Langages Formels – partiel (corrigé)

durée : 2 heures

jeudi 19 mars 2026, 14h

Consignes. Aucun document autorisé. Vous pouvez utiliser les résultats vus en cours ou dans des questions précédentes, si vous les citez de manière claire et non ambiguë. Les résultats vu en TD ou ailleurs qu'en cours, s'ils sont utilisés, doivent êtres redémontrés. Seules les notions et notations du cours et du sujet seront admises dans les réponses. La lisibilité et la concision seront appréciées.

Pour tout ce sujet, on fixe un alphabet fini Σ avec $|\Sigma| \geq 2$.

Définition A. Pour $u, v \in \Sigma^*$, on écrit $u \sqsubseteq v$ si u est un sous-mot de v. C'est-à-dire que si $u = u_1 \dots u_n$, avec $u_1, \dots, u_n \in \Sigma$, alors $u \sqsubseteq v$ si, et seulement si, $v \in \Sigma^* u_1 \Sigma^* \cdots \Sigma^* u_n \Sigma^*$. On écrit $u \sqsubset v$ si $u \sqsubseteq v$ et $u \neq v$. Pour tout $S \subseteq \Sigma^*$, on note $\uparrow S$ la clôture vers le haut de S pour $\sqsubseteq$, c'est-à-dire

$$\uparrow S := \{w \in \Sigma^* \mid \text{ il existe } s \in S \text{ tel que } s \sqsubseteq w\}.$$

Lorsque $S = \{w\}$, on écrit $\uparrow w$ au lieu de $\uparrow\{w\}$.

Question 1. Donner une grammaire linéaire à droite pour le langage $\uparrow abb$, sur l'alphabet $\{a, b\}$.

Solution. $S \to bS$, $S \to aT$, $T \to aT$, $T \to bU$, $U \to aU$, $U \to bV$, $V \to aV$, $V \to bV$, $V \to \epsilon$. □

Définition B. Soit $L \subseteq \Sigma^*$. Le langage L est *positivement testable par morceaux* (ptpm) si L est fermé vers le haut pour $\sqsubseteq$, c'est-à-dire que pour tout $u \in L$ et $v \in \Sigma^*$, si $u \sqsubseteq v$ alors $v \in L$. On note PTPM la classe des langages positivement testable par morceaux (sur l'alphabet fixé Σ).

Définition C. Soit R une relation binaire sur un ensemble S. Une *antichaîne* pour R est un sous-ensemble A de S tel que, pour tous $a, b \in A$, on n'ait pas aRb. On dit que R satisfait la *condition de chaîne descendante* (DCC) s'il n'existe pas de suite $(s_n)_{n\in\mathbb{N}}$ telle que $s_{n+1} R s_n$ pour tout $n \in \mathbb{N}$. On dit que R est un *wqo* (well-quasi-order) si R satisfait DCC et il n'existe pas d'antichaine infinie pour R.

On admet le fait suivant sans preuve.

Lemme D (Higman)**.** *La relation $\sqsubset$ est un wqo.*

Question 2. Prouver que, pour tout langage ptpm L, il existe un ensemble fini F tel que $L = \uparrow F$.

Solution. Soit F l'ensemble des éléments minimaux de L. Comme $\sqsubset$ satisfait la DCC, si $w \in L$, alors il existe $f \in F$ tel que $f \sqsubseteq w$. Ainsi $L = \uparrow F$. De plus, F est fini, sinon il constituerait une antichaîne infinie pour $\sqsubseteq$. □

Question 3. Montrer que l'ensemble des langages ptpm est inclus dans l'ensemble des langages sans étoile, et que cette inclusion est stricte.

Solution. Par la Question 2, il suffit de montrer que $\uparrow u$ est sans étoile pour tout $u \in \Sigma^*$. Par définition de la relation de sous-mot,

$$\uparrow u = \Sigma^* u_1 \Sigma^* \cdots \Sigma^* u_n \Sigma^*$$

où $u = u_1 \dots u_n$ avec $u_1, \dots, u_n \in \Sigma$.

Tout langage fini est sans étoile, mais le seul langage fini et ptpm est le langage vide. □

Définition E. Pour un langage $L \subseteq \Sigma^*$, on définit le *préordre syntaxique* $\preceq_L$ sur Σ^* par, pour $x, y \in \Sigma^*$,

$$x \preceq_L y \iff \text{pour tous } \alpha, \beta \in \Sigma^*, \text{ si } \alpha y \beta \in L \text{ alors } \alpha x \beta \in L.$$

Question 4. Prouver que L est ptpm si, et seulement si, pour tous $u, v \in \Sigma^*$ et $a \in \Sigma$, on a

$$uav \preceq_L uv.$$

Solution. $\Rightarrow$. Soient $\alpha, \beta \in \Sigma^*$ tels que $\alpha uv\beta \in L$. Comme L est ptpm, on a aussi $\alpha uav\beta \in L$. Ainsi $uav \preceq_L uv$.

$\Leftarrow$. Soient $w \in L$ et $w' \in \Sigma^*$ tels que $w \sqsubseteq w'$. Alors w' peut être obtenu à partir de w par un nombre fini de réécritures de la forme $uv \to uav$. L'hypothèse implique que si $uv \in L$, alors $uav \in L$. Par récurrence sur le nombre d'étapes, si $w \in L$ alors $w' \in L$. □

Rappelons que la relation de Green $\mathcal{R}$ est définie par $u\mathcal{R}v$ si, et seulement si, il existe $x, y \in M$ tel que $ux = v$ et $vy = u$, la relation $\mathcal{L}$ est définie par $u\mathcal{L}v$ si, et seulement si, il existe $x, y \in M$ tel que $xu = v$ et $yv = u$, et la relation $\mathcal{J}$ est définie par $u\mathcal{J}v$ si, et seulement si, il existe $x, x', y, y' \in M$ tel que $xux' = v$ et $yvy' = u$.

Définition F. Un monoïde fini M est *$\mathcal{R}$-trivial* si, pour tous $u, v \in M$, $u\mathcal{R}v$ implique $u = v$, *$\mathcal{L}$-trivial* si, pour tous $u, v \in M$, $u\mathcal{L}v$ implique $u = v$, et *$\mathcal{J}$-trivial* si, pour tous $u, v \in M$, $u\mathcal{J}v$ implique $u = v$.

Remarquons que la congruence syntaxique $\sim_L$ est $\preceq_L \cap \succeq_L$, donc le monoïde syntaxique M_L peut être décrit comme $\Sigma^*/(\preceq_L \cap \succeq_L)$.

Question 5. Montrer que, pour tout langage ptpm L, le monoïde syntaxique M_L est $\mathcal{R}$-trivial.

Solution. Soit F fini tel que $L = {\uparrow}F$. Supposons $[x]_{\sim_L} \mathcal{R} [y]_{\sim_L}$. On choisit α, β tels que $x \sim_L y\alpha$ et $y \sim_L x\beta$. On montre que $x \sim_L y$. Supposons $uxv \in L$. Alors $ux\beta v \in L$ car L est ptpm. Comme $y \sim_L x\beta$, on a $uyv \in L$. L'autre direction est analogue. □

Question 6. Soit M un monoïde fini et supposons que $n \in \mathbb{N}_{\geq 1}$ est tel que, pour tout $x \in M$, x^n est idempotent. Montrer que les propriétés suivantes sont équivalentes :

1. M est $\mathcal{R}$-trivial ;
2. pour tous $x, y, u \in M$, si $uxy = u$, alors $ux = u$;
3. pour tous $x, y \in M$, $(xy)^n x = (xy)^n$.

Solution. (1) $\Rightarrow$ (2). Si $uxy = u$, alors $ux\mathcal{R}u$, donc $ux = u$.

(2) $\Rightarrow$ (3). Soient $x, y \in M$. Alors

$$(xy)^n = (xy)^{2n} = (xy)^n xy (xy)^{n-1}.$$

En appliquant (2) avec $u = (xy)^n$, $x = x$ et $y = y(xy)^{n-1}$, on obtient $(xy)^n x = (xy)^n$.

(3) $\Rightarrow$ (1). Soient $u, v, x, y \in M$ tels que $ux = v$ et $vy = u$. Alors $uxy = u$. En itérant cette hypothèse n fois, on obtient $u(xy)^n = u$, donc

$$v = ux = u(xy)^n x = u(xy)^n = u.$$ □

Rappelons que *la logique du premier ordre sur les mots* a un prédicat binaire $<$ et, pour chaque lettre $a \in \Sigma$, un prédicat unaire a.

Définition G. Une *phrase existentielle* est une formule de la logique du premier ordre sur les mots de la forme

$$\exists x_1 \dots \exists x_n \psi$$

où $n \geq 0$ et ψ est une formule ne contenant pas de quantificateurs, et ne contenant que les variables $x_1, \dots, x_n$. Un langage L est *définissable existentiellement* s'il existe une phrase existentielle ϕ telle que

$$L = \{w \in \Sigma^* \mid w \models \phi\}.$$

Remarquons que $\top$ est une phrase existentielle, donc Σ^* est définissable existentiellement.

Question 7. Prouver qu'un langage $L \subseteq \Sigma^*$ est ptpm si, et seulement si, L est définissable existentiellement.

Solution. $\Rightarrow$. Soit $w = w_1 \dots w_k$ avec $w_1, \dots, w_k \in \Sigma$. Le langage $\uparrow w$ est défini par la phrase existentielle

$$\exists x_1 \dots \exists x_k \left(w_1(x_1) \wedge \bigwedge_{i=2}^{k} [x_{i-1} < x_i \wedge w_i(x_i)] \right).$$

Pour un ensemble fini F, le langage $\uparrow F$ est défini par une disjonction finie de telles phrases, et la disjonction peut être ramenée à l'intérieur des quantificateurs existentiels.

$\Leftarrow$. Soit $\phi = \exists x_1 \dots \exists x_n \psi$ une phrase existentielle définissant L, avec ψ sans quantificateurs. Nous utilisons la Question 4. Supposons $uv \in L$. Alors $uv \models \phi$. Choisissons une valuation $V : x_1, \dots, x_k \to |uv|$ telle que $uv, V \models \psi$. Définissons une valuation $V' : x_1, \dots, x_k \to |uav|$ par

$$V'(x_i) := \begin{cases} V(x_i) & \text{si } V(x_i) < |u|, \\ V(x_i) + 1 & \text{si } V(x_i) \geq |u|. \end{cases}$$

Alors $uav, V' \models \psi$, car (uv, V) est (isomorphe à) un sous-modèle de (uav, V'), et ψ est sans quantificateurs. Ainsi $uav \models \phi$. □

Question 8. Montrer que le complémentaire d'un langage ptpm n'est pas nécessairement ptpm.

Solution. Considérons le langage $L := \Sigma^* a \Sigma^*$, qui est clairement ptpm. Son complémentaire $L^{\complement}$ n'est pas ptpm : par exemple, le mot vide appartient à $L^{\complement}$, mais le mot d'une lettre a n'appartient pas à $L^{\complement}$. □

Question 9. Calculer l'automate minimal et le monoïde syntaxique du langage $(\uparrow ab)^{\complement}$.

Solution. Le monoïde syntaxique de $(\uparrow ab)^{\complement}$ a comme éléments $\{1, a, b, ab, ba\}$. Tous les éléments sauf ba sont idempotents, ab est absorbant, et $bab = (ba)^2 = ab$.

L'automate déterministe minimal du langage possède trois états q_1, q_a, q_{ab}, avec q_1 initial, q_1 et q_a acceptants, et les fonctions de transition

$$a := (q_a, q_a, q_{ab}), \quad b := (q_1, q_{ab}, q_{ab}).$$ □

Définition H. Un langage $L \subseteq \Sigma^*$ est *testable par morceaux* (tpm) s'il est une combinaison booléenne finie de langages ptpm. Plus explicitement : pour $\mathcal{K} \subseteq \mathcal{P}(\Sigma^*)$, on note $\mathcal{B}(\mathcal{K})$ le plus petit ensemble de langages fermé par unions finies et complémentation tel que $\mathcal{K} \subseteq \mathcal{B}(\mathcal{K})$. La classe des langages testable par morceaux est alors définie par $\mathsf{TPM} := \mathcal{B}(\mathsf{PTPM})$.

Rappelons qu'un monoïde N *divise* un monoïde M s'il existe un sous-monoïde M' de M et un morphisme surjectif de M' sur N.

Question 10. Soient $\sim_0, \sim_1, \sim_2$ des congruences de monoïde sur Σ^*, et supposons que

$$\sim_1 \cap \sim_2 \subseteq \sim_0 .$$

Montrer que $\Sigma^*/\sim_0$ divise $\Sigma^*/\sim_1 \times \Sigma^*/\sim_2$.

Solution. Soit M' l'image du morphisme

$$\Sigma^* \to \Sigma^*/\sim_1 \times \Sigma^*/\sim_2$$

qui envoie w sur $\langle [w]_{\sim_1}, [w]_{\sim_2} \rangle$. Alors M' est clairement un sous-monoïde. Définissons maintenant $f : M' \to M/\sim_0$ par $f(\langle [w]_{\sim_1}, [w]_{\sim_2} \rangle) := [w]_{\sim_0}$. Cette application est bien définie par l'hypothèse, et elle est surjective et morphisme de monoïdes, par un résultat vu en cours. □

Question 11. Montrer que, pour tous langages $L_1, L_2 \subseteq \Sigma^*$, le monoïde syntaxique $M_{L_1 \cup L_2}$ divise $M_{L_1} \times M_{L_2}$.

Solution. Remarquons que $\sim_{L_1} \cap \sim_{L_2} \subseteq \sim_{L_1 \cup L_2}$: si $w \sim_{L_1} w'$ et $w \sim_{L_2} w'$, et si $\alpha w \beta \in L_1 \cup L_2$, alors $\alpha w \beta \in L_i$ pour un certain $i \in 1, 2$. Par hypothèse, on a alors aussi $\alpha w' \beta \in L_i$, et donc $\alpha w' \beta \in L_1 \cup L_2$. On applique alors la Question 10. □

Question 12. Montrer que, si deux monoïdes finis M_1, M_2 sont $\mathcal{R}$-triviaux, alors leur produit $M_1 \times M_2$ est $\mathcal{R}$-trivial.

Solution. Si $\langle x_1, x_2 \rangle \mathcal{R} \langle y_1, y_2 \rangle$, alors $x_1 \mathcal{R} y_1$ et $x_2 \mathcal{R} y_2$. Puisque M_1 et M_2 sont $\mathcal{R}$-triviaux, on obtient $x_1 = y_1$ et $x_2 = y_2$. □

Nous admettons sans preuve le lemme suivant.

Lemme I. *Si M est $\mathcal{R}$-trivial et N divise M, alors N est $\mathcal{R}$-trivial.*

Question 13. Montrer que, pour tout langage testable par morceaux L, le monoïde syntaxique M_L est $\mathcal{R}$-trivial.

Solution. Considérons l'ensemble

$$\mathcal{S} := \{L \subseteq \Sigma^* \mid M_L \text{ est } \mathcal{R}\text{-trivial}\}.$$

La Question 5 montre que tout langage ptpm appartient à $\mathcal{S}$. Les Question 11, 12 et le Lemme I montrent que si $L_1, L_2 \in \mathcal{S}$, alors $L_1 \cup L_2 \in \mathcal{S}$. Il est immédiat que $\mathcal{S}$ est fermée par complémentation, puisque $M_{L^\complement} = M_L$. □

Nous admettons le fait suivant.

Lemme J. *Un monoïde fini M est $\mathcal{J}$-trivial si, et seulement si, M est à la fois $\mathcal{L}$-trivial et $\mathcal{R}$-trivial.*

Pour le reste de ce sujet, on fixe un monoïde fini $\mathcal{J}$-trivial M et un morphisme $f : \Sigma^* \to M$.

Définition K. Soit $w \in \Sigma^*$ un mot de longueur k. Pour $0 \leq i \leq k$, on note $w^{(i)}$ le préfixe de longueur i de w (en particulier $w^{(0)}$ est le mot vide et $w^{(k)} = w$). On dit qu'un entier $1 \leq i \leq k$ est *stable à droite dans w* si $f(w^{(i-1)}) = f(w^{(i)})$, et qu'il est *instable à droite* sinon. On définit $r(w)$ comme le sous-mot de w constitué uniquement des lettres situées aux positions instables à droite. Formellement, si $w = w_1 \dots w_k$ avec $w_i \in \Sigma$, alors $r(w) := w_{t_1} \dots w_{t_\ell}$, où $1 \leq t_1 < \dots < t_\ell \leq k$ est la liste ordonnée des positions instables à droite. Le mot w est *réduit à droite* si $r(w) = w$, c'est-à-dire si toute position $1 \leq i \leq k$ est instable à droite. Lorsque $k = 0$, on pose $r(\epsilon) = \epsilon$ et on considère ϵ comme réduit à droite.

On définit de manière analogue, en utilisant les suffixes au lieu des préfixes, les notions de *positions stables à gauche*, *instables à gauche*, de mots *réduits à gauche*, et la fonction de *réduction gauche* ℓ.

Question 14. Soit $k \geq 1$ et $w = w_1 \dots w_k$ avec $w_1, \dots, w_k \in \Sigma$. Montrer que, si $1 \leq i < j \leq k$ est tel que $f(w^{(i)}) = f(w^{(j)})$, alors $i+1$ est stable à droite dans w.

Solution. Considérons $u := f(w^{(i)}), x := f(w_{i+1}), y := f(w_{i+2} \dots w_j)$. Alors $uxy = f(w^{(j)}) = f(w^{(i)}) = u$. Par $\mathcal{R}$-trivialité (Question 6), on obtient $ux = u$, c'est-à-dire $f(w^{(i+1)}) = f(w^{(i)})$. □

Question 15. Prouver qu'il existe au plus $|\Sigma|^{|M|}$ mots réduits à droite.

Solution. La Question 14 montre que, si $w \in \Sigma^*$ est réduit à droite, alors $f(w^{(i)}) \neq f(w^{(j)})$ pour tous $1 \leq i < j \leq k$. Ainsi $k \leq |M|$. □

Question 16. Soit $w \in \Sigma^*$ et supposons que $r(w)a \sqsubseteq w$. Montrer que $f(wa) = f(w)$.

Solution. Si $r(w) = \epsilon$, alors $1_M = f(b) = f(w)$ pour toute lettre $b \in \Sigma$ apparaissant dans w, et donc en particulier $f(wa) = f(w) \cdot f(a) = 1_M = f(w)$.

Supposons maintenant que $k > 0$ et $w = a_1 \dots a_k$ avec chaque $a_i \in \Sigma$, et soit t un entier instable dans w. Soit t' le dernier entier instable précédant t. Si t est le premier entier instable dans w, on pose $t' = 0$.

Nous affirmons que la lettre a_t ne peut apparaître nulle part strictement entre les entiers t' et t. En effet, comme pour tout $t' < j < t$ l'entier j est stable dans w, on a $f(w^{(t')}) = f(w^{(j-1)}a_j) = f(w^{(t')}a_j)$, tandis que $f(w^{(t)}) = f(w^{(t-1)}a_t) = f(w^{(t')}a_t) \neq f(w^{(t')})$.

Le paragraphe précédent implique que si $\phi\colon \{1, \dots, \ell\} \to \{1, \dots, k\}$ est un plongement du mot $r(w) = a_{t_1} \dots a_{t_\ell}$ comme sous-mot de w, alors pour tout $1 \leq i \leq \ell$ on a $\phi(i) \geq t_i$. Soit maintenant $\psi\colon \{1, \dots, \ell+1\} \to \{1, \dots, k\}$ un plongement de $r(w)a$ comme sous-mot de w. Alors ψ doit envoyer la dernière position vers un indice $j > t_\ell$. Comme toutes les positions après t_ℓ sont stables, on a en particulier $f(w^{(j-1)}) = f(w^{(j)}) = f(w)$. Ainsi $f(wa) = f(w^{(j-1)}a) = f(w^{(j)}) = f(w)$. □

Définition L. On définit $\mathcal{F} := \{ra\ell \mid r \in RD,\ a \in \Sigma,\ \ell \in RG\}$, où RD est l'ensemble des mots réduits à droite et RG l'ensemble des mots réduits à gauche. Pour $w \in \Sigma^*$, on définit $F(w) := \{u \in \mathcal{F} \mid u \sqsubseteq w\}$.

Question 17. Soient $u, v \in \Sigma^*$ et $a \in \Sigma$. Montrer que si $f(uv) \neq f(uav)$, alors $F(uv) \subsetneq F(uav)$.

Solution. Il est clair que $F(uv) \subseteq F(uav)$. Raisonnons par contraposée et supposons $F(uv) = F(uav)$. Observons que $r(u)a\ell(v) \in F(uav)$. Alors $r(u)a\ell(v) \in F(uv)$, donc $r(u)a\ell(v)$ est un sous-mot de uv. Cela implique que soit $r(u)a \sqsubseteq u$, soit $a\ell(v) \sqsubseteq v$, selon l'endroit où le plongement envoie la position "centrale" a. Supposons $r(u)a \sqsubseteq u$; l'autre cas est symétrique. La Question 16 donne alors $f(ua) = f(u)$, et donc $f(uv) = f(uav)$. □

Définition M. Pour $n \geq 1$, une *n-chaîne alternante* pour L est une suite $w_1, \dots, w_{2n} \in \Sigma^*$ telle que pour tout $1 \leq i < 2n$, $w_i \sqsubseteq w_{i+1}$, et pour tout j impair $1 \leq j < 2n$, $w_j \in L$, tandis que pour tout j pair $1 < j \leq 2n$, $w_j \notin L$.

On admet le fait suivant.

Lemme N. *Si un langage L n'est pas testable par morceaux, alors pour tout $n \geq 1$, il existe une n-chaîne alternante pour L.*

Question 18. Montrer que pour tout $m \in M$, l'ensemble $f^{-1}(m)$ est testable par morceaux.

Solution. Par l'absurde, supposons que $L := f^{-1}(m)$ ne soit pas testable par morceaux. Prenons $n > \frac{|F|+1}{2}$. Par le Lemme N, il existe une n-chaîne alternante $w_1, \dots, w_{2n}$ pour L. Pour tout $1 \leq i < 2n$, exactement l'un de $f(w_i)$ et $f(w_{i+1})$ est égal à m, donc $f(w_i) \neq f(w_{i+1})$. Comme w_{i+1} est obtenu à partir de w_i par insertions successives de lettres, il existe une insertion qui modifie la valeur de f. Par la Question 17, cette insertion augmente strictement la valeur de F. Ainsi $F(w_i) \subsetneq F(w_{i+1})$, donc $|F(w_{i+1})| > |F(w_i)|$. On obtient alors $|F(w_{2n})| \geq 2n - 1 > |\mathcal{F}|$, ce qui est impossible puisque $F(w_{2n}) \subseteq \mathcal{F}$. □

Question 19. En utilisant les résultats précédents, démontrer le théorème de Simon :

Théorème (Simon)**.** *Soit L un langage régulier. Alors L est testable par morceaux si, et seulement si, M_L est $\mathcal{J}$-trivial.*

Solution. $\Rightarrow$. Cela découle de la Question 13, de son dual, et du Lemme J.

$\Leftarrow$. Soit $\phi : \Sigma^* \to M_L$ le morphisme syntaxique. On a $L = \bigcup_{m \in \phi[L]} \phi^{-1}(m)$ par un résultat vu en cours. Ainsi L est testable par morceaux par la Question 18. □

Note. Le théorème de Simon est dû à [1]. La preuve ci-dessus est pour l'essentiel identique à celle qui figure dans [2, ch. 8] et a été initialement élaborée en collaboration avec J. Marquès.

[1] I. Simon, "Piecewise testable events", in : Proc. 2nd GI Conf., H. Brackage (ed.), pp. 214–222, Lecture Notes in Comput. Sci. vol. 33, Springer Verlag, Berlin, Heidelberg, New York, 1975

[2] M. Gehrke et S. van Gool, *Topological duality for distributive lattices : Theory and applications*, Cambridge University Press, 2024