Terminology

- **Discourse**: anything longer than a single utterance or sentence
  - Monologue
  - Dialogue:
    - May be multi-party
    - May be human-machine
Is this text coherent?

“Consider, for example, the difference between passages (18.71) and (18.72). Almost certainly not. The reason is that these utterances, when juxtaposed, will not exhibit coherence. Do you have a discourse? Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book.”

Or, this?

“Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book. Do you have a discourse? Almost certainly not. The reason is that these utterances, when juxtaposed, will not exhibit coherence. Consider, for example, the difference between passages (18.71) and (18.72).”
What makes a text coherent?

- Discourse structure
  - In a coherent text the parts of the discourse exhibit a sensible ordering and hierarchical relationship
- Rhetorical structure
  - The elements in a coherent text are related via meaningful relations ("coherence relations")
- Entity structure ("Focus")
  - A coherent text is about some entity or entities, and the entity/entities is/are referred to in a structured way throughout the text.

Outline

- Discourse Structure
  - TextTiling (unsupervised)
  - Supervised approaches
- Coherence
  - Hobbs coherence relations
  - Rhetorical Structure Theory
  - Entity Structure
- Pronouns and Reference Resolution
Conventions of Discourse Structure

• Differ for different genres
  – Academic articles:
    • Abstract, Introduction, Methodology, Results, Conclusion
  – Newspaper stories:
    • Inverted Pyramid: Lead then expansion, least important last
  – Textbook chapters
  – News broadcasts
  – Arguments
  – NB: We can take advantage of this to ‘parse’ discourse structures

Discourse Segmentation

• Simpler task: Separating document into *linear sequence of subtopics*

• Applications
  – Information retrieval
    • Automatically segmenting a TV news broadcast or a long news story into sequence of stories
    • Audio browsing (e.g. of voicemail)
  – Text summarization
  – Information extraction
    • Extract information from a coherent segment or topic
  – Question Answering
Unsupervised Segmentation

- Hearst (1997): 21-paragraph science news article on “Stargazers”
- Goal: produce the following subtopic segments:

1-3 Intro - the search for life in space
4–5 The moon’s chemical composition
6-8 How early earth-moon proximity shaped the moon
9–12 How the moon helped life evolve on earth
13 Improbability of the earth-moon system
14–16 Binary/trinary star systems make life unlikely
17–18 The low probability of nonbinary/trinary systems
19–20 Properties of earth’s sun that facilitate life
21 Summary

Intuition: Cohesion

- Halliday and Hasan (1976): “The use of certain linguistic devices to link or tie together textual units”

- Lexical cohesion:
  - Indicated by relations between words in the two units (identical word, synonym, hypernym)
    - Before winter I built a chimney, and shingled the sides of my house. I thus have a tight shingled and plastered house.

    - Peel, core and slice the pears and the apples. Add the fruit to the skillet.
Intuition: Cohesion

• Non-lexical: anaphora
  – The Woodhouses were first in consequence there. All looked up to them.
• Cohesion chain:
  – Peel, core and slice the pears and the apples. Add the fruit to the skillet. When they are soft…

• Note: cohesion is not coherence

Cohesion-Based Segmentation

• Sentences or paragraphs in a subtopic are cohesive with each other
• But not with paragraphs in a neighboring subtopic
• So, if we measured the cohesion between every neighboring sentences
  – We might expect a ‘dip’ in cohesion at subtopic boundaries.
TextTiling (Hearst ’97)

1. Tokenization
   - Each space-delimited word
   - Converted to lower case
   - Throw out stop list words
   - Stem the rest
   - Group into pseudo-sentences (windows) of length w=20

2. Lexical Score Determination: cohesion score
   Three part score including
   • Average similarity (cosine measure) between gaps
   • Introduction of new terms
   • Lexical chains

3. Boundary Identification
TextTiling Method

Cosine Similarity

$$\text{sim}_{\text{cosine}}(\vec{b}, \vec{a}) = \frac{\vec{b} \cdot \vec{a}}{|\vec{b}| |\vec{a}|} = \frac{\sum_{i=1}^{N} b_i \times a_i}{\sqrt{\sum_{i=1}^{N} b_i^2} \sqrt{\sum_{i=1}^{N} a_i^2}}$$
Vector Space Model
(recall distributional semantics, following IR)

• In the vector space model for TextTiling, both segments are represented as vectors
• Numbers are derived from the words that occur in the collection
  – Could count bits to get similarity, frequency, weights etc.
• But that favors long documents over shorter ones
  – Cosine to normalize dot product by vector lengths

Lexical Score Part 2: Introduction of New Terms
Lexical Score Part 3: Lexical Chains

![Diagram of lexical chains](image)

Figure 4: Results of the block similarity algorithm on the Starwars text. Internal numbers indicate paragraph numbers, y-axis indicates token-sequence gap number, z-axis indicates similarity between blocks centered at the corresponding token-sequence gap. Vertical lines indicate boundaries chosen by the algorithm; for example, the leftmost vertical line represents a boundary after paragraph 3. Note how these align with the boundary gaps of Figure 3 above.
Supervised Discourse segmentation

• Lexical features as before
• Discourse markers or cue words
  – Broadcast news
    • Good evening, I’m <PERSON>
    • …coming up….
  – Science articles
    • “First,…”
    • “The next topic…..”
  – Subtasks
    • disambiguating predefined connectives
    • learning domain specific markers

Supervised discourse segmentation

• Supervised machine learning
  – Label segment boundaries in training and test set
    • Easy to get automatically for very clear boundaries
      (paragraphs, news stories)
  – Extract features in training
  – Learn a (sequence) classifier
  – In testing, apply features to predict boundaries
Supervised discourse segmentation

• Evaluation: WindowDiff (Pevzner and Hearst 2000) assign partial credit

![Diagram of WindowDiff evaluation](image)

Some Local Work

• Cue phrase disambiguation (Hirschberg & Litman, 1993; Litman, 1996)
• Supervised linear discourse segmentation (Passonneau & Litman, 1997)
• Representation and application of hierarchical discourse segmentation
  – Mihai Rotaru, Computer Science, PhD 2008: Applications of Discourse Structure for Spoken Dialogue Systems (Litman)
• Measuring and applying lexical cohesion (Ward & Litman, 2008)
• Penn Discourse Treebank in argumentative essays and discussion forums
  – (Chandrasekaran et al., 2017; Zhang et al., 2016)

• Swapna Somasundaran, Computer Science, PhD 2010: Discourse-level relations for Opinion Analysis (Wiebe)
Part II of: What makes a text coherent?

- Appropriate sequencing of subparts of the discourse -
  - discourse/topic structure

- **Appropriate use of coherence relations**
  *between subparts of the discourse* --
  *rhetorical structure*

- Appropriate use of referring expressions

Text Coherence, again

The reason is that these utterances, when juxtaposed, will not exhibit coherence. Almost certainly not. Do you have a discourse? Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book.
Or…. Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book. Do you have a discourse? Almost certainly not. The reason is that these utterances, when juxtaposed, will not exhibit coherence.

Coherence

• John hid Bill’s car keys. He was drunk.

• ??John hid Bill’s car keys. He likes spinach.

• Again, not the same as cohesion.
Why Coherence? Summarization

Slide from Mirella Lapata and Regina Barzilay

Applications of Coherence Metrics

- **Text Generation:**
  - concept-to-text generation
  - summarization, text simplification
  - question answering, machine translation
  - essay grading
- **Text Understanding**
  - Improving coreference resolution
- **Potential uses**
  - automatic evaluation tool for text quality
  - during system development (avoids repeated human evaluations)
- **Software**
  - Brown Coherence Toolkit
  - Coh-Metrix: cohmctrx.memphis.edu/CohMetrixDemo/demo.htm
Hobbs ’79: Coherence Relations

• Result
  – Infer that the state or event asserted by S0 causes or could cause the state or event asserted by S1.

  The Tin Woodman was caught in the rain. His joints rusted.

• Explanation
  – Infer that the state or event asserted by S1 causes or could cause the state or event asserted by S0.

  John hid Bill’s car keys. He was drunk.
• Parallel
  – Infer \( p(a_1, a_2..) \) from the assertion of \( S_0 \) and \( p(b_1,b_2...) \) from the assertion of \( S_1 \), where \( a_i \) and \( b_i \) are similar, for all \( i \).

  The Scarecrow wanted some brains. The Tin Woodman wanted a heart.

• Elaboration
  – Infer the same proposition \( P \) from the assertions of \( S_0 \) and \( S_1 \).

  Dorothy was from Kansas. She lived in the midst of the great Kansas prairies.
Coherence Relations

John went to the bank to deposit his paycheck. (S1)
He then took a train to Bill’s car dealership. (S2)
He needed to buy a car. (S3)
The company he works for now isn’t near any public transportation. (S4)
He also wanted to talk to Bill about their softball league. (S5)

Rhetorical Structure Theory

- Another theory of discourse structure, based on identifying relations between segments of the text
  - **Nucleus/satellite** notion encodes asymmetry
    - Nucleus is thing that if you deleted it, text wouldn’t make sense.
  - Some rhetorical relations:
    - **Elaboration**: (set/member, class-instance, whole/part…)
    - **Contrast**: multinuclear
    - **Condition**: Sat presents precondition for N
    - **Purpose**: Sat presents goal of the activity in N
One Rhetorical Relation

- A sample definition
  - Relation: Evidence
  - Constraints on N: H might not believe N as much as S thinks s/he should
  - Constraints on Sat: H already believes or will believe Sat
  - Effect: H’s belief in N is increased

- An example:
  Kevin must be here.
  His car is parked outside.

Some Problems with RST

- How many Rhetorical Relations are there?
- How can we use RST in dialogue as well as monologue?
- Difficult to get annotators to agree on labeling the same texts
- Trees versus directed graphs?
Automatic Labeling / Discourse Parsing

- Supervised machine learning
  - Get a group of annotators to assign a set of relations to a text, and/or tree structure
  - Extract a set of surface features from the text that might signal the presence of the coherence relations
  - Train a supervised ML system based on the training set
  - Some publicly available resources (RST treebank, Penn Discourse Treebank) exist
- Semi and unsupervised approaches
- Inferential / Abduction (my dissertation!)

Penn Discourse Treebank

- The city’s Campaign Finance Board has refused to pay Mr. Dinkins $95,142 in matching funds because his campaign records are incomplete.
  - Causal relation
- So much of the stuff poured into Motorola’s offices that its mail rooms there simply stopped delivering it. Implicit = so. Now, thousands of mailers, catalogs and sales pitches go straight into the trash.
  - Consequence relation
Shallow Features

- Explicit markers / cue phrases: *because, however, therefore, then, etc.*
- Tendency of certain syntactic structures to signal certain relations:
  - Infinitives are often used to signal purpose relations: *Use rm to delete files.*

- Ordering
- Tense/aspect
- Intonation

Part III of: What makes a text coherent?
Entity-based Coherence

- Appropriate sequencing of subparts of the discourse - discourse/topic structure
- Appropriate use of coherence relations between subparts of the discourse -- rhetorical structure
- *Appropriate use of referring expressions*
Pronouns and Reference Resolution
A Reference Joke

Gracie: Oh yeah ... and then Mr. and Mrs. Jones were having matrimonial trouble, and my brother was hired to watch Mrs. Jones.

George: Well, I imagine she was a very attractive woman.

Gracie: She was, and my brother watched her day and night for six months.

George: Well, what happened?

Gracie: She finally got a divorce.

George: Mrs. Jones?

Gracie: No, my brother's wife.

Reference Resolution: Vocabulary

• Process of associating Bloomberg/he/his with particular person and big budget problem/it with a concept
  Giuliani left Bloomberg to be mayor of a city with a big budget problem. It’s unclear how he’ll be able to handle it during his term.

• Referring exprs.: Guilani, Bloomberg, a big budget problem, he, it, his

• Presentational it: non-referential

• Referents: the person named Guilani, the concept of a big budget problem
• Co-referring referring expressions: Guiliani, he, his
• Antecedent: Guiliani
• Anaphors: he, his

Discourse Models

• Needed to model reference because referring expressions (e.g. Guiliani, Bloomberg, he, it, budget problem) encode information about beliefs about the referent
• When a referent is first mentioned in a discourse, a representation is evoked in the model
  – Information predicated of it is stored also in the model
  – On subsequent mention, it is accessed from the model
Types of Referring Expressions

• Entities, concepts, places, propositions, events, ...

According to John, Bob bought Sue an Integra, and Sue bought Fred a Legend.
– But that turned out to be a lie. (a speech act)
– But that was false. (proposition)
– That struck me as a funny way to describe the situation. (manner of description)
– That caused Sue to become rather poor. (event)
– That caused them both to become rather poor. (combination of multiple events)

Reference Phenomena:
5 Types of Referring Expressions

• Indefinite NPs

A homeless man hit up Bloomberg for a dollar.
Some homeless guy hit up Bloomberg for a dollar.
This homeless man hit up Bloomberg for a dollar.

• Definite NPs

The poor fellow only got a lecture.

• Demonstratives

This homeless man got a lecture but that one got carted off to jail.

• Names

Prof. Litman teaches on Tuesday.
Pronouns

A large tiger escaped from the Central Park zoo chasing a tiny sparrow. It was recaptured by a brave policeman.

- Referents of pronouns usually require some degree of salience in the discourse (as opposed to definite and indefinite NPs, e.g.)
- How do items become salient in discourse?

Salience vs. Recency

E: So you have the engine assembly finished. Now attach the rope. By the way, did you buy the gas can today?
A: Yes.
E: Did it cost much?
A: No.
E: OK, good. Have you got it attached yet?
Reference Phenomena: Information Status

- Giveness hierarchy / accessibility scales …
- But complications

Inferables

- I almost bought an Acura Integra today, but a door had a dent and the engine seemed noisy.
- Mix the flour, butter, and water. Knead the dough until smooth and shiny.
Discontinuous Sets

- Entities evoked together but mentioned in different sentence or phrases
  
  John has a St. Bernard and Mary has a Yorkie. They arouse some comment when they walk them in the park.

Generics

I saw two Corgis and their seven puppies today. They are the funniest dogs
Constraints on Pronominal Reference

- Number agreement
  John’s parents like opera. John hates it/John hates them.
- Person agreement
  George and Edward brought bread. They shared it.

- Gender agreement
  John has a Porsche. He/it/she is attractive.
- Syntactic constraints
  John bought himself a new Volvo. (himself = John)
  John bought him a new Volvo (him = not John).
Preferences in Pronoun Interpretation

- Recency
  
  John bought a new boat. Bill bought a bigger one. Mary likes to sail it.

- But...grammatical role raises its ugly head...

  John went to the Acura dealership with Bill. He bought an Integra.
  Bill went to the Acura dealership with John. He bought an Integra.
  ?John and Bill went to the Acura dealership. He bought an Integra.

- And so does...repeated mention

  - John needed a car to go to his new job. He decided that he wanted something sporty. Bill went to the dealership with him. He bought a Miata.
  - Who bought the Miata?
  - What about grammatical role preference?

- Parallel constructions

  Saturday, Mary went with Sue to the farmer’s market.
  Sally went with her to the bookstore.
  Sunday, Mary went with Sue to the mall.
  Sally told her she should get over her shopping obsession.

- Selectional restriction

  John left his plane in the hangar.
  He had flown it from Memphis this morning
• Verb semantics/thematic roles

John telephoned Bill. He’d lost the directions to his house.

John criticized Bill. He’d lost the directions to his house.

Summary: What Affects Reference Resolution?

• Lexical factors
  – Reference type: Inferrability, discontinuous set, generics, one anaphora, pronouns,…
• Discourse factors:
  – Recency
  – Focus/topic structure, digression
  – Repeated mention
• Syntactic factors:
  – Agreement: gender, number, person, case
  – Parallel construction
  – Grammatical role
• Semantic/lexical factors
  – Selectional restrictions
  – Verb semantics, thematic role
Reference Resolution Algorithms

• Given these types of features, can we construct an algorithm that will apply them such that we can identify the correct referents of anaphors and other referring expressions?

Reference Resolution Task

• Finding in a text all the referring expressions that have one and the same denotation
  – Pronominal anaphora resolution
  – Anaphora resolution between named entities
  – Full noun phrase anaphora resolution
Issues

• Which constraints/features can/should we make use of?
• How should we order them? I.e. which override which?
• What should be stored in our discourse model? I.e., what types of information do we need to keep track of?
• How to evaluate?

Some Algorithms

• Hobbs ‘78: syntax tree-based referential search
• Centering: also entity-coherence
• Supervised learning approaches
Hobbs: Syntax-Based Approach

• Search for antecedent in parse tree of current sentence, then prior sentences in order of recency
  – For current S, search for NP nodes to the left of a path $p$ from the pronoun up to the first NP or S node (X) above it in L2R, breadth-first
    • Propose as pronoun’s antecedent any NP you find as long as it has an NP or S node between itself and X
    • If X is highest node in sentence, search prior sentences, L2R breadth-first, for candidate NPs
    • O.w., continue searching current tree by going to next S or NP above X before going to prior sentences

Centering Theory:
Grosz, Joshi and Weinstein, 1995

• The way entities are introduced and discussed influences coherence
• 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 (=topic or focus).
  – Coherent discourses have utterances with common centers.
• Entity transitions capture degrees of coherence
  – (e.g., in Centering theory CONTINUE > SHIFT).
• Unlike Hobbs, has discourse model
Claim: Entity coherence: 
Discourses without a clear ‘central entity’ feel less coherent

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.

Claim: Entity coherence: 
Discourses without a clear ‘central entity’ feel less coherent

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.
Concepts and definitions, I

- Every UTTERANCE U in a discourse (segment) DS updates the LOCAL FOCUS - a PARTIALLY RANKED set of discourse entities, or FORWARD-LOOKING CENTERS (CFs)
- An utterance U in discourse segment DS updates the existing CF set by replacing it with the set of CFs REALIZED in U, CF(U,DS) (usually simplified to CF(U))
- The most highly ranked CF realized in utterance U is CP(U)

<table>
<thead>
<tr>
<th>(1)</th>
<th>u1. Susan gave James a pet hamster.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CF(u1) = [Susan, James, pet hamster]. CP(u1) = Susan</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(2)</th>
<th>u2. She gave Peter a nice scarf.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CF(u2) = [Susan, Peter, nice scarf]. CP(u2) = Susan</td>
</tr>
</tbody>
</table>

Concepts and Definitions, II:
The CB

- The BACKWARD-LOOKING CENTER of utterance Uᵢ, 
  - CB(Uᵢ), 
  - is the highest-ranked element of CF(Uᵢ₋₁) that is realized in Uᵢ
The CB: Examples

(1)  u1. Susan gave James a pet hamster.
    CF(u1) = [Susan, James, pet hamster]. CB = undefined CP=Susan

(2)  u2. She gave Peter a nice scarf.
    CF(u2) = [Susan, Peter, nice scarf]. CB=Susan. CP=Susan

NB: The CB is not always the most ranked entity of the PREVIOUS utterance

(2') u2. He loves hamsters.
    CF(u2) = [James, hamsters]. CB=James. CP=James

… or the most highly ranked entity of the CURRENT one

(2'') u2. Peter gave her a nice scarf.
    CF(u2) = [Peter, Susan, nice scarf]. CB=Susan. CP=Peter

Claims of the theory: Local salience and pronominalization

- Grosz et al (1995): the CB is also the most salient entity. Texts in which other entities (but not the CB) are pronominalized are less felicitous

(1)  a. Something must be wrong with John.
    b. He has been acting quite odd.
    c. He called up Mike yesterday.
    d. John wanted to meet him quite urgently.

(2)  a. Something must be wrong with John.
    b. He has been acting quite odd.
    c. He called up Mike yesterday.
    d. He wanted to meet him quite urgently.
Rule 1

RULE 1: if any CF is pronominalized, the CB is.

Claims of the theory:
Preserving the ranking

• Discourses without a clear ‘central entity’ feel less coherent

(1) a. John went to his favorite music store to buy a piano.
   b. He had frequented the store for many years.
   c. He was excited that he could finally buy a piano.
   d. He arrived just as the store was closing for the day.

(2) a. John went to his favorite music store to buy a piano.
   b. It was a store John had frequented for many years.
   c. He was excited that he could finally buy a piano.
   d. It was closing just as John arrived.
Transitions

- Grosz et al.: utterances are easier to process
  - if they preserve CB of previous utterance or
  - if CB(U) is also CP(U).

**CONTINUE:** Ur is a continuation if CB(Ur) = CB(U_{r-1}),
and CB(Ur) = CP(Ur)

**RETAIN:** Ur is a retain if CB(Ur) = CB(U_{r-1}), but
CB(Ur) is different from CP(Ur)

**SHIFT:** Ur is a shift if CB(Ur) ≠ CB(U_{r-1})

Utterance classification

(0) u0. Susan is a generous person.
   CF(u0) = [ ] CB = CP = .
(1) u1. She gave James a pet hamster.
   CF(u1) = [ ]. CB = CP=

**CONTINUE:**

(2)
Utterance classification

(0) u0. Susan is a generous person.
   CF(u0) = [Susan] CB = undefined CP = Susan.
(1) u1. She gave James a pet hamster.
   CF(u1) = [Susan, James, pet hamster]. CB = Susan CP = Susan
   CONTINUE:

   (2) u2. She gave Peter a nice scarf.
   CF(u2) = [Susan, Peter, nice scarf]. CB = Susan CP = Susan
   CONTINUE
Utterance classification, II

(0) u0. Susan is a generous person.
   CF(u0) = [Susan] CB = undefined CP = Susan.
(1) u1. She gave James a pet hamster.
   CF(u1) = [Susan, James, pet hamster]. CB = Susan CP = Susan

SHIFT:

(2’) u2. He loves hamsters.
   CF(u2) = [James]. CB = James CP = James SHIFT

Utterance classification, III

(0) u0. Susan is a generous person.
   CF(u0) = [Susan] CB = undefined CP = Susan.
(1) u1. She gave James a pet hamster.
   CF(u1) = [Susan, James, pet hamster]. CB = Susan CP = Susan

RETAIN:

(2”) u2. Peter gave her a nice scarf.
   CF(u2) = [Peter, Susan, nice scarf]. CB = Susan CP = Peter RETAIN
Rule 2

RULE 2: (Sequences of) continuations are preferred over (sequences of) retains, which are preferred over (sequences of) shifts.

Summary of the claims

CONSTRAINT 1: All utterances of a segment except for the first have exactly one CB

RULE 1: If any CF is pronominalized, the CB is.

RULE 2: (Sequences of) continuations are preferred over (sequences of) retains, which are preferred over (sequences of) shifts.
Application – anaphora resolution

• Examples in prior slides assumed pronoun resolution, but can use the rules to disambiguate pronouns!
• Centering also used for coherence (entity grids, next)

Original centering theory

• Grosz et al do not provide algorithms for computing any of the notions used in the basic definitions:
  – UTTERANCE (clause? finite clause? sentence?)
  – PREVIOUS UTTERANCE
  – REALIZATION
  – RANKING
  – What counts as a ‘PRONOUN’ for the purposes of Rule 1? (Only personal pronouns? Or demonstrative pronouns as well? What about second person pronouns?)
• One of the reasons for the success of the theory is that it provides plenty of scope for theorizing …
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.
He is awaiting a hearing, his fate in the balance.
American scholars applauded the arrest.
Barzilay and Lapata: 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$_o$.

Constituent parse, then identification of subjects (s), objects (o), and other grammatical roles (x)

Barzilay and Lapata: The Entity Grid

1. Pinochet$_s$ London$_x$ October$_x$
2. Pinochet$_s$ surgery$_x$
3. arrest$_o$ 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$_o$

NPs represented by their head N
Barzilay and Lapata: The Entity Grid

1. Pinochet
2. London
3. October
4. Pinochet
5. hearing
6. scholars

Barzilay and Lapata: The Entity Grid

1. Pinochet
2. London
3. October
4. Pinochet
5. hearing
6. scholars
Barzilay and Lapata: The Entity Grid

<table>
<thead>
<tr>
<th></th>
<th>Pinochet</th>
<th>London</th>
<th>October</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>arrest</td>
<td></td>
<td>response</td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>Pinochet</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td>scholars</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Barzilay and Lapata: The Entity Grid

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

```
S X X
S -- X
-- S X X O
-- O
-- O
-- O X X S
```
A (simplified) parsed text and extracted grid

1. \[ \text{The Justice Department} \] is conducting an [anti-trust trial] against [Microsoft Corp.].
2. [Microsoft] is accused of trying to forcefully buy into [markets] where [its own products] are not competitive enough to unseat [established brands].
4. [Microsoft] claims its tactics are commonplace and good economically.

- A probabilistic model of particular entity transitions can then be trained.
- For example, the transitions \( \{X, O, S, S\} \) for the headword Microsoft exemplify that new entities are often introduced in non-subject position and only later appear in subject position.

Barzilay and Lapata: The Entity Grid

**Definition**

A local entity transition is sequence \( \{S, O, X, \ldots\}^n \) that represents entity occurrences and their syntactic roles in \( n \) adjacent sentences.

**Example (transitions of length 2)**

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>O</th>
<th>X</th>
<th>S</th>
<th>O</th>
<th>X</th>
<th>S</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_1 )</td>
<td>0</td>
<td>0</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
<td>0.07</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>0.02</td>
<td>0</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Feature vector document representation (row per grid)
Barzilay and Lapata: The Entity Grid

**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.
- Coreference resolution tool (Ng and Cardie, 2002).

**Syntax: Does syntactic knowledge matter?**
- Use four categories \{S, O, X, \text{–}\}.
- Reduce categories to \{X, \text{–}\}.

Supervised Anaphora Resolution

- **Input:** pronoun plus current and preceding sentences
- **Training:** hand-labeled corpus of positive examples plus inferred negative examples
- **Features**
  - Strict number
  - Compatible number
  - Strict gender
  - Compatible gender
  - Sentence distance
• Hobbs distance (noun groups between pronoun and candidate antecedent
• Grammatical role
• Linguistic form (proper name, definite, indefinite, pronominal antecedent)

– From pronouns to general coreference
  • Edit distance for full NPs

Summary: Reference Resolution

• Many approaches to reference resolution
• Use similar information/features but different methods
  – Hobbs’ Syntax
  – Centering coherence
  – Supervised and shallow methods use simple techniques with reasonable success
  – Also “deep” inferential approaches