25. Undecidable Problems about Context-Free Languages

Goal: Prove a number of undecidability results for context-free languages.
Note that they carry over to more expressive language classes.

Approach: Reduction from PCP.

Remark: Another important concept is the encoding of the invalid computations of a TM as a context-free language.
→ How to capture the valid computations has been shown in the reduction of HP to PCP.

Theorem:

Given two context-free grammars G_1 and G_2, the following problems are undecidable:

(1) Is $L(G_1) \cap L(G_2) \neq \emptyset$

(2) Is $|L(G_1) \cap L(G_2)| = \infty$?

(3) Is $L(G_1) \cap L(G_2)$ context-free?

(4) Is $L(G_1) \subseteq L(G_2)$?

(5) Is $L(G_1) = L(G_2)$?

Proof:

(1) Consider the PCP instance

$$K = (x_1, y_1) \dots (x_k, y_k) \quad \text{over } \{0,1\}.$$

We construct grammars over the alphabet $\{0, 1, \$, \underbrace{a_1, \dots, a_k}_{\text{one letter per pair.}}\}$

We define G_1 to have the productions

$S \to A \$ B$

$A \to a_1 A x_1 \mid \dots \mid a_k A x_k \qquad B \to y_1^{rev} B a_1 \mid \dots \mid y_k^{rev} B a_k$

$A \to a_1 x_1 \mid \dots \mid a_k x_k \qquad B \to y_1^{rev} a_1 \mid \dots \mid y_k^{rev} a_k$.

Grammar G_1 generates the language

$$L_1 = \{ a_{i_n} \dots a_{i_1} x_{i_1} \dots x_{i_n} \$ y_{j_m}^{rev} \dots y_{j_1}^{rev} a_{j_1} \dots a_{j_m} \mid n, m \geq 1,\ j_*, i_* \in \{1, \dots, k\}\}.$$

Grammar G_2 has the rules

$$S \to a_1 S a_1 \mid \dots \mid a_k S a_k \mid T$$

$$T \to 0T0 \mid 1T1 \mid \$.$$

Grammar G_2 generates

$$L_2 = \{ uv \$ v^{rev} u^{rev} \mid v \in \{0,1\}^*,\ u \in \{a_1, \dots, a_k\}^* \}.$$

We now have

K has a non-empty solution $i_1 \dots i_n$

iff $L_1 \cap L_2 \neq \emptyset$, namely it contains

$$a_{i_n} \dots a_{i_1} x_{i_1} \dots x_{i_n} \$ y_{i_n}^{rev} \dots y_{i_1}^{rev} a_{i_1} \dots a_{i_n}.$$

Hence, f with $f(K) := (G_1, G_2)$
is a reduction from PCP to the intersection non-emptiness problem.

(2) If PCP has a solution, then it has infinitely many solutions, namely by repeating the index sequence indefinitely.
Hence, the above f is even a reduction to infinity of the intersection.

(3) We argue that the above f is even a reduction to the problem of <u>being</u> <u>not context-free</u>:

Is $L(G_1) \cap L(G_2)$ <u>not</u> context-free.

If the problem of being not context-free is undecidable, the problem of being context-free has to be undecidable as well.
Why? Decidable languages are closed under complement!

If $L_1 \cap L_2 = \emptyset$, the language is of course context-free.
If $L_1 \cap L_2 \neq \emptyset$, the language is not context-free.
To see this, apply the pumping lemma.
We would have to pump all four parts of a word to stay in the language.
For long enough words, the pumping lemma only allows us to pump (at most) two parts.

(4) We reduce intersection emptiness (which also has to be undecidable due to complementation) to inclusion.
The point is to note that the above languages are <u>deterministic</u> context-free.
Hence, by Section 12 we can compute grammars $\overline{G_1}$ and $\overline{G_2}$ with

$$L(\overline{G_1}) = \overline{L(G_1)} = \overline{L_1} \quad \text{and} \quad L(\overline{G_2}) = \overline{L(G_2)} = \overline{L_2}.$$

We now have

$$L_1 \cap L_2 = \emptyset \quad \text{iff} \quad L(G_1) \subseteq L(\overline{G_2}).$$

(5) Note that

$$L(G_1) \subseteq L(G_2) \quad \text{iff} \quad \underbrace{L(G_1) \cup L(G_2)}_{L(G_3)} = L(G_2).$$

Grammar G_3 can be computed from G_1 and G_2, because the context-free languages are effectively closed under union.
Hence, we have a reduction from inclusion (undecidable by (4)) to equivalence. $\square$

Indeed, since G_1 and G_2 are deterministic, we get the following corollary:

Corollary:

Problems (1) to (4) are undecidable even for deterministic context-free languages.

Note that G_3 used in the reduction for (5) is no longer deterministic.

Indeed, the following is a big decidability result.

Theorem (Sénizergues, Gödel award 2002):

$L_1 = L_2$ is decidable for deterministic context-free languages.

Theorem:

Given a CFG G, the following are undecidable:

(1) Is G ambiguous?

(2) Is $\overline{L(G)}$ context-free?

(3) Is $L(G)$ regular?

(4) Is $L(G)$ deterministic context-free?

(5) Is $L(G) = \Sigma^*$?

Proof:

Consider G_1 and G_2 from the above PCP reduction.

(1) Let G_3 be such that

$$L(G_3) = L(G_1) \cup L(G_2).$$

Then PCP has a solution iff G_3 is ambiguous, i.e. some word in $L(G_3)$ has two (different) parse trees.

(2) Consider G_4 with

$$L(G_4) = L(\overline{G_1}) \cup L(\overline{G_2}).$$

It can again be computed with the effective closure properties.

Now PCP instance K has a solution

$$\text{iff} \quad L(G_1) \cap L(G_2) = \overline{\overline{L(G_1)} \cup \overline{L(G_2)}} \qquad (*)$$
$$= \overline{L(\overline{G_1}) \cup L(\overline{G_2})}$$
$$= \overline{L(G_4)} \quad \text{is not context-free.}$$

Again, since being not context-free is undecidable, being context-free has to be undecidable.

(3)-(5) Note that $L(G_1) \cap L(G_2) = \emptyset$ iff $L(G_4) = \Sigma^*$.
Since Σ^* is regular (and hence deterministic context-free), and since both language classes are closed under complement, (3) and (4) follow with the above Equivalence (*). $\square$

Corollary:
(1) Given a context-free language L and a regular language R, the problem $L = R$ is undecidable.
(2) Given a context-sensitive language L, $L \neq \emptyset$ and $|L| = \infty$ are undecidable.

Proof:
(1) Choose $R = \Sigma^*$.
(2) Choose $L = L(G_1) \cap L(G_2)$, which we can do because the context-sensitive languages are effectively closed under intersection. $\square$