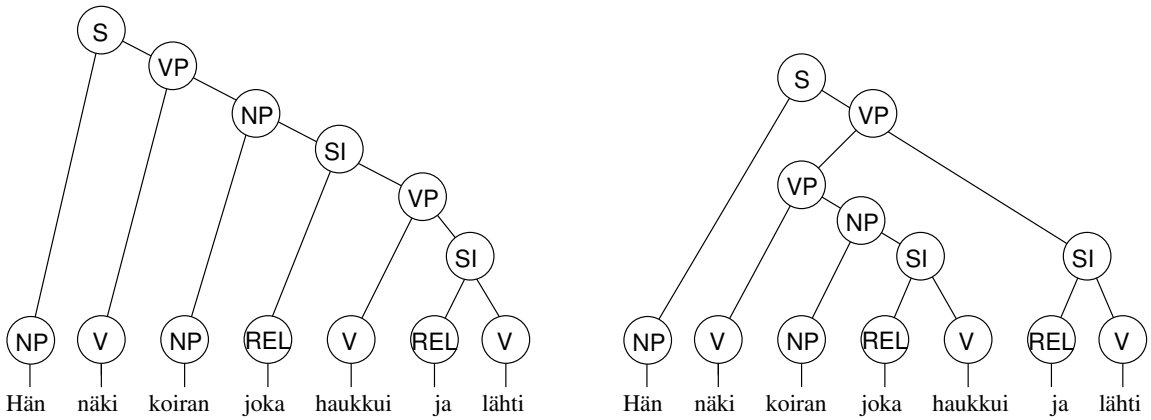
# T-61.281 Luonnollisen kielen tilastollinen käsittely

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

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




Kuva 1: Jäsennykset

| | | | |
|---|---|---|---|
| S | → | NP VP | 1.0 |
| PP | → | A NP | 1.0 |
| VP | → | V NP | 0.7 |
| VP | → | VP PP | 0.2 |
| VP | → | VP SI | 0.08 |
| VP | → | V SI | 0.02 |
| SI | → | REL V | 0.5 |
| SI | → | REL VP | 0.5 |
| NP | → | NP SI | 0.1 |
| NP | → | NP PP | 0.15 |

| | | | |
|---|---|---|---|
| NP | → | hän | 0.3 |
| NP | → | tuulen | 0.15 |
| NP | → | kasvoillaan | 0.15 |
| NP | → | koiran | 0.1 |
| V | → | näki | 0.3 |
| V | → | tunsi | 0.29 |
| V | → | haukkui | 0.1 |
| V | → | tuulen | 0.01 |
| V | → | lähti | 0.3 |
| 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-algoritmilla (inside algorithm). Voit samalla myös helposti pitää kirjaa siitä, mikä on todennäköisin jäsennyspuu tälle lauseelle.

3. Tarkastellaan kielioppia, missä on kaksi sääntöä:

$$P(S \rightarrow w) = 1 - p$$
$$P(S \rightarrow S\ S) = p$$

$S$ on aloitussymboli ja $w$ on terminaalisymboli. 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 $P(w) + P(w\ w) + P(w\ w\ w) + \cdots = 1$.