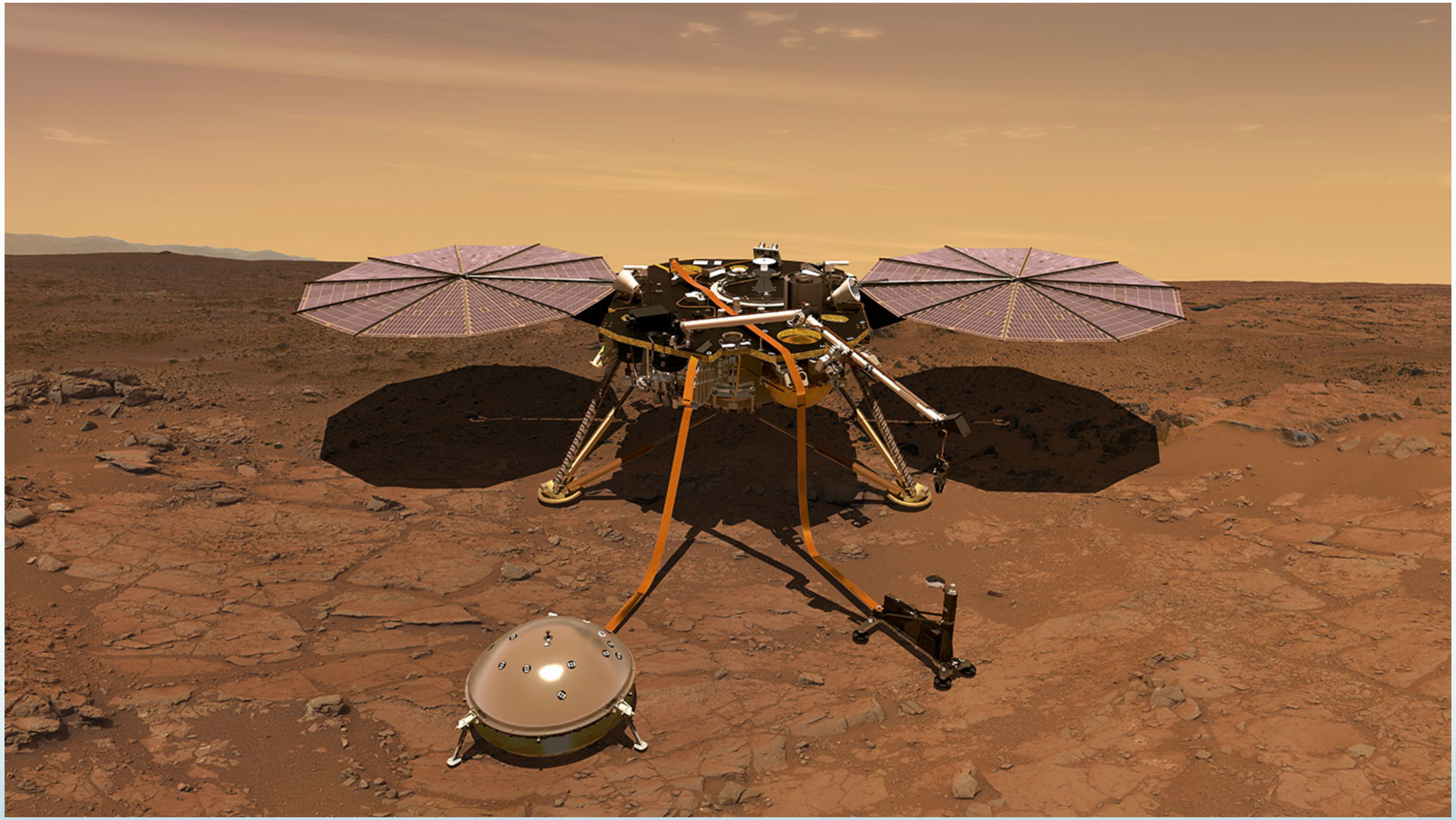
Discourse & Coreference Resolution

LING 571 — Deep Processing Methods in NLP
November 26th 2018
Ryan Georgi

UNIVERSITY OF WASHINGTON

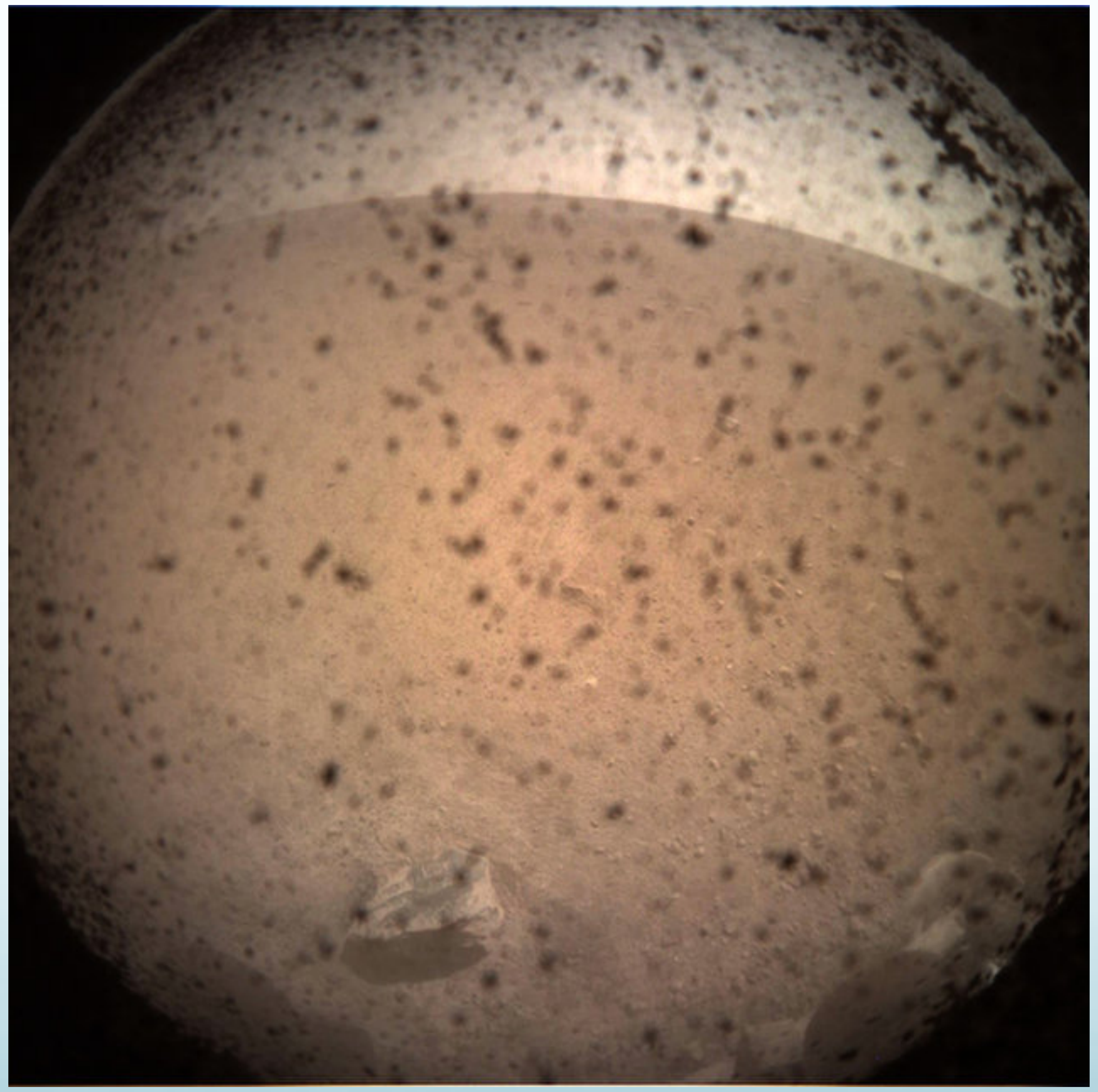
PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

Happy InSight Day!



InSight Coreference

- PASADENA, Calif. — NASA's newest Mars robot has already captured a photo of its rusty, dusty home. The InSight lander touched down on the Red Planet today (Nov. 26) just before 3 p.m. EST (2000 GMT) and beamed home its first image from the surface mere minutes later.



via space.com,

UNIVERSITY OF WASHINGTON

PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

Discourse & Coref Recap

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.

- ***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.

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.

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

Referring Expressions

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

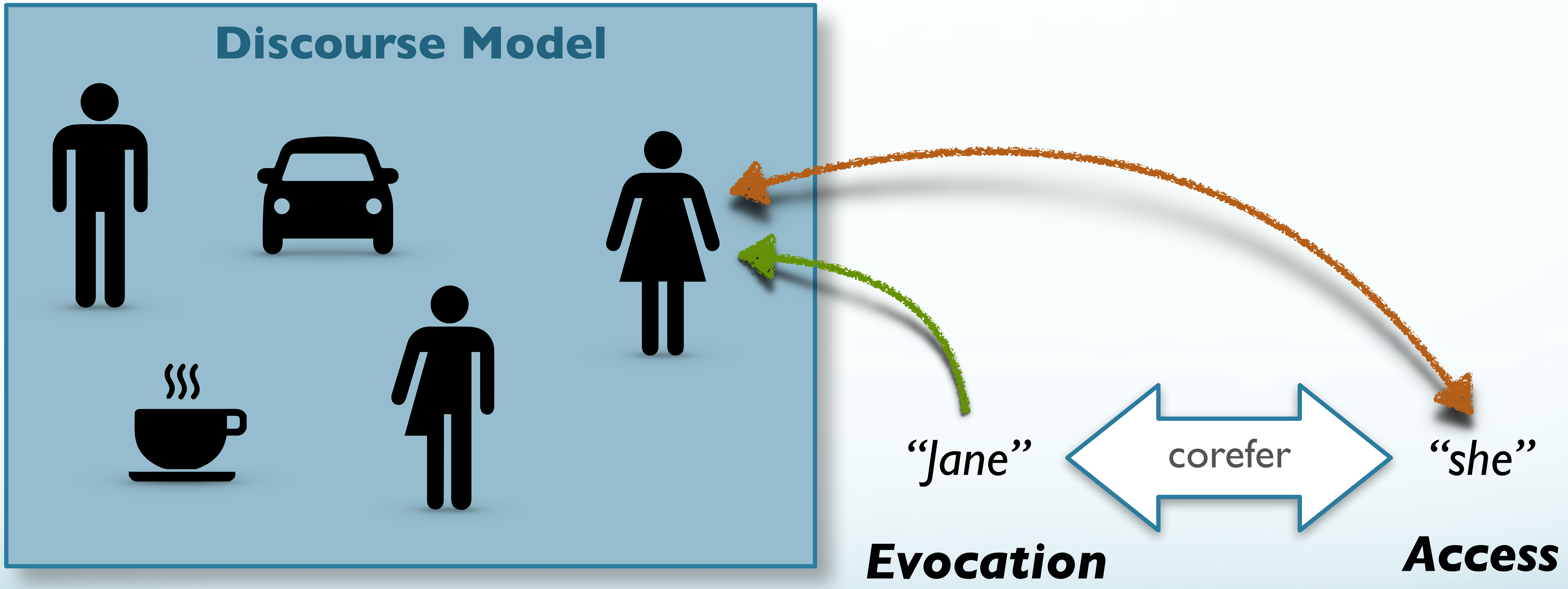
Referring Expressions

- *Queen Elizabeth* – *she/her* – *the Queen* – *HRM* – *the British Monarch*

- "Correct" form depends on discourse context
  - ***she, her*** presume prior mention or presence in the world
  - ***the Queen*** presumes an Anglocentric geopolitical discourse context generally or the UK (or British Commonwealth) specifically

    (…i.e. likely a different interpretation during a RPDR viewing party.)

Reference and Model



UNIVERSITY OF WASHINGTON

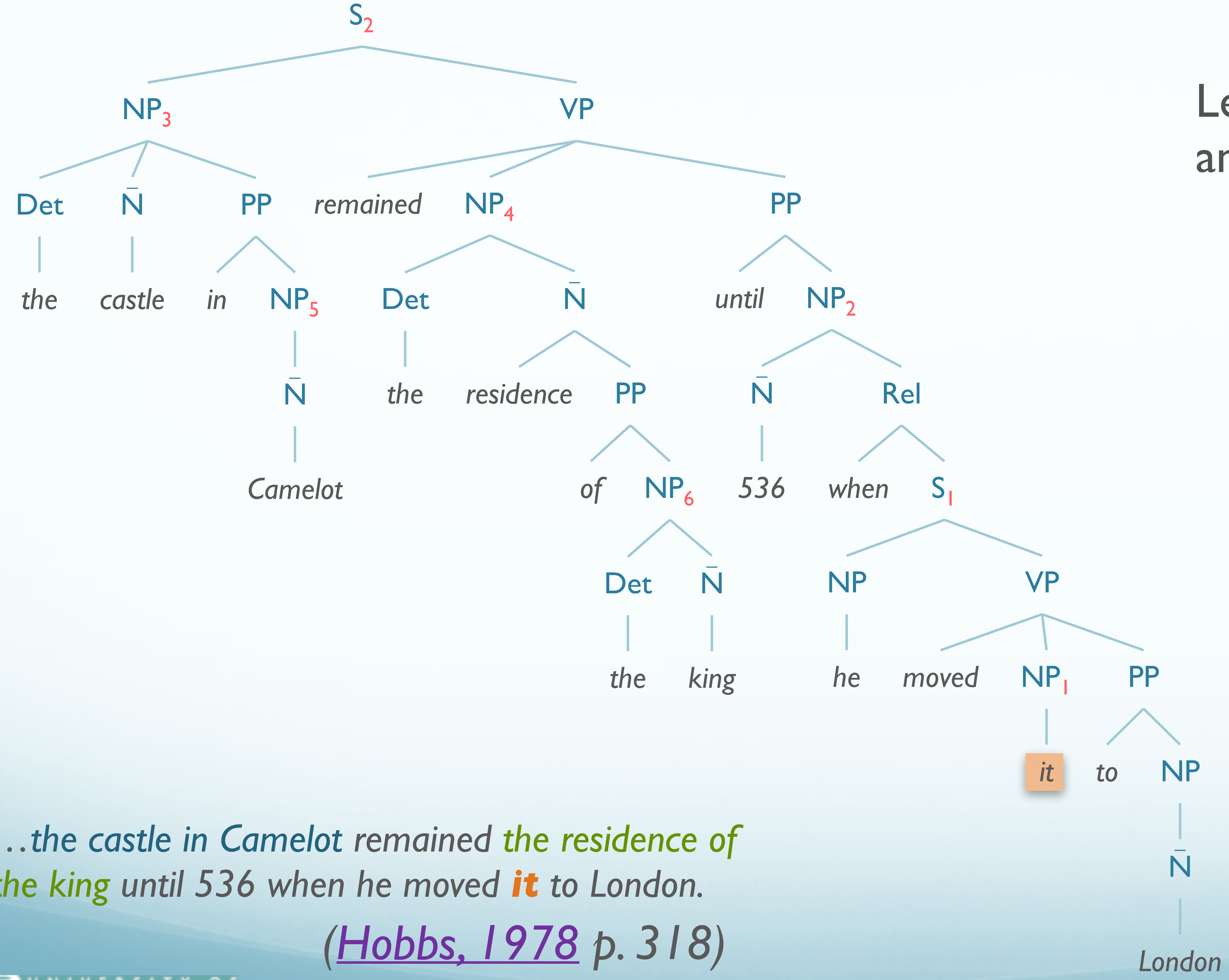
PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

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

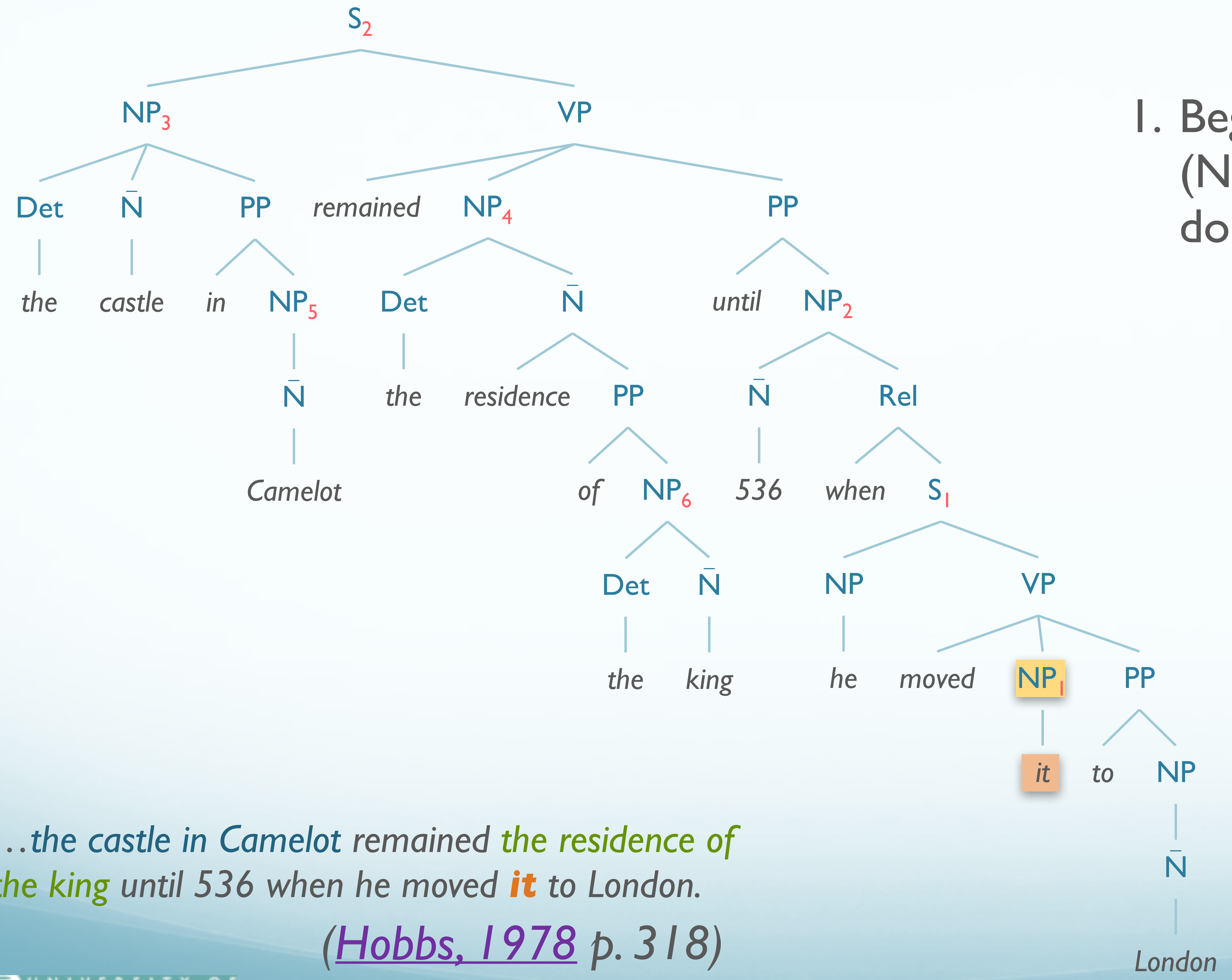
Hobbs Algorithm Walkthrough (Improved)

Let's figure out what the antecedent for "*it*" is



*…the castle in Camelot remained the residence of the king until 536 when he moved **it** to London.*

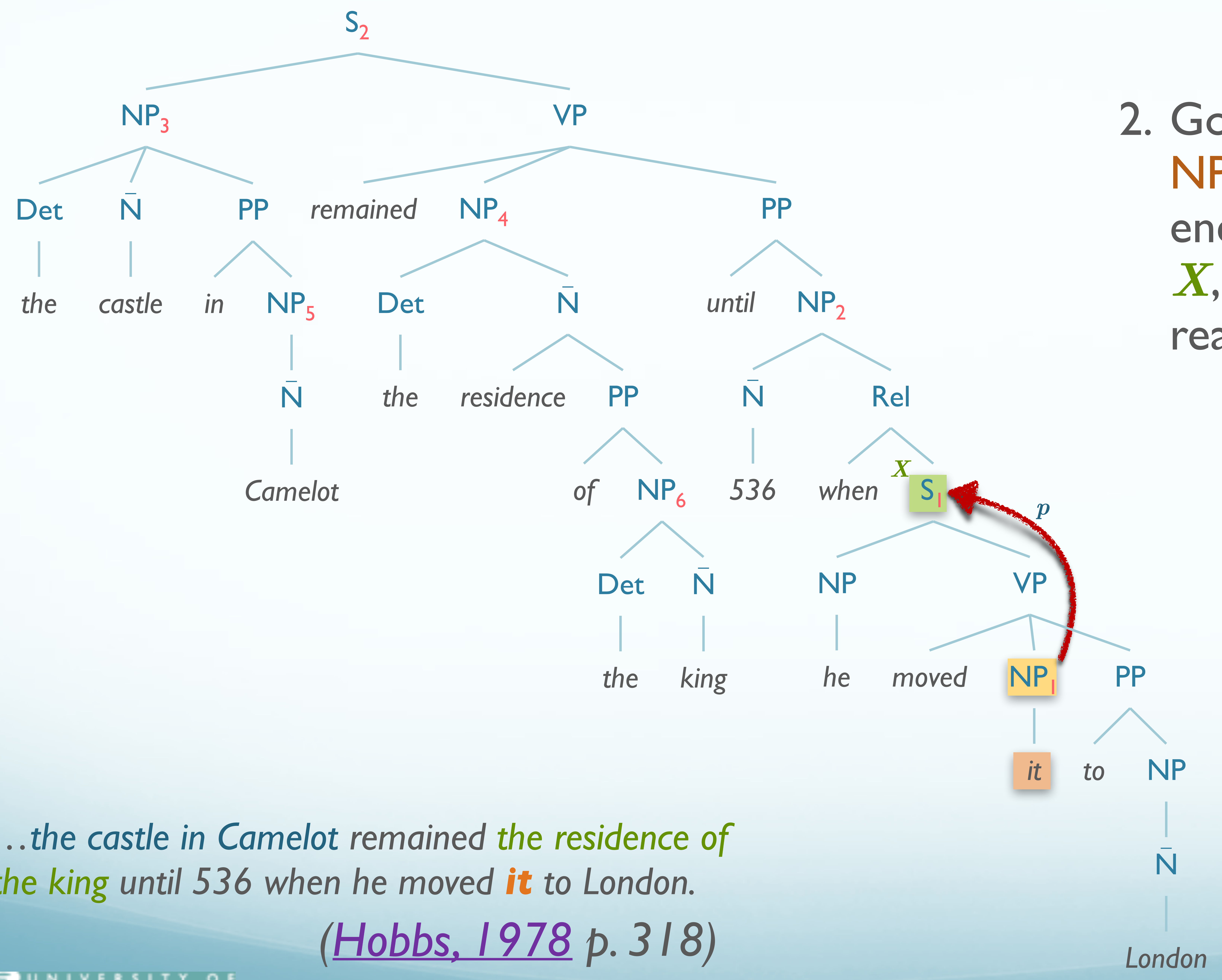
(Hobbs, 1978 p. 318)



1. Begin at the noun phrase (NP) node immediately dominating the pronoun

*…the castle in Camelot remained the residence of the king until 536 when he moved **it** to London.*

([Hobbs, 1978](#) p. 318)



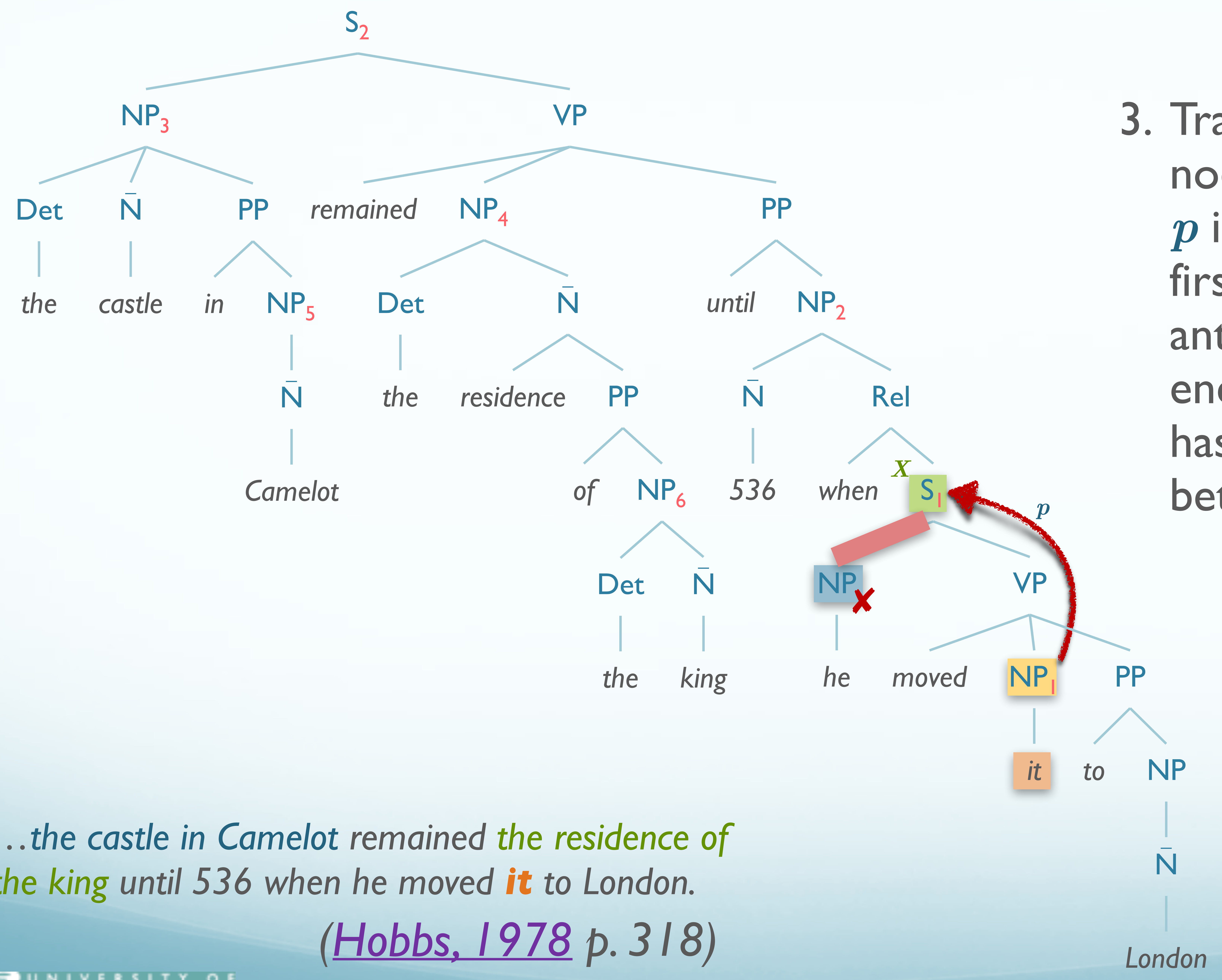
2. Go up the tree to the first NP or sentence (S) node encountered. Call this node ***X***, and call the path used to reach it ***p***.

*…the castle in Camelot remained the residence of the king until 536 when he moved **it** to London.*

(Hobbs, 1978 p. 318)

UNIVERSITY OF WASHINGTON

PROFESSIONAL MASTER'S IN COMPUTATIONAL LINGUISTICS

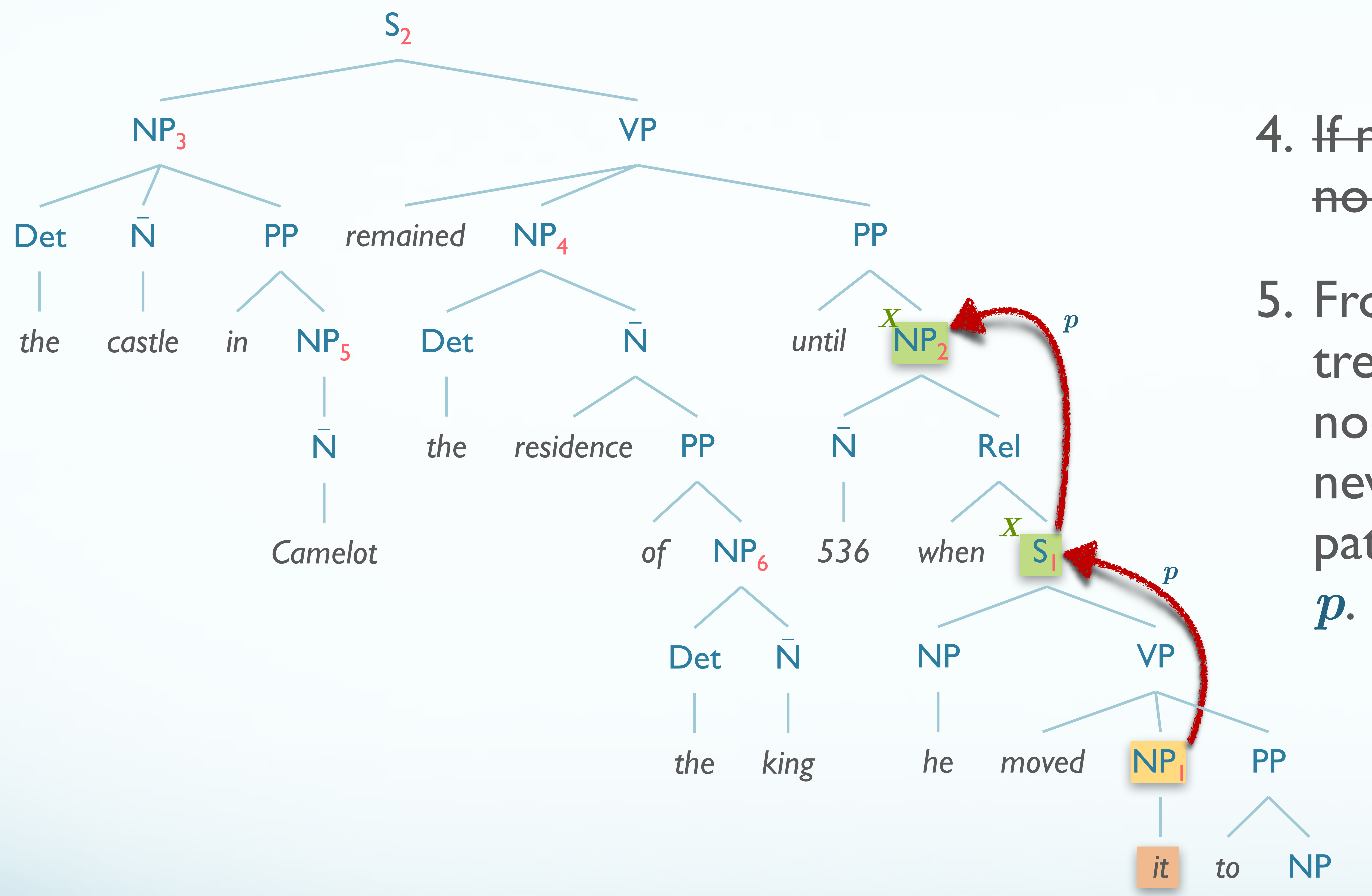


3. Traverse all branches below node X to the left of path p in a left-to-right, breadth-first fashion. Propose as the antecedent any encountered NP node that has an NP or S node between it and X.

No NP or S between "he" NP and X

*…the castle in Camelot remained the residence of the king until 536 when he moved **it** to London.*

(Hobbs, 1978 p. 318)

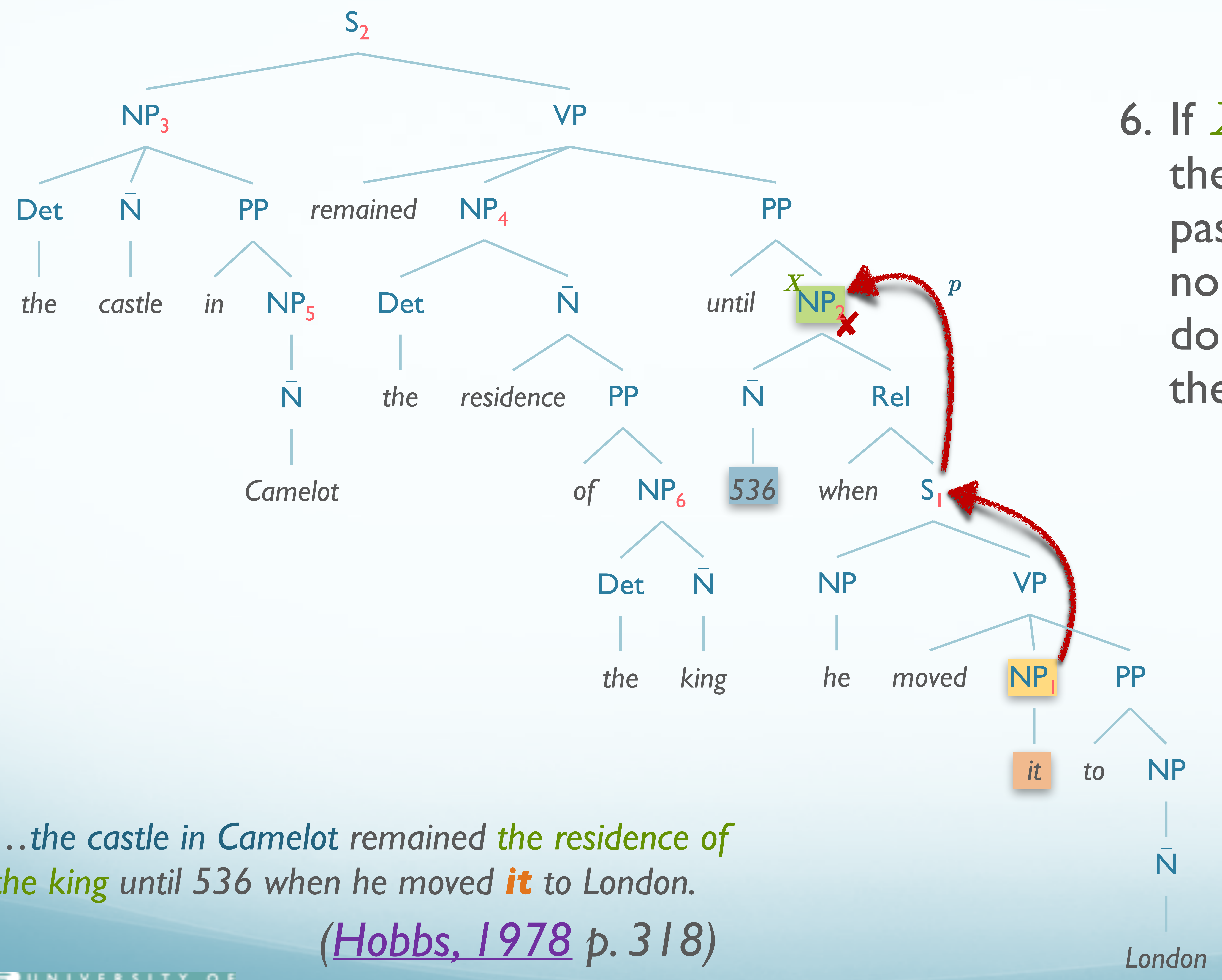


4. ~~If node *X* is the highest S node in the sentence…~~

5. From node *X*, go up the tree to the first NP or S node encountered. Call this new node *X*, and call the path traversed to reach it *p*.

…the castle in Camelot remained the residence of the king until 536 when he moved **it** to London.

([Hobbs, 1978](#) p. 318)

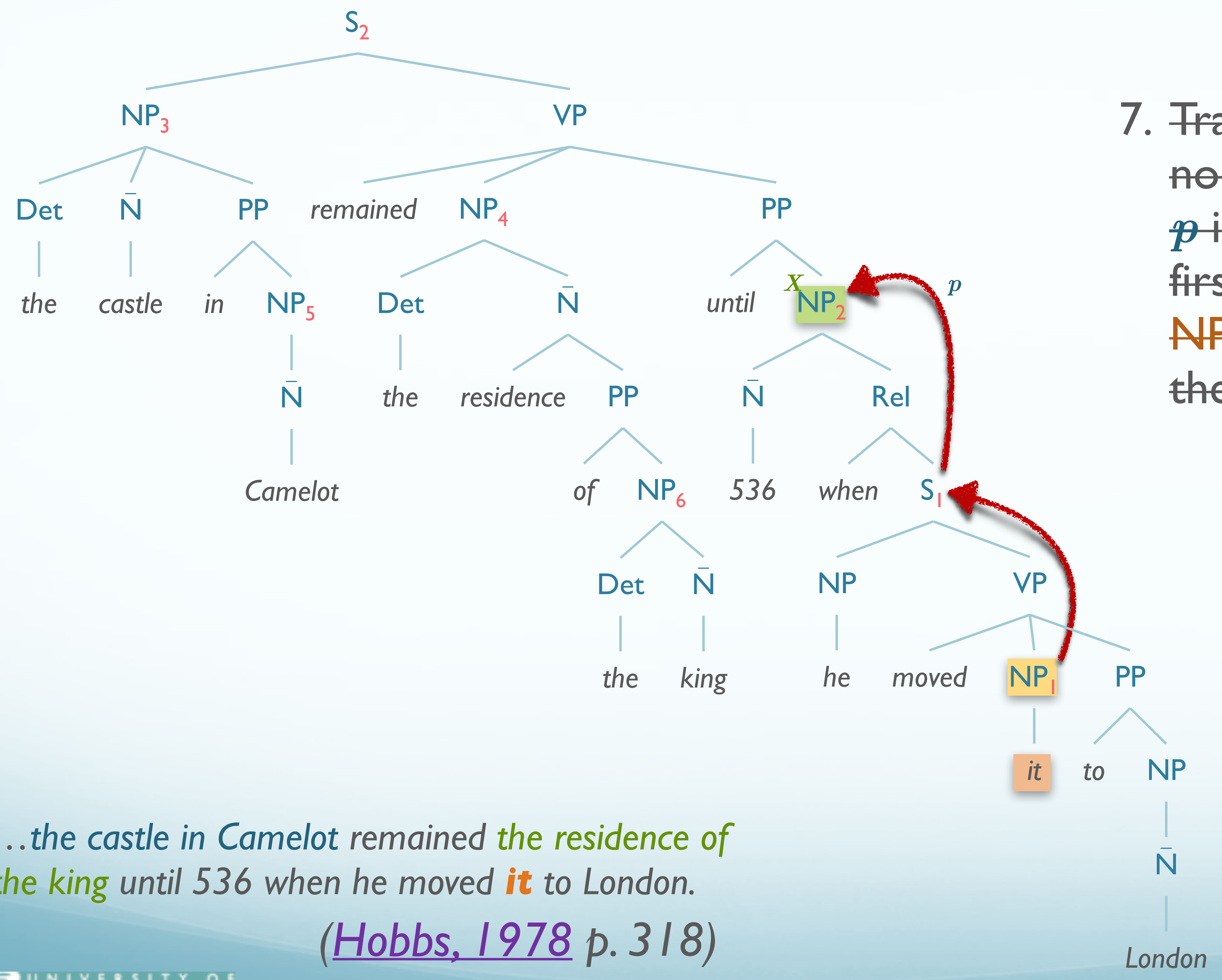


6. If X is an NP node and if the path p to X did not pass through the Nominal node that X immediately dominates, propose X as the antecedent.

"536" can't be "moved"!

…the castle in Camelot remained the residence of the king until 536 when he moved ***it*** *to London.*

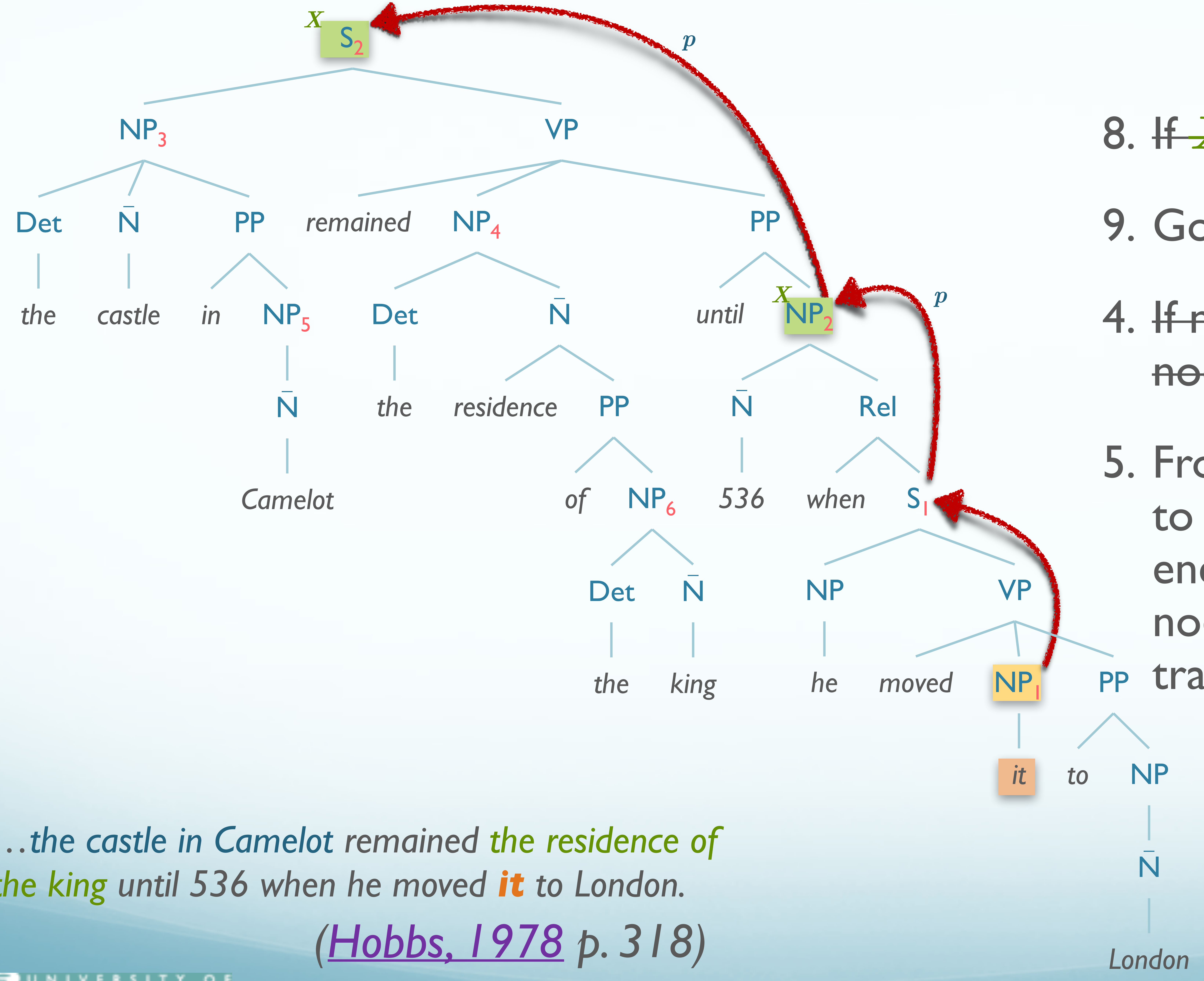
(Hobbs, 1978 p. 318)



7. ~~Traverse all branches below node *X* to the *left* of path *p* in a left-to-right, breadth-first manner. Propose any NP node encountered as the antecedent.~~

…the castle in Camelot remained the residence of the king until 536 when he moved **it** to London.

(Hobbs, 1978 p. 318)



8. ~~If X is an S node…~~

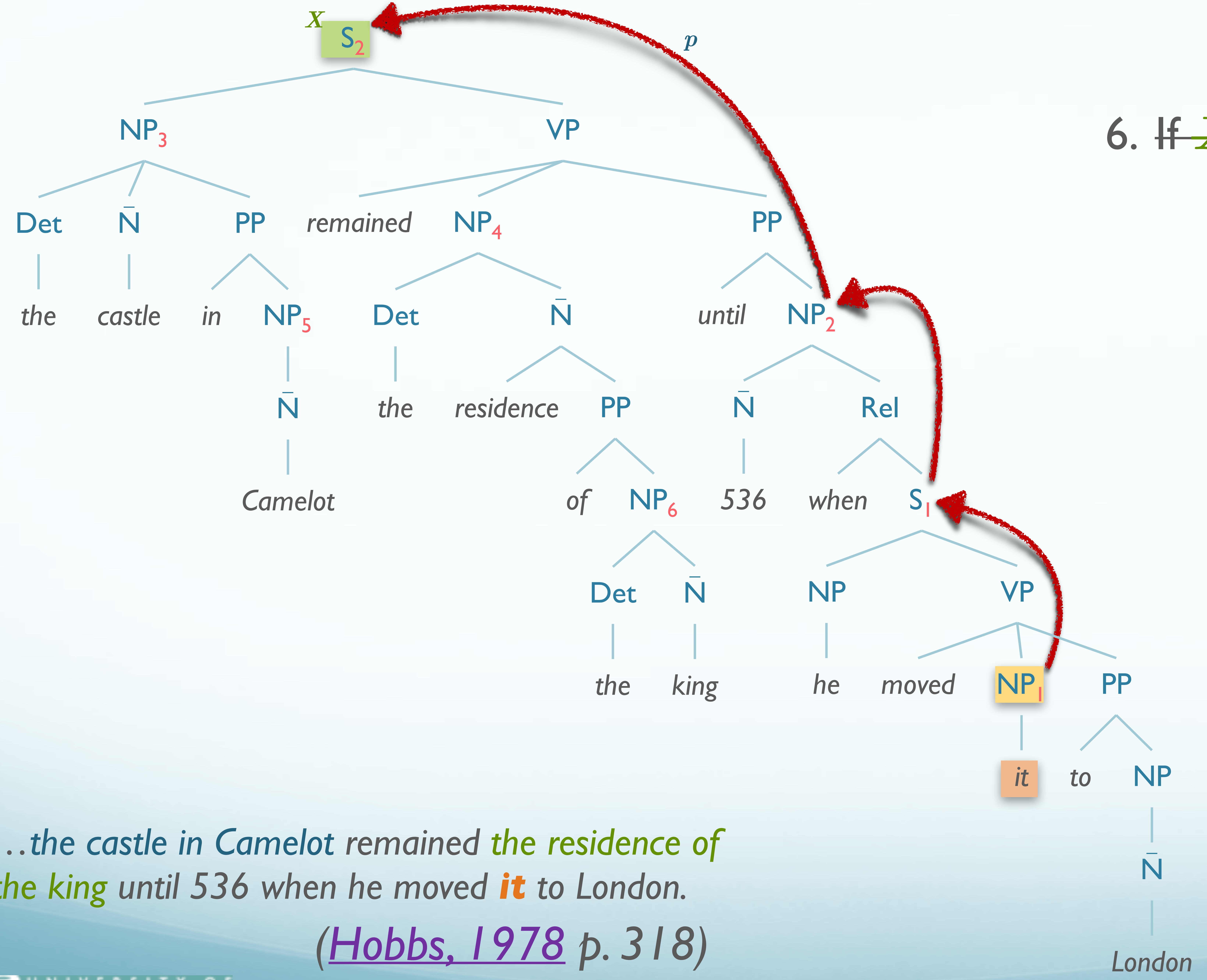
9. Go to step 4.

4. ~~If node X is the highest S node in the sentence…~~

5. From node X, go up the tree to the first NP or S node encountered. Call this new node X, and call the path traversed to reach it p.

*…the castle in Camelot remained the residence of the king until 536 when he moved **it** to London.*

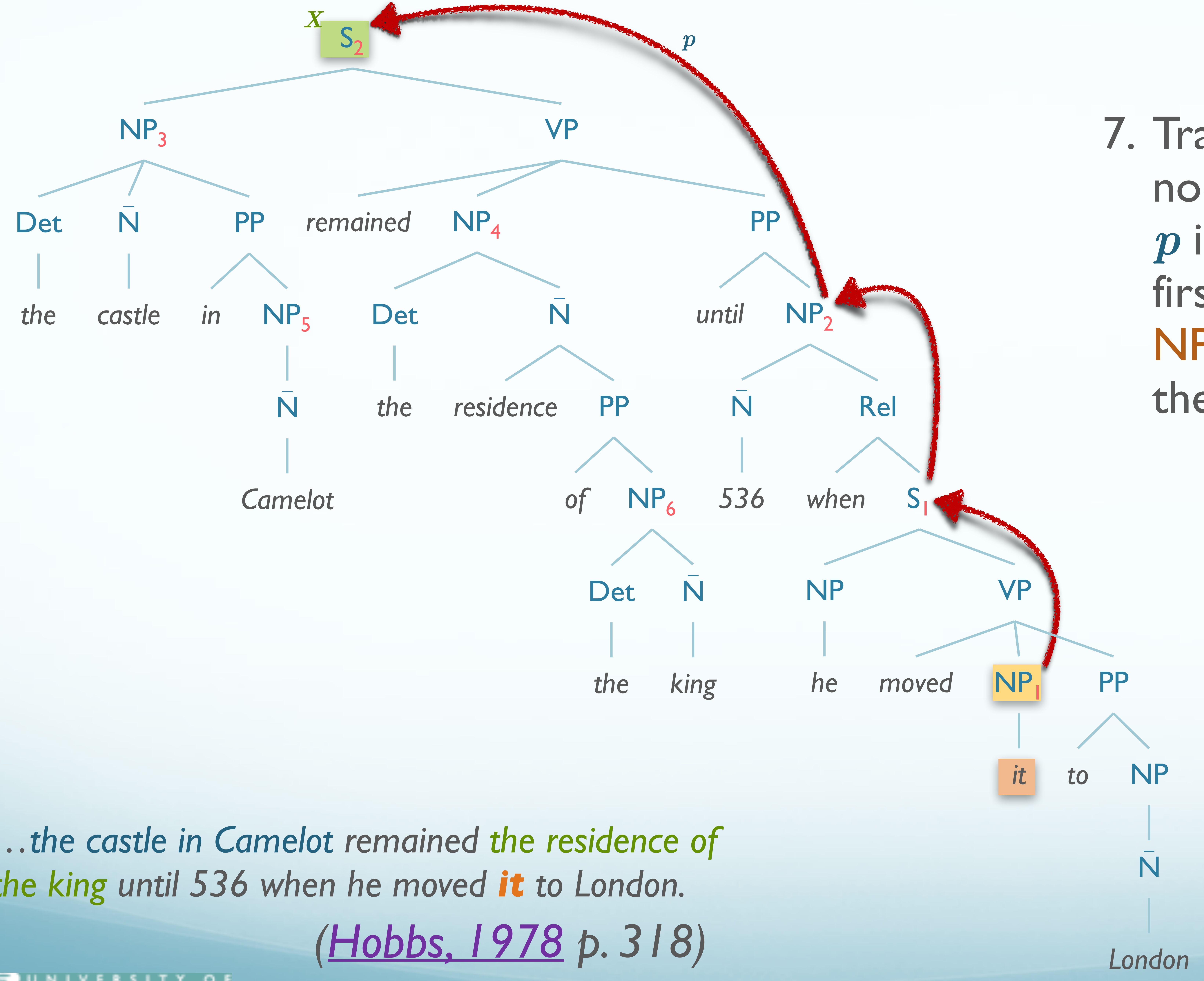
(Hobbs, 1978 p. 318)



6. ~~If *X* is an NP node~~…

…the castle in Camelot remained the residence of the king until 536 when he moved ***it*** *to London.*

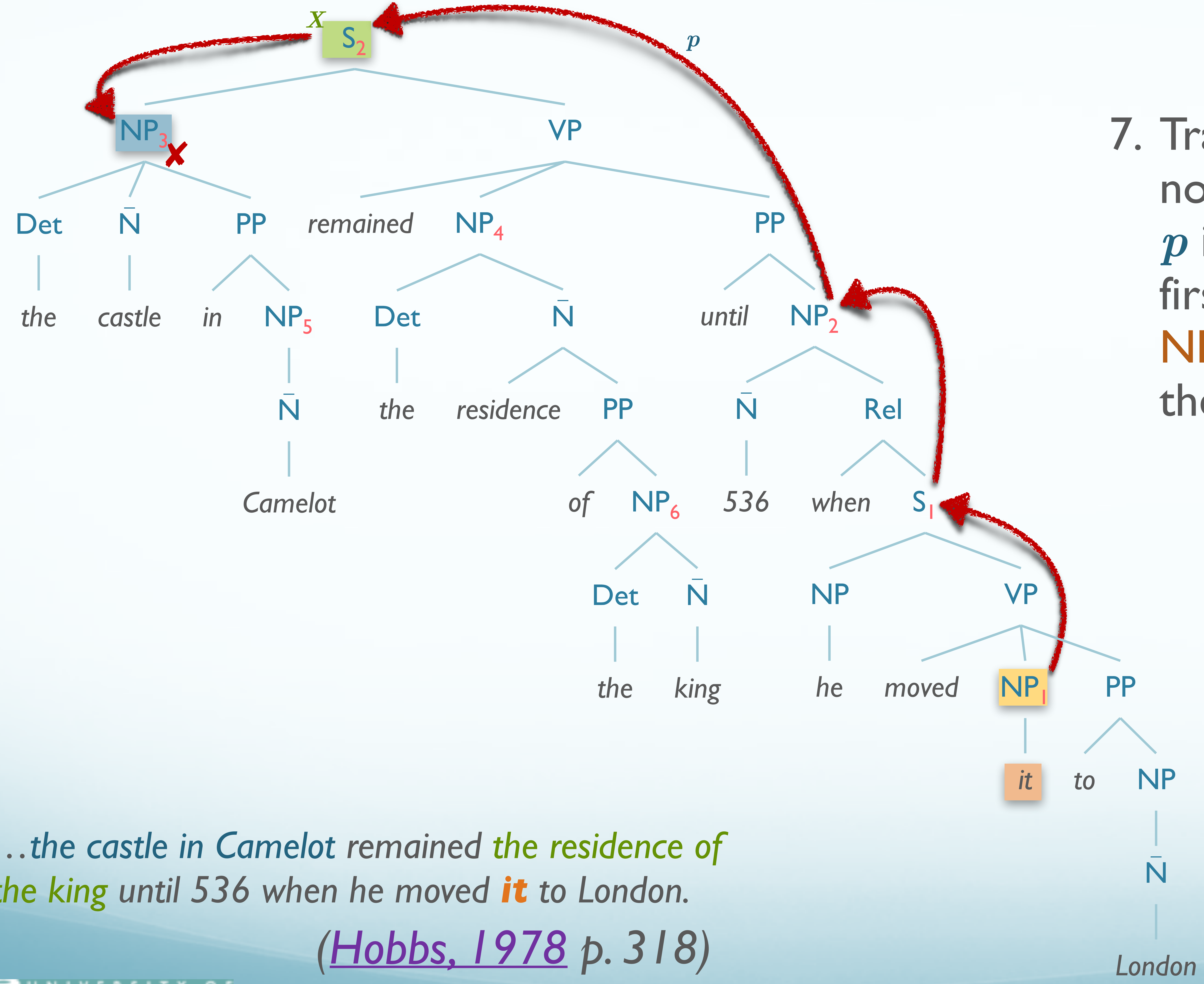
(Hobbs, 1978 p. 318)



7. Traverse all branches below node X to the *left* of path p in a left-to-right, breadth-first manner. Propose any NP node encountered as the antecedent.

*…the castle in Camelot remained the residence of the king until 536 when he moved **it** to London.*

(Hobbs, 1978 p. 318)



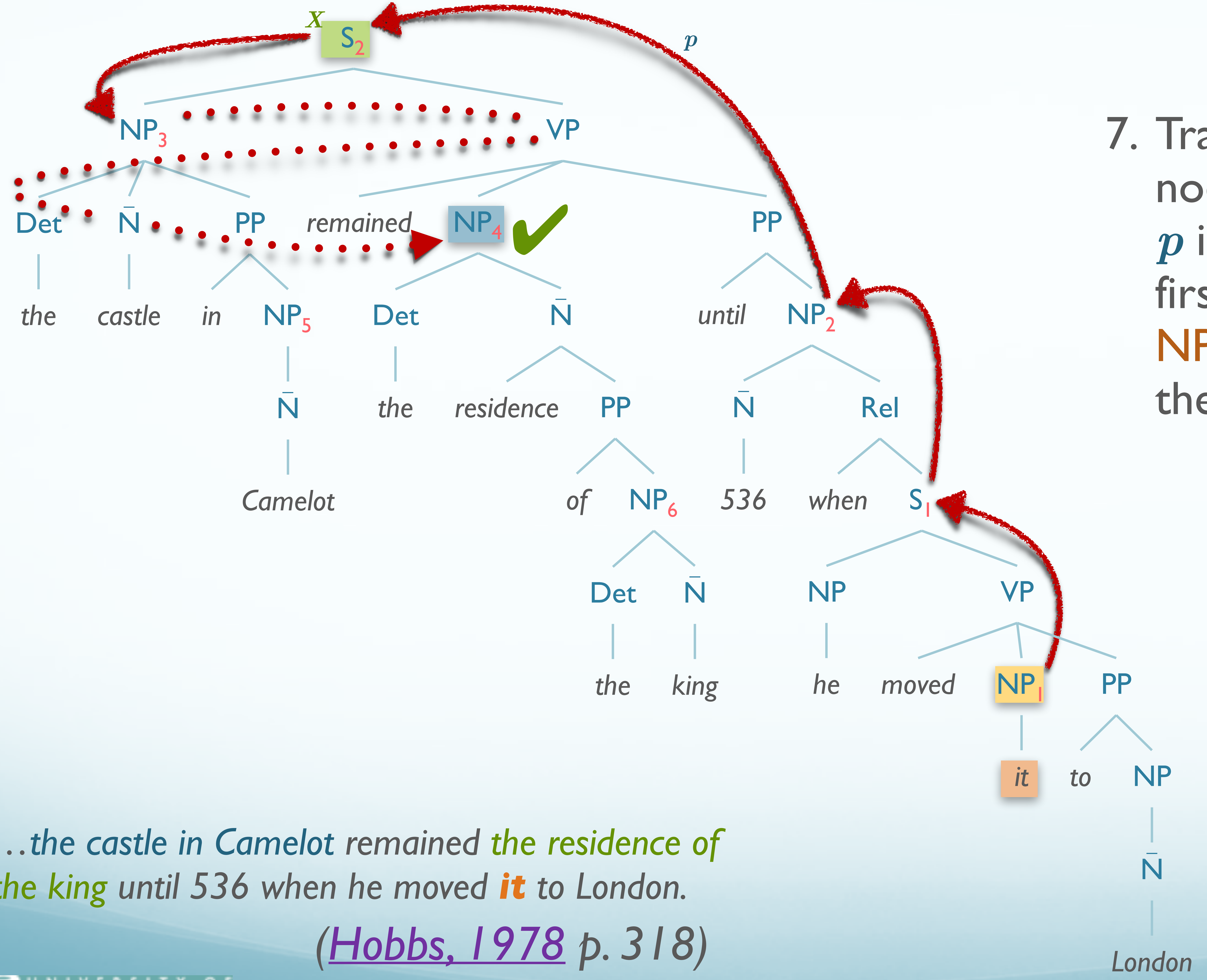
7. Traverse all branches below node X to the *left* of path p in a left-to-right, breadth-first manner. Propose any NP node encountered as the antecedent.

Moving castles? 🤔

…the castle in Camelot remained the residence of the king until 536 when he moved ***it*** *to London.*

(Hobbs, 1978 p. 318)

UNIVERSITY OF WASHINGTON



7. Traverse all branches below node X to the *left* of path p in a left-to-right, breadth-first manner. Propose any NP node encountered as the antecedent.

"the residence of the king"

…the castle in Camelot remained the residence of the king until 536 when he moved ***it*** *to London.*

(Hobbs, 1978 p. 318)

Hobbs Algorithm Detail (Hobbs, 1978)

1. Begin at the noun phrase (NP) node immediately dominating the pronoun
2. Go up the tree to the first NP or sentence (S) node encountered. Call this node X, and call the path used to reach it p.
3. Traverse all branches below node X to the left of path p in a left-to-right, breadth-first fashion. Propose as the antecedent any encountered NP node that has an NP or S node between it and X.
4. If node X is the highest S node in the sentence, traverse the surface parse trees of previous sentences in the text in order of recency, the most recent first; each tree is traversed in a left-to-right, breadth-first manner, and when an NP node is encountered, it is proposed as antecedent. If X is not the highest S node in the sentence, continue to step 5.

Hobbs Algorithm Detail (Hobbs, 1978)

5. From node X, go up the tree to the first NP or S node encountered. Call this new node X, and call the path traversed to reach it p.

6. If X is an NP node and if the path p to X did not pass through the Nominal node that X immediately dominates, propose X as the antecedent.

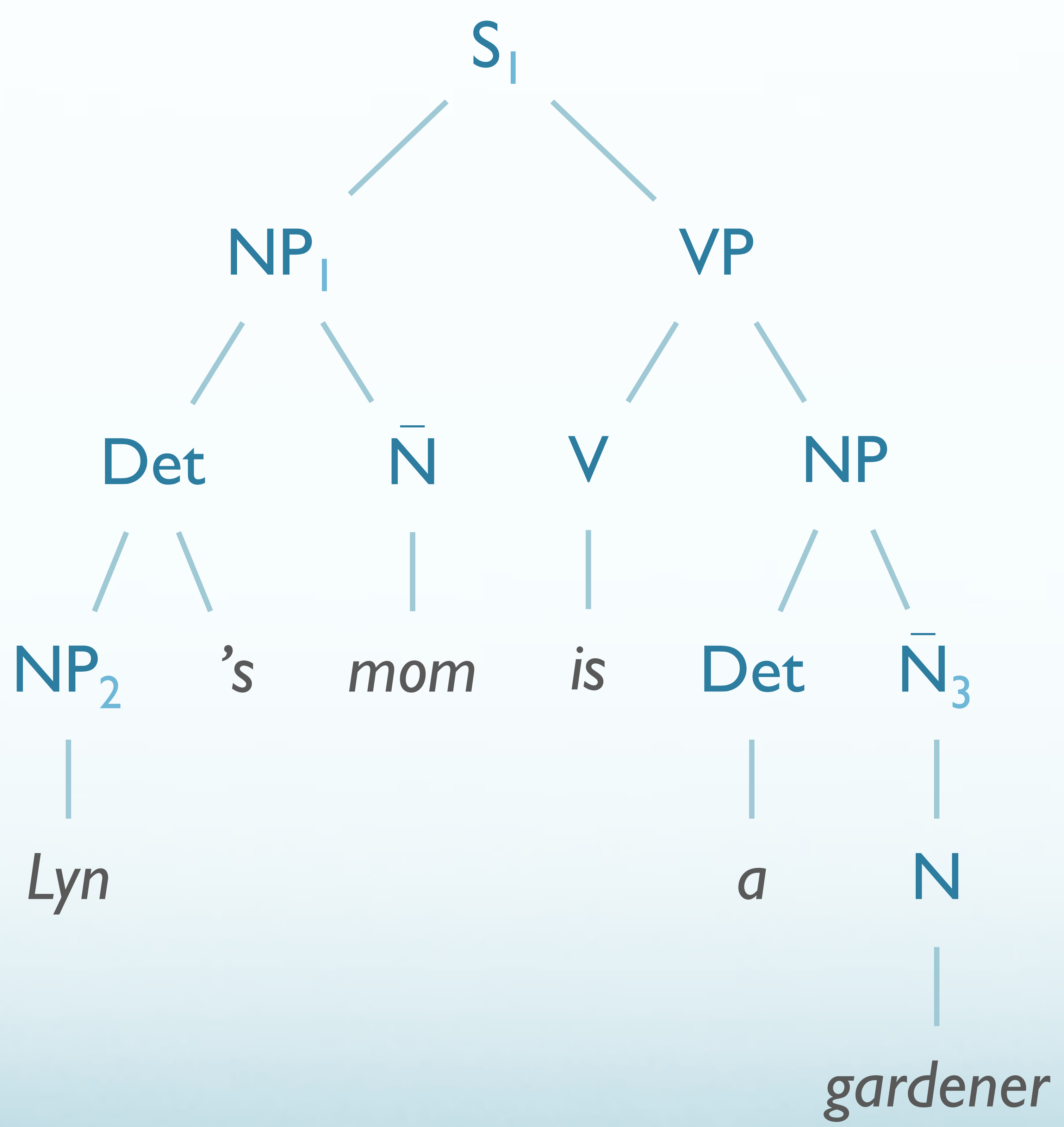
7. Traverse all branches below node X to the *left* of path p in a left-to-right, breadth-first manner. Propose any NP node encountered as the antecedent.

8. If X is an S node, traverse all branches of node X to the ***right*** of path p in a left-to-right, breadth-first manner, but do not go below any NP or S node encountered. Propose any NP node encountered as the antecedent.

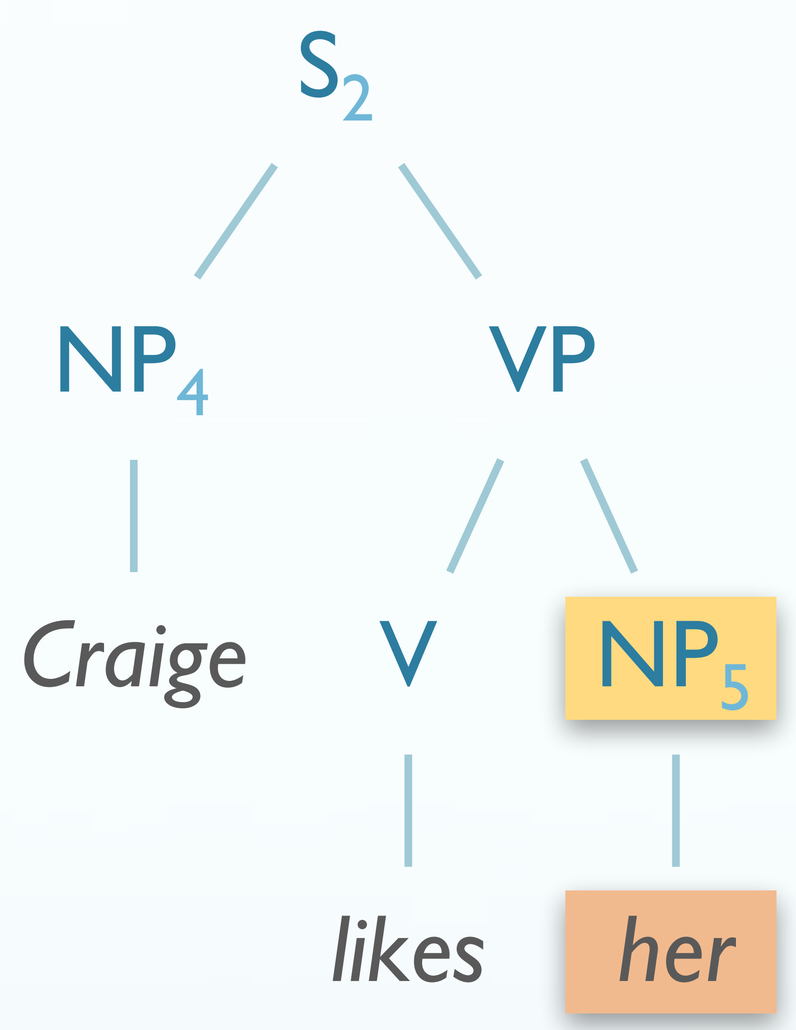
9. Go to step 4.

Hobbs Example

Lyn's mom is a gardener.



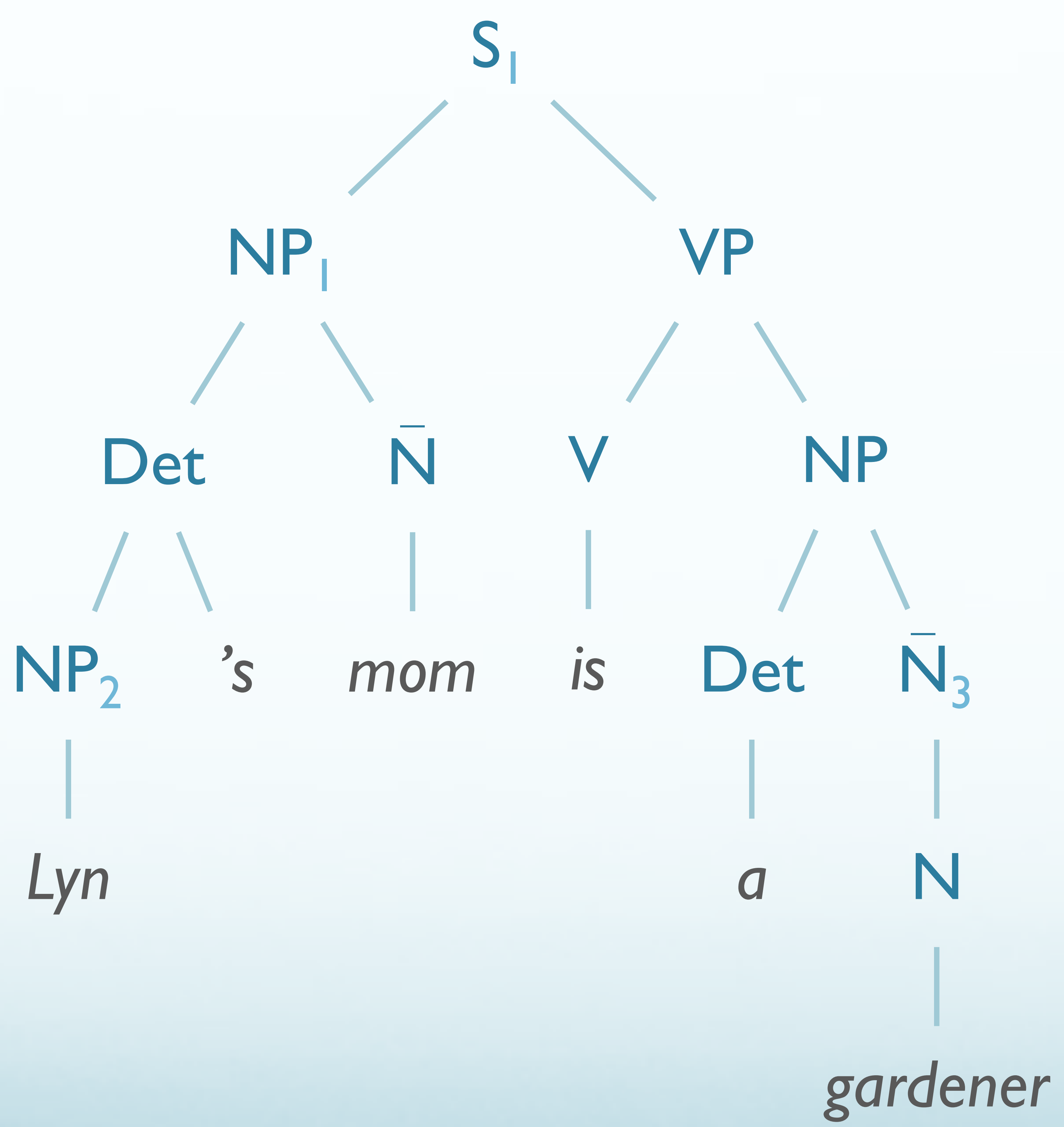
Craige likes ***her****.*



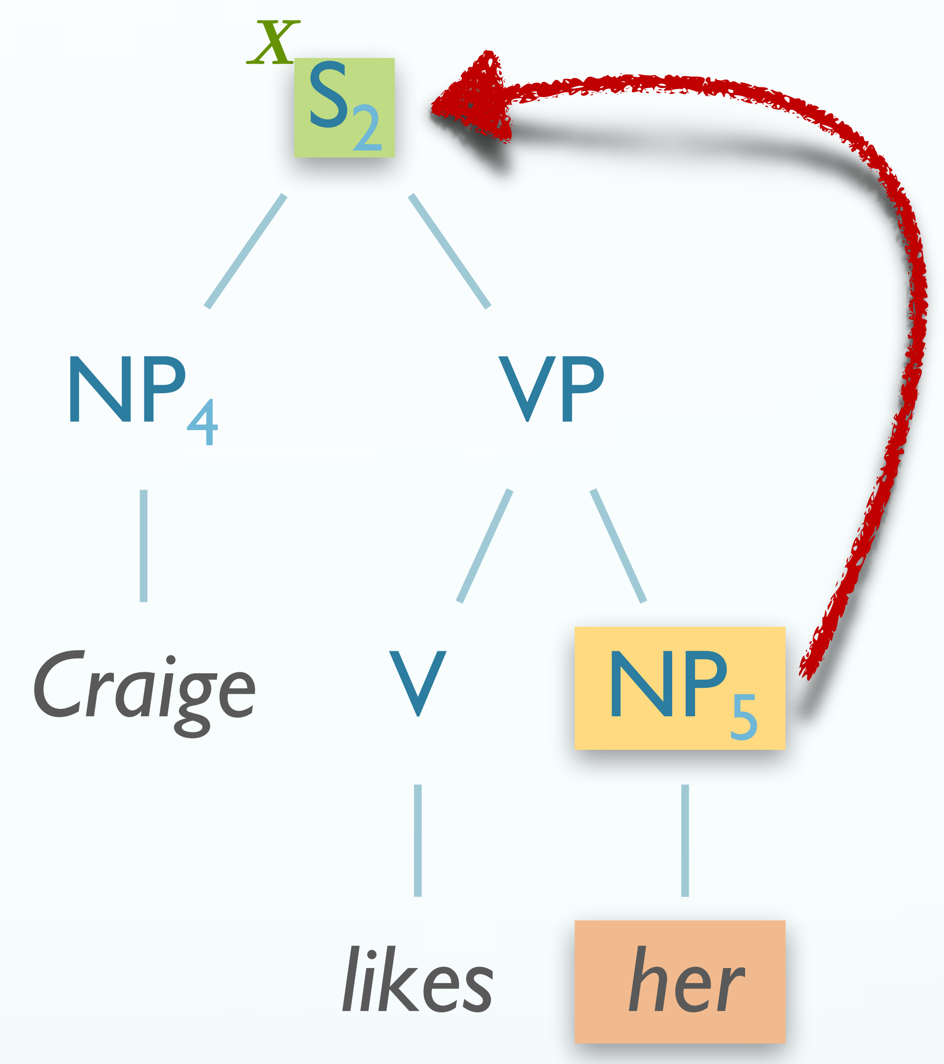
1. Begin at the noun phrase (NP) node immediately dominating the pronoun

Hobbs Example

Lyn's mom is a gardener.



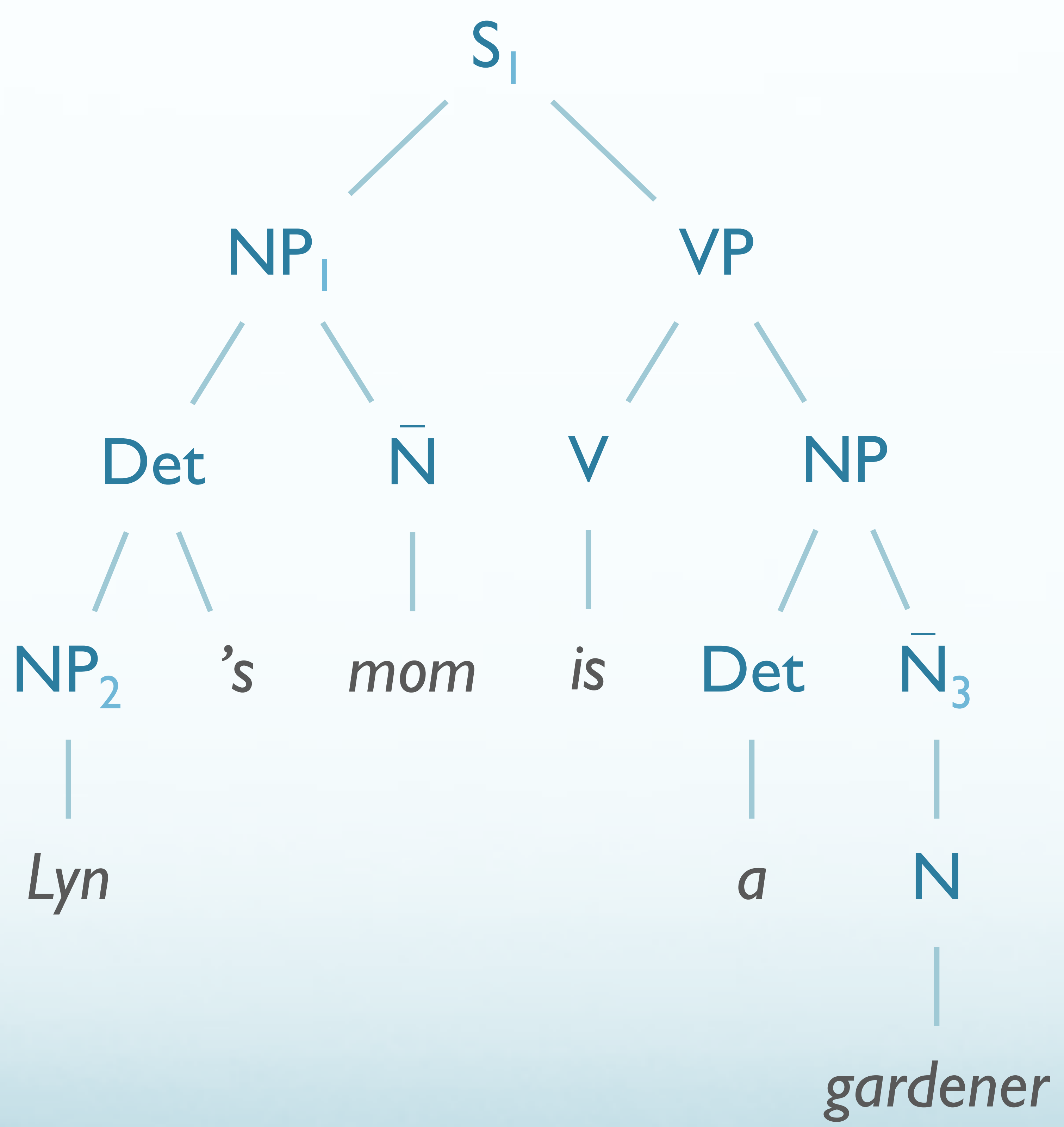
Craige likes ***her***.



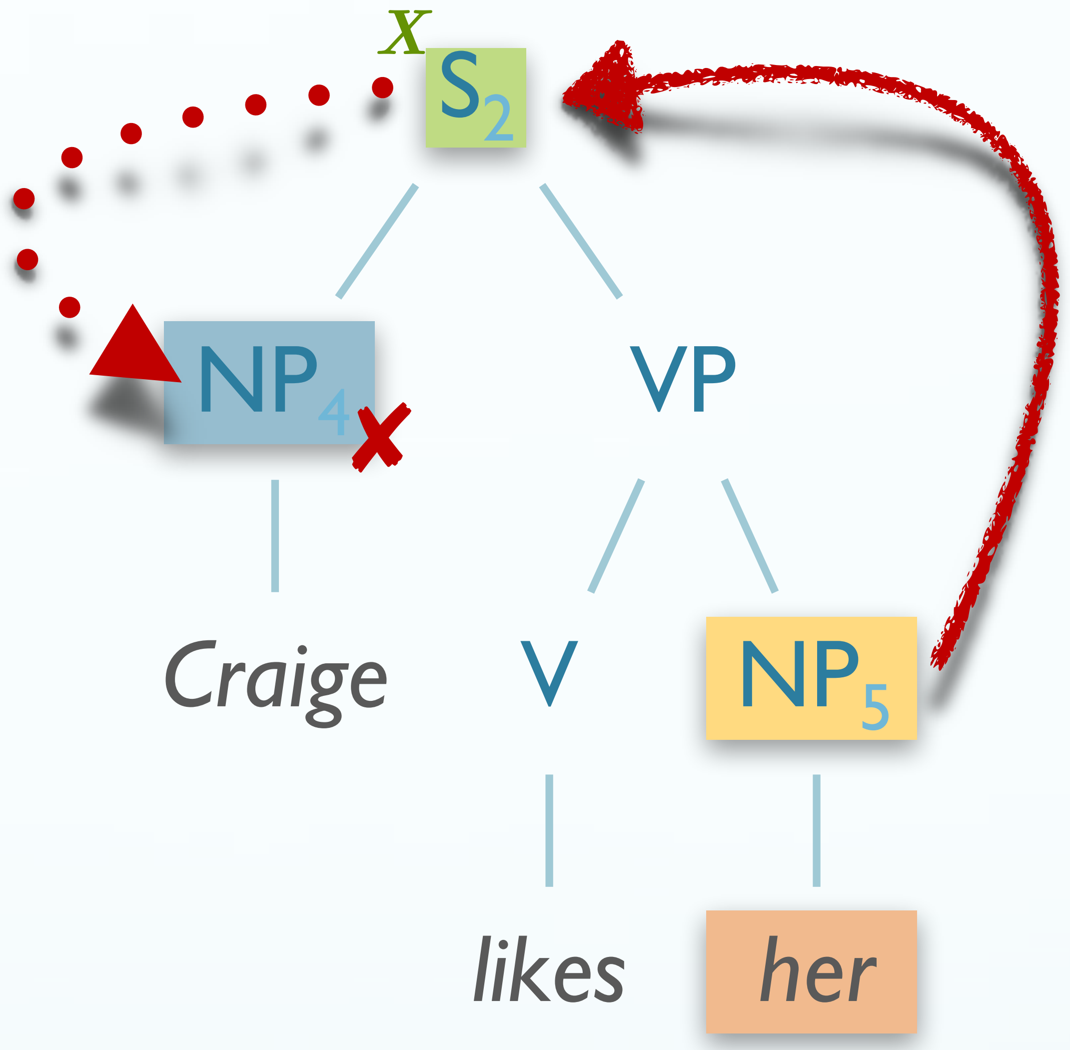
2. Go up the tree to the first NP or sentence (S) node encountered. Call this node X, and call the path used to reach it p.

Hobbs Example

Lyn's mom is a gardener.



*Craige likes **her**.*

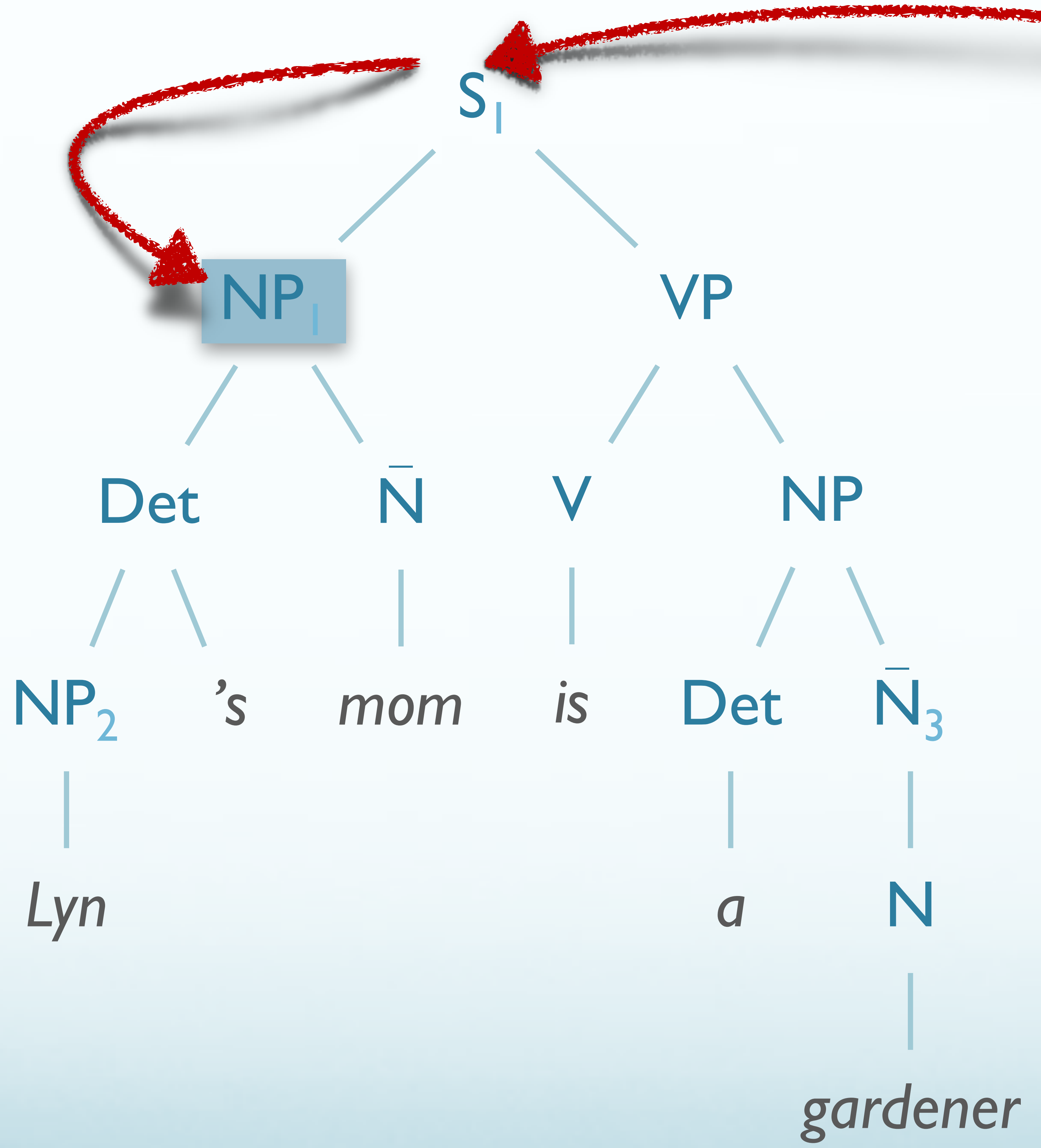


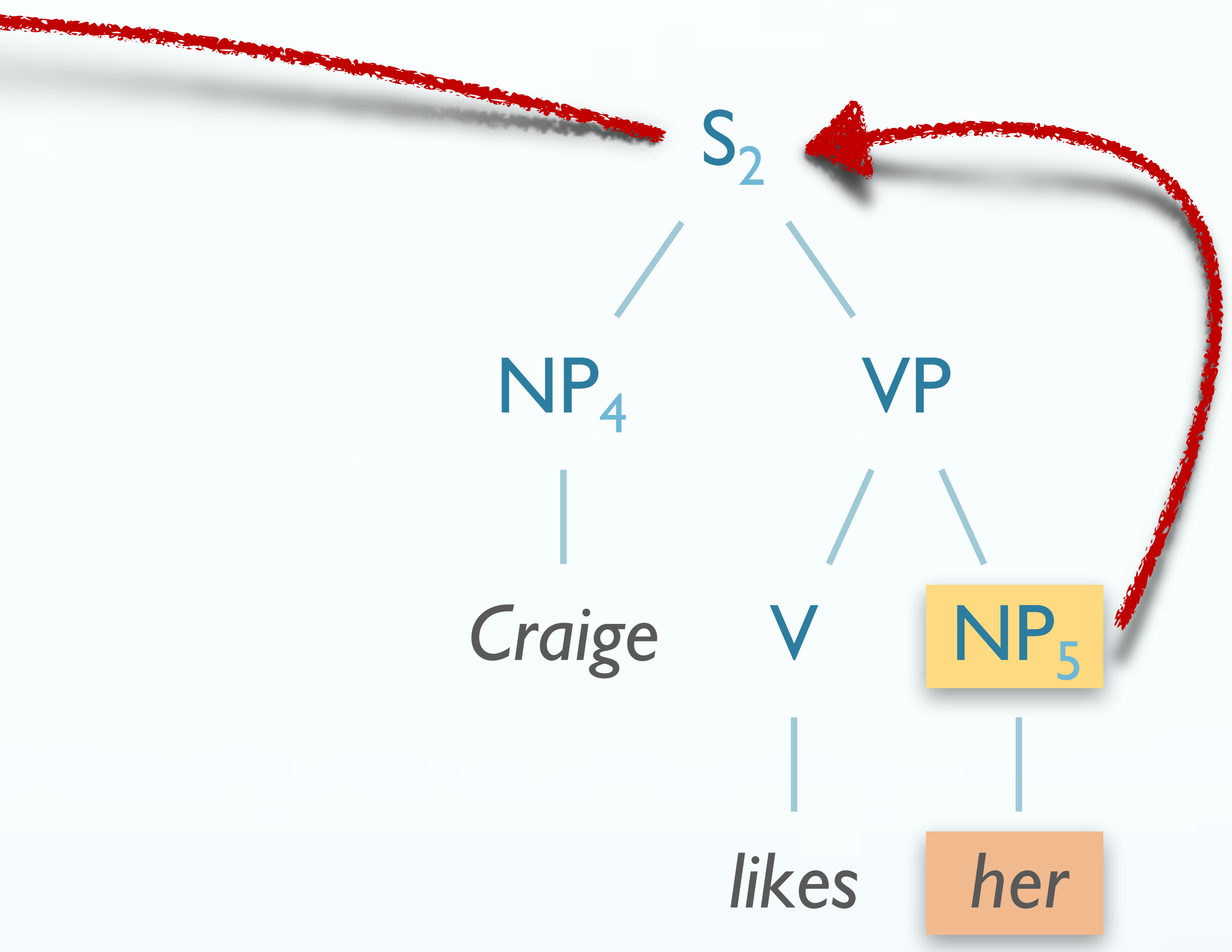
3. Traverse all branches below node X to the left of path p in a left-to-right, breadth-first fashion. Propose as the antecedent any encountered NP node that has an NP or S node between it and X.

Hobbs Example

Lyn's mom is a gardener.

Craige likes ***her***.





4. If node X is the highest S node in the sentence, traverse the surface parse trees of previous sentences in the text in order of recency, the most recent first; each tree is traversed in a left-to-right, breadth-first manner, and when an NP node is encountered, it is proposed as antecedent.

Hobbs Example

- What about…?
  - *Lyn's mom ~~is~~ hired a gardener.*
  - *Craige likes her.*

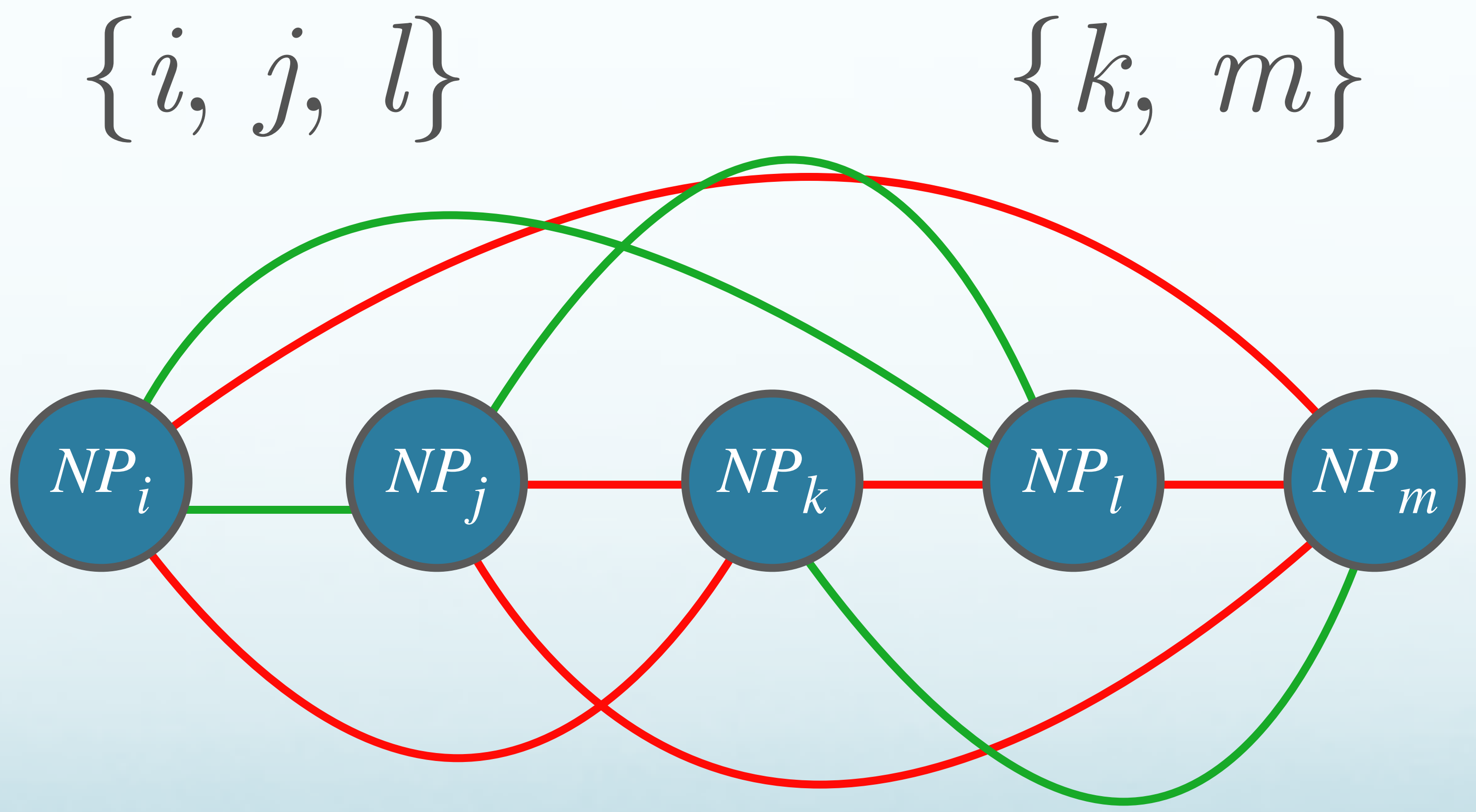
Other Coreference Approaches

Data-driven Reference Resolution

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

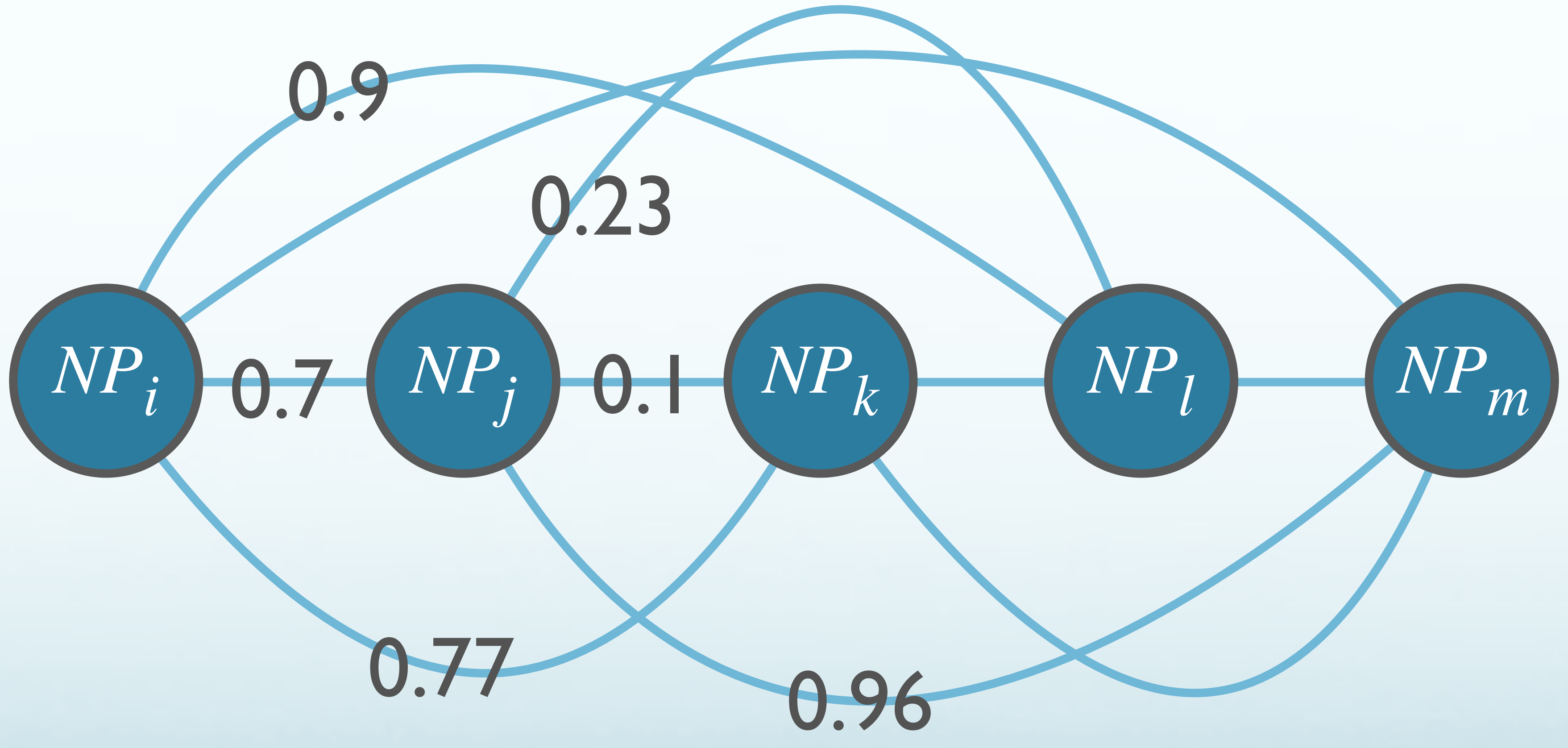
Other kinds of Coreference Models

- **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



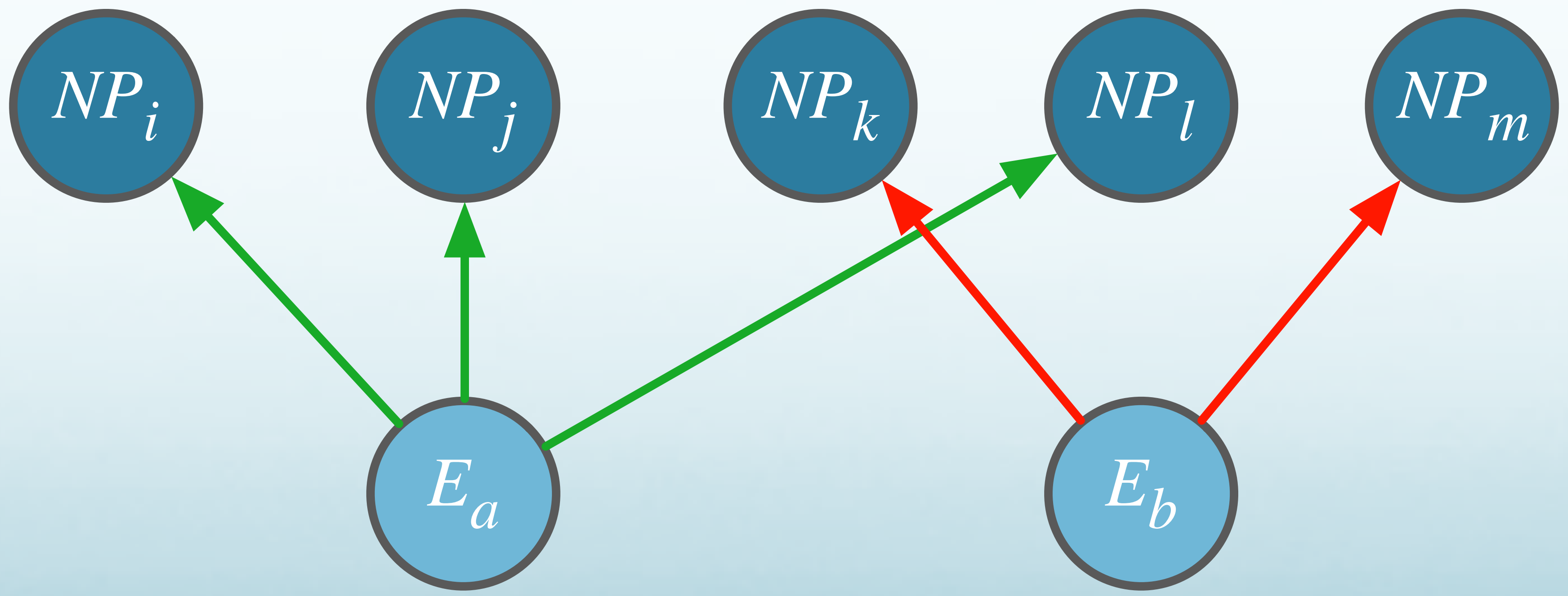
Other kinds of Coreference Models

- **Mention 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



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
  - German, Czech, Japanese, Spanish, Catalalan, Medline

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?

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)

Clustering by Classification

- **Mention-pair** style system:
  - For each pair of NPs, classify +/– coreferent
  - Linked pairs form ***coreferential chains***
  - Process candidate pairs from end "backward" to start
  - All mentions of an entity appear in a single chain.

Mention-Pair Systems: The Locality Problem

- **Problem**:
  - Mention-pair classifier approach makes local decisions w/large number of features
  - Each local decision shouldn't really be considered transitive
  - Local decisions can't exploit global constraints
  - Low precision features may overwhelm less frequent, high precision ones
- **Solution**:
  - Apply ***sieve*** from highest to lowest precision features
  - Make high-precision information available

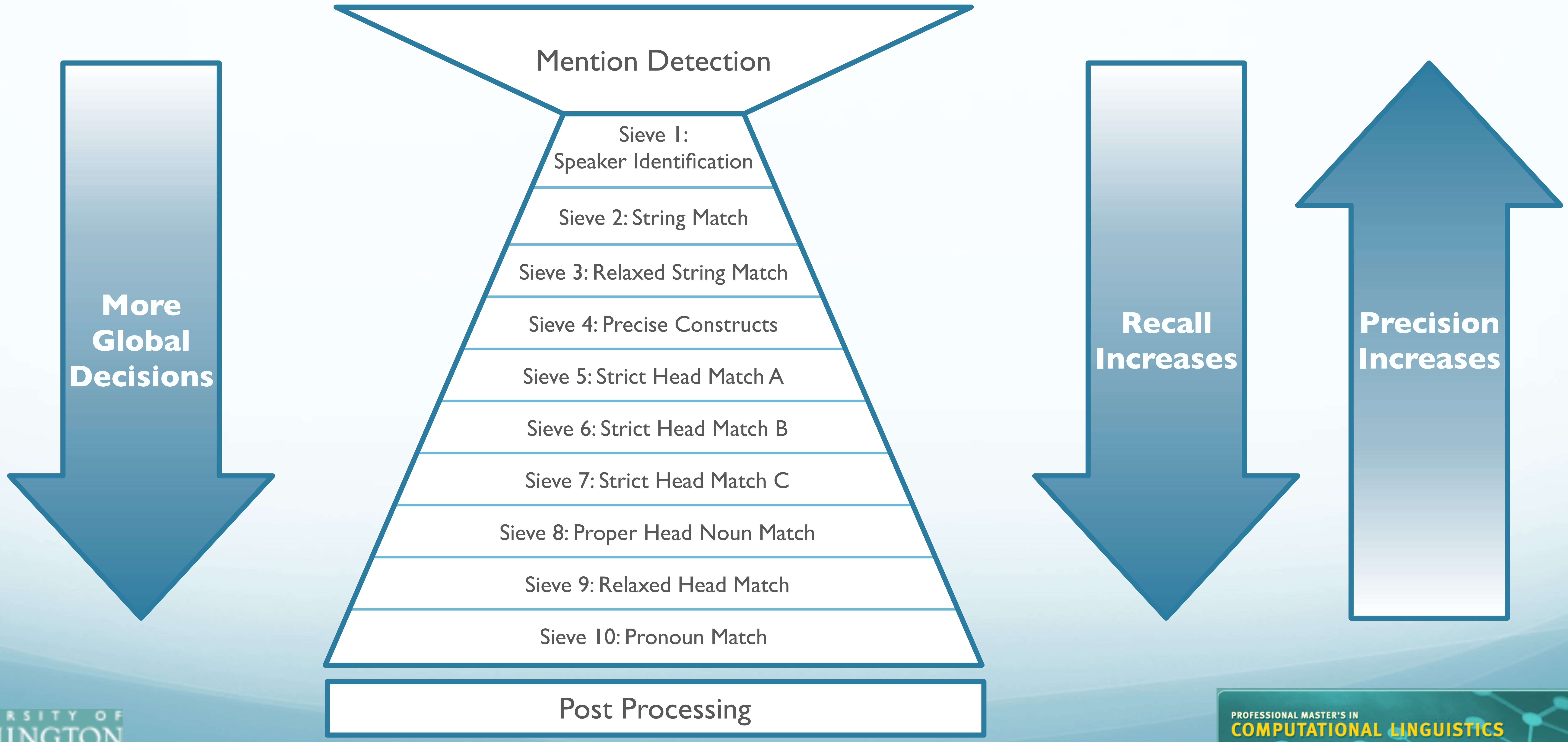
Multi-pass Sieve Approach (Raghunathan et al, 2010)

- Apply tiers of deterministic coreference modules
- Aggregate information across mentions in cluster
  - Share attributes based on prior tiers
- Simple, extensible architecture
  - Outperforms many other approaches

Multi-pass Sieve Approach (Raghunathan et al, 2010)

- Example:
  - *The second attack occurred after some rocket firings aimed, apparently, toward **the israelis**, apparently in retaliation. **we**'re checking our facts on that one. … the president, quoted by ari fleischer, his spokesman, is saying he's concerned the strike will undermine efforts by palestinian authorities to bring an end to terrorist attacks and does not contribute to the security of **israel**.*
- In most local decisions, "*we*" would be ***incorrectly*** linked to "*the israelis*"
- But… exact-string-match nature of *israel* ⇔ *israelis* is very high precision
  - Initializing these as a high-precision link, and asserting that "*israelis*" is referencing the `[animacy=inanimate]` geopolitical entity will prevent merging with ***we*** later.

Multi-pass Sieve Approach (Raghunathan et al, 2010)



Multi-pass Sieve Approach (Raghunathan et al, 2010)

- Pre-processing (mention detection).
  - Gold mention boundaries given, parsed, NE tagged
- For each mention, each module can skip or pick best candidate antecedent
  - Antecedents ordered:
    - Same sentence: by Hobbs algorithm
    - Prev. sentence:
      - For Nominal: by right-to-left, breadth first: proximity/recency
      - For Pronoun: left-to-right: salience hierarchy
    - Within cluster: aggregate attributes, order mentions
    - Prune indefinite mentions: can't have antecedents

Multi-pass Sieve Approach (Raghunathan et al, 2010)

- Pass 1: Exact match (N): 96% Precision
- Pass 2: Precise constructs
  - Predicate nominative, (role) appositive; pronoun, acronym, demonym
- Pass 3: Strict head matching
  - Matches cluster head noun AND all non-stop cluster words AND modifiers AND non-i-within-I (embedded NP)
- Pass 4 & 5:
  - Variant of 3, but iteratively drop one constraint

Multi-pass Sieve Approach (Raghunathan et al, 2010)

- Pass 6: Relaxed head match
  - Head matches any word in cluster AND all non-stop cluster words AND non-i-within-I
- Pass 7: Pronouns
  - Enforce constraints on gender, number, person, animacy, and NER labels

Coreference Evaluation

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)
    - Problem: Link-based — ignores singletons; penalizes large clusters

Coreference Evaluation

- Mention-Pair Based Model: Ng and Cardie (2002)
  - F_1: **70.4** (MUC-6) — **63.4** (MUC-7)
- Multi-pass Sieve results:

	MUC-6		
Passes	**P**	**R**	**F_1**
{1}	95.9	31.8	47.8
{1,2}	95.4	43.7	59.9
{1,2,3}	92.1	51.3	65.9
{1,2,3,4}	91.7	51.9	66.3
{1,2,3,4,5}	91.1	52.6	66.7
{1,2,3,4,5,6}	89.5	53.6	67.1
{1,2,3,4,5,6,7}	83.7	74.1	**78.6**

Questions

- Good 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

Reference Resolution Algorithms

- Other alternative strategies:
  - Linguistically informed, saliency hierarchy
    - Centering theory
  - Machine learning approaches:
    - Supervised: Classification
    - Unsupervised: Clustering
  - Heuristic, high precision
    - CogNIAC

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
    - Non-RNN, essentially just local classification w/some distributional semantics

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!)

Discourse Structure

Why Model Discourse Structure?

Theoretical Concerns

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

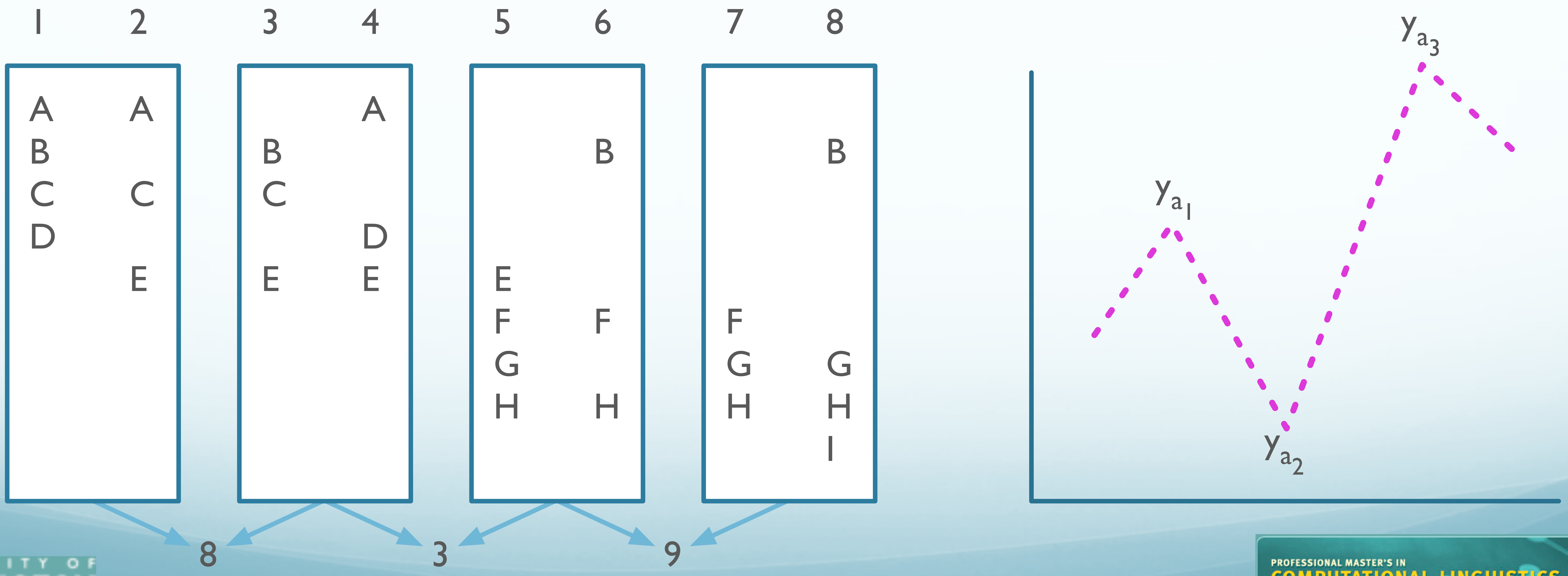
Lexical Cohesion Score

- 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

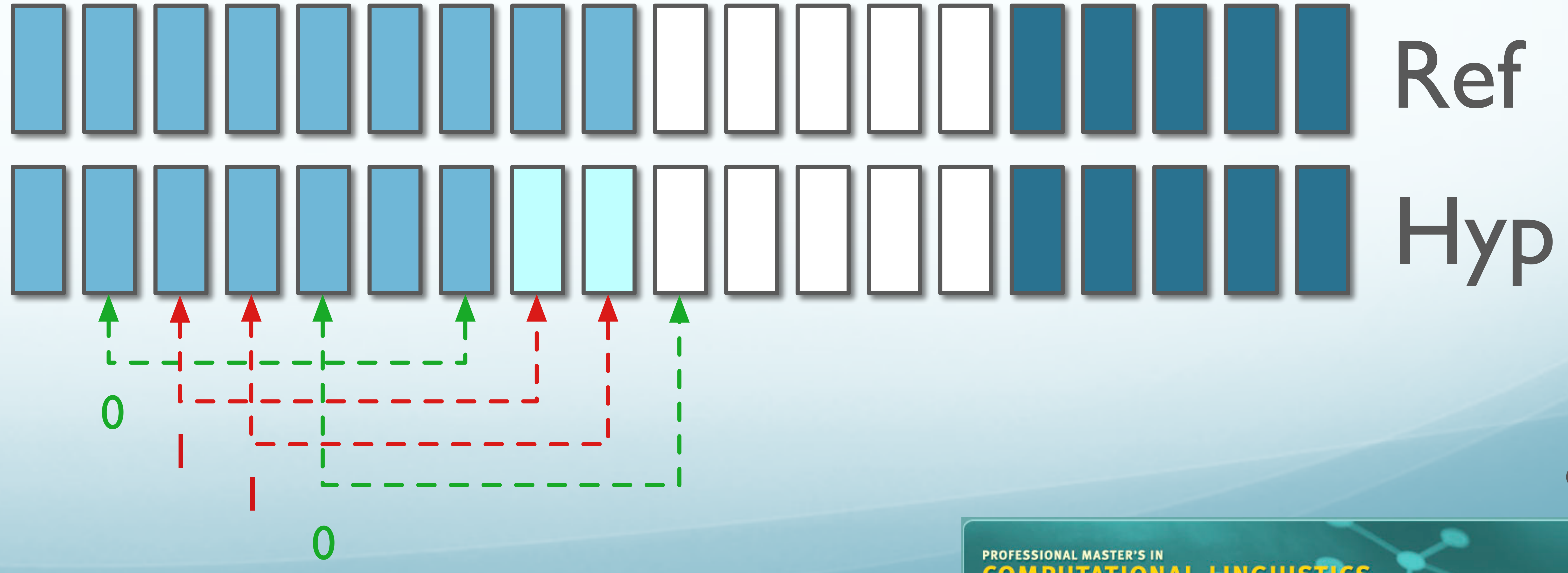
- Depth Score:
  - Difference between position and adjacent peaks
  - e.g. $\left(y_{a_1} - y_{a_2}\right) + \left(y_{a_3} - y_{a_2}\right)$



Evaluation?

- Accuracy?
  - <5% interword positions will be boundary.
- Precision/Recall/F-Measure?
  - No credit for near-misses
- WindowDiff
  - Pevzner & Hearst, 2002

$$WindowDiff(ref,hyp) = \frac{1}{N-k}\sum_{i=1}^{N-k}\left(\left|b\left(ref_i, ref_{i+k}\right) - b\left(hyp_i, hyp_{i+k}\right)\right| \neq 0\right)$$

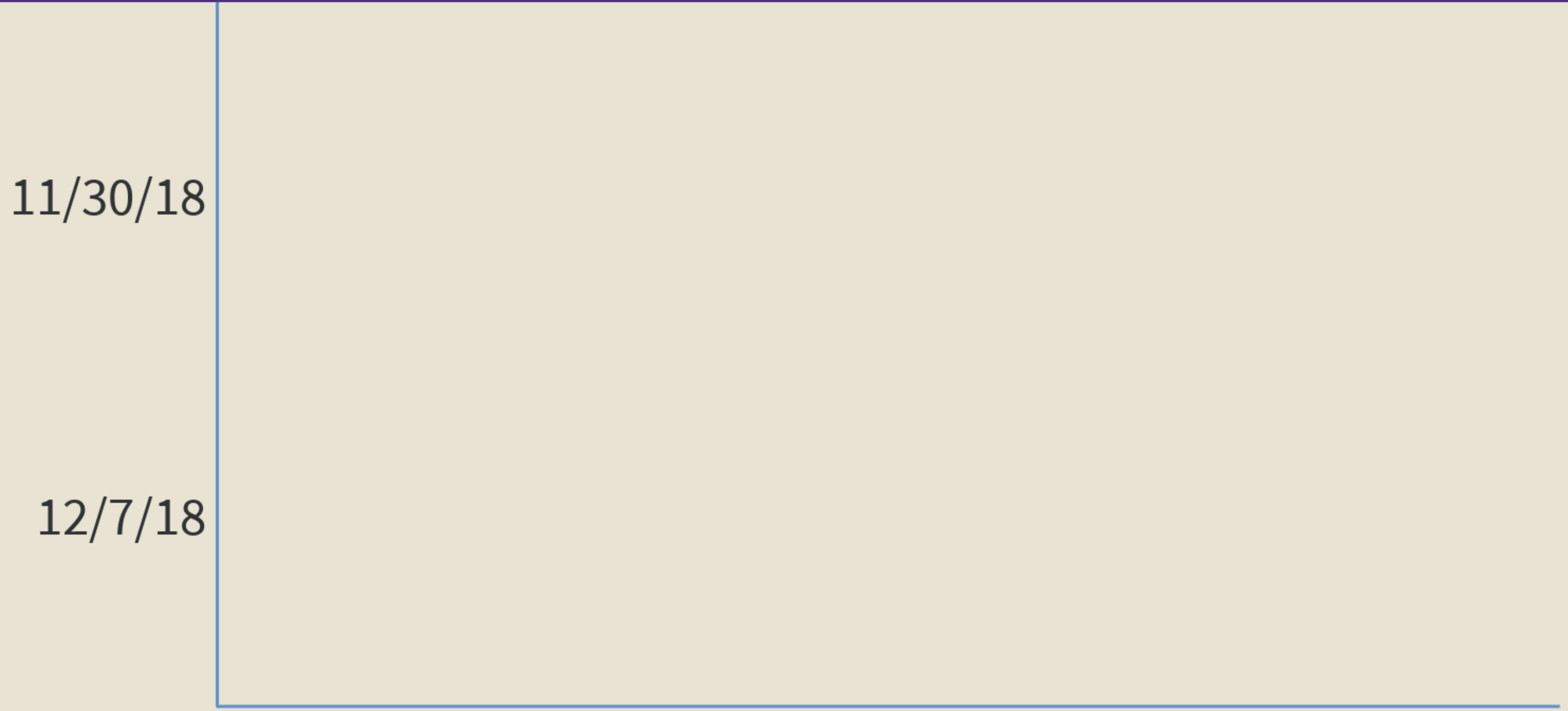


For Next Time

- Coherence:
  - Shallow and deep discourse Parsing
- Case study of shallow and deep NLP: Q&A

HW #9

Due date for HW #9?



Start the presentation to see live content. Still no live content? Install the app or get help at **PollEv.com/app**

Total Results

Goals

- Explore the task of pronomial anaphora resolution
- Gain familiarity with syntax-based resolution techniques
- Analyze the effectiveness of the Hobbs algorithm by applying it to pairs of parsed sentences.

Task

- Given pairs of sentences (S_0, S_1) as context
  - Resolve the pronoun(s) in S_1 using the Hobbs algorithm.
    - J&M p. 704-705
- **Subtasks**:
  - **Parsing Sentences** — Automatic (CKY, Earley, etc)
  - **Hobbs Algorithm** — May be done either:
    - **Manually** — manually mark up the output parse tree
    - **Coded** — implement Hobbs algorithm — will require feature grammar or similar for finding agreement, etc.

Notes

- For implementation
  - May use any NLTK tools for parse tree manipulation
    - …*as long as it doesn't directly implement the Hobbs algorithm!*
  - May create lookup table/dictionary for agreement
- Two results files:
  - One for all parsed output
  - One for remaining manual steps
    - (Based on a copy of the first)

NLTK Tools

- “Climbing” parse trees:
  - NLTK ParentedTree
    - nltk.org/howto/tree.html
  - Conversion from standard tree **t**
    - `parented_tree = nltk.tree.ParentedTree.convert(t)`

- Accessing feature structures

```
fs = nltk.grammar.FeatStructNonterminal(parented_tree.label())
pronoun_agr = fs['agr']
antecedent_agr.subsumes(pronoun_agr)
```