TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond
Tarkvarateaduse Instituut

Eve Veltmann 153071IAPM

PREFIKS-, SUFIKS- JA FAKTOR-KINNISTE REGULAARSETE KEELTE VASAKFAKTORID JA AATOMID

magistritöö

Juhendaja: Hellis Tamm
PhD

Tallinn 2020

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Eve Veltmann

01.08.2020

Annotatsioon

Töö eesmärgiks on uurida prefiks-, sufiks- ja faktor-kinniste keelte vasakfaktoreid ja aatomeid ning leida neile iseloomulikke omadusi.

Analüüsil lähtusime prefiks-, sufiks- ja faktor-kinniste keelte tunnus-NFA'de omadustest ning minimaalse DFA ja atomaadi leidmise protseduuridest. Leitud lõplike automaatide iseloomulike joonte ja vaadeldud keeleklasside keelereeglite alusel tuletasime ja tõestasime vasakfaktorite ja aatomite omadused.

Töö tulemusena tõime välja prefiks-, sufiks- ja faktor-kinniste keelte vasakfaktorite ja aatomite leitud omadused, kirjeldasime vaadeldud keeleklasside omavahelised seosed ning märkisime ära nende keelte minimaalse DFA ja atomaadi omadused.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 61 leheküljel, 5 peatükki, 13 joonist, 5 tabelit.

Abstract

Quotients and atoms of prefix-, suffix-, and factor-closed regular languages.

The purpose of this work is to study the properties of quotients and atoms of prefix-, suffix-, and factor-closed languages.

Our research is based on the knowledge that prefix-, suffix-, and factor-closed languages have special kind of NFAs accepting these languages. We use Brzozowski's double-reversal method and theory of átomata to convert these NFAs into minimal DFAs and átomata. From the characteristic features of minimal DFAs and átomata of these languages, we derive some properties of quotients and atoms of prefix-, suffix-, and factor-closed languages.

As a result of the study, we highlight the characteristics of quotients and atoms of prefix-, suffix-, and factor-closed languages. Also, we describe the relationships between these languages and sketch the minimal DFAs and átomata of the languages.

The thesis is in Estonian and contains 61 pages of text, 5 chapters, 13 figures, 5 tables.

Lühendite ja mõistete sõnastik

DFA	*Deterministic finite automaton,* deterministlik lõplik automaat.
NFA	*Nondeterministic finite automaton,* mittedeterministlik lõplik automaat.
FA	*Finite automaton,* lõplik automaat.
Σ	Automaadi tähestik.
Σ^*	Kõigi võimalike sõnade hulk tähestikus Σ, kaasa arvatud tühi sõna.
ε	Tühi sõna.
$\mathcal{L}$	Keel tähestikus Σ on teatud sõnade alamhulk hulgast Σ^*.
$\overline{\mathcal{L}}$	Keele täiend $\overline{\mathcal{L}}$ on keelde $\mathcal{L}$ mittekuuluvate sõnade alamhulk hulgast Σ^*.
$\mathcal{L}^R$	*Reversal language*, pööratud keel on kõigi keele $\mathcal{L}$ ümberpööratud sõnade hulk.
$\mathcal{N} \xrightarrow{det} \mathcal{N}^D$	*Determinization,* NFA determiniseerimine.
$\mathcal{N} \xrightarrow{trim} \mathcal{N}^T$	*Trimming*, automaadi trimmimine.
$\mathcal{D} \xrightarrow{min} \mathcal{D}^M$	*Minimization,* automaadi minimeerimine.
$\mathcal{N} \xrightarrow{rev} \mathcal{N}^R$	*Reversal,* automaadi ümberpööramine.
K	*Left quotient,* keele vasakfaktor.
$\mathcal{L}_{PC}$	*Prefix-closed language,* prefiks-kinnine keel.
$\mathcal{L}_{SC}$	*Suffix-closed lannuage,* sufiks-kinnine keel.
$\mathcal{L}_{BC}$	*Bifix-closed language,* bifiks-kinnine keel.
$\mathcal{L}_{FC}$	*Factor-closed language,* faktor-kinnine keel.
A_i	Keele aatom – mittetühi regulaarne keel, mis moodustatakse keele vasakfaktorite ja nende täiendite kombinatsiooni ühisosana.
A^-	Negatiivne aatom – vasakfaktorite täiendite ühisosa, kui see ei ole tühi hulk.
Prefiks-kinnise keele tunnus-NFA	NFA, mille kõik olekud on lõppolekud.

Sufiks-kinnise keele tunnus-NFA	NFA, mille kõik olekud on algolekud ja millel on üks lõppolek.
Faktor-kinnise keele tunnus-NFA	NFA, mille kõik olekud on alg- ja lõppolekud.

Sisukord

Jooniste loetelu

Tabelite loetelu

1 Sissejuhatus

1.1 Taust

Regulaarsed keeled ja neid aktsepteerivad lõplikud automaadid on olnud arvutiteaduse üheks oluliseks valdkonnaks juba möödunud sajandi keskpaigast. Tulenevalt erinevatest kasutusaladest on laiaulatuslik ka uurimistööde temaatika.

Käesoleva töö aspektist on oluline esile tõsta kahte uurimissuunda. Esimeseks suunaks on automaatide erinevate keerukuste (*complexity*) ja nende ülemiste ja alumiste tõkete (*lower/upper bound*) määramine ning automaatide lihtsustamise võimalused. Viimasega seoses on leitud, et iga keele jaoks on olemas unikaalne minimaalne deterministlik lõplik automaat (*deterministic finite automaton* – DFA) ja atomaat (*átomaton*) ning kirjeldatud nende lõplike automaatide leidmise protseduurid.

Teiseks oluliseks suunaks on erinevate regulaarsete keelte klasside kirjeldamine ja süstematiseerimine. Üheks tähelepanu leidnud keelte klassiks on sõnaosa suhtes kumerad (*prefix-convex, suffix-convex, factor-convex*) keeled. Kõigil neil klassidel on omakorda alamklassid nagu sõnaosa suhtes kinnised (*-closed*) või vabad (*-free*) keeled, ideaalid jne.

Lisaks keele kui terviku iseloomustamisele saab eraldada keelest mitmeid olulisi komponente. Keele vasakfaktorid ja aatomid on keele komponendid, mis on otseselt seotud eelnimetatud keelt aktsepteerivate lõplike automaatidega. Keele vasakfaktorid on võrdsed keelt aktsepteeriva minimaalse DFA olekute paremkeeltega ja aatomid on võrdsed keele atomaadi ehk erilise mittedeterministliku lõpliku automaadi (*nondeterministic finite automaton* – NFA) olekute paremkeeltega.

1.2 Probleem

Kuna lõplike automaatide kasutuse peamiseks probleemiks on keerukuse kiire kasv tehetes, on uurimistööd pigem keskendunud erinevate keerukusnäitajate ja vastavate ülemiste ja alumiste tõkete leidmisele. Oluliselt vähem on uurimistöödes vaadeldud erinevate keelte ja keele osade omadusi ja sealtkaudu keele alamgrupi määramise võimalusi.

1.3 Eesmärk

Töö eesmärgiks on leida ja kirjeldada prefiks-, sufiks- ja faktor-kinniste keelte vasakfaktorite ja aatomite omadusi ning seeläbi laiendada teoreetilisi teadmisi eelmainitud keeltest ja nendega seotud automaatidest.

1.4 Ülevaade tööst

Käesolev töö on jaotatud neljaks peatükiks.

Teine peatükk on teema käsitlusest teadustöödes ja artiklites. Vaatleme teemaga seotud üldist tausta ja keskendume käesoleva töö seisukohalt olulistele artiklitele.

Kolmandas peatükis on lühike ülevaade automaatide ja regulaarsete keelte teooriast ja käesolevas töös kasutatud tehnikatest.

Neljas peatükk sisaldab minu tulemusi. Selles kirjeldame ja tõestame käsitletavate keelte vasakfaktorite ja aatomite omadusi ning teeme üldistava kokkuvõtte leitust.

Töö lõpetab kokkuvõte.

2 Taust ja kirjanduse ülevaade

2.1 Regulaarsed keeled ja nendega seotud automaadid

Olles lihtsustatud tasandil intuitiivselt hoomatav, on lõplik automaat ja tema poolt aktsepteeritav keel heaks mudeliks paljudes valdkondades. Peamisena võib kindlasti välja tuua keelte analüüsi ning erinevaid mustrite leidmise rakendusi. Sheng Yu [1] loetleb olulisimate regulaarsete keelte ja lõplike automaatide rakendusvaldkondadena programmeerimiskeelte analüsaatoreid kompilaatorites ja kasutajaliideste translaatorites [2], [3], mustrite sobitamist tekstiredaktorites [4], aga ka vooluahela *(circuit*) disainitööriistu [5]. Hiljem lisandunud kasutusvaldkondadest toob ta välja paralleelprotsesside disaini ja analüüsi [6], [7], [8], pildi genereerimise ja kompressiooni vahendeid [9], [10], [11], objekt-orienteeritud keelte tüübiteooriat [12], DNA analüüsirakendusi [13] jne.

Esimesed olulised lõplikke automaate ja regulaarseid keeli käsitlevad teadustööd jäävad eelmise sajandi keskpaika. Üldiselt on tunnustatud valdkonna arengu alguse olulisimate töödena McCulloch'i ja Pitts'i närvivõrkude mudelite kirjeldust [14], Kleene'i teoreemi lõpliku automaadi ja regulaaravaldise seosest [15], Myhill'i [16] ja Nerode [17] töid regulaarsete keelte omadustest ning Rabin'i ja Scott'i mittedeterministliku automaadi esitust [18].

Tänu laialdasele rakendusalale on lõplikud automaadid säilitanud oma keskse koha ka tänases arvutiteaduses. Käesoleva töö vaates keskendume peamiselt regulaarseid kumeraid keeli ja unikaalsete lõplike automaatide leidmist ja omadusi käsitlevatele teadustöödele.

2.2 Automaatide lihtsustamine, keelte unikaalsed automaadid

Automaatide suurus on üheks oluliseks probleemiks lõplike automaatide kasutusel praktilistes rakendustes. Sellest tulenevalt on pööratud erilist tähelepanu automaadi lihtsustamise võimalustele ja ekvivalentse minimaalse automaadi leidmisele.

Mitmetest DFA minimeerimise meetoditest kasutame käesolevas töös Brzozowski kahekordse pööramise meetodit (*Brzozowski's double-reversal method*) [19], kus minimaalne DFA leitakse keele pöördkeele DFA pööramise ja determiniseerimise tulemusena. See meetod on hõlpsasti kasutatav ja töötab ka juhul, kui esialgne automaat on NFA. Sellest tulenevalt on seda uuritud, täiendatud ja arendatud mitmetes artiklites [20], [21], [22].

Samuti otsitakse võimalusi minimaalse NFA konstrueerimise algoritmi [23] parendamiseks. Seda nii eelviidatud Kameda-Weineri meetodi arendamise [24], kui ka uute lähenemisviiside otsimise kaudu [25], [26], [27].

Brzozowski ja Tamm esitasid oma artiklis [20] uued mõisted automaatide teoorias – regulaarse keele aatomid ja nendega seotud automaat ehk atomaat. Autorid näitasid, et keele pöördkeele minimaalse DFA pööramisel saadakse NFA, mille paremkeeled on algse keele aatomid. Artiklis kirjeldavad autorid põhjalikult atomaadi konstrueerimise meetodeid ja aatomite ning atomaadi omadusi.

2.3 Regulaarsete keelte kumerate keelte klass ja selle alamklassid

Regulaarsete keelte klassidele ja klasside hierarhia kirjeldamisele keskendunud töid on vähe. Pigem on regulaarsete keelte klasse ja alamklasse kirjeldatud lõplike automaatide uurimistöödes, märkides, milliste keelerühmade osas püütakse probleemi lahendada või uudset lahendust pakkuda.

Käesolevas töös vaatleme prefiksite, sufiksite ja faktorite suhtes kinniseid keeli (edaspidi ka prefiks-, sufiks- ja faktor-kinnised keeled), mis on kõik vastavate kumerate keelte (edaspidi ka kumerkeel) alamklassid.

Thierrin kirjeldas kumerkeeli 1973-ndal aastal [28]. Mainitud töö keskendub alamsõna (*subword*) suhtes kumeratele keeltele.

Ang ja Brzozowski analüüsisid ja süstematiseerisid kumeraid keeli laiaulatuslikumalt [29]. Töös käsitlevad nad prefiksite, sufiksite, bifiksite, faktorite ja alamsõnade suhtes kumeraid keeli ja nende keelte vastava sõnaosa suhtes kinniseid ja vabu alamklasse. Analüüsi käigus joonistub kumerkeelte klasside hierarhia ning leiab tõestamist ja nimetamist hulk klasside omadusi ja omavahelisi seoseid. Neist käesoleva töö kontekstis olulisi nimetame töö keeleklasse käsitlevas osas.

Kinnised keeled on olnud uurimisobjektiks paljudele erinevate autorite töödele. Kinniste keelte vasakfaktorite keerukust (*quetient complexity*) analüüsivas töös [30] esitavad autorid ka põhjaliku loendi kinniste keelte uurimustest. Kasutan siinkohal siiski ainult viidet oma väite kinnituseks ja ei korda kogu loendit.

Erinevalt enamikust keelte omadusi ja keerukust uurivatest töödest keskenduvad Kao, Rampersad ja Shallit kindla struktuuriga NFA'de analüüsile [31]. Nad vaatlevad NFA'sid, mille kõik olekud on kas algolekud, lõppolekud või siis korraga nii alg- kui ka lõppolekud. Muuhulgas esitavad autorid kolm käesoleva töö seisukohalt väga olulist tõdemust:

- Keel on prefiks-kinnine vaid siis, kui seda aktsepteerib mõni NFA, mille kõik olekud on lõppolekud.
- Keel on sufiks-kinnine vaid siis, kui seda aktsepteerib mõni NFA, mille kõik olekud on algolekud ja millel on ainult üks lõppolek.
- Keel on faktor-kinnine vaid siis, kui seda aktsepteerib NFA, mille kõik olekud on samaaegselt nii alg- kui lõppolekud.

Need määratlused on aluseks mitmetele käesolevas töös esitatud tulemustele.

3 Töös kasutatud mõisted

Selles osas käsitleme üldtuntud mõisteid ja tehnikaid.

3.1 Tähestik, sõna ja keel

3.1.1 Tähestik, sõna ja sõnaosa

Tähestik on kõikide kasutatavate sümbolite hulk. Edaspidi tähistame seda sümboliga Σ.

Sõna on tähestiku sümbolite kombineerimisega saadav sõne. Vahemärkusena pean oluliseks siinkohas täpsustada, et enamikus automaate ja regulaarseid keeli käsitlevates uurimustes kasutatakse antud kontekstis sõna „*word*“ mitte „*string*“. Sellest tulenevalt kasutame ka meie edaspidi sõna „sõna“, mitte „sõne“.

Sõna tähistusena kasutame edaspidi tähti u, v, w, x, z.

Σ^* on kõigi sõnade hulk tähestikus Σ. Σ^* sisaldab alati ka tühja sõna, mida tähistame sümboliga ε.

Sõnaosadest vaatleme töös peamiselt kolme: prefiks, sufiks ja faktor.

Vaatleme sõnu u, v, w ja x tähestikus Σ. Kui kehtib võrdus $w = uv$, siis u on w prefiks ja v on w sufiks. Kui $w = uxv$, siis u, x ja v on sõna w faktorid. Samas on w faktorid ka ux ja xv.

Alamsõna on sõna, mille saame algsest sõnast tähtede eemaldamisel. Kui $w = uxvz$, siis sõnad uvz, uxz, uz ja xz on sõna w alamsõnad. Samas on alamsõnad ka kõik sõna w faktorid.

Kuna kehtib võrdus $w = \varepsilon\mathrm{w} = \mathrm{w}\varepsilon$, siis saame öelda, et w on iseenda prefiks, sufiks ja faktor.

Sõna ümberpööramine muudab sümbolite järjestuse temas vastupidiseks. Sõna w pööramisel saadud sõna tähistame sümboliga w^R. Näiteks kui sõna $w = abc$, siis $w^R = cba$, kus a, b ja c on sümbolid tähestikus.

Sõna pööramisel kehtib reegel: kui $w \in \Sigma^*$ ja $a \in \Sigma$, siis $(wa)^R = aw^R$.

3.1.2 Keel, keele täiend ja pööratud keel

$\boldsymbol{Keel}$ tähestikus Σ on teatud sõnade alamhulk hulgast Σ^*. Edaspidi tähistame keelt sümboliga $\mathcal{L}$. Seega, kui $\mathcal{L}$ on keel tähestikus Σ, siis kehtib seos $\mathcal{L} \subseteq \Sigma^*$.

$\boldsymbol{Keele\ täiend}$ $\overline{\mathcal{L}}$ on keelde $\mathcal{L}$ mittekuuluvate sõnade alamhulk hulgast Σ^*, kui $\mathcal{L}$ on keel tähestikus Σ. Teisisõnu $\overline{\mathcal{L}} = \Sigma^* \setminus \mathcal{L}$.

$\boldsymbol{Pööratud\ keel}$ $\mathcal{L}^R$ on kõigi keele $\mathcal{L}$ pööratud sõnade hulk ehk $\mathcal{L}^R = \{w^R | w \in \mathcal{L}\}$.

Keel on $\boldsymbol{regulaarne\ keel,}$ kui leidub mõni lõplik automaat, mis aktsepteerib täpselt selle keele kõiki sõnu ehk teisiti öeldes aktsepteerib seda keelt.

3.2 Lõplikud automaadid ja regulaarsed keeled

$\boldsymbol{Lõplik\ automaat}$ (*finite automaton* – FA) on viie parameetriga iseloomustatav protsessimudel, mis koosneb lõplikust arvust olekutest ja siiretest nende olekute vahel.

FA aktsepteerib regulaarset keelt, otsustades, kas sisendsõna kuulub sellesse keelde või mitte. Igale regulaarsele keelele leidub FA, mis aktsepteerib seda keelt.

Kaks FA'd on $\boldsymbol{ekvivalentsed}$, kui nad aktsepteerivad ühte ja sama keelt.

Vaadeldakse kahte liiki lõplikke automaate – deterministlik lõplik automaat (*deterministic finite automaton* – DFA) ja mittedeterministlik lõplik automaat (*nondeterministic finite automaton* – NFA).

Igast NFA'st saab alati moodustada ekvivalentse ehk sama keelt aktsepteeriva DFA.

3.2.1 Deterministlik lõplik automaat – DFA

DFA kirjeldamiseks kasutame viisikut $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$, kus

Q on automaadi olekute mittetühi lõplik hulk,

Σ on automaadi tähestik ehk lubatud sümbolite mittetühi lõplik hulk,

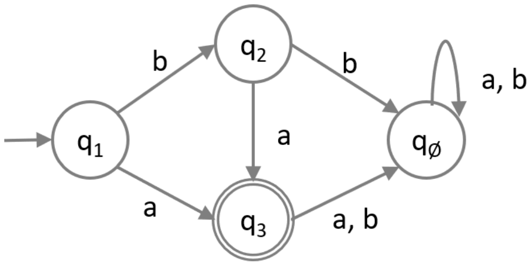
δ on siirdefunktsioon $Q \times \Sigma \rightarrow Q$,

q_1 on automaadi algolek $q_1 \in Q$,

F on automaadi lõppolekute hulk ja $F \subseteq Q$.

Mitte tühja keelt aktsepteerival DFA´l on alati üks algolek ja üks või mitu lõppolekut. Siirdefunktsioon kirjeldab siirdeid kõigist olekutest kõigi tähestiku sümbolite jaoks. DFA puhul on igal tähestiku sümbolil mistahes olekust väljumiseks täpselt üks võimalik siire.

Automaadi kirjeldamiseks kasutatakse nii graafilist kui tabelkuju.



	a	**b**
→q_1	q_3	q_2
q_2	q_3	$q_\emptyset$
←q_3	$q_\emptyset$	$q_\emptyset$
$q_\emptyset$	$q_\emptyset$	$q_\emptyset$

Joonis 1 DFA

Näiteks Joonisel 1 on toodud DFA graafiline kujutis ja siirete tabel.

Selle DFA olekute hulk on $\mathrm{Q} = \{\mathrm{q}_1, \mathrm{q}_2, \mathrm{q}_3, \mathrm{q}_\emptyset\}$, tähestik on $\Sigma\ = \{\mathrm{a}, \mathrm{b}\}$, algolek on q_1, lõppolekute hulk on F = $\{q_3\}$ ja võimalikud siirded on kirjeldatud ülalpool joonisel toodud siirete tabelis.

Skeemil ja siirete tabelis tähistame automaadi algolekut siseneva noolega. Automaadi lõppolekut tähistame skeemil kahekordse ringjoonega ja siirete tabelis väljuva noolega.

Olekut $q_\emptyset$ nimetame tühjaks olekuks. Sellesse olekusse koonduvad kõik negatiivse tulemiga lõppevad siirete jadad ehk tulemused, kus DFA ei aktsepteeri sisendsõna.

Minimaalne DFA on DFA, millele ei leidu väiksema olekute arvuga ekvivalentset DFA'd.

3.2.2 Mittedeterministlik lõplik automaat – NFA

NFA kirjeldamiseks kasutame viisikut $\mathcal{N} = (Q, \Sigma, \delta, I, F)$, kus

Q on automaadi olekute mittetühi lõplik hulk,

Σ on automaadi tähestik ehk lubatud sümbolite mittetühi hulk,

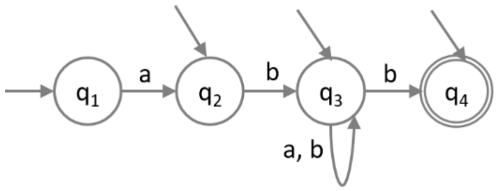
δ on siirdefunktsioon $Q \times \Sigma \to 2^Q$, kus 2^Q tähistab Q kõikide alamhulkade hulka, I on automaadi algolekute hulk ja $I \subseteq Q$,

F on automaadi lõppolekute hulk ja $F \subseteq Q$.

Mittetühja keelt aktsepteerival NFA´l võib olla üks või mitu alg- ja lõppolekut.

Igal tähestiku sümbolil on mistahes NFA olekust väljumiseks siirete hulk, mis võib olla tühi hulk või sisaldada ühte või mitut siiret.

Vaatame ka siin näidet



	a	b
→q_1	{q_2}	-
→q_2	-	{q_3}
→q_3	{q_3}	{q_3,q_4}
←→q_4	-	-

Joonis 2 NFA

Joonisel toodud NFA olekute hulk on $\mathrm{Q} = \{\mathrm{q}_1, \mathrm{q}_2, \mathrm{q}_3, \mathrm{q}_4\}$, tähestik on $\Sigma \ = \{\mathrm{a}, \mathrm{b}\}$, algolekute hulk on $\mathrm{I} = \{\mathrm{q}_1, \mathrm{q}_2, \mathrm{q}_3, \mathrm{q}_4\}$ ja lõppolekute hulk on $F = \{q_4\}$.

NFA joonisel kasutame samu tähiseid kui DFA joonisel.

3.2.3 Siirdefunktsiooni laiendused

Kasutame ka laiendatud siirdefunktsioone :

Sõna siirete jada automaadis on $\delta' \colon Q \times \Sigma^* \rightarrow 2^Q$, kus $\delta'(q, \varepsilon) = \{q\}$ ja $\delta'(q, xa) = \bigcup_{p \in \delta'(q,x)} \delta(p, a)$, kus $q \in Q$ ja $x \in \Sigma^*$.

Olekute alamhulga siirete funktsioon on $\delta'' \colon 2^Q \times \Sigma^* \rightarrow 2^Q$,
kus $\delta''(P, w) = \bigcup_{q \in P} \delta'(q, w)$, kus $P \subseteq Q$ ja $w \in \Sigma^*$.

Lihtsuse mõttes tähistame edaspidi funktsioone δ' ja δ'' tähisega δ.

3.2.4 Lõpliku automaadi keel

Kui automaat aktsepteerib mingit keelt, siis kõigi selle keele sõnade jaoks eksisteerib siirete jada automaadi algolekust lõppolekusse. Kuna NFA ja DFA algolekud on defineeritud erinevalt, on ka keele definitsioon nende jaoks erinev.

NFA $\mathcal{N} = (Q, \Sigma, \delta, I, F)$ aktsepteerib keelt $L(\mathcal{N}) = \{w \in \Sigma^* | \delta\ (I, w) \cap F\ \neq \emptyset\}$.

Sama definitsioon kehtib DFA puhul arvestusega, et algolekute hulk koosneb ainult ühest võimalikust olekust ehk siis $L(\mathcal{D}) = \{w \in \Sigma^* | \delta(q_1, w) \cap F \neq \emptyset\}$.

3.2.5 Lõpliku automaadi oleku vasakkeel ja saavutamatu olek

Automaadi oleku vasakkeel on selle automaadi keele sõnade prefiksite hulk, mis automaadi siirete tulemusel jõuavad algolekust vaadeldava olekuni.

DFA ja NFA olekute vasakkeelte definitsioonid erinevad tulenevalt lubatud algolekute määratlusest.

NFA $\mathcal{N}$ oleku q vasakkeel $L_{I,q}(\mathcal{N})$ on keele $\mathcal{L}$ kõikide sõnaosade hulk, mis jõuavad automaadi siirete tulemusel mingist algolekust olekusse q. Ehk siis NFA $\mathcal{N}$ oleku q vasakkeel $L_{I,q}(\mathcal{N}) = \{w \in \Sigma^* |\ q \in \delta(I, w)\}$.

Peatükis 3.2.2 esitatud näites on NFA oleku q_2 vasakkeel $L_{I,q_2}(\mathcal{N}) = \{\varepsilon, a\}$.

DFA puhul algolekute hulk $I\ = \{q_1\}$, kus q_1 on DFA ainus algolek. Seega DFA $\mathcal{D}$ oleku q vasakkeel $L_{q_1,q}(\mathcal{N}) = \{w \in \Sigma^* |\ q \in \delta(q_1, w)\}$.

Saavutamatu olek on olek, mille vasakkeel on tühi hulk.

3.2.6 Lõpliku automaadi oleku paremkeel, tühi olek ja olekute ekvivalentsus

Automaadi oleku paremkeel on selle automaadi keele sõnade sufiksite hulk, mis automaadi siirete tulemusel jõuavad vaadeldavast olekust lõppolekuni.

DFA ja NFA olekute paremkeelte definitsioonid on ühesugused.

Automaadi $\mathcal{N}$ oleku q paremkeel $L_{q,F}(\mathcal{N})$ on keele $\mathcal{L}$ sõnaosade hulk, mis jõuavad automaadis siirete tulemusel olekust q mõnda lõppolekusse. Ehk siis automaadi $\mathcal{N}$ oleku q paremkeel $L_{q,F}(\mathcal{N}) = \{w \in \Sigma^* | \delta(q, w) \cap F\ \neq \emptyset\}$.

Näiteks peatükis 3.2.2 vaadeldud NFA oleku q_3 paremkeel $\mathrm{L}_{q_3,F}(\mathcal{N}) = (a^* b^*)^* b = \Sigma^* b$.

Tühi olek on olek, mille paremkeel on tühi.

Kaks olekut on ekvivalentsed, kui nende paremkeeled on identsed.

3.2.7 Regulaarse keele vasak- ja paremfaktor

Regulaarse keele $\mathcal{L}$ vasakfaktor (*left quotient*) sõnale w on regulaarne keel paremjääkide x_i hulgast, mis saadakse prefiksi w eraldamisel kõigist keelde $\mathcal{L}$ kuuluvatest sõnadest. Ehk siis keele vasakfaktor sõnale w on regulaarne keel $w^{-1}\mathcal{L} = \{x \in \Sigma^* | wx \in \mathcal{L}\}$.

On üldiselt teada, et regulaarse keele vasakfaktorite hulk on lõplik ja vastavuses minimaalse DFA olekute hulgaga.

Vasakfaktorite hulka tähistame $K = \{K_1, \ldots, K_n\}$, kus n on vasakfaktorite arv.

Kui w on tühi sõna, siis vasakfaktor on võrdne keelega, ehk $\varepsilon^{-1}\mathcal{L} = \mathcal{L}$.

Kui $\mathcal{N}$ on keelt $\mathcal{L}$ aktsepteeriv automaat ja w on sõna oleku q vasakkeeles, siis q paremkeel on alamhulk keele $\mathcal{L}$ vasakfaktorist ehk $L_{q,F}(\mathcal{N}) \subseteq w^{-1}\mathcal{L}$.

Regulaarse keele $\mathcal{L}$ vasakfaktorite DFA on DFA $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$, kus olekute hulk Q on üksüheses vastavuses keele vasakfaktoritega, kõigi võimalike siirete puhul kehtib reegel $\delta(w^{-1}\mathcal{L}, a) = a^{-1}(w^{-1}\mathcal{L})$, algoleku q_1 paremkeel on $\varepsilon^{-1}\mathcal{L} = \mathcal{L}$ ja lõppolekute paremkeeled sisaldavad tühja sõna.

On teada, et keele $\mathcal{L}$ vasakfaktorite DFA on keele $\mathcal{L}$ minimaalne DFA.

Analoogselt vasakfaktoriga saame sõnastada ka paremfaktori mõiste.

Keele $\mathcal{L}$ paremfaktor sõnale w on regulaarne keel vasakjääkide x_j hulgast, mis saadakse sufiksi w eraldamisel kõigist keelde $\mathcal{L}$ kuuluvatest sõnadest. Ehk siis keele paremfaktor sõnale w on regulaarne keel $\mathcal{L}w^{-1} = \{x \in \Sigma^* | xw \in \mathcal{L}\}$.

3.2.8 Regulaarse keele aatomid

Regulaarse keele $\mathcal{L}$ aatomid [20] on mittetühjad regulaarsed keeled, mis moodustatakse keele vasakfaktorite ja nende täiendite kombinatsiooni ühisosana. Keele aatomite hulka tähistame edaspidi $A = \{A_1, \ldots, A_m\}$, kus m on aatomite arv.

Kui keele $\mathcal{L}$ vasakfaktorite hulk $K = \{K_1, \dots, K_n\}$, siis keele $\mathcal{L}$ aatom $A_i = \widetilde{K_1} \cap \widetilde{K_2} \cap \dots \cap \widetilde{K_n}$, kus $\widetilde{K_l}$ on keele $\mathcal{L}$ vasakfaktor K_l või vasakfaktori täiend $\overline{K_l}$ ja A_i ei ole tühi hulk.

Keele kahe suvalise aatomi ühisosa on alati tühi hulk.

Negatiivne aatom on vasakfaktorite täiendite ühisosa $\overline{K_1} \cap \overline{K_2} \cap \dots \cap \overline{K_n}$, kui see ei ole tühi hulk. Negatiivset aatomit tähistame edaspidi A^- .

Positiivne aatom on vasakfaktorite ja nende täiendite kombinatsiooni ühisosa, kui kombinatsioonis on vähemalt üks vasakfaktor ja kombinatsiooni ühisosa on mittetühi hulk.

Positiivne aatom on alati mõne vasakfaktori alamhulk.

Kõik vasakfaktorid on kirjeldatavad positiivsete aatomite ühenditena. Seega iga vasakfaktor on esitatav kujul $K_i = A_{i1} \cup \dots \cup A_{ik}$, kus $A_{i1}, \dots, A_{ik}$ on keele aatomid, mis esinevad vasakfaktoris K_i .

Näiteks, kui keele $\mathcal{L}$ vasakfaktorite hulk on $K = \{K_1, K_2\}$, siis keelel $\mathcal{L}$ on kolm kandidaat-hulka, mis on positiivsed aatomid, kui nad ei ole tühjad hulgad:

$$A_1 = K_1 \cap K_2 \neq \emptyset$$

$$A_2 = K_1 \cap \overline{K_2} \neq \emptyset$$

$$A_3 = \overline{K_1} \cap K_2 \neq \emptyset$$

Keele $\mathcal{L}$ negatiivne aatom on $A^- = \overline{K_1} \cap \overline{K_2}$, kui $\overline{K_1} \cap \overline{K_2} \neq \emptyset$.

Algaatom on aatom, mis on algse keele alamhulk. Seega A_i on keele $\mathcal{L}$ algaatom, kui $A_i \subseteq \mathcal{L}$, kus A_i on keele $\mathcal{L}$ aatom.

Lõppaatom on aatom, mis sisaldab tühja sõna. Seega A_i on keele $\mathcal{L}$ lõppaatom, kui $\varepsilon \in A_i$, kus A_i on keele $\mathcal{L}$ aatom.

3.3 Operatsioonid lõplike automaatidega, minimaalne DFA ja atomaat

3.3.1 NFA determiniseerimine

Determiniseerimine on NFA teisendamine temaga ekvivalentseks ehk sama keelt aktsepteerivaks DFA'ks. NFA determiniseerimise tulemust tähistame tähega D automaadi ülaindeksis, ehk siis NFA $\mathcal{N}$ determiniseerimise tulemus on $\mathcal{N}^D$. Kogu protseduuri tähistame skemaatiliselt $\mathcal{N} \xrightarrow{det} \mathcal{N}^D$.

Determiniseerimiseks kasutame algolekust saavutatavate alamhulkade meetodit.

Olekute alamhulkade moodustamisel kasutame järgmisi teadmisi:

- olekute alamhulga $s = \{q_1, \dots, q_n\}$ siirdefunktsioon sümbolile a on hulka kuuluvate olekute siirdefunktsioonide ühend $\delta(s, a) = \bigcup_{q_i \in s} \delta(q_i, a)$, kus $a \in \Sigma$;

- kui olekute alamhulka kuulub vähemalt üks lõppolek, siis on ka olekute alamhulk lõppolek.

Meetodi selgituseks determiniseerime peatükis 3.2.2 vaadeldud NFA.

(a)

NFA $\mathcal{N}$	a	b
$\rightarrow q_1$	$\{q_2\}$	-
$\rightarrow q_2$	-	$\{q_3\}$
$\rightarrow q_3$	$\{q_3\}$	$\{q_3, q_4\}$
$\leftarrow\rightarrow q_4$	-	-

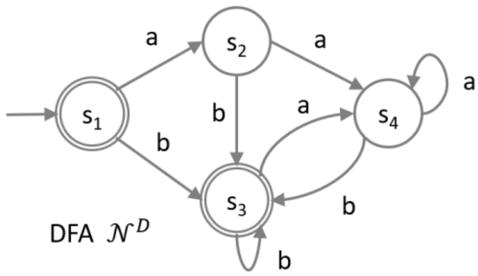
(b)

$\mathcal{N} \rightarrow \mathcal{N}^D$	a	b	uus
$\leftarrow\rightarrow\{q_1q_2q_3q_4\}$	$\{q_2q_3\}$	$\{q_3q_4\}$	s_1
$\{q_2q_3\}$	$\{q_3\}$	$\{q_3, q_4\}$	s_2
$\leftarrow\{q_3q_4\}$	$\{q_3\}$	$\{q_3, q_4\}$	s_3
$\{q_3\}$	$\{q_3\}$	$\{q_3, q_4\}$	s_4

(c)

DFA $\mathcal{N}^D$	a	b
$\leftarrow\rightarrow s_1$	s_2	s_3
s_2	s_4	s_3
$\leftarrow s_3$	s_4	s_3
s_4	s_4	s_3

(d)



Joonis 3 NFA determiniseerimine

Joonis 3 tabel (a) on NFA siirete tabel. Kogume tabelisse (b) kõik algolekute hulgast saavutatavad alamhulgad ja nende siirded. Esmalt ühendame kõik algolekud üheks alamhulgaks ja leiame selle alamhulga siirded kõigi automaadi tähestiku sümbolitega.

Selliselt saadud uued alamhulgad lisame tabelisse uuteks ridadeks ja leiame ka neile kõik siirded. Kordame tegevust, kuni kõik saavutatavad alamhulgad on leitud.

Lihtsustamise eesmärgil kasutame tabeli (b) teisendamisel DFA siirete tabeli reegleid järgivaks tabeliks (c) alamhulkade tähiseid, olekute liittähiste asemel. Joonise 3 osas (d) on leitud DFA graafiline kujutis.

3.3.2 Automaadi trimmimine (*trimming*)

Automaadi trimmimisel eemaldame automaadist kõik tühjad ja saavutamatud olekud.

Automaadi trimmimise tulemit tähistame T tähega automaadi ülaindeksis, ehk siis NFA $\mathcal{N}$ trimmimise tulemus on $\mathcal{N}^T$ ja kogu protseduuri tähistame $\mathcal{N} \xrightarrow{trim} \mathcal{N}^T$.

3.3.3 DFA minimeerimine, minimaalne DFA

DFA on minimaalne, kui ükski tema olekutest ei ole saavutamatu olek või ekvivalentne sama automaadi mõne teise olekuga [20].

Automaadi minimeerimine on minimaalse olekute arvuga ekvivalentse automaadi leidmise protsess. DFA minimeerimisel koondatakse kõik ekvivalentsed olekud üheks ja kõrvaldatakse saavutamatud olekud.

Automaadi minimeerimise tulemust tähistame tähega M automaadi ülaindeksis, seega DFA $\mathcal{D}$ minimeerimise tulemus on $\mathcal{D}^M$ ja kogu protseduuri tähistame $\mathcal{D} \xrightarrow{min} \mathcal{D}^M$.

Näiteks peatükis 3.3.1 determiniseerimise tulemusel saadud DFA's (Joonis 4 (a) allpool) kehtib võrduste paar $\delta(s_4, a) = \delta(s_2, a)$ ja $\delta(s_4, b) = \delta(s_2, b)$. Kuna olekud s_4 ja s_2 kumbki ei ole lõppolekud, saame võrdustest järeldada, et nende paremkeeled on võrdsed ja olekud on ekvivalentsed. DFA minimeerimiseks koondame need kaks seisundit üheks ja valime talle nimeks s_2. Muudatuse tulemusel saame siirete tabeli (b) allpool oleval joonisel.

Joonise tabelist (b) on näha, et $\delta(\mathrm{s}_1, a) = \delta(\mathrm{s}_3, a)$ ja $\delta(\mathrm{s}_1, \mathrm{b}) = \delta(\mathrm{s}_3, \mathrm{b})$ ja olekud s_1 ja s_3 on lõppolekud, seega s_1 ja s_3 on ekvivalentsed ja me saame nad koondada. Kehtivad ka võrdused $\delta(\mathrm{s}_1, a) = \delta(\mathrm{s}_2, a)$ ja $\delta(\mathrm{s}_1, \mathrm{b}) = \delta(\mathrm{s}_2, \mathrm{b})$, aga kuna s_2 ei ole lõppolek, ei sisalda tema paremkeel tühja sõna.

Seega ei ole oleku s_2 paremkeel võrdne olekute s_1 ja s_3 paremkeelega ja s_2 ei ole ekvivalentne eelmainitud olekutega. Koondades olekud s_1 ja s_3 olekuks s_1, saame siirete tabeli (c).

(a)

DFA $\mathcal{N}^D$	**a**	**b**
$\leftarrow\rightarrow s_1$	s_2	s_3
s_2	s_4	s_3
$\leftarrow s_3$	s_4	s_3
s_4	s_4	s_3

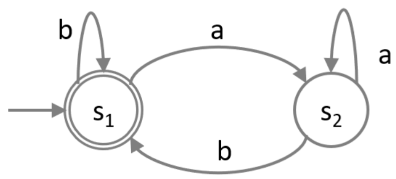
(b)

DFA $\mathcal{N}^D$	**a**	**b**
$\leftarrow\rightarrow s_1$	s_2	s_3
s_2	s_2	s_3
$\leftarrow s_3$	s_2	s_3

(c)

DFA $\mathcal{N}^{DM}$	**a**	**b**
$\leftarrow\rightarrow s_1$	s_2	s_1
s_2	s_2	s_1

(d)



Joonis 4 DFA minimeerimine

Kuna tabelis ei ole enam ekvivalentseid olekuid, oleme leidnud minimaalse DFA (Joonis 4 (c) ja (d)), mis on ekvivalentne esialgse DFA'ga (Joonis 3 (c) ja (d)).

Keele $\mathcal{L}$ minimaalse DFA D olekute paremkeelteks on keele $\mathcal{L}$ vasakfaktorid (3.2.7).

Seega keele $\mathcal{L}$ vasakfaktorid on:

$$K_1 = \Sigma^* b \cup \varepsilon = \mathcal{L}$$
$$K_2 = \Sigma^* b$$

3.3.4 Lõpliku automaadi ümberpööramine (*reversal*)

Automaadi ümberpööramisel vahetame alg- ja lõppolekud ning muudame siirete suuna vastupidiseks. Kui olek on korraga alg- ja lõppolek, siis jääb ta ka ümberpööramise tulemusena alg- ja lõppolekuks.

Ümberpööratud automaati tähistame tähega R automaadi ülaindeksis, seega NFA $\mathcal{N}$ ümberpööramise tulemit tähistame $\mathcal{N}^R$ ja kogu protseduuri tähistame $\mathcal{N} \xrightarrow{rev} \mathcal{N}^R$.

Automaadi pöördautomaat aktsepteerib esialgse automaadi keele pöördkeelt. Kui automaat $\mathcal{N}$ aktsepteerib keelt $\mathcal{L}$, siis tema pöördautomaat $\mathcal{N}^R$ aktsepteerib pöördkeelt $\mathcal{L}^R$. Seega, kui $L(\mathcal{N}) = \mathcal{L}$, siis $L(\mathcal{N}^R) = \mathcal{L}^R$.

Näitena pöörame ümber eelmises näites leitud DFA (Joonis 5 vasak pool). Kuna s_1 on alg- ja lõppolek, siis jääb ta ka pööratud automaadis alg- ja lõppolekuks. Automaadi siirded muudame automaadi pööramisel vastupidiseks. Kuna automaadis oli siire $\delta(s_1, a) = s_2$, siis pööratud automaadis on siire $s_1 \in \delta(s_2, a)$. Kuna esialgsel automaadil on ka siire $\delta(s_2, a) = s_2$, siis pööramise tulemusel tekib siirete hulk $\delta(s_2, a) = \{s_1, s_2\}$.

DFA $\mathcal{N}^{DM}$	**a**	**b**
$\leftarrow\rightarrow s_1$	s_2	s_1
s_2	s_2	s_1

NFA $\mathcal{N}^{DMR}$	**a**	**b**
$\leftarrow\rightarrow s_1$	-	$\{s_2, s_1\}$
s_2	$\{s_1, s_2\}$	-

Joonis 5 Automaadi pööramine

Joonise paremal poolel on pööramise tulemusena saadud pöördautomaat. Kui algne automaat aktsepteerib kõiki sõnu tähestikus $\{a, b\}$ mis lõppevad b tähega ja tühja sõna, siis pöördautomaat aktsepteerib kõiki sõnu tähestikus $\{a, b\}$ mis algavad b tähega ja tühja sõna.

Reeglina saame nii NFA kui ka DFA ümberpööramise tulemusena NFA.

3.3.5 Regulaarse keele atomaat

Kui $A = \{A_1, \dots, A_m\}$ on keele $\mathcal{L}$ aatomite hulk, siis ***keele $\mathcal{L}$ atomaat*** on keelt $\mathcal{L}$ aktsepteeriv NFA $\mathcal{A} = (Q, \Sigma, \delta, I, \{q_m\})$, kus $Q = \{q_1, q_2, \dots, q_m\}$ on üksüheses vastavuses aatomite hulgaga A, I on keele $\mathcal{L}$ algaatomitele vastavate olekute hulk, q_m on keele $\mathcal{L}$ lõppaatomile vastav olek ja $q_j \in \delta(q_i, a)$ siis ja ainult siis, kui $A_j \subseteq a^{-1}A_i$, A_j ja A_i on vastavalt NFA $\mathcal{A}$ olekutega q_j ja q_i seotud keele $\mathcal{L}$ aatomid ja $a \in \Sigma$ [20].

On teada, et keele $\mathcal{L}$ atomaadi olekute paremkeeled on keele $\mathcal{L}$ aatomid ehk

$L_{q_i,F}(\mathcal{A}) = \{w \in \Sigma^* | \delta(q_i, w) \cap F \ \neq \emptyset\} = A_i$ [20].

Keele $\mathcal{L}$ atomaat $\mathcal{A}$ on isomorfne automaatidega $\mathcal{D}^{RDR}$ ja $\mathcal{N}^{RDMR}$, kus $\mathcal{D}$ on keele $\mathcal{L}$ minimaalne DFA ja $\mathcal{N}$ on keelt $\mathcal{L}$ aktsepteeriv NFA. Seega on keele $\mathcal{L}$ atomaat $\mathcal{A}$ isomorfne keele $\mathcal{L}$ pöördkeele $\mathcal{L}^R$ minimaalse DFA pöördautomaadiga [20].

Eeltoodu annab meile mitu võimalust keele atomaadi leidmiseks.

Esimesel juhul alustame keele minimaalsest DFA'st, pöörame automaadi ümber, determiniseerime tulemuse ja seejärel pöörame tulemit veelkord. Peatükis 3.3.4 konstrueerisime juba pööratud automaadi peatükis 3.3.3 leitud minimaalsele DFA'le. Viidatud automaat (Joonis 6 (a)) on NFA, kuna ei ole rahuldatud DFA nõue, et igal seisundil peab olema iga tähestiku sümboli kohta täpselt üks võimalik siire.

(a)

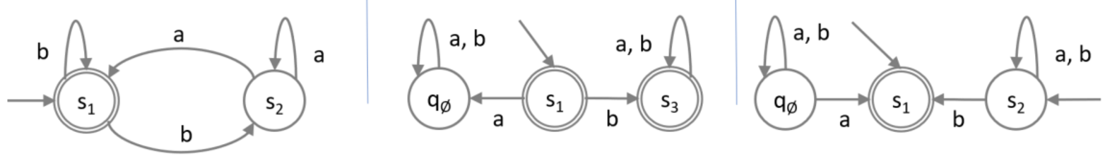
NFA $\mathcal{N}^{DMR}$	**a**	**b**
$\leftarrow\rightarrow s_1$	-	$\{s_2, s_1\}$
s_2	$\{s_1, s_2\}$	-

(b)

NFA $\mathcal{N}^{DMRD}$	**a**	**b**
$\leftarrow\rightarrow s_1$	$q_\emptyset$	s_3
$\leftarrow s_3$	s_3	s_3
$q_\emptyset$	$q_\emptyset$	$q_\emptyset$

(c)

NFA $\mathcal{N}^{DMRDR}$	**a**	**b**
$\leftarrow\rightarrow s_1$	$q_\emptyset$	s_3
$\leftarrow s_3$	s_3	s_3
$q_\emptyset$	$q_\emptyset$	$q_\emptyset$



Joonis 6 Atomaadi konstrueerimine

Seega esimese toiminguna determiniseerime automaadi. Selleks ühendame olekute grupid $s_1 \cup s_2 = s_3$ ja lisame tühja oleku puuduvate siirete jaoks. Uue olekute alamgrupi siirded suunduvad temasse endasse ehk siis $\delta(s_3, a) = \{s_1, s_2\} = s_3$ ja $\delta(s_3, b) = \{s_1, s_2\} = s_3$.

Determiniseerimise tulemus on ülalpool oleva joonise osas (b). Joonise osas (c) on eelmainitud automaadi pööramise tulemus – atomaat.

Eelnevalt leidsime keele $\mathcal{L}$ vasakfaktorid:

$$K_1 = \Sigma^* b \cup \{\varepsilon\} = \mathcal{L}$$
$$K_2 = \Sigma^* b$$

Vasakfaktorite täiendid on vastavalt:

$$\overline{K_1} = \Sigma^* a$$
$$\overline{K_2} = \Sigma^* a \cup \{\varepsilon\}$$

Kuna leidsime kaks vasakfaktorit, siis on neli võimalikku kombinatsiooni vasakfaktoritest ja nende täienditest. Seega keele aatomi kandidaadid on järgmised:

$$A_1 = K_1 \cap K_2 = K_2 = \Sigma^* b$$
$$A_2 = K_1 \cap \overline{K_2} = \{\varepsilon\}$$
$$A_3 = \overline{K_1} \cap K_2 = \emptyset$$
$$A^- = \overline{K_1} \cap \overline{K_2} = \overline{K_1} = \Sigma^* a$$

Kuna aatom peab olema mittetühi hulk, siis A_3 ei ole aatom. A_1 ja A_2 on positiivsed aatomid ja A^- on negatiivne aatom.

Teise võimalusena võib atomaadi leidmist alustada NFA'st, leida minimaalne DFA pöördkeelele ja see ümber pöörata.

Kuna keele $\mathcal{L}$ atomaat on pöördkeele $\mathcal{L}^R$ minimaalse DFA pöördautomaat, on tal alati täpselt üks lõppolek.

3.4 Kumerad keeled ja kinnised keeled

Kumera keele (*convex languages*) mõiste esitas Thierrin [28], [30]. Ang ja Brzozowski analüüsisid ja süstematiseerisid kumerate keelte alamklasse laiaulatuslikumalt [29].

Kumera keele klass määrab ära, millise sõnaosa suhtes keel on kumer. Üldiselt tuntakse viit kumerkeelte klassi: prefiks-, sufiks-, bifiks-, faktor- ja alamsõna-kumerat keelt. Üldistusena kasutatakse ka X-kumera keele mõistet, kus X on siis vastavalt kas prefiks, sufiks, bifiks, faktor või alamsõna.

X-kumeratel keeltel eristatakse nelja erijuhtu ehk alamklassi:

- X-kinnised keeled,
- X-vabad keeled,
- Ideaalid,
- X-kumerad keeled, mis ei kuulu kolme eeltoodud alamklassi.

Ideaal on keel, mis aktsepteerib sõnade prefiksina, sufiksina, või nii prefiksi kui sufiksina tähestiku kõiki võimalikke kombinatsioone ehk siis Σ^*. Täpsemalt, vasakideaal $\mathcal{L}_{LI} = \Sigma^*\mathcal{K}$, paremideaal $\mathcal{L}_{RI} = \mathcal{K}\Sigma^*$ ja kahepoolne ideaal $\mathcal{L}_{BI} = \Sigma^*\mathcal{K}\Sigma^*$, kus $\mathcal{K}$ on regulaarne keel tähestikus Σ [28].

On teada, et kõik X-kinnised keeled on ka X-kumerad keeled. Seega edaspidi me X-kinniste keelte puhul nende X-kumerust ei korda, vaid nimetame neid X-kinnisteks keelteks.

Käesolevas töös käsitleme peamiselt mittetühjasid prefiks-, sufiks- ja faktor-kinniseid keeli.

3.4.1 Prefiks-kinnine keel

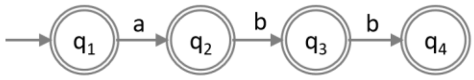
Keel $\mathcal{L}$ on ***prefiks-kumer*** (*prefix-convex*) siis ja ainult siis, kui sõnade u ja uvx kuulumisest keelde $\mathcal{L}$ järeldub, et ka sõna uv kuulub keelde $\mathcal{L}$ [32].

Prefiks-kumer keel $\mathcal{L}$ on ***prefiks-kinnine*** (*prefix-closed*) keel, kui sõna w kuuludes keelde $\mathcal{L}$ kuuluvad sellesse keelde ka kõik w prefiksid [33]. Seega, kui $w = u_i v_i$ ja $w \in \mathcal{L}$, siis ka $u_i \in \mathcal{L}$, kus $i \in \{1, \ldots, n+1\}$ on vabalt valitud lõike koht sõnas w ja n on sõna w pikkus.

Muu hulgas on tõestatud järgmised prefiks-kinniste keelte omadused:

- Prefiks-kinnine keel on parem-ideaali täiend. Ehk siis prefiks-kinnine keel $\mathcal{L}_{PC} = \overline{\mathcal{K}\Sigma^*}$, kus $\mathcal{K}$ on regulaarne keel tähestikus Σ [29].
- Kahe prefiks-kinnise keele ühend on prefiks-kinnine keel ehk kui $\mathcal{L}$ ja $\mathcal{K}$ on prefiks-kinnised keeled, siis ka $\mathcal{L} \cup \mathcal{K}$ on prefiks-kinnine keel [29].
- Kahe prefiks-kinnise keele ühisosa on prefiks-kinnine keel ehk kui $\mathcal{L}$ ja $\mathcal{K}$ on prefiks-kinnised keeled, siis ka $\mathcal{L} \cap \mathcal{K}$ on prefiks-kinnine keel [29].
- Prefiks-kinnise keele vasakfaktorid on prefiks-kinnised keeled [29].
- Igal prefiks-kinnisel keelel on seda aktsepteeriv NFA, mille olekud on kõik lõppolekud [31]. Edaspidi nimetame seda NFA'd ***prefiks-kinnise keele tunnus-NFA***'ks.

Näiteks on lihtne prefiks-kinnine keel $\mathcal{L}_{PC} = \{\varepsilon, a, ab, abb\}$.



	a	b
$\leftarrow\rightarrow q_1$	$\{q_2\}$	-
$\leftarrow q_2$	-	$\{q_3\}$
$\leftarrow q_3$	-	$\{q_4\}$
$\leftarrow q_4$	-	-

Joonis 7 Prefiks-kinnise keele NFA

Lisas 1 leiame sellele keelele vasakfaktorid ja aatomid.

Vasakfaktorid :

$K_1 = \{\varepsilon, a, ab, abb\} = \mathcal{L}$ $\quad$ $K_2 = \{\varepsilon, b, bb\}$

$K_3 = \{\varepsilon, b\}$ $\quad$ $K_4 = \{\varepsilon\}$

$K_\emptyset = \emptyset$

Aatomid:

$A_1 = \{\varepsilon\}$ $\quad$ $A_2 = \{b\}$

$A_3 = \{a, ab, abb\}$ $\quad$ $A_4 = \{bb\}$

$A_0 = \Sigma^* aa + \Sigma^* aab + \Sigma^* aabb + \Sigma^* ba + \Sigma^* bab + \Sigma^* babb + \Sigma^* bbb$

3.4.2 Sufiks-kinnine keel

Keel $\mathcal{L}$ on ***sufiks-kumer*** *(suffix-convex)* siis ja ainult siis, kui sõnade x ja uvx kuulumisest keelde $\mathcal{L}$ järeldub, et ka sõna vx kuulub keelde $\mathcal{L}$.

Sufiks-kumer keel $\mathcal{L}$ on ***sufiks-kinnine*** *(suffix-closed)* keel, kui sõna w kuuludes keelde $\mathcal{L}$ kuuluvad sellesse keelde ka kõik w sufiksid [33]. Seega, kui $w = u_i v_i$ ja $w \in \mathcal{L}$, siis ka $v_i \in \mathcal{L}$, kus $i \in \{1, \dots, n+1\}$ on vabalt valitud lõike koht sõnas w ja n on sõna w pikkus.

Muu hulgas on tõestatud järgmised sufiks-kinniste keelte omadused:

- Sufiks-kinnine keel on vasak-ideaali täiend. Seega sufiks-kinnine keel $\mathcal{L}_{SC} = \overline{\Sigma^* \mathcal{K}}$, kus $\mathcal{K}$ on regulaarne keel tähestikus Σ [29].

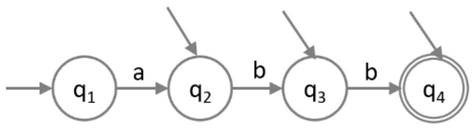
- Kahe sufiks-kinnise keele ühend on sufiks-kinnine keel. Seega, kui $\mathcal{L}$ ja $\mathcal{K}$ on sufiks-kinnised keeled, siis ka $\mathcal{L} \cup \mathcal{K}$ on sufiks-kinnine keel [29].

- Kahe sufiks-kinnise keele ühisosa on sufiks-kinnine keel. Seega kui $\mathcal{L}$ ja $\mathcal{K}$ on sufiks-kinnised keeled, siis ka $\mathcal{L} \cap \mathcal{K}$ on sufiks-kinnine keel [29].

- Sufiks-kinnise keele paremfaktorid on sufiks-kinnised keeled [29].

- Iga sufiks-kinnise keele jaoks leidub seda aktsepteeriv NFA, mille olekud on kõik algolekud ja millel on üks lõppolek [31]. Edaspidi nimetame seda ***sufiks-kinnise keele tunnus-NFA'ks.***

Sufiks-kinnist keelt kasutasime eelmise osa näidetes. Vaatame siiski ka lihtsat sufiks-kinnist keelt $\mathcal{L}_{SC} = \{\varepsilon, b, bb, abb\}$.



	a	b
→q1	{q2}	-
→q2	-	{q3}
→q3	-	{q4}
←→q4	-	-

Joonis 8 Sufiks-kinnise keele NFA

Lisas 2 Leiame sellele keelele vasakfaktorid ja aatomid.

Vasakfaktorid :

$K_1 = \{\varepsilon, b, bb, abb\} \quad = \mathcal{L}$

$K_2 = \{bb\}$

$K_3 = \{\varepsilon, b\}$

$K_4 = \{b\}$

$K_5 = \{\varepsilon\}$

$K_\emptyset = \emptyset$

Aatomid:

$A_1 = \{\varepsilon\}$

$A_2 = \{b\}$

$A_3 = \{bb\}$

$A_4 = \{abb\}$

$A_0 = \Sigma^* a + \Sigma^* ab + \Sigma^* bbb + \Sigma^* aabb + \Sigma^* babb$

3.4.3 Bifiks-kinnine keel

Bifiks-kumer *(bifix-convex)* keel on keel, mis on korraga nii prefiks- kui ka sufiks-kumer.

Bifiks-kinnine (*bifix-closed*) keel on keel, mis on korraga nii prefiks- kui ka sufiks-kinnine keel [29]. Bifiks-kumer keel $\mathcal{L}$ on bifiks-kinnine keel, kui sõna w kuuludes keelde $\mathcal{L}$ kuuluvad sellesse keelde ka kõik w prefiksid ja sufiksid. Ehk siis kui

$w = u_i v_i$ ja $w \in \mathcal{L}$, siis ka $u_i, v_i \in \mathcal{L}$, kus $i \in \{1, \ldots, n+1\}$ on vabalt valitud lõike koht sõnas w.

Muu hulgas on tõestatud järgmised bifiks-kinniste keelte omadused:

- Kahe bifiks-kinnise keele ühend on bifiks-kinnine keel ehk kui $\mathcal{L}$ ja $\mathcal{K}$ on bifiks-kinnised keeled, siis ka $\mathcal{L} \cup \mathcal{K}$ on bifiks-kinnine keel [29].
- Kahe bifiks-kinnise keele ühisosa on bifiks-kinnine keel ehk kui $\mathcal{L}$ ja $\mathcal{K}$ on bifiks-kinnised keeled, siis ka $\mathcal{L} \cap \mathcal{K}$ on bifiks-kinnine keel [29].
- Kõik bifiks-kinnised keeled on alati ka faktor-kinnised [29].

3.4.4 Faktor-kinnine keel

Keel $\mathcal{L}$ on ***faktor-kumer*** *(factor-convex)* siis ja ainult siis, kui sõnade u, x ja uvx kuulumisest keelde $\mathcal{L}$ järeldub, et ka sõnad uv ja vx kuuluvad keelde $\mathcal{L}$.

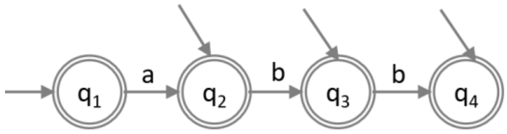
Faktor-kumer keel $\mathcal{L}$ on ***faktor-kinnine*** *(factor-closed)* keel, kui sõna w kuuludes keelde $\mathcal{L}$ kuuluvad sellesse keelde ka kõik w faktorid [33]. Ehk siis kui $w = u_i x_{ij} v_j$ ja $w \in \mathcal{L}$, siis ka $u_i, x_{ij}, v_j \in \mathcal{L}$,kus $i, j \in \{1, \ldots, n+1\}$ on vabalt valitud lõikekohad sõnas w selliselt, et $i \leq j$.

Muu hulgas on tõestatud järgmised faktor-kinniste keelte omadused:

- Faktor-kinnine keel on alati ka bifiks-kinnine keel [29].
- Faktor-kinnine keel on kahepoolse ideaali täiend. Ehk siis faktor-kinnine keel $\mathcal{L}_{FC} = \overline{\Sigma^* \mathcal{K} \Sigma^*}$, kus $\mathcal{K}$ on regulaarne keel tähestikus Σ [34].
- Kahe faktor-kinnise keele ühend on faktor-kinnine keel ehk kui $\mathcal{L}$ ja $\mathcal{K}$ on faktor-kinnised keeled, siis ka $\mathcal{L} \cup \mathcal{K}$ on faktor-kinnine keel [29].
- Kahe faktor-kinnise keele ühisosa on faktor-kinnine keel ehk kui $\mathcal{L}$ ja $\mathcal{K}$ on faktor-kinnised keeled, siis ka $\mathcal{L} \cap \mathcal{K}$ on faktor-kinnine keel [29].

- Igale faktor-kinnisele keelele leidub teda aktsepteeriv NFA, mille olekud on kõik nii alg- kui ka lõppolekud [31]. Edaspidi nimetame seda ***faktor-kinnise keele tunnus-NFA'ks.***

Vaatame näitena jällegi lihtsat faktor-kinnist keelt $\mathcal{L}_{FC} = \{\varepsilon, a, b, ab, bb, abb\}$.



	a	b
←→q_1	$\{q_2\}$	-
←→q_2	-	$\{q_3\}$
←→q_3	-	$\{q_4\}$
←→q_4	-	-

Joonis 9 Faktor-kinnise keele NFA

Lisas 3 on toodud selle keele vasakfaktorite ja aatomite leidmine.

Vasakfaktorid

$K_1 = \{\varepsilon, a, b, ab, bb, abb\}$ $K_2 = \{\varepsilon, b, bb\}$

$K_3 = \{\varepsilon, b\}$ $K_4 = \{\varepsilon\}$

$K_\emptyset = \emptyset$

Aatomid

$A_1 = \{\varepsilon\}$ $A_2 = \{b\}$

$A_3 = \{bb\}$ $A_4 = \{abb, ab, a\}$

$$A_0 = \Sigma^* a + \Sigma^* ab + \Sigma^* bbb + \Sigma^* aabb + \Sigma^* babb$$

3.4.5 Alamsõna-kinnine keel

Keel $\mathcal{L}$ on ***alamsõna-kumer*** *(subword-convex)* siis ja ainult siis, kui sõnade u ja w kuulumisest keelde $\mathcal{L}$ järeldub, et ka sõna v kuulumine keelde $\mathcal{L}$, kus u on v alamsõna ning u ja v on w alamsõnad [30].

Alamsõna-kumer keel $\mathcal{L}$ on ***alamsõna-kinnine*** *(subword-closed)* keel, kui sõna w kuuludes keelde $\mathcal{L}$ kuuluvad sellesse keelde ka kõik w alamsõnad [30].

Muu hulgas on tõestatud järgmised alamsõna-kinniste keelte omadused:

- Kahe alamsõna-kinnise keele ühend on alamsõna-kinnine keel ehk kui $\mathcal{L}$ ja $\mathcal{K}$ on alamsõna-kinnised keeled, siis ka $\mathcal{L} \cup \mathcal{K}$ on alamsõna-kinnine keel [29].

- Kahe alamsõna-kinnise keele ühisosa on alamsõna-kinnine keel ehk kui $\mathcal{L}$ ja $\mathcal{K}$ on alamsõna-kinnised keeled, siis ka $\mathcal{L} \cap \mathcal{K}$ on alamsõna-kinnine keel [29].

- Kõik alamsõna-kinnised keele vasakfaktorid on alamsõna kinnised keeled [29].

4 Kinniste keelte vasakfaktorite ja aatomite omadused

Töös käsitleme ainult mittetühjasid regulaarseid keeli.

Järgnevates peatükkides analüüsime eraldi kinniste keelte alamgruppe ja leiame neist igale vasakfaktorite ja aatomite iseloomulikud omadused.

4.1 Bifiks- ja alamsõna-kinnised keeled

On teada, et kõik bifiks- ja alamsõna-kinnised keeled on samaaegselt ka faktor-kinnised keeled [29].

Kuna kõik faktor-kinnised keeled on ka bifiks-kinnised [29], siis on nende keelte hulgad võrdsed ja kõik faktor-kinniste keelte kohta öeldu kehtib ka bifiks-kinniste keelte kohta.

Kuna kõik alamsõna-kinnised keeled on ka faktor-kinnised keeled, siis on neil ka faktor-kinniste keelte tunnus-NFA ning nende vasakfaktoritel ja aatomitel on faktor-kinniste keelte vastavate komponentide omadused.

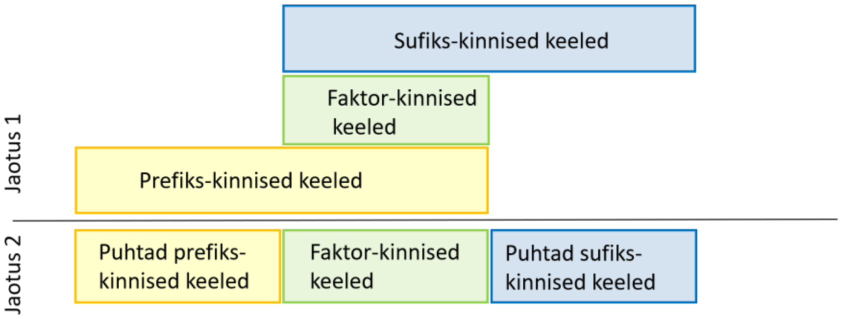
Alamsõna-kinnistel keeltel on omadusi, mida ei ole kõikidel faktor-kinnistel keeltel. Näiteks alamsõna-kinniste keelte vasakfaktorid on alamsõna kinnised keeled [29]. Selliseid varasemalt kirjeldamata alamsõna-kinnise keele omadusi me käesoleva töö ettevalmistamisel ei leidnud.

Lähtuvalt eeltoodust me neid kahte kinniste keelte alamklassi eraldi ei käsitle.

4.2 Prefiks- ja sufiks-kinnised keeled

Prefiks-kinniste keelte klassi kuuluvad ka keeled, mis on samal ajal ka sufiks-kinnised keeled. Sama kehtib ka vastupidisena: sufiks-kinniste keelte klassi kuuluvad ka keeled, mis on ka prefiks-kinnised.

Prefiks- ja sufiks-kinniste keelte ühisosa moodustab faktor-kinniste keelte hulga. Ehk keeled, mis on üheaegselt nii prefiks- kui sufiks-kinnised on ka faktor-kinnised keeled (Joonis 10, jaotus 1).



Joonis 10 Kinniste keelte jaotused

Kogu vaadeldavate kinniste keelte hulga saame leida prefiks-kinniste ja sufiks-kinniste keelte ühendina.

Jagame selle keelte hulga kolme keelte gruppi (Joonis 10, jaotus 2):

- ***puhtad prefiks-kinnised keeled*** ehk keeled, mis on ainult prefiks-kinnised ja ei ole samal ajal sufiks-kinnised ja faktor-kinnised keeled;
- ***puhtad sufiks-kinnised keeled*** ehk keeled, mis on ainult sufiks-kinnised ja ei ole samal ajal prefiks- ja faktor-kinnised keeled;
- ***faktor-kinnised keeled*** ehk prefiks- ja sufiks-kinniste keelte ühisosa.

Nendel keelte gruppidel puuduvad omavahelised ühisosad, seega on keeled omaduste lõikes paremini grupeeritud kui esimesena toodud jaotuses, kus keelte klassidel on ühisosa ja kõigis kolmes grupis on keeli, mis on nii prefiks- kui ka sufiks-kinnised.

Lähtuvalt keelte definitsioonide vastandlikkusest võib oletada, et prefiks- ja sufiks-kinnistel keeltel, nende vasakfaktoritel ja aatomitel on vastandlikke omadusi, mis ei ilmne, kui vaatleme neid keeli ühes grupis. Selliste omaduste leidmisel ja tõestamisel saame kasutada puhta keele mõistet ja omadusi.

4.3 Puhtad kinnised keeled

Vaatame veidi lähemalt grupeerimise tulemusel tekkinud kahte uut keelte gruppi.

Puhas prefiks-kinnine keel on prefiks-kinnine keel, mis ei ole samal ajal sufiks-, bifiks- või faktor-kinnine.

Puhas sufiks-kinnine keel on sufiks-kinnine keel, mis ei ole samal ajal prefiks-, bifiks- või faktor-kinnine.

Lause 1. *Puhas kinnine keel sisaldab lisaks tühjale sõnale vähemalt kahte sõna ja kasutab vähemalt kahte erinevat sümbolit.*

Tõestus

Vaatleme kõigepealt prefiks-kinnist keelt, mis koosneb ühest sõnast ja tühjast sõnast – $\mathcal{L}_{PC} = \{\varepsilon,\ x\}$. Kui sõna x koosneb mitmest sümbolist, siis ei sisalda keel kõiki keelde kuuluvate sõnade prefikseid iseseisvate sõnadena ja keel ei saa olla prefiks-kinnine keel. Seega sõna x saab koosneda vaid ühest sümbolist, olles selliselt ka iseenda prefiks. Samas on ta siis ka iseenda sufiks ja faktor. Sellest tulenevalt on tegu faktor-kinnise keelega.

Teiseks vaatleme prefiks-kinnist keelt, milles on lisaks tühjale sõnale kaks sõna ja mis kasutab ainult ühte sümbolit – $\mathcal{L}_{PC}^{min} = \{\varepsilon,\ x_1,\ x_1x_1\}$. Kuna sümbol x_1 on üheaegselt sõna x_1x_1 prefiks, sufiks kui ka faktor, siis on vaadeldav keel ka sufiks- ning faktor-kinnine keel.

Vaatleme nüüd lausele vastavat prefiks-kinnist keelt $\mathcal{L}_{PC}^{min} = \{\varepsilon,\ x_1,\ x_1x_2\}$, kus $x_1, x_2 \in \Sigma$ ja $x_1 \neq x_2$. Kuna keeles puudub sõna x_2, siis ei saa keel $\mathcal{L}_{PC}^{min}$ olla sufiks- või faktor-kinnine keel. Seega on tegu puhta prefiks-kinnise keelega.

Sama kehtib ka sufiks-kinnise keele puhul. Lausele vastav sufiks-kinnine keel $\mathcal{L}_{SC}^{min} = \{\varepsilon,\ x_2,\ x_1x_2\}$, kus $x_1, x_2 \in \Sigma$ ja $x_1 \neq x_2$, ei saa olla prefiks- ega faktor-kinnine kuna temas puudub sõna x_1.

Seega kaks viimati vaadeldud keelt on vastavalt minimaalsed prefiks- ja sufiks-kinnised keeled. □

4.4 Prefiks-kinnise keele vasakfaktorid

Peatükis 3.3.3 märkisime, et keele vasakfaktorid on võrdsed keele minimaalse DFA olekute paremkeeltega. Seega vaatleme kõigepealt prefiks-kinnise keele minimaalset DFA'd.

4.4.1 Prefiks-kinnise keele minimaalne DFA

On teada, et iga prefiks-kinnise keele jaoks leidub NFA, mille kõik olekud on lõppolekud [31]. Sellest tulenevalt saame väita järgmist:

Lause 2. *Prefiks-kinnise keele minimaalse DFA kõik mittetühjad olekud on lõppolekud.*

Seda väidet saame tõestada kahest erinevast lähtekohast vaadatuna.

Tõestus variant 1:

Esimesel juhul lähtume determiniseerimise ja minimeerimise protseduuridest, mida vaatlesime töö eelmises osas.

Võtame prefiks-kinnist keelt $\mathcal{L}_{\mathrm{PC}}$ aktsepteeriva NFA $\mathcal{N}_{PC} = (Q_{PC}, \Sigma, \delta_{PC}, I_{PC}, F_{PC})$, kus kõik olekud on lõppolekud ehk $F_{PC} = Q_{PC}$ ja leiame talle determiniseerimise tulemi $\mathcal{N}_{PC} \xrightarrow{det} \mathcal{N}_{PC}^{D}$, kus $\mathcal{N}_{PC}^{D} = (Q_{PC}^{D}, \Sigma, \delta_{PC}^{D}, s_1, F_{PC}^{D})$.

NFA determiniseerimisel ühendame tema olekud alamhulkadeks $s_1, s_2, \dots, s_n \subseteq Q_{PC}$. Seega DFA $\mathcal{N}_{PC}^{D}$ olekute hulk $Q_{PC}^{D} = \{s_1, s_2, \dots, s_n, s_\emptyset\}$, kui determiniseeritud automaadil on tühi olek, ja $Q_{PC}^{D} = \{s_1, s_2 \dots s_n\}$, kui seda ei ole.

Ühendamisel kehtib reegel, et kui olekute alamhulgas on mõni lõppolek, siis ka see olekute alamhulk on lõppolek. Seega, kui $q_i \in F_{PC}$ ja $q_i \in s_j$, siis $s_j \subseteq F_{PC}^{D}$. Kuna NFA $\mathcal{N}_{PC}$ kõik olekud on lõppolekud, siis ka kõik determiniseerimisel saadavad olekute alamhulgad on lõppolekud ehk $s_1, s_2, \dots, s_n \subseteq F_{PC}^{D}$ ja lõppolekute hulk on $F_{PC}^{D} = \{s_1, s_2, \dots, s_n\}$ ning $F_{PC}^{D} = Q_{PC}^{D} \setminus s_\emptyset$. Seega $\mathcal{N}_{PC}^{D}$ kõik mittetühjad olekud on lõppolekud.

Determiniseerimise tulemus $\mathcal{N}_{PC}^{D}$ ei ole reeglina minimaalne DFA. Seega peame vaatama ka minimeerimise protseduuri.

Minimeerime determiniseerimise tulemi $\mathcal{N}_{PC}^{D} \xrightarrow{min} \mathcal{N}_{PC}^{DM}$, kus $\mathcal{N}_{PC}^{DM} = (Q_{PC}^{DM}, \Sigma, \delta_{PC}^{DM}, s_1, F_{PC}^{DM})$. Minimeerimisel koondame olekud, millede paremkeeled on võrdsed. Seega, kui $L_{s_i, F_{PC}^{D}}(\mathcal{N}_{PC}^{D}) = L_{s_j, F_{PC}^{D}}(\mathcal{N}_{PC}^{D})$, siis koondamise tulem $s_k = \{s_i, s_j\}$ ja $L_{s_k, F_{PC}^{DM}}(\mathcal{N}_{PC}^{DM}) = L_{s_i, F_{PC}^{D}}(\mathcal{N}_{PC}^{D}) = L_{s_j, F_{PC}^{D}}(\mathcal{N}_{PC}^{D})$.

On üldiselt teada, et ainult lõppoleku paremkeel sisaldab tühja sõna. Seega, lähtuvalt paremkeelte võrdsuse nõudest, saab lõppolekut koondada ainult teise lõppolekuga ning tulemuseks on jällegi lõppolek. Kuna automaadi $\mathcal{N}_{PC}^{D}$ kõik mittetühjad olekud on lõppolekud, siis ka minimeerimisel moodustatavad olekud peavad olema lõppolekud. Seega leitava minimaalse DFA kõik mittetühjad olekud on lõppolekud ehk $\mathcal{N}_{PC}^{DM} = (Q_{PC}^{DM}, \Sigma, \delta_{PC}^{DM}, s_1, F_{PC}^{DM})$, kus $F_{PC}^{DM} = Q_{PC}^{DM} \setminus s_{\emptyset}$. □

Tõestus variant 2:

Teisel juhul lähtume keele iseloomustusest.

Lähtuvalt prefiks-kinnise keele definitsioonist kehtib alati reegel, et kui sõna w kuulub keelde $\mathcal{L}$ ja u on sõna w prefiks, siis ka u peab kuuluma keelde $\mathcal{L}$.

Vaatleme keele $\mathcal{L}$ minimaalset DFA'd $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$, millel on rohkem kui üks mittetühi olek ja üks olek q_i ei ole lõpp- ega tühi olek ning sõna $uv \in \mathcal{L}$, millel on DFA $\mathcal{D}$ aktsepteerib siiretejaga $\delta(q_1, u) = q_i$ ja $\delta(q_i, v) = q_f$, kus $q_f \in F$. Seega $u \in L_{q_1 q_i,}(\mathcal{D})$ ja on sõna uv prefiks.

Selleks, et keel $\mathcal{L}$ rahuldaks prefiks-kinnise keele nõuet, peab ka sõna uv prefiks u kuuluma keelde $\mathcal{L}$. Samas kuna q_i ei ole lõppolek, siis tema vasakkeele sõnad ei saa kuuluda keelde $\mathcal{L}$, sest DFA'd $\mathcal{D}$ ei aktsepteeri seda. Kuna DFA puhul on igal tähestiku sümbolil igast olekust väljumiseks täpselt üks võimalik siire, siis on selge, et kindlal sõnal ei saa olla DFA's kahte võimalikku siirete jada algolekust automaadi läbimiseks. Seega saame teha järelduse, et kui minimaalses DFA's ei ole kõik mittetühjad olekud lõppolekud, siis selle automaadi keel ei saa olla prefiks-kinnine keel. □

4.4.2 Prefiks-kinnise keele vasakfaktorite omadused

Teadmistest, et prefiks-kinnise keele minimaalse DFA kõik mittetühjad olekud on lõppolekud ja minimaalse DFA paremkeeled on keele vasakfaktorid, saame teha mitmeid järeldusi keele vasakfaktorite omaduste kohta.

Lause 3. *Kui keelt $\mathcal{L}$ aktsepteeriva minimaalse DFA kõik mittetühjad olekud on lõppolekud, siis ka keele $\mathcal{L}$ vasakfaktoritel on neid aktsepteerivad NFA'd, mille kõik olekud on lõppolekud.*

Tõestus

On lihtne näha, et kui keele $\mathcal{L}$ minimaalse DFA $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$ kõik mittetühjad olekud on lõppolekud ehk $F = Q \setminus s_\emptyset$, siis ka $\mathcal{D}$ oleku q_i paremkeelt aktsepteerivas osas $\mathcal{D}_{q_{i,F}} = (Q_{q_{i,F}}, \Sigma, \delta, q_i, F_{q_i})$ on kõik olekud $Q_{q_{i,F}} = \{q_i, \dots q_f\}$ lõppolekud ja kehtib võrdus $\mathrm{F}_{q_i} = \mathrm{Q}_{q_{i,F}}$. Me võime eraldada selle osa automaadist. Selliselt saame DFA $\mathcal{D}$ oleku q_i paremkeelt aktsepteeriva NFA $\mathcal{N} = (Q, \Sigma, \delta, I, F)$, kus algolekute hulk $I = \{q_i\}$, ja mille kõik olekud on lõppolekud. Seega igal DFA $\mathcal{D}$ oleku paremkeelel on teda aktsepteeriv NFA, mille kõik olekud on lõppolekud.

Kuna DFA $\mathcal{D}$ on keele $\mathcal{L}$ minimaalne DFA, siis keele $\mathcal{L}$ vasakfaktorid on kõik DFA $\mathcal{D}$ paremkeeled. Seega on ka igal keele $\mathcal{L}$ vasakfaktoril teda aktsepteeriv NFA, mille kõik olekud on lõppolekud. □

Lause 4. *Prefiks-kinnise keele vasakfaktorite ühend on faktor-kinnine keel.*

Tõestus

Lähtuvalt vasakfaktori definitsioonist $w^{-1}\mathcal{L} = \{x \in \Sigma^* | wx \in \mathcal{L}\}$ saame öelda, et iga vasakfaktor on kogum keele sõnade sufiksitest, mis saadakse kindla prefiksi eraldamisel kõigist keele sõnadest.

Olgu prefiks-kinnise keele $\mathcal{L}$ sõnade kõigi võimalike prefiksite hulk $P = \{p_{1p}, p_{2p}, \dots\}$, sufiksite hulk $S = \{s_{1s}, s_{2s}, \dots\}$ ja vasakfaktorite hulk $K = \{K_1, K_2, \dots, K_n\}$. Kuna keel $\mathcal{L}$ on prefiks-kinnine keel, siis $P = \mathcal{L}$.

Tulenevalt definitsioonist $K_i = p_{jp}^{-1}\mathcal{L} = \{s_{js} \in \mathrm{S} | p_{jp}s_{js} \in \mathcal{L}\}$.

Kuna keele $\mathcal{L}$ iga prefiks $p_{jp} \in P$ on seotud eeltoodu kaudu mõne vasakfaktoriga, siis on selge, et igal keele sufiksil $s_{js} \in S$ on vähemalt üks vasakfaktor, millesse ta kuulub. Seega, kui sufiks $s_{is} \in S$, siis ka $s_{is} \in K_i$ ja sellest tulenevalt ka $s_{is} \in K_1 \cup K_2 \cup \ldots \cup K_n$.

Kuna keele $\mathcal{L}$ iga võimalik sufiks kuulub vasakfaktorite ühendisse, siis on kõigi vasakfaktorite ühend sufiks-kinnine keel.

Eelnevalt nägime, et prefiks-kinnise keele kõik vasakfaktorid on prefiks-kinnised keeled. On teada, et prefiks-kinniste keelte ühend on jällegi prefiks-kinnine keel [29].

Seega on prefiks-kinnise keele vasakfaktorite ühend korraga nii prefiks- kui ka sufiks-kinnine keel ehk faktor-kinnine keel. □

Eeltoodust on lihtne näha, et prefiks-kinnise keele vasakfaktoritel on all loetletud omadused.

Järeldused Lausetest 2 - 4

1. Kuna kõigil prefiks-kinnise keele vasakfaktoritel on NFA, mille kõik olekud on lõppolekud, siis kõik prefiks-kinnise keele vasakfaktorid on prefiks-kinnised keeled. Kõnealuse omaduse leidsid ka Ang ja Brzozowski [29] kasutades omaduse tõestuseks teist lähenemist.

2. Kuna prefiks-kinnise keele minimaalse DFA kõik mittetühjad olekud on lõppolekud ja lõppoleku paremkeel sisaldab tühja sõna, siis kõik prefiks-kinnise keele mittetühjad vasakfaktorid sisaldavad endas tühja sõna.

3. Kuna tühi sõna kuulub igasse prefiks-kinnise keele mittetühja vasakfaktorisse ja prefiks-kinnise keele kõigi sõnade kõik prefiksid peavad kuuluma sellesse keelde, siis sõnade arv igas prefiks-kinnise keele mittetühjas vasakfaktoris on suurem kui temasse kuuluva pikima sõna sümbolite arv.

4. Kuna prefiks-kinnise keele minimaalse DFA kõik mittetühjad olekud on lõppolekud, siis ka igast selle DFA olekust mittetühja olekusse suunduv siire jõuab lõppolekusse. Seega prefiks-kinnise keele iga vähemalt kahte sõna sisaldav vasakfaktor sisaldab vähemalt ühte ühetähelist sõna.

5. Eelmisest tulenevalt on selge ka, et kui tähestiku mõni sümbol esineb prefiks-kinnises keeles, siis peab ta sisalduma vähemalt ühes vasakfaktoris ühetähelise sõnana ja vastupidi. Ehk kui tähestiku sümbol ei sisaldu ühetähelise sõnana üheski prefiks-kinnise keele vasakfaktoris, siis ei kasuta see keel seda tähestiku sümbolit.
6. Prefiks-kinnise keele vasakfaktorite ühend on faktor-kinnine keel (Lause 4).
7. Kuna kõik prefiks-kinnise keele mittetühjad vasakfaktorid sisaldavad tühja sõna, siis nende vasakfaktorite ühisosa ei ole tühi hulk.
8. Kuna prefiks-kinnise keele mittetühjad vasakfaktorid on prefiks-kinnised keeled ja prefiks-kinniste keelte ühisosa on prefiks-kinnine keel [29], siis prefiks-kinnise keele mittetühjade vasakfaktorite ühisosa on prefiks-kinnine keel, mis sisaldab vähemalt tühja sõna.

4.5 Sufiks-kinnise keele vasakfaktorid

Nii nagu prefiks-kinnise keele puhul, vaatame ka sufiks-kinnise keele vasakfaktorite leidmiseks kõigepealt nende keelte minimaalset DFA'd.

4.5.1 Sufiks-kinnise keele minimaalne DFA

On teada, et iga sufiks-kinnise keele jaoks leidub NFA, mille kõik olekud on algolekud ja millel on ainult üks lõppolek [31].

Samuti on teada, et keelt aktsepteeriva DFA determiniseeritud pöördautomaat on selle keele pöördkeele minimaalne DFA. Seega, kui $\mathcal{D}$ on keelt $\mathcal{L}$ aktsepteeriv DFA, siis $\mathcal{D}^{RD}$ on keele $\mathcal{L}$ pöördkeele $\mathcal{L}^R$ minimaalne DFA [19].

Lause 5. *Sufiks-kinnine keel on prefiks-kinnise keele pöördkeel.*

Tõestus

Peatükis 4.4.1 tõestasime, et prefiks-kinnise keele $\mathcal{L}_{\mathrm{PC}}$ minimaalse DFA $\mathcal{D} = \{Q, \Sigma, \delta, q_1, F\}$ kõik mittetühjad olekud on lõppolekud ehk $F = Q \setminus q_\emptyset$ (Lause 2). Leiame nüüd sellele DFA'le pöördautomaadi $\mathcal{D}^R$:

$\mathcal{D} \xrightarrow{rev} \mathcal{D}^R$, kus $\mathcal{D}^R = (Q^R, \Sigma, \delta^R, I^R, F^R)$, $I^R = F$ ja $F^R = \{q_1\}$.

Kuna $\mathcal{D}$ on automaat, kus kõik mittetühjad olekud on lõppolekud ja minimaalsele DFA'le omaselt on ainult üks algolek, siis automaadi pööramise tulemusena saame NFA, kus kõik olekud on algolekud ja millel on üks lõppolek.

Leitud pöördautomaat vastab sufiks-kinnise keele tunnus-NFA'le. Seega võime järeldada, et prefiks-kinnise keele pöördkeel on sufiks-kinnine keel. □

Kuna sufiks-kinnine keel on prefiks-kinnise keele pöördkeel, siis eelmise tõestuse käigus leitud NFA determiniseerimisel saame sufiks-kinnise keele $\mathcal{L}_{\text{SC}}$ minimaalse DFA, mille paremkeeled on sufiks-kinnise keele $\mathcal{L}_{\text{SC}}$ vasakfaktorid.

Leiame sellele DFA'le iseloomulikud omadused.

Lause 6. *Sufiks-kinnise keele minimaalse DFA algolek on alati ka lõppolek.*

Tõestus

Vaatame eelpool leitud sufiks-kinnise keele $\mathcal{L}_{\text{SC}}$ tunnud-NFA $\mathcal{D}^R = (Q^R, \Sigma, \delta^R, I^R, F^R)$ determiniseerimist $\mathcal{D}^R \xrightarrow{det} \mathcal{D}^{RD}$, kus $\mathcal{D}^{RD} = (Q^{RD}, \Sigma, \delta^{RD}, s_1, F^{RD})$ on keele $\mathcal{L}_{\text{SC}}$ minimaalne DFA, s_1 on selle algolek ja Q^{RD}on minimaalse DFA $\mathcal{D}^{RD}$ olekute hulk. Q^{RD} koosneb determiniseerimisel automaadi $\mathcal{D}^R$ olekute hulgast Q^R moodustatud alamhulkadest.

Kuna kõik automaadi D^R olekud on algolekud, siis $\mathcal{D}^{RD}$ algoleku s_1 leidmiseks moodustame alamhulga kõigist NFA $\mathcal{D}^R$ algolekutest $s_1 = \{q_1, q_2, \dots, q_n\}$, seega on s_1 kõikide automaadi $\mathcal{D}^R$ saavutatavate olekute hulk.

Kuna automaadi $\mathcal{D}^R$ ainus lõppolek q_1 on ka algolek, siis kuulub ta samuti hulka s_1 ja seega on s_1 alati lisaks algolekule ka lõppolek. □

Lause 7. *Kõigil sufiks-kinnises keeles kasutatud sümbolitel on siire minimaalse DFA algolekust mittetühja olekusse.*

Tõestus

Vaatame eelmise tõestuse determiniseerimise protseduuri $\mathcal{D}^R \xrightarrow{det} \mathcal{D}^{RD}$, kus $\mathcal{D}^R = \left(Q^R, \Sigma, \delta^R, I^R, F^R\right)$ on keele $\mathcal{L}_{\text{SC}}$ tunnus-NFA, $\mathcal{D}^{RD} = (Q^{RD}, \Sigma, \delta^{RD}, s_1, F^{RD})$ on keele $\mathcal{L}_{\mathrm{SC}}$ minimaalne DFA ja DFA $\mathcal{D}^{RD}$ algolek $s_1 = \{q_1, q_2, \dots, q_n\}$ on kõigi automaadi $\mathcal{D}^R$ saavutatavate olekute hulk.

Kui keel $\mathcal{L}_{\mathrm{SC}}$ kasutab sümboleid $X = \{x_1, \dots, x_k\}$, kus $X \subseteq \Sigma$ ja $k \geq 1$, siis igal keeles kasutatud sümbolil x_l on DFA's $\mathcal{D}^{RD}$ igast olekust $s_l \in Q^{RD}$ siirdefunktsioon $\delta(s_l, x_i) = \bigcup_{q_i \in s_l} \delta(q_i, x_i)$.

Oletame, et leidub sümbol $x \in \Sigma$, millel on siirded $\delta(s_1, x) = s_\emptyset$ ja $\delta(s_i, x) = s_j$, kus $s_\emptyset$ on minimaalse DFA $\mathcal{D}^{RD}$ tühi olek ja $s_\mathrm{i}, s_\mathrm{j} \in Q^{RD}$ on mittetühjad olekud.

Kuna NFA $\mathcal{D}^R$ ei sisalda tühja olekut, peab sellisel juhul $\bigcup_{q_i \in s_1} \delta(q_i, x) = \emptyset$. Samas $\delta(s_i, x) = \bigcup_{q_i \in s_i} \delta(q_i, x) = s_j$ ja $\bigcup_{q_i \in s_i} \delta(q_i, x) \neq \emptyset$

Kuna kõik minimaalse DFA $\mathcal{D}^{RD}$ mittetühjad olekud on tunnus-NFA olekute alamhulgad, siis $s_1, s_\mathrm{i} \subseteq Q^R$ ja s_1 on automaadi $\mathcal{D}^R$ kõigi saavutatavate olekute hulk $s_1 = \{q_1, q_2, \dots, q_n\}$, siis peab s_i olema s_1 alamhulk.

Kuna $s_\mathrm{i} \subseteq s_1$, siis ka $\delta(s_i, x) \subseteq \delta(s_1, x)$ ja $\bigcup_{q_i \in s_i} \delta(q_i, x) \leq \bigcup_{q_i \in s_1} \delta(q_i, x)$. Siin tekib vastuolu. $\bigcup_{q_i \in s_1} \delta(q_i, x)$ ei saa olla tühi hulk, sest tema alamhulk on mittetühi.

Seega, kui sümbolil puudub siire minimaalse DFA $\mathcal{D}^{RD}$ algolekust mittetühja olekusse, siis ei saa tal olla ka siiret DFA $\mathcal{D}^{RD}$ mistahes mittetühjast olekust mittetühja olekusse. □

Lause 8. *Vähemalt ühel sufiks-kinnises keeles kasutatud sümbolitest on siire minimaalse DFA algolekust lõppolekusse.*

Tõestus

Vaatleme eelmise tõestuse determiniseerimise protseduuri $\mathcal{D}^R \xrightarrow{det} \mathcal{D}^{RD}$, kus $\mathcal{D}^R = (Q^R, \Sigma, \delta^R, I^R, F^R)$ on keele $\mathcal{L}_{\mathrm{SC}}$ tunnus-NFA ja $\mathcal{D}^{RD} = (Q^{RD}, \Sigma, \delta^{RD}, s_1, F^{RD})$ on keele $\mathcal{L}_{\mathrm{SC}}$ minimaalne DFA. DFA $\mathcal{D}^{RD}$ algolek $s_1 = \{q_1, q_2, \dots, q_n\}$ on kõigi automaadi $\mathcal{D}^R$ saavutatavate olekute hulk ja $\delta(s_1, x_j) = \bigcup_{q_i \in s_1} \delta(q_i, x_j)$, kus $X = \{x_1, \dots, x_k\}$ on kõigi keeles $\mathcal{L}_{\mathrm{SC}}$ kasutatud sümbolite hulk, $X \subseteq \Sigma$ ja $k \geq 1$.

Kuna keel $\mathcal{L}_{\mathrm{SC}}$ ei ole tühi keel, siis automaadis $\mathcal{D}^R$ peab eksisteerima vähemalt ühel sümbolil siire $q_1 \in \delta(q_i, x_j)$, kus $q_1 \in F^R$. Kuna s_1 on automaadi $\mathcal{D}^R$ kõikide olekute hulk, siis ka olek $q_i \in s_1$.

Tulenevalt eelnevast, peab DFA's $\mathcal{D}^{RD}$ olema siire $\delta(s_1, x_j) = s_i$, kus $q_1 \in s_i$. Kuna q_1 on automaadi $\mathcal{D}^R$ lõppolek, siis ka s_i on automaadi $\mathcal{D}^{RD}$ lõppolek.

Seega keele $\mathcal{L}_{\mathrm{SC}}$ DFA's $\mathcal{D}^{RD}$ on vähemalt ühel sümbolil siire algolekust lõppolekusse. □

Lause 9. *Puhta sufiks-kinnise keele minimaalsel DFA'l on alati vähemalt üks mittetühi olek, mis ei ole ei alg- ega lõppolek.*

Tõestus

Keele definitsioonist tulenevalt ei ole keel faktor-kinnine, kui vähemalt ühe sõna üks faktor ei kuulu sellesse keelde iseseisva sõnana. Seega puhtas sufiks-kinnises keeles peab olema vähemalt üks faktor, mis ei kuulu sellesse keelde iseseisva sõnana.

Teisalt sisaldab puhas sufiks-kinnine keel kõikide sõnade kõiki sufikseid iseseisvate sõnadena. Sellest tulenevalt on sufiks-kinnise keele sõnade iga faktor ka mõne selle keele sõna prefiks.

Seega puhtas sufiks-kinnises keeles peab olema vähemalt üks prefiks, mis ei kuulu sellesse keelde iseseisva sõnana.

Vaatleme eelmise tõestuses käsitletud determiniseerimise protseduuri $\mathcal{D}^R \xrightarrow{det} \mathcal{D}^{RD}$, kus $\mathcal{D}^R = (Q^R, \Sigma, \delta^R, I^R, F^R)$ on keele $\mathcal{L}_{\mathrm{SC}}$ tunnus-NFA ja $\mathcal{D}^{RD} = (Q^{RD}, \Sigma, \delta^{RD}, s_1, F^{RD})$ keele $\mathcal{L}_{\mathrm{SC}}$ minimaalne DFA.

Selleks, et keeles oleks sõna $w = uv$, mille prefiks u ei kuulu keelde iseseisva sõnana, peab determiniseerimise tulemi $\mathcal{D}^{RD}$ siirete hulgas olema siirete jada $\delta(s_1, u) = s_j$, kus u on keelde $\mathcal{L}_{\mathrm{SC}}$ kuuluva sõna w prefiks, mis ei kuulu keelde $\mathcal{L}_{\mathrm{SC}}$ iseseisva sõnana. Viimasest tulenevalt ei saa olek s_j olla lõppolek, vastasel juhul aktsepteeriks automaat sõna u keelde $\mathcal{L}_{\mathrm{SC}}$ kuuluvaks.

Kuna u on sõna w prefiks, siis siirete jada $\delta(s_1, u) = s_j$ moodustab osa siirete jadast $\delta(s_1, w) \subseteq s_f$. Seega olek s_j ei saa olla tühi olek.

Eelnevalt tõestasime, et sufiks-kinnise keele minimaalse DFA algolek on alati ka lõppolek (Lause 6). Seega, kui olek s_j ei ole automaadi $\mathcal{D}^{RD}$ lõppolek, siis ei saa ta olla ka selle automaadi algolek.

Lähtuvalt eeltoodust on puhta sufiks-kinnise keele minimaalsel DFA'l alati mittetühi olek, mis ei ole alg- ega ka lõppolek. □

Järeldused lausetest Lause 6 - Lause 9

Eeltoodust saame välja tuua sufiks-kinnise keele minimaalse DFA järgmised omadused:

- Algolek on alati ka lõppolek;
- Kõigil kasutatavatel sümbolitel on siire algolekust mittetühja olekusse;
- Vähemalt ühel sümbolil on siire algolekust lõppolekusse;
- Puhta sufiks-kinnise keele minimaalsel DFA'l leidub vähemalt üks mittetühi olek, mis ei ole alg- ega lõppolek.

4.5.2 Sufiks-kinnise keele vasakfaktorite omadused

Lisaks eeltoodule leiame sufiks-kinnise keele keelereeglitest sufiks-kinnise keele ja selle vasakfaktorite omadusi.

Lause 10. *Iga puhas sufiks-kinnine keel $\mathcal{L}_{SC}$ sisaldab vähemalt ühte ühetähelist sõna.*

Tõestus

Sufiks-kinnise keele keelereeglitest tulenevalt peavad iga sufiks-kinnisesse keelde kuuluva sõna kõik võimalikud sufiksid kuuluma sellesse keelde. Seega kui n-täheline sõna $w = x_1x_2 \ldots x_{n-1}x_n$, kuulub sufiks-kinnisesse keelde $\mathcal{L}_{\mathrm{SC}}$, siis selleks peavad olema selles keeles sõnad $x_2 \ldots x_{n-1}x_n, \ldots, x_{n-1}x_n, x_n$ ja tühi sõna ε. Loendist on näha, et koos n pikkusega sõnaga peab olema keeles n sõna, mis on algse sõna sufiksid ja millede pikkused on vastavalt $n-1$ kuni nullini. Peatüki 4.3 näitest nägime, et minimaalne puhas sufiks-kinnine keel sisaldab lisaks tühjale sõnale vähemalt kahte sõna, millest üks on ühetäheline sõna. Seega iga puhas sufiks-kinnine keel sisaldab endas vähemalt ühte ühetähelist sõna. Sama tulemuse saame järeldada ka varem tõestatud minimaalse DFA omadusest (Lause 8), et minimaalses DFA's on vähemalt ühel kasutatud sümbolil siire algolekust lõppolekusse. □

Lause 11. *Sufiks-kinnise keele $\mathcal{L}_{SC}$ iga vasakfaktor on selle keele sõnade alamhulk.*

Tõestus

On teada, et keele $\mathcal{L}_{\mathrm{SC}}$ vasakfaktori sõnale w leiame, kui eemaldame kõigist keelde $\mathcal{L}_{\mathrm{SC}}$ kuuluvatest sõnadest prefiksi w ehk siis $w^{-1}\mathcal{L}_{\mathrm{SC}} = \{x \in \Sigma^* | wx \in \mathcal{L}_{\mathrm{SC}}\}$.

Teisalt keel $\mathcal{L}_{\mathrm{SC}}$ on sufiks-kinnine keel üksnes siis, kui sõna kuulumisest keelde $\mathcal{L}_{\mathrm{SC}}$ järeldub, et ka kõik selle sõna sufiksid kuuluvad keelde $\mathcal{L}_{\mathrm{SC}}$. Seega kui $wx \in \mathcal{L}_{\mathrm{SC}}$, siis ka $x \in \mathcal{L}_{\mathrm{SC}}$.

Eeltoodust tulenevalt on selge, et iga sõna x, mis kuulub keele $\mathcal{L}_{\mathrm{SC}}$ vasakfaktorisse $\mathcal{K}_i$, peab kuuluma ka keelde $\mathcal{L}_{\mathrm{SC}}$. Seega iga vasakfaktor on algse keele sõnade alamhulk ehk $\mathcal{K}_i \subseteq \mathcal{L}_{\mathrm{SC}}$. □

Lähtudes eeltoodust ja peatükis 4.5.1 leitud minimaalse DFA üldistatud kirjeldusest saame järeldada järgmisi sufiks-kinnise keele vasakfaktorite omadusi:

1. Sufiks-kinnise keele kõik vasakfaktorid on selle keele sõnade alamhulgad. Seega kõik vasakfaktoritesse kuuluvad sõnad peavad kuuluma ka algsesse keelde.

2. Kuna puhta sufiks-kinnise keele minimaalses DFA's on vähemalt üks mittetühi olek, mis ei ole alg- ega lõppolek (Lause 9), siis vähemalt üks puhta sufiks-kinnise keele vasakfaktor ei sisalda endas tühja sõna (Lisa 2).

3. Kuna sufiks-kinnise keele minimaalses DFA's on vähemalt ühel sümbolil siire algolekust lõppolekusse, siis vähemalt üks vasakfaktor sisaldab endas ühetähelist ja tühja sõna.

4. Kuna kõik sufiks-kinnise keele vasakfaktorid on algse keele alamhulgad, siis nende vasakfaktorite ühend on võrdne algse keelega.

5. Kuna puhtal sufiks-kinnisel keelel on vähemalt üks vasakfaktor, mis ei sisalda tühja sõna, siis vasakfaktorite ühisosa võib olla tühi hulk (Lisa 2).

4.6 Faktor-kinnise keele vasakfaktorid

Faktor-kinnised keeled on samaaegselt ka prefiks- ja sufiks-kinnised keeled. Seega on faktor-kinniste keelte minimaalsel DFA'l ja vasakfaktoritel ühiseid omadusi nii prefiks- kui ka sufiks-kinniste keelte minimaalse DFA ja vasakfaktoritega.

4.6.1 Faktor-kinnise keele minimaalne DFA

On teada, et iga faktor-kinnise keele jaoks leidub NFA, mille kõik olekud on korraga nii alg- kui lõppolekud [31].

Peatükis 4.4.1 tõestasime, et kuna prefiks-kinnise keele tunnus-NFA kõik olekud on lõppolekud, siis ka prefiks-kinnise keele minimaalse DFA kõik mittetühjad olekud on lõppolekud (Lause 2). Kuna tõestus põhineb vaid lõppoleku omadusele, siis tulemust ei mõjuta asjaolu, et kõik NFA olekud on samal ajal ka algolekud.

Ka faktor-kinnise keele tunnus-NFA kõik olekud on lõppolekud, seega tulenevalt eeltoodust saame järeldada, et kõik faktor-kinnise keele minimaalse DFA mittetühjad olekud on lõppolekud.

Samas on faktor-kinnise keele tunnus-NFA kõik olekud algolekud nagu on seda ka sufiks-kinnise keele tunnus-NFA'l.

Peatükis 4.5.1 tõestasime, et kuna sufiks-kinnise keele tunnus-NFA kõik olekud on algolekud, siis sufiks-kinnise keele minimaalses DFA algolek on alati ka lõppolek (Lause 6) ja kõigil kasutatud sümbolitel siire sellest algolekust mittetühja olekusse (Lause 7).

Kuna ka faktor-kinnise keele tunnus-NFA's on kõik olekud on algolekud, siis kehtib sufiks-kinnise keele minimaalse DFA algoleku kohta tõestatu ka faktor-kinnise keele minimaalses DFA kohta.

Järeldused:

Faktor-kinnise keele minimaalne DFA on DFA, kus

- algolek on alati ka lõppolek;
- kõik mittetühjad olekud on lõppolekud;
- kõigil kasutatavatel sümbolitel on siire algolekust mittetühja olekusse.

Seega faktor-kinnise keele minimaalse DFA kirjeldus sisaldab endas nii prefiks-kinnise keele minimaalse DFA kui ka sufiks-kinnise keele minimaalse DFA omadusi.

4.6.2 Faktor-kinnise keele vasakfaktorite omadused

Eelmises peatükis nägime, et faktor-kinnise keele minimaalsel DFA'l on sarnaseid jooni nii prefiks-kinnise keele minimaalsele DFA kui ka sufiks-kinnise keele minimaalse DFA'ga. Seega peavad minimaalse DFA nende omaduste kaudu leitud prefiks- ja sufiks-kinniste keelte vasakfaktorite omadused kehtima ka faktor-kinniste keelte puhul.

Sama järelduse saame teha teadmisest, et iga faktor-kinnine keel on samaaegselt ka prefiks- ja sufiks-kinnine keel.

Lause 12. *Faktor-kinnisel keelel võib leiduda vasakfaktor, mis ei ole faktor-kinnine keel.*

Tõestus

Kuna minimaalse DFA algolekule vastav vasakfaktor on võrdne keele endaga, siis faktor-kinnisel keelel on alati vähemalt üks faktor-kinnine vasakfaktor.

Varasemast on selge, et faktor-kinnine keel on alati ka prefiks-kinnine ja sufiks-kinnine. Seega saab eeldada, et ta kannab endas nii prefiks- kui sufiks-kinniste keelte omadusi.

Prefiks-kinnise keele vasakfaktorid on prefiks-kinnised keeled (peatükk 4.4.2). Sufiks-kinnise keele vasakfaktorid on algse keele sõnade alamhulgad (Lause 11). Seega võib eeldada, et faktor-kinnise keele vasakfaktorid on prefiks-kinnised keeled, kuid ei ole alati faktor-kinnised.

Tõestame väite näite abil, leides faktor-kinnise keele, mille vähemalt üks vasakfaktor ei ole faktor-kinnine keel.

Vaatame keelt $\mathcal{L}_{\mathrm{FC}} = \{\varepsilon, a, b, \ ab, bb, ba, \ bba, abb, abba\}$, selle keele vasakfaktorid on järgmised:

$K_1 = \varepsilon^{-1}\mathcal{L}_{\mathrm{FC}} = \{\varepsilon, a, b, \ ab, bb, ba, \ bba, abb, abba\} = \mathcal{L}_{\mathrm{FC}}$

$K_2 = a^{-1}\mathcal{L}_{\mathrm{FC}} = \{\varepsilon, b, bb, bba\}$ $\qquad$ $K_3 = (ab)^{-1}\mathcal{L}_{\mathrm{FC}} = \{\varepsilon, b, ba\}$

$K_4 = (abb)^{-1}\mathcal{L}_{\mathrm{FC}} = (bb)^{-1}\mathcal{L}_{\mathrm{FC}} = \{\varepsilon, a\}$ $\qquad$ $K_5 = b^{-1}\mathcal{L}_{\mathrm{FC}} = \{\varepsilon, a, b, ba\}$

$K_6 = (ba)^{-1}\mathcal{L}_{\text{FC}} = (bba)^{-1}\mathcal{L}_{\text{FC}} = \{\varepsilon\}$

On selgelt näha, et vasakfaktorid K_2 ja K_3 ei ole faktor-kinnised keeled. Vasakfaktoris K_2 puuduvad sõna bba faktorid ba ja a. Vasakfaktoris K_3 puudub sõna ba faktor a. Seega vasakfaktorid K_2 ja K_3 on küll prefiks-kinnised keeled kuid ei ole faktor-kinnised. □

Prefiks-kinniste keelte vasakfaktorite omadustest tulenevad faktor-kinniste keelte vasakfaktorite omadused:

1. Kuna faktor-kinnise keele kõigil vasakfaktoritel on NFA, mille kõik olekud on lõppolekud, siis faktor-kinnise keele kõik vasakfaktorid on prefiks-kinnised keeled.

2. Kuna faktor-kinnise keele minimaalse DFA kõik mittetühjad olekud on lõppolekud ja on teada, et lõppoleku paremkeel sisaldab tühja sõna, siis kõik faktor-kinnise keele mittetühjad vasakfaktorid sisaldavad endas tühja sõna.

3. Kuna faktor-kinnise keele kõik mittetühjad vasakfaktorid on prefiks-kinnised keeled ja neisse kuulub tühi sõna ning prefiks-kinnise keele kõigi sõnade kõik prefiksid peavad kuuluma sellesse keelde, siis sõnade arv faktor-kinnise keele igas mittetühjas vasakfaktoris on suurem kui temasse kuuluva pikima sõna pikkus.

4. Kuna faktor-kinnise keele minimaalse DFA kõik mittetühjad olekud on lõppolekud, siis ka igast selle DFA olekust mittetühja olekusse suunduv siire jõuab lõppolekusse, seega faktor-kinnise keele iga vähemalt kahte sõna sisaldav vasakfaktor sisaldab vähemalt ühte ühetähelist sõna ja tühja sõna.

5. Eelmisest tuleneb ka, et kui tähestiku mõni sümbol esineb faktor-kinnises keeles, siis peab ta sisalduma vähemalt ühes vasakfaktoris ühetähelise sõnana ja vastupidi – kui tähestiku sümbol ei sisaldu ühetähelise sõnana üheski faktor-kinnise keele vasakfaktoris, siis ei kasuta see keel seda tähestiku sümbolit.

Faktor-kinnine keel on samaaegselt ka sufiks-kinnine keel ja faktor-kinnise keele minimaalsel DFA'l on sarnaseid omadusi ka sufiks-kinnise keele minimaalse DFA'ga. Vaatleme järgmiseks sufiks-kinnise keele vasakfaktorite omadusi faktor-kinniste keelte vasakfaktoritel.

Vastav võrdlus on alltoodud tabelis.

Sufiks-kinnise keele vasakfaktori omadus	**Prefiks-kinnise keele vasakfaktori omadus**	**Faktor-kinnise keele vasakfaktori omadus**
Vähemalt üks vasakfaktor sisaldab endas tühja sõna.	Kõik mittetühjad vasakfaktorid sisaldavad endas tühja sõna.	Kõik mittetühjad vasakfaktorid sisaldavad endas tühja sõna.
Vähemalt üks vasakfaktor sisaldab endas ühetähelist ja tühja sõna.	Iga vähemalt kahte sõna sisaldav vasakfaktor sisaldab vähemalt ühte ühetähelist sõna ja tühja sõna.	Iga vähemalt kahte sõna sisaldav vasakfaktor sisaldab vähemalt ühte ühetähelist sõna ja tühja sõna.
Iga mittetühi vasakfaktor on algse keele sõnade alamhulk		Iga mittetühi vasakfaktor on algse keele sõnade alamhulk
Puhta sufiks-kinnise keele vähemalt üks vasakfaktor ei sisalda endas tühja sõna.	Kõik prefiks-kinnise keele mittetühjad vasakfaktorid sisaldavad endas tühja sõna.	Kõik prefiks-kinnise keele mittetühjad vasakfaktorid sisaldavad endas tühja sõna.

Tabel 1 Faktor- ja sufiks-kinniste keelte vasakfaktorite omadused.

Kaks esimest tabelis toodud sufiks-kinnise keele vasakfaktorite omadust on üldisemad variandid eespool prefiks-kinnise keele vasakfaktorite omaduste põhjal sõnastatud faktor-kinnise keele omadustest 2 ja 4. Kuna viimati mainitud omadused on piiravamad ja täpsemad, siis me tabelis toodud kahest esimesest sufiks-kinnise keele vasakfaktorite omadusest uusi faktor-kinnise keele vasakfaktorite omadusi ei leia.

Kolmas omadus põhineb sufiks-kinnise keele keelereeglile. Kuna faktor-kinnine keel on samaaegselt ka sufiks-kinnine keel, siis on see omadus ka faktor-kinniste keelte vasakfaktoritel.

Tabelis toodud viimane omadus kehtib ainult puhta sufiks-kinnise keele puhul. Faktor-kinnise keele vasakfaktoritele see omadus laieneda ei saa, kuna see on vastuolus prefiks-kinnise keel vasakfaktorite omadusega, et kõik prefiks-kinnise keele vasakfaktorid sisaldavad tühja sõna.

Seega saame sufiks-kinniste keelte vasakfaktorite omadustest tulenevalt lisada ühe faktor-kinnise keele vasakfaktorite omaduse:

6. Faktor-kinnise keele iga vasakfaktor on algse keele sõnade alamhulk (Lause 11).

Lõpetuseks omadused faktor-kinniste keelte vasakfaktorite ühisosa ja ühendi kohta:

7. Kuna kõik faktor-kinnise keele vasakfaktorid on prefiks-kinnised keeled ja samal ajal ka algse keele sõnade alamhulgad, siis faktor-kinnise keele vasakfaktorite ühend on faktor-kinnine keel ja võrdne algse keelega.

8. Kuna faktor-kinnise keele vasakfaktorid on prefiks-kinnised keeled ja prefiks-kinniste keelte ühisosa on prefiks-kinnine keel [29], siis faktor-kinnise keele vasakfaktorite ühisosa on prefiks-kinnine keel, mis sisaldab vähemalt tühja sõna. Seega faktor-kinnise keele vasakfaktorite ühisosa ei saa olla tühi hulk.

4.7 Prefiks-kinnise keele aatomid

Keele aatomite üldistatud omaduste leidmiseks on meil kaks teed:

- keele atomaadi ja tema olekute paremkeelte analüüs;
- keele vasakfaktorite ja nende täiendite kombinatsioonide ühisosa analüüs.

Vaatame mõlemaid võimalusi.

Nii nagu vasakfaktorite omaduste otsinguid alustasime minimaalse DFA analüüsist, nii alustame ka aatomite omaduste leidmist keele atomaadi kirjeldusest, kuna on teada, et atomaadi olekute paremkeeled on võrdsed keele aatomitega [20].

4.7.1 Prefiks-kinnise keele atomaat

On teada, et atomaat on keele pöördkeele minimaalse DFA pööratud automaat (NFA) [20].

Varasemalt tõestasime, et prefiks-kinnise keele pöördkeel on sufiks-kinnine keel (Lause 5). Seega prefiks-kinnise keele atomaat on sufiks-kinnise keele minimaalse DFA pöördautomaat.

Sufiks-kinnise keele minimaalse DFA omadustest ja automaadi pööramise protseduuri reeglitest lähtudes, leiame prefiks-kinnise keele atomaadi eriomadused.

Lause 13. *Prefiks kinnise keele atomaadil on alati olek, mis on nii alg- kui lõppolek.*

Tõestus

Vaatleme prefiks-kinnist keelt $\mathcal{L}_{\mathrm{PC}}$ ja sufiks-kinnist keelt $\mathcal{L}_{\mathrm{SC}}$, kus $\mathcal{L}_{\mathrm{PC}} = \mathcal{L}_{SC}^{R}$. Seega keele $\mathcal{L}_{\mathrm{PC}}$ atomaadi saame leida keele $\mathcal{L}_{\mathrm{SC}}$ minimaalse DFA pööramisel $\mathcal{D}_{SC}^{DM} \xrightarrow{rev} \mathcal{A}_{PC}$, kus $\mathcal{D}_{SC}^{DM} = (Q_{SC}^{DM}, \Sigma, \delta_{SC}^{DM}, \mathrm{s}_1^{DM}, F_{SC}^{DM})$ on keele $\mathcal{L}_{\mathrm{SC}}$ minimaalne DFA ja $\mathcal{A}_{PC} = (Q_{PC}^{A}, \Sigma, \delta_{PC}^{A}, \mathrm{I}_{PC}^{A}, F_{PC}^{A})$ on keele $\mathcal{L}_{\mathrm{PC}}$ atomaat.

Peatükis 4.5.1 vaatlesime sufiks-kinnise keele minimaalset DFA'd ja tõestasime muu hulgas, et tema algolek on ka lõppolek (Lause 6).

Seega $\mathcal{D}_{SC}^{DM}$ algolek s_1^{DM} on alati ka lõppolek. Automaadi pööramisel säilitab olek oma seisundi, jäädes edasi alg- ja lõppolekuks. □

Lause 14. *Kõigil prefiks-kinnises keeles kasutatud sümbolitel on siire atomaadi saavutatavast olekust lõppolekusse.*

Tõestus

Vaatleme uuesti atomaadi leidmise protseduuri $\mathcal{D}_{SC}^{DM} \xrightarrow{rev} \mathcal{A}_{PC}$, kus $\mathcal{D}_{SC}^{DM} = (Q_{SC}^{DM}, \Sigma, \delta_{SC}^{DM}, \mathrm{s}_1^{DM}, F_{SC}^{DM})$ on keele $\mathcal{L}_{\mathrm{SC}}$ minimaalne DFA, $\mathcal{A}_{PC} = (Q_{PC}^{A}, \Sigma, \delta_{PC}^{A}, \mathrm{I}_{PC}^{A}, F_{PC}^{A})$ on keele $\mathcal{L}_{\mathrm{PC}}$ atomaat ja $\mathcal{L}_{SC} = \mathcal{L}_{PC}^{R}$.

Tõestasime, et kõigil sufiks-kinnises keeles kasutatud sümbolitel on siire minimaalse DFA algolekust mittetühja olekusse (Lause 7).

Kui $X = \{x_1, \dots, x_n\}$ on keeles $\mathcal{L}_{SC} = \mathcal{L}_{PC}^{R}$ kasutatavate sümbolite hulk ja $\mathcal{X} \subseteq \Sigma$, siis tulenevalt eeltoodust on iga $x_i \in X$ jaoks minimaalses DFA's $\mathcal{D}_{SC}^{DM}$ siire $\delta^{DM}(\mathrm{s}_1^{DM}, x_i) = s_k^{DM}$, kus s_1^{DM} ja s_k^{DM} on vastavalt minimaalse DFA $\mathcal{D}_{SC}^{DM}$ algolek ja mittetühi olek.

Automaadi $\mathcal{D}_{SC}^{DM}$ pööramisel saame olekust s_1^{DM} atomaadi $\mathcal{A}_{PC}$ lõppoleku ehk $\mathrm{s}_1^{A} \in F_{PC}^{A}$. Olekust s_1^{DM} väljuvatest siiretest $\delta^{DM}(\mathrm{s}_1^{DM}, x_i) = s_k^{DM}$ saame pööramisel iga $x_i \in X$ siirde $s_1^{A} \in \delta^{A}(s_k^{A}, x_i)$.

Kuna s_k^{DM} on alati mittetühi olek ja pööramisel muutuvad mittetühjad olekud saavutatavateks olekuteks, siis prefiks-kinnise keele atomaadis on igal selles keeles esineval sümbolil siire saavutatavast olekust lõppolekusse. □

Lause 15. *Vähemalt ühel prefiks-kinnises keeles kasutatud sümbolil on siire atomaadi algolekust lõppolekusse.*

Tõestus

Vaatleme veelkord atomaadi leidmise protseduuri $\mathcal{D}_{SC}^{DM} \xrightarrow{rev} \mathcal{A}_{PC}$, kus $\mathcal{D}_{SC}^{DM} = (Q_{SC}^{DM}, \Sigma, \delta_{SC}^{DM}, \mathrm{s}_1^{DM}, F_{SC}^{DM})$ on keele $\mathcal{L}_{\mathrm{SC}}$ minimaalne DFA, $\mathcal{A}_{PC} = (Q_{PC}^{A}, \Sigma, \delta_{PC}^{A}, \mathrm{I}_{PC}^{A}, F_{PC}^{A})$ on keele $\mathcal{L}_{\mathrm{PC}}$ atomaat ja $\mathcal{L}_{SC} = \mathcal{L}_{PC}^{R}$ ning $X = \{x_1, \dots, x_n\}$ on keeles $\mathcal{L}_{SC} = \mathcal{L}_{PC}^{R}$ kasutatavate sümbolite hulk ja $\mathcal{X} \subseteq \Sigma$.

Tõestasime varasemalt, et vähemalt ühel sufiks-kinnises keeles kasutatud sümbolitest on siire minimaalse DFA algolekust lõppolekusse (Lause 8). Seega minimaalsel DFA'l $\mathcal{D}_{SC}^{DM}$ on vähemalt üks siire $\delta_{SC}^{DM}(\mathrm{s}_1^{DM}, x_j) = \mathrm{s}_i^{DM}$, kus $\mathrm{s}_i^{DM} \in F_{SC}^{DM}$.

Pöördautomaadi leidmise reeglite kohaselt muutuvad automaadi pööramisel lõppolekud algolekuteks ja vastupidi. Seega prefiks-kinnise keele atomaadi algolekute hulk I_{PC}^{A} on võrdne tema pöördkeele minimaalse DFA $\mathcal{D}_{SC}^{DM}$ lõppolekute hulgaga ja vastupidi ehk $I_{PC}^{A} = F_{SC}^{DM}$ ja $F_{PC}^{A} = s_1^{DM}$.

Lähtuvalt eeltoodust tekib pööramise tulemusena atomaadis $\mathcal{A}_{PC}$ siire $\mathrm{s}_1^{A} \in \delta_{PC}^{A}(\mathrm{s}_i^{A}, x_i)$, kus $\mathrm{s}_i^{A} \in I_{PC}^{A}$ ja $\mathrm{s}_1^{A} \in F_{PC}^{A}$.

Seega on prefiks-kinnise keele atomaadis $\mathcal{A}_{PC}$ vähemalt ühel olekul siire algolekust lõppolekusse. □

Lause 16. *Puhta prefiks-kinnise keele atomaadil on alati vähemalt üks saavutatav olek, mis ei ole ei alg- ega lõppolek.*

Tõestus

Kasutame atomaadi leidmise protseduuri $\mathcal{D}_{SC}^{DM} \xrightarrow{rev} \mathcal{A}_{PC}$, kus $\mathcal{D}_{SC}^{DM} = (Q_{SC}^{DM}, \Sigma, \delta_{SC}^{DM}, \mathrm{s}_1^{DM}, F_{SC}^{DM})$ on keele $\mathcal{L}_{\mathrm{SC}}$ minimaalne DFA, $\mathcal{A}_{PC} = (Q_{PC}^{A}, \Sigma, \delta_{PC}^{A}, \mathrm{I}_{PC}^{A}, F_{PC}^{A})$ on keele $\mathcal{L}_{\mathrm{PC}}$ atomaat ja $\mathcal{L}_{SC} = \mathcal{L}_{PC}^{R}$.

Varasemalt tõestasime, et puhta sufiks-kinnise keele minimaalsel DFA'l on vähemalt üks mittetühi olek, mis ei ole ei alg- ega lõppolek (Lause 9).

Seega peab olema olek $s_i^{DM} \in Q_{SC}^{DM}$, mis on saavutatav ja mittetühi ning samas ei ole alg- ega lõppolek. Sellise olekuga peab olema seotud vähemalt kaks siiret : $\delta^{DM}(s_h^{DM}, x_i) = s_i^{DM}$ ja $\delta^{DM}(s_i^{DM}, x_j) = s_j^{DM}$, kus $x_i, x_j \in \Sigma$ ja $s_h^{DM}, s_h^{DM} \in Q_{SC}^{DM}$.

Automaadi pööramisel muutuvad olekutevahelised siirded vastassuunalisteks. Seega pööramisel saame atomaadis siirded $s_h^A \in \delta^A(s_i^A, x_i)$ ja $s_i^A \in \delta^A(s_j^A, x_j)$. Siit näeme, et mittetühi olek ei saa minimaalse DFA pööramisel muutuda tühjaks või saavutamatuks olekuks. Samuti on pööramise reeglitest tulenevalt selge, et kui olek ei olnud enne pööramist ei alg- ega ka lõppolek, siis ei ole ta ka pärast pööramist alg- või lõppolek.

Seega puhta prefiks-kinnise keele atomaadil on alati vähemalt üks saavutatav olek, mis ei ole alg- ega lõppolek. □

Järeldus

Seega prefiks-kinnise keele atomaat on NFA, kus

- üks algolek on samal ajal ka lõppolek;
- kõigil kasutatavatel sümbolitel on siire saavutatavast olekust lõppolekusse;
- vähemalt ühel kasutatud sümbolitest on siire algolekust lõppolekusse.

Puhta sufiks-kinnise keele atomaadil on lisaks eeltoodule alati vähemalt üks saavutatav olek, mis ei ole ei alg- ega lõppolek.

4.7.2 Prefiks-kinnise keele aatomid

On teada, et

- keele $\mathcal{L}$ aatom $A_l = \widetilde{K_1} \cap \widetilde{K_2} \cap \ldots \cap \widetilde{K_n}$, kus $\widetilde{K_l}$ on kas keele $\mathcal{L}$ vasakfaktor K_l või vasakfaktori täiend $\overline{K_l}$ [20];

- aatomid on võrdsed atomaadi olekute paremkeeltega [20];

- iga vasakfaktor on väljendatav aatomite ühendina [20];

Kasutades peatükis 4.4.2 leitud prefiks-kinnise keele omadusi, saame teha mõned järeldused prefiks-kinniste keelte aatomite kohta.

Lause 17. *Prefiks-kinnise keele positiivsete aatomite ühend on faktor-kinnine keel.*

Tõestus

Tõestasime, et prefiks-kinnise keele kõigi vasakfaktorite ühend on faktor-kinnine keel (Lause 4). Kuna vasakfaktorid on kirjeldatavad positiivsete aatomite ühendina, siis vasakfaktori iga sõna peab kuuluma ka mõnda aatomisse.

Kuna aatom on positiivne, kui aatomis väljendatuna $A_i = \widetilde{K_1} \cap \widetilde{K_2} \cap \ldots \cap \widetilde{K_n}$ leidub vähemalt üks $\widetilde{K_l} = K_l$, seega positiivne aatom sisaldab vaid sõnu, mis esinevad vähemalt ühes vasakfaktoris.

Eeltoodust saame järeldada, et positiivsete aatomite ühend on võrdne vasakfaktorite ühendiga ja on seega samuti faktor-kinnine keel. □

Lause 18. *Prefiks-kinnise keele negatiivne aatom on kahepoolne ideaal.*

Tõestus

Eelnevalt leidsime, et prefiks-kinnise keele positiivsete aatomite ühend $\mathcal{A}^+$, mis sisaldab endas kõiki keele vasakfaktoritesse kuuluvaid sõnu, on faktor-kinnine keel.

Negatiivne aatom $\mathcal{A}^-$ leitakse vasakfaktorite täiendite ühisosana, seega koosneb sõnadest, mis ei kuulu ühtegi vasakfaktorisse.

Positiivsete aatomite ühend $\mathcal{A}^+$ on Σ^* alamhulk, mis sisaldab kõiki vähemalt ühte vasakfaktorisse kuuluvaid sõnu ja negatiivne aatom $\mathcal{A}^-$ on Σ^* alamhulk, mis sisaldab vasakfaktoritesse mittekuuluvaid sõnu. Kuna sõna ei saa korraga hulka kuuluda ja mitte kuuluda, siis on alamhulgad teineteist eitavad.

Kuna alamhulkade tingimused on teineteist eitavad, saame öelda, et $\Sigma^* = \mathcal{A}^+ + \mathcal{A}^-$.

Lähtuvalt täiendi definitsioonist saame öelda, et positiivsete aatomite ühend on negatiivse aatomi täiend $\mathcal{A}^+ = \overline{\mathcal{A}^-}$.

On teada, et faktor-kinnine keel on kahepoolse ideaali täiend. Ehk siis faktor-kinnine keel $\mathcal{L}_{FC} = \overline{\Sigma^*\mathcal{K}\Sigma^*}$, kus $\mathcal{K}$ on regulaarne keel tähestikus Σ [34].

Kuna $\mathcal{A}^+$ on faktor-kinnine keel, siis peab negatiivne aatom olema kahepoolne ideaal. □

Eeltoodust ja atomaadi kirjeldusest saame teha järgmised järeldused prefiks-kinnise keele aatomite kohta:

1. Kuna puhta prefiks-kinnise keele atomaat sisaldab vähemalt ühte olekut, mis ei ole lõppolek (Lause 16), siis leidub puhtal prefiks-kinnisel keelel aatom, mis ei ole prefiks-kinnine keel.
2. Prefiks-kinnise keele positiivsete aatomite ühend on faktor-kinnine keel (Lause 17).
3. Prefiks-kinnise keele negatiivne aatom on kahepoolne ideaal (Lause 18).
4. Kuna igal prefiks-kinnises keeles kasutatud sümbolil on atomaadis siire saavutatavast olekust lõppolekusse ja aatom on atomaadi oleku paremkeel, siis järelikult peab iga selline sümbol kuuluma ka ühte aatomisse ühetähelise sõnana. Sama omadus on järeldatav ka prefiks-kinnise keele vasakfaktorite omadustest.
5. Eelmisest tulenevalt kehtib ka vastupidine seos. Kui tähestiku sümbol ei esine üheski aatomis ühetähelise sõnana, siis ei esine ta ka selles keeles.

4.8 Sufiks-kinnise keele aatomid

Alustame jällegi automaadist, kui praktilisest aatomite leidmise abivahendist.

4.8.1 Sufiks-kinnise keele atomaat

Kuna prefiks- ja sufiks-kinnised keeled on teineteise pöördkeeled, siis sufiks-kinnise keele atomaat on prefiks-kinnise keele minimaalse DFA pöördautomaat.
Sellest tulenevalt saame kirjeldada ka sufiks-kinnise keele atomaadi.

Lause 19. *Sufiks-kinnise keele atomaadi kõik saavutatavad olekud on algolekud.*

Tõestus
Teame, et sufiks-kinnise keele atomaadiks on prefiks-kinnise keele minimaalse DFA pöördautomaat.
Võtame prefiks-kinnise keele $\mathcal{L}_{\mathrm{PC}}$ minimaalse DFA ja leiame sufiks-kinnise keele $\mathcal{L}_{\mathrm{SC}}$ atomaadi $\mathcal{D}_{PC}^{DM} \xrightarrow{rev} \mathcal{A}_{SC}$, kus $\mathcal{D}_{PC}^{DM} = (Q_{PC}^{DM}, \Sigma, \delta_{PC}^{DM}, \mathrm{s}_1^{DM}, F_{PC}^{DM})$ on keele $\mathcal{L}_{\mathrm{PC}}$ minimaalne DFA, $\mathcal{A}_{SC} = \{Q_{SC}^{A}, \Sigma, \delta_{SC}^{A}, \mathrm{I}_{SC}^{A}, F_{SC}^{A}\}$ on keele $\mathcal{L}_{\mathrm{SC}}$ atomaat ja $\mathcal{L}_{\mathrm{PC}} = \mathcal{L}_{SC}^{R}$.

Peatükis 4.4.1 tõestasime, et prefiks-kinnise keele minimaalse DFA kõik mittetühjad olekud on lõppolekud ehk $F_{PC}^{DM} = Q_{PC}^{DM} \setminus s_\emptyset$. Minimaalse DFA pööramisel muutuvad kõik lõppolekud algolekuteks ja vastupidi, seega $F_{PC}^{DM} = \mathrm{I}_{SC}^{A} = Q_{SC}^{A} \setminus s_\emptyset$, kus $s_\emptyset$ on saavutamatu olek mis tekib DFA $\mathcal{D}_{PC}^{DM}$ pööramisel, kui sellel DFA'l on tühi olek $s_\emptyset$.
Seega sufiks-kinnise keele atomaat on NFA kus kõik saavutatavad olekud on algolekud.

□

Järeldus

Teades, et atomaadil on alati üks lõppolek, saame Lausest 19 järeldada, et sufiks-kinnise keele atomaadi kirjeldus vastab sufiks-kinnise keele tunnus-NFA'le ja seega toetab ka sufiks-kinnise keele vasakfaktorite omadusi.

4.8.2 Sufiks-kinnise keele aatomid

Kuna eelmises peatükis tõdesime, et sufiks-kinnise keele atomaat vastab keele tunnus-NFA kirjeldusele, siis saame esitada sufiks-kinnise keele vasakfaktorite kohta tõestatuga sarnase väite:

Lause 20. *Sufiks-kinnise keele atomaadi kõigi saavutatavate olekute paremkeeled on algse keele sõnade alamhulgad.*

Tõestus

On teada, et keele $\mathcal{L}_{\mathrm{SC}}$ atomaadi oleku q_i paremkeel on keele $\mathcal{L}_{\mathrm{SC}}$ sufiksite hulk, mis jõuavad automaadis siirete tulemusel olekust q_i lõppolekusse ehk $L_{q_i,F}(\mathcal{A}_{SC}) = \{w \in \Sigma^* | \delta'(q_i, w) \cap F_{SC}^{A} \neq \emptyset\}$.

Kui $\mathcal{A}_{SC}$ on keelt $\mathcal{L}_{\mathrm{SC}}$ aktsepteeriv atomaat, siis kehtivad seosed $L(\mathcal{A}_{SC}) = \mathcal{L}_{\mathrm{SC}} = \{w \in \Sigma^* \, | \delta'(\mathrm{I}_{SC}^{A}, w) \cap F_{SC}^{A} \neq \emptyset\}$.

Lause 19 tõestuses nägime, et atomaadi algolekute hulk vastab tema saavutatavate olekute hulgale ehk $\mathrm{I}_{SC}^{A} = Q_{SC}^{A} \setminus s_\emptyset$. Seega iga atomaadi saavutatav olek $q_i \in \mathrm{I}_{SC}^{A}$ ja tema paremkeel on $L_{q_i,F}(\mathcal{A}_{SC}) = \{w \in \Sigma^* | \delta'(q_i, w) \cap F_{SC}^{A} \neq \emptyset\} \subseteq \mathcal{L}_{\mathrm{SC}}$. Teiste sõnadega – kuna kõik sufiks-kinnise keele $\mathcal{L}_{\mathrm{SC}}$ atomaadi saavutatavad olekud on algolekud, siis kõikide olekute paremkeeled on keele $\mathcal{L}_{\mathrm{SC}}$ alamhulgad.

Kuna atomaadi $\mathcal{A}_{SC}$ paremkeeled on keele $\mathcal{L}_{\mathrm{SC}}$ aatomid, siis on kõik aatomid keele $\mathcal{L}_{\mathrm{SC}}$ alamhulgad. □

On teada, et sufiks-kinnine keel on vasak-ideaali täiend ehk $\mathcal{L}_{SC} = \overline{\Sigma^*\mathcal{K}}$ [29], siis sufiks-kinnise keele täiend on vasak-ideaal ehk $\overline{\mathcal{L}_{SC}} = \Sigma^* \setminus \mathcal{L}_{SC} = \Sigma^* \setminus \overline{\Sigma^*\mathcal{K}} = \Sigma^*\mathcal{K}$.

Kasutades peatükis 4.5.2 leitud sufiks-kinnise keele vasakfaktorite omadusi ja eelmises peatükis leitud sufiks-kinnise keele atomaadi kirjeldust, saame koos eeltooduga teha mõned järeldused sufiks-kinnise keele aatomite kohta:

1. Kuna sufiks-kinnise keele atomaadi kõigi saavutatavate olekute paremkeeled on algse keele alamhulgad, siis kõik sufiks-kinnise keele positiivsed aatomid on selle keele alamhulgad. Seega kõik positiivsetesse aatomitesse kuuluvad sõnad peavad kuuluma ka algsesse keelde ja kõik positiivsed aatomid on algaatomid.

2. Kuna kõik positiivsetesse aatomitesse kuuluvad sõnad kuuluvad algsesse keelde ja kõik algsesse keelde kuuluvad sõnad peavad kuuluma mõnda positiivsesse aatomisse, siis sufiks-kinnise keele positiivsete aatomite ühend on sufiks-kinnine keel ja võrdne algse keelega.

3. Kuna sufiks-kinnise keele $\mathcal{L}_{\mathrm{SC}}$ positiivsete aatomite ühend on $\mathcal{L}_{\mathrm{SC}}$, siis sufiks-kinnise keele negatiivne aatom ehk vasakfaktorite täiendite ühisosa on keele $\mathcal{L}_{\mathrm{SC}}$ täiend. Seega puhta sufiks-kinnise keele negatiivne aatom on vasak-ideaal.

4.9 Faktor-kinnise keele aatomid

Leiame kõigepealt faktor-kinnise keele atomaadi iseloomulikud omadused.

4.9.1 Faktor-kinnise keele atomaat

Eelnevalt nägime, et lihtne ja praktiline viis keele atomaadi leidmiseks on tema pöördkeele minimaalse DFA pöördautomaadi leidmine.

Lause 21. *Iga faktor-kinnise keele pöördkeel on samuti faktor-kinnine keel.*

Tõestus

Teame, et keele pöördkeelt aktsepteerib algset keelt aktsepteeriva automaadi pöördautomaat.

Leiame faktor-kinnist keelt $\mathcal{L}_{\mathrm{FC}}$ aktsepteeriva NFA pöördautomaadi $\mathcal{N}_{FC} \xrightarrow{rev} \mathcal{N}_{FC}^{R}$, kus $\mathcal{N}_{FC} = (Q_{FC}, \Sigma, \delta_{FC}, I_{FC}, F_{FC})$ on keele $\mathcal{L}_{\mathrm{FC}}$ tunnus-NFA ja $\mathcal{N}_{FC}^{R} = \left(Q_{FC}^{R}, \Sigma, \delta_{FC}^{R}, I_{FC}^{R}, F_{FC}^{R}\right)$ on pöördkeele $\mathcal{L}_{FC}^{R}$ NFA.

Igal faktor-kinnisel keelel leidub NFA, mille kõik olekud on alg- ja lõppolekud [31].

Seega, kui $\mathcal{N}_{FC}$ on faktor-kinnise keele tunnus-NFA, on kõik tema olekud alg- ja lõppolekud, see tähendab $Q_{FC} = I_{FC} = F_{FC}$.

Pööramise protseduuri tulemusena vahetame automaadi alg- ja lõppolekud. Seega $I_{FC} = F_{FC}^{R}$ ja $F_{FC} = I_{FC}^{R}$. Kuna $I_{FC} = F_{FC}$, siis ka $I_{FC}^{R} = F_{FC}^{R}$. Seega kõik pöördautomaadi olekud on nii alg- kui ka lõppolekud.
Kuna faktor-kinnisel keelel leidub NFA, mille pöördautomaadi kõik olekud on nii alg- kui ka lõppolekud, siis faktor-kinnise keele pöördkeel on faktor-kinnine keel. □

Järeldus

Kuna faktor-kinnise keele pöördkeel on samuti faktor-kinnine keel, siis faktor-kinnise keele minimaalse DFA pööramisel leiame teise faktor-kinnise keele atomaadi. Kasutame seda omadust faktor-kinnise keele atomaadi leidmisel.

Järeldused faktor-kinnise keele minimaalse DFA omadustest

Peatükis 4.6.1 leidsime, et faktor-kinnise keele minimaalsel DFA'l on kõik mittetühjad olekud lõppolekud. Peatükis 4.8.1 tõestasime sufiks-kinnise keele atomaadi leidmisel, et kui pöördkeele minimaalse DFA kõik mittetühjad olekud on lõppolekud, siis keele atomaadi kõik saavutatavad olekud on algolekud (Lause 19).

Kuna ka faktor-kinnise keele minimaalse DFA kõik mittetühjad olekud on lõppolekud ja faktor-kinnise keele pöördkeel on samuti faktor-kinnine keel, siis saame järeldada, et faktor-kinnise keele atomaadi kõik saavutatavad olekud on algolekud.

Seega faktor-kinnise keele atomaat on NFA, mis vastab sufiks-kinnise keele tunnus-NFA'le, kus kõik saavutatavad olekud on algolekud ja millel on üks lõppolek.

Peatükis 4.6.1 leidsime, et faktor-kinnise keele minimaalses DFA's on igal keeles kasutatud sümbolil siire algolekust mittetühja olekusse. Peatükis 4.7.1 tõestasime prefiks-kinnise keele atomaadi kohta, et sellise minimaalse DFA pööramise tulemusel saame atomaadi, milles igal keeles kasutatud sümbolil siire saavutatavast olekust lõppolekusse (Lause 14).

Seega saame öelda, et ka faktor-kinnise keele atomaadis on kõigil keeles kasutatavatel sümbolitel siire saavutatavast olekust lõppolekusse.

Järeldused

Eelnevast saame järeldada, et faktor-kinnise keele atomaat on NFA,

- mille kõik saavutatavad olekud on algolekud;
- mille ainus lõppolek on alati ka algolek;
- kõigil kasutatavatel sümbolitel on siire saavutatavast olekust lõppolekusse.

4.9.2 Faktor-kinnise keele aatomid

Kuna faktor-kinnise keele atomaadil on sufiks-kinnise keele tunnus-NFA omadused, siis võib oletada, et ka faktor-kinnise keele aatomid kannavad endas sufiks-kinnise keele aatomite omadusi.

Peatükis 4.8.2 tõestasime, et kui atomaadi kõik olekud on algolekud, siis on aatomid algse keele sõnade alamhulgad (Lause 20). Ka faktor-kinnise keele atomaadi kõik saavutatavad olekud on algolekud, seega ka faktor-kinnise keele aatomid on algse keele sõnade alamhulgad.

Kasutades peatükis 4.6.2 esitatud faktor-kinnise keele vasakfaktorite omadusi ja peatükis 4.9.1 toodud faktor-kinnise keele atomaadi kirjeldust, saame koos eelöelduga järeldada faktor-kinniste keelte aatomite kohta järgmist:

1. Kuna faktor-kinnise keele atomaadi kõik paremkeeled on algse keele alamhulgad, siis kõik faktor-kinnise keele positiivsed aatomid on selle keele alamhulgad. Seega kõik positiivsesse aatomitesse kuuluvad sõnad kuuluvad ka algsesse keelde ja kõik positiivsed aatomid on algaatomid.

2. Sama tulemuse saame ka vasakfaktorite omaduste kaudu. Kuna faktor-kinnise keele iga mittetühi vasakfaktor on selle keele sõnade alamhulk (Lause 11) ja vasakfaktor on moodustatav positiivsete aatomite ühendina, siis iga faktor-kinnise keele positiivne aatom on algse keele sõnade alamhulk.

3. Kuna iga faktor-kinnises keeles kasutatud sümbol peab esinema vähemalt ühes vasakfaktoris ühetähelise sõnana ja kõik vasakfaktorid on positiivsete aatomite ühendid, siis esineb iga kasutatud sümbol ka ühes faktor-kinnise keele positiivses aatomis ühetähelise sõnana.

4. Eelmisest tuleneb ka vastupidine seos. Kui tähestiku sümbol ei sisaldu ühetähelise sõnana üheski faktor-kinnise keele positiivses aatomis, siis ei kasuta keel seda tähestiku sümbolit.

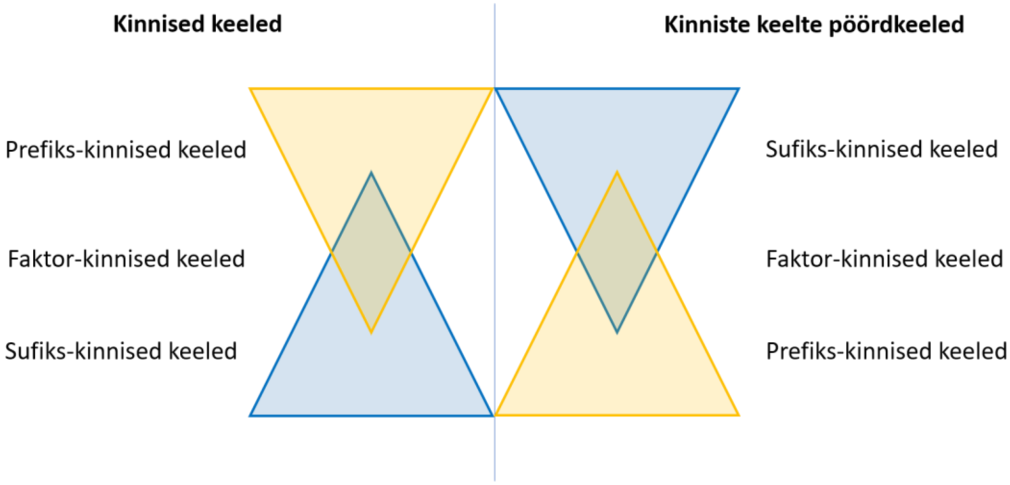
5. Kuna faktor-kinnise keele vasakfaktorite ühend on faktor-kinnine keel ja võrdne algkeelega ning kõik vasakfaktorid on positiivsete aatomite ühendid, on ka faktor-kinnise keele positiivsete aatomite ühend faktor-kinnine keel ja võrdne algse keelega.

6. Kuna faktor-kinnise keele positiivsete aatomite ühend on faktor-kinnine keel, mille täiend on kahepoolne ideaal [34], siis faktor-kinnise keele negatiivne aatom on kahepoolne ideaal.

4.10 X-kinniste keelte vasakfaktorid ja aatomid

Käesolevas töös käsitlesime X-kinniste keelte alamklasse, kus X tähistab kas prefiksit, sufiksit, bifiksit, faktorit või alamsõna.

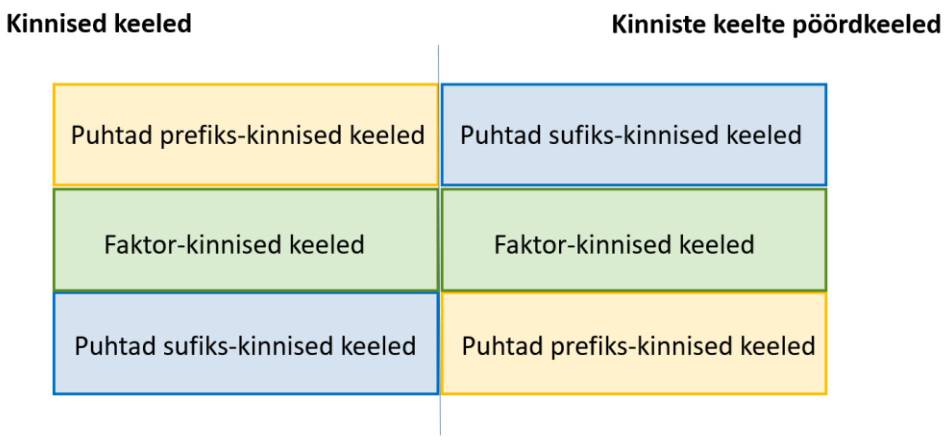
Töös nägime, et käsitletud X-kinnised keeled on omavahel väga tihedalt seotud. Seoste graafiline kujund on esitatud alloleval joonisel.



Joonis 11 Kinniste keelte alamklassid ja pöördkeeled

Kuna prefiks- ja sufiks-kinnised keeled on teineteise pöördkeeled, siis nende keelte ühisosa moodustavate faktor-kinniste keelte pöördkeeled on samuti faktor-kinnised keeled. Samuti on teada, et kõik bifiks- ja alamsõna-kinnised keeled on ka faktor-kinnised keeled. Kusjuures kõik faktor-kinnised keeled on bifiks-kinnised keeled, kuid vaid osa faktor-kinnistest keeltest on alamsõna-kinnised keeled.

Lisaks eelmainitule lisasime töös puhta prefiks- ja sufiks-kinnise keele mõiste.



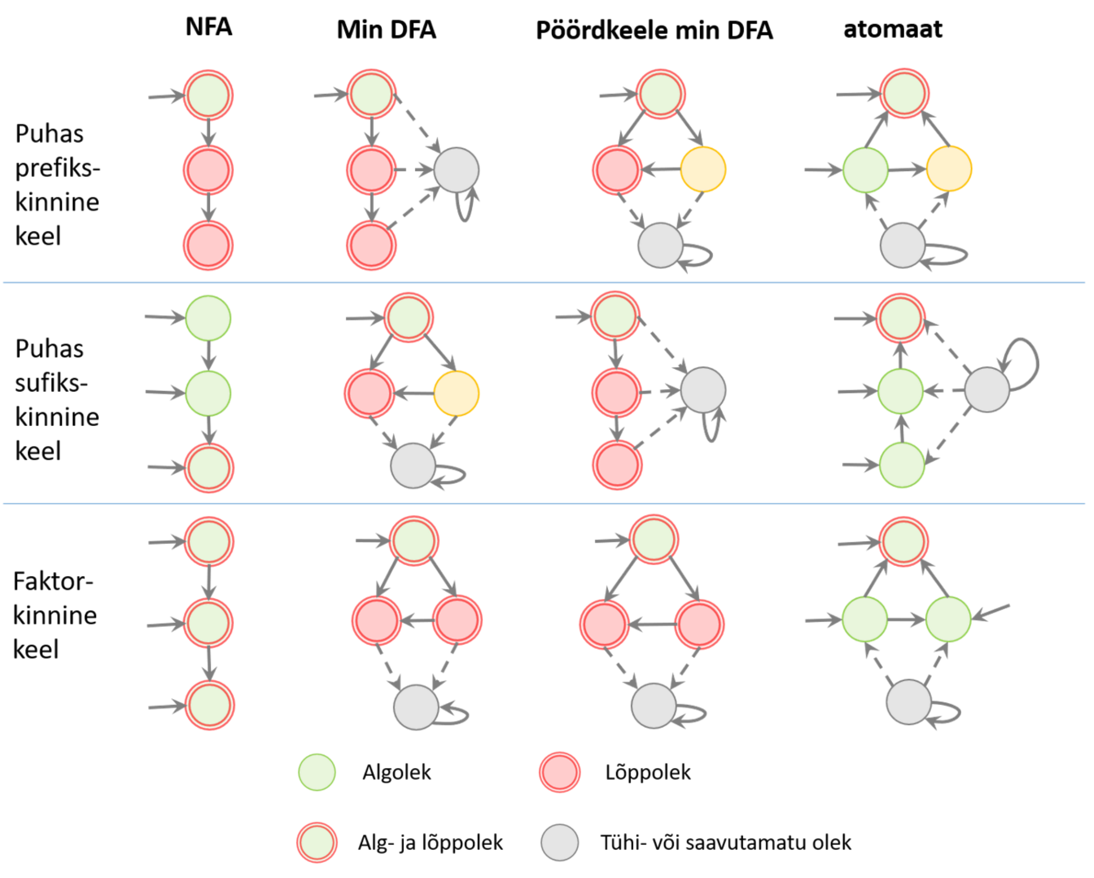
Joonis 12 Puhtad kinnised keeled ja nende pöördkeeled

Erinevalt prefiks- ja sufiks-kinnistest keeltest puudub puhastel prefiks- ja sufiks-kinnistel keeltel ühisosa faktor-kinniste keeltega. Puhastel prefiks- ja sufiks-kinnistel keeltel on omadusi, mis puuduvad faktor-kinnistel keeltel.

Töös lähtusime teadmisest, et kõigil X-kinnistel keeltel leidub tunnus-NFA, mille olekud on kõik kas algolekud, lõppolekud või siis nii alg- kui lõppolekud [31] ja analüüsisime keelte kolme iseloomulikku automaati:

- minimaalne DFA,
- pöördkeele minimaalne DFA,
- atomaat.

Nende automaatide näidis-skeemid on toodud alloleval joonisel.



Joonis 13 X-kinniste keelte automaatide näidis skeemid

Automaatide näidis-skeemidelt näeme, et lähtuvalt tunnus-NFA erinevusest on ka teistel automaatidel igal keele alamklassil talle iseloomulikke jooni.

Ühendavaks jooneks kõigil X-kinniste keelte automaatidel on vähemalt ühe korraga nii alg- kui lõppoleku olemasolu.

4.10.1 X-kinniste keelte vasakfaktorid

Töös kirjeldatud vasakfaktorite omadused saame jagada kolme gruppi:

- Vasakfaktorite keele alamklass;
- Sõnad vasakfaktorites;
- Vasakfaktorite ühend ja ühisosa.

Alltoodud tabelis (Tabel 2) on kokkuvõte töös käsitletud X-kinniste keelte vasakfaktorite kuuluvusest X-kinniste keelte alamklassidesse.

Vasakfaktorite omadus, mis kehtib kõigil alamklassi ja keelegruppi kuuluvatel keeltel.	**Prefiks-kinnine keel**		**Faktor-kinnine keel**	**Sufiks-kinnine keel**	
	puhas	**üldine**		**üldine**	**puhas**
1. Kõik kuuluvad algse keelega samasse keele alamklassi	Jah	Jah	Ei	Ei	Ei
2. On kõik prefiks-kinnised keeled	Jah	Jah	Jah	Ei	Ei

Tabel 2 X-kinniste keelte vasakfaktorite keele alamklassid

Enamiku X-kinniste keelte kõik vasakfaktorid on prefiks-kinnised keeled. Erandiks on puhtad sufiks-kinnised keeled, mille vasakfaktorid ei kuulu kindlasse X-kinniste keelte alamklassi. Seega vaadeldud X-kinniste keelte alamklassidest vaid prefiks-kinnistel keeltel kuuluvad vasakfaktorid algse keelega samasse alamklassi. Puhta prefiks-kinnise keele kõik vasakfaktorid ei kuulu puhaste prefiks-kinniste keelte gruppi, kuid kuuluvad kindlasti prefiks-kinniste keelte alamklassi.

Ülevaate terviklikkuse huvides peab lisama, et ka alamsõna-kinnise keele kõik vasakfaktorid on alamsõna-kinnised keeled, olles samal ajal ka prefiks-, bifiks- ja faktor-kinnised keeled.

Omaduse 2 (Tabel 2) negatiivsete tulemite õigsus on lihtsalt mõistetav Lisas 2 toodud näites.

Järgmiseks koondtabel vasakfaktorite sõnade omadustest. Ka selles osas on töös tõestatud tunnuse esinemine. Kuna negatiivse tulemi kinnituseks piisab ühest kontranäitest, siis kinnitavad negatiivseid tulemeid keelte näited Lisades 1 – 4.

Vasakfaktorite omadus, mis kehtib kõigil alamklassi ja keelegruppi kuuluvatel keeltel.	Prefiks-kinnine keel		Faktor-kinnine keel	Sufiks-kinnine keel	
	puhas	üldine		üldine	puhas
3. On kõik algse keele sõnade alamhulgad	Ei	Ei	Jah	Jah	Jah
4. Sõnade arv mittetühjas vasakfaktoris on suurem kui temasse kuuluva pikima sõna pikkus.	Jah	Jah	Jah	Ei	Ei
5. Kõik mittetühjad vasakfaktorid sisaldavad tühja sõna	Jah	Jah	Jah	Ei	Ei
6. Vähemalt üks vasakfaktor ei sisalda endas tühja sõna	Ei	Ei	Ei	Ei	Jah
7. Iga vähemalt kahte sõna sisaldav vasakfaktor sisaldab vähemalt ühte ühetähelist sõna.	Jah	Jah	Jah	Ei	Ei
8. Vähemalt üks vasakfaktor sisaldab endas mõnda keele poolt kasutatavat sümbolit ühetähelise sõnana ja tühja sõna.	Jah	Jah	Jah	Jah	Jah
9. Iga keele poolt kasutatav sümbol sisaldub vähemalt ühes vasakfaktoris ühetähelise sõnana.	Jah	Jah	Jah	Ei	Ei
10. Vähemalt üks keele poolt kasutatav sümbol sisaldub vähemalt ühes vasakfaktoris ühetähelise sõnana.	Jah	Jah	Jah	Jah	Jah

Tabel 3 Sõnad X-kinniste keelte vasakfaktorites

Vasakfaktorite sõnadest leidsime omadusi kolmest aspektist (Tabel 3):

- Tühja sõna esinemine
- Ühetäheliste sõnade esinemine
- Vasakfaktorite sõnade seos algkeele sõnadega

Kahe esimese omaduste grupi osas on prefiks- ja faktor-kinnised keeled sarnased ja oluliselt erinevad sufiks-kinnise keele vasakfaktorite omadustest.

Kolmanda omaduse osas on aga sarnased sufiks- ja faktor-kinnise keele vasakfaktorid, sisaldades endas ainult algse keele sõnu. Prefiks-kinnise keele vasakfaktorites on lisaks algse keele sõnadele ka nende sõnade sufiksid.

Viimasest tuleneb ka, et lisaks faktor-kinnisele keelele on ka prefiks-kinnise keele vasakfaktorite ühend faktor-kinnine keel (Tabel 4) .

Samas on sarnasus ka sufiks-, bifiks- ja faktor-kinniste keelte vasakfaktorite vahel. Nende keelte vasakfaktorite ühend on võrdne algse keelega.

Vasakfaktorite omadus, mis kehtib kõigil alamklassi või keelegruppi kuuluvatel keeltel.	**Prefiks-kinnine keel**		**Faktor-kinnine keel**	**Sufiks-kinnine keel**	
	puhas	**üldine**		**üldine**	**puhas**
11. Vasakfaktorite ühend on faktor-kinnine keel	Jah	Jah	Jah	Ei	Ei
12. Vasakfaktorite ühend on võrdne algse keelega.	Ei	Ei	Jah	Jah	Jah
13. Vasakfaktorite ühisosa võib olla tühi hulk	Ei	Ei	Ei	Ei	Jah

Tabel 4 X-kinniste keelte vasakfaktorite ühendi ja ühisosa omadused

X-kinniste keelte vasakfaktorite ühisosa puudumine on otseselt seotud tühja sõna esinemisega vasakfaktorites. Kuna igas prefiks- ja faktor-kinnise keele vasakfaktoris on tühi sõna, on nende keelte vasakfaktorite ühisosa vähemalt tühi sõna. Vaadeldud keeltest saab tühi hulk olla ainult puhta sufiks-kinnise keele vasakfaktorite ühisosa. Igal puhtal sufiks-kinnisel keelel on vähemalt üks vasakfaktor, milles puudub tühi sõna ja vähemalt üks vasakfaktor, mis sisaldab tühja sõna, kuna vaatleme ainult mittetühjasid keeli. Seega ei saa vasakfaktorite ühisosasse kuuluda tühi sõna.

4.10.2 X-kinniste keelte aatomid

Aatomite omadus, mis kehtib kõigil alamklassi kuuluvatel keeltel.	**Prefiks-kinnine keel**		**Faktor-kinnine keel**	**Sufiks- kinnine keel**	
	puhas	**üldine**		**üldine**	**puhas**
Kõik positiivsed aatomid kuuluvad algse keelega samasse keelte alamklassi	Ei	Ei	Ei	Ei	Ei
Positiivsed aatomid on kõik X-kinnised keeled	Ei	Ei	Ei	Ei	Ei
Kõik positiivsed aatomid on algaatomid	Ei	Ei	Jah	Jah	Jah
Iga keele poolt kasutatav sümbol sisaldub vähemalt ühes positiivses aatomis ühetähelise sõnana.	Jah	Jah	Jah	Ei	Ei
Vähemalt üks keele poolt kasutatav sümbol sisaldub vähemalt ühes positiivses aatomis ühetähelise sõnana.	Jah	Jah	Jah	Jah	Jah

Aatomite omadus, mis kehtib kõigil alamklassi kuuluvatel keeltel.	**Prefiks-kinnine keel**		**Faktor-kinnine keel**	**Sufiks- kinnine keel**	
	puhas	**üldine**		**üldine**	**puhas**
Positiivsete aatomite ühend on faktor-kinnine keel	Jah	Jah	Jah	Ei	Ei
Positiivsete aatomite ühend on võrdne algse keelega.	Ei	Ei	Jah	Jah	Jah
Negatiivne aatom on vasakideaal	Ei	Ei	Ei	Ei	Jah
Negatiivne aatom on kahepoolne ideaal	Jah	Jah	Jah	Ei	Ei

Tabel 5 X-kinniste keelte aatomite omadused

Erinevalt vasakfaktoritest ei kandu ühegi X-kinnise keele alamklass üle kõigile positiivsetele aatomitele. Samuti ei kuulu X-kinnise keele kõik positiivsed aatomid ühte kindlasse X-kinnisesse keelde.

Ka keele aatomite osas keskendusime töös eelkõige omaduste leidmisele. Negatiivsed tulemid on eeskätt tõestatavad lisades 1 – 3 toodud näidetega.

Kuna kõik positiivsed aatomid on vasakfaktorite alamhulgad, siis kanduvad vasakfaktorite omadustest ka aatomitele üle omadused, mis seovad vasakfaktorite sõnu algse keelega. Kuna prefiks-, ja faktor-kinnise keele puhul on igal kasutatud sümbolil ühetäheline sõna vähemalt ühes vasakfaktoris, siis esineb iga kasutatud sümboli ühetäheline sõna ka ühes aatomis. Samal põhjusel on sufiks-kinnise keele mõnes positiivses aatomis vähemalt üks sümbol esindatud ühetähelise sõnana.

Samuti, kuna sufiks- ja faktor-kinniste keelte vasakfaktorid on algse keele sõnade alamhulgad, siis on algse keele sõnade alamhulgad ka kõik selle keele positiivsed aatomid.

Kuna positiivsete aatomite ühend on võrdne vasakfaktorite ühendiga, siis prefiks- ja faktor-kinniste keelte positiivsete aatomite ühend on faktor-kinnine keel ja nende keelte negatiivne aatom on kahepoolne ideaal.

Sufiks-kinnise keele positiivsete aatomite ühend on võrdne algse keelega. Kui tegemist on puhta sufiks-kinnise keelega, siis on tema positiivsete aatomite ühendiks sufiks-kinnine keel ja seega tema negatiivne aatom on vasak-ideaal.

5 Kokkuvõte

Töö eesmärgiks oli uurida prefiks-, sufiks- ja faktor-kinniste keelte vasakfaktoreid ja aatomeid ja leida neile iseloomulikke omadusi.

Analüüsis lähtusime prefiks-, sufiks- ja faktor-kinniste keelte tunnus-NFA'dest ja minimaalse DFA ning atomaadi leidmise protseduuridest.

Süstemaatiliselt valitud keelenäidete teisenduste tulemusena sõnastasime hüpoteesid, mille üldistatud kehtivust töö neljandas osas tõestasime, lähtudes vaadeldud keeleklasside reeglitest, leitud minimaalse DFA ning atomaadi omadustest ja vasakfaktorite ning aatomite üldistest omadustest.

Lisaks peamiseks eesmärgiks olnud vasakfaktorite ja aatomite omadustele tõestasime uuritavate keelte vahelisi seoseid ning leidsime iseloomulikke omadusi keelte olulistele automaatidele.

Töö tulemusena võib väita, et vaadeldud X-kinniste keelte klassid on omavahel tihedalt seotud. Nende vasakfaktoritel ja aatomitel on sarnaseid, aga ka vastandlikke omadusi.

Prefiks- ja faktor-kinniste keelte vasakfaktorid on prefiks-kinnised keeled ja sellest tulenevalt on ka neil sarnaseid omadusi. Nende keelte nii vasakfaktorite kui ka positiivsete aatomite ühend on faktor-kinnine keel ja negatiivne aatom on kahepoolne ideaal.

Teisalt, sufiks- ja faktor-kinniste keelte vasakfaktorid ja aatomid on algse keele sõnade alamhulgad. Nende keelte nii vasakfaktorite kui ka positiivsete aatomite ühend on võrdne algse keelega.

Töö peatükis 4.10 on kokkuvõte kõigist leitud vasakfaktorite ja aatomite omadustest, aga ka keelte vahelistest seostest ja oluliste automaatide skemaatilised kirjeldused.

Huvitava probleemina kerkis üles alamsõna-kinniste keelte vasakfaktorite ja aatomite omadused. Kuna kõik alamsõna-kinnised keeled on ka faktor-kinnised keeled, on nende vasakfaktoritel ja aatomitel faktor-kinniste keelte vastavate komponentide omadused. Samas on teada, et alamsõna-kinnise keele vasakfaktorid on kõik ka alamsõna-kinnised keeled aga faktor-kinnise keele kõik vasakfaktorid on vaid prefiks-kinnised keeled.

Sellest võib järeldada, et alamsõna-kinnistel keelte vasakfaktoritel ja aatomitel võib olla täiendavalt omadusi, mis ei laiene kõigile faktor-kinnistele keeltele.

Kasutatud kirjandus

[1] S. Yu, "Regular Languages," in *Handbook of Formal Languages*, Berlin Heidelberg, Springer-Verlag, 1997, pp. 41 - 110.

[2] A. V. Aho and J. D. Ullman, The Theory of Parsing, Translation, and Compiling, Englewood Cliffs: Prentice Hall, 1972.

[3] A. V. Aho, M. S. Lam, R. Sethi and J. D. Ullmann, Compilers - Principles, Techniques and Tools, Boston: Pearson Education, Inc, 2007.

[4] D. E. Knuth, J. H. Morris and V. R. Pratt, "Fast pattern matching in strings," *SIAM Journal on Computing,* vol. 6, no. 2, pp. 323-350, 1977.

[5] J. A. Brzozowski and M. Yoeli, Digital Networks, Englewood Cliffs: Prentice Hall, 1976.

[6] J. M. Baeten and W. P. Weijland, Process Algebra, Cambridge: Cambridge University Press, 1990.

[7] V. Diekert and G. Rozenberg, Book of Traces, World Scientific Publishing Co.Pte.Ltd, 1995.

[8] L. Guo, K. Salomaa and S. Yu, "Synchronization Expressions and Languages," in *Parallel and Distributed Processing*, Dallas, TX, USA, 1994.

[9] K. Culik II and S. Dube, "Rational and Affine Expressions for Image Description," *Discrete Applied Mathematics,* vol. 41, pp. 85 - 120, 1993.

[10] K. Culik II and P. C. von Rosenberg, "Generalized Weighted Finite Automata Based Image Compression.," *Journal of Universal Computer Science,* vol. 5, no. 4, pp. 227-242, 1993.

[11] K. Culik and S. Dube, "Affine Automata and Related Techniques for Generation of Complex Images.," *Theoretical Computer Science,* vol. 116, pp. 373 - 398, 1993.

[12] O. Nierstrasz, "Regular Types for Active Objects," in *Object-Oriented Software Composition*, Prentice Hall, 1995, pp. 99-121.

[13] K. Culik II and T. Harju, "Splicing semigroups of dominoes and DNA," *Discrete Applied Mathematics,* vol. 31, pp. 261 - 277, 1991.

[14] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Mathematical Biophysics,* vol. 5, pp. 115 - 133, 1943.

[15] S. C. Kleene, "Representation of events in nerve nets and finite automata," in *Automata Studies*, Princeton, 1956, pp. 3-41.

[16] J. Myhill, "Finite automata and the representation of events," in *WADD TR-57-624*, Ohio, Wright Patterson AFB, 1957, pp. 112-137.

[17] A. Nerode, „Linear automata transformation," *Proceedings of the American Mathematical Society,* kd. 9, pp. 541-544, 1958.

[18] M. O. Rabin and D. Scott, "Finite Automata and their Decision problems," *IBM Journal of Research and Development,* vol. 3, pp. 114 - 125, 1959.

[19] J. Brzozowski, "Canonical regular expressions and minimal state graphs for definite events," in *Proceedings of the Symposium on Mathematical Theory of Automata,* 1963.

[20] J. Brzozowski and H. Tamm, "Theory of átomata," *Theoretical Computer Science,* vol. 539, pp. 13 - 27, 2014.

[21] B. W. Watson, "Directly Constructing Minimal DFAs: Combining Two Algorithms by Brzozowski.," *In: Yu S., Păun A. (eds) Implementation and Application of Automata. CIAA 2000. Lecture Notes in Computer Science,* vol. vol 2088, pp. 311-317, 2001.

[22] M. V. de Parga, P. Garcia and D. Lopez, "A polynomial double reversal minimization algorithm for deterministic finite automata," *Theoretical Computer Science,* vol. 487, pp. 17-22, 2013.

[23] T. Kameda and P. Weiner, "On the State Minimization of Nondeterministic Finite Automata," *IEEE Transactions on Computers,* Vols. C-19, pp. 617 - 627, 1970.

[24] H. Tamm, "New Interpretation and Generalization of the Kameda-Weiner Method," in *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016),* Schloss Dagstuhl--Leibniz-Zentrum fuer Informatik, 2016, pp. 116:1 - 116:12.

[25] L. Polák, "Minimalizations of NFA Using the Universal Automaton," *International Journal of Foundations of Computer Science,* vol. 16, no. 05, pp. 999-1010, 2005.

[26] L. Ilie ja Y. Sheng, „Reducing NFAs by invariant equivalences,“ *Theoretical Computer Science,* kd. 306, pp. 373-390, 2003.

[27] H. Björklund and W. Martens, "The tractability frontier for NFA minimization," *Journal of Computer and System Sciences,* vol. 78, pp. 198-210, 2012.

[28] G. Thierrin, "Convex languages," in *Automata, Languages and Programming,* North Holland, M. Nivat, 1973, pp. 481- 492.

[29] T. Ang and J. Brzozowski, "Continuous languages," David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, 2008.

[30] J. Brzozowski, J. Shallit ja Z. Xu, „Decision problems for convex languages,“ *Information and Computation,* kd. 209, nr 3, pp. 353 - 367, 2011.

[31] J.-Y. Kao, N. Rampersad and J. Shallit, "On NFAs where all states are final, initial, or both," *Theoretical Computer Science,* vol. 410, pp. 5010-5021, 2009.

[32] J. A. Brzozowski and C. Sinnamon, "Complexity of proper prefix-convex regular languages," *Theoretical Computer Science,* vol. 787, pp. 2-13, 2019.

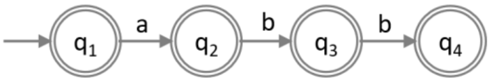
[33] J. Brzozowski, G. Jiraskova and C. Zou, "Quotient Complexity of Closed Languages," *Theory of Computing Systems,* vol. 54, no. 2, p. 277–292, 2014.

[34] A. de Luca and S. Varricchio, "Some Combinatorial Properties of Factorial Languages," in *In: Capocelli R.M. (eds) Sequences,* New York, Springer, 1990, pp. 258 - 266.

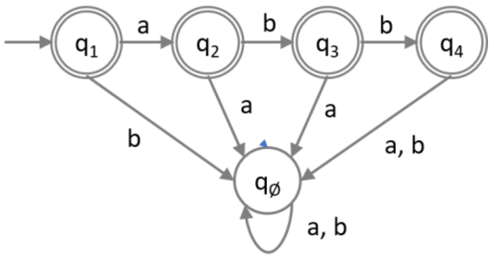
Lisa 1 – Prefiks-kinnine keel

Keel: $\mathcal{L} = \{\varepsilon, a, ab, abb\}$

NFA $\mathcal{N}$	a	b
$\leftarrow\rightarrow q_1$	$\{q_2\}$	-
$\leftarrow q_2$	-	$\{q_3\}$
$\leftarrow q_3$	-	$\{q_4\}$
$\leftarrow q_4$	-	-



$\mathcal{N}^{DM}$	a	b
$\leftarrow\rightarrow q_1$	q_2	$q_\emptyset$
$\leftarrow q_2$	$q_\emptyset$	q_3
$\leftarrow q_3$	$q_\emptyset$	q_4
$\leftarrow q_4$	$q_\emptyset$	$q_\emptyset$
$q_\emptyset$	$q_\emptyset$	$q_\emptyset$



Vasakfaktorid :

$K_1 = \{\varepsilon, a, ab, abb\} = \mathcal{L}$

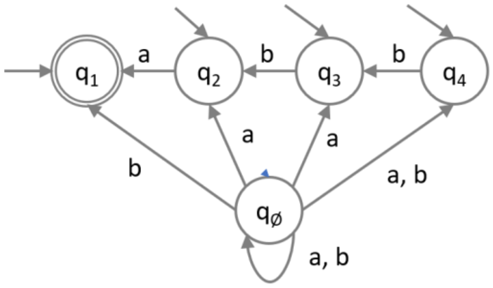
$K_2 = \{\varepsilon, b, bb\}$

$K_3 = \{\varepsilon, b\}$

$K_4 = \{\varepsilon\}$

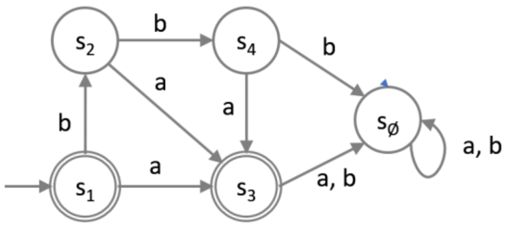
$K_\emptyset = \emptyset$

$\mathcal{N}^{DMR}$	**a**	**b**
←→q_1	-	-
→q_2	q_1	-
→q_3	-	q_2
→q_4	-	q_3

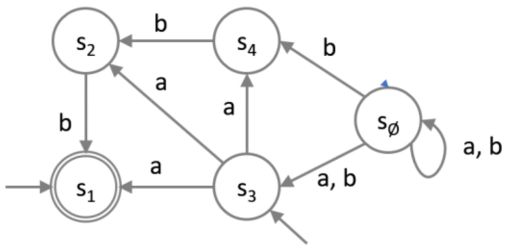


$\mathcal{N}^{DMR} \rightarrow \mathcal{N}^{DMRD}$	**a**	**b**	**uus**
←→$\{q_1q_2q_3q_4\}$	$\{q_1\}$	$\{q_{3,}q_2\}$	s_1
$\{q_2q_3\}$	$\{q_1\}$	$\{q_2\}$	s_2
←$\{q_1\}$	-	-	s_3
$\{q_2\}$	$\{q_1\}$	-	s_4

$\mathcal{N}^{DMRD}$	**a**	**b**
←→s_1	s_3	s_2
s_2	s_3	s_4
←s_3	$s_\emptyset$	$s_\emptyset$
s_4	s_3	$s_\emptyset$



$\mathcal{N}^{DMRDR}$	**a**	**b**
←→s_1	-	-
s_2	-	$\{s_1\}$
→s_3	$\{s_{1,}\ s_{2,}\ s_4\}$	-
s_4	-	$\{s_2\}$



Aatomid:

$A_1 = \{\varepsilon\}$ $\quad$ $A_2 = \{b\}$

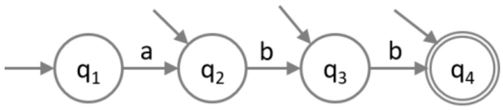
$A_3 = \{a, ab, abb\}$ $\quad$ $A_4 = \{bb\}$

$A_0 = \Sigma^*aa + \Sigma^*aab + \Sigma^*aabb + \Sigma^*ba + \Sigma^*bab + \Sigma^*babb + \Sigma^*bbb$

Lisa 2 – Sufiks-kinnine keel

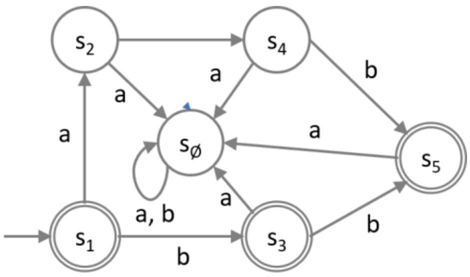
Keel $\mathcal{L} = \{\varepsilon, b, bb, abb\}$

NFA $\mathcal{N}$	a	b
$\rightarrow q_1$	$\{q_2\}$	-
$\rightarrow q_2$	-	$\{q_3\}$
$\rightarrow q_3$	-	$\{q_4\}$
$\leftarrow\rightarrow q_4$	-	-

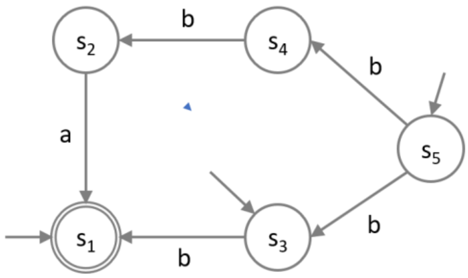


$\mathcal{N} \rightarrow \mathcal{N}^D$	a	b	uus
$\leftarrow\rightarrow\{q_1 q_2 q_3 q_4\}$	$\{q_2\}$	$\{q_3, q_4\}$	s_1
$\{q_2\}$	-	$\{q_3\}$	s_2
$\leftarrow\{q_3\, q_4\}$	-	$\{q_4\}$	s_3
$\{q_3\}$	-	$\{q_4\}$	s_4
$\leftarrow\{q_4\}$	-	-	s_5

$\mathcal{N}^{DM}$	a	b
$\leftarrow\rightarrow s_1$	s_2	s_3
s_2	$s_\emptyset$	s_4
$\leftarrow s_3$	$s_\emptyset$	s_5
s_4	$s_\emptyset$	s_5
$\leftarrow s_5$	$s_\emptyset$	$s_\emptyset$

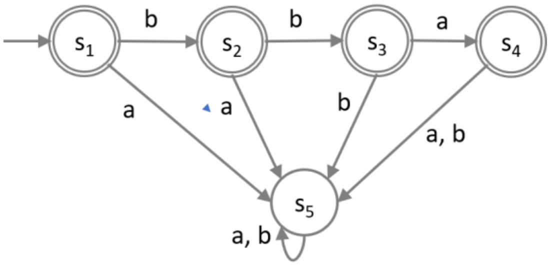


$\mathcal{N}^{DMR}$	a	b
$\leftarrow\rightarrow s_1$	-	-
s_2	$\{s_1\}$	-
$\rightarrow s_3$	-	$\{s_1\}$
s_4		$\{s_2\}$
$\rightarrow s_5$	-	$\{s_3, s_4\}$

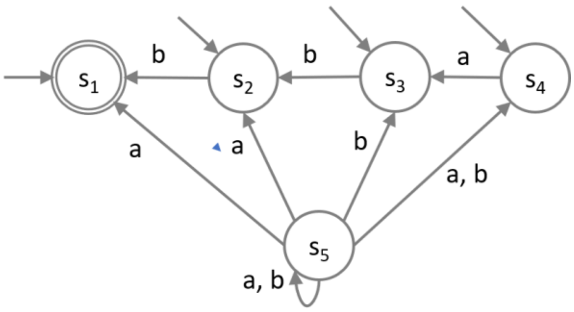


$\mathcal{N}^{DMR} \rightarrow \mathcal{N}^{DMRD}$	a	b	uus
$\leftarrow\rightarrow\{s_1 s_3 s_5\}$	-	$\{s_1, s_3, s_4\}$	s_1
$\leftarrow\{s_1, s_3, s_4\}$	-	$\{s_1, s_2\}$	s_2
$\leftarrow\{s_1, s_2\}$	$\{s_1\}$	-	s_3
$\leftarrow\{s_1\}$	-	-	s_4

$\mathcal{N}^{DMRD}$	a	b
$\leftarrow\rightarrow s_1$	$s_\emptyset$	s_2
$\leftarrow s_2$	$s_\emptyset$	s_3
$\leftarrow s_3$	s_4	$s_\emptyset$
$\leftarrow s_4$	$s_\emptyset$	$s_\emptyset$



$\mathcal{N}^{DMRDR}$	a	b
$\leftarrow\rightarrow s_1$	-	-
$\rightarrow s_2$	-	s_1
$\rightarrow s_3$		s_2
$\rightarrow s_4$	s_3	-



Vasakfaktorid :

$K_1 = \{\varepsilon, b, bb, abb\} = \mathcal{L}$

$K_2 = \{bb\}$

$K_3 = \{\varepsilon, b\}$

$K_4 = \{b\}$

$K_5 = \{\varepsilon\}$

$K_\emptyset = \emptyset$

Aatomid:

$A_1 = \{\varepsilon\}$

$A_2 = \{b\}$

$A_3 = \{bb\}$

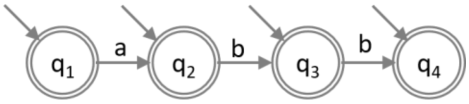
$A_4 = \{abb\}$

$A_0 = \Sigma^* a + \Sigma^* ab + \Sigma^* bbb + \Sigma^* aabb + \Sigma^* babb$

Lisa 3 – Faktor–kinnine keel

Keel $\mathcal{L} = \{\varepsilon, a, b, ab, bb, abb\}$

NFA $\mathcal{N}$	a	b
$\leftarrow\rightarrow\{q_1\}$	$\{q_2\}$	-
$\leftarrow\rightarrow\{q_2\}$	-	$\{q_3\}$
$\leftarrow\rightarrow\{q_3\}$	-	$\{q_4\}$
$\leftarrow\rightarrow\{q_4\}$	-	-



NFA determiniseerimine

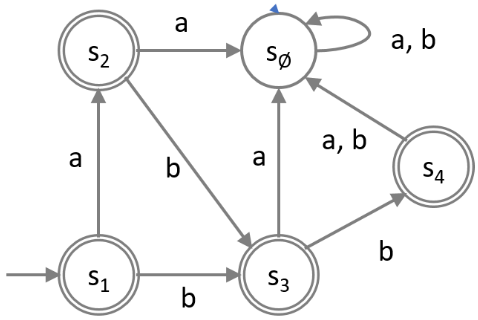
$\mathcal{N} \rightarrow \mathcal{N}^D$	**A**	**B**
$\longleftrightarrow\{q_1 q_2 q_3 q_4\}$	$\{q_2\}$	$\{q_3\ q_4\}$
$\leftarrow\{q_2\}$	-	$\{q_3\}$
$\leftarrow\{q_3\ q_4\}$	-	$\{q_4\}$
$\leftarrow\{q_3\}$	-	$\{q_4\}$
$\leftarrow\{q_4\}$	-	-

Minimeerimine

$\mathcal{N} \rightarrow \mathcal{N}^{DM}$	**a**	**b**	**Uus**
$\longleftrightarrow\{q_1 q_2 q_3 q_4\}$	$\{q_2\}$	$\{q_3 q_4\}$	s_1
$\leftarrow\{q_2\}$	-	$\{q_3 q_4\}$	s_2
$\leftarrow\{q_3 q_4\}$	-	$\{q_4\}$	s_3
$\leftarrow\{q_4\}$	-	-	s_4

Minimaalne DFA

$\mathcal{N}^{DM}$	a	b
←→s_1	s_2	s_3
←s_2	$s_\emptyset$	s_3
←s_3	$s_\emptyset$	s_4
←s_4	$s_\emptyset$	$s_\emptyset$



Vasakfaktorid

$K_1 = \{\varepsilon, a, b, ab, bb, abb\}$

$K_2 = \{\varepsilon, b, bb\}$

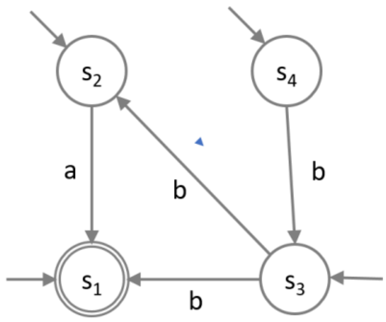
$K_3 = \{\varepsilon, b\}$

$K_4 = \{\varepsilon\}$

$K_\emptyset = \emptyset$

Minimaalse DFA pöördautomaat

$\mathcal{N}^{DMRT}$	a	b
←→$\{s_1\}$	-	-
→$\{s_2\}$	$\{s_1\}$	-
→$\{s_3\}$	-	$\{s_1, s_2\}$
→$\{s_4\}$		$\{s_3\}$

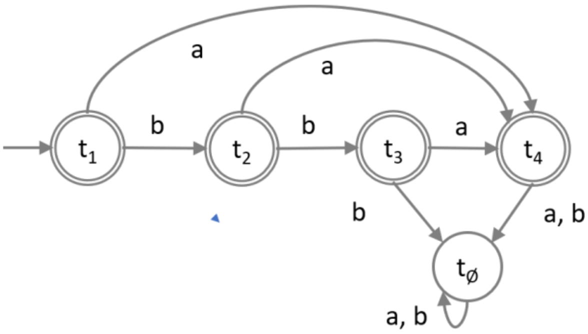


Pöördautomaadi determiniseerimine

$\mathcal{N}^{DMR}$→ $\mathcal{N}^{DMRD}$	a	b	uus
←→ $\{s_1, s_2, s_3, s_4\}$	$\{s_1\}$	$\{s_1, s_2, s_3\}$	t_1
← $\{s_1, s_2, s_3\}$	$\{s_1\}$	$\{s_1, s_2\}$	t_2
← $\{s_1, s_2\}$	$\{s_1\}$	-	t_3
← $\{s_1\}$	-	-	t_4

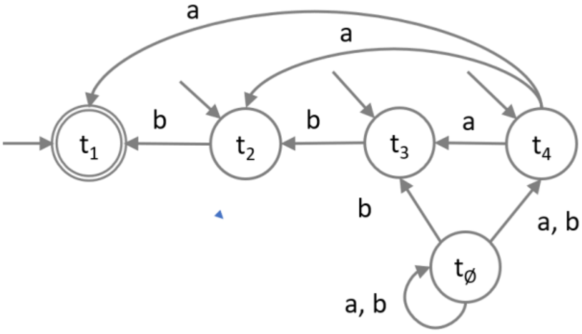
Pöördkeele minimaalne DFA

$\mathcal{N}^{DMRD}$	**a**	**b**
←→t_1	t_4	t_2
←t_2	t_4	t_3
←t_3	t_4	$t_\emptyset$
←t_4	$t_\emptyset$	$t_\emptyset$
$t_\emptyset$	$t_\emptyset$	$t_\emptyset$



Atomaat

$\mathcal{N}^{DMRDR}$	**a**	**b**
←→{t_1}	-	-
→{t_2}	-	{t_1}
→{t_3}	-	{t_2}
→{t_4}	{t_1,t_2,t_3}	-
{$t_\emptyset$}	{$t_\emptyset$, t_4,}	{$t_\emptyset$, t_3, t_4}



Aatomid

$A_1 = \{\varepsilon\}$ $\quad$ $A_2 = \{b\}$

$A_3 = \{bb\}$ $\quad$ $A_4 = \{abb, ab, a\}$

$A_0 = \Sigma^* aa + \Sigma^* ba + \Sigma^* aab + \Sigma^* bbb + \Sigma^* bab + \Sigma^* aabb + \Sigma^* babb$