DECODING DISTRIBUTED TREE STRUCTURES

L. Ferrone[1] F.M. Zanzotto[1] X. Carreras[2]

[1]University of Rome "Tor Vergata" [2]Xerox Research Center Europe

UNIVERSITA' degli STUDI di ROMA TOR VERGATA

Xerox Research Centre Europe

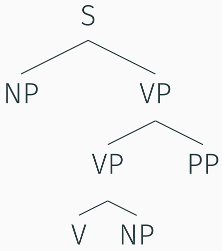
INTRODUCTION

- Natural language processing tasks benefit from syntactic information

- Natural language processing tasks benefit from syntactic information

Directly

- Symbolic Tree Structures



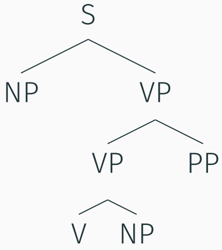
- Tree Kernels (Collins; 2001)

FROM SYNTACTIC TREES TO DISTRIBUTED TREES

- Natural language processing tasks benefit from syntactic information

Directly

- Symbolic Tree Structures



- Tree Kernels (Collins; 2001)

Indirectly

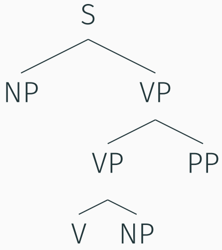
- Distributed Tree Structures

FROM SYNTACTIC TREES TO DISTRIBUTED TREES

- Natural language processing tasks benefit from syntactic information

Directly

- Symbolic Tree Structures



- Tree Kernels (Collins; 2001)

$\longrightarrow$

Indirectly

- Distributed Tree Structures

$$\mathbf{t} = (0.0112, 0.212, \ldots, 0.0081) \in \mathbb{R}^d$$

Distributed Trees (Zanzotto; 2012)

- Approximate tree kernels (Collins; 2001)

$$\langle \mathbf{t}_1, \mathbf{t}_2 \rangle \approx TK(T_1, T_2)$$

Distributed Trees (Zanzotto; 2012)

- Approximate tree kernels (Collins; 2001)

$$\langle \mathbf{t_1}, \mathbf{t_2} \rangle \approx TK(T_1, T_2)$$

- Faster to compute than tree kernels

UNIVERSITA' degli STUDI di ROMA TOR VERGATA

Distributed Trees (Zanzotto; 2012)

- Approximate tree kernels (Collins; 2001)

$$\langle \mathbf{t_1}, \mathbf{t_2} \rangle \approx TK(T_1, T_2)$$

- Faster to compute than tree kernels
- Can be used as input in any algorithm
  - Neural network
  - Support Vector Machines
  - ...

WHAT'S IN A DISTRIBUTED VECTORS?

Question

- How much information is stored in a distributed vector?
- In other words, can we decode the structured representation from a distributed vector?

UNIVERSITA' degli STUDI di ROMA TOR VERGATA

Our Idea

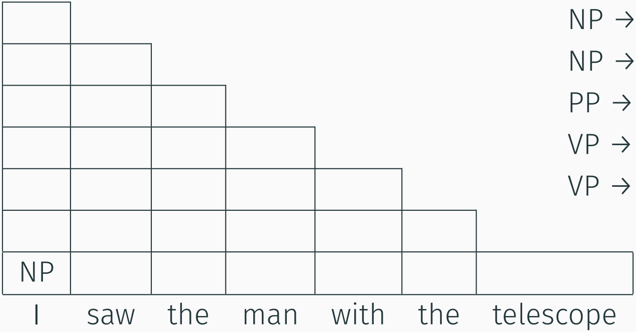
- Traditional parsing:
    - CYK algorithm (and others)
- Use distributed vectors to *"guide"* the algorithm choices

CYK (Cocke, Younger, Kasami; 1967)

Given a sentence *s* of length *n* and a grammar *G*:

- builds a $n \times n$ table which contains the partial parses of the sentence

CYK ALGORITHM



Grammar:

S → NP VP

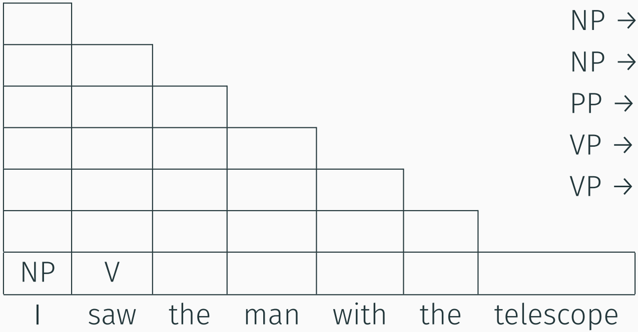
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP



Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

NP	V	DET				
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

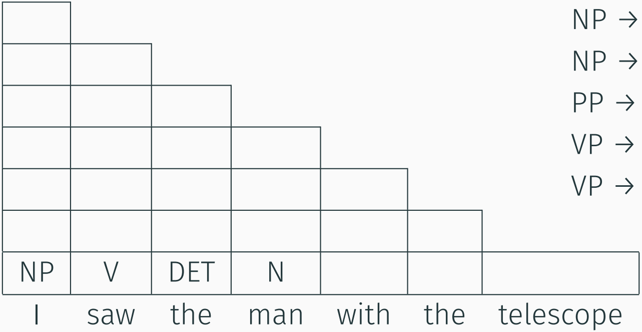
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP



Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

UNIVERSITA' degli STUDI di ROMA TOR VERGATA

NP	V	DET	N	P		
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

NP	V	DET	N	P	DET	
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

·						
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

	·					
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

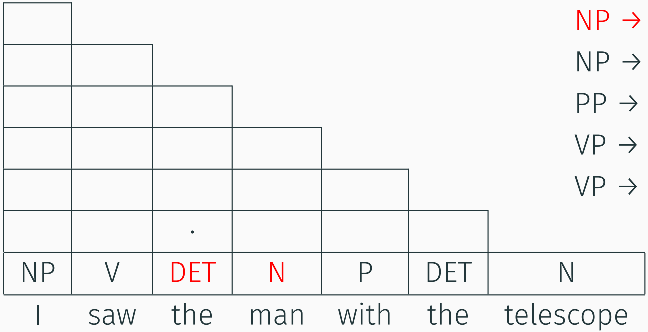
NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM



Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

		NP				
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

We also store backpointers to record what rule we choose

		NP				
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

- S → NP VP
- NP → DET N
- NP → NP PP
- PP → P NP
- VP → V NP
- VP → VP PP

CYK ALGORITHM

		NP	·			
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

		NP		·		
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

·						
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

	VP					
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

	VP	·				
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

	VP		·			
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

S	·					
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

S		·				
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

S			·			
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

·						
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

- S → NP VP
- NP → DET N
- NP → NP PP
- PP → P NP
- VP → V NP
- VP → VP PP

	·					
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

- S → NP VP
- NP → DET N
- NP → NP PP
- PP → P NP
- VP → V NP
- VP → VP PP

CYK ALGORITHM

		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

·						
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

	·					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

	VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

	VP, VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

Again, we store backpointers for the two possibilities

	VP, VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

CYK ALGORITHM

S						
	VP, VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

Ambiguity

- Even in this small example there are two plausible interpretations

Ambiguity

- Even in this small example there are two plausible interpretations
- In general there are (exponentially) many more!

Ambiguity

- Even in this small example there are two plausible interpretations
- In general there are (exponentially) many more!
- Usually parsers use probabilistic grammars to disambiguate
    - Each rule of the grammar has an inherent probability (which must be learned)

UNIVERSITA' degli STUDI di ROMA TOR VERGATA

Idea

We show that a reference distributed vector of the correct parse is enough to eliminate ambiguity

(and thus reconstruct the original parse)

UNIVERSITA' degli STUDI di ROMA
TOR VERGATA

Ingredients:

Ingredients:

- Reference grammar G

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

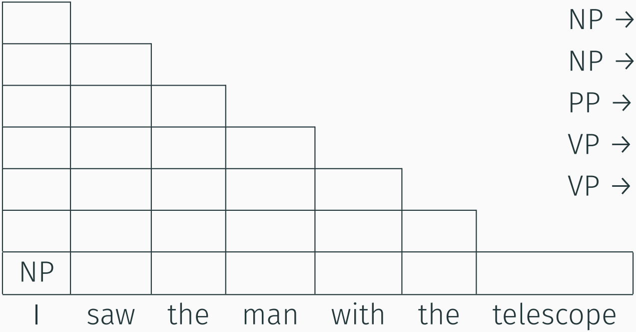
VP → VP PP

Ingredients:

- Reference grammar G
- Distributed vector $\mathbf{t}$

$$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

DISTRIBUTED CYK



Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK

Grammar:

S → NP VP

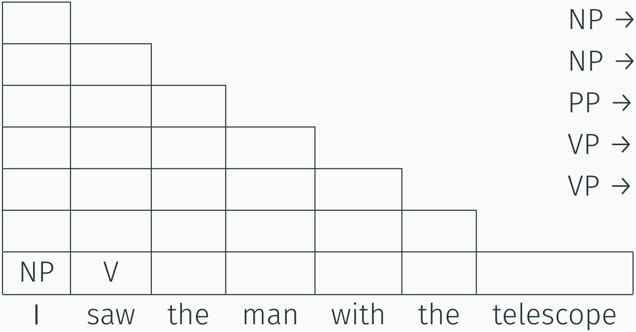
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP



$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK

Grammar:

S → NP VP

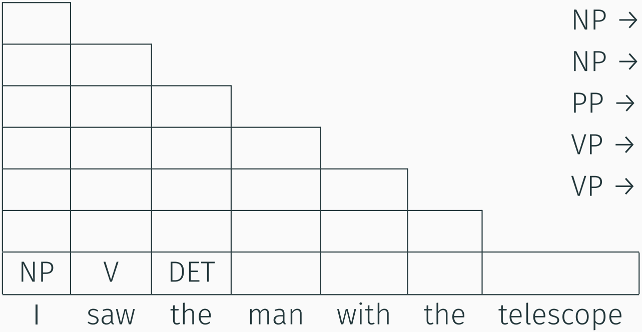
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP



$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK

NP	V	DET	N			
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK

Grammar:

S → NP VP

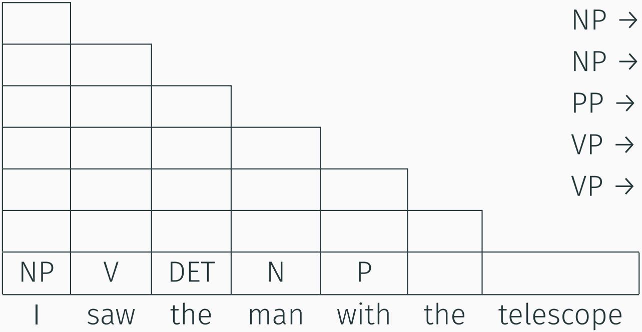
NP → DET N

NP → NP PP

PP → P NP

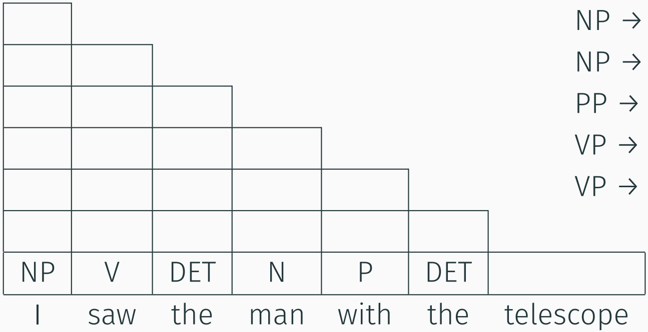
VP → V NP

VP → VP PP



$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK



Grammar:

S → NP VP

NP → DET N

NP → NP PP

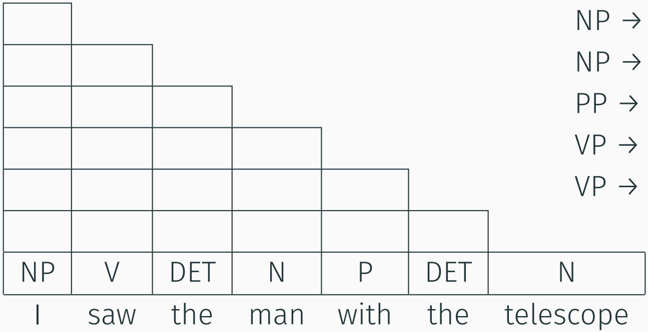
PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK



Grammar:

S → NP VP

NP → DET N

NP → NP PP

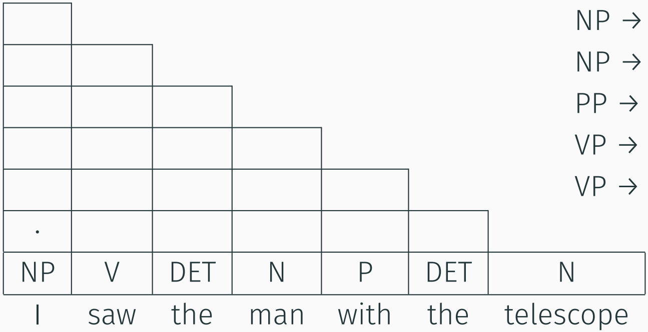
PP → P NP

VP → V NP

VP → VP PP

$t \to \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK



Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK

Grammar:

S → NP VP

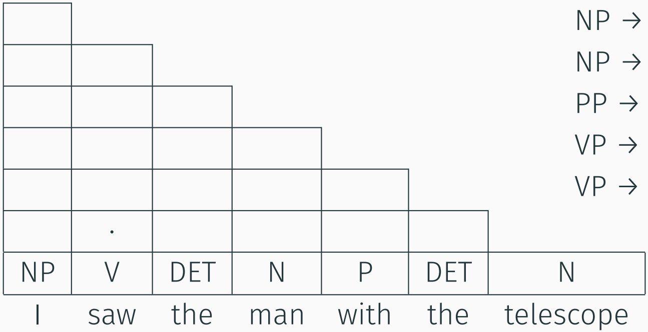
NP → DET N

NP → NP PP

PP → P NP

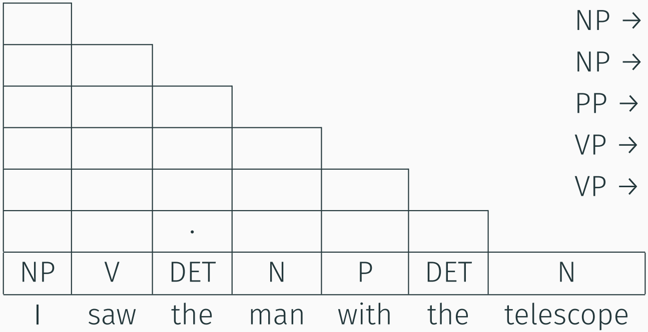
VP → V NP

VP → VP PP



$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK



Grammar:

S → NP VP

NP → DET N

NP → NP PP

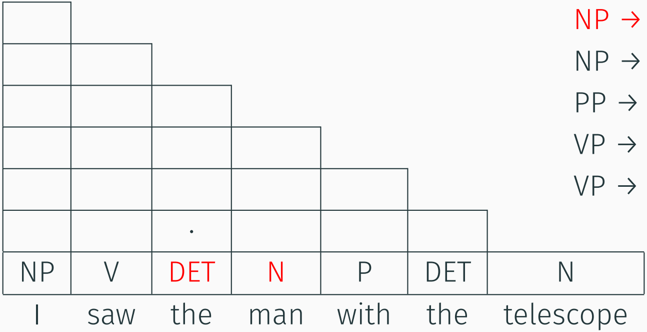
PP → P NP

VP → V NP

VP → VP PP

$$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

DISTRIBUTED CYK



Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

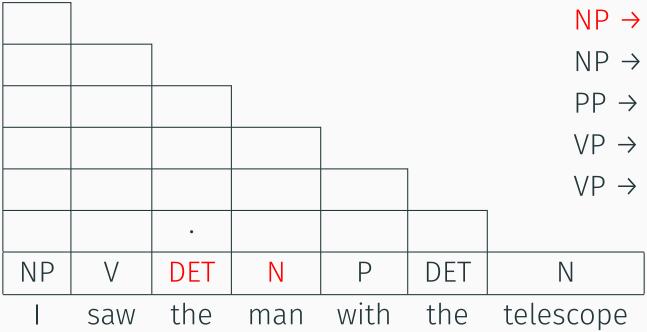
VP → V NP

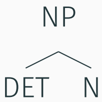
VP → VP PP

$$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

Grammar:

S → NP VP
NP → DET N
NP → NP PP
PP → P NP
VP → V NP
VP → VP PP





$t \to \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$ $\quad$ $\mathbf{r} = (0.005, 0.043, \ldots, 0.016)$

Grammar:

S → NP VP

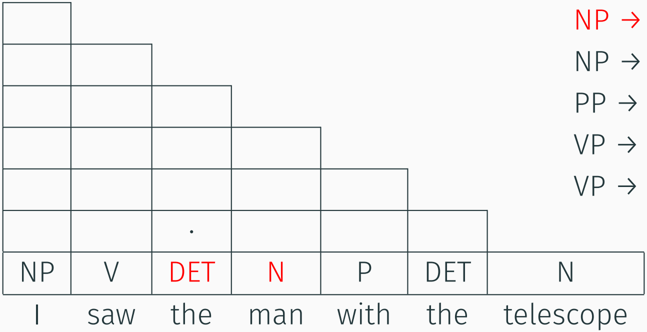
NP → DET N

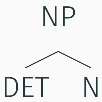
NP → NP PP

PP → P NP

VP → V NP

VP → VP PP





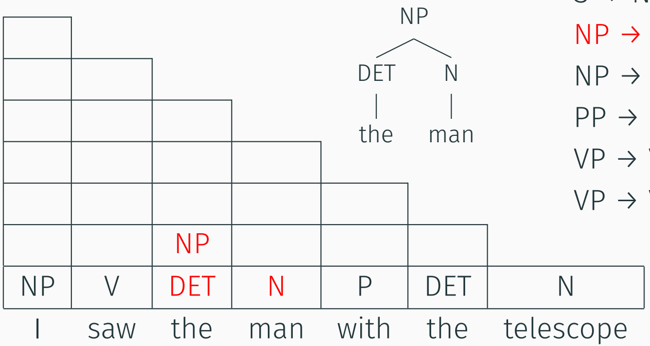
$\langle \mathbf{t}, \mathbf{r} \rangle$

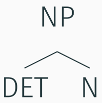
$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$ $\mathbf{r} = (0.005, 0.043, \ldots, 0.016)$

We store the full subtrees here!

Grammar:

S → NP VP
NP → DET N
NP → NP PP
PP → P NP
VP → V NP
VP → VP PP

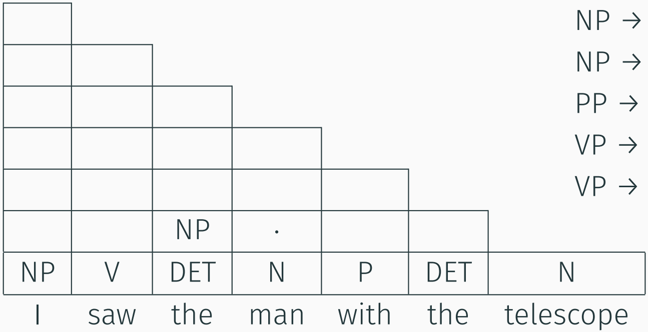




$\langle \mathbf{t}, \mathbf{r} \rangle$

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$ $\quad \mathbf{r} = (0.005, 0.043, \ldots, 0.016)$

DISTRIBUTED CYK



Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$$t \to \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

Grammar:

S → NP VP

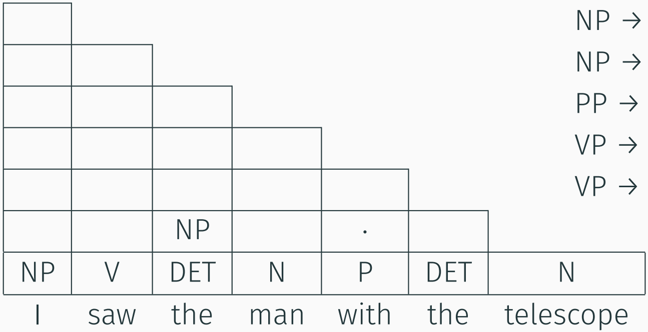
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP



$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK

		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

·						
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

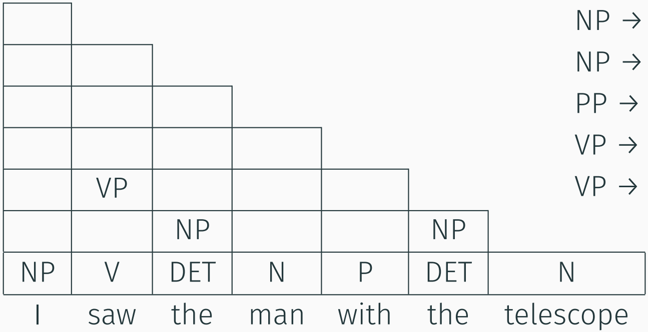
NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$



Grammar:

S → NP VP

NP → DET N

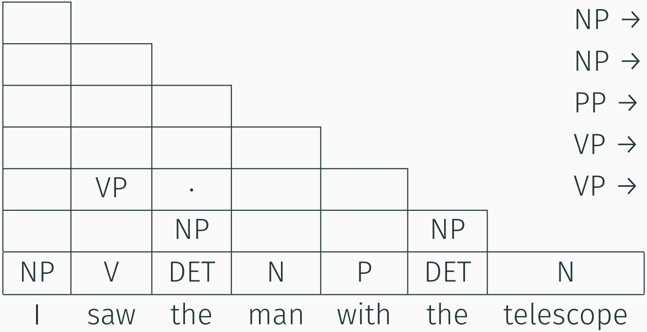
NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$



Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

Grammar:

S → NP VP

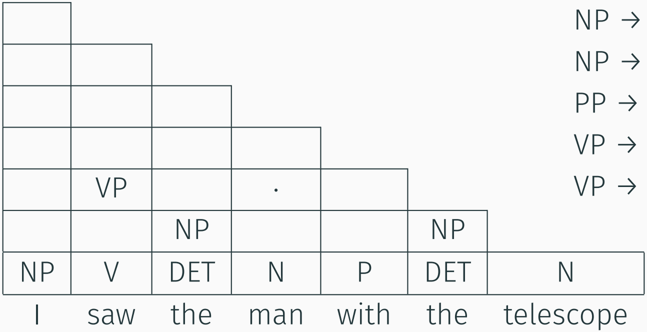
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP



$$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

DISTRIBUTED CYK

	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

S	·					
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK

S		·				
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

S			·			
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK

·						
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

DISTRIBUTED CYK

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

	·					
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK

Grammar:

- S → NP VP
- NP → DET N
- NP → NP PP
- PP → P NP
- VP → V NP
- VP → VP PP

		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

DISTRIBUTED CYK

·						
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

DISTRIBUTED CYK

	·					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

	·					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

	·					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

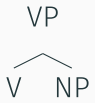
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP



$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$ $\quad \mathbf{r}_1 = (0.005, 0.043, \ldots, 0.016)$

	·					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

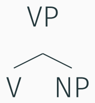
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP



$\langle \mathbf{t}, \mathbf{r}_1 \rangle$

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$ $\quad \mathbf{r}_1 = (0.005, 0.043, \ldots, 0.016)$

	VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

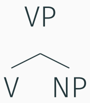
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

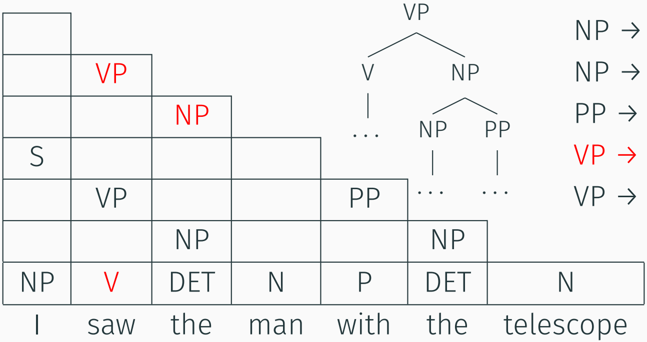


$\langle \mathbf{t}, \mathbf{r}_1 \rangle$

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

$\mathbf{r}_1 = (0.005, 0.043, \ldots, 0.016)$

We store the full subtrees here!



Grammar:

S → NP VP

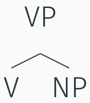
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP



$\langle \mathbf{t}, \mathbf{r}_1 \rangle$

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$ $\quad$ $\mathbf{r}_1 = (0.005, 0.043, \ldots, 0.016)$

DISTRIBUTED CYK

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

	VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

	VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

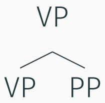
NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

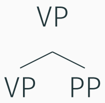


$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$ $\quad$ $\mathbf{r}_2 = (0.001, 0.008, \ldots, 0.024)$

Grammar:

S → NP VP
NP → DET N
NP → NP PP
PP → P NP
VP → V NP
VP → VP PP

	VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope



$\langle \mathbf{t}, \mathbf{r}_2 \rangle$

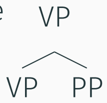
$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$

$\mathbf{r}_2 = (0.001, 0.008, \ldots, 0.024)$

Grammar:

S → NP VP
NP → DET N
NP → NP PP
PP → P NP
VP → V NP
VP → VP PP

	VP,VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope



$\langle \mathbf{t}, \mathbf{r}_2 \rangle$

$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$ $\mathbf{r}_2 = (0.001, 0.008, \ldots, 0.024)$

$\langle \mathbf{t}_1, \mathbf{t} \rangle$

$\langle \mathbf{t}_2, \mathbf{t} \rangle$

	VP,VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

S						
	VP,VP					
		NP				
S						
	VP			PP		
		NP			NP	
NP	V	DET	N	P	DET	N
I	saw	the	man	with	the	telescope

Grammar:

S → NP VP

NP → DET N

NP → NP PP

PP → P NP

VP → V NP

VP → VP PP

$$t \rightarrow \mathbf{t} = (0.011, 0.212, \ldots, 0.008) \in \mathbb{R}^d$$

EXPERIMENTS AND RESULTS

Dataset

Wall Street Journal sections of PennTree Bank:

- Sections 1~23: Grammar extraction
- Section 24: testing

Experimental pipeline

Experimental pipeline

- Parse the dataset and binarize the trees

Experimental pipeline

- Parse the dataset and binarize the trees
- Extract grammar from training set
    - The set of all rules in the grammar, no probabilities learned

Experimental pipeline

- Parse the dataset and binarize the trees
- Extract grammar from training set
    - The set of all rules in the grammar, no probabilities learned
- On test set (1346 sentences):
    - Compute the distributed vector $\mathbf{t}$

Experimental pipeline

- Parse the dataset and binarize the trees
- Extract grammar from training set
    - The set of all rules in the grammar, no probabilities learned
- On test set (1346 sentences):
    - Compute the distributed vector $\mathbf{t}$
    - use $\mathbf{t}$ to parse the sentence

Experimental pipeline

- Parse the dataset and binarize the trees
- Extract grammar from training set
    - The set of all rules in the grammar, no probabilities learned
- On test set (1346 sentences):
    - Compute the distributed vector $\mathbf{t}$
    - use $\mathbf{t}$ to parse the sentence
    - compare the result with the correct tree

Parameters

Parameters

- d: Dimension of the vector representation

Parameters

- d: Dimension of the vector representation
- k: number of partial trees kept in each cell
    - we only report $k = 2$

Results

- Number of exactly reconstructed trees;
- (Labelled) precision, recall and f-measure;

1024	2048	4096	8192	16384
23.5%	52.32%	75.58%	87.5%	**92.79%**

Table 1: Percentage of exactly reconstructed sentence

	1024	2048	4096	8192	16384
precision	0.71	0.85	0.951	0.99	**0.994**
recall	0.477	0.78	0.929	0.967	**0.976**
f-measure	0.57	0.81	0.939	0.974	**0.984**

Table 2: Precision, recall and F-measure

SUMMARY AND FUTURE WORK

Summary

- We showed that distributed trees are able to reconstruct almost always the original symbolic tree

Summary

- We showed that distributed trees are able to reconstruct almost always the original symbolic tree

Future work

- Expand the experimental setting:
    - from CNF to general grammars

Summary

- We showed that distributed trees are able to reconstruct almost always the original symbolic tree

Future work

- Expand the experimental setting:
  - from CNF to general grammars
- Use the reconstruction method on other distributed representations

Questions?