

Markovin piilomallit, Viterbin polut ja Chu-Liu / Edmondsin algoritmi

Näytös 1/3: Kielitiedepläjäys

A black and white photograph of a theater stage. The stage is framed by a large, ornate archway with decorative carvings. Heavy, dark curtains hang across the stage, partially drawn to reveal a bright light source at the bottom. In the foreground, the silhouettes of several rows of theater seats are visible, facing the stage. The text "Näytös 1/3: Kielitiedepläjäys" is overlaid in white, bold, sans-serif font across the upper portion of the image.

Minä

tarvitsen

tähän

esimerkin.

Minä

PRON

tarvitsen

NOUN

tähän

PRON

esimerkin.

NOUN

Sanaluokat

Minä

PRON

tarvitsen

~~NOUN~~

tähän

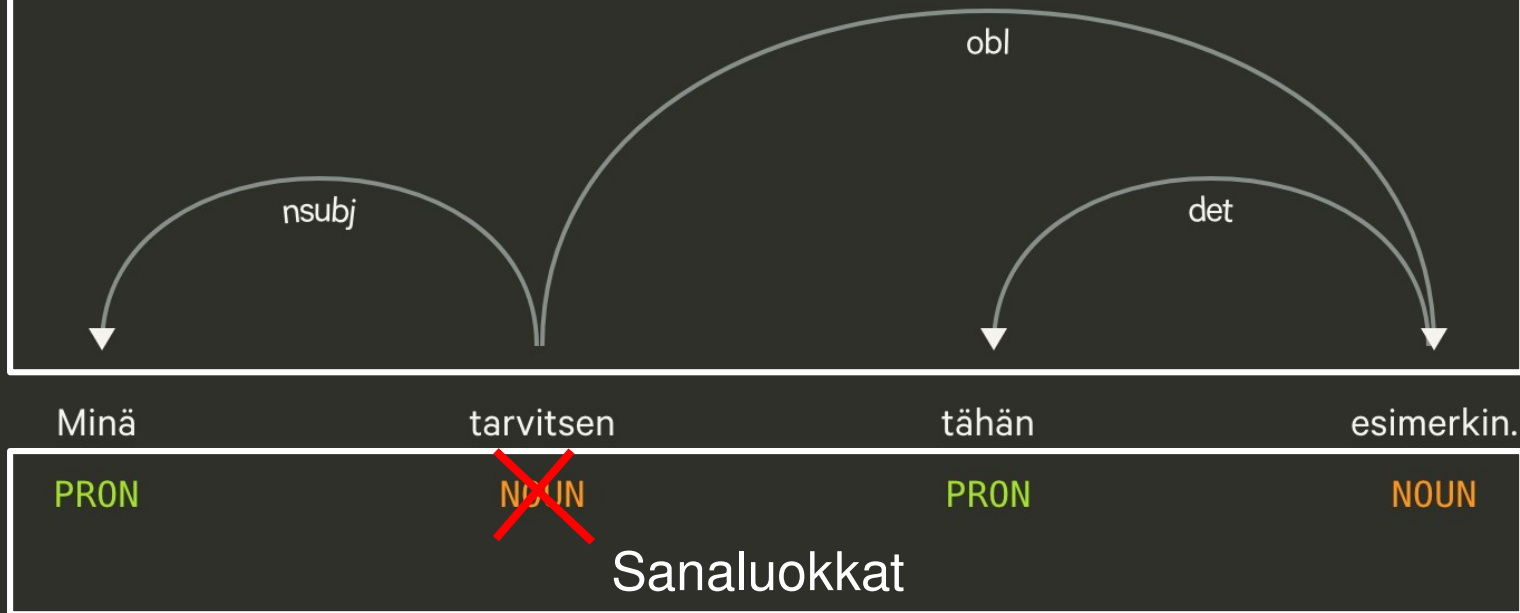
PRON

esimerkin.

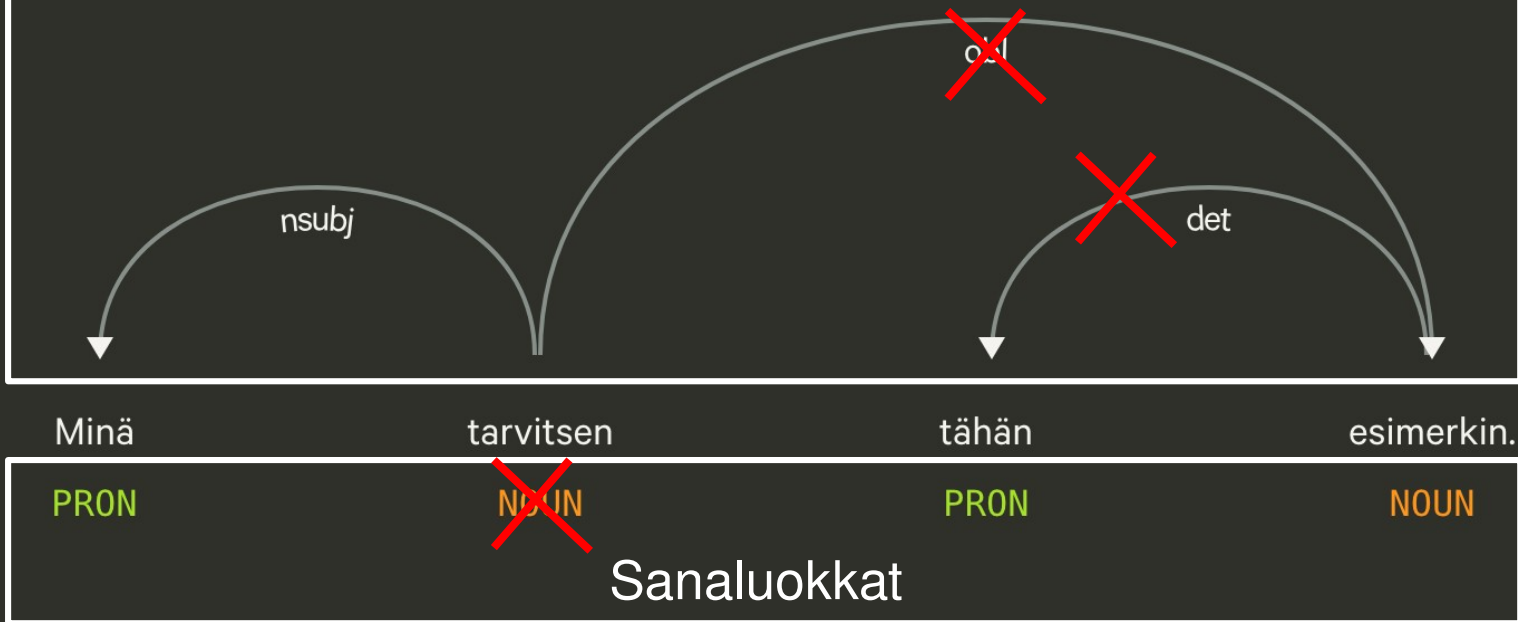
NOUN

Sanaluokat

Lauseen jäsennys



Lauseen jäsennys



Takapakki: mitä olivat...

Sanaluokat

Sanaluokat

Nimi	Esimerkki	Mitä tekee
Verbi (teonsana)	Juosta	Kuvaa tekemisen
Substantiivi (asiasana)	Talo	Kuvaa asian
Adjektiivi (laatusana)	Hyvä	Kuvaa sen, millainen jokin on
Pronomini (asemosana)	Tuo	Viittaa johonkin muuhun
Numeraali (lukusana)	23, toinen	Kuvaa lukumäärään
Taipumattomat sanat	Ehkä, jos, joka	~kaikki muu

Sanaluokat

HUOM: ei tärkeä tietää yksittäisiä.
Ydinajatus:
Tapa jaotella sanoja yksitellen.

Nimi	Esimerkki	Mitä tekee
Verbi (teonsana)	Juosta	Kuvaa tekemisen
Substantiivi (asiasana)	Talo	Kuvaa asian
Adjektiivi (laatusana)	Hyvä	Kuvaa sen, millainen jokin on
Pronomini (asemosana)	Tuo	Viittaa johonkin muuhun
Numeraali (lukusana)	23, toinen	Kuvaa lukumäärään
Taipumattomat sanat	Ehkä, jos, joka	~kaikki muu

Sanaluokat

Riippuvat keltä kysyy :D



HUOM: ei tärkeä tietää yksittäisiä.
Ydinajatus:
Tapa jaotella sanoja yksitellen.

Nimi	Esimerkki	Mitä tekee
Verbi (teonsana)	Juosta	Kuvaa tekemisen
Substantiivi (asiasana)	Talo	Kuvaa asian
Adjektiivi (laatusana)	Hyvä	Kuvaa sen, millainen jokin on
Pronomini (asemosana)	Tuo	Viittaa johonkin muuhun
Numeraali (lukusana)	23, toinen	Kuvaa lukumäärään
Taipumattomat sanat	Ehkä, jos, joka	~kaikki muu

Sanaluokat

Kutakuinkin kuten koulussa käydään →

Riippuvat keltä kysyy :D



HUOM: ei tärkeä tietää yksittäisiä.
Ydinajatus:
Tapa jaotella sanoja yksitellen.

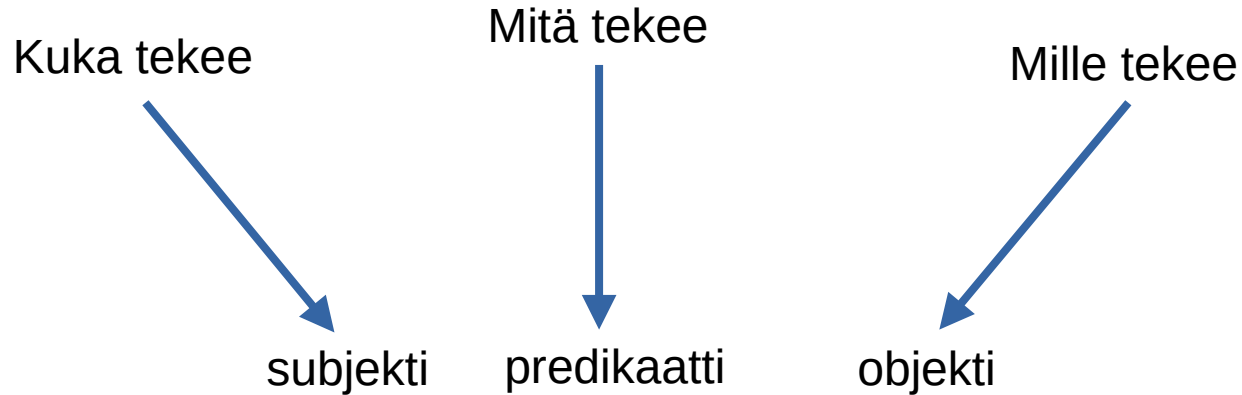
Nimi	Esimerkki	Mitä tekee
Verbi (teonsana)	Juosta	Kuvaa tekemisen
Substantiivi (asiasana)	Talo	Kuvaa asian
Adjektiivi (laatusana)	Hyvä	Kuvaa sen, millainen jokin on
Pronomini (asemosana)	Tuo	Viittaa johonkin muuhun
Numeraali (lukusana)	23, toinen	Kuvaa lukumäärään
Taipumattomat sanat	Ehkä, jos, joka	~kaikki muu

Takapakki: mitä olivat...

Lauseenjäsenet: sanat lauseessa

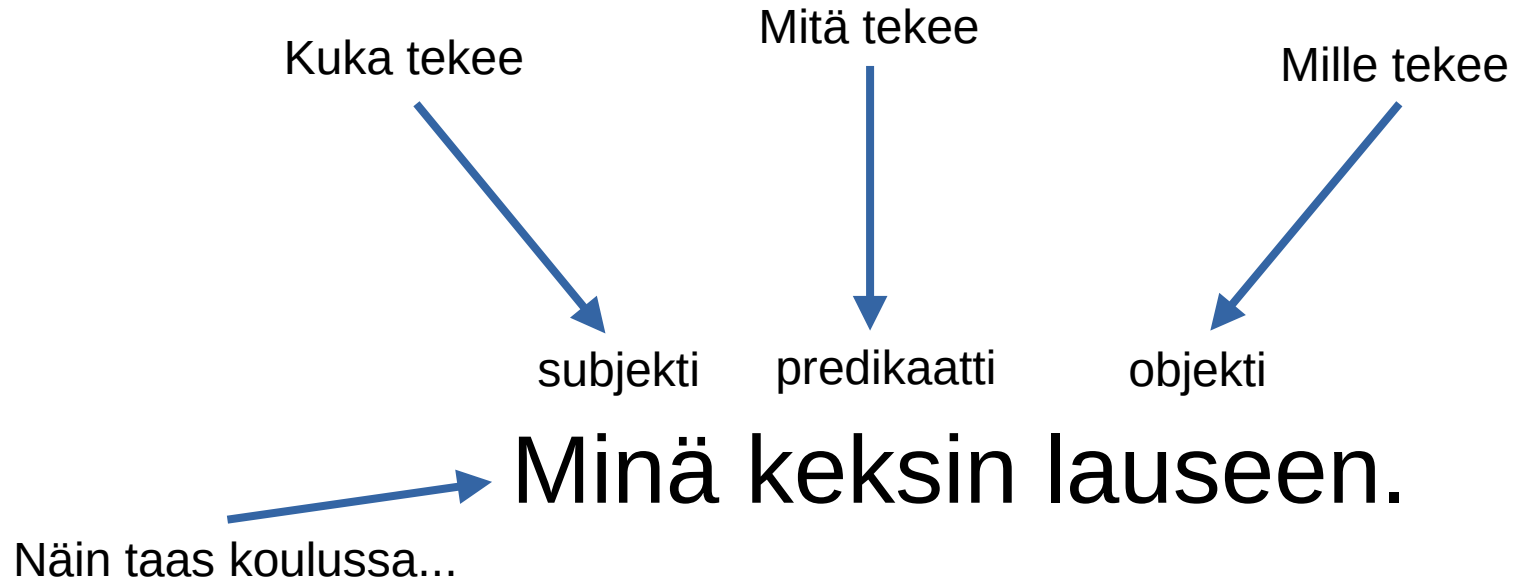
Minä keksin lauseen.

Lauseenjäsenet: sanat lauseessa



Minä keksin lauseen.

Lauseenjäsenet: sanat lauseessa



Lauseenjäsenet: sanat lauseessa

subjekti predikaatti ??? ???
Minä keksin vaikeamman lauseen.

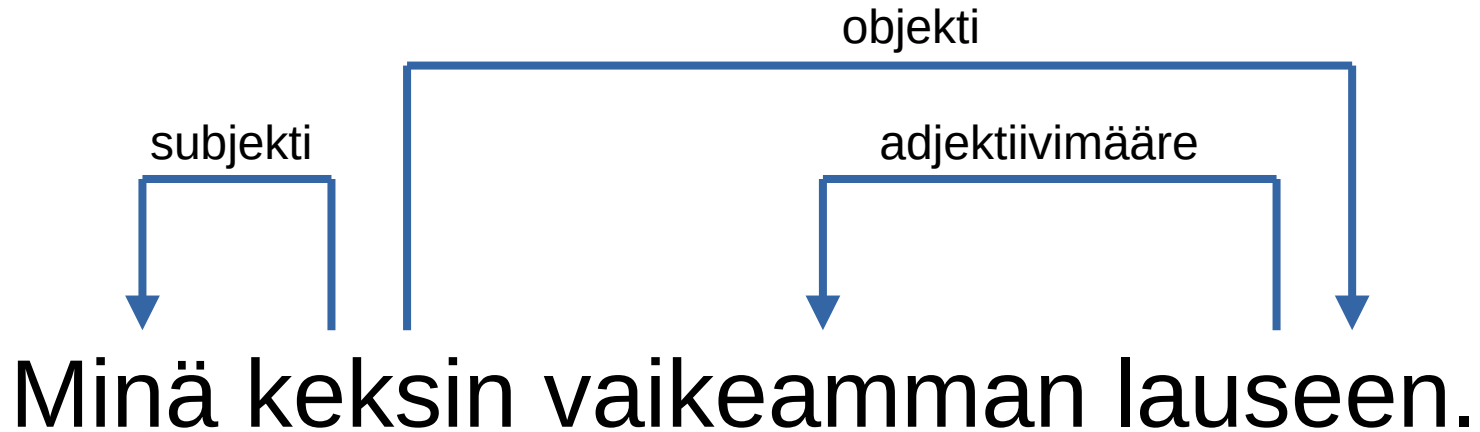
Lauseenjäsenet: sanat lauseessa

Tarvitaan väkevämpi teoria: riippuvuusjäsenitys

subjekti	predikaatti	???	???
Minä	keksin	vaikeamman	lauseen.

Lauseenjäsenet: sanat lauseessa

Esitetään jäsenitys suunnattuna puuna.



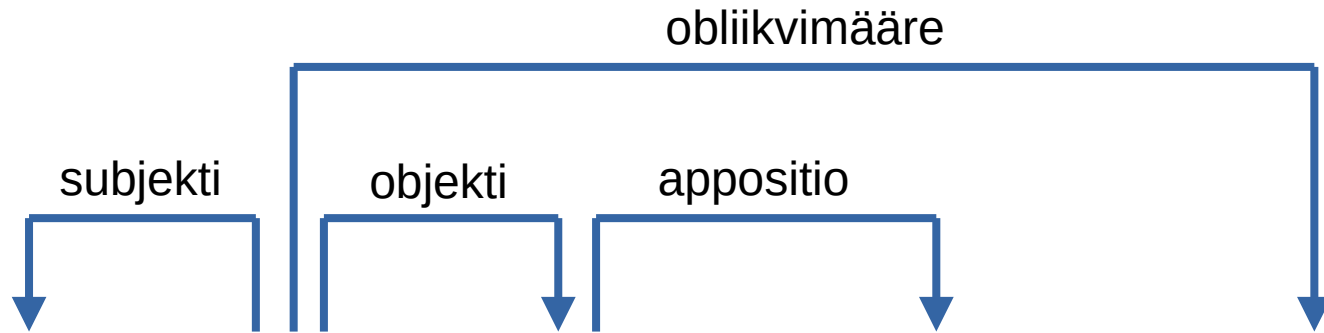
Ambiguity

Article [Talk](#)

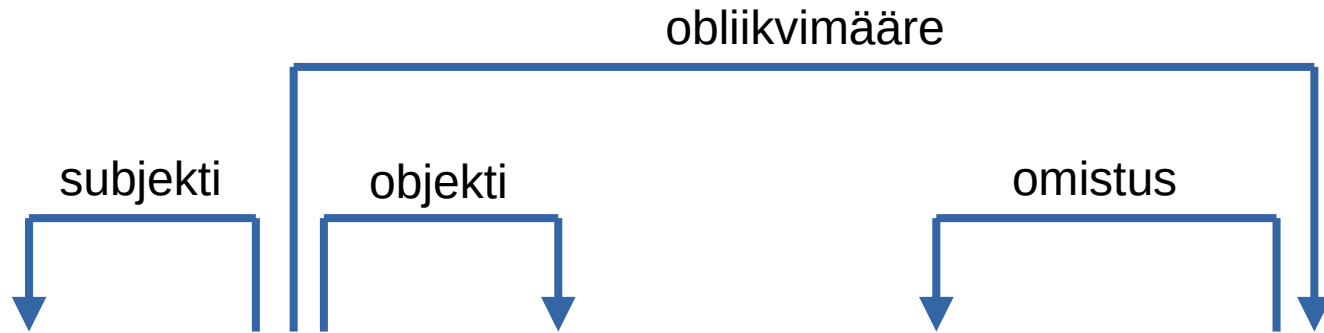
From Wikipedia, the free encyclopedia

For other uses, see [Ambiguity \(disambiguation\)](#).

Hän ilmoitti koiransa Mustin kilpailuun.

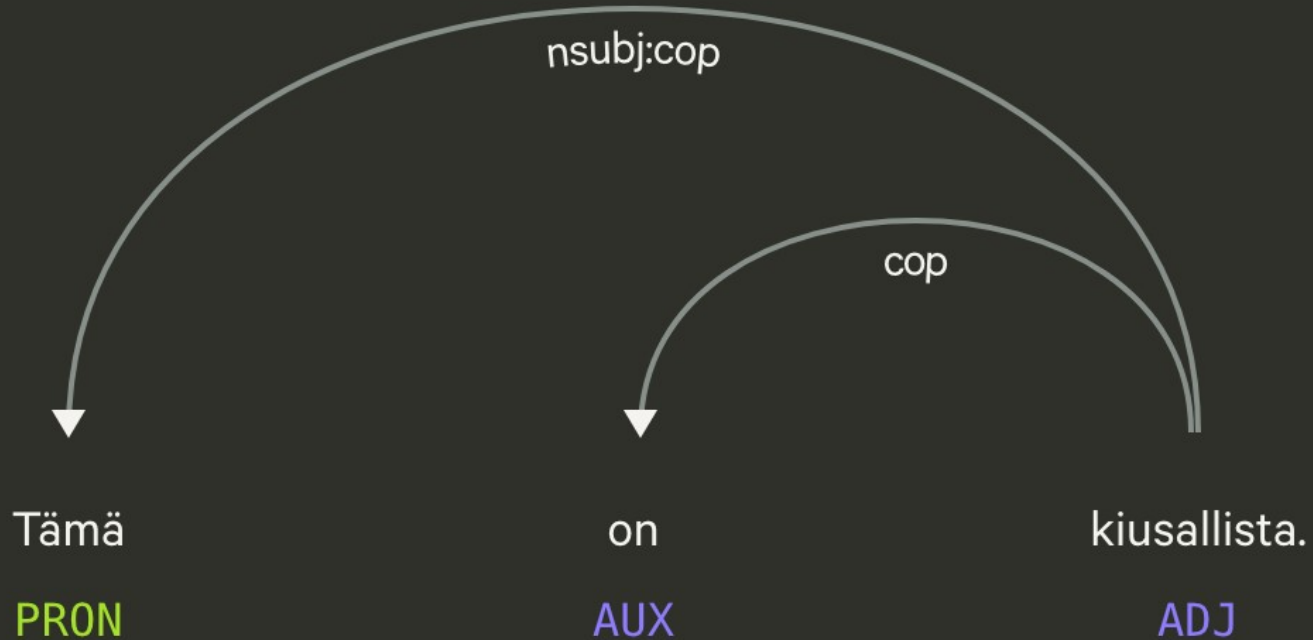


Hän ilmoitti koiransa Mustin kilpailuun.



Hän ilmoitti koiransa Mustin kilpailuun.

Predikatiivi



Predikatiivi

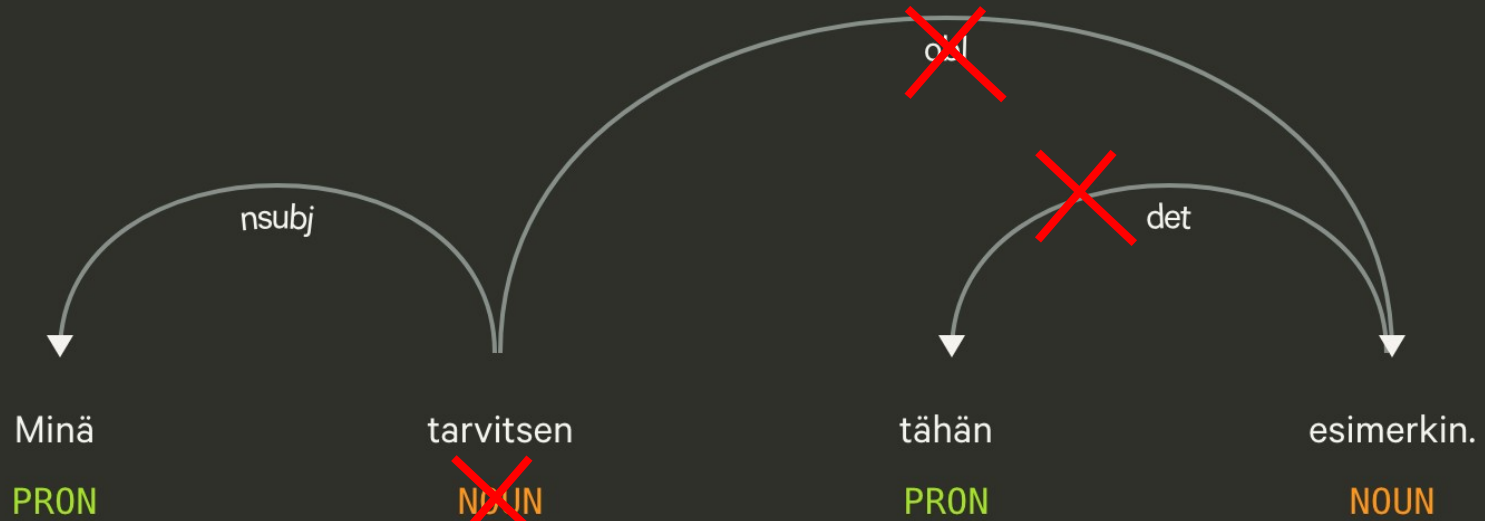


Mitä täl kaikel tekee :D?

- Sanaluokkien tunnistaminen osana jotakin suurempaa.
- Jäsennetty lause sisältää tarkempaa merkitystä.
- Ennen: konekäännökset
- Virkerakenteiden oikoluvussa.
- Rakenteisen tiedon louhinta

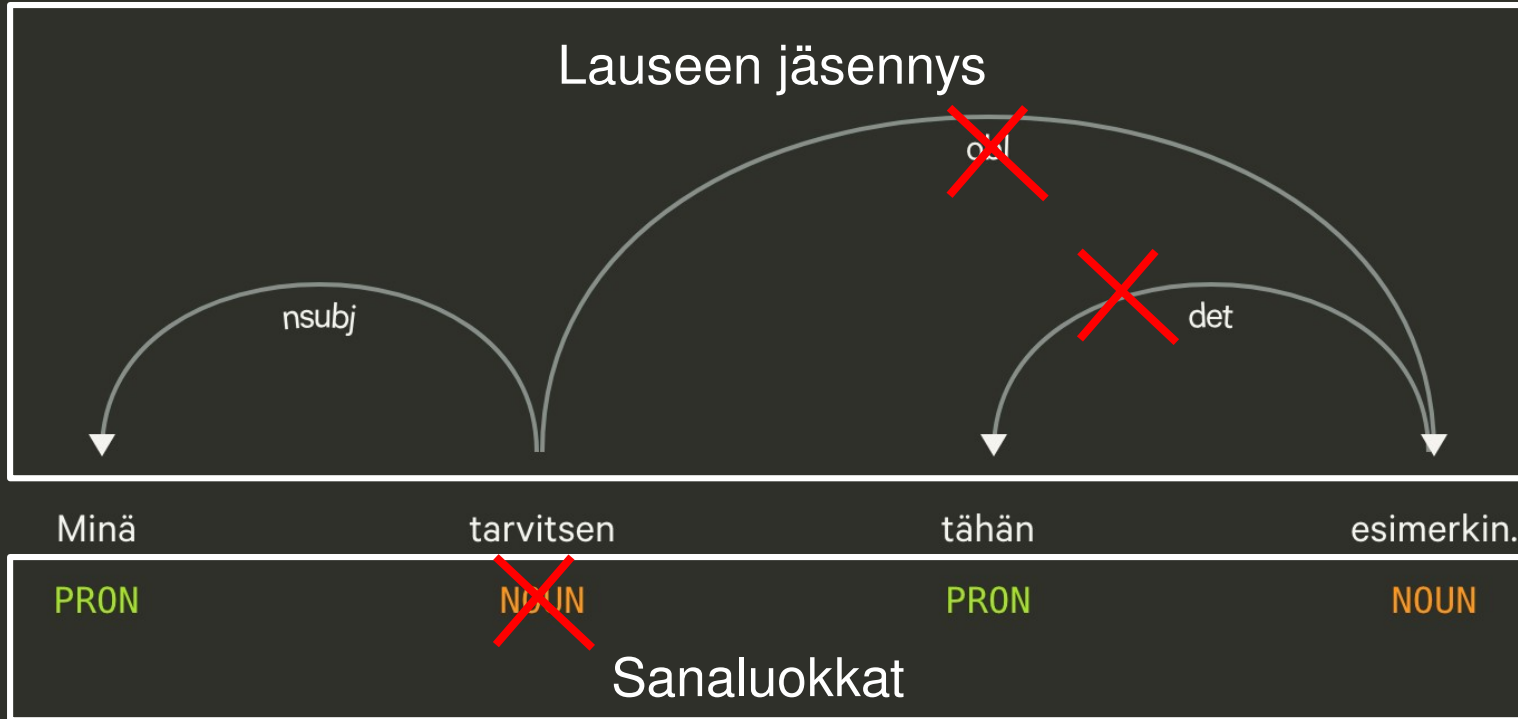
Tarkista, puuttuuko virkkeestä
pääverbi tai -verbejä

Lauseen jäsennys



Sanaluokat

Maksimaalinen virittävä puu



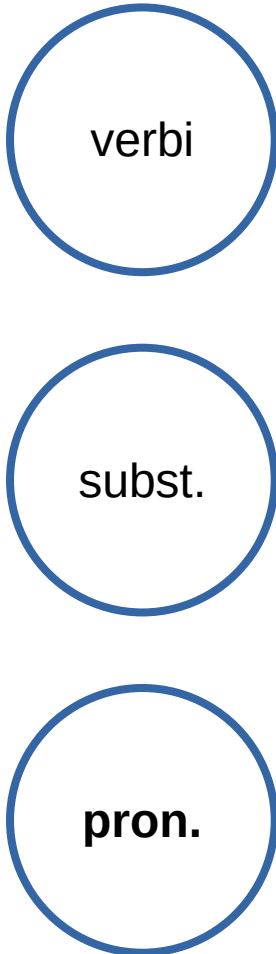
Markovin piilomalli

Näytös 2/3: Markovin piilomalli



Markovin piilomalli

- Käytetään jonon jäsenille merkintöjen antamiseen
- Ajatus on, että merkittävän jonon tuottaa järjestelmä, joka käy läpi jonon tiloja.
- Nämä tilat ovat mallin ”piilossa” oleva osa: tunnemme merkittävän jonon, mutta emme merkintöjä.
- Siirtymisillä tilasta toiseen on todennäköisyydet
- Eri tiloissa on eri todennäköisyydet tuottaa jonon eri jäsenet



verbi

subst.

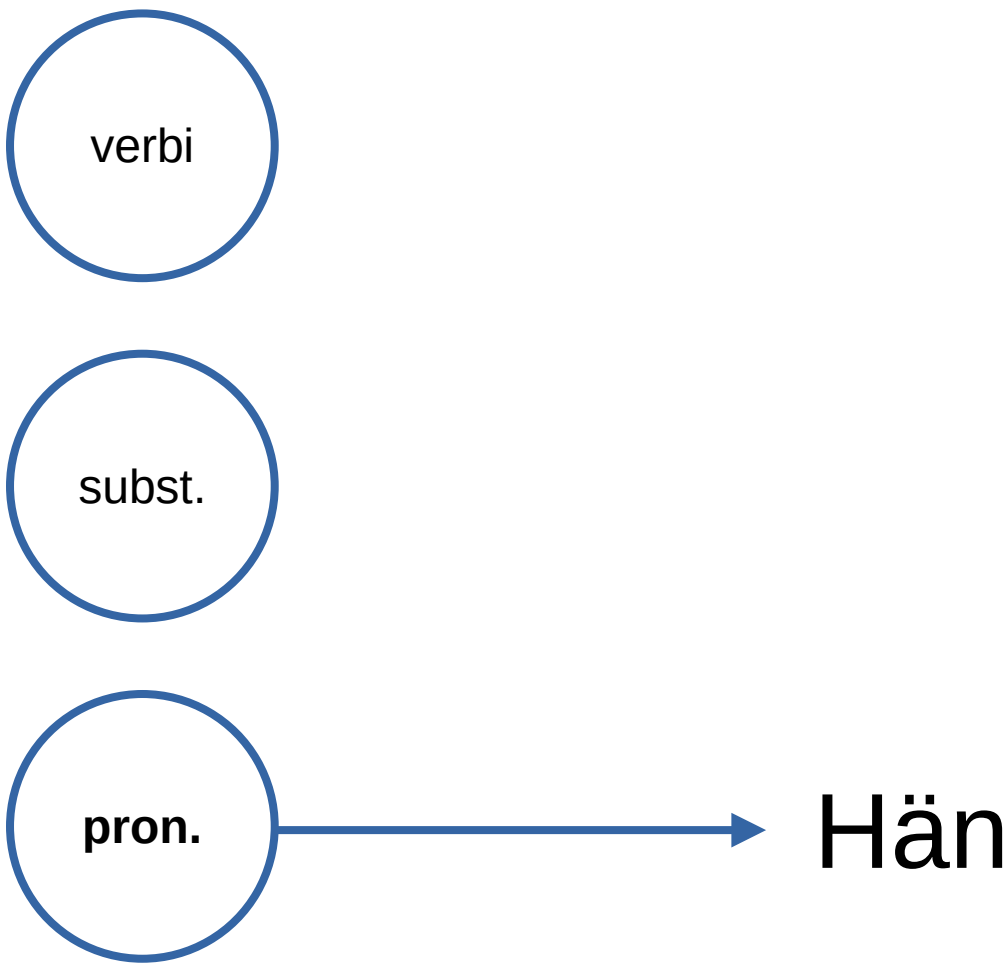
pron.

verbi

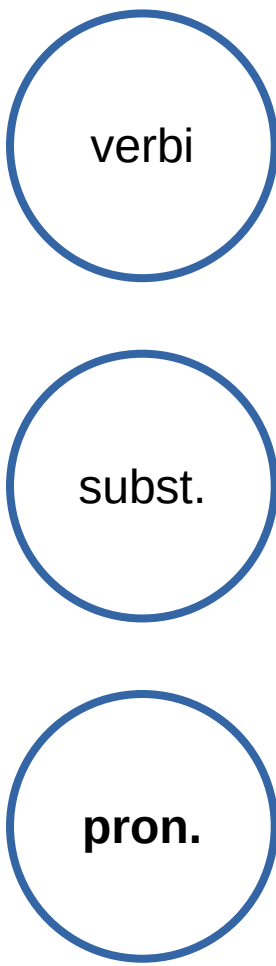
subst.

pron.

Hän



```
graph LR; A((verbi)); B((subst.)); C((pron.)) --> D[Hän];
```

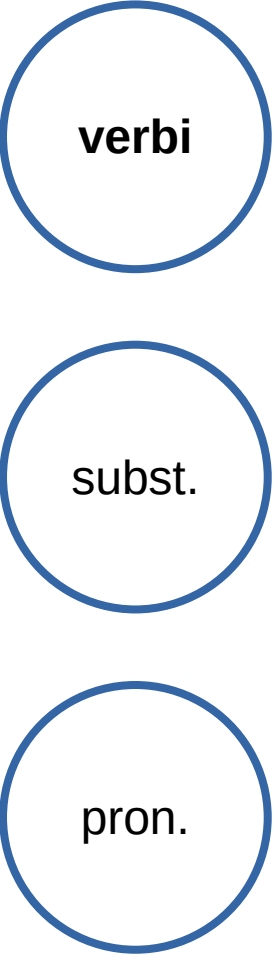


verbi

subst.

pron.

Hän



verbi

subst.

pron.

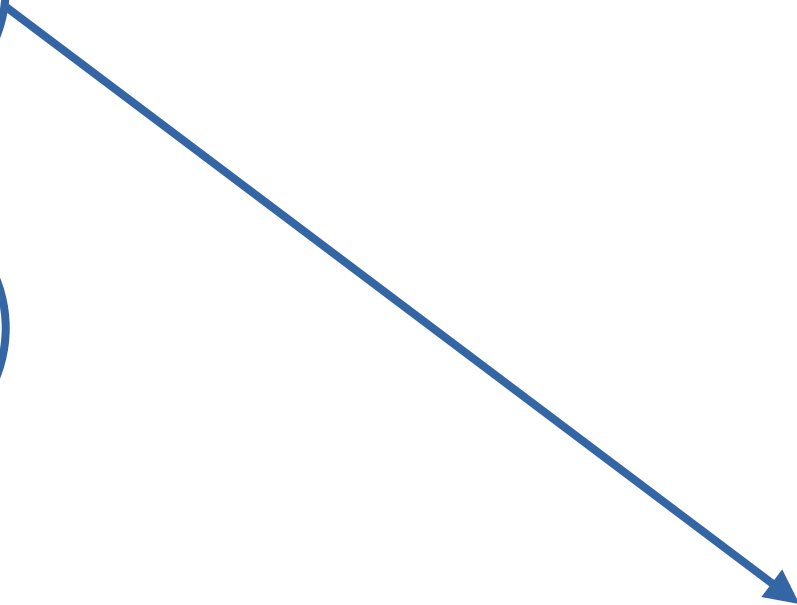
Hän

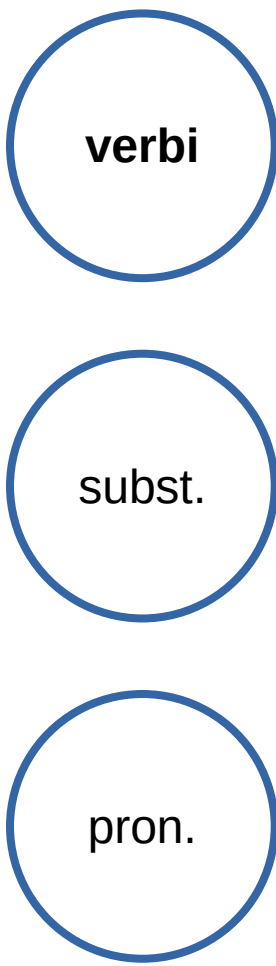
verbi

subst.

pron.

Hän on



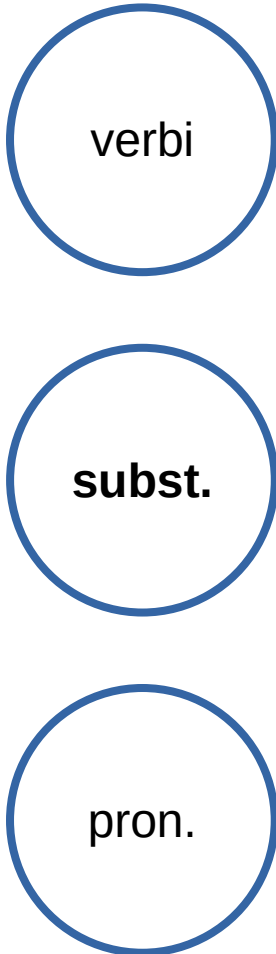


verbi

subst.

pron.

Hän on



verbi

subst.

pron.

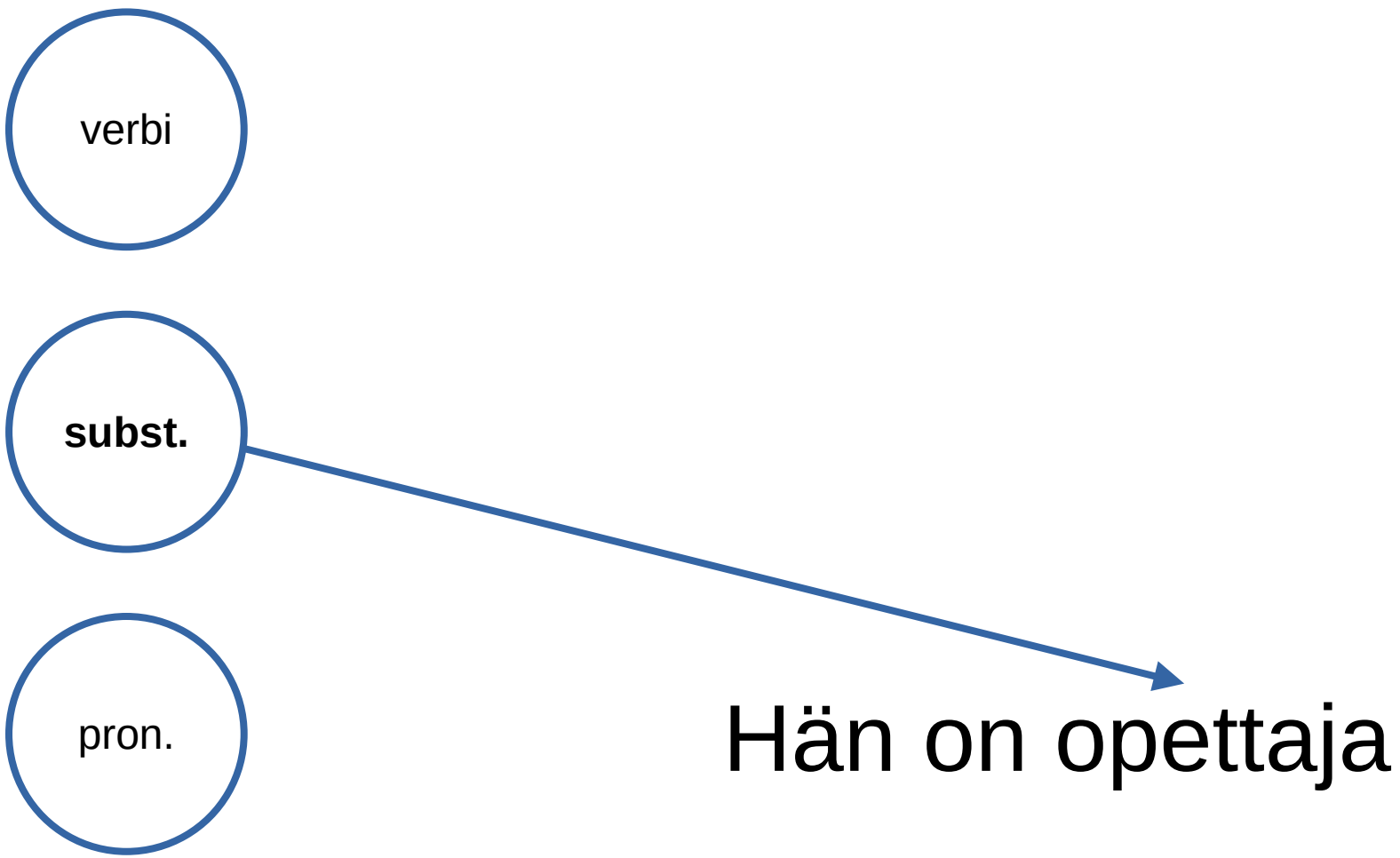
Hän on

verbi

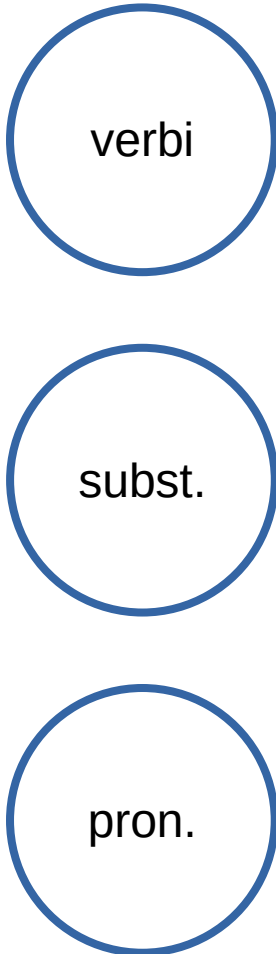
subst.

pron.

Hän on opettaja



```
graph LR; A((verbi)); B((subst.)) --> C[Hän on opettaja]; D((pron.))
```



verbi

subst.

pron.

Hän on opettaja

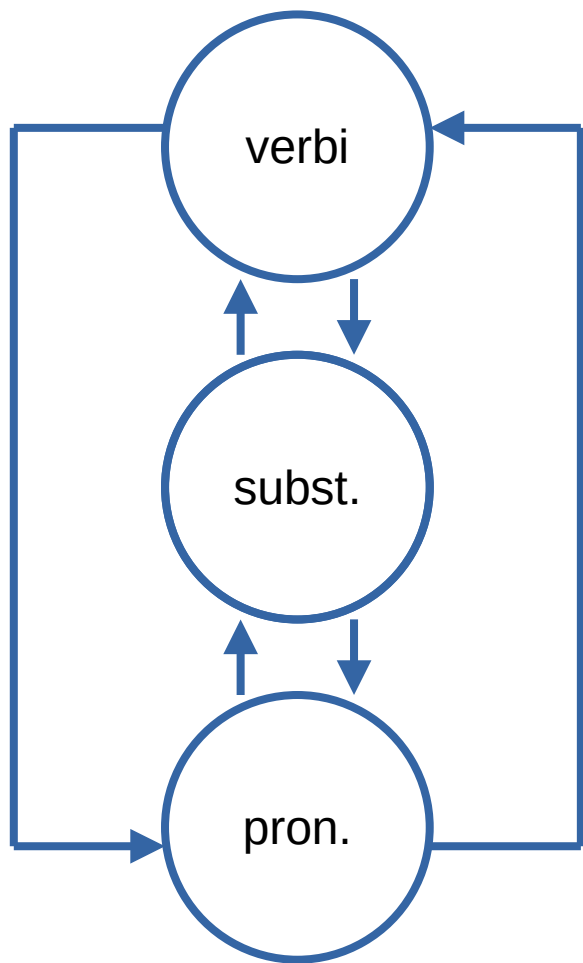
verbi

Mikä on todennäköisin jono tiloja
(sanaluokkia) havaituille
tapahtumille (sanoille)?

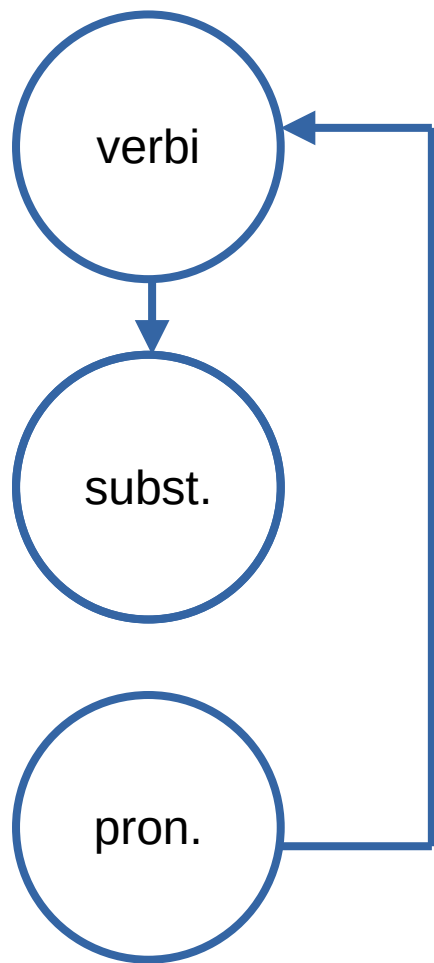
subst.

pron.

Hän on opettaja

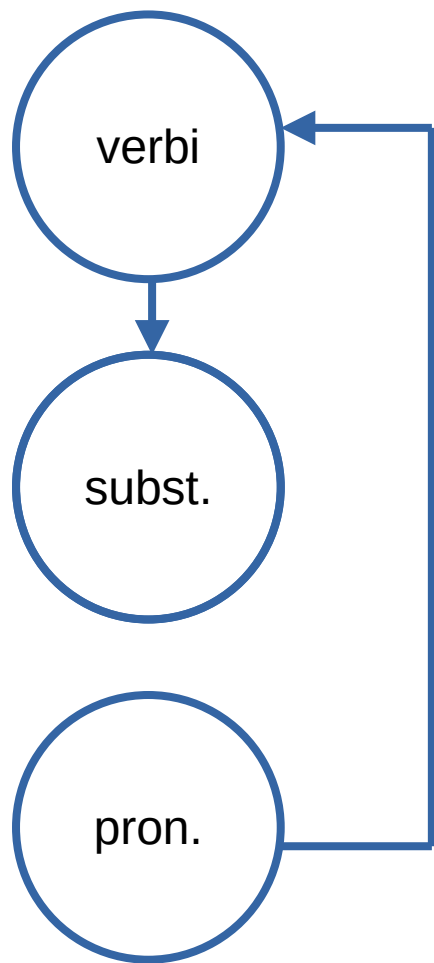


Jos hahmotamme tilat verkkona, niin voimme ajatella todennäköisimmän sarjan siirtymiä polkuna siinä.



Jos hahmotamme tilat verkkona, niin voimme ajatella todennäköisimmän sarjan siirtymiä polkuna siinä.

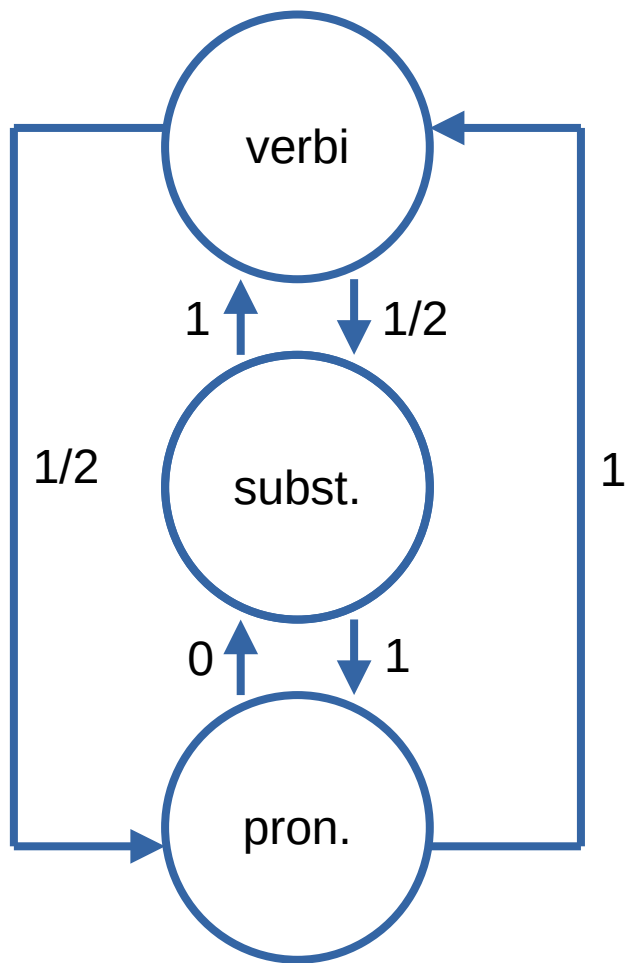
Tämä polku on nimeltään **Viterbin polku** ja se löytyy **Viterbin algoritmilla**.



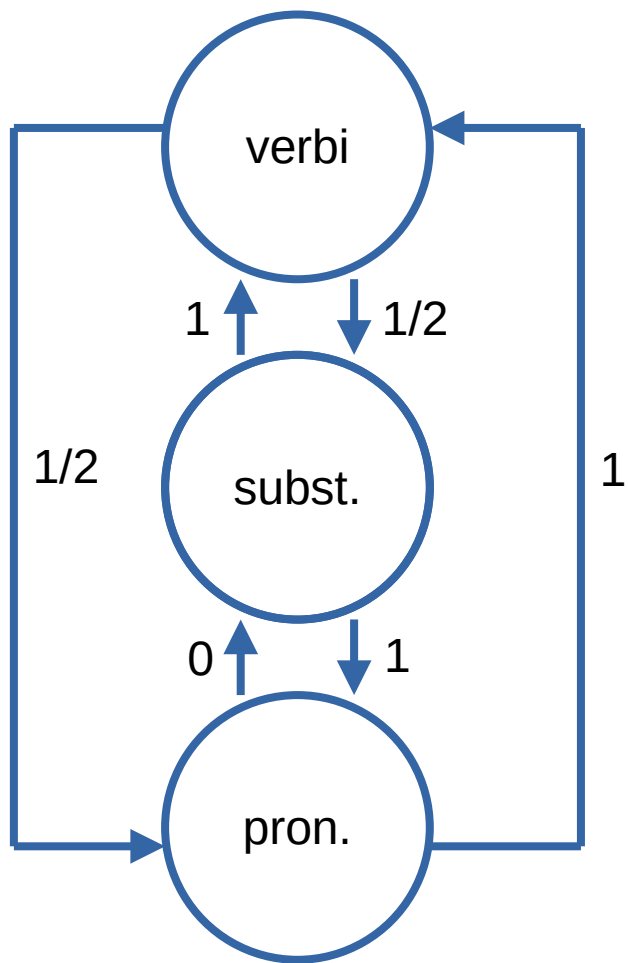
???

Jos hahmotamme tilat verkkona,  niin voimme ajatella **todennäköisimmän** sarjan siirtymiä polkuna siinä.

Tämä polku on nimeltään **Viterbin polku** ja se löytyy **Viterbin algoritmilla**.

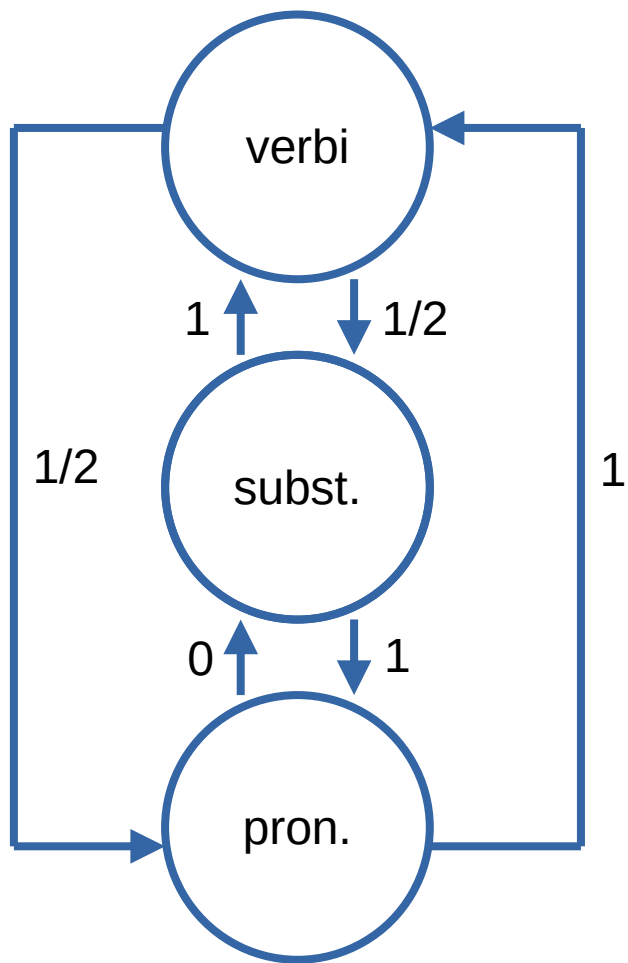


Siirtymätodennäköisyydet A: jokaisella kaarella paino



Havaintotodennäköisyydet B: jokaisella havainto tuntien tila -yhdistelmällä todennäköisyys

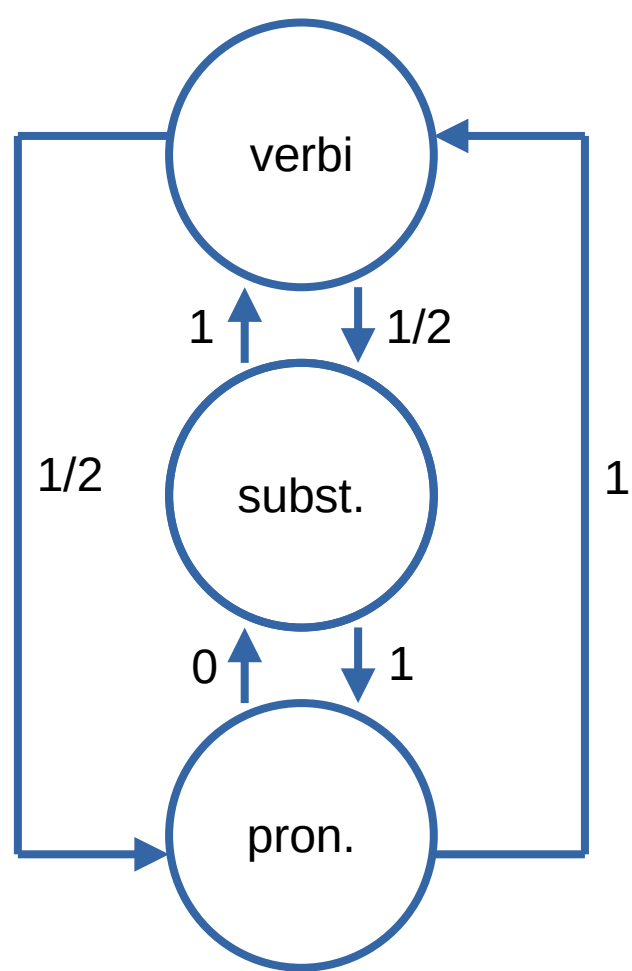
Verbi	korjaa	0.3
Verbi	olen	0.7
Substantiivi	opettaja	0.5
Substantiivi	heinä	0.5
Pronomini	heinä	0.2
Pronomini	minä	0.8



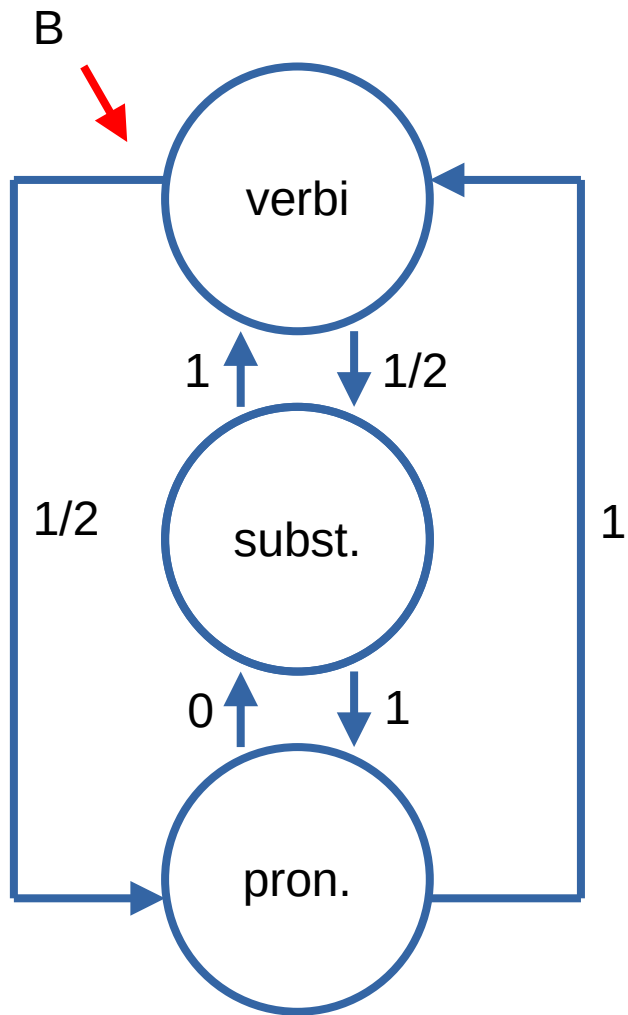
Havaintotodennäköisyydet B: jokaisella havainto tuntien tila -yhdistelmällä todennäköisyys

Verbi	korjaa	0.3	} Yht. 1
Verbi	olen	0.7	
Substantiivi	opettaja	0.5	} Yht. 1
Substantiivi	heinä	0.5	
Pronomini	heinä	0.2	} Yht. 1
Pronomini	minä	0.8	

Aloitustodennäköisyydet: todennäköisyys aloittaa tietyistä tilasta



Verbi	Pronomini	Substantiivi
0	0.5	0.5
Verbi	korjaa	0.3
Verbi	olen	0.7
Substantiivi	opettaja	0.5
Substantiivi	heinä	0.5
Pronomini	heinä	0.2
Pronomini	minä	0.8



Verbi	Pronomini	Substantiivi
0	0.5	0.5
Verbi	korjaa	0.3
Verbi	olen	0.7
Substantiivi	opettaja	0.5
Substantiivi	heinä	0.5
Pronomini	heinä	0.2
Pronomini	minä	0.8

← π

← A

Mitäs me näillä?

Mitäs me näillä?

Minä olen opettaja

Pronomini

Verbi

Substantiivi

Mitäs me näillä?

$\pi(\text{Pron})$

Minä olen opettaja

Pronomini

Verbi

Substantiivi



Mitäs me näillä?

$\pi(\text{Pron}) * A(\text{Minä} | \text{Pron})$

Minä olen opettaja

Pronomini

Verbi

Substantiivi



Mitäs me näillä?

$\pi(\text{Pron})^*A(\text{Minä}|\text{Pron})^*B(\text{Pron} \rightarrow \text{Verbi})$

Minä olen opettaja

Pronomini

Verbi

Substantiivi

Mitäs me näillä?

$\pi(\text{Pron})^*A(\text{Minä}|\text{Pron})^*B(\text{Pron} \rightarrow \text{Verbi})^*A(\text{olen}|\text{Verbi})$

Minä olen opettaja

Pronomini

Verbi

Substantiivi

Mitäs me näillä?

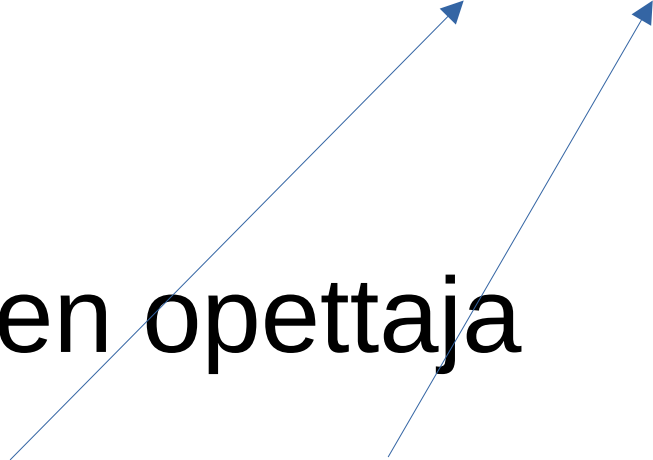
$\pi(\text{Pron}) * A(\text{Minä} | \text{Pron}) * B(\text{Pron} \rightarrow \text{Verbi}) * A(\text{olen} | \text{Verbi}) * B(\text{Verbi} \rightarrow \text{Subst})$

Minä olen opettaja

Pronomini

Verbi

Substantiivi



Mitäs me näillä?

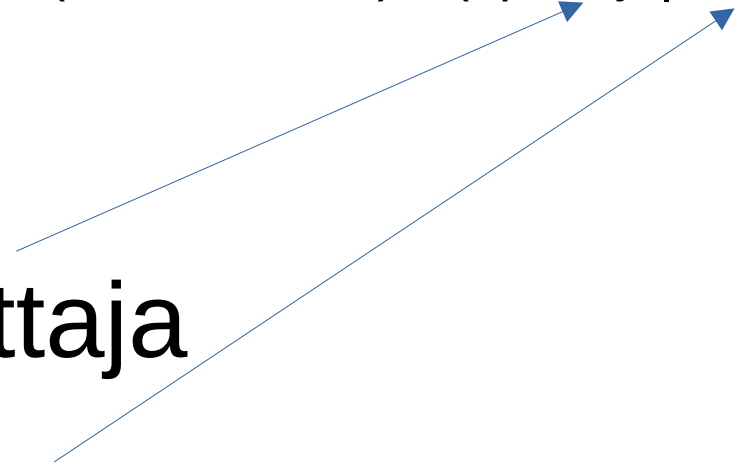
$\pi(\text{Pron}) * A(\text{Minä} | \text{Pron}) * B(\text{Pron} \rightarrow \text{Verbi}) * A(\text{olen} | \text{Verbi}) * B(\text{Verbi} \rightarrow \text{Subst}) * A(\text{opettaja} | \text{Subst})$

Minä olen opettaja

Pronomini

Verbi

Substantiivi



Mitäs me näillä?

$\pi(\text{Pron}) * A(\text{Minä} | \text{Pron}) * B(\text{Pron} \rightarrow \text{Verbi}) * A(\text{olen} | \text{Verbi}) * B(\text{Verbi} \rightarrow \text{Subst}) * A(\text{opettaja} | \text{Subst})$

Minä olen opettaja

Pronomini

Verbi

Substantiivi

Mitäs me näillä?

$\pi(\text{Pron}) * A(\text{Minä} | \text{Pron}) * B(\text{Pron} \rightarrow \text{Verbi}) * A(\text{olen} | \text{Verbi}) * B(\text{Verbi} \rightarrow \text{Subst}) * A(\text{opettaja} | \text{Subst})$

Osataan laskea polun todennäköisyys, mutta miten löydetään paras polku?

Minä olen opettaja

Pronomini

Verbi

Substantiivi

Ahne menettely

Valitaan aina todennäköisin jatke:

Pyörällä

Ahne menettely

Valitaan aina todennäköisin jatke:

Pyörällä

substantiivi

Ahne menettely

Valitaan aina todennäköisin jatke:

Pyörällä päästään

substantiivi

Ahne menettely

Valitaan aina todennäköisin jatke:

Pyörällä päästään

substantiivi

substantiivi

Ahne menettely

Valitaan aina todennäköisin jatke:

Pyörällä päästään perille

substantiivi

substantiivi

Ahne menettely

Valitaan aina todennäköisin jatke:

Pyörällä päästään perille

substantiivi

substantiivi

???

Viterbin algoritmi (DP)

Ylläpidetään tieto todennäköisimmästä tavasta, millä tietty sana on tiettyä luokkaa.

Pyörällä päästään perille

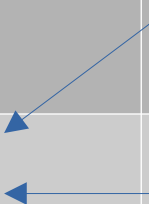
Verbi	0.1		
Substantiivi	0.9		

Viterbin algoritmi (DP)

Ylläpidetään tieto todennäköisimmästä tavasta, millä tietty sana on tiettyä luokkaa.

Pyörällä päästään perille

Verbi	0.1	0.5	
Substantiivi	0.9	0.7	



Viterbin algoritmi (DP)

Ylläpidetään tieto todennäköisimmästä tavasta, millä tietty sana on tiettyä luokkaa.

Pyörällä päästään perille

Verbi	0.1	0.5	0.4
Substantiivi	0.9	0.7	0.6

The diagram illustrates the Viterbi algorithm's state transitions. It consists of a 2x3 grid of cells. The top row is labeled 'Verbi' and the bottom row is labeled 'Substantiivi'. The columns contain numerical values: 0.1, 0.5, and 0.4 in the top row; 0.9, 0.7, and 0.6 in the bottom row. Blue arrows indicate the most probable path from right to left. An arrow points from the 0.4 cell to the 0.5 cell. Another arrow points from the 0.6 cell to the 0.7 cell. A third arrow points from the 0.5 cell to the 0.9 cell. A final arrow points from the 0.7 cell to the 0.1 cell.

Viterbin algoritmi (DP)

Ylläpidetään tieto todennäköisimmästä tavasta, millä tietty sana on tiettyä luokkaa.

Pyörällä päästään perille

Verbi	0.1	0.5	0.4
Substantiivi	0.9	0.7	0.6

The diagram illustrates the Viterbi algorithm's state transitions. It consists of a 2x3 grid of cells. The first row is labeled 'Verbi' and the second row is labeled 'Substantiivi'. The columns contain numerical values representing probabilities. Blue arrows indicate the most likely path from right to left. An arrow points from the '0.4' cell (Verbi, Column 3) to the '0.5' cell (Verbi, Column 2). Another arrow points from the '0.6' cell (Substantiivi, Column 3) to the '0.7' cell (Substantiivi, Column 2). A third arrow points from the '0.5' cell (Verbi, Column 2) to the '0.9' cell (Substantiivi, Column 1). A fourth arrow points from the '0.7' cell (Substantiivi, Column 2) to the '0.9' cell (Substantiivi, Column 1).

State	Column 1	Column 2	Column 3
Verbi	0.1	0.5	0.4
Substantiivi	0.9	0.7	0.6

Tosielämä kohtaa Markovin piilomallin

- Surkea yleistämään uusiin sanoihin
- Aloitustodennäköisyydet
- Ei Viterbilläkään kovin voimakas

Surkea yleistämään

- Ihminen pystyy arvaamaan sanaluokkia:
Ähmyrä suhmuroi kähmyrän. → S V S
- Mitä tehdään uusille sanoille: __harvinainen__
- Hienostuneet versiot

Aloitustodennäköisyydet

- Usein aloitustodennäköisyyksien laskemista helpotetaan niin, että kaikki virkkeet alkavat ja loppuvat erikoismerkillä

Ahne menettely

Valitaan aina todennäköisin jatke:

Muista

Ahne menettely

Valitaan aina todennäköisin jatke:

Muista

verbi

Ahne menettely

Valitaan aina todennäköisin jatke:

Muista kymmenen

verbi

Ahne menettely

Valitaan aina todennäköisin jatke:

Muista kymmenen

verbi

numeraali

Ahne menettely

Valitaan aina todennäköisin jatke:

Muista kymmenen käskyä

verbi

numeraali

Ahne menettely

Valitaan aina todennäköisin jatke:

Muista kymmenen käskyä

verbi

numeraali

substantiivi

Ahne menettely

Valitaan aina todennäköisin jatke:

Muista kymmenen käskyä tuntuivat

verbi

numeraali

substantiivi

???

Ahne menettely

Valitaan aina todennäköisin jatke:

Muista kymmenen käskyä tuntuivat turhilta.

verbi

numeraali

substantiivi



Uudempi tieto kumoaa vanhan

Ahne menettely

Valitaan aina todennäköisin jatke:

Viterbi ei pelastaisi, koska virhe ehti jo hautautua

Muista kymmenen käskyä tuntuivat turhilta.

verbi

numeraali

substantiivi

Uudempi tieto kumoaa vanhan

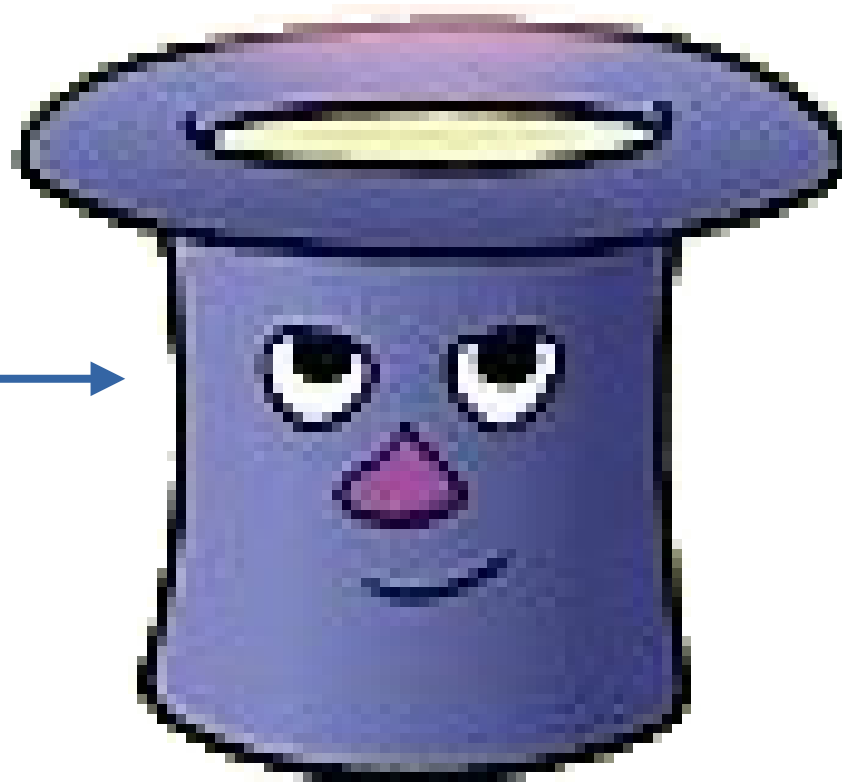


Näytös 3/3: Chu-Liu / Edmondsin algoritmi

Jäsentäminen

Minä laadin kalvoja

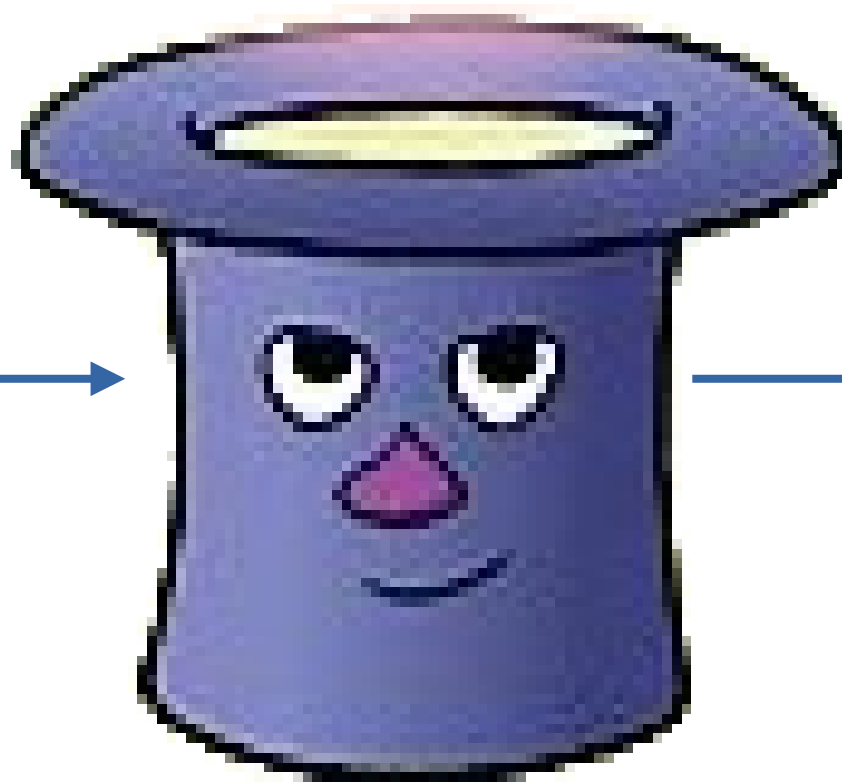
Jäsentäminen



Minä laadin kalvoja

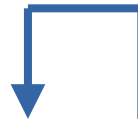
Jäsentäminen

Minä laadin kalvoja



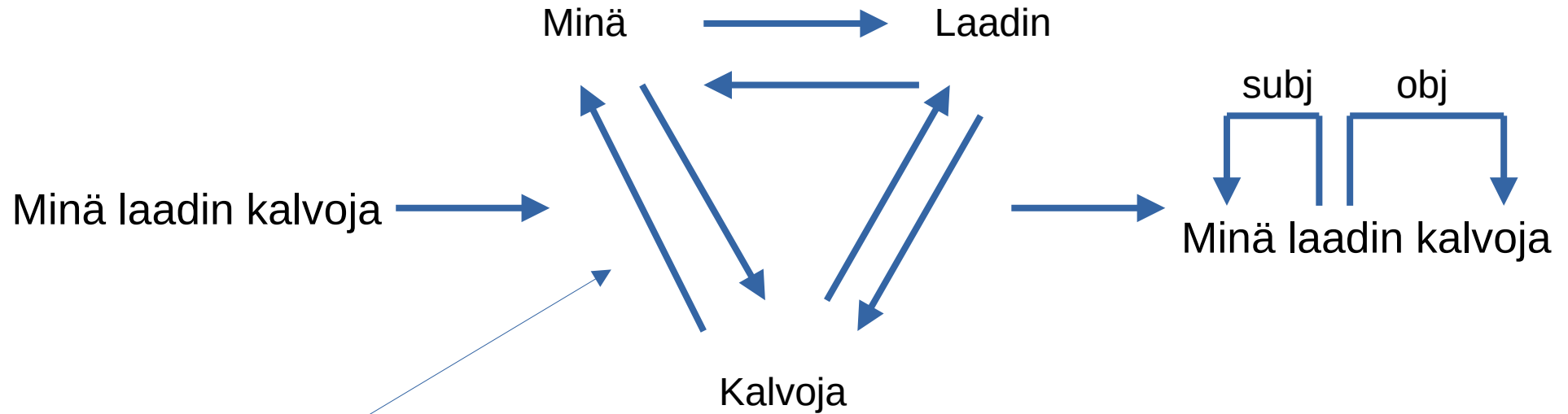
subj

obj

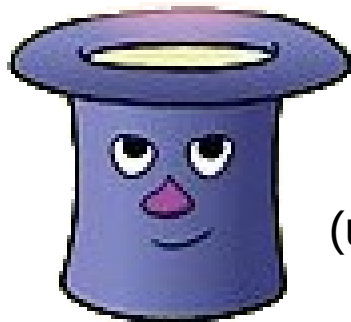


Minä laadin kalvoja

Jäsentäminen

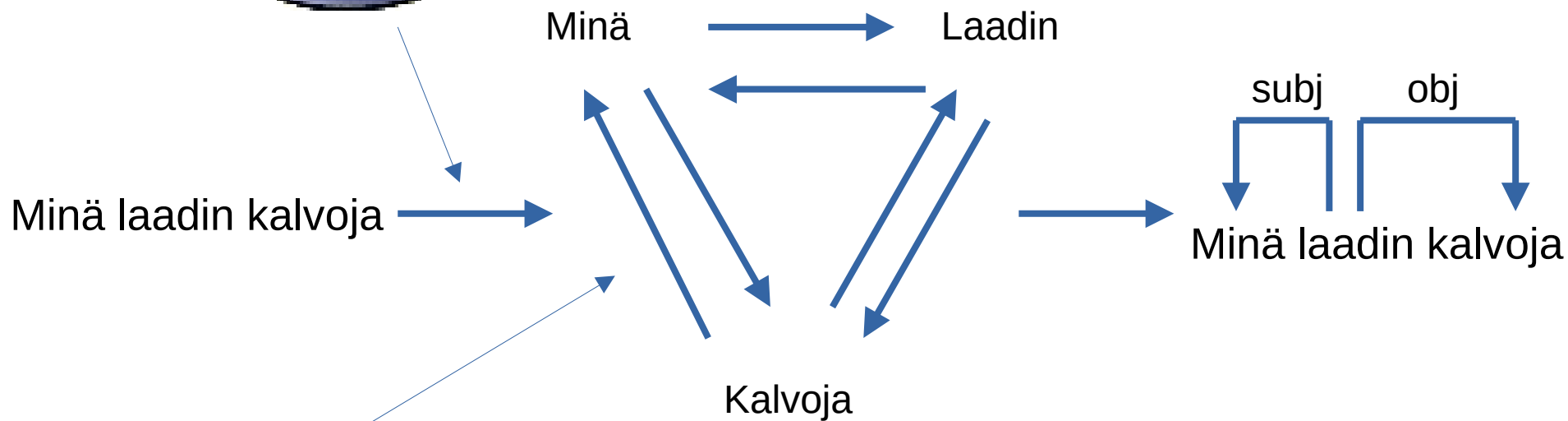


Verkko, jossa on kaikki mahdolliset yhteydet ja laabelit painotettuina

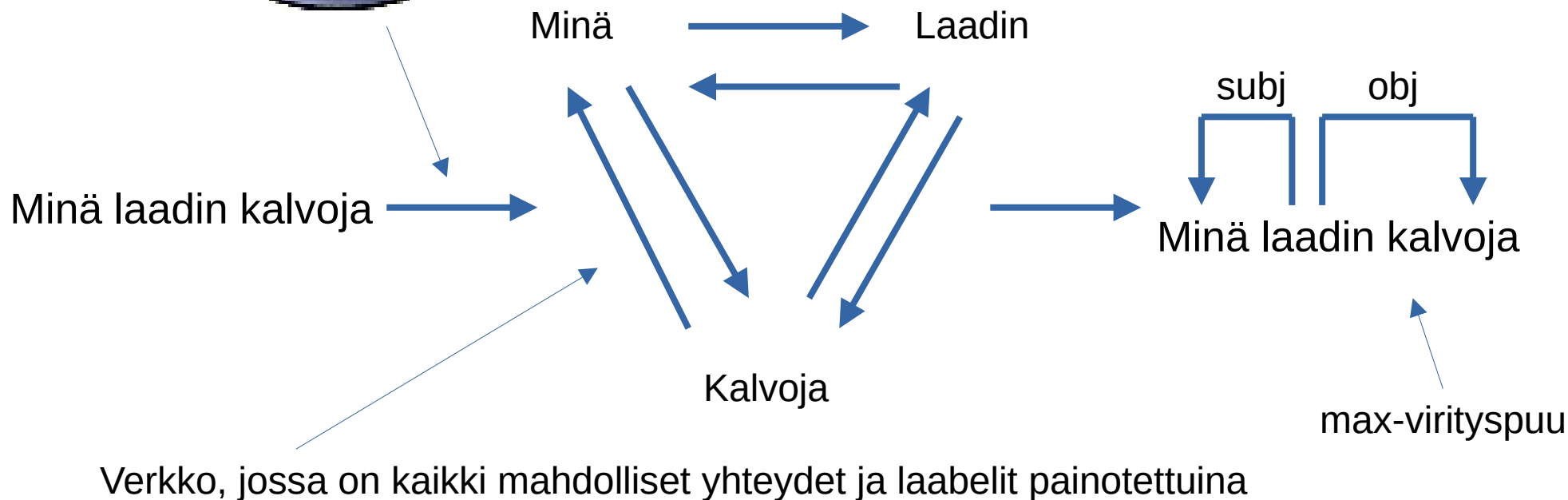


Jäsentäminen

(usein neuroverkko)



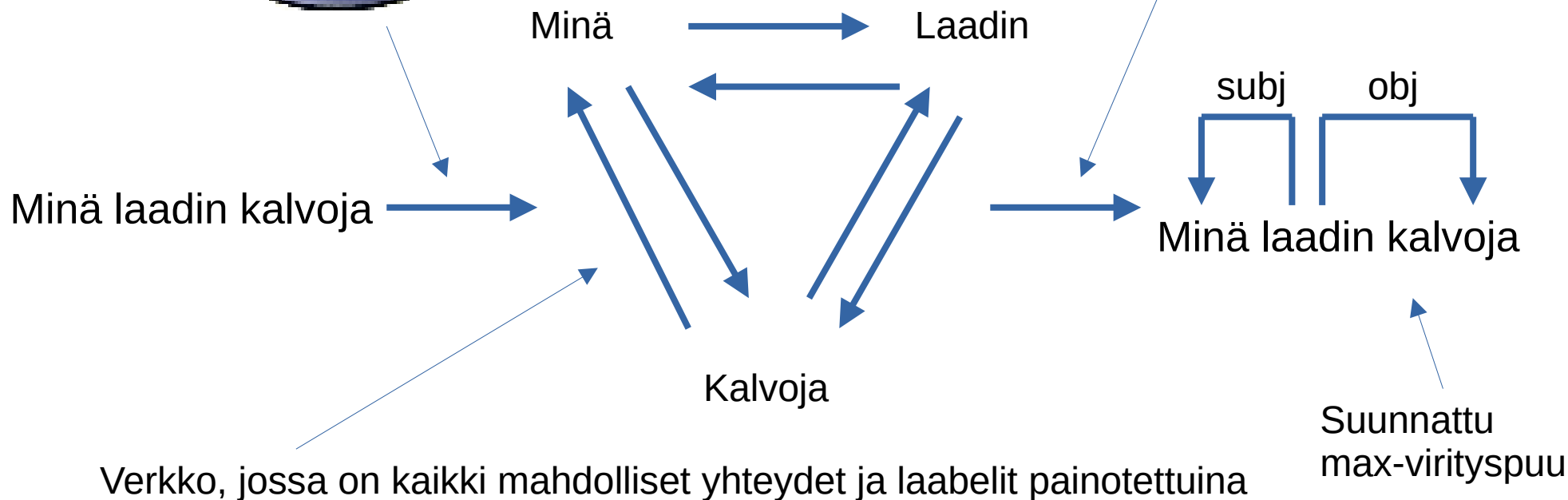
Verkko, jossa on kaikki mahdolliset yhteydet ja laabelit painotettuina





Jäsentäminen

Chu-Liu / Edmondsin algoritmi



Suunnattu max-virityspuu?

- Suunnattu puu, joka:
 - Sisältää kaikki verkon solmut (virittävyys)
 - On tällaisista puista kaikkein suuripainoisin (max)

Roope on hyvä ja jatkaa tästä