Einführung in die Computerlinguistik
Hausaufgabe (CFG 3), Abgabe 15.06.2020

Laura Kallmeyer

Sommer 2020, Heinrich-Heine-Universität Düsseldorf

Aufgabe 1

1. *Aus der Vorlesung wissen Sie, dass $L_1 = \{a^n b^n c^n \mid n \geq 1\}$ nicht kontextfrei ist (s. Folie 18).*

   Zeigen Sie unter Bezugnahme auf die Abgeschlossenheit von CFLs unter Homomorphismen, dass auch $L_2 = \{a^n b^n c^n d^n \mid n \geq 1\}$ nicht kontextfrei ist.

2. *Jetzt nehmen Sie an, dass von L_2 schon bewiesen wurde, dass diese Sprache nicht kontextfrei ist.*

   Zeigen Sie, unter Bezugnahme auf die Abgeschlossenheit von CFLs unter Schnittbildung mit regulären Sprachen, dass auch $L_3 = \{w \mid w \in \{a, b, c, d\}^+$ mit $|w|_a = |w|_b = |w|_c = |w|_d\}$ nicht kontextfrei ist.

Tipp: Die Beweisführung in allen drei Fällen ist ein Widerspruchsbeweis: Man nimmt das Gegenteil von dem an, was man beweisen möchte, und zeigt, dass dies zu einem Widerspruch führt.

Konkret: Man nimmt an, dass die jeweilige Sprache L_i kontextfrei ist. Dann muss auch ihr Bild unter einem geeigneten Homomorphismus (bzw. ihr Schnitt mit einer geeigneten regulären Sprache) kontextfrei sein. Hier muss man Homomorphismus bzw. reguläre Sprache so wählen, dass man bei einer Srache landet, von der man schon gezeigt hat, dass sie nicht kontextfrei ist. Dies ist ein Widerspruch zu dem, was aus der Annahme der Kontextfreiheit von L_i folgt, und somit ist diese Annahme falsch.

Lösung:

1. Annahme: L_2 ist kontextfrei. Dann ist auch $f(L_2)$ kontextfrei, wobei f ein Homomorphismus ist mit $f(a) = a$, $f(b) = b$, $f(c) = c$, $f(d) = \varepsilon$. $f(L_2) = L_1$. Wir wissen aber schon, dass L_1 nicht kontextfrei ist, es ergibt sich also ein Widerspruch. Damit kann L_2 auch nicht kontextfrei sein.

2. Annahme: L_3 ist kontextfrei. Dann ist auch der Schnitt von L_3 mit $L(a^*b^*c^*d^*)$, also mit der von $a^*b^*c^*d^*$ denotierten regulären Sprache, kontextfrei. $L_3 \cap L(a^*b^*c^*d^*) = L_2$. Wir wissen aber schon, dass L_2 nicht kontextfrei ist, es ergibt sich also ein Widerspruch. Damit kann L_3 auch nicht kontextfrei sein.

Aufgabe 2

Betrachten Sie die folgende Sprache: $\{(aaa)^n bcd(aa)^n \mid n \geq 0\}$. Es handelt sich um eine kontextfreie Sprache.

Damit gilt das Pumping Lemma von Folie 17 für diese Sprache.

1. *Wie ist das kleinstmöglichste k, so dass es ein Wort der Länge k gibt und das Pumping Lemma für die Konstante k und diese Sprache gilt?*

2. *Welches Wort der Länge k ist dann das kürzeste, in dem man zwei Teilstrings v_1 und v_2 findet, die wie im Pumping Lemma parallel iteriert werden können (bei dieser Sprache gibt es nur eines), und welche Teilstrings sind v_1 und v_2?*

Lösung:

1. $k = 8$

2. $w = aaabcdaa$ mit $v_1 = aaa$, $v_2 = aa$ und $w = v_1 bcd v_2$.

Aufgabe 3

Betrachten Sie die folgende Sprache: $L = \{a^n b a^m b a^n \mid n \geq m \geq 1\}$.

Beweisen Sie, dass diese Sprache keine kontextfreie Sprache ist indem Sie zeigen, dass sie das Pumping Lemma für kontextfreie Sprachen nicht erfüllt.

Tipp: Nehmen Sie an, dass L das Pumping Lemma mit einer Konstante k erfüllt. Dann müsste man in dem Wort $w = a^k b a^k b a^k$ *zwei Teilstrings (mindestens einer nicht leer) finden können, so dass diese im Sinne des Pumping Lemmas iterierbar sind. Zeigen Sie, dass dies nicht der Fall sein kann.*

(Denken Sie daran, dass auch eine Iteration mit $i = 0$*, also ein Weglassen der iterierbaren Teilstrings laut Pumping Lemma möglich sein sollte.)*

Lösung:

Annahme: L erfüllt das Pumping Lemma mit einer Konstante k. Wir betrachten $w = a^k b a^k b a^k$. Laut Pumping Lemma muss gelten, dass es Strings x, v_1, y, v_2, z gibt mit $w = x v_1 y v_2 z$ und

- $|v_1 v_2| \geq 1$,
- $|v_1 y v_2| \leq k$,
- für alle $i \geq 0$: $x v_1^i y v_2^i z \in L$.

Für v_1 und v_2 gilt, dass beide notwendig keine bs enthalten, gleichlang sind, und in dem ersten und dritten Teilstring von as liegen müssen. Andernfalls würde bei Iteration mit $i > 1$ die Bedingung verletzt werden, dass die Anzahl der a's am Anfang und am Ende gleich sein müssen, oder es würden zu viele b's entstehen, oder zu viel a's im mittleren Teil.

v_1 und v_2 dürfen nicht leer sein (Bedingung im Pumping Lemma).

Dann führt aber die Iteration mit $i = 0$, also das Weglassen von v_1 und v_2 zu einem Wort, bei dem die mittlere Folge von a's länger ist als die beiden am Anfang und am Ende, also einem Wort, das nicht in L ist.

Damit ergbt sich ein Widerspruch zu der Annahme, dass L das Pumping Lemma für die Konstante k erfüllt. L ist somit nicht kontextfrei.