Speech and Language Processing

Discourse: Anaphora Resolution

Dan Jurafsky

Thanks to Diane Litman, Andy Kehler, Jim Martin!!! This material is from J+M Chapter 18, written by Andy Kehler, slides inspired by Diane Litman + Jim Martin

Reference Resolution

- Two examples:
  - John went to Bill's car dealership to check out an Acura Integra. He looked at it for half an hour
  - I'd like to get from Boston to San Francisco, on either December 5th or December 6th. It's ok if it stops in another city along they way
- What is the target of "it"?
  - First example: two possible targets
  - Second example: where is the target?

Why reference resolution?

- Conversational Agents: Airline reservation system needs to know what "it" refers to in order to book correct flight
- Information Extraction: First Union Corp. is continuing to wrestle with severe problems unleashed by a botched merger and a troubled business strategy. According to industry insiders at Paine Webber, their president, John R. Georgius, is planning to retire by the end of the year.

Some terminology

- John went to Bill's car dealership to check out an Acura Integra. He looked at it for half an hour
- Reference: process by which speakers use words John and he to denote a particular person
  - Referring expression: John, he
  - Referent: the actual entity (but as a shorthand we might call "John" the referent).
  - John and he "corefer"
  - Antecedent: John
  - Anaphor: he
- Cataphora: pronoun before the referent
  - Before he bought it, John checked over the Integra very carefully

Many types of reference

- (after Webber, '91)
- According to John, Bob bought Sue an Integra, and Sue bought Fred a Legend
  - But that turned out to be a lie (a speech act)
  - But that was false (proposition)
  - That struck me as a funny way to describe the situation (manner of description)
  - That caused Sue to become rather poor (event)
- But we focus on references to entities

Reference Phenomena

- Indefinite noun phrases: new to hearer
  - I saw an Acura Integra today
  - Some Acura Integras were being unloaded…
- Definite noun phrases: identifiable to hearer because
  - Mentioned: I saw an Acura Integra today. The Integra was white
  - Identifiable from beliefs (common knowledge): The Indianapolis 500
  - Inherently unique: The fastest car in Indianapolis 500…

Reference Phenomena: Pronouns

- **Simple example:**
  - I saw an Acura Integra today. It was white
- **Referent salience, in case of discourse:**
  1. John went to Bob's party, and parked next to a beautiful Acura Integra
  2. He went inside and talked to Bob for more than an hour.
  3. Bob told him that he recently got engaged.
  4. a) He also said that he bought it yesterday.
     b) He also said that he bought the Acura yesterday.

"it"… what?

Reference phenomena

- **Demostratives**
  - I bought an Integra yesterday. It's similar to the one I bought five years ago. That one was really nice, but I like this one even better
- **A non-pronominal anaphora**
  - I saw no less that 6 Acura Integra today. I want one
    - … one (of them)

Reference phenomena

- **Inferrable**
  - I almost bought an Acura Integra today, but the engine seemed noisy.
- **Pronominal**
  - I saw no less than 6 Acura Integras today. They are the coolest cars.

Pronominal Reference Resolution

- Given a pronoun, find the reference
- Constraints to leverage
  - Hard constraints on reference
  - Soft constraints on reference
- Algorithms which use/don't use these constraints

Hard constraints: syntax

- Number agreement
  - *John has an Acura. They??? are red
  - John has an Acura. It is red
- Person and case agreement
  - *John and Mary have Acuras. We love them??? (who/what???)
  - John and I have Acuras. We love them.
- Gender agreement
  - John has an Acura. He / it is attractive.
- Syntactic constraints
  - John bought himself a new Acura (himself == John)
  - John bought him a new Acura (him =/= John)

Soft constraints: Pronoun Interpretation Preferences

- **Selectional Restrictions**
  - John parked his Acura in the garage. He had driven it around for hours.
- **Recency**
  - John has an Integra. Bill has a Legend. Mary likes to drive it.

Soft constraints: Pronoun Interpretation Preferences

- **Grammatical Role: Subject preference**
  - **John went to the Acura dealership with Bill. He bought an Integra.**
  - **John and Bill went to the Acura dealership. He bought an Integra**

Soft constraints: Repeated Mention preference

- John needed a car to get to his new job. He decided that he wanted something sporty. Bill went to the Acura dealership with him. He bought an Integra.

Soft constraints: Parallelism Preference

- **Same structure**
  - **Mary went with Sue to the Acura dealership.**
    Sally went with her to the Mazda dealership.
- **But… with similar structure**
  - **Mary went with Sue to the Acura dealership.**
    Sally told her not to buy anything.

Soft constraints: Verb Semantics Preferences

- John telephoned Bill. He lost the pamphlet on Acuras.
- John criticized Bill. He lost the pamphlet on Acuras.
- Implicit causality
  - Implicit cause of criticizing is object.
  - Implicit cause of telephoning is subject.

Algorithms for pronoun anaphora resolution

- **Knowledge-rich approach**
  - **Syntactic-based: Hobbs' algorithm**
  - **Discourse-based: Centering Theory**
  - **Hybrid approaches: Lappin and Leas**
  - **Corpus-based: Charniak, Hale, and Ge**
- **Knowledge-poor approach**
  - **Machine Learning**

Hobbs

- Hobbs (1978) proposes an algorithm that searches parse trees for antecedents of a pronoun.
- starting at the NP node *immediately dominating the pronoun*
- in a *specified search order*
- looking for the *first match of the correct gender and number*
- Idea: discourse and other preferences will be approximated by search order.

Hobbs

- Extending Hobbs:
- Add "simple" selectional restrictions, e.g. (for the verb "to move"):
  - dates can't move
  - places can't move
  - large fixed objects can't move
- For "they", in addition to accepting plural NPs, collects selectionally compatible entities (somehow), e.g., conjoined NPs.
- Assume some process that recovers elided constituents and inserts them in the tree.

Example:

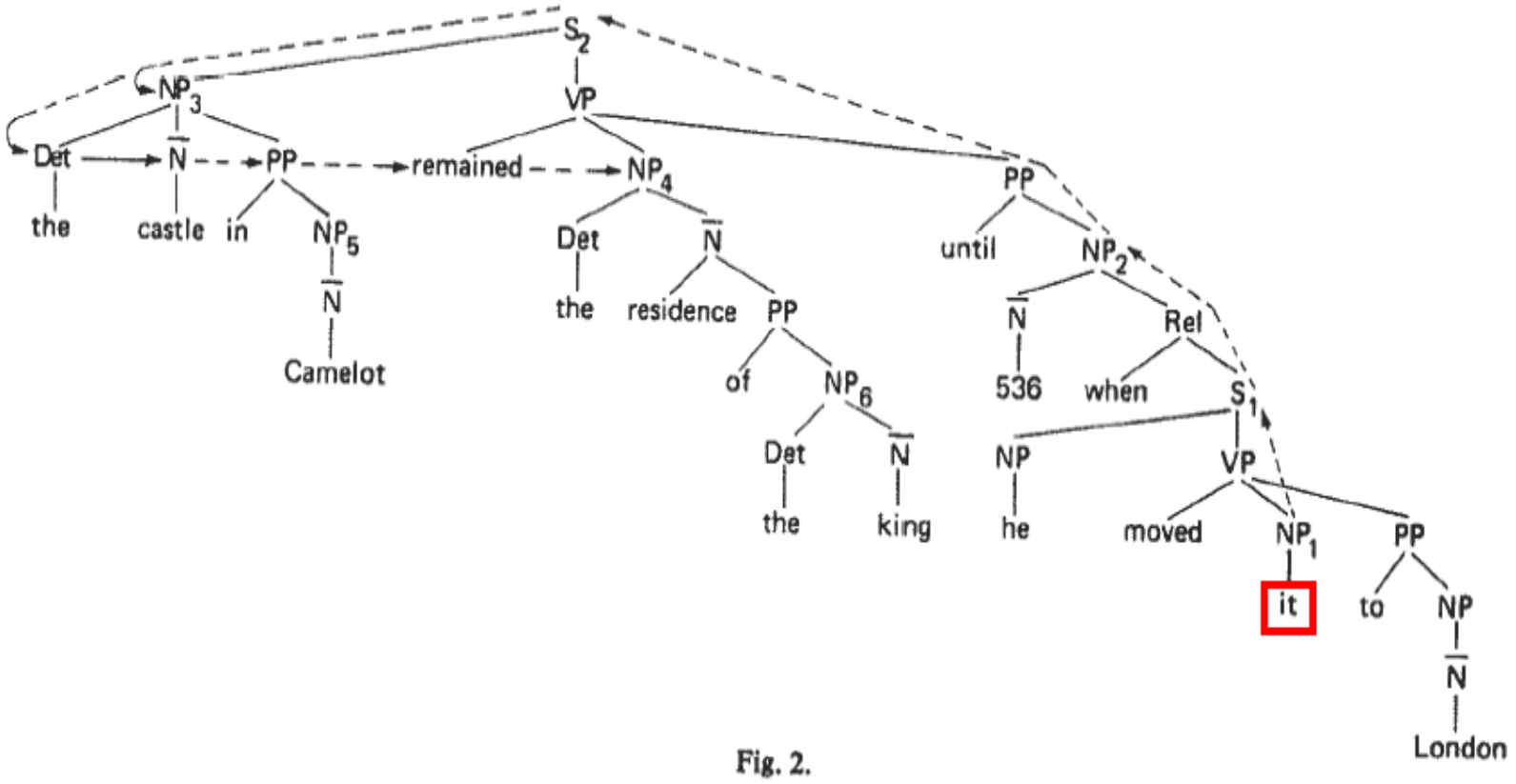

Fig. 2.

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

Hobbs' Algorithm

1. **Begin at the NP immediately dominating the pronoun.**
2. **Go up tree to first NP or S encountered.**
   a. **Call node X, and path to it, p.**
   b. **Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.**
3. **If X is highest S node in sentence,**
   – **Search previous trees, in order of recency, left-to-right, breadth-first, proposing NPs encountered.**
4. **Otherwise, from X, go up to first NP or S node encountered,**
   – **Call this X, and path to it p.**
5. **If X is an NP, and p does not pass through an N-bar that X immediately dominates, propose X.**
6. **Search below X, to left of p, left-to-right, breadth-first, proposing NP encountered.**
7. **If X is an S, search below X to right of p, left-to-right, breadth-first, but not going through any NP or S, proposing NP encountered.**
8. **Go to 2.**

Example:

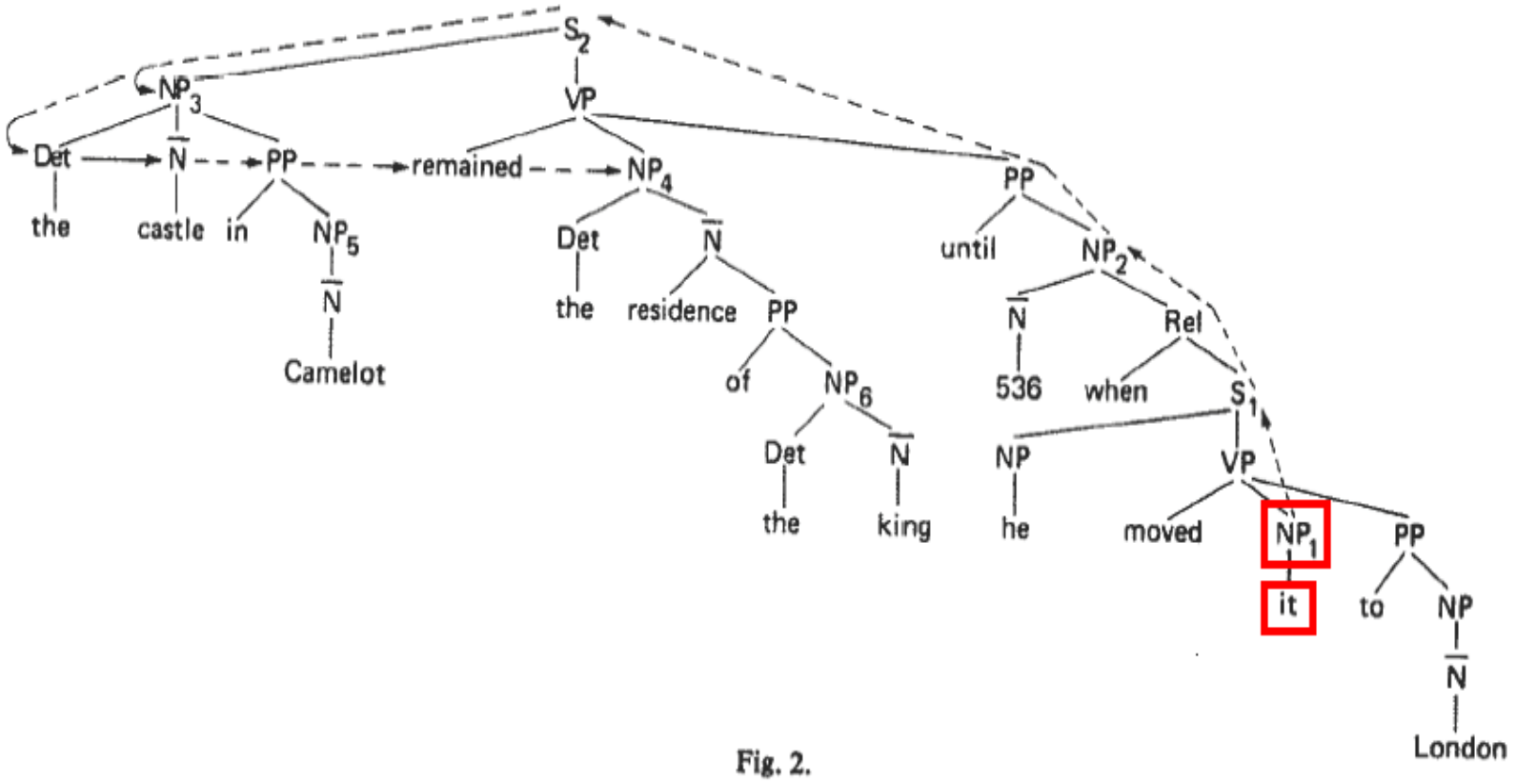

Fig. 2.

1. Begin at the NP immediately dominating the pronoun.

Hobbs' Algorithm

1. **Begin at the NP immediately dominating the pronoun.**
2. **Go up tree to first NP or S encountered.**
   a. **Call node X, and path to it, p.**
   b. **Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.**
3. **If X is highest S node in sentence,**
   – **Search previous trees, in order of recency, left-to-right, breadth-first, proposing NPs encountered.**
4. **Otherwise, from X, go up to first NP or S node encountered,**
   – **Call this X, and path to it p.**
5. **If X is an NP, and p does not pass through an N-bar that X immediately dominates, propose X.**
6. **Search below X, to left of p, left-to-right, breadth-first, proposing NP encountered.**
7. **If X is an S, search below X to right of p, left-to-right, breadth-first, but not going through any NP or S, proposing NP encountered.**
8. **Go to 2.**

Example:

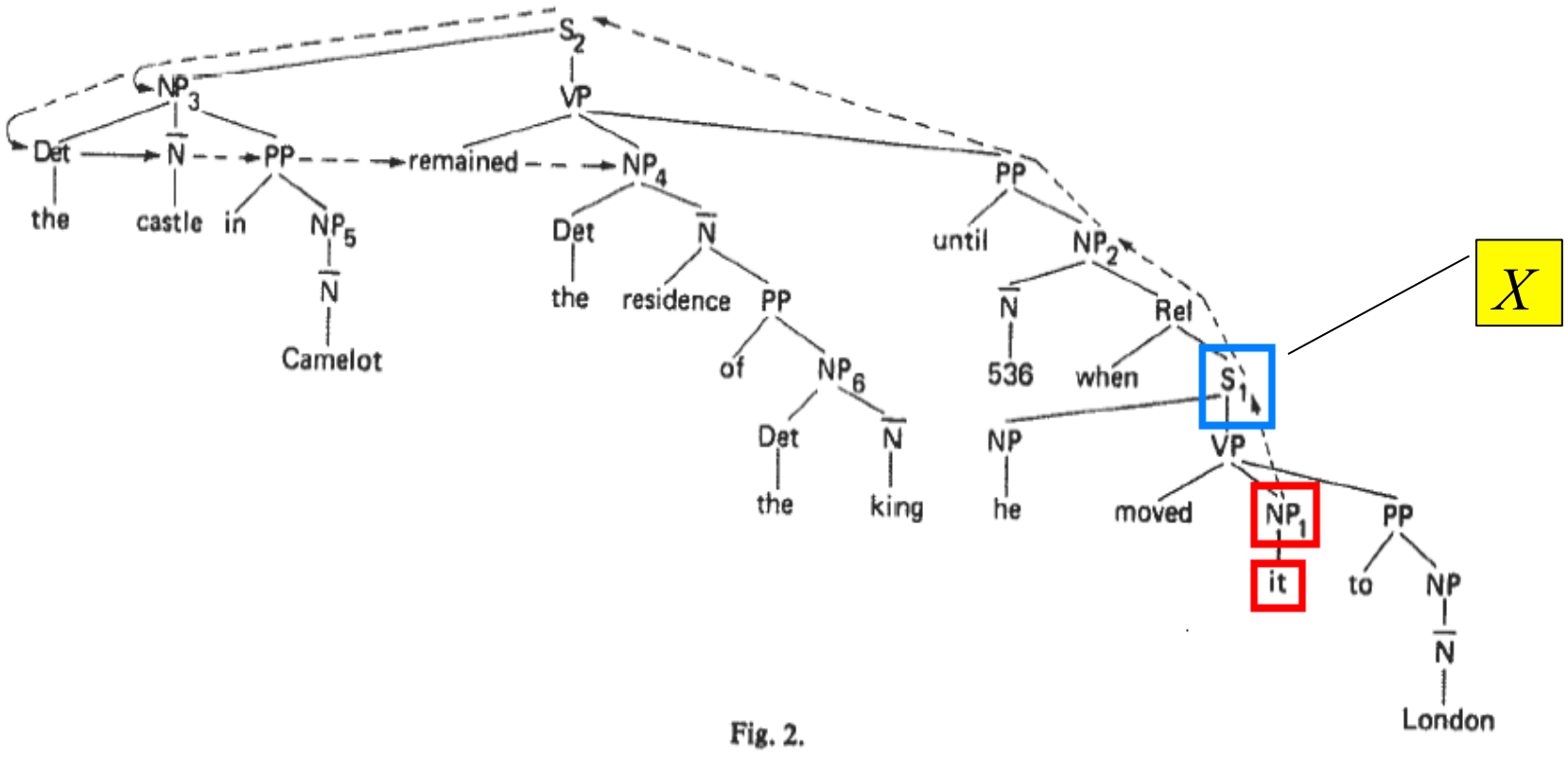

Fig. 2.

2. Go up tree to first NP or S encountered
a. Call node X, and path to it, p.

Example:

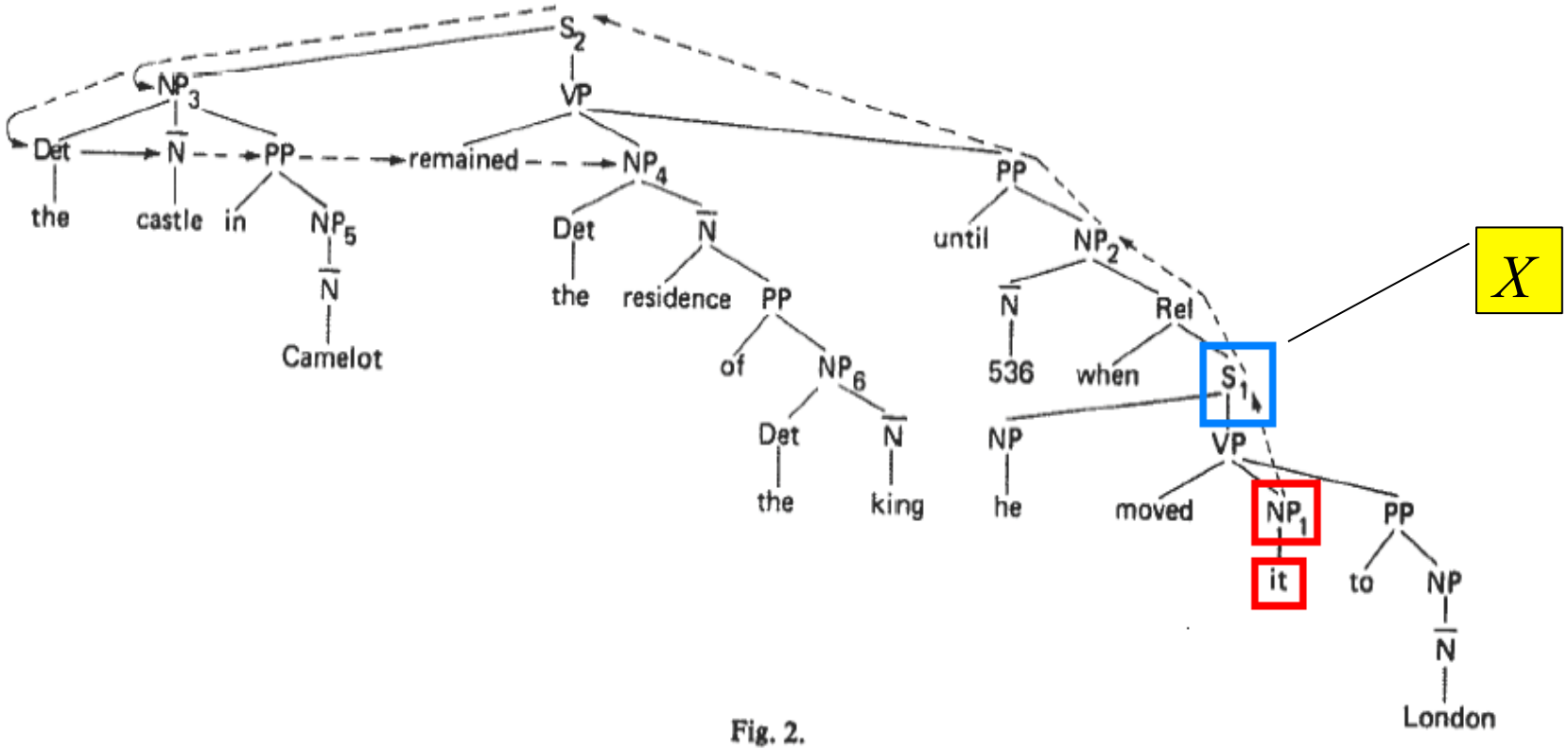

Fig. 2.

b. Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X

→ search yields no candidate. Go to next step of the algorithm.

Hobbs' Algorithm

1. **Begin at the NP immediately dominating the pronoun.**
2. **Go up tree to first NP or S encountered.**
   a. **Call node X, and path to it, p.**
   b. **Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.**
3. **If X is highest S node in sentence,**
   – **Search previous trees, in order of recency, left-to-right, breadth-first, proposing NPs encountered.**
4. **Otherwise, from X, go up to first NP or S node encountered,**
   – **Call this X, and path to it p.**
5. **If X is an NP, and p does not pass through an N-bar that X immediately dominates, propose X.**
6. **Search below X, to left of p, left-to-right, breadth-first, proposing NP encountered.**
7. **If X is an S, search below X to right of p, left-to-right, breadth-first, but not going through any NP or S, proposing NP encountered.**
8. **Go to 2.**

Example:

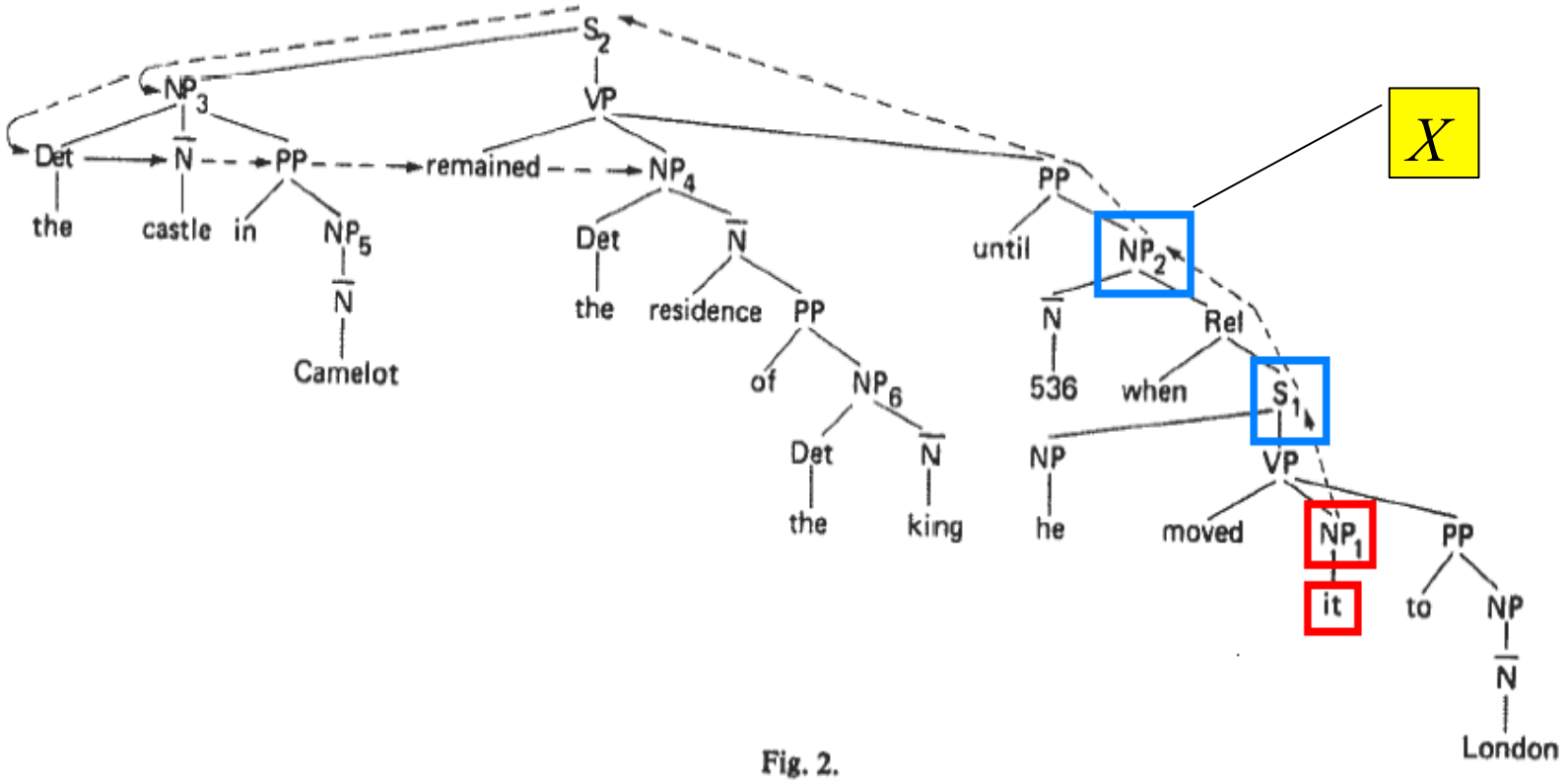

Fig. 2.

3. false

4. Otherwise, from X go up to first NP or S node encountered.

a. Call this X, and path to it p.

Hobbs' Algorithm

1. **Begin at the NP immediately dominating the pronoun.**
2. **Go up tree to first NP or S encountered.**
   a. **Call node X, and path to it, p.**
   b. **Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.**
3. **If X is highest S node in sentence,**
   – **Search previous trees, in order of recency, left-to-right, breadth-first, proposing NPs encountered.**
4. **Otherwise, from X, go up to first NP or S node encountered,**
   – **Call this X, and path to it p.**
5. **If X is an NP, and p does not pass through an N-bar that X immediately dominates, propose X.**
6. **Search below X, to left of p, left-to-right, breadth-first, proposing NP encountered.**
7. **If X is an S, search below X to right of p, left-to-right, breadth-first, but not going through any NP or S, proposing NP encountered.**
8. **Go to 2.**

Example:

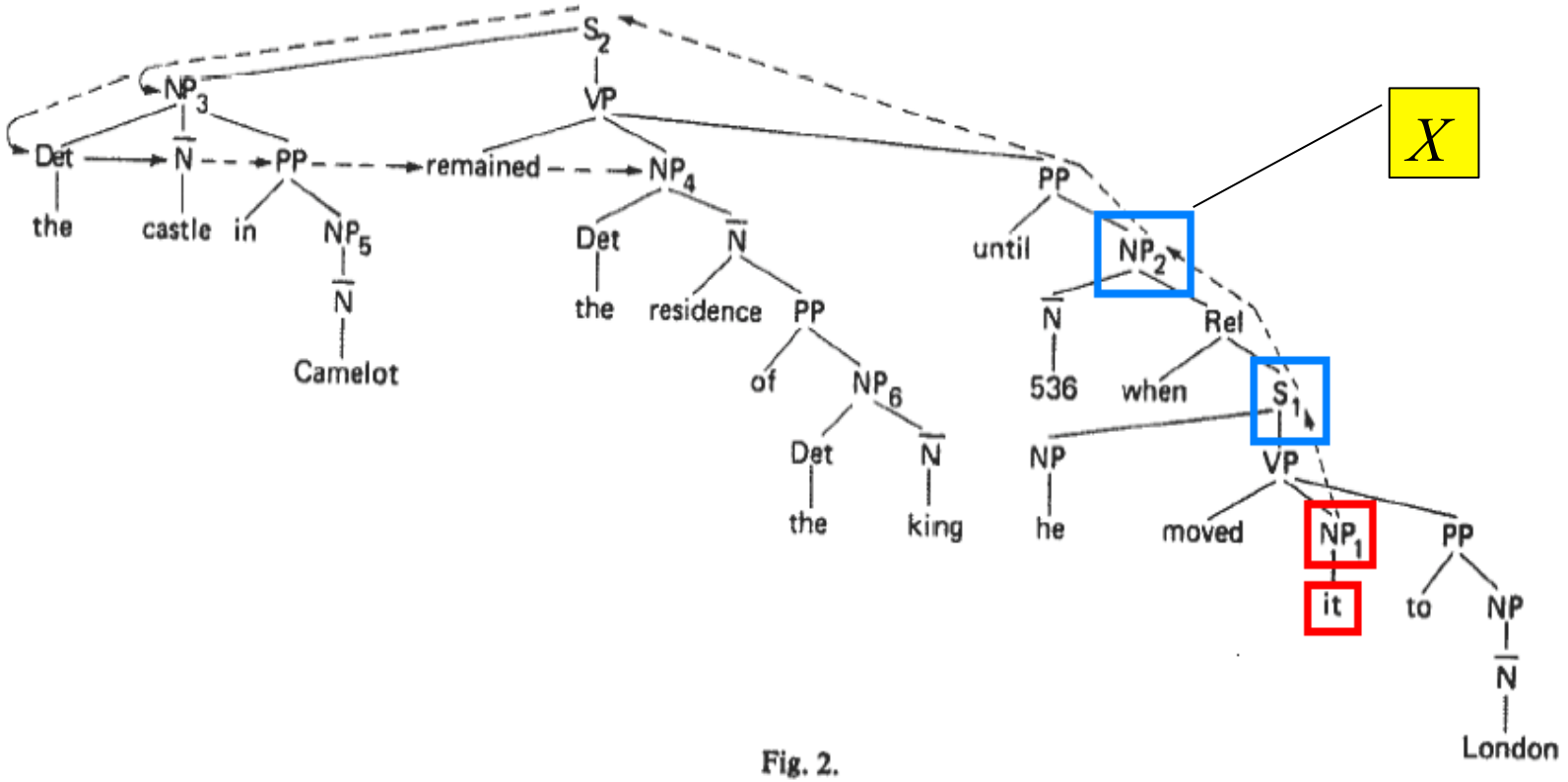

Fig. 2.

5. If X is an NP, and p does not pass through an N-bar that X immediately dominates, propose X

→ NP_2 is proposed. Rejected by selectional restrictions (dates can't move).

Hobbs' Algorithm

1. **Begin at the NP immediately dominating the pronoun.**
2. **Go up tree to first NP or S encountered.**
   a. **Call node X, and path to it, p.**
   b. **Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.**
3. **If X is highest S node in sentence,**
   – **Search previous trees, in order of recency, left-to-right, breadth-first, proposing NPs encountered.**
4. **Otherwise, from X, go up to first NP or S node encountered,**
   – **Call this X, and path to it p.**
5. **If X is an NP, and p does not pass through an N-bar that X immediately dominates, propose X.**
6. **Search below X, to left of p, left-to-right, breadth-first, proposing NP encountered.**
7. **If X is an S, search below X to right of p, left-to-right, breadth-first, but not going through any NP or S, proposing NP encountered.**
8. **Go to 2.**

Example:

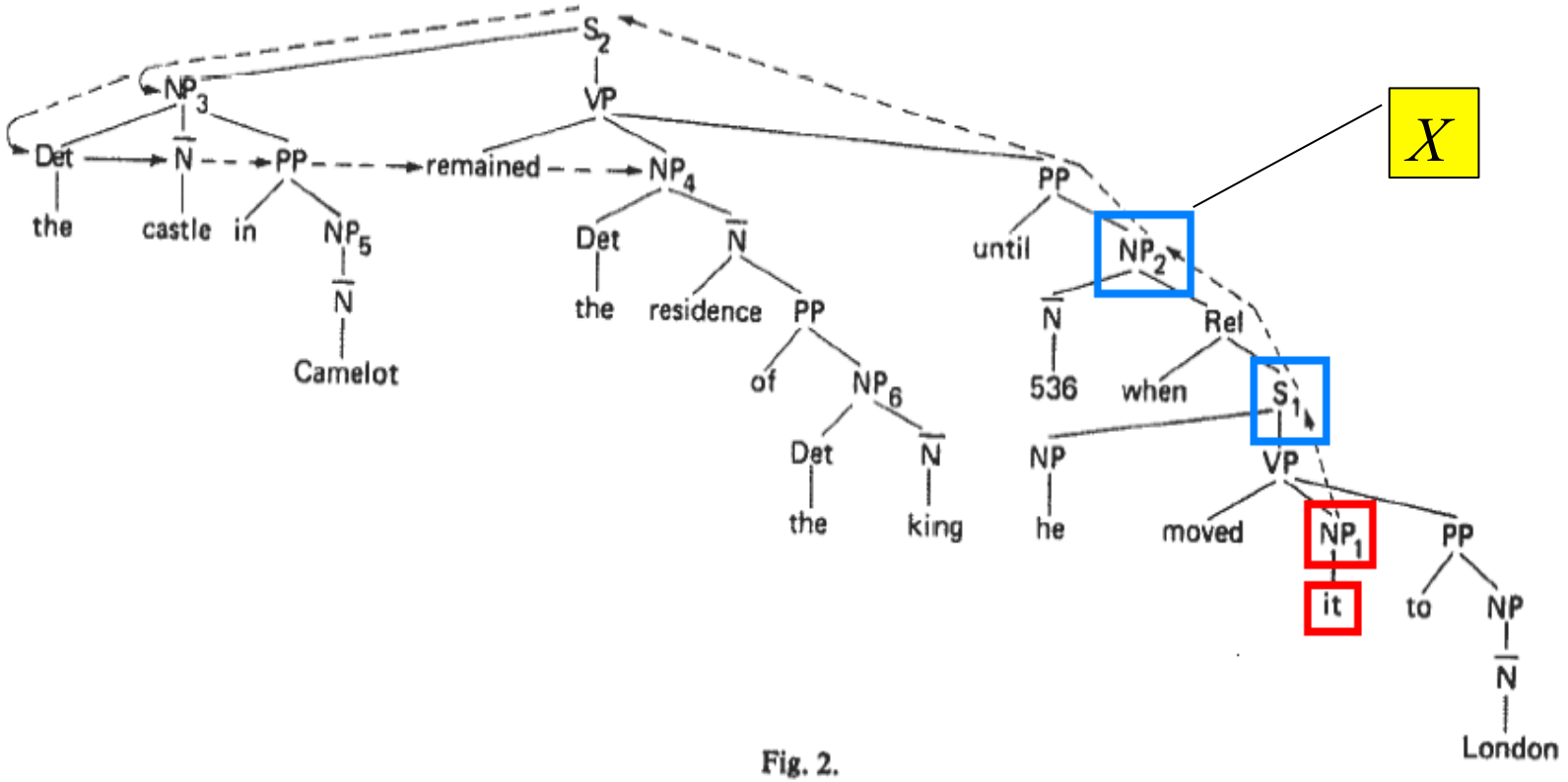

Fig. 2.

6. Search below X, to left of p, left-to-right, breadth-first, proposing NP encountered.

→ search yields no candidate. Go to next step of the algorithm.

Hobbs' Algorithm

1. **Begin at the NP immediately dominating the pronoun.**
2. **Go up tree to first NP or S encountered.**
   - **Call node X, and path to it, p.**
   - **Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.**
3. **If X is highest S node in sentence,**
   - **Search previous trees, in order of recency, left-to-right, breadth-first, proposing NPs encountered.**
4. **Otherwise, from X, go up to first NP or S node encountered,**
   - **Call this X, and path to it p.**
5. **If X is an NP, and p does not pass through an N-bar that X immediately dominates, propose X.**
6. **Search below X, to left of p, left-to-right, breadth-first, proposing NP encountered.**
7. **If X is an S, search below X to right of p, left-to-right, breadth-first, but not going through any NP or S, proposing NP encountered.**
8. **Go to 2.**

Hobbs' Algorithm

1. **Begin at the NP immediately dominating the pronoun.**
2. **Go up tree to first NP or S encountered.**
   - **Call node X, and path to it, p.**
   - **Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.**
3. **If X is highest S node in sentence,**
   - **Search previous trees, in order of recency, left-to-right, breadth-first, proposing NPs encountered.**
4. **Otherwise, from X, go up to first NP or S node encountered,**
   - **Call this X, and path to it p.**
5. **If X is an NP, and p does not pass through an N-bar that X immediately dominates, propose X.**
6. **Search below X, to left of p, left-to-right, breadth-first, proposing NP encountered.**
7. **If X is an S, search below X to right of p, left-to-right, breadth-first, but not going through any NP or S, proposing NP encountered.**
8. **Go to 2.**

Example:

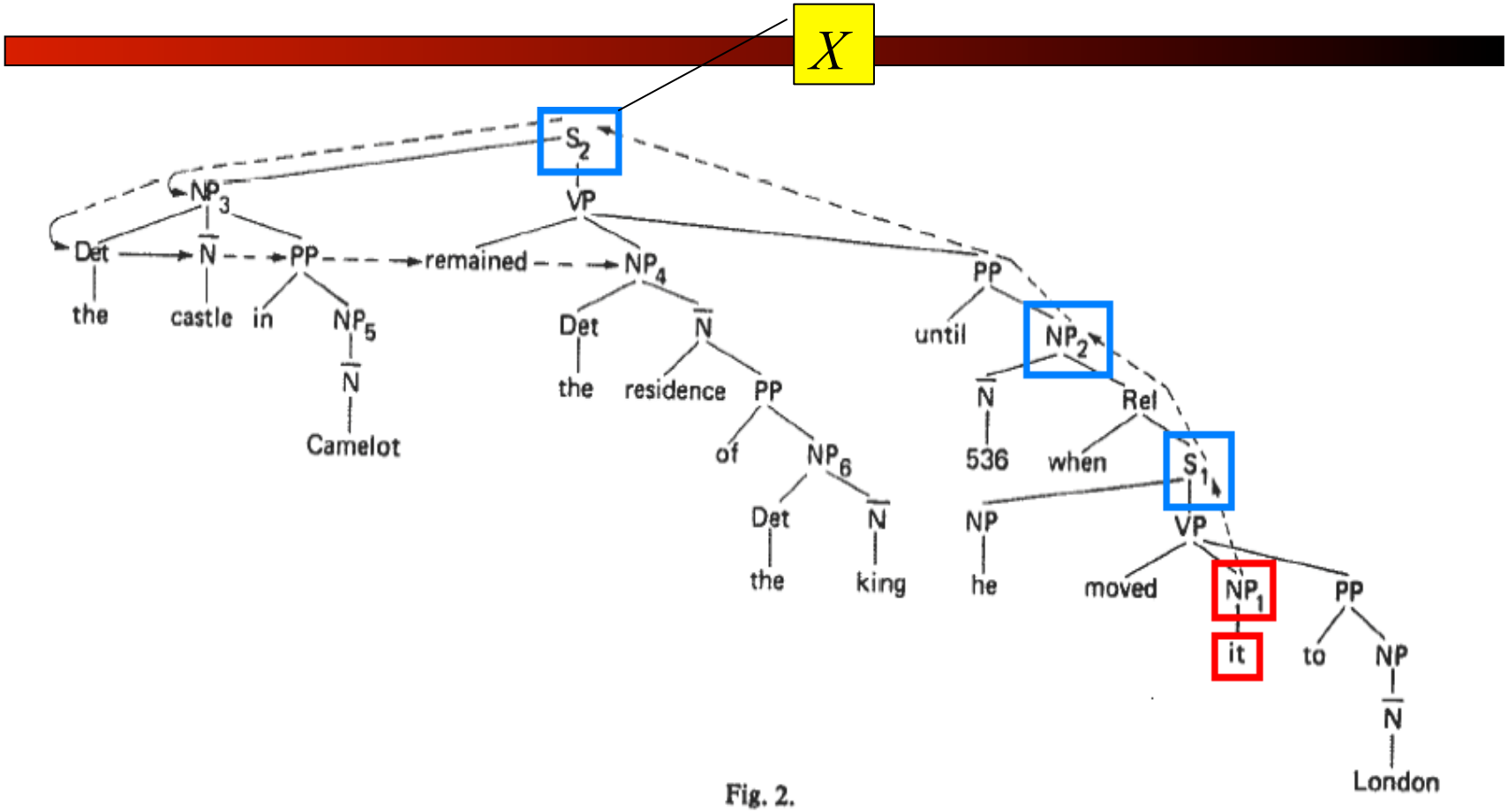

Fig. 2.

7. false

2. Go up tree to first NP or S encountered

a. Call node X, and path to it, p

Example:

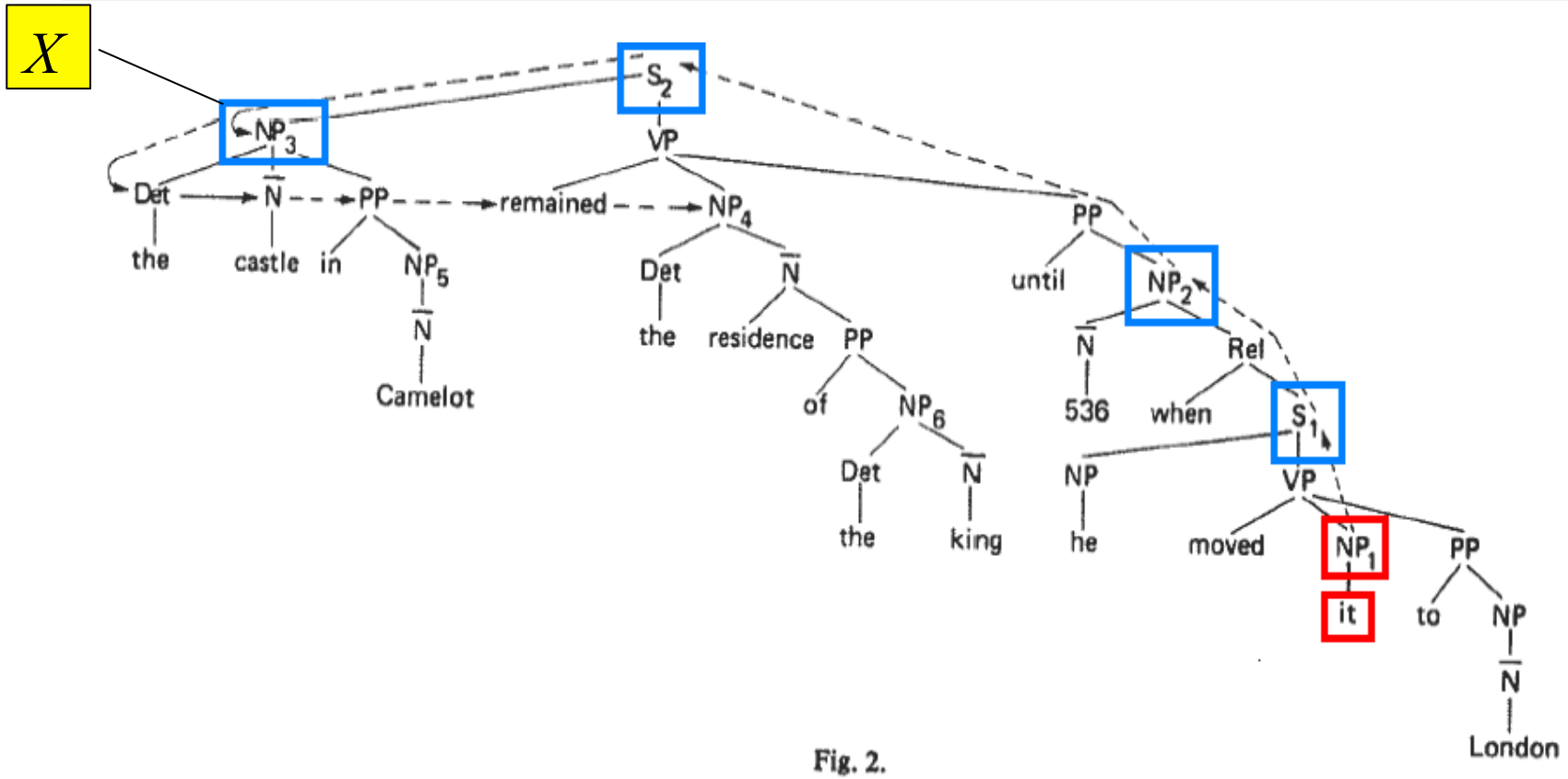

Fig. 2.

b. Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.

NP_3 proposed → Rejected by rejected by selectional restrictions (can't move large fixed objects.)

Example:

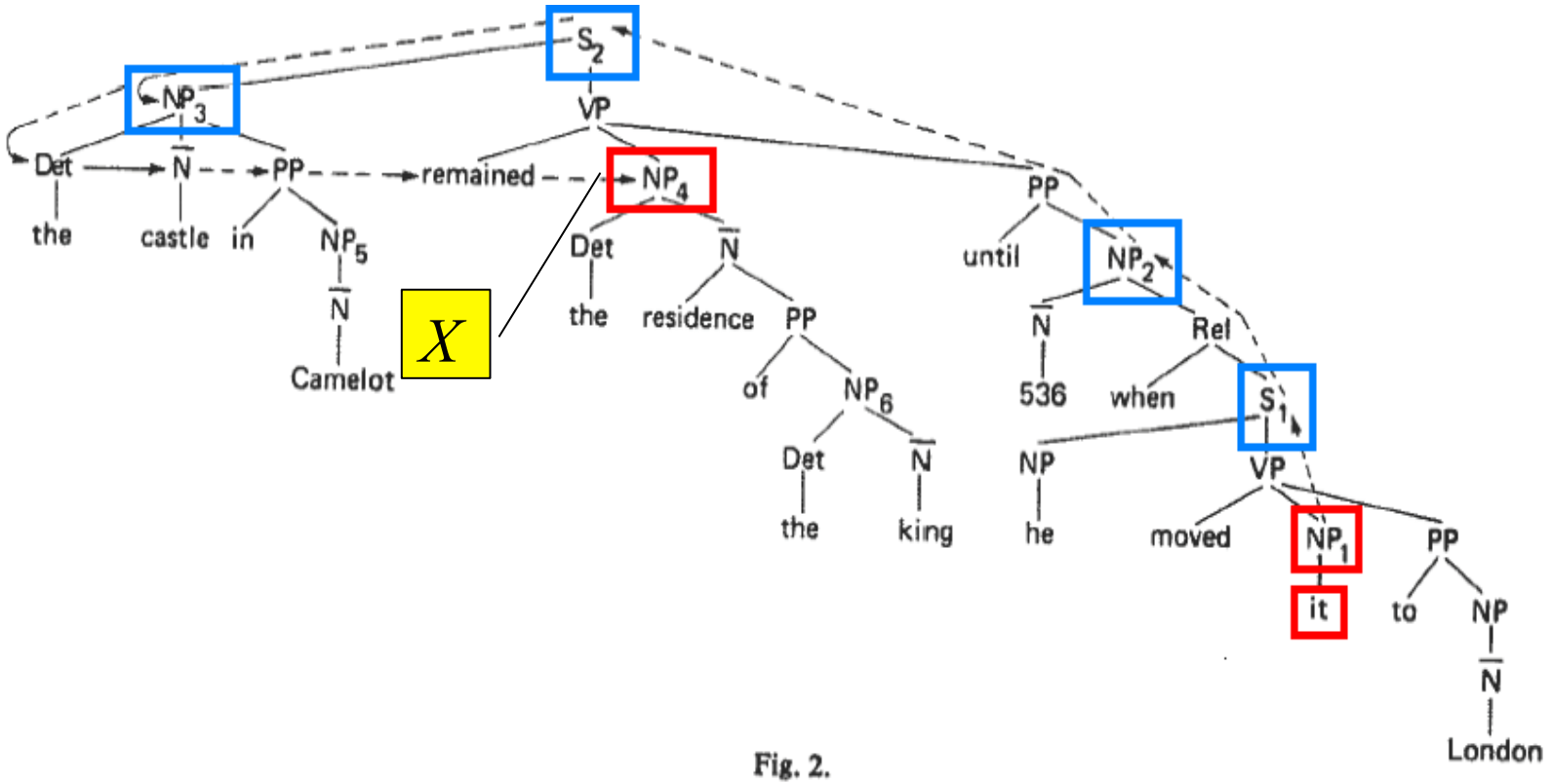

Fig. 2.

b. Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.

→ NP_4 proposed. Accepted.

Lappin and Leass

- Lappin and Leass (1994): Given he/she/it, assign antecedent.
- Implements only recency and syntactic preferences
- Two steps
  - Discourse model update
    - When a new noun phrase is encountered, add a representation to discourse model with a salience value
    - Modify saliences.
  - Pronoun resolution
    - Choose the most salient antecedent

Salience Factors and Weights

- Salience given to NP

1 Recency	100
2 Subject emphasis	80
3 Existential emphasis	70
4 Accusative (direct object) emphasis	50
5 Ind. Obj and oblique emphasis	40
6 Non-adverbial emphasis	50
7 Head noun emphasis	80

Information leveraged

- NP *saliences*
  - according to recency of the NP (1)
  - according to syntactical role (2-7)
  - cut in half after each sentence is processed

Syntactic roles

- Salience factors 2-6: Syntactic role preference
  - Subject > existential predicate nominal > object > indirect object > demarcated adverbial PP
- Examples for 2-5
  - An Acura Integra is parked in the lot (subject)
  - There is an Acura Integra parked in the lot (ex. pred nominal)
  - John parked an Acura Integra in the lot (object)
  - John gave his Acura Integra a bath (indirect obj)
  - In his Acura Integra, John showed Susan his new CD player (demarcated adverbial PP)
- Add salience if 6 holds (not part of demarcated adverbial PP):
  - Inside his Acura Integra, John showed Susan his new CD player (here, it is part of demarcated adverbial PP → no salience)
- Add salience if 7 holds (part of head noun):
  - The owner's manual for an Acura Integra is on John's desk

NP

head of NP

Lappin and Leass Algorithm

- Collect the potential referents (up to 4 sentences back)
- Remove potential referents that do not agree in number or gender with the pronoun
- Remove potential references that do not pass syntactic coreference constraints
- Compute total salience value of referent from all factors, including, if applicable:
  - role parallelism (+35)
  - or cataphora (-175).
- Select referent with highest salience value. In case of tie, select closest.

Example

- **John saw a beautiful Acura Integra at the dealership. He showed it to Bob. He bought it.**

Sentence 1:

Referent	1 recency	2 Subj	3 Exist	4 Obj	5 Ind-obj	6 Non-adv	7 Head N	Total
John	100	80				50	80	310
Integra	100			50		50	80	280
dealership	100					50	80	230

After sentence 1

- **Cut all values in half**

Referent	Phrases	Value
John	{John}	155
Integra	{a beautiful Acura Integra}	140
dealership	{the dealership}	115

Sentence 2:
He showed it to Bob

- He specifies male gender
- So Step 2 reduces set of referents to only John.
- Now update discourse model:
  - He in current sentence (recency=100), subject position (=80), not adverbial (=50) not embedded (=80), so add 310:

Referent	Phrases	Value
John	{John, he_1}	155+310
Integra	{a beautiful Acura Integra}	140
dealership	{the dealership}	115

Sentence 2:
He showed it to Bob

- Need to add "it", which can be "Integra" or "dealership".
- Need to add weights:
  - Parallelism: "it" and "Integra" are objects ("dealership" is not), so +35 for "Integra"
  - Integra: 175, dealership: 115
    → pick Integra
- Update discourse model:
  - "it" is object, gets 100+50+50+80=280:

Sentence 2:
He showed it to Bob

Referent	Phrases	Value
John	{John, he_1}	465
Integra	{a beautiful Acura Integra, it_1}	140+280
dealership	{the dealership}	115

Sentence 2:
He showed it to Bob

- **Bob is a new referent**
- **Update disciurse model:**
  - **Bob is oblique argument, weight is 100+40+50+80=270**

Referent	Phrases	Value
John	{John, he_1}	465
Integra	{a beautiful Acura Integra, it_1}	420
Bob	{Bob}	270
dealership	{the dealership}	115

Sentence 3:
He bought it

- **Drop weights in half:**

Referent	Phrases	Value
John	{John, he_1}	232.5
Integra	{a beautiful Acura Integra, it_1}	210
Bob	{Bob}	135
dealership	{the dealership}	57.5

He_2 will be resolved to John, and it_2 to Integra

Evaluation

- Referential Rate (Byron, 2001)
- RR = C / (T+E)

C: # pronouns correctly resolved
T: all pronouns in the test set
E: all excluded referential pronouns

REFERENCES

References

- **Techniques for Anaphora Resolution: A Survey**
  www.cs.cornell.edu/courses/cs674/2005sp/projects/tejaswini-deoskar.doc
- **Anaphora Resolution**
  https://people.ucsc.edu/~abrsvn/anaphora%20resolution.ppt