CPE 372/641 Natural Language Processing

Chart Parsing

Asst. Prof. Dr. Nuttanart Muansuwan

Review: simple context-free grammar for a fragment of English

CFG example

CFG's are also called phrase-structure grammars.
Equivalent to Backus-Naur Form (BNF).

1. S → NP VP
2. VP → V NP
3. NP → NAME
4. NP → ART N
5. NAME → Beavis
6. V → ate
7. ART → the
8. N → cat

- CFG's are *powerful* enough to describe most of the structure in natural languages.
- CFG's are *restricted* enough so that efficient parsers can be built.

Parsing Strategies: top-down vs. bottom-up

General Parsing Strategies

Grammar	Top-Down	Bottom-Up
1. S → NP VP	S → NP VP	→ NAME ate the cat
2. VP → V NP	→ NAME VP	→ NAME V the cat
3. NP → NAME	→ Beav VP	→ NAME V ART cat
4. NP → ART N	→ Beav V NP	→ NAME V ART N
5. NAME → Beavis	→ Beav ate NP	→ NP V ART N
6. V → ate	→ Beav ate ART N	→ NP V NP
7. ART → the	→ Beav ate the N	→ NP VP
8. N → cat	→ Beav ate the cat	→ S

A top-down parser

A parsing algorithm –

uses grammatical rules to search for a combination that generates a tree describing the structure of the input sentence

A top-down parser –

starts with the S symbol and rewrite it into a sequence of terminal symbols that matches the words in the input sentence

Some problems with top-down parser

- Start with S, so it only focuses on grammaticality.
- Not explore possible derivations or rules.

Chart Parser

Chart Parsers

chart: data structure that stores partial results of the parsing process in such a way that they can be reused. The chart for an n-word sentence consists of:

- $n + 1$ **vertices**
- a number of **edges** that connect vertices

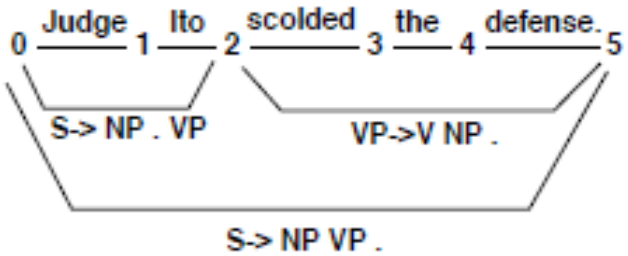


Chart parsing: the general idea

- The process of parsing an n- word sentence consists of a chart with n+1 vertices and adding edges to the chart one at a time.
- Goal: to produce a complete edge that spans from vertex 0 to n and is of category S

Bottom-up chart parsing

1. $S \rightarrow NP\ VP$
2. $NP \rightarrow ART\ ADJ\ N$
3. $NP \rightarrow ART\ N$
4. $NP \rightarrow ADJ\ N$
5. $VP \rightarrow AUX\ VP$
6. $VP \rightarrow V\ NP$

} Grammar 2.

- Example: given Grammar 2 and a sentence that starts with ART, the parser uses ART to matched with rules 2 and 3.
  -- use $\circ$ to indicate what has been seen so far
  -- thus,

$$NP \rightarrow ART \circ ADJ\ N$$
$$NP \rightarrow ART \circ N$$

An example of chart parsing algorithm

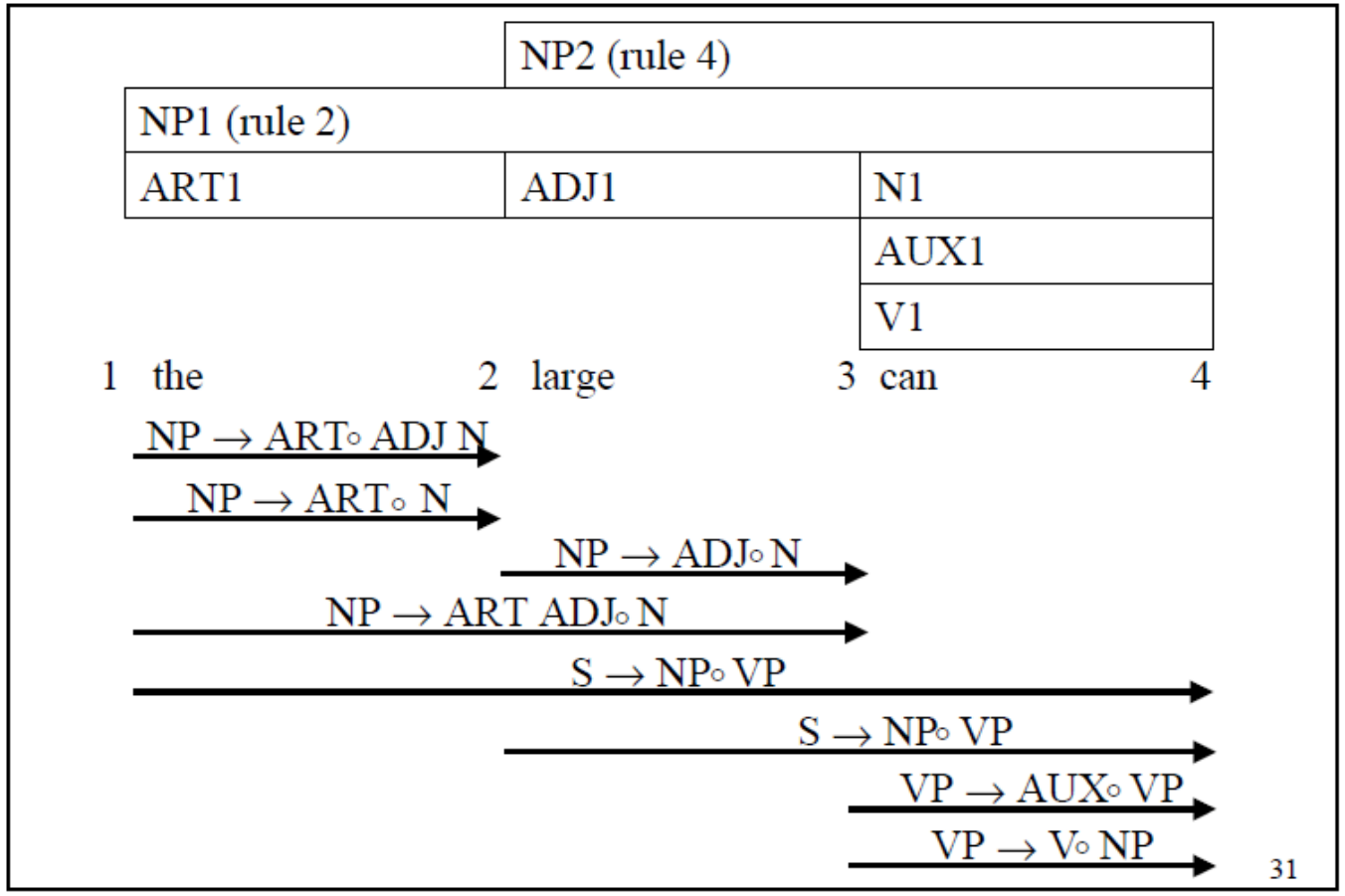
- Consider using the algorithm on the sentence

"*The large can can hold the water*"

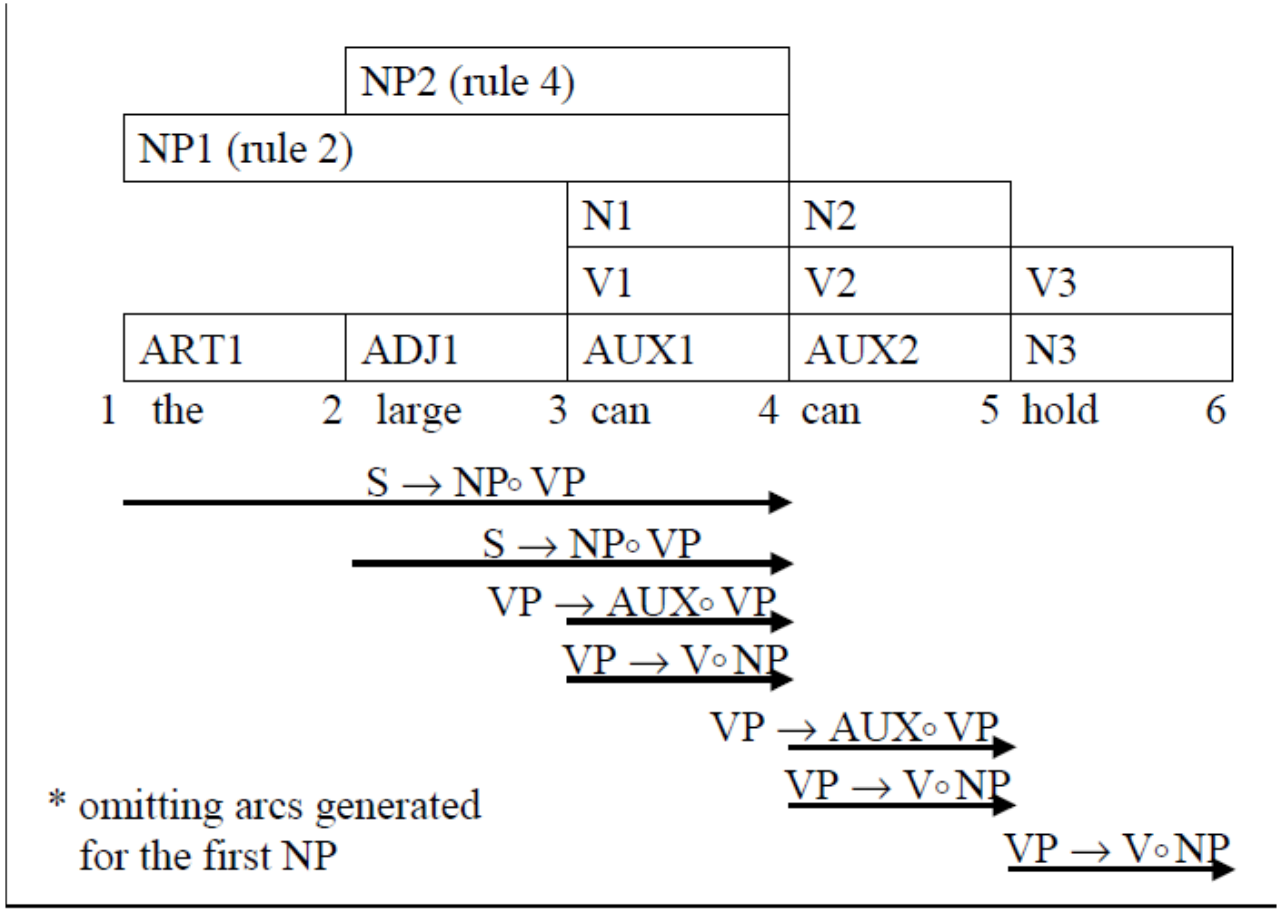
using the following Grammar 2 and following lexicon:

the:	ART
large:	ADJ
can:	N, AUX, V
hold:	N, V
water:	N, V

After parsing ‘the large can’



After adding ‘hold’



The final chart