Undecidable CFG Problems

Theorem 9.20: The problem of deciding whether a given CFG is ambiguous is undecidable.

Proof sketch: We will reduce PCP to this problem.

Aribtrary PCP Instance: $w_1, \ldots, w_k$, $x_1, \ldots, x_k$ over Σ.

Let $a_1, \ldots, a_k$ be symbols not in Σ.

Let G be the CFG given by:

$$S \rightarrow A \mid B$$
$$A \rightarrow w_i A a_i \mid w_i a_i \quad \text{for } 1 \leq i \leq k$$
$$B \rightarrow x_i B a_i \mid x_i a_i \quad \text{for } 1 \leq i \leq k$$

where S, A, and B are distinct symbols not in $\Sigma \cup \{a_1, \ldots, a_k\}$.

If the instance of PCP has a solution, then G is clearly ambiguous.

Suppose G is ambiguous, and let
$z \in (\Sigma \cup \{a_1, \ldots, a_k\})^*$,
$\alpha, \beta, \gamma_1, \gamma_2 \in (\Sigma \cup \{a_1, \ldots, a_k, S, A, B\})^*$, and
$C \in \{S, A, B\}$ such that

- $S \stackrel{*}{\Rightarrow} \alpha C \beta \Rightarrow \alpha \gamma_1 \beta \stackrel{*}{\Rightarrow} z$;
- $S \stackrel{*}{\Rightarrow} \alpha C \beta \Rightarrow \alpha \gamma_2 \beta \stackrel{*}{\Rightarrow} z$; and
- $\gamma_1 \neq \gamma_2$.

It is easily seen that there is at most one derivation $A \Rightarrow \cdots \Rightarrow z$ and at most one derivation $B \Rightarrow \cdots \Rightarrow z$; hence $\alpha C \beta = S$.

The string obtained by removing all a_is from z is a solution to the instance of PCP.

The construction is clearly computable.

Therefore, the ambiguity problem is undecidable.

Lemma: Let $w_1, \ldots, w_k$ be strings over Σ, and let $a_1, \ldots, a_k$ and S be distinct symbols not in Σ. Let $G = (\{S\}, \Sigma \cup \{a_1, \ldots, a_k\}, P, S)$ be the CFG such that P is given by

$$S \to w_i S a_i \mid w_i a_i \quad \text{for } 1 \leq i \leq k.$$

Then we can effectively construct a CFG G' such that $L(G') = \overline{L(G)}$.

Proof sketch:

- We can easily construct a PDA M such that $L(M) = \overline{L(G)}^R$.
- We can construct a CFG G'' such that $L(G'') = L(M) = \overline{L(G)}^R$.
- By reversing the right-hand-sides of all productions in G'', we obtain a CFG G' such that $L(G') = \overline{L(G)}$.

Theorem 9.22: The following problems are undecidable, where G_1 and G_2 denote given CFGs and R denotes a given regular expression:

- Is $L(G_1) \cap L(G_2) \neq \varnothing$?
- Is $L(G_1) \neq L(G_2)$?
- Is $L(G_1) \neq L(R)$?
- Is $L(G_1) \neq T^*$ for every alphabet T?
- Is $L(G_2) - L(G_1) \neq \varnothing$?
- Is $L(R) - L(G_1) \neq \varnothing$?

We will reduce PCP to each of these problems.

Aribtrary PCP Instance: $w_1, \ldots, w_k$, $x_1, \ldots, x_k$ over Σ.

Let $I = \{a_1, \ldots, a_k\}$ be an alphabet such that $\Sigma \cap I = \varnothing$.

Define

G_A**:** $S \rightarrow w_i S a_i \mid w_i a_i$ for $1 \leq i \leq k$.

G_B**:** $S \rightarrow x_i S a_i \mid x_i a_i$ for $1 \leq i \leq k$.

Then $L(G_A) \cap L(G_B) \neq \varnothing$ iff the PCP instance has a solution.

We can construct CFGs G'_A and G'_B such that $L(G'_A) = \overline{L(G_A)}$ and $L(G'_B) = \overline{L(G_B)}$.

Furthermore, we can construct the following:

- a CFG G_1 such that $L(G_1) = L(G'_A) \cup L(G'_B) = \overline{L(G_A) \cap L(G_B)}$;
- a CFG G_2 such that $L(G_2) = (\Sigma \cup I)^*$; and
- a regular expression R such that $L(R) = (\Sigma \cup I)^*$.

Each of the following holds iff the PCP instance has a solution:

- $L(G_1) \neq L(G_2)$
- $L(G_1) \neq L(R)$
- $L(G_1) \neq T^*$ for every alphabet T
- $L(G_2) - L(G_1) \neq \varnothing$
- $L(R) - L(G_1) \neq \varnothing$