UPPSALA
UNIVERSITET

Introduction

Syntactic analysis (5LN455)

2012-11-13

Sara Stymne
Department of Linguistics and Philology

Mostly based on slides from Marco Kuhlmann

UPPSALA
UNIVERSITET

Today

- Introduction to syntactic analysis
- Course information

UPPSALA
UNIVERSITET

What is syntax?

- Syntax addresses the question how sentences are constructed in particular languages.
- The English word *syntax* comes from the Ancient Greek word *sýntaxis* ‘arrangement’.

UPPSALA UNIVERSITET

What is syntax not?

Syntax does not answer questions about …

… how speech is articulated and perceived (phonetics, phonology)

… how words are formed (morphology)

… how utterances are interpreted in context (semantics, pragmatics)

simplified

UPPSALA UNIVERSITET

Why should you care about syntax?

- Syntax describes the distinction between well-formed and ill-formed sentences.
- Syntactic structure can serve as the basis for semantic interpretation and can be used for
  - Machine translation
  - Information extraction and retrieval
  - Question answering
  - ...

UPPSALA
UNIVERSITET

Parsing

The automatic analysis of a sentence
with respect to its syntactic structure.

UPPSALA
UNIVERSITET

Theoretical frameworks

- Generative syntax
  Noam Chomsky (1928–)
- Categorial syntax
  Kazimierz Ajdukiewicz (1890–1963)
- Dependency syntax
  Lucien Tesnière (1893–1954)

UPPSALA
UNIVERSITET

Theoretical frameworks

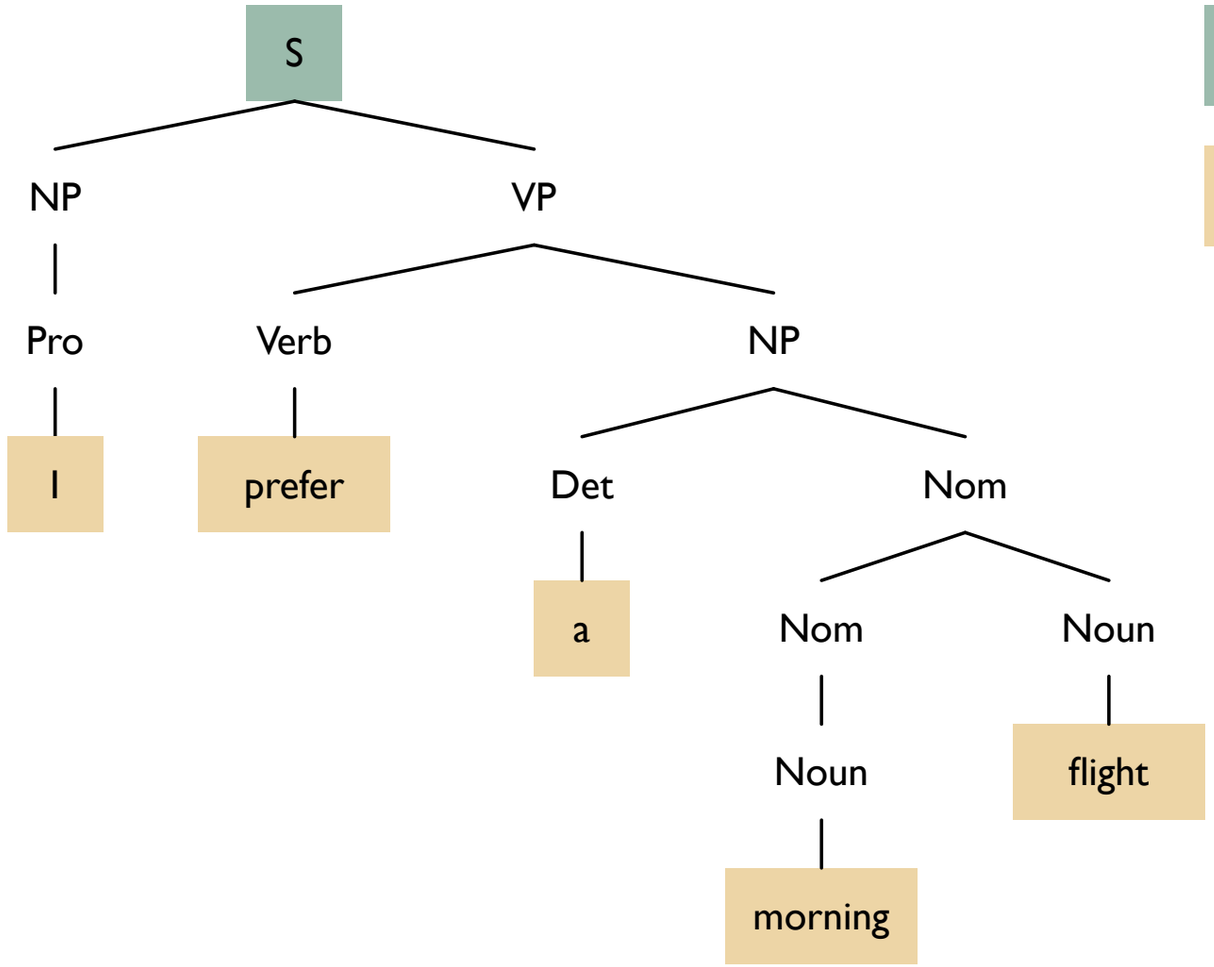
Chomsky

Ajdukiewicz

Tesnière

UPPSALA UNIVERSITET

Phrase structure trees

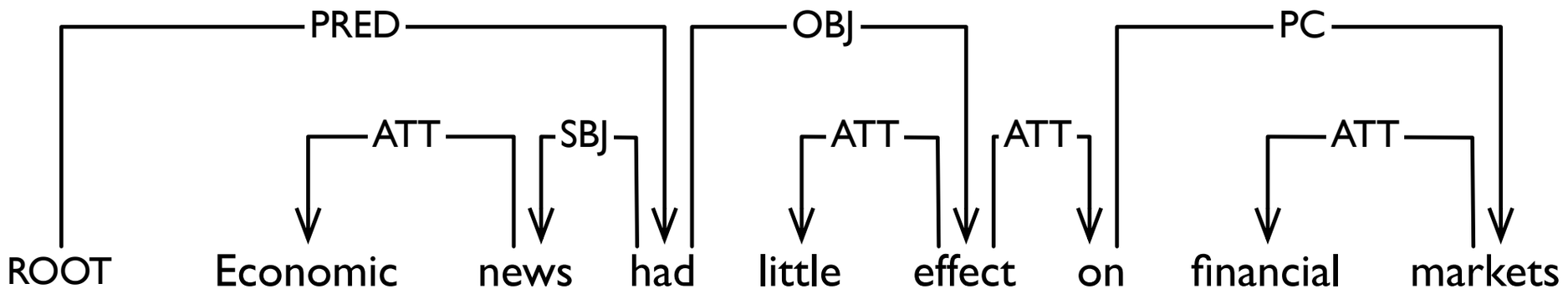


root (top)

leaves (bottom)

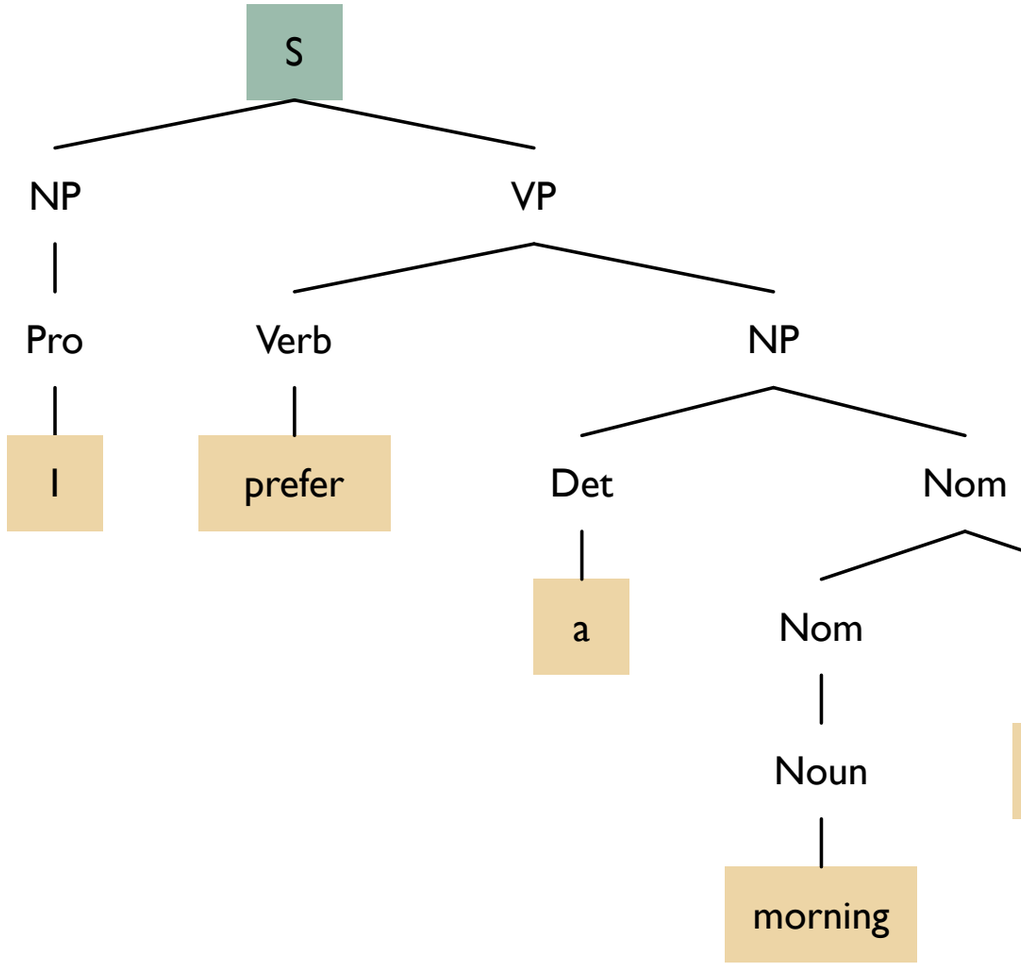
UPPSALA
UNIVERSITET

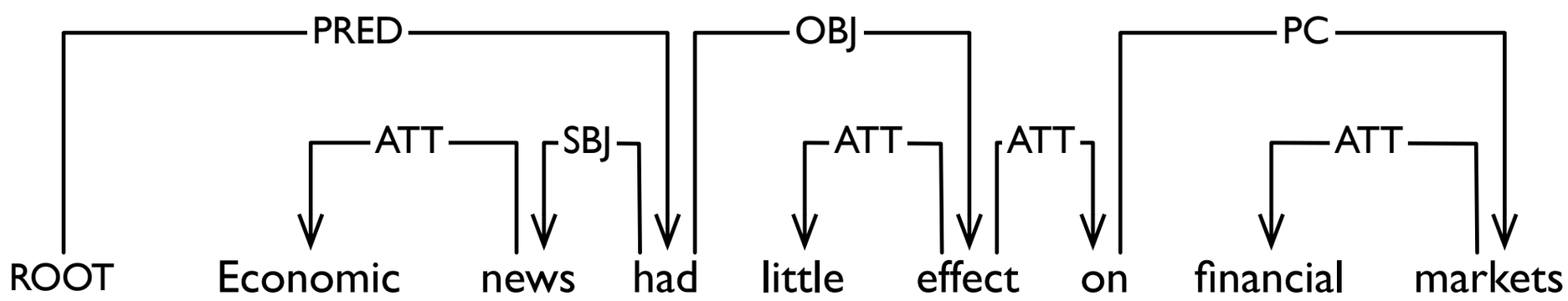
Dependency trees



UPPSALA UNIVERSITET

Phrase structure vs dependency trees





UPPSALA UNIVERSITET

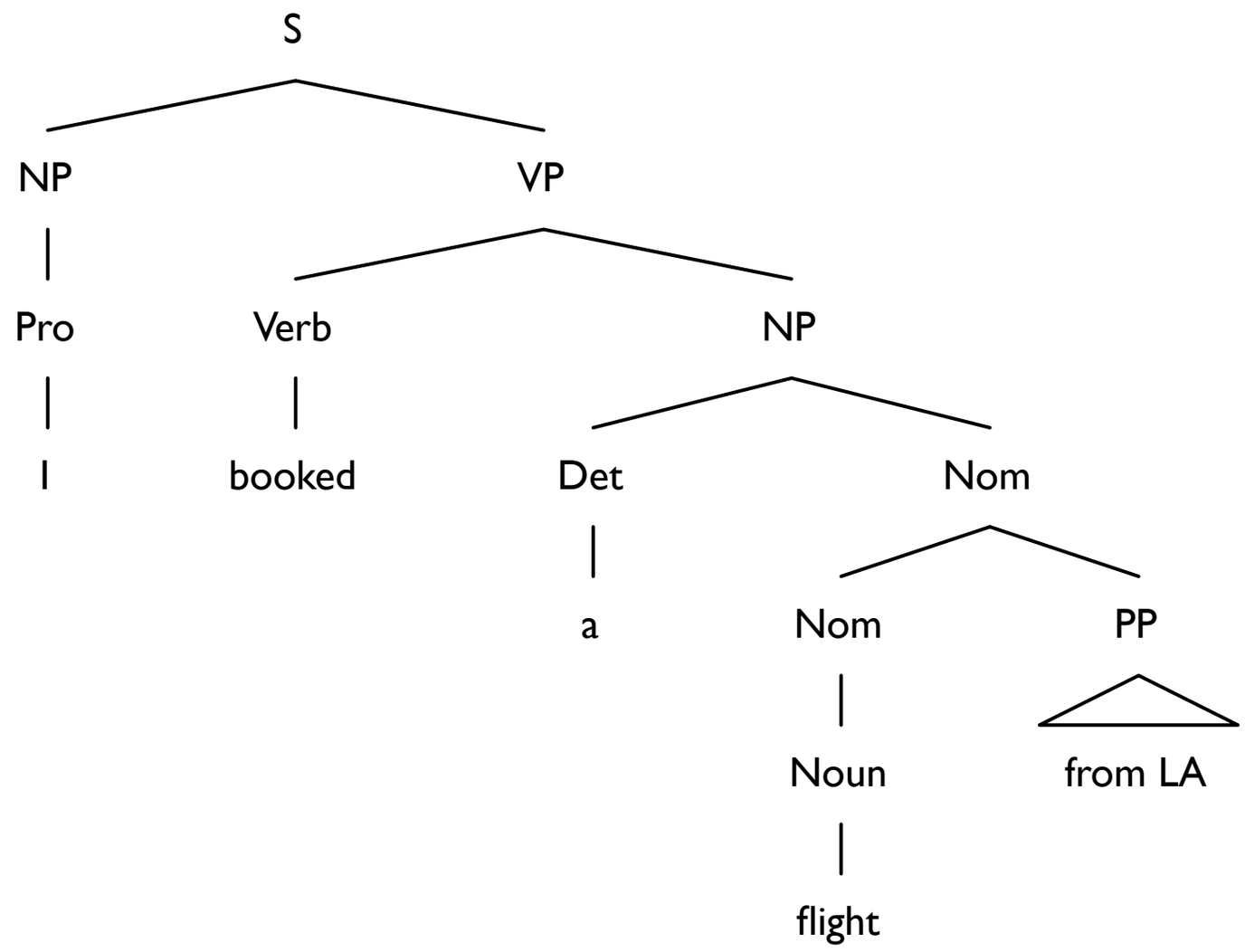
Ambiguity

I booked a flight from LA.

- This sentence is ambiguous. In what way?
- What should happen if we parse the sentence?

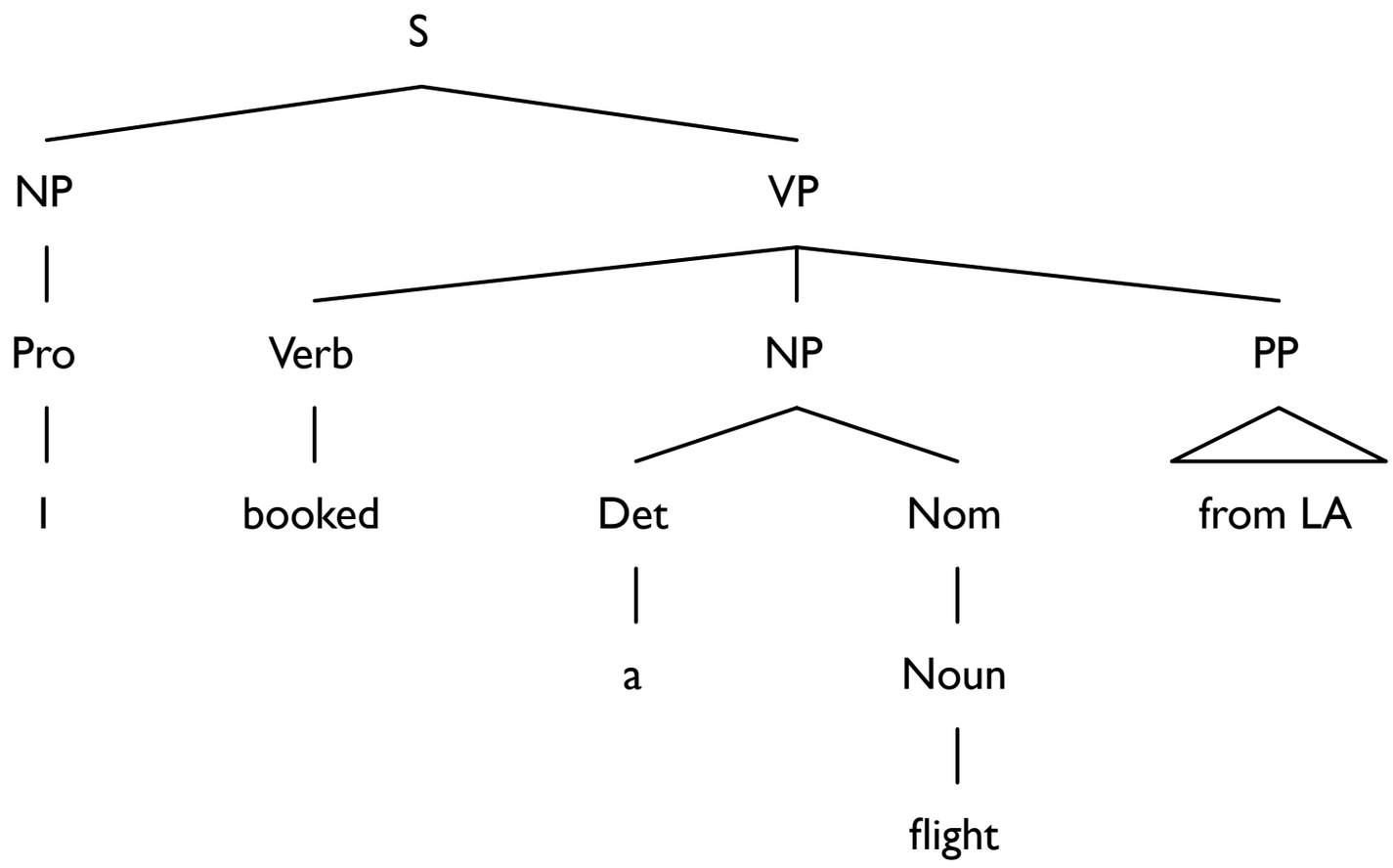
UPPSALA UNIVERSITET

Ambiguity



UPPSALA UNIVERSITET

Ambiguity



UPPSALA
UNIVERSITET

Interesting questions

- Is there any parse tree at all?
- What is the best parse tree?

UPPSALA UNIVERSITET

Parsing as search

- Parsing as search:
  Search through all possible parse trees for a given sentence.
- In order to search through all parse trees we have to ‘build’ them.

UPPSALA
UNIVERSITET

Top–down and bottom–up

top–down

only build trees that are rooted at *S*

may produce trees that do not match the input

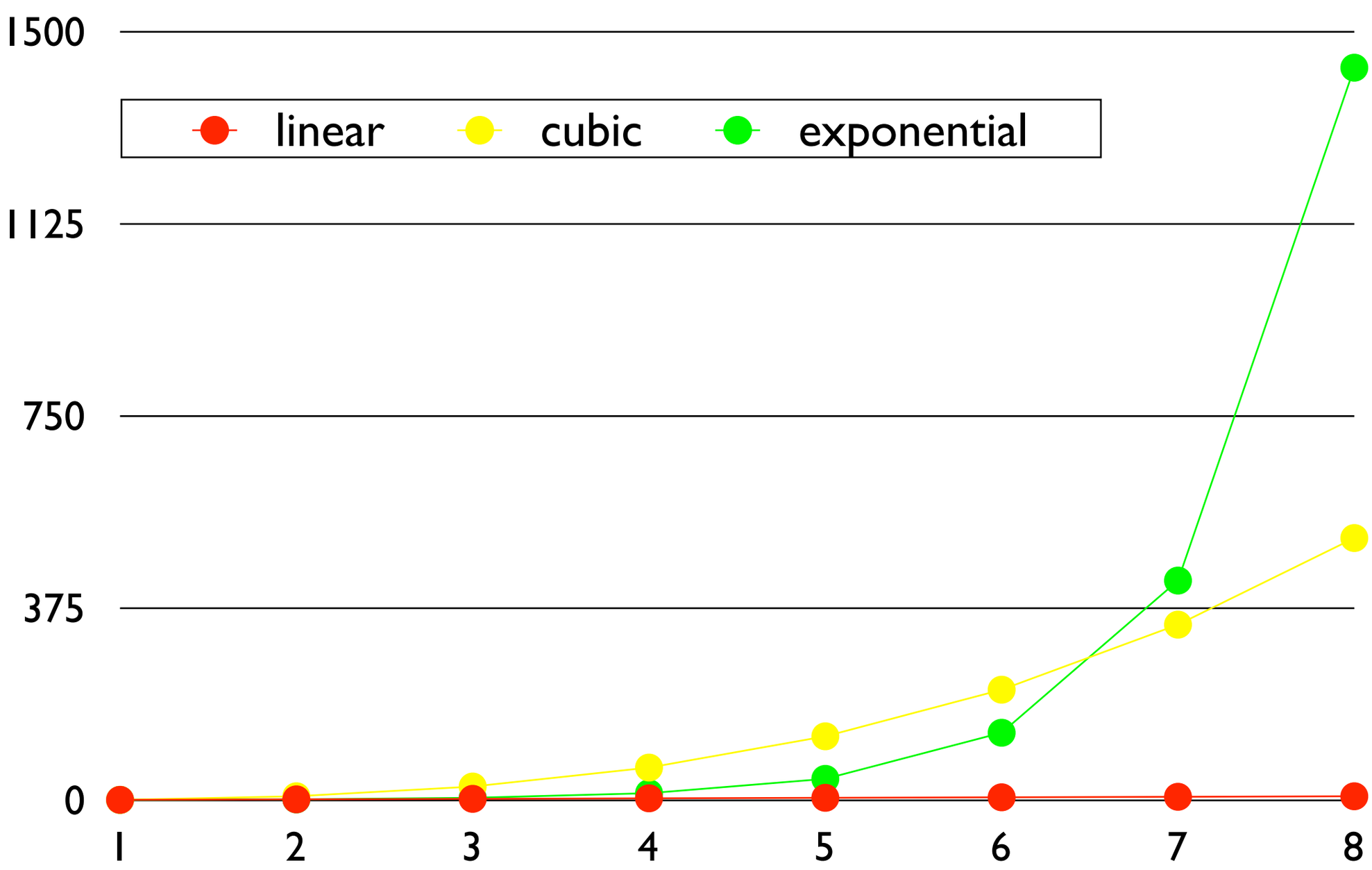
bottom–up

only build trees that match the input

may produce trees that are not rooted at *S*

UPPSALA UNIVERSITET

How many trees are there?



UPPSALA UNIVERSITET

Dynamic programming (DP)

- Divide and conquer:
  In order to solve a problem, split it into subproblems, solve each subproblem, and combine the solutions.

- Dynamic programming:
  Solve each subproblem only once and save the solution in order to use it as a partial solution in a larger subproblem.

- Memoisation:
  Solve only the necessary subproblems and store their solutions for resue in solving other subproblems.

UPPSALA UNIVERSITET

Complexity

- Using DP we can (sometimes) search through all parsetrees in polynomial time.
- That is much better than to spend exponential time!
- But it may still be too expensive!
  In these cases one can use an approximative method such as greedy search or beam search.

UPPSALA
UNIVERSITET

Course information

UPPSALA UNIVERSITET

Intended learning outcomes

At the end of the course, you should be able to

- account for the parsing problem of phrase structure grammar and dependency grammar;
- explain at least two different methods for automatic syntactic analysis: one for phrase structure parsing, one for dependency parsing;
- account for statistical methods for syntactic disambiguation;

UPPSALA UNIVERSITET

Intended learning outcomes

- apply existing systems that use these methods to realistic data and evaluate them with respect to their accuracy and efficiency;
- implement a central component of at least one approach to syntactic analysis in a suitable programming language.

UPPSALA
UNIVERSITET

Examination

- Examination is continuous and distributed over four graded assignments and one seminar.
- Two assignments are one-page papers.
  Time to invest: about 8 hours per assignment.
- The other two assignments are small projects where you need to implement/test parsers.
  Time to invest: about 40 hours per assignment.
- In the final seminar you will discuss some of your assignment results orally, and discuss issues of the other course participants

UPPSALA UNIVERSITET

Assignments

1. Written assignment on phrase structure parsing
2. Programming assignment: implement CKY parsing
3. Written assignment on dependency parsing
4. Use and evaluate an exisiting system for dependency parsing

UPPSALA UNIVERSITET

Learning outcomes and examination

- account for the parsing problem of phrase structure grammar and dependency grammar; paper assignments
- explain at least two different methods for automatic syntactic analysis: one for phrase structure parsing, one for dependency parsing; paper assignments
- account for statistical methods for syntactic disambiguation; paper assignments

UPPSALA UNIVERSITET

Learning outcomes and examination

- apply existing systems that use these methods to realistic data and evaluate them with respect to their accuracy and efficiency; project assignment 1
- implement a central component of at least one approach to syntactic analysis in a suitable programming language. project assignment 2

- All learning outcomes: final seminar

UPPSALA UNIVERSITET

Grading

- The assignments are graded with G and VG
- G on the seminar if present and active. The seminar is obligatory!

- To achieve G on the course:
  - G on all assignments and final seminar
- To achieve VG on the course:
  - Same as for G **and** VG on at least two assignments

UPPSALA
UNIVERSITET

Teachers

- Sara Stymne
  - Course coordinator, most lectures, assignments
- Joakim Nivre
  - Examiner, two lectures

UPPSALA UNIVERSITET

Course workload

- 7.5 hp means about 200 hours work
- 16 h lectures
- 2 h seminar
- 182 h work on your own
  - ~ 96 h assignment work
  - ~ 86 h reading
- No scheduled assignment supervision. Contact Sara when you need help!

UPPSALA UNIVERSITET

Schedule

week 46	(2 hrs taught, 14 hrs reading, 4 hrs assignments)	
13/11	13–15	Lecture: Introduction
week 47	**(4 hrs taught, 8 hrs reading, 8 hrs assignments)**	
18/11	10–12	Lecture: Phrase structure parsing 1
20/11	10–12	Lecture: Phrase structure parsing 2
week 48	**(2 hrs taught, 4 hrs reading, 14 hrs assignments)**	
27/11	13–15	Lecture: Phrase structure parsing 3
week 49	**(2 h taught, 4 hrs reading, 14 hrs assignments)**	
04/12	10–12	Lecture: Phrase structure parsing 4

UPPSALA UNIVERSITET

Schedule

week 50 (2 hrs taught, 8 hrs reading, 10 hrs assignments)		
11/12		Deadline: Assignments 1 and 2
11/12	13–15	Lecture: Dependency parsing 1 (JN)
week 51 (4 hrs taught, 12 hrs reading, 4 hrs assignments)		
16/12	10–12	Lecture: Dependency parsing 2 (JN)
18/12	10–12	Lecture: Dependency parsing 3
week 52-02 (0 hrs taught, 32 hrs reading, 28 hrs assignments)		
week 03 (2 hrs taught, 4 hrs reading, 14 hrs assignments)		
15/01	13–15	Final seminar
15/01		Deadline: Assignments 3 and 4

UPPSALA
UNIVERSITET

Reading

- Daniel Jurafsky and James H. Martin.
  Speech and Language Processing. 2nd edition.
  Pearson Education, 2009.
  Chapters 12-14.
- Sandra Kübler, Ryan McDonald,
  and Joakim Nivre. Dependency Parsing.
  Morgan and Claypool, 2009.
  Chapter 3-4.

UPPSALA UNIVERSITET

Evaluation from last year

- Overall score: 4.25
  - Overall students found the course interesting, and the assignments good
- This year
  - New teachers
  - Seminar at the end of the course
  - Otherwise not much change, since the course was working well