Opció 2:

A partir d'una CFG $G = \langle V, \Sigma, P, S\rangle$ podem construir de la manera següent un NPDA $M = \langle Q, \Sigma, \Gamma, \delta, q_0, Z_0, F\rangle$ tal que $L(M) = L(G)$:

- **el conjunt d'estats $Q = \{q_0, p, f\}$ (estat inicial, estat de procés i estat acceptador);**
- **l'alfabet de pila $\Gamma = \Sigma \cup V \cup \{Z_0\}$ on $Z_0 \notin \Sigma \cup V$ és el símbol de fons de pila;**
- **el conjunt d'estats acceptadors $F = \{f\}$;**
- **la funció de transició δ està format per:**

  $Z_0 q_0 \vdash Z_0 S p$ **(una λ-transició que posa la variable inicial al cim de la pila en l'estat de procés);**

  $\{Zp \vdash \alpha^R p \mid Z \to \alpha \in P\}$ **(una λ-transició per a cada producció de la variable Z al cim de la pila);**

  $\{apa \vdash p \mid a \in \Sigma\}$ **(si el cim de la pila és un terminal, l'autòmat ha de verificar que sigui igual al símbol en curs del mot d'entrada i avançar el capçal);**

  $Z_0 p \vdash f$ **(una λ-transició que permet passar a l'estat acceptador quan tot el mot ha estat processat).**

(Examen de Gener-2008)

Tingueu present en tots els exercicis que les respostes s'han de justificar. La mera presentació d'una proposta de resultat no és suficient.

1. (2.5 punts en total) Atenció, el segon punt d'aquest problema s'ha de contestar i entregar en el propi enunciat.

   - (1) Considerem el llenguatge sobre $\{a, b\}^*$ dels mots que acaben en *abaab*. Definiu formalment aquest conjunt amb la notació clàssica de conjunts, i trobeu l'autòmat determinista mínim que el genera, directament, o per transformació de la seva formalització, aplicant després operacions de tancament i minimitzant.
   - (1.5) Donat un mot w sobre $\{a, b\}^*$, el llenguatge dels mots que acaben en w, és a dir $\{a, b\}^* w$, és regular. De fet, es pot demostrar que existeix un autòmat determinista amb $|w| + 1$ estats que el reconeix. Però ara ens ocuparem de veure que no se'n pot trobar cap de més petit. A continuació hi ha una demostració (amb algunes parts esborrades) que per a qualsevol $w \in \{a, b\}^*$, no hi ha cap autòmat determinista amb menys que $|w| + 1$ estats que reconeix $\{a, b\}^* w$. Afegiu el que hi falta.

     Sigui $n = |w|$. Suposem que hi ha un DFA A amb menys de $n + 1$ estats que reconeix el llenguatge. Sigui q_0 l'estat inicial de A. Com que hi ha $n + 1$ mots a `prefixos`(w), dos d'ells ens porten al mateix estat; els anomenem u i v. Per tant, $q_0 u = q_0 v$. Sense pèrdua de generalitat assumim que u és prefix propi de v (en altre cas v seria prefix propi de u i es procediria anàlogament). Així doncs, v és de la forma uu', per $|u'|$..........0, i w és de la forma $vv' = uu'v'$ per $|v'|$..........0. Com que w acaba en ...., tenim que $(q_0 w)$ és acceptador. Però com que $q_0 w = q_0(vv') = (q_0 v)v' = (..........)v'$, a aquest mateix estat acceptador també s'hi arriba amb .... Donat que $|u'|$..........0 tenim que $|uv'| < |u..........v'| = |w|$. Per tant, uv' no pot acabar en .... perquè és més petit que ...., i per contra s'està acceptant: contradicció.

Resposta:

- **(versió simplificada) Definim el conjunt així: $\{w \in \{a,b\}^* \mid \exists x \in \{a,b\}^* : (w = xabaab)\}$. Aquest llenguatge es pot escriure com la següent concatenació: $\{a,b\}^*\{abaab\}$. Després d'escriure els corresponents autòmats, calcular l'autòmat concatenació i minimitzar, obtenim:**

	a	b
q_0	q_1	q_0
q_1	q_1	q_2
q_2	q_3	q_0
q_3	q_4	q_2
q_4	q_1	q_5
$\dagger q_5$	q_3	q_0

Si ja hem obtingut aquest mateix autòmat bans de minimitzar, en realitat no cal fer-ho, doncs té $|w|+1=6$ estats, i en virtud de l'apartat següent, és mínim.

- **Sigui $n=|w|$. Suposem que hi ha un DFA A amb menys de $n+1$ estats que reconeix el llenguatge. Sigui q_0 l'estat inicial de A. Com que hi ha $n+1$ mots a $\texttt{prefixos}(w)$, dos d'ells ens porten al mateix estat, que anomenem u i v. Per tant, $q_0u = q_0v$. Sense pèrdua de generalitat assumim que u és prefix propi de v (en altre cas v seria prefix propi de u i es procediria anàlogament). Per tant, v és de la forma uu', per $|u'|>0$, i w és de la forma $vv' = uu'v'$ per $|v'| \geqslant 0$. Com que w acaba en w, tenim que (q_0w) és acceptador. Però com que $q_0w = q_0(vv') = (q_0v)v' = (q_0u)v'$, a aquest mateix estat acceptador també s'hi arriba amb uv'. Donat que $|u'|>0$ tenim que $|uv'| < |uu'v'| = |w|$. Per tant, uv' no pot acabar en w perquè és més petit que w, i per contra s'està acceptant: contradicció.**

2. (2 punts en total) Definim la mida d'una gramàtica G, que denotem $|G|$, com el nombre total de símbols que apareixen a les parts dretes de les regles de G. Per exemple, la gramàtica $S \to (S)S|\lambda$ té mida 4 perquè $|(S)S| + |\lambda| = 4+0 = 4$.

   - (1 punt) Demostreu que per a qualsevol gramàtica G existeix una gramàtica G' tal que $\mathtt{L}(G) = \mathtt{L}(G')$, totes les parts dretes de G' tenen mida menor o igual a 2 (més formalment: $\forall X \to \alpha \in \texttt{regles}(G') : |\alpha| \leqslant 2$), i $|G'| \leqslant 2|G|$.
   - (1 punt) Demostreu que, per a qualsevol gramàtica G que compleixi $\lambda \notin \mathtt{L}(G)$, existeix una gramàtica G' sense λ-produccions (és a dir, les de la forma $X \to \lambda$) tal que $\mathtt{L}(G') = \mathtt{L}(G)$ i $|G'| \leqslant 4|G|$.

   Resposta:

   - **Construïm G' a partir de G de la següent manera. Per cada regla $X \to \alpha_1 \dots \alpha_n$ de G amb $n \geqslant 3$, on els α_i són els terminals i no terminals concatenats que apareixen a la part dreta de la regla, ens inventem (n-2) variables noves $X_2, \dots, X_{n-1}$, esborrem la regla anterior, i afegim les regles $X \to \alpha_1X_2,\ X_2 \to \alpha_2X_3,\ X_3 \to \alpha_3X_4,\ \dots,\ X_{n-1} \to \alpha_{n-1}\alpha_n$. És obvi que desde $\alpha_1 \dots \alpha_n$ i α_1X_2 es genera el mateix llenguatge. Per tant, la nostra transformació de G en G' preserva el llenguatge. Además, la mida de la regla original era n, i la suma de mides de les corresponents noves regles és $2*(n-1) < 2n$. Així doncs, $|G'| \leqslant 2|G|$.**
   - **Donada G, obtenim primer G'' segons la transformació de l'apartat anterior. Així doncs, $\mathtt{L}(G) = \mathtt{L}(G'')$, $|G''| \leqslant 2|G|$, i totes les parts dretes de G'' tenen com a molt mida 2. Un cop fet això, apliquem la transformació clàssica per a eliminar λ-produccions. En el nostre cas particular, això correspon a fer el següent. Per cada regla de la forma $X \to \alpha_1\alpha_2$, si α_1 és anulable, llavors hem d'afegir la regla $X \to \alpha_2$, i si α_2 és anulable, llavors hem d'afegir la regla $X \to \alpha_1$. Finalment, hem d'esborrar totes les produccions de la forma $Z \to \lambda$. Aquesta transformació de G'' en una nova gramàtica G' preserva el llenguatge, com ja és sabut. Además, com a molt, per cada regla de mida 2 s'afegeixen dues regles de mida 1. Per tant, $|G'| \leqslant 2|G''| \leqslant 4|G|$.**

3. (1.5 punts) Demostreu que el llenguatge L següent no és regular.

$$L = \{a^x b^y c^z d^t \mid x, y, z, t \geqslant 0 \ \wedge \ ((x = t) \Rightarrow (y = z))\} \cup a^*b^*c^* \cup b^*c^*d^*$$

Resposta: Suposem que L és regular. Intersequem L amb ab^*c^*d, que és regular, obtenint així $L' = \{ab^nc^nd\}$. Donat que la intersecció de llenguatges regulars dona lloc a un llenguatge regular, resulta que L' és regular. Apliquem el morfisme definit per $\sigma(a) = \lambda$, $\sigma(b) = a$, $\sigma(c) = b$, $\sigma(d) = \lambda$ sobre L' obtenint així $\{a^nb^n\}$. Donat que l'aplicació d'un morfisme sobre un llenguatge regular dona lloc a un llenguatge regular, resulta que $\{a^nb^n\}$ és regular. Però $\{a^nb^n\}$ es un conegut llenguatge no regular, contradicció.

Resposta alternativa 1: Suposem que L és regular. Llavors, existeix un DFA A que el reconeix. Sigui N el nombre d'estats de A. Donat que la paraula ab^Nc^Nd és de L, llavors és acceptada per A. En la execució de A amb entrada ab^Nc^Nd, passem almenys per aquests estats: $q_0a, q_0ab, q_0abb, \ldots, q_0ab^N$. En total en són $N+1$ estats, i per tant n'hi ha un de repetit. Així doncs, existeixen i, j cumplint $0 \leqslant i < j \leqslant N$ tals que $q_0ab^i = q_0ab^j$. Donat que $q_0ab^Nc^Nd$ és acceptador, també ho és $q_0ab^Nc^Nd = (q_0ab^i)b^{N-i}c^Nd = (q_0ab^j)b^{N-i}c^Nd = q_0(ab^{N-i+j}c^Nd)$. Així doncs, $ab^{N-i+j}c^Nd$ ens porta a estat acceptador, però donat que $j > i$, aquesta paraula no és del llenguatge, contradicció.

Resposta alternativa 2: Apliquem el lema de bombament. Fixat N, agafem la paraula $w = a^Nbccd^{N+N!}$, que és del llenguatge. Considerem una factorització qualsevol $w = xyz$ cumplint $|xy| \leqslant N$ i $|y| \geqslant 1$. Necessàriament, existeixen j, k cumplint $x = a^j$, $y = a^k$, $k > 1$, $j + k \leqslant N$. Busquem un i tal que $(i-1)k = N!$. Aquest és $i = 1 + (N!)/k$, que és un nombre natural perquè $N!$ és un producte que conté k entre els seus factors. La paraula $xy^iz = a^{N+N!}bccd^{N+N!}$ no és del llenguatge, i això conclou la prova.

4. (1.5 punts) Classifiqueu com a decidible, semi-decidible però no decidible, o no semi-decidible, el problema següent.

$$C = \{\langle x, y\rangle \mid \forall z : (M_x(z)\downarrow \Leftrightarrow M_y(z)\downarrow)\}$$

Resposta: Demostrem que no és ni semi-decidible reduïnt desde $\overline{\mathtt{K}} = \{x | M_x(x)\uparrow\}$. La reducció consisteix en generar, per a cada x, la parella $\langle p(x), q\rangle$, on q és un nombre que codifica un programa que no s'atura per a cap entrada, i $p(x)$ és el programa següent:

```
entrada y
   Simular M_x(x)
   sortida y
```

Si un cert x pertany a $\overline{\mathtt{K}}$, llavors $M_{p(x)}$ no s'atura per a cap entrada, i per tant $\langle p(x), q\rangle$ pertany a C. Si x no pertany a $\overline{\mathtt{K}}$, llavors $M_{p(x)}$ és una màquina que s'atura per a totes les entrades, i per tant $\langle p(x), q\rangle$ no pertany a C. Això conclou la prova.

5. (2.5 punts en total) Durant el curs hem vist que és indecidible el problema de, donades dues gramàtiques, saber si tenen intersecció no buida.

$$\{\langle G_1, G_2\rangle \mid \mathtt{L}(G_1) \cap \mathtt{L}(G_2) \neq \emptyset\}$$

- (1.5) Demostreu que també ho és el problema de, donades dues gramàtiques, saber si tenen com a intersecció algun mot de longitud parell. Feu una reducció des del problema d'intersecció no buida. (anomenem *intersecció parella* a aquest nou problema.)

$$\{\langle G_1, G_2\rangle \mid \exists w : (|w| \in \dot{2} \wedge w \in (\mathtt{L}(G_1) \cap \mathtt{L}(G_2)))\}$$

- (1) Demostreu que també és indecidible el problema de, donades dues gramàtiques, saber si tenen com a intersecció dos o més mots de longitud parell. (anomenem *2-intersecció parella* a aquest nou problema.)

$$\{\langle G_1, G_2\rangle \mid \exists w_1, w_2 : (w_1 \neq w_2 \wedge |w_1|, |w_2| \in \dot{2} \wedge w_1, w_2 \in (\mathtt{L}(G_1) \cap \mathtt{L}(G_2)))\}$$

Resposta:

- **Donada la entrada $\langle G_1, G_2\rangle$ d'intersecció no buida, construïm una nova entrada $\langle G_1', G_2'\rangle$ per a intersecció parella com segueix. Ens inventem un símbol nou $\#$. Si S_1 és el símbol inicial de G_1, llavors G_1' té una nova variable S_1' com a símbol inicial, i les seves regles són les de G_1 més $S_1' \to S_1\#\#S_1$. Si S_2 és el símbol inicial de G_2, llavors G_2' té una nova variable S_2' com a símbol inicial, i les seves regles són les de G_2 més $S_2' \to S_2\#\#S_2$.**

  Si G_1 i G_2 generen una paraula comuna w, llavors G_1' i G_2' generen la paraula comuna $w\#\#w$ de longitud parella. Per a la direcció contrària, si G_1' i G_2' generen una paraula comuna u de longitud parella, llavors, per la forma d'aquestes gramàtiques, u és de la forma $u_1\#\#u_2$, on tant u_1 com u_2 són generables tant amb G_1 com amb G_2. Per tant, G_1 i G_2 generen alguna paraula comuna, i això conclou la prova.

- **Reduïm desde el llenguatge de l'apartat anterior. Donada la entrada $\langle G_1', G_2'\rangle$ d'intersecció parella, construïm una nova entrada $\langle G_1'', G_2''\rangle$ per a 2-intersecció parella com segueix. Ens inventem un símbol nou \$. Si S_1' és el símbol inicial de G_1', llavors G_1'' té una nova variable S_1'' com a símbol inicial, i les seves regles són les de G_1' més $S_1'' \to S_1' | \$\$$. Si S_2' és el símbol inicial de G_2', llavors G_2'' té una nova variable S_2'' com a símbol inicial, i les seves regles són les de G_2' més $S_2'' \to S_2' | \$\$$.**

  Si G_1' i G_2' generen una paraula comuna de longitud parella, llavors G_1'' i G_2'' també la generen, i ademés totes dues generen també \$\$, que és de longitud parella. Per a la direcció contrària, si G_1'' i G_2'' generen dues paraules comunes de longitud parella, llavors almenys una no és \$\$, i ademés és generable desde G_1' i G_2'. Això conclou la prova.