Correct gebruik van het *pumping lemma* voor reguliere talen

Het pumping lemma is een zeer nuttige techniek om aan te tonen dat een bepaalde taal niet-regulier is, op voorwaarde dat het correct wordt toegepast! Herinner je de precieze formulering.

Als een taal $\mathcal{L}$ oneindig groot en regulier is, dan geldt de volgende eigenschap: er bestaat een constante k, zodat voor elk woord $w \in \mathcal{L}$ met $|w| \geqslant k$, een ontbinding $w = xyz$ bestaat met $|xy| \leqslant k$ en $|y| \neq 0$, zodat voor elke $q \in \mathbb{N}$, het woord $w_q = xy^qz$ opnieuw tot de taal $\mathcal{L}$ behoort.

Meer in formulevorm, als $\mathcal{L}$ oneindig groot en regulier is, dan geldt

$$\Big(\exists\, k \in \mathbb{N}\Big)\Big(\forall\, w \in \mathcal{L} \text{ met } |w| \geqslant k\Big)\Big(\exists\, x, y, z \in \Sigma^* \text{ met } w = xyz,\, |y| \neq 0,\, |xy| \leqslant k\Big)\Big(\forall\, q \in \mathbb{N}\Big)\Big(w_q = xy^qz \in \mathcal{L}\Big).$$

Let op de volgorde van de kwantoren $\exists$ (“er bestaat”) en $\forall$ (“voor alle”).

Het pumping lemma wordt gebruikt om te tonen dat een bepaalde oneindige taal $\mathcal{L}$ niet regulier kan zijn, volgens het volgende stappenplan. De achterliggende bedoeling is om aan te tonen dat de conclusies van het pumping lemma niet gelden voor de taal $\mathcal{L}$.

1. Leg uit waarom $\mathcal{L}$ oneindig groot is. Dit spreekt meestal voor zich.
2. Neem aan dat $\mathcal{L}$ tóch regulier is. Dankzij deze aanname kunnen we het pumping lemma toepassen. We proberen uiteindelijk een strijdigheid te vinden, zodat deze aanname niet kan kloppen.
3. Het pumping lemma beweert dat er een constante k bestaat met een bepaalde eigenschap. Hoeveel deze constante precies bedraagt, weten we niet. Om verder te redeneren zijn we dus genoodzaakt om k algemeen te laten.
4. Voor deze k beweert het pumping lemma dat elk woord $w \in \mathcal{L}$ met $|w| \geqslant k$, op- en neergepompt kan worden naar nieuwe woorden in $\mathcal{L}$. Aangezien we op zoek zijn naar een strijdigheid, volstaat het om één concreet woord w te geven dat zal afhangen van k en dat *niet* op- of neergepompt zal kunnen worden. Beschouw dus een goedgekozen woord w in de taal. Dit is de moeilijkste stap!
5. Het pumping lemma verder volgend moet er voor deze w een ontbinding $w = xyz$ met $|y| \neq 0$ en $|xy| \leqslant k$ bestaan met een bepaalde eigenschap. Opnieuw weten we niet wat x, y en z precies zijn; tracht wel om zoveel mogelijk informatie te halen omtrent x, y en z uit je concrete keuze van w.
6. Volgens het pumping lemma behoort elk woord $w_q = xy^qz$ opnieuw tot de taal $\mathcal{L}$ voor elke keuze van $q \in \mathbb{N}$. Aangezien we op zoek zijn naar een strijdigheid, volstaat het om één concreet getal q te kiezen waarvoor we zullen kunnen aantonen dat $w_q \notin \mathcal{L}$. Afhankelijk van de opgave kan het interessant zijn om neerwaarts te pompen (met $q = 0$) of opwaarts (met $q \geqslant 2$). Beschouw nu dus een goedgekozen q.
7. Hopelijk konden we in stap 5 voldoende informatie omtrent x, y en z afleiden om een beschrijving van $w_q = xy^qz$ op te stellen en te argumenteren waarom deze w_q *niet* tot $\mathcal{L}$ behoort. Is dat gelukt, dan hebben we onze gezochte strijdigheid gevonden! Anders moeten we een alternatief woord w of getal q kiezen.

Let op de subtiliteiten: we kunnen de waarde voor k in stap 3 **niet** zelf kiezen, en evenmin de ontbinding van w in stap 5. We moeten immers aantonen dat voor elke mogelijk k en elke mogelijk ontbinding er een strijdigheid optreedt.

Vaak blijkt het handig om op een taal niet rechtstreeks het pumping lemma toe te passen, maar om eerst één of meerdere sluitingseigenschappen van de reguliere talen te gebruiken, om deze taal eenvoudiger te maken of meer structuur te geven. Hoe strenger de structuur van een taal, hoe eenvoudiger het immers wordt om deze structuur kapot te maken door op- of neerpompen. Het stappenplan gaat dan als volgt.

1. Neem aan dat $\mathcal{L}$ tóch regulier is. We proberen een strijdigheid te vinden.
2. Gebruik een sluitingseigenschap van reguliere talen (bijvoorbeeld de eigenschap dat reguliere talen gesloten zijn onder doorsnedes, complementering, lettersubstituties, omkering…) om een nieuwe taal $\mathcal{L}'$ te bekomen. Aangezien we ervan uitgingen dat $\mathcal{L}$ regulier was, zal ook $\mathcal{L}'$ regulier blijven. Herhaal deze stap eventueel op $\mathcal{L}'$.
3. Pas het pumping lemma toe op de gereduceerde taal $\mathcal{L}'$ (zoals hierboven) om aan te tonen dat $\mathcal{L}'$ niet regulier is, wat de gezochte strijdigheid oplevert.

Let erop dat je de sluitingseigenschappen correct toepast en geen essentiële niet-reguliere eigenschappen van $\mathcal{L}$ weggooit. Indien je na een doorsnede of lettersubstitutie bijvoorbeeld een taal $\mathcal{L}'$ bekomt die wél regulier is, dan kun je niks besluiten over het al dan niet regulier zijn van de originele taal $\mathcal{L}$…

Het is niet altijd strikt noodzakelijk om de sluitingseigenschappen te gebruiken: soms kun je rechtstreeks het pumping lemma toepassen en een strijdigheid vinden. Vaak is het echter wel makkelijker om zoveel mogelijk structuur op je taal te introduceren waardoor het pumping lemma makkelijker wordt.

Een illustratief voorbeeld is de taal $\mathcal{L} = \{\mathtt{0}^a\,\mathtt{1}^b : a \neq b\}$. Deze is uiteraard niet-regulier. Het is mogelijk om rechtstreeks het pumping lemma te gebruiken, maar zeker niet voor de hand liggend: we starten dan immers met een woord $w \in \mathcal{L}$ en we moeten voor elke mogelijke ontbinding een woord $w_q \notin \mathcal{L}$ vinden. Met andere woorden, we starten met heel weinig structuur op w (een verschillend aantal $\mathtt{0}$ en $\mathtt{1}$) en we moeten een sterke structuur op een w_q kunnen forceren (essentieel, een gelijk aantal $\mathtt{0}$ en $\mathtt{1}$).

Oefening: gebruik **rechtstreeks** het pumping lemma om aan te tonen dat $\mathcal{L}$ niet-regulier is.

Met de sluitingseigenschappen wordt het een pak eenvoudiger: als $\mathcal{L}$ regulier zou zijn, dan ook de taal

$$\overline{\mathcal{L}} \cap \mathtt{0}^*\,\mathtt{1}^* = \{\mathtt{0}^a\,\mathtt{1}^b : a = b\},$$

maar dit is het klassiek voorbeeld van een niet-reguliere taal (op een substitutie na)—een strijdigheid.

Een nog extremer voorbeeld is de taal $\mathcal{L} = \{\mathtt{0}^a\,\mathtt{1}^b\,\mathtt{2}^c : a \neq 1 \text{ of } b = c\}$. Hier kunnen we zelfs aantonen dat het pumping lemma rechtstreeks toepassen nooit tot een strijdigheid zal leiden, omdat elk woord w in $\mathcal{L}$ (verschillend van het nulwoord) oppompbaar is! Er zijn immers vier mogelijkheden voor $w \in \mathcal{L}$.

- Als $a = 0$, dan zijn er geen restricties op b en c. Kies hier $x = \varepsilon$, y gelijk aan de eerste letter van w ($\mathtt{1}$ of $\mathtt{2}$) en z gelijk aan de rest. Na op- of neerpompen blijft $\#_{\mathtt{0}}(w_q) = 0$, zodat inderdaad $w_q \in \mathcal{L}$ (voor elke q).
- Als $a = 1$, kies dan $x = \varepsilon$, $y = \mathtt{0}$ en z gelijk aan de rest. Na op- of neerpompen blijkt $\#_{\mathtt{0}}(w_q) \neq 1$, zodat inderdaad $w_q \in \mathcal{L}$ (voor elke q).
- Als $a = 2$, kies dan $x = \varepsilon$, $y = \mathtt{00}$ en z gelijk aan de rest. Na op- of neerpompen blijkt $\#_{\mathtt{0}}(w_q)$ steeds even, zodat in het bijzonder $\#_{\mathtt{0}}(w_q) \neq 1$ en inderdaad $w_q \in \mathcal{L}$ (voor elke q).
- Als $a \geqslant 3$, kies dan $x = \varepsilon$, $y = \mathtt{0}$ en z gelijk aan de rest. Na op- of neerpompen blijkt $\#_{\mathtt{0}}(w_q) > 1$, zodat inderdaad $w_q \in \mathcal{L}$ (voor elke q).

Met andere woorden, deze taal $\mathcal{L}$ voldoet aan de conclusie van het pumping lemma: elk (voldoende lang) woord w in de taal is op- en neerpompbaar tot nieuwe woorden in de taal. Het pumping lemma kan dus onmogelijk leiden tot een strijdigheid. Toch is de taal niet-regulier! We kunnen dit hard maken dankzij de sluitingseigenschappen. Onderstel immers dat $\mathcal{L}$ tóch regulier zou zijn, dan ook de doorsnede

$$\mathcal{L} \cap \mathtt{0}\,\mathtt{1}^*\,\mathtt{2}^* = \{\mathtt{0}\,\mathtt{1}^b\,\mathtt{2}^c : b = c\},$$

en op deze taal is het pumping lemma wél toepasbaar (of je kan inzien dat de lettersubstitutie $\mathtt{0} \mapsto \varepsilon$ tot het klassieke voorbeeld van een niet-reguliere taal leidt). Alternatief kun je gebruik maken van de eigenschap dat reguliere talen gesloten zijn onder omkering: op de taal $\mathcal{L}^{\mathsf{R}}$ kun je namelijk wel rechtstreeks het pumping lemma toepassen, met het woord $w = \mathtt{2}^k\,\mathtt{1}^k\,\mathtt{0}$.