Discourse Structure

LING 571 — Deep Processing Methods in NLP

November 28, 2022

Shane Steinert-Threlkeld

Announcements

- HW8:
  - Only caring about disambiguating the probe word:
    - Pseudo-code from slide 42, not slide 41, of these: https://www.shane.st/teaching/571/aut22/slides/14_Word-Sense-Disambiguation.pdf
    - Can also omit the first “for k=1 to num_senses(w i)” loop
    - Normalization **not needed**. Why not?
  - See updated hw8 slides for cleaner/updated pseudo-code

Ambiguity of the Week

The kids were playing Rock Paper Scissors.

: Scissors

: Everything!

: What?

: Nothing beats everything.

: Ok play again, Rock Paper Scissors shoot!

: Everything!

: Nothing!

Breaking Language Technology

https://twitter.com/xkcd/status/1333529967079120896

Roadmap

- Coreference
  - Recap
  - (Hobbs Walkthrough)
  - Other approaches
  - Evaluation
- Discourse Structure
  - Cohesion [Segmentation]
  - Coherence

Discourse & Coref Recap

What is Discourse?

- Discourse is "a ***coherent*** ***structured*** ***group of sentences***." *(J&M p. 681)*

What is Discourse?

- Discourse is "a **coherent structured group of sentences**." *(J&M p. 681)*

- Understanding depends on **context**
  - Word sense — *plant*
  - Intention — *Do you have the time?*
  - Referring expressions — *it*, *that*, *the screen*

Reference: Terminology

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment.

Reference: Terminology

- ***referring expression***: (refexp)

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment.

Reference: Terminology

- ***referring expression***: (refexp)
  - An expression that picks out entity (***referent***) in some knowledge model

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment.

Reference: Terminology

- ***referring expression***: (refexp)
  - An expression that picks out entity (***referent***) in some knowledge model
  - Referring expressions used for the same entity ***corefer***

Queen Elizabeth set about transforming **her husband**, **King George VI**, into **a viable monarch**. **Logue**, **a renowned speech therapist**, was summoned to help **the King** overcome **his speech impediment**.

Reference: Terminology

- ***referring expression***: (refexp)
  - An expression that picks out entity (***referent***) in some knowledge model
  - Referring expressions used for the same entity ***corefer***
    - *Queen Elizabeth, her, the Queen*

Queen Elizabeth set about transforming **her** **husband**, **King George VI**, into **a viable monarch**. **Logue**, **a renowned speech therapist**, was summoned to help **the King** overcome **his** **speech impediment**.

Reference: Terminology

- ***referring expression***: (refexp)
  - An expression that picks out entity (***referent***) in some knowledge model
  - Referring expressions used for the same entity ***corefer***
    - *Queen Elizabeth, her, the Queen*
    - *Logue, a renowned speech therapist*

> **Queen Elizabeth** set about transforming **her husband**, **King George VI**, into a viable monarch. **Logue**, **a renowned speech therapist**, was summoned to help **the King** overcome **his speech impediment**.

Reference: Terminology

- ***referring expression***: (refexp)
  - An expression that picks out entity (***referent***) in some knowledge model
  - Referring expressions used for the same entity ***corefer***
    - *Queen Elizabeth, her, the Queen*
    - *Logue, a renowned speech therapist*
  - Entities in **purple** do not corefer to anything.

> **Queen Elizabeth** set about transforming **her husband**, **King George VI**, into a viable monarch. **Logue**, **a renowned speech therapist**, was summoned to help **the King** overcome **his speech impediment**.

Reference: Terminology

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment.

Reference: Terminology

- ***Antecedent:***
  - An expression that introduces an item to the discourse for other items to refer back to
  - Queen Elizabeth… her

Queen Elizabeth set about transforming **her husband**, **King George VI**, into a viable monarch. **Logue**, **a renowned speech therapist**, was summoned to help **the King** overcome **his speech impediment**.

Reference: Terminology

- ***Anaphora***: An expression that refers back to a previously introduced entity.
  - *cataphora*: Introduction of expression before referent:
    - "Even before *she* saw it, *Dorothy* had been thinking about…"

Queen Elizabeth set about transforming **her husband**, **King George VI**, into a viable monarch. **Logue**, **a renowned speech therapist**, was summoned to help **the King** overcome **his speech impediment**.

Reference: Terminology

- ***Anaphora***: An expression that refers back to a previously introduced entity.
  - ***cataphora***: Introduction of expression before referent:
    - "Even before ***she*** saw it, ***Dorothy*** had been thinking about…"

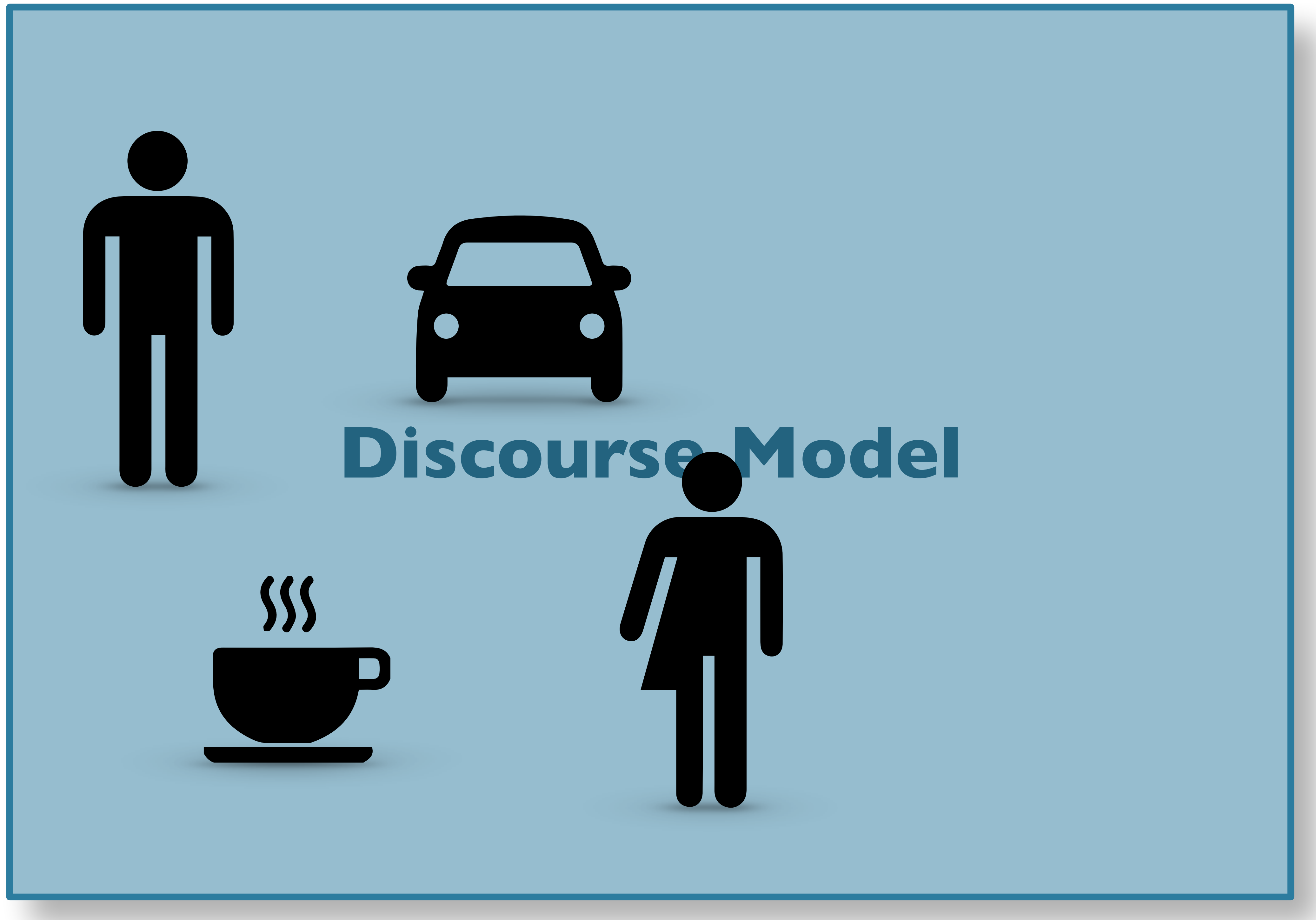
*Not all anaphora is referential! e.g. "*No dancer hurt their knee.*"

Queen Elizabeth set about transforming **her** **husband**, **King George VI**, into **a viable monarch**. **Logue**, **a renowned speech therapist**, was summoned to help **the King** overcome **his** **speech impediment**.

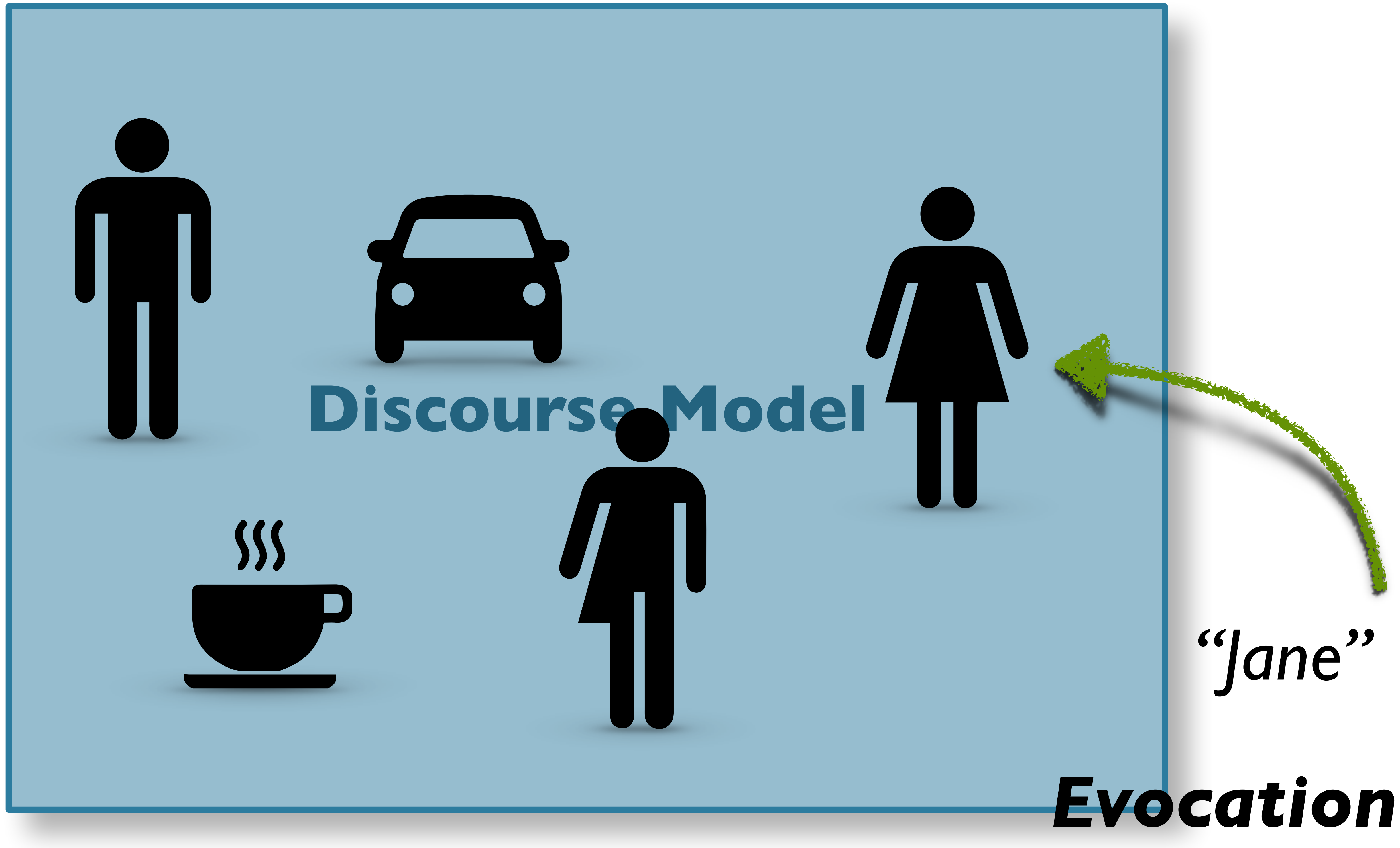
Referring Expressions

- Many forms:
  - *Queen Elizabeth*
  - *she/her*
  - *the Queen*
  - *HRM*
  - *the British Monarch*

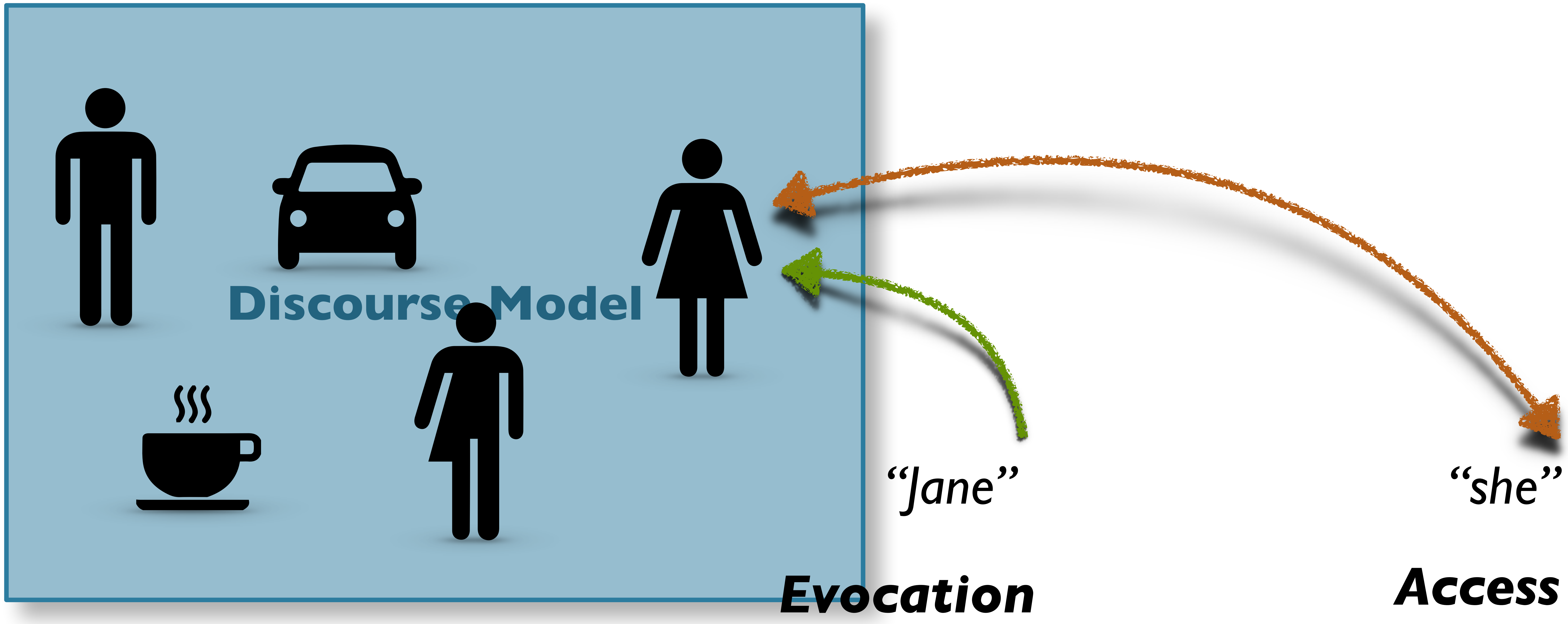
Reference and Model



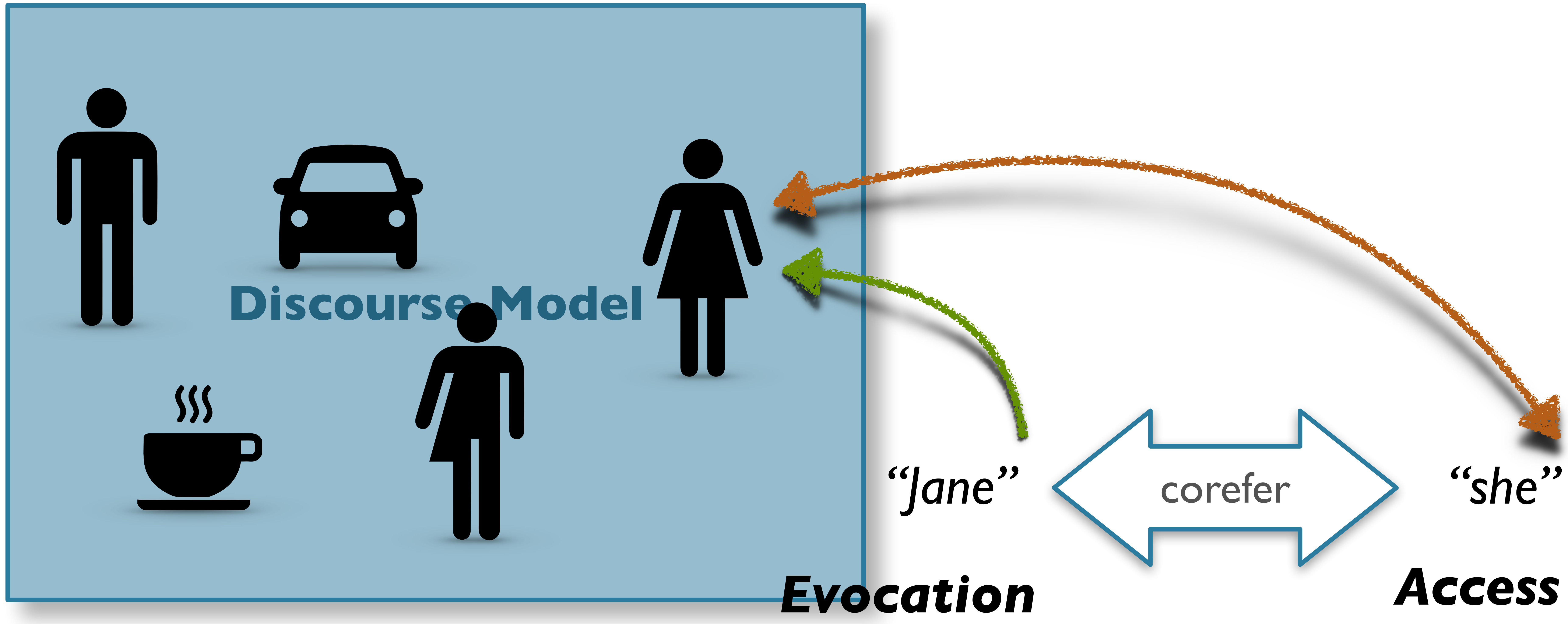
Reference and Model



Reference and Model



Reference and Model



Reference Tasks

- **Coreference resolution***:*
  - Find all expressions referring to the same entity in a text.
  - A set of coreferring expressions is a ***coreference chain***.

Reference Tasks

- **Coreference resolution***:*
  - Find all expressions referring to the same entity in a text.
  - A set of coreferring expressions is a ***coreference chain***.
- **Pronomial anaphora resolution**:
  - Find antecedent for a single pronoun.
  - Subtask of coreference resolution

Other Coreference Approaches

Data-driven Reference Resolution

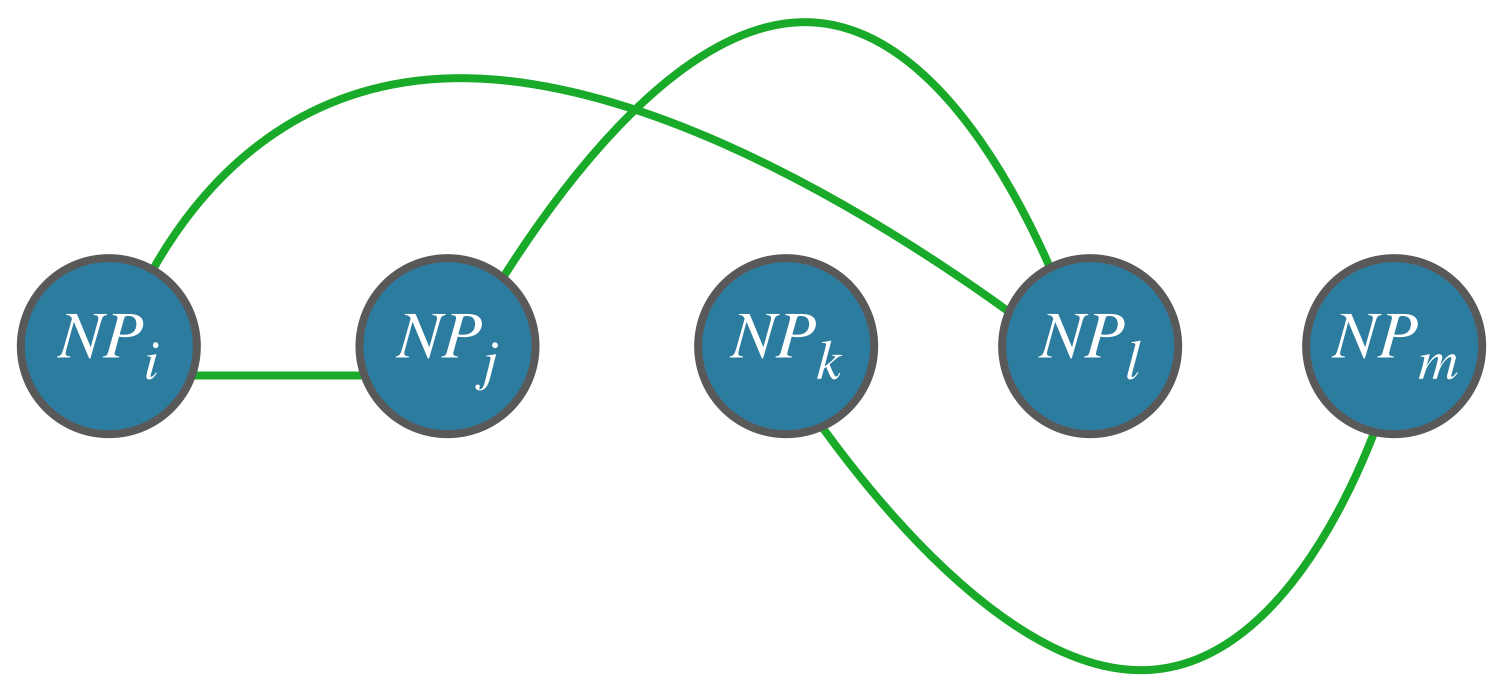
- Prior approaches:
    - Knowledge-based, hand-crafted (e.g. Hobbs' Algorithm)
- Surely, there must be ML methods to approach the problem?

Other kinds of Coreference Models

- **Mention-Pair Models**
  - Treat coreference chain as pairwise decisions (classification task)
  - For each NP_i, NP_j, do they corefer? YES/NO
  - Join together by transitivity

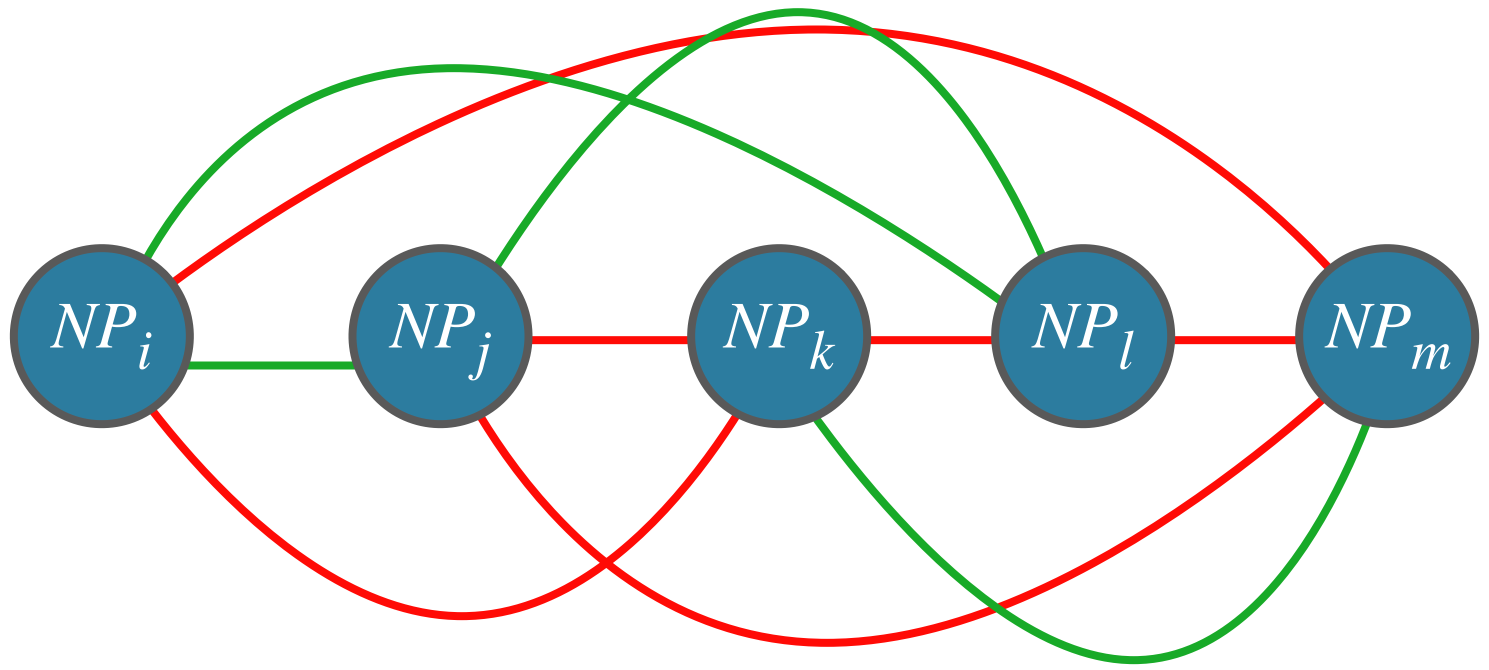
Other kinds of Coreference Models

- **Mention-Pair Models**
  - Treat coreference chain as pairwise decisions (classification task)
  - For each *NP_i*, *NP_j*, do they corefer? YES/NO
  - Join together by transitivity



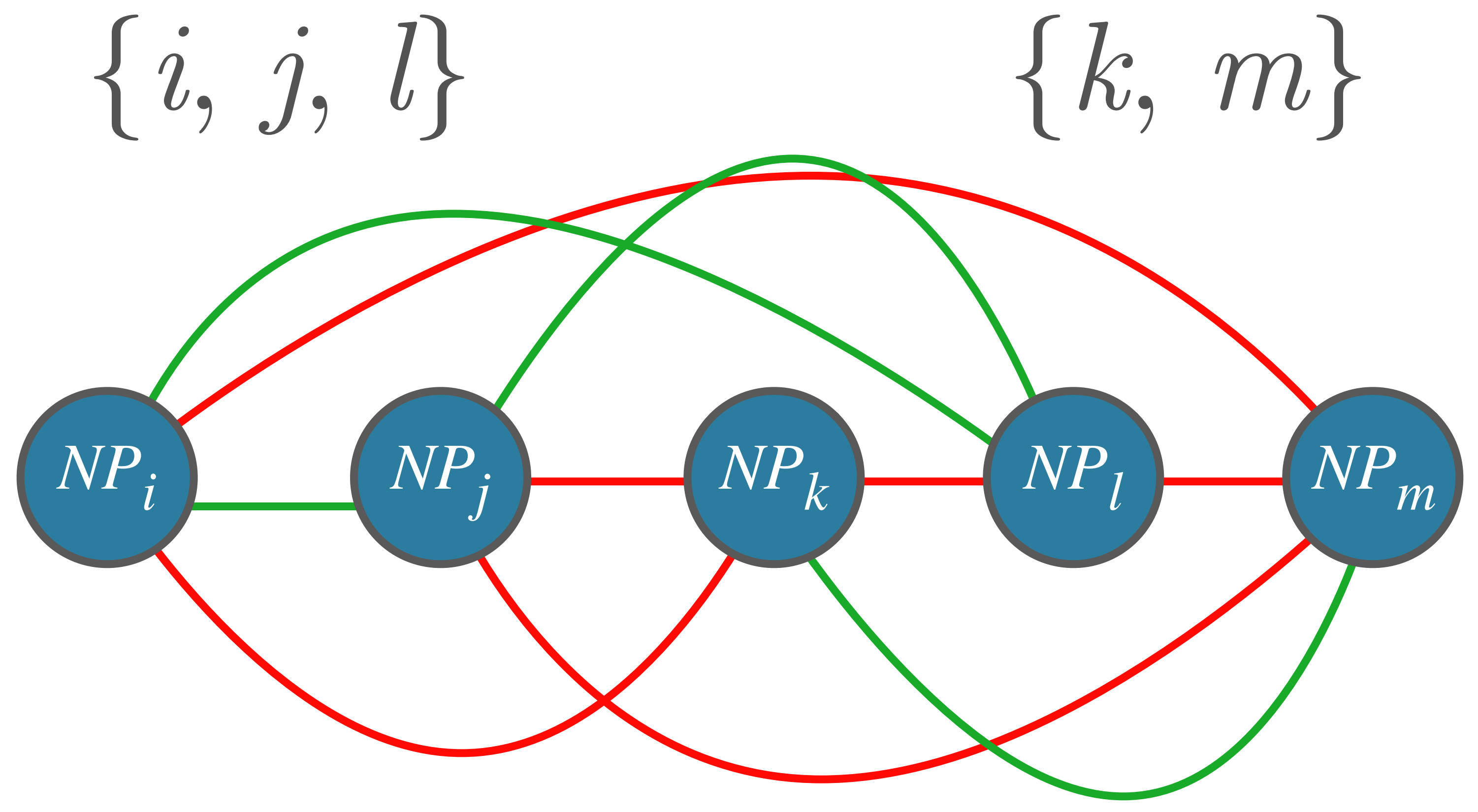
Other kinds of Coreference Models

- **Mention-Pair Models**
  - Treat coreference chain as pairwise decisions (classification task)
  - For each NP_i, NP_j, do they corefer? YES/NO
  - Join together by transitivity



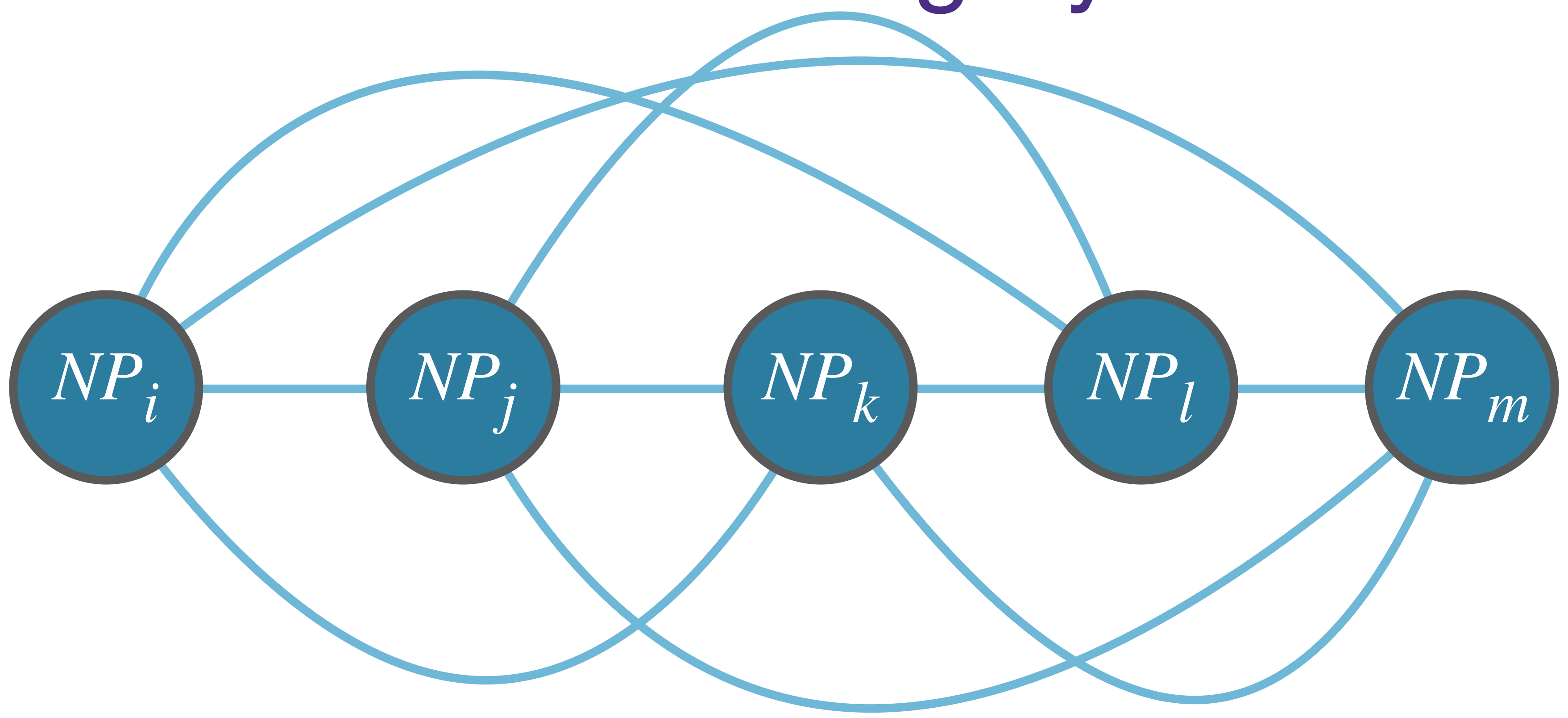
Other kinds of Coreference Models

- **Mention-Pair Models**
  - Treat coreference chain as pairwise decisions (classification task)
  - For each *NP_i*, *NP_j*, do they corefer? YES/NO
  - Join together by transitivity



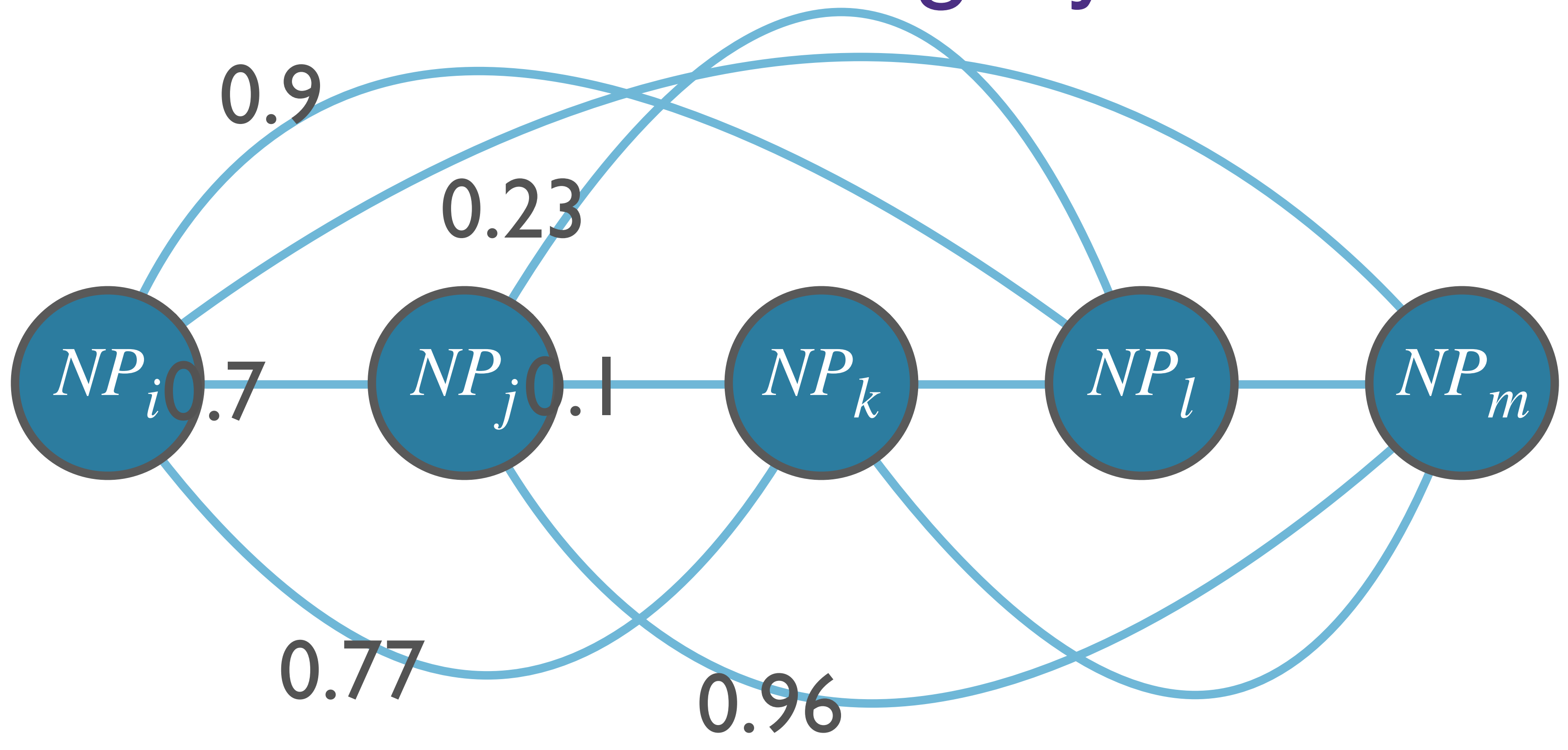
Other kinds of Coreference Models

- **Mention Ranking Models**
  - For each NP_k and all candidate antecedents, which one is the best suggestion?
  - Can be thought of as clustering method
    - Each entity a different cluster
  - Ranking problems, also well-studied category



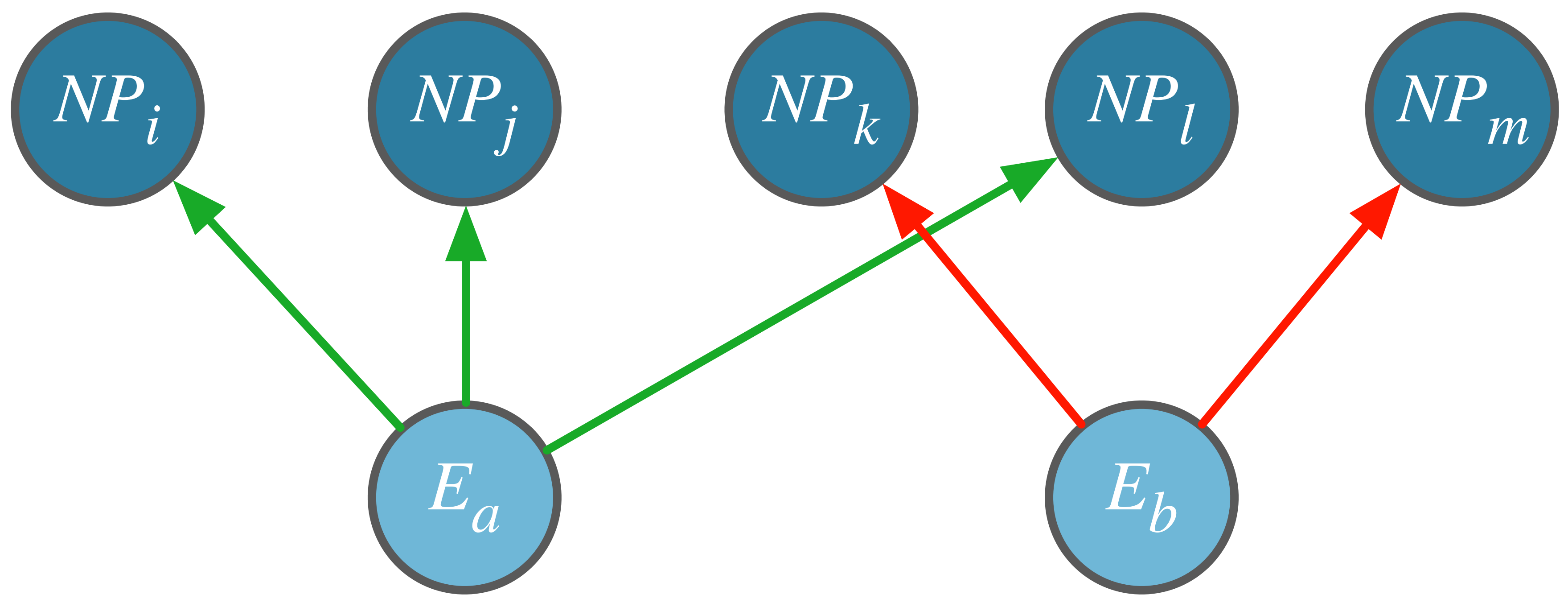
Other kinds of Coreference Models

- **Mention Ranking Models**
  - For each NP_k and all candidate antecedents, which one is the best suggestion?
  - Can be thought of as clustering method
    - Each entity a different cluster
  - Ranking problems, also well-studied category



Other kinds of Coreference Models

- **Entity-Mention Model:**
  - Posit underlying entities in discourse model
  - Each "mention" is linked to a discourse entity
  - More theoretically satisfying, but less successful work done on this approach



ML Methods for Coreference Resolution

- Annotated corpora provide ground truth with which to train supervised ML

ML Methods for Coreference Resolution

- Annotated corpora provide ground truth with which to train supervised ML
- We can take Noun Phrases (NPs) from our corpus and represent them as…

ML Methods for Coreference Resolution

- Annotated corpora provide ground truth with which to train supervised ML
- We can take Noun Phrases (NPs) from our corpus and represent them as…
  - …feature vectors! Hooray!

ML Methods for Coreference Resolution

- Annotated corpora provide ground truth with which to train supervised ML
- We can take Noun Phrases (NPs) from our corpus and represent them as…
  - …feature vectors! Hooray!
  - You know the drill, what are our features?

ML Methods for Coreference Resolution

- Annotated corpora provide ground truth with which to train supervised ML
- We can take Noun Phrases (NPs) from our corpus and represent them as…
  - …feature vectors! Hooray!
  - You know the drill, what are our features?
    - Word embeddings plus…

Typical Feature Set (Soon et. al, 2001)

- **lexical**
  - String Matching (e.g. *Mrs. Clinton* ⇔ *Clinton*)

Typical Feature Set (Soon et. al, 2001)

- **lexical**
  - String Matching (e.g. *Mrs. Clinton* ⇔ *Clinton*)
- **grammatical/syntactic**
  - i-Pronoun, j-Pronoun — Are the NPs pronouns
  - Demonstrative, Definite… — Are the NPs a demonstrative, or definite noun phrase
  - Agreement — number, gender, animacy
  - appositive (*The prime minister of Germany, Angela Merkel…*)
  - binding constraints
  - span, maximal-np, …

Typical Feature Set (Soon et. al, 2001)

- **semantic**
  - Same semantic class (e.g. Person, Organization, Location, etc)
  - Alias (e.g. *1-08-2018, Jan 8*)
- **positional**
  - distance between the NPs in terms of # of words/sentences
- **knowledge-based**
  - Naïve pronoun resolution algorithm (Hobbs)

Reference Resolution Algorithms

- Coreference Models with NNs:
  - (Clark and Manning, 2016)
    - Assign a score to each candidate antecedent
    - Each possible candidate also has possible “new referent” symbol
    - Also utilize word embeddings + avg embeddings
      - Plus ‘manual’ features as well
    - Non-RNN, essentially just local classification w/some distributional semantics

Coreference Evaluation

Coreference Annotated Corpora

- **Available Shared Task Corpora**
  - MUC-6, MUC-7 (Message Understanding Conference)
    - 60 documents each, newswire, English
  - ACE (Automatic Content Extraction)
    - English, Chinese, Arabic
    - blogs, newswire, Usenet, broadcast
- **Treebanks**
  - OntoNotes — English, Chinese (Trad/Simp), Arabic
    - Used in CoNLL 2012 shared task
  - German, Czech, Japanese, Spanish, Catalalan, Medline

Coreference Evaluation

- **Which NPs are evaluated?**
  - Gold standard tagged?
  - Automatically extracted?

Coreference Evaluation

- **Which NPs are evaluated?**
  - Gold standard tagged?
  - Automatically extracted?
- **How good are the coreference chains?**
  - Any cluster-based evaluation could be used
  - MUC scorer (Vilain et al, 1995)
    - F1 for hypothesized vs gold co-reference links
    - Problem: Link-based — ignores singletons; penalizes large clusters

How do the muppets corefer?

D.5 Pairwise Relations (ELMo and OpenAI Transformer)

Pretrained Representation	Syntactic Dep. Arc Prediction PTB	Syntactic Dep. Arc Prediction EWT	Syntactic Dep. Arc Classification PTB	Syntactic Dep. Arc Classification EWT	Semantic Dep. Arc Prediction	Semantic Dep. Arc Classification	Coreference Arc Prediction
ELMo (original), Layer 0	78.27	77.73	82.05	78.52	70.65	77.48	72.89
ELMo (original), Layer 1	89.04	86.46	96.13	93.01	87.71	93.31	71.33
ELMo (original), Layer 2	88.33	85.34	94.72	91.32	86.44	90.22	68.46
ELMo (original), Scalar Mix	89.30	86.56	95.81	91.69	87.79	93.13	73.24
ELMo (4-layer), Layer 0	78.09	77.57	82.13	77.99	69.96	77.22	73.57
ELMo (4-layer), Layer 1	88.79	86.31	96.20	93.20	87.15	93.27	72.93
ELMo (4-layer), Layer 2	87.33	84.75	95.38	91.87	85.29	90.57	71.78
ELMo (4-layer), Layer 3	86.74	84.17	95.06	91.55	84.44	90.04	70.11
ELMo (4-layer), Layer 4	87.61	85.09	94.14	90.68	85.81	89.45	68.36
ELMo (4-layer), Scalar Mix	88.98	85.94	95.82	91.77	87.39	93.25	73.88
ELMo (transformer), Layer 0	78.10	78.04	81.09	77.67	70.11	77.11	72.50
ELMo (transformer), Layer 1	88.24	85.48	93.62	89.18	85.16	90.66	72.47
ELMo (transformer), Layer 2	88.87	84.72	94.14	89.40	85.97	91.29	73.03
ELMo (transformer), Layer 3	89.01	84.62	94.07	89.17	86.83	90.35	72.62
ELMo (transformer), Layer 4	88.55	85.62	94.14	89.00	86.00	89.04	71.80
ELMo (transformer), Layer 5	88.09	83.23	92.70	88.84	85.79	89.66	71.62
ELMo (transformer), Layer 6	87.22	83.28	92.55	87.13	84.71	87.21	66.35
ELMo (transformer), Scalar Mix	90.74	86.39	96.40	91.06	89.18	94.35	75.52
OpenAI transformer, Layer 0	80.80	79.10	83.35	80.32	76.39	80.50	72.58
OpenAI transformer, Layer 1	81.91	79.99	88.22	84.51	77.70	83.88	75.23
OpenAI transformer, Layer 2	82.56	80.22	89.34	85.99	78.47	85.85	75.77
OpenAI transformer, Layer 3	82.87	81.21	90.89	87.67	78.91	87.76	75.81
OpenAI transformer, Layer 4	83.69	82.07	92.21	89.24	80.51	89.59	75.99
OpenAI transformer, Layer 5	84.53	82.77	93.12	90.34	81.95	90.25	76.05
OpenAI transformer, Layer 6	85.47	83.89	93.71	90.63	83.88	90.99	74.43
OpenAI transformer, Layer 7	86.32	84.15	93.95	90.82	85.15	91.18	74.05
OpenAI transformer, Layer 8	86.84	84.06	94.16	91.02	85.23	90.86	74.20
OpenAI transformer, Layer 9	87.00	84.47	93.95	90.77	85.95	90.85	74.57
OpenAI transformer, Layer 10	86.76	84.28	93.40	90.26	85.17	89.94	73.86
OpenAI transformer, Layer 11	85.84	83.42	92.82	89.07	83.39	88.46	72.03
OpenAI transformer, Layer 12	85.06	83.02	92.37	89.08	81.88	87.47	70.44
OpenAI transformer, Scalar Mix	87.18	85.30	94.51	91.55	86.13	91.55	76.47
GloVe (840B.300d)	74.14	73.94	77.54	72.74	68.94	71.84	72.96

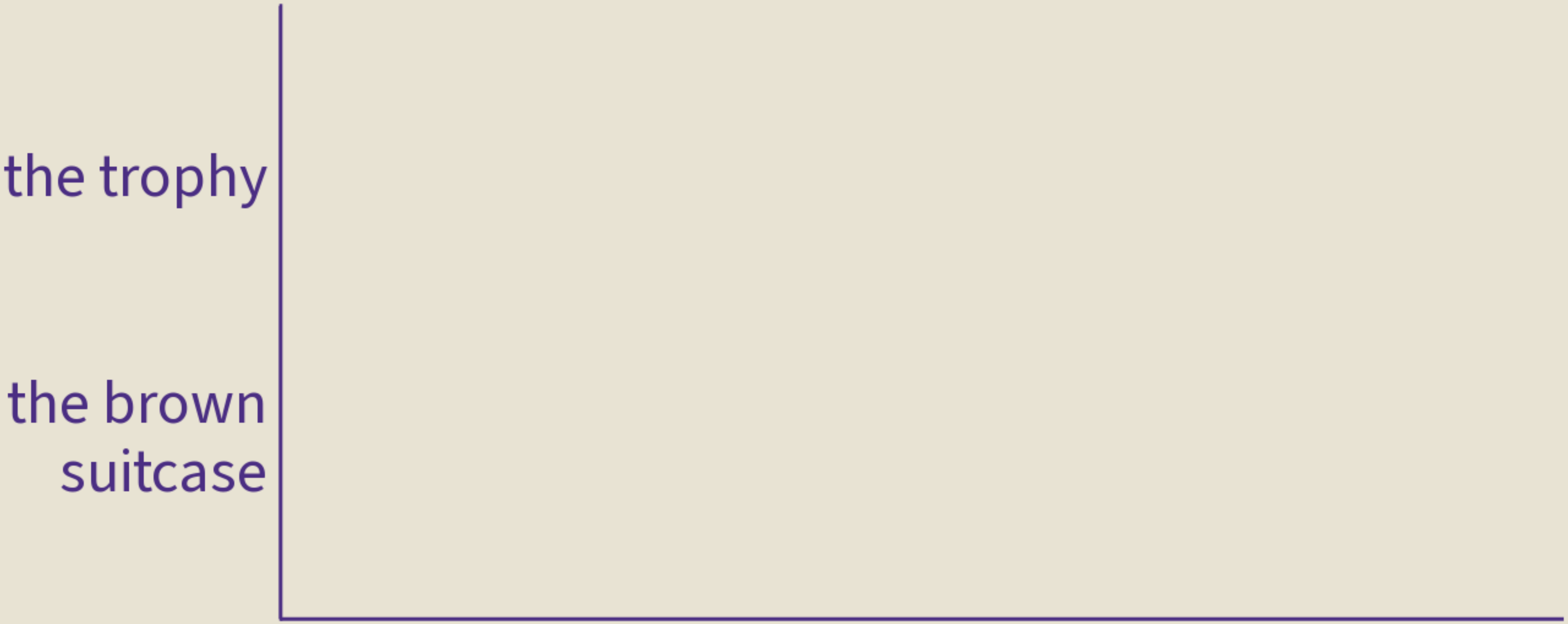
Table 9: Pairwise relation task performance of a linear probing model trained on top of the ELMo and OpenAI contextualizers, compared against a GloVe-based probing baseline.

No significant improvement over global embedding baseline
[BERT slightly better]

Liu et al 2019

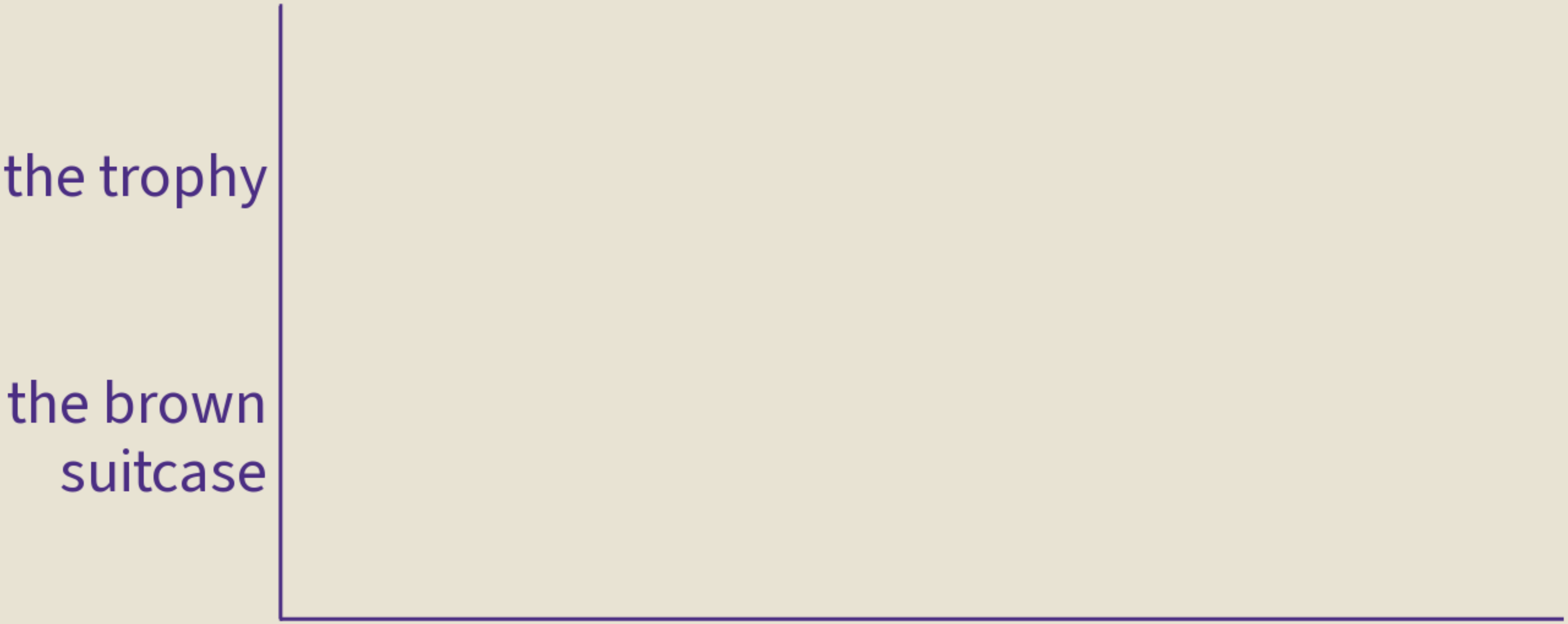
Coreference and World Knowledge

The trophy doesn't fit into the brown suitcase because it's too small. What's too small?



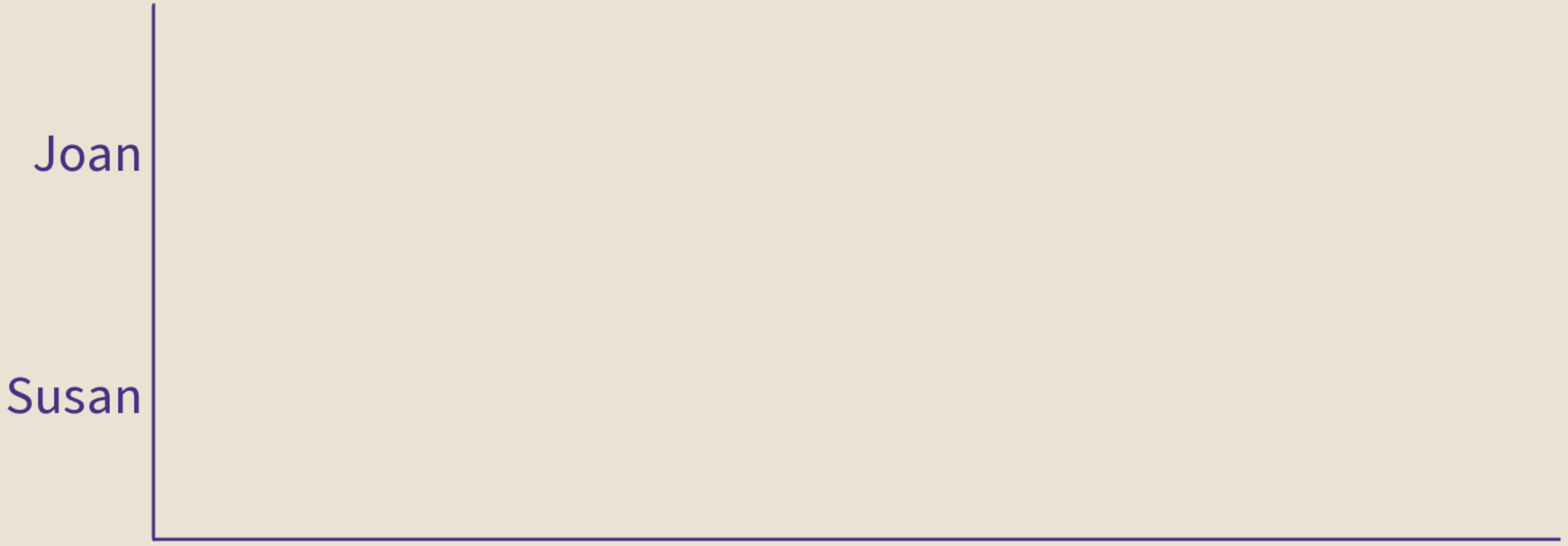
Total Results: 0

The trophy doesn't fit into the brown suitcase because it's too large. What's too large?



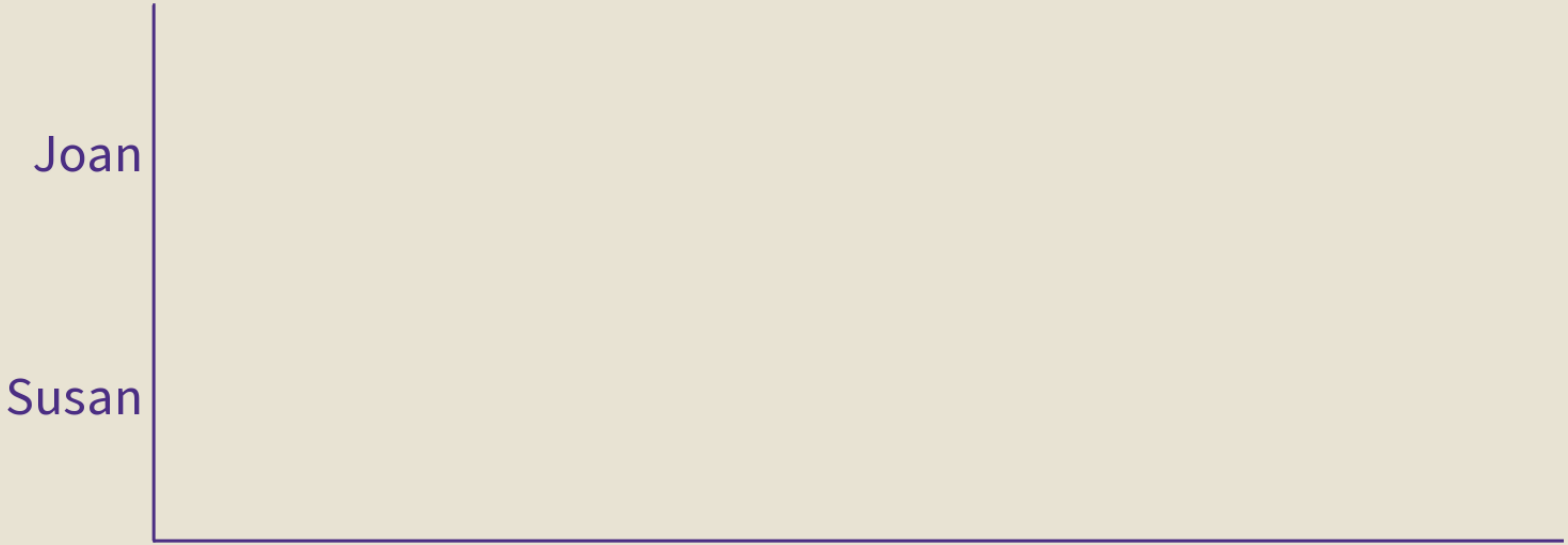
Total Results: 0

Joan made sure to thank Susan for all the help she had given. Who had given help?



Total Results: 0

Joan made sure to thank Susan for all the help she had received. Who had received help?



Total Results: 0

Coreference and World Knowledge

Coreference and World Knowledge

- The trophy doesn't fit into the brown suitcase because it's too [small/large]. What is too [small/large]?

Coreference and World Knowledge

- The trophy doesn't fit into the brown suitcase because it's too [small/large]. What is too [small/large]?
  - Answers:The suitcase/the trophy.

Coreference and World Knowledge

- The trophy doesn't fit into the brown suitcase because it's too [small/large]. What is too [small/large]?
  - Answers:The suitcase/the trophy.
- Joan made sure to thank Susan for all the help she had [given/received]. Who had [given/received] help?

Coreference and World Knowledge

- The trophy doesn't fit into the brown suitcase because it's too [small/large]. What is too [small/large]?
  - Answers:The suitcase/the trophy.
- Joan made sure to thank Susan for all the help she had [given/received]. Who had [given/received] help?
  - Answers: Susan/Joan.

Coreference and World Knowledge

- The trophy doesn't fit into the brown suitcase because it's too [small/large]. What is too [small/large]?
  - Answers:The suitcase/the trophy.
- Joan made sure to thank Susan for all the help she had [given/received]. Who had [given/received] help?
  - Answers: Susan/Joan.
- Paul tried to call George on the phone, but he wasn't [successful/available]. Who was not [successful/available]?

Coreference and World Knowledge

- The trophy doesn't fit into the brown suitcase because it's too [small/large]. What is too [small/large]?
  - Answers:The suitcase/the trophy.
- Joan made sure to thank Susan for all the help she had [given/received]. Who had [given/received] help?
  - Answers: Susan/Joan.
- Paul tried to call George on the phone, but he wasn't [successful/available]. Who was not [successful/available]?
  - Answers: Paul/George.

Coreference and World Knowledge

- The trophy doesn't fit into the brown suitcase because it's too [small/large]. What is too [small/large]?
  - Answers:The suitcase/the trophy.
- Joan made sure to thank Susan for all the help she had [given/received]. Who had [given/received] help?
  - Answers: Susan/Joan.
- Paul tried to call George on the phone, but he wasn't [successful/available]. Who was not [successful/available]?
  - Answers: Paul/George.
- The lawyer asked the witness a question, but he was reluctant to [answer/repeat] it . Who was reluctant to [answer/repeat] the question?

Coreference and World Knowledge

- The trophy doesn't fit into the brown suitcase because it's too [small/large]. What is too [small/large]?
  - Answers:The suitcase/the trophy.
- Joan made sure to thank Susan for all the help she had [given/received]. Who had [given/received] help?
  - Answers: Susan/Joan.
- Paul tried to call George on the phone, but he wasn't [successful/available]. Who was not [successful/available]?
  - Answers: Paul/George.
- The lawyer asked the witness a question, but he was reluctant to [answer/repeat] it . Who was reluctant to [answer/repeat] the question?
  - Answers: The witness/the lawyer.

Winograd Schema Challenge

- Still hard!
  - [WSC]
  - [Winogrande]

SuperGLUE GLUE Paper </> Code Tasks Leaderboard FAQ Diagnostics Submit

Leaderboard Version: **2.0**

Rank	Name	Model	URL	Score	BoolQ	CB	COPA	MultiRC	ReCoRD	RTE	WiC	WSC	AX-g	AX-b
1	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8	89.0	95.8/98.9	100.0	81.8/51.9	91.7/91.3	93.6	80.0	100.0	99.3/99.7	76.6
2	T5 Team - Google	T5		88.9	91.0	93.0/96.4	94.8	88.2/62.3	93.3/92.5	92.5	76.1	93.8	92.7/91.9	65.6
3	Facebook AI	RoBERTa		84.6	87.1	90.5/95.2	90.6	84.4/52.5	90.6/90.0	88.2	69.9	89.0	91.0/78.1	57.9
4	IBM Research AI	BERT-mtl		73.5	84.8	89.6/94.0	73.8	73.2/30.5	74.6/74.0	84.1	66.2	61.0	97.8/57.3	29.6
5	SuperGLUE Baselines	BERT++		71.5	79.0	84.8/90.4	73.8	70.0/24.1	72.0/71.3	79.0	69.6	64.4	99.4/51.4	38.0
		BERT		69.0	77.4	75.7/83.6	70.6	70.0/24.1	72.0/71.3	71.7	69.6	64.4	97.8/51.7	23.0
		Most Frequent Class		47.1	62.3	21.7/48.4	50.0	61.1/0.3	33.4/32.5	50.3	50.0	65.1	100.0/50.0	0.0
		CBoW		44.5	62.2	49.0/71.2	51.6	0.0/0.5	14.0/13.6	49.7	53.1	65.1	100.0/50.0	-0.4
		Outside Best		-	80.4	-	84.4	70.4/24.5	74.8/73.0	82.7	-	-	-	-
-	Stanford Hazy Research	Snorkel [SuperGLUE v1.9]		-	-	88.6/93.2	76.2	76.4/36.3	-	78.9	72.1	72.6	-	47.6

Winograd Schema Challenge

- Still hard!
  - [WSC]
  - [Winogrande]

Heavily supervised
(benchmark "saturated" now)

SuperGLUE GLUE Paper </> Code Tasks Leaderboard FAQ Diagnostics Submit

Leaderboard Version: **2.0**

Rank	Name	Model	URL	Score	BoolQ	CB	COPA	MultiRC	ReCoRD	RTE	WiC	WSC	AX-g	AX-b
1	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8	89.0	95.8/98.9	100.0	81.8/51.9	91.7/91.3	93.6	80.0	100.0	99.3/99.7	76.6
2	T5 Team - Google	T5		88.9	91.0	93.0/96.4	94.8	88.2/62.3	93.3/92.5	92.5	76.1	93.8	92.7/91.9	65.6
3	Facebook AI	RoBERTa		84.6	87.1	90.5/95.2	90.6	84.4/52.5	90.6/90.0	88.2	69.9	89.0	91.0/78.1	57.9
4	IBM Research AI	BERT-mtl		73.5	84.8	89.6/94.0	73.8	73.2/30.5	74.6/74.0	84.1	66.2	61.0	97.8/57.3	29.6
5	SuperGLUE Baselines	BERT++		71.5	79.0	84.8/90.4	73.8	70.0/24.1	72.0/71.3	79.0	69.6	64.4	99.4/51.4	38.0
		BERT		69.0	77.4	75.7/83.6	70.6	70.0/24.1	72.0/71.3	71.7	69.6	64.4	97.8/51.7	23.0
		Most Frequent Class		47.1	62.3	21.7/48.4	50.0	61.1/0.3	33.4/32.5	50.3	50.0	65.1	100.0/50.0	0.0
		CBoW		44.5	62.2	49.0/71.2	51.6	0.0/0.5	14.0/13.6	49.7	53.1	65.1	100.0/50.0	-0.4
		Outside Best		-	80.4	-	84.4	70.4/24.5	74.8/73.0	82.7	-	-	-	-
-	Stanford Hazy Research	Snorkel [SuperGLUE v1.9]		-	-	88.6/93.2	76.2	76.4/36.3	-	78.9	72.1	72.6	-	47.6

Questions

- Decent results on (clean) text.   What about…

Questions

- Decent results on (clean) text.   What about…
- **Conversational speech**?
  - Fragments, disfluencies, etc…

Questions

- Decent results on (clean) text. What about…
- **Conversational speech**?
  - Fragments, disfluencies, etc…
- **Dialogue**?
  - Multiple speakers introduce referents

Questions

- Decent results on (clean) text.   What about…
- **Conversational speech**?
  - Fragments, disfluencies, etc…
- **Dialogue**?
  - Multiple speakers introduce referents
- **Multimodal communication**?
  - How can entities be evoked in other ways?
  - Are all equally salient?

Questions

- **Other languages?**
  - Are salience hierarchies the same?
  - Syntactic constraints?
    - Reflexives in Chinese, Korean…?
- **Zero anaphora?**
  - How do you resolve a pronoun if you can't find it?
  - e.g. "*There are two roads to eternity, a straight and narrow, and a broad and crooked.*"
    - Each indefinite here implies a gap [road], that would be anaphoric, but leaves a gap

Conclusions

- Coreference establishes ***coherence***
- Reference resolution depends on coherence
- Variety of approaches:
  - Syntactic constraints, recency, frequency, role
- Similar effectiveness - different requirements
- Coreference can enable summarization within and across documents (and potentially languages!), question answering, information retrieval, …

Discourse Structure

Why Model Discourse Structure?

Theoretical Concerns

- Discourse: not just constituent utterances
- Creation of joint meaning
- Context guides interpretation of constituents

Why Model Discourse Structure?

Theoretical Concerns

- Understanding how discourse is structured:
  - What are the units of discourse?
  - How do they combine to establish meaning?
  - How can we derive structure from surface forms?
  - What makes discourse coherent vs. incoherent?
  - How do the units of discourse influence reference resolution?

Why Model Discourse Structure?

Applied Concerns

- Design better summarization, understanding systems
- Improve speech synthesis (discourse-contextual intonation, emphasis)
- Develop approach for generation of discourse
- Design dialogue agents for task interaction
- Guide reference resolution

Discourse (Topic) Segmentation

- BBC Global News Podcast 11/26/2018:
- “I’m Valerie Saunderson, and in the early hours of Monday, the 26th of November, these are our main stories. || After forty-five years, both parties call it a day as Britain’s Brexit agreement is signed off by EU leaders. So, what happens next? We hear from our correspondents in Brussels and London. || There’s been a sharp escalation in a Naval dispute near Crimea, with Ukraine accusing Russian special forces of seizing three of its vessels || An investigation discovers many medical implants haven’t been properly tested before they’re put in patients. || Also in this podcast, NASA prepares for “seven minutes of terror,” the latest landing on the Red planet [Voice #2:] Although we’ve done it before, landing on Mars is hard, and this mission is no different. || [Voice #1:] A year and a half after the start of Brexit Negotiations…”

Discourse (Topic) Segmentation

- BBC Global News Podcast 11/26/2018:
- “I’m Valerie Saunderson, and in the early hours of Monday, the 26th of November, these are our main stories. || After forty-five years, both parties call it a day as Britain’s Brexit agreement is signed off by EU leaders. So, what happens next? We hear from our correspondents in Brussels and London. || There’s been a sharp escalation in a Naval dispute near Crimea, with Ukraine accusing Russian special forces of seizing three of its vessels || An investigation discovers many medical implants haven’t been properly tested before they’re put in patients. || Also in this podcast, NASA prepares for “seven minutes of terror,” the latest landing on the Red planet [Voice #2:] Although we’ve done it before, landing on Mars is hard, and this mission is no different. || [Voice #1:] A year and a half after the start of Brexit Negotiations…”

Discourse Segmentation

- Basic form of discourse structure
  - Divide document into linear sequence of subtopics
- Many genres have conventional structures
  - **Academic**: Intro, Hypothesis, Previous Work, Methods, Results, Conclusion
  - **Newspapers**: Headline, Byline, Lede, Elaboration
  - **Patient Reports**: Subjective, Objective, Assessment, Plan
- Can guide summarization, retrieval

Cohesion

- Use of linguistic devices to link text units
  - Lexical cohesion: Link with relations between words
    - Synonymy, Hypernymy
    - *Peel, core, and slice the **pears** and **apples**. Add **the fruit** to the skillet.*
  - Nonlexical Cohesion
    - e.g. anaphora
    - *Peel, core, and slice the **pears** and **apples**. Add **them** to the skillet.*

Cohesion

- Use of linguistic devices to link text units
  - Lexical cohesion: Link with relations between words
    - Synonymy, Hypernymy
    - *Peel, core, and slice the **pears** and **apples**. Add **the fruit** to the skillet.*
  - Nonlexical Cohesion
    - e.g. anaphora
    - *Peel, core, and slice the **pears** and **apples**. Add **them** to the skillet.*
- Cohesion chain establish link through sequence of words

Cohesion

- Use of linguistic devices to link text units
  - Lexical cohesion: Link with relations between words
    - Synonymy, Hypernymy
    - *Peel, core, and slice the **pears** and **apples**. Add **the fruit** to the skillet.*
  - Nonlexical Cohesion
    - e.g. anaphora
    - *Peel, core, and slice the **pears** and **apples**. Add **them** to the skillet.*
- Cohesion chain establish link through sequence of words
- Segment boundary = dip in cohesion.

TextTiling ([Hearst, 1997](#))

- Lexical, cohesion-based segmentation
  - Boundaries at dips in cohesion scores
  - Tokenization, Lexical cohesion score, Boundary ID
- Tokenization
  - Units?
    - Whitespace delimited words
    - Stopped
    - Stemmed
    - 20 words = 1 pseudo-sentence

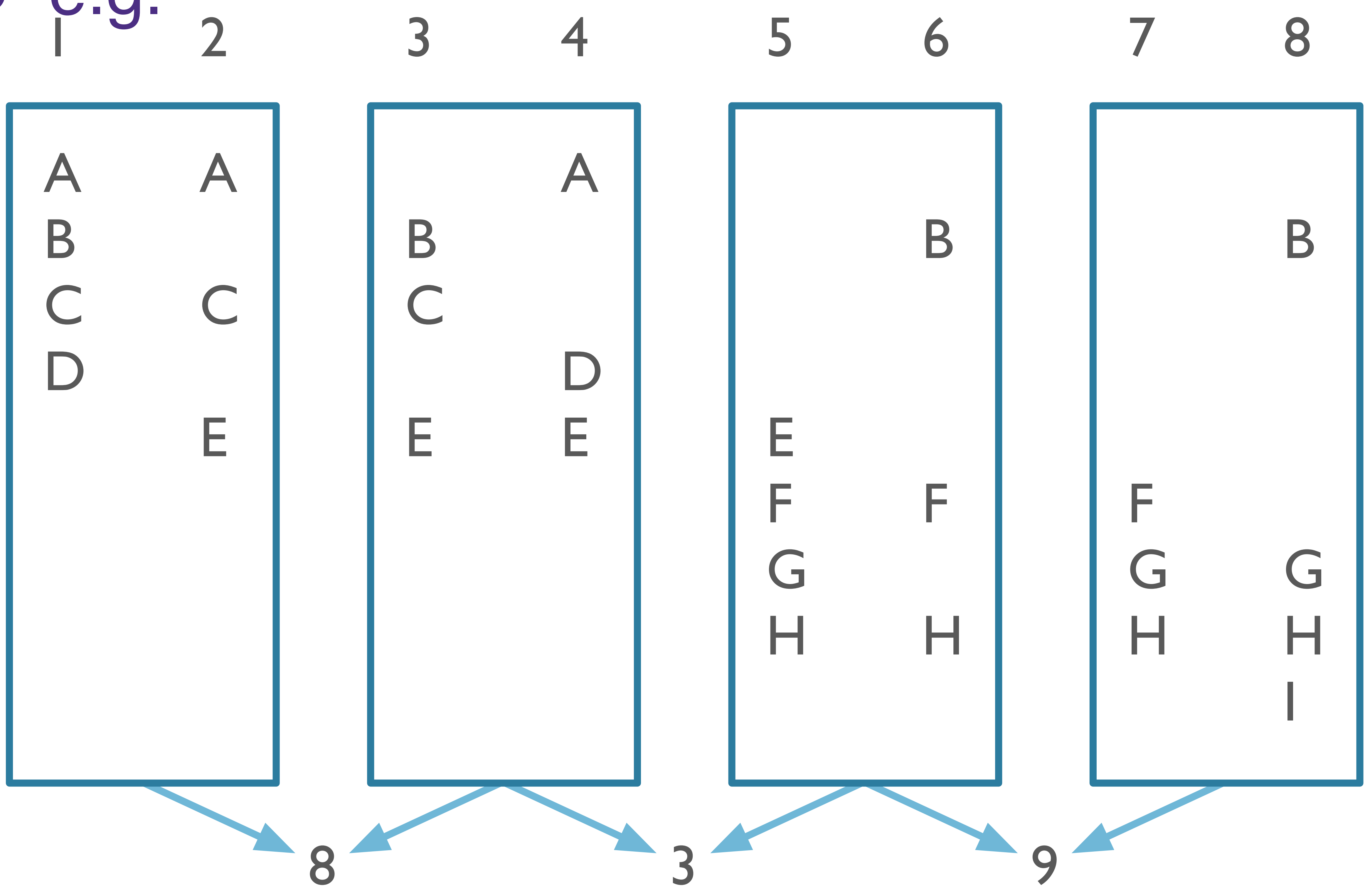
Lexical Cohesion Score

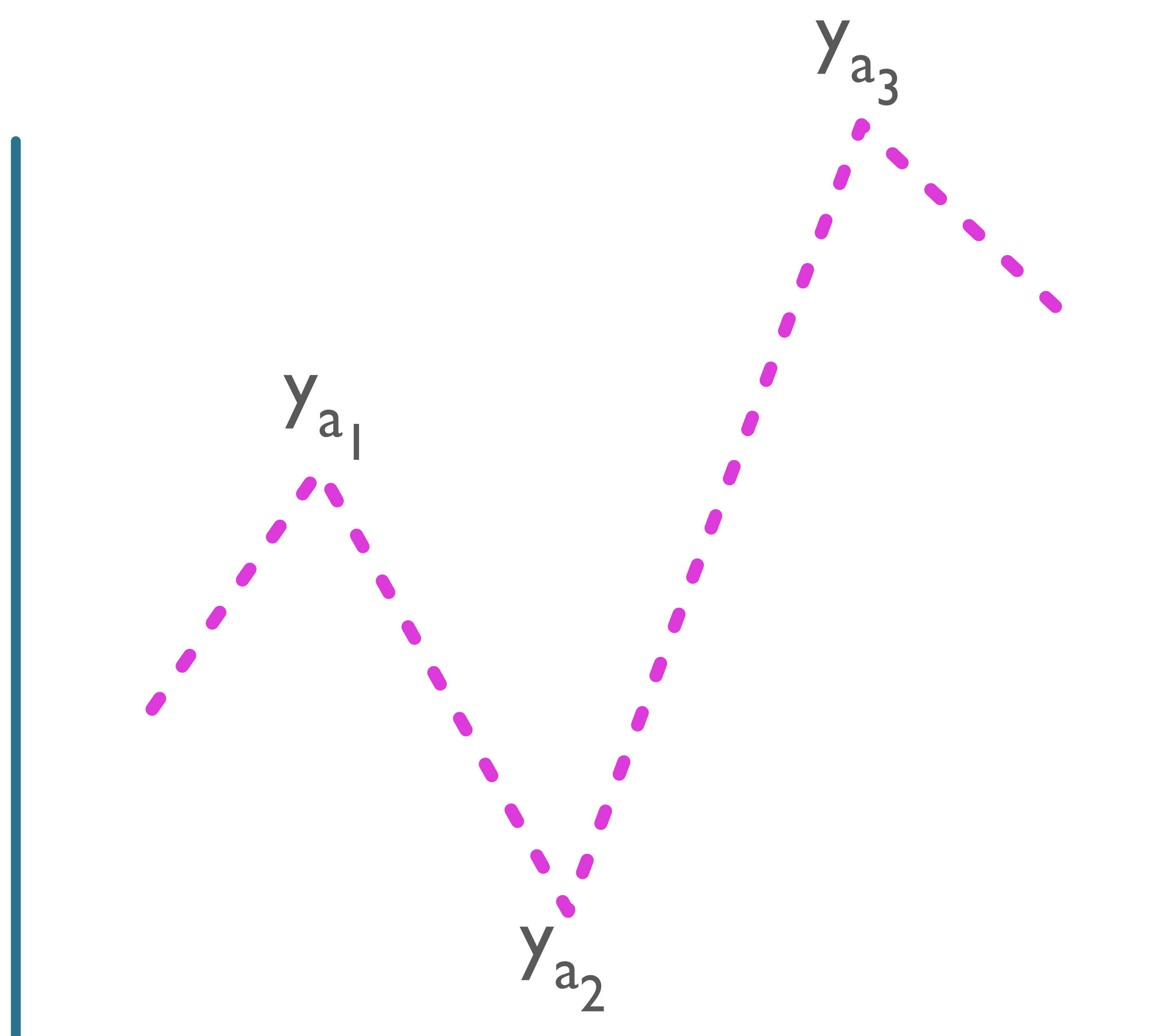
- Similarity between spans of text
  - b = ‘Block’ of 10 pseudo-sentences before gap
  - a = ‘Block’ of 10 pseudo-sentences after gap
  - How do we compute similarity?
    - Vectors and cosine similarity (again!)

$$sim_{cosine}(\vec{b}, \vec{a}) = \frac{\vec{b} \cdot \vec{a}}{|\vec{b}||\vec{a}|} = \frac{\sum_{i=1}^{N} b_i \times a_i}{\sqrt{\sum_{i=1}^{N} b_i^2}\sqrt{\sum_{i=1}^{N} a_i^2}}$$

Segmentation

- Depth Score:
  - Difference between position and adjacent peaks $\left(y_{a_1} - y_{a_2}\right) + \left(y_{a_3} - y_{a_2}\right)$
  - e.g.





Embedding-Based Cohesion

- Aggregation:
  - Sentence similarity
    - Sentence vector: sum of word embedding vectors
  - Pairwise sentence cohesion: $\cos\left(\sum_{w \in s} w, \sum_{w \in t} w\right)$
  - Document cohesion: average pairwise cohesion

$$coherence(T) = \frac{1}{n-1} \sum_{i=1}^{n-1} \cos(s_i, s_{i+1})$$

- Baseline (Xu et al, 2019)
  - Train RNN LM
    - Compute log likelihood of s_i with and without preceding context

Local Coherence Discriminator

- LCD (Xu et al, 2019)
  - Coherence of text = average coherence b/t adj pairs
  - **Supervised** model
    - Trained to distinguish b/t:
      - Adjacent pairs of sentences in training data (pos examples)
      - Randomly associated sentence pairs (assumed negative)
- Approach:
  - Compute sentence embeddings for s, t
  - Concatenate: each vector, diff (s-t); abs diff ls-tl; elementwise product
  - Train FFN s.t. positive examples score higher than neg

LCD

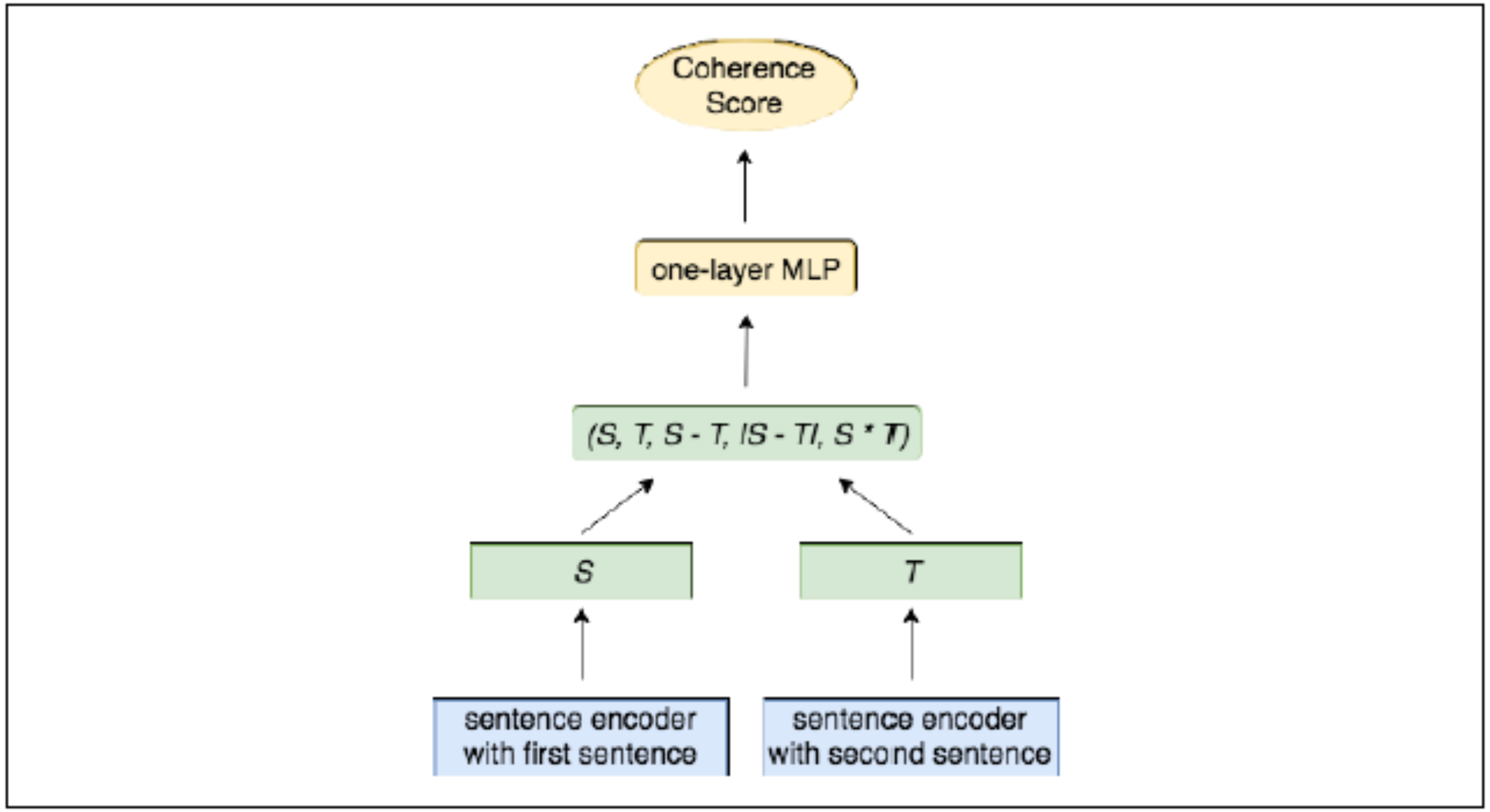


Figure 23.11 The architecture of the LCD model of document coherence, showing the computation of the score for a pair of sentences s and t. Figure from Xu et al. (2019).

Coherence Relations & Discourse Structure

Coherence Relations

John hid Bill's car keys. He was drunk.

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

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

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

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.*

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.*

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.*

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.

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.

- This discourse ***isn't linear***

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.

- 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

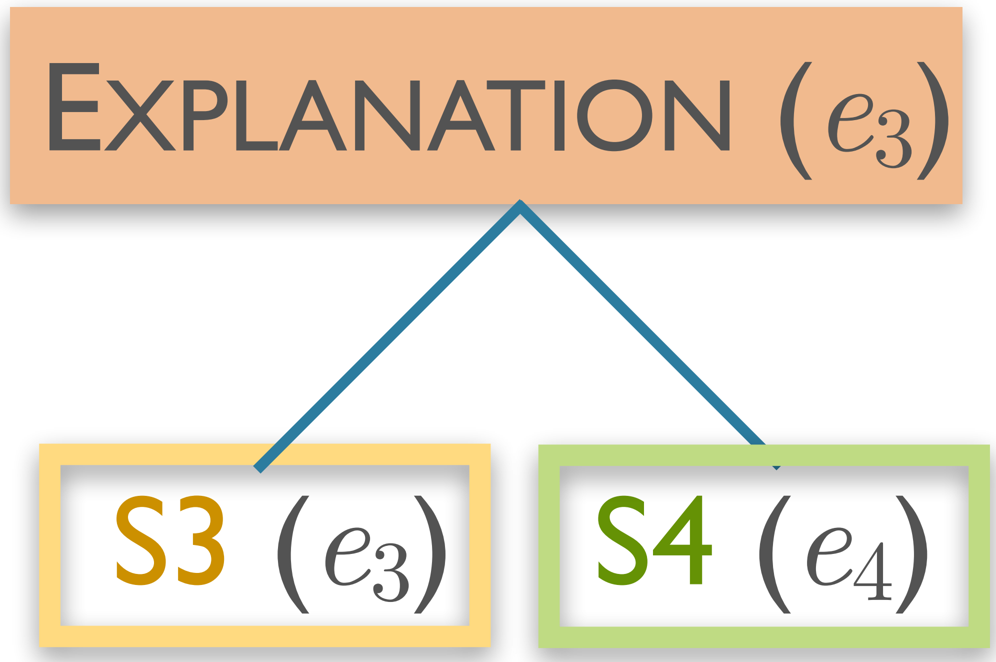
S1 – Armin went to the bank to deposit his paycheck

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.



Adapted from J&M 2nd ed p. 690

Coherence Relation Hierarchy

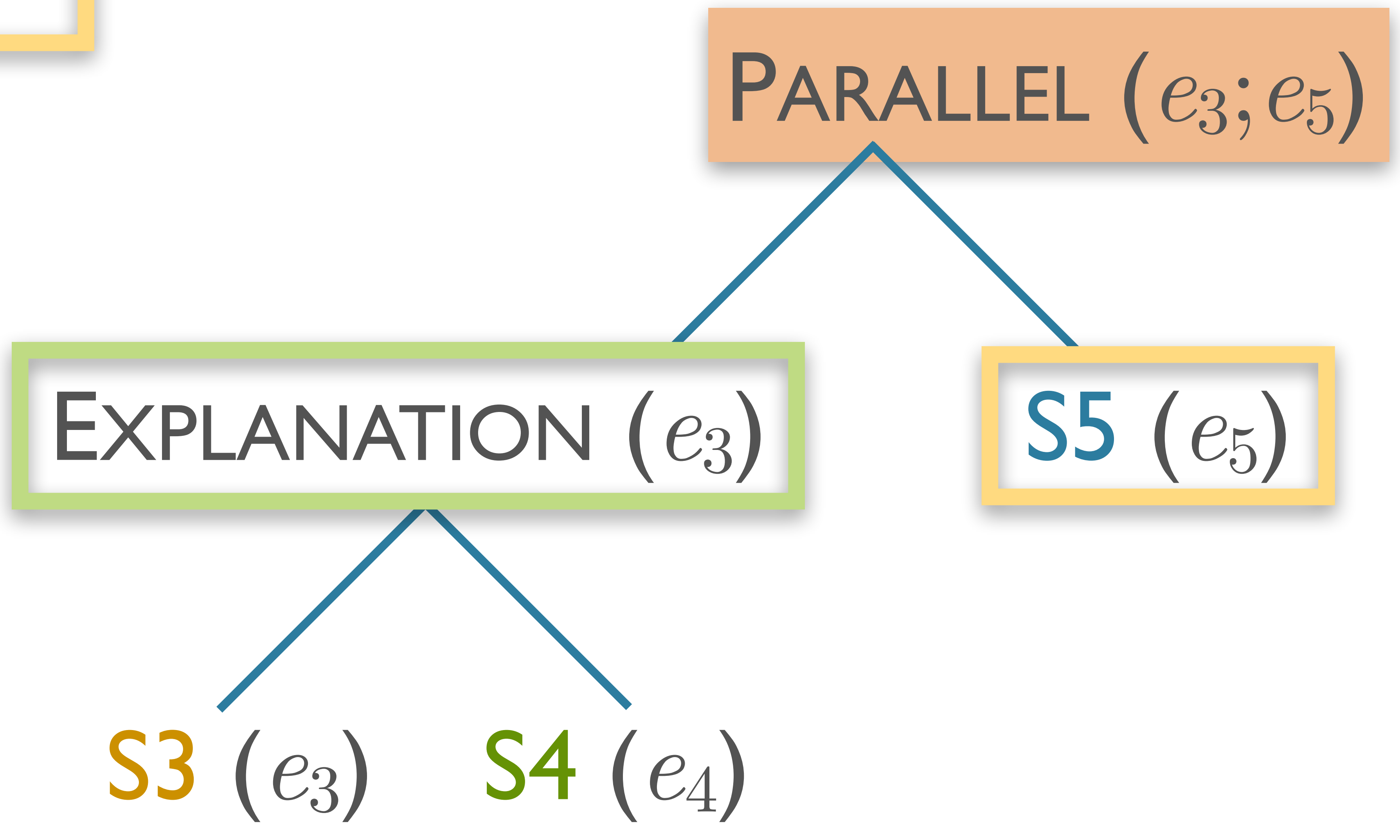
S1 – Armin went to the bank to deposit his paycheck

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

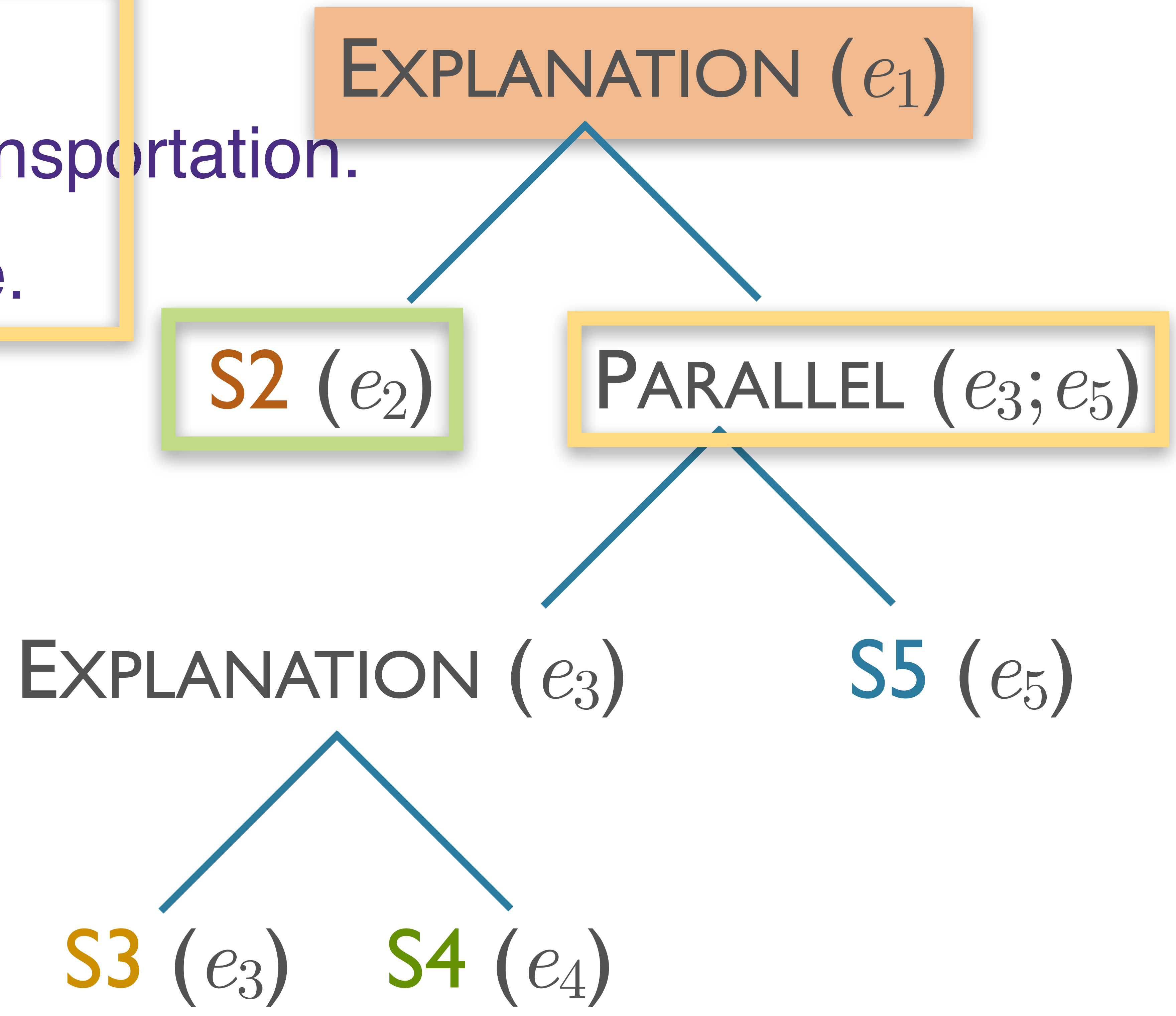
S1 – Armin went to the bank to deposit his paycheck

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.



Adapted from J&M 2nd ed p. 690

Coherence Relation Hierarchy

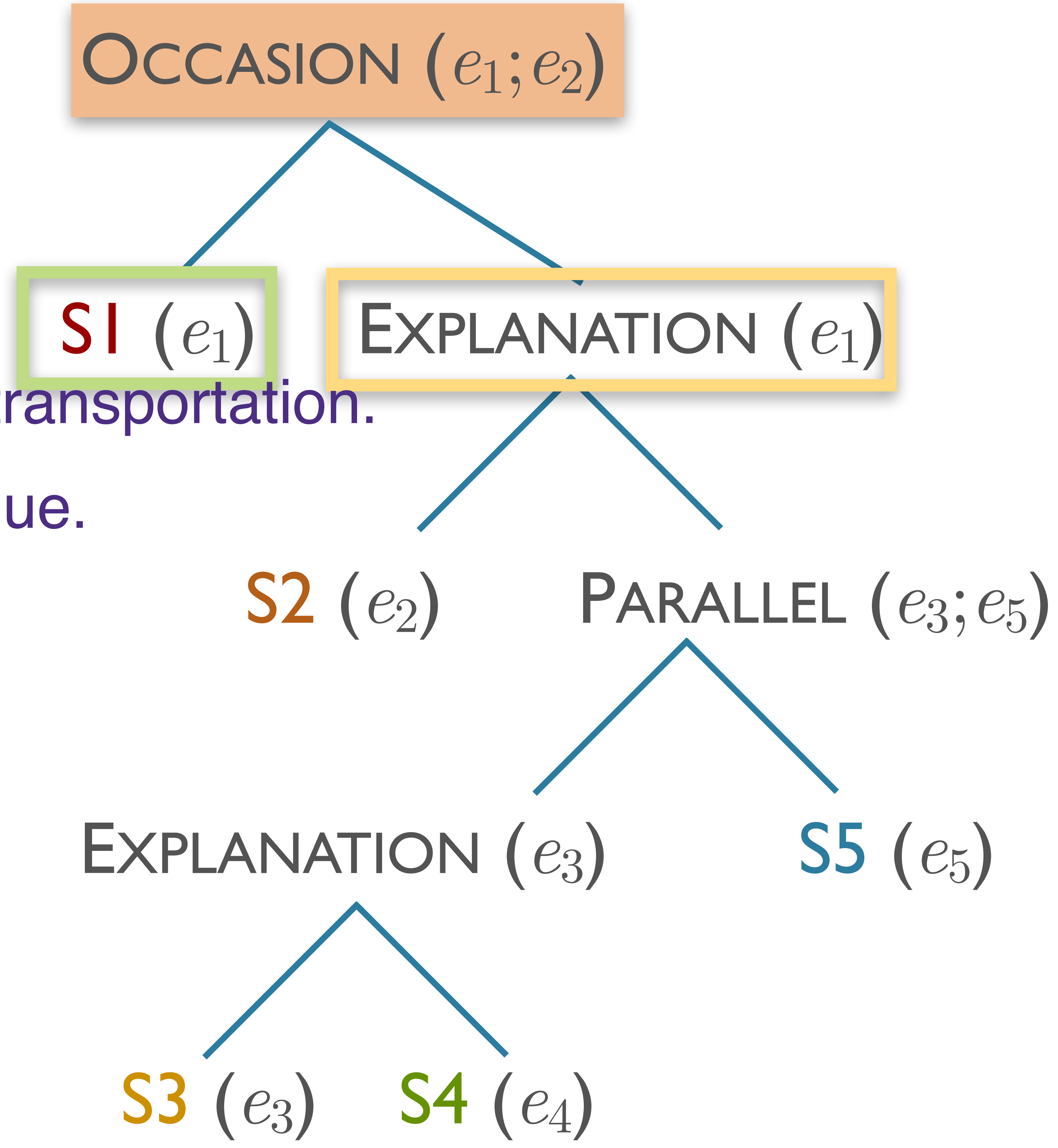
S1 – Armin went to the bank to deposit his paycheck

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

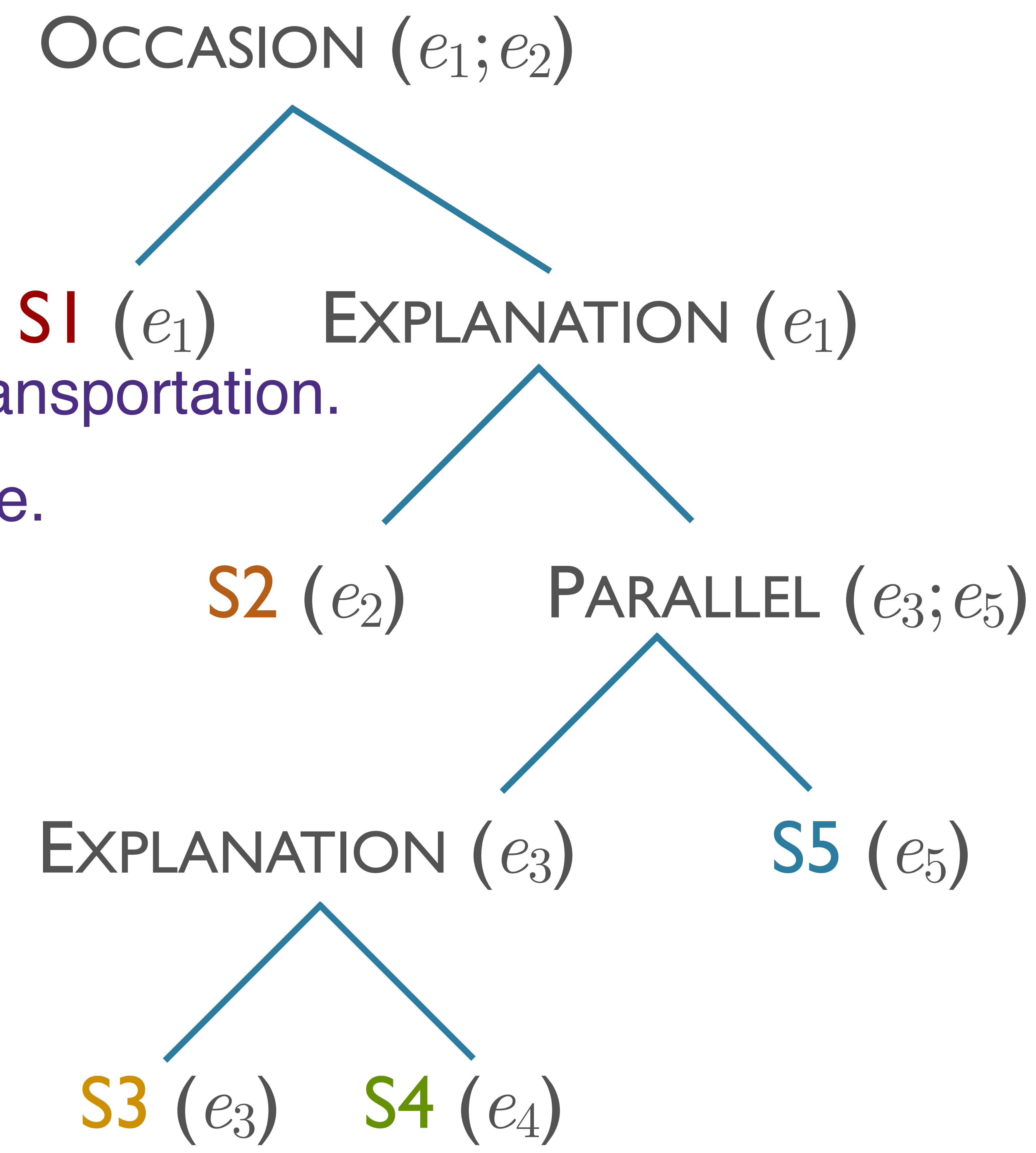
S1 – Armin went to the bank to deposit his paycheck

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.



Adapted from J&M 2nd ed p. 690

Coherence Relations:

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

- “Theory-neutral” discourse model

Coherence Relations:

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

- "Theory-neutral" discourse model
- No stipulation of overall structure, local sequence relations

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.*

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 **S_1** to **S_2** 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

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*’) between spans
  - Arg_2 syntactically bound to connective unit, Arg_1

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*’) between 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
  - Annotators provide implicit discourse unit, label

PDTB

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.

Figure 23.2 The four high-level semantic distinctions in the PDTB sense hierarchy

- PDTB corpus: 18K explicit relations; 16K implicit
- Also Chinese Discourse Treebank,
- ~ half as many explicit discourse connectives

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 test)
    - Explicit discourse connectives: **91**% (**76**% blind test)
    - Non-explicit discourse connectives: **34**% (**36**% blind test)

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

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

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

Relation Classification

- Basic task:
  - Given pair of adjacent sentences, give coherence relation sense label
- Approaches:
  - Employ BoW or sentence embeddings of sentence pairs
  - Pass through some classifier
- Strong approach: (Nie et al, 2019)
  - Represent spans with BERT contextual embeddings
  - Take last layer hidden state for position of <CLS> token
  - Run through 1-layer FFN + softmax for classification
- Other steps use sequence models, heuristics

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.*

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

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?

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

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 U_1

$C_f(U_1)$: {*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 U_2

$C_f(U_1)$: {*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 U_2

$C_f(U_2)$: {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
  - Shallow discourse parsing
  - Also see: Rhetorical Structure Theory (RST)