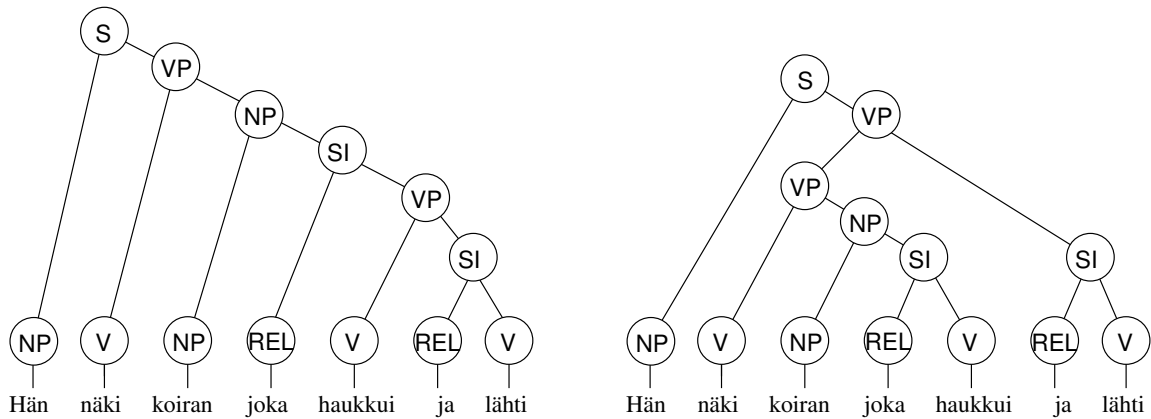


T-61.281 Luonnollisen kielen tilastollinen käsittely

Harjoitus 8, ti 18.3.2003, 16:15-18:00 Tilastolliset yhteydetön kieliopit, Versio 1.0

1. Laske kuvassa 1 olevien jäsenyyspuiden todennäköisyydet käyttäen taulukossa 1 olevia sääntöjä.



Kuva 1: Jäsennykset

S	→	NP VP	1.0	NP	→	hän	0.3
PP	→	A NP	1.0	NP	→	tuulen	0.15
VP	→	V NP	0.7	NP	→	kasvoillaan	0.15
VP	→	VP PP	0.2	NP	→	koiran	0.1
VP	→	VP SI	0.08	V	→	näki	0.3
VP	→	V SI	0.02	V	→	tunsi	0.29
SI	→	REL V	0.5	V	→	haukkui	0.1
SI	→	REL VP	0.5	V	→	tuulen	0.01
NP	→	NP SI	0.1	V	→	lähti	0.3
NP	→	NP PP	0.15	A	→	kalpeilla	1.0
				REL	→	joka	0.5
				REL	→	ja	0.5

Taulukko 1: Tilastollinen yhteydetön kielioppi

2. Laske lauseen “Hän tunsi tuulen kalpeilla kasvoillaan” todennäköisyys, kun mallina on taulukon 1 säännöt. Todennäköisyyden ratkaiseminen sujunee helpoimmin sisäpuoli-algoritmillä (inside algorithm). Voit samalla myös helposti pitää kirjaa siitä, mikä on todennäköisin jäsenyyspuu tälle lauseelle.

3. Tarkastellaan kielioppia, missä on kaksi sääntöä:

$$\begin{aligned}P(S \rightarrow w) &= 1 - p \\P(S \rightarrow S S) &= p\end{aligned}$$

S on aloitussymboli ja w on terminaalisyntoli. On mahdollista, että joillain p :n arvoilla kielimalli ei ole oikea todennäköisyysfunktio, koska osa todennäköisyydestä häviää äärettömien puiden generointiin. Millä p :n arvoilla annettu kielioppi antaa oikean todennäköisyysjakauman eli $P(w) + P(w w) + P(w w w) + \dots = 1$.