UPPSALA
UNIVERSITET

Introduction

Syntactic analysis (5LN455)

2016-11-08

Sara Stymne
Department of Linguistics and Philology

Mostly based on slides from Marco Kuhlmann

UPPSALA UNIVERSITET

Today

- Introduction to syntactic analysis
- Course information
- Exercises

UPPSALA
UNIVERSITET

What is syntax?

- Syntax addresses the question how sentences are constructed in particular languages.
- The English (and Swedish) 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)

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

- Generative syntax
  Noam Chomsky (1928–)

- Categorial syntax
  Kazimierz Ajdukiewicz (1890–1963)

- Dependency syntax
  Lucien Tesnière (1893–1954)

UPPSALA
UNIVERSITET

Theoretical frameworks

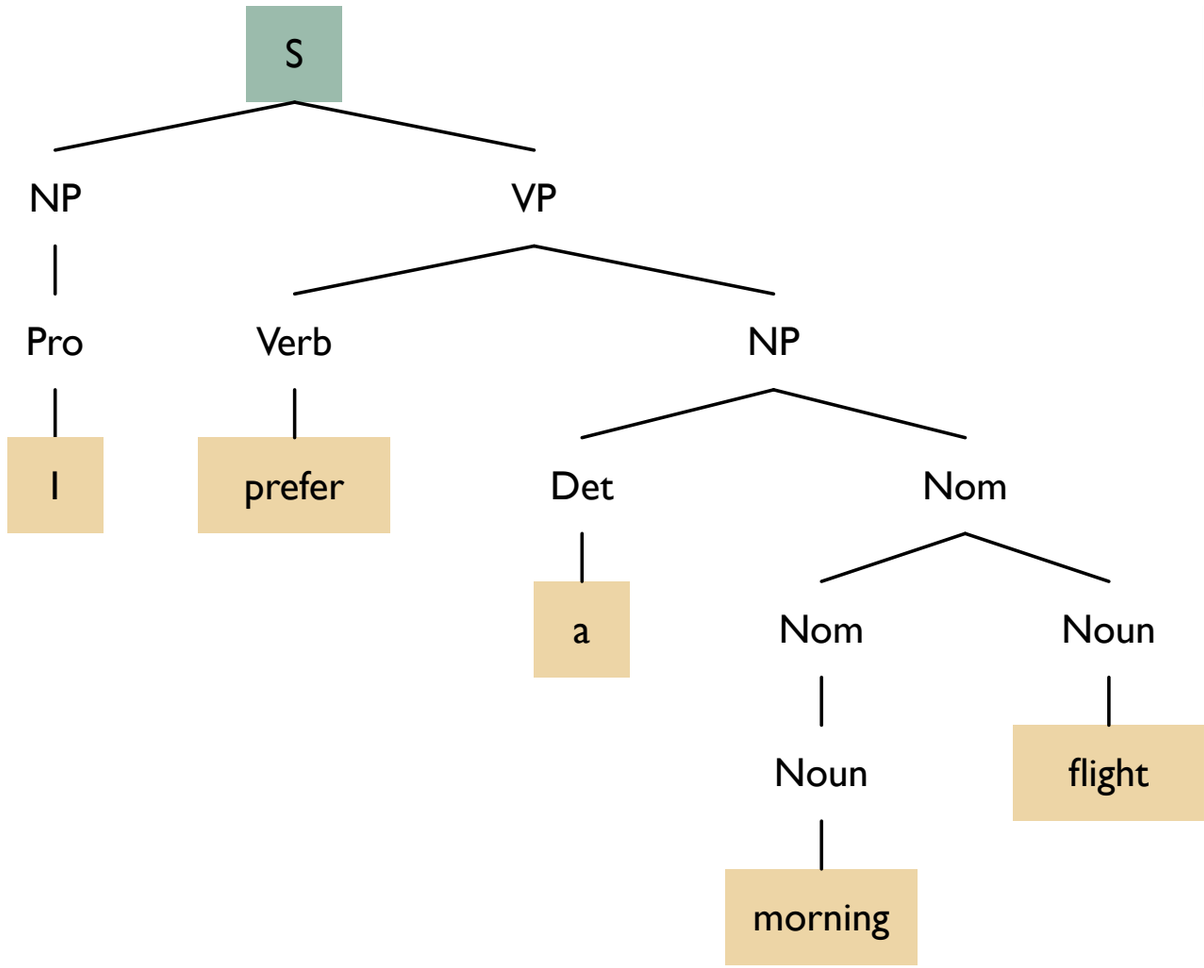
Chomsky

Ajdukiewicz

Tesnière

UPPSALA UNIVERSITET

Phrase structure trees

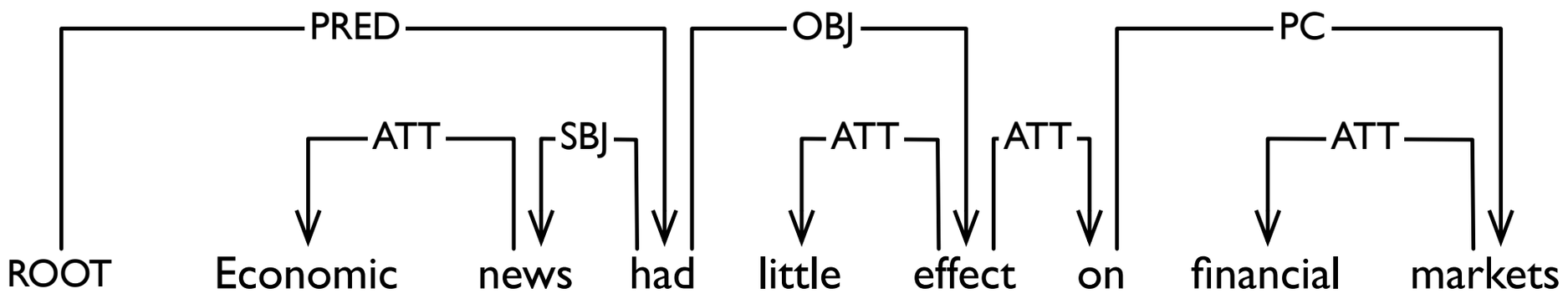


root (top)

leaves (bottom)

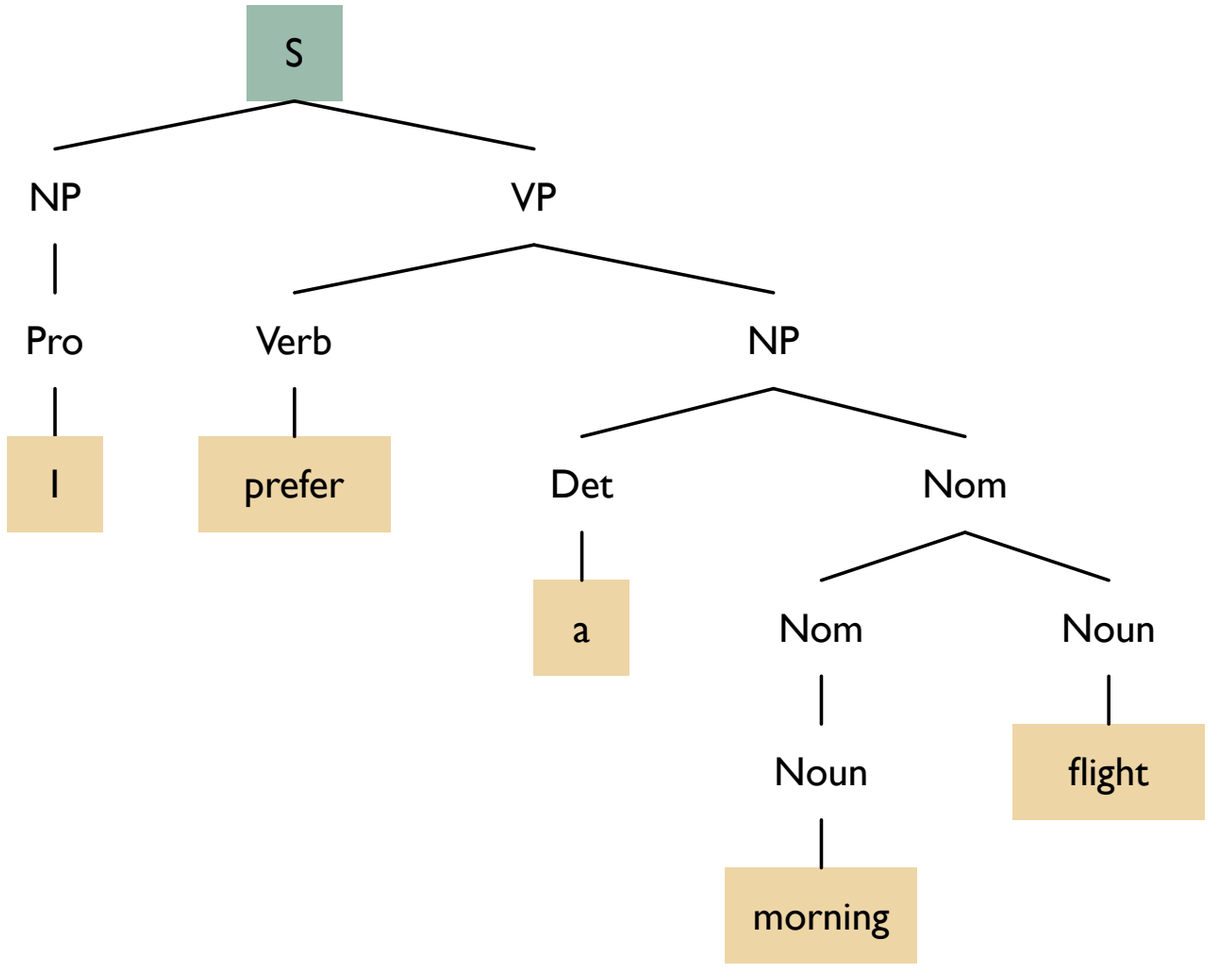
UPPSALA
UNIVERSITET

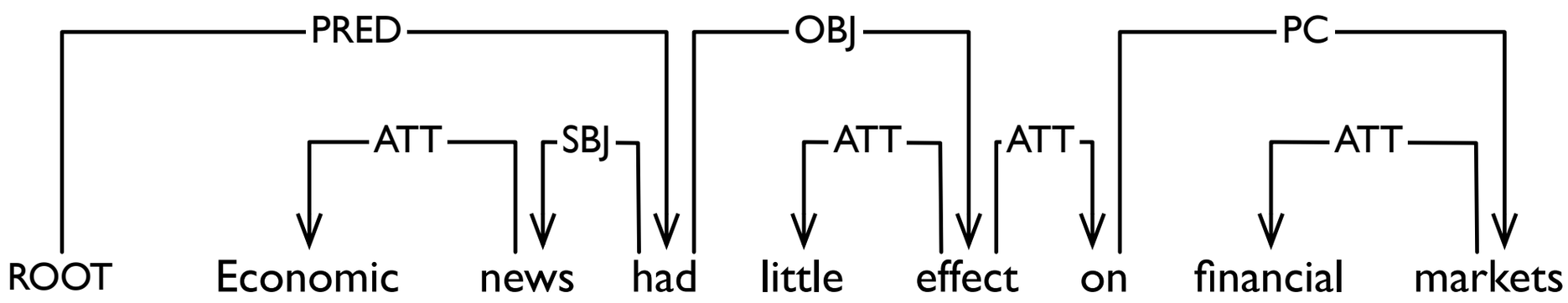
Dependency trees



UPPSALA UNIVERSITET

Phrase structure vs dependency trees





UPPSALA UNIVERSITET

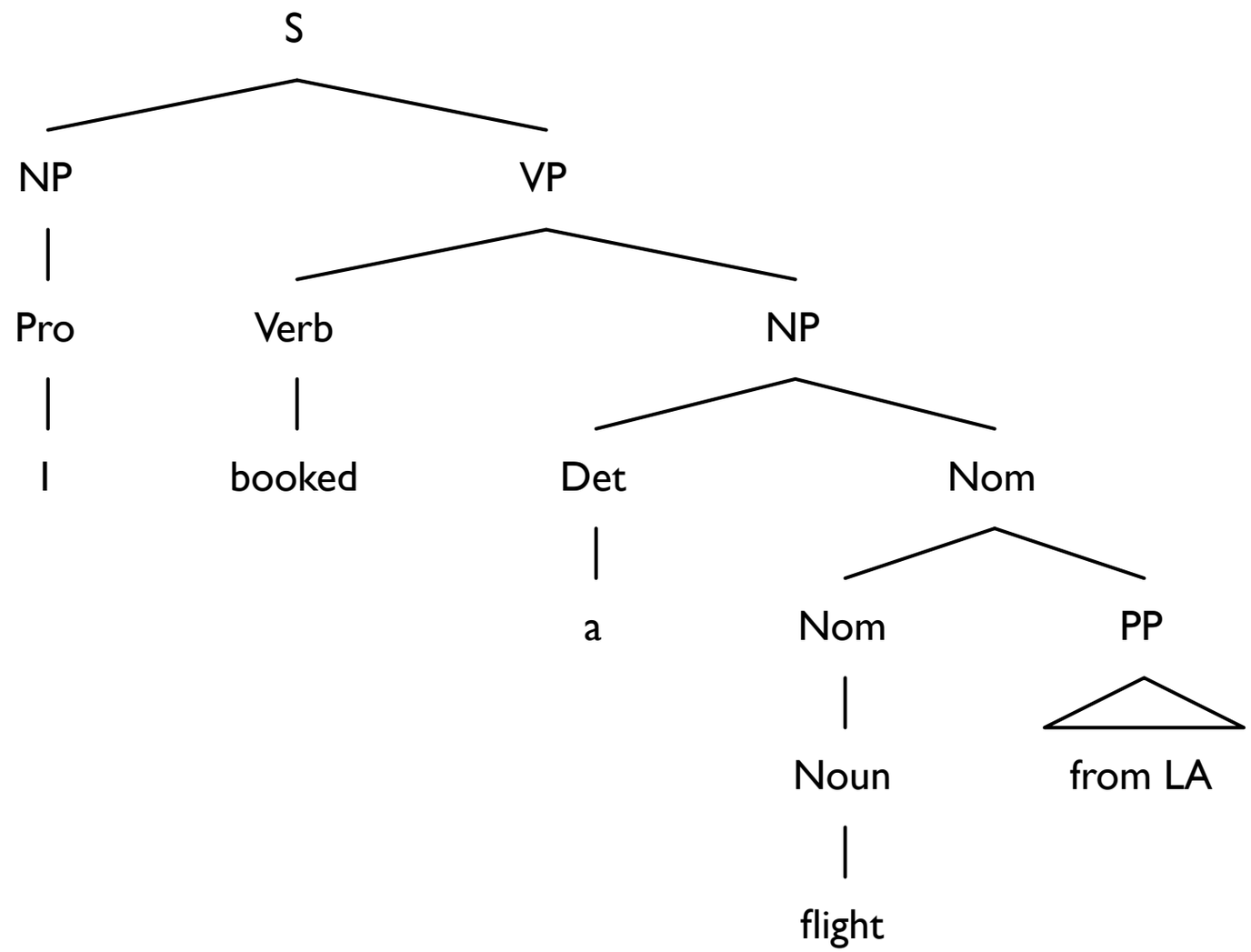
Ambiguity

I booked a flight from LA.

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

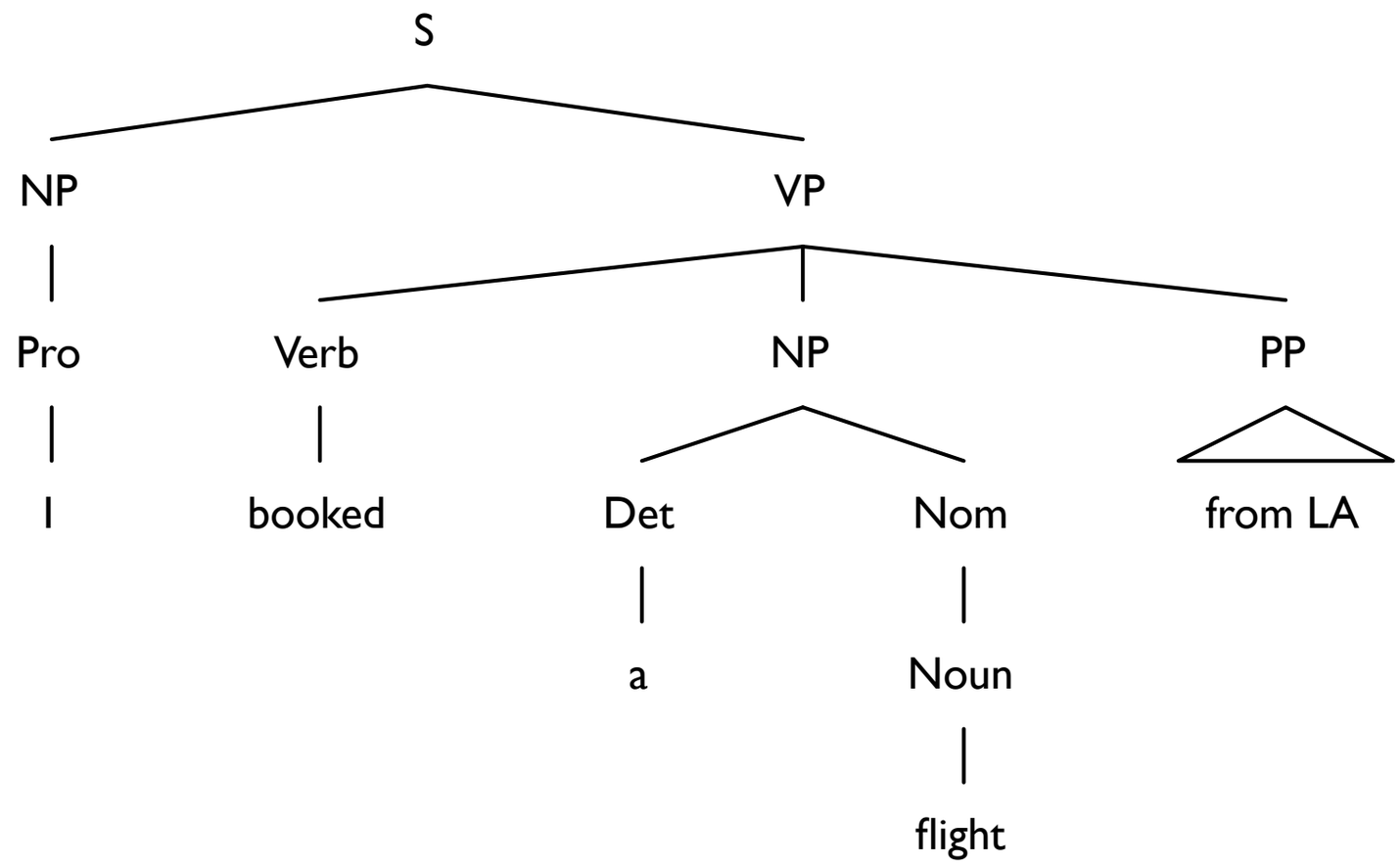
UPPSALA UNIVERSITET

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

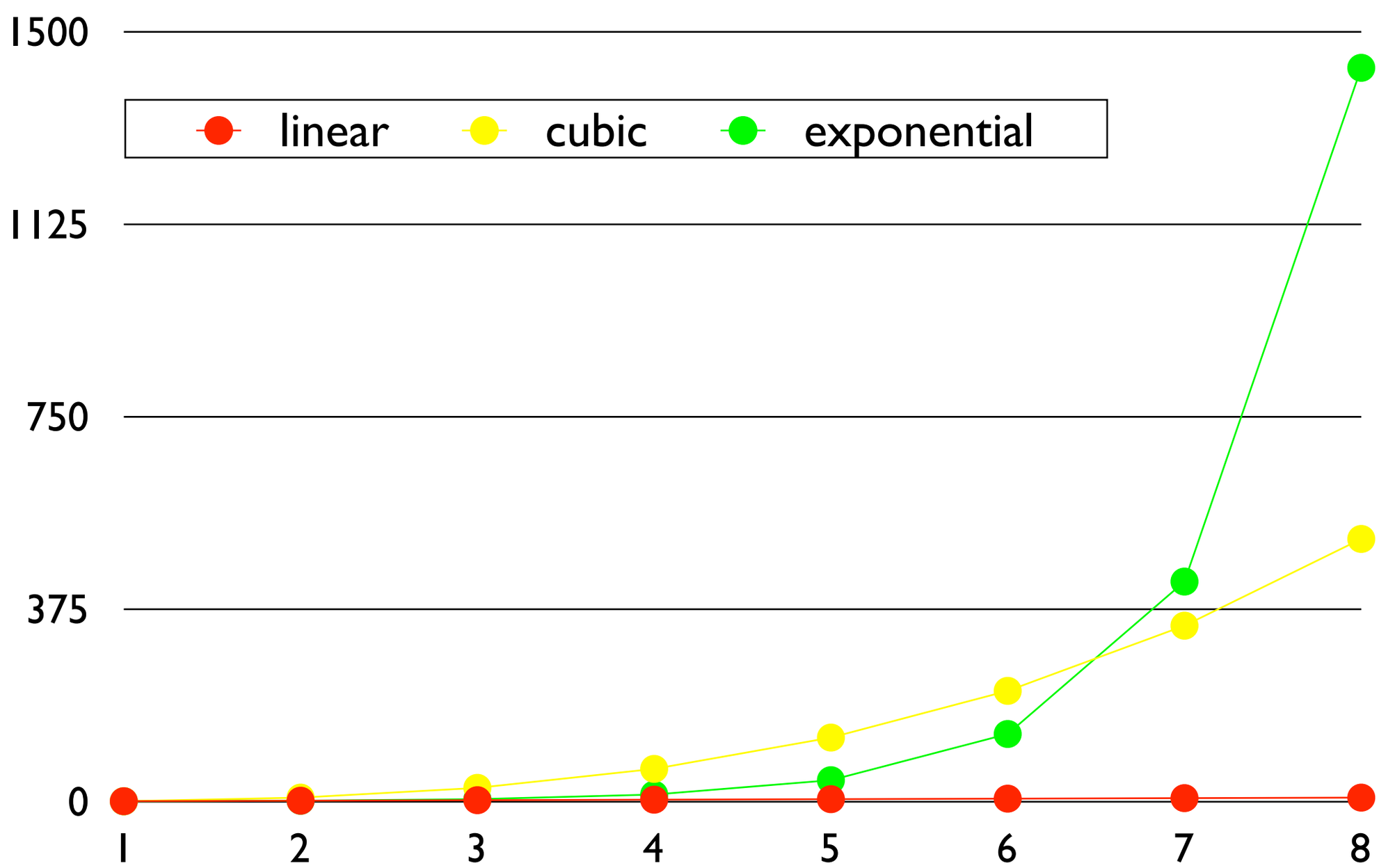
bottom–up

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 (DP) (bottom up):
  Solve each subproblem only once and save the solution in order to use it as a partial solution in a larger subproblem.
- Memoisation (top down):
  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 5LN455

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 5LN455

- 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 5LN455

- Examination is continuous and distributed over four graded assignments and two literature seminars.
- 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 seminars you will discuss scientific articles about parsing.
  Time to invest: about 10 hours per seminar

UPPSALA UNIVERSITET

Assignments 5LN455

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 (MaltParser)

UPPSALA UNIVERSITET

Literature seminars (all)

- Read one given article for each seminar
- Prepare according to the instructions on the homepage
- Everyone is expected to be able to discuss the article and the questions about it
    - It should be clear that you have read and analysed the article, but it is perfectly fine if you have misunderstood some parts
- The seminars are obligatory
    - If you miss a seminar or are unprepared, you will have to hand in a written report.

UPPSALA UNIVERSITET

Learning outcomes and examination 5LN455

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

UPPSALA UNIVERSITET

Learning outcomes and examination 5LN455

- apply existing systems that use these methods to realistic data and evaluate them with respect to their accuracy and efficiency; project assignment 2

- implement a central component of at least one approach to syntactic analysis in a suitable programming language. project assignment 1

UPPSALA UNIVERSITET

Grading 5LN455

- The assignments are graded with G and VG
- G on the seminars if present, prepared and active. The seminars are obligatory!
- To achieve G on the course:
  - G on all assignments and seminars
- To achieve VG on the course:
  - Same as for G **and** VG on at least two assignments, of which at least one is practical

UPPSALA UNIVERSITET

Intended learning outcomes 5LN713/5LN717

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

- explain the standard models and algorithms used in phrase structure and dependency parsing;
- implement and evaluate some of these techniques;
- critically evaluate scientific publications in the field of syntactic parsing,
- design, evaluate, or theoretically analyse the syntactic component of an NLP system (5LN717)

UPPSALA UNIVERSITET

Examination 5LN713/5LN717

- Examination is continuous and distributed over three graded assignments, two literature seminars, and a project (for 7,5 credits)
- Two assignments are small projects where you implement parsers.
- Literature review assignment
- Two literature seminars

UPPSALA UNIVERSITET

Grading 5LN713/5LN717

- The assignments are graded with G and VG
- G on the seminars if present, prepared and active. The seminars are obligatory!
- To achieve G on the course:
  - G on all assignments and seminars
- To achieve VG on the course:
  - Same as for G **and** VG on at least two assignments/project

UPPSALA UNIVERSITET

Teachers

- Sara Stymne
  - Course coordinator (bachelor), lectures, seminars, assignments (bachelor)
- Joakim Nivre
  - Examiner, course coordinator (master), assignments (master)

UPPSALA UNIVERSITET

Teaching

- 8 lectures
- 2 seminars
- No scheduled supervision / lab hours
- Supervision available on demand:
  - Email
  - Knock on office door
  - Book a meeting

Lectures

- Lectures covers basic parsing algorithms, enough material for the bachelor course
- They touch on more advanced material, but master students will need to read up on that independently
- Lectures will usually include small practical tasks
- Do not expect the slides to be self contained! You will not be able to pass the course only by looking at the slides.

UPPSALA
UNIVERSITET

Course workload 5LN455

- 7.5 hp means about 200 hours work
- 16 h lectures
- 4 h seminar
- 180 h work on your own
  - ~ 96 h assignment work
  - ~ 20 h seminar preparation
  - ~ 64 h additional reading

UPPSALA UNIVERSITET

Schedule

week 45	(4 hrs taught, 14 hrs reading, 2 hrs assignments)	
8/11	13–15	Lecture: Introduction
10/11	13-15	Lecture: Phrase structure parsing 1
week 46	**(4 hrs taught, 8 hrs reading, 8 hrs assignments)**	
14/11	13-15	Lecture: Phrase structure parsing 2
15/11	13-15	Lecture: Tree banks
week 47	**(2 hrs taught, 4 hrs seminar prep, 14 hrs assignments)**	
22/11	13–15	Lecture: Phrase structure parsing 3
week 48	**(4 h taught, 6 hrs seminar prep, 10 hrs assignments)**	
29/11	13–16	Literature seminar 1 (in smaller groups, times might change)
1/12	13-15	Dependency parsing 1

UPPSALA UNIVERSITET

Schedule

week 49 (2 hrs taught, 4 hrs reading, 14 hrs assignments)		
6/12		Deadline: Assignments 1 and 2
6/12	13-15	Lecture: Dependency parsing 2
week 50 (2 hrs taught, 12 hrs reading, 6 hrs assignments)		
13/12	13-15	Lecture: Dependency parsing 3
week 52-02 (0 hrs taught, 4 hrs seminar prep, 26 hrs reading, 30 hrs assignments)		
week 03 (2 hrs taught, 6 hrs seminar prep, 12 hrs assignments)		
12/01	13–16	Literature seminar 2 (in smaller groups, times might change)
13/01		Deadline: Assignments 3 and 4

UPPSALA
UNIVERSITET

Reading: course books

- 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 1-4, 6.

UPPSALA UNIVERSITET

Reading: articles

- Seminar 1
  - Mark Johnson. PCFG Models of Linguistic Tree Representations. Computational Linguistics 24(4). Pages 613-632.
- Seminar 2
  - Joakim Nivre and Jens Nilsson. Pseudo-Projective Dependency Parsing. Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05). Pages 99-106. Ann Arbor, USA.

UPPSALA UNIVERSITET

Evaluation from last year

- Overall score: 3.7 (4.75 in 2014)
  - Good seminars and assignments
  - Missed some pre-requisites (mainly with respect to programming)
  - Slides were not self-explanatory
    - Comment: they are not meant to be, they are a complement to the lectures.
- This year
  - Assignment 2 completely changed, to better suit your programming skills
  - Change of teacher (mainly practical reasons)
  - Joint bachelor and master course (no master course before)
  - Otherwise not much change, since the course has been working well for some years. The lectures will be more similar to 2014, when they were more appreciated, and mostly include small exercises.

UPPSALA UNIVERSITET

Work to do this week (5LN455)

- Read chapter 12.1-12.7
- Read chapter 13.1-13.3 in preparation for Thursday
- Read descriptions of assignment 1 and 2
- If you need, repeat:
  - parts of grammar course; phrase structure grammars and dependency grammars
  - programming course; practice some python