

Lasten kielikokemusten kartoitus suuren mittakaavan ääniaineistoilla

Okko Räsänen^{1,2} & Shreyas Seshadri²

¹Tampereen yliopisto, Informaatioteknologian ja viestinnän tdk., Signaalinkäsittelyn laboratorio

²Aalto-yliopisto, Sähkötekniikan korkeakoulu, Signaalinkäsittelyn ja akustiikan laitos

Varhaislapsuudessa tapahtuva kielenoppiminen riippuu lapsen kuulemasta kielestä sekä vuorovaikutuksesta kieltä käyttävien toimijoiden kanssa. Kuullun puheen ja vuorovaikutuksen määrä voi vaihdella esimerkiksi perheestä, sosioekonomisesta ympäristöstä tai kulttuurista toiseen, ja tällä vaihtelulla voi olla vaikutuksia lapsen myöhemmälle kielenkehitykselle (esim. Hart & Risley, 1995; Weisleder & Fernald, 2013; ks. myös Hoff, 2006). Kielelle altistumisen erot voivat myös olla sekä määrällisiä että laadullisia, kuten esimerkiksi puhutaanko lapselle suoraan, vai kuuleeko lapsi pääasiassa aikuisten välisiä keskusteluita. Suurin osa aiheeseen liittyvästä tutkimuksesta on kuitenkin toistaiseksi perustunut varsin pieniin laboratorioissa tai lasten kotona kerättyihin aineistoihin. Lisäksi suurin osa tutkimuksesta on keskittynyt rikkaisiin ja korkeasti koulutettuihin länsimaisiin yhteisöihin. Siksi on epäselvää missä määrin olemassa olevat löydökset ovat yleistettävissä erilaisiin kielenoppimisympäristöihin.

Edellä mainituista syistä varhaisen kielenkehityksen tutkijat ovat viime vuosina alkaneet kerätä pitkiä, koko päivän kestäviä, ääninauhoituksia lasten arkisista elinympäristöistä. Datankeräyksen keskeisenä apuvälineenä on ollut yhdysvaltalaisen Lena Research Foundationin kehittämä niin kutsuttu LENATM-nauhuri (Xu ym., 2008), joka voidaan pukea päälle koko päiväksi tehtävään tarkoitettuun liivin avulla. LENAn mukana tulee myös ohjelmisto, joka analysoi automaattisesti nauhoituksista erilaisia kieleen ja kommunikaatioon liittyviä tunnuslukuja kuten lapsen kuulemia sanamääriä. Ongelmana LENAn käytössä on kuitenkin sen kallis hinta ja teknisten ratkaisujen vanhanaikaisuus, eikä sitä sen suljetun kaupallisen luonteen vuoksi voi kehittää edelleen. LENAn ohjelmistoa ei voi myöskään käyttää muihin kuin LENA-nauhurilla kerättyihin nauhoitteisiin. Lisäksi LENA on optimoitu toimimaan amerikanenglannilla, ja siksi sen kielisyötteen määrää kuvaavien tunnuslukujen luotettavuus vaihtelee kielestä riippuen. Näistä rajoitteista johtuen tutkimuskentällä on kasvava tarve kehittää vaihtoehtoisia menetelmiä suurten ääniaineistojen automaattiseen analyysiin.

Meneillään oleva kansainvälinen yhteistyöprojektimme *Analyzing Child Language Experiences around the World* (ACLEW) pyrkii vastaamaan haasteeseen kehittämällä avoimeen lähdekoodiin perustuvaa työkalupakettia pitkäkestoisten lapsikeskeisten nauhoitusten analyysitarpeisiin. Tavoitteena on tarjota valmiiksi asennetut ja testatut työkalut helposti käytettävän Linux-virtuaalikoneen muodossa, jolloin työkaluja voidaan käyttää laskentaympäristöstä ja käyttöjärjestelmästä riippumatta ilman teknistä erityisosaamista. Tässä esitelmässä kuvataan lyhyesti ACLEW-hankkeen tavoitteita ja menetelmiä, sekä minkälaisia haasteita pitkiin ääninauhointeisiin liittyy. Lisäksi esityksessä kerrotaan miten itse kukin voi ottaa kehitteillä olevan työkalupakin käyttöön omassa tutkimuksessaan.

Viitteet

- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing, Baltimore, MD.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26(1), 55–88.
- Xu., D., Yapanel, U., Gray, S., Gilkerson, J., Richards, J., & Hansen, J. (2008). Signal processing for young child speech language development. *Proc. 1st Workshop on Child, Computer, and Interaction (WOCCI-2008)*, October 23, Chania Crete, Greece.
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143–2152.