CS447: Natural Language Processing

http://courses.engr.illinois.edu/cs447

Lecture 25: A very brief introduction to discourse

Julia Hockenmaier

juliahmr@illinois.edu

3324 Siebel Center

Discourse

What is discourse?

On Monday, John went to Einstein's. He wanted to buy lunch. But the cafe was closed. That made him angry, so the next day he went to Green Street instead.

'Discourse':
any linguistic unit that consists of **multiple sentences**

Speakers describe "some situation or state of the real or some hypothetical world" (Webber, 1983)

Speakers attempt to get the **listener**
to construct a similar **model of the situation**.

Why study discourse?

For natural language **understanding:**

Most information is not contained in a single sentence.

The system has to **aggregate** information
across sentences, paragraphs or entire documents.

For natural language **generation:**

When systems generate text, that text needs to be easy to understand — it has to be **coherent**.

What makes text coherent?

How can we understand discourse?

On Monday, John went to Einstein's. He wanted to buy lunch. But the cafe was closed. That made him angry, so the next day he went to Green Street instead.

Understanding discourse requires (among other things):

1) doing **coreference** resolution:

'the cafe' and '*Einstein's'* **refer to the same entity**

He and *John* refer to the same person.

That refers to *'the cafe was closed'*.

2) identifying **discourse ('coherence') relations**:

'He wanted to buy lunch' is the ***reason*** for

'John went to Bevande.'

Discourse models

An explicit representation of:

— the **events and entities**
that a discourse talks about

— the **relations** between them
(and to the real world).

This representation is often written
in some form of logic.

What does this logic need to capture?

Discourse models should capture...

Physical entities: John, Einstein's, lunch

Events: On Monday, John went to Einstein's

involve entities, take place at a point in time

States: It was closed.

involve entities and hold for a period of time

Temporal relations: afterwards

between events and states

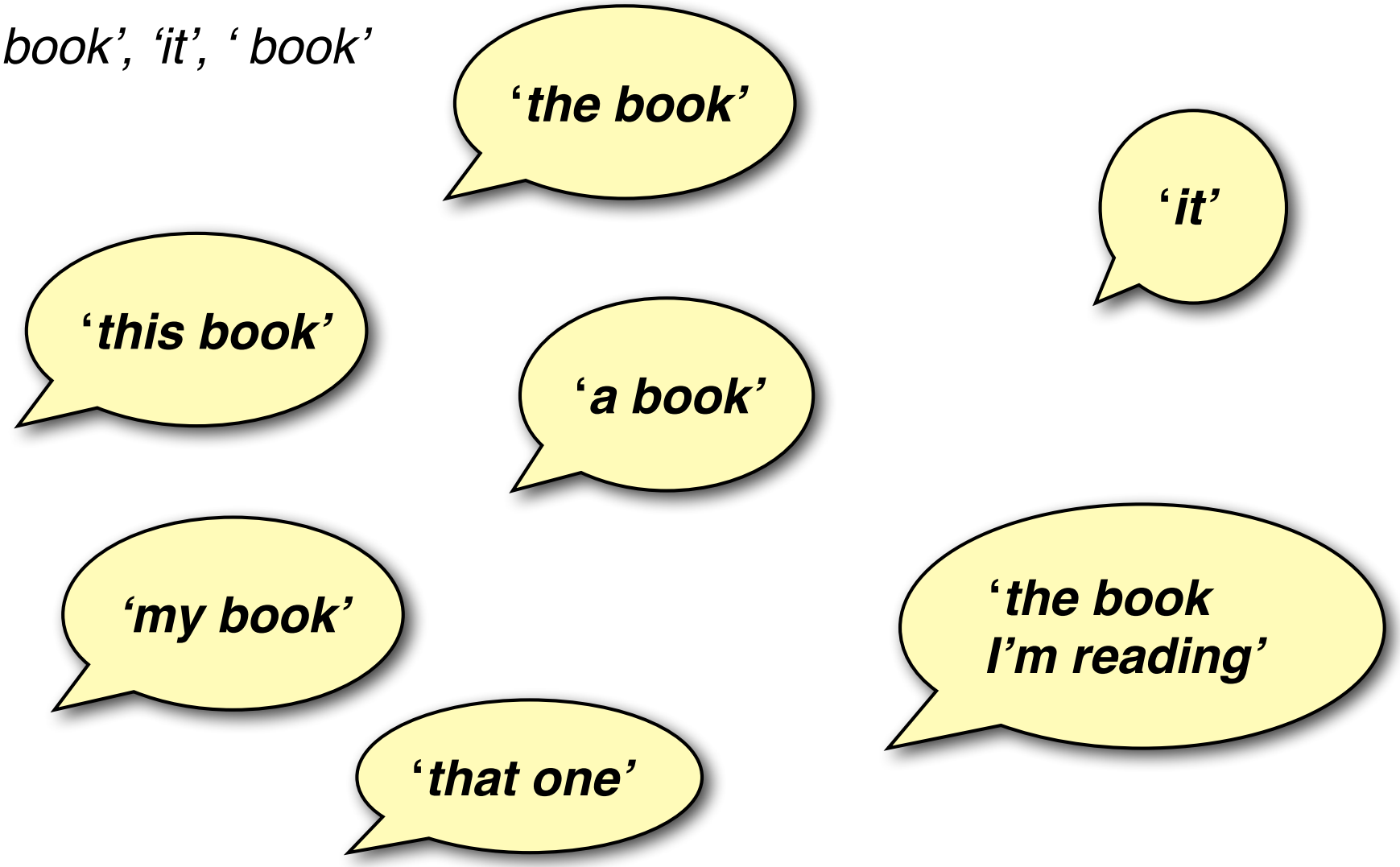
Rhetorical ('discourse') relations: ... so ... instead

between events and states

Referring expressions and coreference resolution

How do we refer to entities?

'a book', 'it', ' book'



Some terminology

Referring expressions (‘*this book*’, ‘it’) refer to some entity (e.g. a book), which is called the **referent.**

Co-reference: two referring expressions that refer to the same entity **co-refer** (are co-referent).
I saw a movie last night. I think you should see it too!

The referent is **evoked** in its first mention, and **accessed** in any subsequent mention.

Indefinite NPs

- **no determiner:**
  *I like **walnuts**.*
- **the indefinite determiner:**
  *She sent her **a** beautiful goose*
- **numerals:**
  *I saw **three** geese.*
- **indefinite quantifiers:**
  *I ate **some** walnuts.*
- **(indefinite) *this*:**
  *I saw **this** beautiful Ford Falcon today*

Indefinites usually **introduce a new discourse entity**.
They can refer to a specific entity or not:
I'm going to buy a computer today.

Definite NPs

- the **definite** article (***the*** *book*),
- **demonstrative articles** (***this/that*** *book*, ***these/those*** *books*),
- **possessives** (***my/John's*** *book*)

Definite NPs can also consist of

- **personal pronouns** (***I, he***)
- **demonstrative pronouns** (***this, that, these, those***)
- universal **quantifiers** (***all, every)***
- (unmodified) **proper nouns** *(**John Smith, Mary, Urbana**)*

Definite NPs **refer to an identifiable entity**
(previously mentioned or not)

Information status

Every entity can be classified along two dimensions:

Hearer-new vs. hearer-old

Speaker assumes entity is (un)known to the hearer

Hearer-old: *I will call Sandra Thompson*.

Hearer-new: *I will call a colleague in California* (=Sandra Thompson)

Special case of hearer-old: **hearer-inferrable**

I went to the student union. The food court was really crowded.

Discourse-new vs. discourse-old:

Speaker introduces new entity into the discourse, or refers to an entity that has been previously introduced.

Discourse-old: *I will call her/Sandra now.*

Discourse-new: *I will call my friend Sandra now.*

Coreference resolution

Victoria Chen, Chief Financial Officer 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 financial services company's president. It has been ten years since she came to Megabucks from rival Lotsabucks.

Coreference chains:

1. {Victoria Chen, Chief Financial Officer...since 2004, her, the 37-year-old, the Denver-based financial services company's president}
2. {Megabucks Banking Corp, Denver-based financial services company, Megabucks}
3. {her pay}
4. {rival Lotsabucks}

Special case: Pronoun resolution

Task: Find the antecedent of an anaphoric pronoun in context

1. ***John*** *saw* ***a beautiful Ford Falcon*** *at* ***the dealership****.*
2. ***He*** *showed* ***it*** *to* ***Bob****.*
3. ***He*** *bought* ***it****.*

he$_{2,}$ it$_2$ = John, Ford Falcon, or dealership?
he$_{3,}$ it$_2$ = John, Ford Falcon, dealership, or Bob?

Anaphoric pronouns

Anaphoric pronouns refer back to some previously introduced entity/discourse referent:
John *showed* ***Bob*** *his car.* ***He*** *was impressed.*
John showed Bob his car. ***This*** *took five minutes.*

The **antecedent** of an anaphor is the previous expression that refers to the same entity.

There are number/gender/person **agreement constraints:** ***girls*** can't be the antecedent of ***he***
Usually, we need some form of **inference** to identify the antecedents.

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 (also captured in Centering Theory)

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

Coref as binary classification

Represent each NP-NP pair (+context) as a feature vector.

Training:
Learn a binary classifier to decide whether NP_i is a possible antecedent of NP_j

Decoding (running the system on new text):

— Pass through the text from beginning to end

— For each NP_i:
Go through $NP_{i-1}...NP_1$ to find best antecedent NP_j.
Corefer NP_i with NP_j.
If the classifier can't identify an antecedent for NP_i, it's a new entity.

Example features for Coref resolution

What can we say about each of the two NPs?

Head words, NER type, grammatical role, person, number, gender, mention type (proper, definite, indefinite, pronoun), #words, …

How similar are the two NPs?

— Do the two NPs have the same head noun/modifier/words?
— Do gender, number, animacy, person, NER type match?
— Does one NP contain an alias (acronym) of the other?
— Is one NP a hypernym/synonym of the other?
— How similar are their word embeddings (cosine)?

What is the likely relation between the two NPs?

— Is one NP an appositive of the other?
— What is the distance between the two NPs?
distance = #sentences, #mentions,..

Lee et al.'s neural model for coref resolution

Joint model for mention identification and coref resolution:

— Use word embeddings + LSTM to get a vector $\mathbf{g}_i$ for each span $i = \text{START}(i)...\text{END}(i)$ in the document (up to a max. span length L)

— Use $\mathbf{g}_i$ + neural net NN_m to get a mention score $m(i)$ for each i (this can be used to identify most likely spans at inference time)

— Use $\mathbf{g}_i\,\mathbf{g}_j$ + NN_c to get antecedent scores $c(i,j)$ for all spans $i,j<i$

— Compute overall score $s(i,j) = m(i) + m(j) + c(i,j)$ for all $i,j<i$
Set overall score $s(i,\varepsilon) = 0$ [i is discourse-new/not anaphoric]

— Identify the most likely antecedent for each span i according to

$$y_i^* = \text{argmax}_{y_i \in \{1,...i-1,\epsilon\}} P(y_i)$$

$$\text{with} \quad P(y_i) = \frac{\exp(s(i,y_i))}{\sum_{y' \in \{1,..i-1,\epsilon\}} \exp(s(i,y'))}$$

— Perform a forward pass over all (most likely) spans to identify their most likely antecedents

Lee et al.'s neural model for coref resolution

Span representation $\mathbf{g}_i$:

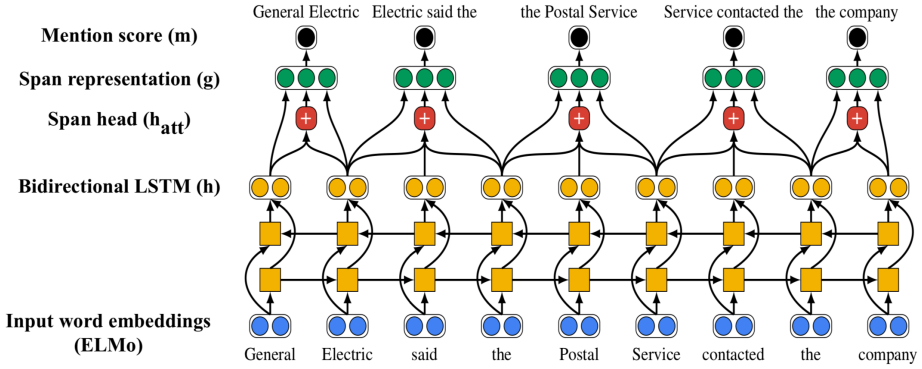
Computed by a biLSTM over word embeddings:

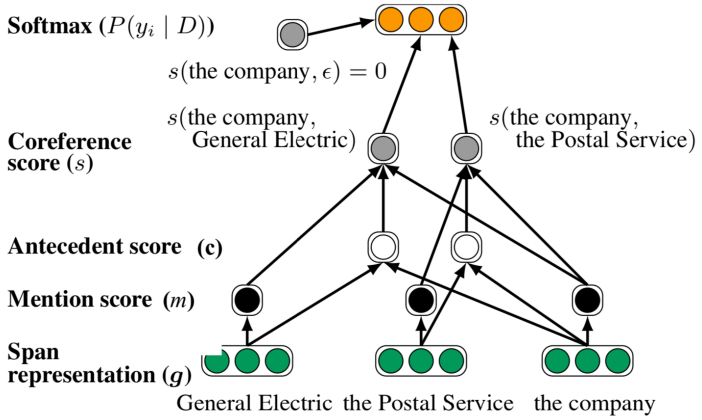
LSTM's hidden state of i's first word,

LSTM's hidden state of i's last,

weighted avg of word embeddings in span i; length of span

$[\mathbf{h}_{\text{START(i)}}, \mathbf{h}_{\text{END(i)}}, \mathbf{h}_{\text{ATT(i)}}, \varphi(i)]$





Scoring function $s(i,j)$:

a) for $j=\varepsilon$ (i has no antecedent): $s(i,\varepsilon) = 0$

b) for $j\neq\varepsilon$: $s(i,j) = m(i) + m(j) + c(i,j)$

$m(i)$: is span i a mention?

binary classifier (feedforward net) with $\mathbf{g}_i$ as input

$c(i,j)$: is j an antecedent of i?

input: $\mathbf{g}_i$, $\mathbf{g}_j$, $\mathbf{g}_i \circ \mathbf{g}_i$ [element-wise multiplication]

Evaluation metrics for coref resolution

Compare hypothesis H against (gold) reference R by:

MUC score:

- Precision/Recall over #coref links
- Ignores singleton mentions
- Rewards long coref chains/clusters

B^3 score:

- Precision/Recall over mentions in same cluster
- may count same mention multiple times

CEAF score:

- Precision/Recall, based on mention alignments

CoNLL F1: combines MUC, B^3, CEAF

Challenge: How to handle predicted mentions?

Entity-based coherence

Entity-based coherence

Discourse 1:
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.

Discourse 2:
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.

Entity-based coherence

Discourse 1:

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.

Discourse 2:

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.

How we refer to entities influences how *coherent* a discourse is (**Centering theory)**

Centering Theory

Grosz, Joshi, Weinstein (1986, 1995)

A linguistic theory of entity-based coherence and salience

It predicts which entities are salient at any point during a discourse.
It also predicts whether a discourse is entity-coherent, based on its referring expressions.

Centering is about **local (=within a discourse segment) coherence and salience**

Centering theory itself is **not a computational model** or an algorithm: many of its assumptions are not precise enough to be implemented directly. (Poesio et al. 2004)

But many algorithms have been developed based on specific instantiations of the assumptions that Centering theory makes. The textbook presents a centering-based pronoun-resolution algorithm

Rhetorical (Discourse) relations

Rhetorical relations

Discourse 1:
John hid Bill's car keys. He was drunk.

Discourse 2:
John hid Bill's car keys. He likes spinach.

Discourse 1 is more coherent than Discourse 2 because "He(=Bill) was drunk" provides an **explanation** for "John hid Bill's car keys"

What **kind of relations** between two consecutive utterances (=sentences, clauses, paragraphs,…) make a discourse coherent?

Rhetorical Structure Theory; also lots of recent work on **discourse parsing (Penn Discourse Treebank)**

Example: The *Result* relation

The reader can infer that the **state/event described in S0 causes** (or: could cause) **the state/event asserted in S1**:

S0: The Tin Woodman was caught in the rain.

S1: His joints rusted.

This can be rephrased as:
"S0. As a result, S1"

Example: The *Explanation* relation

The reader can infer that **the state/event in S1 provides an explanation** (reason) **for the state/event in S0**:

S0: John hid Bill's car keys.

S1: He was drunk.

This can be rephrased as:
"S0 because S1"

Rhetorical Structure Theory (RST)

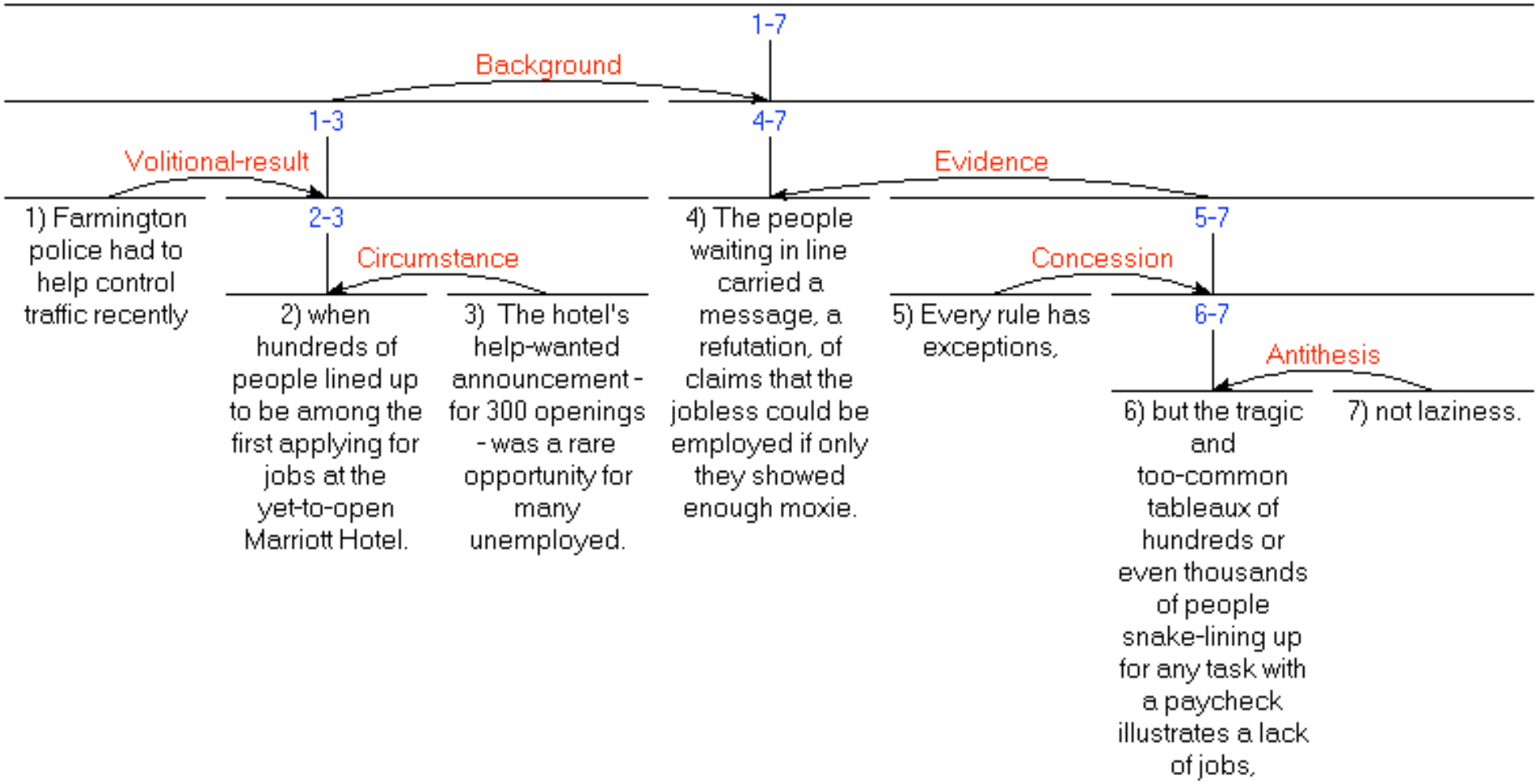
RST (Mann & Thompson, 1987) describes **rhetorical relations** between utterances:
Evidence, Elaboration, Attribution, Contrast, List,…

Different variants of RST assume different sets of relations.

Most relations hold between a **nucleus** (N) and a **satellite** (S). Some relations (e.g. *List*) have **multiple nuclei** (and no satellite).

Every relation imposes certain **constraints** on its arguments (N,S), that describe the goals and beliefs of the **reader** R and **writer** W, and the effect of the utterance on the reader.

Discourse structure is hierarchical



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

Happy fall break!