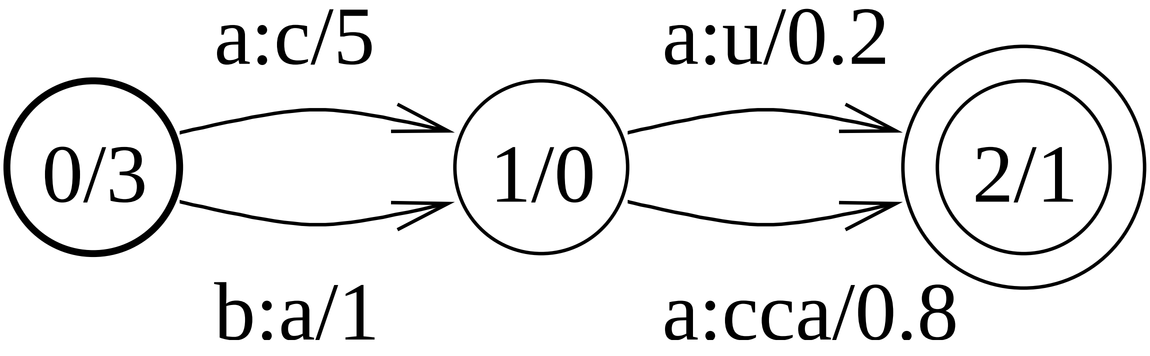
Äärellisien tilakoneiden käyttö puheentunnistuksessa

perustuen Mehryar Mohrin artikkeleihin

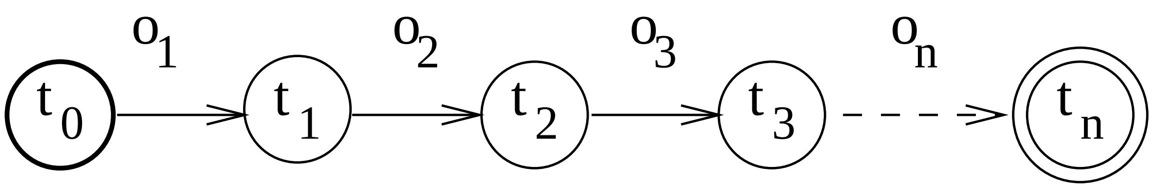
Vesa Siivola, 20. marraskuuta 2001

Mikä on äärellinen tilamuunnin ?



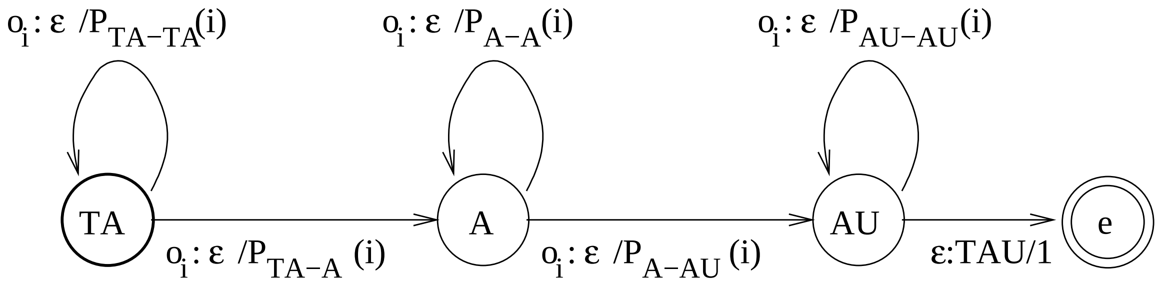
- Epäformaalisti: Syö symboleja, tulostaa toisia symboleja ja mahdollisesti painoja.
- deterministinen versio laskennallisesti tehokas: O(syötteen pituus)
- tilankäyttö mahdollista minimoida deterministisille muuntimille
- Helppo esittää lingvistisiä ja morfologisiä sääntöjä intuitiivisessä muodossa

Akustinen malli



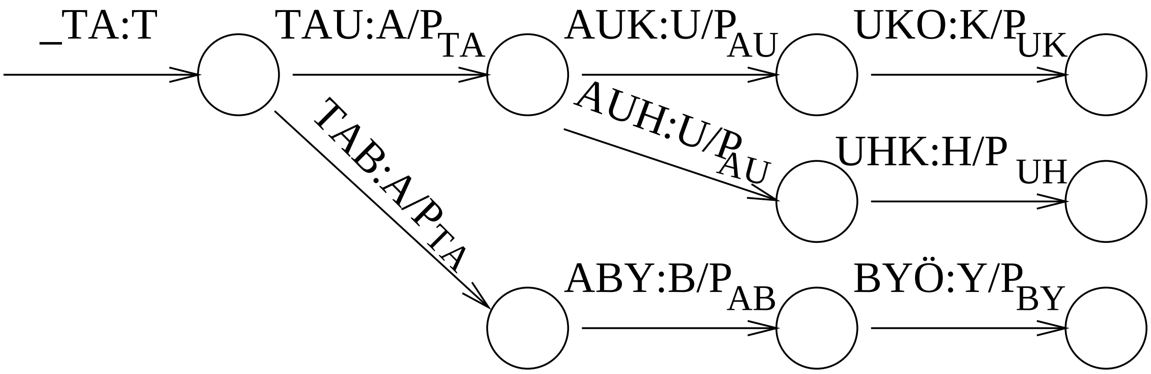
- Käsitellään puheentunnistinta, jossa diskreetit tilat (vektorikvantisaatio).
- Kullakin ajan hetkellä akustinen malli emittoi lähimmän kvantisointivektorin indeksin.

Vektori-indekseistä kontekstiriippuviksi foneemeiksi



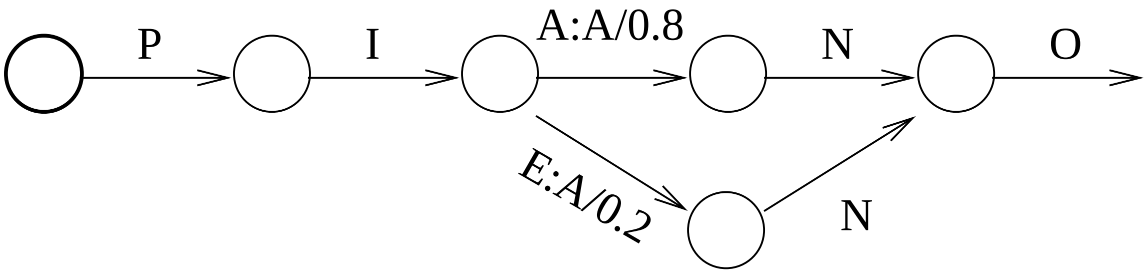
- Muunnin syö lähimmän kvantisaattorin indeksin ja vähentää todennäköisyyttä kullekin tilasiirtymä/havainto-parille
- Päästyään loppuun, tulostaa annetun foneemin konteksteineen (tässä foneemi A, jota on edeltänyt T ja seuraa U)

Kontekstiriippuvat foneemit yksittäisiksi foneemeiksi



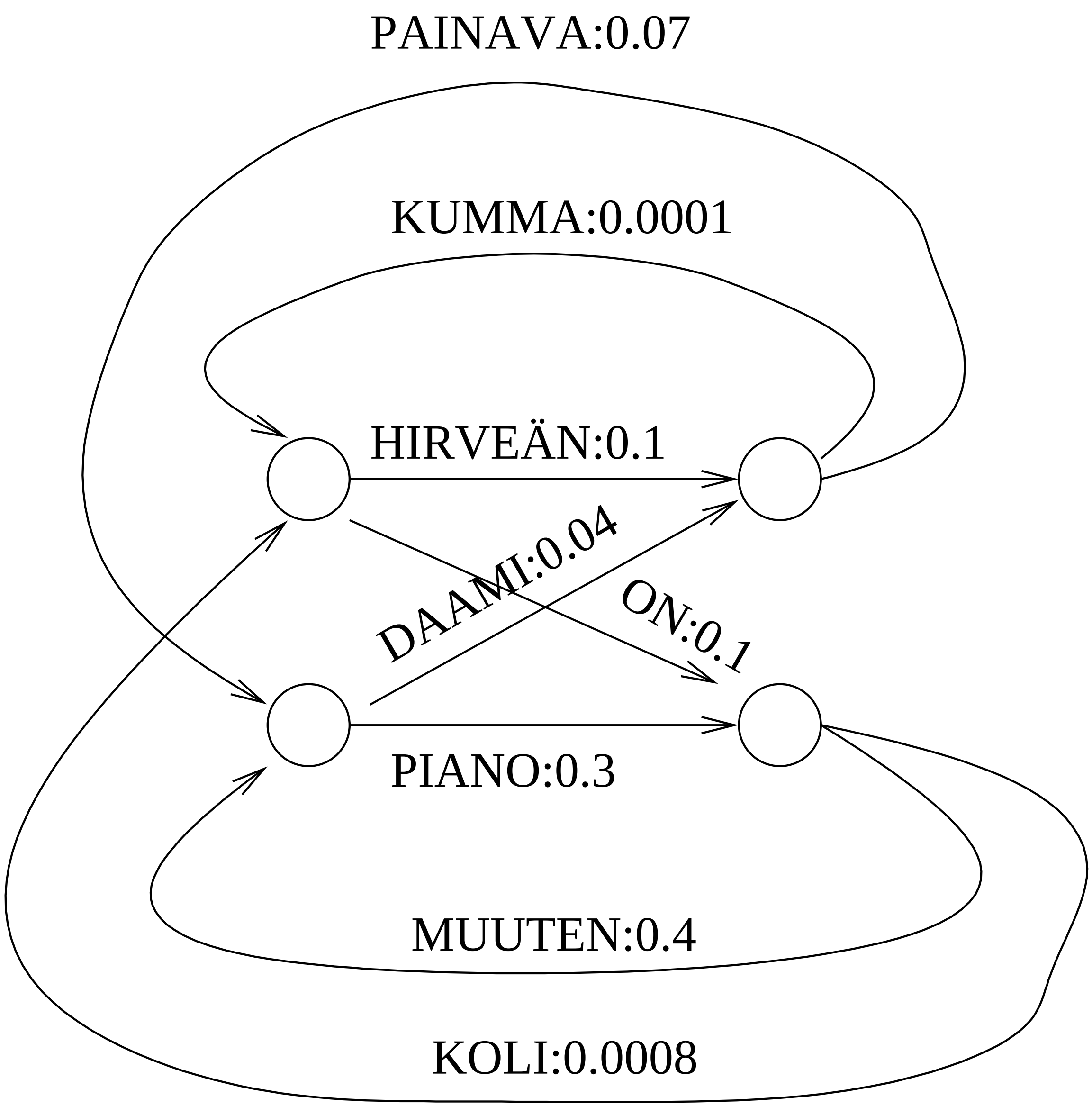
- Esittää kaikki sallitut kontekstiriippuvien foneemien ketjut. Kuvassa vain erittäin pieni osa esimerkkinä.
- Syö foneemin konteksteineen ja palauttaa yksittäisen foneemin

Ääntämissanakirja



- Mallittaa saman sanan monta ääntämystä eri murteissa.
- Syö foneemeja, palauttaa sanoja
- Turkulaisen sanoman “PIANO”:n todennäköisyys 0.2. (joka viides tässäkin luokassa turkulainen ?)

Kielioppi

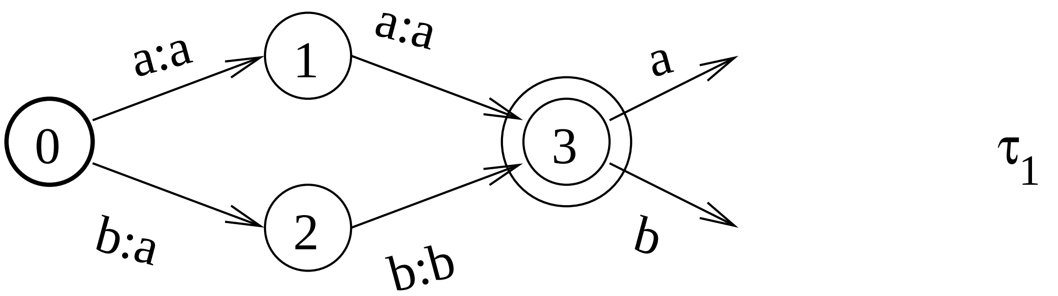


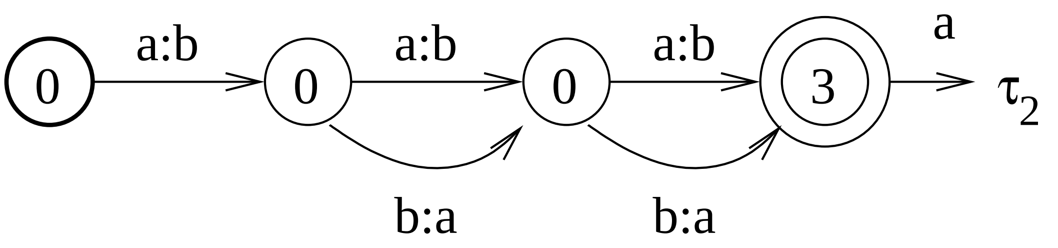
- Kuvassa hämmästyttävän pieni bigram-kielioppi
- Syö sanoja, palauttaa bigram-todennäköisyyksiä

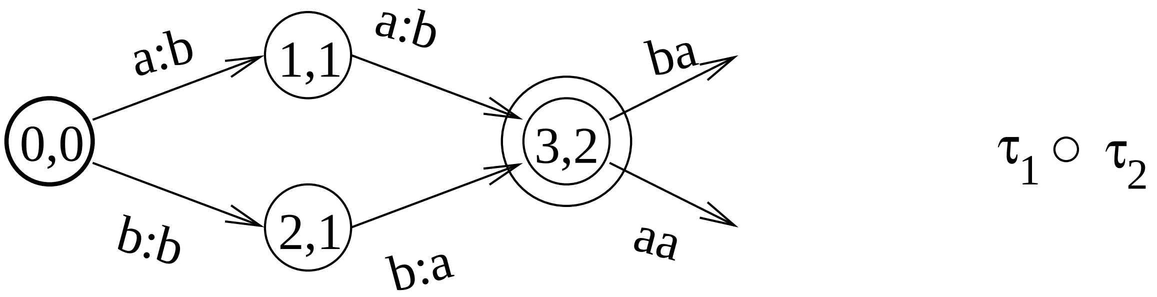
Kaiken ydistäminen

- Kaikki mallit saman formuloinnin alla, voidaan käyttää tehokkaita optimointimenetelmiä
- Saman tason rinnakkaiset mallit voidaan yhdistää unioni-operaattorilla $+$.
- Siispä kaikkien kontekstiriippuvien foneemien mallin yhdistelmä on $\tau_{AAA} + \tau_{AAB} + \tau_{AAC} + \dots$
- Mallit voidaan yhdistää kompositio-operaattorilla $\circ$.
- Koko malli: Kielioppi $\circ$ Ääntömalli $\circ$ foneemimallit $\circ$ kontekstifoneemit $\circ$ Havainnot
- Yhdistämisen tuloksena hillittömän kokoinen muunnin
- Jo rakennettaessa tarvitaan agressiivista karsintaa

Kompositio $\circ$

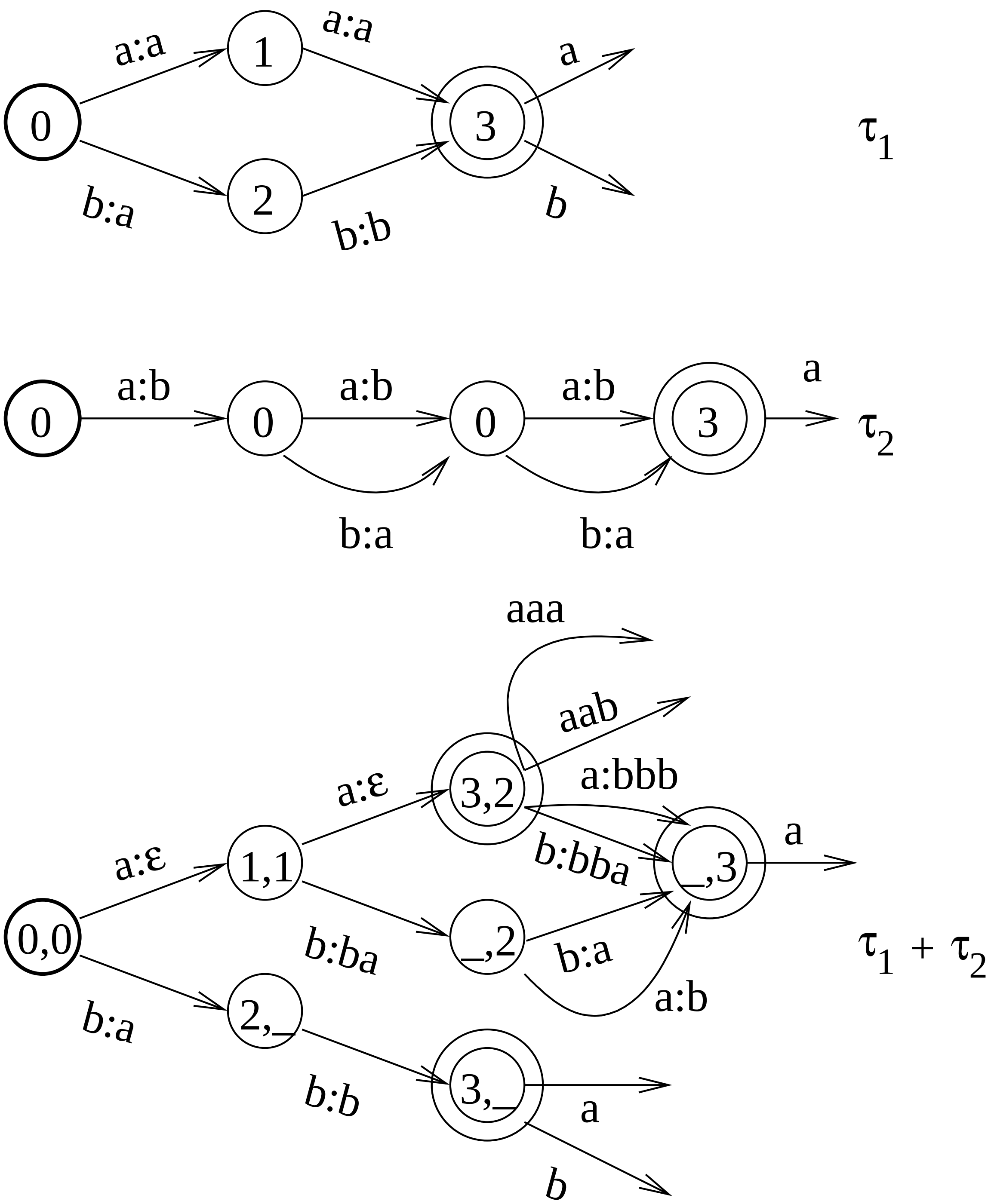






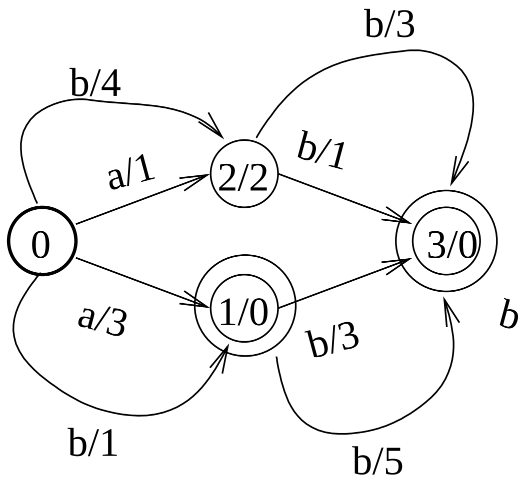
- Kompositiokone tuottaa saman tuloksen kuin jos syöteeseen olisi ensin käytetty muunninta τ_1 ja sitten muunninta τ_2

Unioni +

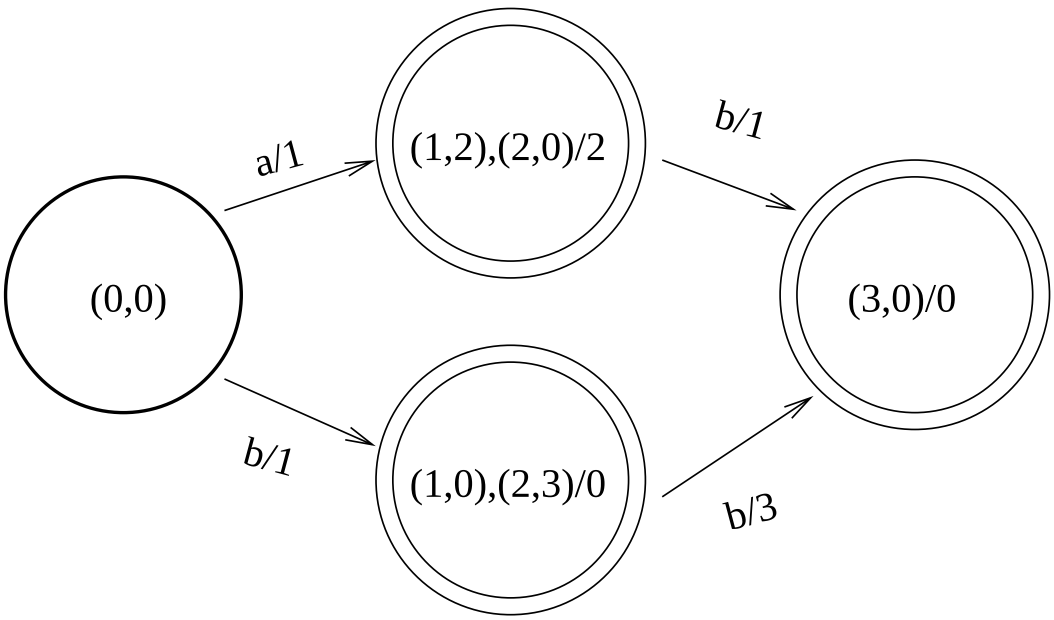


- Unionikone hyväksyy samat syöteet kuin kaikki sen rakentamiseen käytetyt koneet

Determinisointi



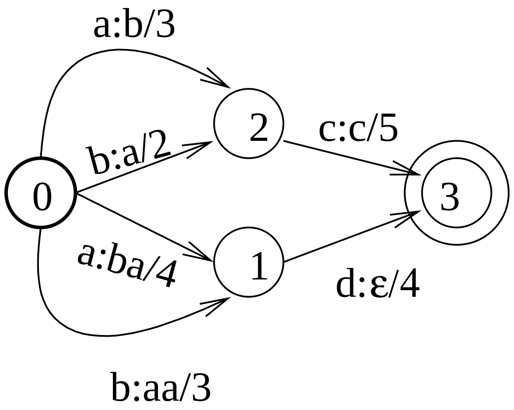
τ_1



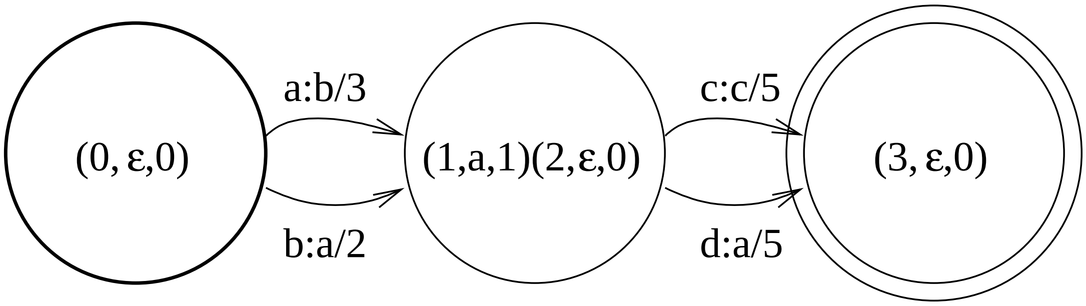
τ_2

- Kuvassa muuntaja τ_1 on determinisoitu puheentunnistuksessa usein käytetyllä trooppisella semirenkaalla $(\mathbb{R} \cup \{\infty\}, \min, +, \infty, 0)$ muuntajaksi τ_2

Determinisointi 2



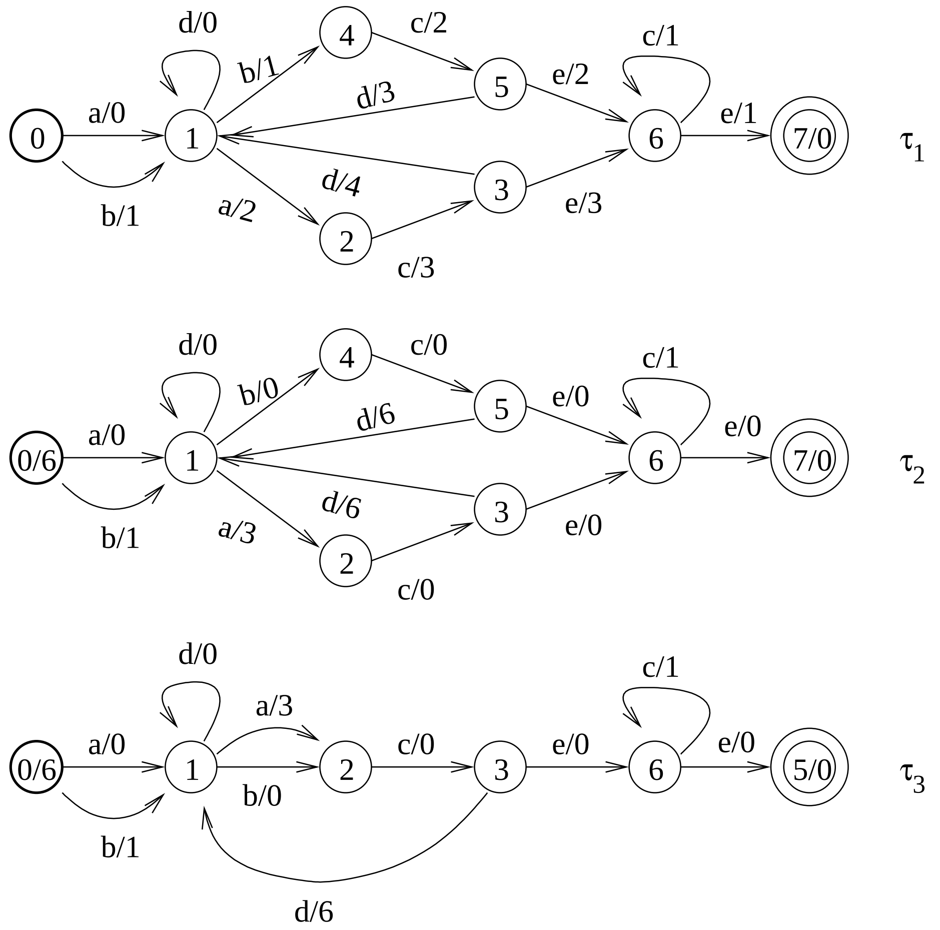
τ_1



τ_2

- Kuvassa muuntaja τ_1 on determinisoitu vasempipuoleisen ketjusemirenkaan $(\Sigma^* \cup \{\infty\}, \wedge, \cdot, \infty, \epsilon)$ ja trooppisen semirenkaan ristitulon muodostamassa semirenkaassa muuntajaksi τ_2

Minimointi



- Kuvassa muuntajan τ_1 painot on ensin työnnetty mahdollisimman pitkälle alkua kohti, saadaan τ_2.
- Painojen alkuun kasaus voi olla hyödyllistä myös puheen dekoodauksen kannalta: tehokkaampi karsiminen
- Lopuksi τ_2 on minimoitu ja saadaan τ_3

Kokeita

Sanahilat

- Puheentunnistimista voidaan ottaa ulos sanahila, jossa on hila melko todennäköisiä sanoja ja transitioita
- Kovin redundantti
- 2000 sanan tunnistimesta pulautettu hila kyselylle "Which flights leave from Detroit and arrive at Saint-Petersburg around nine am ?" sisältää 106 tilaa, 359 transitiota, ja 83 miljoonaa polkua
- Determinisoimalla 38 tilaa, 51 transitiota, 18 polkua
- Minimoimalla 25 tilaa, 33 transitiota, 18 polkua

Kokeita

Ranskan ja Englannin kielen morfologia

Kieli	Ranska	Englanti
Rivejä	828000	145000
Koko	27.9Mb	3.6Mb
tiloja	389650	48770
transitioita	633520	101970
Automaatin koko	3.1Mb	470Kb

- Säännöt saadaan generoitua helposti ja kompaktiin kokoon

Kokeita

Pohjoisamerikan kauppauutiset (NAB News)

- Sanasto 40.0000
- trigram-malli
- 4M transitiota
- Tunnistus reaaliajassa

Isommat mallit

- voidaan käyttää uudelleenpisteyttämään parhaat tulokset
- esim. 40M transitioinen 6gram-malli, 0.05 kertaa reaaliajassa

Yhteenveto

- Yhteinen kehys kaikille malleille
- Valmis muuntaja suhteellisen helppo dekoodata
- Verkon generointi voidaan suorittaa pyydettäessä lennossa: voidaan vähentää muistivaatimuksia kohtuullisiksi
- Löytyy tehokkaita algoritmejä verkon determinisoimiseksi ja minimoimiseksi

Kotitehtävät

1) Mitkä seuraavista ovat semirenkaita ?

$$(\mathbb{R}, +, \cdot, 0, 1) \tag{1}$$

$$(\mathbb{N}, -, \cdot, 0, 1) \tag{2}$$

$$(\mathbb{R} \cup \{\infty\}, \oplus_l, +, \infty, 0) \tag{3}$$

missä operaattori $\oplus_l$ on määritelty

$$x \oplus_l y = -ln(e^{-x} + e^{-y}) \tag{4}$$

Vinkkejä: semirenkaalle $(\mathbb{K}, \oplus, \otimes, \bar{0}, \bar{1})$ operaatioiden $\oplus$ ja $\otimes$ pitää olla suljettuja joukossa $\mathbb{K}$ ja niiden identiteettien pitää olla $\bar{0}$ ja $\bar{1}$. $\bar{0}$ on annihilaattori operaatiolle $\otimes$. Lisäksi kuvausten pitää olla molempien operaattoreiden suhteen assosiativisiä, $\oplus$ suhteen kommutatiivinen ja molempien suhteen distributiivinen. Parempi, pitempi selitys löytyy http://mathworld.wolfram.com/Semiring.html. Heuristisemmatkin ratkaisutavat kelpaavat.

Kotitehtävät

2) Valitse edellisestä tehtävästä oikeat semirenkaat. Jos muuntaja determinisoidaan ja minimoidaan tämän semirenkaan yli, miten muuntajan kaaret ja kustannukset voidaan tulkita ?

Vinkki: Trooppisen semirenkaan $(\mathbb{R} \cup \{\infty\}, min, +, \infty, 0)$ voidaan tulkita etsivän halvimmat polut (Viterbi) kun kaarien kustannukset ovat todennäköisyyksien negatiivisiä logaritmejä.

Vastaukset tehtäviin voi antaa muutamalla lauseella. Jos tehtävät tuntuvat ylitsepääsemättömän vaikeilta, vinkkejä voi udella (vsiivola@cc.hut.fi).