Discourse and Pragmatics

Can you stop talking?

Literally, a **yes/no question**

But we understand it as a **request**

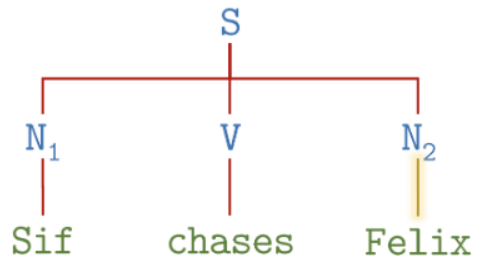
This is an example of an **indirect speech act.**

Semantics: Recap

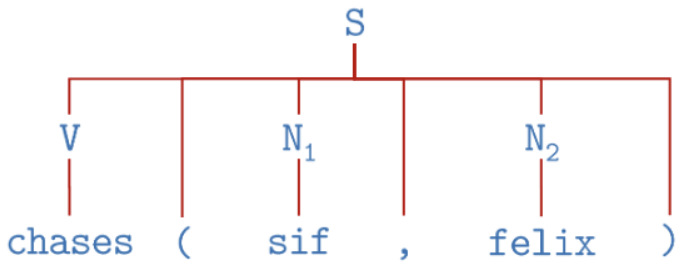
- **Lexical Semantics** meaning of words
- **Compositional Semantics**: combining the meaning of words/phrase

PARSING WITH SYNCHRONOUS GRAMMARS

Generate a logical form capturing the meaning of a sentence.




```
S -> <N1 V N2, V '(' N1 ',' N2 ')'>
N -> <'Sif', 'sif'>
N -> <'Felix', 'felix'>
V -> <'chases', 'chases'>
```



Break a leg!

I just love how they are always late

PRAGMATICS

- Our discussion on semantics has focused on the meaning of individual sentences.
- In reality, sentences almost never appear in isolation.
  - Natural language sentences have context.
  - Surrounding sentences can change the meaning of other sentences.
  - E.g., ‘I was so focused, I didn’t realize how much time had passed. Time flies like an arrow.’
  - vs ‘The wizard studied flies that possessed the ability to travel through time. He held an arrow in his hand and the flies were attracted to it. Time flies like an arrow.’

PRAGMATICS

- The sentence ‘Time flies like an arrow’ has multiple ambiguous interpretations.
  - Each interpretation can be represented with a distinct logical form.
- The surrounding context helps us to determine which interpretation is most likely.
- Consider another example:
  - ‘Everyone looked up.’
  - ‘During gym class, Alice kicked the ball high into the air. Everyone looked up.’
  - ‘The supervillain turned on the machine, and the sky began to change color everywhere. Everyone looked up.’

PRAGMATICS

- ‘Everyone’ in ‘Everyone looked up’ is an example of underspecification.
  - We didn’t specify whether it’s ‘Everyone in the gym class’ or ‘Everyone outside in the world’.
  - We were able to infer that information from the context.
- Why do we leave out this information from sentences?

Pragmatics: **Indirect Speech Acts**

- The literal meaning of a sentence doesn't match the speaker's actual goal.

- **The Request:** "*It's getting a bit chilly in here*."
  - *Literal:* A statement about the temperature.
  - *Pragmatic:* A request for someone to close a window or turn up the heat.

- **The Complaint:** "*Do you know what time it is?*"
  - *Literal:* A question asking for the current time.
  - *Pragmatic:* A way of telling someone they are late.

Pragmatics: **Presuppositions**

- The unspoken assumptions baked into a sentence that must be true for the statement to make sense.

- **Example:** "I'm glad I stopped smoking."
  - *Presupposition:* I used to smoke.

- **Example:** "When did you realize you were wrong?"
  - *Presupposition:* You were, in fact, wrong

Pragmatics: **Conversational Implicature**

- Imply something without saying it directly, often by "flouting" a conversational rule.

Person A: "Do you like the new boss?"
Person B: "She certainly wears interesting ties."

Implication*:* Person B is avoiding the question, likely because they don't like the boss

Pragmatics: **Deixis**

- Words that mean nothing without physical or situational context.

Person A: "Put **that** over **there** by **tomorrow**."

Analysis: This sentence is "deictically centered" on the speaker. Without being in the room to see what Person A is pointing at ("that"), where they are pointing ("there"), and knowing what day it is currently ("tomorrow"), **the sentence has no fixed meaning**

Pragmatics: **Politeness**

- Using specific phrasing to avoid "threatening" someone's social standing or comfort.

  - *Writer:* "What did you think of my first draft?"
  - *Editor:* "It's certainly… ambitious."

The word "**ambitious**" acts as a **hedge**. Instead of saying "it's a mess," the editor uses a "face-saving" strategy to remain polite while indirectly signaling that the draft needs significant work

PRAGMATICS

- We can think of human language as a product of optimization:
  - Humans want to communicate information to each other.
  - But we want to minimize the energy cost of the communication.
    - Talking a lot is tiring.
  - We want to maximize information transfer while minimizing energy cost.
- Speakers want to minimize the number of words they produce.
- Listeners want to make sure they have accurately received the information.
  - But accurate information transfer typically requires more words.
- Thus, natural language strikes a balance between using enough words to convey the correct information, but no more than that.
  - So we avoid repeating information if it can be inferred from context.

GRICE'S MAXIMS

- Paul Grice (1975) posited that human languages follow four 'maxims'.

1. Maxim of quantity:
   - Be informative
   - Make your contribution as informative as required, but no more.
2. Maxim of quality:
   - Be truthful
   - If you are unsure or don't know, say so.
3. Maxim of relevance:
   - Be relevant

GRICE'S MAXIMS

- Paul Grice (1975) posited that human languages follow four 'maxims'.

4. Maxim of manner:
    - Try to ensure the listener understands you.
    - Avoid language that's difficult to understand for the listener.
    - Avoid ambiguity
        - Some ambiguity is okay, if the correct interpretation is obvious to the listener from context.
        - But if not, the listener might ask you to clarify, which increases the cost of communication.
    - Try to consider what the listener knows and doesn't know.

GRICE'S MAXIMS

- These maxims are not axioms or theorems.
  - They are not universally true.
- E.g., What if your goal is deception?
  - You wouldn't want to follow the maxim of quality.
- If you want to change the subject, then you shouldn't follow the maxim of relevance.

GRICE'S MAXIMS

- ***Flouting a maxim*** is an intentional failure to follow conversational maxims, expecting the listener to understand the implicit meaning
  - ***Not Deception***, *rather indirectness to express irony/humor, emphasis, etc.*
- Maxim of **quantity** (Be informative)
  - "How was your interview?" → "the seat was comfy"
  - Conversational **implicature** that it was bad
- Maxim of **quality** (Be truthful)
  - "I love Indiana winters!", using **irony**
- Maxim of **relevance** (Be relevant)
  - "This would taste better with salt" **Indirect speech act**
- Maxim of **manner** (ensure the listener understands you)
  - "I hate to disturb you but..", "I'm not sure, but" **indirect/polite** expressions

PRAGMATICS IN NLP

- “park next to the gray car”

PRAGMATICS IN NLP

- If an NLP model is to accurately perform tasks that involve context, such as long-document question answering, it needs to be able to correctly model how context affects the meaning of sentences.
  - And correctly infer the full meaning of underspecified sentences.

- E.g.,
  - ‘...During gym class, Alice kicked the ball high into the air. Everyone looked up and saw a falcon grab the ball... Jay didn’t come to school that day since he wasn’t feeling well...’

    (the story could be much longer)
  - ‘Did Jay see the falcon?’

Discourse

- Focusing on language interpretation beyond individual sentences.
- **Discourse:** linguistic unit consisting of more than a single sentence.
  - Paragraphs, articles, documents, conversations, etc.

Kim went to MacDonald's last week to get a bite to eat. But the place was closed. That was very disappointing. The following week he went to Chipotle instead.

Hey, what are you up to?

Nothing special

Want to catch a movie?

- **Needed both for analysis and for coherent generation of longer texts.**

Example Discourse Task

1) doing **coreference** resolution:

- *'the place'* and '*MacDonald's'* **refer to the same entity**
- *He* and *Kim* refer to the same person.
- *That* refers to *'the place was closed'*.

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

- *'to get a bite to eat'* is the ***reason*** for
- *'Kim went to MacDonald's.'*

Discourse

- Representing discourse information requires model that can..

..identify **entities, events and states and relations** between them.

- Before attending class yesterday, Kim went to McDonald's since he was hungry.

Entities, both physical and abstract, animate and non-animate.

Discourse

- Representing discourse information requires model that can..

Identify **entities, events and states and relations** between them.

- Before attending class yesterday, Kim went to McDonald's since he was hungry.

Eventualities that involve entities, such as events that happen at a point in time or states that last for a period of time

Discourse

- Representing discourse information requires model that can..

Identify **entities, events and states and relations** between them.

- Before attending class yesterday, Kim went to McDonald's since he was hungry.

Temporal relations between events/states

Discourse

- Representing discourse information requires model that can..

Identify **entities, events and states and relations** between them.

- Before attending class yesterday, Kim went to McDonald's since he was hungry.

Rhetorical relations between propositions

Co-reference

- Referring expressions (e.g., Kim, he, the plate)
- refer to a *real-world* entity which is called the referent.
- Co-reference: two referring expressions that refer to the same entity are co-referent.
- *Kim was **hungry**, **he** hasn't eaten all day!*

- The first mention of the referent is evoked in its first mentioned and access afterwards.

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

Co-reference: identify all the sets of text string that refer to the **same entity.**

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}

Co-reference Resolution

We can encode some preferences about which entities are co-referent:

- **English Constraints**
  - **Number agreement:** I put the packages in the **car,** they are red.
  - **Gender:** The man went into the **car**, he was fast.
  - **Person:** I went to the movie and **she** came with me,

As usual there are exceptions!

The ship moved quickly, “she is a fine vessel” said the captain.

Coref as binary classification

- Represent each NP-NP pair as a binary classification instance.
  - Learn a binary classifier to decide whether NP_i is a possible antecedent of NP_j

- Relevant features rely on: head words, NER type, grammatical role, person, number, gender, mention type, etc.
  - Compare the feature similarity between NP_i and NP_j
  - Include relational information: e.g., appositive (Jon, the mayor of BigCity), as well as distance between NPs

https://aclanthology.org/D17-1018/

End-to-end Neural Coreference Resolution Lee et-al 2017

Joint model for mention identification and coref resolution:

Use word embeddings + LSTM to get a vector $\mathbf{g}_i$ for each span i
$i = \text{START}(i)...\text{END}(i)$ 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
(used to identify most likely mention spans at inference time)

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

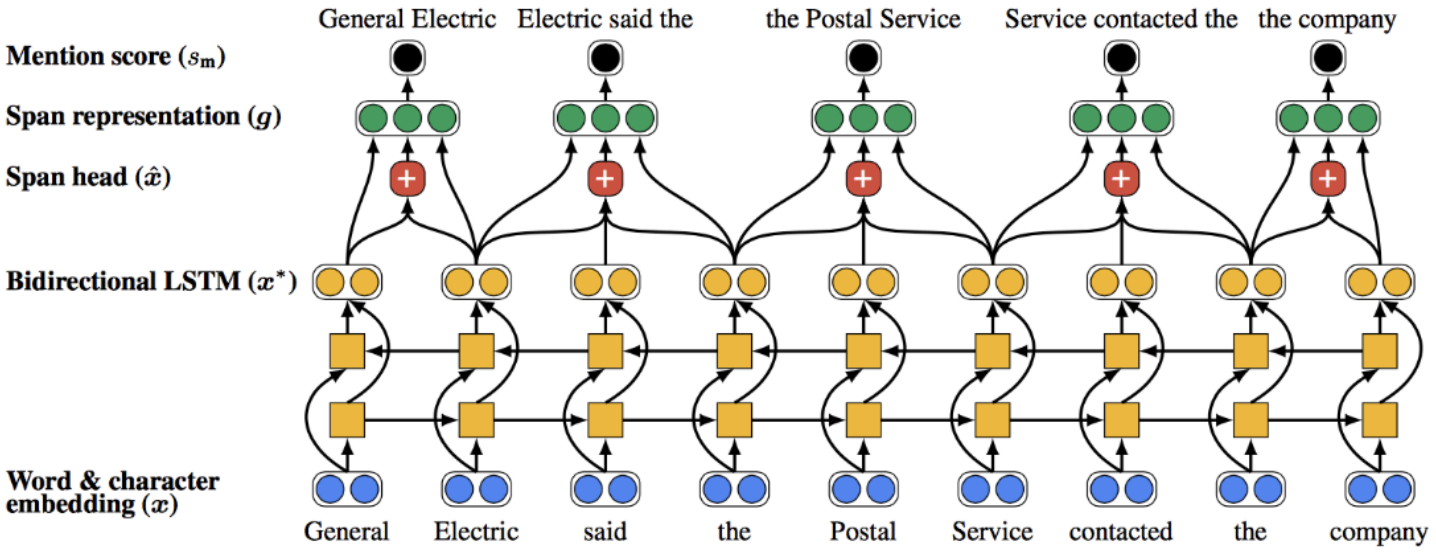
Compute overall score $s(i,j) = m(i)+m(j)+c(i,j)$ for all span pairs $i,j<i$
and set overall score $s(i,\varepsilon) = 0$ [score for i being discourse-new]

Identify the most likely antecedent for each span i according to

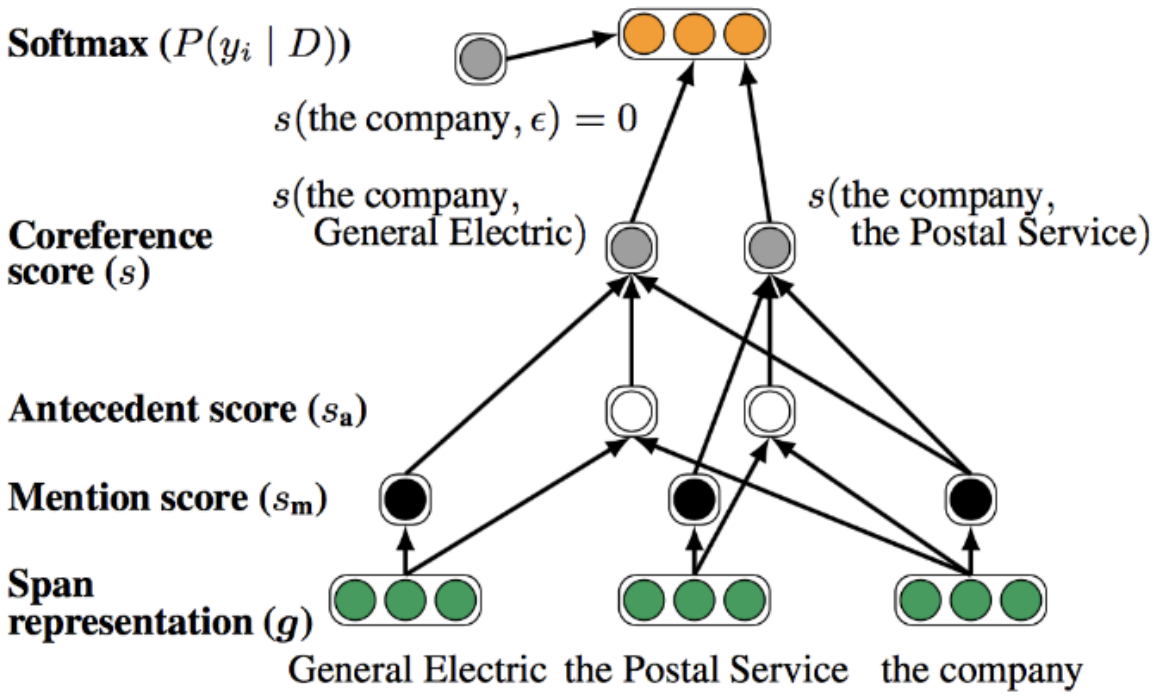
$$y_i^* = \text{argmax}_{y_i \in \{1,...i-1,\epsilon\}} P(y_i) \quad \text{with} \quad P(y_i) = \frac{\exp(s(i, y_i))}{\sum_{y' \in \{1,..i-1,\epsilon\}} \exp(s(i, y'))}$$

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

End-to-end Neural Coreference Resolution Lee et-al 2017



End-to-end Neural Coreference Resolution Lee et-al 2017



Coref Evaluation

- Compare predicted pairs to gold pairs-–
- **MUC score:**
  - Precision/Recall over #coref links
  - Ignores singleton mentions
  - Rewards long coref chains/clusters
- **B3 score:**
  - Precision/Recall over mentions in same cluster
  - May count same mention multiple times
- **CEAF score:**
  - Precision/Recall, based on mention alignments

The importance of world knowledge

- Correct coreference resolution can reflect **knowledge.**
  - E.g., commonsense knowledge, world knowledge, etc.

- The city councilmen refused the demonstrators a permit because they **feared** violence.

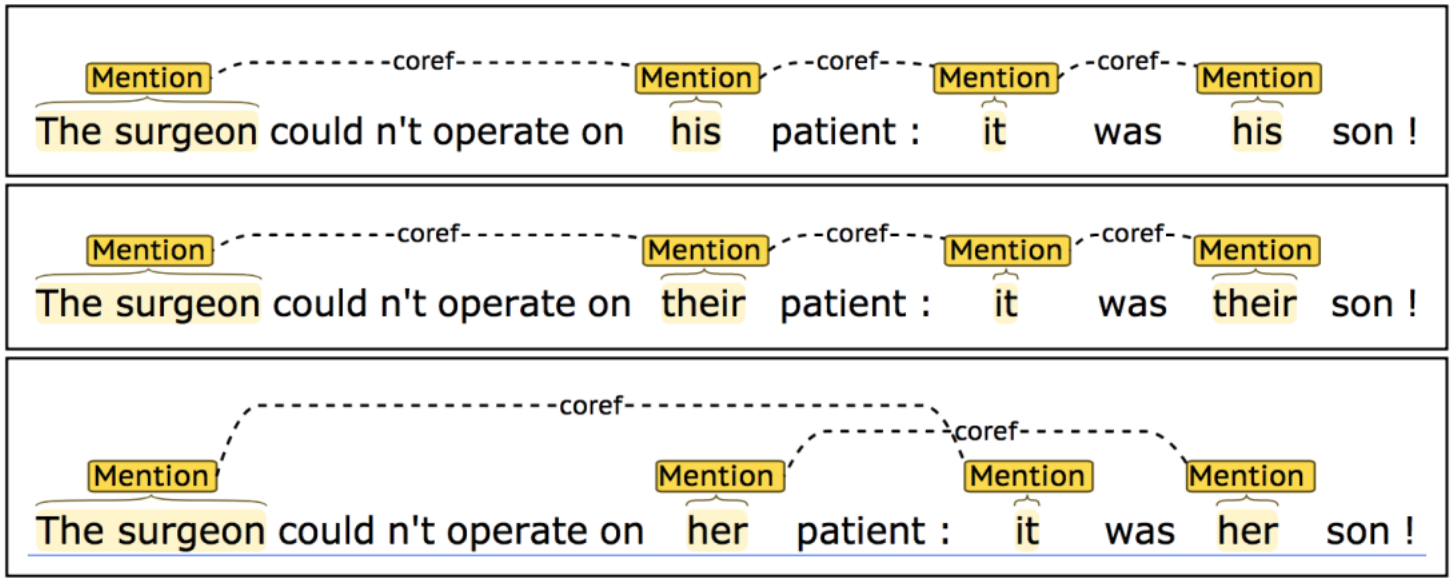
- The city councilmen refused the demonstrators a permit because they **advocated** violence.

- The Winograd Schema Challenge
- https://cs.nyu.edu/faculty/davise/papers/WinogradSchemas/WS.html

Bias

Preferred attachments (both by humans and systems) often reflect stereotypes (e.g. about occupations and gender)

https://www.aclweb.org/anthology/N18-2002/



Modeling Discourse Structure

- Understanding discourse goes beyond individual sentence, it requires using context to create meaning.
- What are discourse units?
- What makes a text coherent?

Local and Global Coherence

- **Local Coherence:** connections between adjacent sentences, ensuring a smooth transition.
  - topical coherence: adjacent sentences relate to similar topic or theme
  - entity-based coherence: adjacent sentences discuss similar entities
  - rhetorical coherence, logical/rhetorical relations between adjacent sentences
- **Global Coherence:** overall structure, including topic consistency and thematic unity. Can be different depending on genre, such as stories, scientific studies, legal texts, etc.

https://web.stanford.edu/~jurafsky/slp3/24.pdf

Topical coherence

Credit cards are convenient, but dangerous. **They** allow people to make large purchases easily without saving up cash in advance. **This** is helpful for items that must be acquired without delay. **However, this convenience** comes at a high cost…

These sentences discuss the same topic, and use a similar words (same semantic field, related to transactions/credit) known as **lexical cohesion**

Rhetorical coherence

John took a train from Paris to Istanbul. He likes spinach.

This discourse is incoherent because there is no apparent rhetorical relation between the two sentences.

Jane took a train from Paris to Istanbul. She had to attend a conference.

This discourse is coherent because there is clear rhetorical relation between the two sentences.

The second sentence provides a REASON or EXPLANATION for the first.

Entity-based coherence

John wanted to buy a piano for his living room. Jenny also wanted to buy a piano.
He went to the piano store.
It was nearby.
The living room was on the second floor. She didn't find anything she liked.
The piano he bought was hard to get up to that floor.

This is **incoherent** because the sentences switch back and forth between **entities** (John, Jenny, the piano, the store, the living room)

Which text reads better?

Jeff went to the store.
He bought some apples.
He walked home quickly.

Jeff went to the store.
The apples were on sale there.
A heavy bag was carried by him.

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 in adjacent sentences 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 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.

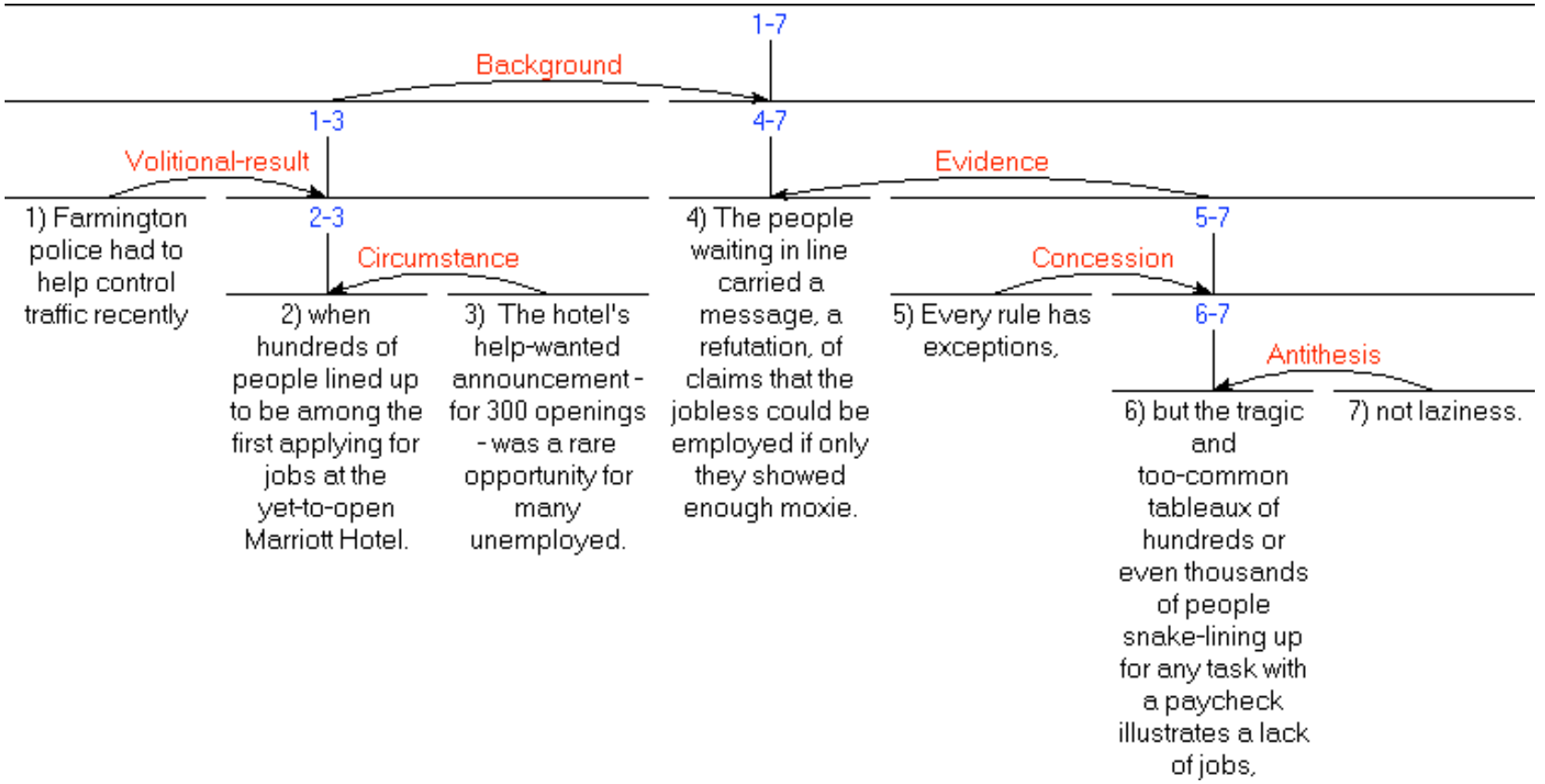
Most relations hold between a **nucleus** (N) and a **satellite** (S).

Some relations (e.g. *List*) have **multiple nuclei**

- **Evidence (N=Claim, S=Evidence):**
  - **Nucleus:** The restaurant has the best pizza in town.
  - **Satellite:** Their Yelp reviews are all five stars

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.

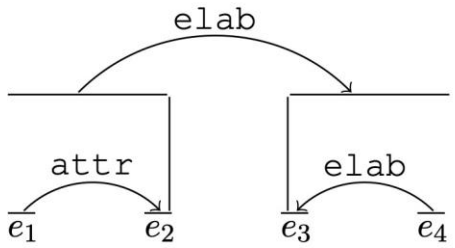
Discourse structure is hierarchical



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

RST Parsing

- Step 1: Identify “elementary discourse units” (EDUs, text spans)
- Step 2: Use a shift-reduce parser to predict RST relations between EDUs



e_1: American Telephone & Telegraph Co. said it
e_2: will lay off 75 to 85 technicians here , effective Nov. 1.
e_3: The workers install , maintain and repair its private branch exchanges,
e_4: which are large intracompany telephone networks.

Figure 27.5 Example RST discourse tree, showing four EDUs. Figure from Yu et al. (2018).

https://aclanthology.org/C18-1047.pdf

RST Parsing

- Step 1: Identify “elementary discourse units” (EDUs, text spans)
- Step 2: Use a shift-reduce parser to predict RST relations between EDUs

Step	Stack	Queue	Action	Relation
1	$\varnothing$	e_1, e_2, e_3, e_4	SH	$\varnothing$
2	e_1	e_2, e_3, e_4	SH	$\varnothing$
3	e_1, e_2	e_3, e_4	RD (attr, SN)	$\varnothing$
4	$e_{1:2}$	e_3, e_4	SH	$\widehat{e_1 \mathbf{e_2}}$
5	$e_{1:2}, e_3$	e_4	SH	$\widehat{e_1 \mathbf{e_2}}$
6	$e_{1:2}, e_3, e_4$	$\varnothing$	RD (elab, NS)	$\widehat{e_1 \mathbf{e_2}}$
7	$e_{1:2}, e_{3:4}$	$\varnothing$	RD (elab, SN)	$\widehat{e_1 \mathbf{e_2}}, \widehat{\mathbf{e_3} e_4}$
8	$e_{1:4}$	$\varnothing$	PR	$\widehat{e_1 \mathbf{e_2}}, \widehat{\mathbf{e_3} e_4}, \widehat{e_{1:2} \mathbf{e_{3:4}}}$

Figure 27.6 Parsing the example of Fig. 27.5 using a shift-reduce parser. Figure from Yu et al. (2018).

https://aclanthology.org/C18-1047.pdf

Penn Discourse Treebank (PDTB)

The PDTB annotates explicit and implicit discourse connectives and their argument spans.

Jewelry displays in department stores were often cluttered and uninspired. And the merchandise was, well, fake. **As a result**, marketers of faux gems steadily lost space in department stores to more fashionable rivals—cosmetics makers.

In July, the Environmental Protection Agency imposed a gradual ban on virtually all uses of asbestos. **(implicit=as a result)** By 1997, almost all remaining uses of cancer-causing asbestos will be outlawed.

https://aclanthology.org/L08-1093/

PDTB semantic distinctions

Class	Type	Example
TEMPORAL	SYNCHRONOUS	The parishioners of St. Michael and All Angels stop to chat at the church door, as members here always have. (Implicit while) In the tower, five men and women pull rhythmically on ropes attached to the same five bells that first sounded here in 1614.
CONTINGENCY	REASON	Also unlike Mr. Ruder, Mr. Breeden appears to be in a position to get somewhere with his agenda. (implicit=because) **As a former White House aide who worked closely with Congress, he is savvy in the ways of Washington.**
COMPARISON	CONTRAST	The U.S. wants the removal of what it perceives as barriers to investment; Japan denies there are real barriers.
EXPANSION	CONJUNCTION	Not only do the actors stand outside their characters and make it clear they are at odds with them, but they often literally stand on their heads.

PDTB sense hierarchy

Temporal	**Comparison**
• Asynchronous	• Contrast (Juxtaposition, Opposition)
• Synchronous (Precedence, Succession)	•*Pragmatic Contrast (Juxtaposition, Opposition)*
	• Concession (Expectation, Contra-expectation)
	• *Pragmatic Concession*
Contingency	**Expansion**
• Cause (Reason, Result)	• *Exception*
• Pragmatic Cause (Justification)	• Instantiation
• *Condition (Hypothetical, General, Unreal Present/Past, Factual Present/Past)*	• Restatement (Specification, Equivalence, Generalization)
• *Pragmatic Condition (Relevance, Implicit Assertion)*	• Alternative (Conjunction, Disjunction, Chosen Alternative)
	• List

Global coherence: argumentation structure

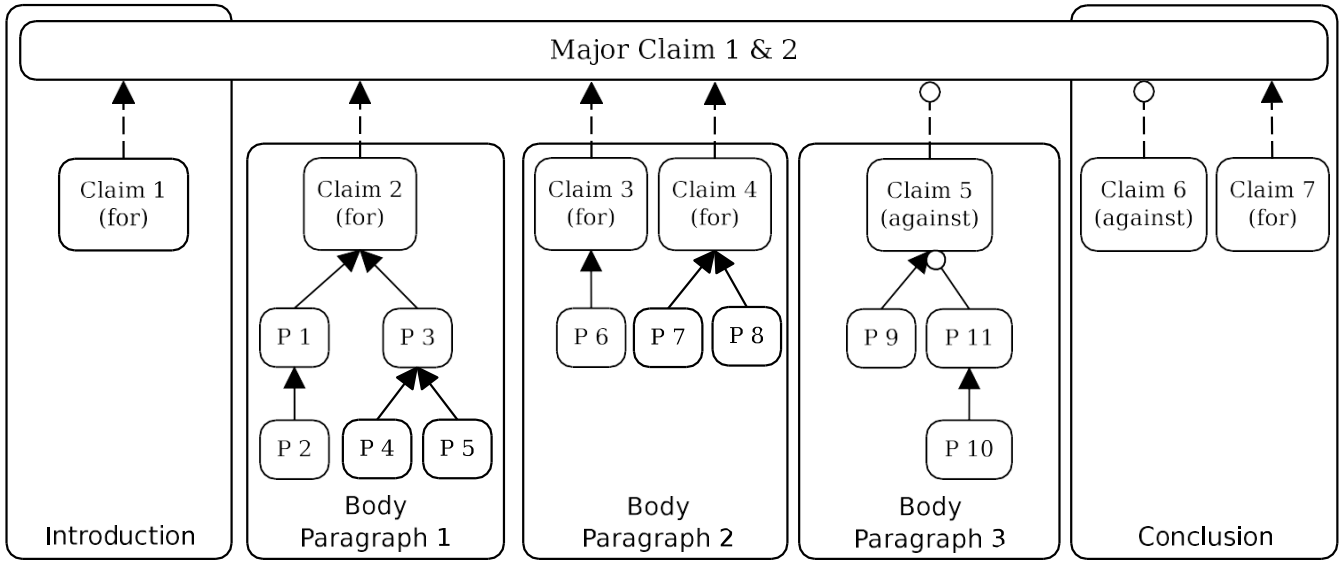
In persuasive essays, **claims** (1) may be followed (or preceded) by **premises** (2,3) that support the claim, (some of which might be supported by their own premises (4)

> (1) Museums and art galleries provide a better understanding about arts than Internet. (2) In most museums and art galleries, detailed descriptions in terms of the background, history and author are provided. (3) Seeing an artwork online is not the same as watching it with our own eyes, as (4) the picture online does not show the texture or three-dimensional structure of the art, which is important to study."

https://aclanthology.org/D14-1006/

Argumentation Mining

- Automatically identify the claims and their supporting premises



Scientific Discourse Structure

Category	Description	Example
AIM	Statement of specific research goal, or hypothesis of current paper	"The aim of this process is to examine the role that training plays in the tagging process"
OWN METHOD	New Knowledge claim, own work: methods	"In order for it to be useful for our purposes, the following extensions must be made:"
OWN RESULTS	Measurable/objective outcome of own work	"All the curves have a generally upward trend but always lie far below backoff (51% error rate)"
USE	Other work is used in own work	"We use the framework for the allocation and transfer of control of Whittaker...."
GAP WEAK	Lack of solution in field, problem with other solutions	"Here, we will produce experimental evidence suggesting that this simple model leads to serious overestimates"
SUPPORT	Other work supports current work or is supported by current work	"Work similar to that described here has been car-ried out by Merialdo (1994), with broadly similar conclusions."
ANTISUPPORT	Clash with other's results or theory; su-periority of own work	"This result challenges the claims of..."

Figure 23.13 Examples for 7 of the 15 labels from the Argumentative Zoning labelset (Teufel et al., 2009).

We can also label spans in scientific papers with the role they play in the overall argumentation of the paper.