LING439/539 Statistical NLP (Fall 2016)

Tuesday 23 August - Introduction

Welcome !

- LING439/539 ***Statistical NLP***
- Tuesday and Thursday, 3:30-4:45PM.

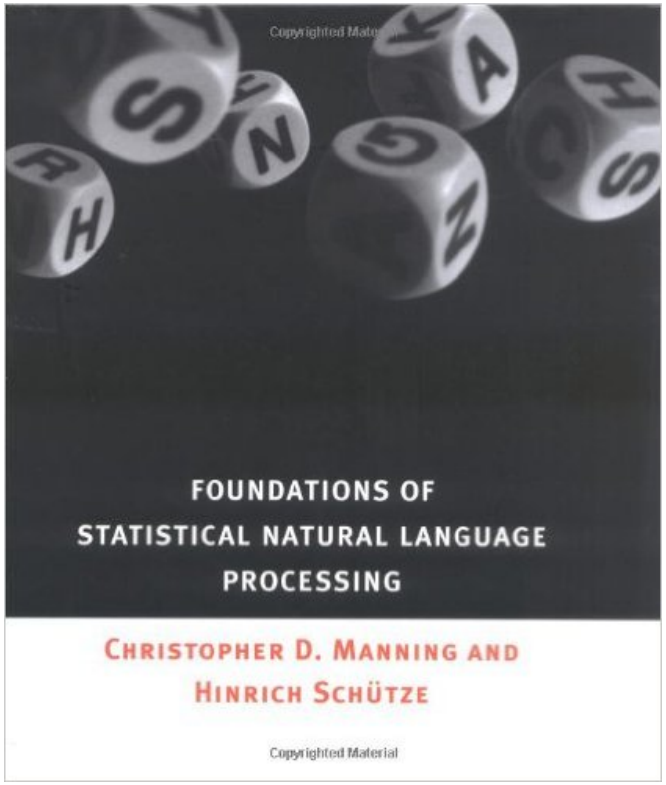
CL/NLP in UA

- LING438/538 Computational linguistics (required 388)
- **LING439/539 Statistical NLP (required 388/438)**
- LING388 Language and computers
- LING581 Advanced CL

Fall 2016

Spring 2017

Text book


FOUNDATIONS OF
STATISTICAL NATURAL LANGUAGE
PROCESSING
CHRISTOPHER D. MANNING AND
HINRICH SCHÜTZE
Copyrighted Material

Prerequisites

- For LING439: LING388 + LING438
- For LING539: No formal prerequisites.

Grading

- For LING439: 100% Homework (+ optional final project)
- For LING539: 75% Homework + 25% final project
  - You can work in groups of up to two people for the final project.
  - For students who take LING438/538 and LING439/539, one final project for both classes is allowed.

Natural language processing…

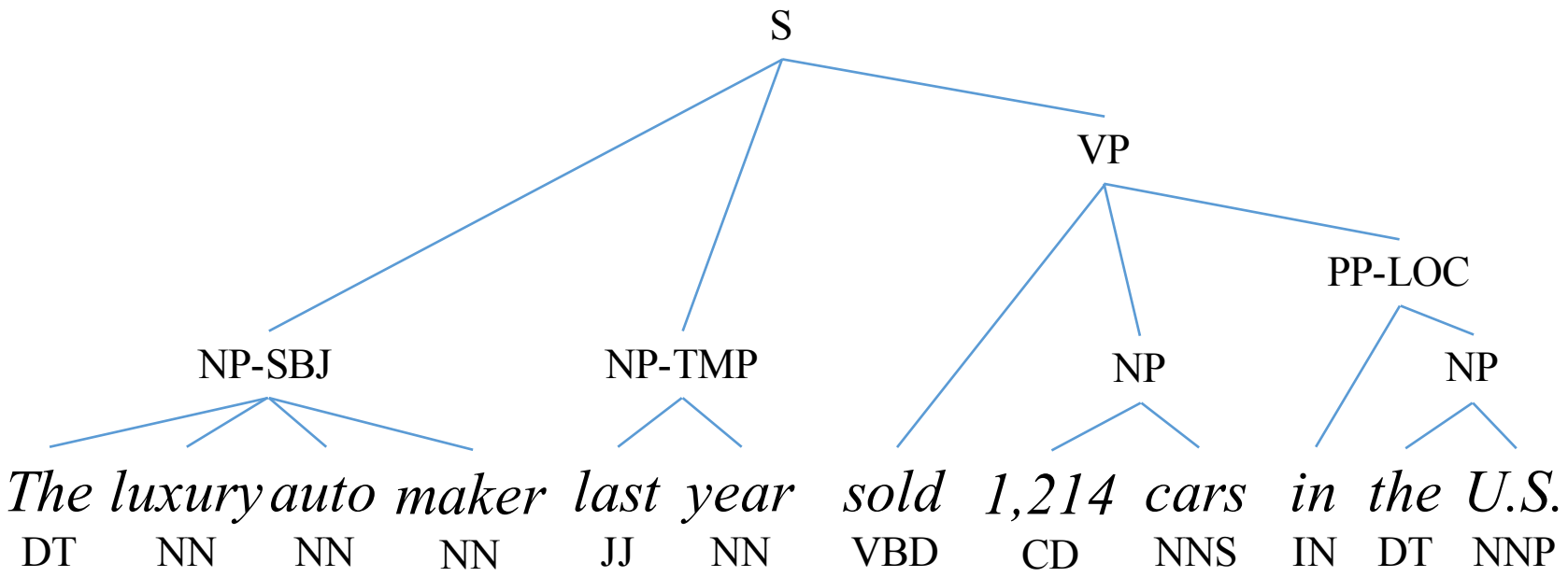
The luxury auto maker last year sold 1,214 cars in the U.S.

Part-of-Speech tagging

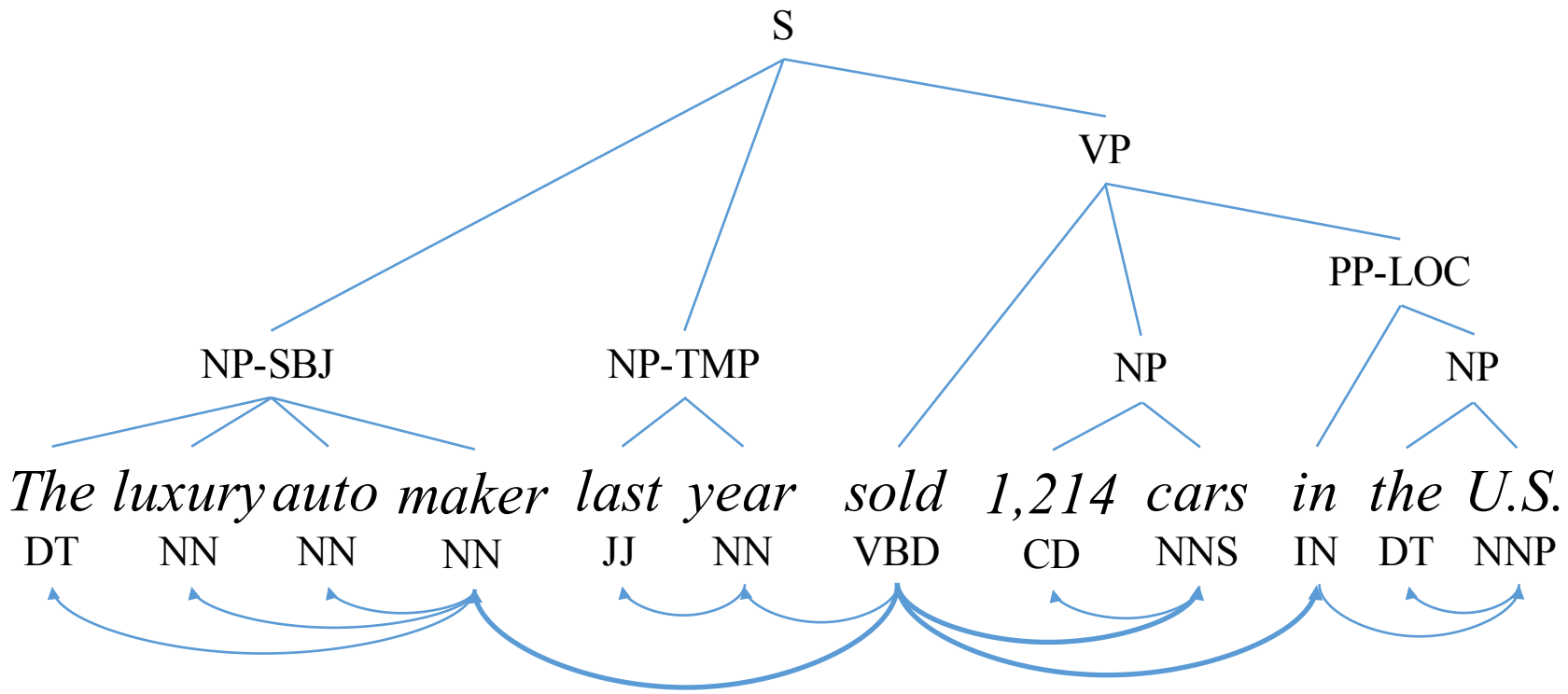
Penn tagset

The	*luxury*	*auto*	*maker*	*last*	*year*	*sold*	*1,214*	*cars*	*in*	*the*	*U.S.*
DT	NN	NN	NN	JJ	NN	VBD	CD	NNS	IN	DT	NNP

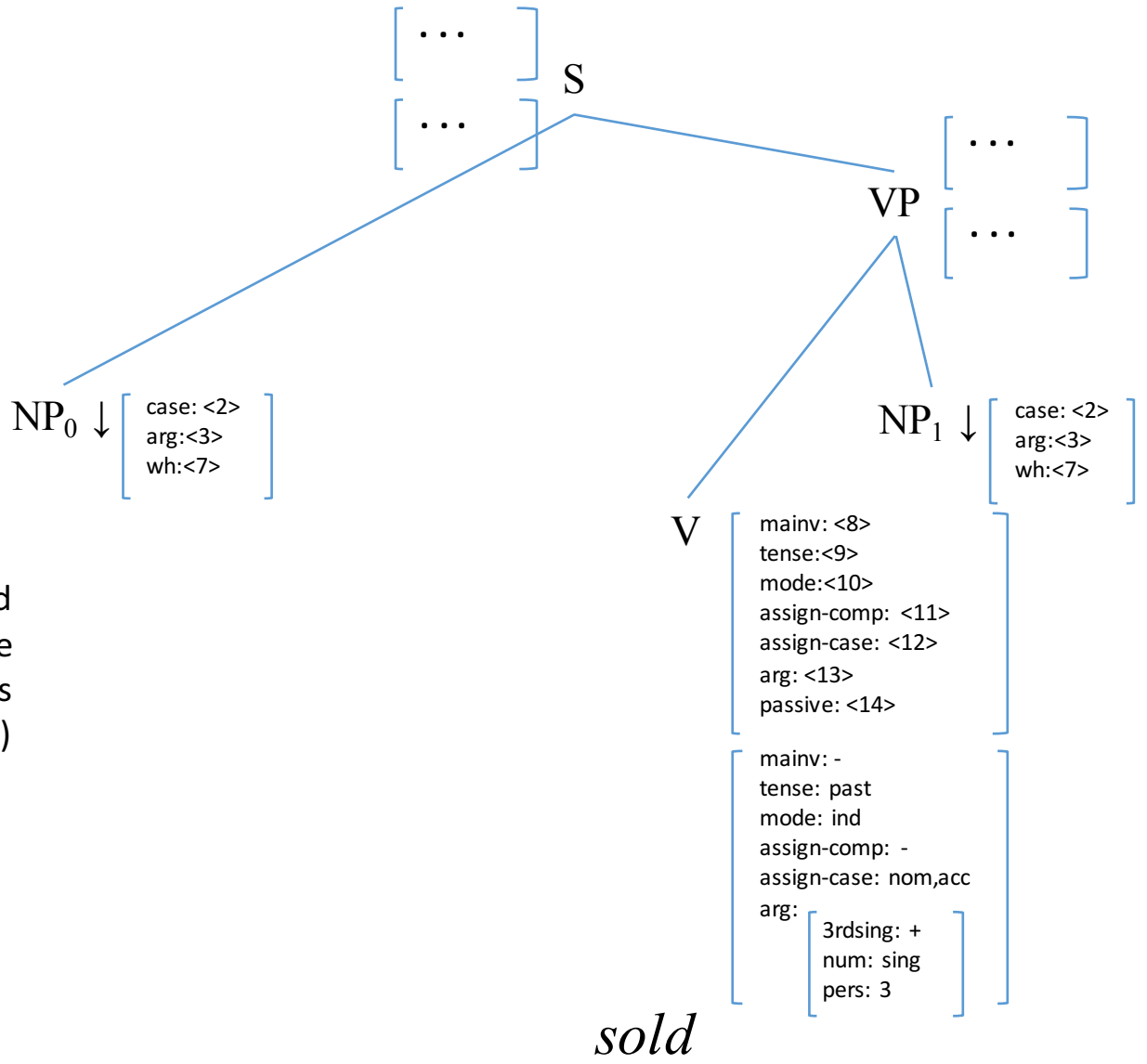
Phrase-structure parsing



Dependency parsing



Feature-based lexicalized tree adjoining grammars (FB-LTAG)



Computational semantics: FrameNet

The luxury auto maker last year **SOLD** *1,214 cars in the U.S.*

Seller **Time** **Commerce _sell** **Goods** **Place**

Named entity recognition

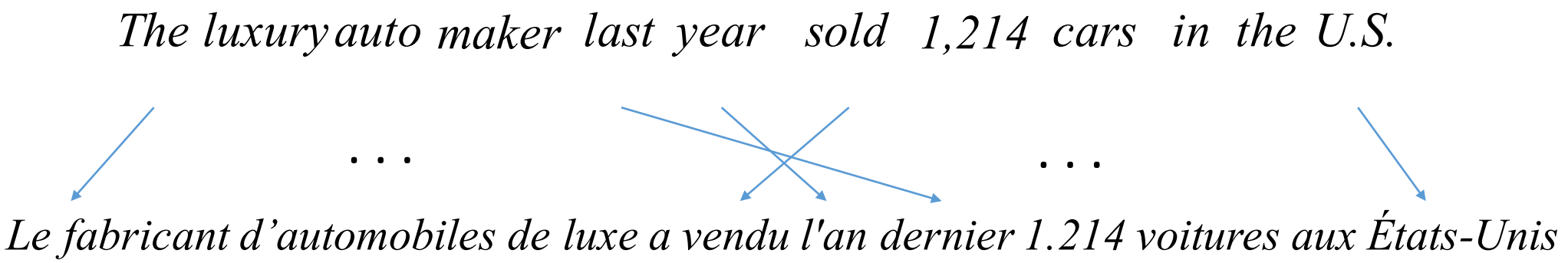
	The	*luxury*	*auto*	*maker*	*last*	*year*	*sold*	*1,214*	*cars*	*in*	*the*	*U.S.*
Named entity	O	O	O	O	O	O	O	O	O	O	O	**LOC**

Machine translation

The luxury auto maker last year sold 1,214 cars in the U.S.

Le fabricant d'automobiles de luxe a vendu l'an dernier 1.214 voitures aux États-Unis

Word alignment

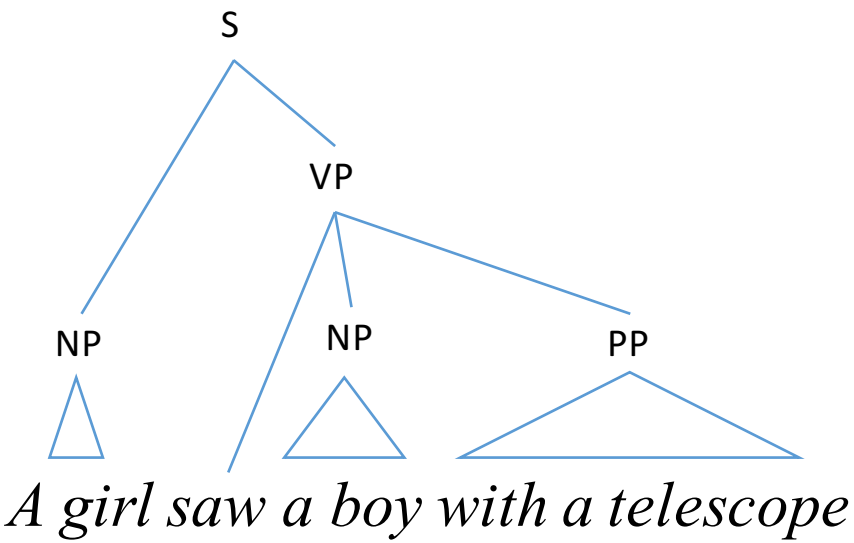


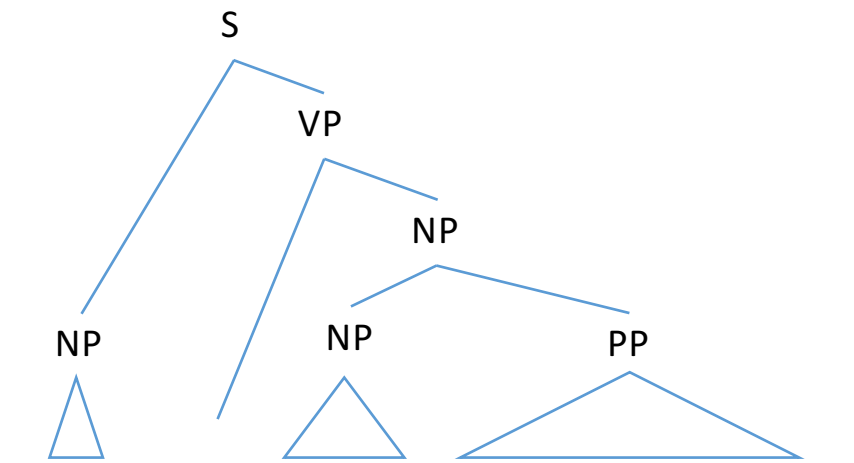
Why natural language processing is difficult ?

Ambiguities !

The luxury auto maker last year sold 1,214 cars in the U.S.

NOUN	*NOUN*
ADJ	*ADJ*
	ADV
	VERB


S
VP
NP
NP
PP
A girl saw a boy with a telescope


S
VP
NP
NP
NP
PP
A girl saw a boy with a telescope

What we will do to remove ambiguities?

Statistical approaches

Contents

Expectation–maximization (EM) algorithm

- POS tagging: forward-backward or Baum-Welch algorithm
- Parsing: Inside-Outside algorithm
- Word alignment in SMT: IBM Model 1