

# Lexique et morphologie

25 novembre 2019

—Kata Gábor

*kata.gabor@inalco.fr*—

## Langue formelle :

une langue  $L$

sur un alphabet/vocabulaire  $V$ ,

est un ensemble de chaînes de symboles produit par concaténation des symboles dans  $V$ .

$L$  est ainsi un sous-ensemble de toutes les chaînes que l'on peut produire par concaténation des éléments de  $V$ .

## Grammaire d'une langue formelle :

un modèle capable de reconnaître / générer toutes les chaînes de caractères qui appartiennent à cette langue.

**Les automates á états finis et les expressions régulières reconnaissent les mêmes langues : les langues régulières.**

## Langue régulière :

$L$  est une langue régulière sur l'alphabet  $V$  :

- $\emptyset$  (l'ensemble vide/langage vide) est une langue régulière ;
- $\forall a \in V, \{a\}$  (l'ensemble composé d'un élément du vocabulaire) est une langue régulière ;
- Si  $L_1$  et  $L_2$  sont des langues régulières,
  - $L_1 \cup L_2$ , l'**union** de  $L_1$  et  $L_2$ ,
  - $L_1 \cdot L_2$ , la **concaténation** de  $L_1$  et  $L_2$ , c'est-à-dire  $\{xy \mid x \in L_1, y \in L_2\}$
  - $L_1^*$ , la **fermeture de Kleene** de  $L_1$ , c'est-à-dire la langue obtenue en prenant un nombre quelconque ( $y$  compris 0) de chaînes de  $L_1$  et en les concaténant.

Les langues régulières sont fermées pour les opérations de

- complémentation
- inversion des chaînes
- intersection
- différence

Reconnaître des formes complexes par automates à états finis

Approche "item and arrangement" :

- définir les radicaux
- définir des classes flexionnels de radicaux
- définir les catégories flexionnelles

- définir les formes d’affixes pour chaque catégorie et chaque classe flexionnelle
- définir l’ordre de concaténation

1. créer des automates pour les (classes de) radicaux
2. créer des automates pour les affixes
3. concaténation des automates de radicaux avec les automates d’affixes

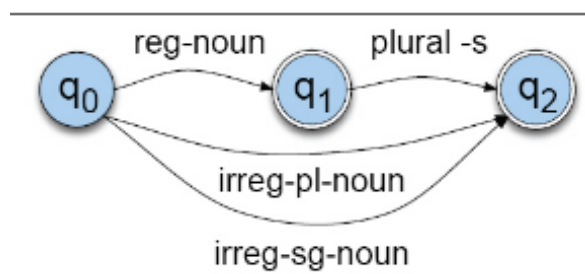


FIGURE 1 – Pluriel régulier et irrégulier des noms

### Automates et expressions régulières

Signification	Expression
séquences de caractères	/chien/, /loup/, /François Hollande/
disjonction	/[aA]/, /[2-5]/, /[a-z]/, / a   b/
négation	/[^a]/
quantification	/a ?/, /b*/, /c+/, /e{3,5}/
parenthèses	/a(ab)+/, /ceci n’(est)était) pas/
n’importe quel caractère	./
repères	^abcdef\$/
alias	^d/, ^s/
caractères spéciaux	^n/, ^t/

opérations sur des automates	notation
concaténation	A B
optionnalité	A ? = ( A)
fermeture Kleene	A* = ( A AA AAA ...)
disjonction	A   B
conjonction	A & B
complément	!A

**Exercice :** Donnez les automates correspondant aux opérations de concaténation, optionnalité, fermeture Kleene, disjonction.

## Exercices

1. Ecrivez des expressions régulières pour les langues suivantes :  
(par mot nous entendons des chaînes de caractères séparés par des espaces ou de la ponctuation)
  - l'ensemble de tous les string alphabétiques,
  - l'ensemble de tous les strings alphabétiques en minuscules finissant par b,
  - tous les strings alphabétiques (en 1) directement précédés et suivis par un b,
  - tous les strings alphabétiques en minuscules finissant par b (en 2), directement précédés et suivis par un b (supplémentaire),
  - tous les strings qui commence en début de ligne par une chiffre (p.ex. 1, 2, 222, ...) et qui termine en fin de ligne par un mot.
2. Considérez le tableau de conjugaison allemande du verbe 'singen' (*chanter*) avec les formes en indicatif présent, passé et subjonctif présent.
  - (a) Construisez un automate *minimal* qui reconnaît toutes ces formes.
  - (b) Pour chaque temps et mode verbal listés, construisez un transducteur individuel qui reconnaît les formes et associe à chaque forme l(es) analyse(s) correspondante(s).

Personne	Indicatif présent (IPR)	Indicatif passé (IPS)	Subjonctif présent (SP)
S1	singe	sang	sänge
S2	singst	sangst	sängest
S3	singt	sang	sänge
PL1	singen	sangen	sängen
PL2	singt	sangt	sänget
PL3	singen	sangen	sängen

3. Considérez le tableau de déclinaison de l'adjectif hongrois *nagy* (grand).
  - (a) Construisez un automate *minimal* qui reconnaît toutes les formes.
  - (b) Construisez un transducteur *minimal* qui reconnaît les formes et associe à chaque forme l(es) analyse(s) correspondante(s).

Cas	Base Singulier	Base Pluriel	Comparatif Singulier	Comparatif Pluriel
Nominatif	nagy	nagyok	nagyobb	nagyobbak
Accusatif	nagyot	nagyokat	nagyobbat	nagyobbakat
Datif	nagynak	nagyoknak	nagyobbnak	nagyobbaknak
Ablatif	nagytól	nagyoktól	nagyobbtól	nagyobbaktól

4. Construisez un modèle morphologique : lexique et automates qui reconnaissent

- la conjugaison des verbes français en -er : à l'indicatif présent, imparfait, futur
- les déclinaisons latines suivantes :

Cas	Singulier	Pluriel
Nominatif	Rosa	Rosae
Accusatif	Rosam	Rosas
Génitif	Rosae	Rosarum
Datif	Rosae	Rosis
Ablatif	Rosa	Rosis

Cas	Singulier	Pluriel
Nominatif	Murus	Muri
Accusatif	Murum	Muros
Génitif	Muri	Murorum
Datif	Muro	Muris
Ablatif	Muro	Muris

- Transformez les automates en transducteurs capables d'associer une analyse aux formes reconnues.