Inleiding Fundamentele Informatica
Antwoorden op geselecteerde opgaven uit Hoofdstuk 8

John Martin: Introduction to Languages and the Theory of Computation

Jetty Kleijn

Najaar 2005

8.1 Show using the pumping lemma lemma, that the given languages are not context-free.

a. $L = \{a^i b^j c^k \mid 0 \leq i < j < k\}$.

Suppose L is a CFL. Then L satisfies the pumping lemma (Theorem 8.1a). Let n be the constant of that lemma. Consider $u = a^n b^{n+1} c^{n+2} \in L$. Then $|u| \geq n$ and thus there exist v, w, x, y, z such that $u = vwxyz$ with $|wy| > 0$, $|wxy| \leq n$, and $vw^i xy^i z \in L$ for every $i \geq 0$. We distinguish two cases:

1. wy contains at least one a. Then, since $|wxy| \leq n$, there are no c's in wy. Consequently, vw^2xy^2z contains at least $n+1$ a's and exactly $n+2$ c's, which implies that vw^2xy^2z is not in L. A contradiction.

2. wy does not contain any a. Then it must contain a b or a c. In this case vw^0xy^0z contains n a's and either at most n b's or at most $n+1$ c's. Thus $vw^0xy^0z \notin L$, again a contradiction.

Since we get in all (both) cases a contradiction we conclude that the pumping lemma is not satisfied and hence L is not context-free.

b. $L = \{x \in \{a,b\}^* \mid n_b(x) = n_a(x)^2\}$.

Examples of words in L are: *aabbbb*, *babbba*, *abbbabbbabbb*.

Suppose L is a CFL. Then L satisfies the pumping lemma (Theorem 8.1a). Let n be the constant of that lemma. Consider $u = a^n b^{n^2} \in L$. Then $|u| \geq n$ and thus there exist v, w, x, y, z such that $u = vwxyz$ with $|wy| > 0$, $|wxy| \leq n$, and $vw^i xy^i z \in L$ for every $i \geq 0$.

Let $n_a(wy) = p$ and $n_b(wy) = q$. Then for each $i \geq 0$ we have $n_a(vw^ixy^iz) = n_a(u)+(i-1)p = n+(i-1)p$ and $n_b(vw^ixy^iz) = n_b(u)+(i-1)q = n^2+(i-1)q$. Since, by our assumption $vw^ixy^iz \in L$ for every $i \geq 0$, it must be the case that $(n+(i-1)p)^2 = n^2+(i-1)q$ for every $i \geq 0$. This however is impossible

as can be seen as follows. Since $|wy| > 0$, at least one of p and q is not 0.
If $p = 0$ and $q \neq 0$, then $n^2 = n^2 + (i-1)q$ for every $i \geq 0$, which clearly is not true if $i \geq 2$.
If $p \neq 0$ and $q = 0$, then $(n + (i-1)p)^2 = n^2$ for every $i \geq 0$, which clearly is not true if $i \geq 2$.
If $p \neq 0$ and $q \neq 0$, then we have (for $i = 2$) that $(n+p)^2 = n^2 + q$ which implies that $q = 2np + p^2$, and (for $i = 3$) that $(n+2p)^2 = n^2 + 2q$ which implies that $q = 2np + 2p^2$. Hence $p = 0$ should hold, a contradiction.
We conclude that the pumping lemma is not satisfied and that, consequently, L is not context-free.
e. $L = \{a^n b^m a^n b^{n+m} \mid m, n \geq 0\}$.
Suppose L is a CFL. Then L satisfies the pumping lemma (Theorem 8.1a). Let n be the constant of that lemma. Consider $u = a^n b^{2n} a^n b^{3n}$. Thus $u \in L$ and $|u| \geq n$. Hence there exist v, w, x, y, z such that $u = vwxyz$ with $|wy| > 0$, $|wxy| \leq n$, and $vw^i xy^i z \in L$ for every $i \geq 0$.
Since $|wxy| \leq n$ it is immediately clear that wxy cannot contain more than two distinct symbols.
Suppose first that wxy consists only of a's or only of b's. If wxy falls within the first group of a's, then $vw^2xy^2z = a^{n+k} b^{2n} a^n b^{3n}$ for some $k \geq 1$, and this word is not from L. We can use the same argument when wxy falls within the first group of b's, the second group of a's or the the second group of b's.
Secondly, if wxy is a subword of $a^n b^{2n}$, then $vw^2xy^2z = a^k s b^l a^n b^{3n}$ with $n_a(s) > n - k$ or $n_b(s) > 2n - l$, which is not in L. We can use the same argument when wxy is a subword of $b^{2n} a^n$ or of $a^n b^{3n}$.
Thus all cases lead to a contradiction. So L is not a CFL.

8.5 Is the given language context-free? Prove your answer.
a. $L = \{a^n b^m a^m b^n \mid n, m \geq 0\}$ is a CFL; give a grammar.
b. $L = \{xayb \mid x, y \in \{a, b\}^* \text{ and } |x| = |y|\}$ is a CFL; give a grammar.
c. $L = \{xcx \mid x \in \{a, b\}^*\}$ is not a CFL; proof similar as in Example 8.2.
d. $L = \{xyx \mid x, y \in \{a, b\}^* \text{ and } |x| \geq 1\}$ is not a CFL;
Assume that L is context-free. Then it satisfies the pumping lemma. Let n be the constant of that lemma. Now consider $u = xyx$ with $x = a^n b^n$ and $y = \Lambda$. Thus $u = a^n b^n a^n b^n$. There must exist words p, q, r, s, t such that $u = pqrst$ such that $|qs| > 0$, $|qrs| \leq n$, and $pq^i rs^i t \in L$ for every $i \geq 0$. We distinguish two cases:
1. qs consists only of a's from the first group of a's in u or it consists only of b's from the second group of b's. Then pq^2rs^2t is either $a^{n+j} b^n a^n b^n$ or $a^n b^n a^n b^{n+j}$ for some $j \geq 1$, which are both not in L. A contradiction.

2. qs contains a b from the first group of b's in u or it contains an a from the second group of a's in u. Then pq^0rs^0t is either $a^kb^la^mb^n$ or $a^nb^ka^lb^m$ with $k, m \geq 1$ and $l < n$. Neither of these words is in L, again a contradiction.
Consequently, we always end up with a contradiction and so L is not a CFL.

The exercises **e**, **f**, and **g** can be solved efficiently using the material from Chapter 7 (see below). When (deterministic) pushdown automata have not yet been considered, one should first see that each of the given languages is context-free and next try to find a CFG for that language!
e. $L = \{x \in \{a,b\}^* \mid n_a(x) < n_b(x) < 2n_a(x)\}$ is a CFL; see Exercise 7.37 b where a PDA is to be given for this language.
f. $L = \{x \in \{a,b\}^* \mid n_a(x) = 10n_b(x)\}$ is a CFL; give a PDA.
g. L is the set of non-balanced strings of parentheses (and). This language is a CFL, which can be proved by giving a DPDA for its complement the language over $\{(,)\}^*$ consisting of all balanced strings. Note that the family of deterministic context-free languages is closed under complementation.

8.6 Generalizing Theorem 5.3 to context-free languages yields:
If L is an infinite context-free language, then there are strings v, w, x, y, and z such that $|wy| > 0$ and $vw^ixy^iz \in L$, for every $i \geq 0$.
Generalizing Theorem 5.4 to context-free languages yields:
If L is an infinite context-free language, then the set of lengthhs of words in L contains an infinite arithmetic progression.
Similar to the way that Theorems 5.3 and 5.4 follow from the pumping lemma for regular languages, the two statements above follow from (are weaker forms of) the pumping lemma for context-free languages.
a. Let $L = \{x \in \{a,b,c\}^* \mid n_a(x) = n_b(x) = n_c(x)\}$. Then using the pumping lemma L can be shown to be not context-free (see Example 8.1). The above generalization of Theorem 5.3 cannot be successfully applied: there is no way to guarantee that the string wy doesn't have an equal number of a's, b's and c's (which allows pumping without leaving the language).
b. Let $L = \{a^ib^ic^i \mid i \geq 0\}$. Then the generalization of Theorem 5.3 can be used to prove that this language is not a CFL, but the generalization of Theorem 5.4 cannot (the length set does contain an infinite arithmetic progression).
c. Let $L = \{a^{n^2} \mid n \geq 0\}$. This language is not a CFL, which can be proved using the generalization of Theorem 5.4 (see also exercise 5.27 h).

8.7 The PDA M constructed in the proof of Theorem 8.4 is a DPDA for $L_1 \cap L_2$ whenever the PDA M_1 accepting L_1 is deterministic (a DPDA).

Thus, it follows that the family of DCFLs is also closed under intersection with regular languages.

8.8 Show that the given languages are CFLs, but their complements not.
a. $L = \{a^i b^j c^k \mid i \geq j \vee i \geq k\}$.
We have $L = L_1 \cup L_2$ with $L_1 = \{a^i b^j c^k \mid i \geq j\}$ and $L_2 = \{a^i b^j c^k \mid i \geq k\}$. It is easy to see that these languages are both context-free.
L_1 is generated by the CFG G_1 with axiom S_1 and set of productions P_1 consisting of $S_1 \to AC, \quad A \to aAb \,|\, aA \,|\, \Lambda, \quad C \to cC\Lambda$;
and L_2 is generated by the CFG G_2 with axiom S_2 and set of productions P_2 consisting of $S_2 \to aS_2c \,|\, aS_2 \,|\, B, \quad B \to bB \,|\, \Lambda$.
Then L is generated by the grammar with axiom S and set of productions $\{S \to S_1,\ S \to S_2\} \cup P_1 \cup P_2$.
The complement of L is $K = K_1 \cup K_2$ with $K_1 = \{a,b,c\}^*\{ba, ca, cb\}\{a,b,c\}^*$ consisting of all words over $\{a,b,c\}$ in which the order a's before b's before c's is not respected and $K_2 = \{a^i b^j c^k \mid i < j \wedge i < k\}$ consisting of all words over $\{a,b,c\}$ in which the symbols appear in the right order but in wrong numbers. Now assume that K is a CFL. Then by Theorem 8.4, also $K_2 = K \cap \{a\}^*\{b\}^*\{c\}^*$ is a CFL. Hence it satisfies the pumping lemma. Let n be the constant of that lemma. Consider the word $u = a^n b^{n+1} c^{n+1}$. Since $u \in K_2$ and $|u| \geq n$, there must exist words v, w, x, y, and z such that $u = vwxyz$ with $|wy| > 0$, $|wxy| \leq n$, and $vw^m xy^m z \in K$ for all $m \geq 0$.
If wy contains at least one a, it cannot contain a c, because $|wxy| < n+1$. Consequently vw^2xy^2z has at least as many a's as c's and is not in K_2. Thus wy consists solely of b's and c's. But this implies that in vw^0xy^0z the number of b's or the number of c's is at most n and so $vw^0xy^0z \notin K$, a contradiction. We conclude that K_2 doesn't satisfy the pumping lemma and consequently, is not context-free. Moreover, our assumption that K is context-free was wrong.
(Try to apply the pumping lemma directly to K instead of K_2.)
b. Similar to a. $L = \{a^i b^j c^k \mid i \neq j \vee i \neq k\} = \{a^i b^j c^k \mid i \neq j\} \cup \{a^i b^j c^k \mid i \neq k\}$ is a union of two (easy) CFLs. The complement of L is $K = K_1 \cup K_2$ with $K_1 = \{a,b,c\}^*\{ba, ca, cb\}\{a,b,c\}^*$ consisting of all words over $\{a,b,c\}$ in which the order a's before b's before c's is not respected and $K_2 = \{a^i b^j c^k \mid i = j = k\}$ consisting of all words over $\{a,b,c\}$ in which the symbols appear in the right order but in wrong numbers. Now assume that K is a CFL. Then by Theorem 8.4, also $K_2 = K \cap \{a\}^*\{b\}^*\{c\}^*$ is a CFL. In Example 8.1 however it has been shown (using the pumping lemma) that K_2 is not a CFL. Thus the assumption that K is a CFL is wrong.
c. $L = \{x \in \{a,b\}^* \mid \nexists w.\, x = ww\}$. Its complement is not context-free as has

been shown in Example 8.2. Hence we are left with the task of proving that L is a CFL. Clearly a word x is not of the form if its length is odd. A word of even length is not of the form ww if and only if it "makes a mistake", which means that it is of the form $uavwbz$ or $ubvwaz$ with $|u| = |w|$ and $|v| = |z|$. Thus we have $L = (\{a,b\}\{a,b\})^*\{a,b\} \cup \{uaybz, ubyaz \mid |y| = |u| + |z|\}$. Now it is easy to give a grammar for L:

$S \to S_1 \,|\, S_2$,

$S_1 \to aaS_1 \,|\, abS_1 \,|\, baS_1 \,|\, bbS_1 \,|\, a \,|\, b$, for words of odd length

$S_2 \to AB \,|\, BA$,

$A \to aAa \,|\, aAb \,|\, bAa \,|\, bAb \,|\, a$ for words uaw with $|u| = |w|$,

$B \to aBa \,|\, aBb \,|\, bBa \,|\, bBb \,|\, b$ for words vbz with $|v| = |z|$.

8.9 Prove, by using Ogden's lemma, that the following languages are not context-free.

a. $L = \{a^i b^{i+k} a^k \mid i, k \geq 0 \text{ and } i \neq k\}$

Assume that L is a CFL. Then it satisfies Ogden's lemma (Theorem 8.2). Let $n \geq 1$ be the integer in that lemma. We now have to find a word $u \in L$ of length at least n that when pumped using the distinguished postions we have chosen will (always) yield a word not in L, thus proving that Ogden's lemma does not hold for L. A contradiction, which implies that L is not a CFL.

Let $u = a^n b^{2n+n!} a^{n+n!}$ which is in L and certainly longer than n. We designate the first n positions in u as distinguished. According to Ogden, there exist v, w, x, y, z such that $u = vwxyz$, both the string wy and the string x contain at least one a from the first group of a's, and for all $m \geq 0$, the word $vw^m xy^m z$ is in L.

Since x contains at least one distinguished position, we have that w consists only of a's from the first group. If y would contain both an a and a b, then $y = y_1 a y_2 b y_3$ or $y = y_1 b y_2 a y_3$ and then $y^2 = y_1 a y_2 b y_3 y_1 a y_2 b y_3$ or $y^2 = y_1 b y_2 a y_3 y_1 b y_2 a y_3$. Then $vw^2 xy^2 z \notin L$ because it has two b's separated by at least an a. Hence y consists only of a's or only of b's, that is $y \in \{a\}^* \cup \{b\}^*$. If $y \in \{a\}^*$, then pumping w and y would involve only a's implying that the relationship between the number of a's and the number of b's would be lost. Thus $y \in \{b\}^*$ and we now know that it must be the case that $w = a^p$ and $y = b^q$ for some integers $p \geq 1$ and $q \geq 0$. If $p \neq q$, then $vw^0 xy^0 z = a^{n-p} b^{2n+n!-q} a^{n+n!} \notin L$. Consequently, $p = q \geq 1$.

Now consider $m = 1 + n!/p$. Thus m is an integer and $mp = p + n!$. Then $vw^m xy^m z = a^{n-p} a^{mp} b^{mp} b^{2n+n!-p} a^{n+n!} = a^{n+n!} b^{2n+2(n!)} a^{n+n!}$ which is not in L. We conclude that there is no possibility to pump u correctly, and L is not a CFL.

b. $L = \{a^i b^i a^j b^j \mid i, j \geq 0 \text{ and } i \neq j\}$
Assume that L is a CFL. Then it satisfies Ogden's lemma (Theorem 8.2). Let $n \geq 1$ be the integer in that lemma. Let $u = a^n b^n a^{n+n!} b^{n+n!}$ which is in L and is certainly longer than n. We designate the first n positions in u as distinguished. According to Ogden, there exist v, w, x, y, z such that $u = vwxyz$, both the string wy and the string x contain at least one a from the first group of a's, and for all $m \geq 0$, the word $vw^m xy^m z$ is in L.
Since x contains at least one distinguished position, we have that w consists only of a's from the first group. If y would contain both an a and a b, then $y = y_1 a y_2 b y_3$ or $y = y_1 b y_2 a y_3$ and then $vw^3 xy^3 z \notin L$ because it has at least three groups of b's. Hence y consists only of a's or only of b's, that is $y \in \{a\}^* \cup \{b\}^*$.
If $y \in \{a\}^*$, then pumping w and y would involve only a's implying that the relationship between the number of a's and the number of b's would be lost. Thus $y \in \{b\}^*$ and if y would be part of the second group of b's, then pumping w and y would violate the relationships between the first a's and b's and between the second a's and b's. Thus y belongs to the first group of b's. We now know that $vwxy = a^n b^k$ for some $0 \leq k \leq n$ and $w = a^p$ and $y = b^q$ for some integers $p \geq 1$ and $q \geq 0$. If $p \neq q$, then $vw^0 xy^0 z = a^{n-p} b^{n-q} a^{n+n!} b^{n+n!} \notin L$. Consequently, $p = q \geq 1$.
Now consider $m = 1 + n!/p$. Thus m is an integer and $mp = p + n!$. Then $vw^m xy^m z = a^{n-p} a^{mp} b^{mp} b^{n-p} a^{n+n!} b^{n+n!} = a^{n+n!} b^{n+n!} a^{n+n!} b^{n+n!}$ which is not in L. We conclude that there is no possibility to pump u correctly, and L is not a CFL.
c. $L = \{a^i b^j a^i \mid i, j \geq 0 \text{ and } i \neq j\}$
Assume that L is a CFL. Then it satisfies Ogden's lemma (Theorem 8.2). Let $n \geq 1$ be the integer in that lemma. Let $u = a^n b^{n+n!} a^n$ which is in L and is certainly longer than n. We designate the first n positions in u as distinguished. According to Ogden, there exist v, w, x, y, z such that $u = vwxyz$, both the string wy and the string x contain at least one a from the first group of a's, and for all $m \geq 0$, the word $vw^m xy^m z$ is in L.
More or less as before, we can now argue that $w = a^p$ belongs to the first group of a's and $y = a^p$ belongs to the second group of a's. Next let $m = 1 + n!/p$. Thus m is an integer and $mp = p + n!$. Then $vw^m xy^m z = a^{n-p} a^{mp} b^{n+n!} a^{mp} a^{n-p} = a^{n+n!} b^{n+n!} a^{n+n!}$ which is not in L. We conclude that there is no possibility to pump u correctly, and L is not a CFL.

8.10
a. Let L be a CFL and F a finite language. Then $L - F$ is a CFL:
Let Σ be an alphabet such that $L, F \subseteq \Sigma^*$.

F is finite implies F is regular which implies that its complement $\Sigma^* - F$ is regular (Theorem 3.4, page 110). Using Theorem 8.4, we conclude that $L \cap (\Sigma^* - F) = L - F$ is a CFL.
b. L is not a CFL and F is a finite language. Then $L - F$ is not a CFL, which we prove by contradiction. Assume that $L - F$ is a CFL.
Note that $L \cap F \subseteq F$ is finite and hence a CFL. Since a union of two CFLs is a CFL (Theorem 6.1), it follows that $(L - F) \cup (L \cap F) = L$ is context-free, a contradiction. We conclude that $L - F$ is not a CFL.
c. L is not a CFL and F is a finite language. Then $L \cup F$ is not a CFL, which we prove by contradiction. Assume that $L \cup F$ is a CFL.
Note that $F - L \subseteq F$ is finite. It then follows from a that $(L \cup F) - (F - L) = L$ is context-free, a contradiction. We conclude that $L - F$ is not a CFL.

8.11 Consider once more exercise 8.10, now with every occurrence of "finite" replaced by "regular".
a. Let L be a CFL and F a regular language. Then $L - F$ is a CFL, which can be proved as in 8.10a.
b. L is not a CFL and F is a regular language. Then $L - F$ may or may not be a CFL. As seen above, if F is a finite language, then $L - F$ is not context-free. On the other hand, if we let $F = \Sigma^*$ where Σ is an alphabet such that $L \subseteq \Sigma^*$, then Then $L - F = \emptyset$ which is a CFL.
c. L is not a CFL and F is a regular language. Similar to b, $L \cup F$ may or may not be CFL. As seen above, if F is a finite language, then $L \cup F$ is not context-free. On the other hand, if we let $F = \Sigma^*$ where Σ is an alphabet such that $L \subseteq \Sigma^*$, then $L \cup F = F = \Sigma^*$ is a CFL.

8.12 Each of the three statements in exercise 8.10 is true when "CFL" is replaced by "DCFL":
a. as before, since the family of DCFLs is also closed under intersection (see exercise 8.7).
b. If $L - F$ is a DCFL, then the complement K of $L - F$ is also a DCFL and thus also $L = K \cap F$, because the family of DCFLs is also closed under intersection (see exercise 8.7). Hence, if L is not a DCFL, then also $L - F$ is not a DCFL.
c. as 8.10a, now using 8.12a.
