SANOJEN JAKSOT JA REUNATTOMAT TEKIJÄT. EHRENFEUCHTIN–SILBERGERIN ONGELMA

Kari Vänttinen

Pro gradu -tutkielma
Kesäkuu 2016

MATEMATIIKAN JA TILASTOTIETEEN LAITOS
TURUN YLIOPISTO

TURUN YLIOPISTO
Matematiikan ja tilastotieteen laitos

VÄNTTINEN, KARI: Sanojen jaksot ja reunattomat tekijät.
Ehrenfeuchtin—Silbergerin ongelma.
Pro gradu -tutkielma, 88 s.
Matematiikka
Kesäkuu 2016

Ehrenfeuchtin–Silbergerin ongelma kysyy, kuinka pitkä sana voi olla sen pisimpien reunattomien tekijöiden pituuden suhteen, ennen kuin sillä on sen lyhimmän jakson pituinen reunaton tekijä. Ongelman ratkaisu on sanan pisimpien reunattomien tekijöiden pituudesta riippuva raja, jolle kaikilla ainakin tämän pituisilla sanoilla on välttämättä lyhimmän jakson pituinen reunaton tekijä. Tutkielmassa esitellään tämän ongelman paras tunnettu ratkaisu. Lisäksi tarkastellaan muita ongelmaan läheisesti liittyviä tuloksia.

Päälauseena todistetaan paras tunnettu raja pisimpien reunattomien tekijöiden suhteen. Todistus on peräisin Štěpán Holubin ja Dirk Nowotkan artikkelista The Ehrenfeucht–Silberger problem (Journal of Combinatorial Theory, Series A) sekä tämän artikkelin alustavasta versiosta ICALP 2009 -konferenssin proceedingsjulkaisussa. Esitetty ratkaisu näytetään vakiotermiä vaille optimaaliseksi vertaamalla sitä parhaaseen tunnettuun esimerkkiin äärettömästä sanajoukosta, jonka jokaisen sanan pisimmät reunattomat tekijät ovat lyhyempiä kuin lyhin jakso ja jonka jokaisen sanan pituus pisimpien reunattomien tekijöiden pituuden suhteen on suurin tunnettu.

Johdatteluna esitellään perustuloksia sanan jaksoista ja reunattomista tekijöistä sekä esitellään eräitä muita ehtoja sille, milloin sanalla on sen lyhimmän jakson pituinen reunaton tekijä.

Toisaalta tarkastellaan myös ongelmaa, joka kysyy vastaavaa rajaa lyhimmän jakson suhteen. Uutena tuloksena parannetaan parasta aiemmin tunnettua rajaa yhtä pienemmäksi, jolloin saatu raja on optimaalinen. Lisäksi todistetaan, mitä muotoa ovat kaikki sanat, joilla ei ole lyhimmän jaksonsa pituista reunatonta tekijää ja jotka ovat lyhimmän jaksonsa suhteen mahdollisimman pitkiä.

Lisäksi tarkastellaan kriittistä tekijöihinjakoa, joka liittää sanan lyhimmän jakson sen paikallisiin jaksoihin. Kriittisen tekijöihinjaon lauseesta esitetään eräs todistus. Tämän lisäksi todistetaan päälauseen todistuksessa tarvittava lemma, joka liittyy sanan konjugaatin tekijöihinjaon kriittisyyteen.

Asiasanat: diskreetti matematiikka, sanojen kombinatoriikka, reuna, jakso, jaksollisuus, Ehrenfeuchtin–Silbergerin ongelma, kriittinen tekijöihinjako

Sisältö

1 Johdanto

Sanoilla tarkoitetaan matematiikassa kirjaimista muodostettua jonoa, johon ei liitetä merkityksiä niin kuin luonnollisessa kielessä tehdään. Sanalla on reuna, jos sillä on jokin toinen sana prefiksinä ja suffiksina. Tällöin sanan lopussa on osa, joka on toistoa sanan alusta. Jaksoksi kutsutaan pituutta, jonka jälkeen sanan kirjaimet alkavat toistua samalla tavalla kuin sanan alussa. Sanan jaksoa ja reunaa voidaan pitää saman asian kahtena puolena, koska jos kuljetaan sanaa vasemmalta oikealle, niin jakson jälkeen tuleva sanan loppuosa on aina reuna.

Tässä tutkielmassa käsitellään sanojen jaksoja ja reunattomia tekijöitä erityisesti Andrzej Ehrenfeuchtin ja D. M. Silbergerin vuonna 1979 julkaistun artikkelin [6] esittämien kysymysten kannalta. On helppoa nähdä, että sanan pisin reunaton tekijä on aina enintään lyhimmän jakson pituinen. Tässä artikkelissa pohdittiin ensimmäisen kerran, milloin sanalla on reunaton tekijä, joka on yhtä pitkä kuin sanan lyhin jakso. Näin tapahtuu ainakin silloin, kun sana on hyvin paljon pidempi kuin sen jakso tai pisimmät reunattomat tekijät. Ehrenfeucht ja Silberger esittivät kysymyksen, kuinka paljon pidempi sanan tulee olla lyhimmän jakson tai toisaalta pisimpien reunattomien tekijöiden pituuden suhteen, jotta sanalla olisi varmasti lyhimmän jakson pituinen reunaton rekijä.

Artikkelissaan Ehrenfeucht ja Silberger saivat rajaksi lyhimmän jakson π suhteen $2\,\pi$. Lisäksi he esittivät otaksuman, että samanlainen raja $2\,\tau$ pätee myös pisimpien reunattomien tekijöiden pituuksille τ. Ronald Assous ja Maurice Pouzet [1] näyttivät kuitenkin samaan aikaan tämän otaksuman vääräksi. He generoivat äärettömän joukon sanoja, jonka pituudet ovat aina $\frac{7}{3}\,\tau - 4$ ja joilla kuitenkaan ei lyhimmän jaksoisia reunattomia tekijöitä. He taas esittivät monimutkaisemman otaksuman, joka yksinkertaistaen tarkoittaisi rajaa $3\,\tau$. Jean-Pierre Duval todisti vuonna 1982 [5] rajan $4\,\tau - 6$ ja esitti toisen otaksuman, josta voidaan johtaa raja $3\,\tau$, jos tämä otaksuma pitäisi paikkansa. Samalla Duval paransi Ehrenfeuchtin ja Silbergerin periodirajaa rajaksi $2\,\pi - k$, missä k on sanassa esiintyvien eri kirjainten lukumäärä.

Ehrenfeuchtin–Silbergerin ongelmaan rajasta suureen τ suhteen tutkittiin 2000-luvun ensimmäisen vuosikymmenen alku- ja keskivaiheilla enemmänkin Duvalin otaksuman seurauksena. Tero Harju ja Dirk Nowotka esittivät artikkelissaan vuonna 2007 [8] todistuksen Duvalin otaksumalle ja esittivät Ehrenfeuchtin–Silbergerin ongelmaan liittyen otaksuman, että Assous'n–Pouzet'n esimerkki on optimaalinen, mikä antaisi rajan $\frac{7}{3}\tau - 3$. Štěpán Holub ja Nowotka todistivat rajan $\frac{7}{3}\tau$ vuoden 2009 konferenssijulkaisussaan [9] ja vuoden 2012 artikkelissaan [10].

Luvussa 2 esitellään sanojen ja jaksollisuuden perustuloksia.

Luvussa 3 näytetään, miten Assous ja Pouzet päätyivät esimerkkiinsä ja pohditaan, millä tavalla se on merkittävä Ehrenfeuchtin–Silbergerin ongelmaan liittyen. Lisäksi pohditaan, että jos olisi olemassa tätä esimerkkiä parempia esimerkkisanoja, niin mitä muotoa niiden pitäisi olla. Tässä tarkastelussa hyödynnetään tietoa, että Ehrenfeuchtin–Silbergerin ongelman raja on ainakin $\frac{7}{3}\tau$.

Luvussa 6 käsitellään kriittistä tekijöihinjakoa, joka on muotoiltu ja todistettu samoihin aikoihin kuin Ehrenfeuchtin–Silbergerin ongelma muotoutui. Kriittinen tekijöihinjako on omana tuloksenaan syvällinen tulos sanan paikallisista jaksoista, ja sitä käytetään myös Ehrenfeuchtin–Silbergerin ongelman ratkaisussa.

Luvussa 5 käsitellään rajaa lyhimmän jakson π suhteen ja parannetaan Duvalin todistamaa rajaa yhdellä pienemmäksi jatkamalla Duvalin esittämästä todistuksesta. Lisäksi näytetään, että näin saatu uusi raja on paras mahdollinen ja todistetaan, mitä muotoa ovat kaikki ne sanat, jotka näyttävät rajan optimaaliseksi.

Luvussa 6 esitetään Holubin ja Nowotkan esittämä ratkaisu Ehrenfeuchtin–Silbergerin ongelmaan.

2 Merkintöjä ja perustuloksia

Tässä luvussa esitellään myöhemmissä luvuissa käytettäviä käsitteitä ja tuloksia sekä esitellään tässä tutkielmassa käytettävät merkintätavat. Sellaiset käsitteet, joita käytetään vain yhdessä luvussa, esitellään kyseisessä luvussa itsessään. Luvussa 2.1 esitellään äärelliset sanat. Luvussa 2.2 esitellään sanojen tekijät. Luvussa 2.3 käsitellään sanojen reunoja ja jaksoja sekä näiden yhteyttä sanojen limittymiseen. Lisäksi siinä esitellään tuloksia sanan lyhimmästä jaksosta ja pisimmästä reunattomasta tekijästä sekä näiden yhteyksiä toisiinsa. Luvussa 2.4 esitellään leksikografiset järjestykset ja maksimisuffiksit sekä Lyndonin sanat: Luvussa 2.4.1 todistetaan enimmäkseen luvuissa 4 ja 6 käytettäviä tuloksia. Luvun 2.4.2 tuloksia käytetään enimmäkseen luvussa 5. Esitys seuraa pääasiassa lähteitä [11] ja [3], mutta terminologiaa on tarvittaessa muokattu nykyistä käyttöä vastaavaksi. Samalla esitellään diagrammit, joilla voidaan havainnollistaa esimerkiksi todistuksissa ilmeneviä sanojen suhteita ilman, että ilmaisujen eksaktius vähenee.

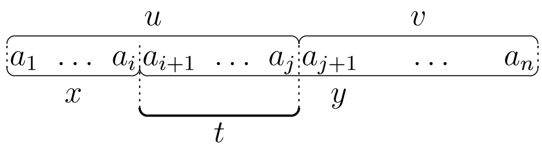
2.1 Sanat

Äärellistä symbolijoukkoa A kutsutaan *aakkostoksi*, ja sen alkioita kutsutaan *kirjaimiksi*. Kun aakkoston A kirjaimet $a_1, a_2, \ldots, a_n$ kirjoitetaan peräkkäin, saadaan *sana* $a_1 a_2 \ldots a_n$ yli aakkoston A. Nollasta kirjaimesta koostuvaa sanaa kutsutaan *tyhjäksi sanaksi*, ja siitä käytetään merkintää ε. Kaikkien sanojen joukosta yli aakkoston A käytetään merkintää A^*. Esimerkiksi a, ba ja $abaab$ ovat sanoja yli aakkoston $\{a, b\}$. *Sanan w aakkostolla* tarkoitetaan aakkostoa, joka koostuu sanassa w esiintyvistä kirjaimista. Esimerkiksi ba on sana myös yli aakkoston $\{a, b, c\}$, mutta sen aakkosto on $\{a, b\}$.

Tässä tutkielmassa muuttujat a, b, c ja d mahdollisten ylä- ja alaindeksien kanssa tarkoittavat kirjaimia, jos ei toisin mainita. Lisäksi muuttujat $i, \ldots, q$ ylä- ja alaindekseineen tarkoittavat epänegatiivisia kokonaislukuja. Muilla muuttujilla, esimerkiksi $f, g, h, p, q, r, \ldots, \alpha, \beta, \ldots$, tarkoitetaan sanoja.

Sanojen u ja v *katenaatio* uv on sana, joka saadaan kirjoittamalla ne peräkkäin:

$$uv = a_1 a_2 \ldots a_m b_1 b_2 \ldots b_n,$$



Kuva 2.1: Levin lemma, kun $|u| \geq |x|$. Sana $t = a_{i+1} \dots a_j$. Tapaus $|x| \geq |u|$ saadaan vaihtamalla u ja x sekä v ja y keskenään.

kun $u = a_1 a_2 \dots a_m$ ja $v = b_1 b_2 \dots b_n$. Esimerkiksi sanojen *ab* ja *ba* katenaatio on *abba*. Katenointi on assosiatiivinen operaatio: jos $u = a_1 a_2 \dots a_m$, $v = b_1 b_2 \dots b_n$ ja $w = c_1 c_2 \dots c_k$, niin

$$(uv)w = a_1 a_2 \dots a_m b_1 b_2 \dots b_n c_1 c_2 \dots c_k = u(vw).$$

Lisäksi $\varepsilon w = w\varepsilon = w$ kaikille sanoille w. Joukko A^* yhdessä katenoinnin kanssa on siis monoidi, jonka ykkösalkio on tyhjä sana ε. Sanan katenoinnista itsensä kanssa käytetään potenssimerkintää: kun $k \geq 0$, niin $w^k = ww \dots w$, missä w toistuu k kertaa.

Seuraavaa ominaisuutta kutsutaan *Levin lemmaksi*. Se on hyvin perustavanlaatuinen vapaan monoidin ominaisuus, joten sitä käytetään tässä tutkielmassa ilman erityistä mainintaa.

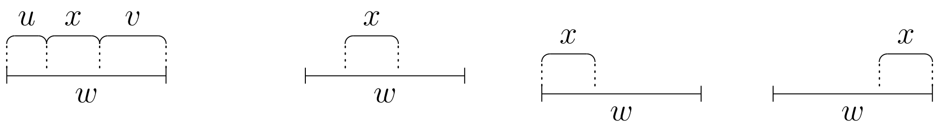
Lemma 2.1. Olkoon $uv = xy$. Tällöin on olemassa sana t, jolle

(i) $u = xt$ ja $y = tv$ tai

(ii) $x = ut$ ja $v = ty$.

Edellinen lemma voidaan perustella ilman tarkkaa todistusta kuvan 2.1 avulla. Jos $|u| \geq |x|$, niin $uv = a_1 \dots a_i a_{i+1} \dots a_j a_{j+1} \dots a_n$, missä $u = a_1 \dots a_j$, $v = a_{j+1} \dots a_n$, $x = a_1 \dots a_i$ ja $y = a_{i+1} \dots a_n$. Tällöin voidaan valita $t = a_{i+1} \dots a_j$.

Olkoon $w = a_1 a_2 \dots a_n$, missä jokainen a_i on kirjain. Sen *peilisana* $\overline{w} = a_n a_{n-1} \dots a_1$ saadaan kirjoittamalla sanan w kirjaimet takaperin. Esimerkiksi sanan *abaabab* peilisana on *babaaba* ja tyhjän sanan peilisana on tyhjä sana. Sanan w *pituus* $|w| = n$ on sen kaikkien kirjainesiintymien lukumäärä. Esimerkiksi sanan *abaabab* pituus on 7 ja tyhjän sanan pituus on 0.



(a) Tekijöihinjako $w = uxv$ (b) Yleinen tekijä (c) Prefiksi (d) Suffiksi

Kuva 2.2: Tekijöiden piirtäminen

Katenaation pituudelle pätee kaava $|uv| = |u| + |v|$, ja peilisanan pituudelle pätee kaava $|\overline{w}| = |w|$.

2.2 Tekijä, prefiksi ja suffiksi

Sana x on sanan w *tekijä*, jos on olemassa sellaiset sanat u ja v, että w voidaan kirjoittaa muodossa

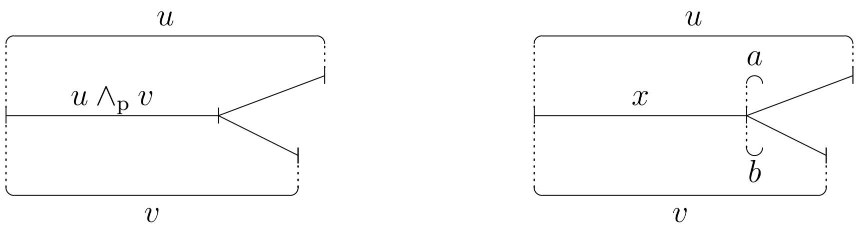
$$w = uxv. \tag{2.1}$$

Tämä tekijöihinjako on esitetty kuvassa 2.2a. Jos lisäksi $x \neq w$, niin x on sanan w *aito* tekijä.

Jos tekijöihinjaossa 2.1 $u = \varepsilon$, niin x on sanan w *prefiksi*. Tästä käytetään merkintää $x \leq_{\mathrm{p}} w$. Jos vastaavasti $v = \varepsilon$, niin x on sanan w *suffiksi*. Tästä käytetään merkintää $x \leq_{\mathrm{s}} w$. Jos lisäksi $x \neq w$, niin käytetään vastaavasti merkintöjä $x <_{\mathrm{p}} w$ ja $x <_{\mathrm{s}} w$. Sanaa x kutsutaan tällöin sanan w *aidoksi prefiksiksi* tai *aidoksi suffiksiksi*.

Erityisesti tyhjä sana ε on jokaisen sanan tekijä, prefiksi ja suffiksi, mutta ei ole itsensä aito tekijä, prefiksi eikä suffiksi. Sanan w tekijöiden joukosta käytetään merkintää $\mathrm{F}(w)$, prefiksien joukosta merkintää $\mathrm{Pref}(w)$ ja suffiksien joukosta merkintää $\mathrm{Suf}(w)$. Kuvissa 2.2b, 2.2c ja 2.2d on esitetty, miten tekijä, prefiksi ja suffiksi esitetään diagrammeina.

Sanojen u ja v *pisin yhteinen prefiksi* $u \wedge_{\mathrm{p}} v$ on pisin sana, joka on sekä sanan u että sanan v prefiksi. Tämä on esitetty kuvassa 2.3a. Pisin yhteinen prefiksi voidaan määritellä tarkemmin seuraavasti: Jos u on sanan v prefiksi, niin $u \wedge_{\mathrm{p}} v = u$, ja jos v on sanan u prefiksi, niin $u \wedge_{\mathrm{p}} v = v$. Jos kumpikaan sana ei ole toisen sanan prefiksi (kuva 2.3b), niin $u \wedge_{\mathrm{p}} v = x$, missä $x \leq_{\mathrm{p}} u, x \leq_{\mathrm{p}} v$ ja $xa \leq_{\mathrm{p}} u, xb \leq_{\mathrm{p}} v$, missä a ja b ovat eri kirjaimia. Sanojen u ja v



(a) Yleinen tapaus, jolloin sanat voivat olla toistensa prefiksejä.

(b) Kumpikaan sanoista u ja v ei ole toisen prefiksi.

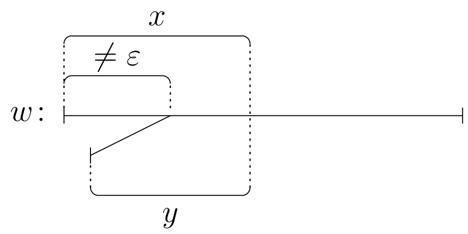
Kuva 2.3: Pisin yhteinen prefiksi

pisin yhteinen suffiksi määritellään vastaavasti, ja siitä käytetään merkintää $u \wedge_{\mathrm{s}} v$. Esimerkiksi $abaab \wedge_{\mathrm{p}} ababbaab = aba$ ja $abaab \wedge_{\mathrm{s}} ababbaab = baab$, sekä $abaab \wedge_{\mathrm{p}} bbaab = \varepsilon$.

Olkoon u sanan w prefiksi ja v sanan w suffiksi. Merkinnällä $u^{-1}w$ tarkoitetaan sanaa u', jolle $uu' = w$. Vastaavasti merkinnällä wv^{-1} tarkoitetaan sanaa v', jolle $v'v = w$. Sanaa $u^{-1}w$ kutsutaan sanan w *vasemmaksi osamääräksi* sanan u suhteen, ja sanaa wv^{-1} kutsutaan sanan w *oikeaksi osamääräksi* sanan v suhteen. Selvästi kumpikin osamäärä on yksikäsitteinen. Määritelmistä nähdään suoraan, että osamäärän pituus saadaan vähentämällä poistettavan sanan pituus: $|u^{-1}w| = |w| - |u|$ ja $|wv^{-1}| = |w| - |v|$.

Jos osamäärämerkintä on termin alussa tai lopussa, on ilmeistä, tarkoitetaanko kulloinkin vasenta vai oikeaa osamäärää. Termin keskellä olevissa osamäärämerkinnöissä käytetään tässä tutkielmassa aina sulkumerkkejä, jotta ilmenisi suoraan, kumpaa osamäärää kulloinkin tarkoitetaan. Muotoa $xy^{-1}z$ olevan merkinnän sijaan tässä tutkielmassa käytetään siis muotoa $(xy^{-1})z$ tai $x(y^{-1}z)$ olevaa merkintää.

Esimerkki 2.1. Kuva 2.4 esittää tapausta, jossa sanalla w on prefiksi x, joka ei ole sanan y suffiksi. Koska $x \leq_{\mathrm{p}} w$, niin $x^{-1}w$ on määritelty, ja edelleen katenaatio $y(x^{-1}w)$ on määritelty Koska $x \not\leq_{\mathrm{s}} y$, niin yx^{-1} ei ole määritelty. Ei siis ole yleisesti mahdollista käyttää muotoa $yx^{-1}w$ olevia merkintöjä, ellei x ole sekä sanan w prefiksi että sanan y suffiksi. Sanalle y on kuitenkin määritelty osamäärä $y(x \wedge_{\mathrm{s}} y)$. Tämä tarkoittaa kuvassa 2.4 sanan y vinolla viivalla merkittyä osaa. Tästä osasta alkava ja sanan w



Kuva 2.4: Monimutkaisempi sanadiagrammi

oikeaan reunaan päättyvä sana voidaan muodostaa kulkemalla diagrammia kyseisestä alkupisteestä loppupisteeseen viivoja pitkin, jolloin saadaan sana $y(x^{-1}w)$.

Prefikseillä ja suffikseilla on ominaisuus, että saman sanan prefiksit (vastaavasti suffiksit) ovat prefiksisuhteessa (vastaavasti suffiksisuhteessa) keskenään. Seuraavaksi kuvataan tämä ominaisuus tarkemmin prefiksirelaatiolle. Olkoot u ja v sanan w prefiksejä. Tällöin $w = u(u^{-1}w) = v(v^{-1}w)$. Jos $|u| \leq |v|$, niin Levin lemman mukaan $v = ut$ jollekin sanalle t, joten u on sanan v prefiksi. Jos taas $|v| \leq |u|$, niin $u = vt$ jollekin sanalle t, joten v on sanan u prefiksi.

Sanat u ja v *kommutoivat*, jos $uv = vu$. Kommutoinnille saadaan heti seuraava tulos.

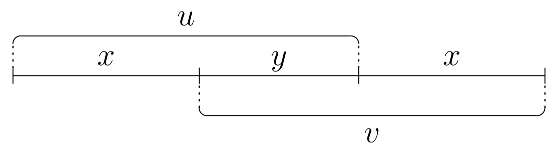
Lemma 2.2. Sanat u ja v kommutoivat jos ja vain jos ne ovat saman sanan potensseja.

Todistus. ($\Leftarrow$) Jos $u = x^i$ ja $v = x^j$, missä $i, j \geq 0$, niin $uv = x^{i+j} = vu$.

($\Rightarrow$) Oletetaan, että $uv = vu$. Todistetaan väite induktiolla sanan uv pituuden suhteen. Jos uv on tyhjä sana, niin u ja v ovat tyhjän sanan potensseja. Induktioaskeleessa voidaan olettaa, että $|u| \leq |v|$, koska $uv = vu$ on sama kuin $vu = uv$. Tällöin Levin lemman 2.1 mukaan on olemassa sana t, jolle $|t| \leq |v|$ ja jolle

$$v = ut = tu. \tag{2.2}$$

Jos $|t| = |v|$, niin $u = \varepsilon$ ja induktioväite on todistettu. Oletetaan siis, että $|t| < |v|$. Tällöin $|ut| < |uv|$, joten induktio-oletuksen mukaan u ja t ovat saman sanan potensseja: $u = x^i$ ja $t = x^j$, missä $i, j \geq 0$. Sijoittamalla nämä



Kuva 2.5: Sanat u ja v ovat konjugaatteja.

potenssit kaavaan 2.2 saadaan $v = x^{i+j}$. Sanat u ja v ovat siis kumpikin sanan x potensseja. $\square$

Sanat u ja v ovat *konjugaatteja*, jos $u = xy$ ja $v = yx$ joillekin sanoille x ja y. Tästä käytetään merkintää $u \sim v$. Voidaan myös sanoa, että v on sanan u *konjugaatti*. Esimerkiksi sanat *abaab* ja *aabab* ovat konjugaatteja. Konjugaatit voidaan merkitä diagrammeissa esimerkiksi kuvan 2.5 tavalla.

Tarkastellaan nyt sanan erisuurten konjugaattien lukumäärää. Tyhjällä sanalla on tarkalleen yksi konjugaatti, nimittäin itsensä. Olkoon nyt w epätyhjä sana. Koska $w = \varepsilon w = w\varepsilon$, voidaan sanan w konjugaatit esittää aina muodossa $u^{-1}wu$, missä $u <_{\mathrm{p}} w$. Jotkut näistä konjugaateista voivat olla yhtäsuuria, joten epätyhjällä sanalla on aina enintään pituutensa verran konjugaatteja. Seuraavaksi käsiteltävillä sanoilla on aina pituutensa verran eri konjugaatteja.

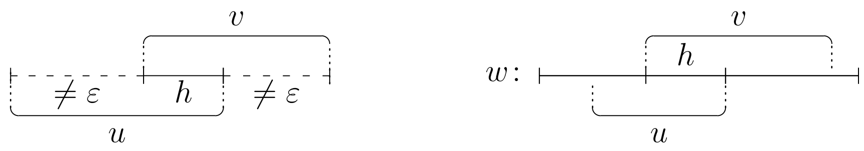
Sana w on *primitiivinen*, jos se ei ole muotoa $w = u^k$, missä $k \geq 2$. Toisin sanoen primitiivinen sana ei ole minkään muun sanan kuin itsensä potenssi. Esimerkiksi sana *ababa* on primitiivinen, mutta sana *ababab* ei ole. Tyhjä sana ei ole primitiivinen, koska $\varepsilon = \varepsilon^2$.

Seuraavan lemman mukaan joko sanan kaikki konjugaatit ovat primitiivisiä tai yksikään sanan konjugaateista ei ole primitiivinen.

Lemma 2.3. Primitiivisen sanan konjugaatit ovat primitiivisiä.

Todistus. Olkoon sana w primitiivinen. Tehdään vastaoletus, että $w = uv$ ja $vu = x^k$, missä $k \geq 2$. Tällöin $v = x^{k_1}x_1$ ja $u = x_2x^{k_2}$, missä $x_1x_2 = x$ ja $k_1 + k_2 = k - 1$. Sijoittamalla ensin u ja v ja sitten x saadaan

$$uv = x_2x^{k_2}x^{k_1}x_1 = x_2(x_1x_2)^{k_2}(x_1x_2)^{k_1}x_1 = (x_2x_1)^{k_2+k_1+1} = (x_2x_1)^k,$$



(a) Sanat u ja v limittyvät. Sanan h tulee olla epätyhjä.

(b) Sanojen u ja v esiintymät ovat limittäin sanassa w. Myös tällöin $h \neq \varepsilon$.

Kuva 2.6: Sanojen limittyminen

missä $k \geq 2$. Tämä on ristiriidassa sanan uv primitiivisyyden kanssa. $\square$

Seuraavan lemman mukaan primitiivisellä sanalla vain triviaalisti konjugoimalla saadaan alkuperäinen sana.

Lemma 2.4. Olkoon sana uv primitiivinen ja u sekä v epätyhjiä. Tällöin $vu \neq uv$.

Todistus. Tehdään vastaoletus, että $vu = uv$. Tällöin lemman 2.2 mukaan u ja v ovat saman sanan potensseja: $u = x^i$ ja $v = x^j$ jollekin sanalle x ja luvuille $i, j \geq 0$. Sijoittamalla saadaan nyt kaava $uv = x^{i+j}$. Koska uv on primitiivinen, on potenssi $i + j = 1$; nollas potenssi antaisi tyhjän sanan. Tämä on ristiriidassa sen kanssa, että sekä $i \geq 1$ että $j \geq 1$. $\square$

Lemman 2.3 mukaan primitiivisen sanan $a_1 a_2 \ldots a_n$ kaikki konjugaatit $(a_1 a_2 \ldots a_i)^{-1} a_1 a_2 \ldots a_n (a_1 a_2 \ldots a_i)$ ovat primitiivisiä. Lemman 2.4 mukaan nämä konjugaatit ovat lisäksi pareittain erisuuria. Primitiivisellä sanalla on siis pituutensa verran erisuuria konjugaatteja. Tätä primitiivisten sanojen ominaisuutta käytetään myöhemmin luvussa 2.4.2 ja edelleen luvussa 5.

2.3 Jaksollisuus, reuna ja limittyminen

Kuva 2.6a esittää yleisesti tilannetta, jossa jokin epätyhjä sana h on sanan u aito suffiksi ja sanan v aito prefiksi. Tätä kutsutaan limittymiseksi. Sanat u ja v *limittyvät*, jos jokin sanan u aito epätyhjä prefiksi h on sanan v aito epätyhjä suffiksi tai toisinpäin. Limittymisen määritelmässä ei siis yksilöidä, kummalta puolelta limittyminen tapahtuu. Tämä yksilöidään tarvittaessa kuvailemalla limittymistä tarkemmin. On tietenkin mahdollista, että sanat

Kuva 2.7: Sanan jakso p on toistuvan tekijän u pituus.

u ja v limittyvät useammalla tavalla. Esimerkiksi sanat *abaabaab* ja *baababba* voivat limittyä kolmella eri tavalla:

$$\begin{cases} \underline{a}baabaab, \\ baababb\underline{a}; \end{cases} \quad \begin{cases} abaabaa\underline{b}, \\ \underline{b}aababba; \end{cases} \quad \begin{cases} abaa\underline{baab}, \\ \underline{baab}abba. \end{cases}$$

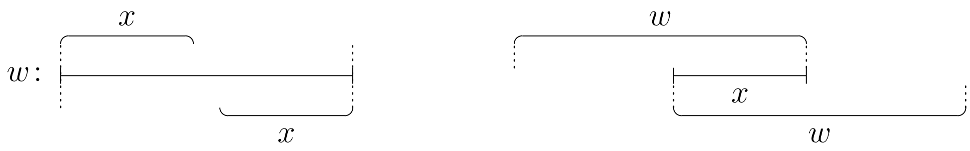
Limittymisellä tarkoitetaan myös myös kuvan 2.6b esittämää tilannetta. Tällöin sanalla w on tekijät u ja v, ja näiden tekijöiden tietyt esiintymät sanassa w ovat *limittäin*. Tällaisessa tapauksessa limittäin olevien tekijöiden esiintymät yksilöidään esimerkiksi kertomalla, mihin sanan w prefikseihin tai suffikseihin esiintymät rajoittuvat.

Positiivinen kokonaisluku p on sanan $w = a_1a_2 \dots a_n$ *jakso* eli *periodi*, jos $p \leq n$ ja $a_i = a_{i+p}$ kaikille $i = 1, \dots, n - p$. Jaksollisuus voidaan esittää kuvan 2.7 tavalla. Sanalla w voi olla useampia jaksoja, ja sen *pienimmästä jaksosta* käytetään merkintää $\pi(w)$. Määritellään lisäksi, että $\pi(\varepsilon) = 0$. Sanan pituus on aina sen jakso, joten $\pi(w) \leq |w|$.

Esimerkki 2.2. Tarkastellaan sanaa *abaabaab*. Koska $abaabaab = (aba)(aba)ab$, sillä on jakso 3. Sillä on myös jakso 6, koska $abaabaab = (abaaba)ab$. Lisäksi jaksona on sanan pituus 8. Näin ollen $\pi(abaabaab) = 3$.

Olkoon $x \leq_{\mathrm{p}} w$ ja $|x| = \pi(w)$. Tällöin sanan x konjugaatteja kutsutaan sanan w *syklisiksi juuriksi*. Sana x on primitiivinen, koska jos $x = y^k$, missä $k \geq 2$, niin $|y|$ olisi lyhintä jaksoa lyhyempi jakso. Lemman 2.3 mukaan siis kaikki sykliset juuret ovat primitiivisiä.

Esimerkki 2.3. Esimerkin 2.2 sanan *abaabaab* sykliset juuret ovat *aba*, *baa* ja *aab*. Voidaan varmistaa, että nämä ovat primitiivisiä.



(a) Reuna on sekä aito prefiksi että suffiksi.

(b) Reunan olemassaolo on yhtäpitävä sen kanssa, että w limittyy itsensä kanssa.

Kuva 2.8: Sanalla w on reuna. Tämä voidaan esittää kahdella tavalla.

Jos epätyhjä sana x on sanan w sekä aito prefiksi että aito suffiksi, kuten kuvassa 2.8a, niin x on sanan w *reuna*. Jos sanalla ei ole reunaa, niin sanotaan, että sana on *reunaton*. Tämä on yhtäpitävä sen kanssa, että sana ei limity itsensä kanssa, eli kuvan 2.8b tapaus ei ole mahdollinen.

Esimerkki 2.4. Olkoon $u = ababbaababb$ ja $v = ababaaababb$, jolloin sana v saadaan sanasta u vaihtamalla sen viides kirjain kirjaimesta a kirjaimeksi a. Sanalla u on tarkalleen yksi reuna $ababb$. Sana v on reunaton, vaikka se eroaa ensimmäisestä sanasta vain yhdellä kirjaimella. Sanan u lyhin jakso $\pi(u) = 6$. Sanan v lyhin jakso on $\pi(v) = 11$, joka on sanan v pituus. Lause 2.8 yhdistää tarkemmin sanan reunojen pituuksia lyhimpään jaksoon.

Seuraavaksi johdetaan sanan reunoihin liittyviä tuloksia. Ensimmäinen tulos kertoo perustavanlaatuisen ominaisuuden reunattomuudesta.

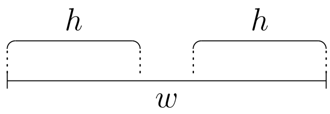
Lemma 2.5. Reunattomat sanat ovat primitiivisiä.

Todistus. Olkoon x epätyhjä sana. Jos nyt $x = y^n$, missä $n \geq 2$, niin y on sanan x reuna, koska y ei voi olla tyhjä sana. $\square$

Jos sana ei ole reunaton, niin sillä voi olla useita reunoja. Seuraavan lauseen mukaan yksi näistä reunoista on helpommin käsiteltävissä, mikä on hyödyksi esimerkiksi todistuksissa.

Lause 2.6 (Duval [5])**.** Sana h on sanan w lyhin reuna, jos ja vain jos h on sanan w reunaton reuna.

Todistus. ($\Rightarrow$) Olkoon h sanan w lyhin reuna. Tehdään vastaoletus, että sanalla h on reuna g. Tällöin g on sanan w aito prefiksi ja aito suffiksi, mikä on ristiriita sen kanssa, että h on sanan w lyhin reuna.



Kuva 2.9: Seurauksen 2.7 mukaan sanan w lyhin reuna h ei limity itsensä kanssa.

($\Leftarrow$) Olkoon nyt h reunaton sana, joka on sanan w reuna. Jos sanalla w olisi reuna g, joka on lyhyempi kuin h, niin g olisi myös sanan h reuna. Näin ollen h on sanan w lyhin reuna. $\square$

Koska reunaton sana ei voi limittyä itsensä kanssa (kuva 2.9), saadaan lauseesta 2.6 seuraava tulos.

Seuraus 2.7. Jos h on sanan w lyhin reuna, niin $w = hgh$ jollekin sanalle g.

Seuraava lause yhdistää sanan lyhimmän jakson ja sen pisimmän reunan. Sanan lyhin jakso saadaan sen mukaan ottamalla sen suurin limittyminen itsensä kanssa ja vähentämällä näin saadun limittyneen osan pituus sanan pituudesta. Tämä on esitetty kuvassa 2.10.

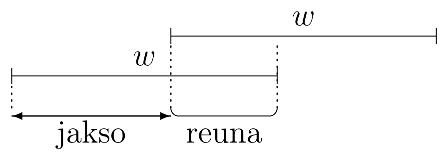
Lause 2.8. Sanalla w on reuna, jonka pituus on k, jos ja vain jos $|w| - k$ on sanan w jakso ja $k \geq 1$. Erityisesti, jos sana w ei ole reunaton, niin sen pisimmän reunan pituus on $|w| - \pi(w)$.

Todistus. ($\Rightarrow$) Olkoon $w = a_1 \ldots a_{|w|}$, ja olkoon h sanan w reuna, jonka pituus on k. Sana h voidaan kirjoittaa muodossa

$$h = a_1 \ldots a_k = a_{|w|-k+1} \ldots a_{|w|},$$

missä kukin a_i on kirjain. Tämä voidaan kirjoittaa muodossa $a_i = a_{i+|w|-k}$, kun $i = 1, \ldots, k$, joten $|w| - k$ on sanan w jakso.

($\Leftarrow$) Olkoon $w = a_1 \ldots a_{|w|}$, ja olkoon $a_i = a_{i+|w|-k}$ kaikille kirjaimille $a_i = 1, \ldots, a_k$. Tällöin $a_1 \ldots a_k = a_{k+1} \ldots a_{|w|}$, mikä tarkoittaa, että sana $a_1 \ldots a_k$ on sanan w prefiksi ja suffiksi eli reuna.



Kuva 2.10: Lauseen 2.8 mukainen yhteys reunan ja jakson välillä

Lopuksi johdetaan sanan pisimmän reunan pituus. Oletetaan, että sana w ei ole reunaton. Edellä todistetun mukaan sanalla w on reuna, jonka pituus on $|w| - \pi(w)$. Koska $\pi(w)$ on sanan w lyhin jakso, niin tämä reuna on lyhin. □

Lauseen 2.8 mukaan sana w on reunaton, jos ja vain jos $\pi(w) = |w|$.

Seuraavaksi esitellään kuvaukseen $\pi(w)$ sekä sanan w reunattomien tekijöiden pituuteen liittyviä epäyhtälöitä.

Sanan w pisimmän reunattoman tekijän pituudesta käytetään merkintää $\tau(w)$. Sana w on siis reunaton, jos ja vain jos $\tau(w) = |w|$. Sanan w pisimmän vähintään kaksi kertaa esiintyvän reunattoman tekijän pituudesta käytetään merkintää $\tau_2(w)$. Sovitaan lisäksi, että $\tau_2(w) = 0$, jos mikään sanan w reunaton tekijä ei esiinny kahdesti sanassa w. Näin ollen siis $0 \leq \tau_2(w) \leq \tau(w)$.

Itse asiassa $\tau_2(w) = 0$, jos ja vain jos jokainen sanan w kirjain esiintyy vain kerran: Jos mikään reunaton tekijä ei esiinny kahdesti sanassa w, niin myös mikään kirjain ei esiinny kahdesti sanassa w. Jos taas jokainen sanan w kirjain esiintyy vain kerran, niin mikään sanan w tekijä ei esiinny kaksi kertaa.

Esimerkki 2.5. Olkoon $w = abaabbaa$. Sen lyhin jakso on 7. Sen tekijä $abaabb$ on reunaton, ja kaikilla kolmella pidemmällä tekijällä on jokin reuna, joten $\tau(w) = |abaabb| = 6$. Tekijä baa on reunaton ja esiintyy kahdesti sanassa w, ja mikään pidempi reunaton tekijä ei esiinny kahdesti. Näin ollen $\tau_2(w) = |baa| = 3$. Tämä on kirjainten uudelleen nimeämistä ja peilausta vaille lyhin binäärinen sana, jolle $\tau(w) \neq \pi(w)$.

Jos sanalla on jokin reunaton tekijä, joka esiintyy kaksi kertaa, niin tämän reunattoman tekijän esiintymät eivät voi olla limittäin. Näin ollen saadaan seuraava raja suureelle $\tau_2(w)$.

Lause 2.9. $\tau_2(w) \leq \dfrac{|w|}{2}$.

Esimerkki 2.6. Kun esimerkissä 2.5 löydettiin reunaton tekijä *baa*, joka esiintyy kahdesti, riittää lauseen 2.9 mukaan tarkistaa, että mikään pituutta 4 oleva reunaton tekijä ei esiinny kahdesti.

Seuraava lemmassa on hyvin intuitiivinen tulos. Sitä käytetään lauseen 2.11 todistuksessa.

Lemma 2.10. $\pi(u) \leq \pi(w)$, kun u on sanan w epätyhjä tekijä.

Todistus. Olkoon $|u| \leq \pi(w)$. Tällöin väite saadaan johdosta

$$\pi(u) \leq |u| \leq \pi(w).$$

Olkoon $\pi(w) = p < |u|$. Olkoon $w = a_1 a_2 \dots a_n$ ja $u = a_i \dots a_j$, missä $1 \leq i < j \leq n$ ja $j \geq i + p$. Koska p on sanan w jakso, niin $a_k = a_{k+p}$ kaikille $k = i, \dots, j - p$. Näin ollen p on myös sanan u jakso, ja sanan u pienin jakso on pienempi tai yhtäsuuri kuin p. □

Sanan jakson ja pisimmän reunattoman tekijän pituuden välille saadaan nyt seuraava tulos.

Lause 2.11. $\tau(w) \leq \pi(w)$.

Todistus. Jos $\pi(w) = |w|$, niin $\tau(w) \leq |w| = \pi(w)$. Olkoon nyt $\pi(w) < |w|$, ja olkoon u sanan w tekijä, joka on pidempi kuin sanan w pienin jakso. Lemman 2.10 mukaan tekijän u pienin jakso on pienempi kuin tekijän u pituus, joten lauseen 2.8 mukaan u ei ole reunaton. □

Seuraava lause on Duvalin artikkelista [5]. Sen todistus sivuutetaan, koska todistusta varten täytyisi todistaa myös useampia teknisiä lemmoja, joita ei tarvita muualla tässä tutkielmassa.

Lause 2.12 (Duval [5])**.** Jos $\tau(w) = \tau_2(w)$, niin $\tau(w) = \pi(w)$.

(a) Sana u on sanan v aito prefiksi.

(b) Sana u ei ole sanan v prefiksi.

Kuva 2.11: Leksikografisen relaation $u \lhd v$ kuvaus riippuu siitä, onko u sanan v prefiksi.

Lauseen 2.11 epäyhtälö muodostaa yhdessä aiemmin mainittujen ilmiselvien epäyhtälöiden kanssa ketjun

$$0 \leq \tau_2(w) \leq \tau(w) \leq \pi(w) \leq |w|. \tag{2.3}$$

Lisäksi, jos $\tau_2(w) = 0$ tai $\pi(w) = |w|$, niin $\tau(w) = \pi(w)$. Lause 2.12 tarkoittaa nyt, että jos $\tau(w) < \pi(w)$, niin ketjun 2.3 epäyhtälöt ovat aitoja ja saadaan ketju

$$0 < \tau_2(w) < \tau(w) < \pi(w) < |w|. \tag{2.4}$$

2.4 Leksikografinen järjestys

Olkoon $\lhd$ aakkoston A täydellinen järjestysrelaatio. Tämä tarkoittaa, että jos a ja b ovat eri kirjaimia aakkoststa A, niin $a \lhd b$ tai $b \lhd a$. Relaatio $\lhd$ voidaan laajentaa joukon A^* *leksikografiseksi järjestykseksi* asettamalla, että $u \lhd v$, jos

(i) $u <_\mathrm{p} v$ tai

(ii) $(u \wedge_\mathrm{p} v)a \leq_\mathrm{p} u$ ja $(u \wedge_\mathrm{p} v)b \leq_\mathrm{p} v$, missä a ja b ovat eri kirjaimia, joille $a \lhd b$.

Tämä määritelmä on esitetty kuvassa 2.11. Merkintää $u \unlhd v$ käytetään, jos $u \lhd v$ tai $u = v$. Merkinnällä $\lhd^a$ tarkoitetaan aakkoston järjestystä, jolla kirjain a on maksimialkio. Vastaavasti merkinnällä $\lhd_b$ tarkoitetaan aakkoston järjestystä, jolla kirjain b on minimialkio.

Seuraavassa lemmassa on sääntöjä leksikografisessa järjestyksessä puolittain kertomiseen ja jakamiseen. Kohdat (i) ja (ii) ovat yleisesti tunnettuja ominaisuuksia, ja ne on mainittu muun muassa Lothairen teoksessa [11]. Kohta (iii) on lisätty näiden kahden mukaan, koska näin saadaan puuttuva oikealta jakamisen sääntö. Lemman jokainen kohta on myös voimassa, kun relaatio $\lhd$ korvataan relaatiolla $\unlhd$.

Lemma 2.13. Olkoot u, v, w, x ja y sanoja. Tällöin

(i) $u \lhd v$ jos ja vain jos $wu \lhd wv$,

(ii) jos $u \not\leq_{\mathrm{p}} v$ ja $u \lhd v$, niin $ux \lhd vy$,

(iii) jos $|u| \leq |v|$ ja $uw \lhd vw$, niin $u \lhd v$.

Todistus. (i) Koska $u <_{\mathrm{p}} v$ jos ja vain jos $wu <_{\mathrm{p}} wv$, riittää käsitellä tapaukset, joissa vasen puoli ei ole oikean puolen aito prefiksi.

Sana w on sanojen wu ja wv yhteinen prefiksi, joten $wu \wedge_{\mathrm{p}} wv = w(u \wedge_{\mathrm{p}} v)$.

Jos $(u \wedge_{\mathrm{p}} v)a \leq_{\mathrm{p}} u$ ja $(u \wedge_{\mathrm{p}} v)b \leq_{\mathrm{p}} v$, missä $a \lhd b$, niin $(wu \wedge_{\mathrm{p}} wv)a = w(u \wedge_{\mathrm{p}} v)a \leq_{\mathrm{p}} u$ ja $(wu \wedge_{\mathrm{p}} wv)b = w(u \wedge_{\mathrm{p}} v)b \leq_{\mathrm{p}} v$.

Jos $(wu \wedge_{\mathrm{p}} wv)a \leq_{\mathrm{p}} wu$ ja $(wu \wedge_{\mathrm{p}} wv)b \leq_{\mathrm{p}} wv$, missä $a \lhd b$, niin tällöin $w(u \wedge_{\mathrm{p}} v)a \leq_{\mathrm{p}} wu$ ja $w(u \wedge_{\mathrm{p}} v)b \leq_{\mathrm{p}} wv$, jolloin $(u \wedge_{\mathrm{p}} v)a \leq_{\mathrm{p}} u$ ja $(u \wedge_{\mathrm{p}} v)b \leq_{\mathrm{p}} v$.

(ii) Olkoon $(u \wedge_{\mathrm{p}} v)a \leq_{\mathrm{p}} u$ ja $(u \wedge_{\mathrm{p}} v)b \leq_{\mathrm{p}} v$, missä $a \lhd b$. Koska $u \leq_{\mathrm{p}} ux$ ja $v \leq_{\mathrm{p}} vy$, saadaan $ux \wedge_{\mathrm{p}} vy = u \wedge_{\mathrm{p}} v$. Näin ollen $(ux \wedge_{\mathrm{p}} vy)a = (u \wedge_{\mathrm{p}} v)a \leq_{\mathrm{p}} u \leq_{\mathrm{p}} ux$ ja $(ux \wedge_{\mathrm{p}} vy)b = (u \wedge_{\mathrm{p}} v)b \leq_{\mathrm{p}} v \leq_{\mathrm{p}} vy$.

(iii) Tehdään vastaoletus, että $v \unlhd u$. Jos $u = v$, niin $uw = vw$, mikä on ristiriidassa oletuksen $uw \lhd vw$ kanssa. Näin ollen $|u| < |v|$, mikä tarkoittaa, että $v \not\leq_{\mathrm{p}} u$. Jos nyt $v \lhd u$, niin kohdan (ii) mukaan $vw \lhd uw$, mikä on ristiriita. □

2.4.1 Maksimisuffiksit

Sanan w suffiksi x on sanan w $\lhd$-*maksimisuffiksi*, jos $y \unlhd x$ kaikille sanan w suffikseille y. Koska sanan jokainen suffiksi on eri sana ja vain samat sanat ovat leksikografisesti yhtäsuuria, on sanan maksimisuffiksi tietyn leksikografisen järjestyksen suhteen aina yksikäsitteinen.

Esimerkki 2.7. Olkoon $w = abaabbaa$ ja olkoon $b \lhd a$. Tällöin sanan w $\lhd$-maksimisuffiksi on $aabbaa$.

Seuraavia tuloksia käytetään erityisesti luvun 6 päättelyissä. Kaksi ensimmäistä tulosta yhdistävät joitakin samanlaisina toistuvia päättelyitä, mutta niitä ei ollut mainittu luvun 6 lähteessä [10].

Lemma 2.14. Olkoon α sanan w $\lhd$-maksimisuffiksi, missä $\lhd$ on jokin leksikografinen järjestys. Jos $\alpha \lhd x$, niin x ei ole sanan w tekijä.

Todistus. Jos x olisi sanan w tekijä, niin olisi olemassa sellainen sana y, että xy on sanan w suffiksi. Koska $\alpha \lhd x$ ja $x \unlhd xy$, saadaan, että $\alpha \lhd xy$. Tämä on ristiriidassa sanan α maksimaalisuuden kanssa. $\square$

Lause 2.15. Olkoon $\lhd$ mielivaltainen leksikografinen järjestys, α sanan w $\lhd$-maksimisuffiksi ja x sanan α epätyhjä prefiksi. Tällöin sana $x\alpha$ ei ole sanan w tekijä.

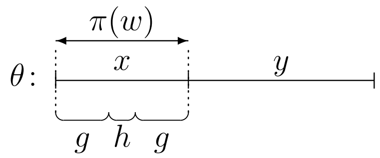
Todistus. Olkoon $\alpha = xy$. Tällöin $y \lhd xy$, koska xy on $\lhd$-maksimisuffiksi ja x on epätyhjä. Lemman 2.13 mukaan nyt $xy \lhd xxy$. Lemman 2.14 mukaan xxy ei ole sanan w tekijä. koska xy on sanan w $\lhd$-maksimisuffiksi. $\square$

Seuraavat kaksi tulosta ovat yleistä tietoa, eikä niiden ensimmäisestä todistuksesta ole varmaa tietoa. Tässä esitettävät todistukset seuraavat Holubia ja Nowotkaa [10]. Ensimmäinen tulos on työkaluna vastaoletuksella todistamisessa.

Lause 2.16. Olkoon $\lhd$ mielivaltainen leksikografinen järjestys. Sanan $\lhd$-maksimisuffiksi esiintyy tarkalleen kerran sanan tekijänä. Se on lisäksi pidempi kuin sanan pisin reuna, jos sana ei ole reunaton.

Todistus. Olkoon α sanan w $\lhd$-maksimisuffiksi, ja olkoon $w = x\alpha y$ joillekin sanoille x ja y. Tällöin $\alpha \unlhd \alpha y$, koska α on sanan αy prefiksi. Toisaalta $\alpha y \unlhd \alpha$, koska α on maksimisuffiksi, joten y on tyhjä sana. Näin ollen α esiintyy tarkalleen kerran sanassa w.

Tehdään vastaoletus, että α on lyhyempi tai yhtä pitkä kuin sanan w pisin reuna, jonka pituus on lauseen 2.8 mukaan $|w| - \pi(w)$. Tällöin sanalla



Kuva 2.12: Lauseen 2.17 todistuksen alku

w on tekijöihinjako $w = uv\alpha$, missä $|u| = \pi(w)$. Nyt $v\alpha \leq_{\mathrm{p}} w$, mikä on ristiriidassa edellisen kohdan kanssa. $\square$

Luvussa 2.3 todettiin, että sanan lyhimmän jakson pituinen prefiksi on aina primitiivinen. Seuraava lause antaa yhden tapauksen, jolloin tämä prefiksi on myös reunaton.

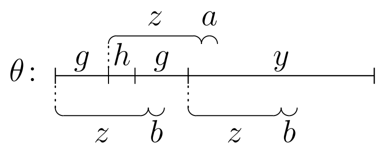
Lause 2.17. Olkoon θ itsensä $\lhd$-maksimisuffiksi jonkin leksikografisen järjestyksen $\lhd$ suhteen. Tällöin sanan θ pienimmän jakson pituinen prefiksi on reunaton.

Todistus. Olkoon x sanan θ pienimmän jakson pituinen prefiksi. Tehdään vastaoletus, että sanalla x on reuna. Olkoon g sellainen sanan w reuna, että $x = ghg$ jollekin sanalle h; seurauksen 2.7 mukaan tällainen reuna on olemassa. Olkoon $y = x^{-1}\theta$, jolloin $\theta = xy = ghgy$, kuten kuvassa 2.12. Koska gy on sanan θ suffiksi, saadaan $gy \unlhd \theta = ghgy$, ja edelleen

$$y \unlhd hgy. \tag{2.5}$$

Jos gy on sanan θ prefiksi, niin se on sanan θ reuna, jolloin lauseen 2.6 mukaan sanalla θ on jakso $|gh|$. Tämä on ristiriidassa sen kanssa, että $|x|$ on sanan θ pienin jakso. Näin ollen gy ei ole sanan θ prefiksi.

Olkoon nyt $z = y \wedge_{\mathrm{p}} hgy$. Koska gy ei ole sanan θ prefiksi, niin y ei ole sanan hgy prefiksi. Näin ollen $zb \leq_{\mathrm{p}} y$ ja $za \leq_{\mathrm{p}} hgy$, missä a ja b ovat eri kirjaimia. Tämä tilanne on esitetty kuvassa 2.13. Nyt relaation (2.5) mukaan pätee relaatio $b \lhd a$. Koska $|x| = \pi(\theta)$ ja $\theta = xy$, on y sanan θ prefiksi. Nyt $zb \leq_{\mathrm{p}} y \leq_{\mathrm{p}} \theta$, ja koska $b \lhd a$, saadaan relaatio $\theta \lhd za \leq_{\mathrm{p}} hgy$. Tämä on ristiriita, koska hgy on sanan θ suffiksi ja θ on itsensä $\lhd$-maksimisuffiksi. $\square$



Kuva 2.13: Lauseen 2.17 todistuksen loppu

2.4.2 Lyndonin sanat

Olkoon $\lhd$ jokin aakkoston täydellinen järjestys. Primitiivinen sana on *Lyndonin sana* järjestyksen $\lhd$ suhteen, jos se on leksikografisesti pienin kaikista konjugaateistaan järjestyksen $\lhd$ suhteen. Kun tässä tutkielmassa puhutaan yleisesti *Lyndonin sanoista* mainitsematta järjestystä, tarkoitetaan Lyndonin sanoja minkä tahansa järjestyksen suhteen.

Koska leksikografinen järjestys on täydellinen järjestys, primitiivisellä sanalla on aina yksikäsitteinen konjugaatti, joka on Lyndonin sana järjestyksen $\lhd$ suhteen. Tämä sana on *Lyndonin konjugaatti* $\lhd$ suhteen.

Seuraava lemma todistetaan yleensä todistamalla vaihtoehtoinen Lyndonin sanojen karakterisaatio. Tässä kuitenkin lemma todistetaan käyttämättä tätä katakterisaatiota. Todistus on Ehrenfeuchtin ja Silbergerin artikkelista [6] muuteltu versio.

Lemma 2.18. Lyndonin sanat ovat reunattomia.

Todistus. Olkoon sana w Lyndonin sana järjestyksen $\lhd$ suhteen. Tehdään vastaoletus, että h on sanan w lyhin reuna. Tällöin seurauksen 2.7 mukaan $w = hgh$, missä $h \neq \varepsilon$. Koska hhg on sanan w konjugaatti, saadaan $hgh = w \lhd hhg$, joten lemman 2.13 (i) mukaan $gh \lhd hg$. Tässä gh ei voi olla sanan hg aito prefiksi, koska gh ja hg ovat yhtä pitkiä. Näin ollen lemman 2.13 (ii) mukaan $ghh \lhd hgh = w$. Tämä on ristiriita, koska sana ghh on sanan w konjugaatti. □

Seuraava lemma on yleistä tietoa. Se on mainittu ainakin Ehrenfeuchtin ja Silbergerin artikkelissa [6], mutta väitettä on muokattu tässä vahvemmaksi.

Lemma 2.19 (Ehrenfeucht, Silberger [6]). Olkoon w primitiivinen sana, a jokin siinä esiintyvä kirjain ja $\lhd_a$ jokin aakkoston järjestys, jossa a on minimaalinen. Tällöin jokin sanan w konjugaatti on Lyndonin sana järjestyksen $\lhd_a$ suhteen, ja tämän konjugaatin ensimmäinen kirjain on a.

Todistus. Edellä on mainittu, että primitiivisellä sanalla on aina Lyndonin konjugaatti minkä tahansa järjestyksen suhteen. Olkoon x sanan w Lyndonin konjugaatti järjestyksen $\lhd_a$ suhteen. Todistetaan, että sanan x ensimmäinen kirjain on a. Sanalla w on jokin konjugaatti y, jonka ensimmäinen kirjain on a. Jos sanan x ensimmäinen kirjain ei ole a, niin $y \lhd x$, mikä on ristiriita, koska x on Lyndonin sana järjestyksen $\lhd_a$ suhteen. Näin ollen sanan x ensimmäinen kirjain on a. $\square$

Edellisen lemman mukaan jokaisella primitiivisellä sanalla, jonka aakkoston koko on k, on ainakin k konjugaattia, jotka ovat Lyndonin sanoja.

3 Assous'n ja Pouzet'n optimaalinen esimerkki

Assous ja Pouzet [1] esittivät tässä luvussa käsiteltävän sanan vastaesimerkkinä Ehrenfeuchtin ja Silbergerin [6] esittämään otaksumaan, että jos $|w| \geq 2\,\tau(w)$, niin $\tau(w) = \pi(w)$. Heidän esittämänsä sana on oikeammin ääretön joukko sanoja w, joille on voimassa sekä $\tau(w) < \pi(w)$ että $|w| > 2\,\tau(w)$. Assous ja Pouzet esittivät lisäksi, millä menetelmällä he olivat päätyneet tähän vastaesimerkkiin. Menetelmässä aloitetaan sopivasta sanasta, jota yritetään jatkaa vasemmalta tai oikealta kasvattamatta pisimmän reunattoman tekijän pituutta. Jos sanaa ei voi pidentää tällä tavalla, niin seuraavaksi sen kirjaimien tilalle sijoitetaan mielivaltaisen pitkiä sanoja, jotka on valittu sopivalla tavalla, jotta alkuperäisen sanan rakenne säilyy. Tätä uutta mielivaltaisen pitkää sanaa yritetään uudestaan jatkaa kasvattamatta pisimmän reunattoman tekijän pituutta. Seuraavassa esitetään tämä menetelmä. Sanojen pisimpiä reunattomia tekijöitä merkitään sanadiagrammeista tutuilla kaarilla ja pisteellä sanan kirjainten välissä merkitään sanan lyhintä jaksoa.

Aloitetaan sanasta

$$w_1 = abaabba.aba, \tag{3.1}$$

jonka pituus on 10 ja $\pi(w_1) = 7$. Sen pisimmät reunattomat tekijät ovat $abaabb$ ja $bbaaba$, joten $\tau(w_1) = 6$. Jos tätä sanaa jatketaan oikealta kirjaimella a, saadaan sana $abaabbaabaa$, jolla on reunaton tekijä $bbaabaa$. Vastaavasti jos sanaa jatketaan oikealta kirjaimella b, saadaan sana $abaabbaabab$, jolla on reunaton tekijä $bbaabab$. Kummankin tekijän pituus on 7, joten sanaa $abaabbaaba$ ei voi jatkaa ilman, että funktion τ arvo ei kasvaisi. Toisin sanoen jokaisella sanalla w', jolla on sana $abaabbaaba$ aitona prefiksinä tai suffiksina, on $\tau(w') \geq 7$.

Vaihdetaan nyt samanaikaisesti jokainen kirjain b kirjaimeen a ja jokainen kirjain a sanaan ba^n, missä $n \geq 1$. Näin saadaan sana

$$w_2 = ba^naba^nba^naaba^n.ba^naba^n, \tag{3.2}$$

jolle $\tau(w_2) = 3n+6$, $\pi(w_2) = |ba^naba^nba^naaba^n| = 4n+7$ ja $|w_2| = 6n+10$. Tätä sanaa voidaan pidentää muuttamatta arvoja $\tau(w)$ ja $\pi(w)$: Funktion π

arvo pysyy samana, kun katenoidaan vasemmalta sanan $ba^naba^nba^naaba^n$ suffiksilla. Jos katenoidaan sanalla ba^n, niin saadulla sanalla on prefiksi $ba^nba^{n+1}ba^nba^{n+2}$, joka on reunaton ja jonka pituus $4n+7 > 3n+6$. Näin ollen mikään pidempi sana kuin a^n ei käy. Katenoimalla vasemmalta sanalla a^n saadaan sana

$$w_{\mathrm{AP}} = a^nba^{n+1}ba^nba^{n+2}b.a^nba^{n+1}ba^n. \tag{3.3}$$

Tälle sanalle vieläkin $\tau(w_{\mathrm{AP}}) = 3n+6$ ja $|w_{\mathrm{AP}}| = 7n+10$. Tätä sanaa ei voi enää pidentää muuttamatta funktioiden τ ja π arvoja: Jatkamalla oikealta kirjaimella a päädytään sanaan $a^nba^{n+1}ba^nba^{n+2}ba^nba.a^nba^{n+1}$, jonka lyhin jakso on $5n+9$. Jatkamalla oikealta kirjaimella b päädytään sanaan $a^nba^{n+1}ba^nba^{n+2}b.a^nba^{n+1}ba^nb$, jonka lyhin jakso on $4n+7$, mutta jolla on reunaton tekijä $a^{n+2}ba^nba^{n+1}ba^nb$, jonka pituus $4n+7 > 3n+6$.

Sanaa w_1 on nyt muutettu lisäämällä tiettyjen kirjaimien väliin ja sanan alkuun ja loppuun lisää kirjaimia. Saadun sanan w_{AP} rakenne ei ole muuttunut, sillä $|w_{\mathrm{AP}}|$, $\pi(w_{\mathrm{AP}})$ ja $\tau(w_{\mathrm{AP}})$ ovat vakio-osiltaan samat kuin sanalla w_1. Lisäksi sana w_1 on eräs w_{AP}-sana: jos $n=0$, niin $w_{\mathrm{AP}} = babbaabbab$, joka on kirjainten vaihtoa vaille sana w_1.

Sana w_{AP} on merkittävä luvun 6 päätuloksen kannalta, koska

$$|w_{\mathrm{AP}}| = 7n+10 = 7(n+2)-4 = \frac{7}{3}\cdot 3(n+2) - 4 = \frac{7}{3}\,\tau(w_{\mathrm{AP}}) - 4.$$

Kerroin $\frac{7}{3}$ on sama kuin tämän tutkielman päätuloksen lauseen 6.15 kerroin. Koska $\tau(w_{\mathrm{AP}}) \neq \pi(w_{\mathrm{AP}})$, sana w_{AP} antaa alarajan lauseen 6.15 ehdolle.

Tarkastellaan seuraavaksi, minkälainen sanan tulisi olla, jotta se antaisi lauseen 6.15 ehdolle alarajan. Olkoon w sana, jolle

(i) $\tau(w) < \pi(w)$ ja

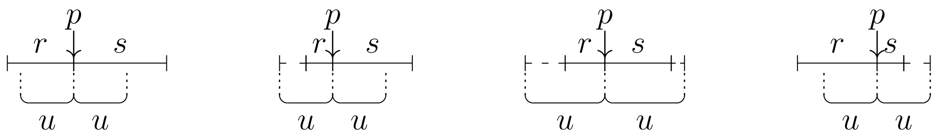
(ii) $|w| = \frac{7}{3}\,\tau(w) - r$, missä $r > 0$.

Ehdon (ii) mukaan $3|w| = 7\,\tau(w) - 3r$. Koska $\mathrm{syt}(3,7) = 1$, ratkaisut ovat muotoa

$$|w| = 7n + k, \tag{3.4}$$

$$\tau(w) = 3n + m, \tag{3.5}$$

missä $k, m, n \geq 0$.



(a) $u \leq_{\mathrm{s}} r$ ja $u \leq_{\mathrm{p}} s$ (b) $r \leq_{\mathrm{s}} u$ ja $u \leq_{\mathrm{p}} s$ (c) $r \leq_{\mathrm{s}} u$ ja $s \leq_{\mathrm{p}} u$ (d) $u \leq_{\mathrm{s}} r$ ja $s \leq_{\mathrm{p}} u$

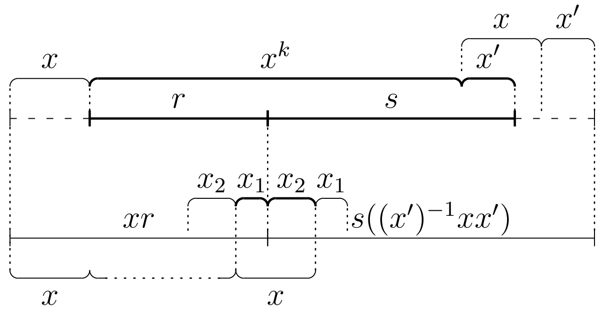
Kuva 4.1: Sanan w toistosana u pisteessä p voi sijoittua neljällä eri tavalla sanan w suhteen riippuen siitä, miten sana u on prefiksi- ja suffiksirelaatioissa sanojen r ja s kanssa.

4 Kriittinen tekijöihinjako

Epätyhjän sanan $w = a_1 a_2 \ldots a_n$ *pisteellä* p tarkoitetaan kohtaa kirjaimien a_p ja a_{p+1} välillä. Toisin sanoen p on kokonaisluku, jolle $1 \leq p \leq |w| - 1$. Sana u on sanan w *toistosana* pisteessä p, jos $w = rs$, missä $|s| = p$, ja lisäksi $u \leq_{\mathrm{s}} r$ tai $r \leq_{\mathrm{s}} u$, ja $s \leq_{\mathrm{p}} t$ tai $s \leq_{\mathrm{p}} u$. Tämä on esitetty kuvassa 4.1. Toistosanan u pituutta kutsutaan tällöin sanan w *lokaaliksi* tai *paikalliseksi jaksoksi* pisteessä p. Sanan w *pienimmästä lokaalista jaksosta* pisteessä p käytetään merkintää $\pi(w, p)$.

Esimerkki 4.1. Tarkastellaan esimerkissä 2.5 käsiteltyä sanaa $w = abaabbaa$. Kokeilemalla nähdään, että lyhimmät toistosanat sanan kussakin pisteessä $1, \ldots, 7$ ovat $ba, aab, a, bbaabaa, b, aabb$ ja a, joten sanan w lyhimmät paikalliset jaksot ovat vastaavasti $2, 3, 1, 7, 1, 4$ ja 1.

Lyhin toistosana on aina reunaton. Tämä nähdään siitä, että toistosanan reuna kelpaa aina lyhyemmäksi toistosanaksi samassa pisteessä. Lyhin toistosana voi kuitenkin olla pidempi kuin mikään sanan reunaton tekijä: Kuvan 4.1 tapauksissa (a), (b) ja (d) lyhin toistosana on sanan tekijä. Koska se on reunaton, niin lauseen 2.11 mukaan se on enintään sanan pienimmän jakson pituinen. Kuvan 4.1c tapauksessa taas toistosana sijoittuu niin, että kumpikaan toistosanan esiintymä ei ole sanan sisällä. Lyhin toistosana ei tässä tapauksessa voi olla sanan tekijä, koska muuten se limittyisi itsensä kanssa, mikä on mahdotonta, koska lyhin toistosana on reunaton. Tällainen tapaus on esimerkkien 2.5 ja 4.1 sanan $abaa.bbaa$ pisteessä 4. Pienin paikallinen jakso tässä pisteessä on suurempi kuin minkään reunattoman tekijän pituus.



Kuva 4.2: Sanalla on aina jakson pituinen toistosana missä tahansa pisteessä.

Lauseen 2.11 mukaan sanan reunattomat tekijät ovat aina enintään lyhimmän jakson pituisia. Seuraavaksi johdetaan samanlainen relaatio lyhimmille toistosanoille.

Olkoon x sanan w prefiksi, jonka pituus on jakso q. Tällöin $w = x^k x'$, missä $k \geq 1$ ja x' on sanan x aito prefiksi. Sana w on sanan $x^{k+1}x'$ aito prefiksi, joten sana xw on sanan $x^{k+2}x'$ aito prefiksi. Olkoon p sanan w piste, ja olkoon $w = rs$, missä $|r| = p$. Tämä tilanne on kuvassa 4.2. Olkoon $x = x_1x_2$, missä $|x_1| = p \pmod{|x|}$. Nyt x_2x_1 on sanan xr suffiksi ja sanan $s((x')^{-1}(xx'))$ prefiksi, joten x_2x_1 on sanan $x^{k+2}x'$ toistosana pisteessä $p + |x|$. Nyt sana x_2x_1 on sanan w toistosana pisteessä p. Näin ollen sanalla w on aina minkä tahansa jakson pituinen toistosana pisteessä p. Lyhimmälle toistosanan pituudelle ja pienimmälle jaksolle saadaan siis epäyhtälö

$$\pi(w, p) \leq \pi(w). \tag{4.1}$$

Jos epäyhtälön sijaan pätee yhtäsuuruus, sanotaan, että p on sanan w *kriittinen piste*. Tällöin tekijöihinjakoa $w = uv$, missä $|u| = p$, sanotaan *kriittiseksi tekijöihinjaoksi*.

Esimerkki 4.2. Esimerkeissä 2.5 ja 4.1 todettiin, että sanan *abaabbaa* lyhin jakso on 7, joka on myös lyhin paikallinen jakso pisteessä 4. Näin ollen piste 4 on sanan *abaabbaa* kriittinen piste.

Jos sanassa w esiintyy vain yhtä kirjainta a, niin $\pi(w) = 1$ ja lyhin toistosana joka pisteessä on a. Tällöin siis jokainen piste on kriittinen.

Kriittisen tekijöihinjaon lause kertoo, että jokaisella epätyhjällä sanalla on vähintään yksi kriittinen tekijöihinjako. Tässä esitetään kriittisen tekijöihinjaon lause sellaisessa muodossa, joka antaa erään kriittisen tekijöihinjaon käyttäen maksimisuffikseja jonkin leksikografisen järjestyksen suhteen. Sen todistus on Harjun ja Nowotkan [7] lyhyempi versio Crochemoren ja Perrinin [4] todistuksesta.

Lause 4.1 (Kriittinen tekijöihinjako)**.** Olkoon w epätyhjä sana, jossa esiintyy vähintään kahta eri kirjainta. Olkoon γ lyhyempi sen $\lhd$- ja $\blacktriangleleft$-maksimisuffikseista, missä $\lhd$ on jokin aakkoston järjestys ja $\blacktriangleleft$ sen käänteisjärjestys. Tällöin $|w| - |\gamma|$ on sanan w kriittinen piste.

Todistus. Olkoon $\lhd$ aakkoston järjestys ja $\blacktriangleleft$ sen käänteisjärjestys. Koska sanassa w esiintyy ainakin kahta eri kirjainta, niin sen $\lhd$- ja $\blacktriangleleft$-maksimisuffiksit alkavat eri kirjaimilla ja ovat siis erisuuria ja siten eripituisia. Koska järjestyksen $\blacktriangleleft$ käänteisjärjestys on $\lhd$, voidaan olettaa, että sanan w $\lhd$-maksimisuffiksi on lyhyempi kuin $\blacktriangleleft$-maksimisuffiksi. Tällöin γ on sanan w $\lhd$-maksimisuffiksi. Olkoon δ sanan w $\blacktriangleleft$-maksimisuffiksi. Tällöin $\delta = u'\gamma$ jollekin epätyhjälle sanalle u'.

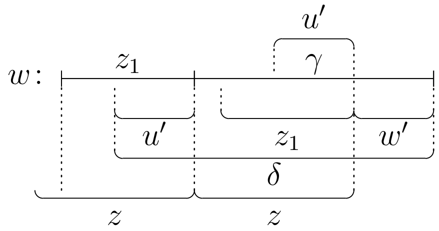
Epäyhtälön (4.1) mukaan riittää todistaa, että $\pi(w, |w| - |\gamma|) \geq \pi(w)$. Olkoon z sanan w sellainen pisteessä $|w| - |\gamma|$ oleva reunaton toistosana, jolle $|z| \leq |w|$. Tällainen toistosana on aina olemassa, koska lyhin toistosana toteuttaa molemmat ehdot. Jos $|z|$ on jokin sanan w jakso, niin erityisesti lyhimmän toistosanan pituus pisteessä $|w| - |\gamma|$ on myös jakso, mikä tarkoittaa, että $\pi(w, |w| - |\gamma|) \geq \pi(w)$. Riittää siis todistaa seuraava väite.

Väite 4.1.1. $|z|$ on jokin sanan w jakso.

Väite todistetaan tarkastelemalla kuvan 4.1 tapauksia erikseen. Kuvien 4.1c ja 4.1b tapauksissa väite todistetaan oikeaksi ja kuvien 4.1d ja 4.1a tapauksissa päädytään ristiriitaan.

(c) Kuvan 4.1c tilanteessa w on sanan z^2 tekijä. Koska z on reunaton, niin lauseen 2.8 mukaan $\pi(z) = |z|$. Edelleen lemman 2.10 mukaan $\pi(z^2) = |z|$. Samoin lemman 2.10 mukaan, koska w on sanan z^2 tekijä, on $|z|$ sanan w jakso.

(b) Olkoon $w = z_1\gamma$, missä $z_1 \leq_{\mathrm{s}} z$, $\gamma = zw'$ ja $w' \neq \varepsilon$. Kun kuvaan lisätään sana δ, saadaan kuva 4.3. Koska $\delta = u'\gamma$ ja $w = z_1\gamma$ ja δ on sanan



Kuva 4.3: Tapaus, jossa toistosana z sijoittuu kuvan 4.1b mukaisesti.

w suffiksi, niin u' on sanan z_1 ja edelleen sanan z suffiksi. Näin ollen sana $u'w'$ on sanan zw' suffiksi, ja koska $zw' = \gamma$, joka on sanan δ suffiksi, sana $u'w'$ on sanan δ suffiksi. Koska δ on $\blacktriangleleft$-maksimisuffiksi, saadaan $u'w' \underline{\blacktriangleleft} \delta = u'\gamma$ ja edelleen $w' \underline{\blacktriangleleft} \gamma$. Toisaalta, koska w' on sanan w suffiksi ja γ on $\lhd$-maksimisuffiksi, saadaan $w' \trianglelefteq \gamma$. Näin ollen w' on sanan γ prefiksi. Koska w' on myös sanan γ aito ja epätyhjä suffiksi, se on sanan γ reuna. Nyt lauseen 2.8 mukaan sanalla w on jakso $|\gamma| - |w'| = |z|$.

(d) Olkoon $w\gamma^{-1} = yz$, missä $y \in A^*$ ja $\gamma \leq_{\mathrm{p}} z$. Koska z on sanan w tekijä, sen prefiksi γ esiintyy kaksi kertaa sanan w tekijänä. Tämä aiheuttaa ristiriidan, koska lauseen 2.16 mukaan $\lhd$-maksimisuffiksi γ esiintyy tarkalleen kerran sanassa w.

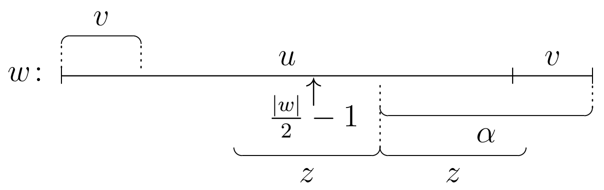
(a) Olkoon $w\gamma^{-1} = yz$, missä $y \in A^*$ ja $z \leq_{\mathrm{p}} \gamma$. Olkoon $\gamma' = z^{-1}\gamma$, jolloin $\gamma = z\gamma'$. Koska γ on $\lhd$-maksimisuffiksi, niin $z\gamma \trianglelefteq \gamma$. Koska tämän relaation vasen puoli on $z^2\gamma'$ ja oikea puoli $z\gamma'$, saadaan relaatio $z\gamma' \trianglelefteq \gamma'$, joka on ristiriidassa sanan $\gamma = z\gamma'$ maksimaalisuuden kanssa. $\square$

Tarkastellaan nyt sanan w lyhintä toistosanaa x jossakin kriittisessä pisteessä. Tällöin $|x| = \pi(w)$. Jos x on sanan w tekijä, niin lauseen 2.11 mukaan

$$\tau(w) \geq |x| = \pi(w) \geq \tau(w),$$

jolloin $\tau(w) = \pi(w)$. Toistosana on aina tekijänä, kun kuvan 4.1c tapaus ei ole kyseessä. Tämä toteutuu, kun $|x| < \frac{|w|}{2} + 1$. Koska $|x| = \pi(w)$, on nyt johdettu seuraava raja sanan pituudelle sen lyhimmän jakson suhteen, jonka yläpuolella pisin reunaton tekijä on lyhimmän jakson pituinen.

Seuraus 4.2. Jos $|w| > 2\,\pi(w) - 2$, niin $\tau(w) = \pi(w)$.



Kuva 4.4: Sana α on pidempi kuin sana v. Sana z on sanan w toistosana pisteessä $|w| - |\alpha|$.

Seuraava esimerkki kertoo, että tämän rajan kerroin 2 on optimaalinen.

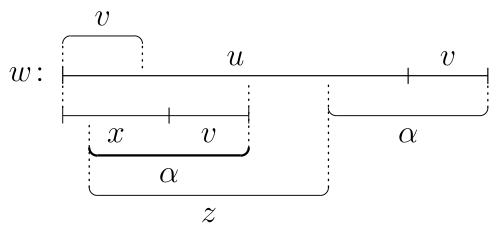
Esimerkki 4.3. Olkoon $w = (aba)^k abba(aba)^k$. Tämän sanan pituus on $6k + 4$ ja lyhin jakso on $3k + 4$, joten $|w| = 2\,\pi(w) - 4$. Sen pisimmät reunattomat tekijät ovat $(aba)^k abb$ ja $bba(aba)^k$, joten $\tau(w) = 3k + 3 = \pi(w) - 1$.

Esimerkissä 4.3 $|w| = 2\,\pi(w) - 4$, mutta seuraus 4.2 antaa rajan $|w| > 2\,\pi(w) - 2$. Tämä johtuu siitä, että seurauksen 4.2 raja ei ole optimaalinen. Duval [5] johti Lyndonin sanoja käyttäen paremman rajan $2\,\pi(w) - k$, missä k on sanan w aakkoston koko. Tämän rajan vakio $-k$ ei kuitenkaan ole optimaalinen. Tämän tutkielman luvussa 5 esitetään uutena tuloksena optimaalinen raja ja lisäksi johdetaan kaikki sanat, joiden pituudet ovat yhtä pienempiä kuin optimaalinen raja ja joilla pisin reunaton tekijä on aidosti lyhyempi kuin lyhin jakso.

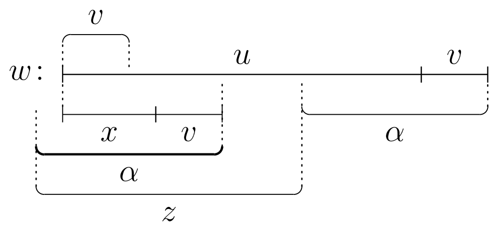
Lauseessa 4.1 on kaksi maksimisuffiksia, joista lyhyempi antaa varmasti kriittisen tekijöihinjaon. Seuraava lause antaa ehdon, jolloin molemmat maksimisuffiksit antavat kriittisen tekijöihinjaon.

Lause 4.3. Olkoon w sana, jolle $|w| \geq 2$, ja olkoon α sen $\triangleleft$-maksimisuffiksi, missä $\triangleleft$ on jokin aakkoston täydellinen järjestys. Jos $|\alpha| \leq \frac{|w|}{2} + \frac{1}{2}$, niin $|w| - |\alpha|$ on sanan w kriittinen piste.

Todistus. Olkoon $w = uv$, missä $|u| = \pi(w)$. Lauseen 2.16 mukaan $|w| - |\alpha| < \pi(w)$. Epäyhtälön (4.1) mukaan pisteessä $|w| - |\alpha|$ on olemassa toistosana z, jolle $|z| \leq |u|$. Tämä tilanne on esitetty kuvassa 4.4. Tehdään nyt vastaoletus, että z on lyhyempi kuin u.



(a) Sana z on lyhyempi tai yhtä pitkä kuin $w\alpha^{-1}$. Sana α esiintyy kaksi kertaa sanassa w.



(b) Sana z on pidempi kuin $w\alpha^{-1}$.

Kuva 4.5: Sana α on sanan z prefiksi. Tapaus (a) on ristiriidassa lauseen 2.16 kanssa.

Jos z on lyhyempi kuin α, niin z on sanan α epätyhjä prefiksi ja $|z\alpha| \leq 2|\alpha| - 1 \leq |w|$. Tällöin $z\alpha$ on sanan w suffiksi, mikä on ristiriidassa lauseen 2.15 kanssa. Näin ollen

$$|z| \geq |\alpha|,$$

joten sana α on sanan z prefiksi. Tämä tilanne on esitetty kuvassa 4.5. Koska $|z| < |u|$, sanalla w on epätyhjä prefiksi $x = (w\alpha^{-1})(v\alpha^{-1}z)^{-1}$. Jos nyt $|z| \leq |w| - |\alpha|$, kuten kuvassa 4.5a, niin α esiintyy kaksi kertaa sanassa w. Tämä on ristiriidassa lauseen 2.16 kanssa, joten $|z| > |w| - |\alpha|$, kuten kuvassa 4.5b.

Koska oletettiin, että $|z| < |u|$, saadaan sanan $v(\alpha^{-1}z)$ pituutta rajoitetuksi johdolla

$$|v(\alpha^{-1}z)| = |z| - |\alpha| + |v| < |u| - |\alpha| + |v| = |w| - |\alpha|.$$

Näin ollen sanojen α ja $w\alpha^{-1}$ esiintymät sanassa z ovat limittäin, ja lisäksi xv on sanan α suffiksi, joka on myös sanan $w\alpha^{-1}$ prefiksi ja joka on pidempi kuin sana v. Sana xv on siis sanan w reuna. Tämä on ristiriidassa sen kanssa, että sanalla w ei ole sanaa v pidempää reunaa. Näin ollen $|z| = |u|$. $\square$

Seuraavan esimerkin mukaan edellisen lauseen raja on paras mahdollinen.

Esimerkki 4.4. Olkoon $w = a^n b a^{n+1}$, missä n on luonnollinen luku. Sen lyhin jakso on $n+2$. Sen maksimisuffiksit molempien leksikografisten järjestysten suhteen ovat $\alpha = a^{n+1}$ ja $\beta = ba^{n+1}$. Koska $|\beta| = n + 2 = \frac{|w|}{2} + 1$, ei edellinen lause kerro, onko maksimisuffiksia β vastaava piste n kriittinen. Lyhin toistosana pisteessä n on ba^n. Se on lyhyempi kuin sanan w lyhin jakso, joten piste n ei ole kriittinen.

Seuraavaa lemmaa käytetään luvun 6.5 lauseen 6.15 todistuksessa. Kohdan i todistus on peräisin Harjulta ja Nowotkalta [8]. Kohdat ii ja iii ovat peräisin Breslauerilta, Jiangilta ja Jiangilta [2], mutta ne on tässä todistettu eri tavalla.

Lemma 4.4. Olkoon w reunaton sana ja rs sen kriittinen tekijöihinjako. Tällöin

(i) r ja s eivät limity eikä kumpikaan ole toisen tekijä,

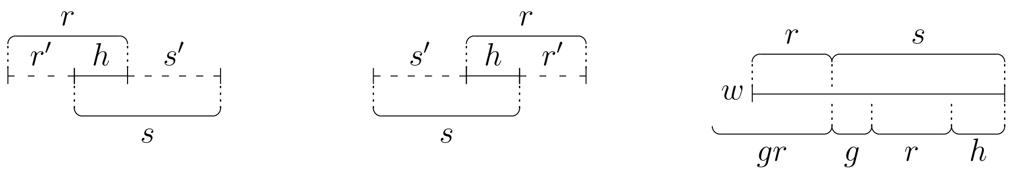
(ii) sana sr on reunaton ja

(iii) sr on kriittinen tekijöihinjako.

Todistus. Oletuksista voidaan heti tehdä huomio sanan w lyhimmästä paikallisesta jaksosta pisteessä r. Koska w on reunaton, niin $\pi(w) = |w|$. Lisäksi, koska rs on kriittinen tekijöihinjako, saadaan johdetuksi $\pi(w, |r|) = \pi(w) = |w|$.

(i) Tehdään vastaoletus, että r ja s limittyvät tai toinen näistä sanoista on toisen tekijä. Jokainen tapaus käsitellään erikseen. Tapaus, jossa s on sanan r tekijä, palautuu tapaukseen, jossa r on sanan s tekijä, tarkastelemalla sanan w peilisanaa $\overline{w} = \overline{s}\,\overline{r}$.

Jos $r = r'h$ ja $s = hs'$ jollekin epätyhjälle sanalle h (kuva 4.6a), niin h on sanan $rs = w$ toistosana pisteessä $|r|$. Mutta tällöin $\pi(w, |r|) \leq |h| < |w|$, mikä on ristiriita, koska $\pi(w, |r|) = |w|$.



(a) Sana h on sanan r suffiksi ja sanan s prefiksi.

(b) Sana h on sanan r prefiksi ja sanan s suffiksi.

(c) Sana r on sanan s tekijä.

Kuva 4.6: Lemman 4.4 kohdan (i) vastaoletukset.

(a) $|h| < |r|$.

(b) $|h| \geq |r|$.

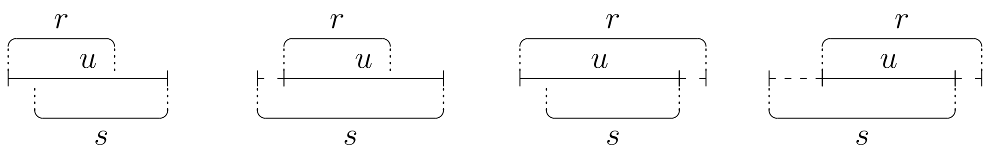
Kuva 4.7: Lemma 4.4 (ii). Vastaoletuksesta seuraa, että sana r joko limittyy sanan s kanssa tai on sanan s tekijä.

Jos $r = hr'$ ja $s = s'h$ jollekin epätyhjälle sanalle h (kuva 4.6b), niin tällöin h on sanan $rs = w$ reuna. Tämä on ristiriita, koska w on reunaton.

Jos $s = grh$ joillekin sanoille g ja h (kuva 4.6c), niin tällöin $w = rgrh$. Nyt gr on sanan w toistosana pisteessä $|r|$, mutta se on lyhyempi kuin koko sana w. Tämä on ristiriidassa sen kanssa, että rs on sanan w kriittinen tekijöihinjako.

(ii) Tehdään vastaoletus, että h on sanan sr lyhin reuna. Voidaan olettaa, että $|r| \leq |s|$, koska toinen tapaus palautuu tähän tapaukseen tarkastelemalla peilisanaa $\overline{w} = \overline{r}\,\overline{s}$. Seurauksen 2.7 mukaan $|h| \leq |s|$. Kuvassa 4.7 on esitetty molemmat tavat, joilla sanojen h ja r esiintymät voivat sijoittua toistensa suhteen: Jos $|h| < |r|$, niin h on sanan r aito suffiksi ja sanan s aito prefiksi. Jos taas $|h| \geq |r|$, niin r on sanan h suffiksina sanan s tekijä. Kumpikin tapaus on ristiriidassa kohdan (i) kanssa.

(iii) Kohdan (ii) mukaan sr on reunaton, joten sen lyhin jakso on $|sr|$. Olkoon u sanan sr lyhin toistosana pisteessä s, ja tehdään vastaoletus, että $|u| < |sr|$. Kuvan 4.8 mukaan joko sanat r ja s limittyvät, r on sanan s tekijä tai s on sanan r tekijä. Tämä on ristiriidassa kohdan (i) kanssa. $\square$



(a) Sanat r ja s limittyvät.

(b) Sana r on sanan s tekijä.

(c) Sana s on sanan r tekijä.

(d) Sanat r ja s limittyvät.

Kuva 4.8: Lemma 4.4 (iii). Kaikki tavat, joilla sanan sr toistosana u voi sijoittua sanojen r ja s suhteen, kun u on lyhyempi kuin sr.

5 Periodiraja

Tässä luvussa tarkastellaan, miten sanan w pituutta täytyy rajoittaa kuvauksen $\pi(w)$ suhteen, jotta $\tau(w) = \pi(w)$. Uutena tuloksena saadaan parasta aiemmin tunnettua rajaa lasketuksi yhdellä, ja näin saatu raja on optimaalinen. Lisäksi uutena tuloksena määritellään kaikki ne sanat, joilla $\tau(w) < \pi(w)$ ja joiden pituus on tarkalleen yhtä pienempi kuin tämä raja.

Seuraava lemma koskee tässä luvussa usein ilmenevää tilannetta, jossa jokin sana on Lyndonin sana mahdollisimman monen aakkoston järjestyksen suhteen.

Lemma 5.1. Olkoon w Lyndonin sana jokaisen järjestyksen $\lhd_a$ suhteen, jolla kirjain a on minimialkio. Olkoon $w = tuv$, missä a ei ole sanan u tekijä, ja olkoon w' sanan w konjugaatti. Jos t on sanan w' prefiksi, niin myös tu on sanan w' prefiksi.

Todistus. Väite on triviaali, kun $u = \varepsilon$. Olkoon u epätyhjä sana. Tehdään vastaoletus, että tu ei ole sanan w' prefiksi. Tällöin $w \wedge_{\mathrm{p}} w' = tu'$, missä u' on sanan u aito prefiksi. Koska w' ei ole sanan w prefiksi eikä toisinpäin,

$$
\begin{aligned}
tu'b &\leq_{\mathrm{p}} w, \\
tu'c &\leq_{\mathrm{p}} w',
\end{aligned}
$$

missä b ja c ovat eri kirjaimia. Koska $u'b \leq_{\mathrm{p}} u$ ja a ei ole sanan u tekijä, saadaan, että $b \neq a$. Kirjain b ei siis ole minkään järjestyksen $\lhd_a$ minimialkio, mutta c voi olla. Valitaan nyt järjestys $\lhd_a$ niin, että $c \lhd_a b$. Tällöin $tu'c \lhd_a tu'b$. Koska $tu'c \leq_{\mathrm{p}} w'$ ja $tu'b \leq_{\mathrm{p}} w$, saadaan lemman 2.13 (ii) mukaan relaatio $w' \lhd_a w$, josta saadaan ristiriita, koska w on Lyndonin sana järjestyksen $\lhd_a$ suhteen. $\square$

Seuraavasta esimerkistä voidaan huomata, että vaikka sanalla olisi pienin mahdollinen määrä Lyndonin konjugaatteja, niin ne eivät välttämättä ole peräkkäisiä konjugaatteja.

Esimerkki 5.1. Tarkastellaan sanaa $u = aabccabbc$, joka on primitiivinen ja jonka aakkosto on $\{a, b, c\}$. Etsitään jokaista aakkosta kohti sen maksimaalinen syklinen toisto, eli pisin potenssi, joka on sanan u syklinen tekijä (eli konjugaatin tekijä). Huomataan, että aa, bb ja cc ovat maksimaaliset sykliset kirjaimen potenssit, ja lisäksi nämä esiintyvät vain yhden kerran sanan u syklisinä tekijöinä. Jokaisella sanan u Lyndonin konjugaatilla x on nyt prefiksinä aa, bb tai cc. Tämä prefiksi määräytyy siitä, mikä on kunkin Lyndonin sanan x määrittämä järjestys $\lhd$, sillä prefiksin kirjain on sama kuin järjestyksen $\lhd$ minimikirjain. Jokainen sanan u Lyndonin konjugaatti on lemman 5.1 mukainen.

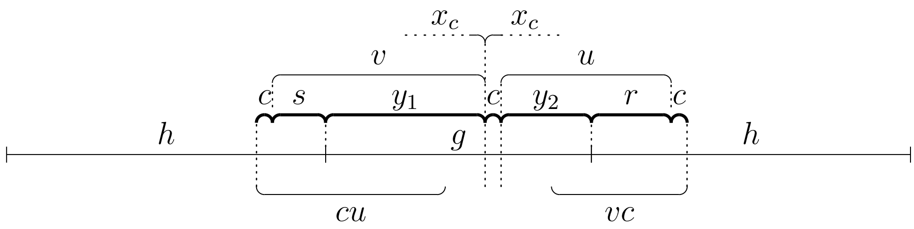
Tarkastellaan sitten sanaa $v = aabaabbc$. Tämäkin sana on primitiivinen, ja sen aakkosto on $\{a, b, c\}$. Kirjaimen a maksimaalinen potenssi sanan v syklisenä tekijänä on aa, ja tämä esiintyy kahdesti. Tästä huolimatta Lyndonin konjugaatit järjestysten $a \leq b \leq c$ ja $a \leq c \leq b$ suhteen eivät jakaudu: $aabaabbc = v$ on Lyndonin konjugaatti kummankin järjestyksen suhteen. Kirjaimen b maksimaalinen potenssi syklisenä tekijänä on bb, ja se esiintyy tarkalleen kerran. Näin ollen $bbc.aabaa$ on Lyndonin konjugaatti sekä järjestyksen $b \leq a \leq c$ että järjestyksen $b \leq c \leq a$ suhteen. Kirjain c esiintyy vain kerran sanassa v, joten kyseistä pistettä vastaava konjugaatti $c.aabaabb$ on Lyndonin sana sekä järjestyksen $c \leq a \leq b$ että järjestyksen $c \leq a \leq b$ suhteen.

Seuraava lemma on lemmaa 5.1 teknisempi, mutta sen rakenne esiintyy kahdessa eri todistuksessa samanlaisena.

Lemma 5.2. Olkoon $w = hgh$, missä h on sanan w pisin reuna, ja olkoon $\tau(w) < \pi(w)$. Jos kirjain c esiintyy tarkalleen kerran sanassa g, niin se ei esiinny sanassa h.

Todistus. Olkoon $g = y_1 c y_2$, missä sanat y_1 ja y_2 eivät sisällä kirjainta c. Sana $x_c = c y_2 h y_1$ on Lyndonin sana kaikkien järjestysten $\lhd_c$ suhteen, koska se on ainoa sanan w syklinen juuri, jonka ensimmäinen kirjain on c ja joka ei ole sanan w tekijä. Samoin myös sana $\overline{y_2 h y_1 c}$ on Lyndonin sana kaikkien järjestysten $\lhd_c$ suhteen.

Tehdään vastaoletus, että kirjain c esiintyy sanassa h. Olkoon r ja s sanan h pisin prefiksi ja suffiksi, joissa ei esiinny kirjainta c. Tällöin siis $rc \leq_{\mathrm{p}} h$ ja $cs \leq_{\mathrm{s}} h$. Tämä tilanne on kuvan 5.1 alkuasetelmana.



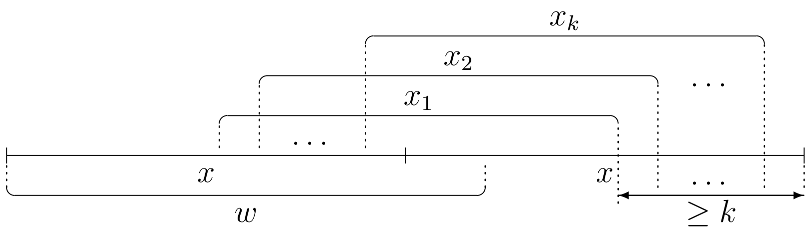
Kuva 5.1: Syklinen juuri x_c on Lyndonin sana kaikkien järjestysten $\lhd_c$ suhteen. Koska $cu \leq_{\mathrm{p}} x_c$ ja kirjain c ei esiinny sanassa u, on lemman 5.1 mukaan syklisellä juurella $(cs)gh(cs)^{-1}$ prefiksinä cu.

Olkoon $u = y_2 r$ ja $v = sy_1$, ja olkoon lisäksi $x_c = cy_2hy_1$. Sana x_c on Lyndonin sana kaikkien järjestysten $\lhd_c$ suhteen, joilla c on minimialkio. Sanalla x_c on nyt prefiksinä sana cu, ja sen konjugaatilla $csgh(cs)^{-1}$ on prefiksinä sana c. Koska sanassa u ei esiinny kirjainta c, on lemman 5.1 mukaan cu sanan $csgh(cs)^{-1}$ prefiksi. Sanat cu ja cvc ovat siis molemmat sanan $csgh(cs)^{-1}$ prefiksejä, mutta kirjain c esiintyy edellisessä vain kerran ja jälkimmäisessä kaksi kertaa. Näin ollen sanan cu täytyy olla sanan cv prefiksi. Tästä saadaan, että sana u on sanan v prefiksi, joten

$$|u| \leq |v|. \tag{5.1}$$

Edellisen kappaleen päättelyt pätevät myös symmetrisesti. Sana $\overline{c^{-1}x_c c}$ on Lyndonin sana kaikkien järjestysten $\lhd_c$ suhteen, joilla c on minimialkio. Sanalla $c^{-1}x_c c$ on suffiksina sana vc, ja sen konjugaatilla $(rc)^{-1}hgrc$ on suffiksina sana c, joten lemman 5.1 mukaan vc on sanan $(rc)^{-1}hgrc$ suffiksi. Koska suffiksissa vc esiintyy vähemmän kirjainta c kuin suffiksissa cuc, on vc sanan uc suffiksi. Näin ollen v on sanan u suffiksi. Epäyhtälön (5.1) mukaan nyt $v = u$. Nyt sanalla x_c on aitona prefiksinä sana cu ja aitona suffiksina sana $cv = cu$, joten sanalla x_c on reuna. Tämä on ristiriita, koska sana x_c on Lyndonin sanana reunaton. Näin ollen vastaoletus ei voi päteä ja sanassa h ei esiinny kirjainta c. $\square$

Seuraava lause parantaa Duvalin [5] todistamaa rajaa $|w| \geq 2\pi(w) - k$ yhtä pienemmäksi. Todistuksessa seurataan ensin Duvalin todistusta, minkä jälkeen käsitellään tapaus, jossa $|w| = 2\pi(w) - k - 1$.



Kuva 5.2: Sanalla w on ainakin k syklistä juurta x_i, jotka ovat Lyndonin sanoja eivätkä näin ollen ole sen tekijöitä. Tämä rajoittaa sanan w pituutta.

Lause 5.3. Olkoon $|w| \geq 2\pi(w) - k - 1$, missä k on sanassa w esiintyvien kirjainten lukumäärä. Tällöin $\tau(w) = \pi(w)$.

Todistus. Olkoon $\{a_1, \ldots, a_k\}$ sanan w aakkosto. Tehdään vastaoletus, että $\tau(w) < \pi(w)$. Olkoon x sanan w prefiksi, jonka pituus on $\pi(w)$. Luvussa 2.3 on todettu, että x on primitiivinen. Lemman 2.19 mukaan sanan x konjugaateista ainakin k on Lyndonin sanoja ja siten reunattomia. Nämä sykliset juuret eivät voi olla sanan w tekijöitä vastaoletuksen mukaan. Olkoon L sanan w niiden syklisten juurten x_i joukko, jotka ovat Lyndonin sanoja. Voidaan olettaa, että $|v_1| > |v_2| > \cdots > |v_{|L|}| > 0$, kun $x_i = v_i u_i$ ja $x = u_i v_i$. Tämä on esitetty kuvassa 5.2. Näin ollen $|v_1| \geq |L|$. Jotta yksikään x_i ei ole sanan w tekijä, tulee sanan w olla sanan xxv_1^{-1} aito prefiksi. Näin ollen

$$|w| < |xx| - |v_1| \leq 2\,\pi(w) - |L| \leq 2\,\pi(w) - k. \tag{5.2}$$

Riittää siis enää tarkastella tapausta, jossa $|w| = 2\,\pi(w) - k - 1$.

Olkoon nyt $|w| = 2\,\pi(w) - k - 1$. Sijoittamalla tämä johtoon (5.2) saadaan joukon L koolle rajat $k + 1 > |L| \geq k$, joista saadaan edelleen, että $|L| = k$. Näin ollen joukko L sisältää tarkalleen kaikki sanan w sykliset juuret, jotka eivät ole sen tekijöitä.

Olkoon h sanan w pisin reuna. Lauseen 2.8 mukaan $|h| = |w| - \pi(w)$. Sijoittamalla tähän yhtälö $|w| = 2\,\pi(w) - k - 1$ saadaan

$$|h| = |w| - \pi(w) = \pi(w) - k - 1 < \pi(w).$$

Näin ollen $x = hg$, missä $g \neq \varepsilon$. Lauseen 2.8 mukaan $|w| = |h| + |x|$, joten

$w = xh = hgh$. Lisäksi $|g| = \pi(w) - |h| = k + 1$. Olkoon $g = g_1 g_2 \dots g_k g_{k+1}$, missä jokainen g_i on kirjain. Edellisen kappaleen mukaan nyt

$$x_i = g_{i+1} \dots g_{k+1} h g_1 \dots g_i$$

kaikilla $i = 1, \dots, k$.

Jokainen joukon L sana alkaa eri kirjaimella, joten $\{g_2, g_3, \dots, g_{k+1}\} = \{a_1, \dots, a_k\}$. Voidaan olettaa, että $g_{i+1} = a_i$, kun $i = 1, \dots, k$. Tällöin $g = g_1 a_1 \dots a_k$.

Oletetaan, että $g_1 \neq a_k$. Tällöin $g_1 = a_i$ jollekin $i \in \{1, \dots, k-1\}$. Koska a_i on sanan gh prefiksi ja sana gh on sanan x_i konjugaatti, on lemman 5.1 mukaan myös $a_i a_{i+1}$ sanan gh prefiksi. Tällöin kuitenkin $a_{i+1} = a_1$, koska $a_i a_1$ on sanan g prefiksi. Tämä on ristiriita, koska $i + 1 \in \{2, \dots, k\}$.

Oletetaan, että $g_1 = a_k$. Tällöin

$$g = a_k a_1 a_2 \dots a_{k-1} a_k.$$

Kirjain a_k on nyt ainoa kirjain, joka esiintyy kaksi kertaa sanassa g. Lemman 5.2 mukaan $h = a_k^n$, missä $n \geq 1$. Sijoittamalla tämä yhtälöön $w = hgh$ saadaan nyt yhtälö $w = a_k^{n+1} a_1 \dots a_{k-1} a_k^{n+1}$. Sanalla w on nyt reuna a_k^{n+1}, mikä on ristiriita, koska sana a_k^n oletettiin sanan w pisimmäksi reunaksi. $\square$

Seuraavat tulokset todistavat, mitä muotoa ovat ne sanat w, joille $\tau(w) < \pi(w)$ ja joiden pituudet ovat lauseen 5.3 periodirajan suhteen maksimaalisia. Ensimmäinen lemma kertoo, että tällaisten sanojen pisin reuna ei ole limittäin sanan prefiksinä ja suffiksina. Lisäksi se kertoo pisimmän reunan väliin jäävän osan muodon.

Lemma 5.4. Olkoon $|w| = 2\,\pi(w) - k - 2$, missä k on sanan w aakkoston koko, ja olkoon $\tau(w) < \pi(w)$. Olkoon h sanan w pisin reuna. Tällöin $w = hgh$, missä

$$g = abyba,$$

$y = a_1 \dots a_{k-2}$ ja sanan w aakkosto on $\{a, b, a_1, \dots, a_{k-2}\}$. Lisäksi $a^n b \leq_{\mathrm{p}} h$ ja $ba^n \leq_{\mathrm{s}} h$, missä $n \geq 0$.

Todistus. Lauseen 2.8 mukaan $|h| = \pi(w) - k - 2 < \pi(w)$, joten sana w on muotoa hgh, missä $|hg| = \pi(w)$. Tästä saadaan, että $|g| = k + 2$.

Olkoon $g = g_1 \dots g_{k+2}$, ja olkoon $G = \{g_i \dots g_{k+2} h g_1 \dots g_{i-1} \mid i = 2, \dots, k + 2\}$. Joukko G koostuu niistä sanan w syklisistä juurista, jotka eivät ole sanan w tekijöitä. Lemman 2.19 mukaan joukossa G on jokaista aakkoston kirjainta a kohti ainakin yksi Lyndonin sana, jonka ensimmäinen kirjain on a. Näin ollen sanassa $g_2 \dots g_{k+2}$ esiintyy jokainen kirjain ainakin kerran. Symmetrisesti, tarkastelemalla sanan g peilisanaa saadaan, että myös sanassa $g_1 \dots g_{k+1}$ esiintyy jokainen kirjain ainakin kerran. Näin ollen sekä g_1 että g_{k+2} esiintyvät ainakin kaksi kertaa sanassa g. On nyt kaksi mahdollisuutta: joko

(i) yksi kirjain a esiintyy kolme kertaa sanassa g, tai

(ii) kaksi eri kirjainta a ja b esiintyy kaksi kertaa sanassa g.

(Tapaus i) Oletetaan, että vain yksi kirjain a esiintyy useamman kerran sanassa g. Tällöin $g = aya$, missä y sisältää jokaisen aakkoston kirjaimen tarkalleen kerran. Lemman 5.2 mukaan $h = a^n$, missä $n \geq 1$. Sana a^{n+1} on nyt sanan w reuna, mikä on ristiriidassa sen kanssa, että h on sanan w pisin reuna.

(Tapaus ii) Oletetaan, että on kaksi eri kirjainta a ja b, jotka esiintyvät kaksi kertaa sanassa g. Tämä jakaantuu kahteen tapaukseen sen mukaan, onko $g_1 = g_{k+2}$ vai ei.

(Tapaus ii.1) Oletetaan, että $g_1 = a$ ja $g_{k+2} = b$, missä $a \neq b$. Tällöin sekä a että b esiintyvät tarkalleen kaksi kertaa sanassa g.

Jos $g_2 = a$, niin sana $g_2 \dots g_{k+2} h g_1$ on Lyndonin sana kaikkien järjestysten $\lhd_a$ suhteen. Tämä on kuitenkin mahdotonta, koska nyt aa esiintyy sanassa w ja tällöin sanalla $g_2 \dots g_{k+2} h g_1$ tulisi olla prefiksinä aa, jolloin a esiintyisi kolme kertaa sanassa g. Näin ollen

$$g = azayb, \tag{5.3}$$

missä $z \neq \varepsilon$ ja a ei esiinny sanoissa z ja y.

Nyt sana $aybhaz$ on Lyndonin sana kaikkien järjestysten $\lhd_a$ suhteen, koska se on joukon $\{vhu \mid u <_{\mathrm{p}} g, v <_{\mathrm{s}} g \text{ ja } h = uv\}$ sanoista ainoa, jonka ensimmäinen kirjain on a. Lemman 5.1 mukaan nyt sana yb on sanan $zaybh$ prefiksi. Koska sanassa yb ei esiinny kirjainta a, on yb sanan z prefiksi. Tämä tarkoittaa, että epätyhjä sana yb esiintyy kahdesti sanassa g. Vain kirjaimet a ja b voivat esiintyä kahdesti sanassa g, joten $yb = b$, mistä seuraa, että $y = \varepsilon$ ja $z = bz'$ jollekin sanalle z'. Sijoittamalla nämä yhtälöön (5.3) saadaan yhtälö

$$g = abz'ab. \tag{5.4}$$

Jos z' on tyhjä sana, niin $g = abab$ ja sanalla $abhab$ on reuna ab. Tämä on ristiriita, koska sana $abhab$ on Lyndonin sanana reunaton. Näin ollen z' on epätyhjä.

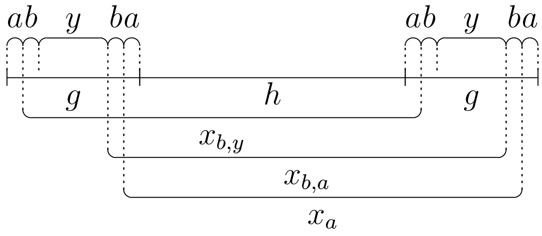
Sana aa ei ole sanan w tekijä, koska ab on Lyndonin sanan $abhabz'$ prefiksi. Symmetrisesti sana bb ei ole sanan w tekijä, koska ba on Lyndonin sanan $\overline{z'abhab}$ prefiksi.

Lemman 5.2 mukaan $h \in \{a,b\}^+$. Koska aa ja bb eivät ole sanan w tekijöitä, on $h = (ab)^n$, missä $n \geq 1$. Tällöin kuitenkin $w = (ab)^{n+1}z'(ab)^{n+1}$, joten sanalla w on sanaa h pidempi reuna $(ab)^{n+1}$. Tämä on ristiriita, koska h oletettiin sanan w pisimmäksi reunaksi. Näin on todistettu, että yhtälö (5.3) ei voi olla voimassa, joten kirjaimet g_1 ja g_{k+2} ovat samat.

(Tapaus ii.2) Oletetaan, että $g_1 = g_{k+2} = a$. Tällöin

$$g = axbybza, \tag{5.5}$$

missä sanoissa x , y ja z ei esiinny kirjaimia a eikä b. Kirjaimet a ja b ovat ainoat kirjaimet, jotka esiintyvät ainakin kahdesti sanassa g, joten lemman 5.2 mukaan sanassa h esiintyy vain kirjaimia a ja b. Kirjain b esiintyy sanassa h, koska muuten h olisi kirjaimen a potenssi ja sanalla w olisi pisintä reunaa h pidempi reuna ha. Olkoot m ja n epänegatiivisia kokonaislukuja, joille $a^n b \leq_{\mathrm{p}} h$ ja $ba^m \leq_{\mathrm{s}} h$. Sana $ahga^{-1}$ on Lyndonin sana, koska se on ainoa syklinen juuri, joka ei ole sanan w tekijä ja jonka ensimmäinen kirjain on a. Koska $a^{n+1}b$ on Lyndonin sanan $ahga^{-1}$ prefiksi, ei tekijä a^{m+1} voi olla pidempi kuin a^{n+1}. Näin ollen $m \leq n$.



Kuva 5.3: Määritelmässä 5.1 nimetyt sykliset juuret

Symmetrisesti myös sana $\overline{a^{-1}gha}$ on Lyndonin sana, koska se on ainoa syklinen juuri, joka ei ole sanan $\overline{w}$ tekijä ja jonka ensimmäinen kirjain on a. Koska ba^{m+1} on sanan $a^{-1}gha$ suffiksi, ei tekijä a^{n+1} voi olla pidempi kuin a^{m+1}. Näin ollen $n \leq m$, joten edellisen kappaleen mukaan $m = n$.

Sana $a^n gh(a^n)^{-1}$ on nyt sanan w syklinen juuri, ja sillä on prefiksinä a^{n+1}. Nyt lemman 5.1 mukaan $a^{n+1}b$ on sanan $a^n gh(a^n)^{-1}$ prefiksi, koska $a^{n+1}b$ on sanan $ahga^{-1}$ prefiksi ja $ahga^{-1}$ on Lyndonin sana kaikkien järjestysten $\lhd_a$ suhteen. Koska $a^n gh(a^n)^{-1} = a^{n+1}xbybzah(a^n)^{-1}$ ja sanassa x ei esiinny kirjainta b, tulee sanan x olla tyhjä sana. Symmetrisesti saadaan, että sanan z tulee olla tyhjä sana. Näin ollen lemma on todistettu. $\square$

Seuraavaksi määritellään tietyt sykliset juuret, jotka eivät ole sanan w tekijöitä.

Määritelmä 5.1. Olkoon $x_a = ahgabyb$, $x_{b,a} = bahgaby$ ja $x_{b,y} = bybaha$. Nämä sanan w sykliset juuret on esitetty kuvassa 5.3.

Seuraava lemma kertoo, mitkä äsken määritellyistä syklisistä juurista ovat Lyndonin sanoja.

Lemma 5.5. Olkoot x_a, $x_{b,a}$ ja $x_{b,y}$ kuten määritelmässä 5.1. Tällöin

(i) x_a on Lyndonin sana kaikkien järjestysten $\lhd_a$ suhteen,

(ii) $x_{b,a}$ on Lyndonin sana kaikkien järjestysten $\lhd_b$ suhteen, joilla b on minimialkio ja $a \lhd_b a_3$, ja

(iii) $x_{b,y}$ on Lyndonin sana kaikkien järjestysten $\lhd_b$ suhteen, joilla b on minimialkio ja $a_3 \lhd_b a$.

Erityisesti, jos w on binäärisana, niin $x_{b,y}$ on Lyndonin sana kaikkien järjestysten $\lhd_b$ suhteen ja $x_{b,a}$ ei ole Lyndonin sana.

Todistus. Olkoon $L = \{vhu \mid u <_{\mathrm{p}} g, v <_{\mathrm{s}} g \text{ ja } g = uv\}$. Jokainen sanan w syklisistä juurista, jotka eivät ole sanan w tekijöitä, on joukossa L. Näin ollen ne sanan w sykliset juuret, jotka ovat Lyndonin sanoja, sisältyvät joukkoon L. Olkoon $\lhd_c$ aakkoston järjestys, jolle kirjain c on minimialkio. Tällöin lemman 2.19 mukaan jokin sanan w primitiivisistä juurista on Lyndonin sana järjestyksen $\lhd_c$ suhteen ja sen ensimmäinen kirjain on c.

Kun $c \neq b$, on joukossa L vain yksi sana, jonka ensimmäinen kirjain on c. Näin ollen erityisesti x_a on Lyndonin sana jokaisen järjestyksen $\lhd_a$ suhteen.

Olkoon $c = b$. Jos $k = 2$, niin $y = \varepsilon$ ja sanalla $x_{b,a}$ on reuna b. Binäärisessä tapauksessa siis $x_{b,y}$ on Lyndonin sana kaikkien järjestysten $\lhd_b$ suhteen.

Olkoon nyt $c = b$ ja $k \geq 3$. Joukon L sanoista kahdella on b ensimmäisenä kirjaimena: $x_{b,y}$ ja $x_{b,a}$. Koska $ba_3 \leq_{\mathrm{p}} x_{b,y}$ ja $ba \leq_{\mathrm{p}} x_{b,a}$, leksikografisesti pienempi sana määräytyy sen mukaan, onko $a_3 \lhd_b a$ (ii) vai $a \lhd_b a_3$ (iii). $\square$

Seuraavaksi todistetaan, että kirjain b esiintyy vain yksittäin, paitsi mahdollisesti tekijässä g.

Lemma 5.6. Olkoon $g = abyba$ kuten lemmassa 5.4. Tällöin sana bb ei ole sanan $bahab$ tekijä.

Todistus. Tehdään vastaoletus, että bb on sanan $bahab$ tekijä, jolloin se on jonkin syklisen juuren x' prefiksi, missä $x' \neq bybaha$.

Jos y on epätyhjä sana, niin bb ei ole sanan $bybaha$ prefiksi ja tällöin $x' \lhd_b bybaha$ kaikille järjestyksille $\lhd_b$, joissa b on minimialkio. Tämä on ristiriita, koska sana $bybaha$ on Lyndonin sana ainakin yhden järjestyksen $\lhd_b$ suhteen.

Oletetaan, että y on tyhjä sana, jolloin $w \in \{a, b\}^*$. Tällöin $bbaha$ on Lyndonin sana kaikkien aakkoston järjestysten $\lhd_b$ suhteen, joissa b on minimialkio. Lemman 5.1 mukaan bba on sanan x' prefiksi. Koska $x' \neq bbaha$ ja ne ovat yhtä pitkiä, on $|x' \wedge_{\mathrm{p}} bbaha| < |x'|$. Olkoon $x' \wedge_{\mathrm{p}} bbaha = bbau$. Tällöin joko

(i) $bbaua \leq_{\mathrm{p}} bbaha$ ja $bbaub \leq_{\mathrm{p}} x'$ tai

(ii) $bbaub \leq_{\mathrm{p}} bbaha$ ja $bbaua \leq_{\mathrm{p}} x'$.

Jos kohta (i) on voimassa, niin $x' \lhd_b bbaha$. Tämä on ristiriita, koska sana $bbaha$ on Lyndonin sana järjestyksen $\lhd_b$ suhteen. Jos taas kohta (ii) on voimassa, niin $ahabb = auavbb$ ja $(bb)^{-1}x'bb = auav'bb$, jolloin $(bb)^{-1}x'bb \lhd_a ahabb$. Tämä on ristiriita, koska sana $ahabb$ on Lyndonin sana järjestyksen $\lhd_a$ suhteen. □

Seuraava lause määrittelee tyhjentävästi jokaisen sanan, jonka pituus on yhtä pienempi kuin lauseen 5.3 raja ja jonka pisin reunaton tekijä on lyhyempi kuin lyhin jakso.

Lause 5.7. Olkoon $|w| = 2\,\pi(w) - k - 2$, missä k on sanan w eri kirjainten lukumäärä, ja olkoon $\tau(w) < \pi(w)$. Tällöin $w = hgh$, missä

$$g = abyba, \tag{5.6}$$

$$h = a^n b(a^{n+1}b)^m a^n, \tag{5.7}$$

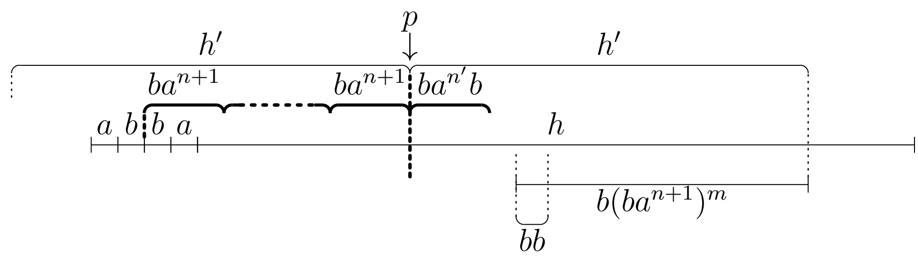
missä $m \geq 0$ ja $n \geq 1$, $y = a_1 \dots a_{k-2}$ ja sanan w aakkosto on $\{a, b, a_1, \dots, a_{k-2}\}$.

Todistus. Lemman 5.4 mukaan $w = hgh$, missä h on sanan w pisin reuna, yhtälö (5.6) on voimassa ja $a^n b \leq_{\mathrm{p}} h$ sekä $ba^n \leq_{\mathrm{s}} h$. Riittää siis todistaa, että yhtälö (5.7) on voimassa. Lemman 5.2 mukaan $h \in \{a, b\}^*$, koska vain kirjaimet a ja b esiintyvät esiintyvät useammin kuin kerran sanassa g. Määritelmän 5.1 mukaisella syklisellä juurella x_a on prefiksi $a^{n+1}b$, ja koska x_a on lemman 5.5 mukaan Lyndonin sana, niin sillä ei ole syklisenä tekijänä pidempää kirjaimen a potenssia kuin a^{n+1}. Lemman 5.6 mukaan sanassa h ei esiinny tekijää bb, joten

$$h = a^n b a^{n_1+1} b a^{n_2+1} b \dots b a^{n_m+1} b a^n,$$

missä $m \geq 0$ ja jokainen $n_i \leq n$. Lauseen väite on nyt, että jokainen $n_i = n$ ja $n \geq 1$, mikä on yhtäpitävä yhtälön (5.7) kanssa.

Oletetaan, että $n = 0$. Tällöin jokainen $n_i = 0$ ja $h = (ba)^m b$. Sanalla w on nyt reuna $(ba)^{m+1}$, mikä on ristiriidassa reunan h maksimaalisuuden



Kuva 5.4: Lauseen 5.7 todistus binäärisessä tapauksessa. Pisteeseen p rajoittuvalla syklisellä juurella on lyhin reuna h'. Sanalla h' pitää olla prefiksinä sana $ba^{n'}b$ ja suffiksina sana $b(ba^{n+1})^m$. Koska h' ei limity itsensä kanssa, on h' sanan h tekijä. Tällöin bb on sanan h tekijä, mikä on ristiriidassa lemman 5.6 kanssa.

kanssa. Näin ollen $n \geq 1$.

Tehdään vastaoletus, että sanalla hg on syklisenä tekijänä sana $ba^{n'}b$, missä $1 \leq n' \leq n$. Lauseen todistus jaetaan kahteen tapaukseen sen mukaan, onko sana y epätyhjä vai tyhjä sana.

Oletetaan, että $y \neq \varepsilon$. Tällöin määritelmän 5.1 sana $x_{b,a}$ on lemman 5.5 mukaan Lyndonin sana. Vastaoletuksen mukaan sana $ba^{n'}b$ on sanan $x_{b,a}$ jonkin konjugaatin x' prefiksi. Koska $ba^{n+1}b \leq_{\mathrm{p}} x_{b,a}$ ja $n' < n+1$, on $x_{b,a} \wedge_{\mathrm{p}} x' = ba^{n'}$. Olkoon $\lhd_b$ aakkoston järjestys, jolla b on minimialkio. Koska $b \lhd_b a$, saadaan nyt

$$x' \lhd_b x_{b,a}. \tag{5.8}$$

Olkoon vielä lisäksi $a \lhd_b a_3$. Tällöin $x_{b,a}$ on Lyndonin sana järjestyksen $\lhd_b$ suhteen, mikä on ristiriidassa relaation (5.8) kanssa.

Oletetaan, että y on tyhjä sana, jolloin $g = abba$ ja sanan w aakkosto on $\{a, b\}$. Vastaoletuksen mukaan sanalla $bahab$ on ainakin yksi muotoa $ba^{n'}b$ oleva tekijä, jolle $1 \leq n' \leq n$. Tarkastellaan vasemmanpuoleisinta tällaista muotoa olevaa tekijää. Tällöin tekijän $ba^{n'}b$ vasemmalla puolella on vain jokin m kappaletta tekijää ba^{n+1}. Tällöin $(ba^{n+1})^m ba^{n'}b$ on sanan $bahab$ prefiksi, missä $m \geq 1$. Seuraava päättely on kuvattu kuvassa 5.4.

Olkoon $h = uv$, missä $u = a^n(ba^{n+1})^{m-1}$. Olkoon nyt $x' = vabbau$. Sana x' ei ole reunaton, koska se on sanan w jakson pituinen tekijä ja $\tau(w) < \pi(w)$. Olkoon h' sanan x' lyhin reuna. Koska b on sanan h' ensimmäinen kirjain ja $ba^{n+1} \leq_{\mathrm{s}} x'$, saadaan $|h'| \geq n+2$. Näin ollen $ba^{n'}b$ on sanan h'

prefiksi. Sana $ba^{n'}b$ ei ole sanan $(ba^{n+1})^m$ tekijä, koska $n' < n+1$. Näin ollen $|h'| > |(ba^{n+1})^m|$ ja siis $b(ba^{n+1})^m$ on sanan h' suffiksi. Koska bb on sanan $b(ba^{n+1})^m$ prefiksi, on se sanan h' tekijä.

Lyhin sana, jolla on $ba^{n'}b$ prefiksinä ja $b(ba^{n+1})^m$ suffiksina, on $ba^{n'}b(ba^{n+1})^m$ eli $ba^{n'}bbau$. Näin ollen

$$|h'| \geq |ba^{n'}bbau| > |abba| + |u|.$$

Seurauksen 2.7 mukaan h' ei limity itsensä kanssa, joten ylläolevaa johtoa käyttäen saadaan johto

$$|h'| \leq |v| + |abba| + |u| - |h'| < |v|.$$

Sana h' on siis sanan v tekijä ja näin ollen sanan h tekijä. Koska bb on sanan h' tekijä, on se myös sanan h tekijä. Tämä on ristiriidassa lemman 5.6 kanssa. $\square$

Tarkastellaan lopuksi lauseen 5.7 mukaista sanaa

$$w = hgh = a^n b(a^{n+1}b)^{m+1}a_3 \dots a_k b(a^{n+1}b)^{m+1}a^n,$$

missä $m \geq 0$, $n \geq 1$ ja $k \geq 2$. Koska $|g| = k+2$ ja $|h| = (m+1)(n+2)+n-1$, saadaan pituudelle kaava

$$|w| = 2|h| + |g| = 2(m+1)(n+2) + 2n + k. \tag{5.9}$$

Sen lyhin jakso on

$$\pi(w) = |h| + |g| = (m+2)(n+2) + k - 1. \tag{5.10}$$

Pisimmät reunattomat tekijät ovat sanat hga^{-1} ja $a^{-1}gh$, joten

$$\tau(w) = \pi(w) - 1 = (m+2)(n+2) + k - 2. \tag{5.11}$$

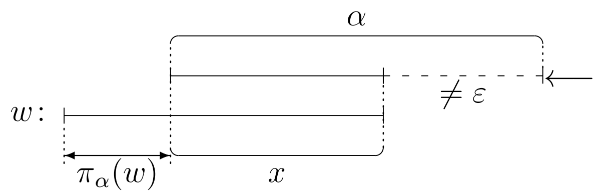
Pisimmät ainakin kahdesti esiintyvät reunattomat tekijät ovat $a^{n+1}b$ ja ba^{n+1}, joten

$$\tau_2(w) = n + 2. \tag{5.12}$$

Sijoittamalla kaava 5.10 yhtälöön $|w| = 2\,\pi(w) - k - 2$ saadaan

$$|w| = 2\,\tau(w) - k = \frac{7}{3}\,\tau(w) - \left(\frac{1}{3}\,\tau(w) - k\right).$$

Tämä on lähinnä arvoa $\frac{7}{3}\,\tau(w)$ silloin, kun $m = 0$, $n = 1$ ja $k = 2$, koska vain tällöin lauseke $\frac{1}{3}\,\tau(w) + k = (m+2)(n+2) + 2k - 2$ saa pienimmän arvonsa 4. Tällöin $w = abaabbaaba$, joka on pienin luvussa 3 mainittu sana.



Kuva 6.1: Sanan w α-suffiksi x ja α-periodi $\pi_\alpha(w)$.

6 Ehrenfeuchtin–Silbergerin ongelman ratkaisu

Tässä luvussa todistetaan tutkielman päälause 6.15, joka antaa parhaan tunnetun rajan sanan w pituudelle kuvauksen $\tau(w)$ suhteen, jotta $\tau(w) = \pi(w)$. Tarkoituksena on muodostaa sanalle tekijöihinjako, josta rakennetaan kahdessa eri tapauksessa pitkä reunaton tekijä. Lopuksi todistetaan, että jos näiden tekijöiden pituudet ovat enintään $\frac{3}{7}|w|$, niin seurauksena on joko ristiriita tai yhtälö $\tau(w) = \pi(w)$. Luku perustuu Holubin ja Nowotkan artikkeleihin [10, 9].

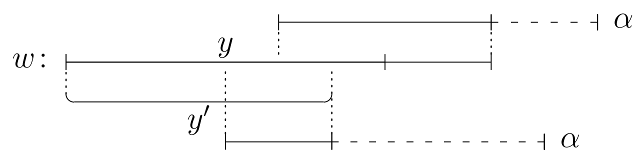
6.1 Suffiksi, periodi ja kriittinen prefiksi sanan suhteen

Tähän mennessä esitellyt käsitteet ovat sanojen kombinatoriikassa yleisiä. Lauseen 6.15 todistuksessa tarvitaan näiden lisäksi kolme uutta käsitettä. Seuraava määritelmä laajentaa sanan pisimmän reunan käsitettä, kun reuna ajatellaan limittymisen kautta.

Määritelmä 6.1. Sanan w pisin suffiksi x, joka on myös sanan α aito prefiksi, on sanan w *α-suffiksi.*

Jos w ei ole reunaton, niin sen w-suffiksi on sanan w pisin reuna. Jos sana w on reunaton, niin sen w-suffiksi on tyhjä sana.

Lauseen 2.8 mukaan sanan lyhin jakso määräytyy pisimmän reunan pituudesta. Seuraavassa määritelmässä laajennetaan pienimmän jakson käsitettä käyttämällä α-suffiksia.



Kuva 6.2: Sana y on sanan w α-kriittinen prefiksi. Sana y' on lyhyempi kuin y, joten sen α-periodi on lyhyempi.

Määritelmä 6.2. Olkoon x sanan w α-suffiksi. Tällöin

$$\pi_\alpha(w) = |wx^{-1}|$$

on sanan w *α-periodi.*

Tämä määritelmä on lyhimmän jakson käsitteen laajennus, koska $\pi_w(w) = \pi(w)$ kaikille epätyhjille sanoille w. Tämä ilmenee myös kuvasta 6.1.

Jos sana y on sanan w prefiksi, niin $\pi_\alpha(y) \leq \pi_\alpha(w)$: jos $|y| > \pi_\alpha(w)$, niin y ja sanan w α-suffiksi ovat limittäin sanassa w, ja jos $|y| \leq \pi_\alpha(w)$, niin $\pi_\alpha(y) \leq |y| \leq \pi_\alpha(w)$. Seuraava määritelmä kuvaa, minkä kohdan vasemmalla puolella sanat limittyvät eri tavalla.

Määritelmä 6.3. Sanan w lyhin prefiksi y, jolle $\pi_\alpha(y) = \pi_\alpha(w)$, on sanan w *α-kriittinen prefiksi*

Sanan w α-kriittinen prefiksi on esitetty kuvassa 6.2. Yllä määritellyistä käsitteistä saadaan heti seuraavat tulokset.

Lause 6.1. Jos α on sanan w suffiksi ja se on reunaton, niin sanan w α-suffiksi on tyhjä sana.

Todistus. Jos sanan w α-suffiksi olisi epätyhjä sana, niin α limittyisi itsensä kanssa. $\square$

Lause 6.2. Olkoon x sanan w α-kriittinen prefiksi ja y sanan w α-suffiksi. Tällöin $\pi_\alpha(w) \leq |x| \leq |w|$. Erityisesti, jos y on tyhjä sana, niin $\pi_\alpha(w) = |x| = |w|$.

Todistus. Määritelmän 6.3 mukaan $\pi_\alpha(w) = \pi_\alpha(x)$, ja määritelmän 6.2 mukaan $\pi_\alpha(x) \leq |x|$. Jälkimmäinen epäyhtälö johtuu siitä, että x on sanan w prefiksi.

Jos sanan w α-suffiksi on tyhjä sana, niin määritelmän 6.2 mukaan $\pi_\alpha(w) = |w|$, mikä pakottaa edellisen kohdan epäyhtälöt yhtälöiksi. □

Edellä todettiin, että α-periodi ei voi kasvaa prefikseillä. Olkoon x sanan w α-suffiksi, ja olkoon y sanan w α-kriittinen prefiksi. Olkoon nyt z sanan y aito prefiksi ja z' sanan w prefiksi, joka on pidempi kuin y. Tällöin saadaan epäyhtälöt

$$\pi_\alpha(z) \leq \pi_\alpha(y) \leq \pi_\alpha(z').$$

Koska y on sanan w α-kriittinen prefiksi, saadaan nämä epäyhtälöt muotoon

$$\pi_\alpha(z) < \pi_\alpha(y) = \pi_\alpha(z').$$

Kun rajoitutaan tarkastelemaan vain tapausta, jossa $|z| = |y| - 1$, saadaan toinen tapa määritellä α-kriittinen prefiksi. Tätä muotoilua käytetään luvun 6 todistuksissa määritelmän 6.3 sijasta.

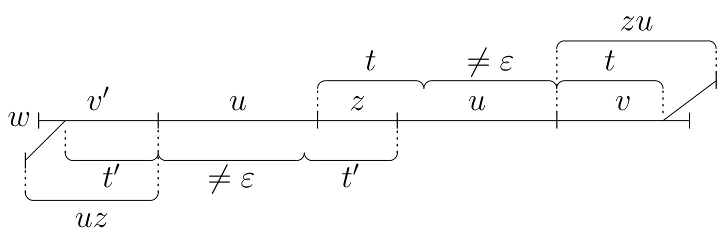
Lause 6.3. Sana za on sanan w α-kriittinen prefiksi, jos ja vain jos za on pisin sanan w prefiksi, jolle $\pi_\alpha(z) < \pi_\alpha(za)$.

Esimerkki 6.1. Olkoon $w = a^9$ ja $\alpha = a^6$. Tällöin sanan w α-suffiksi on a^5 ja $\pi_\alpha(w) = 4$.

Esimerkki 6.2 ([10])**.** Tarkastellaan sanaa $w = ababbaababab$ ja prefiksiä $\alpha = ababb$. Sanan w pituus on 12. Sen α-suffiksi on $abab$, joten $\pi_\alpha(w) = 8$. Sanan w α-kriittinen prefiksi on yhtä lyhyempi sana $ababbaababa$, koska $\pi_\alpha(ababbaababa) = 8$ ja $\pi_\alpha(ababbaabab) = 6$.

6.2 Tekijöihinjako

Tässä luvussa esitellään lauseen 6.15 todistuksessa tarvittava tekijöihinjako sekä esitellään tapa rajoittaa tätä tekijöihinjakoa niin, että saadulla tekijöihinjaolla on hyödyllisiä ominaisuuksia. Lopuksi todistetaan lemma 6.5,



Kuva 6.3: Määritelmän 6.4 tekijöihinjako.

joka antaa ensimmäisen pitkän reunattoman tekijän lauseen 6.15 todistusta varten, sekä esitellään lisää lauseen 6.15 todistuksessa tarvittavia tekijöitä.

Seuraava määritelmä on keskeisessä roolissa koko luvun 6 loppuun asti. Sen määrittelemän tekijöihinjaon oletetaan olevan voimassa jatkuvasti ilman erillistä mainintaa.

Määritelmä 6.4. Olkoon w sana, jossa ainakin yksi kirjain esiintyy ainakin kaksi kertaa. Olkoon sanalla w tekijöihinjako

$$w = v'uzuv, \tag{6.1}$$

missä u on reunaton, $|u| = \tau_2(w)$ ja z on mahdollisimman pitkä. Tällä tarkoitetaan sitä, että ei ole toista vastaavaa tekijöihinjakoa, jolla z olisi pidempi. Olkoot lisäksi

$$t = v \wedge_{\mathrm{p}} zu, \tag{6.2}$$

$$t' = v' \wedge_{\mathrm{s}} uz. \tag{6.3}$$

Seuraavaksi esitetään huomioita määritelmästä 6.4.

Määritelmän 6.4 tekijöistä t ja t' voidaan huomata, että t on aina tekijän zu aito prefiksi ja t' on aina tekijän uz aito suffiksi. Jos nimittäin $t = zu$, niin u on sanan v tekijä, jolloin z ei ole mahdollisimman pitkä. Vastaavasti jos $t' = uz$, niin tekijöihinjako (6.1) ei ilmaisisi tekijän u vasemmanpuoleisinta esiintymää. Tämä on yhtäpitävä sen kanssa, että zu ei ole tekijän v prefiksi ja uz ei ole tekijän v' prefiksi.

Tekijöihinjako (6.1) on aina olemassa, jos ainakin yksi kirjain esiintyy ainakin kaksi kertaa sanassa w, koska tällöin sanalla w on reunaton tekijä, joka esiintyy ainakin kahdesti. Tekijöihinjako ei välttämättä ole yksikäsitteinen, koska voi olla, että jokin toinen reunaton tekijä $\hat{u}$, joka on yhtä pitkä kuin u ja joka esiintyy kahdesti, ja nämä tekijän $\hat{u}$ esiintymät ovat lisäksi yhtä kaukana toisistaan kuin tekijän u esiintymät.

Huomautus 6.1. Määritelmä 6.4 on peilauksen suhteen symmetrinen. Tällä tarkoitetaan sitä, että jos kaavat (6.1), (6.2) ja (6.3) ovat voimassa, niin peilisanalla $\overline{w}$ on tekijöihinjako

$$\overline{w} = \overline{v}\,\overline{u}\,\overline{z}\,\overline{u}\,\overline{v'},$$

joka on määritelmän 6.4 mukainen. Lisäksi $\overline{t} = \overline{v} \wedge_{\mathrm{s}} \overline{u}\,\overline{z}$ ja $\overline{t'} = \overline{v'} \wedge_{\mathrm{p}} \overline{z}\,\overline{u}$, joten myös $\overline{t'}$ ja $\overline{t}$ ovat määritelmän 6.4 mukaisia. Tällöin kuitenkin sanan $\overline{t}$ rooli vastaa sanan t' roolia ja sanan $\overline{t'}$ rooli vastaa sanan t roolia.

Esimerkki 6.3. Sanalla *ababaaaabbaababbaba* on kolme määritelmän 6.4 mukaista tekijöihinjakoa. Seuraavassa nämä tekijöihinjaot on merkitty pisteillä:

$$\begin{aligned} & aba.baa.aab.baa.babbaba \\ = {} & ababaaa.abb.aab.abb.aba \\ = {} & ababaaaa.bba.aba.bba.ba. \end{aligned}$$

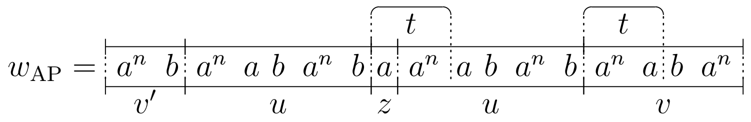
Ensimmäisessä tekijöihinjaossa t ja t' ovat tyhjiä sanoja. Toisessa tekijöihinjaossa $t = a$ ja $t' = \varepsilon$. Kolmannessa tekijöihinjaossa $t = \varepsilon$ ja $t' = a$.

Esimerkki 6.4. Tarkastellaan luvussa 3 käsiteltyä sanaa

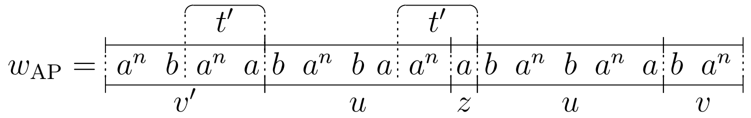
$$w_{\mathrm{AP}} = a^n ba^{n+1} ba^n ba^{n+2} ba^n ba^{n+1} ba^n,$$

missä $n \geq 1$. Sen määritelmän 6.4 mukainen tekijöihinjako on kuvassa 6.4a. Tässä tekijöihinjaossa t' on tyhjä sana. Koska $\overline{w_{\mathrm{AP}}} = w_{\mathrm{AP}}$ ja kyseisessä tekijöihinjaossa vzv ei ole sanan keskellä, on sanalla w_{AP} myös toinen määritelmän 6.4 mukainen tekijöihinjako, joka on kuvassa 6.4b. Tässä toisessa tekijöihinjaossa symmetrisesti t on tyhjä sana.

Myöhemmin ei siis käsitellä lemmojen peilattujen versioiden seurauksia

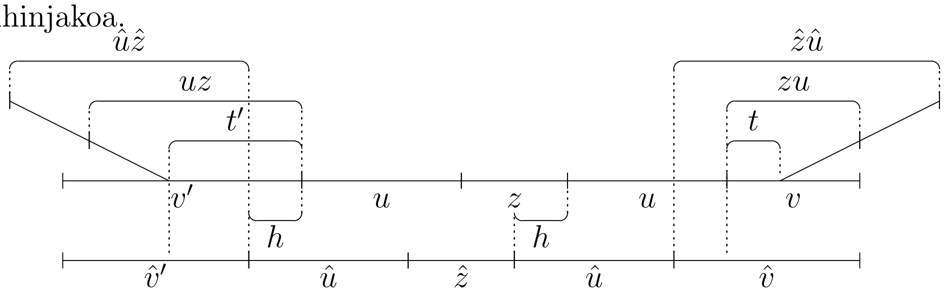


(a) Tässä $t' = \varepsilon$.



(b) Tässä $t = \varepsilon$.

Kuva 6.4: Assous'n–Pouzet'n sanalla on kaksi määritelmän 6.4 mukaista tekijöihinjakoa.



Kuva 6.5: Jos $\hat{u}\hat{z}\hat{u}$ on sanan $t'uzut$ tekijä, niin sanojen t' ja $\hat{t}'$ tulee rajoittua vasemmalta samaan pisteeseen, koska tässä pisteessä erkanevat sekä v' ja uz että $\hat{v}'$ ja $\hat{u}\hat{z}$. Vastaavasti sanat t ja $\hat{t}$ rajoittuvat oikealta samaan pisteeseen.

sanassa w_{AP}, koska tämä palautuu takaisin peilaamattoman sanan tekijöihinjakoon vaihtamalla tekijöihinjakoa.

Määritelmän 6.4 tekijöihinjako ei siis välttämättä ole yksikäsitteinen. On kuitenkin olemassa tapa koota tietyllä tavalla toisiaan lähellä olevia tekijöihinjakoja yhteen. Oletetaan, että sanalla w on toinen määritelmän 6.4 mukainen tekijöihinjako $w = \hat{v}'\hat{u}\hat{z}\hat{u}\hat{v}$ ja että $\hat{u}\hat{z}\hat{u}$ on sanan $t'uzut$ tekijä. Tällöin siis $\hat{u}\hat{z}\hat{u} \neq uzu$, joten $\hat{u}\hat{z}\hat{u}$ on joko sanan $t'uzu$ tai $uzut$ tekijä. Nämä tapaukset ovat symmetrisiä, joten voidaan olettaa, että $\hat{u}\hat{z}\hat{u}$ on sanan $t'uzu$ tekijä. Tämä tilanne on kuvassa 6.5. Sanat $\hat{u}\hat{z}$ ja uz ovat nyt konjugaatteja, joten sanalla t' on suffiksi h, jolle $v' = \hat{v}'h$ ja $\hat{u}\hat{z} = h(uzh^{-1})$. Koska t' on sanan uz aito suffiksi, on se myös sanan huz aito suffiksi. Näin ollen t' on sanojen v' ja huz pisin yhteinen suffiksi, jolloin saadaan seuraava kaava:

$$t' = v' \wedge_{\mathrm{s}} uz = v' \wedge_{\mathrm{s}} huz = \hat{v}'h \wedge_{\mathrm{s}} \hat{u}\hat{z}h = (\hat{v}' \wedge_{\mathrm{s}} \hat{u}\hat{z})h = \hat{t}'h.$$

Sijoittamalla tämä saadaan nyt, että

$$v'(t')^{-1} = \hat{v}'h(\hat{t}'h)^{-1} = \hat{v}'(\hat{t}')^{-1}.$$

Vastaavasti saadaan, että

$$t^{-1}v = \hat{t}^{-1}\hat{v}.$$

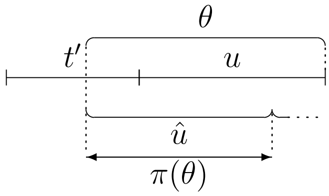
Näin ollen sanan w tekijöihinjakojen joukko voidaan partitioida sen mukaan, mikä on tekijöihinjaosta saatava tekijä $t'uzut$. Seuraavalla määritelmällä voidaan valita vasemmanpuoleisin tekijöihinjako tällä tavalla ekvivalenteista tekijöihinjaoista.

Määritelmä 6.5. Olkoon t' ja u kuten määritelmässä 6.4. Tekijä t' on *mahdollisimman lyhyt*, jos sanalla $t'u$ ei ole reunatonta tekijää $\hat{u} \neq u$, jolle $|\hat{u}| = |u|$.

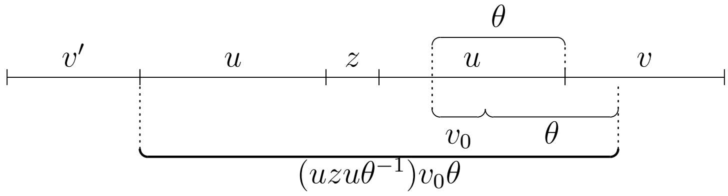
Seuraavan lemman mukaan sanan u maksimisuffiksi jonkin järjestyksen suhteen on myös sanan $t'u$ maksimisuffiksi, jos t' on mahdollisimman lyhyt.

Lemma 6.4. Olkoon t' ja u kuten määritelmässä 6.4. Olkoon lisäksi t' mahdollisimman lyhyt ja θ sanan $t'u$ $\lhd$-maksimisuffiksi jollekin leksikografiselle järjestykselle $\lhd$. Tällöin $|\theta| \leq |u|$.

Todistus. Tehdään vastaoletus, että $|\theta| > |u|$. Tällöin u on sanan θ aito prefiksi. Lauseen 2.15 mukaan sanan θ pisimmän jakson pituinen prefiksi on reunaton. Olkoon $\hat{u}$ tämä prefiksi. Tämä tilanne on esitetty kuvassa 6.6. Todistetaan, että $\hat{u}$ on yhtä pitkä kuin u. Koska u on reunaton, niin $\pi(u) = |u|$. Koska u on sanan θ tekijä, niin lemman 2.10 mukaan saadaan, että $|\hat{u}| \geq |u|$. Koska $\hat{u}$ on sanan $t'u$ tekijä ja $t'u$ esiintyy ainakin kahdesti sanassa w, niin $\hat{u}$ esiintyy ainakin kahdesti sanassa w. Näin ollen $|\hat{u}| \leq \tau_2(w) = |u|$, koska $\hat{u}$ on reunaton. Tästä saadaan, että $|\hat{u}| = |u|$. Sanalla $t'u$ on siis reunaton tekijä $\hat{u} \neq u$, joka on yhtä pitkä kuin u. Tämä on ristiriidassa sen kanssa, että t' on mahdollisimman lyhyt. □



Kuva 6.6: Lemman 6.4 todistus vastaoletuksella $|\theta| > |u|$.



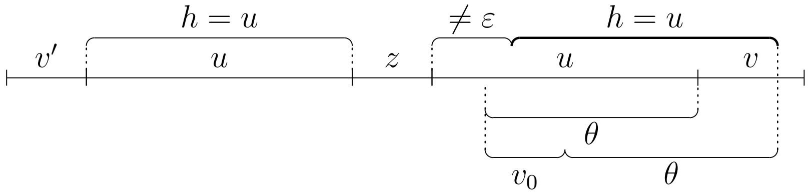
Kuva 6.7: Lemman 6.5 reunaton sana.

Esimerkki 6.5. Tarkastellaan uudelleen sanaa w_{AP} ja sen kahta tekijöihinjakoa, jotka on esitetty kuvassa 6.4 sivulla 51. Kuvan 6.4a tekijöihinjaossa t' on mahdollisimman lyhyt, koska se on tyhjä. Lemma 6.4 ei tässä tapauksessa kerro mitään uutta, koska $t'u = u$. Kuvan 6.4b tekijöihinjaossa t' ei ole mahdollisimman lyhyt, koska edellisen tekijöihinjaon u on tämän tekijöihinjaon tekijän $t'u$ tekijä. Tekijöihinjaot ovat siis ekvivalentteja määritelmää 6.5 edeltävän tarkastelun mielessä.

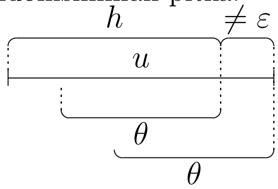
Esimerkki 6.6. Esimerkin 6.3 tekijöihinjaoissa *ababaaa.abb.aab.abb.aba* ja *ababaaaa.bba.aba.bba.ba* tekijät $t'uzut$ ovat samoja. Näistä kahdesta tekijöihinjaosta ensimmäinen on se, jonka tekijä t' on mahdollisimman lyhyt. Tekijöihinjaossa *aba.baa.aab.baa.babbaba* taas t' on tyhjä sana, joten tässä tekijöihinjaossa t' on mahdollisimman lyhyt.

Seuraava lemma antaa pitkän reunattoman tekijän, jos tekijän u jokin maksimisuffiksi esiintyy kaksi kertaa tekijässä uv. Tämä reunaton tekijä on esitetty kuvassa 6.7.

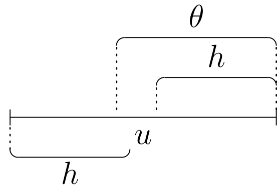
Lemma 6.5. Olkoot u ja v kuten määritelmässä 6.4, ja olkoon $\lhd$ jokin aakkoston järjestys. Olkoon θ tekijän u $\lhd$-maksimisuffiksi ja v_0 epätyhjä sana. Jos $v_0\theta \leq_{\mathrm{p}} \theta v$, niin $(uzu\theta^{-1})v_0\theta$ on reunaton.



(a) Jos $h = u$, niin u esiintyy kaksi kertaa sanassa uv. Tämä on mahdotonta, koska z on mahdollisimman pitkä.



(b) Jos $\theta \leq_{\mathrm{s}} h$ ja $h <_{\mathrm{p}} u$, niin θ esiintyy kahdesti sanassa u.



(c) Jos $h \leq_{\mathrm{s}} \theta$, niin h on sanan u reuna.

Kuva 6.8: Lemman 6.5 todistus. Jos h on sanan $uz(u\theta^{-1})v_0\theta$ lyhin reuna, niin se on sanan u prefiksi. Tämä johtaa ristiriitaan joko sanan z maksimaalisuuden tai sanan u reunattomuuden kanssa.

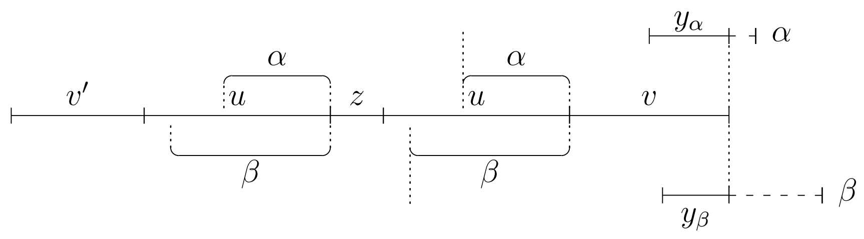
Todistus. Tehdään vastaoletus, että h on sanan $(uzu\theta^{-1})v_0\theta$ lyhin reuna. Sana h on lauseen 2.6 mukaan reunaton ja esiintyy kahdesti sanassa w, joten $|h| \leq \tau_2(w) = |u|$. Jos $|h| = |u|$, niin $h = u$, jolloin tekijä u esiintyy ainakin kaksi kertaa sanassa uv. Tämä tilanne on esitetty kuvassa 6.8a. Tämä on ristiriidassa sen kanssa, että z on mahdollisimman pitkä. Näin ollen $|h| < |u|$, jolloin h on sanan u aito prefiksi.

Koska h ja θ ovat sanan $(uzu\theta^{-1})v_0\theta$ suffikseja, niin $\theta \leq_{\mathrm{s}} h$ tai $h \leq_{\mathrm{s}} \theta$.

Oletetaan, että θ on sanan h suffiksi, kuten kuvassa 6.8b. Koska h on sanan u aito prefiksi, niin θ esiintyy sanassa u myös muunakin kuin suffiksina. Tämä tarkoittaa, että maksimisuffiksi θ esiintyy kaksi kertaa sanassa u. Tämä on ristiriidassa lauseen 2.16 kanssa.

Oletetaan, että h on sanan θ suffiksi, kuten kuvassa 6.8c. Tällöin h on myös sanan u suffiksi. Koska h on myös sanan u aito prefiksi, niin h on sanan u reuna. Tämä on ristiriidassa sanan u reunattomuuden kanssa. $\square$

Edellisessä lemmassa on keskeisessä osassa tekijän u maksimisuffiksi jonkin leksikografisen järjestyksen suhteen. Tällaiset maksimisuffiksit ovat myös myöhemmin tärkeässä asemassa, koska niistä johdetaan lisää pitkiä reunatto-



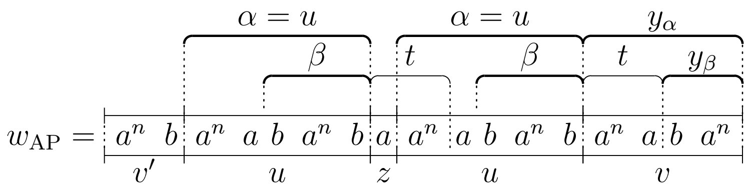
Kuva 6.9: Määritelmän 6.6 tekijät.

mia tekijöitä. Seuraavassa määritelmässä nimetään kriittisen tekijöihinjaon lauseen 4.1 tapaan kaksi tekijän u maksimisuffiksia ja määritellään lisäksi näistä saatavia sanan w suffikseja luvun 6.1 tapaan.

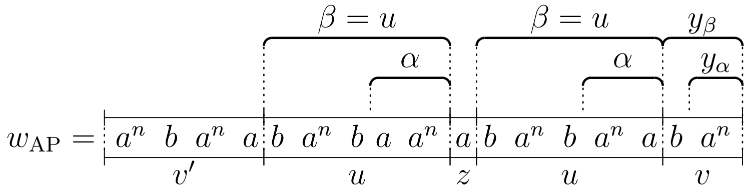
Määritelmä 6.6. Olkoon $w = v'uzuv$ kuten määritelmässä 6.4. Olkoon $\lhd$ aakkoston järjestys ja $\blacktriangleleft$ sen käänteisjärjestys. Olkoot lisäksi

- α sanan u $\lhd$-maksimisuffiksi,
- β sanan u $\blacktriangleleft$-maksimisuffiksi,
- y_α sanan uv α-suffiksi,
- y_β sanan uv β-suffiksi,
- $y = \begin{cases} y_\alpha, & \text{jos } y_\alpha \leq_{\mathrm{s}} y_\beta, \\ y_\beta, & \text{jos } y_\beta \leq_{\mathrm{s}} y_\alpha, \end{cases}$
- $\xi \in \{\alpha, \beta\}$ ja lisäksi $y_\xi = y$, sekä
- γ lyhyempi sanoista α ja β.

Sana y on aina määritelty, koska sekä y_α että y_β ovat sanan uv suffikseja. Sanat α ja β ovat erisuuret, kun sanassa u esiintyy ainakin kahta eri kirjainta, toisin sanoen kun $|u| \geq 2$. Tällöin α ja β myös alkavat eri kirjaimilla, joten y_α ja y_β ovat erisuuret, jos ne eivät ole molemmat tyhjiä sanoja. Sanan ξ määritelmä on siis yksikäsitteinen, jos y_α ja y_β ovat erisuuret. Seuraavan esimerkin mukaan on mahdollista, että y_α ja y_β ovat tyhjiä sanoja, jolloin voidaan valita $\xi = \alpha$ tai $\xi = \beta$.



(a) Tässä $\xi = \gamma = \beta$ ja $y = y_\beta$.



(b) Tässä $\xi = \gamma = \alpha$ ja $y = y_\alpha$.

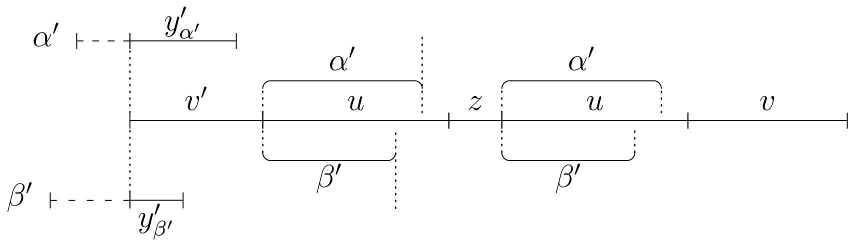
Kuva 6.10: Määritelmä 6.6 sovellettuna Assous'n–Pouzet'n sanaan w_{AP}.

Esimerkki 6.7. Tarkastellaan esimerkin 6.3 tekijöihinjakoa, jossa $u = bba$ ja $v = ba$. Kun järjestykset $\lhd$ ja $\blacktriangleleft$ määritellään niin, että $a \lhd b$ ja $b \blacktriangleleft a$, niin $\alpha = bba = u$ ja $\beta = a$. Näin ollen sekä α-suffiksi y_α että β-suffiksi y_β ovat tyhjiä sanoja.

Jos määritelmässä 6.6 oletetaan lisäksi, että t' on mahdollisimman lyhyt, niin tällöin lemman 6.4 mukaan α ja β ovat myös sanan $t'u$ $\lhd$- ja $\blacktriangleleft$-maksimisuffiksit.

Esimerkki 6.8. Assous'n–Pouzet'n sanan tekijöihinjaoille saadaan nyt kuvassa 6.10 esitetyt tekijät, kun aakkoston $\{a, b\}$ kaksi järjestystä nimetään niin, että $b \lhd a$ ja $a \blacktriangleleft b$. Lemmaa 6.5 ei voi soveltaa kumpaankaan sanan w_{AP} tekijöihinjakoon, koska kummassakin tekijöihinjaossa α ja β esiintyvät vain kerran tekijässä uv. Lemmaa 6.4 taas voidaan käyttää kuvassa 6.10a, koska t' on mahdollisimman lyhyt. Lemma ei kuitenkaan ilmaise tässä yhtään mitään, koska t' on tyhjä sana.

Määritelmästä 6.6 voidaan saada peilikuvaversio tarkastelemalla sanan w peilausta.



Kuva 6.11: Määritelmän 6.7 tekijät vastaavat määritelmän 6.6 tekijöitä sanan w peilaukselle (vertaa kuva 6.9).

Määritelmä 6.7 (Määritelmän 6.6 peilattu versio)**.** Olkoon $w = v'uzuv$ kuten määritelmässä 6.4. Olkoon $\lhd$ aakkoston järjestys ja $\blacktriangleleft$ sen käänteisjärjestys. Olkoot lisäksi α', β', $y'_{\alpha'}$, $y'_{\beta'}$ ne sanat, joiden peilauksille pätevät seuraavat väitteet:

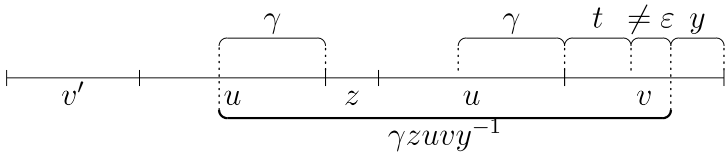
- $\overline{\alpha'}$ on sanan $\overline{u}$ $\lhd$-maksimisuffiksi,
- $\overline{\beta'}$ on sanan $\overline{u}$ $\blacktriangleleft$-maksimisuffiksi,
- $\overline{y'_{\alpha'}}$ on sanan $\overline{v'u}$ $\overline{\alpha'}$-suffiksi ja
- $\overline{y'_{\beta'}}$ on sanan $\overline{v'u}$ $\overline{\beta'}$-suffiksi.

Olkoot lisäksi

- $y' = \begin{cases} y'_{\alpha'}, & \text{jos } y'_{\alpha'} \leq_{\mathrm{p}} y'_{\beta'}, \\ y'_{\beta'}, & \text{jos } y'_{\beta'} \leq_{\mathrm{p}} y'_{\alpha'}, \end{cases}$
- $\xi' \in \{\alpha', \beta'\}$ ja lisäksi $y'_{\xi'} = y'$, sekä
- γ' lyhyempi sanoista α' ja β'.

Nämä sanat on esitetty kuvassa 6.11.

Koska määritelmän mukaan y on sanan ξ aito prefiksi, on se itse asiassa myös sanan yv ξ-suffiksi. Tarkastelu jakautuu nyt kahteen osaan sen mukaan, onko $|v| < |t| + |y|$ vai ei. Myös Assous'n–Pouzet'n sanan tekijöihinjaot jakautuvat näihin tapauksiin: kuvan 6.10a tekijöihinjaossa $|v| = |t| + |y|$ ja kuvan 6.10b tekijöihinjaossa $|v| > |t| + |y|$.



Kuva 6.12: Lemman 6.6 mukaan jos $|v| > |t| + |y|$, niin sanalla w on ainakin tekijän $\gamma zuvy^{-1}$ pituinen reunaton tekijä.

6.3 Tapaus $|v| > |t| + |y|$

Tässä luvussa käsitellään tapaus, jossa $|v| > |t| + |y|$. Seuraava lemma antaa pitkän reunattoman tekijän tässä tapauksessa. Tämä reunaton tekijä saadaan tarkastelemalla sanan w ξ-kriittistä prefiksiä. Lemman muotoilu ei kerro, mikä saatu reunaton tekijä on, mutta antaa sen pituudelle alarajan. Tämä alaraja on kuvassa 6.12 esitetyn tekijän pituus.

Lemma 6.6. Olkoot $w = v'uzuv$, t ja t' kuten määritelmässä 6.4 ja olkoot y ja γ kuten määritelmässä 6.6. Jos $|v| > |t| + |y|$, niin $\tau(w) \geq |\gamma zuvy^{-1}|$.

Todistus. Todistetaan, että sanalla w on reunaton tekijä, jonka pituus on ainakin tekijän $\gamma zuvy^{-1}$ pituus. Olkoon $|v| > |t| + |y|$. Tällöin y on lyhyempi kuin v. Lauseen 6.2 mukaan sanan w ξ-kriittinen prefiksi on ainakin yhtä pitkä kuin wy^{-1}. Koska $y <_{\mathrm{s}} v$, voidaan sanan w ξ-kriittinen prefiksi kirjoittaa muodossa

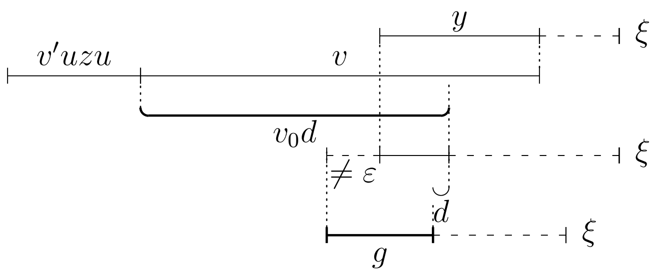
$$v'uzuv_0d,$$

missä v_0 on sanan v aito prefiksi, d on kirjain ja $|v_0d| \geq |v| - |y|$. Olkoon g sanan $v'uzuv_0$ ξ-suffiksi. Lauseen 6.3 mukaan tällöin

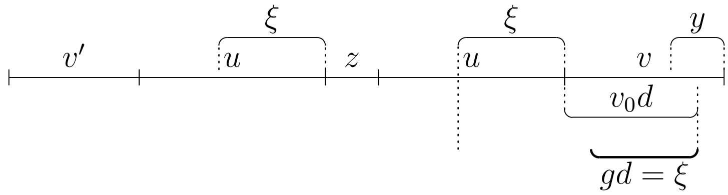
$$|v'uzuv_0| - |g| = \pi_\xi(v'uzuv_0) < \pi_\xi(v'uzuv_0d) = |v'uzu| + |v| - |y|.$$

Tämä tilanne on esitetty kuvassa 6.13. Todistus jakautuu nyt eri osiin sen mukaan, onko ξ sanan $v'uzuv_0d$ suffiksi vai ei.

Oletetaan ensin, että $gd = \xi$. Tällöin saadaan kuvan 6.14 tilanne. Koska γ esiintyy nyt kahdesti sanassa γv, lemman 6.5 mukaan sana $uzuv_0d$ on



Kuva 6.13: Lemman 6.6 todistuksen alku.



Kuva 6.14: Kun $gd = \xi$, niin ξ on tekijän $\xi v_0 d$ reuna. Tällöin lemman 6.5 mukaan $uzuv_0d$ on reunaton.

reunaton. Koska $|u| \geq |\gamma|$ ja $|v_0d| \geq |vy^{-1}|$, saadaan nyt

$$\tau(w) \geq |uzuv_0d| \geq |\gamma zuvy^{-1}|.$$

Oletetaan sitten todistuksen loppuun asti, että $gd \neq \xi$. Tällöin sanalla ξ on prefiksi gc, missä c on kirjain ja $c \neq d$. Nyt joko $c \lhd d$ tai $d \lhd c$, missä jälkimmäinen tapaus itse asiassa palautuu edelliseen tapaukseen.

Oletetaan, että $c \lhd d$. Tarkastellaan tekijää βzuv_0d. Jos se on reunaton, niin sijoittamalla epäyhtälöt $|\beta| \geq |\gamma|$ ja $|v_0d| \geq |vy^{-1}|$ saadaan

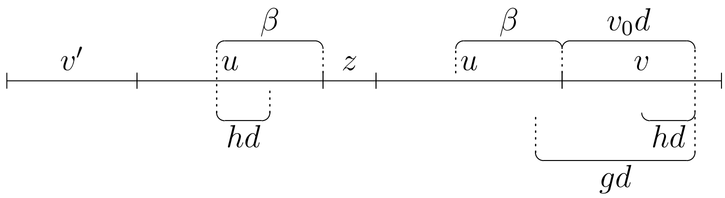
$$\tau(w) \geq |\beta zuv_0d| \geq |\gamma zuvy^{-1}|,$$

mikä on lauseen väite. Jatkossa oletetaan, että tekijällä βzuv_0d on reuna.

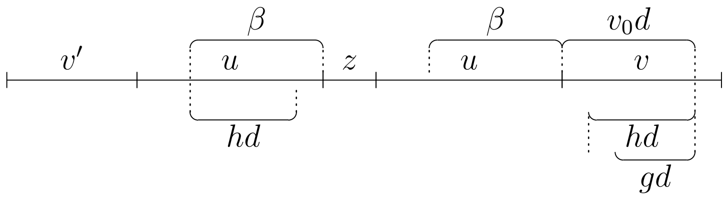
Oletetaan nyt, että $c \lhd d$ ja tekijä βzuv_0d ei ole reunaton. Olkoon hd tekijän βzuv_0d lyhin reuna. Sanat g ja h ovat saman sanan suffikseja, sekä β ja h ovat saman sanan prefiksejä. Todistus jakautuu nyt kolmeen tapaukseen:

(i) $h \leq_{\mathrm{s}} g$ ja $h <_{\mathrm{p}} \beta$,

(ii) $g <_{\mathrm{s}} h$ ja $h <_{\mathrm{p}} \beta$, sekä



Kuva 6.15: Lemman 6.6 todistus, kun $h \leq_{\mathrm{s}} g$ ja $hd \leq_{\mathrm{p}} \beta$. Tämä tapaus johtaa ristiriitaan.



Kuva 6.16: Lemman 6.6 todistus, kun $g <_{\mathrm{s}} h$ ja $hd \leq_{\mathrm{p}} \beta$. Tämä tapaus johtaa ristiriitaan.

(iii) $\beta \leq_{\mathrm{p}} h$.

(i) Oletetaan, että $h \leq_{\mathrm{s}} g$ ja $h <_{\mathrm{p}} \beta$. Tämä on esitetty kuvassa 6.15. Koska $hd \leq_{\mathrm{p}} \beta$ ja $d \blacktriangleleft c$, niin $\beta \blacktriangleleft hc$. Lemman 2.14 mukaan hc ei voi olla sanan u tekijä, koska β on sanan u $\blacktriangleleft$-maksimisuffiksi. Toisaalta saadaan prefiksi- ja suffiksiketju

$$hc \leq_{\mathrm{s}} gc \leq_{\mathrm{p}} \xi \leq_{\mathrm{s}} u,$$

jonka mukaan hc on sanan u tekijä. Näin on saatu ristiriita.

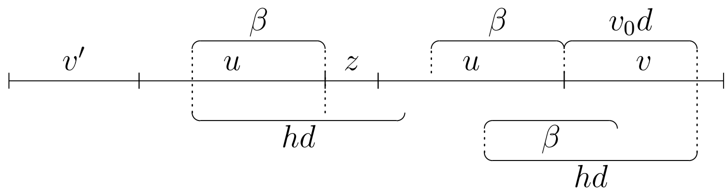
(i) Oletetaan, että $g <_{\mathrm{s}} h$ ja $h <_{\mathrm{p}} \beta$. Tämä tilanne on esitetty kuvassa 6.16. Nyt saadaan prefiksi- ja suffiksiketju

$$gd <_{\mathrm{s}} hd \leq_{\mathrm{p}} \beta \leq_{\mathrm{s}} u,$$

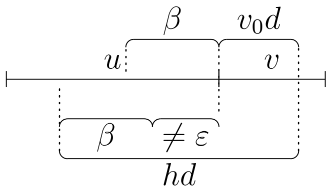
joten gd on sanan u (aito) tekijä. Lemman 2.14 mukaan nyt $gd \trianglelefteq \alpha$ ja $gd \underline{\blacktriangleleft} \beta$. Seuraavaksi näytetään, että riippumatta sanan ξ valinnasta päädytään ristiriitaan.

Oletetaan, että $\xi = \alpha$. Koska nyt $gc \leq_{\mathrm{p}} \alpha$ ja $c \triangleleft d$, niin $\alpha \triangleleft gd$. Tämä on ristiriidassa kaavan $gd \trianglelefteq \alpha$ kanssa.

Oletetaan, että $\xi = \beta$. Tällöin hd on sanan ξ prefiksi, joten h on sanan



Kuva 6.17: $\beta \leq_{\mathrm{p}} h$



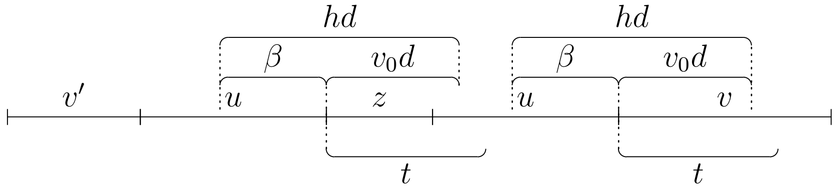
Kuva 6.18: $\beta v_0 <_{\mathrm{s}} h$

ξ aito prefiksi. Lisäksi h on sanan $v'uzuv_0$ suffiksi, joka on pidempi kuin g. Tämä on ristiriidassa sen kanssa, että g on sanan $v'uzuv_0$ ξ-suffiksi.

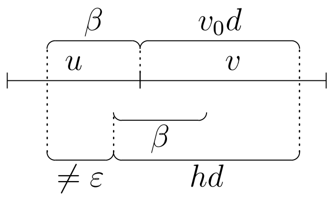
(iii) Oletetaan, että $\beta \leq_{\mathrm{p}} h$. Tämä on esitetty kuvassa 6.17. Sana hd on lyhimpänä reunana reunaton ja se esiintyy kahdesti sanassa w, joten $|hd| \leq |u|$. Sanat hd ja $\beta v_0 d$ ovat kumpikin sanan uv_0d suffikseja.

Oletetaan, että $\beta v_0 <_{\mathrm{s}} h$, kuten kuvassa 6.18. Tällöin β esiintyy kaksi kertaa sanassa u, koska β on sanan hd prefiksi ja $|hd| \leq |uv_0|$. Tämä on ristiriidassa sen kanssa, että β on sanan u $\blacktriangleleft$-maksimisuffiksi.

Oletetaan, että $h = \beta v_0$. Koska $t = v \wedge_{\mathrm{p}} zu$, niin $\beta t = \beta v \wedge_{\mathrm{p}} \beta zu$. Koska hd on sanojen βzu ja βv yhteinen prefiksi, niin hd on sanan βt prefiksi. Koska nyt $hd = \beta v_0 d$, saadaan, että $v_0 d$ on sanan t prefiksi. Näin ollen $|v_0 d| \leq |t|$. Koska sanan $v_0 d$ määrittelyn mukaan $|v_0 d| \geq |v| - |y|$, saadaan



Kuva 6.19: $h = \beta v_0 d$



Kuva 6.20: $h <_{\mathrm{s}} \beta v_0$

nyt sijoittamalla johdetuksi kaava

$$|v| \leq |v_0 d| + |y| \leq |t| + |y|,$$

joka on ristiriidassa lemman oletuksen kanssa.

Oletetaan, että $h <_{\mathrm{s}} \beta v_0$. Tällöin maksimisuffiksi β esiintyy kahdesti sanassa βv, kuten kuvasta 6.20 ilmenee. Lemman 6.5 mukaan sana $uz(uv_0h^{-1})\beta$ on reunaton. Nyt käyttämällä epäyhtälöitä $|u| \geq |hd|$, $|\beta| \geq |\gamma|$ ja $|v| \leq |v_0d| + |y|$ saadaan johdetuksi

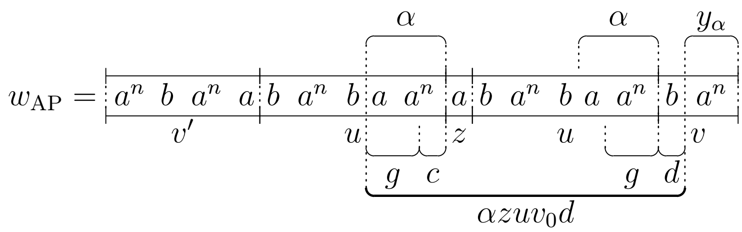
$$\begin{aligned}
\tau(w) &\geq |uz(uv_0h^{-1})\beta| && (uz(uv_0h^{-1})\beta \text{ on reunaton}) \\
&= |u| + |zuv_0| - |h| + |\beta| \\
&\geq |h| + |d| + |zuv_0| - |h| + |\beta| && (|u| \geq |hd|) \\
&= |\beta zuv_0 d| \\
&\geq |\gamma zuvy^{-1}|, && (|\beta| \geq |\gamma| \text{ ja } |v| \leq |v_0d| + |y|)
\end{aligned}$$

mikä on lauseen väite.

Lemman väite on nyt todistettu tapauksessa, jossa $c \lhd d$. Tapaus $d \lhd c$ on samanlainen, koska tällöin $c \blacktriangleleft d$ ja todistus saadaan vaihtamalla β ja α keskenään sekä $\lhd$ ja $\blacktriangleleft$ keskenään. Näin ollen lemma on todistettu. $\square$

Lemmasta 6.6 saadaan myös peilaamalla seuraava versio.

Lemma 6.7 (Lemman 6.6 peilattu versio)**.** Olkoot $w = v'uzuv$, t ja t' kuten määritelmässä 6.4 ja olkoot y' ja γ' kuten määritelmässä 6.7. Jos $|v'| > |t'| + |y'|$, niin $\tau(w) \geq |(y')^{-1}v'uz\gamma'|$.



Kuva 6.21: Assous'n–Pouzet'n sanalla w_{AP} on lemman 6.6 mukainen reunaton tekijä.

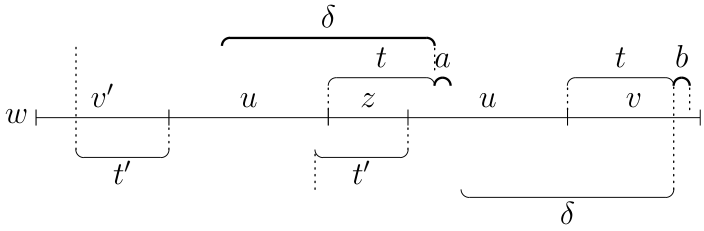
Esimerkki 6.9. Assous'n–Pouzet'n sanan kuvan 6.21 tekijöihinjaossa pätee tämän luvun ehto $|v| > |t| + |y|$. Seuraavassa käydään läpi lemman 6.6 todistus tälle tekijöihinjaolle. Tässä tekijöihinjaossa $\xi = \gamma = \alpha = a^{n+1}$ ja $y = a^n$. Sanan w_{AP} α-kriittinen suffiksi on $v'uzub$, joten $v_0 = \varepsilon$ ja $d = b$. Sanan $v'uzu$ α-suffiksi on $g = a^n$. Koska $a^n b \neq a^{n+1}$, ollaan tilanteessa $gd \neq \xi$. Tällöin $c = a$. Koska $b \lhd a$, niin $d \lhd c$. Joudutaan siis vaihtamaan todistuksen merkintöjä. Reunaton sana saadaan heti ensimmäisenä vaihtoehtona, koska sana $\alpha zuv_0 d = a^{n+2}ba^nba^{n+1}b$ on reunaton. Tämä sana on sama kuin lemmassa mainittu $\gamma zuvy^{-1}$.

6.4 Tapaus $|v| \leq |t| + |y|$ ja $t \neq v$

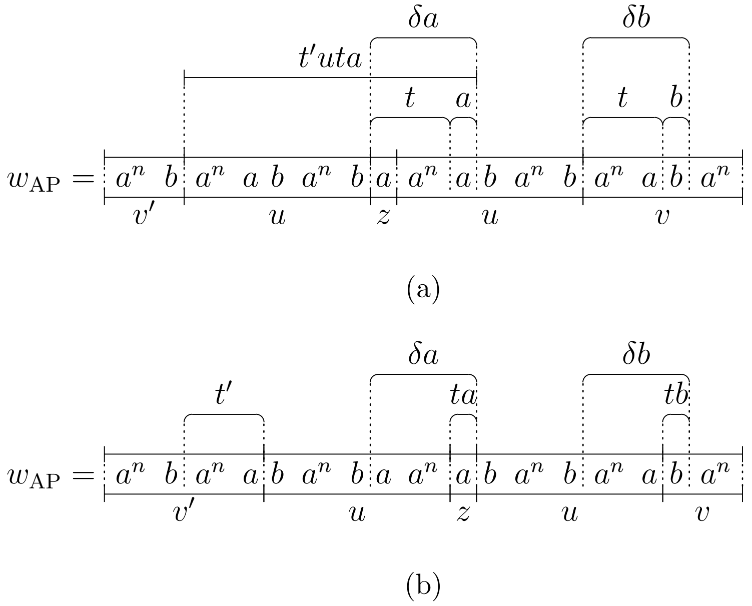
Tässä luvussa käsitellään tapaus, jossa $|v| \leq |t| + |y|$. Lisäksi oletetaan, että v ei ole sanan zu prefiksi. Pitkä reunaton tekijä annetaan lemmassa 6.9. Tämä tekijä saadaan tarkastelemalla sanan w erään pitkän suffiksin kriittistä prefiksiä tietyn maksimisuffiksin suhteen. Juoksevana esimerkkinä tarkastellaan Assous'n–Pouzet'n sanan tekijöihinjakoa.

Seuraava määritelmä on voimassa ilman oletusta $|v| \leq |t| + |y|$. Koska v ei ole sanan zu prefiksi ja zu ei voi olla sanan v prefiksi, voidaan nimetä ensimmäiset kirjaimet, joilla kumpikin tekijä eroaa yhteisen prefiksin jälkeen.

Määritelmä 6.8. Olkoon $t \neq v$. Olkoon a ja b ne kirjaimet, joille $ta \leq_{\mathrm{p}} zu$ ja $tb \leq_{\mathrm{p}} v$. Olkoon $\lhd^a$ jokin järjestys, jonka suhteen a on aakkoston A maksimialkio. Olkoon δ sana, jolle δa on tekijän $t'uta$ $\lhd^a$-maksimisuffiksi. Nämä tekijät ovat kuvassa 6.22.



Kuva 6.22: Määritelmä 6.8. Sana δ valitaan niin, että δa on sanan $t'uta$ $\lhd^a$-maksimisuffiksi.



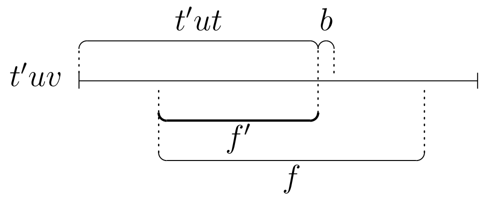
Kuva 6.23: Määritelmä 6.8 sovellettuna Assous'n–Pouzet'n sanaan w_{AP}.

Esimerkki 6.10. Kuvassa 6.23 on määritelmän 6.8 tekijät sanalle w_{AP}.

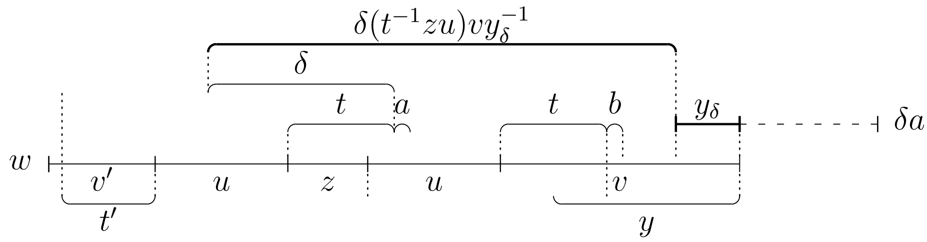
Tämän luvun loppun asti oletetaan, että $|v| \leq |t|+|y|$. Seuraavan lemman mukaan tällöin jokainen sanan $t'uv$ tekijä on leksikografisesti vähemmän kuin δa järjestyksen $\lhd^a$ suhteen.

Lemma 6.8. Olkoon $|v| \leq |t| + |y|$ ja $t \neq v$, sekä olkoot δ, a ja $\lhd^a$ kuten määritelmässä 6.8. Jos f on sanan $t'uv$ tekijä, niin $f \lhd^a \delta a$.

Todistus. Sanalla $t'uv$ on prefiksinä $t'ut$ ja suffiksina y. Jos f on prefiksin $t'ut$ tekijä, niin väite seuraa suoraan lemmasta 2.14. Jos taas f on suffiksin y tekijä, niin se on myös sanan ξ tekijä ja edelleen sanan u tekijä, jolloin väite



Kuva 6.24: Tapaus, jossa sanan $t'uv$ prefiksi $t'ut$ on limittäin tekijän f esiintymän kanssa.



Kuva 6.25: Lemman 6.9 mukaan sanan $t'uta$ δa-maksimisuffiksi y_δ on lyhyempi kuin $t^{-1}v$ ja sana $\delta(t^{-1}zu)vy_\delta^{-1}$ on reunaton.

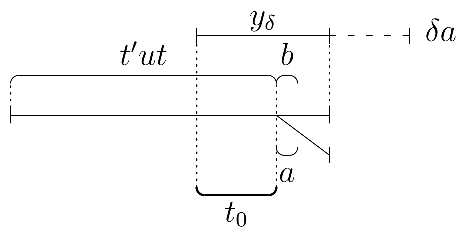
seuraa jälleen lemmasta 2.14. Jäljelle jää nyt tapaus, jossa f alkaa tekijässä $t'ut$ ja loppuu tekijässä y. Itse asiassa todistetaan yleisempi tapaus, jossa f ei ole sanojen $t'ut$ eikä $t^{-1}v$ tekijä.

Oletetaan, että $t'ut$ ja f ovat limittäin kuvan 6.24 mukaisesti ja f' on näiden sanojen limittyvä osa. Toisin sanoen sanalla $t'ut$ on suffiksi f', jolle $f'b$ on sanan f prefiksi. Koska $b \lhd^a a$, niin tällöin $f \lhd^a f'a$. Toisaalta $f'a$ on sanan $t'uta$ suffiksi, joten $f'a \unlhd^a \delta a$, koska δa on sanan $t'uta$ $\lhd^a$-maksimisuffiksi. Väite saadaan nyt ketjusta

$$f \lhd^a f'a \unlhd^a \delta a. \qquad \square$$

Esimerkki 6.11. Kuvan 6.23a tekijöihinjaossa $\delta a = a^{n+2}$, ja tämä on koko sanan pisin kirjaimen a potenssi. Sanan $t'uv$ pisin kirjaimen a potenssi on a^{n+1}. On siis selvää, että jokainen sanan $t'uv$ tekijä on järjestyksen $\lhd^a$ suhteen pienempi kuin δa.

Seuraava lemma antaa pitkän reunattoman tekijän, kun tämän luvun oletukset $|v| \leq |t|+|y|$ ja $t \neq v$ ovat voimassa. Lemman tilanne on kuvassa 6.25.



Kuva 6.26: Lemman 6.9 vastaoletus, että $|y_\delta| \geq |v| - |t|$.

Lemma 6.9. Olkoon $|v| \leq |t| + |y|$ ja $t \neq v$, sekä olkoot δ ja a kuten määritelmässä 6.8. Olkoon y_δ sanan w δa-suffiksi. Tällöin $|y_\delta| < |v| - |t|$ ja sana $\delta(t^{-1}zu)vy_\delta^{-1}$ on reunaton.

Todistus. Aluksi todistetaan väitteen epäyhtälö $|y_\delta| < |v| - |t|$. Tehdään vastaoletus, että $|y_\delta| \geq |v| - |t|$, kuten kuvassa 6.26. Tällöin sanalla $t'ut$ on suffiksi t_0, joka on myös sanan y_δ prefiksi. Näin myös $t_0 b$ on sanan y_δ prefiksi ja edelleen sanan δa aito prefiksi. Koska $b \lhd^a a$, niin tällöin $\delta a \lhd^a t_0 a$. Tämä on ristiriidassa sen kanssa, että δa on sanan $t'uta$ $\lhd^a$-maksimisuffiksi, koska $t_0 a$ on sanan $t'uta$ suffiksi. Näin on todistettu epäyhtälö $|y_\delta| < |v| - |t|$.

Sana y_δ on siis sanan $t^{-1}v$ aito suffiksi. Nyt y_δ on myös sanan $\delta(t^{-1}zu)v$ δa-suffiksi. Olkoon x sanan $\delta(t^{-1}zu)v$ δa-kriittinen prefiksi. Lauseen 6.2 mukaan se on tällöin pidempi kuin $\delta(t^{-1}zu)t$, joten voidaan kirjoittaa

$$x = \delta(t^{-1}zu)tv_\delta d,$$

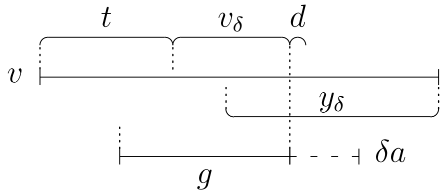
missä v_δ on sanan $t^{-1}v$ aito prefiksi, d on kirjain ja $|tv_\delta d| \geq |v| - |y_\delta|$. Olkoon nyt g sanan xd^{-1} δa-suffiksi. Lauseen 6.3 mukaan tällöin

$$|xd^{-1}| - |g| < |\delta(t^{-1}zu)v| - |y_\delta|.$$

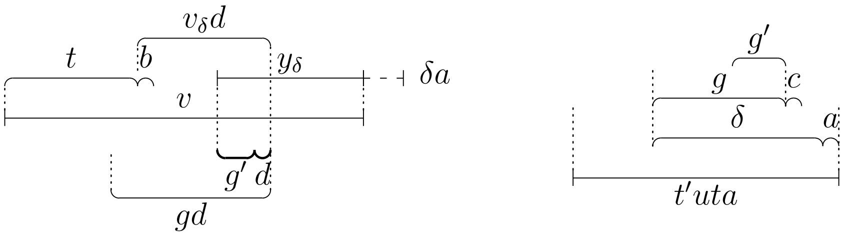
Tämä tilanne on esitetty kuvassa 6.27.

Lemman 6.8 mukaan δa ei ole sanan $t'uv$ tekijä. Koska toisaalta gd on sanan $t'uv$ tekijä, niin lemman 6.8 mukaan

$$gd \lhd^a \delta a. \tag{6.4}$$



Kuva 6.27: Sanan g esiintymä ulottuu enemmän vasemmalle kuin y_δ.



(a) Tapaus, jossa $v_\delta d$ ja y_δ ovat limittäin.

(b) Sana g' on sanan $t'uta$ tekijä.

Kuva 6.28: Jos $|tv_\delta d| > |v| - |y_\delta|$, niin päädytään ristiriitaan.

Sanan g määritelmän mukaan on olemassa jokin kirjain c, jolle $gc \leq_{\mathrm{p}} \delta a$. Lisäksi lauseen 6.3 mukaan $c \neq d$, koska muuten sanojen x ja xd^{-1} δa-periodit olisivat samat. Näin ollen gd ei ole sanan δa prefiksi, joten kaavan (6.4) mukaan $d \lhd^a c$.

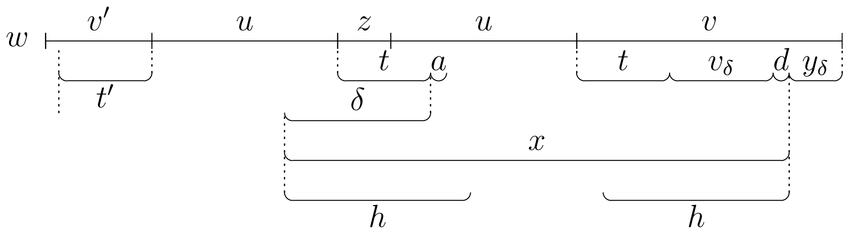
Todistetaan seuraavaksi, että epäyhtälössä $|tv_\delta d| \geq |v| - |y_\delta|$ vain yhtäsuuruus on mahdollinen. Oletetaan, että $|tv_\delta d| > |v| - |y_\delta|$. Tällöin $v_\delta d$ ja y_δ ovat limittäin kuvan 6.28a mukaisesti. Tällöin on jokin sana g', jolle $(v_\delta d(g'd)^{-1})y_\delta = t^{-1}v$. Koska $g'd$ on sanan y_δ prefiksinä sanan δa aito prefiksi ja $d \lhd^a c$, saadaan kaava $\delta a \lhd^a g'c$. Tämä on ristiriidassa lemman 2.14 kanssa, koska $g'c$ on sanan $t'uta$ tekijä ketjun

$$g'c \leq_{\mathrm{s}} gc \leq_{\mathrm{p}} \delta a \leq_{\mathrm{s}} t'uta$$

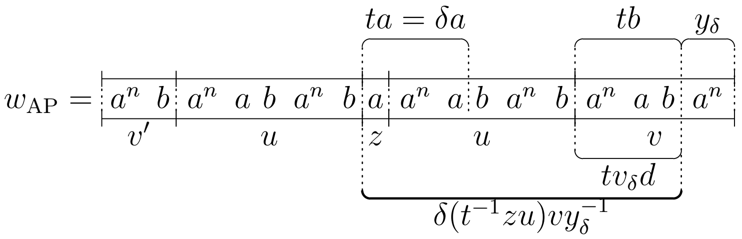
mukaan (kuten kuvassa 6.28b) ja δa on sanan $t'uta$ $\lhd^a$-maksimisuffiksi.

Näin ollen $|tv_\delta d| = |v| - |y_\delta|$, mikä tarkoittaa, että $v = tv_\delta d y_\delta$. Tämä on esitetty kuvassa 6.29. Nyt $x = \delta(t^{-1}zu)vy_\delta^{-1}$.

Lopuksi todistetaan, että x on reunaton. Tehdään vastaoletus, että x ei ole reunaton. Olkoon tällöin h sen lyhin reuna. Tällöin $|h| \leq |u|$, koska h on



Kuva 6.29: Tapaus, jossa $|tv_\delta d| = |v| - |y_\delta|$.



Kuva 6.30: Lemman 6.6 reunaton tekijä Assous'n–Pouzet'n sanan tekijöihinjaolle, jossa oletus $|v| \leq |t| + |y|$ on voimassa.

reunaton ja esiintyy kahdesti sanassa w. Tällöin $|h| \leq |uvy_\delta^{-1}|$, joten

$$h \leq_{\mathrm{s}} uvy_\delta^{-1}. \tag{6.5}$$

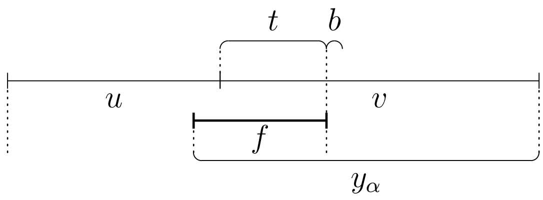
Määritelmän 6.3 mukaan

$$\pi_{\delta a}(x) = |x|, \tag{6.6}$$

koska x on itsensä δa-kriittinen prefiksi. Sanat δa ja h ovat nyt sanan x prefiksejä. Jos $h <_{\mathrm{p}} \delta a$, niin sanan x δa-suffiksi on epätyhjä, mikä on ristiriidassa yhtälön 6.6 kanssa. Näin ollen $\delta a \leq_{\mathrm{p}} h$. Yhdessä kaavan 6.5 kanssa tämä tarkoittaa, että δa on sanan uv tekijä. Tämä on ristiriidassa lemman 6.8 kanssa. Näin on saatu todistetuksi lemman molemmat väitteet. □

Esimerkki 6.12. Kuvassa 6.30 on edellisen lemman pitkä reunaton tekijä sanalle w_{AP}. Tässä tapauksessa $g = \delta = a^{n+1}$, $d = b$ ja $c = a$. Saatu tekijä $a^{n+2}ba^nba^{n+1}b$ on sama kuin kuvan 6.21 tekijä, joka saatiin toisesta tekijöihinjaosta lemmalla 6.6. Tämä tekijä on itse asiassa toinen sanan w_{AP} pisimmistä reunattomista tekijöistä.

Lopuksi esitetään tuloksia sanan w tekijöiden pituuksista.



Kuva 6.31: Määritelmä 6.9

Lemma 6.10. Olkoon $t \neq v$, ja olkoon δ kuten määritelmässä 6.8. Tällöin

$$|\delta| > |t| + |t'| - z.$$

Todistus. Jos $|t| + |t'| - |z| < 0$, niin väite seuraa suoraan. Oletetaan nyt, että $|t| + |t'| - |z| \geq 0$. Tällöin sanan uzu prefiksi uta ja suffiksi $t'u$ ovat limittäin, joten on olemassa jokin epätyhjä sana h, joka on sanan uta suffiksi ja sanan $t'u$ prefiksi ja jolle $|h| = |ta| + |t'| - |z|$. Näin ollen h on sanan $t'uta$ reuna, koska $t'u$ on sanan $t'uta$ aito prefiksi. Lauseen 2.16 mukaan δa on pidempi kuin h, koska δa on sanan $t'uta$ $\lhd^a$-maksimisuffiksi. Näin ollen

$$|\delta| \geq |h| > |t| + |t'| - |z|. \qquad \square$$

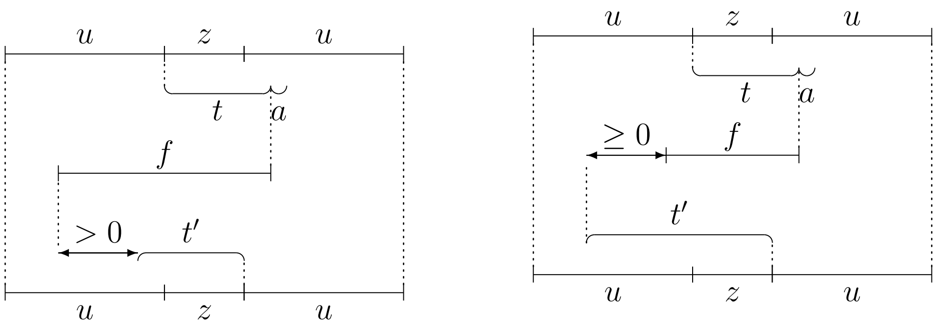
Ehdot $|v| \leq |t| + |y|$ ja $t \neq v$ ovat ekvivalentteja sen kanssa, että sanan uv prefiksi utb ja suffiksi y ovat limittäin. Erityisesti sanat utb ja y_α ovat myös limittäin. Näin ollen voidaan tehdä seuraava määrittely.

Määritelmä 6.9. Olkoon $|v| \leq |t| + |y|$ ja $t \neq v$, sekä olkoon y_α kuten määritelmässä 6.6. Olkoon tällöin sana f sanan y_α prefiksi ja sanan ut suffiksi, jolle $uv = utf^{-1}y_\alpha$. Toisin sanoen $f = y_\alpha(t^{-1}v)^{-1}$. Tämä on esitetty kuvassa 6.31.

Edellisen määritelmän mukaan $|t| = |v| - |y_\alpha| + |f|$. Koska y_α on sanan α aito prefiksi, saadaan seuraava lemma.

Lemma 6.11. $|t| > |v| - |\alpha| + |f|$.

Seuraavassa lemmassa oletetaan, että ξ ja y on saatu määritelmässä 6.6 käyttämällä järjestystä, jossa kirjain a on maksimi. Lemma antaa sanan



(a) Lemman 6.12 väite

(b) Vastaoletus, jonka mukaan $|f| \leq |t| + |t'| - |z|$.

Kuva 6.32: Lemman 6.12 väite (a) ja sen vastaoletus (b). Pituus $|t|+|t'|-|z|$ on sanan t' vasemman päätepisteen ja sanan t oikean päätepisteen välinen etäisyys.

f pituudelle saman alarajan kuin lemma 6.10 sana δ pituudelle, jos t' on mahdollisimman lyhyt. Lemman väite on esitetty kuvassa 6.32a.

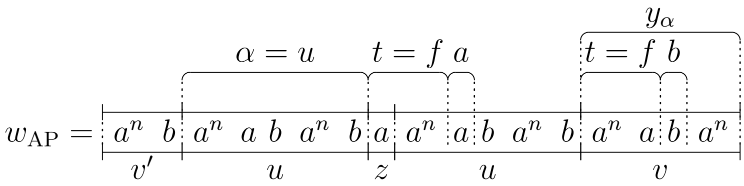
Lemma 6.12. Olkoon $|v| \leq |t| + |y|$ ja $t \neq v$, missä y on määritelmän 6.6 mukainen ξ-suffiksi, kun vaadittuna järjestyksenä käytetään määritelmän 6.8 mukaista järjestystä $\lhd^a$. Olkoon f kuten määritelmässä 6.9. Tällöin jos t' on mahdollisimman lyhyt, niin

$$|f| > |t| + |t'| - |z|.$$

Todistus. Tehdään vastaoletus, että $|f| \leq |t|+|t'|-|z|$, kuten kuvassa 6.32b. Tällöin fa on sanan $t'u$ tekijä. Koska fb on sanan y_α prefiksi ja y_α on sanan α (aito) prefiksi sekä $b \lhd^a a$, saadaan, että $\alpha \lhd^a fa$. Tämä on ristiriidassa lemman 2.14 kanssa, koska lemman 6.4 mukaan α on myös sanan $t'u$ $\lhd^a$-maksimisuffiksi. □

Esimerkki 6.13. Kuvassa 6.33 on esitetty tekijä f sanan w_{AP} vasemmanpuoleiselle tekijöihinjaolle. Voidaan nähdä, että lemman 6.11 väite on voimassa, koska $|t| = n+1 > n = |v|-|\alpha|+|f|$. Lemmojen 6.10 ja 6.12 väitteet ovat myös voimassa, koska $|\delta| = |f| = n+1$ ja $|t| + |t'| - |z| = n$.

Kuten lemmalla 6.6 myös lemmalla 6.9 on peilikuvaversio, joka saadaan



Kuva 6.33: Määritelmän 6.9 tekijä f sanalle w_{AP} on sama kuin t.

tarkastelemalla sanan w peilauksen tekijöihinjakoa $\overline{w} = \overline{v}\,\overline{u}\,\overline{z}\,\overline{u}\,\overline{v'}$ ja määrittelemällä sille samanlaiset tekijät kuin sanalle w.

Lemma 6.13 (Lemman 6.9 peilattu versio)**.** Olkoon $t' \neq v'$. Olkoon tällöin

- a' se kirjain, jolle $a't' \leq_{\mathrm{s}} uz$,
- δ' sana, jolle $\overline{\delta'}a'$ on sanan $\overline{a't'ut} = \overline{t}\,\overline{u}\,\overline{t'}\,a'$ $\lhd^{a'}$-maksimisuffiksi,
- $y'_{\delta'}$ sana, jolle $\overline{y'_{\delta'}}$ on sanan $\overline{w}$ $\overline{\delta'}a'$-suffiksi ja
- $\lhd$ jokin aakkoston järjestys.

Olkoon $|v'| \leq |t'| + |y'|$, missä $\overline{y'}$ on lyhyempi sanan $\overline{w}$ suffikseista sanan $\overline{u}$ $\lhd$- ja $\blacktriangleleft$-maksimisuffiksien suhteen. Tällöin $|y'_{\delta'}| < |v'| - |t'|$ ja sana $(y'_{\delta'})^{-1}v'(uz(t')^{-1})\delta'$ on reunaton.

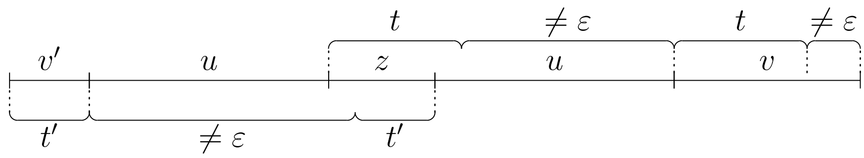
6.5 Päätulos

Tässä luvussa todistetaan luvun 6 ja koko tutkielman päälause 6.15. Seuraavassa lemmassa käsitellään lukujen 6.3 ja 6.4 tapaukset. Sen mukaan kaava $|w| \geq \frac{7}{3}\,\tau(w)$ ei voi olla voimassa, jos $t \neq v$ tai $t' \neq v'$.

Lemma 6.14. Olkoon $w = v'uzuv$ määritelmän 6.4 mukainen tekijöihinjako, $t = zu \wedge_{\mathrm{p}} v$ ja $t' = uz \wedge_{\mathrm{s}} v'$. Olkoon lisäksi $t \neq v$ tai $t' \neq v'$. Tällöin $\tau(w) > \frac{3}{7}|w|$.

Todistus. Tehdään vastaoletus, että

$$\tau(w) \leq \frac{3}{7}|w|. \tag{6.7}$$



Kuva 6.34: Tapaus, jossa $t <_{\mathrm{p}} v$ ja $t' = v'$

Todistus jakautuu osiin yhtäältä sen mukaan, ovatko oletukset $t \neq v$ ja $t' \neq v'$ voimassa samanaikaisesti, ja toisaalta sen mukaan, onko $|v| > |t| + |y|$ (tai symmetrisesti $|v'| > |t'| + |y'|$). Kussakin tapauksessa todistetaan, että kyseisistä oletuksista seuraavat epäyhtälöt eivät voi olla voimassa samaan aikaan kuin vastaoletus.

Tapaus 1: joko $t \neq v$ tai $t' \neq v'$. Tapaukset ($t \neq v$ ja $t' = v'$) ja ($t = v$ ja $t' \neq v'$) ovat symmetrisiä, koska huomautuksen 6.1 mukaan tapauksesta toiseen päästään käsittelemällä peilisanaa $\overline{w}$. Voidaan siis olettaa, että $t \neq v$ ja $t' = v'$, kuten kuvassa 6.34. Oletetaan myös, että t' on mahdollisimman lyhyt. Koska $t' = v'$, on nyt valittu tekijöihinjako, jossa muotoa uzu oleva tekijä esiintyy mahdollisimman vasemmalla.

Tapaus 1.1: $|v| > |t|+|y|$. Lemman 6.6 mukaan $\tau(w) \geq |\gamma zuvy^{-1}|$. Oletetaan ensin, että $|v'| \leq |v|$. Määritelmän 6.6 mukaan $|\gamma| > |y|$. Sijoittamalla saadaan nyt johdetuksi

$$\begin{aligned}
\tau(w) &\geq |\gamma zuvy^{-1}| \\
&> |zuv| \\
&= \frac{1}{2}(|zuv| + |zuv|) \\
&\geq \frac{1}{2}(|zuv| + |z| + |u| + |v'|) \\
&\geq \frac{1}{2}(|v'| + |u| + |zuv|) \\
&= \frac{1}{2}|w| \\
&> \frac{3}{7}|w|,
\end{aligned}$$

mikä on ristiriidassa vastaoletuksen kanssa.

Oletetaan sitten, että $|v'| > |v|$. Oletetaan tällöin aluksi, että $|\gamma z| \leq |v'|$, toisin sanoen $|\gamma z| \leq |t'|$. Koska γz ja t' ovat kumpikin sanan uz suffikseja, on nyt γz sanan t' suffiksi. Näin ollen $\gamma z u$ on sanan $t'u$ suffiksi. Olkoon $\lhd$ järjestys, jolle γ on sanan u $\lhd$-maksimisuffiksi. Lemman 6.4 mukaan γ on myös sanan $t'u$ $\lhd$-maksimisuffiksi. Tämä on ristiriidassa sen kanssa, että $\gamma \lhd \gamma z u$, koska $\gamma z u$ on sanan $t'u$ suffiksi.

Oletetaan nyt, että $|v'| > |v|$ ja $|\gamma z| > |v'|$. Nyt saadaan johdetuksi

$$\begin{aligned}
\tau(w) &\geq |\gamma z u v y^{-1}| \\
&= \frac{1}{2}(2|\gamma z| + 2(|u| + |v| - |y|)) \\
&> \frac{1}{2}(|v'| + |\gamma| + |z| + 2|u| + 2|v| - 2|y|) \quad (\text{koska } |\gamma z| > |v'|) \\
&> \frac{1}{2}(|v'| + |y| + |z| + 2|u| + 2|v| - 2|y|) \quad (\text{koska } |\gamma| > |y|) \\
&> \frac{1}{2}(|v'| + |z| + 2|u| + |v| + |y| - |y|) \qquad (\text{koska } |v| > |t| + |y| \geq |y|) \\
&= \frac{1}{2}|w| \\
&> \frac{3}{7}|w|,
\end{aligned}$$

mikä on ristiriidassa vastaoletuksen kanssa.

Tapaus 1.2: $|v| \leq |t| + |y|$. Lemmasta 6.10 saadaan nyt epäyhtälö

$$L_1 := |\delta| - |t| - |t'| + |z| - 1 \geq 0.$$

Lemman 6.9 mukaan

$$\begin{aligned}
\tau(w) &\geq |\delta(t^{-1} z u) v y_\delta^{-1}| \\
&= |\delta| - |t| + |z| + |u| + |v| - |y_\delta| \qquad (6.8) \\
&\geq |\delta| - |t| + |z| + |u| + |v| - |v| + |t| + 1 \quad (\text{koska } |y_\delta| \leq |v| - |t| - 1) \\
&= |\delta| + |z| + |u| + 1.
\end{aligned}$$

Koska vastaoletuksen mukaan $\tau(w) \leq \frac{3}{7}|w|$, saadaan epäyhtälö

$$3|v'| + 6|u| + 3|z| + 3|v| \geq 7|\delta| + 7|z| + 7|u| + 7,$$

joka saadaan termejä siirtelemällä muotoon

$$L_2 := 3|v'| - |u| - 4|z| + 3|v| - 7|\delta| - 7 \geq 0.$$

Toisaalta koska määritelmän mukaan $|y_\delta| \leq |\delta|$, niin sijoittamalla kaavaan 6.8 saadaan epäyhtälö

$$\tau(w) \geq -|t| + |z| + |u| + |v|,$$

joka vastaoletuksen $\tau(w) \leq \frac{3}{7}|w|$ kanssa tarkoittaa, että

$$3|v'| + 6|u| + 3|z| + 3|v| \geq -7|t| + 7|z| + 7|u| + 7|v|.$$

Termejä siirtelemällä saadaan nyt epäyhtälö

$$L_3 := 3|v'| - |u| - 4|z| - 4|v| + 7|t| \geq 0.$$

Kun lausekkeeseen L_1 sijoitetaan oletus $|t'| = |v'|$, saadaan

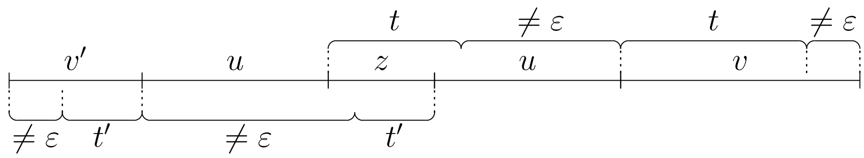
$$L_1' := |\delta| - |t| - |v'| + |z| - 1 \geq 0.$$

Nyt voidaan huomata, että

$$21L_1' + 4L_2 + 3L_3 = -7(|u| + |z| + |\delta| + 7).$$

Tämä on ristiriita, koska yhtälön vasen puoli on epänegatiivinen ja oikea puoli negatiivinen.

Tapaus 2: $t \neq v$ ja $t' \neq v'$. Symmetrian vuoksi voidaan olettaa, että $|v'| \leq |v|$. Oletetaan ensin, että $|v| > |t| + |y|$. Tällöin lemman 6.6 mukaan $\tau(w) \geq$



Kuva 6.35: Tapaus, jossa sekä $t \neq v$ että $t' \neq v'$.

$|\gamma zuvy^{-1}|$. Käyttämällä nyt epäyhtälöä $|\gamma| > |y|$ ja oletusta $|v'| \leq |v|$ saadaan

$$\begin{aligned}|\gamma zuvy^{-1}| > |zuv| &= \frac{1}{2}(2|z| + 2|u| + 2|v|)\\ &> \frac{1}{2}(|z| + 2|u| + 2|v|)\\ &\geq \frac{1}{2}(|z| + 2|u| + |v| + |v'|) = \frac{1}{2}|w|\\ &> \tau(w).\end{aligned}$$

Tämä on ristiriita, koska nyt on johdettu, että $\tau(w)$ on itseään pienempi.

Voidaan siis olettaa, että $|v| \leq |t| + |y|$, jolloin voidaan käyttää luvun 6.4 lemmoja 6.8–6.12. Tapauksessa 1.2 johdettiin lausekkeet L_1, L_2 ja L_3 käyttämättä yhtälöä $|v'| = |v|$. Näin ollen ne ovat epänegatiivisia aina kun $|v| \leq |t| + |y|$. Toisin sanoen $L_1, L_2, L_3 \geq 0$ ovat voimassa myös tässä tapauksessa. Lemmasta 6.11 saadaan käyttämällä epäyhtälöä $|\alpha| \leq |u|$ epäyhtälö

$$L_4 := |t| - |v| + |u| - |f| - 1 \geq 0.$$

Oletetaan nyt, että t' on mahdollisimman lyhyt. Tällöin lemmasta 6.12 saadaan

$$L_5 := |f| - |t| - |t'| + |z| - 1 \geq 0.$$

Tarkastelu jakautuu nyt kahteen osaan sen mukaan, onko $|v'| > |t'| + |y'|$. Näissä käytetään lemmojen symmetrisiä versioita.

Tapaus 2.1: $|v'| > |t'| + |y'|$. Nyt lemman 6.7 mukaan $\tau(w) \geq |(y')^{-1}v'uz\gamma'|$. Käyttämällä vastaoletusta $\tau(w) \leq \frac{3}{7}|w|$ ja suoraan määritelmästä saatavaa epäyhtälöä $|\gamma'| \geq |y'| + 1$ saadaan epäyhtälö

$$3|v'| + 6|u| + 3|z| + 3|v| \geq 7|v'| + 7|u| + 7|z| + 7,$$

josta termejä uudelleenjärjestelemällä saadaan epäyhtälö

$$L_6 := -4|v'| - |u| - 4|z| + 3|v| - 7 \geq 0.$$

Nyt on tarkoitus saada yhtälöistä

$$\begin{array}{rrrrrrrrrrr}
L_1 = & -|t| & -|t'| & & & & +|z| & +|\delta| & & -1 \\
L_2 = & & & 3|v'| & +3|v| & -|u| & -4|z| & -7|\delta| & & -7 \\
L_3 = & 7|t| & & +3|v'| & -4|v| & -|u| & -4|z| & & & \\
L_4 = & |t| & & & -|v| & +|u| & & & -|f| & -1 \\
L_5 = & -|t| & -|t'| & & & & +|z| & & +|f| & -1 \\
L_6 = & & & -4|v'| & +3|v| & -|u| & -4|z| & & & -7
\end{array}$$

kertomalla positiivisilla luvuilla ja summaamalla ristiriitainen kaava. Voidaan nähdä, että

$$14L_1 + 2L_2 + 2L_3 + 7L_4 + 7L_5 + 3L_6 = -21t' - 7z - 49.$$

Tästä saadaan ristiriita, koska yhtälön vasen puoli on epänegatiivisten arvojen summa ja oikea puoli on negatiivinen.

Tapaus 2.2: $|v'| \leq |t'| + |y'|$. Lemmasta 6.13 saadaan käyttämällä vastaoletusta $\tau(w) \leq \frac{3}{7}|w|$ ja epäyhtälöä $|\delta'| \geq |y_{\delta'}|$ johto

$$\begin{aligned}
\frac{3}{7}|w| &\geq \tau(w) \\
&\geq -|y_{\delta'}| + |v'| + |u| + |z| - |t'| + |\delta'| \\
&\geq |v'| + |u| + |z| - |t'|.
\end{aligned}$$

Sijoittamalla tähän $w = v'uzuv$ ja kertomalla puolittain luvulla 7 saadaan epäyhtälö

$$3|v'| + 3|v| + 6|u| + 3|z| \geq 7|v'| + 7|u| + 7|z| - 7|t'|,$$

joka saadaan muotoon

$$L_7 := -4|v'| + 3|v| - |u| - 4|z| + 7|t'| \geq 0.$$

Nyt yhtälöiden

$$\begin{array}{lrrrrrrrrr}
L_1 = & -|t| & -|t'| & & & & +|z| & +|\delta| & & -1 \\
L_2 = & & & 3|v'| & +3|v| & -|u| & -4|z| & -7|\delta| & & -7 \\
L_3 = & 7|t| & & +3|v'| & -4|v| & -|u| & -4|z| & & & \\
L_4 = & |t| & & & -|v| & +|u| & & & -|f| & -1 \\
L_5 = & -|t| & -|t'| & & & & +|z| & & +|f| & -1 \\
L_7 = & & +7|t'| & -4|v'| & +3|v| & -|u| & -4|z| & & &
\end{array}$$

lausekkeet ovat epänegatiivisia. Nyt voidaan nähdä, että

$$14L_1 + 2L_2 + 2L_3 + 7L_4 + 7L_5 + 3L_7 = -35|z| - 42.$$

Tämän yhtälön vasen puoli on epänegatiivinen, mutta oikea puoli on negatiivinen, joten on päädytty ristiriitaan.

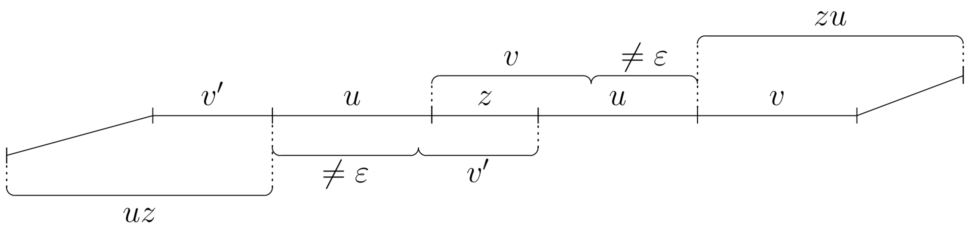
Jokaisessa tapauksessa on nyt päädytty ristiriitaan, joten lemma on todistettu. □

Esimerkki 6.14. Lemman 6.14 oletukset $t \neq v$ ja $t' \neq v'$ ovat voimassa kummassakin sanan w_{AP} tekijöihinjaossa. Väite $\tau(w_{\mathrm{AP}}) > \frac{3}{7}|w_{\mathrm{AP}}|$ pitää paikkansa, koska $\tau(w_{\mathrm{AP}}) = 3n + 6$ ja

$$\frac{3}{7}|w_{\mathrm{AP}}| = \frac{3}{7}(7n + 10) = 3n + 4\frac{2}{7}.$$

Nyt voidaan todistaa tämän tutkielman päälause. Yllä todistetun lemman 6.14 mukaan jos $t \neq v$ tai $t' \neq v'$, niin sanalla w on reunaton tekijä, joka on pidempi kuin $\frac{3}{7}|w|$. Lauseen todistuksessa riittää siis käsitellä tapaukset, joissa $t = v$ ja $t' = v'$.

Lause 6.15. Olkoon w epätyhjä sana. Jos $|w| \geq \frac{7}{3}\,\tau(w)$, niin $\tau(w) = \pi(w)$.



Kuva 6.36: Jos $v <_{\mathrm{p}} zu$ ja $v' <_{\mathrm{s}} uz$, niin sana $v'uzuv$ on sanan $uzuzuzu$ tekijä.

Todistus. Jos jokainen sanan w kirjain esiintyy vain kerran, niin w on reunaton ja $\tau(w) = \pi(w) = |w|$. Voidaan siis olettaa, että ainakin jokin sanan w kirjaimista esiintyy ainakin kaksi kertaa. Tällöin sanalla w on määritelmän 6.4 mukainen tekijöihinjako $w = v'uzuv$.

Oletetaan, että $|w| \geq \frac{7}{3}\tau(w)$. Tällöin lemman 6.14 mukaan $t = v$ ja $t' = v'$. Tämä tarkoittaa, että v on tekijän zu aito prefiksi ja v' on tekijän uz aito suffiksi. Tämä tilanne on esitetty kuvassa 6.36. Sana w on tällöin sanan $uzuzuzu$ tekijä.

Sanalla $uzuzuzu$ on jakso $|uz|$, joten $\pi(uzuzuzu) \leq |uz|$. Tällöin lemman 2.10 mukaan $\pi(w) \leq |uz|$. Jos $\pi(w) \leq |u|$, niin saadaan epäyhtälöketju

$$\pi(w) \leq |u| = \tau_2(w) \leq \tau(w)\, \tau\, \pi(w),$$

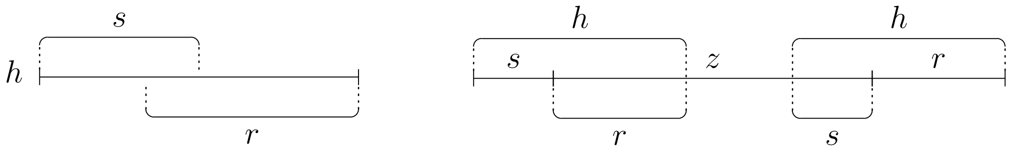
jonka mukaan $\pi(w) = \tau(w) = |u|$. Voidaan siis olettaa, että $|u| < \pi(w) \leq |uz|$. Tämä tarkoittaa erityisesti, että z ei ole tyhjä sana.

Olkoon rs sanan u kriittinen tekijöihinjako. Koska u on reunaton, niin lemman 4.4 mukaan sanan u konjugaatti sr reunaton, sillä on kriittinen tekijöihinjako $s.r$ ja sanat s ja r eivät limity eivätkä ole toistensa tekijöitä. Seuraavaksi siirrytään tarkastelemaan sanan uz konjugaattia szr.

Oletetaan ensin, että szr on reunaton. Tällöin $|szr| \leq \tau(w)$, joten saadaan epäyhtälöketju

$$\pi(w) \leq |uz| = |szr| \leq \tau(w) \leq \pi(w).$$

Näin ollen $\tau(w) = \pi(w) = |uz|$, mikä on lauseen väite.



(a) Jos $|h| < |u|$, niin s ja r limittyvät. (b) Jos $|h| = |u|$, niin $r \leq_{\mathrm{p}} z$ ja $s \leq_{\mathrm{s}} h$.

Kuva 6.37: Tapaukset, joissa $s <_{\mathrm{p}} h$ ja $r <_{\mathrm{s}} h$.

Oletetaan todistuksen loppuun asti, että sana szr ei ole reunaton. Olkoon h sen lyhin reuna. Tällöin $|h| \leq |u|$, koska $|u| = \tau_2(w)$. Sanalla szr on nyt yhteisinä prefikseinä sanat h ja s sekä yhteisinä suffikseina h ja r. Jos h on sekä sanan s prefiksi että sanan r suffiksi, niin h on sanan sr reuna. Sana sr tiedetään kuitenkin reunattomaksi, joten s on sanan h aito prefiksi tai r on sanan h aito suffiksi.

Jos $s <_{\mathrm{p}} h$ ja $h \leq_{\mathrm{s}} r$, niin s on sanan r tekijä. Jos vastaavasti $h \leq_{\mathrm{p}} s$ ja $r <_{\mathrm{s}} h$, niin r on sanan s tekijä. Kumpikin tapaus on ristiriidassa lemman 4.4 kanssa. Näin ollen sekä $s <_{\mathrm{p}} h$ että $r <_{\mathrm{s}} h$. Jos $|h| < |u|$, niin s ja r ovat limittäin sanassa h kuvan 6.37a mukaisesti, koska $|s| + |r| = |u|$. Näin ollen $|h| = |u|$, jolloin r on sanan z prefiksi ja s on sanan z suffiksi kuvan 6.37b mukaisesti.

Väite 6.15.1. Joko ainakin toinen sanoista uz ja zu on reunaton tai u on sanan z prefiksi ja suffiksi.

Väitteen todistus. Oletetaan, että uz on reunaton. Koska z on epätyhjä sana, niin u ei sanan uz suffiksi. Näin ollen erityisesti u ei ole sanan z suffiksi. Vastaavasti saadaan, että jos zu on reunaton, niin u ei ole sanan z prefiksi.

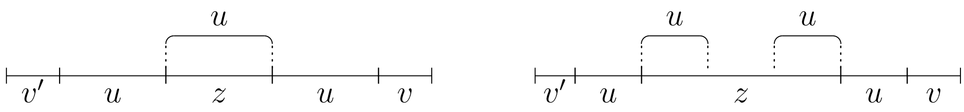
Oletetaan, että $u \leq_{\mathrm{p}} z$ ja $u \leq_{\mathrm{s}} z$. Tällöin u on sanojen zu ja uz reuna, koska sanat uz ja zu onat pidempiä kuin u. □

Lauseen todistus jatkuu nyt käsittelemällä apuväitteen 6.15.1 tapaukset.

Tapaus 1:. Oletetaan, että uz tai zu on reunaton. Tällöin saadaan epäyhtälöketju

$$\pi(w) \leq |u| + |z| \leq \tau(w) \leq \pi(w),$$

josta seuraa lauseen väite $\tau(w) = \pi(w)$.



(a) $u = z$ (b) $u \neq z$

Kuva 6.38: Jos u on sanan z prefiksi ja suffiksi, niin joko $u = z$ tai u on sanan z lyhin reuna.

Tapaus 2:. Oletetaan, että u on sanan z prefiksi ja suffiksi. Mahdolliset tapaukset on nyt esitetty kuvassa 6.38.

Oletetaan, että $u = z$. Tällöin w on sanan u^7 tekijä, jolloin lemman 2.10 mukaan $\pi(w) \leq |u|$. Tämä on ristiriita, koska on oletettu, että $\pi(w) > |u|$.

Oletetaan todistuksen loppuun asti, että $u \neq z$. Tällöin u on sanan z lyhin reuna. Tämä tilanne on esitetty kuvassa 6.38b. Koska u ei voi limittyä itsensä kanssa, voidaan w nyt kirjoittaa muodossa

$$w = v'u^i z' u^j v, \tag{6.9}$$

missä $i, j \geq 2$ ja u ei ole sanan $z'u$ prefiksi eikä sanan uz' suffiksi. Sana z' ei voi olla tyhjä, koska tällöin w olisi sanan u potenssin tekijä ja $\pi(w) \leq |u|$. Sijoittamalla yhtälö $z = u^{i-1}z'u^{j-1}$ kaavoihin $v <_{\mathrm{p}} zu$ ja $v' <_{\mathrm{s}} uz$ saadaan, että v on sanan $u^{i-1}z'u^j$ aito prefiksi ja v' on sanan $u^i z' u^{j-1}$ aito suffiksi. Koska z oletettiin määritelmässä 6.4 mahdollisimman pitkäksi, sana u ei voi olla sanan v prefiksi eikä sanan v' suffiksi. Näin ollen v on sanan u aito prefiksi ja v' on sanan u aito suffiksi, jolloin

$$|v'| < |u| \text{ ja } |v| < |u|. \tag{6.10}$$

Voidaan olettaa, että $i \leq j$, koska tapaus $i > j$ on symmetrinen. Edellä saatiin seuraava sanaa szr koskeva tulos.

Väite 6.15.2:. Sana szr on reunaton tai jompikumpi seuraavista väitteistä on tosi:

(i) sana uz tai sana zu on reunaton,

(ii) sana u on sanan z prefiksi ja suffiksi.

Tämä väite pätee myös tekijälle $sz'u^{j-1}$ sijoittamalla sanan z tilalle sana $z'u^{j-1}$, jolloin väite saadaan muotoon

Väite 6.15.3:. Sana $sz'u^{j-1}r$ on reunaton tai jompikumpi seuraavista väitteistä on tosi:

(i) sana $uz'u^{j-1}$ tai sana $z'u^j$ on reunaton,

(ii) sana u on sanan $z'u^{j-1}$ prefiksi ja suffiksi.

Kohdassa (i) sana $uz'u^{j-1}$ ei voi olla reunaton, koska u on aina sanan $uz'u^{j-1}$ reuna. Kohdassa (ii) sana u voi olla sanan $z'u^{j-1}$ prefiksi, koska u ei ole sanan z' prefiksi ja u ei voi reunattomana sanana limittyä itsensä kanssa. Näin ollen edellinen väite saadaan seuraavaan muotoon.

Väite 6.15.4:. Sana $sz'u^{j-1}r$ sana $z'u^j$ on reunaton.

Väitteen 6.15.4 sanat ovat konjugaatteja, joten ne ovat yhtä pitkiä. Näin ollen saadaan epäyhtälö $|z'u^j| \leq \tau(w)$, josta edelleen lauseen oletusta $\tau(w) \leq \frac{3}{7}|w|$ käyttämällä saadaan epäyhtälö

$$|z'u^j| \leq \frac{3}{7}|w|. \tag{6.11}$$

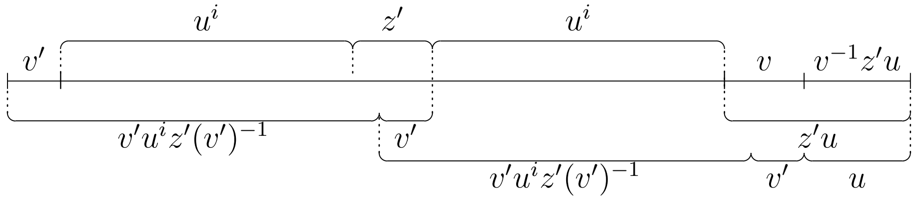
Nyt sijoittamalla yhtälö (6.9) saadaan seuraava päättelyketju:

$$\begin{aligned} 7|z'| + 7j|u| &\leq 3|v'| + 3(i+j)|u| + 3|z'| + 3|v| \\ 4|z'| + (4j-3i)|u| &\leq 3|v'| + 3|v| \\ |v'| + |v| &\geq \left(\frac{4}{3}j - i\right)|u| + \frac{4}{3}|z'| \\ &> \left(\frac{4}{3}j - i\right)|u|. \end{aligned} \tag{6.12}$$

Jos $|i| < |j|$, niin $|j| \geq |i| + 1$ ja $|j| \geq 3$, jolloin epäyhtälö (6.12) muuntuu muotoon

$$|v'| + |v| > \left(\frac{4}{3}j - i\right)|u| \geq \left(\frac{1}{3}j + 1\right)|u| \geq 2|u|.$$

Tämä on ristiriidassa epäyhtälöiden (6.10) kanssa.



Kuva 6.39: Tapaus, jossa $i = j$, $v' \leq_{\mathrm{s}} uz'$ ja $v \leq_{\mathrm{p}} z'u$.

Oletetaan todistuksen loppuun asti, että $i = j$. Sanalla w on nyt tekijöihinjako

$$w = v'u^iz'u^iv.$$

Oletetaan aluksi, että v' on sanan uz' suffiksi ja v on sanan $z'u$ prefiksi. Kuvan 6.39 tapaan saadaan, että

$$\begin{aligned} w &= v'u^iz'u^iv \\ &\leq_{\mathrm{p}} v'u^iz'u^iz'u \\ &\leq_{\mathrm{p}} v'(u^iz')^3 \\ &= (v'u^iz'(v')^{-1})^3v' \\ &\leq_{\mathrm{p}} (v'u^iz'(v')^{-1})^4. \end{aligned}$$

Näin ollen sanalla w on jakso $|v'u^iz'(v')^{-1}| = |u^iz'|$. Nyt käyttämällä epäyhtälöä (6.11) saadaan epäyhtälöketju

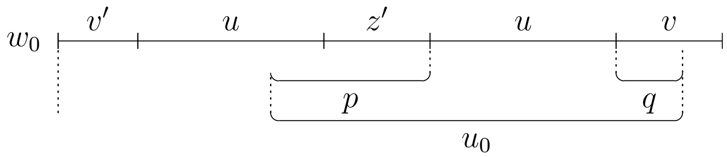
$$\pi(w) \leq |u^iz'| \leq \tau(w) \leq \pi(w),$$

jonka mukaan $\tau(w) = \pi(w)$.

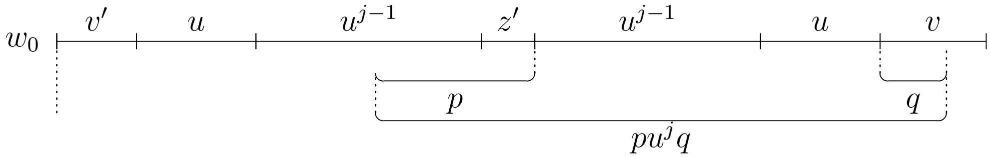
Oletetaan todistuksen loppuun asti, että v' ei ole sanan uz' suffiksi tai v ei ole sanan $z'u$ prefiksi. Siirrytään nyt tarkastelemaan sanan w tekijää

$$w_0 = v'uz'uv.$$

Tämä sanan w_0 tekijöihinjako on määritelmän 6.4 mukainen, koska sanalla w_0 ei ole tekijää u pidempiä kahdesti esiintyviä reunattomia tekijöitä ja epäyhtälöiden (6.10) mukaan z' on mahdollisimman pitkä. Näin ollen mää-



Kuva 6.40: Sanan u_0 valinta.



Kuva 6.41: Sana pu^jq on sanan $v'u^jz'u^jv = w$ tekijä.

ritelmästä 6.4 voidaan saada tekijät

$$
\begin{aligned}
t_0 &= v \wedge_{\mathrm{p}} z'u \text{ ja}\\
t_0' &= v' \wedge_{\mathrm{s}} uz'.
\end{aligned}
$$

Oletuksesta $v' \not\leq_{\mathrm{s}} uz'$ tai $v \not\leq_{\mathrm{p}} z'u$ seuraa, että $t_0' \neq v'$ tai $t_0 \neq v$. Lemman 6.14 mukaan $\tau(w_0) > \frac{3}{7}|w_0|$. Sanalla w_0 on siis reunaton tekijä u_0, jolle $|u_0| > \frac{3}{7}|v'uz'uv|$. Lisäksi sanalla u_0 on tekijänä u.

Oletetaan, että u_0 on sanan $uz'u$ tekijä. Koska u_0 on reunaton, saadaan epäyhtälöketju

$$|u_0| \leq \tau(uz'u) \leq \pi(uz'u) \leq |uz'|.$$

Näin ollen $|uz'| > \frac{3}{7}|v'uz'uv|$, josta saadaan

$$|u^jz'| > \frac{3}{7}|v'uz'uv| + |u^{j-1}| \geq \frac{3}{7}|v'uz'uv| + \frac{6}{7}|u^{j-1}| = \frac{3}{7}|v'uz'uv| = \frac{3}{7}|w|.$$

Tämä on ristiriidassa epäyhtälön (6.11) kanssa.

Oletetaan sitten, että u_0 ei ole sanan $uz'u$ tekijä. Koska u on sanan u_0 tekijä, voidaan symmetrian takia valita, että u_0 sisältää sanan u jälkimmäisen esiintymän. Tällöin $u_0 = puq$, missä p on sanan $v'uz'$ suffiksi ja q on sanan v epätyhjä prefiksi. Tämä on esitetty kuvassa 6.40. Tällöin pu^jq on sanan $v'u^jz'u^jv = w$ tekijä, kuten kuvassa 6.41. Sanan pu^jq pituus saadaan nyt näytetyksi rajaa $\tau(w)$ suuremmaksi johdolla

$$\begin{aligned}
|pu^jq| &= |u^{j-1}| + |u_0| \\
&> |u^{j-1}| + \frac{3}{7}|v'uz'uv| \\
&\geq \frac{6}{7}|u^{j-1}| + \frac{3}{7}|v'uz'uv| \\
&= \frac{3}{7}|v'u^jz'u^jv| \\
&= \frac{3}{7}|w| \\
&\geq \tau(w).
\end{aligned}$$

Näin ollen pu^jq ei ole reunaton.

Olkoon siis h sanan pu^jq lyhin reuna. Koska h on reunaton ja se esiintyy kahdesti, sen pituudelle saadaan yläraja $|h| \leq \tau_2(w) = |u|$. Tällöin h on sanan uq aito suffiksi ja sanan pu prefiksi. Näin ollen h on sanan $puq = u_0$ reuna. Tämä on ristiriidassa sanan u_0 reunattomuuden kanssa. Lause on näin saatu todistetuksi.

$\square$

Lauseen 6.15 todistuksen mukaan yhtälö $\tau(w) = \pi(w)$ on voimassa vain silloin, kun $t = v$ ja $t' = v'$. Näin ollen saadaan seuraava tulos.

Seuraus 6.16. Olkoon $w = v'uzuv$ määritelmän 6.4 mukainen tekijöihinjako ja $|w| \geq \frac{7}{3}\tau(w)$. Tällöin $v' <_{\mathrm{s}} uz$ ja $v <_{\mathrm{p}} zu$.

7 Yhteenveto

Tässä tutkielmassa on esitelty tuloksia sanan lyhimmästä jaksosta $\pi(w)$ ja pisimpien reunattomien tekijöiden pituudesta $\tau(w)$. Erityisenä kohteena on ollut kysymys, milloin nämä suureet pakottuvat yhtäsuuriksi. Luvussa 2.3 on esitelty ketju

$$0 \leq \tau_2(w) \leq \tau(w) \leq \pi(w) \leq |w|, \tag{2.3}$$

jossa kolmas epäyhtälö $\tau(w) \leq \pi(w)$ muuttuu yhtäsuuruudeksi, jos yksikin muista epäyhtälöistä on yhtäsuuruus. Tämä voidaan ilmaista myös niin, että jos $\pi(w) < |w|$, niin ketju

$$0 < \tau_2(w) < \tau(w) < \pi(w) < |w| \tag{2.4}$$

on voimassa.

Suurempina tuloksina on käsitelty ongelmaa, kuinka pitkä sanan tulee olla toisaalta lyhimmän jakson $\pi(w)$ suhteen ja toisaalta pisimpien reunattomien tekijöiden pituuden $\tau(w)$ suhteen, jotta $\tau(w) = \pi(w)$. Uutena tuloksena on todistettu lause 5.3, jossa parantaa rajaa lyhimmän jakson suhteen niin, että parasta aiemmin tunnettua rajaa $2\pi(w) - k$ [5] on laskettu yhdellä pienemmäksi. Saatu raja on optimaalinen:

$$\text{Jos } |w| \geq 2\pi(w) - k - 1, \text{ niin } \tau(w) = \pi(w).$$

Lisäksi todistettiin lause 5.7, joka määrittelee kaikki sanat, joille $\tau(w) < \pi(w)$ ja joiden pituus on yhden pienempi kuin lauseen 5.3 alaraja. Lauseiden 5.3 ja 5.7 todistuksissa käytettiin apuna sellaisia Lyndonin sanoja, jotka ovat Lyndonin sanoja aakkoston jokaisen järjestyksen suhteen, jossa sanan ensimmäinen kirjain on minimaalinen.

Vastaavasta rajasta lyhimmän jakson suhteen on esitelty Holubin ja Nowotkan [10] todistama lause 6.15:

$$\text{Jos } |w| \geq \frac{7}{3}\tau(w), \text{ niin } \tau(w) = \pi(w).$$

Tämä ratkaisee Ehrenfeuchtin–Silbergerin ongelman, koska Assous'n ja

Pouzet'n [1] esittelemälle sanalle

$$w_{\mathrm{AP}} = a^n b a^{n+1} b a^n b a^{n+2} b a^n b a^{n+1} b a^n \tag{3.3}$$

pätee $|w_{\mathrm{AP}}| = \frac{7}{3}\,\tau(w_{\mathrm{AP}}) - 4$ ja $\tau(w_{\mathrm{AP}}) < \pi(w_{\mathrm{AP}})$. Assous'n ja Pouzet'n sana on paras tunnettu esimerkki sanasta, jolla ei ole lyhimmän jakson pituista reunatonta tekijää. Lauseen 6.15 raja on siis kertoimen $\frac{7}{3}$ osalta optimaalinen, mutta vakiotermi ei ole optimaalinen. Assous'n ja Pouzet'n vastaesimerkin mukaan optimaalinen raja voi olla seuraava:

$$\text{Jos } |w| > \frac{7}{3}\tau(w) - 4, \text{ niin } \tau(w) = \pi(w).$$

Holub ja Nowotka mainitsevat, että luvun 6 todistuksia on mahdollista parantaa niin, että ehdoksi saadaan $|w| > \frac{7}{3}\,\tau(w) - \frac{8}{3}$. Yksi tapa parantaa todistusta saadaan huomioimalla, että useassa kohdassa ilmeni aitoja epäyhtälöitä ketjutettuina, mikä vähentää näin saatujen epäyhtälöiden ilmaisuvoimaa. Eräs parannuskeino olisi siis muuttaa jokainen aito epäyhtälö $|x| > |y|$ muotoon $|x| \geq |y| + 1$. Tämä kuitenkin tuskin riittää optimaalisen rajan saamiseksi, vaan lienee tarpeellista myös lemmoissa 6.6 ja 6.9 konstruoida saatavat pitkät reunattomat tekijät aidosti pidemmiksi.

Viitteet

[1] Assous, Roland ja Maurice Pouzet: *Une caracterization des mots periodiques.* Discrete Mathematics, 25(1) :1–5, 1979, ISSN 0012-365X. http://dx.doi.org/10.1016/0012-365X(79)90146-8.

[2] Breslauer, Dany, Tao Jiang ja Zhigen Jiang: *Rotations of periodic strings and short superstrings.* Journal of Algorithms, 24(2):340–353, 1997, ISSN 0196-6774. http://www.sciencedirect.com/science/article/pii/S0196677497908610.

[3] Choffrut, Christian ja Juhani Karhumäki: *Combinatorics of words.* Teoksessa Rozenberg, Grzegorz ja Arto Salomaa (toimittajat): *Handbook of Formal Languages*, nide 1. Word, Language, Grammar, luku 6, sivut 329–438. Springer-Verlag, Berliini/Heidelberg, 1997, ISBN 3-540-60420-0.

[4] Crochemore, Maxime ja Dominique Perrin: *Two-way string-matching.* Journal of the ACM, 38(3):651–675, 1991.

[5] Duval, Jean Pierre: *Relationship between the period of a finite word and the length of its unbordered segments.* Discrete Mathematics, 40(1):31–44, 1982, ISSN 0012-365X. http://dx.doi.org/10.1016/0012-365X(82)90186-8.

[6] Ehrenfeucht, Andrzej ja Donald Morison Silberger: *Periodicity and unbordered segments of words.* Discrete Mathematics, 26(2):101–109, 1979, ISSN 0012-365X. http://dx.doi.org/10.1016/0012-365X(79)90116-X.

[7] Harju, Tero ja Dirk Nowotka: *Density of critical factorizations.* Theoretical Informatics and Applications. Informatique Théorique et Applications, 36(3):315–327, 2002, ISSN 0988-3754. http://dx.doi.org/10.1051/ita:2002016.

[8] Harju, Tero ja Dirk Nowotka: *Periodicity and unbordered words: A proof of the extended Duval conjecture.* Journal of the ACM, 54(4), heinä-

kuu 2007, ISSN 0004-5411. http://doi.acm.org/10.1145/1255443.1255448.

[9] Holub, Štěpán ja Dirk Nowotka: *The Ehrenfeucht–Silberger problem.* Teoksessa Albers, S., A. Marchetti-Spaccamela, Y. Matias, S. Nikoletseas ja W. Thomas (toimittajat): *Automata, Languages and Programming*, nide 5555 sarjassa *Lecture Notes in Computer Science*, sivut 537–548, Berliini/Heidelberg, 2009. Springer-Verlag.

[10] Holub, Štěpán ja Dirk Nowotka: *The Ehrenfeucht—Silberger problem.* Journal of Combinatorial Theory, Series A, 119:668–682, 2012.

[11] Lothaire, M.: *Combinatorics on words*, nide 17 sarjassa *Encyclopedia of mathematics and its applications.* Addison-Wesley, Reading, Massachusetts, 1983. Uusintapainos sarjassa *Cambridge mathematical library*, Cambridge University Press, Cambridge, New York, Melbourne, 1997.