Parsing

CYK Algorithm

CYK Algorithm

Cocke-Younger-Kasami Algorithm is a dynamic programming parsing algorithm.

Follows bottom up parsing approach.

Builds a parse tree incrementally.

Each entry in the table is based on previous entries.

The process is iterated until the entire sentence is parsed.

Algo. Assumes the sentence to be in CNF

CYK Algorithm

Chomsky Normal Form

- A -> BC
- A -> w, where w is a word

Algorithm

Let w=w1 w2 w3 wi …… wj ….. wn

and wij = wi….wi+j-1

//initialization step

For i=1to n do

for all rules A → wi do

chart [i,1] = {A}

//recursive step

For j=2 to n do

for i=1 to n-j+1 do

begin

chart [i,j] = ϕ

for k=1 to j-1 do

chart [i,j] = chart [i,j] $\cup$ {A|A → BC is a production

and

B $\in$ chart [i,k] and C $\in$ chart [i+k,j-k]}

end

If S $\in$ chart [1,n] then accept else reject

CYK Algorithm

- TO get a better understanding of the whole idea, we work out an example.
- Consider the following grammar

- S → NP VP
- VP → Verb NP
- NP → Det Noun
- Det → an | then
- Verb → wrote
- Noun → girl
- Noun → easy

CYK Example

- The sentence to be parsed is

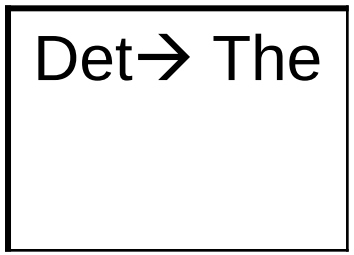
  The girl wrote an essay.

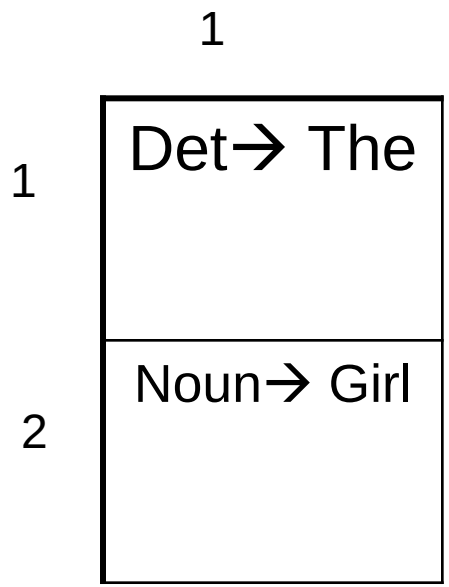
- Since we are using a bottom up approach

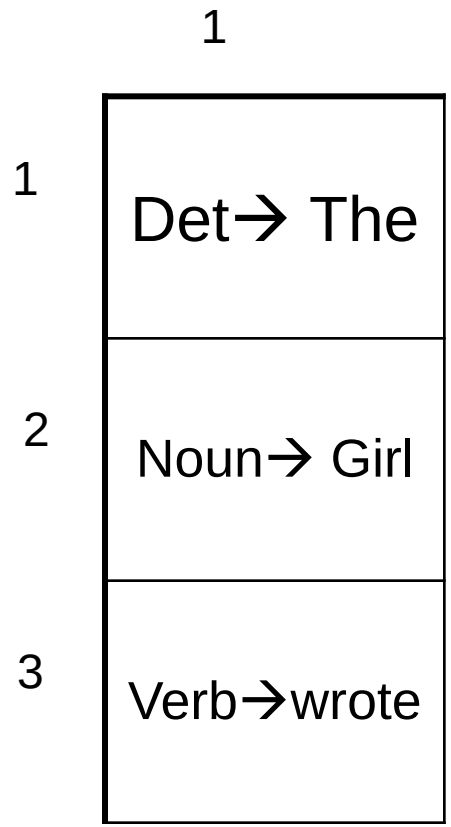
- We start from terminal nodes

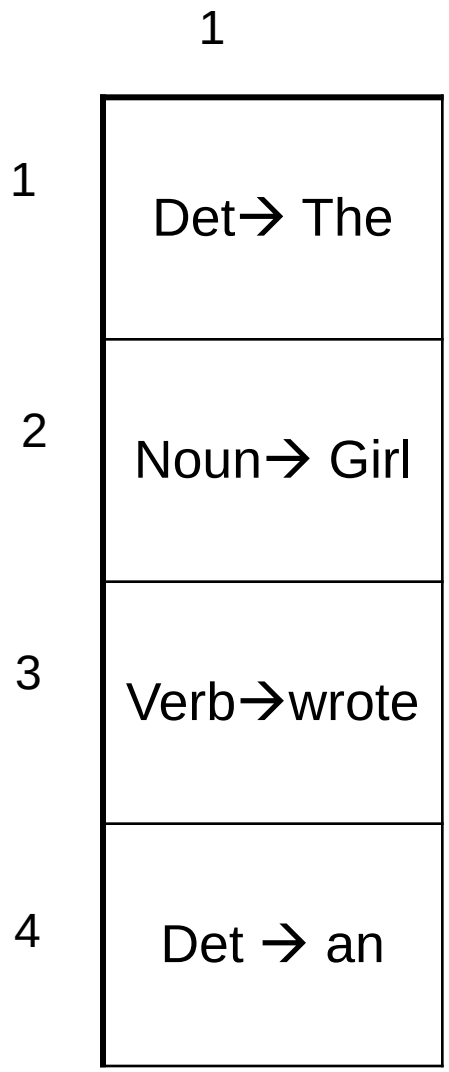
1

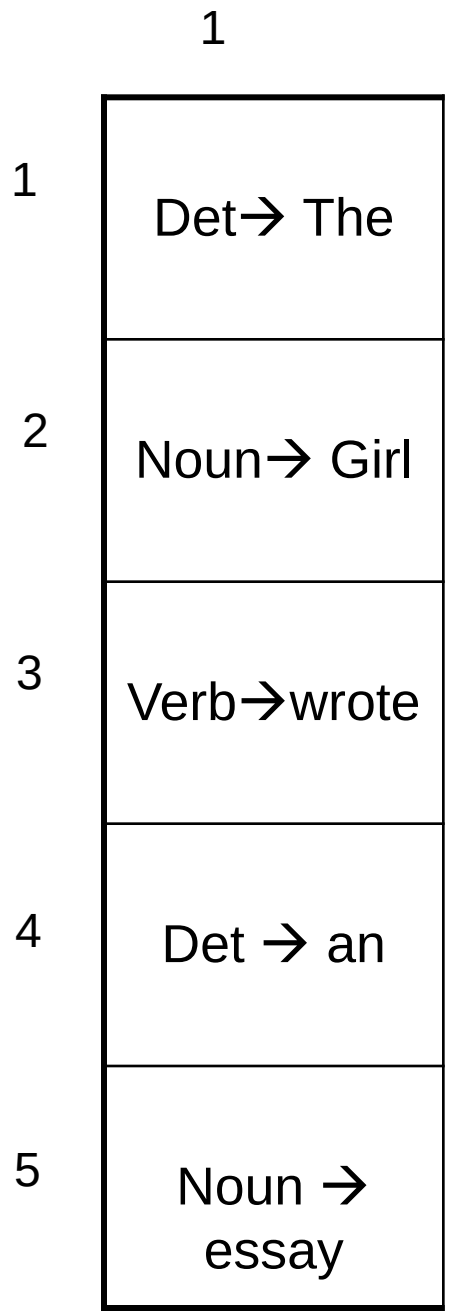
1




1
1
Det→ The
2
Noun→ Girl


1
1
Det→ The
2
Noun→ Girl
3
Verb→wrote


1
1
Det→ The
2
Noun→ Girl
3
Verb→wrote
4
Det → an


1
1
Det→ The
2
Noun→ Girl
3
Verb→wrote
4
Det → an
5
Noun → essay

	1	2
1	Det→ The	NP → Det Noun
2	Noun→ Girl	
3	Verb→wrote	
4	Det → an	
5	Noun → essay	

	1	2
1	Det→ The	NP → Det Noun
2	Noun→ Girl	
3	Verb→wrote	
4	Det → an	NP → Det Noun
5	Noun → essay	

	1	2	3
1	Det→ The	NP → Det Noun	
2	Noun→ Girl		
3	Verb→wrote		VP → Verb NP
4	Det → an	NP → Det Noun	
5	Noun → essay		

	1	2	3	4
1	Det→ The	NP → Det Noun		
2	Noun→ Girl			
3	Verb→wrote		VP → Verb NP	
4	Det → an	NP → Det Noun		
5	Noun → essay			

	1	2	3	4	5
1	Det→ The	NP → Det Noun			S→ NPVP
2	Noun→ Girl				
3	Verb→wrote		VP → Verb NP		
4	Det → an	NP → Det Noun			
5	Noun → essay				

CYK Example

- The entry in the [1,n]th cell contains a start symbol which indicates that S* ➔ w1n
- i.e. the parse is successful
- It is possible to have multiple entries in the cell.
- This shows multiple parsed trees are produced.
- Ambiguity has occurred and needs to be resolved.