



Grammaires formelles : Poser les bases

Karën Fort

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



Quelques sources d'inspiration

par ordre d'importance décroissant

- ▶ cours de B. Habert (ENS Lyon)
- ▶ *Notions sur les grammaires formelles* (Maurice Gross et André Lentin) – Gauthier-Villars, 1967
- ▶ (excellent) cours en ligne de J-F. Perrot (Paris 6), avec son accord : <http://pagesperso-systeme.lip6.fr/Jean-Francois.Perrot/inalco/Automates/Cours2.html>
- ▶ cours de L. Audibert (Paris 13)
- ▶ *Introduction à la calculabilité* (Pierre Wolper) – InterEditions, 1991

Sources

Où la langue parle de la langue

Langue et métalangue

Terminologie

Définitions

Opérations sur les langages

Pour finir

Métalangage de définition

*« Il s'agit de pouvoir faire la différence de manière non ambiguë entre le langage de définition et le langage défini »
[B. Habert]*

Or, pour les langues naturelles, la langue est définie par le biais de la langue !

⇒ « chocs » terminologiques, surtout pour des linguistes :
2♥4♠5 est un « mot »

Terminologie(s)

La théorie des grammaires formelles est appliquée dans différents domaines : langages artificiels, logique, langues naturelles, ...
⇒ terminologie « hésitante »

Sources

Où la langue parle de la langue

Définitions

Plus petit élément manipulé

Mots et langages

Opérations sur les langages

Pour finir

Alphabet

ou ensemble d'atomes

« Un alphabet est un *ensemble fini non vide* dont les éléments sont appelés lettres, symboles, mots »
[L. Audibert, p. 7]

= ce qui entre dans un regroupement

« La seule caractéristique importante d'un alphabet est sa taille (le nombre d'éléments qu'il contient) »
[Introduction à la calculabilité, p. 20]

Mots

ou molécules possibles

« Un mot est une *suite finie* d'éléments de l'alphabet.
Si l'on utilise le terme mot pour désigner les éléments de l'alphabet on utilisera le terme phrase pour désigner une suite finie d'éléments de l'alphabet. » [L. Audibert, p. 7]

Mot = regroupement possible = chaîne de caractères :

- ▶ on parle de **longueur d'un mot**
- ▶ on distingue les différentes occurrences des lettres qui composent le mot
- ▶ un mot particulier, le **mot vide**, ne contient aucune occurrence et il est noté : "" ou λ ou ε

Exemples de mots

- ▶ a , abs , zt , $bbssnbzzyyyyddtra$, $ordinateurrapide$ sont des mots sur l'alphabet $\{a, \dots, z\}$
- ▶ $4\clubsuit 3\diamond 5\heartsuit 2\spadesuit$, 12765 , $\spadesuit\heartsuit$ sont des mots sur l'alphabet $\{0, \dots, 7, \spadesuit, \clubsuit, \heartsuit, \diamond\}$

Concaténation de mots

Les mots sont faits pour être enchaînés, d'abord deux à deux...

Exemple : ab raca dabra = abracadabra



De quel(s) alphabet peuvent provenir ces mots ?

Langage

on appelle langage un **ensemble de mots** quelconque défini sur un alphabet donné

Exemples :

- ▶ $\{aab, aaaa, \varepsilon, a, b, ababbabababbbbbb\}$,
 $\{\varepsilon, aaaaaaaaa, a, bbbbbb\}$ et \emptyset sont des langages sur l'alphabet $\{a, b\}$
- ▶ pour l'alphabet $\{0, 1\}$,
 $\{0, 1, 00, 01, 10, 11, 000, 001, 010, 011, \dots\}$ est un langage

Ne pas confondre

- ▶ le langage \emptyset : langage vide ne contenant aucun mot
- ▶ le langage $\{\varepsilon\}$: langage contenant un mot, qui est le mot vide

Sources

Où la langue parle de la langue

Définitions

Opérations sur les langages

Opérations ensemblistes

Opérations spécifiques

Pour finir

- ▶ opérations ensemblistes (vues précédemment)
- ▶ opérations spécifiques :
 - ▶ concaténation
 - ▶ +
 - ▶ étoile de Kleene

Union de langages

*« Étant donnés deux langages L_1 et L_2 , on appelle union de ces langages et on désigne par $L_1 \cup L_2$ l'ensemble des mots que l'on trouve dans l'un (au moins) de ces langages »
[Notions sur les grammaires formelles, p. 17]*

Exemple :

- ▶ $L_1 = \{a,b,c,d\}$
- ▶ $L_2 = \{d,e,f\}$
- ▶ $L_1 \cup L_2 = \{a,b,c,d,e,f\}$

Intersection de langages

« Dans les mêmes conditions on appelle intersection de ces langages et l'on désigne par $L_1 \cap L_2$ l'ensemble des mots qui appartiennent simultanément aux deux langages »
[Notions sur les grammaires formelles, p. 18]

Exemple :

- ▶ $L_1 = \{a,b,c,d\}$
- ▶ $L_2 = \{d,e,f\}$
- ▶ $L_1 \cap L_2 = \{d\}$

Complémentaire d'un langage

« Le complémentaire du langage L (...) est l'ensemble des mots (...) qui n'appartiennent pas à L »

[Notions sur les grammaires formelles (adapté), p. 18]

Produit de langages

ou concaténation de langages

« Dans les mêmes conditions on appelle produit de deux langages et l'on désigne par L_1L_2 l'ensemble des mots obtenus en prenant un mot de L_1 et en concaténant à sa droite un mot de L_2 »

[Notions sur les grammaires formelles, p. 18]

Exemple :

soient $A = \{ab, raca\}$ et $B = \{ab, raca, dabra\}$

$AB = \{abab, abraca, abdabra, racaab, racaraca, racadabra\}$

Petit exercice



Si $A = B = \{a, aa\}$, que vaut AB ?

Lien avec les langages de programmation

« *En termes de calculs, l'ensemble des séquences décrites par le produit AB est naturellement celui des séquences obtenues en effectuant un calcul de B après un calcul de A .*

Le produit des langages symbolise donc l'exécution séquentielle des programmes. »

[JF Perrot, cours 2]

Étoile de Kleene

ou opération étoile, ou fermeture de Kleene

« Étant donné un langage L , L^ désigne le langage formé de tous les produits de mots de L , en nombre quelconque, de 0 à l'infini »*

[JF Perrot, cours 2 (adapté)]

Exemple :

$\{ \text{ara, cara} \}^* = \{ "", \text{ara, cara, araara, aracara, caraara, caracara, araaraara, araaracara, aracaraara, aracaracara, caraaraara, caraaracara, caracaraara, caracaracara, ...} \}$

Lien avec les langages de programmation

« En termes de calculs, l'étoile A^* s'obtient en répétant - on dit aussi en itérant - un nombre quelconque de fois le calcul décrit par A .

Dans les langages de programmation, cette itération s'écrit sous la forme d'une *boucle*. »

[JF Perrot, cours 2]

Étoile stricte de Kleene

ou +

On appelle **étoile stricte** de Kleene de L l'ensemble des mots obtenus par concaténation d'un nombre **non nul** de mots de L . On note ce langage L^+ .

Sources

Où la langue parle de la langue

Définitions

Opérations sur les langages

Pour finir

CQFR : Ce Qu'il Faut Retenir

TD



- ▶ langue et métalangue
- ▶ alphabet vs mot vs langage
- ▶ concaténation
- ▶ étoile de Kleene

Exercice : à faire, à rendre, sera noté (ou pas)

Soient les langages :

- ▶ $L_1 = \{\text{Walking Dead, Sherlock Holmes, The Handmaid's Tale}\}$
- ▶ $L_2 = \{a, b, c, d\}$
- ▶ $L_3 = \{\text{titi, toto, toutou}\}$

Quel est le langage résultant de :

- ▶ l'union de L_1 et L_2
- ▶ l'union de L_2 et L_3
- ▶ la concaténation de L_1 et L_2
- ▶ la concaténation de L_2 et L_3
- ▶ donnez 10 mots distincts de la fermeture de Kleene du langage L_3

Exercice : à faire, à rendre, sera noté (ou pas)

Soit l'alphabet $A = \{a, b\}$

- ▶ étant donnés les mots $u = ab$ et $v = aba$, écrire les mots
 - ▶ uv
 - ▶ $(vu)^3$
 - ▶ v^2u
- ▶ énoncer tous les mots de longueur 2 définis sur A

Exercice : à faire, à rendre, sera noté (ou pas)

Soit le vocabulaire (ensemble de mots) $V = \{a, m, t, ma, ta, mat\}$

Donner les différentes longueurs possibles des chaînes suivantes ainsi que leurs découpages respectifs :

- ▶ *mama*
- ▶ *tamtam*

Exercice : à faire, à rendre, sera noté (ou pas)

Soit le vocabulaire (ensemble de mots) $V = \{a, m, t, s, e\}$ et les langages suivants, construits à partir de ce vocabulaire :

- ▶ $L_1 = \{ma, ta, se\}$
- ▶ $L_2 = \{ma, me, ta, te\}$
- ▶ $L_3 = \{ma, mat, mata, \lambda\}$
- ▶ $L_4 = \{a, m, ma, am\}$

Définir les langages suivants :

- ▶ $L_1 \cup L_2$
- ▶ $L_1 \cap L_2$
- ▶ $L_1 L_2 \cap L_3$
- ▶ $L_2 * \cap L_3 *$

Vrai ou faux ?

- ▶ $mata \in L_1 *$
- ▶ $amam \in L_3 *$
- ▶ $tamtam \in L_3 *$
- ▶ $mamama \in L_1 * \cap L_2 * \cap L_3 * \cap L_4 *$

Exercice

Soient les langages :

- ▶ $T = \{ A..Z, a..z \}$
- ▶ $C = \{ 1..0 \}$

Que représente (expliquez) :

- ▶ $T \cup C$
- ▶ TC
- ▶ T^4
- ▶ T^*
- ▶ $T(T \cup C)^*$
- ▶ C^+