Kaikki mitä olet aina halunnut tietää pumppauslemmoista, mutta mitä et ole kehdannut kysyä

Tommi Syrjänen

1 Yleistä pumppauslemmoista

Pumppauslemmalla voidaan todistaa, että kieli ei kuulu johonkin kieliluokkaan. Useimmiten käytetään säännöllisten ja yhteydettömien kielten pumppauslemmoja, mutta muitakin on olemassa.

Pumppauslemma esittää ehdon, jonka jokainen luokkaan kuuluva kieli täyttää. Mikäli voidaan osoittaa, että kieli ei toteuta tätä ehtoa, niin se ei kuulu luokkaan. Eli:

$$L \text{ on säännöllinen} \quad \Rightarrow \quad L \text{ toteuttaa säännöllisten kielten pumppauslemman,}$$

joten

$$L \text{ ei toteuta säännöllisten kielten pumppauslemmaa} \quad \Rightarrow \quad L \text{ ei ole säännöllinen.}$$

Tässä on käytetty hyväksi logiikan *modus tollens* sääntöä, jonka mukaan lauseesta $P \rightarrow Q$ seuraa $\neg Q \rightarrow \neg P$ (jos Q on välttämättä tosi aina kun P on tosi ja jos tiedetään, että Q on epätosi, niin myöskään P ei voi olla tosi).

Sitä vastoin siitä, että kieli toteuttaa pumppauslemman ehdot ei voida päätellä vielä mitään. Tästä esitetään myöhemmin (sivulla 10) esimerkki, jossa kieli ei ole säännöllinen, mutta kaikkia sen pitkiä sanoja voidaan kuitenkin pumpata. Näin ollen pumppauslemmoja voidaan käyttää aina vain yhteen suuntaan: osoittamaan, että kieli ei kuulu annettuun luokkaan. Jos kieli kuuluu luokkaan, niin

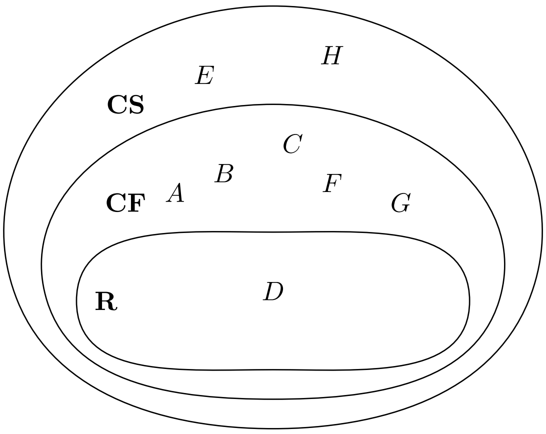
Tiivistelmä

Halutaan todistaa kieli L ei-säännölliseksi.

1. Oletetaan L säännölliseksi. Tällöin $\exists p \geq 1$ siten, että pumppauslemman ehdot toteutuvat.
2. *Valitaan* $x \in L$ siten, että $|x| \geq p$.
3. *Osoitetaan*, että *kaikki* x:n ositukset $x = uvw$ rikkovat jotain lemman ehtoa.

Kieli	Käytetty lemma	Todistustapa
$A = \{a^m b^m \mid m \geq 0\}$	Säännöllinen	$x = a^p b^p$
$B = \{a^m b^n \mid m \geq n\}$	Säännöllinen	$x = a^p b^p$
$C = \{a^m b^n \mid m \leq n\}$	Säännöllinen	$x = a^p b^p$
$D = L(bb \cup a^*)$	—	Ei todistusta
$E = \{a^i b^j c^n \mid i \leq j \leq n\}$	Yhteydetön	$x = a^p b^p c^p$
$F = \{a^i b^n c^n \mid i > 0, n \geq 0\} \cup L(b^* c^*)$	Säännöllinen	Leikkaus $L(aa^*b^*c^*)$ kanssa, $x = ab^p c^p$
$G = \{a^i b^j \mid i \neq j\}$	Säännöllinen	Komplementti
$H = \{a^n b^n a^m b^m c^m \mid n > 0, m \geq 0\}$ $\cup L(a^* b^* c^*)$	Yhteydetön	Leikkaus $L(aa^*bb^*a^*b^*c^*)$ kanssa, $x = aba^p b^p c^p$

Taulukko 1: Tekstissä tarkasteltavat kielet ja todistustavat



R: Säännölliset kielet
CF: Yhteydettömät kielet
CS: Yhteysherkät kielet

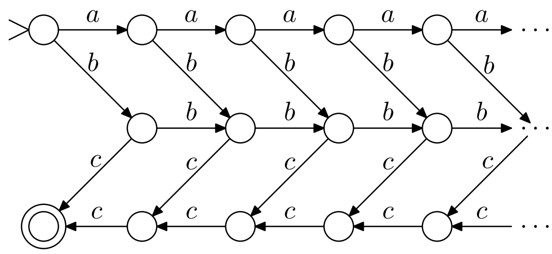
Kuva 1: Tekstissä tarkasteltavien kielten kuuluminen eri kieliluokkiin.

se osoitetaan esittämällä kieli luokan määrittelevällä formalismilla. Esimerkiksi kieli voidaan todistaa säännölliseksi muodostamalla sen tunnistava äärellinen automaatti.

Pumppauslemmat hyödyntävät kuhunkin kieliluokkaan kuuluvien kielten rakenteellista samankaltaisuutta. Esimerkiksi säännölliset kielet määritellään äärellisillä automaateilla, joilla on rajallisesti muistia, joten niillä ei voida tunnistaa kieltä, joka vaatii mielivaltaisen suuriksi kasvavien syötteiden muistamista. Tällainen kieli on vaikkapa:

$$L = \{a^i b^j c^k \mid i, j, k \in \mathbb{N} \text{ ja } i + j = k\} \ .$$

Tarkastaessaan kuuluko sana x kieleen L automaatti joutuu muistamaan sekä a- että b-merkkien lukumäärän. Äärellisten automaattien muistina toimivat automaatin tilat, joten kone joutuu siirtymään aina uuteen tilaan uuden a:n tai b:n luettuaan seuraavaan tapaan:



Tähän koneeseen tarvitaan kuitenkin ääretön määrä tiloja, sillä kielen määrittely ei aseta rajoja sille, montako a- ja b-merkkiä siinä saa olla. Koska kieltä ei voida tunnistaa äärellisellä automaatilla, se ei ole säännöllinen.

Yllä oleva todistelu jättää kuitenkin vielä avoimeksi mahdollisuuden, että jollain muulla periaatteella rakennettu äärellinen automaatti voisi tunnistaa sen. Todistus saadaan täsmälliseksi osoittamalla pumppauslemmalla, että mikä tahansa kielen tunnistava automaatti on väistämättä ääretön.

2 Säännöllisten kielten pumppauslemma

Säännöllisten kielten pumppauslemma kuuluu seuraavasti:

Määritelmä 1 *Jos L on säännöllinen kieli, niin on olemassa jokin $p \geq 1$ siten, että kaikille sanoille $x \in L$ pätee: jos $|x| \geq p$, niin x voidaan esittää muodossa $x = uvw$, missä:*

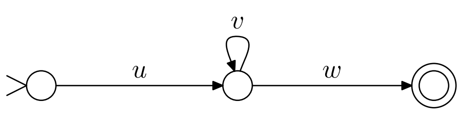
1. $|uv| \leq p$;
2. $|v| > 0$; *ja*
3. $uv^k w \in L$ *kaikilla* $k \in \mathbb{N}$.

Suorasanaisesti lemman ehdot voidaan selittää yhdellä lauseella:

> *Äärettömän säännöllisen kielen tunnistavassa tilakoneessa on silmukka.*

Seuraavaksi käydään läpi lemman ehdot tarkemmin. Lemmassa esiintyvä parametri p vastaa pienimmän L:n tunnistavan deterministisen äärellisen automaatin tilojen määrää. Deterministinen automaatti käyttää jokaisella laskenta-askeella täsmälleen yhden merkin syötteestä. Jos kieleen kuuluvassa sanassa x on enemmän merkkejä kuin automaatissa tiloja, niin silloin automaatissa täytyy olla jossain kohtaa silmukka, josta päästään vielä hyväksyvään lopputilaan.

Sana x jaetaan kolmeen osaan $x = uvw$. Näistä u on sanan alkuosa ennen silmukkaa, v silmukassa kierretty osa ja w loppuosa silmukan jälkeen. Tilakoneena asian voi esittää seuraavasti:



Ositukselle annettujen ehtojen perustelut ovat seuraavat:

1. $|uv| \leq p$: laskennan täytyy joutua silmukkaan ennen kuin automaatista loppuvat tilat kesken;
2. $|v| > 0$: silmukka ei voi olla tyhjä vaan siihen täytyy kuulua ainakin yksi ei-tyhjä siirtymä; ja
3. $uv^k w \in L$ kaikilla $k \in \mathbb{N}$: silmukka v voidaan kiertää mielivaltaisen monta kertaa ja silti voidaan päästä w:llä hyväksyvään lopputilaan.

Osituksen v ei ole välttämättä ainoa x:n osajono, joka luetaan silmukassa. Esimerkiksi kielen $L(a^*b^*)$ tunnistavassa koneessa on kaksi silmukkaa, yksi kummallekin tähtilausekkeelle.

2.1 Lemman soveltaminen

Säännöllisten kielten pumppauslemmaa sovelletaan siten, että etsitään jokin kieleen kuuluva tarpeeksi pitkä sana, jota ei voida osittaa lemman ehtojen mukaisesti.

Lemma sanoo, että *kaikille* sanoille x on olemassa *jokin* ositus. Todistuksessa täytyy osoittaa, että *jonkin* sanan *kaikki* ositukset johtavat ristiriitaan.

Todistetaan esimerkin vuoksi, että kieli

$$A = \{a^m b^m \mid m \geq 0\} \tag{1}$$

ei ole säännöllinen.

Tarkasteltava sana kannattaa valita siten, että lemman ensimmäisen ehdon täyttäviä osituksia on mahdollisimman vähän. Parhaiten tämä toteutuu siten, että otetaan sana, jonka p ensimmäistä merkkiä (missä p on pumppauslemmassa esiintyvä kielestä riippuva parametri) ovat kaikki samoja. Yleensä kannattaa valita näistä sanoista kaikkein yksinkertaisin.

Kielen A tapauksessa yksinkertaisin vastaesimerkki on $x = a^p b^p \in A$. Koska $|x| = 2p > p$, on sana tarpeeksi pitkä. Lemman ehdosta $|uv| \leq p$ seuraa, että kaikki x:n mahdolliset ositukset ovat muotoa:

$$\begin{aligned} u &= a^i \\ v &= a^j \\ w &= a^{p-(i+j)} b^p, \end{aligned} \tag{2}$$

missä $i + j \leq p$. Ehdosta $|v| > 0$ seuraa, että $j > 0$. Tässä täytyy erikoisesti kiinnittää huomiota siihen, että (2) ei ole yksi ainoa ositus, vaan se on joukko ehtoja, jotka x:n ositusten täytyy täyttää. Varsinaiset ositukset saadaan siitä sijoittamalla muuttujien i ja j paikalle sopivat lukuarvot.

Todistus on jo aika lähellä loppuaan, mutta vielä täytyy osoittaa, että mikään (2):n täyttävä ositus ei voi täyttää lemman kolmatta ehtoa ($\forall k : uv^k w \in L$). Käytännössä tämä tapahtuu siten, että etsitään jokin k:n arvo, jolla pumpattu sana ei kuulu kieleen. Yleensä on helpointa kokeilla ensin arvoa $k = 0$:

$$uv^0 w = uw = a^i a^{p-(i+j)} b^p = a^{p-j} b^p$$

Koska kaikilla osituksilla $j > 0$, on $p - j < p$ ja $a^{p-j} b^p \notin A$. Tästä nähdään, että A ei ole säännöllinen.

Kertaus: löydettiin kieleen A kuuluva tarpeeksi pitkä sana, jonka kaikki ositukset rikkovat ainakin yhtä pumppauslemman ehtoa. Jos A olisi säännöllinen, niin kaikki sen pitkät sanat voitaisiin osittaa, joten A ei voi olla säännöllinen.

2.2 Yleisiä virheitä lemman käytössä

Pumppauslemmaa käytettäessä kannattaa erityisesti varoa seuraavien tyypillisten virheiden tekemistä:

1. sanan x:n ositusten puutteellinen läpikäynti;
2. pumppautuvan sanan x käyttäminen vastaesimerkkinä;
3. väärään suuntaan pumppaaminen;
4. liian lyhyen x:n käyttäminen;
5. kielen L mielivaltaisen osajoukon L' tarkastelu; ja
6. väärän pumppauslemman soveltaminen.

2.2.1 Puutteellinen ositus

Tarkastellaan uudelleen kieltä $A = \{a^m b^m \mid m \geq 0\}$, mutta tällä kertaa tutkitaan sanaa $x = a^q b^q$, missä $q = \lceil p/2 \rceil$.[1] Koska $|x| \geq p$, on x tarpeeksi pitkä.

Nyt kohdan (2) mukaiset ositukset muuttuvat muotoon:

$$\begin{aligned} u &= a^i \\ v &= a^j \\ w &= a^{q-(i+j)} b^q, \end{aligned} \tag{2'}$$

missä $i + j \leq q$, $j > 0$. Kuten yllä, $uv \notin A$.

Tällä kertaa todistusta ei kuitenkaan saa jättää tähän, sillä x voidaan osittaa muillakin ehdot $|uv| \leq p$ ja $|v| > 0$ toteuttavilla tavoilla.

Nämä muut ositukset ovat muotoa:

$$\begin{aligned} u &= a^q b^i \\ v &= b^j \\ w &= b^{q-(i+j)}, \end{aligned} \tag{3}$$

missä $i + j \leq q$, ja

$$\begin{aligned} u &= a^{q-i} \\ v &= a^i b^j \\ w &= b^{q-j}, \end{aligned} \tag{4}$$

missä $i + j > 0$ ja $i, j \leq q$.

Todistettaessa kieltä A epäsäännölliseksi sanalla $a^{\lceil p/2 \rceil} b^{\lceil p/2 \rceil}$ täytyy siis erikseen tarkistaa se, että *kaikki* muotoja (2'), (3) ja (4) olevat ositukset johtavat ristiriitaan. Tämän takia tutkittava sana kannattaakin valita huolella, niin että sillä on mahdollisimman vähän ehdon $|uv| \leq p$ täyttäviä osituksia.

[1] Merkintä $\lceil p/2 \rceil$ tarkoittaa pienintä kokonaislukua y, jolle pätee $p/2 \leq y$

2.2.2 Pumppautuva sana vastaesimerkkinä

Pumppauslemman perusteella kieli ei ole säännöllinen, jos siihen kuuluu jokin tarpeeksi pitkä sana, jota ei voida pumpata. Tämä ei kuitenkaan tarkoita sitä, että kaikki kieleen kuuluvat pitkät sanat olisivat välttämättä pumppauskelvottomia. Tarkastellaan kieltä:

$$B = \{a^m b^n \mid m \geq n\}.$$

Sana $x = a^{p+1}b^p$ kuuluu kieleen ja $|x| > p$. Tälle löytyy kuitenkin ositus:

$$\begin{aligned} u &= a^{p-1} \\ v &= a \\ w &= ab^p, \end{aligned}$$

joka toteuttaa kaikki ehdot. Suoraan nähdään, että $|uv| = |a^p| = p$ ja $|v| = |a| = 1 > 0$. Tarkastellaan v:n pumppaamista muutamalla k:n arvolla:

$$\begin{aligned} uv^0w &= a^p b^p \in B \\ uv^2w &= a^{p+2}b^p \in B \\ uv^3w &= a^{p+3}b^p \in B \\ &\vdots \end{aligned}$$

Koska k:n kasvattaminen ainoastaan lisää sanan alkuun a-kirjaimia, on selvää että kaikki näin saatavat sanat kuuluvat kieleen.

Kieltä B ei voida osoittaa epäsäännölliseksi sanalla $x = a^{p+1}b^p$. Sillä, että jotkin x:n ositukset johtavat ristiriitaan (esim. $u = a$, $v = a^{p-1}$ ja $w = ab^p$) ei ole väliä. Jos on olemassa ainakin yksi lemman ehdot täyttävä ositus, niin sana täyttää lemman ehdot.

Kielen B voi todistaa ei-säännölliseksi vaikkapa sanalla $a^p b^p$.

2.2.3 Väärään suuntaan pumppaaminen

Yleensä on yksinkertaisinta kokeilla ensin 0-pumppausta, eli poistaa v kokonaan sanasta. Tämä ei kuitenkaan välttämättä riitä todistuksen tekemiseen, vaan pumppaus joudutaankin tekemään ylöspäin. Näin käy esimerkiksi kielellä:

$$C = \{a^m b^n \mid m \leq n\} \ . \tag{5}$$

Jälleen kerran yksinkertaisin kandidaatti on $a^p b^p$, jonka kaikki ositukset ovat tuttua muotoa (2).

Nyt kuitenkin $uw = a^{p-j}b^p \in C$, joten pelkkä 0-pumppaus ei riitä osoittamaan epäsäännöllisyyttä. Sitä vastoin kasvattamalla k:n arvoa saadaan:

$$uv^2w = a^{p+j}b^p \notin C,$$

joten C ei ole säännöllinen.

2.2.4 Liian lyhyen sanan pumppaaminen

Yksi pumppauslemman ehdoista on, että tarkasteltavan sanan täytyy olla pidempi kuin kielestä riippuva kokonaisluku p. Säännölliseen kieleen voi kuitenkin kuulua joukko pumppauskelvottomia sanoja, jotka ovat lyhyempiä kuin p (tällaisia sanoja voi kuitenkin olla vain äärellinen määrä). Tarkastellaan säännöllistä kieltä:

$$D = L(bb \cup a^*) = \{\varepsilon, bb, a, aa, aaa, \ldots\} \tag{6}$$

Kielen sanaa bb ei voida pumpata. Tämä ei kuitenkaan ole ongelma, sillä pienimmässä kielen tunnistavassa äärellisessä automaatissa on 5 tilaa, joten tälle kielelle $p = 5$ ja $|bb| = 2 < 5 = p$.

Vaatimus $|x| \geq p$ aiheuttaa sen, että esimerkiksi kieltä A ei voi todistaa epäsäännölliseksi sanalla ab, sillä ei voida olettaa, että $|ab| \geq p$.

2.2.5 Kielen mielivaltaisen osajoukon tarkastelu

Joskus kieli L voidaan osoittaa epäsäännölliseksi ainoastaan tarkastelemalla sen osajoukkoa L', ja osoittamalla L' epäsäännölliseksi. Tästä menetelmästä annetaan esimerkkejä luvussa 4.

Kuitenkaan ei ole niin, että mikä tahansa säännöllisen kielen osajoukko olisi säännöllinen. Esimerkiksi kieli $\Sigma^* = \{a, b\}^*$ on säännöllinen, sillä $L((a \cup b)^*) = \Sigma^*$, mutta $A \subset \Sigma^*$.

Tämän takia todistuksessa täytyy erikseen perustella, miksi L':n epäsäännöllisyys johtaa välttämättä L:n epäsäännöllisyyteen.

2.2.6 Väärän pumppauslemman soveltaminen

Niin yllättävältä kuin se kuulostaakin, varsin yleinen virhe pumppauslemmojen käytössä on väärän pumppauslemman valitseminen. Jos tutkittava kieli on yhteydetön, niin yritys todistaa se epäsäännölliseksi käyttämällä yhteydettömien kielten pumppauslemmaa ei johda kovin hyviin tuloksiin.

3 Yhteydettömien kielten pumppauslemma

Yhteydettömien kielten pumppauslemma määritellään seuraavasti:

Määritelmä 2 *Jos L on yhteydetön kieli, niin on olemassa jokin $p \geq 1$ siten, että kaikille sanoille $z \in L$ pätee: jos $|z| \geq p$, niin z voidaan esittää muodossa $z = uvwxy$, missä:*

1. $|vx| > 0$;
2. $|vwx| \leq p$; *ja*
3. $uv^k wx^k y \in L$ *kaikilla* $k \in \mathbb{N}$.

Vaikka lemma on monimutkaisempi kuin säännöllisten kielten vastaava, on sen perusajatus samankaltainen. Yhdellä lauseella ilmaistuna se kuuluu:

> *Ääretöntä yhteydetöntä kieltä vastaavassa kieliopissa jokin välike tuottaa itsensä.*

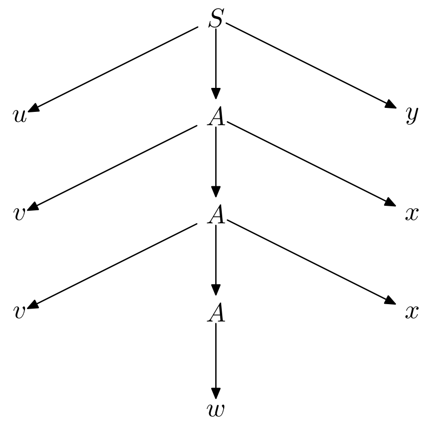
Koska yhteydettömässä kieliopissa on aina äärellinen määrä sääntöjä ja kukin sääntö on äärellisen mittainen, voidaan mielivaltaisen pitkiä sanoja tuottaa ainoastaan muotoa $A \Rightarrow^* \alpha A \beta$ olevilla johdoilla, eli jokin välike A täytyy voida johtaa itsestään. Lemmaa voi havainnollistaa seuraavalla yksinkertaisella kieliopilla:

$$\begin{aligned} S &\to uAy \\ A &\to vAx \mid w \end{aligned} \tag{7}$$

Korvattaessa välikettä A voidaan aina valita käytetäänkö sääntöä $A \to vAx$ johtamaan uusi A-välike, tai sitten voidaan lopettaa johto säännöllä $A \to w$. Johdon aikana A:n vasemmalle puolelle kasaantuu v-jonoja ja oikealle puolelle x-jonoja.

$$S \Rightarrow uAy \Rightarrow uvAxy \Rightarrow uvvAxxy \Rightarrow uvvvAxxxy \Rightarrow uvvvwxxxy$$

Jäsennyspuuna sama asia näyttää seuraavalta:



3.1 Lemman soveltaminen

Yhteydettömien kielten pumppauslemmaa sovelletaan täsmälleen samoin kuin säännöllisten kielten vastaavaa: etsitään kieleen kuuluva sana ja näytetään, että sitä ei voi osittaa annettujen ehtojen mukaisesti. Lemman monimutkaisemman rakenteen vuoksi mahdollisia osituksia on yleensä enemmän kuin säännöllisten kielten lemman tapauksessa.

Todistetaan, että kieli:

$$E = \{a^i b^j c^n \mid i \leq j \leq n\} \tag{8}$$

ei ole yhteydetön.

Tarkastellaan sanaa $z = a^p b^p c^p \in E$. Ehdosta $|vwx| \leq p$ seuraa, että osajono vwx voi sisältää korkeintaan kahta eri merkkiä. Tarkastellaan näitä tapauksia erikseen.

Jos vwx sisältää vain yhtä merkkiä, niin mahdollisia osituksia ovat:

1.

$$\begin{aligned} u &= a^i \\ vwx &= a^j \\ y &= a^{p-(i+j)} b^p c^p, \end{aligned} \tag{9}$$

missä $i + j \leq p$ ja $j > 0$. Tällöin $uv^2wx^2y = a^{p+i+j}b^pc^p \notin E$. (Tällä kertaa 0-pumppaus ei auta, sillä a-merkkien lukumäärä on rajoitettu vain ylhäältä päin, joten niitä voi poistaa sanoista ilman ongelmia.)

2.

$$\begin{aligned} u &= a^p b^i \\ vwx &= b^j \\ y &= b^{p-(i+j)}c^p, \end{aligned} \tag{10}$$

$(i + j \leq p, j > 0)$. Nyt $uwy = a^p b^{p-(i+j)}c^p \notin E$.

3.

$$\begin{aligned} u &= a^p b^p c^i \\ vwx &= c^j \\ y &= c^{p-(i+j)}, \end{aligned} \tag{11}$$

$(i + j \leq p, j > 0)$. Nollapumppauksella saadaan $uwy = a^p b^p c^{p-(i+j)} \notin E$.

Vastaavasti kahta eri merkkiä sisältäviä osituksia ovat:

1.

$$\begin{aligned} u &= a^i \\ vwx &= a^j b^k \\ y &= b^{p-k}c^p, \end{aligned} \tag{12}$$

missä $i+j = p$, $k < p$ ja $k, j > 0$. Tämä tapaus jakautuu edelleen kolmeen mahdolliseen tilanteeseen.

(a) Jos $v = a^m$, $w = a^{j-m}b^k$ ja $x = \varepsilon$ $(1 \leq m \leq j)$, niin ositus palautuu ylläolevaan tapaukseen (9). Samoin ositus $v = \varepsilon$, $w = a^j b^m$, $y = b^{k-m}$ $(0 \leq m < k)$ palautuu tapaukseen (10).

(b) Jos $v = a^m$, $w = a^{j-m}b^{n-k}$ ja $x = b^k$ $(1 \leq m \leq j,\ 1 \leq n \leq k)$, niin $uv^2wx^2y = a^{p+m}b^{p+k}c^p \notin E$.

(c) Jos $v = a^j b^m$ tai $x = a^n b^k$ $(1 \leq m \leq k,\ 1 \leq n \leq j)$, niin sanaa ylöspäin pumpatessa muodostuu osajonoja ba, joita ei esiinny yhdessäkään kielen sanassa.

2.

$$\begin{aligned} u &= a^p b^i \\ vwx &= b^j c^k \\ y &= c^{p-k}, \end{aligned} \tag{13}$$

missä $i + j = p$, $k < p$ ja $k, j > 0$. Tässä tapauksessa käy samoin kuin edellisessä.

Koska sanaa $z = a^p b^p c^p \in E$ ei voida osittaa yhteydettömien kielten pumppauslemman ehtojen mukaisesti, ei E ole yhteydetön.

L_1 ja L_2 säännöllisiä	$\Rightarrow$	$L_1 \cap L_2$ säännöllinen
L säännöllinen	$\Rightarrow$	$\overline{L}$ säännöllinen
L_1 yhteydetön, L_2 säännöllinen	$\Rightarrow$	$L_1 \cap L_2$ yhteydetön

Taulukko 2: Todistuksissa käytetyt sulkeumaominaisuudet

4 Sulkeumaominaisuudet

Joskus tutkittavan kielen muokkaaminen ennen pumppauslemman käyttämistä helpottaa todistuksen tekemistä, ja eräissä harvinaisissa tapauksissa se on jopa välttämätöntäkin.

Kielen muokkaamisessa käytetään apuna kieliluokkien sulkeumaominaisuuksia. Esimerkiksi säännölliset kielet ovat suljettuja leikkauksen suhteen, joten jos L_1 ja L_2 ovat säännöllisiä, niin myös $L_1 \cap L_2$ on säännöllinen. Todistusmenetelmä ominaisuudesta saadaan seuraavalla säännöllä:

$$L_1 \text{ ja } L_2 \text{ säännöllisiä} \quad \Rightarrow \quad L_1 \cap L_2 \text{ säännöllinen,}$$

joten

$$L_1 \cap L_2 \text{ ei ole säännöllinen} \quad \Rightarrow \quad \text{joko } L_1 \text{ tai } L_2 \text{ ei ole säännöllinen.}$$

Jos kieli L_2 valitaan siten, että se on varmasti säännöllinen, niin tutkittava kieli L_1 voidaan osoittaa epäsäännölliseksi tarkastelemalla leikkausta $L_1 \cap L_2$.

Yleensä pumppauslemmatodistuksissa käytetään vain seuraavia sulkeumaominaisuuksia:

$$\begin{aligned} L_1 \text{ ja } L_2 \text{ säännöllisiä} &\quad \Rightarrow \quad L_1 \cap L_2 \text{ säännöllisiä} \\ L \text{ säännöllinen} &\quad \Rightarrow \quad \overline{L} \text{ säännöllinen} \\ L_1 \text{ yhteydetön ja } L_2 \text{ säännöllinen} &\quad \Rightarrow \quad L_1 \cap L_2 \text{ yhteydetön} \end{aligned}$$

Tarkastellaan seuraavaksi esimerkkien avulla, miten niitä käytetään.

4.1 Säännöllisten kielten leikkaus

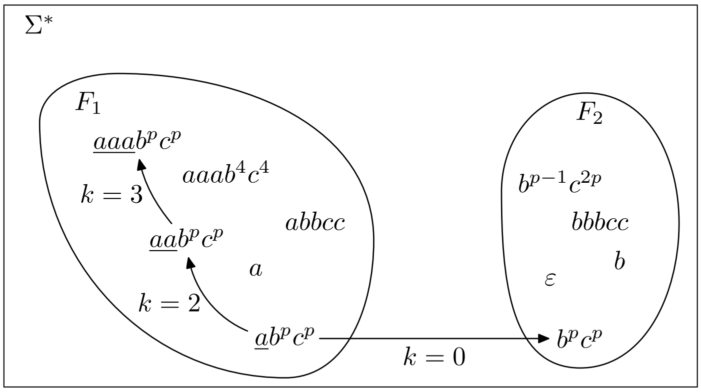
Yksinkertaisin esimerkki ei-säännöllisestä kielestä, johon ei voida suoraan soveltaa pumppauslemmaa on kieli:

$$F = \{a^i b^n c^n \mid i > 0, n \geq 0\} \cup L(b^* c^*) \ . \tag{14}$$

Ongelmana on se, että mitä tahansa kieleen kuuluvaa ei-tyhjää merkkijonoa voidaan pumpata. Tämä nähdään tarkastelemalla unionin osia erikseen:

$$\begin{aligned} F &= F_1 \cup F_2 \\ F_1 &= \{a^i b^n c^n \mid i > 0, n \geq 0\} \\ F_2 &= L(b^* c^*) \ . \end{aligned}$$

Tarkastellaan sanaa $x \in F_1$. Tällöin $x = aa^j b^n c^n$ jollain $j, n \geq 0$. Sana x voidaan osittaa siten, että valitaan pumpattavaksi osuudeksi yksi ainoa a-merkki. Koska F_1:n sanojen alussa saa olla mielivaltaisen monta a:ta, ei ylöspäin pumppaamalla voida saada aikaan kieleen kuulumatonta sanaa. Mikäli $j = 0$ saadaan alaspäin



Kuva 2: Kielen F sanan ab^pc^p pumppaaminen merkkijonolla $x = ab^pc^p$.

pumppaamalla aikaan sana $b^nc^n \notin F_1$. Tämä ei kuitenkaan riitä osoittamaan F:n epäsäännöllisyyttä, sillä $b^nc^n \in F_2$, joten $b^nc^n \in F$.

Tässä tapauksessa tutkittava kieli siis koostuu kahdesta erillisestä sanajoukosta joista toinen (F_2) on säännöllinen ja toinen (F_1) ei, mutta ei-säännöllistä osaa pumpatessa syntynyt vastaesimerkki sisältyy säännölliseen osaan.

Jotta F voidaan todistaa epäsäännölliseksi, niin täytyy jotenkin päästä eroon alikielestä F_2. Tämä voidaan tehdä ottamalla F:n leikkaus kielen $L(aa^*b^*c^*)$ kanssa:

$$F \cap L(aa^*b^*c^*) = \{a, abc, abbcc, aa, aabc, aabbcc, \ldots\} = F_1$$

Kieli $L(aa^*b^*c^*)$ on säännöllinen. Jos F olisi säännöllinen, niin silloin välttämättä myös niiden leikkaus F_1 olisi välttämättä säännöllinen. Joten, jos F_1 ei ole säännöllinen, niin ei myöskään F voi olla säännöllinen.

Kielen F_1 voi todistaa epäsäännölliseksi suoraviivaisesti käyttäen sanaa ab^pc^p.

4.1.1 Säännöllisen kielen komplementti

Tarkastellaan kieltä:

$$G = \{a^ib^j \mid i \neq j\} \tag{15}$$

Tämän kielen voi todistaa epäsäännölliseksi suoraan pumppauslemmalla, mutta sopivan vastaesimerkin löytäminen ei ole aivan yksinkertaista.

Todistus syntyy helpommin huomaamalla, että

$$\overline{G} = \{a^mb^m \mid m \geq 0\} = A \ .$$

Koska säännölliset kielet ovat suljettuja komplementin suhteen, voi G olla säännöllinen vain jos A on säännöllinen. Koska A ei ole säännöllinen, seuraa tästä suoraan G:n epäsäännöllisyys.

Suoraan pumpattava vastaesimerkki kielen säännöllisyydelle on $z = a^pb^{p!}$, missä $p!$ tarkoittaa p:n kertomaa $p! = 1 \times 2 \times \cdots \times p$. Sitä käyttäessä täytyy tosin vielä erikseen osoittaa, että kaikilla $1 \leq i \leq p$ on olemassa jokin $k \in \mathbb{N}$ siten, että $(p - i) + k \cdot i = p!$.

4.1.2 Yhteydettömän ja säännöllisen kielen leikkaus

Yhteydettömät kielet eivät ole suljettuja leikkauksen suhteen, sillä esimerkiksi kieli $\{a^n b^n c^n \mid n \geq 0\}$ voidaan esittää kahden yhteydettömän kielen $\{a^n b^n c^i\}$ ja $\{a^i b^n c^n\}$ leikkauksena.

Sitävastoin yhteydettömän ja säännöllisen kielen leikkaus on aina yhteydetön, ja tätä seikkaa voidaan käyttää apuna todistuksissa. Esimerkiksi kieli:

$$H = \{a^n b^n a^m b^m c^m \mid n > 1, m \geq 0\} \cup L(a^* b^* c^*), \tag{16}$$

joka on muodostettu samalla periaatteella kuin luvun 4.1 kieli F, ei ole yhteydetön, mutta kaikkia siihen kuuluvia sanoja voidaan pumpata yhteydettömien kielten pumppauslemman ehtojen mukaisesti. Tälläkin kertaa täytyy todistusta tehtäessä leikata pois unionin oikean puolen tuoma säännöllinen alikieli:

$$H' = H \cap L(aa^*bb^*a^*b^*c^*) = \{a^n b^n a^m b^m c^m \mid n > 1, m \geq 0\} \ . \tag{17}$$

Kieli H' voidaan osoittaa ei-yhteydettömäksi esimerkiksi sanalla $aba^p b^p c^p \in H'$. Koska H' saatiin tutkittavan kielen ja jonkin säännöllisen kielen leikkauksena, sen ei-yhteydettömyydestä seuraa suoraan se, että myöskään H ei voi olla yhteydetön.

4.1.3 Sulkeuma-argumenttien vaaroista

Eri kieliluokkien sulkeumaominaisuuksia käyttettäessä täytyy olla tarkkana siitä, mihin suuntaan argumenttia käyttää. Jos L_1 ja L_2 ovat säännöllisiä, niin myös $L_1 \cap L_2$ on säännöllinen, mutta $L_1 \cap L_2$ voi olla säännöllinen, vaikka L_1 ja L_2 eivät ole. Esimerkiksi yllä esiteltyjen kielten A ja G:

$$A = \{a^m b^m \mid m \geq 0\}$$
$$G = \{a^i b^j \mid i \neq j\}$$

tapauksessa $A \cap G = \emptyset = L(\emptyset)$. Näin siis kahden ei-säännöllisen kielen leikkaus voi olla säännöllinen.

Myöskään se, että kieli sisältää ei-säännöllisen osajoukon ei ole riittävä ehto sen epäsäännöllisyydelle. Esimerkiksi kieli:

$$\{b^m c^m \mid m \geq 0\} \cup L(a^* b^* c^*)$$

on säännöllinen, vaikka sen määritelmässä esiintyy ei-säännöllinen kieli. Tämä siksi, että kaikki kielen $\{b^m c^m \mid m \geq 0\}$ sanat kuuluvat myös kieleen $L(a^* b^* c^*)$, joten unioni ei itseasiassa lisää tähän yhtään uutta sanaa.

Todistuksissa käytetään sulkeumaominaisuuksista oikeastaan vain leikkausta ja komplementtia. Unionia ja katenaatiota ei käytetä juuri lainkaan, sillä ne yleensä vain muuttavat kielen hankalammaksi käsitellä. Erityisesti katenaation kanssa täytyy olla varovainen, sillä sitä käyttää varsin helposti väärään suuntaan. Turvallisinta onkin pitäytyä ainoastaan leikkauksessa ja komplementissa. Tällöinkin on muistettava, että yhteydettömät kielet eivät ole suljettuja kummankaan operaation suhteen.

5 Yhteenveto

Pumppauslemmoilla voidaan osoittaa, että kieli ei kuulu johonkin kieliluokkaan. Toiseen suuntaan niitä ei voi käyttää. Niitä käytetään siten, että:

1. Valitaan jokin sana $x \in L$, $|x| \geq p$, missä p on kielestä L riippuva pumppauspituus.

2. Osoitetaan, että sanaa x ei voida osittaa siten, että pumppauslemman ehdot toteutuvat.

Säännöllisten kielten pumppauslemman intuitiivinen tulkinta on se, että äärettömän kielen tunnistavassa tilakoneessa täytyy olla silmukka. Lemmalla todistettaessa osoitetaan, että kielen sanojen rakenne on sellainen, että niitä ei voida tunnistaa käyttäen silmukoita. Yhteydettömien kielten pumppauslemma puolestaan sanoo, että ääretöntä kieltä vastaavassa yhteydettömässä kieliopissa jostain välikkeestä saadaan johdettua se itse.

Joskus todistuksen teossa käytetään apuna kielten leikkausta. Tällöin otetaan tutkittavan kielen leikkaus jonkin säännöllisen kielen kanssa, ja todistetaan että se ei ole säännöllinen tai yhteydetön. Tätä ei kuitenkaan kannata tehdä turhaan, sillä se lisää uuden mahdollisen virhepaikan todistukseen. Samoin on joskus helpompi todistaa kieli ei-säännölliseksi tekemällä todistus sen komplementille.

6 Harjoitustehtäviä

1. Osoita, että sivulla 2 määritelty kieli L on yhteydetön, mutta ei säännöllinen.

2. Todista täsmällisesti, että kieli F_1 ei ole säännöllinen käymällä läpi kaikki annetun vastaesimerkin ositukset.

3. Todista täsmällisesti, että kieli H' ei ole yhteydetön käymällä läpi kaikki annetun vastaesimerkin ositukset.

4. Laadi kielen D tunnistava deterministinen minimiautomaatti.

5. Osoita, että kielet A, B, C, F ja G ovat yhteydettömiä laatimalla ne tuottavat yhteydettömät kieliopit.

5.* Osoita, että kielet E ja H ovat yhteysherkät laatimalla ne tuottavat yhteysherkät kieliopit.