmentation





Segmentation from cohesion

- Halliday and Hasan (1976) introduced the notion of **cohesion**. It refers to the links that hold text together.
- In coherent texts, cohesive links occur within topic segments, not between them.
- Types of cohesion:
  - **Reference**: use of pronouns to refer to entities defined elsewhere in the text (usually anaphora)
    My brother is a great driver. **He** learned from the best.
  - **Substitution**: replacement of general phrases for more specific elements, when the meaning is clear
    You should visit Pittsburgh, if you haven't already **done so**.
  - **Ellipsis**: the omission of words to avoid repeated phrases
    The younger child was outgoing, the older more reserved.
  - **Conjunction** between phrases with and, so, however
  - **Lexical**: repeated words, including synonyms etc

Cohesion versus Coherence

- Cohesion without coherence
  - Wash and core six apples.
  - Use the apples to cut out material for your new suit.

Cohesion versus Coherence

- Cohesion without coherence
  - Wash and core six apples.
  - Use the apples to cut out material for your new suit.
- Coherence without cohesion
  - I came home from work at 6:00pm.
  - Dinner consisted of two pieces of fish and a bowl of rice.

Textiling

Hearst (1993) showed that lexical cohesion could be quantified and used to support unsupervised topic topic segmentation.

```
-------------------------------------------------------------------------------------------------------------
Sentence:        05   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95
-------------------------------------------------------------------------------------------------------------
14       form    1     111 1   1                               1 1     1   1       1     1     1    1
 8  scientist                   11            1   1               1          1      1  1
 5      space 11    1     1                                                            1
25       star    1             1                                       11 22  111112  1 1  1   11 1111     1
 5     binary                                                           11  1          1                    1
 4    trinary                                                            1   1         1                    1
 8 astronomer 1                1                                        1 1           1   1    1  1
 7      orbit   1                    1                                     12    1 1
 6       pull                          2     1 1                               1  1
16     planet   1    1       11               1           1                21  11111                  1     1
 7     galaxy    1                                                  1                1  11     1           1
 4      lunar            1  1     1       1
19       life 1  1  1                              1    11 1  11  1      1                 1 1    1 111  1 1
27       moon        13  1111   1 1 22 21  21    21           11 1
 3       move                                      1   1   1
 7  continent                                      2 1 1 2 1
 3  shoreline                                            12
 6       time                      1               1  1  1     1                                          1
 3      water                                 11            1
 6        say                               1 1        1        11                1
 3    species                                        1  1  1
-------------------------------------------------------------------------------------------------------------
Sentence:        05   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95
-------------------------------------------------------------------------------------------------------------
```

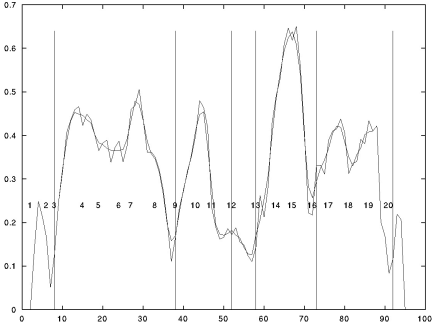
Unsupervised topic segmentation

Components for unsupervised topic segmentation:

- Tokenization
- A similarity metric for adjacent spans of text. Textiling uses cosine similarity,

$$\mathrm{sim}(\mathbf{x}, \mathbf{y}) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}}$$

- A boundary identification heuristic.
- (Optionally) a smoothing technique.



Optimal text segmentation

- Assume a metric for scoring each segment, $f(i, j) \in \mathbb{R}$.
- The optimal segmentation is

$$\arg\max_{\mathbf{y}} \sum_{k=0}^{\#|\mathbf{y}|} f(y_k, y_{k+1}) \tag{1}$$

- How can we optimize this? How many possible segmentations are there?

Features for unsupervised topic segmentation

- Lexical chains (Galley et al, 2004)
- Latent semantic analysis (a.k.a. SVD; Choi et al, 2001)
- Thesaurus relations (Morris and Hirst, 1991)
- Probabilistic models (Utiyama & Isahara, 2001; Eisenstein and Barzilay, 2008)
- Auditory similarity (Malioutov et al, 2007)

Outline

Overview

Coherence

Segmentation

Zoning and ordering

Centering and Reference

Relational models of discourse structure

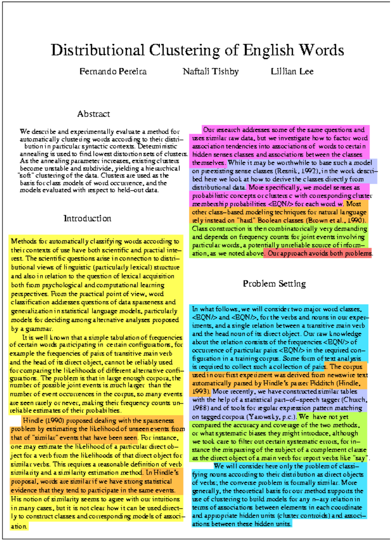
- Rhetorical structure theory
- The Penn Discourse Treebank

Applications

Argumentative zoning

- BKG: General scientific background (yellow)
- OTH: Neutral descriptions of other people's work (orange)
- OWN: Neutral descriptions of own, new work (blue)
- AIM: Statements of the particular aim of the current paper (pink)
- TXT: Statements of textual organization of the current paper (red)
- CTR: Contrastive or comparative statements about other work; explicit mention of weaknesses of other work (green)
- BAS: Statements that own work is based on other work (purple)

(From Teufel, 1999)



Distributional Clustering of English Words

Fernando Pereira Naftali Tishby Lillian Lee

Abstract

We describe and experimentally evaluate a method for automatically clustering words according to their distribution in particular syntactic contexts. Deterministic annealing is used to find lowest distortion sets of clusters. As the annealing parameter increases, existing clusters become unstable and subdivide, yielding a hierarchical "soft" clustering of the data. Clusters are used as the basis for class models of word occurrence, and the models evaluated with respect to held–out data.

Introduction

Methods for automatically classifying words according to their contexts of use have both scientific and practial interest. The scientific questions arise in connection to distributional views of linguistic (particularly lexical) structure and also in relation to the question of lexical acquisition both from psychological and computational learning perspectives. From the practical point of view, word classification addresses questions of data sparseness and generalization in statistical language models, particularly models for deciding among alternative analyses proposed by a grammar.

It is well known that a simple tabulation of frequencies of certain words participating in certain configurations, for example the frequencies of pairs of transitive main verb and the head of its direct object, cannot be reliably used for comparing the likelihoods of different alternative configurations. The problem is that in large enough corpora, the number of possible joint events is much larger than the number of event occurrences in the corpus, so many events are seen rarely or never, making their frequency counts unreliable estimates of their probabilties.

Hindle (1990) proposed dealing with the sparseness problem by estimating the likelihood of unseen events from that of "similar" events that have been seen. For instance, one may estimate the likelihood of a particular direct object for a verb from the likelihoods of that direct object for similar verbs. This requires a reasonable definition of verb similarity and a similarity estimation method. In Hindle's proposal, words are similar if we have strong statistical evidence that they tend to participate in the same events. His notion of similarity seems to agree with our intuitions in many cases, but it is not clear how it can be used directly to construct classes and corresponding models of association.

Our research addresses some of the same questions and uses similar raw data, but we investigate how to factor word association tendencies into associations of words to certain hidden senses classes and associations between the classes themselves. While it may be worthwhile to base such a model on preexisting sense classes (Resnik, 1992), in the work described here we look at how to derive the classes directly from distributional data. More specifically, we model senses as probabilistic concepts or clusters c with corresponding cluster membership probabilities <EQN/> for each word w. Most other class–based modeling techniques for natural language rely instead on "hard" Boolean classes (Brown et al., 1990). Class constuction is then combinatorically very demanding and depends on frequency counts for joint events involving particular words, a potentially unreliable source of information, as we noted above. Our approach avoids both problems.

Problem Setting

In what follows, we will consider two major word classes, <EQN/> and <EQN/>, for the verbs and nouns in our experiments, and a single relation between a transitive main verb and the head noun of its direct object. Our raw knowledge about the relation consists of the frequencies <EQN/> of occurrence of particular pairs <EQN/> in the required configuration in a training corpus. Some form of text analysis is required to collect such a collection of pairs. The corpus used in our first experiment was derived from newswire text automatically parsed by Hindle's parser Fidditch (Hindle, 1993). More recently, we have constructed similar tables with the help of a statistical part–of–speech tagger (Church, 1988) and of tools for regular expression pattern matching on tagged corpora (Yarowsky, p.c.). We have not yet compared the accuracy and coverage of the two methods, or what systematic biases they might introduce, although we took care to filter out certain systematic errors, for instance the misparsing of the subject of a complement clause as the direct object of a main verb for report verbs like "say".

We will consider here only the problem of classifying nouns according to their distribution as direct objects of verbs; the converse problem is formally similar. More generally, the theoretical basis for our method supports the use of clustering to build models for any n–ary relation in terms of associations between elements in each coordinate and appropriate hidden units (cluster centroids) and associations between these hidden units.

Functional discourse structure

Some genres with conventionalized functional organization:

- **research papers**:
  abstract, background, methods, results, discussion
- **inverted pyramid**: lede paragraph, body, tail
- **French rhetoric**: thesis, antithesis, synthesis

Computational approaches have focused on recognizing functional patterns in conventionalized domains.

- biomedical abstracts (e.g., McKnight and Srinivasan 2003)
- legal documents (e.g., Palau and Moens 2009)

Conventionalized topic structures

	Wisconsin	Louisiana	Vermont
1	Etymology	Etymology	Geography
2	History	Geography	History
3	Geography	History	Demographics
4	Demographics	Demographics	Economy
5	Law and government	Economy	Transportation
6	Economy	Law and government	Media
7	Municipalities	Education	Utilities
8	...	...	...

Wikipedia articles about US states

Chen et al (2009) used probability distributions over permutations to model conventionalized topic sequences.

Intentional discourse structure

- Goal: explain discourse in terms of the author's **intentions**.
  - Webber et al (2011) view intentional structure as a type of functional structure.
  - text as the outcome of a planning process (Grosz and Sidner 1986, 1990; Moore and Paris 1993)
- Computationally...
  - Little work on **detecting** intentional discourse structure,
  - but lots of work on **generating** text from intentional structures (e.g., Moore and Paris 1993).

Intentional discourse structure

System's Utterance:

You should replace `(SETQ X 1)` with `(SETF X 1)`. `SETQ` can only be used to assign a value to a simple-variable. `SETF` can be used to assign a value to any generalized-variable. A generalized-variable is a storage location that can be named by any access function.

Intentional Structure:

I0: (Intend E (Intend U (Replace U (SETQ X 1) (SETF X 1))

I1: (Intend E (Recommend E U
(Replace U (SETQ X 1) (SETF X 1))))

I2: (Intend E (Persuaded E U
(Replace U (SETQ X 1) (SETF X 1))))

I3: (Intend E (Believe U
(Someref (Differences-wrt-goal SETQ SETF enhance-maintainability))))

I4: (Intend E (Believe U
(Use SETQ assign-value-to-simple-variable)))

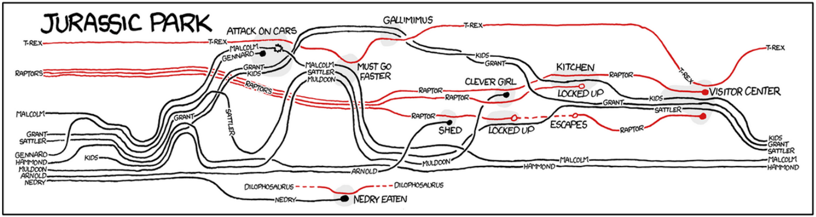
I5: (Intend E (Believe U
(Use SETF assign-value-to-generalized-variable)))

I6: (Intend E (Know-About U (Concept generalized-variable)))

I7: (Intend E (Believe U (Isa generalized-variable (storage-location (restrict named-by symbol)))))

Eventualities

- Goal: explain discourse in terms of **events and states**.
- Some simple examples
  - navigational directions (go right at Peachtree, ...)
  - the “methods” section of a scientific paper
  - news stories, especially low-scoring sports and crime stories



http://xkcd.com/657

Outline

Overview

Coherence

Segmentation

Zoning and ordering

Centering and Reference

Relational models of discourse structure
- Rhetorical structure theory
- The Penn Discourse Treebank

Applications

Salience/Focus

Only some recently mentioned entities can be referred to by pronouns:

John went to Bob's party and parked next to a classic Ford Falcon.
He went inside and talked to Bob for more than an hour.
Bob told him that he recently got engaged.
He also said he bought it (???)/ the Falcon yesterday.

Key insight:

Capturing **which entities are salient** (in focus) **reduces the amount of search** (inference) **necessary to interpret pronouns**!

Centering theory: definitions

Utterance:

A sequence of words (typically a sentence or clause) at a particular point in a discourse.

The centers of an utterance:

Entities (semantic objects) which link the utterance to the previous and following utterances.

Centering: assumptions

In each utterance, **some discourse entities are more salient** than others.

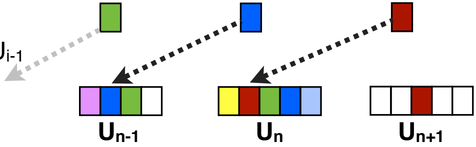
We maintain a list of discourse entities, ranked by salience.
The position in this list determines how easy it is to refer back to an entity in the next utterance.
Each utterance updates this list.

This list is called the **local attentional state.**

The two centers of an utterance

Backward-looking:
mentioned in U_i and U_{i-1}

Forward-looking:
mentioned in U_i



The **forward looking center** of an utterance $\mathbf{U_n}$ is a partially ordered **list of the entities** mentioned in $\mathbf{U_n}$.
The ordering reflects **salience** within $\mathbf{U_n}$:

subject > direct object > object,….

The **backward looking center** of an utterance $\mathbf{U_n}$ is the highest ranked **entity** in the forward looking center of $\mathbf{U_{n-1}}$ that is mentioned in $\mathbf{U_n}$.

Center realization and pronouns

Observation: Only the most salient entities of $\mathbf{U_{n-1}}$ can be referred to by pronouns in $\mathbf{U_n}$.

Constraint/Rule 1:

If ***any*** **element of FW($\mathbf{U_{n-1}}$)** is realized as **a pronoun** in U_n, then the **BW($\mathbf{U_n}$)** has to be realized **as a pronoun** in U_n as well.

Sue *told* ***Joe*** *to feed* ***her dog***.
BW=Susan, **FW**={Sue, Joe, dog}

He *asked* ***her*** *what to feed* ***it****.*
BW=Sue, **FW**={Joe, Sue, dog}

✔ Constraint obeyed

He *asked* ***Sue*** *what to feed* ***it****.*
BW=Sue, **FW**={Joe, Sue, dog}

✘ Constraint violated:
Sue should be a pronoun as well.

Transitions between sentences

Center continuation:

$BW(U_n) = BW(U_{n-1})$. $BW(U_n)$ is highest ranked element in $FW(U_n)$

Sue *gave* ***Joe a dog.***

She *told* ***him*** *to feed* ***it*** *well.* **BW**=Sue, **FW**={Sue, Joe, dog}

She *asked* ***him*** *whether he liked the gift.* **BW**=Sue, **FW**={Sue, Joe, gift}

Center retaining:

$BW(S_n) = BW(S_{n-1})$. $BW(S_n) \neq$ highest ranked element in $FW(S_n)$

Sue *gave* ***Joe a dog****.*

She *told* ***him*** *to feed* ***it*** *well.* **BW**=Sue, **FW**={Sue, Joe, dog}

John *asked* ***her*** *what to feed him.* **BW**=Sue, **FW**={Joe, Sue, dog}

Center shifting:

$BW(S_n) \neq BW(S_{n-1})$

Susan *gave* ***Joe a dog.***

She *told* ***him*** *to feed* ***it*** *well.* **BW**=Sue, **FW**={Sue, Joe,dog}

The dog *was very cute.* **BW**=dog, **FW**={dog}

Local coherence: preferred transitions

Rule/Constraint 2:

- Center **continuation** is preferred over center **retaining**.
- Center **retaining** is preferred over center shifting.

Local coherence is achieved by maximizing the number of center continuations.

Example: Coherent discourse

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

backward-looking center: ? (no previous discourse)

forward-looking center: {*John'*, *store'*, *piano'*}

He *had frequented* ***the store*** *for many years.*

backward-looking center: {*John'*}

forward-looking center: {*John'*, *store'*}

Continuation

He *was excited that* ***he*** *could finally buy* ***a piano****.*

backward-looking center: {*John'*}

forward-looking center: {*John'*, *piano'*}

Continuation

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

backward-looking center: {*John'*}

forward-looking center: {*John'*, *store'*}

Example: incoherent discourse

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

backward-looking center: ? (no previous discourse)

forward-looking center: {*John'*, *store'*, *piano'* }

It *was* ***a store John*** *had frequented for many years.*

backward-looking center: {*John'* }

forward-looking center: {*store'*, *John'* }

Continuation

He *was excited that* ***he*** *could finally buy* ***a piano****.*

backward-looking center: {*John'* }

forward-looking center: {*John'*, *piano'* }

Retention

It *was closing just as* ***John*** *arrived.*

backward-looking center: {*John'* }

forward-looking center: {*store'*, *John'* }

Coherence from entities

- Barzilay and Lapata (2005, 2008) showed how to build computational models of centering and local coherence, using shallow NLP.
- Their system did not fully solve the entity coreference problem, but was still accurate enough to support useful applications.
- Key idea: represent centering through an **entity grid**.

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)

Sentence #:	Walter	Hall	Somerset	Bt.age	day	wife	history	family
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)

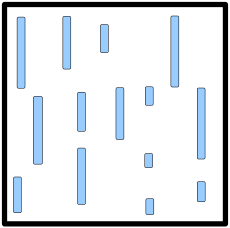
Sentence #:	Walter	Hall	Somerset	Bt.age	day	wife	history	family
1	S	X	X	O				
2	S			O	X	O		
3							X	X

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.

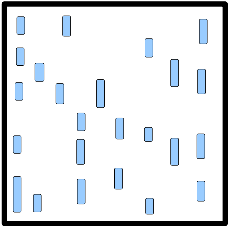
There followed the history of the ancient family.

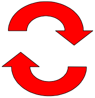
Local coherence

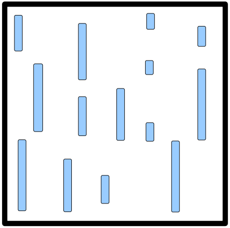


Very low zoom:
entities in long contiguous columns.

A randomly
permuted document:

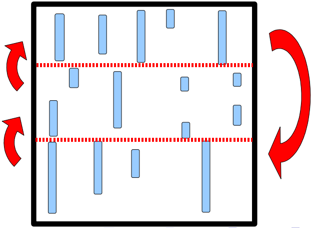






Backwards?

Move the
paragraphs?



Applications of coherence modeling

- **Sentence ordering**
  - In extractive summarization, a set of sentences are selected to cover the content of a collection of documents.
  - Coherence modeling can order these sentences into a useful, readable summary.

Applications of coherence modeling

- **Sentence ordering**
  - In extractive summarization, a set of sentences are selected to cover the content of a collection of documents.
  - Coherence modeling can order these sentences into a useful, readable summary.
- **Readability assessment**
  - B&L show that the entity grid approach accurately distinguishes *Encyclopedia Britannica* from *Britannica Elementary*.
  - Also used for essay scoring by ETS (Burstein et al 2010)

Outline

Overview

Coherence

Segmentation

Zoning and ordering

Centering and Reference

Relational models of discourse structure
- Rhetorical structure theory
- The Penn Discourse Treebank

Applications

Discourse relations

- Goal: identify the discourse **relations** that hold between two *units of discourse*.
- Often discourse relations are signalled by discourse connectors:

  *The kite was created in China, about 2,800 years ago. **Later** it spread into other Asian countries, like India, Japan and Korea. **However**, the kite only appeared in Europe by about the year 1600.*[2]

  - **later** indicates a SUCCESSION relation
  - **however** indicates a CONTRAST relation

[2] http://simple.wikipedia.org/wiki/Kite via Webber et al (2011).

Discourse relations

- Discourse relations can also be signalled implicitly.

  Clouds are heavy. The water in a cloud can have a mass of several million tons.[3]

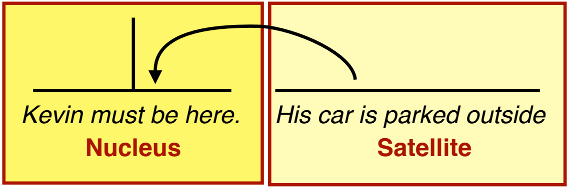
- The second utterance implicitly ELABORATES or INSTANTIATES the claim in the first sentence.

[3] `http://simple.wikipedia.org/wiki/Cloud` via Webber et al (2011).

Rhetorical structure theory

- RST (Mann & Thompson, 1987) describes coherence relations between **utterances**
  - Utterances are not exactly sentences and not exactly clauses...
  - More technically: *elementary discourse units* (EDUs). Identifying them is a preprocessing step.
- Utterances are connected by rhetorical relations: Evidence, elaboration, contrast, attribution, list, ...
- Asymmetric relations have a nucleus and a satellite.
- Symmetric relations have multiple nucleii.

The Evidence relation



Informally: writer is trying to convince reader that nucleus is true.

Constraints on Nucleus:

Reader might not believe Nucleus as much as Writer would like.

Constraints on Satellite:

Reader believes Satellite or will find it credible.

Constraints on Nucleus and Satellite:

Reader's understand of Satellite increases his belief in Nucleus

Effect: Reader's belief of Nucleus is increased.

Nucleus-satellite relations

Elaboration:

Satellite gives more information about **nucleus**:

[The company wouldn't elaborate], [citing competitive reasons].

Attribution:

Nucleus is reported speech/belief.

Satellite attributes Nucleus to somebody:

[Analysts estimated] [that sales at US stores declined]

Background:

Satellite gives background for **nucleus**:

[T is the pointer to the root.] [Initialize T.]

Multi-nucleus relations

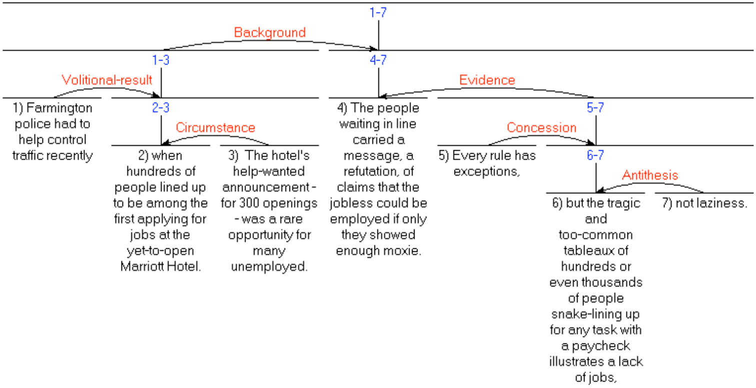
Contrast:

Two or more nuclei contrast along some dimension:
[The priest was in a very bad temper], [but the lama was quite happy].

List:

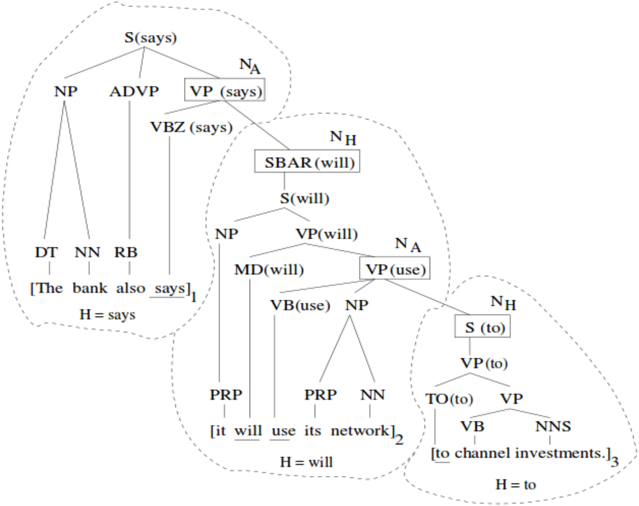
A series of nuclei is given, without contrast or comparision:
[Billy Bones was the mate]; [Long John, he was quartermaster]

Discourse structure is hierarchical



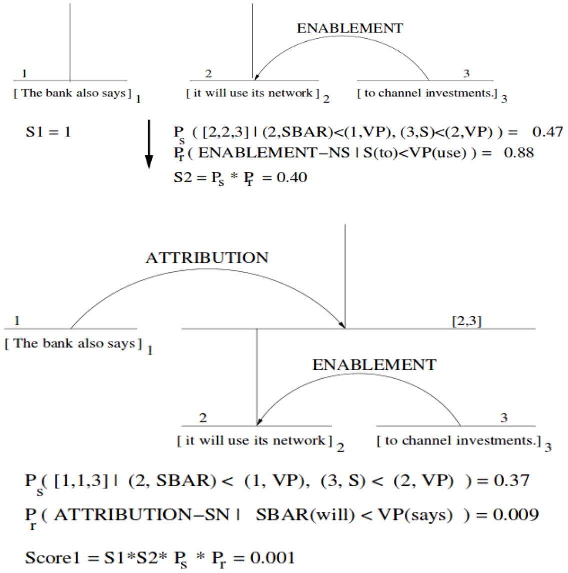
RST website: http://www.sfu.ca/rst/

RST Parsing: EDU segmentation



D = { (2, SBAR(will)) < (1, VP(says)) , (3, S(to)) < (2, VP(use)) }

RST Parsing



ENABLEMENT
1
2
3
[The bank also says] 1
[it will use its network] 2
[to channel investments.] 3
S1 = 1
P_s ([2,2,3] | (2,SBAR)<(1,VP), (3,S)<(2,VP)) = 0.47
P_r (ENABLEMENT–NS | S(to)<VP(use)) = 0.88
S2 = P_s * P_r = 0.40
ATTRIBUTION
1
[2,3]
[The bank also says] 1
ENABLEMENT
2
3
[it will use its network] 2
[to channel investments.] 3
P_s ([1,1,3] | (2, SBAR) < (1, VP), (3, S) < (2, VP)) = 0.37
P_r (ATTRIBUTION–SN | SBAR(will) < VP(says)) = 0.009
Score1 = S1*S2* P_s * P_r = 0.001

Summarization from discourse parsing

With its distant orbit — 50% farther from the sun than Earth — and slim atmospheric blanket, $C1$ Mars experiences frigid weather conditions. $C2$ Surface temperatures typically average about -60 degrees C at the equator and can dip to -123 degrees C near the poles. $C3$ Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, $C4$ but any liquid water formed in this way would evaporate almost instantly $C5$ because of the low atmospheric pressure. $C6$ Although the atmosphere holds a small amount of water, and water-ice clouds sometimes develop, $C7$ most Martian weather involves blowing dust or carbon dioxide. $C8$ Each winter, for example, a blizzard of frozen carbon dioxide rages over one pole, and a few meters of this dry-ice snow accumulate as previously frozen carbon dioxide evaporates from the opposite polar cap. $C9$ Yet even on the summer pole, where the sun remains in the sky all day long, temperatures never warm enough to melt frozen water. $C10$

Summarization from discourse parsing

With its distant orbit — 50% farther from the sun than Earth — and slim atmospheric blanket, $C1$ Mars experiences frigid weather conditions. $C2$ Surface temperatures typically average about -60 degrees C at the equator and can dip to -123 degrees C near the poles. $C3$ Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, $C4$ but any liquid water formed in this way would evaporate almost instantly $C5$ because of the low atmospheric pressure. $C6$ Although the atmosphere holds a small amount of water, and water-ice clouds sometimes develop, $C7$ most Martian weather involves blowing dust or carbon dioxide. $C8$ Each winter, for example, a blizzard of frozen carbon dioxide rages over one pole, and a few meters of this dry-ice snow accumulate as previously frozen carbon dioxide evaporates from the opposite polar cap. $C9$ Yet even on the summer pole, where the sun remains in the sky all day long, temperatures never warm enough to melt frozen water. $C10$

Summarization from discourse parsing

With its distant orbit — 50% farther from the sun than Earth — and slim atmospheric blanket, $C1$ Mars experiences frigid weather conditions. $C2$ Surface temperatures typically average about -60 degrees C at the equator and can dip to -123 degrees C near the poles. $C3$ Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, $C4$ but any liquid water formed in this way would evaporate almost instantly $C5$ because of the low atmospheric pressure. $C6$ Although the atmosphere holds a small amount of water, and water-ice clouds sometimes develop, $C7$ most Martian weather involves blowing dust or carbon dioxide. $C8$ Each winter, for example, a blizzard of frozen carbon dioxide rages over one pole, and a few meters of this dry-ice snow accumulate as previously frozen carbon dioxide evaporates from the opposite polar cap. $C9$ Yet even on the summer pole, where the sun remains in the sky all day long, temperatures never warm enough to melt frozen water. $C10$

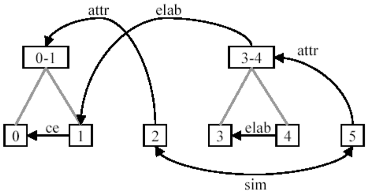
Summarization from discourse parsing

With its distant orbit — 50% farther from the sun than Earth — and slim atmospheric blanket, $C1$ Mars experiences frigid weather conditions. $C2$ Surface temperatures typically average about -60 degrees C at the equator and can dip to -123 degrees C near the poles. $C3$ Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, $C4$ but any liquid water formed in this way would evaporate almost instantly $C5$ because of the low atmospheric pressure. $C6$ Although the atmosphere holds a small amount of water, and water-ice clouds sometimes develop, $C7$ most Martian weather involves blowing dust or carbon dioxide. $C8$ Each winter, for example, a blizzard of frozen carbon dioxide rages over one pole, and a few meters of this dry-ice snow accumulate as previously frozen carbon dioxide evaporates from the opposite polar cap. $C9$ Yet even on the summer pole, where the sun remains in the sky all day long, temperatures never warm enough to melt frozen water. $C10$

Discourse graphs

Wolf and Gibson (2003) argue that many discourses cannot be fully described by a tree, and that graphs are more appropriate.

0. Farm prices in October edged up 0.7% from September
1. as raw milk prices continued to rise,
2. the Agriculture Department said.
3. Milk sold to the nations dairy plants and dealers averaged $14.50 for each hundred pounds,
4. up 50 percent from September and up $1.50 from October 1988,
5. the department said.



(see http://www.isi.edu/~marcu/discourse/Discourse%20structures.htm)

The Penn Discourse Treebank

- Rhetorical Structure Theory (RST) builds a **tree** that completely covers all the text.
- The Penn Discourse Treebank (PDTB) takes a different approach.
  - Spans of text are related by discourse connectors (e.g., however), which may be explicit or implicit.
  - Not every span may participate in a relation.
  - Some spans may participate in multiple relations.
- Officially: Lexicalized Tree-Adjoining Grammar for Discourse (D-LTAG)

Serving as an arg to multiple relations

(9) In times past, life-insurance salesmen targeted heads of household, meaning men, but ours is a two-income family and accustomed to it. So if anything happened to me, I'd want to leave behind enough so that my 33-year-old husband would be able to pay off the mortgage . . . [Lee et al., 2006]

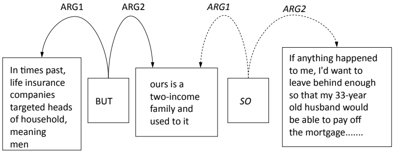
Serving as an arg to multiple relations

(9) In times past, life-insurance salesmen targeted heads of household, meaning men, but ours is a two-income family and accustomed to it. So if anything happened to me, I'd want to leave behind enough so that my 33-year-old husband would be able to pay off the mortgage . . .

[Lee et al., 2006]

Serving as an arg to multiple relations

Fully Shared Arg: Example



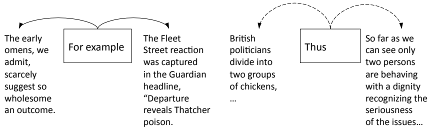
Partial connectivity – Disconnected structures

(10) The early omens, we admit, scarcely suggest so wholesome an outcome. The Fleet Street reaction was captured in the Guardian headline, "Departure Reveals Thatcher Poison." British politicians divide into two groups of chickens, those with their necks cut and those screaming the sky is falling. So far as we can see only two persons are behaving with a dignity recognizing the seriousness of the issues: Mr. Lawson and Sir Alan Walters . . . . [wsj_0553]

Partial connectivity

(11) The early omens, we admit, scarcely suggest so wholesome an outcome. Implicit=FOR EXAMPLE The Fleet Street reaction was captured in the Guardian headline, "Departure Reveals Thatcher Poison." NoRel British politicians divide into two groups of chickens, those with their necks cut and those screaming the sky is falling. Implicit=THUS So far as we can see only two persons are behaving with a dignity recognizing the seriousness of the issues: Mr. Lawson and Sir Alan Walters . . . . [wsj_0553]

Partial connectivity



Automatically recognizing coherence relations

Task involves:

- Identifying the evidence for the discourse relation – ie, evidence for the "discourse predicate";
- Identifying the arguments related by that predicate;
- Identifying the sense of the relation.

[Elwell & Baldridge, 2008; Lin et al, 2010; Pitler & Nenkova, 2009; Prasad et al. 2008; Prasad, Joshi & Webber, 2010; Wellner & Pustejovsky, 2007]

Outline

Overview

Coherence

Segmentation

Zoning and ordering

Centering and Reference

Relational models of discourse structure

- Rhetorical structure theory
- The Penn Discourse Treebank

Applications

Applications of discourse parsing

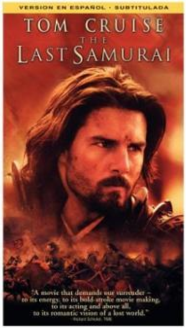
- Automated essay scoring
  - measuring coherence
  - identifying thesis statement
  - relatedness to the prompt
- Summarization
  - extraction: identifying key statements and arguments
  - ordering: assembling extracted sentences into a coherent argument
- Conversation thread disentanglement
- Information extraction
- **Sentiment analysis**

Discourse processing for sentiment analysis

- Standard approach: count evaluative words and compute overall score for a text.
- Refined version: upweight the words at the end of the text. This gives $\sim$ 65% accuracy (Voll & Taboada 2007)

The need for discourse

It could have been a **great** movie. It could have been **excellent**, and to all the people who have forgotten about the older, **greater** movies before it, will think that as well. It does have **beautiful** scenery, some of the **best** since Lord of the Rings. The acting is **well** done, and I really **liked** the son of the leader of the Samurai. He was a **likeable** chap, and I **hated** to see him die... But, other than all that, this movie is nothing more than hidden **rip-offs**.



Discourse processing for sentiment analysis

Discourse to the rescue?

- The top-level structure is CONCESSION, with the nucleus this movie is nothing more than hidden rip-offs.
- The negativity of the nucleus of the top-level element should outweigh all the other positive terms.
- Taboada et al (2009)
  - Hand-labeled RST boosts accuracy from 65% to 79%.
  - But no positive results with automatic discourse parsers.
- An alternative: identify the **functional** discourse structure, thus distinguishing descriptive and evaluative content.

Resources: data

- The RST discourse treebank (LDC2002T07): 385 WSJ articles from the Penn TreeBank
- The Penn discourse treebank 2.0 (LDC2008T05)
  - covers entire Penn TreeBank
  - explicit and implicit discourse connectors
- The Discourse GraphBank (LDC2005T08): edges allowed between any pair of units

Overview

Discourse structure describes the high-level organization of text.

- **Coherence** is the discourse equivalent of **grammaticality**.
- Some discourse processing tasks:
  - linear segmentation
  - hierarchical parsing and chunking
  - coherence estimation

Next time: towards state-of-the-art RST parsing