Introduction to Tree Language Theory

Hitoshi Ohsaki


National Institute of
Advanced Industrial Science and Technology (AIST)

seminar talk (2/10)

2009

II. Grammar

Grammar

grammar : $\mathcal{G} = (\Sigma, T, N, \mathsf{q}_0, \Delta)$

Σ	:	alphabet
T	:	set of terminal symbols such that $T \subseteq \Sigma$
N	:	set of non-terminal symbols such that $N = \Sigma - T$
q_0	:	start symbol such that $\mathsf{q}_0 \in N$
Δ	:	finite set of production rules with the following forms

$$\alpha \rightarrow \beta \qquad (\alpha, \beta \in \Sigma^*)$$

Cf. regular grammar if production rules are in the following forms:

$$p \rightarrow a \qquad p \rightarrow a\,q \qquad p \rightarrow \varepsilon \qquad (p, q \in N, \quad a \in T)$$

Generated languages

Given grammar $\mathcal{G} = (\Sigma, T, N, \mathsf{q}_0, \Delta)$, define for words u, w over Σ,

$u \to_{\mathcal{G}} w$: $\exists\ \alpha \to \beta$ in Δ such that

- u is decomposed to words u_1, α, u_2
- w is decomposed to words u_1, β, u_2

$u_0 \to^*_{\mathcal{G}} u_n$: if $u_0 \equiv u_n$, or
if $u_0 \to_{\mathcal{G}} u_1$ & $u_1 \to^*_{\mathcal{G}} u_n$
(the reflexive and transitive closure of $\to_{\mathcal{G}}$)

$\mathcal{L}(\mathcal{G})$: set of words over T such that for all w in $\mathcal{L}(\mathcal{G})$, $\mathsf{q}_0 \to^*_{\mathcal{G}} w$
(language generated by $\mathcal{G}$)

Note

Question if $w \in \mathcal{L}(\mathcal{G})$ is undecidable in general. (Cf. Rice's theorem in seminar 5)

Decidable sub-classes

grammar whose production rules are in the following forms is called

context-sensitive : $\alpha \rightarrow \beta$ ($\alpha, \beta \in \Sigma^*$ such that $|\alpha| \leqslant |\beta|$,

$q_0 \rightarrow \varepsilon$ start symbol q_0 never appears on right-hand side of any rule)

context-free : $p \rightarrow \beta$ ($p \in N$, $\beta \in \Sigma^*$)

language is called

context-sensitive if it is generated by context-sensitive grammar

context-free if it is generated by context-free grammar

Note

Membership problem, i.e. the question if $w \in \mathcal{L}(\mathcal{G})$, is decidable for the classes of context-sensitive grammar and context-free grammar. Why? (Exercise)

Example

Define the production rules

$$\begin{array}{lcl}
\Delta_1 & : & \mathsf{q}_0 \to \mathsf{a}\,\mathsf{q}_0\mathsf{b} \qquad \mathsf{q}_0 \to \varepsilon \\
\Delta_2 & : & \mathsf{q}_0 \to \mathsf{a}\,\mathsf{q}_1 \qquad \mathsf{q}_0 \to \mathsf{b}\,\mathsf{q}_2 \qquad \mathsf{q}_0 \to \varepsilon \\
 & & \mathsf{q}_1 \to \mathsf{a}\,\mathsf{q}_1\,\mathsf{q}_1 \qquad \mathsf{q}_1 \to \mathsf{b}\,\mathsf{q}_0 \\
 & & \mathsf{q}_2 \to \mathsf{a}\,\mathsf{q}_0 \qquad \mathsf{q}_2 \to \mathsf{b}\,\mathsf{q}_2\,\mathsf{q}_2
\end{array}$$

for the grammar

$$\mathcal{G}_1 = (\Sigma_1, \{\mathsf{a},\mathsf{b}\}, \{\mathsf{q}_0\}, \mathsf{q}_0, \Delta_1)$$

$$\mathcal{G}_2 = (\Sigma_2, \{\mathsf{a},\mathsf{b}\}, \{\mathsf{q}_0,\mathsf{q}_1,\mathsf{q}_2\}, \mathsf{q}_0, \Delta_2)$$

then

$$\mathcal{L}(\mathcal{G}_1) = \{\mathsf{a}^n\,\mathsf{b}^n \mid n \geqslant 0\} \qquad \mathcal{L}(\mathcal{G}_2) = \{w \in \{\mathsf{a},\mathsf{b}\}^* \mid |w|_\mathsf{a} = |w|_\mathsf{b}\}$$

where $|w|_\mathsf{a}$ (resp. $|w|_\mathsf{b}$) is the number of occurrences of a (b) in w

Eliminating ε-derivations

Given grammar $\mathcal{G} = (\Sigma, T, N, \mathsf{q}_0, \Delta)$, for non-terminal symbol q in N, q is called nullable if q admits $q \rightarrow^*_{\mathcal{G}} \varepsilon$.

Claim 1

One can compute the set N_ε of nullable non-terminal symbols if $\mathcal{G}$ is context-free grammar

Claim 2

If $\mathcal{G}$ is context-free grammar, define $\mathcal{G}' = (\Sigma, T, N, \mathsf{q}_0, \Delta')$ where

$$\begin{aligned}\Delta' \;=\; \{\; & q \rightarrow \alpha_0\,\beta_1\,\alpha_1 \cdots \beta_n\,\alpha_n \;\mid\; \exists\, p \rightarrow \alpha_0\,p_1\,\alpha_1 \cdots p_n\,\alpha_n \in \Delta \;: \\ & \exists\, \alpha_i \in (T \cup N - N_\varepsilon)^*,\;\; \exists\, p_1, \ldots, p_n \in N_\varepsilon,\;\; \beta_i \in \{\, p_i, \varepsilon \,\} \\ & (1 \leqslant i \leqslant n) \;\}\end{aligned}$$

Then

$$\mathcal{L}(\,\mathcal{G}'\,) = \mathcal{L}(\,\mathcal{G}\,) - \{\,\varepsilon\,\}$$

... show this claim (Exercise)

Proof of Claim 1

We show that for each $q \in N$, q is nullable if and only if $q \in N_\varepsilon$, where N_ε is obtained by the following procedure. This procedure halts on any context-free grammar.

$N_\varepsilon := \varnothing$
while $S = \varnothing$ do
  $S := \{ q \in N - N_\varepsilon \mid \exists q \to \alpha \in \Delta \text{ such that } \alpha \in {N_\varepsilon}^* \}$;
  $N_\varepsilon := N_\varepsilon \cup S$
od
return N_ε

The "if" part is shown by induction on the number n of loops. For induction step, let S_n be the set S obtained at the n-th loop ($n \geqslant 1$). For $q \in S_n$, there is $q \to \alpha$ with $\alpha \in {N_\varepsilon}^*$. By induction hypothesis, non-terminals of α are all nullable, so $\alpha \to_{\mathcal{G}}^* \varepsilon$, and hence $q \to_{\mathcal{G}}^* \varepsilon$.

The "only if" part is shown by the length of $q \to_{\mathcal{G}}^* \varepsilon$. If the length is 1, Δ contains $q \to \varepsilon$, and thus, q is added to N_ε at 1st loop. For induction step, suppose $q \to_{\mathcal{G}} \alpha$ and $\alpha \to_{\mathcal{G}}^* \varepsilon$ whose length is n ($n \geqslant 1$) and $\alpha \in {N_\varepsilon}^*$. By assumption, Δ contains $q \to \alpha$. If $q \notin N_\varepsilon$, $q \in S$ at some loop, because it satisfies that $q \in N - N_\varepsilon$, $q \to \alpha \in \Delta$ and $\alpha \in {N_\varepsilon}^*$. Hence, $q \in N_\varepsilon$. □

Proposition

For context-free grammar $\mathcal{G}$ with the set T of terminal symbols, one can construct context-free grammar $\mathcal{G}' = (\Sigma, T, N, \mathsf{q}_0, \Delta')$ such that $\mathcal{L}(\mathcal{G}') = \mathcal{L}(\mathcal{G})$ and Δ' contains transition rules with the following forms only :

$$q \rightarrow \alpha \qquad (q \in N, \quad \alpha \in (\Sigma - \{\mathsf{q}_0\})^+,$$

$$\mathsf{q}_0 \rightarrow \varepsilon \qquad \mathsf{q}_0 \text{ never appears in the right-hand side of any rule})$$

Proof

This is an easy consequence of the previous Claims 1,2 except the treatment of the start symbol and the production rule of the form $\mathsf{q}_0 \rightarrow \varepsilon$. First, replace the start symbol, say p_0, of $\mathcal{G}$ by q_0 and add $\mathsf{q}_0 \rightarrow \mathsf{p}_0$ to the set of production rules of $\mathcal{G}$. After constructing $\mathcal{G}'$, test if $\mathcal{G}$ generates ε, which is the decidable problem for context-free grammar. If yes, then add $\mathsf{q}_0 \rightarrow \varepsilon$ to Δ' of $\mathcal{G}'$. □

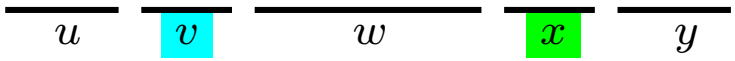
Corollary

Context-free languages are generated by context-sensitive grammar. (This is not obvious only from grammar definitions on page 4, but from the above Proposition.)

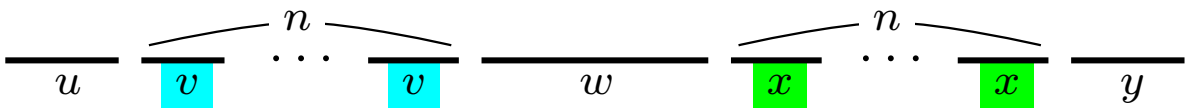
Pumping lemma [Bar-Hillel 1961]

Given context-free language L,

$\exists\ k \geqslant 0$: if $z \in L$ and $|z| \geqslant k$, then z is formed by u, v, w, x, y

as 

such that $|vx| \geqslant 1$ and $u\,v^n\,w\,x^n\,y \in L \quad (n \geqslant 0)$

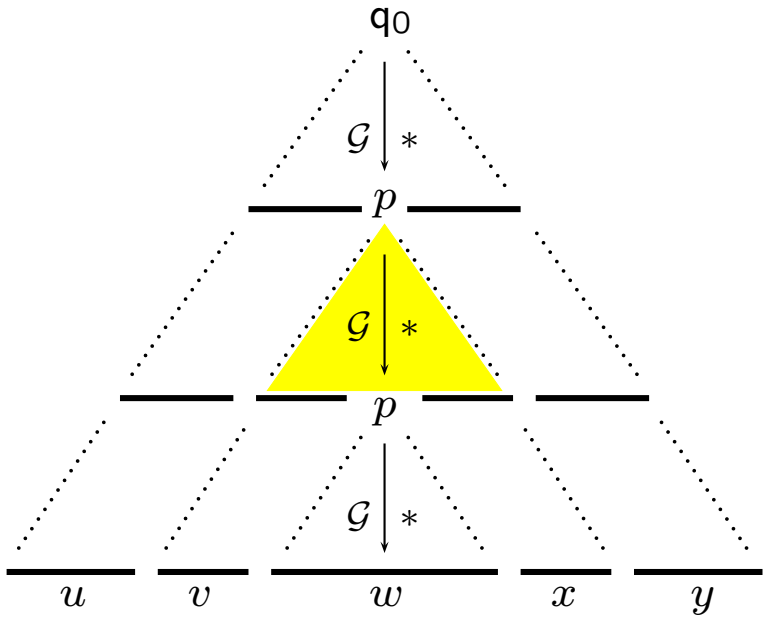
as 

Proof

Suppose $\mathcal{G} = (\Sigma, T, N, \mathsf{q}_0, \Delta)$ is context-free grammar whose production rules are already in the forms of $\mathsf{q}_0 \to \varepsilon$ and $q \to \alpha$ for $\alpha \in (\Sigma - \{\mathsf{q}_0\})^+$. For $p, q \in \Sigma$, we say p is a descendant of q if $q \to^*_{\mathcal{G}} u\,p\,w$. If the length of the derivation $q \to^*_{\mathcal{G}} u\,p\,w$ is more than 0, p is a proper descendant of q. Observe that the (proper) descendant relation is closed under contexts, i.e. p is a descendant of q such that $q \to^*_{\mathcal{G}} u\,p\,w$ if and only if for all s, t in Σ^*, $s\,q\,t \to^*_{\mathcal{G}} s\,u\,p\,w\,t$. For every word z generated by $\mathcal{G}$, one can construct a directed graph of the proper descendant relation associated to z, whose node is marked by an element in Σ and the edge is the proper descendant relation obtained from the derivation $\mathsf{q}_0 \to^*_{\mathcal{G}} z$. (proof cont'd)

Proof (cont'd)

If $\mathcal{G}$ generates a word z whose directed graph of the proper descendant relation admits the path starting from q_0 to a in T such that the length of this path is more than $|N|$, then the path contains a sub-path from p to p. (This situation corresponds to the figure on the right.) If the yellow part is removed, $\mathsf{q}_0 \to_{\mathcal{G}}^{*} u\,w\,y$; if this part is repeated n-times, it turns out $\mathsf{q}_0 \to_{\mathcal{G}}^{*} u\,v^n\,w\,x^n\,y$. Since p is the proper descendant of p in this case, by the shape of rules, $|v\,x| \geqslant 1$. □



Proposition 1

Given context-free grammar $\mathcal{G}$, question if $\mathcal{L}(\mathcal{G}) = \varnothing$ is decidable

Proof

Suppose $\mathcal{G} = (\Sigma, T, N, \mathsf{q}_0, \Delta)$, each of whose rules is already in the form of $\mathsf{q}_0 \to \varepsilon$ or $q \to \alpha$ for $q \in N$, $\alpha \in (\Sigma - \{\mathsf{q}_0\})^+$. Let $\ell = \max\{\,|\alpha| \mid p \to \alpha \in \Delta\,\}$ and $k = \ell^{|N|+1}$, then define $L_{\leqslant k} = \{\,w \in T^* \mid |w| \leqslant k\,\}$. From the proof of the previous lemma, $L_{\leqslant k} = \varnothing$ if and only if $\mathcal{L}(\mathcal{G}) = \varnothing$. Because $L_{\leqslant k}$ is finite, question if $L_{\leqslant k} = \varnothing$ is decidable. □

Example

Using Pumping Lemma, we show that $L = \{ \mathrm{a}^\ell \, \mathrm{b}^\ell \, \mathrm{a}^\ell \mid \ell \geqslant 1 \}$ is not context-free. For leading the contradiction, we suppose below that L is context-free.

Take $z = \mathrm{a}^k \, \mathrm{b}^k \, \mathrm{a}^k$ in L. If k is sufficiently large, z can be decomposed to u, v, w, x, y such that $|v\,x| \geqslant 1$ and $u\,v^n\,w\,x^n\,y \in L$ for all $n \geqslant 0$.

If $|v| = 0$, then $x = \mathrm{a}^i$ or $x = \mathrm{b}^i$ with $i \leqslant k$. However, $u\,v^2\,w\,x^2\,y$ can not be an element in L. Thus, $|v| \neq 0$.

If $|v| \geqslant 1$, then $v = \mathrm{a}^i$ or $v = \mathrm{b}^i$ with $i \leqslant k$. If $|x| = 0$, then it immediately leads to the contradiction, so $|x| \neq 0$.

If $|x| \geqslant 1$ also, then $x = \mathrm{a}^j$ or $x = \mathrm{b}^j$ with $j \leqslant k$. However, in any combination of the possibility of v and x, $u\,v^2\,w\,x^2\,y$ can not be an element in L, leading to the contradiction.

Proposition 2

Given two context-free grammars $\mathcal{G}_1, \mathcal{G}_2$, the intersection-emptiness problem, i.e. the question if $\mathcal{L}(\mathcal{G}_1) \cap \mathcal{L}(\mathcal{G}_2) = \varnothing$, is undecidable.

Proof

Use the reduction from PCP. For an instance of PCP $I = \{\langle u_1, w_1\rangle, \ldots, \langle u_n, w_n\rangle\}$ over Σ, define the grammar $\mathcal{G}_I = (\Sigma \cup \{\mathsf{q}_0\}, \Sigma, \{\mathsf{q}_0\}, \mathsf{q}_0, \Delta_I)$, provided that $\mathsf{q}_0 \notin \Sigma$:

$$\Delta_I \;:\; \mathsf{q}_0 \to \overline{u}_1 \mathsf{q}_0 w_1 \quad \ldots \quad \mathsf{q}_0 \to \overline{u}_n \mathsf{q}_0 w_n \quad \mathsf{q}_0 \to \varepsilon$$

where $\overline{u}_i$ denotes the reverse of word u_i, e.g. $\overline{\mathsf{abcd}} = \mathsf{dcba}$. Moreover, if $\Sigma = \{\mathsf{a}_1, \ldots, \mathsf{a}_k\}$, define the grammar $\mathcal{G}_{\mathsf{eq}} = (\Sigma \cup \{\mathsf{q}_0, \mathsf{q}_1\}, \Sigma, \{\mathsf{q}_0, \mathsf{q}_1\}, \mathsf{q}_0, \Delta_{\mathsf{eq}})$, where

$$\begin{aligned} \Delta_{\mathsf{eq}} \;:\; & \mathsf{q}_0 \to \mathsf{a}_1 \mathsf{q}_1 \mathsf{a}_1 \quad \ldots \quad \mathsf{q}_0 \to \mathsf{a}_k \mathsf{q}_1 \mathsf{a}_k \\ & \mathsf{q}_1 \to \mathsf{a}_1 \mathsf{q}_1 \mathsf{a}_1 \quad \ldots \quad \mathsf{q}_1 \to \mathsf{a}_k \mathsf{q}_1 \mathsf{a}_k \quad \mathsf{q}_1 \to \varepsilon \end{aligned}$$

Note that $\mathcal{L}(\mathcal{G}_{\mathsf{eq}})$ does not contain the empty word ε. Then, by construction, $\mathcal{L}(\mathcal{G}_I) \cap \mathcal{L}(\mathcal{G}_{\mathsf{eq}}) \neq \varnothing$ if and only if I has a solution. However, question if an instance of PCP has a solution is undecidable. □

Closure properties (1)

Let $\mathsf{C}(\mathsf{CF}_T)$ be the set of context-free languages over T of terminal symbols

Proposition

The class $\mathsf{C}(\mathsf{CF}_T)$ is closed under union. However, the class $\mathsf{C}(\mathsf{CF}_T)$ is not closed under intersection or complement if $|T| \geqslant 2$

Proof

For union, let $\mathcal{G}_1 = (\Sigma_1, T, N_1, \mathsf{p}_0, \Delta_1)$ $\mathcal{G}_2 = (\Sigma_2, T, N_2, \mathsf{q}_0, \Delta_2)$ be context-free grammar. We suppose $N_1 \cap N_2 = \varnothing$. Let r_0 be fresh symbol, then define $\mathcal{G} = (\Sigma_1 \cup \Sigma_2 \cup \{\mathsf{r}_0\}, T, N_1 \cup N_2 \cup \{\mathsf{r}_0\}, \mathsf{r}_0, \Delta_1 \cup \Delta_2 \cup \{\mathsf{r}_0 \to \mathsf{p}_0, \mathsf{r}_0 \to \mathsf{q}_0\})$. By construction, for word $w \in T^*$, $\mathcal{G}$ generates w if and only if $\mathcal{G}_1$ or $\mathcal{G}_2$ generates w.

For not being closed under intersection, it suffices to show that $L_1 = \{\, \mathsf{a}^\ell \, \mathsf{b}^\ell \, \mathsf{a}^m \mid \ell, m \geqslant 1 \,\}$ and $L_2 = \{\, \mathsf{a}^m \, \mathsf{b}^\ell \, \mathsf{a}^\ell \mid \ell, m \geqslant 1 \,\}$ are generated by context-free grammar.

For not being closed under complement, use de Morgan's law, because $L_1 \cap L_2 = (\,(L_1)^{\mathsf{C}} \cup (L_2)^{\mathsf{C}}\,)^{\mathsf{C}}$. □

Closure properties (2)

Let $\mathrm{C}(\mathrm{CS}_T)$ be the set of context-sensitive languages over T of terminal symbols

Proposition

The class $\mathrm{C}(\mathrm{CS}_T)$ is closed under union, intersection, complement

Proof

For union, one can apply the same proof of the previous proposition.

For intersection, let $\mathcal{G}_1 = (\Sigma_1, T, N_1, \mathsf{p}_0, \Delta_1)$ and $\mathcal{G}_2 = (\Sigma_2, T, N_2, \mathsf{q}_0, \Delta_2)$ are context-sensitive grammar. Assume that production rules in $\mathcal{G}_1, \mathcal{G}_2$ are Kuroda normal form except the rules $\mathsf{p}_0 \to \varepsilon$, $\mathsf{q}_0 \to \varepsilon$ if each of them exists in Δ_1 or Δ_2, respectively. This assumption does not lose the generality of the discussion. (See Exercise). Define the grammar $\mathcal{G} = (\Sigma_3, T, N, \langle \mathsf{p}_0, \mathsf{q}_0 \rangle, \Delta)$. The set N of non-terminal symbols is defined below, and production rules in Δ are defined in the next slide.

$$N \ : \ N_1 \ \cup \ \{ \langle p, q \rangle \mid p \in N_1, q \in N_2 \}$$

(proof cont'd)

Proof (cont'd)

Δ :

$$
\begin{array}{rcll}
\langle p_1, \mathsf{q}_0 \rangle & \to & \langle q_1, \mathsf{q}_0 \rangle\, r_1 & \text{if } p_1 \to q_1\, r_1 \text{ in } \Delta_1 \\
\langle p_1, \mathsf{q}_0 \rangle\, q_1 & \to & \langle r_1, \mathsf{q}_0 \rangle\, s_1 & \text{if } p_1\, q_1 \to r_1\, s_1 \text{ in } \Delta_1 \\
\langle p_1, \mathsf{q}_0 \rangle & \to & \langle q_1, \mathsf{q}_0 \rangle & \text{if } p_1 \to q_1 \text{ in } \Delta_1 \\
\langle p_1, \mathsf{q}_0 \rangle\, q_1 & \to & p_1\, \langle q_1, \mathsf{q}_0 \rangle & \text{if } p_1,\, q_1 \text{ in } N_1 \\
p_1\, \langle q_1, \mathsf{q}_0 \rangle & \to & \langle p_1, \mathsf{q}_0 \rangle\, q_1 & \text{if } p_1,\, q_1 \text{ in } N_1 \\
\langle p_1, p_2 \rangle\, q_1 & \to & \langle p_1, q_2 \rangle \langle q_1, r_2 \rangle & \text{if } p_1,\, q_1 \text{ in } N_1,\ p_2 \to q_2\, r_2 \text{ in } \Delta_2 \\
\langle p_1, p_2 \rangle \langle q_1, q_2 \rangle & \to & \langle p_1, r_2 \rangle \langle q_1, s_2 \rangle & \text{if } p_1,\, q_1 \text{ in } N_1,\ p_2\, q_2 \to r_2\, s_2 \text{ in } \Delta_2 \\
\langle p_1, p_2 \rangle & \to & \langle p_1, q_2 \rangle & \text{if } p_1 \text{ in } N_1,\ p_2 \to q_2 \text{ in } \Delta_2 \\
\langle p_1, p_2 \rangle & \to & a & \text{if } p_1 \to a \text{ in } \Delta_1,\ p_2 \to a \text{ in } \Delta_2 \\
\langle \mathsf{p}_0, \mathsf{q}_0 \rangle & \to & \varepsilon & \text{if } \mathsf{p}_0 \to \varepsilon \text{ in } \Delta_1,\ \mathsf{q}_0 \to \varepsilon \text{ in } \Delta_2
\end{array}
$$

One can show that for each $p_1, p_2, \ldots, p_n$ in N_1 and $q_1, q_2, \ldots, q_n$ in N_2,

$\mathsf{p}_0 \to^*_{\mathcal{G}_1} p_1\, p_2 \cdots p_n$ iff $\langle \mathsf{p}_0, \mathsf{q}_0 \rangle \to^*_{\mathcal{G}} \langle p_1, \mathsf{q}_0 \rangle\, p_2 \cdots p_n$

$\mathsf{q}_0 \to^*_{\mathcal{G}_2} q_1\, q_2 \cdots q_n$ iff $\langle p_1, \mathsf{q}_0 \rangle\, p_2 \cdots p_n \to^*_{\mathcal{G}} \langle p_1, q_1 \rangle \langle p_2, q_2 \rangle \cdots \langle p_n, q_n \rangle$

Corollary

For the class of $\mathsf{C}(\mathsf{CS}_T)$, question if $\mathcal{L}(\mathcal{G}) = \varnothing$ is undecidable

(Cf. The intersection-emptiness problem for $\mathsf{C}(\mathsf{CF}_T)$ on page 12)

Exercise

1. Show that for the class of context-sensitive grammar, membership problem is decidable.
2. Show Claim 2 .
3. Chomsky normal form if production rules are in the following forms:

   $$p \to q\,r \qquad p \to a \qquad \mathsf{q}_0 \to \varepsilon \qquad (p \in N,\ q, r \in N - \{\mathsf{q}_0\},\ a \in T)$$

   Show that a language is generated by a grammar in Chomsky normal form if and only if it is generated by a context-free grammar.
4. A grammar $\mathcal{G} = (\Sigma, T, N, \mathsf{q}_0, \Delta)$ is called Kuroda normal form if production rules are in the following forms:

   $$p\,q \to r\,s \qquad p \to q\,r \qquad p \to q \qquad p \to a \qquad (p, q, r, s \in N,\ a \in T)$$

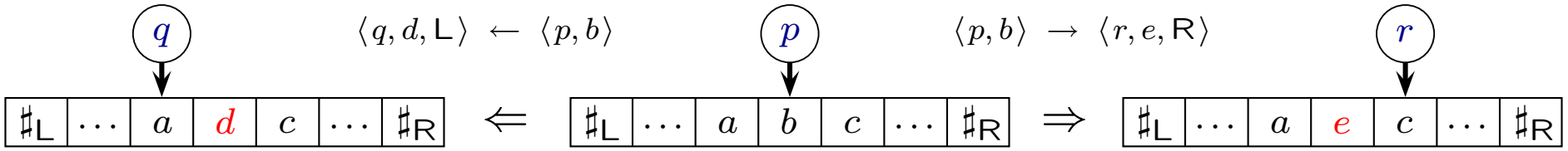
   Show that languages generated by grammar in Kuroda normal form are context-sensitive, and conversely, context-sensitive languages which does not include the empty word are generated by grammar in Kuroda normal form.

Exercise (cont'd)

5. [Pumping lemma for regular grammar] Given regular language L, there exists $k \geqslant 0$ such that if $z \in L$ and $|z| \geqslant k$, then z is formed by u, v, w with $|v| \geqslant 1$ and $u\,v^n\,w \in L \quad (n \geqslant 0)$. Verify this statement.
6. Show that the language $L = \{\, \mathrm{a}^\ell\, \mathrm{b}^\ell \mid \ell \geqslant 0 \,\}$ is not generated by regular grammar.
7. Show that context-free languages over a singleton set of a symbol are generated by regular grammar.
8. For each of languages $L_1 = \{\, \mathrm{a}^\ell\, \mathrm{b}^\ell\, \mathrm{a}^m \mid \ell, m \geqslant 1 \,\}$ and $L_2 = \{\, \mathrm{a}^m\, \mathrm{b}^\ell\, \mathrm{a}^\ell \mid \ell, m \geqslant 1 \,\}$, construct context-free grammar, respectively.
9. [Liu & Weiner 1973] For each $n \geqslant 0$, let $\mathsf{C}(\mathsf{CF}_T^n)$ be the class of languages over T defined as follows: $\mathsf{C}(\mathsf{CF}_T^0)$ is the set of context-free languages whose set of terminal symbols is T; $\mathsf{C}(\mathsf{CF}_T^{n+1})$ is $\{\, L_1 \cap L_2 \mid L_1 \in \mathsf{C}(\mathsf{CF}_T^0),\, L_2 \in \mathsf{C}(\mathsf{CF}_T^n) \,\}$. Show that for all $i \geqslant 0$, $\mathsf{C}(\mathsf{CF}_T^i) \subsetneqq \mathsf{C}(\mathsf{CF}_T^{i+1})$.
10. Show that $\mathsf{C}(\mathsf{CF}_T^n) \subsetneqq \mathsf{C}(\mathsf{CS}_T)$ for all $n \geqslant 0$.

Appendix : Linear-bounded automata (LBA)

LBA is a non-deterministic Turing machine with rewritable length-bounded single-tape as input : $(\Sigma \cup \{\sharp_{\mathsf{L}}, \sharp_{\mathsf{R}}\}, Q, \mathsf{q}_0, Q_{\mathsf{fin}}, \Delta)$, where $\sharp_{\mathsf{L}}, \sharp_{\mathsf{R}}$ are special tape symbols meaning left- right-ends. Transition rules in Δ are in the form of $\langle p, a\rangle \to \langle q, b, x\rangle \quad (p, q \in Q,\ a, b \in \Sigma,\ x \in \{\mathsf{L}, \mathsf{R}\})$.



The class of languages accepted by linear-bounded automata, denoted by NSPACE(n), and C(CS$_T$) coincide [1]. From space complexity observation, it can be shown that NSPACE(n) = co-NSPACE(n) [2]. This implies that for all $L \in$ C(CS$_T$), $(L)^{\mathsf{c}} \in$ C(CS$_T$), stating that C(CS$_T$) is closed under complement.

---

[1] S.-Y. Kuroda: *Class of Languages and Linear-Bounded Automata*, Information and Control 7, pp.207–223, 1964

[2] N. Immerman: *Nondeterministic Space is Closed Under Complementation*, SIAM Journal of Computing 17, pp.935–938, 1988

National Institute of Advanced Industrial Science and Technology (AIST) – Senri-site, AIST Kansai.

Office: Shin-Senri Nishi 1–2–14 (MSK bldg. 5th floor), Toyonaka, Osaka 560–0083, Japan

URL: `http://staff.aist.go.jp/hitoshi.ohsaki/`