



# Grammaires formelles : grammaires régulières

Karën Fort

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



# Quelques sources d'inspiration

par ordre d'importance décroissant

- ▶ cours de D. Battistelli (Paris 3), grâce aux notes de C. Riquier (Master 2, Paris 4)
- ▶ cours d'A. Rozenknop (Paris 13)
- ▶ cours de B. Habert (ENS de Lyon)
- ▶ *Language as a cognitive process - Syntax* (Terry Winograd) – Addison Wesley
- ▶ *Introduction à la calculabilité* (Pierre Wolper) – InterEditions, 1991
- ▶ cours en ligne de J-F. Perrot (Paris 6), avec son accord :  
<http://pagesperso-systeme.lip6.fr/Jean-Francois.Perrot/inalco/Automates/Cours17.html>

## Sources

### Grammaires régulières

- Type de grammaire

- Définition

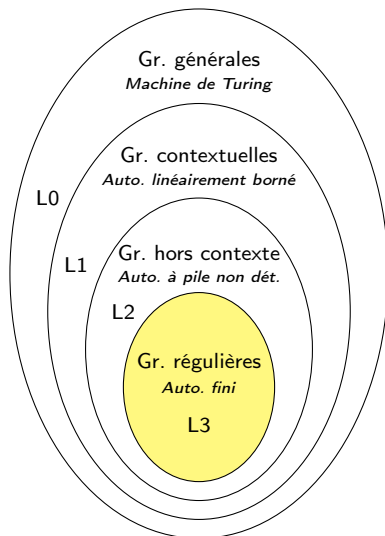
- Automate et grammaire régulière

- Expressivité

- Limites des grammaires régulières

## Pour finir

# Place dans la hiérarchie



## Type 3 : grammaires régulières (ou rationnelles)

### Définition

Une grammaire est régulière si toutes ses productions ont une des formes :

$$\left. \begin{array}{l} A \longrightarrow w B \\ A \longrightarrow w \end{array} \right\} \text{ avec } \left\{ \begin{array}{l} A, B \in V_N \\ w \in V_T \end{array} \right.$$

Toute production d'une grammaire régulière a donc :

- ▶ un membre de gauche constitué d'un seul non-terminal et
- ▶ un membre de droite constitué d'un mot de symboles terminaux suivi éventuellement d'un seul non-terminal

# Grammaires linéaires à droite|gauche

La définition donnée précédemment concerne les grammaires  
linéaires à droite :

$$\left. \begin{array}{l} A \longrightarrow w B \\ A \longrightarrow w \end{array} \right\} \text{ avec } \left\{ \begin{array}{l} A, B \in V_N \\ w \in V_T \end{array} \right.$$

# Grammaires linéaires à droite|gauche

La définition donnée précédemment concerne les grammaires linéaires à droite :

$$\left. \begin{array}{l} A \longrightarrow w B \\ A \longrightarrow w \end{array} \right\} \text{ avec } \left\{ \begin{array}{l} A, B \in V_N \\ w \in V_T \end{array} \right.$$

Dans une grammaire linéaire à gauche, les règles ont l'une des formes suivantes :

$$\left. \begin{array}{l} A \longrightarrow B w \\ A \longrightarrow w \end{array} \right\} \text{ avec } \left\{ \begin{array}{l} A, B \in V_N \\ w \in V_T \end{array} \right.$$


## Grammaires linéaires à droite|gauche

La définition donnée précédemment concerne les grammaires linéaires à droite :

$$\left. \begin{array}{l} A \longrightarrow w B \\ A \longrightarrow w \end{array} \right\} \text{ avec } \left\{ \begin{array}{l} A, B \in V_N \\ w \in V_T \end{array} \right.$$

Dans une grammaire linéaire à gauche, les règles ont l'une des formes suivantes :

$$\left. \begin{array}{l} A \longrightarrow B w \\ A \longrightarrow w \end{array} \right\} \text{ avec } \left\{ \begin{array}{l} A, B \in V_N \\ w \in V_T \end{array} \right.$$

 Il est possible de montrer que tout langage pouvant être défini par une grammaire linéaire à droite peut aussi être défini par une grammaire linéaire à gauche (et inversement)



# Exemple de grammaire régulière


Expressions figées commençant par « carte » :

- ▶ carte à jouer
- ▶ carte de France
- ▶ carte de crédit
- ▶ carte bleue
- ▶ carte à puce
- ▶ carte sur table

 Règles de réécriture permettant de générer ces expressions ?

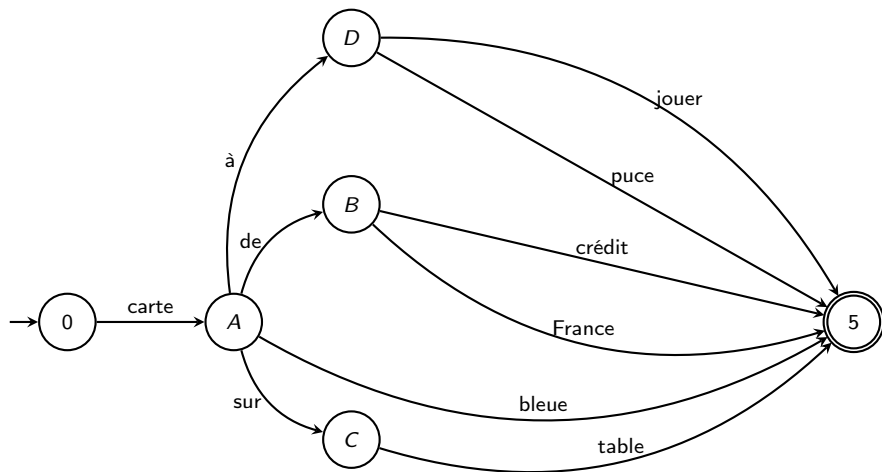
## Exemple de grammaire régulière : le dessous des cartes

$$\left\{ \begin{array}{l} S \longrightarrow \text{carte } A \\ A \longrightarrow \text{de } B \mid \text{bleue} \mid \text{sur } C \mid \text{à } D \\ B \longrightarrow \text{crédit} \mid \text{France} \\ C \longrightarrow \text{table} \\ D \longrightarrow \text{jouer} \mid \text{puce} \end{array} \right.$$

 Automate correspondant ?

# Exemple de grammaire régulière : le dessous des cartes

Automate



# D'un automate à une grammaire régulière

Le passage d'un automate aux règles correspondant à une grammaire régulière est immédiat :

- ▶ les noms d'états correspondent aux non-terminaux
- ▶ les étiquettes des arcs correspondent aux terminaux

## Expressivité des grammaires régulières

« *Regular expressions can be used to denote only a fixed number of repetitions or an unspecified number of repetitions of a given construct.* »

[Aho et al., 1986]



Les grammaires de type 3 permettent d'illustrer l'ordre dans lequel certains éléments sont concaténés : elles sont utiles en morphologie, en phonologie, mais aussi pour de l'analyse syntaxique simple (*chunking*).

## Exemple (d'un sous-ensemble) des insultes du Capitaine Haddock

{	<i>insulte</i>	→	accapareur scolopendre vermicelle ophicléide bougre de1
	<i>de1</i>	→	d' n1
	<i>n1</i>	→	extrait de2
	<i>de2</i>	→	d' n2
	<i>n2</i>	→	hydrocarbure



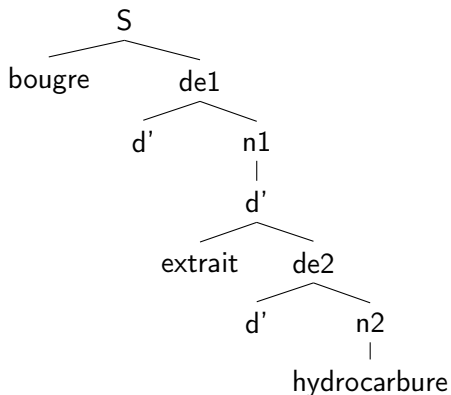
la description des insultes est atomisée et ne rend pas compte des parentés de structure entre elles :

*Bande de Canaques / Espèce de babouin*

*Bougre d'extrait d'hydrocarbure / Espèce de marchand de guano*

# Structure des grammaires régulières

Exemple de *bougre d'extrait d'hydrocarbure*



Structure en **peigne**

## Expressivité des grammaires régulières

« Les règles de Kleene [de réécriture des grammaires régulières] formalisent la notion intuitive de contrainte syntagmatique "dépendance de droite à gauche". On rencontre ce genre de contrainte dans, par exemple, des énoncés où des relatives s'enfilent de gauche à droite, comme dans :

*"Noémon adore Marie qui aime Jean qui admire..."* »

*[Desclés, 1973, p. 312]*



## Expressivité des grammaires régulières : limites

« Prenons maintenant un énoncé comme :

"Noémon qui adore Marie qui admire Jean ... est vieux et sage"

Contrairement à l'énoncé précédent où les relatives s'enfilaient de gauche à droite, ici les relatives "qui adore Marie qui admire Jean" sont emboîtées à l'intérieur de l'énoncé : "Noémon est vieux et sage". Ce processus d'emboîtement de relatives doit pouvoir être poursuivi un nombre quelconque de fois [...]

Il n'existe pas de  $K$ -grammaire [grammaire régulière] qui puisse non seulement décrire adéquatement ce type de contrainte d'emboîtement, mais encore qui énumère tous les énoncés comprenant un nombre quelconque de relatives emboîtées. » [Desclés, 1973, p. 313]

# Limitations

Les grammaires de type 3 ne peuvent traiter :

- ▶ les chaînes  $a^n b^n$  de type  $a_1 \dots a_n b_1 \dots b_n$  (“respectivement”)

## Limitations

Les grammaires de type 3 ne peuvent traiter :

- ▶ les chaînes  $a^n b^n$  de type  $a_1 \dots a_n b_1 \dots b_n$  (“respectivement”)
- ▶ les chaînes  $a^n b^n$  de type  $a_1 \dots a_n b_n \dots b_1$  (expressions **parenthésées**, structures enchâssées, HTML, XML) :

En anglais :

*The dog the stick the fire burned beat bit the cat*

En français :

*Le chat que le voisin que le maire que le préfet qui a été condamné a félicité a attrapé est blanc.*

# Limitations

Les grammaires de type 3 ne peuvent traiter :

- ▶ les chaînes  $a^n b^n$  de type  $a_1 \dots a_n b_1 \dots b_n$  (“respectivement”)
- ▶ les chaînes  $a^n b^n$  de type  $a_1 \dots a_n b_n \dots b_1$  (expressions **parenthésées**, structures enchâssées, HTML, XML) :
- ▶ les chaînes  $abac$   
*soit ... soit ...*  
*ni ... ni ...*

# Limitations

Les grammaires de type 3 ne peuvent traiter :

- ▶ les chaînes  $a^n b^n$  de type  $a_1 \dots a_n b_1 \dots b_n$  (“respectivement”)
- ▶ les chaînes  $a^n b^n$  de type  $a_1 \dots a_n b_n \dots b_1$  (expressions **parenthésées**, structures enchâssées, HTML, XML) :
- ▶ les chaînes  $abac$
- ▶ le **rejet** en fin de phrase des prépositions (en anglais) ou des particules séparables (en allemand) :

*The girl that I do not want to be caught with.*

# Limitations

Les grammaires de type 3 ne peuvent traiter :

- ▶ les chaînes  $a^n b^n$  de type  $a_1 \dots a_n b_1 \dots b_n$  (“respectivement”)
- ▶ les chaînes  $a^n b^n$  de type  $a_1 \dots a_n b_n \dots b_1$  (expressions **parenthésées**, structures enchâssées, HTML, XML) :
- ▶ les chaînes  $abac$
- ▶ le **rejet** en fin de phrase des prépositions (en anglais) ou des particules séparables (en allemand) :
- ▶ Les **dépendances à longue distance** (interrogatives, clivées. . .)  
*Jean veut savoir quelle fille Marie croit que Paul a vue.*

Sources

Grammaires régulières

Pour finir

CQFR : Ce Qu'il Faut Retenir

Bibliographie



Grammaires régulières :

- ▶ définition
- ▶ passer d'un automate à une grammaire (et *vice versa*)
- ▶ structure, limites





Aho, A., Sethi, R., and Ullman, J. (1986).

Compilers : Principles, Techniques, and Tools.

Addison-Wesley series in computer science and information processing. Addison-Wesley Publishing Company.