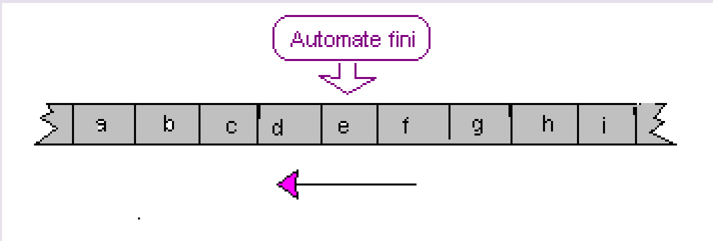
TD2

Modéliser des connaissances linguistiques à l'aide d'automates à états finis

A propos des automates finis (AF)

Un *automate fini* (ou *automate à états finis* ou *automate à nombre fini d'états*) est un système formel qui peut prendre un nombre fini d'*états*. L'un de ces états est distingué et est appelé *état initial*. L'automate a la propriété de pouvoir lire, un à un, les symboles d'une chaîne constituée sur un certain vocabulaire Vt. Lorsque l'automate lit un symbole, il a la propriété de pouvoir changer d'état. Le nouvel état qu'il prend dépend à la fois du symbole qu'il lit et de l'état dans lequel il est lorsqu'il le lit.



Définition formelle :

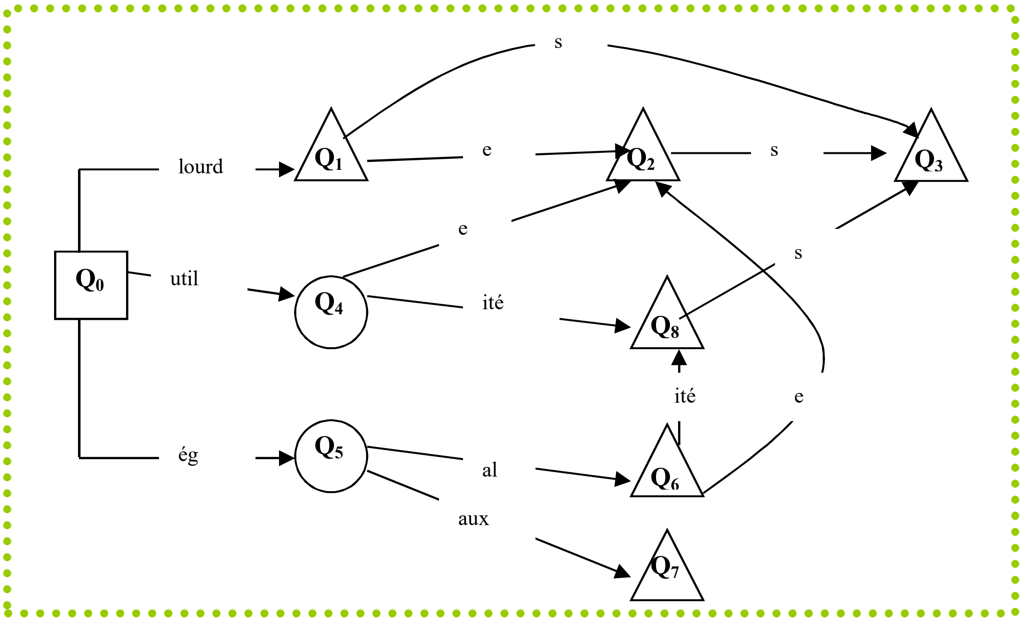
Un automate fini A est un quintuplet **(Q, Vt, P, Q_0, T)** dans lequel :

- -Q est un ensemble fini d'états
- -Vt est un vocabulaire fini de symboles
- - P est un ensemble de règles de transitions du type : $(Q_i, \alpha) \rightarrow Q_j$ où $\alpha \in$ Vt et $Q_i, Q_j \in Q$
- -Q_0 est l'état initial
- -T est l'ensemble d'états terminaux ($T \subseteq Q$)

Le langage défini par un automate fini est **L(A) = { x / x $\in$ V^{+}t et A accepte x }**

Exercice 1.

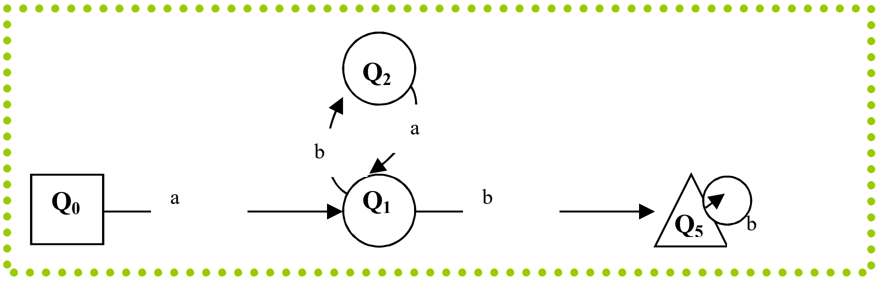
Etant donné l'automate suivant :



Quel est son Vt ? Quel est le langage qu'il définit ?

Exercice 2.

Etant donné l'automate suivant :



Quel est son Vt ? Quel est le langage qu'il définit ?

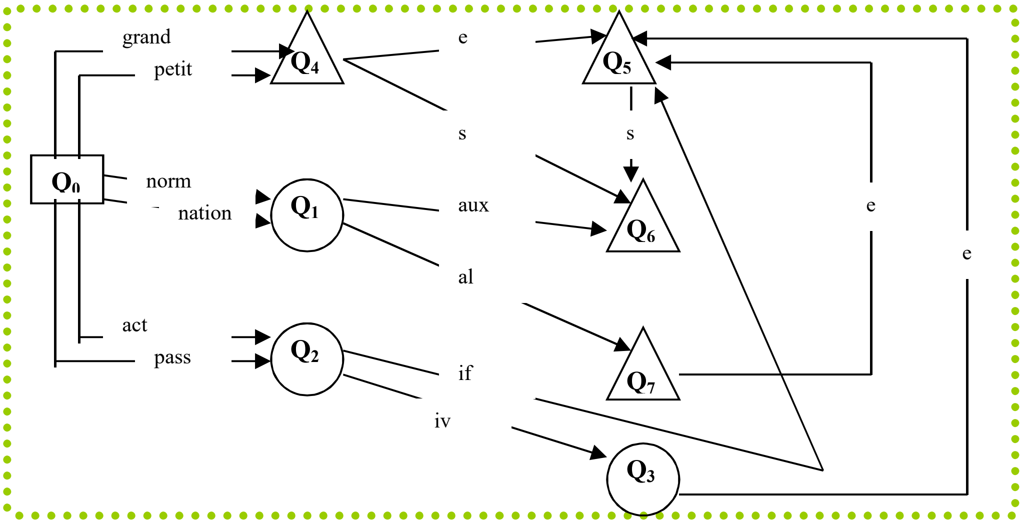
Exercice 3.

Comme pour les grammaires, il est toujours possible d'imaginer différents automates pour reconnaître un même langage. Plusieurs critères peuvent nous guider dans le choix d'un automate plutôt qu'un autre :

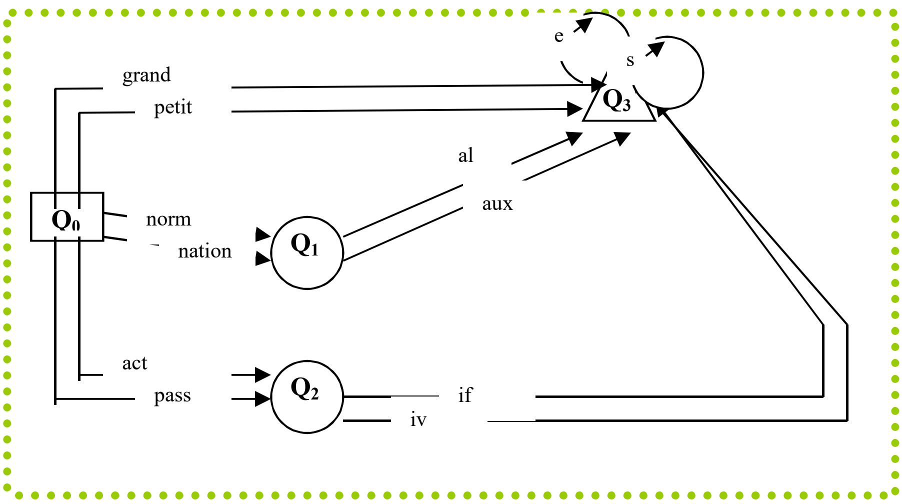
- On peut vouloir minimiser le nombre d'états (critère d'*économie*) ;
- On peut préférer un AFD à un AFND (critère de *simplicité* de fonctionnement) ;
- On peut exiger, ou non, qu'il ne reconnaisse que le langage qui nous intéresse (critère d'*adéquation*).

Cherchons à représenter la morphologie de certains adjectifs comme *petit, grand, …, normal, national, …, actif, passif, …* Voici trois automates finis AF1, AF2 et AF3. Préciser ce qui les distingue au regard des trois critères rappelés ci-dessus.

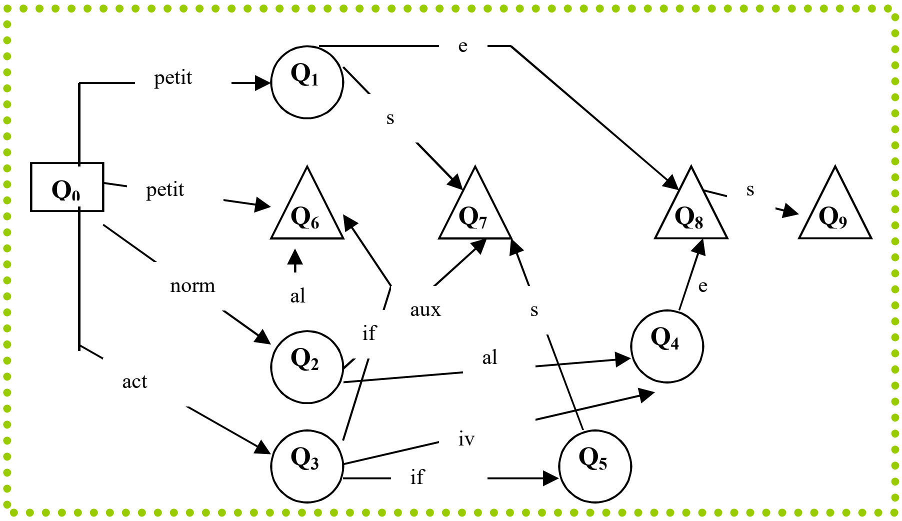
AF_1



AF_2



AF₃



Exercice 4.

Construire un automate, avec le plus petit nombre possible d'états, qui définit le langage suivant :

L = {exciter, surexciter, excitation, surexcitation, excitable, surexcitable, estimer, surestimer, estimation, surestimation, estimable, surestimable}

Exercice 5.

Soit le langage L(G_{adj}) tel que :

L(G_{adj}) = {petit, petite, petits, petites, grand, grande, grands, grandes}

Déterminer un automate qui définit le langage L(G_{adj})

Exercice 6.

Proposer un automate fini qui rende compte de la morphologie de quelques familles de mots avec ou sans préfixe (con-, in-, dé-, *etc*.) et avec différents suffixes (-er, -ation, *etc*.).

Exercice 7.

Soit L1 = { $(ab)^n c^p$ / $n \geq 0$ et $p \geq 1$}. Quel est l'automate fini qui le décrit ?

Exercice 8.

Le langage L2 défini comme l'ensemble de tous les mots sur {a, b} qui contiennent exactement cinq 'a' est-il régulier ? Si oui, donner un automate fini le décrivant.

Exercice 9.

Déterminer un automate fini défini sur Vt = {a} qui permette d'engendrer le langage suivant :

L = {a^n / n pair ≥ 2}

Exercice 10.

Proposez un AF pour le langage de tous les mots de {a, b, c}* donc *bac* est un sous-mot.

Remarque : un *sous-mot* est une sous-suite de lettres – non nécessairement contiguës. A distinguer d'un *facteur*. Par exemple, *pis* est un sous-mot de *produits*.

A propos des langages réguliers

Pour montrer qu'un langage est régulier, il suffit d'exhiber un AF qui l'engendre. L'union de deux langages réguliers est un langage régulier. L'intersection de deux langages réguliers est aussi un langage régulier.

Plus généralement, on dit qu'il y a fermeture quand le langage, résultat d'une opération, est toujours de même type que les langages avec lesquels on a fait l'opération. Il existe bien des cas où des types de langages (au sens strict) sont fermés par rapport à certaines opérations. Les propriétés de fermeture sont résumées dans le tableau suivant :

Types de langage	**Fermeture par rapport à l'opération**					
	$\cup$	$\cap$	complément	produit	*	inverse
0	oui	oui	non	oui	oui	oui
1	oui	oui	?	oui	oui	oui
2	oui	non	non	oui	oui	oui
3	oui	oui	oui	oui	oui	oui

Ainsi, la réunion (ou le produit) de deux langages de type *i* est également un langage de type *i*. Le type du complément d'un langage de type 1 (marqué « ? ») est un problème non encore résolu.

Exercice 11.

Le langage L3 défini comme l'ensemble de tous les mots sur {a, b} tels que ils aient autant de a que de b est-il régulier ?

Exercice 12.

Montrez que les langages suivants sont réguliers :

L4 = $\{ ca^m / m \geq 0\} \cup \{b^n c / n \geq 0\}$

L5 = $\{ x / x \in \{a,b\}^* /$ nombre impair de a et nombre impair de b$\}$

Exercice 13.

Soient des suites de catégories lexico-syntaxiques que l'on peut rencontrer en français (avec des exemples) :

N V D N (*Marie arrose les salades*)
D N P **N V D N** (*la sœur de Joe arrose les salades*)
D A N P **N V D N** (*la jeune sœur de Joe arrose les salades*)
D Ad A N P **N V D N** (*la très jeune sœur de Joe arrose les salades*)
D A C A N P **N V D N** (*la jeune et gentille sœur de Joe arrose les salades*)
D A A C A N P **N V D N** (*la jeune, gentille et sympathique sœur de Joe arrose les salades*)
D A A … A C A N P **N V D N** (*la jeune, drôle, …, gentille et sympathique sœur de Joe arrose les salades*)

Représentez par un seul AF l'ensemble de ces schémas de phrases.

Exercice 14.

Reprenez les occurrences d'adverbiaux calendaires rencontrés dans le texte biographique sur de Villepin et proposez un seul AF à même de les repérer et de les catégoriser selon un mode d'ancrage absolu *vs* relatif (anaphorique *vs* déictique)