# Grammaires formelles :
# Règles de réécriture et grammaires

Karën Fort

karen.fort@sorbonne-universite.fr / https://members.loria.fr/KFort/

LETTRES
SORBONNE
UNIVERSITÉ

# Quelques sources d'inspiration
## par ordre d'importance décroissant

- *Introduction à la calculabilité* (Pierre Wolper) – InterEditions, 1991
- cours de D. Battistelli (Paris 3), grâce aux notes de C. Riquier (Master 2, Paris 4)
- cours d'A. Rozenknop (Paris 13)
- cours en ligne de J-F. Perrot (Paris 6), avec son accord : http://pagesperso-systeme.lip6.fr/Jean-Francois.Perrot/inalco/Automates/Cours17.html

# Une « grammaire » innée

Comment sommes-nous capables de

- former des phrases jamais entendues auparavant ?
- savoir que telle phrase appartient à la langue (= est correcte) ?

**Selon Chomsky (qui a évolué sur le sujet)**

L'humain possède une compétence linguistique
= savoir linguistique implicite indépendant des facteurs qui peuvent venir influencer l'acte concret de parole (performance)

# Définition de la langue

La langue est pour Chomsky :

- un ensemble infini de phrases
- produit grâce à un ensemble fini d'éléments
- jugées grammaticales par les sujets parlants

# Projet de Chomsky

Mettre en lumière des règles formelles sous-jacentes à la langue

$\rightarrow$ comprendre le mécanisme :

- comment se combinent entre eux les différents constituants d'une phrase : ces combinaisons ne sont pas arbitraires
- structure de la phrase

Grammaires formelles = modèles théoriques qui acceptent ou rejettent une certaine suite d'éléments

# Exemple : exercice

Quelques règles permettant de construire des phrases du français :

- une `phrase` est de la forme `sujet verbe`
- un `sujet` est un `pronom`
- un `pronom` est **il** ou **elle**
- un `verbe` est **dort** ou **écoute**

Quelles sont les phrases que permet cet ensemble de règles ?

# Exemple : solution

1. il écoute
2. il dort
3. elle écoute
4. elle dort

# Grammaire et automate

Une grammaire :

- est un ensemble de règles (du type de celles présentées dans l'exemple précédent)
- donne une description générative d'un langage
  - $\rightarrow$ comment **construire** des éléments appartenant au langage

Un automate :

- donne une description analytique d'un langage :
  - $\rightarrow$ procédé pour **reconnaître** les éléments du langage

$\rightarrow$ complémentaires

# Grammaire et règles de réécriture (ou de production)

Quand peut-on décider qu'un assemblage de mots est une phrase ou pas ?

$\rightarrow$ on peut énoncer des règes qui permettent de distinguer des phrases grammaticales des phrases non grammaticales
= règles de grammaires assimilables à des règles de réécriture qui entrent dans un système formel

$\rightarrow$ construire la grammaire d'une langue donnée revient à formuler de façon explicite les règles d'un système de réécriture qui engendrent cette langue

# Formalisation des règles de production

$x \longrightarrow w$ ($x$ se réécrit en $w$)

où :

- $x$ et $w$ sont des chaînes sur $V$ (vocabulaire)
- $x$ ne peut être vide ($x \in V^+$)

# Exemple

$V = \{a, b, c\}$

$$P = \begin{cases} ba \longrightarrow ab & (1) \\ ca \longrightarrow ac & (2) \\ cb \longrightarrow bc & (3) \end{cases}$$

# Réécriture et dérivation

**Définition 1**

Une chaîne $u_1$ se dérive en une chaîne $u_2$ ($u_1 \Longrightarrow^* u_2$) si une succession de réécritures permet d'obtenir $u_2$ à partir de $u_1$.

**Définition 2**

Une dérivation est une succession de réécritures.

# Exercice 1

12675 est-il pair ou impair ?

$$V = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, I, P\}$$

$$P = \begin{cases} 0 \longrightarrow P & (1) \\ 1 \longrightarrow I & (2) \\ 2 \longrightarrow P & (3) \\ 3 \longrightarrow I & (4) \\ 4 \longrightarrow P & (5) \\ 5 \longrightarrow I & (6) \\ 6 \longrightarrow P & (7) \\ 7 \longrightarrow I & (8) \\ 8 \longrightarrow P & (9) \\ 9 \longrightarrow I & (10) \\ IP \longrightarrow P & (11) \\ PI \longrightarrow I & (12) \\ PP \longrightarrow P & (13) \\ II \longrightarrow I & (14) \end{cases}$$

# Correction de l'exercice 1

## 12675 est-il pair ou impair ?

$$12675 \Longrightarrow^* IPPII \xrightarrow{(11)} PPII \xrightarrow{(12)} II \xrightarrow{(14)} I$$

# Exercice 2 : à vous !

$V = \{(,),[,],\{,\},<,>\}$

Les règles de réécriture sont les suivantes :

$$P = \begin{cases} () \longrightarrow \varepsilon & (1) \\ [] \longrightarrow \varepsilon & (2) \\ \{\} \longrightarrow \varepsilon & (3) \\ <> \longrightarrow \varepsilon & (4) \end{cases}$$

L'ensemble des mots qui peuvent se réécrire en $\varepsilon$ est exactement l'ensemble des mots bien parenthésés

Décomposez le processus pour le mot "([]{<<>>}([(){<>}]))"

# Exercice 3 : à vous !

$$V = \{C, V, E, +\}$$

$$P = \begin{cases} CV \longrightarrow VC & (1) \\ VC \longrightarrow E & (2) \\ VEC \longrightarrow E & (3) \\ VE \longrightarrow +V & (4) \\ V + V \longrightarrow +V & (5) \\ EC \longrightarrow C+ & (6) \\ C + C \longrightarrow C+ & (7) \end{cases}$$

1. Donner la dérivation détaillée et la longueur de la dérivation de : $CVCVCV \Longrightarrow^* E$

# Notations

- $V$ : vocabulaire

# Notations

- $V$ : vocabulaire
- $V_T$ : vocabulaire terminal fini (c'est sur ce vocabulaire que sont formées les chaînes définies par la grammaire)

# Notations

- $V$ : vocabulaire
- $V_T$ : vocabulaire terminal fini (c'est sur ce vocabulaire que sont formées les chaînes définies par la grammaire)
- $V_N$ : vocabulaire non-terminal fini (vocabulaire auxiliaire, contenant les éléments du métalangage)

# Notations

- $V$ : vocabulaire
- $V_T$ : vocabulaire terminal fini (c'est sur ce vocabulaire que sont formées les chaînes définies par la grammaire)
- $V_N$ : vocabulaire non-terminal fini (vocabulaire auxiliaire, contenant les éléments du métalangage)

$\rightarrow$ $V = V_T \cup V_N$

# Notations

- $V$ : vocabulaire
- $V_T$ : vocabulaire terminal fini (c'est sur ce vocabulaire que sont formées les chaînes définies par la grammaire)
- $V_N$ : vocabulaire non-terminal fini (vocabulaire auxiliaire, contenant les éléments du métalangage)

$\rightarrow$ $V = V_T \cup V_N$

$\rightarrow$ $V_T \cap V_N = \emptyset$ (aucun élément en commun)

# Notations

- $V$ : vocabulaire
- $V_T$ : vocabulaire terminal fini (c'est sur ce vocabulaire que sont formées les chaînes définies par la grammaire)
- $V_N$ : vocabulaire non-terminal fini (vocabulaire auxiliaire, contenant les éléments du métalangage)

→ $V = V_T \cup V_N$

→ $V_T \cap V_N = \emptyset$ (aucun élément en commun)

- $P$ : axiome ou symbole de départ, unique (élément de $V_N$)

# Notations

- $V$ : vocabulaire
- $V_T$ : vocabulaire terminal fini (c'est sur ce vocabulaire que sont formées les chaînes définies par la grammaire)
- $V_N$ : vocabulaire non-terminal fini (vocabulaire auxiliaire, contenant les éléments du métalangage)

$\rightarrow$ $V = V_T \cup V_N$

$\rightarrow$ $V_T \cap V_N = \emptyset$ (aucun élément en commun)

- $P$ : axiome ou symbole de départ, unique (élément de $V_N$)
- $R$ : ensemble de règles de réécriture de la forme : $\alpha \longrightarrow \beta$ avec

# Notations

- $V$ : vocabulaire
- $V_T$ : vocabulaire terminal fini (c'est sur ce vocabulaire que sont formées les chaînes définies par la grammaire)
- $V_N$ : vocabulaire non-terminal fini (vocabulaire auxiliaire, contenant les éléments du métalangage)

$\rightarrow$ $V = V_T \cup V_N$

$\rightarrow$ $V_T \cap V_N = \emptyset$ (aucun élément en commun)

- $P$ : axiome ou symbole de départ, unique (élément de $V_N$)
- $R$ : ensemble de règles de réécriture de la forme : $\alpha \longrightarrow \beta$ avec
  - $\alpha, \beta \in V^*$

# Notations

- $V$ : vocabulaire
- $V_T$ : vocabulaire terminal fini (c'est sur ce vocabulaire que sont formées les chaînes définies par la grammaire)
- $V_N$ : vocabulaire non-terminal fini (vocabulaire auxiliaire, contenant les éléments du métalangage)

$\rightarrow$ $V = V_T \cup V_N$

$\rightarrow$ $V_T \cap V_N = \emptyset$ (aucun élément en commun)

- $P$ : axiome ou symbole de départ, unique (élément de $V_N$)
- $R$ : ensemble de règles de réécriture de la forme : $\alpha \longrightarrow \beta$ avec
  - $\alpha, \beta \in V^*$
  - $\alpha \neq \emptyset$

# Grammaire formelle

- $V$ : vocabulaire
- $V_T$ : vocabulaire terminal
- $V_N$ : vocabulaire non-terminal
- $P$ : axiome, ou symbole de départ (élément de $V_N$ )
- $R$ : ensemble de règles de réécriture de la forme : $\alpha \longrightarrow \beta$ avec
  - $\alpha, \beta \in V^*$
  - $\alpha \neq \emptyset$

**Définition**

$G = (V_N, V_T, R, P)$

# Détail des notations

- $V_T$ (vocabulaire terminal) : minuscules
- $V_N$ (vocabulaire non terminal) : majuscules
- $P$ (symbole de départ) est parfois noté $S$
- $\varepsilon$ note le mot vide

# Intuition

*« L'originalité des grammaires parmi les systèmes de réécriture est cette nécessité de "**chasser les non-terminaux**" dans le processus de génération des mots du langage »*

*[J-F. Perrot]*

# Exemple

- $V = \{S, A, B, a, b\}$
- $S$ : symbole de départ
- $V_T$ : $\{a, b\}$

$$P = \begin{cases} S \longrightarrow A & (1) \\ S \longrightarrow B & (2) \\ B \longrightarrow bB & (3) \\ A \longrightarrow aA & (4) \\ A \longrightarrow \varepsilon & (5) \\ B \longrightarrow \varepsilon & (6) \end{cases}$$

$aaaa$ fait-il partie du langage défini par cette grammaire ?

# Décomposons

Le mot *aaaa* fait partie du langage défini par cette grammaire et il est obtenu comme suit :

| | |
|---|---|
| $A$ | règle $S \longrightarrow A$ |
| $aA$ | $A \longrightarrow aA$ |
| $aaA$ | $A \longrightarrow aA$ |
| $aaaA$ | $A \longrightarrow aA$ |
| $aaaaA$ | $A \longrightarrow aA$ |
| $aaaa$ | $A \longrightarrow \varepsilon$ |

# Langage

**Définition**

Le langage généré par une grammaire $G$, dénoté $L(G)$ est l'ensemble des mots qui peuvent être générés par $G$.

**Équivalence**

On dit que deux grammaires $G_1$ et $G_2$ sont faiblement équivalentes si :
$L(G_1) = L(G_2)$

# Décidabilité

**Définition**

Un langage est décidable si pour toute phrase on peut savoir au bout d'un temps fini si elle appartient ou nom au langage.

Sources

Introduction

Règles de réécriture

Grammaire formelle

Pour finir

- CQFR : Ce Qu'il Faut Retenir
- TD


- théorie de Chomsky
- règles de réécriture :
  - définition
  - manipulation
- grammaires formelles :
  - définition
  - manipulation
- lien langage / grammaire

# Exercice 1 : à faire, à rendre à la fin du TD

Soit la grammaire $G_1$ définie par :
$V_N = \{A, S\}$
$V_T = \{a, b, c\}$

$$P = \begin{cases} S \longrightarrow AcA & (1) \\ A \longrightarrow a & (2) \\ A \longrightarrow b & (3) \end{cases}$$

$L(G_1)$ ?

# Exercice 2 : à faire, à rendre à la fin du TD

Soit la grammaire $G_2$ définie par :
$V_N = \{S\}$
$V_T = \{a, b, c\}$

$$P = \begin{cases} S \longrightarrow aca & (1) \\ S \longrightarrow acb & (2) \\ S \longrightarrow bcb & (3) \\ S \longrightarrow bca & (4) \end{cases}$$

Que pouvez-vous dire de $L(G_2)$ ? et de $G_2$ ?

# Exercice 3 : à faire, à rendre à la fin du TD

Soit la grammaire $G_3$ définie par :
$V_N = \{S, GN, GV, D, N, V\}$
$V_T = \{\text{le}, \text{la}, \text{gâteau}, \text{fille}, \text{mange}, \text{déguste}\}$

$$P = \begin{cases} S \longrightarrow \text{GN GV} & (1) \\ GV \longrightarrow \text{V GN} & (2) \\ GN \longrightarrow \text{D N} & (3) \\ D \longrightarrow \text{le} & (4) \\ D \longrightarrow \text{la} & (5) \\ N \longrightarrow \text{fille} & (6) \\ N \longrightarrow \text{gâteau} & (7) \\ V \longrightarrow \text{déguste} & (8) \\ V \longrightarrow \text{mange} & (9) \end{cases}$$

$L(G_3)$ ?

# Exercice 4 : à faire, à rendre à la fin du TD

Soit la grammaire $G_4$ définie par :
$V_N = \{S, A, B\}$
$V_T = \{a, b\}$

$$P = \begin{cases} S \longrightarrow AB & (1) \\ S \longrightarrow AS & (2) \\ A \longrightarrow a & (3) \\ B \longrightarrow b & (4) \end{cases}$$

$L(G_4)$ ?

# Exercice 5 : à faire, à rendre à la fin du TD

Soit la grammaire $G_5$ définie par :
$V_N = \{S\}$
$V_T = \{z\}$

$$P = \left\{ \begin{array}{ll} S \longrightarrow zSz & (1) \\ S \longrightarrow z & (2) \end{array} \right.$$

$L(G_5)$ ?

# Exercice 6 : à faire, à rendre à la fin du TD

Soit la grammaire $G_6$ définie par :
$V_N = \{S, GN, GV, Df, Dm, Nf, Nm, V\}$
$V_T =$
$\{\text{un}, \text{une}, \text{le}, \text{la}, \text{enfant}, \text{garçon}, \text{fille}, \text{cerise}, \text{haricot}, \text{cueille}, \text{mange}\}$

$$P = \begin{cases} S \longrightarrow \text{GN GV} & (1) \\ GN \longrightarrow \text{Df Nf}|\text{Dm Nm} & (2) \\ GV \longrightarrow \text{V GN} & (3) \\ Df \longrightarrow \text{une}|\text{la} & (4) \\ Dm \longrightarrow \text{un}|\text{le} & (5) \\ Nf \longrightarrow \text{fille}|\text{cerise} & (6) \\ Nm \longrightarrow \text{enfant}|\text{garçon}|\text{haricot} & (7) \\ V \longrightarrow \text{cueille}|\text{mange} & (8) \end{cases}$$

« un haricot cueille un enfant » appartient-elle à $L(G_6)$ ?

# Exercice 7 : à faire, à rendre à la fin du TD

Construire une grammaire $G_7$ pour le langage
$L(G_7) = \{ab^na | n \in \mathbb{N}\}$

# Exercice 8 : à faire, à rendre à la fin du TD

Construire une grammaire $G_8$ pour le langage
$L(G_8) = \{0^{2n}1^n | n \geq 0\}$

# Exercice 9 : à faire, à rendre à la fin du TD

Soit la grammaire $G_9$ définie par :
$V_N = \{S, A, B\}$
$V_T = \{a, c\}$

$$P = \begin{cases} S \longrightarrow ASA & (1) \\ AAS \longrightarrow B & (2) \\ AB \longrightarrow B & (3) \\ AA \longrightarrow a & (4) \\ aA \longrightarrow c & (5) \\ Ba \longrightarrow ca & (6) \\ Bc \longrightarrow ac & (7) \end{cases}$$

1-1-2-4-6 et 1-1-1-1-1-2-4-5-4-4-5-6 vont donner ?