CORRIGÉ

I Langage de Dyck

► **Question 1** On remarque pour commencer que $T^k \Rightarrow^* u$ si et seulement si $u \in \mathcal{L}_{G_n}(T)^k$.

- Une récurrence immédiate montre que $S \Rightarrow^* ST^k \Rightarrow T^k$ pour tout $k \in \mathbb{N}$. Ensuite, si $u \in \mathcal{L}_{G_n}(T)^*$, alors $u \in \mathcal{L}_{G_n}(T)^k$ pour un certain k, et donc $T^k \Rightarrow^* u$, ce qui montre $\mathcal{L}_{G_n}(T)^* \subseteq \mathcal{L}_{G_n}(S)$.
- Inversement, si $u \in \mathcal{L}_{G_n}(S)$, alors une dérivation gauche de u est nécessairement de la forme $S \Rightarrow ST \Rightarrow \dots \Rightarrow ST^k \Rightarrow T^k \Rightarrow^* u$. On a donc $u \in \mathcal{L}_{G_n}(T)^*$, ce qui donne l'autre inclusion.

► **Question 2** Laissé au lecteur. . .

► **Question 3** La factorisation d'un mot w de D_n^* en mots de D_n est unique car elle est entièrement déterminée par l'appariement de chaque parenthèse ouvrante a_i avec la parenthèse fermante $\overline{a_i}$ qui lui correspond. À l'intérieur de chaque bloc, la règle à la racine est déterminée par le premier caractère du mot. La grammaire n'est donc pas ambiguë.

► **Question 4** On procède par descente récursive : l'appel `s mot` consomme le préfixe de `mot` généré par S et renvoie le reste, et de même pour `t mot`. Ce préfixe est unique s'il existe (même principe qu'à la question précédente) ; s'il n'existe pas, on lève une exception. On vérifie à la fin qu'on a bien consommé l'intégralité du mot.

```
type parenthese = O of int | F of int

exception SyntaxError

let est_parenthese u =
  let rec s mot =
    match mot with
    | O i :: xs -> s (t mot)
    | _ -> mot
  and t mot =
    match mot with
    | O i :: xs ->
      begin match s xs with
        | F j :: ys when i = j -> ys
        | _ -> raise SyntaxError
      end
    | _ -> raise SyntaxError in
  try
    s u = []
  with
  | SyntaxError -> false
```

II Morphismes

► **Question 5** Un morphisme est, de manière évidente, entièrement déterminé par son action sur Σ_1. Pour chaque lettre $a \in \Sigma_1$, on choisit librement son image dans $\Sigma_2 \cup \{\varepsilon\}$: il y a donc $(1 + n_2)^{n_1}$ morphismes alphabétiques de Σ_1^* vers Σ_2^*.

▶ **Question 6** Soit $G = (\Sigma_1, V, P, S)$ une grammaire sans contexte engendrant L. On étend φ à $(V \cup \Sigma_1)^*$ en posant $\varphi(X) = X$ pour $X \in V$, et l'on définit $G' = (\Sigma_2, V, P', S)$ en prenant pour P' l'ensemble des $X \to \varphi(\alpha)$ avec $X \to \alpha \in P$.
Une récurrence immédiate sur $n \geqslant 0$ montre que si $\alpha \Rightarrow_G^n \beta$ (avec $\alpha, \beta \in (\Sigma_1 \cup V)^*$), alors $\varphi(\alpha) \Rightarrow_{G'}^n \varphi(\beta)$, d'où l'on déduit que si $S \Rightarrow_G^* u \in \Sigma_1^*$, alors $\varphi(S) = S \Rightarrow_{G'}^* \varphi(u)$. Donc $\varphi(\mathcal{L}(G)) \subseteq \mathcal{L}(G')$
Inversement, montrons par récurrence sur n que si $\varphi(\alpha) \Rightarrow_{G'}^n \beta$, alors il existe α' tel que $\beta = \varphi(\alpha')$ et $\alpha \Rightarrow_G^n \alpha'$.

- $n = 0$: la dérivation est vide et il n'y a rien à prouver.
- $n = 1$: si $\varphi(\alpha) \Rightarrow_{G'} \beta$, on décompose α en $u_1X_1u_2 \dots u_nX_nu_{n+1}$ avec les u_i composés uniquement de terminaux. On a alors $\varphi(\alpha) = \varphi(u_1)X_1 \dots \varphi(u_n)X_n\varphi(u_{n+1})$. D'après les règles de G', l'un des X_i a été remplacé par $\varphi(\eta)$, avec $X_i \to \eta \in P$. On a donc $\beta = \varphi(u_1)X_1 \dots \varphi(u_{i-1})\varphi(\eta)\varphi(u_i) \dots X_n\varphi(u_n)$, et donc $\beta = \varphi(\underbrace{u_1X_1 \dots u_{i-1}\eta u_i \dots X_nu_n}_{\alpha'})$. On a bien $\alpha \Rightarrow_G \alpha'$ en utilisant la règle $X_i \to \eta$.
- Si $\varphi(\alpha) \Rightarrow_{G'}^{n+1} \beta$, alors $\varphi(\alpha) \Rightarrow_{G'} \gamma \Rightarrow_{G'}^n \beta$. On utilise H_1 pour obtenir $\gamma = \varphi(\alpha')$ avec $\alpha \Rightarrow_G \alpha'$, puis H_n pour $\beta = \varphi(\alpha'')$ avec $\alpha' \Rightarrow_G^n \alpha''$, et donc $\alpha \Rightarrow_G^{n+1} \alpha''$.

Si $u \in \mathcal{L}(G')$, alors $\varphi(S) = S \Rightarrow^* u$, et u est donc de la forme $\varphi(\alpha)$ avec $\alpha \in (\Sigma_1 \cup V)^*$. Comme les variables sont envoyées sur elles-mêmes et que $u \in \Sigma_2^*$, on a en fait $\alpha \in \Sigma_1^*$, donc $\alpha \in \mathcal{L}(G)$ et $u \in \varphi(\mathcal{L}(G))$, ce qui termine la démonstration.

▶ **Question 7** On obtient une démonstration plus simple en supposant G en forme normale de Chomsky. On pose $V' = V \sqcup \{T\}$, et l'on crée pour T les règles suivantes :

$$T \to \varepsilon \mid aT \qquad\qquad a \in A_0$$

On garde les règles binaires $X \to YZ$ inchangées, et l'on remplace chaque règle unaire $X \to a$ par l'ensemble des règles

$$X \to TbT \qquad\qquad b \in \Sigma_1 \text{ et } \varphi(b) = a$$

Si jamais la règle $S \to \varepsilon$ existe, on ajoute une règle $S \to T$. On considère alors l'application ψ qui à un arbre de dérivation pour G' associe l'arbre obtenu en :

- remplaçant toutes les feuilles b par des feuilles $\varphi(b)$;
- supprimant tous les nœuds étiquetés T et les sous-arbres correspondants (sauf si la racine correspond à la règle $S \to T$, auquel cas on associe l'arbre de dérivation de $S \to \varepsilon$).

On voit immédiatement que si A est un arbre de dérivation pour u dans G', alors $\psi(A)$ est un arbre de dérivation pour $\varphi(u)$: les parties effacées ont pour image ε par φ, et les autres feuilles ont la bonne image. On a donc $u \in \mathcal{L}(G') \implies \varphi(u) \in \mathcal{L}(G)$, c'est-à-dire $\mathcal{L}(G') \subseteq \varphi^{-1}(L)$.

Inversement, soit $u \in \varphi^{-1}(L)$, il existe donc $v \in L$ tel que $\varphi(u) = v$. En notant $v = v_1 \dots v_n$ (avec les v_i des lettres), on a alors $u = u_1a_1u_2a_2 \dots u_na_nv_n$, avec les u_i dans A_0^* et les a_i dans A_1 telles que $\varphi(a_i) = v_i$. À partir d'un arbre de dérivation de v pour G, on peut alors obtenir un arbre de dérivation de u pour G' en remplaçant chaque feuille a_i par une feuille v_i et en lui ajoutant deux frères : des sous-arbres de racine T que l'on utilise pour générer les u_i. On a donc $u \in \mathcal{L}(G')$, ce qui conclut la démonstration.

III Théorème de représentation de Chomsky-Schützenberger

▶ **Question 8** On suit l'indication de l'énoncé, et l'on prend comme règles :

$$\begin{array}{ll}
S' \to S_{q_0q_f} & q_f \in F \\
S' \to \varepsilon & \text{si } \varepsilon \in K \cap L \\
X_{pq} \to Y_{pr}Z_{rq} & p, q, r \in Q \text{ et } X \to YZ \in P \\
X_{pq} \to a & \delta(p, a) = q \text{ et } X \to a \in P
\end{array}$$

Le mot vide est bien dans $\mathcal{L}(G')$ si et seulement si il est dans $K \cap L$. Par construction, on a $u \in \mathcal{L}_{G'}(X_{pq})$ si et seulement si $p.u = q$ et $u \in \mathcal{L}_G(X)$ pour tout $u \neq \varepsilon$ (mériterait d'être détaillé). Or $\mathcal{L}(G') \setminus \{\varepsilon\} = \cup_{q_f \in F} \mathcal{L}_{G'}(S_{q_0q_f})$, donc $u \in \mathcal{L}(G')$ si et seulement si $u \in \mathcal{L}_G(S) = L$ et il existe $q_f \in F$ tel que $q_0.u = q_f$, c'est-à-dire $u \in K$.

▶ **Question 9** D_n est algébrique par définition, donc $D_n \cap K$ aussi d'après la question précédente (la condition même alphabet ne pose pas de problème, il suffit de prendre l'union des deux alphabets). D'après la question 6, $\varphi(D_n \cap K)$ est donc algébrique, ce qui fournit un sens du théorème.

▶ **Question 10** En notant n_1 le nombre de règles unaires $X \to a$ et n_2 le nombre de règles binaires $X \to YZ$, on a $n = n_1 + 3n_2$.

▶ **Question 11** Laissé au lecteur.

▶ **Question 12**

▶ **Question 13**

▶ **Question 14**

▶ **Question 15**

▶ **Question 16**

▶ **Question 17**