

Prosovar-hankkeen väliraportti

Puheaineiston keruusta verkossa sekä havaintoja aineistosta

Tommi Nieminen² Tommi Kurki¹ Hamid Behravan¹
Heini Kallio³

¹Turun yliopisto, ²Itä-Suomen yliopisto, ³Helsingin yliopisto

Fonetiikan päivät, 5.–6. 3. 2015 Otaniemi

Jäsennys

- 1 Johdanto
- 2 Keruukäsi
- 3 Kertynyt aineisto ja sen laatu
- 4 Pari tehtyä havaintoa

Prosovar-hanke

- Prosovar = *Suomen prosodian alueellinen ja sosiaalinen variaatio*
- Tommi Kurjen (TY) ja Tommi Niemisen (ISY) ideoima hanke, jonka johtajana toimii Kurki ja jota sen mukaisesti hallinnoidaan Turun yliopistosta
- lisäksi hankkeessa ovat toimineet Hamid Behravan (ISY > TY) hankkeen keruusivuston suunnittelijana sekä Heini Kallio (HY) väitöskirjantekijänä

Mistä hankkeessa on ollut kyse?

- hankkeen päämääränä on ollut:
 - 1) kerätä suomesta prosodian tutkimukseen soveltuva puhetietokanta
 - 2) kehittää tähän uusi, joukkoistukseen perustuva keruumenetelmä
 - 3) tutkia suomen prosodian alueellista ja sosiaalista variaatiota
- hanketta rahoitti Koneen säätiö 1. 8. 2013–31. 5. 2015; hankkeelle on haettu ja haetaan uutta rahoitusta

Jäsennys

- 1 Johdanto
- 2 Keruukone**
- 3 Kertynyt aineisto ja sen laatu
- 4 Pari tehtyä havaintoa

Keruukone verkossa

Tervetuloa Suomen kielen prosodian tutkimushankkeen sivuille!



KÄYTTÄJIEN TUOTTAMA AINEISTO

Aineistoa sivustalla
kartuttaneiden henkilöiden määrä
asuinseuduittain (2.3.2015)

ALUE MÄÄRÄ

■	23
■	34
■	4
■	19
■	2
■	7
■	18
■	37
■	8
■	62

ulkomaat 2



[muuta tilastotietoja tästä](#)

kuuntele

pysäytä

Karttojen pohjakartta © Maanmittauslaitos

Auta meitä selvittämään, miten suomea puhutaan nykyään!

Osallistu tutkimukseen

Joukkoistusta?

- Puhu.utu.fi on ollut aito yritys joukkoistaa keruuta – ei vain hyödyntää verkkoa keruun menetelmänä
- tavoitteena on ollut sivusto, jonne ihmiset voisivat palata:
 - erilaisia tehtäviä
 - ”pelimäisyyttä”
 - karttuvaa tietoa
 - ”paluukanava” tutkijoilta kerääjille
- täysin tässä ei toistaiseksi ilmeisesti ole onnistuttu, koska kerran äänittäneet eivät juurikaan palaa
- miten lisätä vuorovaikutteisuutta? miten tehdä joukkoistamisesta aidompaa?

Erilaisia tehtäviä

- sivustolla on erilaisia tehtäväsarjoja
- käyttäjä voi aina keskeyttää tehtävien tekemisen ja palatessaan saa eteensä tekemättömiä niin kauan kuin niitä on tarjota
- tehtävissä pyritään vakioimaan taustatekijät niin, että puhunnokset olivat mahdollisimman pitkälle vertailukelpoisia
- tavallisia tilanteita, arkista toimintaa – ”sama asia eri murtein”
- tehtävien tyyppejä:
 - marjojenostokeskustelu
 - videon tapahtumien kuvailu
 - henkilön opastus (kuvitteellisen) kartan avulla
 - *ja niin edelleen – älkää minua uskoko vaan käykää itse katsomassa!*

Jäsennys

- 1 Johdanto
- 2 Kertuukone
- 3 Kertynyt aineisto ja sen laatu**
- 4 Pari tehtyä havaintoa

Kertynyt aineisto

- noin 700 rekisteröitynyttä mutta (oikeasti) äänittäneitä käyttäjiä 270
- useimmat äänittäneistä ovat kuitenkin tehneet useamman kuin yhden äänityksen
- pelleilijät ovat olleet hyvin harvassa
- toistaiseksi on turha puuttua vielä sosiaaliseen variaatioon:
 - aineistoa on siihen liian vähän
 - informanttien sosiaalinen jakauma on vinoutunut
 - sosiaalisia muuttujia on Suomessa perinteisestikin ollut vaikea määrittää

Mistä käyttäjähävikki?

- rekisteröityneitä käyttäjiä on ollut selvästi äänittäneitä enemmän: mistä tämä voi johtua?
- tehdäänkö rekisteröityminen vain, jotta päästään katsomaan, millaisia tehtävät ovat?
- onko joukossa ehkä murtautumis- tai muita häirintäyrityksiä?
- vai eikö sivuston tekniikka kaikille käyttäjille edelleenkaan toimi?

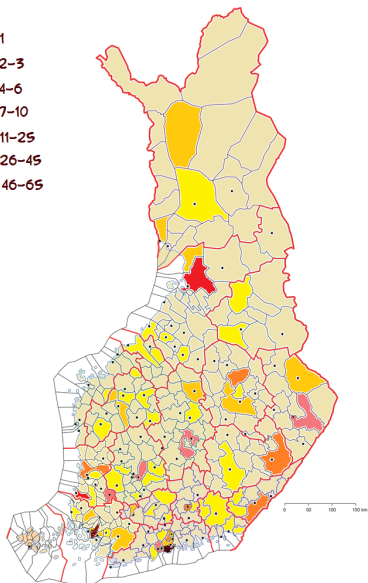
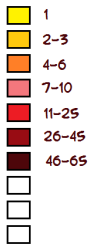
Jakauma murrealueittain

(Perinteinen jako pääkaupunkiseutukorjauksella)

lounaismurteet	23
siirtymämurteet	44
hämäläismurteet	31
Etelä-Pohjanmaan murre	8
keski- ja pohjoispohjalaiset murteet	21
peräpohjalaismurteet	5
savolaismurteet	47
kaakkoismurteet	8
pääkaupunkiseutu	61
perint. ruotsinkieliset alueet	1
ulkomaat	3

Jakauma asuinmaakunnittain

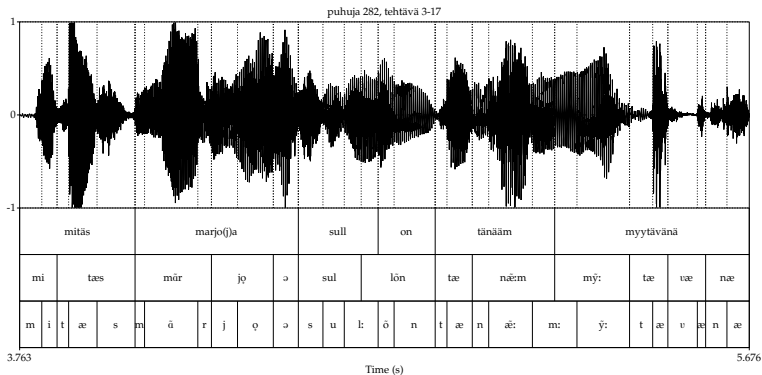
Varsinais-Suomi	41		
Satakunta	26		
Etelä-Karjala	7	Pohjois-Karjala	12
Etelä-Savo	7	Pohjois-Savo	10
Kainuu	2		
Kanta-Häme	4	Pirkanmaa	15
Päijät-Häme	5	Keski-Suomi	13
Kymenlaakso	4		
Pohjanmaa	1		
Etelä-Pohjanmaa	7	Keski-Pohjanmaa	2
Pohjois-Pohjanmaa	18	Lappi	5
Uusimaa	70		
muu	3		



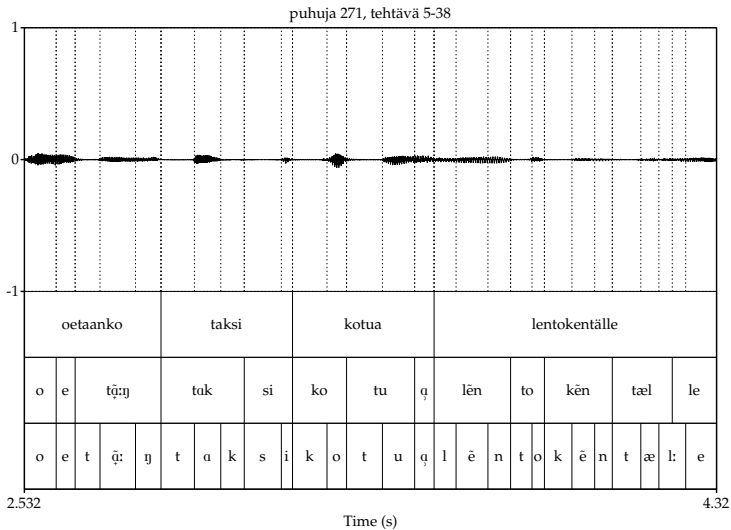
Riittääkö laatu?

- lähtökohtana on alkuaankin ollut tutkittavien prosodisten piirteiden (F_0 , A_0 , d) robustius: suhteet riittävät, absoluuttinen arvo merkityksetön
- äänitystasoa on arvioitu kuvailevalla asteikolla; keskiarvo on tyydyttävä
- äänitystaso-ongelmien lisäksi muita äänitykseen laatuun liittyviä ongelmia:
 - huono S/N-suhde (taustahäly, usein jo itse tallentavasta laitteesta)
 - huono taajuusvaste (etenkin korkeat taajuudet huonoja: esim. frikatiivien analyysi mahdotonta)
 - muita, oudompia ongelmia

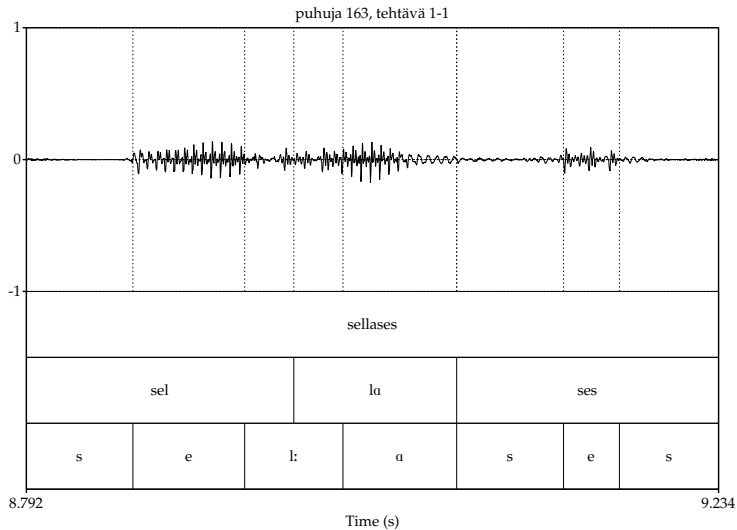
Äänitastaso erinomaisesta...



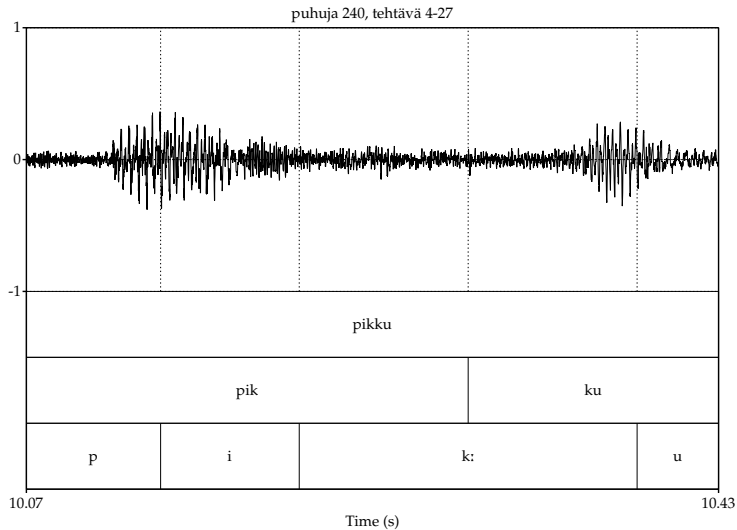
...heikkoon



Korkeiden taajuuksien katoaminen



Outoja ongelmia



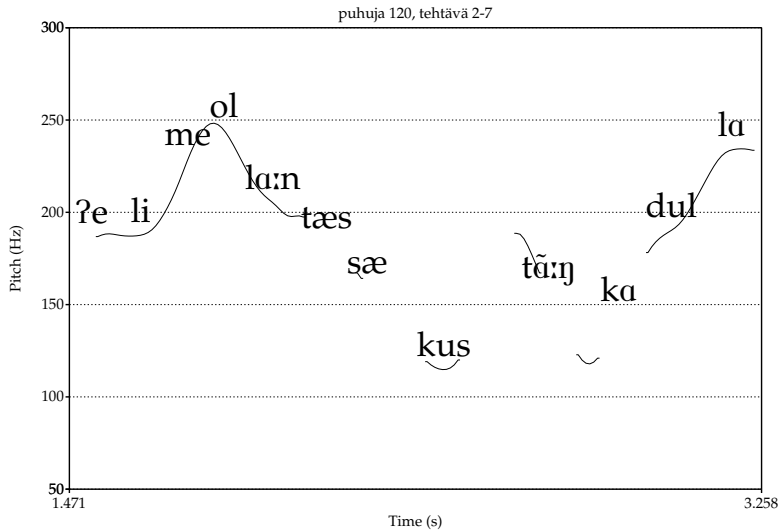
Jäsennys

- 1 Johdanto
- 2 Keruukäsi
- 3 Kertynyt aineisto ja sen laatu
- 4 Pari tehtyä havaintoa**

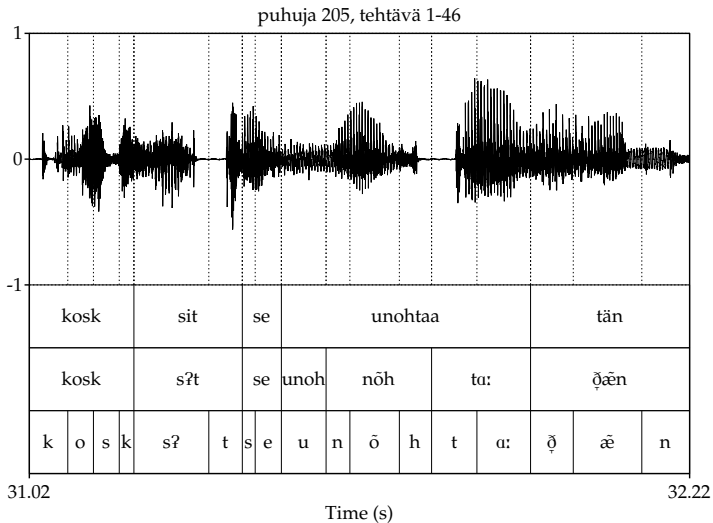
Missä mennään?

- ensimmäinen osatavoite on saada kultakin äänittäneeltä luotettavasti segmentoitua sanat, tavut ja foonit vähintään yhdestä äänitteestä
- (pienä hävikkiä on luvassa siitä, etteivät aivan kaikki äänitteet ole ehkä segmentoitavissa laatusyistä)
- tämä segmentointi riittää jo kesto-suhteiden (ja niiden mahdollisen variaation) analyysiin
- seuraava tavoite on merkitä prominenssit: aksentit eli lausepainot

Fo:n loppunousut

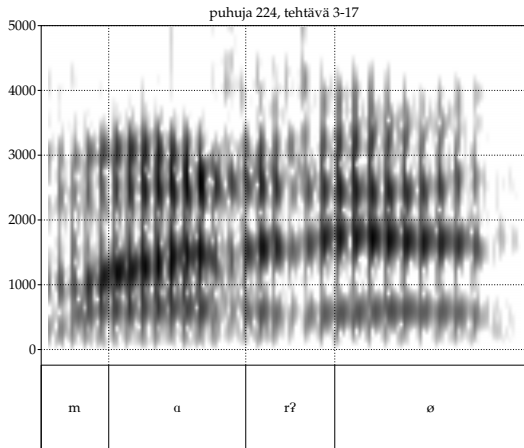


Vokaali osana s:ää



”Puolivokaaliprosodia”

Sana marjoja



KIITOS!