Computational Discourse

11-711 Algorithms for NLP

8 December 2015

Introduction

- Discourse, monologue, dialogue, (conversation)
  - Discourse (*SLP* Ch. 21) vs. (Spoken) Dialogue Systems (*SLP* Ch. 24)
- "Longer-range" analysis (discourse) vs. "deeper" analysis (real semantics):
  - *John bought a car from Bill*
  - *Bill sold a car to John*
  - *They were both happy with the transaction*

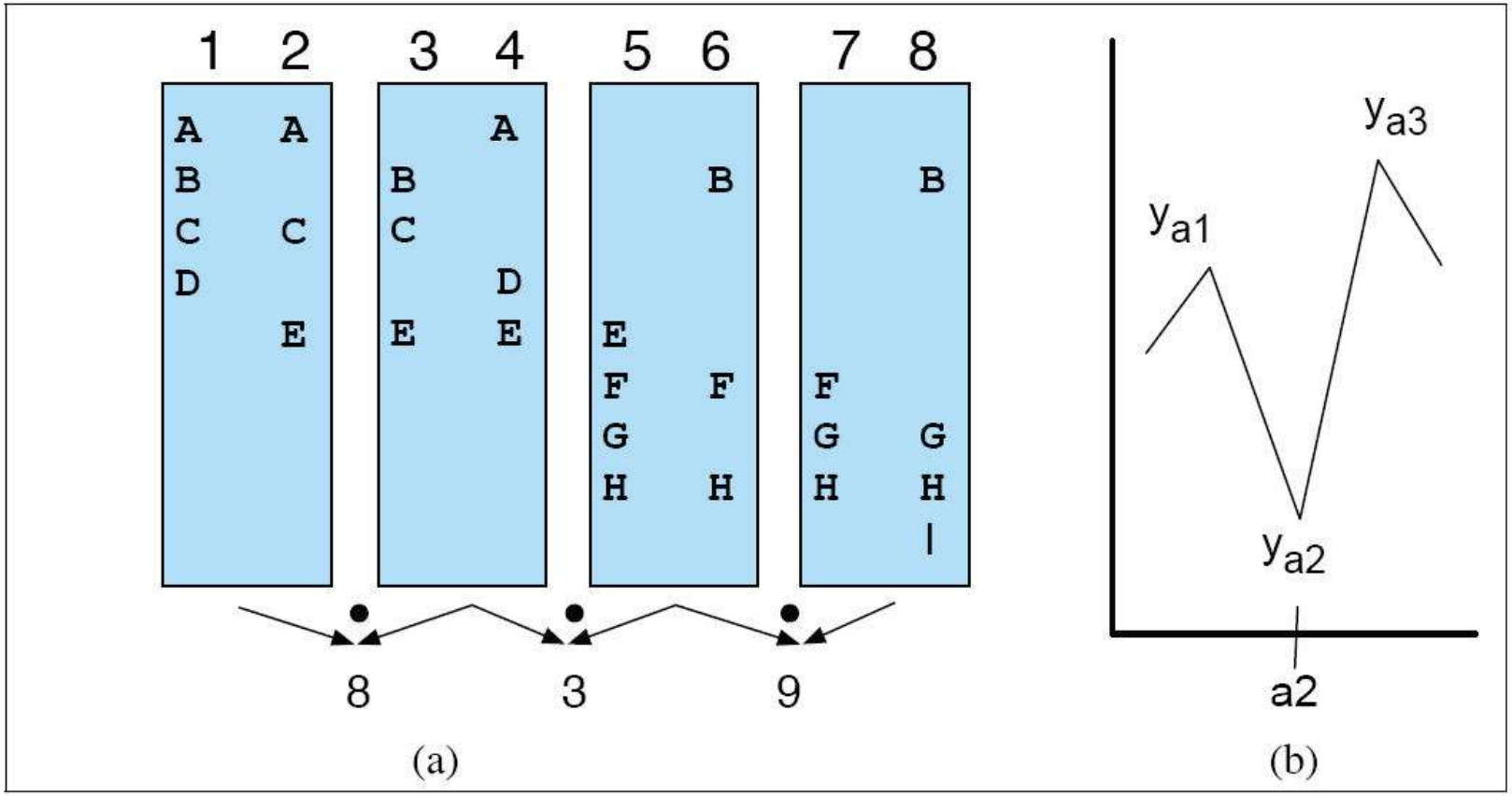
Coherence, Cohesion

- Coherence relations:
  - *John hid Bill's car keys. He was drunk.*
  - *John hid Bill's car keys. He likes spinach.*

- Entity-based coherence (Centering) and lexical cohesion:
  - *John went to the store to buy a piano*
  - *He had gone to the store for many years*
  - *He was excited that he could finally afford a piano*
  - *He arrived just as the store was closing for the day*

  versus
  - *John went to the store to buy a piano*
  - *It was a store he had gone to for many years*
  - *He was excited that he could finally afford a piano*
  - *It was closing for the day just as John arrived*

Discourse segmentation: TextTiling



- Using dips in **cohesion** to segment text.

Coherence Relations

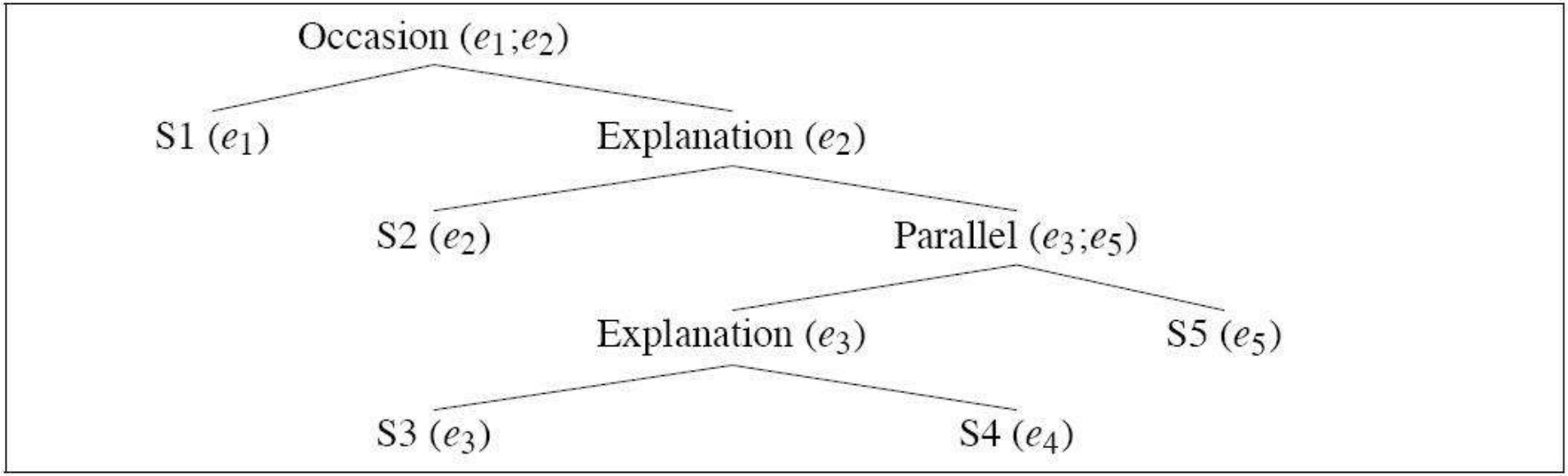
S1: *John went to the bank to deposit his paycheck*

S2: *He then took a bus to Bill's car dealership*

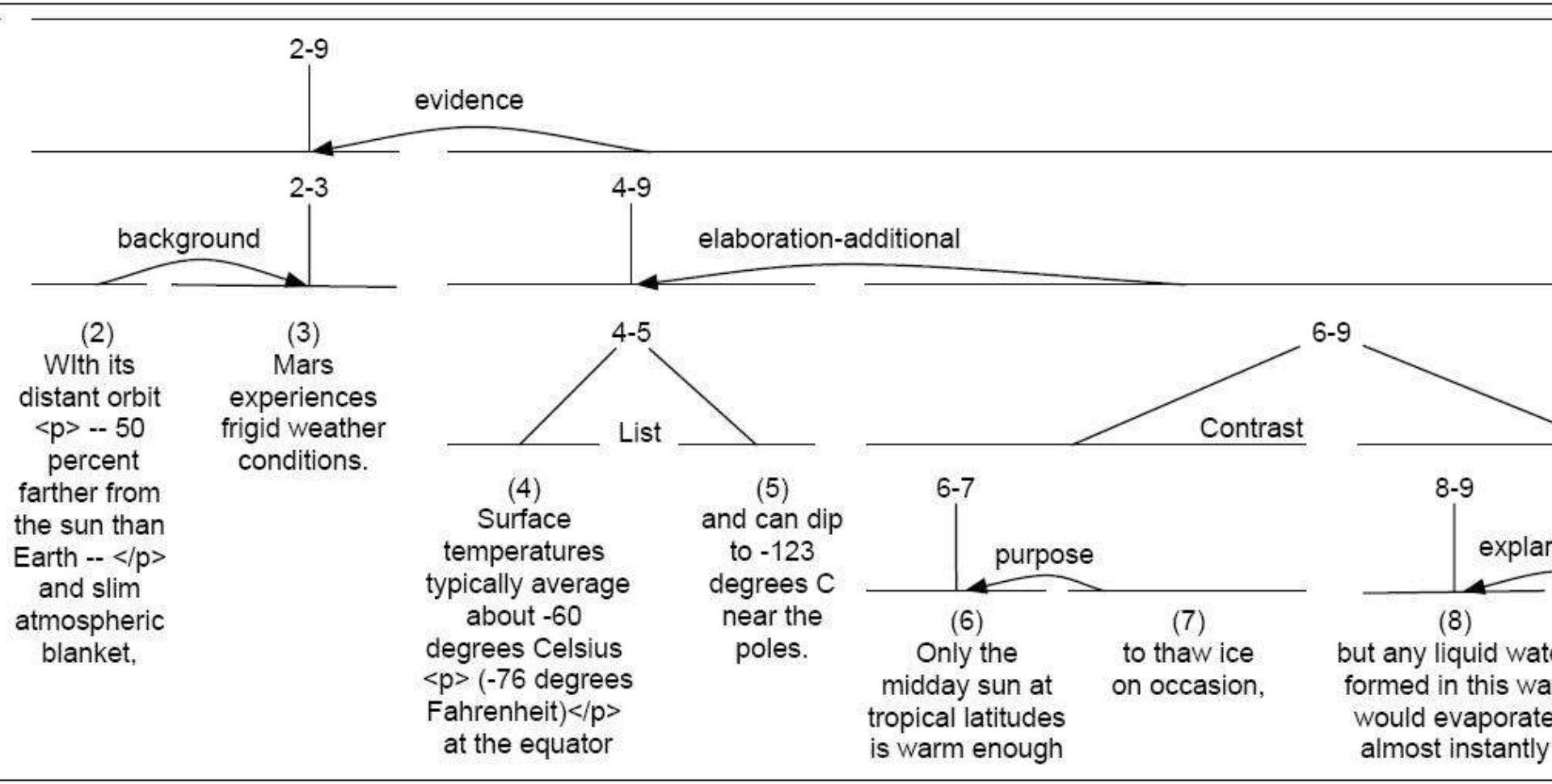
S3: *He needed to buy a car*

S4: *The company he works for now isn't near a bus line*

S5: *He also wanted to talk with Bill about their soccer league*



RST Coherence Relations



RST formal relation definition

- Relation name: **Evidence**
- Constr on N: R not believing N enough for W
- Constr on S: R believes S, or would
- Constr on N+S: R's believing S would increase R's believing N
- Effects: R's belief of N is increased

Automatic Coherence Assignment

- "Discourse parsing"?
- Use **cue phrases**/discourse markers
  - *although, but, because, yet, with, …*
  - but often implicit, as in car key example
- Use **abduction**, defeasible inference
  - All men are mortal
  - Max was mortal
  - **Maybe** Max was a man
- *The city denied the demonstrators a permit because they (feared/advocated) violence*

Reference Resolution: example

- Victoria Chen, CFO of Megabucks Banking Corp since 2004, saw her pay jump 20%, to $1.3 million, as the 37-year-old also became the Denver-based company's president. It has been ten years since she came to Megabucks from rival Lotsaloot.

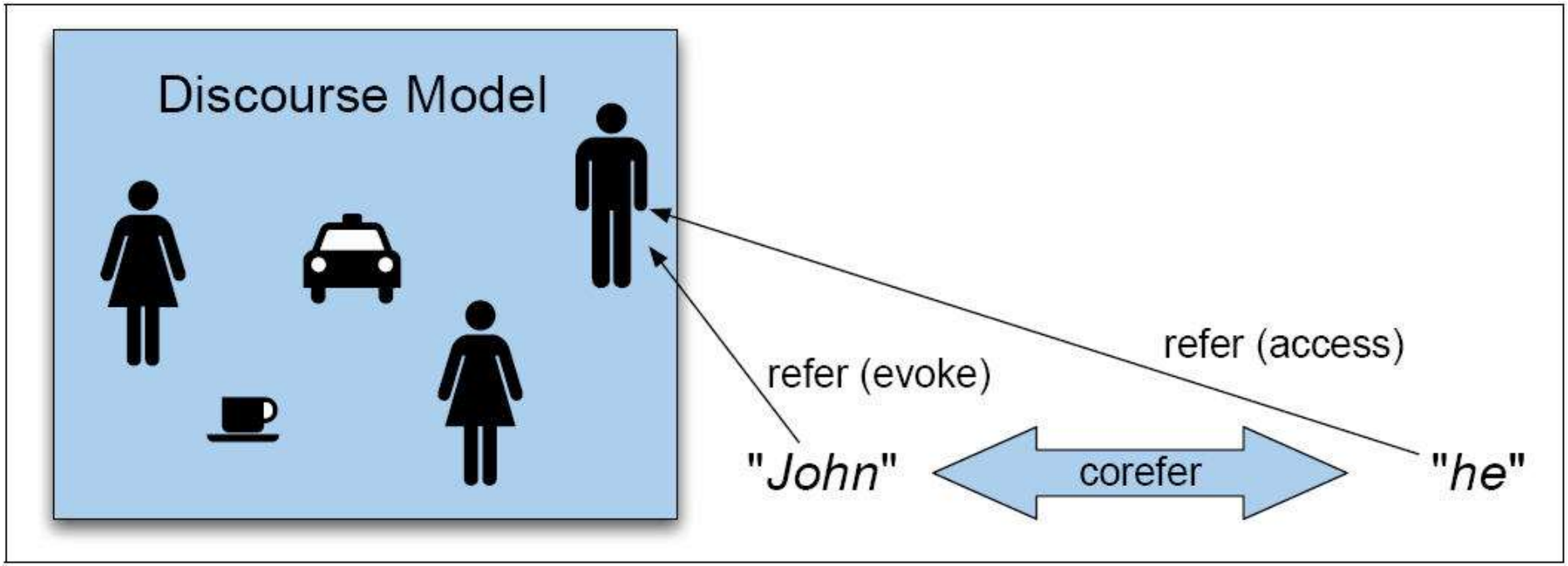
- Should give 4 coreference chains:
  - {Victoria Chen, CFO of Megabucks Banking Corp since 2004, her, the 37-year-old, the Denver-based company's president, she}
  - {Megabucks Banking Corp, the Denver-based company, Megabucks}
  - {her pay}
  - {Lotsaloot}

  .

Reference Resolution

- Determining the referent of a referring expression. Anaphora, antecedents corefer.
- 1961 Ford Falcon: *it, this, that, this car, the car, the Ford, the Falcon, my friend's car, …*
- Coreference chains are part of cohesion
- Note: other kinds of referents:
  - *According to Doug, Sue just bought the Ford Falcon*
    - *But **that** turned out to be a lie*
    - *But **that** was false*
    - ***That** struck me as a funny way to describe the situation*
    - ***That** caused a financial problem for Sue*

Discourse Models



- Discourse context, situational context

Types of Referring Expressions

- Indefinite NPs: *a/an, some, this*, or nothing
  - new entities; specific/non-specific ambiguity
- Definite NPs: usually *the*
  - an entity identifiable by the hearer
- Pronouns: *he, them, it*, etc. Also **cataphora**.
  - strong constraints on their use
  - can be bound: *Every student improved his grades*
- Demonstratives: *this, that*
- Names: construed to be unique, but they aren't
  - *Is that the Bob in LTI or the Bob in the Lane Center?*

Information structure: given/new

- *Where are my **shoes**? Your shoes are in the **closet***
- *What's in the **closet**?*
  - *??Your shoes are in the **closet**.*
  - *Your **shoes** are in the closet*.
- Definiteness/pronoun, length, position in S
- Inferrables: *Some car. … a door … the engine …*
- Generics: *At CMU you have to work hard.*
- Pleonastic/clefts/extraposition:
  - *It is raining. It was me who called. It was good that …*

Pronoun reference resolution: filters

- Agreement in number, person, gender
  - Pittsburgh dialect: *yinz=youse=y'all*
  - UK dialect: *Newcastle are a physical team*.
  - L can have >2 numbers, >3 persons, or >3 genders
- Binding theory: **reflexive** required/prohibited:
  - *John bought himself a new Ford.* [himself=John]
  - *John bought him a new Ford*. [him!=John]
  - *John said that Bill bought him a new Ford*. [him!=Bill]
  - *J said that B bought himself a new F*. [himself=Bill]
  - *He said that he bought J a new Ford*. [*both* he!=J]

Pronoun reference resolution: preferences

- Recency: preference for most recent referent
- Grammatical Role: subj>obj>others
  - *Billy went to the bar with Jim. He ordered rum*.
- Repeated mention: *Billy had been drinking for days. He went to the bar again today. Jim went with him. He ordered rum.*
- Parallelism: *John went with Jim to one bar. Bill went with him to another.*
- Verb semantics: *John phoned/criticized Bill. He lost the laptop.*
- Selectional restrictions: *John parked his car in the garage after driving it around for hours.*

Pronoun ref.res.: Hobbs Algorithm

- Algorithm for walking through parses of current and preceding sentences
- Simple, often used as baseline

- Requires parser, morph gender and number
  - plus head rules and WordNet for NP gender
- Implements binding theory, recency, and grammatical role preferences

Pronoun ref.res.: Centering theory

- Claim: a single entity is “centered” in each S
- Backward-looking center, Forward-looking centers
- C_b = most highly ranked C_f used from prev. S
- Rank: Subj>ExistPredNom>Obj>IndObj-Obl>DemAdvPP
- Defined transitions: (Cp is front of Cf list)

	$C_b(U_{n+1}) = C_b(U_n)$ or undefined $C_b(U_n)$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	Continue	Smooth-Shift
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	Retain	Rough-Shift

Rule 1: If any C_f used as Pro_{n+1}, then $C_{b(n+1)}$ must be Pro too

Rule 2: Rank: Continue>Retain>Smooth>Rough

U1: *John saw a Ford at the dealership*

Cb: NIL

Cf: John, Ford, dealership

U2: *He showed it to Bob* [Bob!=he]

He=John, it={Ford, dealership}

Cb=John

- (it->Ford) => Cf: {John,Ford,Bob} => CONTINUE [tie-winner]
- (it->dealership) => Cf: {John,dealer,Bob} => CONTINUE

U3: *He bought it* [*dealership* is now unavailable]

He={John,Bob}, it=Ford

- (he->John) => Cb=John, Cf={John,Ford} => CONTINUE [Win]
- (he->Bob) => Cb=Bob, Cf={Bob,Ford} => SMOOTH

Centering theory

- Same requirements as Hobbs
- Implements Grammatical Role, Recency, and Repeated Mention

- Can make mistakes:
  - *Bob opened a new dealership last week*
  - *John took a look at the Fords in his lot* [Cb=Bob]
  - *He ended up buying one*
    - He=Bob => CONTINUE, He=John => SMOOTH

Pronoun ref.res.: Log-linear model

- Supervised: hand-labelled coref corpus
- Rule-based filtering of non-referential pronouns
- Features, values for *He* in U3:

	He (U_2)	it (U_2)	Bob (U_2)	John (U_1)
strict number	1	1	1	1
compatible number	1	1	1	1
strict gender	1	0	1	1
compatible gender	1	0	1	1
sentence distance	1	1	1	2
Hobbs distance	2	1	0	3
grammatical role	subject	object	PP	subject
linguistic form	pronoun	pronoun	proper	proper

General Coreference Resolution

- Victoria Chen, CFO of Megabucks Banking Corp since 2004, saw her pay jump 20%, to $1.3 million, as the 37-year-old also became the Denver-based company's president. It has been ten years since she came to Megabucks from rival Lotsaloot.

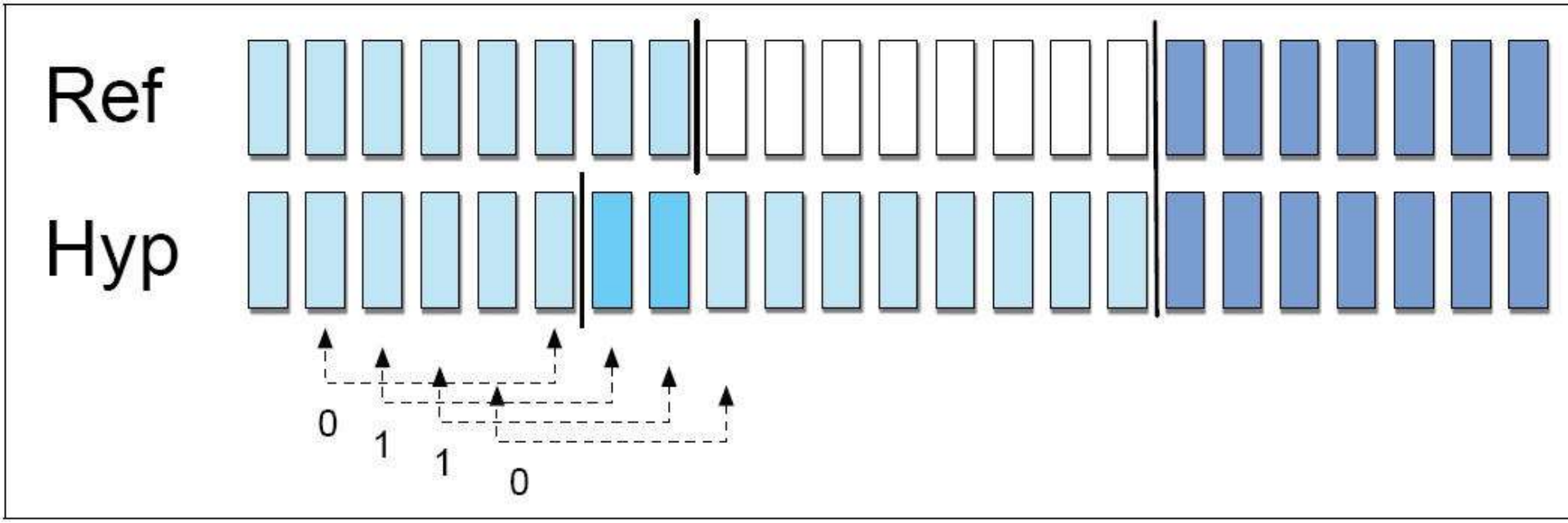
- Should give 4 coreference chains:
  - {Victoria Chen, CFO of Megabucks Banking Corp since 2004, her, the 37-year-old, the Denver-based company's president, she}
  - {Megabucks Banking Corp, the Denver-based company, Megabucks}
  - {her pay}
  - {Lotsaloot}

  .

General Coreference Resolution

- Can use a classifier to classify each pair of NPs as coreferent or not, trained from labelled corpus
- All the earlier features, plus:
  - anaphor edit distance
  - antecedent edit distance
  - alias (rule-based, per type, using NE tagger)
  - appositive
  - linguistic form: proper, def, indef, pronoun
- Combine best: ENCORE (Bo Lin et al 2010)
- ML for Cross-Doc Coref (Rushin Shah et al 2011)

Questions?


Ref
Hyp
0
1
1
0

$$S \rightarrow NP\ VP$$

$$NP \rightarrow \left\{ \begin{array}{l} (Det)\quad Nominal\quad \left(\left\{ \begin{array}{l} PP \\ Rel \end{array} \right\} \right)^* \\ \\ pronoun \end{array} \right\}$$

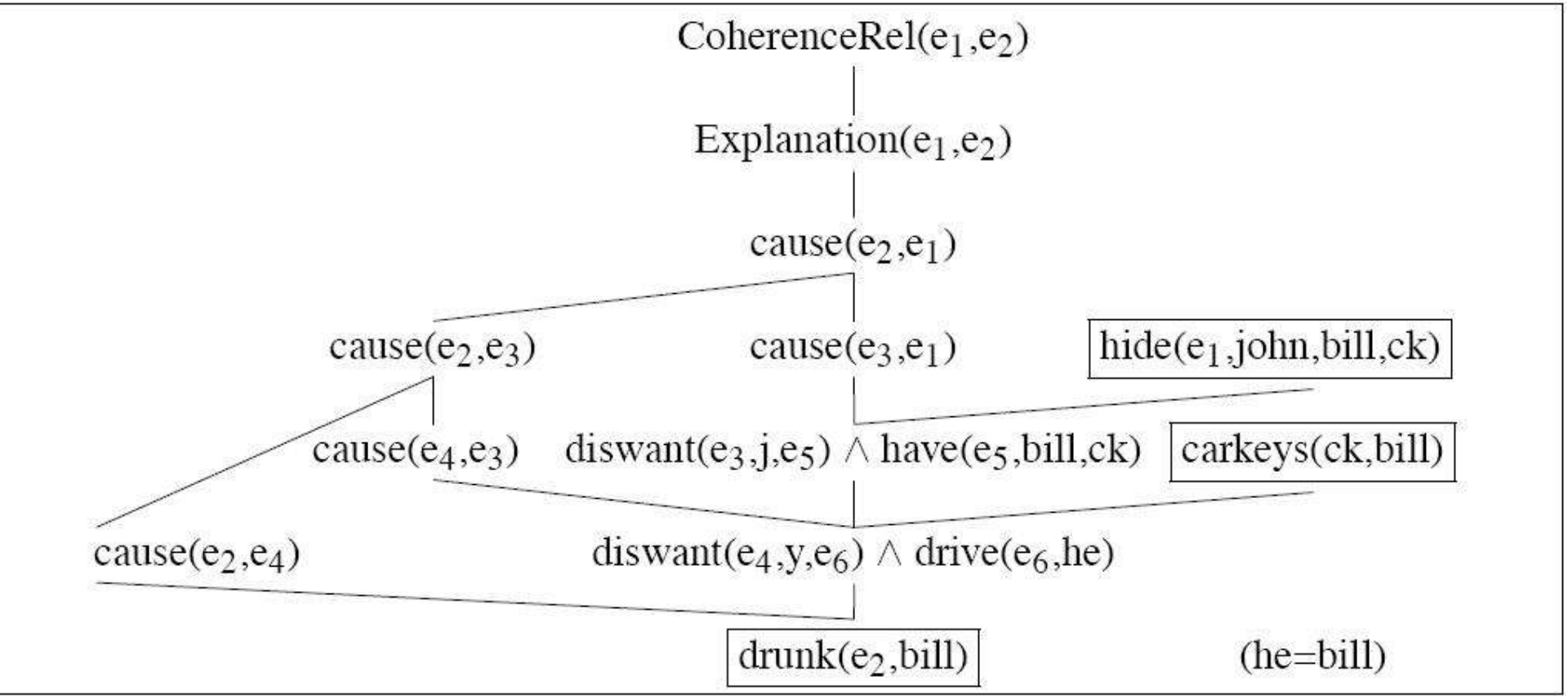
$$Det \rightarrow \left\{ \begin{array}{l} determiner \\ NP\ 's \end{array} \right\}$$

$$PP \rightarrow preposition\ NP$$

$$Nominal \rightarrow noun\ (PP)^*$$

$$Rel \rightarrow wh\text{-}word\ S$$

$$VP \rightarrow verb\ NP\ (PP)^*$$


CoherenceRel(e1,e2)
Explanation(e1,e2)
cause(e2,e1)
cause(e2,e3)
cause(e3,e1)
hide(e1,john,bill,ck)
cause(e4,e3)
diswant(e3,j,e5) ∧ have(e5,bill,ck)
carkeys(ck,bill)
cause(e2,e4)
diswant(e4,y,e6) ∧ drive(e6,he)
drunk(e2,bill)
(he=bill)

Evaluating Coreference Resolution

- B-CUBED:
  - Human-labelled “true” coreference chains
  - Compare hypothesis chains with true chains
  - Compute Precision and Recall for all entities, weighting each entity:
    - P: $\sum_{i=1}^{N} w_i \frac{\#\ correct\ in\ hypo\ chain\ containing\ entity\ i}{\#\ all\ in\ hypo\ chain\ containing\ i}$
    - R: $\sum_{i=1}^{N} w_i \frac{\#\ correct\ in\ hypo\ chain\ containing\ entity\ i}{\#\ all\ in\ ref\ chain\ containing\ i}$

- Or, don’t use a gold-standard: CONE (Bo Lin et al 2010)

- Truth: {E1-E5},{E6,E7},{E8-E12}
- Hypo1: {E1-E5},{E6-E12}
- Hypo2: {E1-E5,E8-E12},{E6,E7}

- Precision-oriented weighting:
  - weight is 1/(number-of-entities)

  H1: P= 1/12 * ((5*5/5)+(2*2/7)+(5*5/7)) = 0.76

  H2: P= 1/12 * ((5*5/10)+(2*2/2)+(5*5/10)) = 0.58
- Class-balancing weighting:
  - weight is 1/(chains-in-hypo*length-hypo-chain-of-entity)

  H1: P= $(\frac{1}{10}(5*5/5)+ \frac{1}{14}(2*2/7)+ \frac{1}{14}(5*5/7))$ = 0.796

  H2: P= $(\frac{1}{20}(5*5/10)+ \frac{1}{4}(2*2/2)+ \frac{1}{20}(5*5/10))$ = 0.75

- (from B-CUBED paper, Baldwin et al)