

# Modèles de Langage et Analyse Syntaxique

## Cours 3 - Algorithmes d'analyse syntaxique - grammaires hors-contexte

Antoine Rozenknop  
antoine.rozenknop@lipn.univ-paris13.fr

21 octobre 2010

# Plan

## Algorithme Cock -Younger-Kasami

Principes de l'algorithme CYK

Complexité pire cas

## Algorithme de Earley

Fondements

Interprétation

Un exemple très simple

# Analyse syntaxique

Intérêt des CFG : existence d'algorithmes efficaces de production de dérivations/arbres d'analyse.

**efficace** — complexité pire cas  $\mathcal{O}(n^3)$

Les deux algorithmes les plus usuels :

- ▶ algorithme **CYK** (Cocke-Younger-Kasami) (milieu des années 60)
- ▶ algorithme **Earley** (1970)

# Plan

Algorithme Cock -Younger-Kasami  
Principes de l'algorithme CYK  
Complexité pire cas

Algorithme de Earley  
Fondements  
Interprétation  
Un exemple très simple

# Algorithme CYK

L'algorithme CYK est un algorithme **tabulaire ascendant** présentant deux avantages :

- ▶ sa complexité pire cas est optimale ( $\mathcal{O}(n^3)$ );
- ▶ très simple à comprendre et implémenter

Il présente cependant deux inconvénients :

- ▶ Nécessite la mise de la CFG sous forme normale de Chomsky
- ▶ Reste  $\mathcal{O}(n^3)$  sur des grammaires de complexité inférieure (par ex. régulières)

## Algorithme CYK : principes

CYK est un analyseur tabulaire (« **chart-parser** ») : calcul efficace (en espace et en temps) de **toutes les interprétations** possibles de **toutes les sous-séquences** de la séquence à analyser).

L'efficacité du calcul est fondée sur la propriété suivante :

*Si la grammaire est sous forme normale, le calcul des interprétations d'une séquence  $W$  de longueur  $l$  nécessite seulement l'exploration de toutes les décompositions de  $W$  en deux sous-séquences exactement.*

*Le nombre de paires de sous-séquences à explorer est donc  $l - 1$ .*

Idee : mettre les analyses des sous-séquences dans une table.

## Algorithme CYK : principes (2)

L'analyse syntaxique d'une séquence de  $n$  mots  $W = w_1 \dots w_n$  est représentée dans une table triangulaire de cellules  $C_{i,l}$  ( $1 \leq i \leq n$ ,  $1 \leq l \leq n - i + 1$ ).

La cellule  $C_{i,l}$  contient **tous les non-terminaux pouvant dériver** la sous-séquence  $w_i \dots w_{i+l-1}$  (la séquence de  $l$  mots commençant au  $i^{\text{e}}$  mot de  $W$ ).

Remplissage de bas en haut

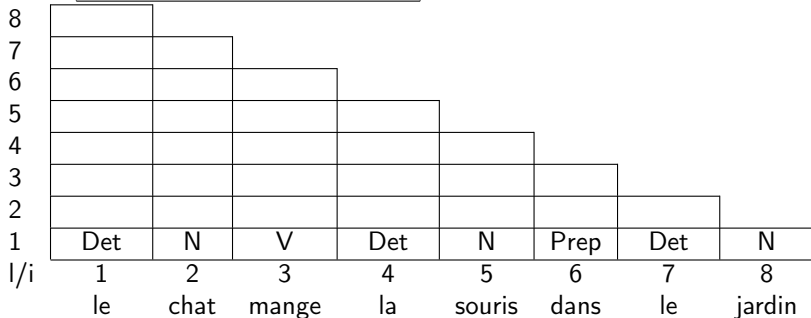
On reconsidère la grammaire mise sous CNF :

Règles de la grammaire originale	Grammaire sous forme normale
R1 : $S \rightarrow SN SV$	R1.1 : $S \rightarrow SN SV$
	R1.2 : $S \rightarrow SN V$
R2 : $SN \rightarrow Det N$	R2 : $SN \rightarrow Det N$
R3 : $SN \rightarrow Det N SNP$	R3.1 : $SN \rightarrow X_1 SNP$
	R3.2 : $X_1 \rightarrow Det N$
R4 : $SNP \rightarrow Prep SN$	R4 : $SNP \rightarrow Prep SN$
R5 : $SV \rightarrow V$	
R6 : $SV \rightarrow V SN$	R6 : $SV \rightarrow V SN$
R7 : $SV \rightarrow V SN SNP$	R7.1 : $SV \rightarrow X_2 SNP$
	R7.2 : $X_2 \rightarrow V SN$



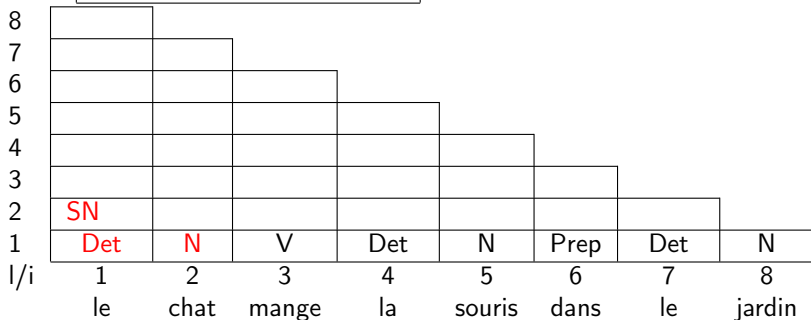
R1.1 :	$S$	$\rightarrow$	$SN SV$
R1.2 :	$S$	$\rightarrow$	$SN V$
R2 :	$SN$	$\rightarrow$	$Det N$
R3.1 :	$SN$	$\rightarrow$	$X_1 SNP$
R3.2 :	$X_1$	$\rightarrow$	$Det N$

R4 :	$SNP$	$\rightarrow$	$Prep SN$
R6 :	$SV$	$\rightarrow$	$V SN$
R7.1 :	$SV$	$\rightarrow$	$X_2 SNP$
R7.2 :	$X_2$	$\rightarrow$	$V SN$



R1.1 :	<i>S</i>	→	<i>SN SV</i>
R1.2 :	<i>S</i>	→	<i>SN V</i>
R2 :	<i>SN</i>	→	<i>Det N</i>
R3.1 :	<i>SN</i>	→	<i>X<sub>1</sub> SNP</i>
R3.2 :	<i>X<sub>1</sub></i>	→	<i>Det N</i>

R4 :	<i>SNP</i>	→	<i>Prep SN</i>
R6 :	<i>SV</i>	→	<i>V SN</i>
R7.1 :	<i>SV</i>	→	<i>X<sub>2</sub> SNP</i>
R7.2 :	<i>X<sub>2</sub></i>	→	<i>V SN</i>



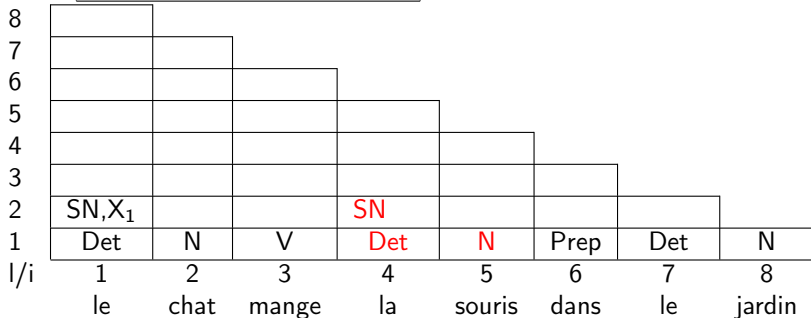
R1.1 :  $S \rightarrow SN SV$   
 R1.2 :  $S \rightarrow SN V$   
 R2 :  $SN \rightarrow Det N$   
 R3.1 :  $SN \rightarrow X_1 SNP$   
 R3.2 :  $X_1 \rightarrow Det N$

R4 :  $SNP \rightarrow Prep SN$   
 R6 :  $SV \rightarrow V SN$   
 R7.1 :  $SV \rightarrow X_2 SNP$   
 R7.2 :  $X_2 \rightarrow V SN$

8								
7								
6								
5								
4								
3								
2	SN, X <sub>1</sub>							
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

R1.1 :	<i>S</i>	→	<i>SN SV</i>
R1.2 :	<i>S</i>	→	<i>SN V</i>
R2 :	<i>SN</i>	→	<i>Det N</i>
R3.1 :	<i>SN</i>	→	<i>X<sub>1</sub> SNP</i>
R3.2 :	<i>X<sub>1</sub></i>	→	<i>Det N</i>

R4 :	<i>SNP</i>	→	<i>Prep SN</i>
R6 :	<i>SV</i>	→	<i>V SN</i>
R7.1 :	<i>SV</i>	→	<i>X<sub>2</sub> SNP</i>
R7.2 :	<i>X<sub>2</sub></i>	→	<i>V SN</i>



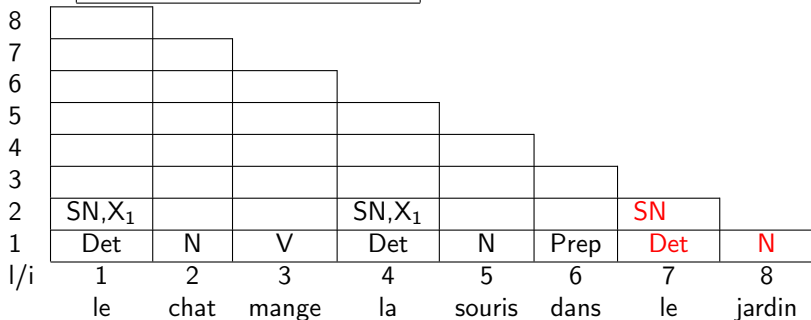
R1.1 :  $S \rightarrow SN SV$   
 R1.2 :  $S \rightarrow SN V$   
 R2 :  $SN \rightarrow Det N$   
 R3.1 :  $SN \rightarrow X_1 SNP$   
 R3.2 :  $X_1 \rightarrow Det N$

R4 :  $SNP \rightarrow Prep SN$   
 R6 :  $SV \rightarrow V SN$   
 R7.1 :  $SV \rightarrow X_2 SNP$   
 R7.2 :  $X_2 \rightarrow V SN$

8								
7								
6								
5								
4								
3								
2	SN, X <sub>1</sub>			SN, X <sub>1</sub>				
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

R1.1 :	$S$	$\rightarrow$	$SN SV$
R1.2 :	$S$	$\rightarrow$	$SN V$
R2 :	$SN$	$\rightarrow$	$Det N$
R3.1 :	$SN$	$\rightarrow$	$X_1 SNP$
R3.2 :	$X_1$	$\rightarrow$	$Det N$

R4 :	$SNP$	$\rightarrow$	$Prep SN$
R6 :	$SV$	$\rightarrow$	$V SN$
R7.1 :	$SV$	$\rightarrow$	$X_2 SNP$
R7.2 :	$X_2$	$\rightarrow$	$V SN$



R1.1 :  $S \rightarrow SN SV$   
 R1.2 :  $S \rightarrow SN V$   
 R2 :  $SN \rightarrow Det N$   
 R3.1 :  $SN \rightarrow X_1 SNP$   
 R3.2 :  $X_1 \rightarrow Det N$

R4 :  $SNP \rightarrow Prep SN$   
 R6 :  $SV \rightarrow V SN$   
 R7.1 :  $SV \rightarrow X_2 SNP$   
 R7.2 :  $X_2 \rightarrow V SN$

8								
7								
6								
5								
4								
3								
2	SN, X <sub>1</sub>			SN, X <sub>1</sub>			SN, X <sub>1</sub>	
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

R1.1 :  $S \rightarrow SN SV$   
**R1.2** :  $S \rightarrow SN V$   
 R2 :  $SN \rightarrow Det N$   
 R3.1 :  $SN \rightarrow X_1 SNP$   
 R3.2 :  $X_1 \rightarrow Det N$

R4 :  $SNP \rightarrow Prep SN$   
 R6 :  $SV \rightarrow V SN$   
 R7.1 :  $SV \rightarrow X_2 SNP$   
 R7.2 :  $X_2 \rightarrow V SN$

8								
7								
6								
5								
4								
3	<b>S</b>							
2	<b>SN, X<sub>1</sub></b>			<b>SN, X<sub>1</sub></b>			<b>SN, X<sub>1</sub></b>	
1	Det	N	<b>V</b>	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin



R1.1 :	$S$	$\rightarrow$	$SN SV$
R1.2 :	$S$	$\rightarrow$	$SN V$
R2 :	$SN$	$\rightarrow$	$Det N$
R3.1 :	$SN$	$\rightarrow$	$X_1 SNP$
R3.2 :	$X_1$	$\rightarrow$	$Det N$

R4 :	$SNP$	$\rightarrow$	$Prep SN$
R6 :	$SV$	$\rightarrow$	$V SN$
R7.1 :	$SV$	$\rightarrow$	$X_2 SNP$
R7.2 :	$X_2$	$\rightarrow$	$V SN$

8								
7								
6								
5								
4								
3	$S$		$SV$					
2	$SN, X_1$			$SN, X_1$			$SN, X_1$	
1	$Det$	$N$	$V$	$Det$	$N$	$Prep$	$Det$	$N$
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

R1.1 :	$S$	$\rightarrow$	$SN SV$
R1.2 :	$S$	$\rightarrow$	$SN V$
R2 :	$SN$	$\rightarrow$	$Det N$
R3.1 :	$SN$	$\rightarrow$	$X_1 SNP$
R3.2 :	$X_1$	$\rightarrow$	$Det N$

R4 :	$SNP$	$\rightarrow$	$Prep SN$
R6 :	$SV$	$\rightarrow$	$V SN$
R7.1 :	$SV$	$\rightarrow$	$X_2 SNP$
R7.2 :	$X_2$	$\rightarrow$	$V SN$

8								
7								
6								
5								
4								
3	S		SV, X <sub>2</sub>					
2	SN, X <sub>1</sub>			SN, X <sub>1</sub>			SN, X <sub>1</sub>	
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

R1.1 :	$S$	$\rightarrow$	$SN SV$
R1.2 :	$S$	$\rightarrow$	$SN V$
R2 :	$SN$	$\rightarrow$	$Det N$
R3.1 :	$SN$	$\rightarrow$	$X_1 SNP$
R3.2 :	$X_1$	$\rightarrow$	$Det N$

R4 :	$SNP$	$\rightarrow$	$Prep SN$
R6 :	$SV$	$\rightarrow$	$V SN$
R7.1 :	$SV$	$\rightarrow$	$X_2 SNP$
R7.2 :	$X_2$	$\rightarrow$	$V SN$

8								
7								
6								
5								
4								
3	S		SV, X <sub>2</sub>			SNP		
2	SN, X <sub>1</sub>			SN, X <sub>1</sub>			SN, X <sub>1</sub>	
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

R1.1 :  $S \rightarrow SN SV$   
 R1.2 :  $S \rightarrow SN V$   
 R2 :  $SN \rightarrow Det N$   
 R3.1 :  $SN \rightarrow X_1 SNP$   
 R3.2 :  $X_1 \rightarrow Det N$

R4 :  $SNP \rightarrow Prep SN$   
 R6 :  $SV \rightarrow V SN$   
 R7.1 :  $SV \rightarrow X_2 SNP$   
 R7.2 :  $X_2 \rightarrow V SN$

8								
7								
6								
5	S							
4								
3	S		SV, X <sub>2</sub>			SNP		
2	SN, X <sub>1</sub>			SN, X <sub>1</sub>			SN, X <sub>1</sub>	
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

R1.1 :	$S$	$\rightarrow$	$SN SV$
R1.2 :	$S$	$\rightarrow$	$SN V$
R2 :	$SN$	$\rightarrow$	$Det N$
R3.1 :	$SN$	$\rightarrow$	$X_1 SNP$
R3.2 :	$X_1$	$\rightarrow$	$Det N$

R4 :	$SNP$	$\rightarrow$	$Prep SN$
R6 :	$SV$	$\rightarrow$	$V SN$
R7.1 :	$SV$	$\rightarrow$	$X_2 SNP$
R7.2 :	$X_2$	$\rightarrow$	$V SN$

8								
7								
6								
5	S			SN				
4								
3	S		SV, X <sub>2</sub>			SNP		
2	SN, X <sub>1</sub>			SN, X <sub>1</sub>			SN, X <sub>1</sub>	
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

R1.1 :	$S$	$\rightarrow$	$SN SV$
R1.2 :	$S$	$\rightarrow$	$SN V$
R2 :	$SN$	$\rightarrow$	$Det N$
R3.1 :	$SN$	$\rightarrow$	$X_1 SNP$
R3.2 :	$X_1$	$\rightarrow$	$Det N$

R4 :	$SNP$	$\rightarrow$	$Prep SN$
R6 :	$SV$	$\rightarrow$	$V SN$
R7.1 :	$SV$	$\rightarrow$	$X_2 SNP$
R7.2 :	$X_2$	$\rightarrow$	$V SN$

8								
7								
6			SV					
5	S			SN				
4								
3	S		SV, X <sub>2</sub>			SNP		
2	SN, X <sub>1</sub>			SN, X <sub>1</sub>			SN, X <sub>1</sub>	
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

R1.1 :	$S$	$\rightarrow$	$SN SV$
R1.2 :	$S$	$\rightarrow$	$SN V$
R2 :	$SN$	$\rightarrow$	$Det N$
R3.1 :	$SN$	$\rightarrow$	$X_1 SNP$
R3.2 :	$X_1$	$\rightarrow$	$Det N$

R4 :	$SNP$	$\rightarrow$	$Prep SN$
R6 :	$SV$	$\rightarrow$	$V SN$
R7.1 :	$SV$	$\rightarrow$	$X_2 SNP$
R7.2 :	$X_2$	$\rightarrow$	$V SN$

8								
7								
6			SV					
5	S			SN				
4								
3	S		SV, X <sub>2</sub>			SNP		
2	SN, X <sub>1</sub>			SN, X <sub>1</sub>			SN, X <sub>1</sub>	
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

R1.1 :	$S$	$\rightarrow$	$SN SV$
R1.2 :	$S$	$\rightarrow$	$SN V$
R2 :	$SN$	$\rightarrow$	$Det N$
R3.1 :	$SN$	$\rightarrow$	$X_1 SNP$
R3.2 :	$X_1$	$\rightarrow$	$Det N$

R4 :	$SNP$	$\rightarrow$	$Prep SN$
R6 :	$SV$	$\rightarrow$	$V SN$
R7.1 :	$SV$	$\rightarrow$	$X_2 SNP$
R7.2 :	$X_2$	$\rightarrow$	$V SN$

8								
7								
6			SV, X <sub>2</sub>					
5	S			SN				
4								
3	S		SV, X <sub>2</sub>			SNP		
2	SN, X <sub>1</sub>			SN, X <sub>1</sub>			SN, X <sub>1</sub>	
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin



R1.1 :  $S \rightarrow SN SV$   
 R1.2 :  $S \rightarrow SN V$   
 R2 :  $SN \rightarrow Det N$   
 R3.1 :  $SN \rightarrow X_1 SNP$   
 R3.2 :  $X_1 \rightarrow Det N$

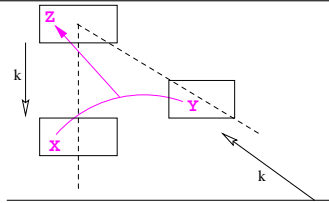
R4 :  $SNP \rightarrow Prep SN$   
 R6 :  $SV \rightarrow V SN$   
 R7.1 :  $SV \rightarrow X_2 SNP$   
 R7.2 :  $X_2 \rightarrow V SN$

8	<b>S</b>							
7								
6			<b>SV, X<sub>2</sub></b>					
5	<b>S</b>			<b>SN</b>				
4								
3	<b>S</b>			<b>SV, X<sub>2</sub></b>			<b>SNP</b>	
2	<b>SN, X<sub>1</sub></b>			<b>SN, X<sub>1</sub></b>			<b>SN, X<sub>1</sub></b>	
1	Det	N	V	Det	N	Prep	Det	N
l/i	1	2	3	4	5	6	7	8
	le	chat	mange	la	souris	dans	le	jardin

# Algorithme formel

```

pour tout  $2 \leq i \leq n$  (ligne) faire
  pour tout  $1 \leq j \leq n - i + 1$  (colonne) faire
    pour tout  $1 \leq k \leq i - 1$  (décomposition) faire
      pour tout  $X \in C[i - k][j]$  faire
        pour tout  $Y \in C[k][j + i - k]$  faire
          pour tout  $Z \rightarrow X Y \in \mathcal{R}$  faire
            Ajouter  $Z$  à  $C[i][j]$ 
  
```



# Plan

## Algorithme Cock -Younger-Kasami

Principes de l'algorithme CYK

Complexité pire cas

## Algorithme de Earley

Fondements

Interprétation

Un exemple très simple

## Complexité (1)

- Le calcul des interprétations d'une cellule  $C_{i,j}$  nécessite  $(i - 1)$  explorations de paires de cellules ( $1 \leq k \leq i - 1$ ), le nombre total d'explorations est donc :

$$\sum_{i=2}^n \sum_{j=1}^{n-i+1} (i - 1) = \sum_{i=1}^n (n - i + 1) \cdot (i - 1) \sim \mathcal{O}(n^3).$$

## Complexité (1)

- Le calcul des interprétations d'une cellule  $C_{i,j}$  nécessite  $(i-1)$  explorations de paires de cellules ( $1 \leq k \leq i-1$ ), le nombre total d'explorations est donc :

$$\sum_{i=2}^n \sum_{j=1}^{n-i+1} (i-1) = \sum_{i=1}^n (n-i+1) \cdot (i-1) \sim \mathcal{O}(n^3).$$

- Une cellule contient au plus  $|\mathcal{NT}|$  interprétations (nombre de non-terminaux de la grammaire)

## Complexité (1)

- ▶ Le calcul des interprétations d'une cellule  $C_{i,j}$  nécessite  $(i - 1)$  explorations de paires de cellules ( $1 \leq k \leq i - 1$ ), le nombre total d'explorations est donc :

$$\sum_{i=2}^n \sum_{j=1}^{n-i+1} (i - 1) = \sum_{i=1}^n (n - i + 1) \cdot (i - 1) \sim \mathcal{O}(n^3).$$

- ▶ Une cellule contient au plus  $|\mathcal{NT}|$  interprétations (nombre de non-terminaux de la grammaire)
- ▶ le coût de l'exploration croisée d'une paire de cellules est d'au plus  $|\mathcal{NT}|^2$  accès à la grammaire.

## Complexité (2)

D'où la complexité :

$$\mathcal{O}(n^3 |\mathcal{NT}|^2).$$

Inconvénient de la forme normale, qui augmente la taille de  $\mathcal{NT}$ .

Remarque : une fois la table remplie, on peut extraire un arbre d'analyse en  $\mathcal{O}(n)$ .

# Algorithme de Earley

Algorithme **top-down** (prédicatif).

**3 avantages :**

- ▶ meilleure complexité pire cas connue (de même que CYK) ;
- ▶ s'adapte à des langages moins complexes (par ex. les langages réguliers) ;
- ▶ ne requière pas une forme particulière de grammaire (pas de CNF).

**deux inconvénients :**

- ▶ pas très intuitif ;
- ▶ pas moyen de corriger / reconstruire des phrases incorrectes syntaxiquement, ni d'en donner des analyses partielles.



# Plan

Algorithme Cock -Younger-Kasami  
Principes de l'algorithme CYK  
Complexité pire cas

Algorithme de Earley  
Fondements  
Interprétation  
Un exemple très simple

## Algorithme de Earley (2)

Idée : binarisation dynamique (en cours d'analyse) de la grammaire

Règle pointée : Objet de la forme

$$X \rightarrow X_1 \dots X_k \bullet X_{k+1} \dots X_m$$

où  $X \rightarrow X_1 \dots X_m$  est une règle de la grammaire.

Earley item : une règle pointée et un entier  $i$  (avec  $0 \leq i \leq n$ ,  $n$  : taille de la phrase).

☞ représente la portion d'une règle compatible avec le segment de phrase commençant à la position  $i + 1$ .

Exemple :  $(SV \rightarrow V \bullet SN, 2)$  et  $(SV \rightarrow V \bullet SN \text{ SNP}, 2)$  sont des items pour la chaîne initiale :

le	chat	mangea	la	souris.
1	2	3	4	5

## Algorithme de Earley (3)

Principe de l'algorithme : Construction en parallèle de toutes les règles pointées compatibles avec des sous-chaînes de plus en plus grandes de la phrase à analyser.

☞ construction d'ensembles d'items  $E_j$  tels que :

$$(X \rightarrow \alpha \bullet \beta, i) \in E_j \iff \exists \gamma, \delta : \begin{cases} S \Rightarrow^* \gamma X \delta \\ \gamma \Rightarrow^* w_1 \dots w_i \\ \alpha \Rightarrow^* w_{i+1} \dots w_j \end{cases}$$

Exemple : dans l'exemple précédent,  $(SV \rightarrow V \bullet SN, 2) \in E_3$ .

☞ une phrase est **syntactiquement correcte** s'il existe au moins un  $(S \rightarrow X_1 \dots X_m \bullet, 0)$  dans  $E_n$

## Algorithme d'Earley (4)

### ❖ Initialisation : Construction de $E_0$

1. Pour toute règle  $S \rightarrow X_1 \dots X_n$  :  
ajouter  $(S \rightarrow \bullet X_1 \dots X_n, 0)$  à  $E_0$ .
2. Pour tous les items  $(X \rightarrow \bullet Y \beta, 0)$  dans  $E_0$ , et pour toute règle  $Y \rightarrow \gamma$  :  
ajouter  $(Y \rightarrow \bullet \gamma, 0)$  à  $E_0$ .
3. Répéter (2) jusqu'à stabilisation de  $E_0$ .

## Algorithme d'Earley (5)

✚ **Itérations** : Construction des  $E_j$  ( $1 \leq j \leq n$ ) :

1. Pour tout  $(X \rightarrow \alpha \bullet w_j \beta, i) \in E_{j-1}$  :  
ajouter  $(X \rightarrow \alpha w_j \bullet \beta, i)$  à  $E_j$ .

## Algorithme d'Earley (5)

✚ **Itérations** : Construction des  $E_j$  ( $1 \leq j \leq n$ ) :

1. Pour tout  $(X \rightarrow \alpha \bullet w_j \beta, i) \in E_{j-1}$  :  
ajouter  $(X \rightarrow \alpha w_j \bullet \beta, i)$  à  $E_j$ .
2. Pour tout  $(X \rightarrow \gamma \bullet, i)$  de  $E_j$   
et tout  $(Y \rightarrow \alpha \bullet X \beta, k)$  de  $E_i$  :  
ajouter  $(Y \rightarrow \alpha X \bullet \beta, k)$  à  $E_j$ .

## Algorithme d'Earley (5)

✚ **Itérations** : Construction des  $E_j$  ( $1 \leq j \leq n$ ) :

1. Pour tout  $(X \rightarrow \alpha \bullet w_j \beta, i) \in E_{j-1}$  :  
 ajouter  $(X \rightarrow \alpha w_j \bullet \beta, i)$  à  $E_j$ .
2. Pour tout  $(X \rightarrow \gamma \bullet, i)$  de  $E_j$   
 et tout  $(Y \rightarrow \alpha \bullet X \beta, k)$  de  $E_i$  :  
 ajouter  $(Y \rightarrow \alpha X \bullet \beta, k)$  à  $E_j$ .
3. Pour tous les  $(X \rightarrow \alpha \bullet Y \beta, i) \in E_j$   
 et toute règle  $Y \rightarrow \gamma$  :  
 ajouter  $(Y \rightarrow \bullet \gamma, j)$  à  $E_j$ .

## Algorithme d'Earley (5)

✚ **Itérations** : Construction des  $E_j$  ( $1 \leq j \leq n$ ) :

1. Pour tout  $(X \rightarrow \alpha \bullet w_j \beta, i) \in E_{j-1}$  :  
 ajouter  $(X \rightarrow \alpha w_j \bullet \beta, i)$  à  $E_j$ .
2. Pour tout  $(X \rightarrow \gamma \bullet, i)$  de  $E_j$   
 et tout  $(Y \rightarrow \alpha \bullet X \beta, k)$  de  $E_i$  :  
 ajouter  $(Y \rightarrow \alpha X \bullet \beta, k)$  à  $E_j$ .
3. Pour tous les  $(X \rightarrow \alpha \bullet Y \beta, i) \in E_j$   
 et toute règle  $Y \rightarrow \gamma$  :  
 ajouter  $(Y \rightarrow \bullet \gamma, j)$  à  $E_j$ .
4. Répéter (2) et (3) jusqu'à stabilisation de  $E_j$ .



# Plan

## Algorithme Cock -Younger-Kasami

Principes de l'algorithme CYK

Complexité pire cas

## Algorithme de Earley

Fondements

**Interprétation**

Un exemple très simple

## Algorithme d'Earley (Interprétation)

- + **Itérations** : Construction des interprétations de  $w_1 \dots w_n (E_j)$ 
  1. Ancrage vers les mots :  
 Introduire le mot  $w_j$  si on a une interprétation de  $w_1 \dots w_{j-1}$  qui peut « consommer »  $w_j$  (c.-à-d. « un  $\bullet$  avant le symbole  $w_j$  »)
  2. Progression dans l'interprétation :  
 Si le non-terminal  $X$  a été produit depuis  $w_{i+1}$  et si on a une interprétation terminant en  $w_i$  qui peut « consommer »  $X$ , alors le faire !
  3. Prédiction (des choses utiles) :  
 Si on cherche à consommer  $Y$  pour la suite de l'analyse, alors introduire toutes les règles susceptibles de produire (plus tard)  $Y$ .

# Plan

## Algorithme Cock -Younger-Kasami

Principes de l'algorithme CYK

Complexité pire cas

## Algorithme de Earley

Fondements

Interprétation

Un exemple très simple

## Algorithme d'Earley (Exemple complet)

analyse de la phrase « *Je pense* » avec la grammaire suivante :

$S \rightarrow SN SV$	$Pron \rightarrow Je$
$SN \rightarrow Pron$	$V \rightarrow pense$
$SN \rightarrow Det N$	
$SV \rightarrow V$	
$SV \rightarrow V S$	
$SV \rightarrow V SN$	

$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$(SN \rightarrow \bullet Pron, 0)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$(SN \rightarrow \bullet \text{Pron}, 0)$      $(SN \rightarrow \bullet \text{Det N}, 0)$

$(\text{Pron} \rightarrow \bullet \text{Je}, 0)$



$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$(SN \rightarrow \bullet Pron, 0)$

$(Pron \rightarrow \bullet Je, 0)$

$(SN \rightarrow \bullet Det N, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$(SN \rightarrow \bullet \text{Pron}, 0)$      $(SN \rightarrow \bullet \text{Det N}, 0)$

$(\text{Pron} \rightarrow \bullet \text{Je}, 0)$

$E_1$

$(\text{Pron} \rightarrow \text{Je} \bullet, 0)$

$(SN \rightarrow \text{Pron} \bullet, 0)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$

$(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$

$(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$

$(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$

$(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$

$(SV \rightarrow \bullet V, 1)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$(SN \rightarrow \bullet Pron, 0)$        $(SN \rightarrow \bullet Det N, 0)$

$(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$

$(SN \rightarrow Pron \bullet, 0)$        $(S \rightarrow SN \bullet SV, 0)$

$(SV \rightarrow \bullet V, 1)$        $(SV \rightarrow \bullet V S, 1)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$(SN \rightarrow \bullet Pron, 0)$

$(Pron \rightarrow \bullet Je, 0)$

$(SN \rightarrow \bullet Det N, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$

$(SN \rightarrow Pron \bullet, 0)$

$(SV \rightarrow \bullet V, 1)$

$(SV \rightarrow \bullet V SN, 1)$

$(S \rightarrow SN \bullet SV, 0)$

$(SV \rightarrow \bullet V S, 1)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$

$(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$

$(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$

$(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$

$(SV \rightarrow \bullet V, 1)$      $(SV \rightarrow \bullet V S, 1)$

$(SV \rightarrow \bullet V SN, 1)$      $(V \rightarrow \bullet pense, 1)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$   
 $(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$   
 $(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$   
 $(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$   
 $(SV \rightarrow \bullet V, 1)$      $(SV \rightarrow \bullet V S, 1)$   
 $(SV \rightarrow \bullet V SN, 1)$      $(V \rightarrow \bullet pense, 1)$

$E_2$

$(V \rightarrow pense \bullet, 1)$



$E_0$

$(S \rightarrow \bullet SN SV, 0)$   
 $(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$   
 $(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$   
 $(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$   
 $(SV \rightarrow \bullet V, 1)$      $(SV \rightarrow \bullet V S, 1)$   
 $(SV \rightarrow \bullet V SN, 1)$      $(V \rightarrow \bullet pense, 1)$

$E_2$

$(V \rightarrow pense \bullet, 1)$   
 $(SV \rightarrow V \bullet, 1)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$   
 $(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$   
 $(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$   
 $(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$   
 $(SV \rightarrow \bullet V, 1)$      $(SV \rightarrow \bullet V S, 1)$   
 $(SV \rightarrow \bullet V SN, 1)$      $(V \rightarrow \bullet pense, 1)$

$E_2$

$(V \rightarrow pense \bullet, 1)$   
 $(SV \rightarrow V \bullet, 1)$      $(SV \rightarrow V \bullet S, 1)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$   
 $(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$   
 $(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$   
 $(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$   
 $(SV \rightarrow \bullet V, 1)$      $(SV \rightarrow \bullet V S, 1)$   
 $(SV \rightarrow \bullet V SN, 1)$      $(V \rightarrow \bullet pense, 1)$

$E_2$

$(V \rightarrow pense \bullet, 1)$   
 $(SV \rightarrow V \bullet, 1)$      $(SV \rightarrow V \bullet S, 1)$   
 $(SV \rightarrow V \bullet SN, 1)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$   
 $(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$   
 $(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$   
 $(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$   
 $(SV \rightarrow \bullet V, 1)$      $(SV \rightarrow \bullet V S, 1)$   
 $(SV \rightarrow \bullet V SN, 1)$      $(V \rightarrow \bullet pense, 1)$

$E_2$

$(V \rightarrow pense \bullet, 1)$   
 $(SV \rightarrow V \bullet, 1)$      $(SV \rightarrow V \bullet S, 1)$   
 $(SV \rightarrow V \bullet SN, 1)$      $(S \rightarrow SN SV \bullet, 0)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$   
 $(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$   
 $(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$   
 $(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$   
 $(SV \rightarrow \bullet V, 1)$      $(SV \rightarrow \bullet V S, 1)$   
 $(SV \rightarrow \bullet V SN, 1)$      $(V \rightarrow \bullet pense, 1)$

$E_2$

$(V \rightarrow pense \bullet, 1)$   
 $(SV \rightarrow V \bullet, 1)$      $(SV \rightarrow V \bullet S, 1)$   
 $(SV \rightarrow V \bullet SN, 1)$      $(S \rightarrow SN SV \bullet, 0)$   
 $(S \rightarrow \bullet SN SV, 2)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$   
 $(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$   
 $(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$   
 $(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$   
 $(SV \rightarrow \bullet V, 1)$      $(SV \rightarrow \bullet V S, 1)$   
 $(SV \rightarrow \bullet V SN, 1)$      $(V \rightarrow \bullet pense, 1)$

$E_2$

$(V \rightarrow pense \bullet, 1)$   
 $(SV \rightarrow V \bullet, 1)$      $(SV \rightarrow V \bullet S, 1)$   
 $(SV \rightarrow V \bullet SN, 1)$      $(S \rightarrow SN SV \bullet, 0)$   
 $(S \rightarrow \bullet SN SV, 2)$      $(SN \rightarrow \bullet Pron, 2)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$   
 $(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$   
 $(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$   
 $(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$   
 $(SV \rightarrow \bullet V, 1)$      $(SV \rightarrow \bullet V S, 1)$   
 $(SV \rightarrow \bullet V SN, 1)$      $(V \rightarrow \bullet pense, 1)$

$E_2$

$(V \rightarrow pense \bullet, 1)$   
 $(SV \rightarrow V \bullet, 1)$      $(SV \rightarrow V \bullet S, 1)$   
 $(SV \rightarrow V \bullet SN, 1)$      $(S \rightarrow SN SV \bullet, 0)$   
 $(S \rightarrow \bullet SN SV, 2)$      $(SN \rightarrow \bullet Pron, 2)$   
 $(SN \rightarrow \bullet Det N, 2)$

$E_0$

$(S \rightarrow \bullet SN SV, 0)$   
 $(SN \rightarrow \bullet Pron, 0)$      $(SN \rightarrow \bullet Det N, 0)$   
 $(Pron \rightarrow \bullet Je, 0)$

$E_1$

$(Pron \rightarrow Je \bullet, 0)$   
 $(SN \rightarrow Pron \bullet, 0)$      $(S \rightarrow SN \bullet SV, 0)$   
 $(SV \rightarrow \bullet V, 1)$      $(SV \rightarrow \bullet V S, 1)$   
 $(SV \rightarrow \bullet V SN, 1)$      $(V \rightarrow \bullet pense, 1)$

$E_2$

$(V \rightarrow pense \bullet, 1)$   
 $(SV \rightarrow V \bullet, 1)$      $(SV \rightarrow V \bullet S, 1)$   
 $(SV \rightarrow V \bullet SN, 1)$      $(S \rightarrow SN SV \bullet, 0)$   
 $(S \rightarrow \bullet SN SV, 2)$      $(SN \rightarrow \bullet Pron, 2)$   
 $(SN \rightarrow \bullet Det N, 2)$      $(Pron \rightarrow \bullet Je, 2)$



$E_0$

(S  $\rightarrow$ • SN SV, 0)  
 (SN  $\rightarrow$ • Pron, 0) (SN  $\rightarrow$ • Det N, 0)  
 (Pron  $\rightarrow$ • Je, 0)

$E_1$

(Pron  $\rightarrow$ Je •, 0)  
 (SN  $\rightarrow$ Pron •, 0) (S  $\rightarrow$ SN• SV, 0)  
 (SV  $\rightarrow$ • V, 1) (SV  $\rightarrow$ • V S, 1)  
 (SV  $\rightarrow$ • V SN, 1) (V  $\rightarrow$ • pense, 1)

$E_2$

(V  $\rightarrow$ pense •, 1)  
 (SV  $\rightarrow$ V •, 1) (SV  $\rightarrow$ V • S, 1)  
 (SV  $\rightarrow$ V • SN, 1) (S  $\rightarrow$ SN SV •, 0)  
 (S  $\rightarrow$ • SN SV, 2) (SN  $\rightarrow$ • Pron, 2)  
 (SN  $\rightarrow$ • Det N, 2) (Pron  $\rightarrow$ • Je, 2)