Exam

Duration: 3 hours. All paper documents permitted. The numbers $[n]$ in the margin next to questions are indications of duration and difficulty, not necessarily of the number of points you might earn from them. You must **fully justify** all your answers.

Exercise 1. Let us consider the ranked alphabet $\mathcal{F} \stackrel{\text{def}}{=} \{f^{(2)}, a^{(0)}\}$, and the language L of trees where every branch is of even length; for instance $f(a,a) \in L$ but $f(a, f(a,a)) \notin L$.

[1] 1. Give a PDL($\downarrow_0, \downarrow_1$) node formula φ over the set of atomic predicates $A \stackrel{\text{def}}{=} \{f, a\}$ s.t. for all $t \in T(\mathcal{F})$, $t, \varepsilon \models \varphi$ if and only if $t \in L$.

We simply check that there are no leaves at odd depth.

$$\varphi \stackrel{\text{def}}{=} [(\downarrow;\downarrow)^*]\neg\mathsf{leaf}\ ,$$

where

$$\mathsf{leaf} \stackrel{\text{def}}{=} [\downarrow]\bot\ , \qquad\qquad \downarrow \stackrel{\text{def}}{=} (\downarrow_0 + \downarrow_1)\ .$$

[2] 2. Give a wMSO($\downarrow_0, \downarrow_1, P_f, P_a$) sentence ψ over $T(\mathcal{F})$ s.t. for all $t \in T(\mathcal{F})$, $t \models \psi$ if and only if $t \in L$.

One way is to select the even-depth positions using a second-order variable, and to ensure that it contains all the leaves.

$$\begin{aligned}\psi \stackrel{\text{def}}{=} \exists X\ .\ \forall x.\mathsf{root}(x) &\Longrightarrow x \notin X\\ &\wedge x \in X \Longrightarrow (\forall y\ .\ x \downarrow y \Longrightarrow y \notin X)\\ &\wedge x \notin X \Longrightarrow (\forall y\ .\ x \downarrow y \Longrightarrow y \in X)\\ &\wedge \mathsf{leaf}(x) \Longrightarrow x \in X\ ,\end{aligned}$$

where

$$x \downarrow y \stackrel{\text{def}}{=} x \downarrow_0 y \vee x \downarrow_1 y\ , \qquad \mathsf{root}(x) \stackrel{\text{def}}{=} \neg\exists z\ .\ z \downarrow x\ , \qquad \mathsf{leaf}(x) \stackrel{\text{def}}{=} \neg\exists z\ .\ x \downarrow z\ .$$

Alternatively, one could translate the PDL solution:

$$\psi \stackrel{\text{def}}{=} \exists x\ .\ \mathsf{root}(x) \wedge \forall y\ .\ (x\ [\mathrm{TC}_{u,v}\mathsf{two}(u,v)]\ y) \Longrightarrow \neg\mathsf{leaf}(y)\ ,$$

where

$$\begin{aligned}[\mathrm{TC}_{u,v}\varphi(u,v)]\ y &\stackrel{\text{def}}{=} \forall X.(x \in X \wedge \forall u \forall v\ .\ (u \in X \wedge \varphi(u,v)) \Longrightarrow v \in X) \Longrightarrow y \in X\ ,\\ \mathsf{two}(x,y) &\stackrel{\text{def}}{=} \exists z\ .\ x \downarrow z \wedge z \downarrow y\ .\end{aligned}$$

[2] 3. Give a minimal DFTA $\mathcal{A}$ that recognises L.

$\mathcal{A} \stackrel{\text{def}}{=} (Q, \mathcal{F}, \Delta, \{q_1\})$ where $Q \stackrel{\text{def}}{=} \{q_0, q_1, q_\perp\}$ and Δ is defined by

$$\Delta \stackrel{\text{def}}{=} \{(q_\perp, f, q_0, q_1), (q_\perp, f, q_1, q_0), (q_0, f, q_1, q_1), (q_1, f, q_0, q_0), (q_0, a)\} \\ \cup \{(q_\perp, f, q, q') \mid q = q_\perp \text{ or } q' = q_\perp\}$$

The automaton $\mathcal{A}$ is deterministic and complete by definition, hence the languages of the states form a partition of $T(\mathcal{F})$: $L(q_0)$ is the set of trees with all branches of odd length, $L(q_1)$ the set of trees with all branches of even length, and $L(q_\perp)$ the set of tree with a branch of even length and one of odd length. We show this by induction over the trees in $T(\mathcal{F})$.

Base case a: a leaf a is a tree with all its branches of odd length and is indeed recognised by q_0.

Inductive step $f(t_1, t_2)$: If both t_1 and t_2 have all their branches of even length, then by induction hypothesis, $t_i \to^*_{\mathcal{A}} q_1$ for $i \in \{1, 2\}$ and then $f(t_1, t_2) \to^*_{\mathcal{A}} f(q_1, q_1) \to q_0$ as desired. If t_1 and t_2 have all their branches of odd length, then by induction hypothesis $t_i \to^*_{\mathcal{A}} q_0$ for $i \in \{1, 2\}$ and then $f(t_1, t_2) \to^*_{\mathcal{A}} f(q_0, q_0) \to q_1$ as desired. If t_1 has a branch of even length and one of odd length, then by induction hypothesis $t_1 \to^*_{\mathcal{A}} q_\perp$ and then $f(t_1, t_2) \to^*_{\mathcal{A}} q_\perp$ and the same if $t_2 \in L(q_\perp)$. Finally, if $t_1 \to^*_{\mathcal{A}} q_0$ and $t_2 \to^*_{\mathcal{A}} q_1$ or vice-versa, then $f(t_1, t_2) \to^*_{\mathcal{A}} q_\perp$.

As $L(q_0)$, $L(q_1)$, and $L(q_\perp)$ form a partition of $T(\mathcal{F})$, the automaton is minimal.

One could also show that for all $q \neq q'$ in Q, there are contexts $C_{q,q'}$ distinguishing some trees $t \in L(q)$ from $t' \in L(q')$, i.e. with $C_{q,q'}[t] \in L$ and $C_{q,q'}[t'] \in T(\mathcal{F}) \setminus L$. The empty context $\square$ can be used as both C_{q_1,q_0} and $C_{q_1,q_\perp}$: it distinguishes any tree $t \in L(q_1) = L$ from any tree $t' \notin L$. The context $C_{q_0,q_\perp} \stackrel{\text{def}}{=} f(\square, a)$ distinguishes $a \in L(q_0)$ from $f(f(a, a), a) \in L(q_\perp)$.

Inspired by TATA Ex. 2.5

Exercise 2 (Local Unranked Tree Languages)**.** We work throughout this exercise with unranked trees over some finite alphabet Σ. For a tree $a(t_1 \cdots t_n) \in T(\Sigma)$, its *root label* is $\mathrm{root}(a(t_1 \cdots t_n)) \stackrel{\text{def}}{=} a$; we lift this to tree languages by $\mathrm{root}(L) \stackrel{\text{def}}{=} \bigcup_{t \in L} \mathrm{root}(t)$. For a letter $a \in \Sigma$ and a tree $t \in T(\Sigma)$, the *horizontal set* for a in t is defined by induction over t by

$$\mathrm{hs}_a(b(t_0 \cdots t_{n-1})) \stackrel{\text{def}}{=} \begin{cases} \{\mathrm{root}(t_0) \cdots \mathrm{root}(t_{n-1})\} \cup \bigcup_{0 \leq i < n} \mathrm{hs}_a(t_i) & \text{if } b = a \\ \bigcup_{0 \leq i < n} \mathrm{hs}_a(t_i) & \text{if } b \neq a\ . \end{cases}$$

We can also define this set by

$$\mathrm{hs}_a(t) \stackrel{\text{def}}{=} \{\mathrm{root}(t_0) \cdots \mathrm{root}(t_{n-1}) \mid \exists C \in \mathcal{C}(\Sigma)\ .\ t = C[a(t_0 \cdots t_{n-1})]\}$$

where $\mathcal{C}(\Sigma)$ denotes the set of contexts over Σ. For instance, $\mathrm{hs}_a(b(a(ab)a(b))) = \{\varepsilon, b, ab\}$ and $\mathrm{hs}_b(b(a(ab)a(b))) = \{\varepsilon, aa\}$. For a tree language $L \subseteq T(\Sigma)$, $\mathrm{hs}_a(L) \stackrel{\text{def}}{=} \bigcup_{t \in L} \mathrm{hs}_a(t)$.

[1] 1. Consider the tree language $L_1 \stackrel{\text{def}}{=} L(\mathcal{A}_1)$ for the NFHA $\mathcal{A}_1 \stackrel{\text{def}}{=} (Q_1, \{a, b\}, \Delta_1, I_1)$ where $Q_1 \stackrel{\text{def}}{=} \{q, q'\}$, $\Delta_1 \stackrel{\text{def}}{=} \{a((qq')^*) \to q, b(Q_1^* q' Q_1^*) \to q'\}$, and $I_1 \stackrel{\text{def}}{=} \{q\}$. Give $\text{hs}_a(L_1)$ and $\text{hs}_b(L_1)$. *Hint: Don't be surprised by the answer.*

We have $\text{hs}_a(L_1) = \{\varepsilon\}$ and $\text{hs}_b(L_1) = \emptyset$. Indeed, $L_1 = \{a()\}$.

[1] 2. Consider the tree language $L_2 \stackrel{\text{def}}{=} L(\mathcal{A}_2)$ for the NFHA $\mathcal{A}_2 \stackrel{\text{def}}{=} (Q_2, \{a\}, \Delta_2, I_2)$ where $Q_2 \stackrel{\text{def}}{=} \{q_0, q_1\}$, $\Delta_2 \stackrel{\text{def}}{=} \{a(q_0^*) \to q_1, a(q_1^+) \to q_0\}$, and $I_2 \stackrel{\text{def}}{=} \{q_0\}$. Give $\text{hs}_a(L_2)$.

We have $\text{hs}_a(L_2) = \{a\}^*$. Indeed, L_2 is the set of trees in $T(\{a\})$ where every branch is of even length. Thus for all $n \in \mathbb{N}$, $a^n \in \text{hs}_a(L_2)$: for $n = 0$ we can consider the tree $a(a) \in L_2$ seen as $a(\square)[a]$, and for $n > 0$ we can consider the tree $a(a^n) \in L_2$ seen as $\square[a(a^n)]$.

[2] 3. Show that, if L is a recognisable tree language, then for all $a \in \Sigma$, $\text{hs}_a(L)$ is a recognisable word language. *Hint: You may assume without loss of generality that $L = L(\mathcal{A})$ for a trim NFHA $\mathcal{A} = (Q, \Sigma, \Delta, I)$: for all states $q \in Q$, there exist a tree t and a context C such that $t \to_{\mathcal{A}}^* q$ and $C[q] \to_{\mathcal{A}}^* q_r$ for some $q_r \in I$.*

Recall that if $\sigma\colon \Sigma \to \text{Rec}(\Gamma^*)$ maps letters from Σ to recognisable word languages over Γ, then it defines a rational substitution*, and then $\sigma(R)$ for a recognisable language $R \subseteq \Sigma^*$ is also a recognisable language over Γ.*

Let $L = L(\mathcal{A})$ for a trim NFHA $\mathcal{A} = (Q, \Sigma, \Delta, I)$. Let us consider the map $\sigma\colon Q \to 2^{\Sigma^*}$ where $\sigma(q) \stackrel{\text{def}}{=} \{a \in \Sigma \mid \exists R \neq \emptyset \,.\, (q, a, R) \in \Delta\}$; since $\sigma(q)$ is a finite set for all q, it is in particular a recognisable set in $\text{Rec}(\Sigma^*)$. Therefore σ defines a rational substitution.

Let us show that, for all $a \in \Sigma$, $\boldsymbol{\text{hs}_a(L) = \bigcup_{(q,a,R)\in\Delta} \sigma(R)}$. This will imply the result, since the right-hand expression is a finite union of rational substitutions applied to recognisable word languages, and is therefore a recognisable word language.

⊆: Let $a_0 \cdots a_{n-1} \in \text{hs}_a(L)$. Then there exists a tree $C[a(t_0 \cdots t_{n-1})] \in L$ for some context C and trees $t_0, \cdots, t_{n-1}$ such that $\text{root}(t_i) = a_i$ for all $0 \le i < n$. Since $C[a(t_0 \cdots t_{n-1})] \in L$ there exist states $q, q_0, \ldots, q_{n-1}$ and a transition $(q, a, R) \in \Delta$ such that $t_i \to_{\mathcal{A}}^* q_i$ for all $0 \le i < n$ and $q_0 \cdots q_{n-1} \in R$. Furthermore, for all $0 \le i < n$, since $\text{root}(t_i) = a_i$ and $t_i \to_{\mathcal{A}}^* q_i$, there exists $(q_i, a_i, R_i) \in \Delta$ with $R_i \neq \emptyset$ that allows to recognise t_i. Thus for all $0 \le i < n$, $a_i \in \sigma(q_i)$, hence $a_0 \cdots a_{n-1} \in \sigma(R)$.

⊇: Let $a_0 \cdots a_{n-1} \in \sigma(R)$ for some $(q, a, R) \in \Delta$. By definition of σ, for all $0 \le i < n$, there exists $q_i \in Q$ such that $(q_i, a_i, R_i) \in \Delta$ with $R_i \neq \emptyset$, and such that overall $q_0 \cdots q_{n-1} \in R$. For all $0 \le i < n$, since $R_i \neq \emptyset$ it contains some sequence $q_{i,0} \cdots q_{i,n_i-1}$ for some n_i. Since $\mathcal{A}$ was assumed to be trim, there exists a context C such that $C[q] \to_{\mathcal{A}}^* q_r$ for some $q_r \in I$ and there exist trees $t_{i,j}$ such that $t_{i,j} \to_{\mathcal{A}}^* q_{i,j}$ for all $0 \le i < n$ and $0 \le j < n_i$. Let $t_i \stackrel{\text{def}}{=}$

$a_i(t_{i,0} \cdots t_{i,n_i-1})$ for all $0 \leq i < n$: then $t_i \to^*_{\mathcal{A}} q_i$, thus $a(t_0 \cdots t_{n-1}) \to^*_{\mathcal{A}} q$, and $C[q] \to^*_{\mathcal{A}} q_r$ for some $q_r \in I$. Thus $\mathrm{root}(t_i) = a_i$ and $C[a(t_0 \cdots t_{n-1})] \in L$: this shows $a_0 \cdots a_{n-1} \in \mathrm{hs}_a(L)$.

⚠ The condition $R \neq \emptyset$ in the definition of σ is really needed. Consider for instance the trim NFHA $\mathcal{A} \stackrel{\text{def}}{=} (\{q\}, \{a, b\}, \{a(q^*) \to q, b(\emptyset) \to q\}, \{q\})$. Then $L(\mathcal{A}) = T(\{a\})$ the set of unranked trees with only a labels, and $\mathrm{hs}_a(L(\mathcal{A})) = \{a\}^*$ and $\mathrm{hs}_b(L(\mathcal{A})) = \emptyset$. However, if one defines $\sigma'(q) = \{c \in \Sigma \mid (q, c, R) \in \Delta\}$, then $\sigma'(q) = \{a, b\}$, and then $\bigcup_{(q,a,R)\in\Delta} \sigma'(R) = \sigma'(\{q\}^*) = \{a, b\}^* \neq \mathrm{hs}_a(L(\mathcal{A}))$.

4. Given a set $S \subseteq \Sigma$ and a collection $(R_a)_{a\in\Sigma}$ of word languages $R_a \subseteq \Sigma^*$, we define the tree language

$$\mathscr{L}(S, (R_a)_{a\in\Sigma}) \stackrel{\text{def}}{=} \{t \in T(\Sigma) \mid \mathrm{root}(t) \in S \text{ and } \forall a \in \Sigma \,.\, \mathrm{hs}_a(t) \subseteq R_a\} \;.$$

[0] Show that, for all $L \subseteq T(\Sigma)$, $L \subseteq \mathscr{L}(\mathrm{root}(L), (\mathrm{hs}_a(L))_a)$.

Let $t \in L$. Then $\mathrm{root}(t) \in \mathrm{root}(L)$ and for all $a \in \Sigma$, $\mathrm{hs}_a(t) \subseteq \mathrm{hs}_a(L)$, thus $t \in \mathscr{L}(\mathrm{root}(L), (\mathrm{hs}_a(L))_a)$.

5. A language L is called *local* if there exist $S \subseteq \Sigma$ and $(R_a)_{a\in\Sigma}$ where each $R_a \in \mathrm{Rec}(\Sigma^*)$ is a recognisable word language such that $L = \mathscr{L}(S, (R_a)_{a\in\Sigma})$.

[1] Show that a recognisable tree language $L \subseteq T(\Sigma)$ is local if and only if $L = \mathscr{L}(\mathrm{root}(L), (\mathrm{hs}_a(L))_a)$.

If L is local then there exist S and $(R_a)_{a\in\Sigma}$ such that $L = \mathscr{L}(S, (R_a)_{a\in\Sigma})$. Then $\mathrm{root}(L) \subseteq S$ and for all $a \in \Sigma$, $\mathrm{hs}_a(L) \subseteq R_a$. Since the $\mathscr{L}$ operator is monotone, this entails $\mathscr{L}(S, (\mathrm{hs}_a(L))_{a\in\Sigma}) \subseteq L$. The converse inclusion holds by Question 4, thus $L = \mathscr{L}(s, (\mathrm{hs}_a(L))_a)$.

Conversely, if $L = \mathscr{L}(\mathrm{root}(S), (\mathrm{hs}_a(L))_a)$, then since L is recognisable, by Question 3, $\mathrm{hs}_a(L)$ is a recognisable word language for every $a \in \Sigma$, thus L is local.

[1] 6. Show that **not** all recognisable tree languages are local.

L_2 is recognisable but not local: $\mathrm{root}(L_2) = \{a\}$ and as seen earlier $\mathrm{hs}_a(L_2) = \{a\}^*$, thus $\mathscr{L}(\mathrm{root}(L_2), (\mathrm{hs}_a(L_2))_{a\in\Sigma}) = T(\{a\}) \supsetneq L_2$.

Inspired by TATA Sec. 8.7.1; DTDs normally only allow a single start symbol

Exercise 3 (Document Type Definitions)**.** A *document type definition* (DTD) is a tuple $\mathcal{D} = (\Sigma, (R_a)_{a\in\Sigma}, I)$ with Σ a finite alphabet, $R_a \in \mathrm{Rec}(\Sigma^*)$ for all $a \in \Sigma$, and $I \subseteq \Sigma$ a set of start symbols.

Its *language* $L(\mathcal{D}) \subseteq T(\Sigma)$ is defined as the language of the NFHA $\mathcal{A}_{\mathcal{D}} \stackrel{\text{def}}{=} (Q, \Sigma, \Delta, I)$ with states $Q \stackrel{\text{def}}{=} \Sigma$ (thus the symbols from Σ serve both as states and as tree labels) and transition rules $\Delta \stackrel{\text{def}}{=} \{a(R_a) \to a \mid a \in \Sigma\}$ (the left 'a' is seen as a letter from Σ, the right one is seen as a state from Q).

[0] 1. Give a DTD for the set of unranked trees over the alphabet $\{a, b\}$ where every 'b'-labelled node is the child of an 'a'-labelled node.

$\mathcal{D} = (\{a, b\}, (R_a = \{a, b\}^*, R_b = \{a\}^*), \{a\}$.

[3] 2. Show that a recognisable tree language is local if and only if it is the language of a DTD.

(a) First show that, for any DTD $\mathcal{D} = (\Sigma, (R_a)_{a\in\Sigma}, I)$, for all trees $t \in T(\Sigma)$ and $a \in \Sigma$, $t \to^*_{\mathcal{A}_\mathcal{D}} a$ implies $\mathrm{root}(t) = a$.

The tree t can be written as $t = b(t_0 \cdots t_{n-1})$ where $b = \mathrm{root}(t)$ for some subtrees $t_0, \ldots, t_{n-1}$. Since $t \to^*_{\mathcal{A}_\mathcal{D}} a$, there exist $(b(R_b) \to a) \in \Delta$ and a sequence of states $a_0 \cdots a_{n-1} \in R_b$ such that $t_i \to^*_{\mathcal{A}_\mathcal{D}} a_i$ for all $0 \leq i < n$. By definition of Δ, $b = \mathrm{root}(t) = a$.

(b) Show that, for any DTD $\mathcal{D} = (\Sigma, (R_a)_{a\in\Sigma}, I)$, $\mathrm{root}(L(\mathcal{D})) \subseteq I$ and for all $a \in \Sigma$, $\mathrm{hs}_a(L(\mathcal{D})) \subseteq R_a$.

Let $L \stackrel{\text{def}}{=} L(\mathcal{D})$.

$\mathrm{root}(L) \subseteq I$: For all trees $t \in L$, $t \to^*_{\mathcal{A}_\mathcal{D}} s \in I$, thus $\mathrm{root}(t) \in I$ by Question 2a.

$\forall a \in \Sigma\,.\,\mathrm{hs}_a(L) \subseteq R_a$: Consider a tree $C[a(t_0 \cdots t_{n-1})] \in L$ for some context C and trees $t_0, \ldots, t_{n-1}$. Then there exist a transition rule $(a(R_a) \to a) \in \Delta$ and states $a_0, \ldots, a_{n-1} \in Q$ such that $C[a] \to^*_{\mathcal{A}_\mathcal{D}} s$, $t_i \to^*_{\mathcal{A}_\mathcal{D}} a_i$ for all $0 \leq i < n$, and $a_0 \cdots a_{n-1} \in R_a$. By Question 2a, $\mathrm{root}(t_i) = a_i$ for all $0 \leq i < n$. Thus $\mathrm{root}(t_0) \cdots \mathrm{root}(t_{n-1}) \in R_a$.

(c) Show that, for any DTD $\mathcal{D} = (\Sigma, (R_a)_{a\in\Sigma}, I)$, $\mathscr{L}(I, (R_a)_{a\in\Sigma}) \subseteq L(\mathcal{D})$.

Let $L \stackrel{\text{def}}{=} L(\mathcal{D})$. We show more generally by induction over $t \in T(\Sigma)$ that, for all $a \in \Sigma$, if $t \in \mathscr{L}(\Sigma, (R_a)_{a\in\Sigma})$, then $t \to^*_{\mathcal{A}_\mathcal{D}} \mathrm{root}(a)$. This entails the result since $t \in \mathscr{L}(I, (R_a)_{a\in\Sigma})$ implies $\mathrm{root}(t) \in I$ and $t \in \mathscr{L}(\Sigma, (R_a)_{a\in\Sigma})$.

Consider any $t \in \mathscr{L}(I, (R_a)_{a\in\Sigma})$. Let $a \stackrel{\text{def}}{=} \mathrm{root}(t)$; then $t = a(t_0 \cdots t_{n-1})$ for some trees $t_0, \ldots, t_{n-1}$. For all $0 \leq i < n$, let us set $a_i \stackrel{\text{def}}{=} \mathrm{root}(t_i)$; then $t_i \in \mathscr{L}(\Sigma, (R_a)_{a\in\Sigma})$ and by induction hypothesis $t_i \to^*_{\mathcal{A}_\mathcal{D}} a_i$. Since $t \in \mathscr{L}(\Sigma, (R_a)_{a\in\Sigma})$, $a_0 \cdots a_{n-1} \in R_a$, thus we have $t = a(t_0 \cdots t_{n-1}) \to^*_{\mathcal{A}_\mathcal{D}} a(a_0 \cdots a_{n-1}) \to_{\mathcal{A}_\mathcal{D}} a = \mathrm{root}(t)$.

(d) Deduce that, for any DTD $\mathcal{D} = (\Sigma, (R_a)_{a\in\Sigma}, I)$, $\mathscr{L}(\mathrm{root}(L(\mathcal{D})), (\mathrm{hs}_a(L(\mathcal{D})))_{a\in\Sigma}) = \mathscr{L}(I, (R_a)_{a\in\Sigma}) = L(\mathcal{D})$.

Let $L \stackrel{\text{def}}{=} L(\mathcal{D})$. By Question 2b and monotonicity of $\mathscr{L}$, Question 2c, and Question 4 of Exercise 2,

$$\mathscr{L}(\mathrm{root}(L), (\mathrm{hs}_a(L))_{a\in\Sigma}) \subseteq \mathscr{L}(I, (R_a)_{a\in\Sigma}) \subseteq L \subseteq \mathscr{L}(\mathrm{root}(L), (\mathrm{hs}_a(L))_{a\in\Sigma})\,.$$

(e) Conclude.

First assume that $L = L(\mathcal{D})$ for a DTD $\mathcal{D}$. Then $L = \mathscr{L}(\mathrm{root}(L), (\mathrm{hs}_a(L))_{a\in\Sigma})$ by Question 2d, thus L is local by Question 5 of Exercise 2. Since $L = L(\mathcal{A}_\mathcal{D})$, it is also recognisable.

Conversely, let L be a recognisable local tree language. Then by questions 3 and 5 of Exercise 2, $L = \mathscr{L}(\mathrm{root}(L), (\mathrm{hs}_a(L))_{a\in\Sigma})$ where each $\mathrm{hs}_a(L)$ is recognisable. Let us define the DTD $\mathcal{D} \stackrel{\text{def}}{=} (\Sigma, (\mathrm{hs}_a(L))_{a\in\Sigma}, \mathrm{root}(L))$. By Question 2d, $L = L(\mathcal{D})$.

3. A *projection* is a mapping $h\colon \Sigma \to \Sigma'$ for Σ, Σ' two alphabets, which can be lifted to a function $h\colon T(\Sigma) \to T(\Sigma')$ by $h(a(t_0 \cdots t_{n-1})) \stackrel{\text{def}}{=} h(a)(h(t_0) \cdots h(t_{n-1}))$. Recognisable tree languages are closed under projections [TATA Thm. 8.3.9].

[2] Show that $L \subseteq T(\Sigma)$ is recognisable if and only if there exist a finite alphabet Σ', a projection $h\colon \Sigma' \to \Sigma$, and a recognisable local tree language $L' \subseteq T(\Sigma')$ such that $L = h(L')$.

$\Longleftarrow$: Since L' is assumed to be recognisable and recognisable tree languages are closed under projections, $h(L')$ is recognisable.

$\Longrightarrow$: Let $\mathcal{A} = (Q, \Sigma, \Delta, I)$ be an NFHA such that $L = L(\mathcal{A})$. Define $\sigma\colon Q \to \mathrm{Rec}(\Delta^*)$ by $\sigma(q) \stackrel{\text{def}}{=} \{\delta \in \Delta \mid \exists R \exists a \,.\, \delta = (a(R) \to q)\}$. We construct a DTD $\mathcal{D} \stackrel{\text{def}}{=} (\Delta, (R_\delta)_\delta, \sigma(I))$ where $R_\delta \stackrel{\text{def}}{=} \bigcup_{(a(R)\to q)\in\Delta} \sigma(R)$. Then by Question 2, $L' \stackrel{\text{def}}{=} L(\mathcal{D})$ is recognisable and local. Finally, define two projections $h\colon \Delta \to \Sigma$ by $h(a(R) \to q) \stackrel{\text{def}}{=} a$ and $g\colon \Delta \to Q$ by $g(a(R) \to q) \stackrel{\text{def}}{=} q$.

It remains to show that $L = h(L')$ in order to conclude. We show by induction over $t \in T(\Sigma)$ that $t \to_\mathcal{A}^* q$ if and only if there exist $t' \in T(\Delta)$ and $\delta \in \Delta$ such that $h(t') = t$, $g(\delta) = q$, and $t' \to_{\mathcal{A}_\mathcal{D}}^* \delta$.

$\Longleftarrow$: Let $t = a(t_0 \cdots t_{n-1})$ for some $a \in \Sigma$ and $t_0, \ldots, t_{n-1} \in T(\Sigma)$. Then $t \to_\mathcal{A}^* q$ implies that there exist $\delta = (a(R) \to q)$ in Δ and states $q_0, \ldots, q_{n-1} \in Q$ such that $t_i \to_\mathcal{A}^* q_i$ for all $0 \le i < n$ and $q_0 \cdots q_{n-1} \in R$. By induction hypothesis, for all $0 \le i < n$ there exist t'_i and δ_i such that $h(t'_i) = t_i$, $g(\delta_i) = q_i$, and $t'_i \to_{\mathcal{A}_\mathcal{D}}^* \delta_i$. Since $q_0 \cdots q_{n-1} \in R$, $\delta_0 \cdots \delta_{n-1} \in \sigma(R)$. Then $t' \stackrel{\text{def}}{=} \delta(t'_0 \cdots t'_{n-1})$ is such that $h(t') = a(h(t'_0) \cdots h(t'_{n-1})) = a(t_0 \cdots t_{n-1})$, $g(\delta) = q$, and $t' \to_{\mathcal{A}_\mathcal{D}}^* \delta(\delta_0 \cdots \delta_{n-1}) \to_{\mathcal{A}_\mathcal{D}} \delta$ using the rule $\delta(\sigma(R)) \to \delta$ of $\mathcal{A}_\mathcal{D}$.

$\Longrightarrow$: If $t' \to_{\mathcal{A}_\mathcal{D}}^* \delta = (a(R) \to q)$, by Question 2a, $t' = \delta(t'_0 \cdots t'_{n-1})$ for some $t'_0, \ldots, t'_{n-1} \in T(\Delta)$, thus there exist a rule $\delta(\sigma(R)) \to \delta$ of $\mathcal{A}_\mathcal{D}$ and a sequence $\delta_0 \cdots \delta_{n-1} \in \sigma(R)$ such that $t'_i \to_{\mathcal{A}_\mathcal{D}}^* \delta_i$ for all $0 \le i < n$. By induction hypothesis, $h(t'_i) \to^* g(\delta_i)$ for all $0 \le i < n$. Since $\delta_0 \cdots \delta_{n-1} \in \sigma(R)$, $g(\delta_0) \cdots g(\delta_{n-1}) \in R$, hence there is a derivation $h(t') = a(h(t'_0) \cdots h(t'_{n-1})) \to_\mathcal{A}^* a(g(\delta_0) \cdots g(\delta_{n-1})) \to_\mathcal{A} q = g(\delta)$.

Thus $t \in L$ if and only if there exist $t' \in T(\Delta)$ and $\delta \in \Delta$ such that $h(t') = t$, $g(\delta) \in I$, and $t' \to_{\mathcal{A}_\mathcal{D}}^* \delta$, which is if and only if there exists $t' \in L'$ since

$g(\delta) \in I$ if and only if $\delta \in \sigma(I)$.

[4] 4. Show that, from a DTD $\mathcal{D} = (\Sigma, (R_a)_{a\in\Sigma}, I)$, we can construct a PDL($\downarrow, \rightarrow$) state formula $\varphi_{\mathcal{D}}$ with atomic propositions in Σ that defines the same tree language: for all trees $t \in T(\Sigma)$, $t, \varepsilon \models \varphi_{\mathcal{D}}$ iff $t \in L(\mathcal{D})$.

(a) First show that, for any rational expression E defined by the abstract syntax

$$E ::= \varepsilon \mid a \mid E + E \mid E \cdot E \mid E^* \tag{†}$$

where a ranges over Σ, one can construct a PDL($\rightarrow$) path formula π_E with atomic propositions in Σ such that, for all $t \in T(\Sigma)$, $p \in \mathbb{N}^*$ and $i \in \mathbb{N}$ such that $p \cdot i \in \operatorname{dom} t$, and $w \in \operatorname{dom} t$,

$$t, p \cdot i, w \models \pi_E \iff \exists j \geq i \,.\, w = p \cdot j \text{ and } t(p \cdot (i+1)) \cdots t(p \cdot j) \in L(E)\ . \tag{‡}$$

We define π_E by induction over E.

$$\pi_\varepsilon \stackrel{\text{def}}{=} \top? \qquad \pi_a \stackrel{\text{def}}{=} \rightarrow; a?$$

$$\pi_{E+F} \stackrel{\text{def}}{=} \pi_E + \pi_F \qquad \pi_{E\cdot F} \stackrel{\text{def}}{=} \pi_E; \pi_F$$

$$\pi_{E^*} \stackrel{\text{def}}{=} (\pi_E)^*$$

It remains to prove (‡) by induction over E.

ε: $t, p \cdot i, w \models \top?$ iff $w = p \cdot i$.

a: $t, p \cdot i, w \models \rightarrow; a?$ iff there $w = p \cdot (i+1)$ and $t, p \cdot (i+1) \models a$, i.e. $t(p \cdot (i+1)) = a$.

$E + F$: $t, p{\cdot}i, w \models \pi_E + \pi_F$ iff $t, p{\cdot}i, w \models \pi_E$ and $t, p{\cdot}i, w \models \pi_F$, which by ind. hyp. is iff there exists $j \geq i$ such that $w = p{\cdot}j$, $t(p{\cdot}(i+1)) \cdots t(p{\cdot}j) \in L(E)$ and $t(p \cdot (i+1)) \cdots t(p \cdot j) \in L(F)$ and the latter two conditions are iff $t(p \cdot (i+1)) \cdots t(p \cdot j) \in L(E + F)$.

$E \cdot F$: $t, p \cdot i, w \models \pi_E; \pi_F$ iff there exist $w' \in \operatorname{dom} t$, $t, p \cdot i, w' \models \pi_E$ and $t, w', w \models \pi_F$. By ind. hyp., the first one is iff there exists $j \geq i$ such that $w' = p \cdot j$ and $t(p \cdot (i+1)) \cdots t(p \cdot j) \in L(E)$. The second one is thus iff there exists $k \geq j$ such that $w = p \cdot k$ and $t(p \cdot (j+1)) \cdots t(p \cdot k) \in L(F)$. The language conditions are iff $t(p \cdot (i+1)) \cdots t(p \cdot k) \in L(E \cdot F)$.

E^*: $t, p \cdot i, w \models (\pi_E)^*$ iff there exist $n \in \mathbb{N}$ and $w_0, \ldots, w_n \in \operatorname{dom} t$ such that $w_0 = p \cdot i$, $w_n = w$, and $t, w_k, w_{k+1} \models \pi_E$ for all $0 \leq k < n$. Using the ind. hyp. this is iff either $n = 0$ and then $w = p \cdot i$ or there exist $i = i_0 \leq i_1 \leq \cdots \leq i_n$ s.t. $w_k = p \cdot i_k$ for all $0 \leq k \leq n$ and $t(p \cdot i_k + 1) \cdots t(p \cdot i_{k+1}) \in L(E)$ for all $k < n$. The language conditions are iff $t(p \cdot i + 1) \cdots t(p \cdot i_n) \in L(E^*)$.

(b) Conclude, assuming that R_a is given as a rational expression for each $a \in \Sigma$. *Hint: You may use the fact (due to Antimirov) that, for any rational expression E and $a \in \Sigma$, we can construct a finite set $\partial_a E$ of rational expressions over the syntax (†) such that $L(E) = \lambda(E) \cup \bigcup_{a\in\Sigma}(\{a\} \cdot \bigcup_{F\in\partial_a E} L(F))$ where $\lambda(E) \stackrel{def}{=} \{\varepsilon\}$ if $\varepsilon \in L(E)$ and $\lambda(E) \stackrel{def}{=} \emptyset$ otherwise.*

For a rational expression E with $L(E) = \lambda(E) \cup \bigcup_{a\in\Sigma}(\{a\} \cdot \bigcup_{F\in\partial_a E} L(F))$, we define a node formula

$$\varphi_E \stackrel{\text{def}}{=} \varphi_{\lambda(E)} \vee \bigvee_{a\in\Sigma} \langle\downarrow\rangle\big(\mathsf{first} \wedge a \wedge \bigvee_{F\in\partial_a E} \langle\pi_F\rangle \mathsf{last}\big)$$

where $\varphi_{\{\varepsilon\}} \stackrel{\text{def}}{=} \mathsf{leaf}$ and $\varphi_\emptyset \stackrel{\text{def}}{=} \bot$. Then $t, w \models \varphi_E$ if and only if the sequence of child positions of w in t is labelled by a word in $L(E)$. For example, if $E = (ab)^*$, then $\lambda(E) = \{\varepsilon\}$, $\partial_a E = \{b \cdot (ab)^*\}$, and $\partial_b E = \emptyset$, and we obtain the formula

$$\varphi_{(ab)^*} = \mathsf{leaf} \vee \langle\downarrow\rangle(\mathsf{first} \wedge a \wedge \langle\rightarrow; b?; (\rightarrow; a?; \rightarrow; b?)^*\rangle \mathsf{last}) \vee \langle\downarrow\rangle(\mathsf{first} \wedge b \wedge \bot)\ ,$$

which is equivalent to:

$$\mathsf{leaf} \vee \langle\downarrow\rangle(\mathsf{first} \wedge a \wedge \langle\rightarrow; b?; (\rightarrow; a?; \rightarrow; b?)^*\rangle \mathsf{last})\ .$$

Finally, given a DTD $\mathcal{D} = (\Sigma, (R_a)_{a\in\Sigma}, I)$ where for all $a \in \Sigma$, $R_a = L(E_a)$ for a rational expression E_a, we define

$$\varphi_{\mathcal{D}} \stackrel{\text{def}}{=} \big(\bigvee_{a\in I} a\big) \wedge [\downarrow^*]\big(\bigwedge_{a\in\Sigma} a \Longrightarrow \varphi_{E_a}\big)\ .$$