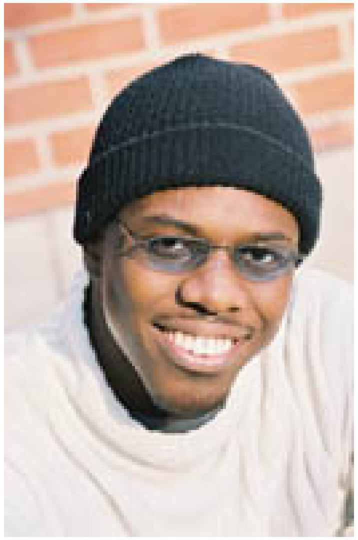
Given/new information and the discourse coherence problem

Micha Elsner

joint work with:

Eugene Charniak
Joseph Browne

BLIP BROWN



Given/new information

- Unfamiliar information:
  - **Sir Walter Elliot, of Kellynch Hall, in Somersetshire**, was a man who... never took up any book but the Baronetage...
- Now it's familiar:
  - **Sir Walter** had improved it...
- We also care about salience:
  - **He** had been remarkably handsome in his youth.

Prince '81

Discourse coherence problem

- Relationship between sentences in a discourse.
  - Earlier sentences make later ones more intelligible.

He had been remarkably handsome.
Sir Walter had improved it.
Sir Walter Elliot, of Kellynch Hall, in Somersetshire never took up any book but the Baronetage.

X

Useful for generation, summarization, &c.
Insights for pragmatics (coreference, importance and temporal order of events).

Discriminative task

- Binary judgement between random permutation and original document.

Sentence 2
Sentence 1
Sentence 4
Sentence 3

vs

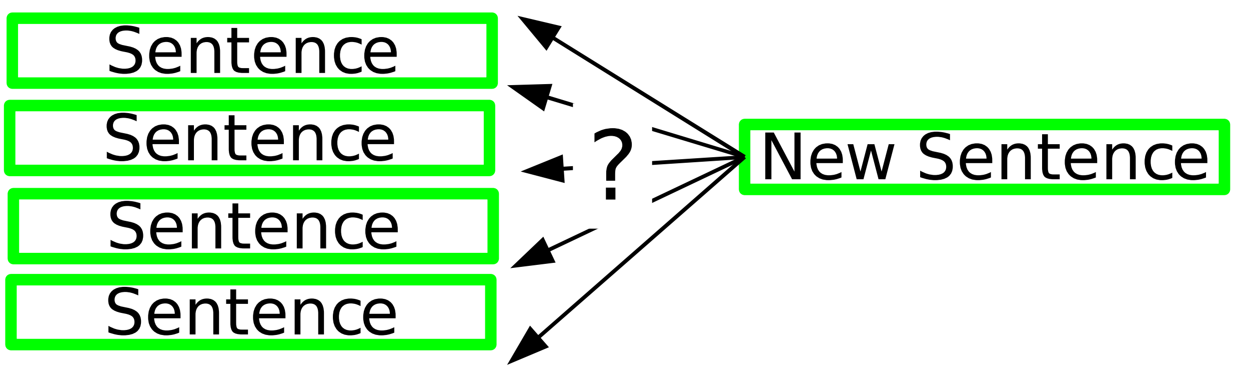
Sentence 1
Sentence 2
Sentence 3
Sentence 4

- Fast, convenient test.
- Longer documents are *much* easier!
- F-score (classifier can abstain).

Barzilay+Lapata '05

Insertion task

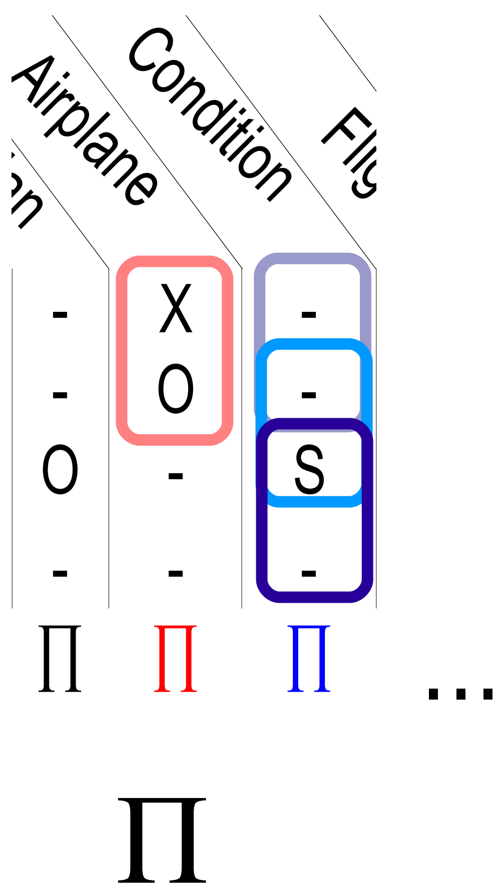
- Remove and re-insert one sentence at a time.
- Examines permutations closer to the original ordering.
  - Hard even for long documents.



Chen+Snyder+Barzilay '07
Elsner+Charniak '07

Baseline (Entity Grid)

- Entity grid: repeated nouns



- Deals only with **previously given** information and **salience.**
  - Nothing to say about **new** information.

disc (F)	ins (prec)
73.2	18.1

Lapata+Barzilay '05

Models

- Noun phrase syntax (NP)
- Pronoun coreference (Prn)
- Quotations (Qt)

	disc (F)	Ins (prec)
Entity Grid (Baseline)	73.2	18.1
EG, NP, Prn, Qt	**78.7**	**23.9**

- Inferrables (Ongoing work)

Anatomy of an unfamiliar NP

Sir Walter Elliot, of Kellynch Hall, in Somersetshire,

was a man who...

- Lots of linguistic markers to introduce this guy...
  - because you don't know who he is.

Anatomy of an unfamiliar NP

full name and title

Sir Walter Elliot, of Kellynch Hall, in Somersetshire,

was a man who...

- Lots of linguistic markers to introduce this guy...
  - because you don't know who he is.

Anatomy of an unfamiliar NP

full name and title

long phrasal modifier

Sir Walter Elliot, of Kellynch Hall, in Somersetshire,

was a man who...

- Lots of linguistic markers to introduce this guy...
  - because you don't know who he is.

Anatomy of an unfamiliar NP

full name and title

long phrasal modifier

Sir Walter Elliot, of Kellynch Hall, in Somersetshire,

was a man who...

copular verb

- Lots of linguistic markers to introduce this guy...
  - because you don't know who he is.

Lots of features!

- **Appositives**: Mr. Shepherd, *a civil, cautious lawyer...*
- **Restrictive relative clauses**: *the first man to...*
- **Syntactic position**: subject, object &c
- **Determiner / quantifier**: *a* (new), *the* (complicated!)
- **Titles and abbreviated titles**:
  – *Sir*, *Professor* (usually new); *Prof.*, *Inc.* (usually old)
- **How many modifiers?:** More implies newer.

- Most important feature: **same head occurred before?**

Vieira+Poesio '00
Ng+Cardie '02
Uryupina '03 ...

Previous work (linguistics)

- When **can** we use “the” (a, this, that...&c)?
  - Linguists (Hawkins '78, Gundel '93 and others)
  - A question of **rules**.
- When **do** we use:
  - Relatives (Fox+Thompson '90)
  - Various modifiers (Fraurud '90, Vieira+Poesio '98, Nenkova+McKeown '03 and others)
  - A question of **typicality**.

Previous work (classifiers)

- Used for coreference resolution:
  - Don't resolve the **new NPs**.
  - Do resolve the **old ones**.

Joint decisions:
Denis+Baldridge '07

Sequential:
Poesio+al '05
Ng+Cardie '02

- Almost any machine learning algorithm available...
- But they all score about 85%.

Modeling coherence

Sir Walter Elliot, of Kellynch Hall, in Somersetshire

he

his

Walter Elliot

Sir Walter

himself

Sir Walter

Sir Walter Elliot

VS

Sir Walter

he

his

Sir Walter

Walter Elliot

Sir Walter Elliot, of Kellynch Hall, in Somersetshire

himself

Sir Walter Elliot

Now some computation...

P(**Sir Walter Elliot, of Kellynch Hall, in Somersetshire** , **new**)

P(**he** , **old**)

P(**his** , **old**)

P(**Walter Elliot** , **old**)

P(**Sir Walter**, **old**)

P(**himself** , **old**)

P(**Sir Walter**, **old**)

P(**Sir Walter Elliot** , **old**)

Using a **generative** system,
P(syntax , label).

Where do the labels come from?
Full coreference!

$$P(\text{chain}) = \Pi\ P(\text{np})$$

$$P(\text{doc}) = \Pi\ P(\text{chain})$$

Full coreference is hard!

- For a disordered document, it's harder.
  – (I'll talk more about this later).
- We use 'same head' heuristic to fake coreference.
  – Works about 2/3 of the time (Poesio+Vieira).
  – Means we can't use the same head feature to build the classifier.

More realistic computation...

P(Sir Walter Elliot, of Kellynch Hall, in Somersetshire , new)

P(Walter Elliot , old)

P(Sir Walter Elliot , old)

One coreferential chain turns into two.
(Bad, but surviveable.)

P(Sir Walter, new)

P(Sir Walter, old)

P(he , old)

P(his , old)

P(himself , old)

And what about the pronouns?
We'll come back to them later.

What else can go wrong?

- Not all new NPs are unfamiliar.
  - Unique referents: **The FBI, the Golden Gate Bridge, Thursday**
  - Our technique will mislabel these.
- We can reduce error by distinguishing three classes: **new**, **old**, **singleton**
  - **singleton**: no subsequent coreferent NPs
  - often look more like old than new

corpus study: Fraurud '90
classifiers: Bean+Riloff '91
Uryupina '03

Results

- Combine systems by multiplication...
  - to construct a joint generative model.
  - Principled, but mixtures might improve?

	disc (F)	ins (prec)
Entity Grid	73.2	18.1
NP syntax	**72.7**	**16.7**
EG, NP	**77.6**	**21.5**

Generative classifier

- Distribution over P(syntax, label)
  - P(label) P(syntax | label)
  - Modifiers generated by Markov chains.
- State-of-the-art performance!
  - As a classifier.
  - And as a coherence model.
- Took a fair amount of time to develop, though.

For the lazy among us...

- We can also use a **conditional** system:
  - P(chain) = Π P(syntax , label)
  - Π P(label | syntax) P(syntax)
    - But different permutations of the document contain the same NPs, so...
      Π P(syntax) is a constant!
  - P(chain) ~ Π P(label | syntax)
- Logistic regression, max-ent...
  - Can't use non-probabilistic systems (boosting, SVM).

Pronoun coreference

- Pronouns occur close after their antecedent nouns.

Marlow sat cross-legged right aft, leaning against the mizzen-mast.
He had sunken cheeks, a yellow complexion, a straight back, an ascetic aspect, and... resembled an idol.
The **director**, satisfied the anchor had good hold, made his way aft and sat down amongst us.
We exchanged a few words lazily. Afterwards there was silence on board the yacht. For some reason or other we did not begin that game of dominoes. We felt meditative, and fit for nothing but placid staring. The day was ending in a serenity of still and exquisite brilliance.

Pronoun coreference

- Pronouns occur close after their antecedent nouns.

Marlow sat cross-legged right aft, leaning against the mizzen-mast.
He had sunken cheeks, a yellow complexion, a straight back, an ascetic aspect, and... resembled an idol.
The **director**, satisfied the anchor had good hold, made his way aft and sat down amongst us.
We exchanged a few words lazily. Afterwards there was silence on bo
did not begi
and fit for nothing but placid staring. The day was ending in a serenity of still and exquisite brilliance.

No possible antecedents here!

Violations cause incoherence

Marlow sat cross-legged right aft, leaning against the mizzen-mast.

The **director**, satisfied the anchor had good hold, made his way aft and sat down amongst us.

We exchanged a few words lazily. Afterwards there was silence on board the yacht. For some reason or other we c

felt medit

staring. The day was ending in a serenity of still and exquisite brilliance.

No possible antecedents here!

He had sunken cheeks, a yellow complexion, a straight back, an ascetic aspect, and... resembled an idol.

What sort of a model?

- Typical coreference models are conditional: P(antecedent | text)

Marlow sat ...

P(Marlow | he) = .99

He had sunken cheeks...

- Probability of linking the pronoun to each available referent.
- High for unambiguous texts...

What sort of a model?

- Typical coreference models are conditional: P(antecedent | text)

Marlow sat ...

P(Marlow | he) = .99 (still!)

We exchanged a few **words** lazily.

P(words | he) ≈ 0

There was **silence** on board the **yacht**.

P(yacht | he) ≈ 0

He had sunken cheeks...

Generative coreference

- Not only tell good *coreference assignments* from bad ones...
- But good *texts* from bad ones.
  - So we need P(text | antecedent)
- Luckily we can do that (sort of)...
  - Ge+Hale+Charniak '98
  - Accuracy 79.1% (on markables)

The probability of an Antecedent and the Pronoun given the Antecedent

Probability that the antecedent is a given how far away a is, and how often it has been mentioned

$$P_p(A=a, S_i | S_{i-1} S_{i-2}) = P(A=a | dist(a), mentions(a)) \cdot$$
$$P(gender(pronoun) | a) \cdot P(number(pronoun) | a)$$

The probability of an Antecedent and the Pronoun given the Antecedent

Probability that the antecedent is a given how far away a is, and how often it has been mentioned

$$P_p(A=a, S_i | S_{i-1} S_{i-2}) = P(A=a | dist(a), mentions(a)) \cdot P(gender(pronoun) | a) \cdot P(number(pronoun) | a)$$

Probability of the pronoun gender given the antecedent.

The probability of an Antecedent and the Pronoun given the Antecedent

Probability that the antecedent is a given how far away a is, and how often it has been mentioned

$$P_p(A=a, S_i | S_{i-1} S_{i-2}) = P(A=a | dist(a), mentions(a)) \cdot P(gender(pronoun) | a) \cdot P(number(pronoun) | a)$$

Probability of the pronoun gender given the antecedent.

Probability of the pronoun number given the antecedent.

The probability of an Antecedent and the Pronoun given the Antecedent

Probability that the antecedent is a given how far away a is, and **how often it has been mentioned**

$$P_p(A=a, S_i | S_{i-1} S_{i-2}) = P(A=a | dist(a), mentions(a)) \cdot$$
$$P(gender(pronoun)|a) \cdot P(number(pronoun)|a)$$

Not a Markov chain!
So no dynamic program to sum over all possible antecedents...

Intractability

- Best order: maximum probability of the document (summing over coreference):

$$P_p(D) = \sum_a P_p(A=a, D)$$

- Exponential sum over structures.

$$P_p(D) \approx \text{argmax}_a P(A=a, D)$$

- Solve this greedily.
  - Usually one structure has all the mass anyway.

Results (part II)

- Improvements continue...
  - On its own, this model is not as strong as the syntactic one.

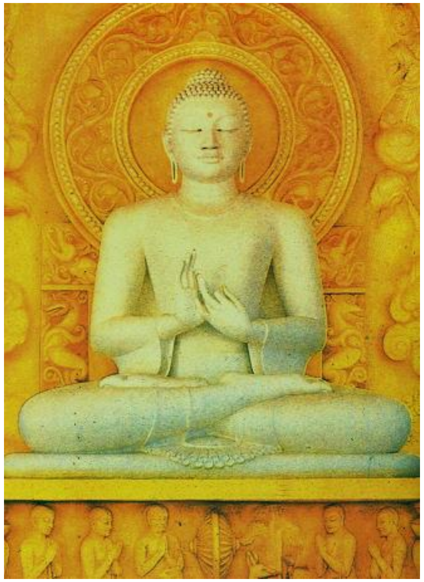
	disc (F)	ins (prec)
Entity Grid	73.2	18.1
Pronoun	**63.1**	**13.9**
EG, NP	77.6	21.5
EG, NP, Prn	**78.2**	**22.7**

Pipe dreams...

- Pronouns can find referents nearly anywhere...

Marlow sat cross-legged right aft.
He resembled an idol.
The **director** made his way aft.

Marlow sat cross-legged right aft.
The **director** made his way aft.
He resembled an idol.



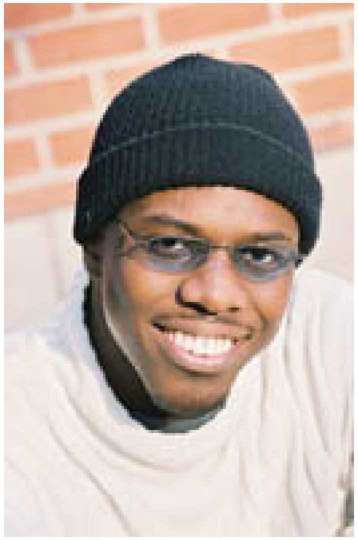
- Semantics could disambiguate:
  - Not all the cases are this hard.
  - But so far, no advantage.

More pipe dreams!

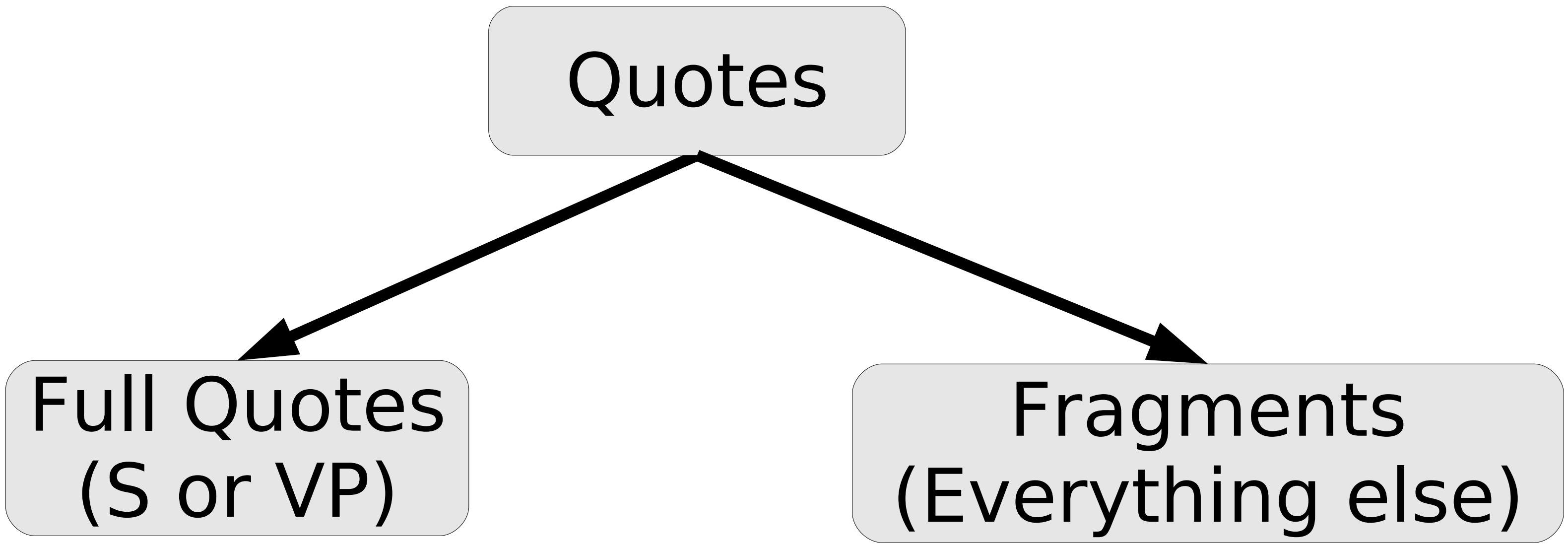
- Full coreference?
  - A generative model now exists: Haghighi+Klein '07 (non-parametric Bayes)
- An “easy” first step:
  - Model the decision to generate pronoun or full NP.
  - Doesn't work! We don't know why...

Quotations

- Some easy typographical stuff:
  - Open quote “ comes before close quote “
  - The stuff inside should be relatively short.
  - We can model this...
- More interesting aspects as well...
  - Based on discourse patterns.
  - *Not* just typography!

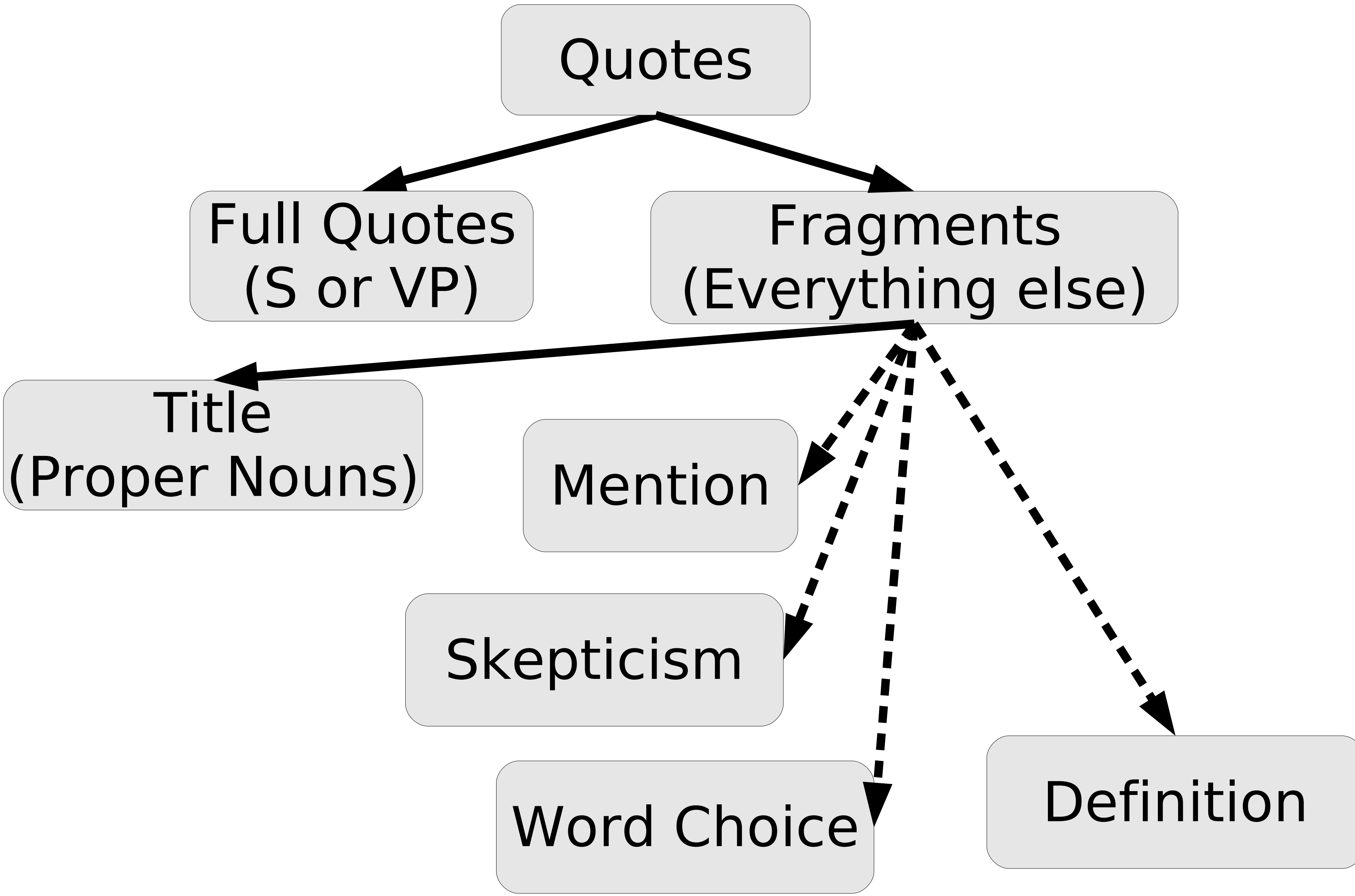


Types of quote



- Full quotes:
  - Almost always “real” speech.
  - Unlikely in first sentence.
- Quote fragments are more complicated...

Types of quote



“Definitional” quotes

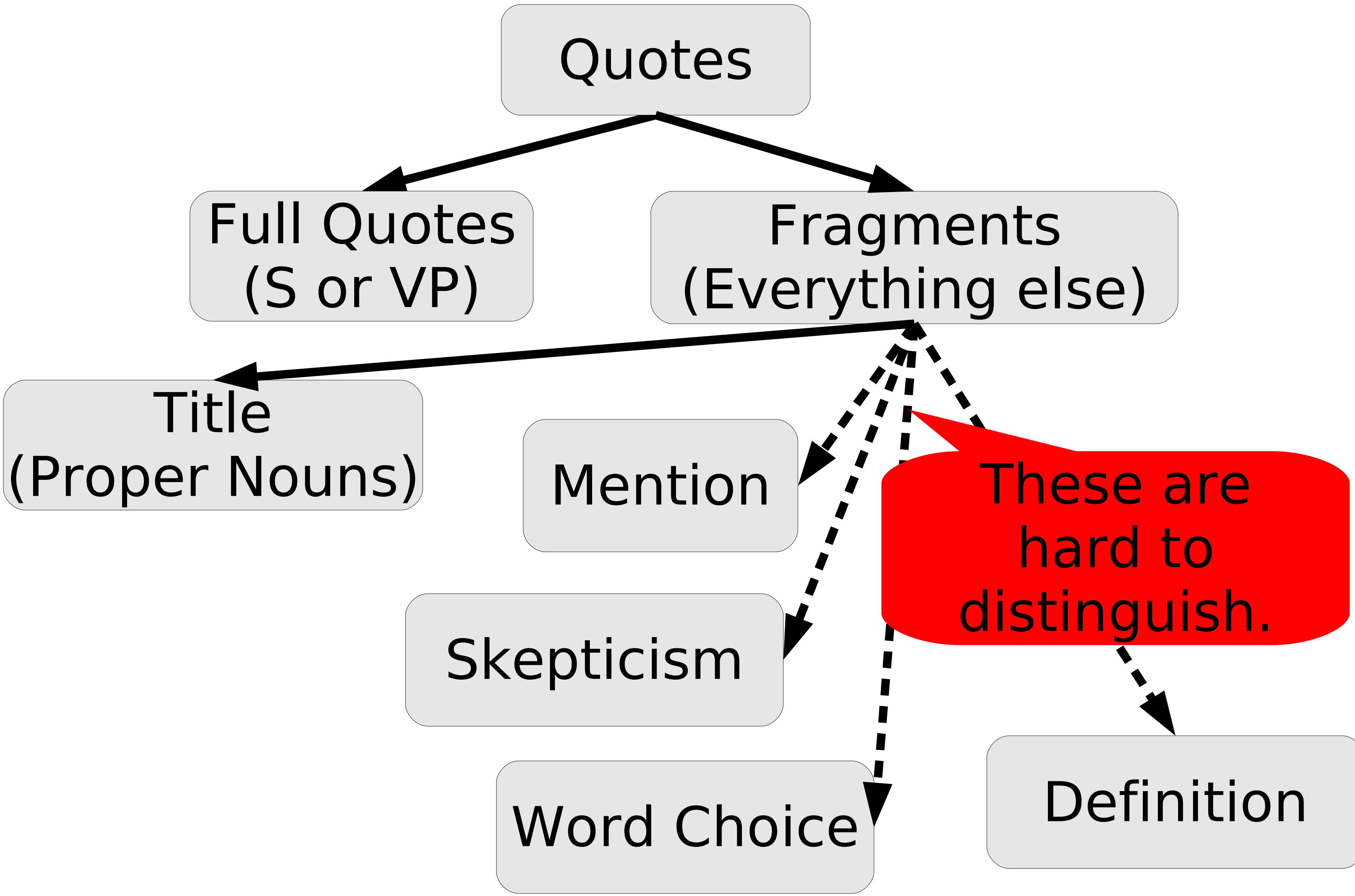
- Used to *define* an unfamiliar word.
- A giant “laser”...

- When you've defined the term, you should stop quoting it.
  - Dr. Evil doesn't do this, which is part of the joke.

Definitional quotes

- Another newness marker.
  - Works for things other than nouns.
  - “recombinant” DNA
  - The Fed appears to be “sterilizing” the intervention.
- Not a new entity, but a new piece of language.
- But we can be fooled…

Types of quote



Other uses for fragment-quotes

- Call attention to word choice:
  - Bush called Mr. Clymer a “major league asshole”.
- Mention rather than use:
  - “You” is a second person pronoun.
- Express skepticism or contempt:
  - Yeah, that's really “helpful”!
- Mark a title:
  - Chaucer's “Book of the Duchess”

Results (part III)

	disc (F)	ins (prec)
Entity Grid	73.2	18.1
Quotes	**38.1**	**?**
EG, NP	77.6	21.5
EG, NP, Prn	78.2	22.7
EG, NP, Prn, Qt	**78.7**	**23.9**

- Poor results are deceptive:
  - Precision 92, recall 24
  - Works well, but only on a few documents.

Conclusion

	disc (F)	ins (prec)
Entity Grid	73.2	18.1
EG, NP	77.6	21.5
EG, NP, Prn	78.2	22.7
EG, NP, Prn, Qt	**78.7**	**23.9**

- Given-new information leads to a series of improvements.

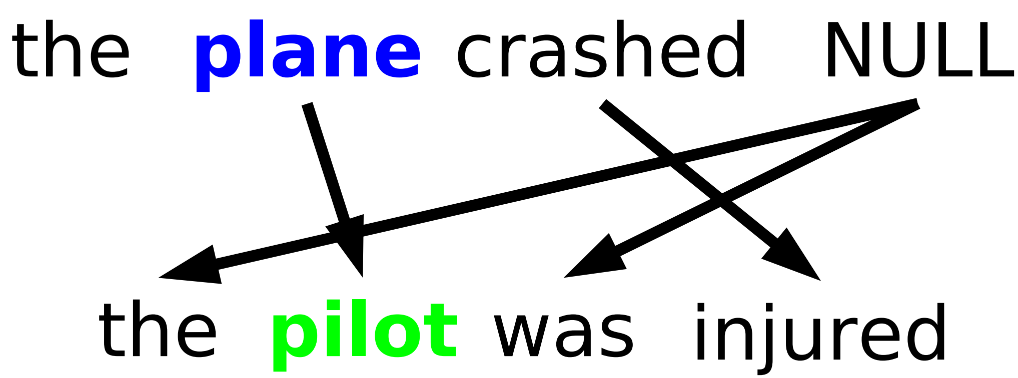
Context-dependent NPs

- The classic *inferrable* (Prince '81)
  - **The plane** crashed. **The pilot** was injured.
  - Looks like a familiar (discourse-old) NP.
  - But really a new entity.
  - Similar to unique NPs (**the FBI**), but licensed by a previous **anchor** (or **target**).
- Looser than coreference, tighter than topic similarity.

Poesio+Vieira+Teufel '97
Poesio+al '04

Alignment models

- IBM model 1: align each new word with a context word.
  - Soricut+Marcu '06, related to Lapata '03



Some preliminary results

- Max-probability words generated by:

airplanes	author	accident
land	book	technology
use	friends	site
restaurants	death	clients
priority	wife	neuromri
transportation	writer	radiologists
planes	life	life
experiences	readers	time
industry	interviews	home
enticements	part	reporters
airports	story	investigation

More preliminary results

- Syntactically biased alignment function:
  - Ex: words prefer to align to subjects.
  - Biases learned during EM (IBM model 2).

	disc (F)
Entity Grid	73.2
IBM model 1	71.8
Syntactic bias	74.4
Bias, 2 prev ss	76.3

Thanks!

- Regina Barzilay, Erdong Chen
- Olga Uryupina
- all of BLLIP
- DARPA GALE
- Everyone here!

Code is available:
http://www.cs.brown.edu/people/melsner