TD Théorie des Langages 1 — Feuille 1
Induction structurelle, Notions de langage

Exercice 1 Soit V un vocabulaire et soit un sous-ensemble $A \subseteq V$. Dans cet exercice on considérera la fonction $|.|_A : V^* \to \mathbb{N}$ qui à tout mot $w \in V^*$ associe le nombre d'occurrences d'éléments de A présents dans w. Ainsi, si $V = \{a, b, c, d\}$ et $A = \{a, b\}$, alors :
- $|cabdbacc|_A = 4$,
- $|bbaba|_A = 5$,
- $|ccdc|_A = 0$.

On pose $V = \{\wedge, \vee, \neg,), (, \perp, \top\}$, et on considère l'ensemble E défini par induction structurelle de la façon suivante :

Base. $\top \in E$ et $\perp \in E$.

Induction. Si w, w_1 et w_2 sont dans E, alors :
- $(\neg\, w) \in E$,
- $(w_1 \wedge w_2) \in E$,
- $(w_1 \vee w_2) \in E$.

▷ Question 1 Est-ce que les deux mots suivants sont dans E ? On justifiera en quelques mots les réponses.

1. $(\top \wedge \neg)$
2. $(\top \wedge \top \wedge \top)$

▷ Question 2 Montrer que $((\top \wedge \top) \vee (\perp \vee \top)) \in E$.

On définit les ensembles de symboles suivants :

$$\begin{aligned} U &= \{\neg\}, \\ B &= \{\wedge, \vee\}, \\ S &= \{\top, \perp\}, \\ N &= U \cup B \cup S. \end{aligned}$$

▷ Question 3 Soit $w_1 = (\top \wedge ((\neg\perp) \vee (\top \wedge (\perp \vee (\neg\top)))))$. Calculer $|w_1|_U$, $|w_1|_B$ et $|w_1|_S$.

▷ Question 4 Démontrer par induction structurelle que pour tout $w \in E$, on a $|w|_N = 2|w|_B + |w|_U + 1$.

▷ Question 5 **[Avancé]** Soit $w \in E$. Exprimer $|w|$ en fonction de $|w|_B$ et $|w|_U$.

▷ Question 6 **[Avancé]** Utiliser la question 5 pour justifier formellement les réponses à la question 1.

Solution de l'Exercice 1.

▷ Question 1 Le symbole $\neg$ ne peut apparaître qu'après $'('$ et il manque des parenthèses pour le deuxième mot (pour chaque $\wedge$ il doit y avoir un $'('$).

▷ QUESTION 2 Pour montrer qu'un terme appartient à un ensemble défini par induction, il faut le construire en utilisant les constructeurs et les case de base. Notons $u = (\top \wedge \top)$ et $v = (\perp \vee \top)$. On a $u \in E$ par application du deuxième constructeur inductif à deux fois le premier cas de base. De même, $v \in E$ par application du troisième constructeur inductif au deuxième et premier cas de base. Au final, le mot $((\top \wedge \top) \vee (\perp \vee \top))$ appartient à E par application du troisième constructeur inductif à u et v.

▷ QUESTION 3 On a $|w_1|_U = 2$, $|w_1|_B = 4$ et $|w_1|_S = 5$.

▷ QUESTION 4 **Rappel :** Pour faire une démonstration par induction structurelle d'une propriété P sur un ensemble E défini inductivement, on doit démontrer P pour les cas de base et démontrer que pour chaque constructeur inductif κ, si on suppose P pour tous les arguments $u_1, \ldots, u_k$ de κ (d'arité k), alors on peut démontrer P sur le mot $\kappa(u_1, \ldots, u_k)$.

Ici, on a deux cas de base et trois constructeurs inductifs, donc cinq cas à considérer.

Montrons la propriété[1] $P[w] = |w|_N = 2|w|_B + |w|_U + 1$ par induction structrelle sur E.

— Pour $\top$: $|\top|_N = 1 = 2*0+0+1 = 2|\top|_B + |\top|_U + 1$.

— Pour $\perp$: $|\perp|_N = 1 = 2*0+0+1 = 2|\perp|_B + |\perp|_U + 1$.

— Pour $(\neg u)$: Soient u dans E et supposons $P[u]$. On a

$$\begin{aligned} |(\neg u)|_N &= 1 + |u|_N \\ &\overset{HI}{=} 1 + (2|u|_B + |u|_N + 1) \\ &= 2|u|_B + +(|u|_U + 1) + 1 \\ &= 2|(\neg u)|_B + |(\neg u)|_U + 1 \end{aligned}$$

Ainsi, on a montré $P[(\neg u)]$.

— Pour $(u \wedge v)$: Soient u, v dans E et supposons $P[u]$ et $P[v]$. On a

$$\begin{aligned} |(u \wedge v)|_N &= 1 + |u|_N + |v|_N \\ &\overset{HI}{=} 1 + (2|u|_B + |u|_N + 1) + (2|v|_B + |v|_N + 1) \\ &= (2|u|_B + 2|v|_B + 2) + (|u|_U + |v|_U) + 1 \\ &= 2|(u \wedge v)|_B + |(u \wedge v)|_U + 1 \end{aligned}$$

Ainsi, on a montré $P[(u \wedge v)]$.

— Pour $(u \vee v)$: Soient u, v dans E et supposons $P[u]$ et $P[v]$. On a

$$\begin{aligned} |(u \vee v)|_N &= 1 + |u|_N + |v|_N \\ &\overset{HI}{=} 1 + (2|u|_B + |u|_N + 1) + (2|v|_B + |v|_N + 1) \\ &= (2|u|_B + 2|v|_B + 2) + (|u|_U + |v|_U) + 1 \\ &= 2|(u \vee v)|_B + |(u \vee v)|_U + 1 \end{aligned}$$

Ainsi, on a montré $P[(u \vee v)]$.

▷ QUESTION 5 On a $|w| = 4|w|_B + 3|w|_U + 1$. En effet, $|w| = |w|_N + |w|_{\{),(\}} = (2|w|_B + |w|_U + 1) + (2|w|_B + 2|w|_U)$.

▷ QUESTION 6 L'égalité de la question précédente n'est pas vérifiée pour les mots de la question 1, ils ne font donc pas partie de E.

1. J'utilise les crochets [_] car les parenthèses font partie du vocabulaire V.

Exercice 2 Définir la fonction $|.|_A$ de l'exercice 1 par induction structurelle sur V^*.

Solution de l'Exercice 2.

Base : $|\varepsilon|_A = 0$.

Induction : Pour tout mot w et tout symbole x,

$$|xw|_A = \begin{cases} 1 + |w|_A & \text{si } x \in A \\ |w|_A & \text{sinon} \end{cases}$$

Exercice 3 Définitions inductives d'ensembles.

▷ Question 1 Donner des définitions inductives des langages suivants :

1. L'ensemble L_1 des mots sur $\{a, b\}$ de longueur paire.
2. L'ensemble L_2 des mots sur $\{a, b\}$ ne contenant pas deux a consécutifs.
3. L'ensemble L_3 des palindromes sur $\{a, b\}$.
4. **[Avancé]** L'ensemble L_4 des mots sur $\{a, b\}$ contenant un nombre pair de a.
5. **[Avancé]** L'ensemble L_5 des mots sur $\{a, b\}$ contenant autant de a que de b.

▷ Question 2 **[Avancé]** Prouver que ces définitions inductives sont correctes.

Solution de l'Exercice 3.

▷ Question 1

1. Soit M_1 l'ensemble défini par induction de la façon suivante :

   Base : $\varepsilon \in M_1$.

   Induction : si $w \in M_1$, alors aaw, abw, baw et bbw sont également dans M_1.

   Remarque : Pour les cas inductifs, on aurait pu choisir de placer les deux lettres derrière w (waa, wab, wba, wbb) voire une de chaque côté (awa, awb, bwa, bwb).
2. Soit M_2 l'ensemble défini par induction de la façon suivante :

   Base : $\{\varepsilon, a\} \subseteq M_2$.

   Induction : si $w \in M_2$, alors bw et abw sont également dans M_2.

   Remarque : Idem, on aurait pu prendre wb, wba pour les cas inductifs.
3. Soit M_3 l'ensemble défini par induction de la façon suivante :

   Base : $\{\varepsilon, a, b\} \in M_3$.

   Induction : si $w \in M_3$, alors awa et bwb sont également dans M_3.
4. Soit M_4 l'ensemble défini par induction de la façon suivante :

   Base : $\varepsilon \in M_4$.

   Induction : si $w \in M_4$, alors bw, wb et awa sont également dans M_4.

5. Soit M_5 l'ensemble défini par induction de la façon suivante :

   Base : $\varepsilon \in M_5$.

   Induction : si w_1 et w_2 sont des éléments de M_5, alors aw_1b, bw_1a et w_1w_2 sont également dans M_5.
   Variante (celle présentée en cours) : si w_1 et w_2 sont de éléments de M_5, alors aw_1bw_2 et bw_1aw_2 sont dans M_5

▷ QUESTION 2 Les preuves que les M_i sont inclus dans les L_i se font par induction structurelle. Nous détaillons la preuve pour $i = 1$, il s'agit donc de prouver que $M_1 \subseteq L_1$ par induction structurelle.

— On montre tout d'abord que la propriété est vérifiée pour tous les cas de base. On a $|\varepsilon| = 0$, donc $\varepsilon \in L_1$.

— On montre ensuite que la propriété est préservée par les règles de construction. Soit $w \in M_1$, et supposons que $w \in L_1$. Alors $|aaw| = |w| + 2$ est pair, donc aaw est bien élément de L_1. On prouve de la même manière que abw, baw et bbw sont dans L_1.

Les preuves que les définitions inductives engendrent bien tous les mots des langages correspondants se feront par récurrence bien fondée sur la longueur des mots.

1. Prouvons que $L_1 \subseteq M_1$. Soit $w \in L_1$ de longueur n, et supposons que pour tout $w' \in L_1$, si $|w'| < n$, alors $w' \in M_1$.

   Si $n = 0$, alors $w = \varepsilon$ et il est clair que $w \in M_1$. Il est également clair que si $w \in L_1$, alors on ne peut pas avoir $n = 1$. On suppose maintenant que $w = uvw'$, où u, v sont des lettres dans $\{a, b\}$. Sans perte de généralité, on peut supposer que $u = a$ et $v = b$, les autres cas sont similaires. Comme $w' \in L_1$ est nécessairement de longueur paire et que $|w'| = |w| - 2 < |w|$, par hypothèse d'induction, $w' \in M_1$. Donc, le mot abw' est également dans M_1 d'après les règles de construction de M_1, ce qui prouve que $w \in M_1$.

   On en déduit que $L_1 \subseteq M_1$.

2. Prouvons que $L_2 \subseteq M_2$. Soit $w \in L_2$ de longueur n, et supposons que pour tout $w' \in L_2$, si $|w'| < n$, alors $w' \in M_2$.

   Si $n \in \{0, 1\}$, alors $w \in \{\varepsilon, a, b\}$, et il est aisé de vérifier que $w \in M_2$. Supposons maintenant $n > 1$ et distinguons deux cas.

   — Si $w = bw'$, alors w' ne peut pas contenir deux a consécutifs et est donc élément de L_2. Comme $|w'| = n - 1$, on en déduit que $w' \in M_2$, puis que $bw' \in M_2$.

   — Sinon $w = aw'$, et nécessairement $w' = bw''$ car $n > 1$ par hypothèse. Donc $w = abw''$ et w'' ne peut pas contenir deux a consécutifs. Donc $w'' \in M_2$, et $abw'' \in M_2$.

3. Prouvons que $L_3 \subseteq M_3$. Soit $w \in L_3$ de longueur n, et supposons que pour tout $w' \in L_3$, si $|w'| < n$, alors $w' \in M_3$.

   Si $n \in \{0, 1\}$, alors il est aisé de vérifier que $w \in M_3$.

   Supposons maintenant que la première lettre de w soit un a, le cas où cette première lettre est un b est similaire. Alors comme w est un palindrôme, sa dernière lettre est également un a, donc w est de la forme $aw'a$. Comme w' est nécessairement un palindrôme, de longueur strictement inférieure à n, on a $w' \in M_3$, et donc, $aw'a \in M_3$.

4. Prouvons que $L_4 \subseteq M_4$. Soit $w \in L_4$ de longueur n, et supposons que pour tout $w' \in L_4$, si $|w'| < n$, alors $w' \in M_4$.

   Il est clair que si $w = \varepsilon$, alors $w \in M_4$.
   - Si $w = bw'$, alors w' doit contenir un nombre pair de a, donc par hypothèse d'induction $w' \in M_4$ et on en déduit que $bw' \in M_4$.
   - Si $w = w'b$, alors le même raisonnement que précédemment prouve que $w'b \in M_4$.
   - Supposons maintenant que $w = aw'a$. Alors une fois encore, w' doit contenir un nombre pair de a et est bien élément de M_4 ; on en déduit que $aw'a \in M_4$.

5. Prouvons que $L_5 \subseteq M_5$. Soit $w \in L_5$ de longueur n, et supposons que pour tout $w' \in L_5$, si $|w'| < n$, alors $w' \in M_5$.

   Pour la première version : il est clair que si $w = \varepsilon$, alors $w \in M_5$. Supposons maintenant que la première lettre de w est a, la preuve dans le cas où la première lettre est b est similaire.
   - Si w est de la forme aw_1b, alors nécessairement, w_1 contient autant de a que de b. Ce mot étant de longueur $n - 2$, on en déduit que $w_1 \in M_5$, et qu'on a bien $aw_1b \in M_5$.
   - Supposons que w est de la forme aw_1a, et considérons l'ensemble

     $$A \stackrel{\text{def}}{=} \{w' \mid w' \text{ est un préfixe strict de } w \text{ et } |w'|_a \leq |w'_b|\}.$$

     L'ensemble A est non-vide car aw_1 en est un élément. Prenons l'élément de A de longueur minimale ; cet élément est de la forme aw_2, où $w_2 \in \{a,b\}^*$, et w est de la forme aw_2w_3a. Comme aw_2 est de longueur minimale dans A, on en déduit que $|aw_2|_a = |aw_2|_b$ (sinon en enlevant sa dernière lettre, on obtient toujours un élément de A). Nécessairement, on a aussi $|w_3a|_a = |w_3a|_b$, puisque $|w|_a = |w|_b$. Ceci signifie que aw_2 et w_3a sont tous deux éléments de L_5, et comme $|aw_2| < |w|$ et $|w_3a| < |w|$, ces mots sont également éléments de M_5. Par la suite, aw_2w_3a est nécessairement dans M_5 d'après la dernière règle de construction de M_5.

   Pour la variante (preuve non présentée en cours) : si $|w| = 0$, alors $w = \varepsilon \in M_5$. Pour $|w| > 0$, supposons $w = aw'$ (le cas $w = bw'$ est identique, en remplaçant tous les a par b). Soit x le plus court des préfixes y de w' tels que $|y|_b > |y|_a$ (forcément $x \neq \varepsilon$). Noter que x existe car $|w'|_b > |w'|_a$. On a donc $w = axw_2$. x ne peut pas terminer par a, sinon $x = za$ et $|z|_b > |z|_a$. avec z plus court que x. Donc x termine par b : $x = w_1b$. Maintenant, $|x|_b = |x|_a + 1$ car sinon $|w_1|_b > |w_1|_a$, avec w_1 plus court que x. On en déduit $|w_1|_b = |w_1|_a$, donc $w_1 \in L_5$, et donc aussi $w_2 \in L_5$. On a donc $w = aw_1bw_2$. L'HI nous dit que w_1 et w_2 sont dans M_5, donc par induction w aussi.