MPRI 2-27-1 Exam

Duration: 3 hours
Written documents are allowed. The numbers in front of questions are indicative of hardness or duration.

1 Multiple Context-Free Grammars

Multiple Context-Free Grammars are a mildly context-sensitive formalism defined by Seki, Matsumura, Fuji, and Kasami in 1991. The purpose of this section is to instantiate the 'parsing as intersection' framework in their case.

Exercise 1 (Multiple Context-Free Grammars)**.** Let $\mathcal{X}$ be an infinite countable set of variables. A *multiple context-free grammar* (MCFG) of rank m and degree d is a tuple $\mathcal{G} = \langle N, \Sigma, P, S\rangle$ where N is a finite alphabet of nonterminals $A^{(r)}$ with positive ranks $0 < r \leq m$, Σ a finite alphabet of terminals, $S^{(1)} \in N$ is the start symbol with rank 1, and P is a finite set of productions p of form

$$A^{(r_0)}(\alpha_1, \ldots, \alpha_{r_0}) :\!-\; B_1^{(r_1)}(x_{1,1}, \ldots, x_{1,r_1}), \ldots, B_k^{(r_k)}(x_{k,1}, \ldots, x_{k,r_k}). \tag{p}$$

where $0 \leq k \leq d$ is the degree of the production, $A^{(r_0)}, B_1^{(r_1)}, \ldots, B_k^{(r_k)}$ are nonterminals in N, $x_{1,1}, \ldots, x_{k,r_k}$ are distinct variables from $\mathcal{X}$, and $\alpha_1, \ldots, \alpha_{r_0}$ are *linear strings* α over $\Sigma \uplus \{x_{1,1}, \ldots, x_{k,r_k}\}$, i.e. strings where each variable $x_{i,j}$ occurs exactly once. Note that if $k = 0$, this entails $\alpha_j \in \Sigma^*$ for all $1 \leq j \leq r_0$.

An MCFG $\mathcal{G}$ defines a deduction system over *judgements* of form $\vdash_{\mathcal{G}} A^{(r)}(w_1, \ldots, w_r)$ where $A^{(r)}$ is a nonterminal from N and $w_1, \ldots, w_r$ are finite strings over Σ. This deduction system has a rule

$$\frac{\vdash_{\mathcal{G}} B_1^{(r_1)}(w_{1,1}, \ldots, w_{1,r_1}) \quad \cdots \quad \vdash_{\mathcal{G}} B_k^{(r_k)}(w_{k,1}, \ldots, w_{k,r_k})}{\vdash_{\mathcal{G}} A^{(r_0)}(\alpha_1\sigma, \ldots, \alpha_{r_0}\sigma)} \tag{$\vdash_{\mathcal{G}}$}$$

for every p in P and strings $w_{1,1}, \ldots, w_{k,r_k}$ in Σ^*, where σ is the substitution $x_{i,j} \mapsto w_{i,j}$.

The *language* of a nonterminal $A^{(r)}$ is the set of r-tuples of strings defined by

$$L_{\mathcal{G}}(A) \stackrel{\text{def}}{=} \{(w_1, \ldots, w_r) \in (\Sigma^*)^r \mid \vdash_{\mathcal{G}} A^{(r)}(w_1, \ldots, w_r)\} .$$

The *language* generated by $\mathcal{G}$ is accordingly the language of its start symbol $S^{(1)}$:

$$L(\mathcal{G}) \stackrel{\text{def}}{=} \{w \in \Sigma^* \mid \vdash_{\mathcal{G}} S^{(1)}(w)\} .$$

Example 1 (Copy Language)**.** The MCFG $\mathcal{G}_{\text{copy}} = \langle\{S^{(1)}, A^{(2)}\}, \{a, b\}, P, S\rangle$ with productions

$$\begin{aligned} S^{(1)}(xy) &:\!-\ A^{(2)}(x, y).\\ A^{(2)}(ax, ay) &:\!-\ A^{(2)}(x, y).\\ A^{(2)}(bx, by) &:\!-\ A^{(2)}(x, y).\\ A^{(2)}(\varepsilon, \varepsilon) &:\!-\ . \end{aligned}$$

generates the language $L(\mathcal{G}_{\text{copy}}) = \{ww \mid w \in \{a, b\}^*\}$.

[1] 1. Give a MCFG for the language $L_{\text{cross}} \stackrel{\text{def}}{=} \{a^n b^m c^n d^m \mid n, m \geq 0\}$.

[2] 2. Show that any context-free language is generated by an MCFG of rank 1.

Exercise 2 (Emptiness of MCFGs)**.** The first main ingredient in the 'parsing as intersection' framework is to prove that the emptiness problem is decidable for MCFGs. In order to consider complexity questions, we define the size of a MCFG $\mathcal{G} = \langle N, \Sigma, P, S\rangle$ by summing $k + \sum_{j=1}^{r_0}(|\alpha_j| + 1)$ over all the productions p in P.

[3] 1. Show that there exists a linear-time algorithm that inputs an MCFG $\mathcal{G}$ and returns whether $L(\mathcal{G}) = \emptyset$.

Exercise 3 (Intersection with a Regular Language)**.** The second ingredient of the 'parsing as intersection' framework is to show that the class of languages generated by MCFGs is closed under intersection with regular languages.

[2] 1. As a preliminary, show that for any MCFG, one can construct in linear time an equivalent MCFG where the productions p in P with $k > 0$ enforce $\alpha_j \in \mathcal{X}^*$, i.e. no terminal symbol appears in such productions, and each α_j is of form $y_1 \cdots y_{n_j}$ for $y_1, \ldots, y_{n_j}$ variables taken among $x_{1,1}, \ldots, x_{k,r_k}$.

[5] 2. Show that, given an MCFG $\mathcal{G} = \langle N, \Sigma, P, S\rangle$ and a nondeterministic finite automaton $\mathcal{A} = \langle Q, \Sigma, \delta, I, F\rangle$, we can compute an MCFG $\mathcal{G}'$ such that $L(\mathcal{G}') = L(\mathcal{G}) \cap L(\mathcal{A})$ and $|\mathcal{G}'| \in O(|\mathcal{G}| \cdot |Q|^{m \max(k+1,2)})$.

   Hint: Use nonterminals $A^{(r)}_{q_1,q'_1,\ldots,q_r,q'_r}$ such that $\vdash_{\mathcal{G}'} A^{(r)}_{q_1,q'_1,\ldots,q_r,q'_r}(w_1, \ldots, w_r)$ if and only if $\vdash_{\mathcal{G}} A^{(r)}(w_1, \ldots, w_r)$ and $q_j \xrightarrow{w_j}_{\mathcal{A}} q'_j$ for all $1 \leq j \leq r$.

[1] 3. Deduce an algorithm for the *membership problem*, which given an MCFG $\mathcal{G} = \langle N, \Sigma, P, S\rangle$ and a string w in Σ^*, returns whether $w \in L(\mathcal{G})$.

2 Covert Movements in Second-Order ACGs

In the exercises that follow, one only considers 2nd-order ACGs. This allows one not to bother about linearity constraints, and to work in the setting of the simply-typed λ-calculus.

Exercise 4. One considers the three following signatures:

$$(\Sigma_{\text{ABS}}) \qquad \begin{aligned} \textsc{pierre} &: NP \\ \textsc{maison} &: N \\ \textsc{une} &: N \to QNP \\ \textsc{acheter} &: QNP \to VP \\ \textsc{veux} &: VP \to NP \to S \end{aligned}$$

$$(\Sigma_{\text{S-FORM}}) \qquad \begin{aligned} /Pierre/ &: string \\ /maison/ &: string \\ /une/ &: string \\ /acheter/ &: string \\ /veux/ &: string \end{aligned}$$

where, as usual, *string* is defined to be $o \to o$ for some atomic type o.

$$(\Sigma_{\text{L-FORM}}) \qquad \begin{aligned} \mathbf{p} &: \mathsf{ind} \\ \mathbf{house} &: \mathsf{ind} \to \mathsf{prop} \\ \mathbf{buy} &: \mathsf{ind} \to \mathsf{ind} \to \mathsf{prop} \\ \mathbf{want} &: \mathsf{ind} \to \mathsf{prop} \to \mathsf{prop} \end{aligned}$$

One then defines two morphisms ($\mathcal{L}_{\text{SYNT}} : \Sigma_{\text{ABS}} \to \Sigma_{\text{S-FORM}}$, and $\mathcal{L}_{\text{SEM}} : \Sigma_{\text{ABS}} \to \Sigma_{\text{L-FORM}}$) as follows:

$$(\mathcal{L}_{\text{SYNT}}) \qquad \begin{aligned} NP &:= string \\ N &:= string \\ QNP &:= string \\ VP &:= string \\ S &:= string \end{aligned}$$

$$\begin{aligned} \textsc{pierre} &:= /Pierre/ \\ \textsc{maison} &:= /maison/ \\ \textsc{une} &:= \lambda x.\, /une/ + x \\ \textsc{acheter} &:= \lambda x.\, /acheter/ + x \\ \textsc{veux} &:= \lambda xy.\, y + /veux/ + x \end{aligned}$$

where, as usual, the concatenation operator $(+)$ is defined as functional composition.

$$
(\mathcal{L}_{\mathrm{SEM}}) \qquad
\begin{aligned}
NP &:= (\mathsf{ind} \rightarrow \mathsf{prop}) \rightarrow \mathsf{prop} \\
N &:= \mathsf{ind} \rightarrow \mathsf{prop} \\
QNP &:= (\mathsf{ind} \rightarrow \mathsf{prop}) \rightarrow \mathsf{prop} \\
VP &:= \mathsf{ind} \rightarrow \mathsf{prop} \\
S &:= \mathsf{prop}
\end{aligned}
$$

$$
\begin{aligned}
\textsc{Pierre} &:= \lambda x.\, x\, \mathbf{p} \\
\textsc{maison} &:= \mathbf{house} \\
\textsc{une} &:= \lambda xy.\, \exists z.\, (x\, z) \wedge (y\, z) \\
\textsc{acheter} &:= \lambda xy.\, x\, (\lambda z.\, \mathbf{buy}\, y\, z) \\
\textsc{veux} &:= \lambda xy.\, y\, (\lambda z.\, \mathbf{want}\, z\, (x\, z))
\end{aligned}
$$

[2] 1. Check that $\mathcal{L}_{\mathrm{SEM}}$ is such that the interpretation it gives to ACHETER is consistent with the interpretation it gives to the types.

[1] 2. Give a term, say t, such that:

$$\mathcal{L}_{\mathrm{SYNT}}(t) = /\mathit{Pierre}/ + /\mathit{veux}/ + /\mathit{acheter}/ + /\mathit{une}/ + /\mathit{maison}/$$

Then, compute $\mathcal{L}_{\mathrm{SEM}}(t)$.

Exercise 5. One extends Σ_{ABS} with the following constants (and types):

$$
\begin{aligned}
\textsc{trace} &: XNP \\
\textsc{x-acheter} &: XNP \rightarrow XVP \\
\textsc{x-veux} &: XVP \rightarrow NP \rightarrow XS \\
\textsc{qr} &: QNP \rightarrow XS \rightarrow S
\end{aligned}
$$

Accordingly, one extends $\mathcal{L}_{\mathrm{SYNT}}$ as follows:

$$
\begin{aligned}
XNP &:= \mathit{string} \rightarrow \mathit{string} \\
XVP &:= \mathit{string} \rightarrow \mathit{string} \\
XS &:= \mathit{string} \rightarrow \mathit{string}
\end{aligned}
$$

$$
\begin{aligned}
\textsc{trace} &:= \lambda x.\, x \\
\textsc{x-acheter} &:= \lambda xy.\, /\mathit{acheter}/ + (x\, y) \\
\textsc{x-veux} &:= \lambda xyz.\, y + /\mathit{veux}/ + (x\, z) \\
\textsc{qr} &:= \lambda xy.\, y\, x
\end{aligned}
$$

[1] 1. Compute the interpretation of the following term (according to the above extension of $\mathcal{L}_{\mathrm{SYNT}}$):

$$\textsc{qr}\,(\textsc{une}\,\textsc{maison})\,(\textsc{x-veux}\,(\textsc{x-acheter}\,\textsc{trace})\,\textsc{pierre}) \qquad (t_{\mathrm{re}})$$

Exercise 6. One also extends $\mathcal{L}_{\text{SEM}}$ as follows:

$$
\begin{aligned}
XNP &:= \mathsf{ind} \to (\mathsf{ind} \to \mathsf{prop}) \to \mathsf{prop} \\
XVP &:= \mathsf{ind} \to \mathsf{ind} \to \mathsf{prop} \\
XS &:= \mathsf{ind} \to \mathsf{prop}
\end{aligned}
$$

$$
\begin{aligned}
\textsc{trace} &:= \lambda xy.\, y\, x \\
\textsc{x-acheter} &:= \lambda wxy.\, w\, x\, (\lambda z.\, \mathbf{buy}\, y\, z) \\
\textsc{x-veux} &:= \cdots \\
\textsc{qr} &:= \cdots
\end{aligned}
$$

[3] 1. Complete the above extension (i.e., provide the interpretations of X-VEUX and QR) in such a way that $\mathcal{L}_{\text{SEM}}(t_{\text{re}})$ yields a *de re* interpretation.