Éléments de théorie des langages et de théorie des ensembles pour informaticien·es

Quelles sont les « *données* » que l'on peut représenter en informatique ?
Quelles sont les « *bonnes* » représentations de ces données ?

Feuille « 02 » (du 6 septembre 2024)

NB seules les sections 1 et 2 sont à maîtriser dans le cadre du cours « Automates finis ». Les sections suivantes fournissent juste un complément utile pour les plus motivé·es d'entre vous par l'informatique théorique (voir l'explication en entête de la section 3 page 8).

Ce document [1] mélange notions de cours avec exercices à réaliser sur papier et/ou machine (en PYTHON). En particulier, les paragraphes intitulés « **RENDU** n » sont des exercices de programmation à rendre sur https://teide.ensimag.fr/ (connection via un PC de l'école ou via le VPN Ensimag - voir https://intranet.ensimag.grenoble-inp.fr/fr/informatique/vpn-ensimag). Le format et la date limite du rendu sont indiqués sur Teide.

Les autres exercices sont corrigés, dans un document séparé, qu'il ne faut consulter qu'après avoir vraiment trouvé au moins un début de solution (même imparfait ou en partie faux). NE PAS CRAINDRE DE FAIRE DES ERREURS. Au contraire, rencontrer nos erreurs est nécessaire pour corriger nos images mentales floues ou fausses.

> *C'est quand nous craignons de nous tromper que l'erreur qui est en nous se fait immuable comme un roc.*
>
> Alexandre Grothendieck - https://fr.wikipedia.org/wiki/Alexandre_Grothendieck

1 Introduction (assez informelle) à la théorie des langages

On peut vouloir représenter les entiers naturels (de l'ensemble noté $\mathbb{N}$) par des séquences finies non-vides d'éléments de $\{0, 1\}$, comme « 1101 » qui représente l'entier écrit « 13 » en notation décimale. La notation décimale représente elle-même les entiers naturels comme des séquences finies non-vides d'éléments de $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Ceci conduit à la définition suivante, à la base de la **théorie des langages**.

Définition d'un mot w sur un vocabulaire V Soit V un ensemble fini de *symboles* (e.g. lettres, chiffres, caractères, idéogrammes, etc) appelé *vocabulaire*. Un mot est une séquence finie éventuellement vide d'éléments de V séparés par un *méta-symbole* dédié, généralement noté « . » (ou parfois avec juste un espace), qui ne doit pas figurer dans V. Le mot vide (i.e. la séquence vide) est noté par un autre méta-symbole, généralement « ε », qui ne doit pas non plus être dans V.

Exemple 1. Quatres mots sur $V = \{\mathtt{A}, \mathtt{T}, \mathtt{G}, \mathtt{C}\}$ (le vocabulaire des bases nucléiques de l'ADN [2]) :

ε T A.C C.T.T.A.T.G

Ce dernier mot pourrait peut-être coder un gène...

Exemple 2. On peut utiliser un vocabulaire V lui-même constitué de mots (V doit juste être un ensemble fini). Les mots sur V représentent alors des phrases. Typiquement, l'ensemble des mots de la langue française étant en nombre fini, on pourrait modéliser les phrases du français comme des mots (au sens de la théorie des langages) sur le vocabulaire défini par le dictionnaire, modulo conjugaison et passage au pluriel, étendu éventuellement avec les caractères de ponctuation.

1. N'hésitez pas à signaler les erreurs ou les points obscurs à sylvain.boulme@imag.fr, merci d'avance !
2. https://fr.wikipedia.org/wiki/Base_azot%C3%A9e

Notation sans séparateur Quand ce n'est pas ambigu, on peut omettre les « . ». Les 4 mots de l'exemple 1 peuvent aussi s'écrire :

ε `T` `AC` `CTTATG`

Exercice 1. Sur le vocabulaire $\{\texttt{C}, \texttt{I}, \texttt{CI}\}$ alors l'écriture sans séparateur est ambiguë. Par exemple, « `ICI` » peut correspondre à plusieurs mots. Lesquels ?

Exemple 3. On peut considérer les chaînes de caractères Python comme des mots sur le vocabulaire des caractères UTF-8 [3]. Comme « . » et « ε » appartiennent à ce vocabulaire, cela nécessite d'adapter les notations mathématiques usuelles. Le plus simple est d'utiliser celles de Python en utilisant une notation sans séparateur et entre apostrophes (*quote* en anglais). Ainsi le mot vide sera simplement écrit $''$. Dans cette notation, $'\texttt{ab}'$ représente un mot de deux symboles. [4]

Concaténation de deux mots Soient w_1 et w_2 deux mots non vides sur le même vocabulaire V. On note $w_1.w_2$ leur concanétation (formant un seul mot sur V obtenu en ajoutant le « . » entre les deux mots). Cette opération est étendue pour accepter des mots vides avec les égalités $\varepsilon.w = w$ et $w.\varepsilon = w$. Lorsque ce n'est pas ambigu, on s'autorise à laisser le symbole « . » implicite et à écrire $w_1 w_2$ à la place.

Exemple 4. Sur le vocabulaire $V = \{\texttt{a}, \texttt{b}\}$, on a ainsi $\texttt{bab}.\texttt{ba} = \texttt{babba}$. En Python, l'opérateur de concaténation sur les chaînes s'écrit `+`. [5] Donc `'bab'+'ba'` retourne `'babba'`.

Itéré d'un mot Soit w un mot sur un vocabulaire V. Soit n un entier naturel. On définit le mot w^n par récurrence sur n avec $w^0 \stackrel{def}{=} \varepsilon$ et $w^{n+1} \stackrel{def}{=} w.w^n$.

Exercice 2. Sur le vocabulaire $V = \{\texttt{a}, \texttt{b}\}$, que vaut $(\texttt{abb})^2.(\texttt{ba})^3$? En Python, l'opérateur d'itération est noté multiplicativement (ce qui est cohérent avec sa notation additive de la concaténation). Vérifiez en calculant `'abb'*2+'ba'*3`.

Longueur des mots On note $|w|$, la longueur du mot w, c'est-à-dire son nombre de symboles dans V (en comptant les occurrences multiples). Ainsi $|w_1.w_2| = |w_1|+|w_2|$. En Python, l'opérateur `len(w)` calcule la longueur de la chaîne `w`.
De façon similaire si $x \in V$, alors on note $|w|_x$ le nombre d'occurrences de x dans w. Et on a aussi $|w_1.w_2|_x = |w_1|_x + |w_2|_x$.

Exercice 3. Donner $|w|$ et $|w|_{\texttt{T}}$ de chacun des mots w de l'exemple 1. Puis, calculer $|w^n|$ en fonction de $|w|$ et n.

Exercice 4. On appelle B_0 l'ensemble des mots non vides sur le vocabulaire $\{\texttt{0}, \texttt{1}\}$. Chaque mot $w \in B_0$ représente un entier naturel noté $[\![w]\!]$, appelé **sémantique** de w, et défini par récurrence sur la longueur des mots :

$$\begin{array}{rcll} [\![a]\!] & \stackrel{def}{=} & a & \text{pour } a \in \{\texttt{0}, \texttt{1}\} \\ [\![w.a]\!] & \stackrel{def}{=} & (2 \times [\![w]\!]) + a & \text{pour } w \in B_0 \text{ et } a \in \{\texttt{0}, \texttt{1}\} \end{array}$$

On propose d'encoder un mot B_0 en Python comme un uplet non vide formé de 0 et de 1.

3. Voir https://fr.wikipedia.org/wiki/UTF-8.
4. Comme l'apostrophe « $'$ » fait elle-même partie du vocabulaire, pour représenter ce symbole à l'intérieur d'un mot, il faudra utiliser un mécanisme d'échappement de Python.
5. Les concepteurs de Python ont donc choisi de noter l'opération de concaténation de façon « additive », c'est-à-dire avec un `+`. Les mathématicien·nes préfèrent réserver les notations additives à des opérations *commutatives*, c'est-à-dire vérifiant « $x + y = y + x$ ». Ce n'est pas le cas de la concaténation. Pour les opérations non commutatives (comme la multiplication de matrices ou la concaténation de mots), les mathématicien·nes préfèrent une notation multiplicative (avec « $\times$ » ou « . »). Voir https://fr.wikipedia.org/wiki/Notation_multiplicative. Plus généralement, la notation additive de Python n'est pas très adaptée quand on passe de la concaténation des mots à la concaténation des langages (comme on va le voir bientôt).

Exemple : « `(1,1,0,1)` » encode le mot « `1.1.0.1` ». Définir une fonction PYTHON « `semT(w)` » qui calcule l'entier $[\![\texttt{w}]\!]$. Par exemple, elle vérifie :

```
assert (semT((1,1,0,1))==13)
```

Exercice 5. Au lieu de représenter les mots de B_0 par des tuples, on veut les représenter directement dans des chaînes de caractères PYTHON dans la notation sans « . ». Ainsi la chaîne PYTHON `'1101'` encode le mot « `1.1.0.1` ». Adapter « `semT` » en « `semS` » de sorte que :

```
assert (semS('1101')==13)
```

On pourra utiliser la fonction prédéfinie **int** pour convertir les chiffres `'0'` et `'1'` en nombre.

RENDU 1. Généraliser `semS` en une fonction `semF` qui retourne un flottant PYTHON sur les nombres binaires à virgule (avec un nombre fini de chiffres) – écrits sur le vocabulaire V à trois éléments : « 0 » et « 1 » et « , ». Testez que votre fonction produit des résultats corrects avec :

```
assert (semF('1101')==13)
assert (semF('1101,')==13)
assert (semF('1101,000')==13)
assert (semF('11,101')==3.625)    # voir le cours 1
assert (semF(',' + '1'*53)==0.9999999999999999)
assert (semF(',' + '1'*54)==1)
assert (semF(',' + '01'*27)==1/3)
```

Remarquons que le codage des flottants PYTHON sur 64 bits (voir https://en.wikipedia.org/wiki/Double-precision_floating-point_format) induit des résultats mathématiquement approximatifs.

Langage sur un vocabulaire V L'ensemble des mots sur V est noté V^*. Un langage L sur V est par définition un sous-ensemble de V^*.

Précision importante cette définition mathématique assimile un langage à une *syntaxe* (c'est-à-dire une notation, comme la notation décimale ou la notation binaire). Alors qu'au niveau informel (par exemple, quand on parle d'un « langage de programmation » donné), un langage correspond souvent à la combinaison d'une syntaxe avec une sémantique. Dans la suite du cours d'automates finis, on va se focaliser dans un premier temps sur les syntaxes. La prise en compte des sémantiques sera abordée en période 2.

Exercice 6. Expliciter le lien entre $\{0,1\}^*$ et B_0 de l'exo 4.

Exemple 5. Soit L un langage sur V. Étant donné un entier naturel n, on note $\lfloor L \rfloor_n$ l'ensemble des mots de L qui ont une longueur égale à n. Ainsi, $\lfloor L \rfloor_n$ est aussi un langage (fini) sur V. Avec les notations de la section 2, on définira simplement $\lfloor L \rfloor_n \stackrel{def}{=} \{w \in L \mid |w| = n\}$.

Exercice 7. Écrire en PYTHON une fonction `filtreB0(n)` qui affiche l'ensemble des mots de $\lfloor B_0 \rfloor_{\texttt{n}}$ (un par ligne). Par exemple, `filtreB0(3)` produit l'affichage ci-contre.

```
000
001
010
011
100
101
110
111
```

Indication : on écrira en fait une fonction `filtreB(n, pref)` qui, lorsque `pref` vaut un mot u, affiche l'ensemble des mots de la forme $u.w$ où w est un mot de $\lfloor \{0,1\}^* \rfloor_{\texttt{n}}$. Autrement dit `filtreB(n, u)` permet d'afficher *récursivement* l'ensemble des mots de $\lfloor \{0,1\}^* \rfloor_{\texttt{n}}$ en insérant un préfixe `u` devant chaque mot affiché. Avec les notations de la section 2, l'ensemble affiché est simplement $\{u\}.\lfloor \{0,1\}^* \rfloor_{\texttt{n}}$.

Dans la représentation B_0 de l'exo 4, tout mot w correspond au même entier que le mot $0.w$. En particulier, chaque entier peut donc avoir une représentation arbitrairement grande. On peut donc préférer la représentation sans zéro superflu, autrement dit définir l'ensemble B_1 des mots de B_0 tel que tout mot de longueur supérieur ou égale à 2 ne commence pas par `0`. Comme B_1 est un sous-ensemble de B_0, cela ne nécessite même pas d'adapter la sémantique.

100
101
110
111

RENDU 2. En utilisant directement la fonction `filtreB(n, pref)` introduite précédemment, définir une fonction PYTHON `filtreB1(n)` qui affiche $\lfloor B_1 \rfloor_n$. On vérifiera que `filtreB1(3)` produit l'affichage ci-contre.

Comme les langages sont des ensembles, la théorie des langages a besoin de s'appuyer sur une théorie des ensembles. Avant de continuer à formaliser la théorie des langages, il faut davantage formaliser la théorie des ensembles.

2 Théorie des ensembles

On définit ci-dessous les concepts mathématiques de base de la théorie des ensembles [6]. Pour faire le lien avec l'informatique, on programme aussi certains d'entre eux sur les ensembles finis de PYTHON3, et plus précisément les « *frozenset* » [7]. On observe que PYTHON3 offre des constructions syntaxiques directement inspirées de la théorie des ensembles, notamment les définitions par compréhension. [8] Ce type de constructions existent en fait de nombreux langages de programmation modernes comme HASKELL ou RUST. [9]

Une des difficultés dans la formalisation de la notion intuitive d'ensemble est d'éviter les ensembles *paradoxaux* qui mènent à une contradiction logique. Notamment le célèbre ensemble paradoxal de Russel [10] définit comme "*l'ensemble des ensembles qui n'appartiennent pas à eux-mêmes*" [11] qu'on pourrait noter informellement « $\{x \mid x \notin x\}$ ». Par exemple, un tel ensemble pourrait être construit par « $\{x \in \mathbb{E} \mid x \notin x\}$ » dans la théorie ci-dessous, en supposant l'existence d'un ensemble $\mathbb{E}$ de tous les ensembles. Ce qu'on évite donc de faire ! Ainsi, ce paradoxe s'avère très utile pour démontrer que certains ensembles, comme $\mathbb{E}$, n'existent pas (voir aussi à l'exo 25).

Pour éviter un tel risque de contradiction logique, on se limite à supposer l'existence de quelques constructions « *intuitives* » dont on se convainc à l'usage qu'elles ne sont pas contradictoires. On construit les autres ensembles à partir de ces constructions de base. C'est la méthode axiomatique. [12]

Égalité sur les ensembles (extensionnalité) On **suppose** que deux ensembles E et F sont indistinguables dans nos énoncés sur les ensembles si et seulement si $\forall\, x, x \in E \Leftrightarrow x \in F$. Dans ce cas, on note $E = F$ (l'égalité mathématique usuelle).

En PYTHON, lorsque `E` et `F` sont deux « *frozenset* », alors on peut tester leur égalité au sens mathématique avec l'expression « `E == F` ». Attention, dans d'autres cas, l'opérateur `==` de PYTHON ne correspond pas toujours à l'égalité mathématique (ce sera précisé à l'exemple 9).

Ensemble vide On **suppose l'existence** de l'ensemble vide, noté $\varnothing$ tel que $\forall\, x, x \notin \varnothing$.
En PYTHON, c'est « **frozenset**`()` ». Parfois, on notera aussi $\{\}$ au lieu de $\varnothing$.

6. Voir aussi https://fr.wikipedia.org/wiki/Th%C3%A9orie_des_ensembles.
7. Ce sont des ensembles finis **immutables**, comme ceux des ensembles mathématiques.
Voir https://docs.python.org/3/library/stdtypes.html#frozenset.
8. Voir le tutoriel https://docs.python.org/3/tutorial/datastructures.html#list-comprehensions.
9. PYTHON est aussi bien adapté pour programmer en théorie des ensembles, car il n'a pas de typage statique, comme la théorie des ensembles. Et il offre des fonctions de réflexion, exploitées en section 6.
10. Voir https://fr.wikipedia.org/wiki/Paradoxe_de_Russell
11. Le paradoxe est le suivant : cet ensemble appartient à lui-même si et seulement si il n'appartient pas à lui-même !
12. Voir https://fr.wikipedia.org/wiki/Syst%C3%A8me_axiomatique#M%C3%A9thode_axiomatique.
Historiquement, il y a eu une tentative de dériver toutes les constructions mathématiques à partir de cette théorie des ensembles. Mais la théorie des ensembles présentant certaines limites (notamment lorsqu'on cherche à considérer les « *structures algébriques* » comme objets mathématiques), d'autres théories alternatives ont été ensuite proposées. Une des dernières en date, apparue dans les années 2010, est l'univalence ; trouver une « bonne » théorie sur-laquelle fonder formellement les mathématiques est donc toujours un sujet de recherche.
Voir https://en.wikipedia.org/wiki/Univalent_foundations.

Ensemble des parties Pour tout ensemble E, on **suppose l'existence** de l'ensemble, noté $\mathcal{P}(E)$, des parties de E tel que $\forall\, A,\ A \in \mathcal{P}(E) \Leftrightarrow (\forall\, x, x \in A \Rightarrow x \in E)$.
Usuellement, on note aussi $A \subseteq E$ la proposition logique $A \in \mathcal{P}(E)$. En PYTHON, on écrit « `A <= E` ».
(On dit aussi que A est un sous-ensemble de E ou que A est une partie de E).

Exemples 6. $\mathcal{P}(\varnothing) = \{\varnothing\}$; $\mathcal{P}(\{1\}) = \{\varnothing, \{1\}\}$; $\mathcal{P}(\{1,2\}) = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$
et $\mathcal{P}(\{1,2,3\}) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$

Exercice 8. Calculer $\mathcal{P}(\{0,1,2,3\})$.

Ensemble par compréhension Pour tout ensemble E et toute proposition logique « P_x » portant sur une variable « x », on **suppose l'existence** de l'ensemble « $\{x \in E \mid P_x\}$ » qui est la partie $A \in \mathcal{P}(E)$ telle que $(\forall\, x, x \in A \Leftrightarrow P_x)$.

Exemple 7. Soit $E \subseteq \mathbb{N}$.
L'ensemble des entiers impairs de E est défini par $\{x \in E \mid x \bmod 2 = 1\}$.
Si E vaut $\{x \in \mathbb{N} \mid x < 10\}$, on peut calculer cet ensemble en PYTHON par :

```
E=frozenset(range(10))
frozenset(x for x in E if x % 2 == 1)
```

On peut aussi écrire plus simplement `frozenset(x for x in range(10) if x % 2 == 1)`.

Exercice 9. Étant donné n et m dans $\mathbb{Z}$, on note $\{n, \ldots, m\} \stackrel{def}{=} \{x \in \mathbb{Z} \mid n \leq x \leq m\}$.
(1) Donner une définition plus simple de « $\{n, \ldots, m\}$ » quand $n > m$.
(2) Donner une représentation en PYTHON de $\{n, \ldots, m\}$ (en tant que *frozenset*).

Union de deux ensembles Pour tous ensembles E et F, on **suppose l'existence** de l'ensemble, noté $E \cup F$, tel que $\forall\, x, x \in E \cup F \Leftrightarrow (x \in E \vee x \in F)$.
En PYTHON, l'union de « *frozenset* » est calculée par l'opérateur « `|` ». On peut vérifier qu'elle correspond à l'ensemble attendu avec :

```
assert (E | F == frozenset(x for b in range(2) for x in (E if b==0 else F)))
```

Intersection et différence Soient E et F deux ensembles. On définit :

intersection $E \cap F \stackrel{def}{=} \{x \in E \mid x \in F\}$.
En PYTHON, l'intersection de « *frozenset* » est calculée par l'opérateur « `&` ».

différence $E \backslash F \stackrel{def}{=} \{x \in E \mid x \notin F\}$.
En PYTHON, la différence de « *frozenset* » est calculée par l'opérateur « `-` ».

Exercice 10. Sur le modèle de l'union ci-dessus, donner deux assertions PYTHON testant une égalité d'ensembles qui vérifient que « `&` » et « `-` » sont corrects vis-à-vis de la définition mathématique.

Exercice 11. Soient E un ensemble avec x et y tels que $x \in E$ et $y \in E$. Définir :
(1) par compréhension l'ensemble singleton $\{x\}$, qui contient comme x unique élément.
(2) $\{x, y\}$, la plus petite partie de E contenant x et y.

Produit cartésien Pour tous ensembles E et F, on **suppose l'existence** de l'ensemble, noté $E \times F$, des paires d'éléments de E et F, tel que $\forall\, p,\ p \in E \times F \Leftrightarrow \exists\, x \in E, \exists\, y \in F, p = (x, y)$.

Exercice 12. Définir une fonction PYTHON « `prod(E,F)` » qui calcule le produit (en tant que « *frozenset* ») des ensembles finis `E` et `F`. Elle vérifie par exemple :

```
assert(prod(range(3),{'a','b'}) == frozenset({(0,'a'),(0,'b'),(1,'a'),(1,'b'),(2,'a'),(2,'b')}))
```

Compréhension dérivée de l'image d'une expression Soient E_1, …, E_n et F des ensembles. Soit $e_{x_1,\ldots,x_n}$ une expression qui dépend (éventuellement) de variables $x_1 \ldots x_n$, telle que $\forall\, x_1 \in E_1, \ldots \forall\, x_n \in E_n, e_{x_1,\ldots,x_n} \in F$.

On définit $\{e_{x_1,\ldots,x_n} \mid x_1 \in E_1 \wedge \ldots \wedge x_n \in E_n\} \stackrel{def}{=} \{y \in F \mid \exists\, x_1 \in E_1, \ldots, \exists\, x_n \in E_n, y = e_{x_1,\ldots,x_n}\}$ (où y est une variable qui n'apparaît pas dans le contexte).

Exercice 13. Soit $n \in \mathbb{N}$. On pose $P_n \stackrel{def}{=} \{\ (i,j) \mid i \in \{0,\ldots,n\} \wedge j \in \{0,\ldots,n-i\}\ \}$.

1. Écrire une fonction Python "**def** `P(n)`" qui retourne la liste des éléments de P_n (classée pour l'ordre croissant sur les couples de Python). Vérifier que `P(3)` donne le résultat attendu.
2. Calculer le nombre d'éléments de P_n (en fonction de n).
3. Montrer que $P_n = \{\ p \in \mathbb{N} \times \mathbb{N} \mid \exists\, i \in \mathbb{N}, \exists\, j \in \mathbb{N}, p = (i,j) \wedge i + j \leq n\ \}$.
4. Soit $P \stackrel{def}{=} \{\ p \in \mathbb{N} \times \mathbb{N} \mid \exists\, n \in \mathbb{N}, p \in P_n\ \}$ qu'on notera aussi $P = \bigcup_{n\in\mathbb{N}} P_n$.
   Montrer que $P = \mathbb{N} \times \mathbb{N}$.

Concaténation de langages Soient L_1 et L_2 deux langages sur un vocabulaire V. On pose $L_1.L_2 \stackrel{def}{=} \{w_1.w_2 \mid w_1 \in L_1 \wedge w_2 \in L_2\}$. On appelle $L_1.L_2$ la concaténation des langages L_1 et L_2 (il s'obtient en concaténant chacun des mots de L_1 avec chacun des mots de L_2).

Exercice 14. Donner la définition par compréhension primitive de $L_1.L_2$ (il s'agit essentiellement de trouver le bon F). Montrer ensuite comment simplifier $\{\varepsilon\}.L$ et $L.\{\varepsilon\}$.

Exercice 15. Montrer que $\lfloor V^* \rfloor_1 = V$ (cf. la notation de l'exemple 5). Remarquons que V est ici à la fois le vocabulaire et un langage sur V.

Itération de langages Soit n un entier naturel et L un langage (sur un vocabulaire V). On définit le langage (sur V) L^n par récurrence sur n, avec $L^0 \stackrel{def}{=} \{\varepsilon\}$ et $L^{n+1} \stackrel{def}{=} L.L^n$.

Itération de Kleene Soit L un langage (sur un vocabulaire V).

On définit le langage (sur V) L^* par $L^* \stackrel{def}{=} \{w \in V^* \mid \exists\, n \in \mathbb{N}, w \in L^n\}$ (qu'on aurait aussi pu noter $L^* = \bigcup_{n\in\mathbb{N}} L^n$).

On pose aussi $L^+ \stackrel{def}{=} L.L^*$.

Remarquons que cette définition « surcharge » la définition de V^* (V étant à la fois le vocabulaire et un langage sur V). Mais ce n'est pas un souci puisque cette deuxième (basée sur l'itération de Kleene) est cohérente avec la première (l'ensemble des mots sur V). [13]

Exercice 16. On considère le vocabulaire $V = \{\mathtt{0}, \mathtt{1}\}$. En utilisant uniquement des sous-ensembles finis de V et les opérateurs de concaténation et d'itération sur les langages, définir l'ensemble L des mots de V dont chaque symbole `0` est immédiatement suivi d'un symbole `1` comme dans `1101011101` ou `01111011`.

Exercice 17. Est-ce que les égalités suivantes sont valides (quelque soit L) ? Sinon, quelles sont les inclusions qui sont valides ? Justifier (avec des contre-exemples le cas échéant).

1. $L^* = \{w^n \mid w \in L \wedge n \in \mathbb{N}\}$?
2. $L^* = L^+ \cup \{\varepsilon\}$?
3. $L^+ = L^* \backslash \{\varepsilon\}$?

Exercice 18. En considérant le B_0 de l'exo 4, montrer que $B_0 = \{0,1\}^+$.

Exercice 19. En considérant que le B_1 du rendu 2 est formellement défini par $B_1 \stackrel{def}{=} B_0 \backslash (\{0\}.B_0)$, montrer que

$$B_1 = \{0\} \cup \{1\}.\{0,1\}^*$$

13. Autrement dit, on aurait pu appeler $\alpha(V)$ l'ensemble des mots sur V. Puis ayant défini l'itération de Kleene L^* d'un langage L en fonction de $\alpha(V)$, on aurait pu montrer $V^* = \alpha(V)$; et finalement oublier $\alpha(V)$ pour n'utiliser que V^*.

ce qui se lit par convention (le $\cup$ étant vu comme un opérateur additif moins prioritaire que les symboles multiplicatif) $B_1 = \{0\} \cup (\{1\}.(\{0,1\}^*))$.
En utilisant les égalités $(L_1 \cup L_2).L_3 = (L_1.L_3) \cup (L_2.L_3)$ et $L_3.(L_1 \cup L_2) = (L_3.L_1) \cup (L_3.L_2)$, on appliquera un raisonnement algébrique (c'est-à-dire exploitant des propriétés d'égalités) dans les langages (et les ensembles).

Exercice 20. On considère le vocabulaire $V = \{\texttt{0}, \texttt{1}, \texttt{'}, \texttt{'}\}$ (à trois symboles, le dernier représentant une virgule qu'on note entre apostrophes pour éviter la confusion avec le méta-symbole). En utilisant uniquement des sous-ensembles finis de $V \cup \{\varepsilon\}$ et les opérateurs d'union, de concaténation et d'itération sur les langages, définir un langage B_v des nombres à virgules, écrits en binaires avec un nombre fini de chiffres (dont la sémantique est donnée au rendu 1). On impose que tout mot de B_v contienne *au moins* un chiffre binaire (avant ou après la virgule) et *au plus* une virgule. Donc les 4 mots suivants sont dans B_v : « `00,00` », « `1` », « `10,` » et « `,010` ». Mais les deux mots suivants n'y sont pas : « `0,,` » et « `,` ».

Compréhension généralisée La forme précédente est généralisable avec une proposition $P_{x_1,\ldots,x_n}$ qui dépend de variables $x_1 \ldots x_n$ telle que $\forall\, x_1 \in E_1, \ldots \forall\, x_n \in E_n, P_{x_1,\ldots,x_n} \Rightarrow e_{x_1,\ldots,x_n} \in F$,
avec $\{e_{x_1,\ldots,x_n} \mid x_1 \in E_1 \wedge \ldots \wedge x_n \in E_n \wedge P_{x_1,\ldots,x_n}\}$
$\stackrel{def}{=} \{y \in F \mid \exists\, x_1 \in E_1 \ldots \exists\, x_n \in E_n, y = e_{x_1,\ldots,x_n} \wedge P_{x_1,\ldots,x_n}\}$
(où y est une variable qui n'apparaît pas dans le contexte).

Par exemple, l'item 3 de l'exo 13 aurait pu s'écrire :
« Montrer que $P_n = \{(i,j) \mid i \in \mathbb{N} \wedge j \in \mathbb{N} \wedge i + j \leq n\}$ ».

Exercice 21. Soit $V = \{\texttt{a}, \texttt{b}, \texttt{c}\}$. On pose $A = \{0,1,2,3\}$ et $B = \{\texttt{b}, \texttt{c}\}$. On définit le langage L sur V par

$$L \stackrel{def}{=} \{\texttt{a}^n.c.\texttt{b}^p \mid n \in A \wedge p \in \mathbb{N} \wedge c \in B \wedge n \leq p \leq 2n\}$$

Donner la définition par compréhension primitive de L et calculer son nombre d'éléments.

RENDU 3. Définir L de l'exo 21 à l'aide d'un schéma de compréhension Python en faisant références aux définitions de `A` et `B` ci-dessous :

```
A=range(4)
B=frozenset(('b','c'))
```

Pour faciliter le débogage, on pourra représenter L par un **tuple** (appelé `L`) plutôt qu'un **frozenset**. Vérifiez que le nombre d'éléments est bien celui attendu (via le **len** de Python).

Les deux variantes de définitions par compréhension introduites ci-dessus sont présentées d'une façon légèrement différente à la section suivante, dans le paragraphe "Image d'une application implicite".