



# Johdatus uusiin luonnollisen kielen mallinnusmenetelmiin automaattisessa puheentunnistuksessa



**Mihin tarvitaan kielimallia?**





**Miksi?**



**Rahaa tarjolla?**



Puheentunnistimia **on käytössä** mm.

-yleiset puhelinpalvelut

-käyttöliittymät

-vammaisten apuvälineet

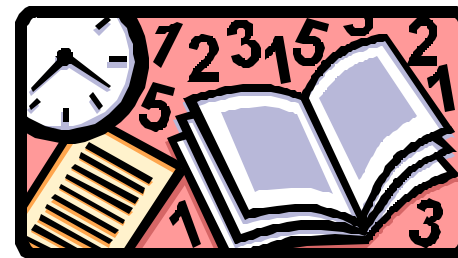
-puheen tallennus ja indeksointi

**Toimiiko** riittävän hyvin?



Miksi ei toimi?  $\Longrightarrow$  Onko liian vaikea...

1. **Signaali** (häiriötä, vaihtelevat olosuhteet)
2. **Puhe** (epäselvää, katkoja, puuttuvia foneemeja)
3. **Tehtävä** (paljon vaihtoehtoja, vaihtelevat aiheet)





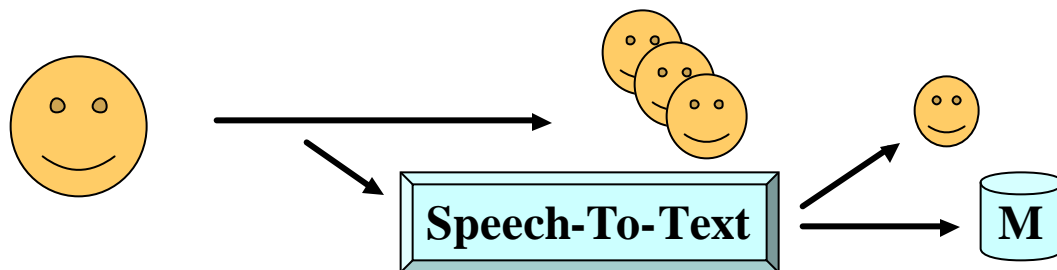
# Kielen mallinnus



Arvataan se mitä ei kuulla?

## Luonnollinen kieli?

- toisille ihmisille suunnattua (TV, radio, kokoukset)
- sanasto laaja ja vaihtuu aiheen mukaan
- **mallit** yleensä vain **tilastollisia**
- opetusaineistoa helposti saatavilla?





Mitä tekee kielimalli?



**(HS 3.9.2001 / Reuters-AP-STT)**  
**Kaksi palestiinalaista surmattiin Hebronissa**  
**Peresin ja Arafatin tapaamisen valmistelut jatkuivat**  
**...Israelin ulkoministeri (A?) kehotti sunnuntaina olemaan**  
**odottamatta liikoja hänen ja palestiinalaisten presidentin**  
**Jasser Arafatin (B?). ...**

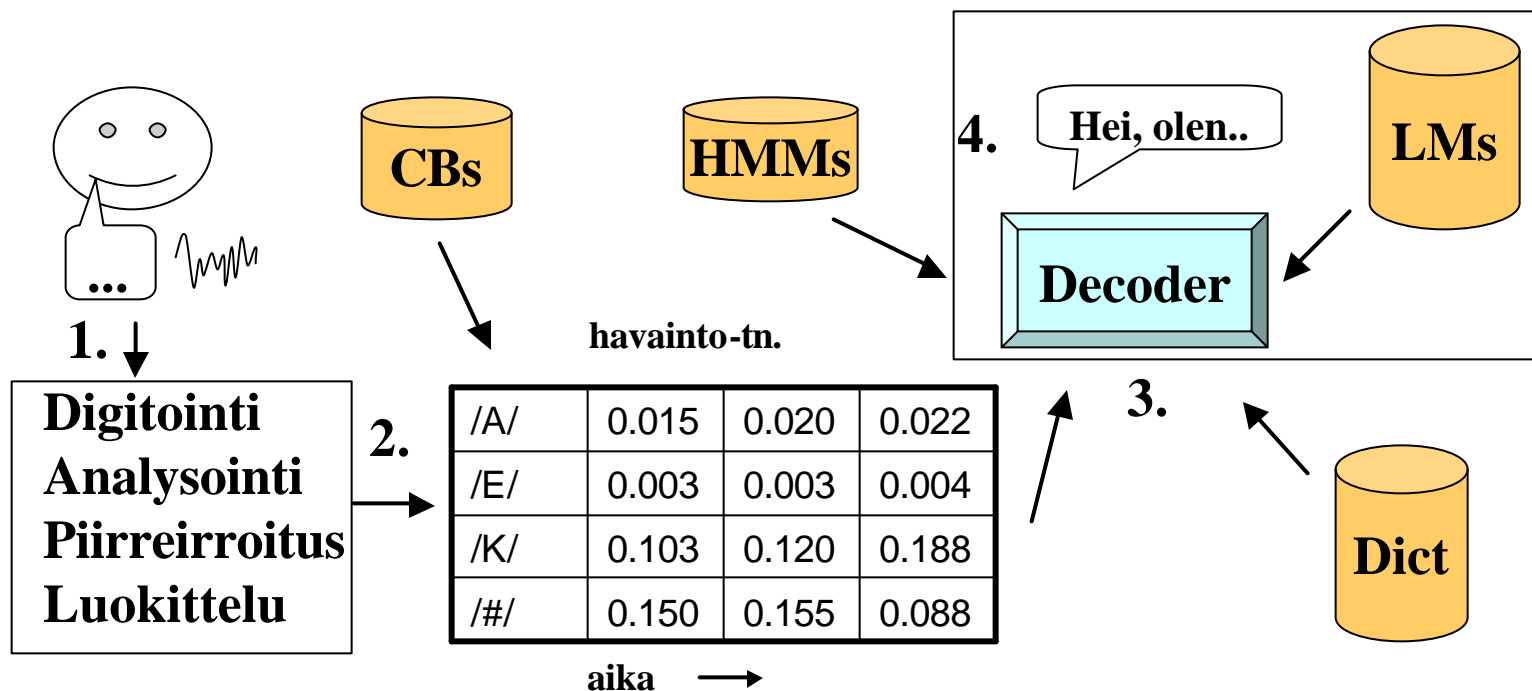
- A?**
- 1. Tampere**
  - 2. Simonperä**
  - 3. Shimon Peres**

- B?**
- 1. aavistetulta tappamiselta**
  - 2. valmis tulevalta aamiaiselta**
  - 3. valmisteilla olevalta tapaamiselta**



# Laaja sanasto, jatkuva puhe = LVCSR

⇒ Miten puheentunnistus toimii?





# Tilastolliset (ja oppivat) mallit



Ennustetaan **sanajonojen todennäköisyyksiä**:

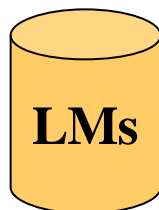
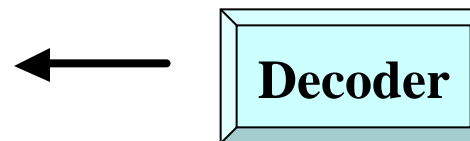
- $P(w(t), \dots, w(1) \mid \text{malli})$

**Seuraavan sanan tn. :**

- $P(w(t) \mid w(t-1), \dots, w(1), \text{malli})$

Yhdistettävissä akustisten mallien tn:iin (O):

- $\text{Max}_W P(O|W)P(W)$



+



+



+

/A/	0.015	0.020	0.022
/E/	0.003	0.003	0.004
/K/	0.103	0.120	0.188
/#/	0.150	0.155	0.088



# Tavallisia malleja



## 1. Word N-gram

$$P( w(t) | w(t-1), \dots, w(t-N+1) )$$

**N edellistä sanaa** huomioon

- + English, Easy, Estimation
- + Extensions
- sanasto, taivutukset
- sanajärjestys, sivulauseet
- data harvaa

## 2. Word cache

$$P( w(t) | w(t) = w(t-i) )$$

**Sanankäytön** aiempi **tiheys**

## 3. Topic model

$$P( w(t) | C(W) = c_k )$$

Joka **puheenaiheella**  
oma **sanajakaumansa**





# Mallien käyttö



Decoder

## 1. Viterbi –haku

- Haetaan **N parasta hypoteesia valmiiksi**
- Valitaan se joka parhaiten sopii kielimalleihin

	<b>1. Välikohtaus <u>heikensi</u> näkymiä entisestään.</b> <b>2. Välikohtaus <u>heikkeni</u> näkymiä entisestään.</b>

## 2. Stack decoder

- Säilytetään **pinossa osittaisia hypoteeseja**
- Laajennetaan aina parasta hypoteesia eteenpäin sana kerrallaan

	<b>Välikohtaus heikensi (35)</b> <b>Välikohtaus heikkeni (33)</b> <b>Välikohta usein (21)</b> <b>Valikko (7)</b>



Menetelmien parannuksia... →

**Yhdistelmämallit**

**Taivutusmuotojen  
käyttö**

**Automaattinen  
puheenaiheen  
valinta**

**yhdyssanat  
multi N-gram  
variable N-gram**

**Yleisemmät  
sanayhdistelmien  
ja aihealueiden  
esitystavat**

-Dr. Erik McDermott pitää vierailuesitelmän **to 13.9.**

-T-61.184 Informaatiotekniikan erikoiskurssi 4 (Lagus, Kurimo) **alkaa to 20.9.**