More On Discourse

LING 571 — Deep Processing Techniques for NLP
November 28th, 2018
Ryan Georgi

Discourse Segmentation Recap

- Mars. With its distant orbit—50 percent farther from the sun that Earth—and slim atmospheric blanket, Mars experiences frigid weather conditions. Surface temperatures typically average about -60 degrees Celsius (-76 degrees Farenheit) at the equator and can dip to -123 degrees C near the poles. Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, but any liquid water formed in this way would evaporate almost instantly because of the low atmospheric pressure.

Discourse Segmentation Recap

- Mars. With its distant orbit—50 percent farther from the sun that Earth—and slim atmospheric blanket, Mars experiences frigid weather conditions. Surface temperatures typically average about -60 degrees Celsius (-76 degrees Farenheit) at the equator and can dip to -123 degrees C near the poles. Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, but any liquid water formed in this way would evaporate almost instantly because of the low atmospheric pressure.

- Different discourse segments are related by *coherence relations*

Text Coherence

- Cohesion — repetition, transitions — does not imply coherence
- Coherence relations:
  - Possible meaning relations between utterances in discourse
  - Example (Eisenstein, 2015 via Benioff & Weiss, 2012)
    - The more people you love, the weaker you are.
    - (?) You'll do things for them you know you shouldn't.
    - (?) You'll act the fool to make them happy, to keep them safe.
    - (?) Love no one but your children.
    - (?) On that front, a mother has no choice.

Text Coherence

- Cohesion — repetition, etc. — does not imply coherence
- Coherence relations:
  - Possible meaning relations between utterances in discourse
  - Example (Eisenstein, 2015 via Benioff & Weiss, 2012)
    - (*Nucleus/Premise*) The more people you love, the weaker you are.
    - (*Expansion*) You'll do things for them you know you shouldn't.
    - (*Expansion*) You'll act the fool to make them happy, to keep them safe.
    - (*Contingency*) Love no one but your children.
    - (*Contingency*) On that front, a mother has no choice.

Coherence Relations & Discourse Structure

Coherence Relations

John hid Bill's car keys. He was drunk.
?? *John hid Bill's car keys. He likes spinach.*

- Why is this odd?
  - No obvious relation between sentences
    - Readers often try to construct relations
- How are the first two related?
  - Explanation/cause
- Utterances should have meaningful connection
  - Establish through ***coherence relations***

Coherence Relations

- **Result**: Infer that the state or event asserted by S_0 causes, or could cause the state asserted by S_1.
  - *The Tin Woodman was caught in the rain. His joints rusted.*
- **Explanation:** Infer that the state or event asserted by S_1 causes or could cause the state or event asserted by S_0.
  - *John hid Bill's car keys. He was drunk.*
- **Parallel:** Infer $p(a_1, a_2, \ldots)$ from the assertion of S_0 and $p(b_1, b_2, \ldots)$ 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.*

Coherence Relations

- **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.*
- **Occasion:** A change of state can be inferred from the assertion of S_0 whose final state can be inferred from S_1, or a change of state can be inferred from the assertion of S_1.
  - *Dorothy picked up the oil-can. She oiled the Tin Woodman's joints.*

Coherence Relation Hierarchy

S1 – Armin went to the bank to deposit his paycheck

S2 – He then took a train to Kim's car dealership.

S3 – He needed to buy a car.

S4 – The company he works for now isn't near any public transportation.

S5 – He also wanted to talk to Kim about their softball league.

- This discourse ***isn't linear***
- Primarily about S1, S2
  - S3-S5 relate to different parts of S1, S2

Adapted from J&M 2nd ed p. 690

Coherence Relation Hierarchy

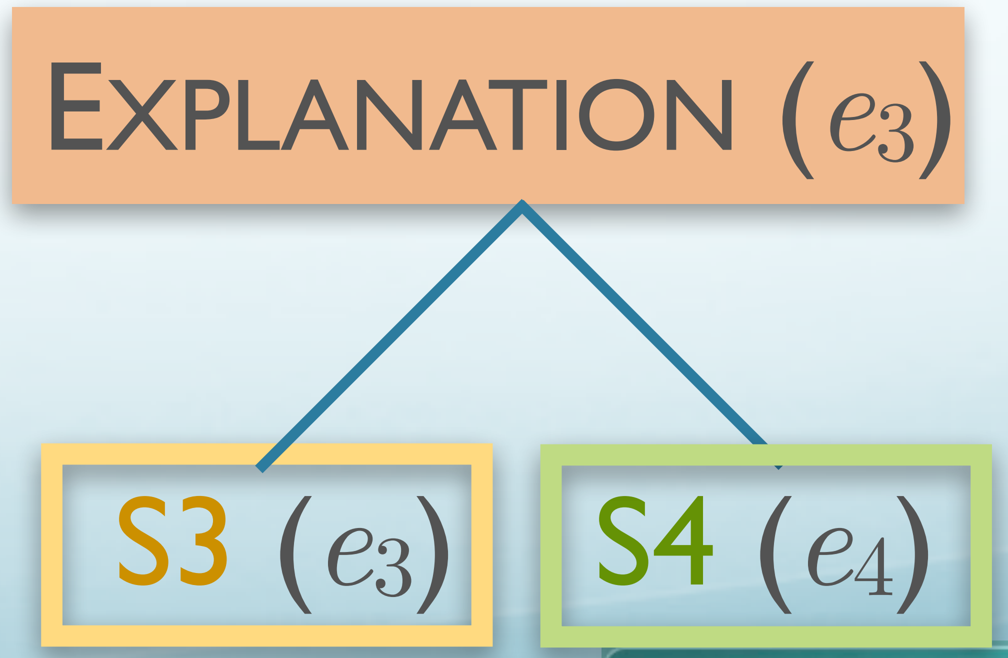
S1 – Armin went to the bank to deposit his paycheck

S2 – He then took a train to Kim's car dealership.

S3 – He needed to buy a car.

S4 – The company he works for now isn't near any public transportation.

S5 – He also wanted to talk to Kim about their softball league.



Coherence Relation Hierarchy

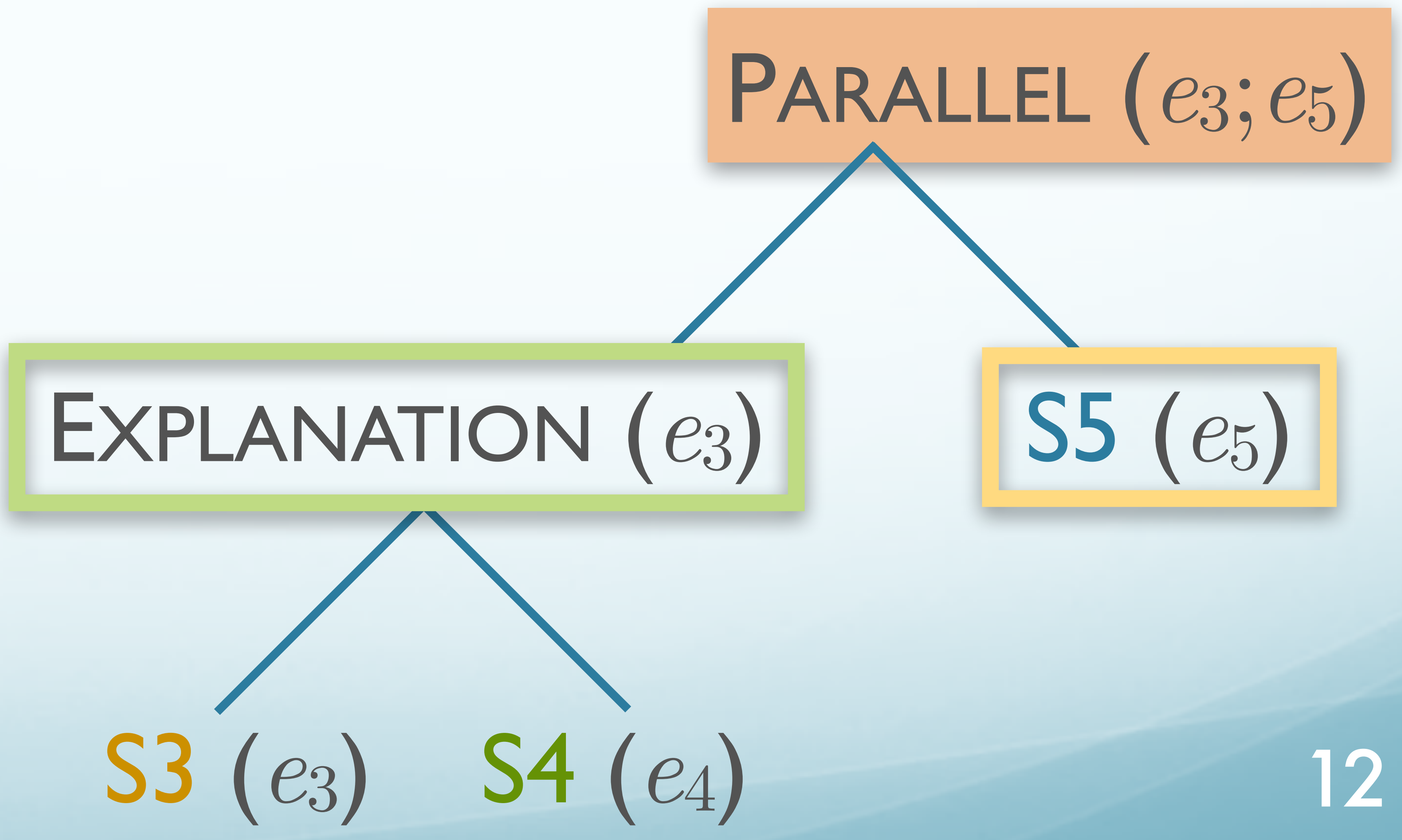
S1 – Armin went to the bank to deposit his paycheck

S2 – He then took a train to Kim's car dealership.

S3 – He needed to buy a car.

S4 – The company he works for now isn't near any public transportation.

S5 – He also wanted to talk to Kim about their softball league.



Adapted from J&M 2nd ed p. 690

Coherence Relation Hierarchy

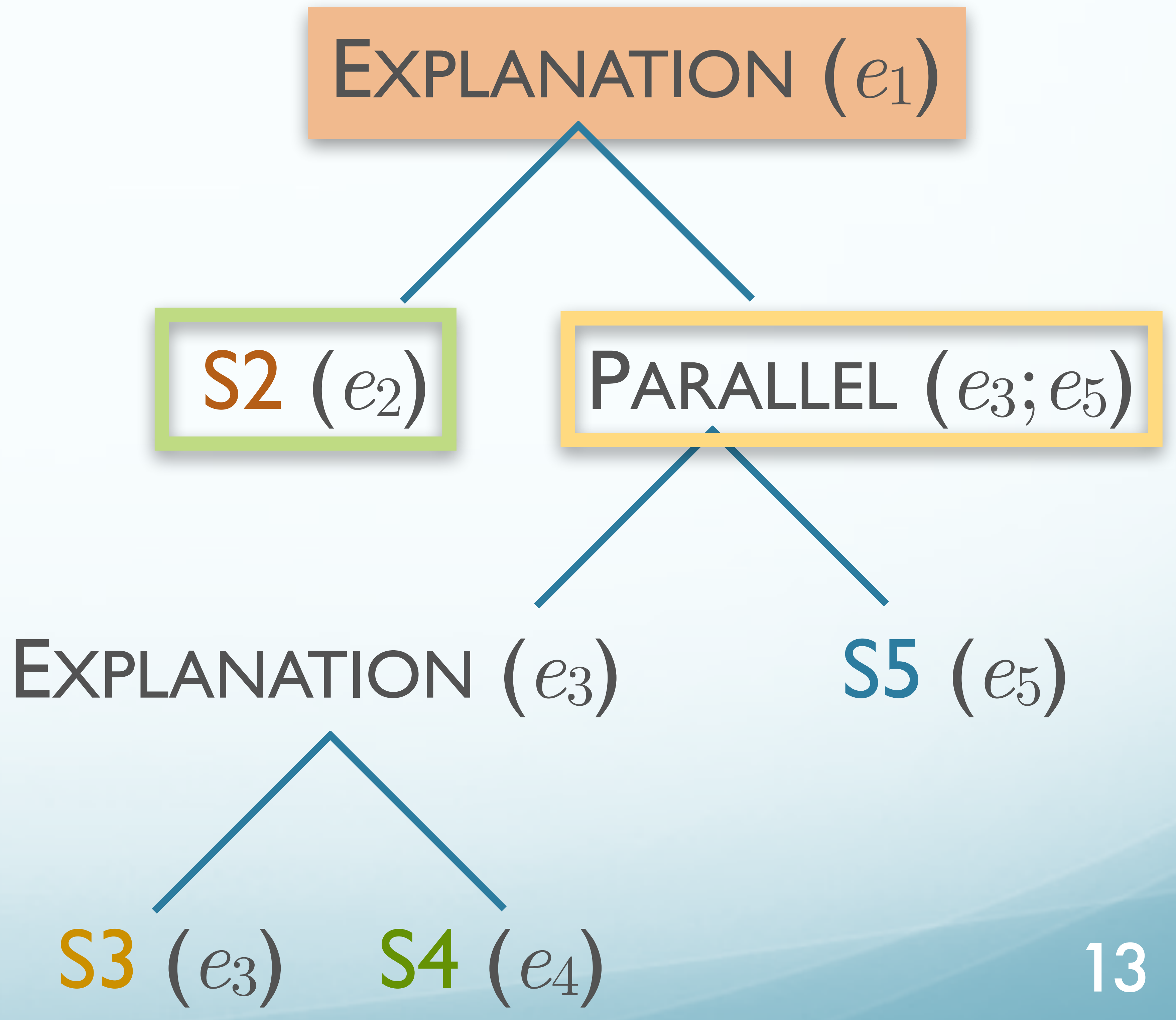
S1 – Armin went to the bank to deposit his paycheck

S2 – He then took a train to Kim's car dealership.

S3 – He needed to buy a car.

S4 – The company he works for now isn't near any public transportation.

S5 – He also wanted to talk to Kim about their softball league.



Adapted from J&M 2nd ed p. 690

Coherence Relation Hierarchy

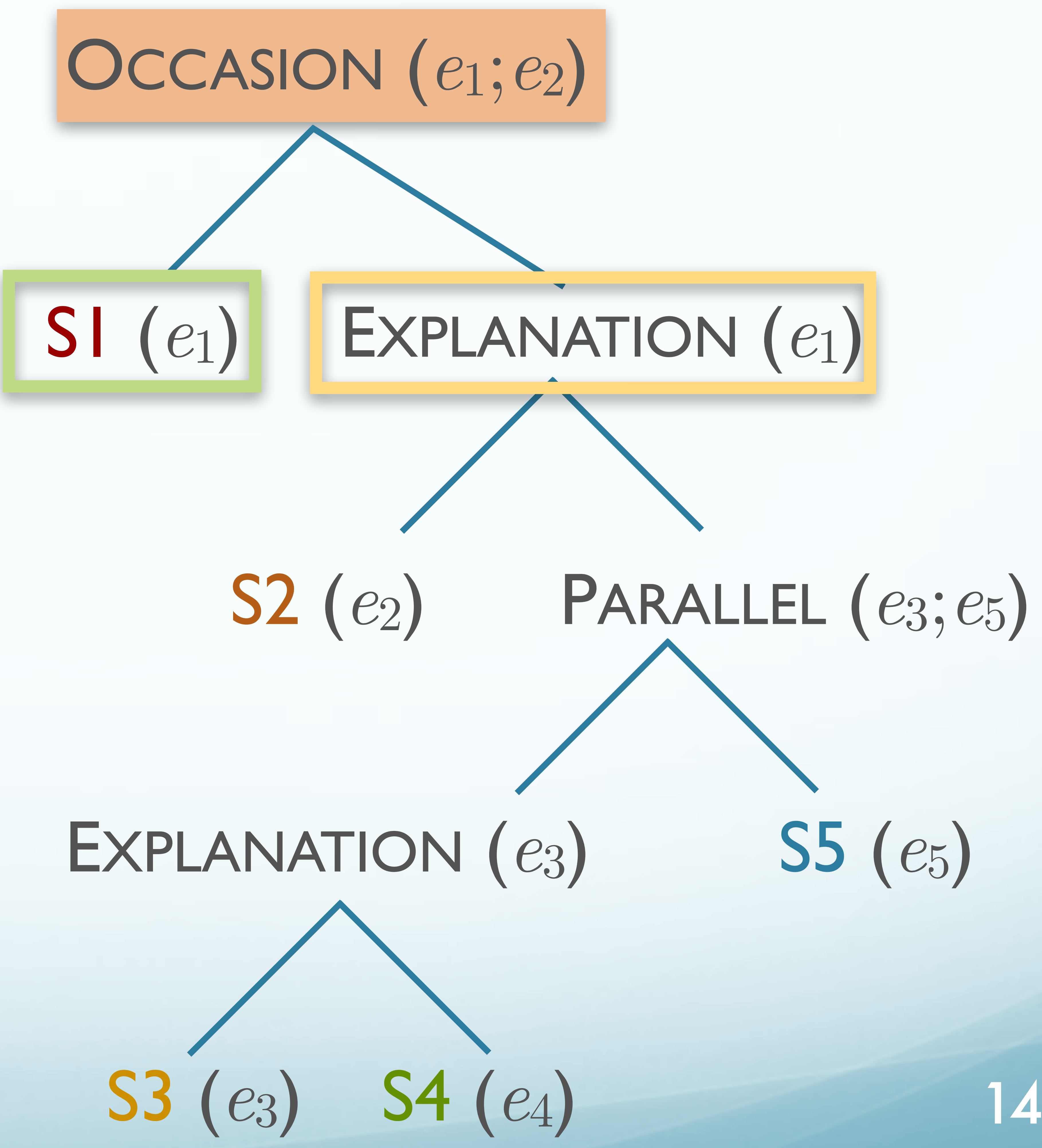
S1 – Armin went to the bank to deposit his paycheck

S2 – He then took a train to Kim's car dealership.

S3 – He needed to buy a car.

S4 – The company he works for now isn't near any public transportation.

S5 – He also wanted to talk to Kim about their softball league.



UNIVERSITY OF WASHINGTON

Adapted from J&M 2nd ed p. 690

PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

Coherence Relation Hierarchy

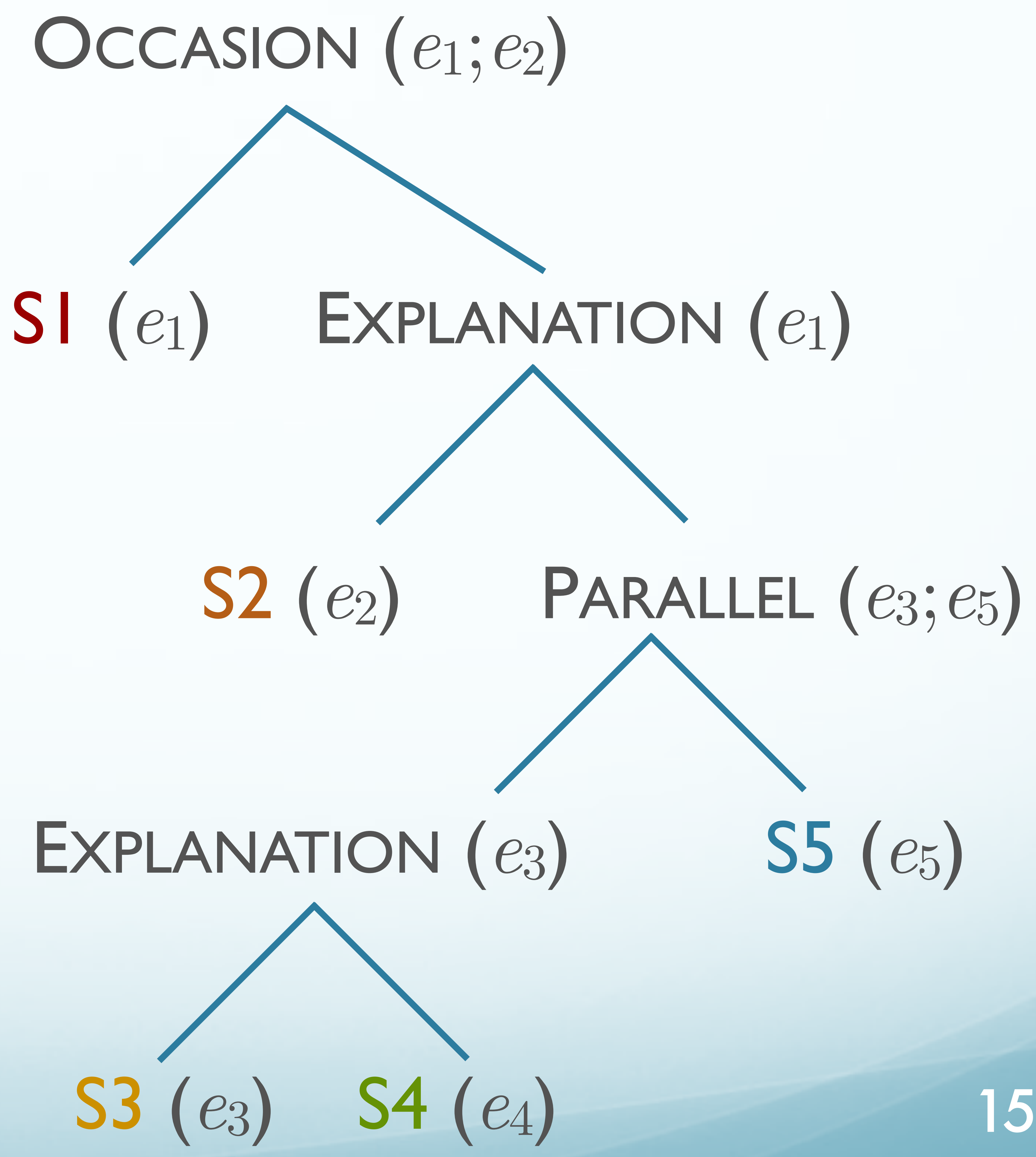
S1 – Armin went to the bank to deposit his paycheck

S2 – He then took a train to Kim's car dealership.

S3 – He needed to buy a car.

S4 – The company he works for now isn't near any public transportation.

S5 – He also wanted to talk to Kim about their softball league.



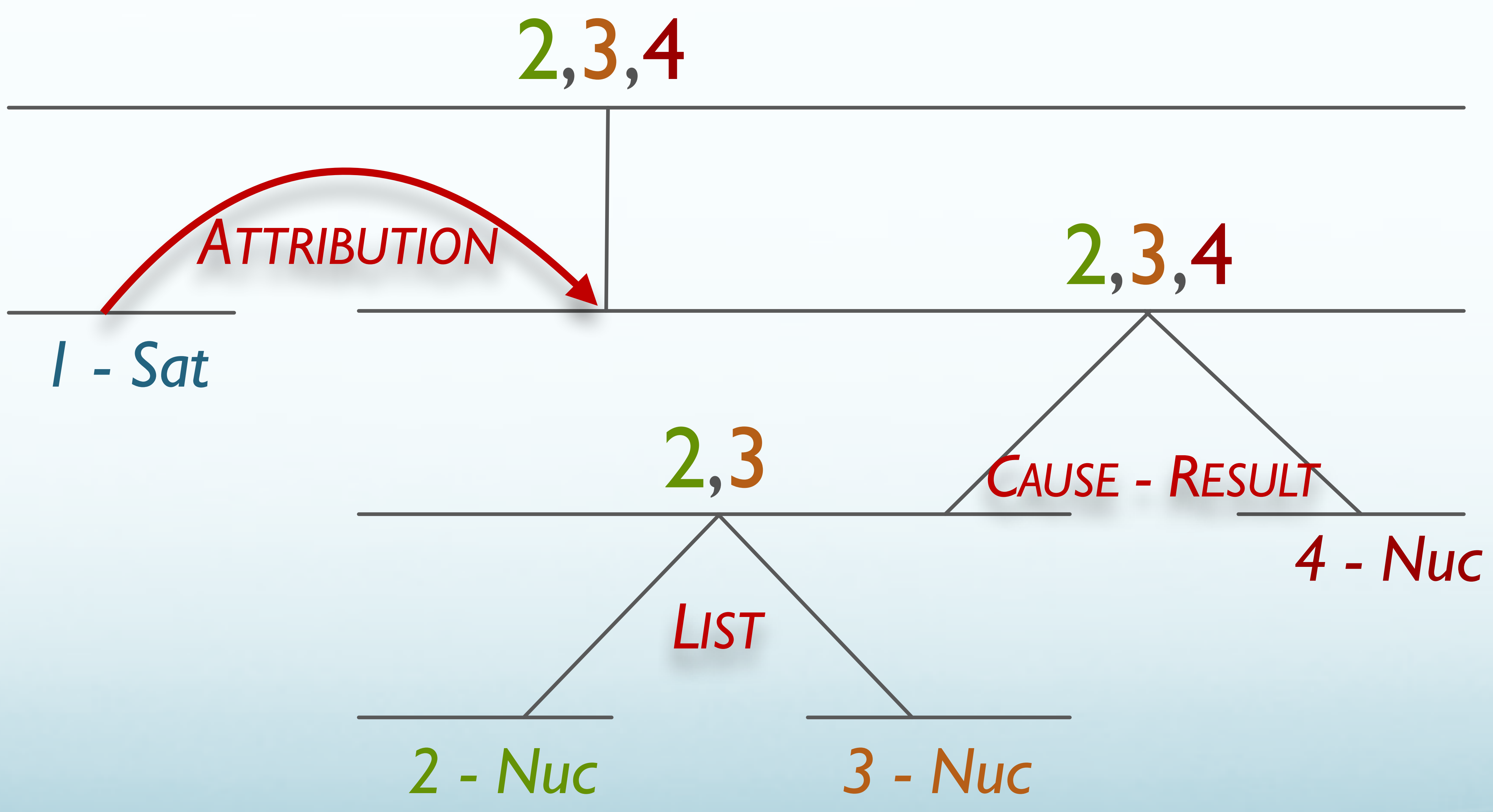
Adapted from J&M 2nd ed p. 690

Rhetorical Structure Theory (RST)

- Mann & Thompson (1987)
- Goal: Identify hierarchical structure of text
- Posits that discourse has:
  - **Nuclei**: Basic information, core concepts
  - **Satellites**: Units of discourse reliant on nuclei, unintelligible without
- Derives from functional relations b/t clauses

Rhetorical Structure Theory (RST)

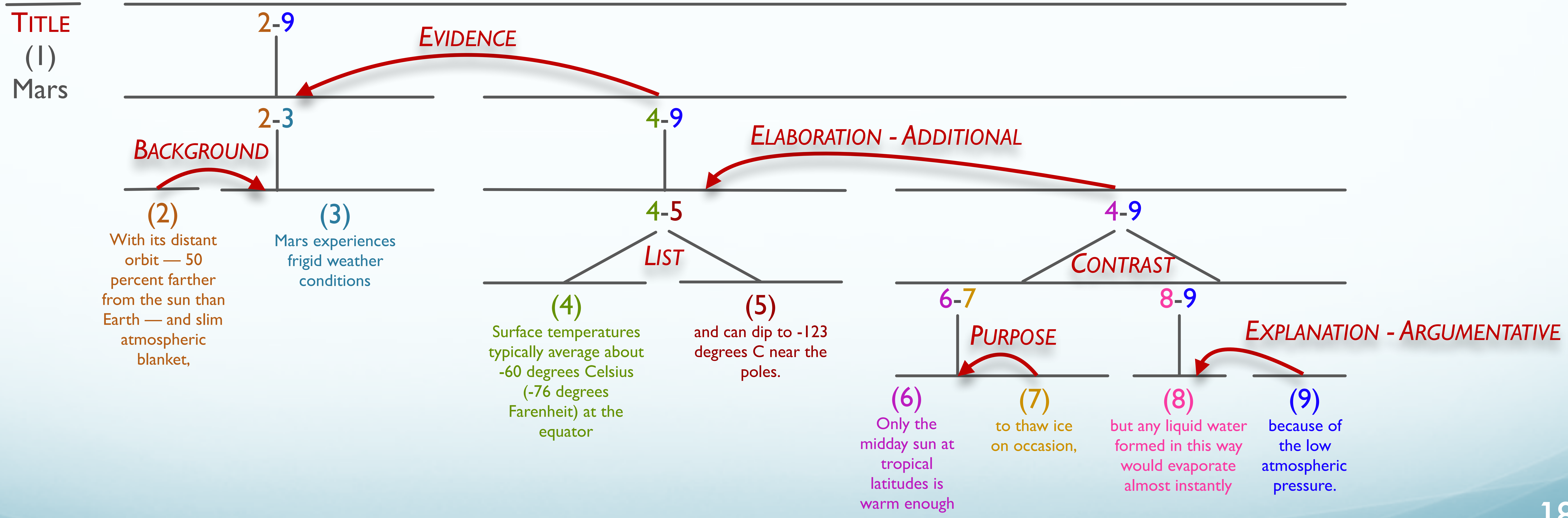
- *Mr. Watkins said volume on Interprovincial's system is down about 2% since January and is expected to fall further making expansion unnecessary until perhaps the mid-1990s*



via [Louis et al, 2010]

Mars Example

Mars. With its distant orbit—50 percent farther from the sun that Earth—and slim atmospheric blanket, Mars experiences frigid weather conditions. Surface temperatures typically average about -60 degrees Celsius (-76 degrees Farenheit) at the equator and can dip to -123 degrees C near the poles. Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, but any liquid water formed in this way would evaporate almost instantly because of the low atmospheric pressure.



Rhetorical Structure Theory:
Relations

- ELABORATION — Satellite gives further information about content of nucleus
- ATTRIBUTION — Satellite gives source of attribution in reported speech
- CONTRAST — Multinuclear, with nuclei contrasting along a dimension.
- LIST — Multinuclear, without nuclei contrasting

- …etc. See RST Treebank (Carlson et. al, 2001)
  - for all 78 relations in 16 classes.

RST Parsing

1. Learn and apply classifiers for segmentation and parsing of discourse
2. Assign coherence relations between spans
3. Create a representation over whole text ⇒ parse

- Discourse structure
  - RST trees: Fine-grained, hierarchical structure, clause-based units
- State-of-the-art: Ji & Eisenstein, 2014 [github]
  - Shift-reduce parser w/jointly trained word embeddings on RST Treebank
  - Span: **82.1**; Nuclearity: **71.1**; Relation: **61.6** (Human, IAA: **65.8**)

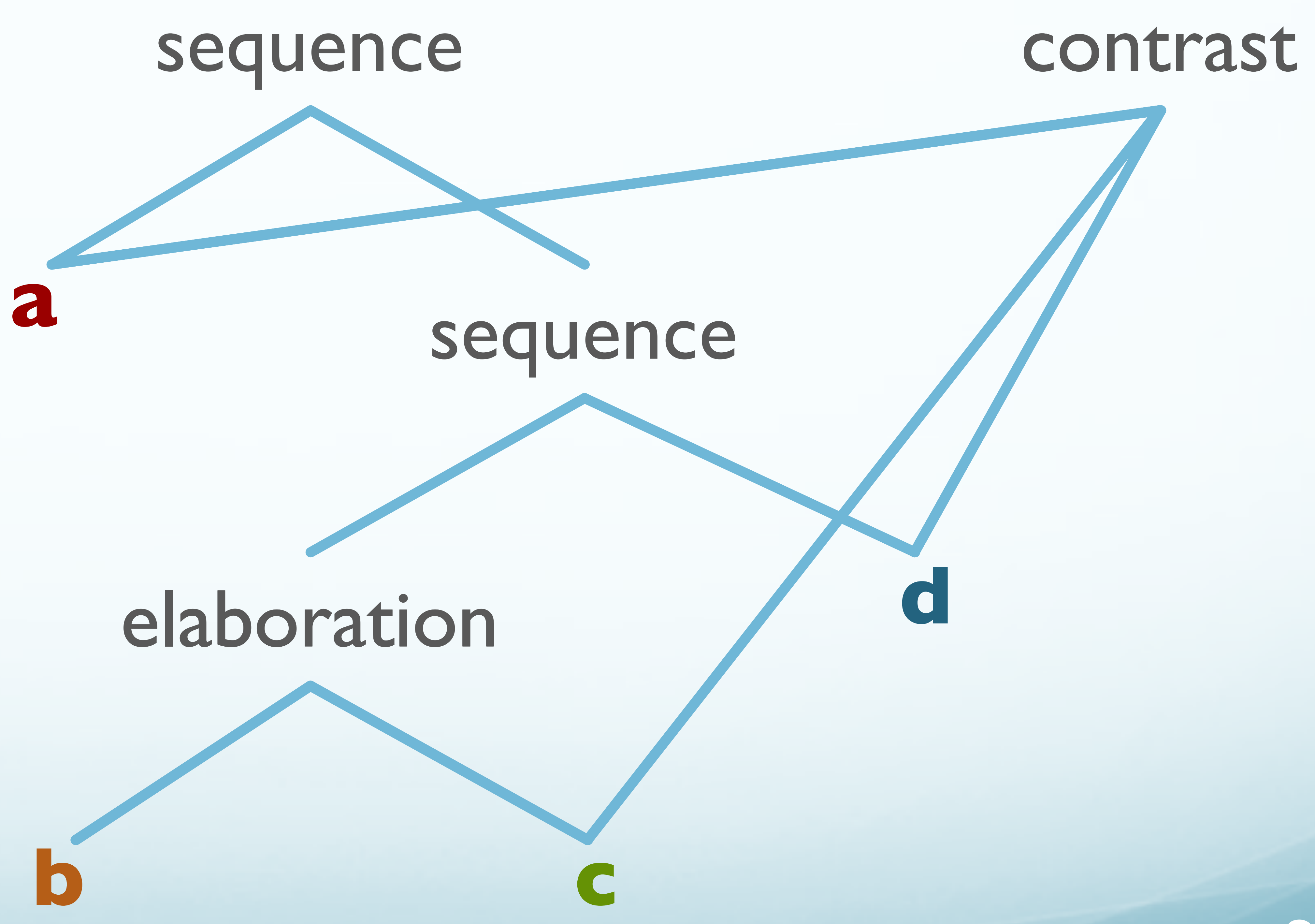
Why Represent As a Hierarchy?

- Helpful for determining main concepts
- Useful for downstream tasks
  - Summarization*
  - Information Retrieval
  - Sentiment Analysis — particularly for items with multiple facets

**Will return to this in a bit*

But… Is Discourse Hierarchical?

- In structuring discourse units, is this strictly hierarchical?
  a) The car was finally coming toward him.
  b) He [Chee] finished his diagnostic tests,
  c) feeling relief.
  d) ***But then*** the car started to turn right.



- ***But*** is linking **a**, **c**, **d** in a *contrast* relation
- ***then*** is linking **b**+**c**, **d** *sequentially*

Coherence Relations:
The Penn Discourse Treebank (PDTB) ([Prasad et al, 2008])

- "Theory-neutral" discourse model
- No stipulation of overall structure, local sequence relations
- *U.S. Trust, a 136-year-old institution that is one of the earliest high-net worth banks in the U.S., has faced intensifying competition from other firms that have established, and heavily promoted, private-banking businesses of their own.* ***As a result***, *U.S. Trust's earnings have been hurt.*
- PDTB annotation links **S1** to **S2** by way of connective
  - Provides sense label

Coherence Relations:
The Penn Discourse Treebank (PDTB) (Prasad et al, 2008)

- Discourse units (sentential, or sub-sentential) marked in pairs:
  - Arg_1, Arg_2
- **Explicit Relations**:
  - triggered by lexical markers (‘*but*’, ‘*as a result*’) b/t spans
  - Arg_2 syntactically bound to connective unit, Arg_1
- **Implicit Relations**:
  - Adjacent sentences assumed related
  - Arg_1: first sentence (can be anywhere in discourse)
  - Arg2: second sentence, in linear sequence

Shallow Discourse Parsing

- For extended discourse
- …for each clause/sentence pair in sequence
- …identify discourse relation, Arg_1, Arg_2
- CoNLL15 Shared task Results:
  - **61**% overall (**55**% blind)
    - Explicit discourse connectives: **91**% (**76**% blind)
    - Non-explicit discourse connectives: **34**% (**36**% blind)

Basic Methodology

- Pipeline:
  1. Identify discourse connectives
  2. Extract arguments for connectives (Arg_1, Arg_2)
  3. Determine presence/absence of relation in context
  4. Predict sense of discourse relation
- Resources: Brown clusters, lexicons, parses
- Approaches:
  - 1,2: Sequence labeling techniques
  - 3,4: Classification (4: multiclass)
    - Some rule-based or most common class

Identifying Relations

- Key source of information:
  - Cue phrases
    - aka: discourse markers, cue words, clue words
    - *although, but, for example, however, yet, with, and…*
      - John hid Bill's keys **because** he was drunk

Identifying Relations: Issues

- Ambiguity: discourse vs. sentential use
  - ***With*** *its distant orbit, Mars exhibits frigid weather.*
  - *We can see Mars* ***with*** *a telescope.*
- Ambiguity: cue multiple discourse relations
  - ***Because***: CAUSE, or EVIDENCE
  - ***But***: CONTRAST, or CONCESSION
- Sparsity:
  - Only **15-25**% of relations marked by cues

Entity-Based Coherence and Centering Theory

Entity-Based Coherence

*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.

- Versus:

*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.

- Which is better? Why?
  - First focuses on a single entity
  - Second interleaves entities *John* and the ***music store***

Centering Theory

- Entity-based coherence is inspiration for ***Centering theory*** (Grosz et al, 1995)
  - Explicitly encodes a discourse model
  - Different entities are uniquely "*centered*" at different points in discourse

Centering Theory Details

- Two adjacent utterances:
  - U_n
  - U_{n+1}
- Two ideas of "centers"
  - **backward-looking** center — $C_b(U_n)$
  - **forward-looking** centers — $C_f(U_n)$

Centering Theory Details

- **backward-looking** center — $C_b(U_n)$
  - The entity that is currently being focused ("centered") after U_n is interpreted
- **forward-looking** centers — $C_f(U_n)$
  - A list of all entities mentioned in U_n which could be focused in subsequent utterances
  - Order with precedence list:
    - subject > existential predicate nominal > object > indirect object or oblique > demarcated adverbial PP
- C_p — shorthand for highest-ranked forward-looking candidate

Centering Theory Hand-wavy Algorithm

- John saw a beautiful 1961 Ford Falcon at the used car dealership. (U_1)
- He showed it to Bob. (U_2)
- He bought it. (U_3)

Centering Theory Hand-wavy Algorithm

- John saw a beautiful 1961 Ford Falcon at the used car dealership. (U_1)
- He showed it to Bob. (U_2)
- He bought it. (U_3)

After U1

$C_f(U_1)$: {*John, Ford, dealership*}

$C_p(U_1)$: *John*

$C_b(U_1)$: undefined

Centering Theory Hand-wavy Algorithm

- John saw a beautiful 1961 Ford Falcon at the used car dealership. (U_1)
- He showed it to Bob. (U_2)
- He bought it. (U_3)

Processing U₂

$C_f(U_1)$: *{John, Ford, dealership}*

$C_p(U_1)$: *John*

$C_b(U_1)$: undefined

he=John, it=Ford

Centering Theory Hand-wavy Algorithm

- John saw a beautiful 1961 Ford Falcon at the used car dealership. (U_1)
- He showed it to Bob. (U_2)
- He bought it. (U_3)

After U₂

$C_f(U_2)$: *{John, Ford, Bob}*

$C_p(U_2)$: *John*

$C_b(U_2)$: *John*

Computational Discourse:
Summary

- **Cohesion**
  - Modeled with linking lexical terms and thematic overlap
- **Coherence**
  - Determine relevance of discourse units to one another
  - Can add structure to discourse to model relations and their importance

Computational Discourse:
Key Tasks

- **Reference resolution**
  - Constraints and preferences
  - Heuristic, learning and sieve models
- **Discourse structure modeling**
  - Linear topic segmentation
  - Rhetorical Structure Theory (RST)
  - shallow discourse parsing

Case Study:
Discourse & Summarization

Motivation: Why Summarization?

- Information Retrieval is Very Powerful
  - Search engines index and search enormous doc sets
  - Retrieve billions of documents in tenths of seconds
- But still very limited!
  - Technically: keyword search (mostly)
  - Conceptually: User seeks information
    - Sometimes a website or document
    - Sometimes the answer to a question
    - Sometimes, a summary of a document or document set

Why Summarization?

- Even web search relies on simple summarization
  - Provide "thumbnail" of relevant information in ranked document.

Caldera - Wikipedia
https://en.wikipedia.org/wiki/Caldera
A *caldera* is a large cauldron-like depression that forms following the evacuation of a magma chamber/reservoir.
Yellowstone Caldera · Bárðarbunga · Cauldron · Long Valley Caldera

Caldera: Crater Formed by Volcanic Collapse or Explosion
https://geology.com/articles/caldera/
They are large volcanic craters that form by two different methods: 1) an explosive volcanic eruption; or, 2) collapse of surface rock into an empty magma chamber. The accompanying image is a satellite view of one of the most famous *calderas* - Crater Lake in Oregon.

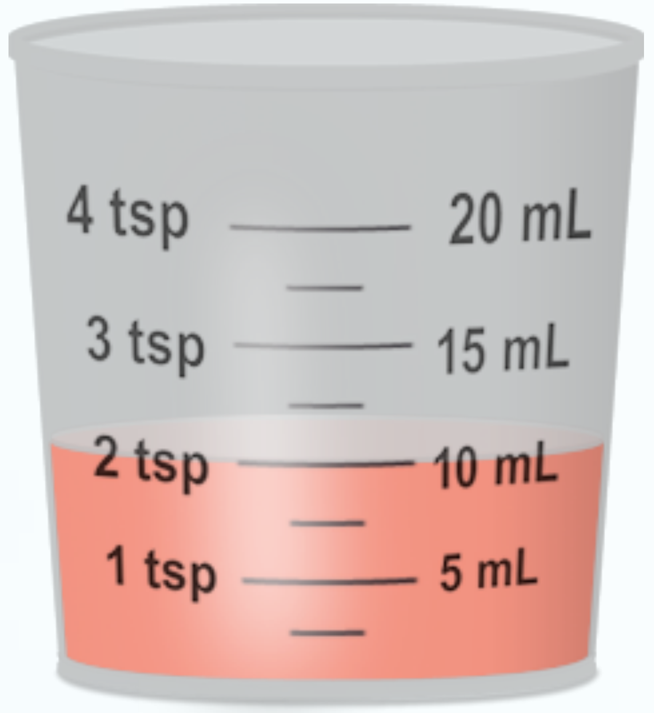
How Volcanoes Work - Calderas - SDSU geology
www.geology.sdsu.edu/how_volcanoes_work/Calderas.html
CALDERAS. When an erupting volcano empties a shallow-level magma chamber, the edifice of the volcano may collapse into the voided reservoir, thus forming a steep, bowl-shaped depression called a *caldera* (Spanish for kettle or cauldron). These features are highly variable in size, ranging from 1-100 km in diameter.

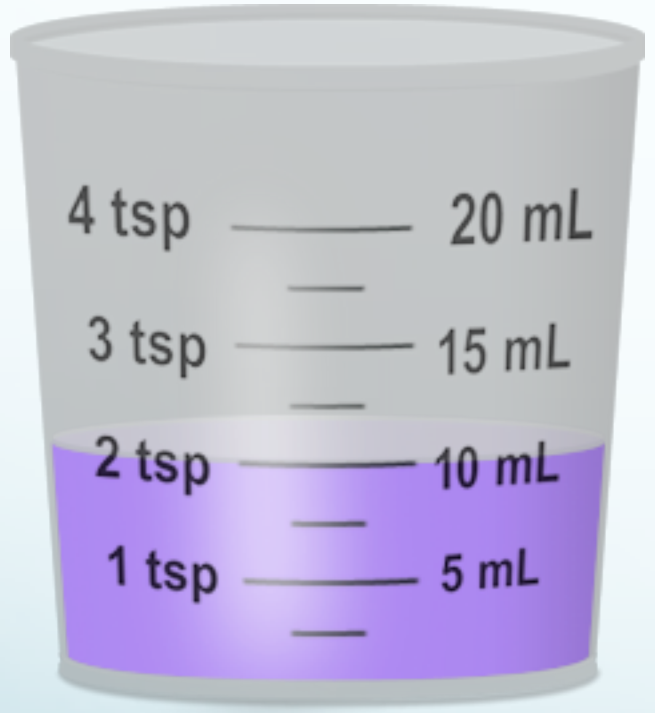
Why Summarization?

- Complex questions go beyond factoids & infoboxes
  - Require explanations, analysis
- For instance: *Is acetaminophen or ibuprofen better for reducing fever in kids?*
- Following is top-ranked page:

Acetaminophen or Ibuprofen?

top hit: parents.com/health/hygiene/childrens-health-myths/

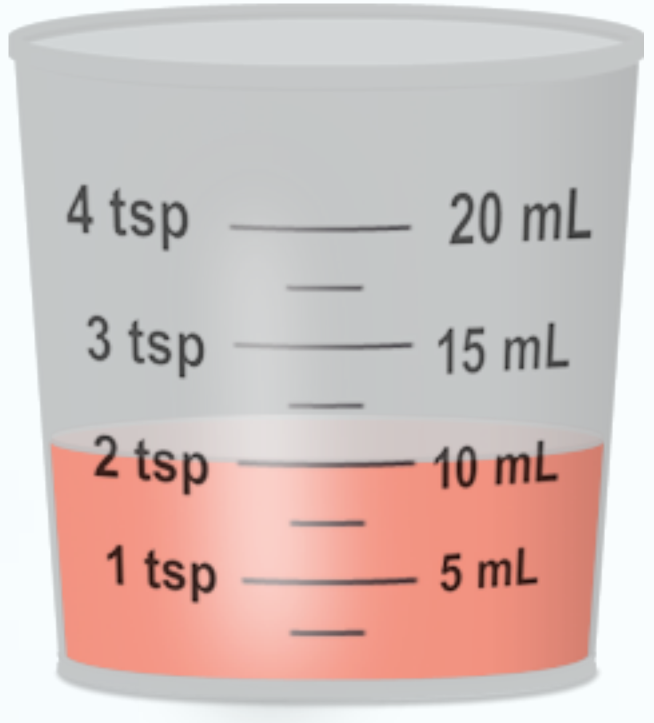


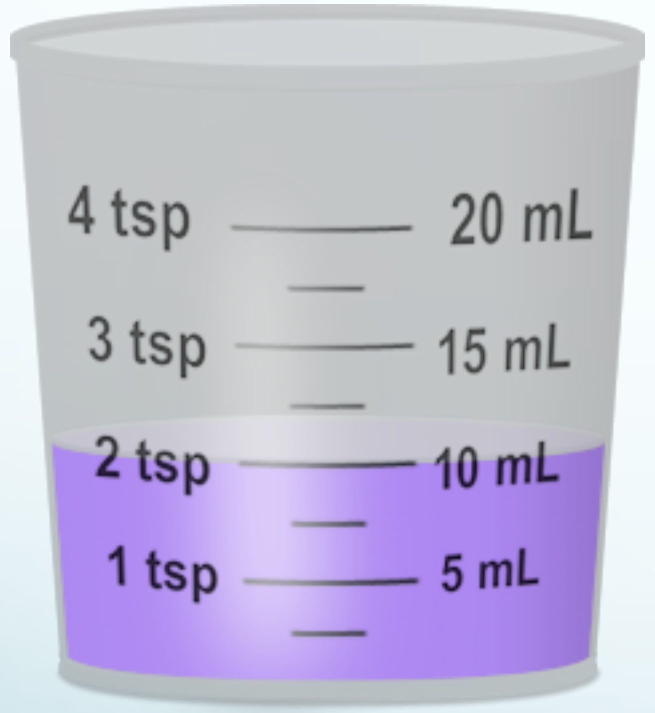


Ibuprofen beats acetaminophen for treating both pain and fever, according to recent research. One large study in the Archives of Pediatrics & Adolescent Medicine found that ibuprofen lowered kids' fevers better than acetaminophen at two, four, and six hours after taking the medicine. Another study of kids in the ER concluded that ibuprofen provided significantly better pain relief than acetaminophen (or codeine) for broken bones, bruises, and sprains. Ibuprofen and acetaminophen both act on pain receptors in the brain, but ibuprofen also has an anti-inflammatory effect that helps decrease swelling. "That may also make it a better choice for ear infections, which typically involve inflammation," says Richard Rosenfeld, MD, a pediatric ear, nose, and throat specialist at Long Island College Hospital, in Brooklyn. Ibuprofen also lasts longer than acetaminophen, making it more likely that your child will sleep through the night, especially in the early stages of an ear infection. Interestingly, a new study also found that children who took acetaminophen before age 1 were almost 50 percent more

Acetaminophen or Ibuprofen?

top hit: parents.com/health/hygiene/childrens-health-myths/



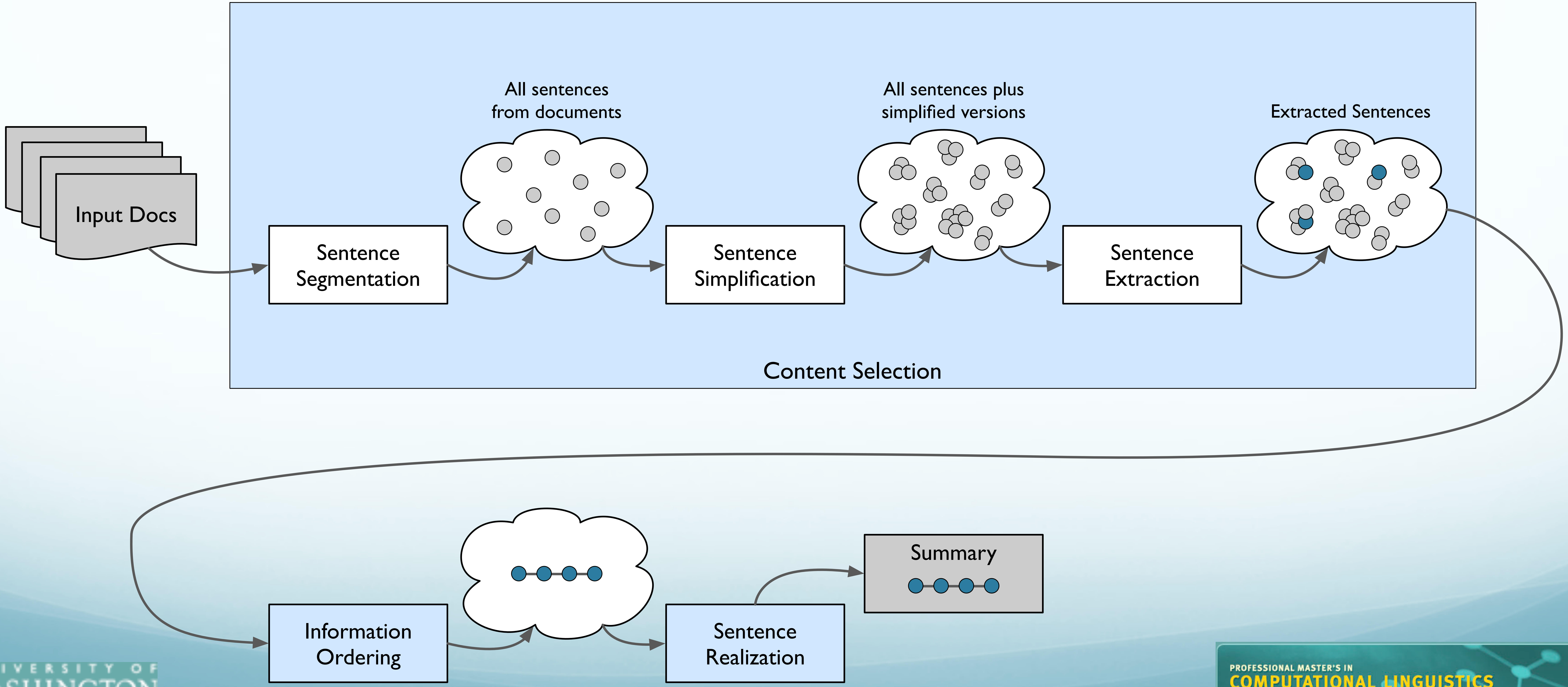


- Summary:
  - Ibuprofen beats acetaminophen for treating both pain and fever, according to recent research.

Why Summarization?

- As Torres-Moreno (2014) puts it:
  - “too much information kills information”

General Architecture



Why Discourse?
Content Selection

- **Relations**
  - Different discourse relations have different relevance for inclusion in summary
  - e.g. *ELABORATION* likely less helpful than *RESULT* or *EXPLANATION*

- **Structure**
  - Some information more "core"
  - nucleus vs. satellite; depth in tree

Why Discourse?
Information Ordering

- Make sure that nucleus is introduced before satellites

Why Discourse?
Summary Realization

- Ensure sequential sentences are coherent, in additional to cohesive

Analyzing Discourse Features

Louis et al (2010)

- Design different features, both **discourse-related** and **non-discourse**
  - Using model summaries (human-generated)
    - Perform statistical significance tests on included vs. non-included sentences
    - χ^2 (categorical) t-test (continuous)
- Use features in logistic regression classifier (MaxEnt)
  - Use to select sentences for extraction
- Evaluate against model, human-written summaries

Experimental Setup

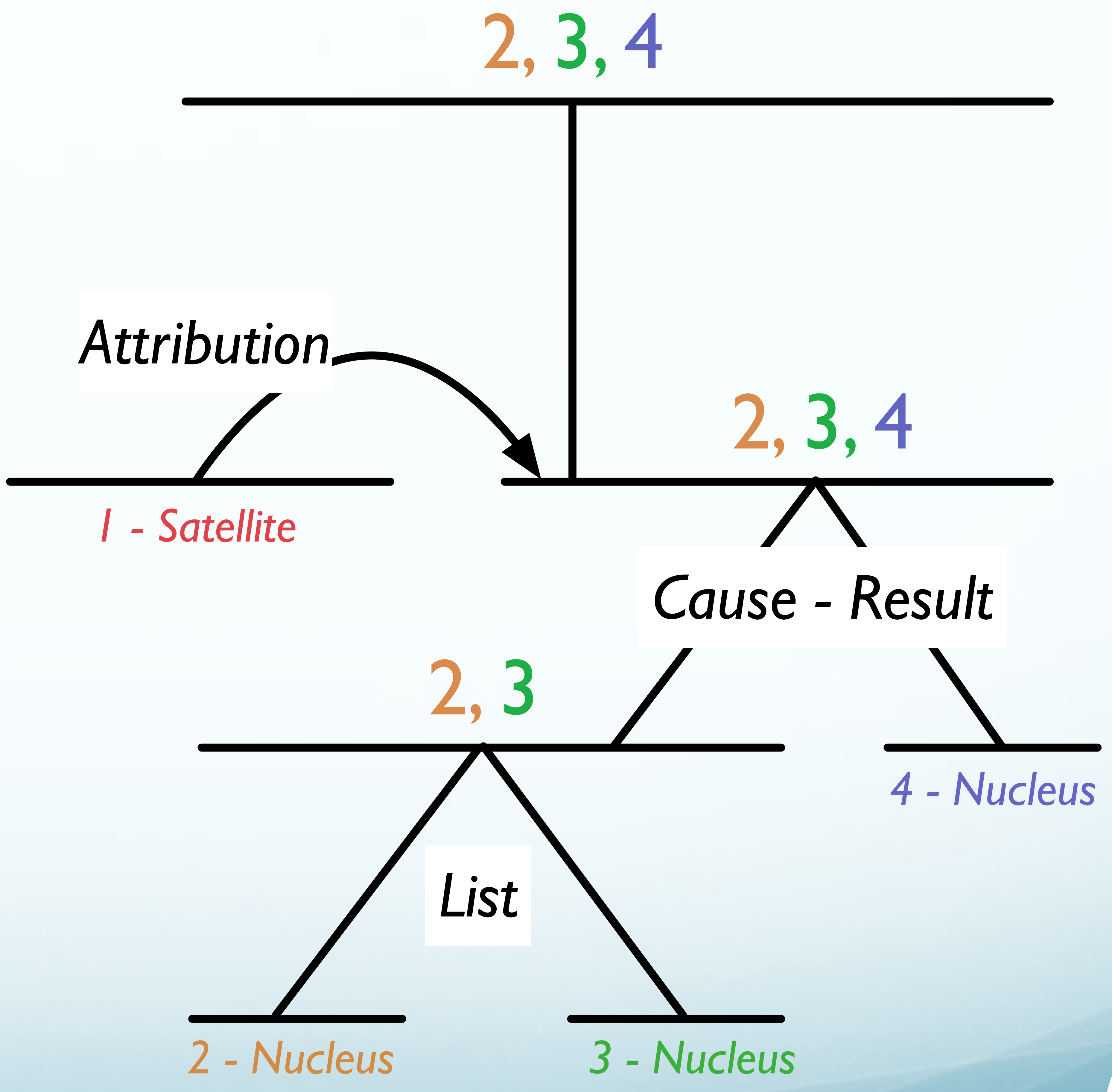
- Caveat:
  - Experimental approach is using human-created discourse analyses
  - Authors do not attempt using automatic discourse parsers for analyses
  - Purely a study of how well discourse features correlate in an idealized setting

How Would This be Applied?

- Learn and apply classifiers for segmentation and parsing of discourse
- Assign coherence relations between spans
- Create a representation over whole text → parse
- Use parsed representations as features in classifier for content selection

Discourse (RST) Structure Example

- 1. [Mr. Watkins said]
- 2. [volume on Interprovincial's system is down about 2% since January]
- 3. [and is expected to fall further,]
- 4. [making expansion unnecessary until perhaps the mid-1990s.]

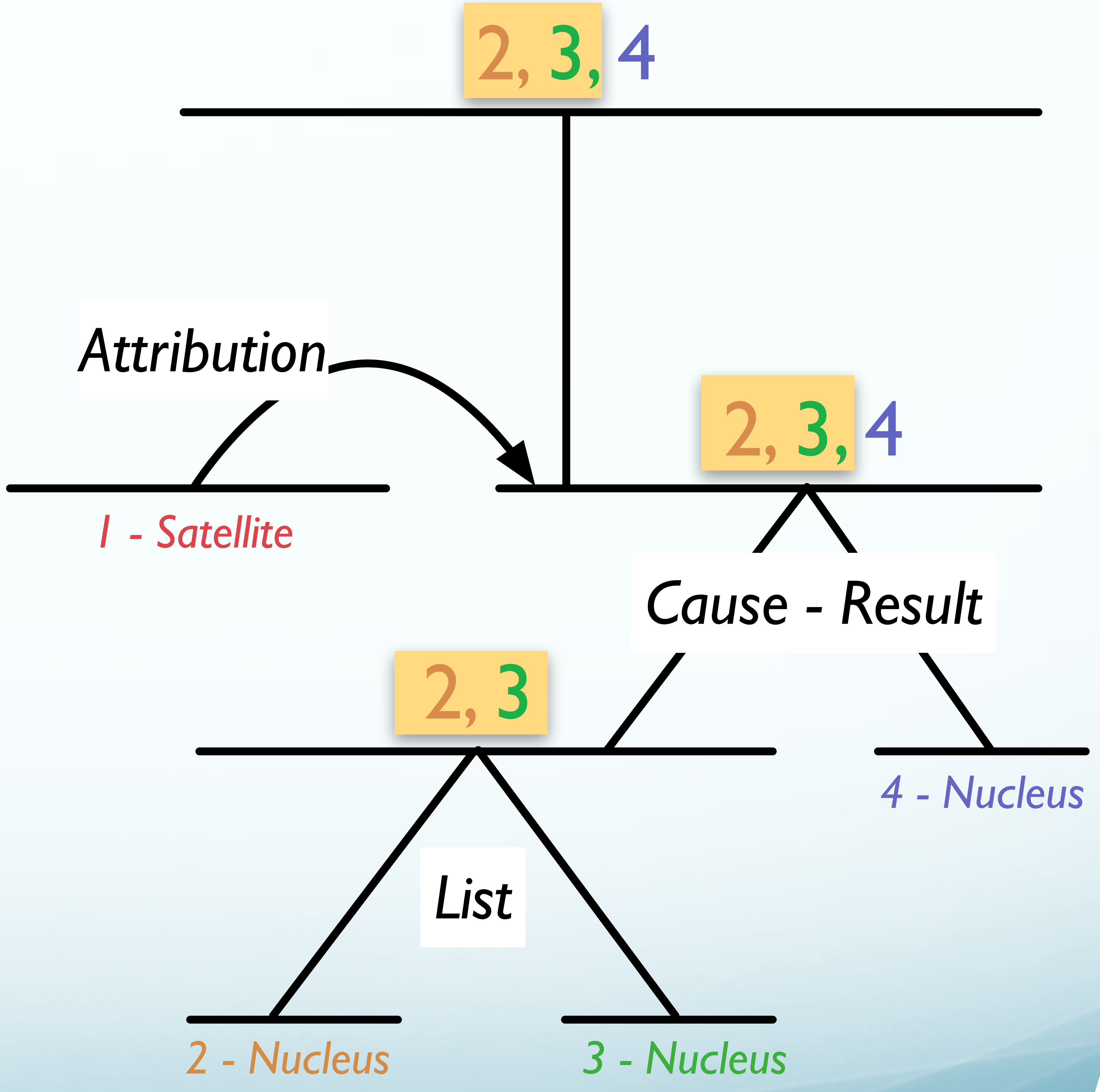


Discourse Structure Features

- **Satellite penalty**
  - For each EDU — number of satellite nodes between EDU and root
    - 1 satellite in tree: one step to root: penalty = 1
  - **Intuition**: Helpful summary content will be closely related to nucleus.

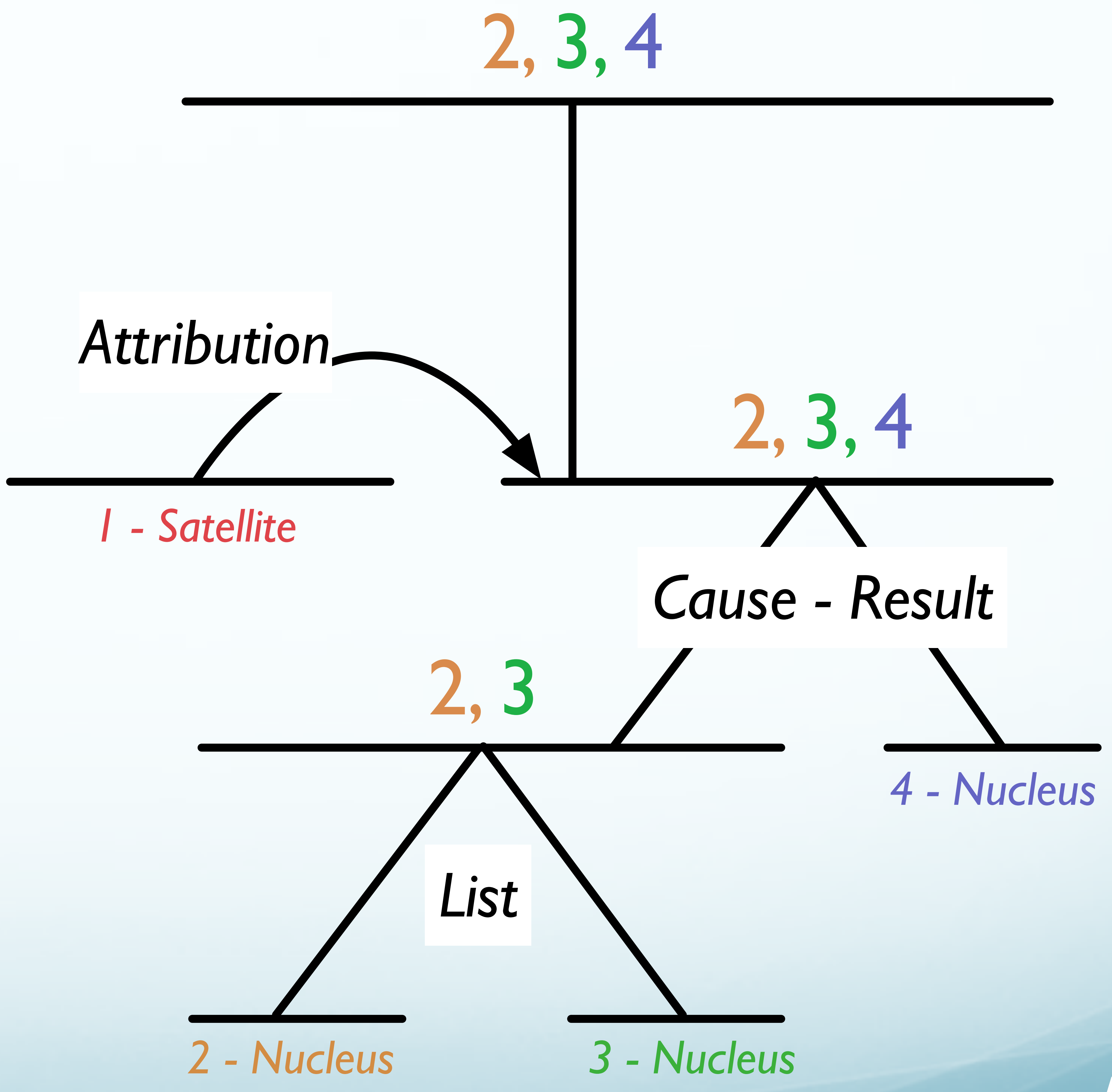
Discourse Structure Features

- **Promotion set:**
  - Nuclear units at some level of tree
    - At leaves, EDUs are themselves nuclear
  - Intuition:
    - The more times a unit is promoted in the tree, the more necessary its concepts to understanding the whole discourse



Discourse Structure Features

- **Depth score:**
  - Distance from lowest tree level to EDU's highest rank
    - **2,3,4**: score=4
    - **1**: score=3
- **Promotion score:**
  - # of levels span is promoted
    - **1**: score = 0
    - **4**: score = 2
    - **2,3**: score = 3



Converting to Sentence Level

- Each feature has:
  - Raw score
  - Normalized score: $\frac{\text{raw score}}{\text{sentence length}}$
- Sentence score for a feature:
  - Maximum value over all EDUs in sentence

"Semantic" Features

- Represent sentences purely in terms of their discourse relationships
- **Binary features:**
  - Implicit vs. Explicit
  - sentence_in_{RELATION_NAME}
  - sentence_**contains**_{ARG$_1$|ARG$_2$}_of_{RELATION_NAME} **(multi-sentential)**
  - sentence_**expresses**_{RELATION_NAME} **(both args in single sent)**
- **Real-valued features:**
  - Number of relations
  - Distance between arguments within sentence

Example 1

- *In addition, its machines are easier to operate, so customers require less assistance from software.*

- Is there an explicit discourse marker?
  - Yes, "**so**"

- Discourse relation?
  - ***Contingency***

Example II

- *(1) Wednesday's dominant issue was Yasuda & Marine Insurance, which continued to surge on rumors of speculative buying. (2) It ended the day up 80 yen to 1880 yen.*

- Is there an explicit discourse marker?
  - **No**

- Is there a relation?
  - **Yes, Implicit**.

- What relation?
  - **Expansion**. (More specifically, restatement).

Non-Discourse Features

- Sentence length
- Sentence position
- Probabilities of words in sentence
  - mean, sum, product
- # of “signature” words
  - (Words that are statistical outliers, likely indicating a topic)

Significant Features: Summary Sentences

- Structure:
  - depth score
  - promotion score
- Semantic:
  - Arg1 of Explicit Expansion
  - Implicit Contingency
  - Implicit Expansion
  - Distance to Arg
- Non-discourse:
  - length
  - 1st in paragraph
  - offset from end of paragraph
  - # signature terms
  - mean
  - sum word probabilities

Significant Features: Non-Summary Sentences

- Structure:
  - satellite penalty
- Semantic:
  - Explicit expansion
  - Explicit contingency
  - Arg2 of implicit temporal
  - Arg2 of implicit contingency
  - # of shared relations
- Non-discourse:
  - offset from paragraph start
  - offset from article start
  - sentence probability

Observations

- Non-discourse features good cues to summary
- Structural features match intuition
- Semantic features
  - Relatively few useful features for selecting summaries
  - Most features associated with non-summary… but most sentences are non-summary

Evaluation

- Structural is best, both alone and in combination
- Best overall combines all types

Features used	Acc	P	R	F
structural	78.11	63.38	22.77	33.50
semantic	75.53	44.31	5.04	9.05
non-discourse (ND)	77.25	67.48	11.02	18.95
ND + semantic	77.38	59.38	20.62	30.61
ND + structural	78.51	63.49	26.05	36.94
semantic + structural	77.94	58.39	30.47	40.04
structural + semantic + ND	78.93	61.85	34.42	44.23