

- life. – *Language & Communication* 23 s. 193–229.
- SKNA = Suomen kielen nauhoitearkisto. Kotimaisten kielten keskus. Helsinki.
- SKST = Suomalaisen Kirjallisuuden Seuran Toimituksia.
- SURKKA, SANNI 2016: *Tyylikästä lässyttystä. Homoseksuaali mies kielellisenä stereotyyppinä ja identiteettinä*. Suomen kielen pro gradu -tutkielma. Helsingin yliopisto.
- SUST = Suomalais-Ugrilaisen Seuran Toimituksia.
- TAMMINEN, HENRY 2017: Tämän kuluman murrettä jotenkii. *Kuinka savonlinnalaiset tunnistavat kotimurteensa erityispiirteet?* Suomen kielen pro gradu -tutkielma. Itä-Suomen yliopisto.
- TOIVANEN, JAANA 2012: Rautavaaralaismaallikoiden murrekäsityksiä. Suomen kielen seminaariesitelmä. Itä-Suomen yliopisto.
- VAATTOVAARA, JOHANNA – HALONEN, MIA 2015: Missä on ässä? ”Stadilaisen s:n” helsinkiläisyydestä. – Marja-Leena Sorjonen, Anu Rouhikoski & Heini Lehtonen (toim.), *Helsingissä puhuttavat suomet. Kielen indeksisyys ja sosiaaliset identiteetit* s. 40–83. SKST 1310. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VAATTOVAARA, JOHANNA – SOININEN-STOJANOV, HENNA 2006: Pääkaupunkiseudulla kasvaneiden kotiseuturajaukset ja kielelliset asenteet. – Kaisu Juusela & Katariina Nisula (toim.), *Helsinki kieliyhteisönä* s. 223–254. Helsinki: Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.
- Villitystä Savonlinnan seudulla. Blogi. <http://notkolla.blogspot.fi/2010/04/vihersijo-viime-viikolla-windmill-in.html>. (Blogikirjoitus päivätty 20.4.2010; luettu 15.3.2017.)
- VIRTARANTA, PERTTI 1946: *Länsiyläsatakuntalaisten murteiden äännehistoria I. Konsonantit*. SKST 230. Helsinki: Suomalaisen Kirjallisuuden Seura.
- 1958: *Pääpainollisen tavun jälkeisen soinnillisen dentaalspirantin edustus suomen murteissa*. SKST 252. Helsinki: Suomalaisen Kirjallisuuden Seura.
- YLI-LUUKKO, EEVA 1994: Jämsän äijän oma kuva. – Ilkka Savijärvi & Eeva Yli-Luukko, *Jämsän äijän murrekirja* s. 45–200. SKST 618. Helsinki: Suomalaisen Kirjallisuuden Seura.

Puheentunnistus ja kielenoppiminen työssä: Uusia mahdollisuuksia kartoittamassa

Luonnollista kieltä käsittelevä tietokone on ollut pitkään fennististen vitсийen kohde. Tietokonekäännöksistä on etsitty päättömiä esimerkkejä, puheentunnistusohjelmien väärille tulkinnoille on naureskeltu ja termeille *konekäännös-kieli* ja *Google Translate -kieli* on vakiintunut vähemmän mairittelevia merkityksiä. Tietotekniikan kehittyminen – erityi-

sesti suuriin datamääriin perustuva tutkimus ja sovellukset – on kuitenkin muuttanut tilannetta merkittävästi. Humanismi ja tietokoneet ovat kohdanneet toisensa, minkä osoittavat viime aikoina perustetut digitaalisen humanismin oppituolit. Fennistit ovat perinteisesti käyttäneet laajoja korpuksia, mutta nyt on alettu pohtia myös, miten koneiden kielenprosessointi-

kykyä voisi hyödyntää kielentutkimuksen apuna mahdollisimman monipuolisesti. Olisiko esimerkiksi puheentunnistus jo kehittynyt tasolle, jossa siitä voisi olla hyötyä kielenoppijalle?

Pilottihankkeemme Digitaalinen sovellus ammatillisen suomen oppimiseen työssä (lyhyesti Applla) kokeilee kielenoppimisen ja puheentunnistamisen yhdistämistä. Koneen Säätiön rahoittamassa monitieteisessä hankkeessa hyödynnetään tietoa kielenoppimisesta (kirjoittajat ja opiskelija Johanna Olkku Helsingin yliopistosta), puheentunnistusratkaisuista ja sovelluskehityksestä (Harri Airaksinen, Olli Alm ja Bao Nguyen Metropolia-ammattikorkeakoulusta) sekä johtamisesta (Hanna Rajalahti Haaga-Helia-ammattikorkeakoulusta). Hankkeessa kehitetään ja testataan tabletissa tai älypuhelimessa toimivaa digitaalista sovellusta (Nguyen 2017), joka tuottaa tavallisista vuorovaikutustilanteista automaattisen keskustelulitteraatin. Projektissa ei kehitetä puheentunnistuksen teknologioita sinänsä, vaan siinä testataan olemassa olevien ratkaisujen soveltuvuutta kielenoppimisen tarpeisiin. Pohdimme, miten kielenoppija voi hyödyntää sovelluksen tuottamia litteraatteja ja mitä muuta oppimista edistävää tietoa oppijalle voitaisiin keskustelulitteraatin perusteella tarjota. Sovellus kartoittaa ainakin suomalaisessa kontekstissa uutta maastoa: Tekes on rahoittanut hankkeita Company training on the go (Tekes 2014) ja 4M system (Nyrkkö, Carlson, Keijola, Ahonen-Myka, Niemi, Piitulainen, Viitanen, Meri, Seitsonen, Mannonen & Juvonen 2007), joissa kummassakin on käytetty puheentunnistusta työssä oppimisen tukena. Kumpikaan hankkeista ei kuitenkaan ole keskitynyt kielenoppimiseen.

Appla-sovelluksen ideana on tukea aikuista suomen kielen oppijaa työpaikalla tapahtuvassa kommunikaatiossa ja siten itsenäisessä kielenoppimisessa. Kohdeyryhmänä ovat suomen kielen oppijat,

joilla on jo peruskielitaito ja jotka joko ovat kouluttautumassa toimihenkilö- tai asiantuntijatehtäviin tai jo työskentelevät näissä tehtävissä. Kyse ei siis ole oppijan auttamisessa alkeistason kommunikaatio-ongelmien yli, vaan pyrkimyksenä on tukea häntä haastavissa kielenkäyttötilanteissa kuten puhelinkeskusteluissa (ks. esim. Kela & Komppa 2011) ja aikapaineisissa kokoustilanteissa (esim. Jäppinen 2011). Tällaisiin kielellisesti vaativampiin tehtäviin on tällä hetkellä tarjolla vain vähän koulutusta. Kotoutumiskoulutuksessa tavoitteena on saavuttaa Eurooppalaisen viitekehyksen taitotaso B.1.1, mikä on kuitenkin moniin tehtäviin riittämätön. Esimerkiksi asiantuntijatehtäviin sisältyy usein paljon kielen ja kommunikaation varaan rakentuvia työtehtäviä, jolloin niissä vaaditaan vähintään B2-tason kielitaitoa. (Komppa, Jäppinen, Herva & Hämäläinen 2014; Jäppinen 2011.) Kun formaali kielenopetus loppuu tätä alemmalle tasolle, muodostuu kielitaidosta kynnys, joka estää maahanmuuttajia pääsemästä tai siirtymästä urapolullaan esimies- ja korkeakoulutusta vaativiin tehtäviin (ks. *Kielimuurin yli* 2013). Yhtenä ratkaisuna ongelmaan on nähty kielen oppiminen työssä, koska monien työtehtävien vaatimaa kielitaitoa on vaikea hankkia muuten kuin käytännössä (ks. myös Seilonen, Suni, Härmälä & Neittaanmäki 2016: 137). Tähän tarpeeseen Applla-sovellusta on lähdetty kehittämään: sen avulla käyttäjä voi ”kesyttää” työpaikan kielellisiä käytänteitä.

Käytäntö: hyvää ja huonoa puheentunnistusta

Appla-sovelluksen käytännön kehitystyöstä ja ohjelmoinnista on vastannut Metropolia-ammattikorkeakoulun opiskelija Bao Nguyen opettajiensa Harri Airaksisen ja Olli Almin johdolla. Sovellus toimii Android-käyttöjärjestelmässä, ja se perustuu Googlen tarjoamaan puheentunnistus-

teknologiaan (Google Speech API). Sovellus on nykytilassaan hyvin yksinkertainen. Se avaa mobiililaitteen ruudulle näkymän, jossa on kielenvalitsin ja tallennuspainike. Kun käyttäjä käynnistää tallennuksen, sovellus alkaa litteroida puhetta laitteen ruudulle reaaliajassa. Automaattinen litteraatti tarjoaa kielenoppijalle tukea keskustelun kulun seuraamiseen, ja siihen voidaan myös palata myöhemmin: litteraatio välittyy verkkoalustalle, josta sen voi tarvittaessa ottaa käyttöön, kun haluaa palata tilanteeseen vaikkapa kysyäkseen ongelmallisista kohdista opettajalta tai kollegalta.

Hankkeessa on käynnissä testausvaihe. Testaajat ovat olleet suomea toisena kielenä puhuvia korkeakouluopiskelijoita, ja he ovat olleet suomen kielen taidoiltaan arviolta Eurooppalaisen viitekehityksen tasolla A2–C2. Testejä on videoitu elokuuhun 2017 mennessä yhteensä noin 3,5 tuntia.

Esittelemme seuraavaksi sovelluksen tuottaman tekstin tasoa. Seuraavalla sivulla olevassa katkelmassa (1a ja 1b) ohjaaja antaa testaajille suullisesti ohjeet tehtävää varten. Testaajat kuuntelevat ohjeen, joka samalla tallentuu Appla-sovellukseen. Kyseisessä testitilanteessa tavoitteenamme oli selvittää, miten opiskelijat käyttävät hyväkseen sovelluksen tuottamaa tekstiä. Keskusteluanalyysin konventioiden mukaisesti litteroitu näyte 1a osoittaa, miten ohje tilanteessa annettiin (ks. merkintätavoista esim. Stevanovic & Lindholm toim. 2016). Näyte 1b on Appla-sovelluksen tuottama litteraatti samasta tilanteesta.

Esimerkki sekä kumoaa että vahvistaa joitakin puheentunnistukseen liittyviä ennakkokäsityksiä. Kone tekee virheitä: se tekee Kelomäestä *Kellomäen*, täyden-tää kesken jääneen sanamuodon *opiskelijoi* muotoon *opiskelijoiden* (vrt. *opiskelijoil*) ja tulkitsee puhekielisen konditionaali-muodon *sais* englannin kielen sanaksi *size*. Pääosin virheet ovat harmittomia, mutta muutamissa kohdissa merkitys muuttuu olennaisesti. Esimerkiksi katkelman alku-

puolella *teidän tehtävä* muuttuu muotoon *heidän tehtävänään* ja loppupuolella *nii* muuttuu muodoksi *Ei*, mikä muuttaa lauseen merkityksen käänteiseksi. Kun kone tulkitsee kaksi lekseemiä (*sinne niit*, 1a) yhdeksi (*synnit*, 1b), lauseen merkityksen ymmärtäminen jälkikäteen vaikeutuu.

Toisaalta ohjelma tekstittää monin kohdin hämmästyttävän hyvin puhekielisiäkin sanamuotoja. Monet tyypilliset puhekieliset muodot, kuten *mut*, *tää*, *oo* ja *tota*, sovellus osaa kirjoittaa oikein. Myös sellaiset puhekieliset muodot, joissa on sanansisäinen kato, kuten *sellanen* (vrt. *sellainen*), voivat toistua koneen litteraatissa puhutussa muodossaan, mutta näitä muotoja kone myös korjaa: litteraatissa (1a) on muoto *tällasen*, mutta koneen litteraatissa (1b) *tällaisen*. Suurin osa yksityiskohdista ja yhdyssanoista on kohdallaan.

Testitilanteissa olemme pyrkinet selvittämään muun muassa sitä, kuinka paljon edellä mainittujen kaltaiset virheet lopulta häiritsevät opiskelijoita. Ne testaajat, jotka ovat olleet arviomme mukaan vähintään taitotasolla B1, ovat usein tunnistanet koneen tekemät virheet tai väärät tulkinnat. Alustavien havaintojen mukaan sen sijaan se, ettei koneen tuottamassa litteraatissa ole esimerkiksi välimerkkejä, häiritsee joitakin testaajia. Toisaalta mahdollisuus tarkistaa yksityiskohtaisia tietoja, vaikkapa kellonaikoja tai osoitteita, on tuntunut käyttökelpoiselta.

Mitä haluamme saavuttaa?

Tavoitteenamme on lisätä sovellukseen toimintoja, jotka tarjoavat oppijalle hyödyllistä lisätietoa vuorovaikutuksesta. Sovellus voisi kertoa käyttäjälle reaaliajassa, mikä on käynnissä olevan keskustelun tema, tai se voisi tarjota käydystä vuorovaikutuksesta yleisimmin esiintyneiden sanojen listan, josta käyttäjä voisi poimia opeteltavakseen uudet ja merkitykselliset sanat.

Jatkossa pohdittavaksi tulee myös uusia eettisiä kysymyksiä. Tässä vaiheessa sovel-

oke:i (.) jälleen (0.4) sellanen (.) tehtävä (0.5) #e# (0.8) meidän suomen kielen kulttuurin opiskelijoi- (0.4) opiskelijoille ↓kutsu? (0.6) .thh ja (0.5) teidän tehtävänä ois nyt siis (1.0) mieltä että mi- minkälainen (0.6) miten me saatais kutsuttua suomen kielen >ja< kulttuurin (0.3) opiskelijoita (0.2) tällaseen tapahtumaan (0.3) ää, (0.5) sen ↑tapahtuman nimi on (.) fennistiikan nykytila ja tulevaisuus? (0.8) se on (.) yhden meidän opettajan (.) Tapani Kelomäen (0.3) ↑juhlaseminaari? (0.9) öömm (1.1) ja se on suomeksi? (0.6) ↓öö ja siellä käsitellään =siel on opiskelijoitaki puhumassa (.) ja sit siellä (.) käsitellään? niinku suomen kieltä (.) eri >näkökulmista?< .hhh (0.5) ja tota: (0.2) öää (0.4) se on (0.4) se on tota: (1.0) kolmas päivä ↑helmikuuta? (0.6) kakstuhattaseittemäntoista °nyttten?° (0.5) ja se on (.) .h perjantaipäivä (.) kello (.) kolmetoista viiva ↓seittemäntoista? (0.3) .mthh (.) ja paikkana on (.) pääarakennuksen (.) luentosali kolmetoista? (0.4) ja (0.4) tästä sais opintopisteitä taas? (0.6) eli jos kirjottaa: (0.3) semmosen (.) ↓esseen tai vähä niinku luentopäiväkirjan? (0.2) kokonaisuuden? (0.4) nii siitä vois saada ↓kaks pistettä? (0.5) ja (.) sitte pitä: (.) mieltä että (1.5) m:inkälainen=minkälaisen? ↓tekstin te tekisitte että ne (0.2) opiskelijat (0.3) ↓tulis sinne. (0.3) #ku tää on °vähä niinku°(0.3) tää ↑ei oo pakollinen mut_olis kiva saada sinne niit opiskelijoi(ta)#.(0.9).nss ä (.) krhmm

Näyte 1a. Keskustelunanalyttinen litteraatti.

lusta on testattu niin, ettei kenenkään testaajan tietoja ole tallentunut sovellukseen. Käyttäjä voi luoda profiilin nimimerkillä, eikä hänen tarvitse antaa mitään yksilöivää informaatiota. Sovellusta käytettäessä materiaali tallentuu Googlelle, mikä on otettu huomioon testitilanteissa valitsemalla ai-

Okei jälleen sellanen tehtävä meidän suomen kielen kulttuurin opiskelijoiden opiskelijoille kutsu ja heidän tehtävänään ois nyt siis mieltä että minkälainen Miten me saatais kutsuttua suomen kielen ja kulttuurin opiskelijoita tällaiseen tapahtumaan sen tapahtuman nimi

on fennistiikan nykytila ja Tulevaisuus se on yhden meidän opettajan Tapani Kellomäen juhlaseminaari se on suomeksi tai siellä käsitellään siel on opiskelijoita kin puhumassa ja sit siellä käsitellään niinku suomen kieltä eri näkökulmista ja tota se on se on totta

kolmas päivä helmikuuta 2017 nyttten Ja se on perjantai päivä kello 13 viiva 17.00 ja paikkana on

rakennuksen luentosali 13 ja tästä size opintopisteitä taas eli jos kirjoittaa semmosen esseen tää vähän niinku luentopäiväkirjan kokonaisuuden Ei siitä voi saada kaksi pistettä ja sitte pitäis mieltä että minkälainen Minkälaisen tekstitä tekisitte että ne opiskelijat tuli sinne kun tää on vähän niinku tää ei oo pakollinen Mut Olis kiva saada synnit opiskelijat

Näyte 1b. Sovelluksen tuottama litteraatti.

heita, jotka eivät vaaranna testaajien yksityisyyttä. Lienee selvää, ettei sovellusta voi tulevaisuudessakaan käyttää sairaiden tapaisissa työyhteisöissä, joita koskevat hyvin tiukat salassapitomääräykset. Sen sijaan esimerkiksi kaupan ja hallinnon aloilla on työtehtäviä, joita jo nyt tallenne-

taan palvelun kehittämistarkoituksiin. Lisäksi tulevaisuudessa ratkaistavana on se, kuinka kauan, missä muodossa ja kenelle sovellukseen tallentuva materiaali on käytettävissä.

Tärkeimmät tavoitteet liittyvät kuitenkin kielenopetuksen tehostamiseen. Yksi yhteiskuntamme tulevaisuuden haasteita on muualta muuttaneiden saaminen mahdollisimman joustavasti työelämään. Ammatin vaatimaa kielitaitoa ei opi kuin ammatissa, mutta ongelmana on, ettei kielitaitoon työn ohessa keskitytä tietoisesti. Pelkän suomenkielisessä ympäristössä oleilun kehittävä vaikutus on kuitenkin rajallinen, eikä äidinkielistä useinkaan ole kielellistä apua, ellei kielenoppija sitä osaa määrätietoisesti vaatia (ks. esim. Eskildsen & Theodórsdóttir 2017; Wagner 2015; Kurhila 2006). Applä-sovellus on askel pyrki- myksessä ratkoa tätä ongelmaa. Tässä vaiheessa voisimme tietenkin pysähtyä siihen huomioon, että litteraatissa on monenlaisia virheitä. Toisaalta puutteellisestakin litteraatista voi olla hyötyä, ja sovellusta tähän mennessä testanneista opiskelijoista valtaosa on pitänyt Applaa kiinnostavana ja potentiaalisena. On hyvä muistaa, että puheentunnistusteknologia on kehittynyt viime aikoina nopeasti. Myös humanistien kannattaa pohtia uusille teknologioille järkevää käyttöä.

JOHANNA KOMPPA
LARI KOTILAINEN
etunimi.sukunimi@helsinki.fi

Lähteet

- ESKILDSSEN, SØREN W. – THEODÓRSDÓTTIR, GUÐRÚN 2017: Constructing L2 learning spaces. Ways to achieve learning inside and outside the classroom. – *Applied Linguistics* 38 s. 143–164. <https://doi.org/10.1093/applin/amv010>.
- Eurooppalainen viitekehys = Eurooppalainen kielten oppimisen, opettamisen ja arvioinnin yhteinen viitekehys*. Helsinki: WSOY 2003.
- JÄPPINEN, TUULA 2011: Suomen kielen taidon riittävyys yritysten aikapaineisissa puhetilanteissa esimiesten ja työharjoittelijoiden kuvaamana. – *Puhe ja kieli* 31 s. 193–214.
- KELA, MARIA – KOMPPA, JOHANNA 2011: Sairaanhoidajan työkieli – yleiskieltä vai ammattikieltä? Funktionaalinen näkökulma ammattikielen oppimiseen toisella kielellä. – *Puhe ja kieli* 31 s. 173–192.
- Kielimuurin yli* 2013. Suositukset korkeakoulutettujen maahanmuuttajien suomen kielen opintojen kehittämiseksi ja työllistymisen edistämiseksi pääkaupunkiseudulla. ONE Baltic Sea Region -hanke. Verkkodokumentti. http://www.helsinki.fi/urapalvelut/ONE_BSR/kielimuurin_yli.pdf (10.4.2017).
- KOMPPA, JOHANNA – JÄPPINEN, TUULA – HERVA, MARJA – HÄMÄLÄINEN, TAIJA 2014: *Korkeakoulutuksen ammatilliset suomi toisena kielenä -viitekehykset*. Aatos-artikkelit 16. Helsinki: Metropolia-ammattikorkeakoulu.
- KURHILA, SALLA 2006: *Second language interaction*. Amsterdam: John Benjamins.
- NGUYEN, BAO 2017: Assisting language learning with Applä application. Metropolia-ammattikorkeakoulun opinnäytetyö. <http://urn.fi/URN:NBN:fi:amk-201705239648>.
- NYRKKÖ, SEPPÖ – CARLSON, LAURI – KEIJOLA, MATTI – AHONEN-MYKA, HELENA – NIEMI, JYRKI – PIITULAINEN, JUSSI OLAVI – VIITANEN, SIRKE – MERI, MARTTI – SEITSONEN, LAURI – MANNONEN, PETRI – JUVONEN, JANI 2007: Ontology-based knowledge in interactive maintenance guide. – *Proceedings of the 40th Hawaii International Conference on System Sciences*. <https://doi.org/10.1109/HICSS.2007.411>.
- SEILONEN, MARJA – SUNI, MINNA – HÄRMÄLÄ, MARITA – NEITTAANMÄKI, REETTA 2016: Ammatillisen kielitaidon arviointikoikeilu tervey-

denhuollon alalla. – Ari Huhta & Raili Hildén (toim.), *Kielitaidon arviointitutkimus 2000-luvun Suomessa* s. 110–141. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA.

STEVANOVIC, MELISA & LINDHOLM, CAMILLA 2016: *Keskusteluanalyysi. Kuinka tutkia sosiaalista toimintaa ja vuorovaikutusta*. Tampere: Vastapaino. Tekes 2014. Learning solutions. [\[tumat.tekes.fi/uploads/352fe399/Sundquist_Suvi_alku-2041.pdf\]\(https://tumat.tekes.fi/uploads/352fe399/Sundquist_Suvi_alku-2041.pdf\) \(22.6.2017\).](https://tapah-</p></div><div data-bbox=)

WAGNER, JOHANNES 2015: Designing for language learning in the wild. Creating social infrastructures for second language learning. – Teresa Cadierno & Søren W. Eskildsen (toim.), *Usage-based perspectives on second language learning*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110378528-006>.