

Toimittaneet

Minnaleena Toivola, Pekka Lintunen & Leena Maria Heikkola

**PUHEEN TUTKIMUKSEN
UUSIA SUUNTIA – AINEISTONA
VAPAASTI TUOTETTU PUHE**

**NEW DIRECTIONS IN
SPEECH RESEARCH
– FREELY PRODUCED
SPEECH AS DATA**

AFinLA- teema / n:o 17

ISBN: 978-951-9388-78-6
ISSN: 2814-6298

AFinLA-teema 2024 / n:o 17

**Puheen tutkimuksen uusia suuntia
– Aineistona vapaasti tuotettu puhe**
**New directions in speech research
– Freely produced speech as data**

Toimittaneet

Minnaleena Toivola, Pekka Lintunen & Leena Maria Heikkola



Suomen soveltavan kielitieteen yhdistys AFinLA, Jyväskylä 2024
© kirjoittajat ja Suomen soveltavan kielitieteen yhdistys AFinLA

Sisältö

- *Minnaleena Toivola, Leena-Maria Heikkola & Pekka Lintunen*
Vapaasti tuotettu puhe tutkimuksen kohteena 5
Freely produced speech as the focus of research

Osa I

- *Lotta Aarikka*
"Tavoitteena on tarkastella luontevaa arkikeskustelua" – Vapaa, spontaani, luonnollinen ja aito (arki-)puhe fennistisissä murteita käsittelevissä väitöskirjoissa23
"The aim is to examine natural everyday conversation" - Free(ly produced), spontaneous, natural and authentic (everyday) speech in dialect dissertations in the field of Finnish language

Osa II

- *Riikka Ullakonoja*
Venäläisen aksentin tunnistaminen suomen suullisen kielitaidon arvioinnissa49
The identification of the Russian accent in Finnish oral proficiency assessment
- *Katja Haapanen, Antti Saloranta, Kimmo U. Peltola, Henna Tamminen, Lannie Uwu-khaeb & Maija S. Peltola*
The production of voicing in Namibian English stops by speakers of Khoekhoegowab.....77
- *Anna von Zansen & Heini Kallio*
DigiTala – Moodle-sovellus suullisen kielitaidon automaattiseen arviointiin.....91
DigiTala – An online tool for automated assessment of oral language skills
- *Leena Maria Heikkola, Minnaleena Toivola, Jeongdo Kim & Antti Saloranta*
Kielivalinnat ja kielenvaihtaminen ensikielisten ja ei-ensikielisten suomenpuhujien keskusteluissa – Kyselytutkimus 117
Language choices and language switching in conversations between L1 and L2 Finnish speakers – A survey

Osa III

- ✦ *Pauliina Peltonen & Pekka Lintunen*
Word searches and fluency during problem-solving tasks: a mixed-methods study of L1 Finnish and L2 English interaction..... 143
- ✦ *Sanna Olkkonen, Maarit Mutta & Pekka Lintunen*
Fluency profiles in L2 monologue production: the role of underlying cognitive factors..... 167
- ✦ *Elina Lehtilä, Pauliina Peltonen & Pekka Lintunen*
Listener perceptions of L1 speech in L2 and L3 fluency assessment contexts: a qualitative approach..... 191

Toivola, M., P. Lintunen & L. M. Heikkola (toim.) 2024. Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. *New directions in speech research – Freely produced speech as data.* AFinLA-teema / n:o 17, 5–21.

Minnaleena Toivola

Helsingin yliopisto

Leena-Maria Heikkola

UiT Norjan arktinen yliopisto
Åbo Akademi

Pekka Lintunen

Turun yliopisto

Vapaasti tuotettu puhe tutkimuksen kohteena

Nostot

- Vapaasti tuotettua puhetta käytetään usein aineistona puheentutkimuksessa.
- Vapaasti tuotetun puheen ja kontrolloidun puheen välillä on jatkumo.
- Tässä johdannossa esittelemme teemanumeron artikkeleiden teemoja.



VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Abstract

Freely produced speech is a common source of data for speech research. This article is an introduction to this thematic volume and discusses freely produced speech from different perspectives. Freely produced speech is often seen as the opposite of controlled, often phonetic experiments in laboratory conditions. However, it is very difficult to determine which data collection methods are sufficient for freely produced speech to be considered authentic and natural. The topics of the articles included vary from dialectological research in Finland and the phonetic features of Namibian English to second language fluency studies as well as proficiency or speaker assessments in a second language. As this thematic issue will show, it is essential to conduct research on freely produced speech in a comprehensive way, as well as to understand the importance of the results obtained. The way freely produced speech is assessed can have a significant impact on an individual's life, for example, in gatekeeper evaluations.

Keywords: speech, speech research, language learning, language use

Asiasanat: puhe, puheentutkimus, kielen oppiminen, kielen käyttö

1 Johdanto

Mitä on vapaasti tuotettu puhe? Sen voi määritellä ainakin osittain olevan suunnittelematonta, dynaamista ja tilanteista. Ääneen luettuun tekstiin verrattuna vapaasti tuotettuun puheeseen liittyy enemmän kielellisestä prosessoinnista johtuvia tekijöitä kuten epäröintiä tai sujuvuuden vaihtelua. Kun keskustelemme toistemme kanssa arkipäivän tilanteissa, usein dialogisesti ja multimodaalisesti, puhumme yleensä kiinnittämättä tietoisesti huomiota esimerkiksi siihen, miten tuotamme ääntä fysiologisesti. Emme myöskään välttämättä ole tietoisia siitä, miten alitajuisesti arvioimme keskustelukumppanin kielitaitoa ja miten olemme vuorovaikutuksessa sellaisten puhujien kanssa, joiden tuotoksessa kuuluu omasta puheestamme poikkeavia piirteitä kuten vierasta aksenttia. Tutkimuksen näkökulmasta on erityisen kiinnostavaa tarttua juuri tähän suunnitteleamattomuuteen ja pyrkiä vangitsemaan ohimenevä hetki, puhetilanne, tarkastelun kohteeksi.

Tässä julkaisussa tuomme esiin näkökulmia vapaasti tuotettuun puhuttuun kieleen tutkimuskohteena. Pyysimme tutkijoita Puheen ja puheentutkimuksen teemaverkoston kirjoituskutsussa pohtimaan vapaasti tuotetun puheen tutkimuksen mahdollisuuksia. Halusimme rajata teemaksi puheen erityisesti siinä muodossa, että tutkittava aineisto olisi mahdollisimman vähän kontrolloitua, jolloin päästään lähemmäksi autenttista kielenkäyttöä, kuten spontaania vuorovaikutusta, puhekielisiä ilmaisuja ja keskustelun dynaamista luonnetta. Kontrolloiduilla koeasetelmilla on puheen- ja kielentutkimuksessa kuitenkin nyt ja jatkossa oma merkityksensä, joka on luonnollisesti yhteydessä tutkimuksen tarkoitukseen. Käsittelemme eri aineistonkeruumenetelmien käyttöä myöhemmin tässä johdannossa.

Puheen tutkimusta, erityisesti suullista kielitaitoa ja ääntämistä, on tarkasteltu aiemmassa Kurosen ym. (2017a) toimittamassa AFinLA-teeman numerossa 10. Tässä teemanumerossa keskityimme osin samoihin aihepiireihin, kuten toisen tai vieraan kielen taitoon ja sen arviointiin, puheen sujuvuuteen, kielenvaihtoon sekä ensikielen vaikutukseen toisen kielen tuottamisessa. Kurosen ym. (2017a, ks. myös Kuronen ym. 2017b) julkaisussa keskityttiin erityisesti tutkimusasetelmiin, joissa kohteena oli toisen tai vieraan kielen oppiminen. Myös monet tässä teemanumerossa esitetyt tulokset ovat sovellettavissa kielenoppimiseen ja -opetukseen.

Käsillä oleva teemanumero sisältää tämän johdannon lisäksi kahdeksan tutkimusartikkelia, joissa vapaasti tuotettu puhe on ollut tutkittavana aineistona. Eri tilanteissa äänitettyä puheaineistoa ovat tutkineet seuraavat kirjoittajat:

- Haapanen, Saloranta, Peltola, Tamminen, Uwu-khaeb ja Peltola
- Lehtilä, Peltonen ja Lintunen
- Olkkonen, Mutta ja Lintunen,
- Peltonen ja Lintunen
- Ullakonoja

Aarikka sen sijaan käsittelee aihepiiriä toisesta näkökulmasta: hän tarkastelee, miten käsitteeseen vapaa puhe on suhtauduttu murteentutkimuksen perinteessä. Heikkolan, Toivolan, Kimin ja Salorannan tutkimuksen kohteena ovat puolestaan käsitykset tilanteista, joissa tuotetaan vapaata puhetta keskusteluissa ei-ensikielisen suomenpuhujan kanssa. Lisäksi von Zansen ja Kallio esittelevät kielenopetukseen suunnitellun DigiTala-työvälineen, jonka avulla on mahdollista arvioida puhujan kielitaitoa suomeksi tai ruotsiksi vapaasti tuotetussa puheessa.

Edellä kuvatut, käytettyjen tutkimusaineistojen mukaan jaotellut artikkelit muodostavat kolme kokonaisuutta, joita esittelemme tässä johdannossa seuraavasti: Esittelemme luvussa 2 Aarikan tutkimuksen, joka käsittelee puhuttujen aineistojen määritelmiä julkaisun osassa I sekä käymme läpi suhtautumista eri tavoin tuotettuun puheeseen, puheaineistojen laatuun ja määrään sekä siihen, miten hyvin niiden ajatellaan kuvaavaan tutkimuksessa ”aitoa” puhetta. Luvussa 3 esittelemme artikkelit seuraavilta kirjoittajilta:

- Ullakonoja
- Haapanen, Saloranta, Peltola, Tamminen, Uwu-khaeb ja Peltola
- Zansen ja Kallio
- Heikkola, Toivola, Kim ja Saloranta.

Nämä muodostavat julkaisun osan II, jonka yhdistävänä teemana on vapaasti tuotettu puhe kielenoppimisessa ja sen arvioinnissa sekä vuorovaikutuksessa.

Näiden artikkelien yhdistävänä teemana on vapaasti tuotettu puhe kielenoppimisessa ja sen arvioinnissa sekä vuorovaikutuksessa.

Luvussa 4 käsitellään vieraan kielen oppimiseen liittyvää puheen sujuvuutta eri näkökulmista. Se muodostaa julkaisun osan III, joka sisältää Lehtilän, Peltosen ja Lintusen; Olkkosen, Muttan ja Lintusen sekä Peltosen ja Lintusen artikkelit.

2 Tutkimusta puheesta

2.1 Vapaasti tuotettu puhe aineistona

Puheaineistoja käytetään monilla, erityisesti humanistisilla, mutta myös muilla tieteenoaloilla, joilla työskennellään esimerkiksi haastattelujen ja kenttätöiden parissa erilaisten kieleen ja puheeseen liittyvien ilmiöiden dokumentoimiseksi. Kielitieteilijöistä murteentutkijoiden ja foneetikkojen voidaan sanoa olleen ensimmäisiä, jotka ovat käyttäneet äänitettuja aineistoja. (ks. Mielikäinen 2017). Murteentutkimuksessa pyrittiin dokumentoimaan katoavaa kansankielistä puhetta, ja foneetikot tarvitsivat puheaineistoja voidakseen analysoida puhetta ja tehdä akustisia analyysejä myös näönvaraisesti. (Calamai & Frontini 2018.) Kurki ja Mustajoki (2019) nostavat keskiöön puhutun kielen ja variaation tutkimuksen fennistisessä kielentutkimuksessa. Jo 1800-luvun lopusta lähtien suomalaisessa murteentutkimuksessa on kerätty puhuttuja aineistoja haastattelumenetelmin (Kurki & Mustanoja 2020). He kuitenkin muistuttavat, että haastattelemalla on kerätty vuosikymmenten aikana hyvin monenlaisia murreaineistoja ja sanaa *haastattelu* on käytetty yleisnimityksenä, vaikka haastatteluiden sisältämässä keskustelupuheessa on voinut olla hyvin monitasoista variaatiota.

Tämän teemanumeron ensimmäisessä tutkimusartikkelissa Lotta Aarikka käsittelee vapaasti tuotetun puheen käsitettä ja sen määrittelyä rajatussa ja historiallisessa kontekstissa. Hänen vuonna 2023 ilmestyneen väitöstutkimuksensa aineistosta nousi esiin, että Suomessa vuosina 1871–2017 julkaistuissa murreaiheisissa väitöskirjoissa tavoiteltavaa puheaineistoa kuvattiin ”luonnolliseksi, luontevaksi, tuttavalliseksi tai spontaaniksi arkikieleksi”. Puheaineistojen kuvailun tarkempi käsittely rajautui pois hänen väitöskirjatutkimuksestaan (Aarikka 2023), mutta tämän teemanumeron julkaisussa Aarikka paneutuu aiheeseen syvemmin.

Murteentutkimuksessa aineistonkeruu on tyypillisesti perustunut tilanteisiin, joissa puhujan on toivottu tuottavan puhetta mahdollisimman vapaasti. Vaikka aineistoa on aiemmin pidetty vapaasti tuotettuna puheena, sen aitoutta ja institutionaalisuutta on myöhemmässä tutkimuksessa alettu pohtia tarkemmin. Analyysissaan suomen murteita käsittelevistä väitöskirjoista Aarikka pohtii vapaan puheen käsitettä ja sen määritelmiä muun muassa seuraavista näkökulmista: Onko puheen tuottotapa provosoitu vai provosoimaton? Miten kielentutkimuksen menetelmät soveltuvat tiettyyn puhetilanteeseen? Vapaan puheen määrittely ei ole yksinkertaista, kuten Aarikka artikkelissaan toteaa. Selkeimpinä vastakohtina vapaalle puheelle hän nostaa esiin

lukupuhunnan ja kehyslauseisiin asetetut minimiparit. Kirjoittaja pohtii myös, miten vapaan puheen määritelmä vaikuttaa tutkimuskysymyksiin, tutkimuskohteiden määrittelyyn ja siten tutkimustuloksiin. Toisaalta Aarikka nostaa esiin, että pelkkä vapaaseen ja epämuodolliseen puheeseen keskittyminen ei ole riittävää, vaan myös muodollisia puhetilanteita tulisi tutkia lisää. Lisäksi Aarikka ehdottaa, että tulevaisuudessa tutkimus voisi kohdistua erityisesti arjessa käytettävään kieleen.

Suomenkielisen keskustelupuheen fonetiikkaa on ensimmäisten joukossa tutkinut Ogden (2001), joka käytti tutkimusmateriaalina radiosta äänitettyä puhetta. Ohjelman kuuntelijat soittivat lähetykseen ja keskustelivat juontajan kanssa, joten käytetty kieli vastasi hyvin luonnollista keskustelupuhetta. Haittapuolena tällaisissa äänityksissä on aineiston huono laatu, mikä vaikeuttaa erityisesti akustisen analyysin tekemistä. Lennes ym. (2006) sen sijaan tutkivat erittäin korkealaatua aineistoa, joka oli äänitetty Teknillisen korkeakoulun akustiikan ja audiosignaalien käsittelyn laboratorion suuressa, kaiuttomassa tilassa. Tilanteessa jäljiteltiin puhelinkeskustelua, sillä puhujat kuulivat sekä oman että toisen puhujan äänen kuulokkeilla, mutta eivät nähneet toisiaan. Kukin osallistujapari tunsu toisensa hyvin ja sai keskustella tilanteessa vapaasti. Voidaan kuitenkin kysyä, minkälainen merkitys tutkimusjärjestelyillä ja poikkeavalla tilalla oli puheen tuottoon. Lenneksen ym. (2006) mukaan aineisto oli runsasta ja puhujat keskustelivat vilkkaasti äänityksen aikana, joten on mahdollista, että oudossakin tilanteessa voidaan tuottaa spontaania puhetta, joka ulkopuolisen korvaan kuulostaa ”normaalilta” tai ”aidolta”. Teknisesti korkealaatuisen aineiston suurin hyöty näkyy luotettavuuden paranemisena, kun puheesta voidaan tehdä tarkkoja akustisia analyysejä myös automaattisia menetelmiä käyttämällä. Huonompilaatuisessa aineistossa esiintyvää taustahälyä poistettaessa saatetaan samalla hävittää puhe-signaaliin oleellisesti kuuluvia piirteitä, mikä mahdollisesti vaikuttaa heikentävästi analyysin onnistumiseen. Puheaineiston laatu voi olla hyvinkin vaihtelevaa, mutta oleellista on, että tutkijat ovat aineistoa hankkiessaan selvillä siitä, kuinka hyvin se sopii tutkimustarkoitukseen.

Erytyisesti sosiofoneettisessa tutkimuksessa ollaan kiinnostuneita erilaisista tutkimustarkoituksiin suunnitelluista keskustelutehtävistä, joiden avulla puheesta saadaan elisitoitua tiettyjä tutkimuksen kohteeksi haluttavia piirteitä muutoin kuin ääneen lukemalla. Elisitoidun puheen voidaan ajatella olevan ainakin lähellä spontaania keskustelupuhetta. Esimerkiksi yleisesti käytetyssä Brownin ym. (1983) luomassa karttatehtävässä (*maptask*) saadaan elisitoitua tietynlaista keskustelua, kun toinen osallistuja antaa ohjeita hänellä karttaan merkityn reitin perusteella ja toinen henkilö pyrkii löytämään omasta kartastaan saman reitin ja määränpään. Sittemmin kyseistä karttatehtävää on muokattu monella tavalla. Sen uusimmat versiot ovat digitaalisia, ja sen tutkimuskäyttöä on laajennettu myös kirjoitetun viestinnän puolelle, kun osallistujat ovat voineet kommunikoida toistensa kanssa myös muuten kuin puhumalla (ks. esim. Berríos ym. 2023).

Myös muunlaista kuin vapaasti tuotettua aineistoa tarvitaan osana foneettista tutkimusta, sillä intra- ja intersubjektivinen vertailu puheen segmentaalisisissa ja toisaalta myös prosodisissa piirteissä olisi käytännössä erittäin vaikeaa, jos aineistona olisi vain spontaania puhetta. Erilaisia luettuja tekstejä käytetäänkin edelleen foneettisessa tutkimuksessa, sillä niiden etuna on hyvä kontrolloitavuus sekä vertailtavuus. Yhtenäisten materiaalien käyttö lukupuhunnassa myös kansainvälisesti (ks. esim. Deterding 2006; Baird ym. 2022) tekee tutkimustulosten vertailusta parempaa. Kontrolloidun laboratoriopuheen puolesta on puhunut erityisesti Xu (2010). Hänen mielestään manipuloimalla puheesta tutkittavia tekijöitä ja pitämällä muut tekijät vakiona voidaan keskittyä tiettyihin ja ennalta määrättyihin puheen piirteiden havaitsemiseen. Vaihtelu puheessa voidaan tällöin suoraan liittää manipuloituihin tekijöihin. Xu (2010) esittää omana näkemyksenään, että voidaksemme edistyä sen ymmärtämisessä, mitä puhe puhe on ja mistä se koostuu, pitää kaiken foneettisen puheetutkimuksen nojata vahvasti kontrolloituun puheeseen, jonka avulla voidaan testata hypoteeseja kiinnostavista puheen ilmiöistä.

Foneettisen tutkimuksen kohteena ovat usein puheen piirteet, jotka esiintyvät segmentaalisisella tai suprasegmentaalisisella tasolla, mutta voidaan kysyä, riittävätkö ne kuvaamaan vapaasti tuotettua puhetta ja siihen liittyvää tilanteisuutta riittävästi. Muun muassa Klaus J. Kohler (2000) on tuonut jo yli kaksi vuosikymmentä sitten voimakkaasti esiin, että foneettisten peruspiirteiden tutkimisen lisäksi tulisi huomioida myös erilaiset puhetyylit sekä puhujien mukautuminen viestintätilanteeseen. Lisäksi yksittäisen piirteiden tutkiminen osana kehysanaa ei välttämättä ole sovellettavissa tuottamiseen ja havaitsemiseen edes lausetasolle, saati laajemmin (Kohler 2000). Tällä hetkellä foneettisessa tutkimuksessa ollaan puheentuoton ja -havaitsemisen lisäksi kiinnostuneita myös puheeseen liittyvistä multimodaalisista elementeistä, kuten eleistä, ilmeistä ja kehon asennosta osana viestintää (ks. esim. Danner ym. 2018).

Käsitys siitä, nimitetäänkö eri tavoin elisitoitua aineistoa osittain vapaasti tuotetuksi, semispontaaniksi tai jollain muulla tavoin määritellyksi puheeksi, on pitkälti mielipidekysymys. Voidaan myös kyseenalaistaa keskustelu siitä, onko tiukasti kontrolloitu ja laboratorio-olosuhteissa tuotettu puhe huonompaa vai parempaa kuin vapaasti tuotettu puhe, sillä niiden välinen vertailu on ongelmallista. Jossain määrin tätä vastakkainasettelua voidaan tutkia tietyin menetelmin tuotetuista puheaineistoista. Uusinta puheaineistojen vertailevaa tutkimusta puheen prosodian alueelta ovat tehneet de Ruiter (2015) sekä vokaaleista Hall-Lew ja Boyd (2020). De Ruiter (2015) on tutkinut saksankielistä kuvakerrontaa (jota hän tutkimuksessaan nimittää spontaaniksi puheeksi) ja luettua puhetta. Luettava teksti oli muodostettu samojen puhujien tuottaman, ensimmäisen tutkimusvaiheen kuvakerronnan perusteella. Intonaatio oli tulosten mukaan kuvakerronnassa ja sitä vastaavassa luetussa puheessa selvästi erilaista. (De Ruiter 2015.) Hall-Lew ja Boyd (2020) vertailivat, millaisia akustisia mittaustuloksia saadaan englannin vokaaleista, kun aineisto kerättiin samoilta puhujilta haastatteleamalla, elisitoitintehävän avulla sekä tekstejä ääneen lukemalla.

Tulosten mukaan vokaalien laadussa oli vain vähän eroja erityisesti elisointitehtävän ja haastattelussa tuotetun puheen välillä. Lisäksi haastattelupuheesta ja ääneen luetusta puheesta mitatut vokaalien formanttien erot olivat pienempiä kuin tutkijat olivat odottaneet. (Hall-Lew & Boyd 2020.) Tuloksiin on kuitenkin syytä suhtautua varauksella, sillä aineisto kerättiin vain muutamalta puhujalta ja se on sovellettavissa lähinnä yksittäisten äänteiden tasolla.

Akustisten analyysien osoittamien mittaustulosten lisäksi eri tavoin tuotetusta puheesta voidaan tehdä myös kuulonvaraista, vertailevaa tutkimusta. Vakkilaisen ja O'Dellin (2022) tutkimus osoitti, että suomenkielisen sanan lopussa olevan vokaalin ei tarvitse olla yhtä pitkä spontaanissa puheessa kuin lukupuhunnassa tullakseen havaituksi pitkänä. Tulosta pitää tarkastella suuntaa antavasti, sillä tutkimuksessa ei varsinaisesti pyritty vertaamaan, miten kuulija erottaa lukupuheen spontaanista puheesta. Tutkimusta eri puheaineistojen välisistä eroista tarvitaankin lisää, mutta sitä rajoittavat jo vertailtavuuteen itseensä liittyvät haasteet. Wagner ym. (2015) ovat ehdottaneet, että puheilmioita tutkittaessa olisi tärkeää lisätä tietoa erilaisista menetelmistä, kerätä runsaasti aineistoa erilaisilla menetelmillä ja pyrkiä ymmärtämään syvemmin niiden vaikutus saatujen tulosten tulkinnassa. Tämä edistäisi paremmin ymmärrystä puheesta ja sen tutkimisesta kuin argumentointi tietynlaisen puheaineiston puolesta tai sitä vastaan.

Nykyään spontaanin keskustelupuheen äänittäminen on teknisesti huomattavasti helpompaa kuin vuosikymmeniä sitten, mutta aineistojen kerääminen on yhä edelleen vaativaa ja siihen liittyy myös monia eettisiä kysymyksiä. Euroopan Unionin tietosuojasäädös (General Data Protection Regulation; Yleinen tietosuojasäädös (GDPR) 2016) rinnastaa puheäänän suoraksi tunnustetiedoksi, kuten nimen tai osoitteen, jolloin sen kerääminen on tarkkaan säädeltyä. Eettisesti voidaan äänittää vain luvanvaraista puhetta, joten voidaankin kysyä, onko tutkimukseen saatava aineisto koskaan täysin spontaania, vapaata puhetta vai liittyykö tilanteeseen aina ns. "tutkimuksen maku". Käytännössä puhuja voi kuitenkin nopeasti unohtaa, että häntä äänitetään.

2.2 Vapaasti tuotetun puheen määrä tutkimuksessa

Vaikka eri tavoin tuotettua puhetta on nykyään saatavilla myös erilaisten avointen puhekorpusten (ks. esim. Kielipankki) kautta helpommin kuin aiemmin, on silti epäselvää, kuinka paljon vapaasti tuotettua puhetta tutkitaan. Tähän ei ole löydettävissä kovin paljon tutkimustietoa, mutta esimerkiksi Wagner ym. (2015) tarkastelivat vuosien 2007 (Trouvain ja Barry 2007) ja 2011 (Lee ja Zee 2011) fonetiikan alan suurimman kansainvälisen konferenssin (International Congress of Phonetic Sciences, ICPhS) aiheita. Tarkastelu osoitti, että kontrolloimattomaan puheaineistoon perustuvien tutkimusten määrä oli kasvanut vuodesta 2007 vuoteen 2011, mutta se oli yhä huomattavasti pienempi kuin kontrolloitujen aineistojen määrä.

Tarkastelimme vastaavalla tavalla, vaikkakin suppeammin, kuinka paljon ja minkälaisia puheentutkimukseen liittyviä aineistoja oli mukana AFinLAN syysseminaarissa vuonna 2022, jossa myös esiteltiin suurin osa tämän julkaisun artikkeleiden aiheista. Kävimme ensin läpi kaikki symposiumin tiivistelmien otsikot (145) puheen- ja kielentutkimukseen liittyvillä suomen-, englannin- ja ruotsinkielisillä lyhennetyillä hakusanoilla: *puh, kesk, vuor, suull, ään, haas, spe, con, flu, inter, dia, språk* ja *rör*. Otsikoista 34 valikoitui sellaisiksi, joissa esiintyi jokin lyhennetyistä hakusanoista. Tämän jälkeen tarkistimme tiivistelmistä, oliko kyseisten tutkimusten aineistona ollut puhetta. Näistä osa jäi abstraktin perusteella epäselväksi tai aineisto perustui johonkin muuhun kuin puheeseen (mutta saattoi olla myös metapuhetta puheesta). Poistimme epäselvät tapaukset, jolloin jäljelle jäi 24 tutkimusta, joissa aineistona oli puhetta, joka oli tuotettu muutoin kuin luettuna. Tiivistelmissä tuotiin selvennyksenä myös lähes aina esiin, että puheaineisto oli jotenkin tallennettua. Tämän jälkeen kävimme kaikki loput symposiumin abstraktit läpi, ja niiden joukosta löytyi vielä 41 muuta tutkimusta, joissa oli mainittu puheaineistot ja menetelmänä oli käytetty esimerkiksi etnografista tutkimusta. Lisäksi aineistona oli muuta havainnointia ja keskusteluiden seuraamista sekä YouTubea katsottuja videoita. Osassa näistä tiivistelmistä tuotiin selvästi esiin, että tutkimuksessa käytetty puhuttu aineisto oli äänitetty tai videoitu tutkimusta varten. Lisäksi yhdestätoista muusta abstraktista ei saanut täysin selvää, mikä aineisto oli kyseessä, mutta tutkimus perustui mahdollisesti puheeseen.

Kaikkissa tutkimuksissa, joissa tuotiin selvästi esille puheaineiston käyttö, koostui aineisto selkeässä enemmistössä tutkimuksista vapaasti tuotetusta puheesta. AFinLAN vuoden 2022 syysseminaarissa esityksiä tarkastellessa voi ainakin 65 (45 %) esitelmän sanoa perustuneen eri tavoin tuotettuun puheaineistoon. Saatu tulos osoittaa, että erilaisten puheaineistojen käyttö, erityisesti vapaasti tuotettu puhe, on yleistä soveltavan kielitieteen tutkimuksissa Suomessa.

3 Kielitaidon vai puhujan arviointia?

Usein kielitaidon arvioinnin tarkoitus on varmistua siitä, että kielenoppija on saavuttanut vaadittavan taidon johonkin ammattiin tai opiskeluun päästäkseen. Tällaista kielitaidon arviointia voidaan kutsua portinvartija-arvioinniksi (ks. tarkemmin Huhta ja Hilden 2016), josta esimerkkinä Suomessa ovat YKI-kokeet eli yleiset kielitutkinnot. Vaikka portinvartija-arvioinnissa pyritään objektiivisuuteen, yksittäisten arvioijien näkemyksillä voi olla yllättävän suuri rooli. Näkemyksiin voivat vaikuttaa myös muut kuin varsinaiseen kielitaitoon liittyvän arvioinnin ulkopuoliset tekijät, kuten puhujan ensikielen tunnistaminen. Aholan (2020) tutkimuksessa havaittiin, että yleiseen kielitutkintoon osallistuvien thainkielisten puhujien suorituksia arvioitiin ankarammin ääntämisen ja rakenteiden osalta, jos arvioijat tunnistivat puhujien olevan thainkielisiä.

Toisaalta thainkieliseksi tunnistaminen vaikutti siihen, että heille annettiin parempia arvioita sujuvuudesta ja yleiskriteerin hallinnasta (Ahola 2020).

Riikka Ullakonoja puolestaan tutkii artikkelissaan yleisen kielitutkinnon arvioiteja ääntämisen näkökulmasta. Tutkimuksen aineisto on kerätty kyseisen tutkinnon suomen kielen kokeeseen osallistuneilta venäläistaustaisilta puhujilta ja heidän puhutuoksiaan kuuntelevilta yleisen kielitutkinnon koulutetuilta arvioijilta. Ullakonoja keskittyy erityisesti siihen, millä perusteella puhujan ensikieli tunnistetaan pelkästään puhujan toisen kielen puheesta. Tällöin tunnistamiseen voivat vaikuttaa hyvinkin pienet vierasta aksenttia indikoivat puheen ominaisuudet. Tutkimukseen osallistuneet kuuntelijat tunnistivat puhujan ensikielen tiettyjen suomen kielen ääntämiseen liittyvien piirteiden perusteella. Tulos vahvistaa aiempia havaintoja siitä, että venäjänkielisille puhujille on haastavaa tuottaa suomen vokaalien kestoja sekä erityisesti suomen konsonantteja /l/ ja /h/. Lisäksi venäjän kieleen kuuluva palatalisaatio ja suomen kielestä poikkeava intonaatio liitetään tutkimuksessa siihen, että puhujan äidinkielenä tunnistetaan venäjä. Ullakononon tutkimus ei kuitenkaan vahvista aikaisempien tutkimusten havaintoja, joiden mukaan painottomien vokaalien redusoituminen ja narinan puuttuminen olisivat erityisesti venäläisen aksentin piirteitä. Sen sijaan tutkimuksessa nousee esiin mahdollisia uusia venäläisen aksentin erityispiirteitä sibilanteissa ja tremulanteissa, jotka saivat arvioijilta mainintoja niiden suomesta poikkeavan artikulaation vuoksi. Ullakonoja myös muistuttaa, että lyhyetkin spontaanit puhenäytteet riittävät vieraan aksentin kuulonvaraiseen arvioimiseen, vaikka syvälliseen ja monitahoiseen foneettiseen analyysiin ne saattavat olla liian suppeita. Tutkimuksesta saatuja tuloksia voidaan hyödyntää esimerkiksi venäjän- ja ukrainankielisten oppijoiden suomen kielen opetuksessa.

Ääntämisen arviointi on tärkeä osa kielitaidon portinvartija-arvioinneissa. Vieraaseen aksenttiin suhtautumisella voi olla myös sosiaalisia vaikutuksia, erityisesti silloin, jos kielenoppija kokee, että hänen ääntämiseensä suhtaudutaan kielteisesti. Ääntämistä pidetään osana identiteettiä, ja siksi se onkin kielen oppimisessa monimutkainen tekijä (Setter & Jenkins 2005). Jos kielenoppijan tavoitteena on saavuttaa kohdekieleessä hyvä ääntämistaito, on oleellista tietää, mihin piirteisiin kielenopetuksessa tulisi keskittyä. Tämän selvittämiseksi tarvitaan tutkimusta eri kohdekielten ja lähtökielten puhujien tuottamasta ääntämisestä. Katja Haapanen, Antti Saloranta, Kimmo Peltola, Henna Tamminen, Lannie Uwu-khaeb ja Maija Peltola tutkivat artikkelissaan, kuinka khoekhoegowabia ensikielenään puhuvat tuottavat soinnillisuutta sananalkuisissa klusiilikonsonanteissa vapaasti tuotetussa Namibian englannissa. Soinnillisuutta tarkastellaan akustisesti mitattavien VOT-arvojen perusteella vertaamalla niitä tyypillisiin englannin ja afrikaansin tuottomalleihin. Haapanen ym. osoittavat, että khoekhoegowabia ensikielenään puhuvat tuottavat Namibian englannissa selkeitä, britti-englannin kielen mallin mukaisia klusiileja, joiden sointikontrastit eivät heijasta afrikaansin ja khoekheongowabin malleja. Jos puhuja ei pysty tuottamaan soinnillisia ja soinnittomia klusiileja englannin mallin mukaisesti, se voi johtaa vääri-

nymmäryksiin yksittäisten foneemien tasolla, koska soinnillisuus voi erottaa sanojen merkityksiä toisistaan. Kirjoittajat ehdottavat, että tutkimuksen tulokset voivat auttaa kehittämään foneettisia harjoitusmalleja englannin kielen opetukseen Namibiassa. Koska tutkimuksessa on tarkasteltu sointia vain sananalkuisesti, nostavat Haapanen ym. esiin, että sitä tulisi tarkastella myös muissa foneettisissa konteksteissa ja eri ensikielten puhujilla.

Koska kielitaidon arviointiin voivat vaikuttaa monet muutkin tekijät kuin varsinainen puhetuotos, on tärkeää kehittää ihmisten tekemien arvioiden tueksi uusia menetelmiä, jotta arvioinnista saataisiin neutraalimpaa sekä tarvittaessa nopeampaa. Anna von Zansenin ja Heini Kallion artikkelissa esitellään puheentunnistus- ja arviointiteknologiaa käyttävää DigiTala-työkalua, jota voidaan käyttää apuna kielitaidon arvioinnissa. Työkalu tarjoaa suomen ja suomenruotsin oppijoille mahdollisuuden harjoitella ääneen lukemista sekä spontaania puhetta, joista se tekee automaattista arviointia. Oppijalta äänitetään kohdekielellä puhetta, josta saadaan työkalun avulla palautetta kielitaidon eri osa-alueista sekä kootusti yleisen kielitaitotason arvio. Palaute voi auttaa kielenoppijaa tunnistamaan omat kehitystarpeensa. Von Zansen ja Kallio pohtivat automaattisen arvioinnin mukanaan tuomia hyötyjä erityisesti portinvartijatilanteissa, kuten kansalaisuuteen ja työnhakuun liittyvissä arvioinneissa, ja tuovat myös esiin tekoälyyn pohjautuvan arvioinnin rajoitukset erityisesti pienissä kielissä kuten suomessa ja suomenruotsissa. Tekoälypohjaisen arvioinnin tueksi tarvitaan laaja puhetietokanta, joka edustaisi kaikkia kielitaitotasoja ja kielenoppijoiden taustoja, eikä aineisto olisi vinoutunut. Tällaisen aineiston kerääminen on haasteellista. Lisäksi ohjelmistoa tulisi kehittää, ja eri käyttäjäryhmät tulisi osallistaa kehitystyöhön. Kirjoittajien mukaan olisi myös tärkeää pohtia arvioiden luotettavuutta, validiteettia ja oikeudenmukaisuutta automaattista arviointia käytettäessä. Von Zansen ja Kallio ehdottavat tuleviksi kehittämiskohteiksi dialogipuheen ja ei-sanallisen kommunikation sisällyttämisen automaattiseen kielitaidonarviointiin.

Spontaaneissa keskusteluissa tapahtuu monia asioita samanaikaisesti. Oman puheen monitoroinnin lisäksi kiinnitämme keskusteluissa hyvin vaihtelevasti huomiota muihin seikkoihin. Esimerkiksi voimme tehdä oletuksia siitä, mitä kieltä keskustelukumppanimme haluaa puhua tai voimme muuttaa omaa tapaamme puhua tekemiemme havaintojen tai oletusten mukaan. Leena Maria Heikkolan, Minnaleena Toivolan, Jeongdo Kimin ja Antti Salorannan artikkelissa tarkastellaan ensikielisten suomenpuhujien näkemyksiä vapaasti tuotetusta puheesta keskusteluissa puhujan kanssa, jonka ensikieli on joku muu kuin suomi. Tutkimuksen tavoitteena on selvittää, mitä kieltä ei-ensikielisen suomenpuhujan kanssa puhutaan ja vaihtuuko keskustelun kieli englanniksi, vaikka aloitus olisi tapahtunut suomen kielellä. Kyselytutkimuksessa saatujen tulosten mukaan ensikieliset suomenpuhujat ilmoittivat aloittavansa keskustelut oletetun ei-ensikielisen suomenpuhujan kanssa suomeksi, jos keskustelu käytiin suomenkielisessä ympäristössä. Heikkolan ym. artikkelissa kiinnitetään huomiota siihen, että heidän tuloksensa poikkeaa aiemmasta tanskalaisesta tutkimuksesta:

Jespersen ja Hejnán (2021) mukaan ei-ensikieliset tanskanpuhujat ilmoittivat ensikielisten tanskanpuhujien aloittavan keskustelun hyvin usein englanniksi. Heidän tutkimuksessaan ei-ensikieliset tanskanpuhujat pitivät keskustelun aloittamista englanniksi tai kielenvaihtamista englanniksi toiseuttavana. Samalla he kokivat tulevansa luokitelluksi ”toisiksi” vieraan aksenttinsa tai ulkonäkönsä vuoksi. Heikkolan ym. tutkimuksessa ensikieliset suomenpuhujat olettivat, että ei-ensikieliset suomenpuhujat osaavat englantia paremmin kuin suomea, ja yli puolet kyselyyn vastaajista ilmoitti vaihtavansa kielen suomesta Englantiin kommunikaation sujuvoittamiseksi. Ensikieliset suomenpuhujat myös olivat sitä mieltä, että ei-ensikieliset suomenpuhujat reagoivat kielenvaihtamiseen positiivisesti ja näkivät sen kohteliaana ja hyödyllisenä. Tutkimuksen tulokset korostavat tarvetta lisätä tietoisuutta omien kielivalintojen ja kielenvaihtamisen seurauksista sekä suomen kielen puhumisen tärkeydestä kielenoppimisen ja osallisuuden mahdollistajana. Heikkola ym. esittävät, että jatkossa kielivalintoja ja kielenvaihtamista sekä näiden syitä tulisi tutkia myös ei-ensikielisten suomenpuhujien näkökulmasta.

Luvussa 3 esittelemämme tutkimukset osoittavat, että on tärkeää tutkia havaitsemiseen liittyviä, usein tiedostamattomia tekijöitä, jotka saattavat vaikuttaa toimintaamme erityisesti puhujien kanssa, jotka oppivat ensikieltämme. Vieraskielisyys on ilmiö, jonka ensikieliset puhujat havaitsevat herkästi, joskus hyvin pienistäkin ääntämisen piirteistä segmentaalisisella tasolla. Toisaalta yhtä tarkasti ei välttämättä tunnusteta puhujan ensikieltä tai sen tunnustaminen aiheuttaa ennako-oletuksia, millä taas voi olla laajempia merkityksiä esimerkiksi kielitaitoarvioinneissa. Onkin tärkeää tiedostaa, miten vieraskielisyys vaikuttaa ja miten suomenoppijoiden kielitaitoa arvioidaan tai olemmeko valmiita puhumaan omaa ensikieltämme vieraskieliseltä kuulostavan puhujan kanssa.

4 Vapaasti tuotettu puhe vieraan kielen sujuvuuden tutkimuksessa

Vieraita kieliä opiskeltaessa sujuvuutta pidetään tärkeänä tavoitteena: oppija haluaa usein itsekin tulla kohdekielen sujuvaksi käyttäjäksi. Käsitteenä sujuvuus on kuitenkin monitulkintainen ja voi viitata laajasti koko kielitaidon yleistason, puhutun kielen yleistason tai kapeammassa merkityksessä puheen ajallisesti mitattaviin ilmiöihin kuten taukoihin ja puhenopeuteen (Lennon 1990, 2000; Tavakoli & Hunter 2018). Tämän teemanumeron kolmannen kokonaisuuden artikkeleissa keskitytään sujuvuuteen erityisesti kapean määrittelyn näkökulmasta. Puheen sujuvuuden tutkimus alkoi varsinaisesti 1950-luvulla ja keskittyi aluksi erityisesti ensikieliin, kun taas vieraan kielen sujuvuutta ryhdyttiin tutkimaan enemmän Suomessa ja kansainvälisesti 1970-luvun lopulla (ks. Lintunen ym. 2020). Tätä ennen tutkijat olivat usein kiinnostuneita puheen

tarkkuudesta. Kiinnostus sujuvuuteen lisäsi myös kiinnostusta vapaasti tuotettuun puheeseen aineistona samaan aikaan, kun vieraan kielen oppimisen tutkijat kiinnostuivat oppijoiden kommunikatiivisesta kompetenssista.

Vieraan kielen sujuvuus on monitahoinen ilmiö, joka jaetaan usein Segalowitzin (2010) mukaan kolmeen ulottuvuuteen, eli puhetuotoksen sujuvuuteen (*utterance fluency*), kognitiiviseen eli prosessoinnin sujuvuuteen (*cognitive fluency*) sekä kuulijan havaitsemaan sujuvuuteen (*perceived fluency*). Tässä teemanumerossa jokaista ulottuvuutta käsittelee yksi artikkeli. Sujuvuusaineistoja kerätään eri metodein riippuen siitä, mikä sujuvuuden osa-alue on tarkastelun kohteena. Tavallista kuitenkin on, että puhetuotoksella on analyysissa jonkinlainen rooli. Tuotosta voidaan käyttää esimerkiksi havaitun sujuvuuden arvioinnin perusteena tai vertailukohtana kognitiivisen sujuvuuden testeihin.

Tuotettua puhetta voidaan kerätä eri tavoin. Ellis ja Barkhuizen (2005: 22–49) kuvaavat mahdollisuuksia kerätä suullista aineistoa kielenoppijoilta. Heidän mukaansa keräystilanteen vaatima kontrollointi vaikuttaa siihen, kuinka luonnollista tai vapaata puhetta tutkija pystyy keräämään (ks. myös osa 2 foneettisesta tutkimuksesta): vapain puhe esiintyy luonnollisissa kielenkäyttötilanteissa ja täyttää yleensä jonkin kommunikatiivisen tarpeen. Tutkija taas usein haluaa osallistujien tuottavan puhetta äänitystilanteessa melko luonnollisesti, mikäli tutkittavia piirteitä ei ole täysin määritelty etukäteen, tai hyvin kontrolloidusti, jos tavoitteena on saada esimerkkejä tietyistä kielen piirteistä. Puhetta saadaan tavallisesti jonkin tehtävän perusteella, mikä vaikuttaa siihen, kuinka tallennettu puhe sijoittuu vapaan ja kontrolloidun puheen jatkumolle. Jokaisessa äänitystilanteessa ns. tarkkailijan paradoksi (Labov 1972) vaikuttaa kuitenkin puheen luonnollisuuteen: audio- tai videolaitteet voivat vaikuttaa puhujan toimintaan ja tuotettuun kieleen (ks. myös ns. Hawthorne-ilmiö vieraan kielen tutkimuksessa esim. Mackey & Gass 2005:187–188). Kun kyseessä ovat kielenoppijat, täytyy myös huomioida osallistujien ikä ja kielitaidon taso: matalampi kielitaito saattaa esimerkiksi luokkahuoneahdistukseen yhdistettynä tehdä puhetilanteesta haastavan, eikä se välttämättä tuota edustavaa näytettä puhujan taidoista (ks. Horwitz ym. 1986). Sujuvuustutkimuksen aineistoiksi kerätään usein lyhyitä dialogisia ongelmanratkaisutehtäviä tai kuvakerrontaa, jossa osallistuja kertoo tarinan omin sanoin esimerkiksi videon, sarjakuvan tai erillisten kuvien pohjalta (ks. lisää Peltonen 2020: 40–41). Tällä tavoin koottujen aineistojen ongelma on, etteivät ne ole täysin vapaita vaan perustuvat tehtävänantoon, mutta toisaalta ne tarjoavat vertailukelpoista aineistoa eri puhujilta (ks. myös osa 2).

Tässä teemanumeron artikkelissaan Pauliina Peltonen ja Pekka Lintunen tutkivat artikkelissaan puhetuotoksen sujuvuutta. Tätä tarkasteltiin analysoimalla ensikielen (suomi) ja toisen kielen (englanti) sanahakuja ja näiden yhteyttä sujuvuuteen. Aineistona käytettiin ongelmanratkaisutehtävässä tuotettua vapaata puhetta. Tutkimuksen mukaan ensikielellä tehtiin enemmän yksittäisiä sanahakuja kuin toisella kielellä. Ensikielen ja toisen kielen sanahakujen välillä ei ollut merkittävää korrelaatiota, vaik-

ka laadullinen analyysi viittasikin siihen, että puhujien kommunikaatiotyylit olivat samanlaiset eri kielissä. Merkittävä uutta luova tulos artikkelissa on se, että erityisesti yhteistyötä edistävät sanahaut itse asiassa paransivat vuorovaikutuksen sujuvuutta, eivätkä heikentäneet sitä, kuten aiemmassa tutkimuksessa on esitetty. Kirjoittajat tarkastelivat myös ongelmanratkaisutaitoja ja niiden yhteyksiä sanahakuihin: jotkut käyttivät strategiana enemmän yksittäisiä sanahakuja, kun taas toiset käyttivät tukena vuorovaikutuskumppania. Kirjoittajat korostavatkin, että ongelmanratkaisustrategiat saattavat olla kielikohtaisia. Tuloksia voidaan käyttää apuna ongelmanratkaisustrategioiden harjoittelussa kielenopetuksessa.

Sanna Olkkonen, Maarit Mutta ja Pekka Lintunen taas keskittyvät artikkelissaan kognitiiviseen sujuvuuteen. Tutkimuksen kohteena on edistyneiden kielenoppijoiden sujuvusprofiilit, joiden avulla tarkastellaan kognitiivisen sujuvuuden, toisen kielen kielitaidon ja puheen tuottamisen sujuvuuden välisiä suhteita. Osallistujien ensikieli on suomi ja kohdekielenä englanti. Kielenoppijoiden sujuvusprofileja tutkimalla kirjoittajat haluavat syventää tietoa puheen tuoton taustalla olevista mekanismeista. Sujuvusprofileita tarkastellaan monimenetelmäisesti käyttämällä kuvankerrontatehtävällä kerättyjä puhenäytteitä ja kognitiivisia sanankäyttötehtäviä ensi- ja toisessa kielessä. Odotusten vastaisesti kognitiivinen sujuvuus kertoi enemmän yksilöllisistä eroista kuin toisen kielen kielitaidosta, eikä näiden välinen korrelaatio ollut korkea. Kielitaito- ja sujuvustehtävien tulosten pohjalta Olkkonen ym. tekivät klusterianalyysin, jonka pohjalta luotiin sujuvusprofiilit. Profiilit paljastivat, että korkea kielitaito ei välttämättä ole aina yhteydessä korkeaan sujuvuuteen toisessa kielessä ja että ensikielen yksilölliset piirteet voivat vaikuttaa toisen kielen puheen tuottamiseen. Tutkimus osoittaa myös käänteisen suhteen tarkkuuden ja sujuvuuden välillä: toiset kielenoppijat antoivat nopeita mutta epätarkempia vastauksia, kun taas toiset puolestaan hitaampia mutta tarkempia vastauksia. Kirjoittajat esittävät, että erilaiset sujuvusprofiilit tulisi ottaa huomioon kielenopetuksessa räätälöimällä yksilöllisempiä tehtäviä sujuvuuden kehittämiseksi.

Elina Lehtilä, Pauliina Peltonen ja Pekka Lintunen tarkastelevat artikkelissaan kuulijan havaitsemaa sujuvuutta. Artikkelin tuo sujuvuustutkimukseen uutena näkökulmana sen, että sujuvuutta arvioidessa oli mahdollista tutustua myös arvioitavan puhujan ensikieliseen puhenäytteeseen. Arvioijilta kysyttiin, kuinka hyödylliseksi he kokivat puhujan ensikielisen (suomi) tuotoksen kuuntelun, kun arvioitavana oli toisen (englanti) ja kolmannen (ruotsi) kielen sujuvuus. Vaikka vain osa arvioijista koki ensikielen kuuntelun hyödylliseksi, korreloivat ensikielen koettu hyödyllisyys ja sen vaikutusaste sujuvuusarvioon keskenään. Laadullisen analyysin perusteella sujuvuuden arvioijat kokivat ensikielisen näytteen kuuntelun vapauttavan kognitiivisia resursseja toisen ja kolmannen kielen sujuvuuden arviointiin, vaikka sen kuuntelun koettu hyödyllisyys vaihtelikin arvioijien välillä. Ensikielen sujuvuutta arvioitaessa tärkeimmäksi piirteeksi nousi puhenopeus, kun taas epäroinnit ja korjaukset mainittiin harvemmin. Tämä voi viitata siihen, että näiden arviointi saattaa olla alitajuista tai että ne jäävät

arvioiden ulkopuolelle, koska arvioijilla ei ole juurikaan aiempaa arviointikokemusta. Lehtilä ym. nostavat esiin kiinnostavan tuloksen, että osa sujuvuuden arvioijista vertailee ensikielisen tuotoksen piirteitä toisen tai kolmannen kielen piirteisiin, kun taas osa arvioijista jättää tietoisesti ensikielen huomioimatta. Tutkijat korostavatkin, että on tärkeä huomata kielenoppijoiden yksilöllisten piirteiden mahdollinen, usein alitajuinenkin vaikutus sujuvuuden ja laajemmin kielitaidon arviointiin.

5 Lopuksi

Tässä teemanumerossa lähestytään vapaasti tuotettua puhetta eri näkökulmista ja pohditaan yhtäältä sitä, mitä vapaasti tuotettu puhe on, ja toisaalta sitä, kuinka yleisesti vapaasti tuotettua puhetta käytetään tutkimuksen aineistona. Pienoistutkimuksemme vuoden 2022 AFinLA-syysseminariumin esitelmistä osoittaa, että puheaineistojen käyttö soveltavan kielentutkimuksen aineistona on runsasta. Mielestämme tästäkin syystä on tärkeää nostaa esiin, mitä oikeastaan tarkoitetaan, kun tutkimukseen on valittu tietynlainen puheaineisto, erityisesti tämän teemanumeron aiheena oleva vapaasti tuotettu puhe. Ennen kaikkea on tärkeää kuvailla aineiston hankintaa mahdollisimman tarkasti, luotettavasti ja läpinäkyvästi. Puheaineistojen tekniseen laatuun vaikuttavat puolestaan usein melko yksinkertaiset asiat, kuten sopivan mikrofonin valinta. Hyvälaatuisia aineistoa on myös helpompi analysoida automaattisin menetelmin. Lisäksi kuulonvarainen arviointi on sekä miellyttävämpää että tehokkaampaa, jos aineistossa on vain vähän taustahälyä. Teemanumeron artikkeleiden tulokset antavat arvokasta uutta tietoa vapaasti tuotetusta puheesta. Lisäksi ne tuovat esiin uusia tutkimusaiheita, joihin tarttua tulevaisuudessa.

Kiitokset

Tätä julkaisua on avustanut Suomen tiedekustantajien liitto. Julkaisu on myös saanut TSV:n kautta tieteellisen julkaisutoiminnan avustusta, jota opetus- ja kulttuuriministeriö myöntää.

Kirjallisuus

- Aarikka, L. 2023. *Murre ja sen tutkimus: Näkökulmia fennistisen murteen tutkimuksen historiaan ja kieli-ideologioihin 1871–2017*. Annales Universitatis Turkuensis, sarja B, osa 626. Turku: Turun yliopisto.
- Ahola, S. 2020. Yleisten kielitutkintojen arvioijien käsityksiä thainkieliseksi tunnistettujen suomenoppijoiden suullisesta kielitaidosta. *Puhe ja kieli*, 40 (4), 203–224. <https://doi.org/10.23997/pk.103307>
- Baird L., N. Evans & S. J. Greenhill 2022. Blowing in the wind: using ‘North Wind and the Sun’ texts to sample phoneme inventories. *Journal of the International Phonetic Association*, 52 (3), 453–494. <https://doi.org/10.1017/S002510032000033X>
- Berrios, J., A. Swain & M. Fricke 2023. Implementing the map task in applied linguistics research: what, how, and why. *Research Methods in Applied Linguistics*, 2 (3). <https://doi.org/10.1016/j.rmal.2023.100081>.
- Brown, G., A. Anderson, R. Shillcock & G. Yule 1983. *Teaching talk: strategies for production and assessment*. Cambridge: Cambridge University Press
- Calamai, S. & F. Frontini 2018. FAIR data principles and their application to speech and oral archives. *Journal of New Music Research*, 47 (4), 339–354. <https://doi.org/10.1080/09298215.2018.1473449>
- Danner S., A. Barbosa & L. Goldstein 2018. Quantitative analysis of multimodal speech data, *Journal of Phonetics*, 71, 268–283. <https://doi.org/10.1016/j.wocn.2018.09.007>.
- Deterding, D. 2006. The North Wind versus a Wolf: short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association*, 36 (2), 187–196. <http://www.jstor.org/stable/44526411>
- Ellis, R. & G. Barkhuizen 2005. *Analysing learner language*. Oxford: Oxford University Press.
- Hall-Lew, L. & Z. Boyd 2020. Sociophonetic perspectives on stylistic diversity in speech research. *Linguistics Vanguard*, 6 (1). <https://doi.org/10.1515/lingvan-2018-0063>
- Horwitz, E. K., M. B. Horwitz & J. Cope 1986. Foreign language classroom anxiety. *The Modern Language Journal*, 70 (2), 125–132. <https://doi.org/10.2307/327317>
- Huhta, A. & R. Hilden 2017. Kielitutkinnot ja muu laajamittainen kielitaidon arviointi Suomessa. Teoksessa A. Huhta & R. Hilden (toim.) *Kielitaidon arviointitutkimus 2000-luvun Suomessa*. AFinLA-e Soveltavan kielitieteen tutkimuksia 9. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 3–26.
- Jespersen, A. B. & A. Hejná 2021. Second language learners of Danish as the linguistic other. *Otherness: Essays & Studies*, 8 (3), 129–161. https://otherness.dk/fileadmin/www.othernessandthearts.org/Publications/Journal_Otherness/Otherness_Essays_and_Studies_8.3/Second_Language_Learners_of_Danish_as_the_Linguistic_Other_-_Jespersen_and_Hejna__with_abstract_.pdf
- Kielipankki www.kielipankki.fi. Luettu [28.3.2024]
- Kohler, K. J. 2000. Investigating unscripted speech: implications for phonetics and phonology. *Phonetica*, 57 (2–4), 85–94. <https://doi.org/10.1159/000028464>
- Kurki, T. & L. Mustanoja 2019. Käänteitä, aaltoja ja risteymiä fennistisessä variaation-tutkimuksessa. Teoksessa M. Hamunen, T. Nieminen, T. Kelomäki & H. Dufva (toim.) *Käänteitä ja käsitteitä: näkökulmia kielitieteelliseen keskusteluun*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1454. Helsinki: Suomalaisen kirjallisuuden Seura, 87–120.
- Kurki, T. & L. Mustanoja 2020. Haastattelu aineistonkeruun metodina sosiolingvistiikassa. Teoksessa M. Luodonpää-Manni, M. Hamunen, R. Konstenius, M. Miestamo, U. Nikanne & K. Sinnemäki (toim.) *Kielitutkimuksen menetelmiä II*. Helsinki: Suomalaisen Kirjallisuuden Seura, 274–312. <https://doi.org/10.21435/skst.1457>

- Kuronen, M., P. Lintunen & T. Nieminen (toim.) 2017a. *Näkökulmia toisen kielen puheeseen – Insights into second language speech*. AFinLA-e. Soveltavan kielitieteen tutkimuksia n:o 10. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinla/issue/view/5195>
- Kuronen, M., P. Lintunen & T. Nieminen 2017b. Suullisen kielitaidon ja ääntämisen tutkimuksesta soveltavan kielentutkimuksen alalla Suomessa. Teoksessa M. Kuronen, P. Lintunen & T. Nieminen (toim.) *Näkökulmia toisen kielen puheeseen – Insights into Second Language Speech*. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 3–17. <https://journal.fi/afinla/article/view/73121>
- Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania.
- Lee, W.-S. & E. Zee (toim.) 2011. *Proceedings of the 17th International congress of the phonetic sciences 2011*. Hong Kong: City University of Hong Kong.
- Lennes, M., E. Aho, M. Toivola & L. Wahlberg 2006. On the use of the glottal stop in Finnish conversational speech. Teoksessa Aulanko, R., Wahlberg, L., & Vainio, M. (toim.) *Fonetiikan päivät 2006: The phonetics symposium 2006*. Helsingin yliopisto, 93–102.
- Lennon, P. 1990. Investigating fluency in EFL: a quantitative approach. *Language Learning*, 40, 387–417. <http://dx.doi.org/10.1111/j.1467-1770.1990.tb00669.x>.
- Lennon, P. 2000. The lexical element in spoken second language fluency. Teoksessa H. Riggensbach (toim.) *Perspectives on fluency*. Ann Arbor, MI: The University of Michigan Press, 25–42.
- Lintunen, P., M. Mutta & P. Peltonen 2020. Sujuvuustutkimuksen käännteitä: katsaus sujuvuustutkimuksen menetelmiin. Teoksessa S. Grasz, T. Keisanen, F. Oloff, M. Rauniomaa, I. Rautiainen & M. Siromaa (toim.) *Menetelmällisiä käännteitä soveltavassa kielentutkimuksessa – Methodological Turns in Applied Language Studies*. AFinLAN vuosikirja 78. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 174–197. <https://journal.fi/afinlavk/article/view/89441>
- Mackey A. & S. Gass 2005. *Second language research: methodology and design*. Mahwah: Lawrence Erlbaum Associates.
- Mielikäinen, A. (2017). Ensimmäisten suomenkielisten murre tutkimusten metakieli. *Virittäjä*, 121 (3), 355–381. <https://doi.org/10.23982/vir.58386>
- Ogden R. 2001. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association*, 31 (1), 139–152. <https://doi.org/10.1017/S0025100301001116>
- Peltonen, P. 2020. *Individual and interactional speech fluency in L2 English from a problem-solving perspective: a mixed-methods approach*. Turku: University of Turku. <https://urn.fi/URN:ISBN:978-951-29-8137-3>
- de Ruiter, L. 2015. Information status marking in spontaneous vs. read speech in story-telling tasks – Evidence from intonation analysis using GToBI. *Journal of Phonetics*, 48, 29–44. <https://doi.org/10.1016/j.wocn.2014.10.008>.
- Segalowitz, N. 2010. *The cognitive bases of second language fluency*. New York: Routledge
- Setter, J. & J. Jenkins 2005. Pronunciation. *Language Teaching*, 38, 1–17.
- Tavakoli, P. & A.-M. Hunter 2018. Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, 22 (3), 330–349. <http://dx.doi.org/10.1177/1362168817708462>
- Trouvain, J. & W. J. Barry (toim.) 2007. *Proceedings of the 16th International congress of the phonetic sciences 2007*. Dudweiler: Pirrot GmbH.
- Vakkilainen, J. & M. O'Dell 2022. Suomen pitkien painottomien vokaalien havaitseminen spontaanissa ja luetussa puheessa. *Puhe ja kieli*, 41 (4), 247–264. <https://doi.org/10.23997/pk.113851>

- Wagner, P., J. Trouvain & F. Zimmerer 2015. In defense of stylistic diversity in speech research. *Journal of Phonetics*, 48, 1–12. <https://doi.org/10.1016/j.wocn.2014.11.001>.
- Xu, Y. 2010. In defense of lab speech. *Journal of Phonetics*, 38 (3), 329–336. <https://doi.org/10.1016/j.wocn.2010.04.003>.
- Yleinen tietosuoja-asetus (GDPR) 2016/679. <https://eur-lex.europa.eu/Fl/legal-content/summary/general-data-protection-regulation-gdpr.html>

Osa I

Toivola, M., P. Lintunen & L. M. Heikkola (toim.) 2024. Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. *New directions in speech research – Freely produced speech as data*. AFinLA-teema / n:o 17, 23–47.

Lotta Aarikka
Itä-Suomen yliopisto

”Tavoitteena on tarkastella luontevaa arkikeskustelua” – *Vapaa, spontaani, luonnollinen ja aito (arki-)puhe* fennistisissä murteita käsittelevissä väitöskirjoissa

Nostot

- Tutkimuksessa tarkastellaan aineistopohjaisesti *vapaan puheen* käsitettä.
- Analyysin mukaan *vapaa puhe* määrittyy suhteessa kielentutkimuksen menetelmiin kuten foneettisiin mittauksiin sekä erilaisiin kyselymenetelmiin.
- Analyysissa käsitellään puhetilanteita (haastattelu ja keskustelu) ja niiden suhdetta puheen *vapauden* määrittelyyn.
- Artikkelin päätelmissä pohditaan, miten ja miksi tietynlaista puhetta arvotetaan toista aidommaksi, luonnollisemmaksi ja vapaammaksi.

Abstract

The study examines the concept of *free(ly produced)* ('vapaa') *speech*. The expressions *free* ('vapaa'), *spontaneous* ('spontaani'), *natural* ('luonnollinen'), *authentic* ('aito') and *everyday* ('arki-') are the subject of the analysis. The research material consists of 51 doctoral dissertations dealing with dialects published in the field of Finnish language studies between 1871 and 2017, from which previously mentioned expressions have been searched qualitatively by means of close reading and mechanically by word searches. The research shows that *free(ly produced)* ('vapaa') speech can be defined in different ways and that speech produced in different speech situations appears at different contexts as both *free(ly produced)* and *non-free(ly produced)* speech. The study discusses phonetic experiments, elicitation methods, and interview and conversation as data acquisition methods and speech situations in language research. The study shows that although the scientific field has emphasized that no speech is more *natural* and *authentic* than another, in practice such assessments are made in research discourses (dissertations).

Keywords: research history, spoken language research, concept analysis, critical discourse research

Asiasanat: tutkimushistoria, puhutun kielen tutkimus, käsiteanalyysi, kriittinen diskursintutkimus

1 Tutkimuksen tausta, tutkimuskysymykset ja tutkimuksen tarkoitus

Puhuttu kieli on kiinnostanut suomen kielen tutkijoita aivan tieteenalan historian alusta saakka (esim. Aminoff 1871). Puhuttua kieltä on tutkittu pelkästään fennistisen murteentutkimuksen alalla monesta näkökulmasta: kantakielimäisinä ehyeksi konstruoituna kielimuotoina (esim. Lindén 1942), varioivana yksilön kielenkäyttönä (esim. Nuolijärvi 1986), kielihistoriallisesti (esim. Rapola 1919) sekä synkronisen variaation ja käynnissä olevan muutoksen näkökulmasta (esim. Kurki 2005). Murteita puhuttuina kielinä on tutkittu myös kielen eri rakennetasoihin keskittyen – äännerakenteen (esim. Genetz 1877; Ylitalo 2009), morfologian (esim. Paunonen 1974), sanaston (esim. Ruoppila 1943), syntaksin (esim. Väänänen 2016) sekä vuorovaikutuksen (esim. Priiki 2017) näkökulmista.

Kielentutkimuksessa on toisinaan tapana erotella tutkijan kielitajuun perustuva kieliaineiston tai esimerkkien tuottaminen "aidosta" kielenkäytöstä eli kieliaineistosta, joka perustuu muiden kuin tutkijan kielen käyttöön (ks. esim. Hamunen & Huumo 2020: 200; Väänänen 2016: 16). Puhutun kielen tutkimuksen¹ näkökulmasta kieliaineiston

1 Tarkoitan tässä artikkelissa *puhutulla kielellä* kieliaineistoa, jonka kielenkäyttäjät tai kielenkäyttäjät ovat tuottaneet puhumalla eli äänneiden virtana. *Puhutun kielen tutkimuksella* tarkoitan tutkimusta, jossa tallennetaan puhetta ja puheentuottotilanteita esimerkiksi muistiinpanolitte-roimalla, nauhoittamalla tai tallentamalla niitä digitaalisesti sekä käsittelemällä puheaineistoa esimerkiksi litteroimalla ääni- ja videotallenteita tai tekemällä foneettisia mittauksia.

tai puheen nimeäminen ”aidoksi” tuntuu tarpeettomalta todennäköisesti siksi, että puhutun kielen tutkimus on lähtökohtaisesti kieltä käytössä tarkastelevaa².

Väitöstutkimukseni (Aarikka 2023: 147) analyysiprosessissa kuitenkin havainnoin, että tutkimuksen (tavoiteltu) kohde saatetaan myös puhutun kielen tutkimuksessa nimetä esimerkiksi *luonnolliseksi*, *luontevaksi*, *tuttavalliseksi* tai *spontaaniksi arkikieliksi*, kuten esimerkissä 1 (esimerkkien lihavoinnit kirjoittajan lisäämiä, esimerkkien esittämistävasta ks. artikkelin luku 2).

- (1) Itse puhekielen tutkimuskohde – **arkikieli, luonnollinen (puhe)kieli tai spontaani puhekieli (*vernacular*)** – on käsitteenä sumea.

VK 41 (2005), s. 64

Siihen, että puhutun kielen tutkijoiden keskeisenä kiinnostuksen kohteena on *spontaan*i ja *luonnollinen*, *vapaamuotoinen* puhe, on kiinnitetty huomiota myös tutkimuskirjallisuudessa – samalla todeten, että käsite on monimerkityksinen abstraktio (ks. Kurki & Mustanoja 2020: 284–285; Laakso 1996: 222–223).

Väitöskirjassani (Aarikka 2023) tutkin fennistisen murteentutkimuksen historiaa ja kieli-ideologioita neljästä näkökulmasta: tarkastelin murreaineistoja, murteenpuhujia, murteen käsitettä sekä murteentutkijaa toimijana. Tässä tutkimuksessa pureudun väitöskirja-aineistoni valossa visaisiin ja perustavanlaatuisiin kysymyksiin: Mitä *vapaa* (myös *aito*, *luonnollinen*, *spontaan*i ja *arkinen*) *puhe* oikeastaan on? Miten *vapaan puheen* käsite on fennistisen murteentutkimuksen historian aikana muuttunut? Miten käsitykset siitä, mikä on *vapaata puhetta*, vaikuttavat tutkimusten kysymyksenasetteluihin, tutkimuskohteiden määrittelyyn ja siten tutkimustuloksiin?

Tutkimukseni tarkoitus on tarkastella aineistopohjaisesti sitä, millaisia käsityksiä fennistisessä murteentutkimuksessa on esitetty *aidosta*, *vapaasta*, *luonnollisesta*, *spontaanista* ja *arkisesta* puheesta ja kielenkäyttötilanteista. Aineistopohjaisen analyysin lisäksi tukeudun analyysissäni tutkimuskirjallisuudessa esitettyihin määritelmiin. Analyysini tarkoitus tuottaa tiedeyhteisölle välineitä reflektoida sitä, millaisia käsityksiä tutkimuksessa tuotetaan puhutusta kielestä tutkimuskohteena. Tavoitteeni on lisätä kielentutkimuksen ja kielentutkijoiden itseyttä siitä, miten tutkimuskohdetta – *vapaata puhetta* – määritellään ja mitä seurauksia sillä on.

2 Aineisto, menetelmät ja analyysin prosessi

Vastaan edellisessä luvussa esittämiini tutkimuskysymyksiin tarkastelemalla monografiamuotoisessa väitöskirjatutkimuksessani *Murre ja sen tutkimus: Näkökulmia fen-*

² Toisaalta voi ajatella, että kielentutkijoiden esimerkkien keksiminen on myös kielenkäyttöä.

nistisen murteentutkimuksen historiaan ja kieli-ideologioihin 1871–2017 (Aarikka 2023) käyttämäni aineistoa: vuosina 1871–2017 ilmestyneiden murreväitöskirjojen (N=51) osista koostettua tekstiaineistoa. Tekstiaineistoni väitöskirjoja, rajausperusteita sekä aineiston käsittelyä kuvaan tarkemmin väitöskirjassani (Aarikka 2023: 24–31, 36–37).

Esimerkkejä esittäessäni viitataan tekstiaineistooni väitöskirjojen (VK) tekijöiden ja nimikkeiden sijaan tunnistenumeroilla (0–50) ja väitöskirjojen ilmestymisvuosilla. Ratkaisuun olen päätenyt siksi, että tarkoitukseni ei ole tehdä arvottavia huomioita yksittäisten tutkijoiden yksittäisissä tutkimuksissa tekemistä valinnoista vaan tarkastella aineistoa kokonaisuutena. Esitän tässä artikkelissa esimerkkejä neljästätoista väitöskirjasta, joiden täydet tekijä- ja julkaisutiedot löytyvät artikkelin lopusta aineslähdeluettelosta.

On tärkeää huomata, että tutkimuksen kohteena olevat väitöskirjat keskittyvät nimenomaan murteisiin tai alueelliseen variaatioon ja aineisto on siten painottunutta, mikä selittää osaa tekemistäni havainnoista. Aineiston painottuneisuuden vaikutusta reflektoin tarkemmin artikkelini analyysiluvuissa sekä päätelmä- ja pohdintaluvussa 4. Aineiston painottuneisuudesta huolimatta tutkimieni väitöskirjojen näyte on ajallisesti kattava. Tekstiaineistoni väitöskirjat ovat myös moninaisia, eikä niitä kaikkia voi kuvailla yksinomaan dialektologisiksi tai sosiolingvistiksi tutkimuksiksi (tarkemmin, ks. Aarikka 2023: 26, 206–254). Jotkin väitöskirjoista asemoituvat pragmatiikan (esim. Kuiri 1984) tai vuorovaikutuslingvistiikan (esim. Priiki 2017) alaan. Vaikka aineistoni ei sisällä kaikkia fennistiikan alan puhuttua kieltä käsitteleviä väitöskirjoja, tutkimastani väitöskirjojen näytteestä heijastuu eri tutkimusparadigmojen kuten keskustelunanalyysin ja tilanteisen vaihtelun tutkimuksen vaikutus fennistiseen puhutun kielen tutkimukseen (tarkemmin, ks. Aarikka 2023: 118–120, 243).

Artikkelini on lähiluvun ja sisällönanalyysin menetelmin ja kriittisen diskurssintutkimuksen diskurssihistoriallisen lähestymistavan keinoin toteutettu tutkimus. Kriittinen diskurssintutkimus on monipuolinen tutkimusala, jonka avulla tutkitaan kielen, ideologioiden ja siten vallankäytön suhteita (kriittisen diskurssintutkimuksen eri suuntauksista ks. esim. Pietikäinen 2000: 194–196; Reisigl & Wodak 2016; *ideologian* ja *diskurssin* käsitteistä kielitieteessä kokoavasti ja taustoittavasti ks. esim. Pietikäinen & Määttä 2014).

Sitoutuessani diskurssihistorialliseen lähestymistapaan metodina sitoudun myös sen diskurssikäsitteeseen ja diskurssin määritelmään. Martin Reisigl (2018: 51–52) esittää kymmenen diskurssihistoriallisen lähestymistavan diskurssimääritelmälle ominaista piirrettä, joista tutkimukseni kannalta keskeisiä ovat seuraavat: Diskurssilla on sekä sosiaalinen että semioottinen luonne ja diskurssit representoivat, luovat, uusintavat sekä muuttavat sosiaalista todellisuutta. Diskurssin semioottinen luonne aktualisoituu teksteissä eli tutkimissani väitöskirjoissa. Väitöskirjatekstit ovat tietty kielenkäyttötilanne (*discursive unit*), joilla on tietty sosiaalinen konteksti ja tehtävä: väitöskirjat ovat opinnäytteitä, joilla saavutetaan paitsi tohtorin tutkinto myös tohtorin asema tiedeyhteisössä. Tämä sosiaalinen konteksti ja tehtävä yhdistää väitöskirjatekstejä siitä

huolimatta, että väitöskirjatekstit ja tiedeyhteisö sekä niihin liittyvät konventiot ovat kiistatta muuttuneet (tarkemmin, ks. Aarikka 2023: 51–56). Useimmilla kielenkäyttötilanteilla kuten väitöskirjateksteillä on intertekstuaalinen luonne, eli väitöskirjatekstit ovat yhteydessä toisiinsa. Diskursseilla on tietyt toimintakenttensä (*field of action*), jotka kehystävät niitä. Näissä toimintakentissä diskurssit osallistuvat vallan ja tietynlaisen sosiaalisen järjestyksen (*social order*) ylläpitämiseen eli ne jäsentävät, tuottavat, uusintavat ja muokkaavat käsityksiä, tietoa sekä ideologioita. Eri topiikkien ympärille voi muodostua kilpailevia diskursseja, joita esitän tämän tutkimuksen analyysiluvuissa. Yritän aineistoni avulla ymmärtää kyseisiä kilpailevia diskursseja ja pohtia myös niiden välisiä ristiriitoja sekä ajallista muutosta.

Olen analysoinut tekstiaineistoni Nvivo-ohjelmalla. Analyysin prosessi on edennyt siten, että olen lukenut tekstiaineistoni kahteen kertaan. Ensimmäisellä kerralla pohdin teemoja ja näkökulmia, jotka aineistossani toistuvat. Lukiessani tekstiaineistoani toiseen otteeseen koodasin sitä. Käytännössä koodaaminen tarkoittaa Nvivo-ohjelmassa sitä, että tekstiosioita valitaan esimerkiksi maalaamalla ja raahaamalla ne tutkijan määrittelemään temaattiseen koodiin (*code*), jolloin kaikki kyseiseen koodiin merkityt tekstiosiot päätyvät yhteen näkymään, josta niitä voi koodata edelleen alakodeihin (*child code*).

Laadullisen lähilukuun perustuvan analyysiprosessin aikana tunnistin tämän artikkelin fokuksen kannalta keskeisiä ilmauksia, joita etsin tekstiaineistostani myös sanahauilla. Tällaisia ilmauksia ovat: *vapaa*, *spontaani*, *luonnollinen*, *aito* ja *arki*. Koodatessani pyrin tunnistamaan tekstiaineistostani diskurssihistoriallisen lähestymistavan diskursiivisia strategioita (*discursive strategies in DHA*, Reisigl 2018: 52), erityisesti nimeämistä (*nomination*) ja predikointia (*predication*) eli sitä, miten ja missä tekstiyhteyksissä eli koteksteissä kieliaineistoa, kieltä, puhetta tai puhetilanteita³ nimetään edellä mainituin ilmauksin.

Edellä kuvatun analyysiprosessin lopputulos on aineistonäyte, joka sisältää em. ilmausten ilmenemiskotekstejä yhteensä 31:stä tekstiaineistoni väitöskirjasta (N = 51). Tekstiaineistoni (N = 51) ensimmäinen väitöskirja on ilmestynyt vuonna 1871, mutta ensimmäinen väitöskirja, jossa mainitaan jokin koodaamistani ilmauksista (*spontaani*) on väitöskirja 13, joka on ilmestynyt vuonna 1963. Yksittäisestä tekstiaineistoni väitöskirjasta koodattujen ilmenemiskotekstien lukumäärä vaihtelee yhdestä 19:än.

Olen koonnut taulukkoon 1 aineistoani kuvailevia tietoja. Taulukosta käy ilmi, miten monta ilmenemiskotekstiä olen kustakin ilmauksesta koodannut sekä se, miten monessa väitöskirjassa (n = 31) ilmausta käytetään. Lisäksi taulukossa on tieto siitä,

3 En tässä artikkelissa tarkastele sellaisia *vapaa*, *spontaani*, *luonnollinen*, *aito* ja *arki*-ilmenemiä, joissa kuvailtava ilmiö on jokin muu kuin kieliaineisto, kieli, puhe tai puhetilanne. Tällaisia ilmauksia ovat esimerkiksi johto-opin alalla käytetty *spontaani johtaminen* sekä kielikäsitustutkimuksissa käytetty ilmaus *spontaani murteen kuvailu*. Olen rajannut tarkastelun ulkopuolelle myös ilmauksen *luonnollinen kieli*, kun se viittaa kieleen, joka ei ole konekieltä.

missä ja koska ilmestyneessä väitöskirjassa kutakin ilmausta käytetään ensimmäisen ja toisaalta viimeisimmän kerran.

TAULUKKO 1. Analyysin kohteena olevat ilmaukset ja niiden ilmenemiskotekstien määrä, sekä tiedot siitä, miten monessa tekstiaineistoni väitöskirjassa (n = 31) ilmausta käytetään ja missä väitöskirjassa ilmaukset ilmenevät ensimmäisen ja viimeisimmän kerran.

Ilmaus	Koodattuja ilmenemiskotekstejä	Ilmenee väitöskirjassa (n = 31)	Ensimmäinen ilmenemä	Viimeisin ilmenemä
<i>vapaa</i>	24	12	VK 18 (1972)	VK 48 (2016)
<i>spontaani</i>	50	20	VK 13 (1963)	VK 49 (2017)
<i>luonnollinen</i>	17	7	VK 22 (1984)	VK 48 (2016)
<i>aito</i>	50	21	VK 15 (1971)	VK 49 (2017)
<i>arki</i>	44	14	VK 22 (1984)	VK 49 (2017)

Käytyäni kaikki ilmauksittain kootut koodaukset koodasin niitä edelleen ja pyrin hahmottamaan, mitkä diskurssit koteksteissä muotoutuvat. Analyysin prosessia voi kuvata hermeneuttiseksi kehäksi. Hermeneuttinen kehä tarkoittaa tutkimusmenetelmää, jossa ymmärrys tutkimuksen kohteesta syvenee vähitellen (Pietikäinen & Mäntynen 2009: 141–144; Tieteen termipankki 22.10.2022: Filosofia:hermeneuttinen kehä).

3 Aineiston analyysi

Lähiluvun ja analyysin prosessissani havaitsin kolme keskeistä *aidon*, *vapaan*, *spontaanin*, *luonnollisen* ja *arkisen* ilmenemiskotekstien tyyppiä, joita käsittelem tässä analyysiluvussa. Ensimmäinen niistä on menetelmien vaikutus puheen *vapauteen* ja *spontaaniuteen* sekä kieliaineiston *aitouden* arviointi (luku 3.1). Toinen (luku 3.2) on puhetilanteiden (haastattelu ja keskustelu) *spontaanius* ja *vapaus*. Kolmantena (luku 3.3) käsittelem kokoavasti *vapaata* ja *luontevaa (arki)puhetta* tutkimuksen kohteena, mitä sivusin jo tämän artikkelin johdannossa.

3.1 Kyselymenetelmien ja koeasetelmien vaikutus puheen vapau- teen ja spontaaniuteen sekä puheaineiston aitouden arviointi

AFinLAN Puheen ja puheentutkimuksen teemaverkoston työpajan abstraktikutsussa (6.5.2022) *vapaasti tuotettu puhe* määriteltiin melko laveasti: sen ulkopuolelle rajattiin vain lukupuhe.⁴ Myös aineistostani (esimerkki 2) käy ilmi, että kehyslauseissa⁵ tuotetut puhunnokset erotetaan *spontaanista puheesta* ja pohditaan myös molempiin puhe-
muotoihin kohdistuvan tutkimuksen etuja:

- (2) Kokeellisia mittauksia olen tutkimuksessani käyttänyt kuuloarvioiden tukena. Mittausaineistona on **spontaanialpuhetta** neljältä kielenoppaalta sekä **kehyslausemateriaalia** yhdeltä. Kummassakin aineistossa on etunsa. Spontaanialpuhetta mittaamalla voidaan selvittää, millaisia ovat todelliset idiolektierot ja miten suurta on idiolektinsisäinen vaihtelu. Kehyslauseisiin sijoitettujen minimiparien avulla taas voidaan tarkemmin tutkia konsonanttien kestoa eri asemissa, spontaanialpuheestahan tällaisia vain yhden äännepiirteen verran toisistaan poikkeavia muotoja on vaikea löytää suurestakaan aineistosta.

VK 26 (1987), s. 19

Kehyslauseiden avulla tuotettu puhe on kuitenkin tutkimuksen kohteena vain muutamassa tekstiaineistoni väitöskirjassa, joissa hyödynnetään myös fonetiikan mittausmenetelmiä (tarkemmin, ks. Aarikka 2023: 134–137). Sen sijaan useassa tekstiaineistoni väitöskirjassa kommentoinnin kohteena ovat muut provokaatioksi tulkittavat kyselymenetelmät, joita kielentutkimuksessa on hyödynnetty. Esimerkistä 3 käy ilmi, että kyselyä on pyritty välttämään (*on ollut pakko turvautua*) ja että *spontaanialpuheen* kuuntelua on pidetty *luotettavana* menetelmänä suhteessa kyselyyn:

- (3) Käyttämäni materiaali, joka on peräisin vuosilta 1954–1962, perustuu valtaosaltaan omiin muistiinpanoihini ja nauhoitteisiini. Havaintoni olen yleensä tehnyt **kielenoppaiden spontaanialpuheesta; vain muutamissa tapauksissa on ollut pakko turvautua suoranaiseen kyselyyn**. Menetelmä on hidas ja vaivalloinen, mutta sen **tulokset ovat luotettavia**.

VK 13 (1963), s. 7

4 Puheen ja puheentutkimuksen teemaverkoston AFinLAN syysseminariumin abstraktikutsu (6.5.2022): ”Tarkoitamme vapaasti tuotetulla puheella kaikkea tutkimustarkoituksiin kerättyä aineistoa, jossa tutkittavat eivät lue valmista tekstiä. Tutkimuksen kohteena voi olla esimerkiksi foneettiset segmentaaliset ja prosodiset piirteet tai havaitsemiseen liittyvät tekijät vapaasti tuotetussa puheessa.”

5 Kehyslause (myös *muottilause*) on Tieteen termipankin (15.9.2023: Kielitiede:muottilause) määritelmän mukaan ’foneettisissa tutkimuksissa käytetty tutkittavan ilmiön ympärille konstruoitu kehys’.

Kaikki analysoimani ilmenemiskotekstit, joissa käsitellään kyselemällä hankitun aineiston ja *vapaan puheen* eroja, eivät liity tutkijan itse keräämään aineistoon. Tekstiaineistoni väitöskirjoissa otetaan myös kantaa erityisesti kahden suuren suomen kielen aineistokokoelman luotettavuuteen kyselymenetelmien näkökulmasta. Tällaisia aineistoja ovat Suomen murteiden sana-arkisto (SMSA, esimerkki 4) ja Muoto-opin arkisto (MA, esimerkki 5), joita hyödynnetään useassa tekstiaineistoni väitöskirjassa (ks. myös Aarikka 2023: 79–87).

- (4) Verrattuani kokoelmatietoja Keräilysanastoon ja Lönnrotin sanakirjaan olen päättänyt pitääni **tutkimukseni kannalta epäluotettavina** useita SS:n vanhoja pitäjänkokoelmia (liite 2). Näiden kokoelmien tietoja olen kelpuuttanut aineistooni vain erityisestä syystä. Yleensä olen pitänyt **muistiinpanoa luotettavana**, mikäli asianomaista sanaa ei mainita Lönnrotin sanakirjassa eikä Keräilysanastossa. **Aidoiksi** olen katsonut myös sellaiset muistiinpanot, joissa sana muodoltaan tai merkitykseltään poikkeaa tavanomaisesta tai joissa on **poikkeuksellisen aidontuntuinen esimerkki**. Osoituksena **muistiinpanon aitoudesta** olen yleensä pitänyt myös sitä, jos sana on myöhemmin samassa asussa merkitty samasta pitäjistä tai lähiympäristöstä.

VK 19 (1972), s. 29

- (5) Suomen murteiden sanakirjan kokoelmat (SMSA) taas ovat morfosyntaktisen tutkimuksen tarpeisiin usein **erittäin epäluotettavia, ainakin silloin kun vastauksia ohjaa valmis kysely**. Tämän osoittivat itselleni mm. Sanastaja-lehden modaalisesta *tarvita*-verbistä tekemän kyselyn (n:o 2027) vastaukset, mutta ongelma koskee kyselyin koottua arkistoinenista yleensäkin. **Parhaan osan arkistokokoelmista ovat koulutetut kerääjät poimineet tai litteroineet spontaanista puheesta, joskus aidon keskustelunkin sivustakuulijoina**. Jo nesessiiviverbien polysemia vaatisi kuitenkin tilannekontekstia koskevan tiedon lisäksi yleensä arkistolipuille merkittyä laajemman tekstiyhteyden tuntemista. Tämä on MA:n aineistonkin ongelma.

VK 32 (1992), s. 30–31

Esimerkeissä 4 ja 5 toistuvat *luotettavuuden* ja *epäluotettavuuden* arviointi. Esimerkissä 4 tutkija pohtii Suomen murteiden sana-arkiston kokoelmien eli puheesta tehtyjen muistiinpanojen *aitoutta* ja selvittää, millä perustein *aitoutta* arvioi (*poikkeuksellisen aidontuntuinen erimerkki*, ks. myös Laakso 1996). Esimerkissä 5 huomionarvoista on myös tutkijan käyttämä ilmaus *joskus aidon keskustelunkin sivustakuulijoina*, jonka tulkitsen kannanotoksi liittyen siihen, että *aito keskustelu* edustaa parasta mahdollista *spontaanin puheen* alakategoriaa⁶. Puhetilanteisiin suhteessa *aitoutteen* ja *spontaanitteen* palaan seuraavassa analyysiluvussa 3.2.

6 Tulkintaani vaikuttaa *-kin* -liitepartikkeli, jonka ymmärrän esimerkin kontekstissä merkityksessä 'jopa': "Kun *-kin* kiinnittyy muuhun sanaan kuin verbiin tai pikkusanaan eli partikkeliin, kuten *varsin* tai *silloin*, sen tavallisin merkitys on 'myös': minäkin 'minä myös'. Usein sävy on paljoksuva 'jopa'; 'peräti'" (Kielitoimiston ohjepankki 2024, sähköinen lähde.)

On tärkeää huomata, että sekä SMSA:n että MA:n kokoelmat on koostettu muistiinpanolitteroiden eli kuulonvaraisesti puhetilanteessa muistiinpanoja tekemällä, joskin myöhemmin perustetun ja koostetun Muoto-opin arkiston keruiden yhteydessä on tehty nauhoituksiakin (ks. esim. Juusela 1989: 26) mutta epäsystemaattisesti. Kerääjät merkitsivät puhetta ja sen sisältämiä kielenilmiöitä standardoiduille korteille tai lipuille, jotta aineiston arkistointi ja siten hyödyntäminen olisi mahdollisimman nopeaa ja mutkatonta työlääseen litterointiin verrattuna (ks. esim. MO keruuopas: 9–11, 14–19; Strandberg 2004: 27–34). Muistiinpanolitteroinnista johtuen arkistoaineistoa on siis jouduttu arvioimaan myös sen luotettavuuden – ja joskus tiettyjen kokoelmien kerääjien tai tutkijoiden luotettavuuden – näkökulmasta siitä huolimatta, että kyseiset arkistoaineistot ovat keruuohjeiden perusteella *spontaanin puheen* tarkkailun ja kyselyn yhdistelmämenetelmällä koostettuja (Nuolijärvi 1988: 118–122; Strandberg 2004: 28; Kurki & Mustanoja 2020: 277–278). Kiinnostavaa on, että muoto-opin keruuohjeissa ensisijaisena keruumenetelmänä mainitaan juuri *spontaanin puheen* kuuntelu ja kielenkäyttäjän puheen tarkkaileminen ”jokapäiväisen elämän tilanteissa” (MO keruuopas: 11–12).

Aineistostani sekä arkistojen keruuohjeista voi saada sellaisen käsityksen, että erilaisissa kyselymenetelmissä sekä koeasetelmissä on pelkkiä heikkouksia ja että tutkijoiden ainoa kiinnostuksen kohde on *spontaanin puhe* sekä sen tarkkailun avulla havaittavat ilmiöt (Aarikka 2023: 92–97). Kahdessa tekstiaineistoni väitöskirjassa kuitenkin tuodaan myös esiin, että erilaisilla kyselymenetelmillä ja koeasetelmilla on mahdollista saada tietoa sellaista kielenilmiöistä, jotka eivät ole *spontaanissa puheessa* kovinkaan frekventtejä eli suurtaajuisia (esimerkki 6, ks. myös esimerkki 2), tai että aineisto voi olla vinoutunut siitä johtuen, että *luontevassa arkikeskustelussa* puhujat eivät tuota kaikkien kielenilmiöiden esiintymiä tasaisesti (esimerkki 7).

- (6) Mon. genetiivi ei ole kovinkaan suurtaajuinen muotoryhmä (ks. esim. Haikola, Vir. 1950 s. 400 ja Åstedt, Vir. 1957 s. 427; Haikolan ja Åstedtin tarkkailemissa murreteksteissä on esim. mon. partitiiveja noin kolminkertainen määrä mon. genetiiveihin verrattuna). Vaikeutena on kuitenkin erityisesti se, että **spontaanin puheeseen perustuva mon. genetiiviaineisto jakautuu vartalotyypeittäin epätasaisesti.**

VK 20 (1974), s. 15 (alaviite)

- (7) Kielitieteessä kuitenkin on melko tavallista, että dataa ei ole saatavilla niin paljon kuin olisi ihanteellista ja tarkasteltavat ryhmät ovat epätasaisia. Kun **kyse ei ole koeasetelmasta vaan tavoitteena on tarkastella luontevaa arkikeskustelua**, äänitetyt puhujat eivät tuota tarkastelun kohteena olevia esiintymiä tasaisesti. Tarkasteltava data voi siis sisältää tyhjiä ja pieniä soluja. (Tagliamonte & Baayen 2012.) **Kaikki puhujat eivät esimerkiksi tuota spontaanisti keskustelussa viittauksia** kaikenikäisiin henkilöihin.

VK 49 (2017), s. 39

Kiinnostavaa on, että aineiston arvioinnissa käytetään *luotettavuuden* (ks. esimerkit 3, 4 ja 5) lisäksi ilmausta *aito*. Samaa ilmausta – joskin lainausmerkeissä – käyttää myös Johanna Laakso (1996) artikkelissaan, jossa hän selvittelee "aitouden" ja "epä-aitouden" käsitteitä sekä "aitouden" lingvistisiä kriteerejä, kun tarkoitus on arvioida muistiinpanolitteroidun aineiston käytettävyyttä ja luotettavuutta. On totta, että jos aineistokokonaisuuden on määrä sisältää kielen käyttöön perustuvaa kielenainesta (kuten SMSA:n ja MA:n), kyselymenetelmien hyödyntämisen voi ajatella tuottavan epäluotettavaa tietoa kielen provosoimattomasta käytöstä. Toisin sanoen: Jos tutkimuskysymykset koskevat *spontaania* kielenkäyttöä, ovat provosoimalla eli kyselemällä tuotetut kieliaineistot tietenkin kyseiseen tutkimukseen soveltumaton aineisto. Jos aineistolle halutaan esittää kysymyksiä siitä, miten kielenkäyttäjät pystyvät produktiivisesti johtamaan sanoista tai muodoista toisia, provosointi ei näyttäydy ongelmana. Erilaisia kyselymenetelmiä myös käytetään kielentutkimuksessa jatkuvasti (ks. esim. Vanhatalo & Vehkalahti 2020), esimerkiksi tuottamassa tietoa ihmisten kielikäsitteistä. Kysymys ei siten nähdäkseni ole varsinaisesti aineiston *aitoudesta* vaan siitä, ettei aineisto edusta sitä, mitä sen ajatellaan edustavan ja mitä tutkija haluaa tutkia.

3.2 *Spontaanius ja luonnollisuus puhetilanteissa*

Ilmauksella **puhetilanne** tarkoitan kielentutkimuksessa käytettyjä käsitteitä ja termejä, joilla erilaisia kielenkäyttötilanteita määritellään ja eritellään toisistaan. Puhetilanteen sijaan aineistossani käytetään myös termejä ja käsitteitä kuten **tekstilaji**, **tilanneytppi** ja **diskurssirakenne** – käytetty terminologia voikin olla moninaista (ks. esim. Lappalainen 2004: 42). Tieteen termipankin (Kielitiede:puhetilanne, 14.9.2023) määritelmän mukaan puhetilanne on 'kielenkäyttötilanne, jossa on mukana vähintään puhuja ja kuulija'.

Edellisessä luvussa käsittelemäni puheen keruumenetelmiin ja kieliaineiston tallennusmenetelmiin liittyvä keskustelu kytkeytyy osittain myös teknologian kehitykseen: muistiinpanojen ja siten puheen *aitoutta* on arvioitu, koska saatavilla ei ole ollut tallennetta, jota voisi verrata muistiinpanolitteraattiin. Muistiinpanoja käytettiin aineistonkeruumenetelmänä vielä 1980-luvulle saakka siitä huolimatta, että puheen tallentaminen kelanauhoille alkoi yleistyä jo 1950-luvulla, kun nauhoitearkistot perustettiin (Aarikka 2023: 98–106). Näkemykseni mukaan puhetilannetta ja eritoten haastattelua ja keskustelua aletaan teoretisoida ja kuvailla käsitteinä 1970–1980-luvulla, jolloin puheen tilanteista vaihtelua ja keskustelua alettiin tutkia suomen kielen alalla (keskustelunalyysin historiasta ja tutkimuskohteista, ks. Hakulinen 1995). Tilanteisen vaihtelun tarkasteleminen ei olisi ollut mahdollista ilman teknologista kehitystä ja koneavusteista puheen tallennusta, mutta on tärkeä muistaa, että tutkijat ovat olleet tietoisia ihmisten puhuvan eri tavoin eri tilanteissa jo kauan ennen tilanteisen vaihtelun

tutkimusta (ks. esim. Kettunen 1910: 59), vaikka tilanteista vaihtelua sinänsä on ollut vaikeaa ellei lähes mahdotonta tutkia systemaattisesti ilman tallenteita.

Aineistossani puhetilanteiden eroja pohditaan ensimmäistä kertaa väitöskirjassa 29 (1984, esimerkki 8), jossa tutkija viittaa Barbara Sandigin (1975) tutkimuksiin erilaisten puhetilanteiden (*tekstilajien*) ominaispiirteistä⁷:

- (8) Sandigin vertailu antaa aiheen olettaa, että haastattelu ei itse asiassa ole niin paljon "luonnollisesta" keskustelusta poikkeava tekstilaji kuin etukäteen ehkä on taipumus olettaa.

VK 29 (1984), s. 57

Esimerkissä 8 kiinnostavaa on, että sen muotoilusta (*etukäteen ehkä on taipumus olettaa*) on tulkittavissa, että haastattelu ja keskustelu on nähty väitöskirjan ilmestymisen aikaan eli vuonna 1984 toisistaan poikkeavina puhetilanteina, joista keskustelu nimetään (tosin lainausmerkeissä⁸) "luonnolliseksi" (vrt. Seppänen 1995: 18). Näin ollen *luonnolliseksi* ei arvioida ja nimetä pelkästään puhetta vaan myös puhetilannetta.

Tutkimissani väitöskirjoissa toisinaan myös *spontaaniksi* nimetty *haastattelu* on paitsi puhetilanne myös kielentutkimuksen menetelmä (Aarikka 2023: 106–120; ks. myös Kurki & Mustanoja 2020), jolla puheaineistoa kerätään. Keskusteluanalyttisen tutkimuksen piirissä on korostettu eroa tutkijan järjestämän puhetilanteen (kuten haastattelu) ja ("aidon") vuorovaikutuksen eli keskustelun välillä, kuten Eeva-Leena Seppänen (1995: 18) kirjoittaa:

Keskusteluanalyttinen tutkimus kohdistuu "aitoon" vuorovaikutukseen: vuorovaikutustapahtumiin, joilla on oma, tutkimuksesta erillinen tavoitteensa ja jotka olisivat tapahtuneet ilman tutkimustarvettakin. Tällaisia vuorovaikutustapahtumia ovat vaikkapa ystävien välinen puhelinkeskustelu, lääkärin ja potilaan välinen keskustelu ja perheen ruokapöytäkeskustelu. Keskusteluanalyttikot eivät siis tutki koeasetelman kautta järjestettyjä vuorovaikutustapahtumia, vaan heitä kiinnostavat erilaiset arjen vuorovaikutustilanteet.

Tämä puhuttuun kieleen kohdistuvan tutkimuksen monimuotoistuminen laukaisee tulkintani mukaan aineistossani erottuvan mielenkiintoisen ja paikoin perin ristiriitaisen diskurssien kilpailun siitä, onko haastattelu *vapaa* ja *spontaani* puhetilanne vai ei.

7 Aineistossani puhetilanteiden "haastattelu" ja "keskustelu" ominaispiirteitä ja/tai eroavaisuuksia arvioidaan seuraavien muuttujien näkökulmasta, mutta harvoin systemaattisesti kaikista tässä mainituista näkökulmista: 1) monologisuus ja dialogisuus sekä vuorojen vaihtuminen, 2) kysymys-vastaus -rakenne, 3) keskustelijoiden määrä, 4) puheenaiheet ja 5) tasavertaisuus puhujien kesken.

8 Kiitän nimitöntä artikkelikäsitelmän kirjoittajaa siitä huomiosta, että lainausmerkkien käyttö tuo aineistoesimerkin 8 tulkintaan moniäänisyyttä: kirjoittaja saattaa osoittaa lainausmerkeillä, ettei sitoudu näkemykseen (Nordlund 2005: 344, 349–345).

34 "TAVOITTEENA ON TARKASTELLA LUONTEVAA ARKIKESKUSTELUA" – VAPAA, SPONTAANI, LUONNOLLINEN JA AITO (ARKI-)PUHE FENNISTISISSÄ MURTEITA KÄSITTELEVISSÄ VÄITÖSKIRJOISSA

Esimerkissä 9 tutkija tuo ensin esiin haastatteluun kohdistuvaa kritiikkiä. Kiintoisaa on, että kritiikin ytimestä hän nostaa esiin kaksi seikkaa: nauhoittamisen sekä haastattelun kahden ihmisen välisenä puhetilanteena (vrt. alaviite 7). Esimerkin 9 toisesta virkkeestä käy selkeästi ilmi, että tutkimuksen tavoiteltu kohde on puhujan *luonnollinen puhetapa*.

- (9) Luotettavan aineiston saamista **nauhoitetun, kahdenkeskisen** haastattelun avulla on epäilty (ks. esim. Gumperz 1972: 208; Douglas-Cowie 1978: 39–40). Useimmiten haastattelussa silti päästäneen tilanteeseen, jossa kielenopas käyttää **luonnollista puhetapaansa**.

VK 26 (1987), s. 18

Esimerkissä 10 tutkija puolestaan argumentoi aineistoesimerkkien avulla, miten haastattelu ei välttämättä eroa *arkipäiväisestä keskustelusta*, sillä haastattelijan ja haastateltavan roolit voivat murtua tavalla, jossa haastateltava johdattelee keskustelua tai esittää haastattelijalle kysymyksiä (vrt. alaviite 7):

- (10) Monesti **haastattelusta kehkeytyy** niin tutkijan kuin tilanteeseen osallistuvien puhujienkin näkökulmasta **varsin arkipäiväinen keskustelu**, jonka aikana alkuasetelmassa vallinneet roolit saattavat unohtua (esimerkki 1) tai kokonaan murtua (esimerkki 2) (vrt. Milroy & Gordon 2003: 61–62).

VK 46 (2011), s. 81

Haastattelupuheen luonnollisuuden (esimerkki 9) ja keskustelunomaisuuden (esimerkki 10) lisäksi useassa tutkimistani väitöskirjoista tuodaan esiin, että jokaisella kielenkäyttötilanteella on tietynlaiset rajansa (ks. myös Kurki & Mustanoja 2020: 285), eikä minkäänlainen puheaineisto kata koko idiolektin eli yksilön kielenkäytön kirjoa, kuten esimerkiksi 11:

- (11) **Vaikka haastattelutilanne onnistuisikin olemaan suhteellisen vapaamuotoinen, on selvää, että siinä tavoitetaan idiolektista vain osa.** Haastatteluaineisto ei luonnollisestikaan anna riittävästi informaatiota siitä, miten samainen puhuja käyttäytyy kioskillä asioidessaan (vrt. Lappalainen & Raevaara 2009). Se ei liioin kerro siitä, millaisiin sanankääntein puhuja suoriutuu julkisesta puhetilanteesta tai miten hän rupattelee lähimpien ystäviensä tai perheensä parissa. **Samalla tavoin asiointipuheaineisto tai julkisten puhetilanteiden nauhoitteet eivät riitä kuvaamaan idiolektin koko variaatiota.** (Kurki 2007: 151.)

VK 46 (2011), s. 84

Tästä huolimatta tulkintani mukaan useassa tutkimistani väitöskirjoista tehdään selkeä eron teko haastattelun ja keskustelun välillä ja saatetaan jopa todeta, että haastattelu on menetelmänä vanhentunut verrattuna keskustelutaltiointeihin (esimerkki 12) tai

ettei haastattelu ole (*sanan tiukassa merkityksessä*) vuorovaikutusta (esimerkki 13, vrt. Seppänen 1995: 18).

- (12) Dialoginakin etenevä **murrehaastattelu eroaa toisaalta spontaanista ja tuttavallisesta keskustelusta**: sen teema on etukäteen suunniteltu eikä puhekumppaneiden suhde ole tasaveroinen (ks. Kuiri 1984: 55–60 viitteineen). Spontaanin **murrekeskustelun nauhoittaminen** olisi jo pitkään ollut mahdollista, mutta viimeisten parinkymmenen vuoden aikana **tallennusvälineiden kehitys on ollut kielitieteellisten aineistonkeruumenetelmien kehitystä nopeampaa**.

VK 32 (1992), s. 31–32

- (13) Voi lisäksi todeta, että **murrehaastattelu ei kenties ole vuorovaikutusta** sanan tiukassa merkityksessä, koska vuorot eivät vaihdu kuten arkikeskustelussa, mutta se on omanlaistaan puhuttua kieltä, koska puhujalla on siinä mahdollisuus kertoa asioista laajasti.

VK 48 (2016), s. 13

Erityisen kiinnostavana pidän sitä, että aineistossani toisaalta korostetaan haastattelun elastisuutta puhetilanteena (ks. esimerkki 10) ja kiinnitetään huomiota siihen, että haastattelut ovat keskenään erilaisia vuorovaikutustilanteita, jotka sisältävät myös lyhyempiä, toisistaan eroavia jaksoja⁹, ja toisaalta tuodaan esiin, että haastattelu on kategorisesti tietynlainen puhetilanne (ks. esimerkit 12 ja 13). Tarkastelen tätä artikkelia laajemmin aineistoni väitöskirjoissa esitetyjä *murrehaastatteluun* liittyviä käsityksiä (Aarikka 2023: 106–120) sekä puhetilanteiden *haastattelu ja keskustelu* elastisuutta (Aarikka 2023: 143–147) väitöskirjassani.

Useassa aineistoni väitöskirjassa mainitaan myös *havainnoijan paradoksi* (*observer's paradox*), joka on sosiolinguisti William Labovin (1972) kehittänyt käsite. Esimerkissä 14 tutkija avaa havainnoijan paradoksin merkitystä ja tähdentää, että haastattelu voi sisältää *pilkahduksia vapaasta puheesta* tai tyylilaji voi *vapautua arkipuheen tasolle*:

- (14) Nauhoitettua haastattelua pidetään yleisesti käyttökelpoisena menetelmänä puhekielen aineiston hankkimiseksi; nimenomaan kahdenkeskiset haastattelut antavat usein hyvää aineistoa. **Kun tutkijan tarkoituksena on tavallisesti saada puhujalta tietoa vapaan puheen ilmiöistä**, nousee ongelmaksi ns. havainnoijan paradoksi (*the observer's paradox*): pyritään selvittämään, miten ihmiset puhuvat, kun heidän kielenkäyttöään ei tarkkailla, mutta tavallinen tapa saada tätä tietoa puhujilta on järjestää haastattelu, jossa heidän kielenkäyttöään tarkkaillaan (ks. esim. Labov 1975a: 509). (---) Usein tämäntapainen haastattelurekisteri määritellään puoliformaa-

9 Tutkimuskirjallisuudessa kyseisiä jaksoja kutsutaan esimerkiksi *tilanteen sisäiseksi vaihteluksi* (Lappalainen 2004: 49–52 ja siinä mainitut lähteet) ja *mikrotilanteiksi* (Mustanoja 2011: 115–123 ja siinä mainitut lähteet).

liksi tai puolivapautuneeksi: tämä on formaaleista tyyllilajeista vapautunein, ja lisäksi **haastattelijalla on mahdollisuus saada kesken haastattelun myös pilkahduksia vapaasta puheesta** (Labov 1975a: 79–94). **Mikäli haastattelu sujuu luontevasti, saattaa tyyllilaji vapautua arkipuheen tasolle** (Suojanen & Salomaa & Vuorinen 1985: 137). Keski-ikäiset puhujat ovat tutkijan kannalta hankalimpia, sillä heillä on käytettävissään yleensä laajempi rekisterivalikoima kuin vanhoilla tai nuorilla; lisäksi he ovat ilmeisen herkkiä sociolingvistisille normeille (esim. Chambers & Trudgill 1980: 91–95). Juuri keski-ikäisiltä onkin aineistoa joskus yritetty kerätä **ryhmähaastattelu**in (esim. Thelander 1979: 17–50, Labov 1973: 85), **joissa puhe kyllä saattaa muuttua vapautuneemmaksi**, mutta riittävän ja laadullisesti hyvän idiolekteittaisen aineiston kerääminen on tällä tavoin työlästä. Selvää on, että **kielellisesti parhaat näytteet vapaasta puheesta saa salanauhoituksella** tai muulla "suoralla" tarkkailulla, mutta niissä törmätään melkoisiin teknisiin ja eettisiinkin vaikeuksiin (ks. Labov 1975a: 509–518; laajemmin Samarin 1967).

VK 25 (1987), s. 9–10

Kiinnostavaa on, että *havainnoijan paradoksin* mukaan puheen tarkkailu saattaa vaikuttaa nimenomaan puheen *luontevuuteen* (ks. myös Tieteen termipankki 15.9.2023: Kielitiede:havainnoijan paradoksi) ja että havainnoijan paradoksiin liittyvät maininnat kytkeytyvät aineistossani aina nimenomaan haastatteluun menetelmänä ja puhetilanteena (ks. esimerkki 14; vrt. kuitenkin Nuolijärvi 1986: 34–35), vaikka myös keskusteluja tallennettaessa puhetilanteessa on läsnä laitteisto ja tutkittavien on eettisistäkin syistä tiedettävä, että heidän puhettaan tallennetaan ja tutkitaan.

Kokoavasti voi todeta, että aineistostani tekemieni havaintojen perusteella haastattelua kuvataan siis toisaalta *spontaanina* puhetilanteena, joka sisältää *spontaania* puhetta, mutta toisaalta järjestettynä puhetilanteena, joka ei ole yhtä *spontaani* ja *vapaa* kuin (*tuttavallinen arki-*) keskustelu. Edellä mainituista kilpailevista määritelmistä huolimatta yksimieliseltä vaikuttaa suhtautuminen siihen, että tutkimuksen tavoiteltava kohde on *aito, arkinen, luonteva* ja *luonnollinen* puhe, joka on mahdollisimman epäformaalia (ks. myös Laakso 1996: 222–223; Kurki & Mustanoja 2020: 278–281, 284–285).

Epäformaaliuden ja *spontaaniuden* ihanne käy ilmi myös esimerkistä 15, jossa tutkijan menetelmiin liittyvä metateksti keskittyy kuvailemaan sitä, miten aineistonkeruun tavoitteeksi tuli *spontaanisti etenevä keskustelu*, vaikka tarkoitus on analysoida haastattelukysymyksiin saatuja vastauksia sisällöllisesti ja vertailla niitä toisiinsa:

- (15) Aivan ensimmäisissä haastatteluissa noudatin ennalta laatimaani kysymysrunkoa melko tarkasti, mutta haastattelukokemuksen karttuessa pyrin sen tarkasta noudattamisesta eroon ja pyrin etenemään enemmän informantini ehdoilla. **Tavoitteekseni tuli edetä keskusteluissamme entistä spontaanimm**in ja kerätä talteen kaikenlaisia havaintoja, joita informanteilleni johtui keskustelun kuluessa mieleen.

VK 47 (2011), s. 23

Seuraavassa analyysiluvussa 3.3 käsittelenkin sitä, mistä nostin esimerkkejä jo tämän artikkelin johdannossa: *arkista, luontevaa ja luonnollista* puhetta puheentutkimuksen keskeisenä mielenkiinnon kohteena.

3.3 Tutkimuksen kohteena vapaa ja luonteva (arki)puhe

Kuten tämän artikkelin johdannossa kirjoitin, kiinnostuin väitöstutkimukseni (Aarikka 2023: 147) aikana tutkimuksen tavoitellusta kohteesta, joka nimetään väitöskirja-aineistossani esimerkiksi *luontevaksi ja tuttavalliseksi arkikieleksi*:

- (16) Puhetottumuksissa tämä hallinta ei välttämättä ilmene jyrkinä tyylinvaihdoksina, vaan pikemminkin jatkumonluonteisena sarjana toisaalta **enemmän murteellisuuksia tai arkipuheelle ominaisia piirteitä, toisaalta formaalisempien kielimuotojen aineksia**. Kielentutkijan ongelmista tunnetusti kouraantuntuvimpia on se, **miten saada talteen noista kielenpuhujan hallitsemista kielimuodoista kiinnostavinta, luontevaa ja tuttavallista arkikieltä**.

VK 29 (1989), s. 28

Tässä analyysiluvussa pohdin aineistoni sekä tutkimuskirjallisuuden avulla, mitä *arkininen, luonteva ja luonnollinen* puhe oikein on. Esimerkissä 16 nimetään myös kaksi kiintoisaa käsitettä: *arkipuhe* sekä *arkikieli*. Käsittelen tämän luvun lopussa siis myös *arkisuuden, arkipuheen, arkikielen ja arkikeskustelun* käsitteitä aineistoni valossa.

Kuten edellisessä luvussa totesin, William Labovin (1972) *havainnoijan paradoksi*-käsitteen keskeisenä ajatuksena on, että kielentutkijat ovat kiinnostuneita kielestä, jota kielenkäyttäjät ei pyri kontrolloimaan siksi, että puhetta tarkkaillaan. Labov (1972: 208) kutsuu tätä ”minimaalisesti monitoroidun puheen” kielimuotoa *vernakulaariksi* (*vernacular*, ks. myös esimerkki 1) ja määrittelee sen seuraavalla tavalla:

Not every style or point on the stylistic continuum is of equal interest to linguists. Some styles show irregular phonological and grammatical patterns, with a great deal of ‘hypercorrection’. In other styles, we find more systematic speech, where the fundamental relations which determine the course of linguistic evolution can be seen most clearly. This is the ‘vernacular’ – the style in which the minimum attention is given to the monitoring of speech. Observation of the vernacular gives us the most systematic data for our analysis of linguistic structure.

Aineistossani sekä luvussa 3.1 käsittelemäni kyselymenetelmien luotettavuuden epäileminen ja sen esittäminen, ettei kyselemällä saatu kieliaineisto ole *aitoa*, on hyvin linjassa Labovin edellä esitettyjen ajatusten kanssa. Puhetilanteisiin *haastattelu* ja *keskustelu* liittyvät aitouden ja vapauden kysymykset, joita käsitelin edellisessä luvussa

3.2, kytkeytyvät mielestäni Labovin määritelmään: tietyissä kielenkäyttötilanteissa eli puhetilanteissa tuotettu puhe on sitä aidompaa, mitä vapautuneempaa puhe on.

Tässä kohden on oleellista tuoda esiin aineistoni painottuneisuus, joka vaikuttaa tekemiini havaintoihin. Koska tutkimani väitöskirjat keskittyvät juurikin alueellisten kielimuotojen ja alueellisen vaihtelun (eli murteiden) tutkimukseen, epäformaaliutta ja "vapautuneisuutta" tavoitellaan myös siksi, että teoreettinen oletus on ollut – ja jossain määrin on edelleen – että formaaleissa puhetilanteissa puhujat käyttävät formaaleja puheen rekisterejä tai tyylejä (ks. esim. Labov 1972: 208; Suojanen 1985: 17–20; kritiikistä, ks. myös Lappalainen 2004: 43–45). Tämä teoreettinen oletus käy ilmi esimerkiksi 17, jossa tutkija pohtii nuorten puhujien vastauksien lyhyttä, sen vaikutusta puhekielisyyteen ja lyhyiden vastauksien merkitystä puhujan yleiskielisyyden kannalta:

- (17) Monet nuorten haastattelut ovat luonteeltaan sellaisia, että ne soveltuvat huonosti muun aineiston kanssa vertailtavaksi. 15–20-vuotiailla haastateltavillani oli taipumus suhtautua haastatteluun kuin tenttiin, eli juuri siten kuin yksilöhaastattelujen arvostelijat ovat halunneet tämän aineistonhankintamenetelmän nähdä. Lyhyet vastaukset eivät kuitenkaan välttämättä tee nuorten puheesta **äänne- ja muotopiirteiltään sen formaalimpaa kuin on heidän normaali arkipuheensa**, eikä vastausten lyhyys ole merkki yleiskielistymisestä.

VK 29 (1989), s. 30

Aineistossani on myös viitteitä siitä, että vapaamuotoisen ja epäformaalin puheen virran tavoittelemisessa ei ole kyse pelkästään sosiolingvistiikkaan kytkeytyvästä ajattelumallista ja diskurssista (vrt. Kurki & Mustanoja 2020: 284–285), sillä myös tutkimuksessa, jota tutkija ei määrittele sosiolingvistiseksi, tuodaan esiin vapaan puheen *aitous* ja siten ihanteellisuus. Esimerkistä 18 käy ilmi, miten tutkija korostaa henkilökohtaista suhdettaan puhujiin, joiden puhetta hän on havainnoinut esimerkiksi illalla työpäivän päätyttyä, ja nimittää näitä havainnoinnin hetkiä kiintoisasti paitsi *vapaaksi puheeksi myös vuolaaksi keskusteluksi*:

- (18) Erikaisen lämpimästi muistelen kielenoppaitani, niitä yli 300 vanhusta, joiden puheesta olen tutkimusaineistoni tallentanut. Enimmät keruumatkani olen tehnyt polkupyörällä siten, että tukikohtanani on ollut jokin kirkonkylän koulu ja laajoilla pitäjillä olen yöpynyt usein kielenoppaitteni kotitalojen pirteissä ja kamareissa. **Kirjani aidoimmat murre-esimerkit ovatkin vapaasta puheesta merkittyjä.** Antoisia ovat olleet varsinkin illat, **kun koko talonväki on tullut jo tutuksi vieraan kanssa ja on keskusteltu vuolaasti työpäivän päätyttyä.**

VK 18 (1972), s. 6

Esimerkissä 19 puolestaan tutkija kertoo yleisellä tasolla murrenauhoitusten tekemisen periaatteista ja toteaa, että tiettyjen topiikkien valinta (*lähipiirin vanhat asiat*) saa puhujat innostumaan niin, että *puhe pääsee vuolaasti virtaamaan*. Esimerkin 19 lopussa tutkija tuo esiin, että hänen kuvaamansa haastattelumenetelmä vastaa William Labovin sosiolingvististä haastattelua.

- (19) Kun heitä lisäksi haastatetaan [sic] heidän oman lähipiirinsä vanhoista asioista, katsotaan, että aihe innostaa heitä sen verran, että tilanteen ehkä aiheuttama jännittyneisyys katoaa ja **puhe pääsee vuolaasti virtaamaan aidolla, vanhalla murteella**. Kielenopas saa toimia haastattelussa ikään kuin asiantuntijana, kun haastattelija usein tahallises-tikin heittäytyy tietämättömän rooliin; tätä menetelmään on käyttänyt myös William Labov sosiolingvistikissa tutkimuksissaan.

VK 29 (1984), s. 53–54

Lopuksi tarkastelen sitä, miten aineistossani käsitellään *arkea*, *arkipuhetta* ja *arkikeskustelua*. Analyysini mukaan *arki* suhteessa kieleen näyttäytyy neljällä tapaa: 1) puhetilanteen ”arkisuutena” (esimerkki 20), 2) keskustelukumppaneiden tuttuutena (esimerkki 21), 3) topiikkien eli puheenaiheiden ”arkisuutena” (esimerkki 22) ja 4) jonakin kiinnostavana vaikeasti hahmotettavana ”yleisyytenä” ja ”jokapäiväisyytenä”, ”yleisenä” puhekielenä (esimerkki 23).

- (20) En ole myöskään hakeutunut suositusten mukaisesti nauhoittamaan informanttia mihinkään erilliseen tilaan, jossa olisimme olleet muilta ihmisiltä rauhassa kahden (vrt. Yli-Paavola 1970: 40), vaan olen pyrkinyt keskustelemaan heidän kanssaan **mahdollisimman normaalissa arkitilanteessa**, jossa on ympärillä muita ihmisiä.

VK 41 (2005), s. 66–67

- (21) Vaikka useimmilla varmasti onkin se tunne, että haastattelujen kieli ei ole aivan samantyyppistä kuin **perheenjäsenten välinen arkikeskustelu**, ei haastattelua pidetä vierasena; senhän osoittaa jo sekin, että Suomessa on niin helppo saada haastateltavia erilaisiin tutkimuksiin.

VK 23 (1986), s. 35

- (22) TRE-aineiston nauhoitukset on toteutettu sekä vuonna 1977 että vuonna 1997 niin, että haastateltavalle entuudestaan tuntematon yliopisto-opiskelija on käynyt **jututtamassa häntä arkipäivän asioista**.

VK 46 (2011), s. 80

- (23) **Arkipuheaineiston perusteella saadaan tietoa siitä, kuinka eri tavoin pronominit puhekielessä käyttäytyvät**. Esimerkiksi jos *se*-pronomini muistuttaa keskustelukäytössä enemmän persoonapronominia *hän* kuin muita demonstratiivipronominien

40 "TAVOITTEENA ON TARKASTELLA LUONTEVAA ARKIKESKUSTELUA" – VAPAA, SPONTAANI, LUONNOLLINEN JA AITO (ARKI-)PUHE FENNISTISISSÄ MURTEITA KÄSITTELEVISSÄ VÄITÖSKIRJOISSA

luokkaan perinteisesti ryhmiteltyjä pronomineja, onko sen luokittelu demonstratiiviksi perusteltua?

VK 49 (2017), s. 35

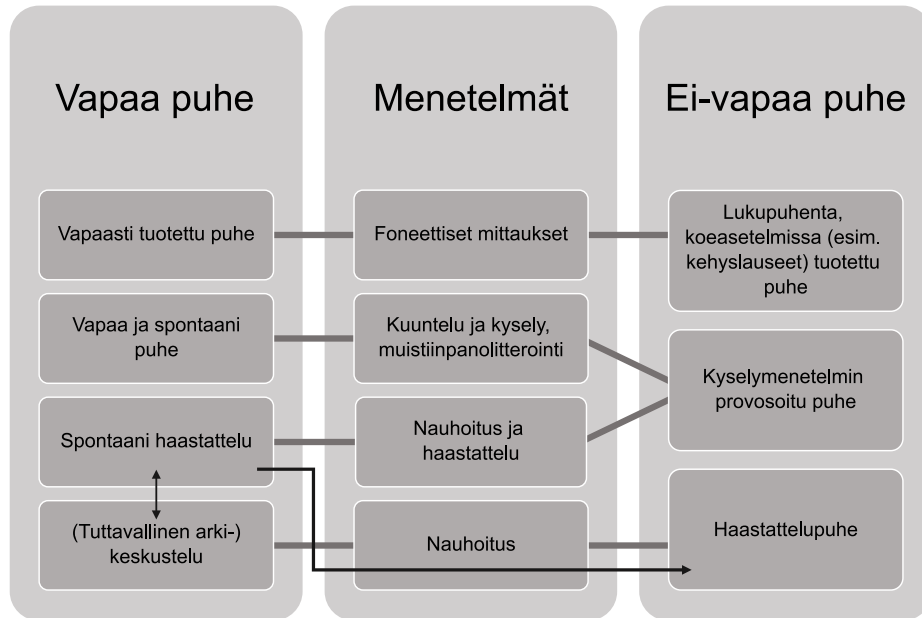
Edellä esittämieni esimerkkien kautta on mahdollista saada käsitys siitä, mitä kaikkea *arki* voi kielentutkimuksessa puhekielen kannalta tarkoittaa. On kuitenkin oleellista tuoda esiin, että *arkikeskustelu* (*everyday conversation*) on myös kielentutkimuksen termi, joka tarkoittaa keskustelua, jossa ei ole institutionaalisia rooleja¹⁰ (Hakulinen 2009; Suomalainen 2020: 31–32), jollaiseksi esimerkiksi haastattelijan ja haastateltavan roolit haastattelutilanteessa voi ymmärtää. Auli Hakulinen (2009: 58) kuitenkin tuo esiin, että jako institutionaalisen ja arkikeskustelun välillä on osittain menetelmällis-teoreettinen ja hypoteettinen, eivätkä erot keskustelutyypin eli puhetilanteiden välillä ole selviä ja kategorisia. Keskustelutyypit siinä missä muutkin puhetilanteet ovat tutkijoiden työkaluja – abstraktioita, jotka ovat hyödyllisiä ja käyttökelpoisia tutkimuksen tekemisen kannalta (ks. myös Hakulinen 2009: 58; Suomalainen 2020: 31–32). Aineistoanalyysistäni käy kuitenkin mielestäni ilmi, että puhetilanteisiin ja niiden kautta jäsenyvään puheeseen, sen *luontevuuteen*, *vapauteen* ja *arkisuuteen*, ei kuitenkaan ole suhtauduttu yksin abstrakteina ajattelun ja tutkimuksen apuvälineinä vaan hyvin konkreettisina hahmotelmina siitä, onko puhe *vapaata* ja siten kiinnostavaa tutkittavaa vai ei.

4 Päätelmät ja pohdintaa

Olen tässä artikkelissa käsitellyt aineistolähtöisesti sitä, kuinka *vapaaseen puheeseen* liittyviin käsityksiin ovat tutkimushistorian aikana vaikuttaneet paitsi keruumenetelmien ja niihin liittyvien teknologioiden muutos myös teoreettiset ja tutkimusalakohtaiset painotukset siitä, mihin puhutun kielen tutkimuksen tulisi keskittyä – toisin sanoen mikä puhutussa kielessä ja kielenkäytössä on tärkeää ja kiinnostavaa.

Olen koonnut asetelmaan 1 analyysini pohjalta laatimani synteessin, jossa olen pyrkinyt hahmottelemaan *vapaata puhetta* suhteessa paitsi fennistisessä murteen-tutkimuksessa ja kielentutkimuksessa hyödynnettyihin menetelmiin myös siihen, minkälainen puhe rajautuu *vapaan puheen* ulkopuolelle. Aineistoanalyysini ja sen perusteella laatimani asetelman (ks. kuvio 1) avulla vastaan ensimmäiseen tutkimuskysymykseeni: Mitä *vapaa puhe* oikeastaan on?

10 Tieteen termipankin (Kielitiede:arkikeskustelu, 18.9.2023) mukaan *arkikeskustelua* on 'puhuttu ja muu (ei-kielellinen) vuorovaikutus, jolla ihmiset toimivat jokapäiväisessä elämässään ilman erityisiä institutionaalisia rooleja'.



KUVIO 1. Vapaa puhe, menetelmät ja ei-vapaa puhe.

Luvussa 3.1 toin esiin, että *vapaan puheen* vastinpariksi (ja sen ulkopuolelle) muotoutuvat erilaiset koeasetelmissa tallennetut puhunnokset, kuten lukupuhenta ja kehyslauseisiin asetetut minimiparit. Samassa luvussa käsitelin erilaisia provokaatio- eli kyselymenetelmiä, jotka kytkeytyvät puheen kuuntelun ja kyselyn menetelmäperinteeseen, jossa puhetta on muistiinpanolitteroitu eli merkitty muistiin suoraan puhetilanteessa. Myös tallennetuissa haastatteluissa *vapaaksi* asemoitu puhe kontrastoituu kyselymenetelmään: haastattelussa tuotetun puheen voi nähdä *vapaana*, sillä haastattelija tyypillisesti pyrkii ohjaamaan puheen sellaisiin puheenaiheisiin, joiden ansioista puhujat unohtavat puhettaan tarkkailtavan ja antavat puheen ”virrata vuolaasti” ja vapaasti (ks. esimerkit 9, 10, 18 ja 19).

1970–1980-luvuilla puhetilanteen käsitettä aletaan teoretisoida ja kiinnittää huomiota puheen tilanteeseen vaihteluun. Tämä laukaisee tulkintani mukaan diskurssien kilpailun ja jossain määrin myös muutoksen, jolloin haastattelussa tuotettu puhe ei enää näyttäydykään yksiselitteisesti *vapaana puheena*, vaan osin ”epäaitona” vuorovaikutustilanteena, osin institutionaalisen vuorovaikutustilanteena verrattuna erityisesti *arkikeskusteluun*. Näitä havaintoja käsitelin luvussa 3.2. Kiinnostavaa on, että erityisesti haastatteluihin kielentutkimuksen menetelmänä kohdistuva kritiikki johtaa siihen, että kyselyn ja kuuntelun menetelmällä muistiinpanolitteroitu puhe saattaa alkaa näyttäytyä ”vapaampana” kuin haastattelupuhe (esimerkki 5, ks. myös Aarikka 2023: 97), sillä tutkija on saattanut viettää puhujien kanssa pitkäjäksoisesti

aikaa erilaisissa tilanteissa ja siten tehdä huomioita puheesta myös sellaisissa puhe-tilanteissa, joita ei kutsuta haastatteluiksi (ks. esimerkki 18). Nämä kaksi havaintoa ovat keskeisimmät vastaukset tutkimuskysymykseeni: Miten *vapaan puheen* käsite on fennistisen murteentutkimuksen ja puhutun kielen tutkimuksen historian aikana muuttunut?

Analyysini tulokset, joiden mukaan puheen *vapaus* määrittyy toisaalta suhteessa puheen tuottotapaan (provosoitu/provosoimaton) ja kielentutkimuksen menetelmiin, toisaalta suhteessa puhetilanteeseen ja sen laukaisemaan puheen rekisteriin, vastaavat kielentutkijoille tehtyä kyselyä vuodelta 1995, josta Johanna Laakso (1996: 223) raportoi artikkelissaan:

Spontaaniutta puheessa on yritetty määritellä monin tavoin. Valaiseva tulos, jonka foneetikko Susan Fagyal sai keväällä 1995 LINGUIST-sähköpostilistalla esittämäänsä kyselyyn "spontaaniin puheen" käsitteen määrittelystä ja merkityksestä (selonteko kyselyn tuloksista: LINGUIST 6-607, 25.4.1995¹¹). Vastanneista 15 kielentutkijasta suurin osa piti spontaaniutta tärkeänä ja omassa työssään keskeisenä käsitteenä, mutta sen määritelmät perustuivat useisiin näkökulmiin: tuottotapaan (spontaania on reaaliajassa tuotettu tai ennalta suunnitteleminen, ei paperista luettu tai etukäteen harjoiteltu) tai sitten tyyliin tai rekisteriin (spontaania on epämuodollisissa puhetilanteissa käytetty, "casual speech"!). Useimmat vastaajat määrittelivät spontaaniksi yksinkertaisesti kaiken "luonnossa", siis koetilanteen ulkopuolella esiintyvän puheen.

Lopuksi pohdin tutkimuskysymyksistä vaikeinta ja moniulotteisinta: Miten käsitykset siitä, mikä on *vapaata puhetta*, vaikuttavat tutkimusten kysymyksenasetteluihin, tutkimuskohteiden määrittelyyn ja siten tutkimustuloksiin?

Kuten analyysiluvuissa 3.2 ja 3.3 osoitin, kiistattomalta näyttää, että puhutun kielen tutkimuksen keskeisin tutkimuksen kohde on mahdollisimman *vapaa* ja myös epäformaali puhe. Tutkimuskohde tuntuu määräytyvän erityisesti siitä teoreettisesta olettamasta käsin, että mahdollisimman *vapaa puhe* on myös mahdollisimman murteellista ja puhekielistä, verrattuna puhekielen varieteetin eli vernakulaarin vastinpariin **standardiin** (myös *yleiskieli*, *yleispuhekieli*).

Nähdäkseni Labovin (1972) näkökulma ei ole ainoa oikea: mahdollisimman epäformaalissa puhetilanteessa tuotettu puhe ja siten mahdollisimman epäformaali puhetapa ei ole kielentutkimuksen kannalta mitenkään objektiivisesti kaikkein kiinnostavinta, vaikka aineistoni perusteella vaikuttaa, että aineistoni puhutun kielen tutkijat haluavat kohdentaa tutkimuksensa nimenomaan *vapaaseen puheeseen* (miten se ikinä määritelläänkään). On tosin todettava, että Nykysuomalaisen puhekielen murros-hankkeen Turun osahankkeessa tehtiin erilaisia tilannetallennuksia sekä formaaleista että epäformaaleista puhetilanteista (Suojanen 1985: 17–23) – hankkeen esitutkimusraportissa (Pauononen ym. 1976: 22–27) mainitaan Turun osatutkimuksen osalta

11 En ole löytänyt alkuperäislähdettä.

jopa, että itse haastatteluun puhetilanteena on yritetty rakentaa sekä formaalia että epäformaalia osuutta. On aiheellinen kysymys, mitä jää tutkimuksen katveeseen, jos alueellisesta variaatiosta kiinnostuneet puhutun kielen tutkijat keskittyvät pelkästään mahdollisimman *vapaan* puheen ja puhetilanteiden tutkimukseen. Asia erikseen on myös se, ovatko formaalit puhetilanteet ja formaalit rekisterit todella *ei-vapaita*. Miksi puheen tuottaminen muodollisuudeltaan eriasteisissa puhetilanteissa olisi vähemmän *vapaata* kuin toisissa, jos hyväksymme sen tosiseikan, että kaikissa puhetilanteissa ja siten kaikessa puheentuottamisessa on läsnä erilaisia muuttujia, positioita ja rajoitteita, jotka vaikuttavat puhetilanteen sekä puheen muotoutumiseen? Leslie Milroy (1997: 59) kirjoittaa:

[W]e shall not fall into the trap of attempting to record the vernacular of a given speaker, defining this as his most natural and unconstrained linguistic code. For it is clear that *any* speech varies considerably in response to situational context, not merely in response to self-monitoring on a single dimension of casual to formal.

Vaikka Milroyn artikuloima ansa on siis tutkijoiden hyvin tiedostama (esimerkki 5; ks. myös Hakulinen 2009; Mustanoja 2011: 85), aineistoanalyysini osoittaa mielestäni, ettei ansaa aina pystytä välttämään. Toisin sanoen *vapaan puheen* tavoittelu tuntuu toisinaan johtavan tilanteeseen, jossa kielentutkimuksen diskursseissa vahvistetaan ja uusinnetaan ajatusta, että tietyssä puhetilanteessa tuotettu tietynlainen puhe on aidompaa, vapaampaa ja luonnollisempaa kuin toisissa.

Viimeisessä analyysiluvussa käsittelemäni *arjen*, *arkisuuden* ja *arkipuheen* käsitteet kielentutkimuksen kontekstissa tarjoaisivat jatkotutkimuksen aihetta. Käsitystä siitä, millaista ihmisten arki on, millaista kieltä arjessa käytetään, mistä arjessa puhutaan ja keiden kanssa, tuotetaan myös kielentutkimuksen keinoin – ikään kuin huomaamatta. Kyseiseen teemaan olisi hedelmällistä pureutua tätä artikkelia syvällisemmin esimerkiksi semanttisen analyysin kautta (vrt. Teeri-Niknamoghdam 2023).

Aineslähteet

- VK 13 = Leskinen, H. 1963. *Luoteis-Laatokan murteiden äännehistoria. I. Konsonantit*. Suomalaisen Kirjallisuuden Seuran toimituksia 275. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VK 18 = Räisänen, A. 1972. *Kainuun murteiden äännehistoria. I. Vokaalisto*. Suomalaisen Kirjallisuuden Seuran toimituksia 307. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VK 19 = Rintala, P. 1972. *Suomen läntä-loppuiset adjektiivit*. Suomalaisen Kirjallisuuden Seuran toimituksia 306. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VK 20 = Paunonen, H. 1974. *Monikon genetiivin muodostus suomen kielessä. I. Johdanto. Yksivartaloisten kaksitavujen monikon genetiivi suomen murteissa*. Suomalaisen Kirjallisuuden Seuran toimituksia 317. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VK 23 = Nuolijärvi, P. 1986. *Kolmannen sukupolven kieli. Helsinkiin muuttaneiden suurten ikäluokkien eteläpohjalaisten ja pohjoissavolaisten kielellinen sopeutuminen*. Suomalaisen Kirjallisuuden Seuran toimituksia 436. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VK 25 = Nahkola, K. 1987. *Yleisgeminaatio. Äännevuokseksen synty ja vaiheet kielisysteemissä erityisesti Tampereen seudun hämäläismurteiden kannalta*. Suomalaisen Kirjallisuuden Seuran toimituksia 457. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VK 26 = Palander, M. 1987. *Suomen itämurteiden erikoisgeminaatio*. Suomalaisen Kirjallisuuden Seuran toimituksia 455. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VK 29 = Juusela, K. 1989. *Törmäysmurteen variaatiosta. Jälkitavun i-loppuisen diftongin edustus Töysän murteessa*. Suomalaisen Kirjallisuuden Seuran toimituksia 496. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VK 32 = Laitinen, L. 1992. *Välttämättömyys ja persoona. Suomen murteiden nesessiivisten rakenteiden semantiikkaa ja kielioppia*. Suomalaisen Kirjallisuuden Seuran toimituksia 569. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VK 41 = Kurki, T. 2005. *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen. Kielenmuutosten seuraamisesta ja niiden tarkastelussa käytettävistä menetelmistä*. Suomalaisen Kirjallisuuden Seuran toimituksia 1036. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VK 46 = Mustanoja, L. 2011. *Idiolekti ja sen muuttuminen. Reaaliaikatutkimus Tampereen puhekielestä*. Tampere: Tampere University Press. <https://urn.fi/urn:isbn:978-951-44-8417-9>
- VK 47 = Nupponen, A-M. 2011. *"Savon murre" savolaiskorvin. Kansa murteen havainnoijana*. Joensuu: University of Eastern Finland. <http://urn.fi/URN:ISBN:978-952-61-0354-9>
- VK 48 = Väänänen, M. 2016. *Subjektin ilmaiseminen yksikön ensimmäisessä persoonassa. Tutkimus suomen vanhoista murteista*. Turku: Turun yliopisto. <https://urn.fi/URN:ISBN:978-951-29-6664-6>
- VK 49 = Priiki, K. 2017. *Hän, se, tää vai toi? Vuorovaikutussosiolingvistinen tutkimus henkilöviittauksista Kaakkois-Satakunnan nykypuhekielessä*. Turku: Turun yliopisto. <https://urn.fi/URN:ISBN:978-951-29-6719-3>

Kirjallisuus

- Aarikka, L. 2023. *Murre ja sen tutkimus: Näkökulmia fennistisen murteen tutkimuksen historiaan ja kieli-ideologioihin 1871–2017*. Turku: Turun yliopisto.
<https://urn.fi/URN:ISBN:978-951-29-9301-7>
- Aminoff, T. 1871. *Etelä-Pohjanmaan kielimurteesta tutkimus*. Helsinki: J. C. Frenckellin ja Pojan kirjapaino.
- Genetz, A. 1877. *Lautphysiologische Einführung in das Studium der vestfinnischen Sprachen mit besonderer Berücksichtigung des Karelischen*. Helsinki.
- Hakulinen, A. 1995. Johdanto. Teoksessa L. Tainio (toim.) *Keskusteluanalyysin perusteet*. Tampere: Vastapaino, 18–31.
- Hakulinen, A. 2009. Conversation types. Teoksessa S. D'hondt, J.-O. Östman & J. Verschueren (toim.) *The pragmatics of interaction*. Amsterdam: John Benjamins, 55–65.
- Hamunen, M. & T. Huumo 2020. Kielitaju kielitutkijan työkaluna. Teoksessa M. Luodonpää-Manni, M. Hamunen, R. Konstenius, M. Miestamo, U. Nikanne & K. Sinnemäki (toim.) *Kielitutkimuksen menetelmiä. II*. Helsinki: Suomalaisen Kirjallisuuden Seura, 199–239. <https://doi.org/10.21435/skst.1457>
- Juusela, K. 1989. *Törmäysmurteen variaatiosta. Jälkitavun i-loppuisen diftongin edustus Töysän murteessa*. Suomalaisen Kirjallisuuden Seuran toimituksia 496. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kettunen, L. 1910. Väärä suhteemme kirjakieleen. Syytä kirjakielessä itsessäänkin. *Virittäjä* 14, 57–63.
- Kielitoimiston ohjepankki 2024. Liitteet -kin ja -kaan/-kään: yölläkin, ei ollutkaan. <https://kielitoimistonohjepankki.fi/ohje/liitteet-kin-ja-kaan-kaan-yollakin-ei-ollutkaan/#merkitys> [luettu 29.2.2024]
- Kuiri, K. 1984. *Referointi Kainuun ja Pohjois-Karjalan murteissa*. Suomalaisen Kirjallisuuden Seuran toimituksia 405. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kurki, T. 2005. *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen. Kielenmuutosten seuraamisesta ja niiden tarkastelussa käytettävistä menetelmistä*. Suomalaisen Kirjallisuuden Seuran toimituksia 1036. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kurki, T. & L. Mustanoja 2020. Haastattelu aineistonkeruun metodina sosiolinguistisessä. Teoksessa M. Luodonpää-Manni, M. Hamunen, R. Konstenius, M. Miestamo, U. Nikanne & K. Sinnemäki (toim.) *Kielitutkimuksen menetelmiä. II*. Helsinki: Suomalaisen Kirjallisuuden Seura, 274–312. <https://doi.org/10.21435/skst.1457>
- Laakso, J. 1996. Haisee provosoidulta. Aineiston arvioinnin ongelmia Dmitri Tsvetkovin vatjan sanaston valossa. *Virittäjä*, 100 (2), 219–238.
<https://journal.fi/virittaja/article/view/38866>
- Labov, W. 1972. *Sociolinguistic patterns*. Oxford: Blackwell.
- Lappalainen, H. 2004. *Variaatio ja sen funktiot. Erään sosiaalisen verkoston jäsenten kielellisen variaation ja vuorovaikutuksen tarkastelua*. Suomalaisen Kirjallisuuden Seuran toimituksia 964. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Lindén, E. 1942. *Kaakkois-Hämeen murteiden äännehistoria. I. Konsonantisto*. Suomalaisen Kirjallisuuden Seuran toimituksia 219. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Milroy, L. 1987. *Observing & analysing natural language*. New York: Basil Blackwell.
- MO keruuopas = Muoto-opin keruuopas 1983 [1969] lisävihkoineen (1. 1975, 2. 1978).
http://www.kielipankki.fi/wp-content/uploads/ma_keruuopas.pdf
- Mustanoja, L. 2011. *Idiolekti ja sen muuttuminen. Reaaliaikatutkimus Tampereen puhekielestä*. Tampere: Tampere University Press. <https://urn.fi/urn:isbn:978-951-44-8417-9>

- Nordlund, T. 2005. Suomen kielen kvasirakenne ja kieliopillistunut moniäänisyys. Teoksessa M. Haakana & J. Kalliokoski (toim.) *Referointi ja moniäänisyys*. Helsinki: Suomalaisen Kirjallisuuden Seura, 338–359.
- Nuolijärvi, P. 1986. *Kolmannen sukupolven kieli. Helsinkiin muuttaneiden suurten ikäluokkien eteläpohjalaisten ja pohjoissavolaisten kielellinen sopeutuminen*. Suomalaisen Kirjallisuuden Seuran toimituksia 436. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Nuolijärvi, P. 1988. Kielenulkoisten taustamuuttujien huomioon ottaminen 1800-luvun ja 1900-luvun alun murretutkimuksissa. Teoksessa J. Kalliokoski, P. Leino & P. Pyhtilä (toim.) *Kieli 3*. Helsinki: Helsingin yliopisto, 117–160.
- Paunonen, H. 1974. *Monikon genetiivin muodostus suomen kielessä. I. Johdanto. Yksivartaloisten kaksitavujen monikon genetiivi suomen murteissa*. Suomalaisen Kirjallisuuden Seuran toimituksia 317. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Paunonen, H., A. Mielikäinen & M. K. Suojanen (toim.) 1976. *Nyky-suomalaisen puhekielen murroksen tutkimus. Esitutkimusraportti*. Helsinki: Valtion humanistinen toimikunta.
- Pietikäinen, S. 2000. Kriittinen diskurssintutkimus. Teoksessa K. Sarjavaara & A. Piirainen-Marsh (toim.) *Kieli, diskurssi & yhteisö. Soveltavan kielentutkimuksen teoriaa ja käytäntöä 2*. Jyväskylä: Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto, 191–218.
- Pietikäinen, S. & A. Mäntynen 2009. *Kurssi kohti diskurssia*. Tampere: Vastapaino.
- Pietikäinen, S. & S. Määttä 2014. Ideology. Teoksessa J.-O. Östman, J. Verschueren, J. Blommaert & C. Butcaen (toim.) *Handbook of pragmatics online*. Amsterdam: John Benjamins.
- Priiki, K. 2017. Hän, se, tää vai toi? Vuorovaikutussosiolingvistinen tutkimus henkilöviittauksista Kaakkoi-Satakunnan nykypuhekielessä. Turku: Turun yliopisto.
<https://urn.fi/URN:ISBN:978-951-29-6719-3>
- Rapola, M. 1919. *Kantasuomalaiset pääpainottomain tavujen i-loppuiset diftongit suomen murteissa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Reisigl, M. 2018. The discourse-historical approach. Teoksessa J. Flowerdew & J. E. Richardson (toim.) *The Routledge handbook of critical discourse studies*. New York: Routledge, 44–59.
- Reisigl, M. & R. Wodak 2016. The discourse-historical approach. Teoksessa R. Wodak & M. Meyer (toim.) *Methods of critical discourse studies*. 3. painos. Los Angeles: Sage, 87–121.
- Ruoppila, V. 1943. *Kotieläinten nimitykset Suomen murteissa. I. Hevonen, nauta, lammas, vuohi*. Suomalaisen Kirjallisuuden Seuran toimituksia 222. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Sandig, B. 1975 (1972). Zur Differenzierung gebrauchssprachlicher Textsorten im Deutschen. Teoksessa E. Gülich & W. Raible (toim.) *Textsorten. Differenzierungskriterien aus linguistischer Sicht*. 2. painos. Wiesbaden: Athenaion, 113–124.
- Seppänen, E.-L. 1995. Vuorovaikutus paperilla. Teoksessa L. Tainio (toim.) *Keskusteluanalyysin perusteet*. Tampere: Vastapaino, 18–31.
- Strandberg, J. 2004. *Ei sanat salahan joua. Fennistiikan murteenkeruun historiaa 1868–1925*. Pro gradu -tutkielma. Helsingin yliopisto.
- Suojanen, M. K. 1985. *Mitä Turussa puhutaan? Raportti Turun puhekielen tutkimuksesta*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 23. Turku: Turun yliopisto.
- Suomalainen, K. 2020. *Kuka sinä on? Tutkimus yksikön 2. persoonan käytöstä ja käytön variaatiosta suomenkielisissä arkikeskusteluissa*. Turku: Turun yliopisto.
<https://urn.fi/URN:ISBN:978-951-29-8238-7>
- Teeri-Niknammoghadam, K. 2023. Termi *kieliopillinen* Virittäjän artikkelissa. *Virittäjä*, 127 (2), 270–281. <https://doi.org/10.23982/vir.121424>
- Tieteen termipankki 2023. <https://tieteentermipankki.fi>

- Vanhatalo, U. & K. Vehkalahti 2020. Kysely. Teoksessa M. Luodonpää-Manni, M. Hamunen, R. Konstenius, M. Miestamo, U. Nikanne ja K. Sinnemäki (toim.) *Kielentutkimuksen menetelmiä. II*. Helsinki: Suomalaisen Kirjallisuuden Seura, 240–273.
<https://doi.org/10.21435/skst.1457>
- Väänänen, M. 2016. *Subjektin ilmaiseminen yksikön ensimmäisessä persoonassa. Tutkimus suomen vanhoista murteista*. Turku: Turun yliopisto.
<https://urn.fi/URN:ISBN:978-951-29-6664-6>
- Ylitalo, R. 2009. *The realisation of prominence in three varieties of standard spoken Finnish*. Oulu: Oulun yliopisto. <https://urn.fi/URN:ISBN:9789514291142>

Osa II

Toivola, M., P. Lintunen & L. M. Heikkola (toim.) 2024. Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. *New directions in speech research – Freely produced speech as data.* AFinLA-teema / n:o 17, 49–76.

Riikka Ullakonoja

Jyväskylän yliopisto

Venäläisen aksentin tunnistaminen suomen suullisen kielitaidon arvioinnissa

Nostot

- Venäläinen aksentti on usein hyvin tunnistettava, mutta se vaihtelee puhujittain.
- Venäläinen aksentti tunnistetaan useimmiten ääntämisen perusteella.
- Ääntämisessä venäläinen aksentti tunnistetaan yksittäisistä äänneistä, liudennuksesta ja intonaatiosta.



VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Abstract

The paper focuses on the Russian accent in the Finnish oral proficiency test. The data consist of spontaneous monologues on an intermediate proficiency level from 10 speakers (about 90 s each). Raters ($n=44$) were asked to identify the accent in the speech samples as well as to give justifications for their recognition in their own words. The analysis is based on a classification of the raters' comments as well as auditory and acoustic analysis by the author. The results show that the raters paid most attention to individual sounds, but also palatalization and intonation were mentioned as the reasons for detection of the Russian accent. To conclude, further and more detailed acoustic analysis is needed to gain more understanding of the phonetic features of Russian accented Finnish.

Asiasanat: arviointi, suomen kieli, ääntäminen

Keywords: assessment, Finnish, pronunciation

1 Vieras aksentti

Vieraalla aksentilla tarkoitetaan sitä, että kuulija tunnistaa puheesta piirteitä, jotka poikkeavat ko. kielen ensikielisestä ääntämyksestä ja joiden perusteella tulee vaikutelma, että puhujan ensikieli on jokin muu (esim. Flege 1988; Saito ym. 2016). Tässä tutkimuksessa viitataan vieraalla aksentilla kuulijoiden havaintoon venäläisestä korostuksesta suomessa ja akustisiin mittauksiin venäjän kielelle ominaisista piirteistä suomenkielisissä puhenäytteissä. Vieras aksentti on usein hyvin helposti kuultavissa. Se pystytään tunnistamaan hyvinkin lyhyistä, jopa yksittäisen äänteen mittaisista puhenäytteistä (Flege 1984). Tunnistaminen on mahdollista jopa silloin, kun puhenäytteet soitetaan takaperin tai niistä on poistettu paljonkin prosodisen tason informaatiota tai äännetason informaatio kokonaan (Munro 1995; Munro ym. 2010).

Tutkijat ovat melko yksimielisiä siitä, että vieraan kielen suullisen kielitaidon oppimisen tavoitteena tulisi olla on puheen ymmärrettävyys ensikielisen kielenpuhujan kaltaisen ääntämisen sijaan (Munro & Derwing 1995; Tergujeff 2022). Myöskään eurooppalaisessa viitekehyksessä (EVK) puhujalta ei ylimmälläkään (C2) tasolla edellytetä ensikielisen kaltaista ääntämystä (Council of Europe 2020). Monelle maahanmuuttajalle vieraasta aksentista voi kuitenkin olla haittaa. Voimakas vieras aksentti voi aiheuttaa kielteisiä asenteita tai kielteistä kohtelua esim. työelämässä niin, että vieraalla aksentilla puhuva työntekijä koetaan epämiellyttävänä ja jopa ammattitaidottomana. Se voi siten olla este myös työllistymiselle ja kansalaisuuden saamiselle. (Kokkonen 2007; Grosjean 2010: 77–84; Teló ym. 2022; Tsunemoto ym. 2022; Halonen ym. 2024.) Toki on huomattava, että harva pystyy välttämään vierasta aksenttia puheessaan, jos on aloittanut kielen opiskelun murrosiän jälkeen (ks. esim. Mennen & de Leeuw 2014).

Suurin osa aiemmista tutkimuksista on keskittynyt englannin kieleen opittavana kielenä (esim. Munro & Derwing 1995; Moyer 2013). Englantia käytetään englanninkielisissä maissa ja maailmalla myös yleisemmin erittäin paljon lingua francana. Tämä

on todennäköisesti johtanut siihen, että kuulijat, olivat he sitten kadunmiehiä tai kielitaidon arvioinnin ammattilaisia, ovat ehkä erittäin tottuneita vieraalla aksentilla puhuttuun englantiin. Ehkä siksi tutkimus vierasta aksenttia puhuvan tuottamasta vaikutelmasta ja asenteista puhujaa kohtaan on jäänyt taka-alalle. Suomessa tilanne on kuitenkin toinen. Suomenkieliset ovat herkkiä havaitsemaan ei-ensikieliset suomen piirteet, arkailevat käyttää suomea lingua francana ja vaihtavat helposti englantiin kuullessaan vieraalla aksentilla puhuttua suomea (esim. Toivola & Ullakonoja 2017; Ruuska 2020; Kim ym. 2022). Lisäksi suomalaiset kokevat vieraalla aksentilla puhutun suomen häiritseväksi ja koomisenakin (esim. Dufva & Halonen 2016). Aiempi tutkimus on osoittanut, että jopa koulutetut kielitaidon arvioijat Yleisissä kielitutkinnoissa kiinnostuvat vieraaseen aksenttiin huomiota niin, että puhujan oletettu ensikieli vaikuttaa siihen, miten puhujan suomen suullista kielitaitoa arvioidaan. Ensikielen tunnistaminen näyttäisi vaikuttavan erityisesti ääntämisen ja sujuvuuden arviointiin: jos puhujan ensikieli tunnistettiin, hänen ääntämistään arvioitiin ankarammin, jos taas ei tunnistettu, sujuvuus arvioitiin heikommaksi. Lisäksi ensikielten välillä oli eroja siinä, miten ankarasti puhujien kielitaitoa arvioitiin. Esimerkiksi venäjää ensikielenään puhuvilla tunnistaminen vaikutti arviointiin niin, että sujuvuuden arvio laski merkitsevästi kun venäjänkielisen ensikieli tunnistettiin. (Ahola & Halonen 2021; Ahola 2022; Hirvelä ym. 2022; Halonen ym. 2024.) Koska kielitaidon arvioinnin pitäisi kuitenkin olla objektiivista ja pohjautua samoihin kriteereihin puhujan ensikielestä riippumatta, on tärkeää tutkia tekijöitä, jotka vaikuttavat arviointiin varsinaisten arviointikriteerien ulkopuolella. Yksi tällainen tekijä on ääntämisestä syntyvä vieraan aksentin vaikutelma.

Yleisissä kielitutkinnoissa arvioidaan aikuisten toiminnallista kielitaitoa samantapaisissa tilanteissa, mihin kielenoppija viestintätilanteissaan tyypillisesti joutuu. Yleisissä kielitutkinnoissa ei mitata tietyn oppimäärän edellyttämien taitojen hallintaa, vaan tutkinnossa osoitettavat taidot on voinut saada eri tavoin. Puhumisen osakokeen arviointi on kriteerivitteistä, eli se pohjautuu ennalta määriteltyihin kriteereihin, eikä osallistujien suoritusta verrata toisiin osallistujiin. (Yleisten kielitutkintojen perusteet 2011.) Kielitaidon arvioinnin alan aiempi tutkimus arviointikriteerien ulkopuolisista tekijöistä on keskittynyt lähinnä arvioijan taustoihin (esimerkiksi arviointikokemuksen määrään, tottumukseen kuunnella kyseistä aksenttia) (Isaacs & Thomson 2013; Ahola 2016; Kahng 2023), mutta siitä, tunnistako arvioija kyseisen aksentin tai miten hänen havaintonsa puhujan ääntämisestä vaikuttavat suullisen kielitaidon arviointiin, on toistaiseksi vähän tutkimustietoa.

Tässä artikkelissa raportoidussa tutkimuksessa keskitytään venäläisellä aksentilla puhuttuun suomeen, koska venäjänkieliset ovat merkittävä vähemmistökieliryhmä Suomessa ja suurin kieliryhmä Yleisten kielitutkintojen suomen kielen tutkinnossa. On tärkeää selvittää, mitkä piirteet heidän puhumassaan suomessa tekevät siitä koulutettujen kielitaidon arvioijien mielestä venäjänkieliseksi tunnistettavan. Aiemmat tutkimukset venäjää ensikielenään puhuvien suomen kielen ääntämisestä ovat keskittyneet prosodiaan (Aho & Toivola 2008) tai raportoineet ääntämispiirteistä muuta-

man arvioijan kuulohavainnon perusteella (Aho ym. 2016). Näistä poiketen keskityn tutkimuksessani sekä äänne- että prosodisen tason ilmiöihin ja tarkastelen niitä myös akustisen analyysin valossa. Lisäksi käsittelen puhenäytteistä annettuja sanallisia arvioita suurehkolta joukolta (n=44) arvioijia, jotka osallistuivat *Rikkinäistä suomea: Aksenttien arviointi yhteiskunnallisena portinvartijana* -hankkeeseen. Hankkeessa keskitymme vieraan aksentin rooliin suomen kielen puhumistaidon arvioinnissa, esimerkiksi aksentin herättämiin stereotypioihin (ks. Halonen, Huhta ym. 2020). Olemme myös tutkineet sitä, miten venäjänkielisten osallistujien puhumisen taito on muuttunut vuosien saatossa ja mitkä taustamuuttajat puhumisen taitoon vaikuttavat (Ahola & Hirvelä 2024). Yksi hankkeen yleisistä tutkimuskysymyksistä on, miten hyvin ja millä perusteella arvioijat tunnistavat osallistujien ensikielen, mihin tunnistaminen perustuu ja onko tunnistamisella merkitystä arvioille kielitaidosta. Tässä artikkelissa keskityn siihen, mitkä arvioijien kuvaamat ja foneettisen (kuulonvaraisen ja akustisen) analyysin perusteella havaitut ääntämispiirteet vaikuttavat siihen, että puhuja tunnistetaan ensikieleltään juuri venäjänkieliseksi ja kuinka helppoa tunnistaminen on.

2 Venäläinen aksentti suomen kielessä

2.1 Aiempi tutkimus venäläisestä aksentista suomen kielessä

Venäläinen aksentti suomessa on usein hyvin tunnistettava. Toivolan ja Ullakononjan (2017) tutkimuksessa jopa venäjänkieliset, jotka eivät osanneet lainkaan suomea, pystyivät erottamaan lyhyet venäläisellä aksentilla puhutut suomen kielen näytteet ensikielisten suomenpuhujien näytteistä melko hyvin. Kyseisessä tutkimuksessa näytteet tosin olivat alkeisoppijoilta, joiden venäläisen aksentin voidaan olettaa olleen vahvempi kuin tässä tutkimuksessa, jossa puhujat ovat keskitason puhujia. Hyvin edistyneitä eritaustaisia suomen puhujia käsittelevässä Kurosen ja Kautosen (2018) tutkimuksessa venäjänkielinen puhuja arvioitiin vahvimmin vieraalla aksentilla puhuvaksi. Ei kuitenkaan voida sanoa, että venäjänkielisillä olisi erityisen vahva aksentti. Leinosen (2015) väitöskirjassa suomalaisnuoret arvioivat opettajan samalle kielitaidon tasolle sijoittamista puhenäytteistä venäjänkieliset aksentiltaan keskivahvoiksi. Kuitenkin Ahon ja Toivolan (2008) tutkimuksessa venäjänkieliset maahanmuuttajat itse pitivät muita ensikieliryhmiä tärkeämpänä sitä, että he oppisivat ääntämään suomea "samalla tavalla kuin suomalaiset".

Kattavin tutkimus venäläisestä aksentista suomen kielessä on Toivolan (2011) väitöskirja, jossa noin sata arvioijaa arvioi vieraan aksentin kymmenen venäjänkielisen puhumasta suomesta. Puhujat olivat tasoltaan keski- tai edistyneellä tasolla (mutta tasoa ei ollut määritelty esim. EVK-asteikolla). Äänteellisiä poikkeamia tarkasteltiin kahden tutkijan kuulonvaraisen arvion perusteella, mutta tuloksia ei esitetty niin

tarkasti, että niistä kävisi ilmi, missä äänneissä venäjänkielisillä puhujilla oli haasteita. Tuloksena oli muun muassa, että äänneiden kestot tuottavat vaikeuksia (esim. liian lyhyet tai pitkät kestot sanan sisällä tai liian lyhyt kesto sanan lopussa). Aiempien tutkimusten perusteella venäjää ensikielenään puhuvien suomen oppijoiden puheesta tiedetään myös, että vokaalien kvantiteettioppositio (eli pitkän ja lyhyen vokaalin ero) ja suomen painottomat vokaalit (esim. /i/:n ja /e/:n eron tuottaminen ja pitkät painottomat vokaalit) ovat vaikeita oppia. Yksittäisistä vokaaleista haasteita tuottavat /æ/, /ø/ ja /y/ ja diftongeista /ie/ ja /yø/, ja konsonanteista haasteellisia ovat puolestaan /h/ ja /l/ ja konsonanttien kestot. Venäjän kielestä suomeen siirtyvät helposti myös konsonanttien palatalisoituminen (liudennus), prosodinen jaksottelu ja intonaatiomallit. Lisäksi suomen kielelle tyypillisen lausuman loppuisen narinan puuttumista on havaittu puhujilla, joiden venäläinen aksentti on vahva. (Ylinen 2006; Aho & Toivola 2008; de Silva & Ullakonoja 2009; Toivola 2011; Ullakonoja & Kuronen 2015; Aho ym. 2016; Aho ym. 2019.) Nämä haasteet johtunevat ainakin osittain venäjän ja suomen erilaisista fonologisista järjestelmistä, joita tarkastelen seuraavaksi.

2.2 Suomen ja venäjän fonologian kohtaaminen

Suomen ja venäjän fonologiset järjestelmät eroavat toisistaan paljon, ja kaikkia eroja on mahdoton kuvata tässä. Kaikki järjestelmien väliset erot eivät myöskään todennäköisesti aiheuta oppijoille ongelmia tai tuota vieraan aksentin vaikutelmaa kuulijalle. Tässä aluvuossa kuvataan kuitenkin keskeiset järjestelmien väliset erot tutkimuskirjallisuuden pohjalta. Suurin osa venäjän konsonanteista on erilaisia kuin suomessa, ja kaksoiskonnantit ovat harvinaisia. Venäjän konsonanteille erityinen piirre on palatalisoituminen eli liudentuminen, joka koskee lähes kaikkia konsonantteja: etuvokaalin edessä konsonantti liudentuu. Se tarkoittaa sitä, että konsonantti saa j-mäisen sävyn tässä äänneympäristössä. Venäjänkieliselle suomen oppijalle tämä saattaa tuottaa haasteita, koska suomen fonotaksissa vastaavaa sääntöä ei ole, vaikkakin liudennus ilmiönä on mahdollinen esimerkiksi joissain murteissa. Erityisen kuultava tämä haaste voi olla /l/-äänneen kohdalla, koska venäjän liudentumaton /l/-äänne on voimakkaasti velaaristunut. Konsonanttien liudentuminen vaikuttaa venäjän kielessä myös vokaalien ääntämiseen siten, että painolliset vokaalit (/i/:n variantteja lukuun ottamatta) saavat i-mäisen alkusävyn esiintyessään liudentuneen konsonantin jälkeen. Kun painollinen vokaali (/i/:n variantteja lukuun ottamatta) on täysin liudentuneessa äänneympäristössä, eli sen edellä ja jäljessä on liudentunut konsonantti, vokaalin alussa ja lopussa on i-mäinen liukuma-aines. (de Silva 1999: 46; Bondarko 2009; Iivonen 2009a.)

Konsonanttien välillä on muitakin eroja. Suomessa konsonanttien soinnillisuudella on venäjään verrattuna hyvin vähän merkitystä. Venäjässä useimmat konsonantit esiintyvät soinnillisuuden suhteen pareina ja soinnilliset konsonantit ovat voimakkaasti

soinnillisia. Suomessa on vain yksi sibilantti /s/, joka mukautuu vapaassa puheessa jonkin verran ääntöympäristön, lähinnä sitä ympäröivien vokaalien mukaan. Tätä äännettä voidaan ajatella vastaavan jopa yhdeksän venäjän äännettä: sibilantit /s/, si/, /z/, /zʲ/, /ʃ/, /ʃʲ/ ja /ʒʲ/ ja affrikaatat /ts/ ja /tʃ/. Useat suomen kielen apikaaliartikulaatiolla ääntyvät konsonantit ääntyvät venäjässä dentaalidorsaaleina, esim. /t/ ja /d/. Suomen /h/:n ääntäminen vaihtelee äänneympäristön ja aseman mukaan niin, että se voidaan luokitella frikatiiviluokan lisäksi myös approksimantteihin (eli äänneisiin, joilla on frikatiiveja heikompi kohina) ja lisäksi se esiintyy usein soinnillisena soinnillisten äänneiden välissä. Venäjässä vastaavaa vaihtelua ei tapahdu ja vastaava äänne on velaarifrikatiivi /x/ tai sen liudentunut variantti /xʲ/. Suomen kielessä on puolestaan yksi ainut konsonantti /ŋ/, jota ei ole venäjässä ja joka saattaa venäjänkielisen suomen oppijan puheessa korvautua helposti konsonanttiyhtymällä /nk/. (Bondarko 1998; de Silva 1999: 46–52, 72–73; Bondarko 2009; livonen 2009a; Lennes 2009; Raimo & Ojala 2009; de Silva & Ullakonoja 2009.)

Venäjässä konsonantit ovat kahta poikkeusta lukuun ottamatta aina lyhyitä, mutta vokaalien kesto vaihtelee, koska se toimii sanapainon parametrinä eli sanan painollinen vokaali on painottomia vokaaleja pidempi. Suomessa taas sekä konsonanttien että vokaalien pituus on merkitystä erottava. Venäjässä vain painollinen vokaali voi esiintyä kestoltaan pitkänä, kun taas suomessa fonologisesti pitkiä vokaaleita on myös painottomassa asemassa. Venäjänkielisillä on siis käsitys siitä, että vokaaleiden kesto on tärkeä, mutta ongelmaksi muodostuu se, että suomessa painottomat vokaalit eivät redusoidu² niin kuin venäjässä. Venäjässä painoton /i/ ja /e/ redusoituvat liudentumattomassa äänneympäristössä samaksi /ɨ/ foneemiksi ja painoton /a/ ja /o/ yleensä joko a-maiseksi /ʌ/ tai svaavokaaliksi /ə/. Suomen etuvokaaleilla /æ/, /y/ ja /ø/ ei ole vastinetta venäjän kielessä, vaan kaikki venäjän pyöreät vokaalit ovat takavokaaleja ja väljät keski- tai takavokaaleja. (de Silva 1999: 46, 63–69; Bondarko 2009; livonen 2009a; de Silva & Ullakonoja 2009; de Silva ym. 2010: 11–37, 130–133.) Suomen kielessä on paljon diftongeja ja vokaaliyhtymiä, kun taas venäjän kielessä ei ole diftongeja ja vokaaliyhtymiäkin esiintyy melko harvoin. Myös diftongit voivat siis tuottaa haasteita venäjänkieliselle suomen oppijalle (de Silva & Ullakonoja 2009).

Venäjän prosodia eroaa suomen prosodiasta muun muassa siten, että intonaatiolla on venäjässä merkityksiä erottava tehtävä (Volskaya 2009) ja esimerkiksi kysymyslause voidaan erottaa väitelauseesta pelkän intonaation avulla. Suomessa kysymykset erotetaan väitelauseesta tyypillisesti leksikaalisin tai kieliopillisin keinoin, vaikka ne eroavatkin jonkin verran myös perustaaajuudeltaan (Hirvonen 1970; livonen 2009b). Aiempaa tutkimusta näiden kielten perustaaajuusjakaumista on vähän. Niiden

1 Moskovalaisessa ääntämyksessä esiintyy lisäksi pitkä palatoalveolaarinen sibilantti [ʒʲ].

2 Redusoitumisella tarkoitetaan sitä, että äänteen laatu muuttuu ja kesto lyhenee painottomassa asemassa.

perusteella on mahdollista, että sekä perustaajuuden yleinen ja yksilökohtainen vaihtelu on suurempaa venäjän kielessä kuin suomen kielessä ja että suomea puhutaan keskimäärin matalammalta kuin venäjää (de Silva ym. 2003; Ullakonoja 2007; Lennes ym. 2009; Lennes & Toivola 2023). Suomen kielelle on tyypillistä narina ilmauksen lopussa, kun taas venäjän kielelle ei – etenkin naispuhujilla tätä ei havaita (Aho & Toivola 2008; Lennes ym. 2009). Prosodian tasolla venäjä ja suomi eroavat toisistaan myös sanapainon suhteen. Suomessa sanapaino on kiinteä ja sen ajatellaan asettuvan aina sanan ensimmäiselle tavulle, jolloin se toimii myös sanarajan merkinä. Venäjässä puolestaan paino on vaihteleva eli se voi asettua mille tahansa sanan tavuista. Lisäksi se voi myös olla liikkuva eli vaihtaa paikkaa saman sanan eri taivutusmuodoissa. (de Silva 1999: 58–69; Bondarko 2009; de Silva ym. 2010: 35–37.)

Osa venäjän kielelle tyypillisistä piirteistä aiheuttaa suomen kielessä käytettynä merkityksen väärintymmärrystä (esimerkiksi, jos puhuja liioittelee painollisen vokaalin kestoa sanassa *tuli*, se saatetaan kuulla sanana *tuuli*). Osa taas tuottaa pelkästään vieraan aksentin vaikutelman sanojen tai lauseen merkitystä muuttamatta. Venäläisestä aksentista tehdyn aiemman tutkimuksen perusteella sekä venäjän ja suomen fonologian kontrastiivisen tarkastelun avulla voidaan olettaa, että monella venäjää ensikielenään puhuvalla Yleisten kielitutkintojen suomen keskitason testiin tulevalla osallistujalla on puheessaan piirteitä, jotka paljastavat hänen ensikielensä. On siis kiinnostavaa tarkastella sitä, mitkä piirteet ovat sellaisia, joihin koulutetut kielitaidon arvioijat kiinnittävät huomiota ja tunnistavat venäjänkieliseksi.

3 Aineisto

Puhujan venäjänkieliseksi tunnistamista lähestyn kahdella eri menetelmällä:

- 1) tarkastelemalla arvioijien näytteille antamia ensikielioletuksia ja niiden sanallisia perusteluita ja luokittelemalla ne, sekä
- 2) kuulonvaraisen ja akustisen analyysin keinoin.

Puheaineistona käytetään *Rikkinäistä suomea: Aksenttien arviointi yhteiskunnallisena portinvartijana* -hankkeessamme kerättyä kielitestiaineistoa. Aineisto on kerätty todellisessa Yleisten kielitutkintojen testilanteessa, jossa osallistajat ovat antaneet luvan käyttää suoritustaan tutkimustarkoitukseen. Tässä tutkimuksessa keskityttiin hankkeen laajemmasta aineistosta vain venäjää ensikielenään puhuviin osallistujiin, joita oli viisi miestä ja viisi naista. Aineiston keruu ja hankkeen teoreettiset lähtökohdat on kuvattu tarkemmin toisaalla (Halonen, Huhta ym. 2020; Halonen ym. 2024), joten seuraavassa kuvataan vain tämän tutkimuksen kannalta olennaisimmat asiat.

3.1 Puhujat

Aineistona käytettiin kymmenen Yleisten kielitutkintojen suomen keskitason testiin osallistunutta venäjää ensikielenään puhuvaa puhujaa, viisi miestä (Rusm1–Rusm5) ja viisi naista (Rusf1–Rusf5). Heidän suorituksensa vastasi Yleisten kielitutkintojen keskitason kriteerijä, eli he olivat saaneet kolmen tehtävän perusteella puhumisen taitotasoarvioiksi 3 tai 4. Taso 3 vastaa EVK-asteikolla tasoa B1 ja taso 4 tasoa B2 tai yli B2³. Suurin osa kuului ikäryhmään 25–44 vuotta, olivat ammattikorkeakoulu- tai yliopistotutkinnon suorittaneita henkilöitä ja opiskelleet suomea 1–3 vuotta. Vertailuaineistoksi nauhoitettiin puhetta kuudelta suomea ensikielenään puhuvalta mieheltä (3) ja naiselta (3), jotka eivät puhuneet vahvasti mitään tiettyä suomen alueellista varianttia. He edustivat eri koulutustasoja ja ikäryhmiä.

3.2 Puhetehtävä

Jokaiselta puhujalta tallennettiin tutkimusta varten yksi monologimuotoinen spontaani kerrontatehtävä, joka käsitteli arkipäivän aihetta. Aineistoa on noin 90 s per puhuja. Venäjää ensikielenään puhuvien aineisto tallennettiin kielistudio-olosuhteissa todellisessa kielitestitilanteessa, kun taas ensikieleltään suomenkielisten aineisto nauhoitettiin studio-olosuhteissa.

3.3 Arvioijat

Arvioijat (n=44) olivat koulutettuja yleisten kielitutkintojen arvioijia, joista suurin osa oli naispuolisia (93 %) ylemmän korkeakoulututkinnon suorittaneita suomi toisena kielenä -opettajia ja ilmoitti ensikielekseen suomen (Ahola 2022: 25). Yksi arvioija mainitsi ensikielekseen venäjän, kolme muun kuin suomen ja seitsemän ei ilmoittanut ensikieltään. Arvioijat olivat eri vaiheissa uraansa. Heidän ikänsä vaihteli 31–69 ikävuoden välillä (keski-ikä 49 vuotta), opetuskokemusta heillä oli 8–43 vuotta (keskiarvo 21 vuotta) ja yleisten kielitutkintojen arviointikokemusta heillä oli 0–22 vuotta (keskimäärin 10 vuotta). Arvioijista 55 % kertoi venäläisen aksentin olevan heille entuudestaan tuttu (mutta läheskään kaikki eivät olleet vastanneet tähän kysymykseen). Arvioijiin viitataan pseudonyymeillä Arv1–Arv44.

3.4 Arviointitehtävä

Vain venäjää ensikielenään puhuvien puhujien puhenäytteet arvioitiin. Näytteet esitettiin arvioijille satunnaisessa järjestyksessä ja he vastasivat sähköisellä lomak-

3 _____
Henkilön pitäisi osallistua ylimmän tason testiin, jotta saataisiin selville, onko hän tasolla C1 tai C2.

keella useaan kysymykseen per näyte. Heitä pyydettiin arvioimaan näytteet yleisten kielitutkintojen arviointimatriisilla asteikolla 1–6, joka vastaa EVK-asteikkoa A1–C2 (Council of Europe 2020). Yleiskriteerin lisäksi heitä pyydettiin arvioimaan numeerisesti kaikki osakriteerit ja kommentoimaan puhujaa/arviota. Lisäksi arvioijilta kysyttiin avokysymyksenä ”Minkä oletat olevan puhujan ensikieli/lähtökieli?” ja heitä pyydettiin perustelevaan vastauksensa. Nämä kaksi kysymystä muodostavat tämän artikkelin arviointiaineiston. Puhenäytteiden arviointialusta ja arviointilomake kokonaisuudessaan on esitelty tarkemmin toisaalla (Halonen, Huhta ym. 2020).

4 Analyysimenetelmät

Ensin arvioijien ensikielioletukset luokiteltiin, ja niistä laskettiin tunnistamisprosentit (ks. tarkempi kuvaus luokittelusta luku 5.1). Tunnistettujen ja tunnistamattomien näytteiden ryhmien välistä eroa näytteelle annetun taitotasoarvion suhteen tutkittiin ristiintaulukoinnilla, Khiin neliötestillä ja sovitetun standardoidun jäännöksen avulla. Khiin neliötestillä selvitetään, onko ryhmien välillä riippuvuutta ja sovitettu standardoitu jäännös kertoo siitä, missä riippuvuutta on. (Metsämuuronen 2005: 330–335; Field 2009: 686–700.) Sitten arvioijien antamat ensikieliluokitteluperusteet analysoitiin laadullisen sisällönanalyysin keinoin luokittelemalla ne aineistosta nouseviin luokkiin. Kehitin menetelmää niin, että käytin luokittelussa painokertoimia sen mukaan, miten montaa eri luokkaa arvioija oli käyttänyt. Jokaisen arvioijan vastaus oli kokonaisuudessaan 100 %, ja riippuen siitä, kuinka monta luokkaa vastaus sisälsi, nämä luokat saivat painokertoimet. Jos arvioija oli vastannut pelkästään *ääntäminen*, sille annettiin kerroin 1 (eli 100 % vastauksesta oli ääntämistä). Jos arvioija puolestaan oli vastannut *puhetapa*, *ääntäminen*, eli antanut kaksi perustelua, *puhetapa* sai kertoimen 0,5 (50 %) ja *ääntäminen* samoin 0,5 (50 %). Niin ikään, jos arvioija oli maininnut kolmeen eri kategoriaan kuuluvaa seikkaa, kukin luokka sai kertoimen 0,33 (30 %). Tällä tavoin vältin sen, että pitemmät ja monipuolisemmat vastaukset saisivat lyhyempiä enemmän painoarvoa aineiston kokonaistarkastelussa. Tämä painokertoimiin perustuva luokittelu vaikuttaa esitettyihin tuloksiin niin, että jokaisella yksittäisellä arvioijalla on aineiston kokonaistarkastelussa sama painoarvo, riippumatta siitä, kuinka pitkä (ja moneen luokkaan jakautuva) vastaus hänellä oli.

Analysoin puhenäytteet kahdella tapaa: kuulonvaraisesti kirjaamalla näytteistä ylös venäjän kielelle tyypillisiä piirteitä ja akustisesti Praat-ohjelmalla (Boersma & Weenink 2023). Sillä ei tehty laajoja akustisia mittauksia, vaan kuulonvaraisia havaintoja tukemaan piirrettiin havainnollistavia kuvioita. Kuvioissa äänneiden⁴ taajuutta suhteessa

4 Käytän yleisen tavan mukaisesti IPA-transkriptiossa foneemeista kenoviivoja ja allofoneista hakasulkeita.

aikaan on havainnollistettu kapeakaistaisilla spektrogrammeilla (aikaikkuna 33 ms), jossa näkyvät selkeästi eri formantit ja energian lyhytaikainen jakautuminen (tummina pystyjuovina). Spektri-kuvioilla on puolestaan kuvattu spektrogrammin poikkileikkaus tietystä äänneestä. Formanttianalyyssissä on kuvattu spektrogrammin päälle neljä alinta formanttia. Praat käyttää näiden kaikkien laskemisessa FFT (Fast Fourier Transform) -algoritmia. Perustaajuusanalyyssissä on käytetty Praatin autokorrelaatiomenetelmää ja sen tuottamat pitch-tiedostot on tarkastettu ja korjattu spektrogrammin pohjalta. (Ks. tarkemmin Boersma 1993; Raimo & Ojala 2009; Vainio ym. 2009; Johnson 2011: 60–67, 77–79; Boersma 2013.)

5 Ensikieleltään venäjänkieliseksi tunnistaminen

5.1 Venäjänkieliseksi tunnistamisen helppous

Ensimmäiseksi tarkastelen sitä, miten hyvin arvioijat (n=44) tunnistivat ensikieleltään venäjänkieliset puhujat (n=10) (taulukko 1), eli mitä he vastasivat kysymykseen ”minkä oletat olevan puhujan ensikieli/lähtökieli?”. Ensimmäiseen luokkaan (taulukossa *venäjä*) eli ns. varmaan tunnistamiseen luokiteltiin vain vastaukset, jossa arvioija oli antanut yksiselitteisen vastauksen *venäjä* (tai kielen sijaan *venäläinen* tai *Venäjä*) ilman epävarmuuden ilmaisua tai jonkun toisen kielen mainintaa⁵. Niin sanotun epävarman tunnistamisen luokka (taulukossa *ehkä venäjä*) tarkoittaa vastauksia, jossa on vastattu pelkästään *venäjä*, mutta ilmaistu jotenkin epävarmuutta. Toinen epävarmuutta ilmaiseva luokka on *venäjä* mainittuna jonkun muun tai joidenkin muiden kielten kanssa (taulukossa *venäjä tai joku muu kieli tai muita kieliä*). Vääristä ensikielioletuksista eriteltiin taulukkoon vain eniten mainintoja saaneet kielet eli viron, arabian ja englannin, vaikka ne olisi mainittu toisten kielten kanssa. Luokka *muu väärä kieli* tarkoittaa muita oletettuja kieliä. Luokkaan *en osaa sanoa* luokiteltiin epävarmuuden ilmausten lisäksi tyhjät vastaukset eli 10 %:ssa tapauksista arvioija ei nimennyt ensikieltä.

5 Mukaan otettiin kolme arviota, joissa venäjää oli tarkennettu esim. vironvenäläinen, venäjä, inkeriläinen tausta ja inkerinsuomalaiset sukujuuret; venäjä.

TAULUKKO 1. Puhujien tunnistaminen ensikieleltään venäjänkielisiksi koko aineistossa.

Oletettu ensikieli	n	%
Venäjä (ns. varma tunnistaminen)	256	58 %
Ehkä venäjä (ns. epävarma tunnistaminen)	23	5 %
Venäjä tai joku muu kieli tai muita kieliä	21	5 %
Arabia yhtenä mainituista kielistä ilman venäjää	11	3 %
Englanti yhtenä mainituista kielistä ilman venäjää	14	3 %
Viro yhtenä mainituista kielistä ilman venäjää	26	6 %
Muu kieli ilman venäjää	45	10 %
En osaa sanoa	44	10 %
Yhteensä	440	100 %

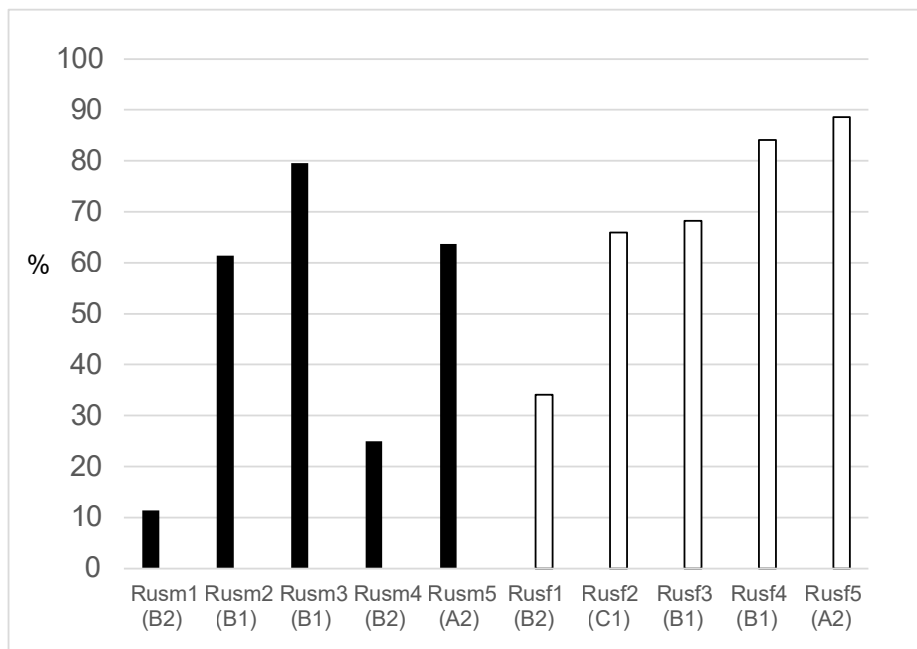
Suurimmassa osassa (noin 2/3) vastauksia venäjä oli mainittu, mutta osassa venäläinen aksentti oli sekoitettu viron, englannin tai arabian puhujiin ja 10 %:ssa vastauksista ensikieltä ei ollut nimetty. Ensikieleltään venäjänkielisiksi tunnistettiin kaiken kaikkiaan vähän yli puolet (58 %) (ns. varma tunnistaminen). Jos mukaan otetaan nekin vastaukset, joissa venäjän lisäksi oli mainittu joku muu kieli tai joitain muita kieliä ja ilmaistu epävarmuutta (ns. epävarma tunnistaminen), tunnistamisprosentti oli 68 %. Jatkossa tarkastelen vain ensimmäistä luokkaa, jossa arvioija ei ollut maininnut muita kieliä kuin venäjän vastauksessaan, eli 58 %:a (n=256) kaikista (n=440) vastauksista (ns. varma tunnistaminen)⁶.

Riippumatta tunnistamisen määrittelystä puhujat vaihtelivat siinä, kuinka hyvin heidät oli tunnistettu. Kuviossa 1 näkyy, että naispuhujat (valkoisella) tunnistettiin keskimäärin miehiä (mustalla) paremmin, esimerkiksi yksi miespuhujaa (Rusm1) tunnistettiin vain harvoin, kun taas kaksi naispuhujaa (Rusf4 ja Rusf5) lähes aina (95 % arvioijista). Puhujan sukupuolta ei kuitenkaan voida pitää ainoana syynä tunnistamiselle, sillä aiemman nimenomaan ensikieleltään venäjänkielisiä suomen puhujia koskevan tutkimuksen mukaan näyttää siltä, että näytteen taitotasolla on yhteys tunnistamiseen (Halonen ym. 2024). Tässäkin aineistossa yksi miespuhujaa (Rusm3) tunnistetaan lähes yhtä hyvin kuin kaksi eniten tunnistettua naispuhujaa. Pelkän yksittäisen näytteen perusteella⁷ arvioitu taitotason keskiarvo on merkitty kuvioon 1 puhujan pseudonyymien alle. Kuvion perus-

6 Näin ollen jatkotarkastelusta jää pois yksi arvioija, joka ei ollut tunnistanut yhtään venäjänkielistä näytettä oikein.

7 Yleisten kielitutkintojen tutkimuksessa puhumisen taitoa arvioidaan useamman puhetehtävän perusteella yhdeltä osallistujalta.

teella näyttää siltä, että taitotasolla saattaisi olla jonkin verran yhteyttä tunnistamiseen, sillä parhaan yleistaitotasoarvion saaneet miespuhujat (Rusm1 ja Rusm4) ovat vaikeimmin tunnistettavia ja heikoimmaksi naispuhujaksi arvioitu puhuja (Rusf5) on helpoiten tunnistettava. Kuitenkin heikoin miespuhujaksi arvioitu Rusm5 ei ole helpommin tunnistettava, eikä paras naispuhujaksi arvioitu Rusf2 vaikeimmin tunnistettavana. Khiin neliö -testin mukaan tunnistettujen ($n=182$) ja tunnistamattomien ($n=251$) näytteiden⁸ taitotasoarvioissa on merkitsevää eroa $\chi^2(5) = 43,481$; $p < 0,001$. Näyttäisi siltä, että tunnistaminen liittyy heikompiin (eli A2 ja B1) taitotasoarvioihin (tunnistettujen näytteiden sovitettu standardoitu jäännös tasolla A2 oli 3,1 ja tasolla B1 4,2, kun taas tasolla B2 se oli -5,5). Kun miesten ja naisten ryhmiä tarkasteltiin erikseen tunnistettavuuden suhteen Khiin neliö -testin mukaan naisilla tunnistettujen ($n=69$) ja tunnistamattomien ($n=148$) näytteiden taitotasoarvioissa on eroa $\chi^2(5) = 16,915$; $p < 0,005$, samoin kuin miehillä (tunnistettujen $n=103$, tunnistamattomat $n=113$). $\chi^2(5) = 48,847$; $p < 0,001$. Lisäksi saatiin selville, että naisilla riippuvuus näkyy erityisesti tasolla A2 (tunnistettujen näytteiden sovitettu standardoitu jäännös tasolla A2 oli 3,4) ja miehillä tasoilla B1 ja B2 (tunnistettujen näytteiden sovitettu standardoitu jäännös tasolla B1 oli 5,3 ja tasolla B2 -6,4).



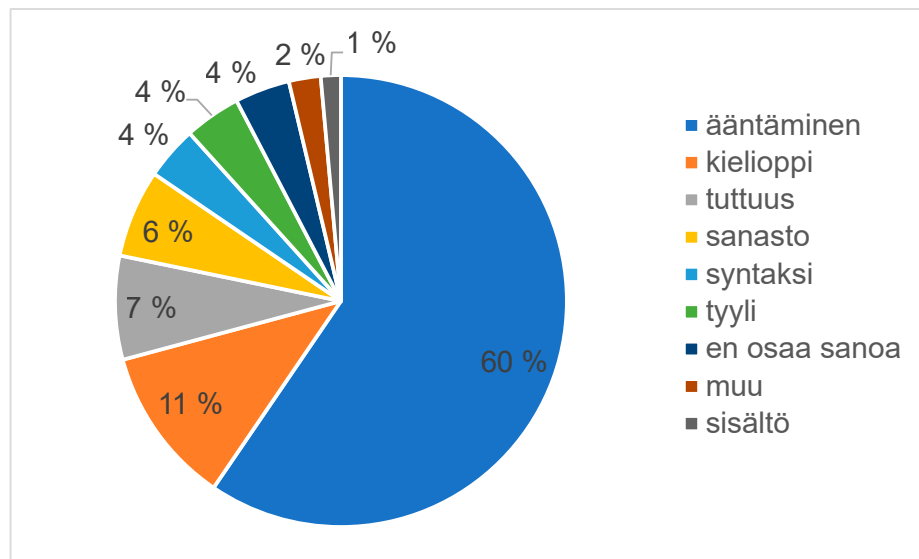
KUVIO 1. Eri puhujien tunnistaminen venäjää ensikielenään puhuviksi prosentteina arvioijien ($n=44$) mukaan.

⁸ Ristiintaulukoinnista puuttuu 7 arviota, joissa arvioija oli vastannut joko en käyttänyt kriteeriä tai kriteeriä ei pystynyt käyttämään/soveltamaan näytteessä.

Seuraavassa luvussa tarkastellaan artikkelin päätavoitetta eli sitä, liittyvätkö tietyt venäläisen aksentin ääntämispiirteet ensikielen tunnistamiseen.

5.2 Arvioijien antamat tunnistamisen perustelut

Luokittelin arvioijien antamat sanalliset perustelut ensikielen tunnistamiselle seuraaviin alaluokkiin: *ääntäminen, sanasto, kielioppi, tyyli, syntaksi, sisältö, tuttuus ja muu syy*. Annoin vastauksille painokertoimet sen mukaan, esiintyikö se perusteluissa ainoana luokkana vai yhdessä jonkun muun kanssa (ks. luku 5.1). Kategoriaan *en osaa sanoa* luokittelin tyhjien vastausten lisäksi muut epävarmuutta ilmaisevat vastaukset. Kuvioista 2 näkyy, että ääntäminen oli eniten (60 %) mainittu perustelu puhujan tunnistamiselle ja seuraavaksi eniten mainintoja saivat kieliopilliset seikat (11 %) ja tuttuus (7 %). Tuttuudella tarkoitetaan sitä, että arvioija raportoi venäläisen aksentin olevan itselleen tuttu esimerkiksi opetuskokemuksen kautta. Huomionarvoista on, että vaikka perustelut usein painottuivat virheisiin ja ensikielisestä ääntämisestä poikkeamiseen, mukana on myös vastauksia, joissa jonkun asian osaaminen hyvin oli merkinä siitä, että puhuja on venäläinen, esim. ”Perustan arvioni siihen, että vaikka puhujan kielitaito on vielä puutteellinen, hänellä on suomalainen intonaatio, minkä venäjänkieliset omaksuvat joskus hyvin.” (Arv27).

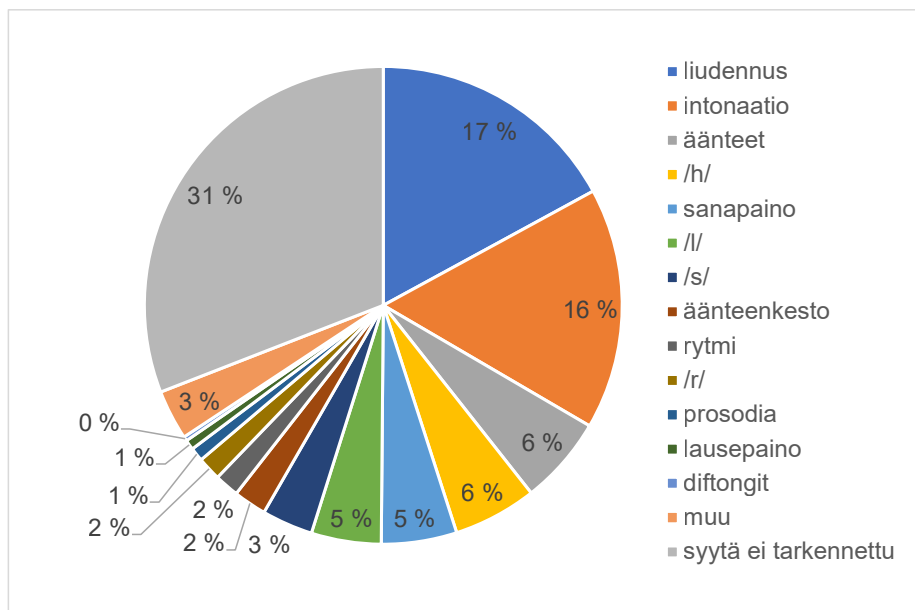


KUVIO 2. Arvioijien sanalliset perustelut (n=256) ensikieleltään venäjänkieliseksi tunnistamiselle prosentteina alaluokittain.

6 Ääntäminen perusteena ensikieleltään venäjänkieliseksi tunnistamiselle

Koska ääntäminen oli tärkein peruste puhujan ensikielen tunnistamiselle, luokittelin mainitut ääntämispiirteet edelleen seuraaviin luokkiin: *äänteet (ilman tarkennusta tietyistä äänneistä)*, *liudennus*, *h, r, s, l*, *diftongit*, *prosodia*, *sanapaino*, *intonaatio*, *äänteenkesto*, *rytmi*, *lausepaino* ja *muu syy*. Luokat ovat tosin osin päällekkäisiäkin, koska ne sisältävät ylä- ja alakäsitteitä ja rinnakkaisia käsitteitä. Esimerkiksi äänteenkestolla saatetaan viitata sanapainon tuottamiseen liioitellusti ja prosodially intonaatioon, mutta pyrin noudattamaan luokittelussa arvioijan sanallistamista mahdollisimman tarkasti.

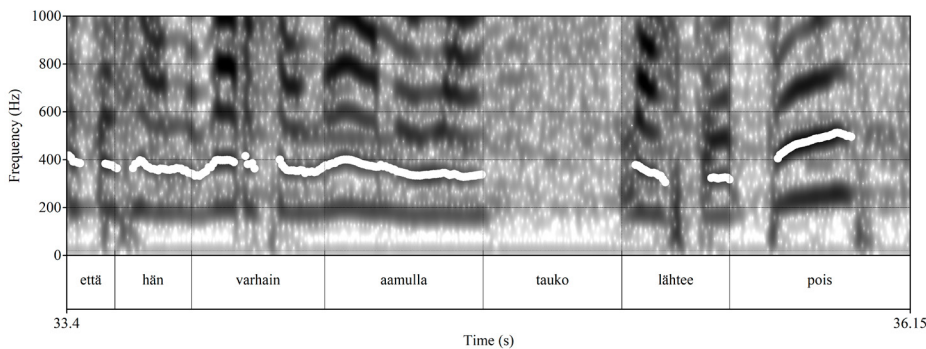
Kuviossa 3 näkyy eri ääntämisperusteiden prosentuaaliset osuudet. Suurin luokka (31 %) on se, missä ääntämisen mainitsemista ei tarkennettu. Tämän jälkeen suurimmat luokat ovat *liudennus* (17 %) ja *intonaatio* (16 %). Seuraavassa otan ääntämisperusteista esimerkkejä ja vertailen venäläisellä aksentilla puhuttua suomea ensikieleltään suomenkielisen puheeseen.



KUVIO 3. Arvioijien ääntämiseen liittyvät sanalliset perustelut (n=197) venäjänkieliseksi tunnistamiselle prosentteina alaluokittain.

6.1 Intonaatio

Intonaatio oli 16 %:ssa vastauksista ensikieleltään venäjänkieliseksi tunnistamisen perusteena. Arvioijien vastauksissa sanaa ”intonaatio” ei useimmiten ole avattu sen tarkemmin, mutta voinee olettaa, että useimmiten tarkoitetaan venäjälle tyypillistä intonaatiota kuten esimerkiksi ”intonaatio venäläinen ”laulava”” (Rusf4, Arv44), ”naisellinen” sävelkorkeus” (Rusf4, Arv14). Kuulonvarainen analyysini paljasti tämän lisäksi ehkä suomen kieltä jäljittelemään pyrkivän liioitellun monotonisen intonaation, josta tuli epäluonteva vaikutelma (Rusf3, Rusf4). Kuulonvaraisessa analyysissä kiinnitin huomiota myös yleisesti korkeaan perustaajuuteen (Rusm1) ja venäjän kielelle tyypilliseen nousevaan intonaatioon lauseen lopussa (Rusf2), kun puhuja aikoo vielä jatkaa aiheesta (Kuvio 4). Lisäksi kiinnitin kahdella puhujalla (Rusf4 ja Rusf5) huomiota siihen, että he käyttivät narinaa ensikielisten puhujien tapaan puhunnoksen lopun merkkinä, vaikka myös venäjän kielelle tyypillistä ilmaisun lopun intonaatiota esiintyy.

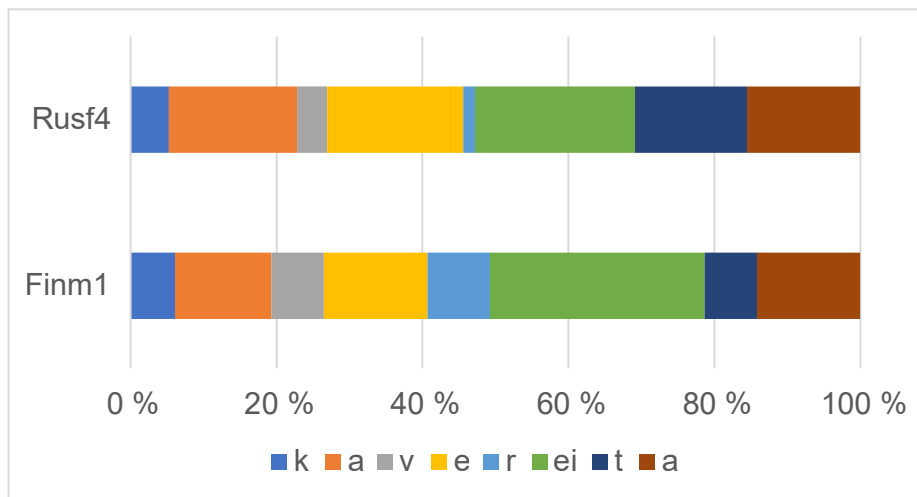


KUVIO 4. Spektrogrammin päällä näkyvä ensikieleltään venäjänkielisen puhujan (Rusf2) perustaajuuskäyrä, jossa perustaajuuskäyrän nousu puhunnoksen lopussa venäjän kielelle tyypillisellä tavalla.

6.2 Muut prosodiset piirteet

Muita prosodisia piirteitä kuin intonaatio mainittiin 11 %:ssa ääntämiseen liittyvistä perusteluista. Tarkemmin mainittiin *sanapaino*, *rytmi*, *ääntenkesto* ja *lausepaino*. Ääntenkestolla tarkoitetaan ehkä usein vokaalin kestoa, joka puolestaan usein liittyy sanapainoon. Sanapainosta kommentoitiin mm. ”painottaa ja venyttää joidenkin sanojen ensimmäisiä tavuja” (Rusm1, Arv33), ”joissakin sanoissa erikoinen painotus: turvallinen - paino LI-tavulla” (Rusm5, Arv14), ”painottomat vokaalit lyhenevät” (Rusf4, Arv23) ja ”toisen tavun pitkien vokaalien lyhentyminen” (Rusf5, Arv12). Sanapainon

tuottamisen vaikeuteen kiinnitin huomiota kuulonvaraisessa analyysissä samoilla neljällä puhujalla (Rusm1, Rusm3, Rusf3, Rusf4). Tyypillisesti ensikieleltään venäjänkieliset lausuvat ensimmäisen eli painollisen tavun vokaalin liioitellun pitkänä ja vastaavasti painottomat vokaalit liioitellun lyhyinä, vaikka ne olisivatkin fonologisesti pitkiä. Tämä vaikuttaa myös sanan rytmisen rakenteen horjumiseen. Kuviossa 5 on esimerkki sanasta *kavereita*. Kunkin äänteen (tai diftongin) kesto on suhteutettu sanan kokonaiskestoan. Tästä nähdään, että ensikieleltään venäjänkielisellä puhujalla ensitavun vokaali on tarpeettoman pitkä. Tosin sanan muutkin vokaalit ovat melkein yhtä pitkiä, eivätkä lyhyempiä, kuten taustakirjallisuuden perusteella voisi olettaa. Monasti ensitavun vokaali on suhteessa vieläkin pidempi kuin tässä esimerkissä, mutta spontaanista puheesta on vaikea löytää eri puhujilta täysin vertailukelpoisia esimerkkejä. Ensitavun vokaalin keston liioittelu lienee venäjän vaikutusta, koska venäjässä sanapainon tärkein parametri on kesto, kun taas suomessa näin ei ole.

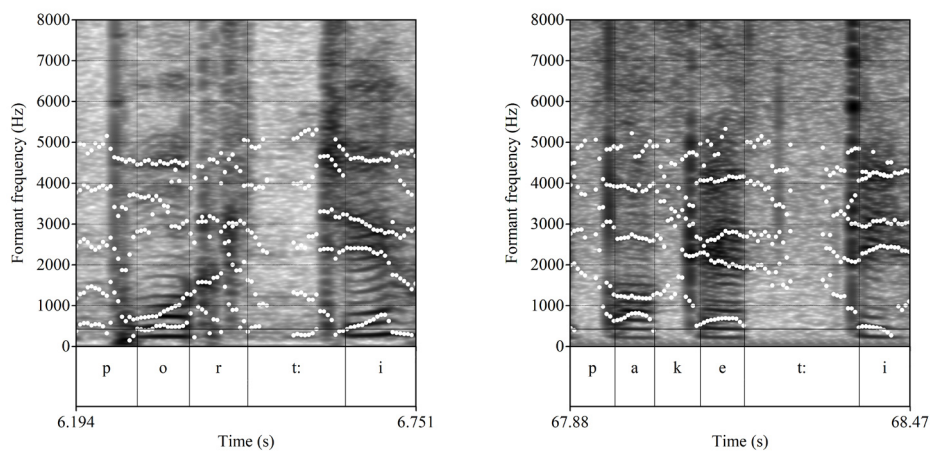


KUVIO 5. Äänteiden ja diftongin suhteelliset kestot sanassa *kavereita* suomenkielisellä (Finm1) ja ensikieleltään venäjänkielisellä (Rusf4) puhujalla.

6.3 Liudennus

17 % ääntämiseen liittyvistä perusteluista koski liudennusta. Arvioijien kommentoissa kuvattiin esimerkiksi ”vahvoja pehmennyksiä ja liudennuksia” (Rusf3, Arv41) ja ”pehmeitä j-äänteitä vokaalien edellä” (Rusf5, Arv41). Itse kiinnitin liudennukseen huomiota erityisesti kolmella puhujalla (Rusm3, Rusf2 ja Rusf5), joilla se esiintyi systemaattisesti.

Kuviossa 6 on kuvattu ensikieleltään venäjänkielisen puhujan (Rusf2) sanan *portti* loppuvokaali, jossa liudennus kuuluu selvästi ja kielen selän lähentyminen kitalakea näkyy toisen formantin (F2) arvoissa, jotka ovat korkeammat heti klusiilin jälkeen kuin vokaalin lopussa. Liudennus tämän sanan lopputavussa johtunee siitä, että venäjän kielessä konsonantti liudentuu vokaalin /i/ edellä. Vertailukohtana on suomenkielisen puhujan (Fif3) vastaava äänne, jossa F2 pysyy melko tasaisena läpi vokaalin.



KUVIO 6. Vasemmalla ensikieleltään venäjänkielisen puhujan (Rusf2) ja oikealla suomenkielisen puhujan (Fif3) tuottama sananloppuinen [i] sanoissa *portti* ja (*keksi*) *paketti*.

6.4 Tietyt yksittäiset äänneet

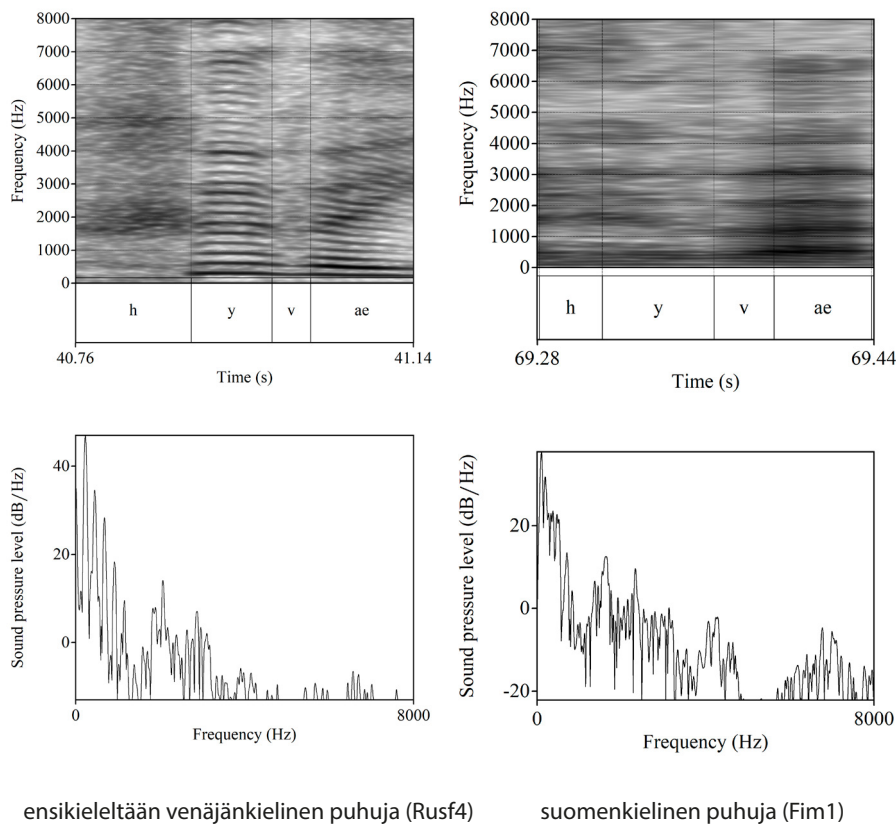
Yksittäisistä äänneistä venäjänkieliseksi tunnistettavista mainittiin useimmin /l/ (5 %), /h/ (6 %), /s/ (noin 3 %) ja /r/ (noin 2 %).

6.4.1 /l/:n ääntyminen

Lateraaliin /l/ kiinnitin huomiota kuulonvaraisessa analyysissä erityisesti vain yhdellä puhujalla (Rusf5), mutta korvani ei ollut tässä yhtä tarkka kuin arvioijilla, jotka kiinnittivät /l/-äänneeseen huomiota kahdeksalla puhujalla kymmenestä (vain Rusm1 ja Rusm5 eivät saaneet tästä mainintaa). Tämä johtuu ehkä siitä, että omassa analyysissäni luokittelin haasteet /l/-ääntymisessä yleisemmin liudennuksen alle. Arvioijat kommentoivat tätä äännettä seuraavasti: ”paksu l” (Rusf1, Arv8) ja ”pehmeä l” (Rusf2, Arv37).

6.4.2 /h/:n ääntyminen

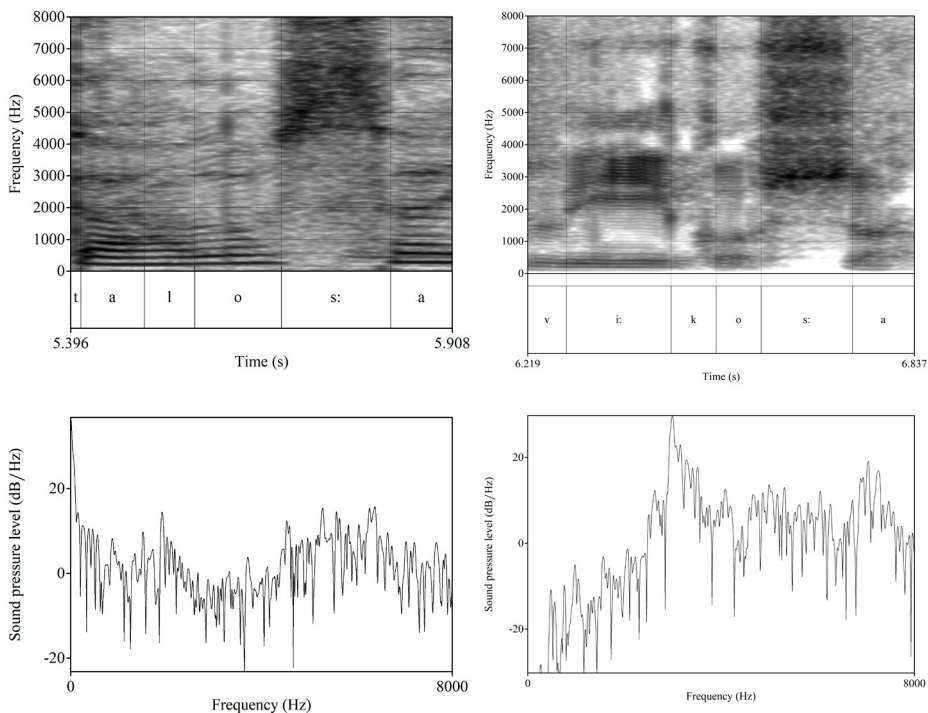
Kuulonvaraisessa analyysissä /h/ erottui suurimmalla osalla (7/10) puhujista (Rusm2, Rusm3, Rusm5, Rusf1, Rusf2, Rusf3, Rusf4 ja Rusf5) pitkällä kestollaan ja voimakkaalla hankaushälyllään. Arvioijat mainitsivat tämän äänteen ainakin kerran puhujilla Rusm2, Rusm5, Rusf3, Rusf4 ja kun luonnehtivat sitä, mainitsivat ”taka-h” (Rusf4, Arv22), ”vahva H” (Rusm5, Arv17) ja ”pehmeä h” (Rusf4, Arv6). Otan esimerkin sanasta *hyvä* venäjänkielisellä (Rusf4) ja suomenkielisellä puhujalla (Finm1) (Kuvio 7). Rusf4 sai useita mainintoja [h]:n ääntämisestä ja hän äänsi sanan *hyvä* useita kertoja näytteen aikana. Hänellä äänne on hyvin voimakas frikatiivi, voimakkain energia keskittyy 2000 Hz:n ympärille. Suomenkielisellä puhujalla [h] on heikommin äännetty. Tämä on tyypillistä sananalkuisessa asemassa suomen kielessä. Spektrogrammissa energia on keskittynyt 1000–3000 Hz:n välille.



KUVIO 7. Oikealla suomenkielisen (Fim1) ja vasemmalla ensikieleltään venäjänkielisen (Ruf4) puhujan [h] sanassa *hyvä*. Spektri on otettu [h]:n keskikohdasta.

6.4.3 /s/:n ääntyminen

Kuulonvaraisessa analyysissä kiinnitin huomiota /s/:n ääntymiseen etenkin kolmella puhujalla kymmenestä (Rusf2, Rusf3 ja Rusf4). Arvioijat mainitsivat ”ässän” viidellä puhujalla (Rusm2, Rusf2, Rusf3, Rusf4 ja Rusf5) ja yhteensä 3 % ääntämisperusteluita. Useimmiten äännettä kuvattiin ”terävänä”, mutta oli myös tapauksia, joissa se oli äännetty ”liian suhisevana” eli todennäköisesti ääntäminen oli lähempänä varianttia [ʃ]. Lisäksi ainakin yhdellä puhujalla (Rusf4) esiintyi soinnillinen variantti [z]. Hän äänsi sanan *musiikki* ensin [muzikas] (venäjäksi *musiikki* on [ˈmuzɪkə]) ja korjasi sitten [musi:k:iɔ]. Ässän maininta yleensäkin on kiinnostavaa ja hieman yllättävääkin, sillä voisi ajatella, että kun suomessa on vain yksi ”pakollisesti käytössä oleva” s-äänne, se sallisi paljonkin vaihtelua realistumisissaan, ainakin niin, ettei sen ”terävyys” häiritisi kuulijaa.



ensikieleltään venäjänkielinen puhuja (Rusf4)

suomenkielinen puhuja (Fim3)

KUVIO 8. Oikealla suomenkielisen (Fim3) ja vasemmalla ensikieleltään venäjänkielisen (Ruf4) puhujan [s:] sanoissa *viikossa* ja *talossa*. Spektri on otettu [s:]n keskikohdasta.

Kuviossa 8 on havainnollistettu suomenkielisen ja ensikieleltään venäjänkielisen puhujan [s:]. Ensikieleltään venäjänkielisen puhujan (Rusf2) ”ässän” mainitsi useampi arvioija luonnehtimalla sitä teräväksi. Yksi arvioija (Arv27) pitikin sitä ainoana kyseisen puhujan venäjänkielisyteen viittaavana piirteenä:

Mitään erityistä venäläisyyttä en puheessa havaitse. Tein arvioni puhettavan perusteella. Hyvin kieltä osaavat, hyvinkoulutetut venäläiset naiset puhuvat terävästi ja täsmällisesti. Ehkä terävä s viittaa venäjään. Tämä tosin on käytännön havaintoni, en itse osaa venäjää analysoidakseni tarkemmin.

Suomenkielisellä puhujalla voimakkain energia sijoittuu 3000 Hz:n molemmin puolin, kun taas ensikieleltään venäjänkielisellä vasta noin 4000 Hz kohdalta. Kun seuraavasta vokaalista (molemmilla puhujilla sananloppuinen [ɑ]) mitattiin F2-arvot vokaalin alusta ja keskeltä, saatiin tulokseksi ensikieleltään venäjänkieliselle puhujalle 1827 Hz ja 1951 Hz ja suomenkieliselle 1631 Hz ja 1267 Hz. Tämä kertoo todennäköisesti siitä, että ensikieleltään venäjänkielisen puhujan sibilantti on ääntynyt edempänä (eli lähempänä venäjän dentaalia ääntöpaikkaa) kuin suomenkielisen, vaikkakin kuulohavainnon perusteella myös vokaalin laatu muistuttaa suomenkielisellä puhujalla enemmän svaavokaalia.

6.4.4 /r/:n ääntyminen

Arvioijat mainitsivat myös /r/-ääntymisen perusteena ensikieleltään venäjänkieliseksi tunnistamiselle neljällä puhujalla kymmenestä (Rusm4, Rusf2, Rusf3 ja Rusf4). Erityisesti yksi arvioija (Arv14) kommentoi sitä näin: ”vahva r-äänne vie huomion” (Rusm4) ja ”Jäin kuuntelemaan voimakasta r-äännettä, joka tuntui vievän huomion muusta ääntämisestä” (Rusf3). Kuulonvaraisessa analyysissäni en kiinnittänyt siihen huomiota, joten en tehnyt akustista analyysiä tästä äänneestä.

6.4.5 Vokaalit

Vokaalit tuotiin esiin yleisesti, mutta ei mainittu, missä vokaalissa oli erityisesti huomattu poikkeamaa. Vain vokaalit /æ/ ja /y/ nimettiin muutamia kertoja. Nämä olivat samat, joihin kiinnitin itsekkin huomiota kuulonvaraisessa analyysissäni muutamien puhujien kohdalla. Luokittelin kaikki nämä maininnat äänneet-luokkaan. Jonkin verran oli myös ”j-mäiset i- ja e-kirjaimet” (Arv42) tyyppisiä vastauksia, jotka luokittelin luokan liudennus alle. Arvioijat mainitsivat myös diftongit muutamia kertoja. Kuulonvaraisessa analyysissäni kiinnitin huomiota diftongeihin kolmella puhujalla (Rusm5, Rusf2, Rusf5).

Otan esimerkin sanasta *ystävä*, jonka ääntäminen [jʏstävɑ] paljasti arvioijien mukaan puhujan Rusf5 ensikieleltään venäjänkieliseksi. Kompastuskivenä on paitsi sananalkuisen [y]:n ääntäminen [j]-alkuisena, myös [æ]:n ääntäminen [ɑ]-maisena. Tämä johtunee siitä, että venäjän kielessä ei ole laveaa etuvokaalia /æ/ (vaan lavea

keskivokaali /a/ ja laeva takavokaali /ɑ/) ja venäjistä puuttuu niin ikään pyöreä etuvokaali /y/ (ks. esim. Bondarko 2009; de Silva ym. 2010). Ensikieleltään venäjän puhujat käyttänevät usein suomen /y/:n tilalla venäjän /ɯ/ foneemia (kirjoituksessa ю). Sanassa *ystävä* [y] ääntyy heillä [jɯ], koska venäjän /ɯ/ foneemi reaalistuu sananalkuisessa asemassa äännejonona [jɯ] (ks. esim. Bondarko 2009; de Silva ym. 2010).

7 Puhujakohtainen tarkastelu

Harvimmin⁹ ensikieleltään venäjänkieliseksi tunnistetulla miespuhujalla (Rusm1) oli kuulonvaraisen arvioni mukaan vahva aksentti, ja kiinnitin huomiota erityisesti puhujan intonaatioon ja liioiteltuun ensitavun vokaalin kestoan. Arvioijille hän ei kuitenkaan näyttäytynyt tyypillisenä ensikielisenä venäjän kielen puhujana, koska vain viisi tunnisti hänet varmasti (11 oletti hänet ensikieleltään romanisten kielten puhujaksi ja kuusi englannin). Sanallisissa kuvauksissa arvioijat perustelivat muita kieliä kuin venäjää mm. aspiraatiolla ja r:n laadulla. Harvimmin¹⁰ tunnistetulla naispuhujalla Rusf1 puolestaan erottuivat kuulonvaraisessa analyysissä lähinnä intonaatiopiirteet. Arvioijat olivat maininneet hänen puheessaan myös liudennuksen, sanapainon ja rakenteet.

Parhaiten tunnistettu miespuhujalla Rusm3 on ”selkeä tapaus” (Arv17), ja hän on usean arvioijan mielestä tyypillinen ensikielinen venäjänkielinen puhuja. Hänet tunnistettaneen siksi, että hänen puhenäytteessään esiintyy arvioijien mukaan liudennusta, sanapainon tuottamisen haasteita (äännekestojen horjumista, joidenkin tavujen tarpeetonta venytystä, toisen tavun pitkän vokaalin lyhentymistä) ja venäjän kielelle tyypillistä intonaatiota. Lisäksi hänen kohdallaan mainitaan perusteluina myös sanajärjestys ja kielen rakenteeseen liittyviä asioita. Kuulonvaraisessa analyysissä kiinnitin huomiota osin samoihin asioihin kuin arvioijat: liudennukseen, h-äänteeseen ja intonaatioon.

Erittäin hyvin tunnistettujen naispuhujien (Rusf4 ja Rusf5) näytteet oli arvioitu tasoille B1 ja A2. Heillä oli ääntämisessään useita venäjän kielelle tyypillisiä piirteitä, kuten Rusf4:llä intonaatio, sanapaino (ja siihen liittyvät äännekestit) ja äännepoikkeamat (h, s, r) ja Rusf5:llä sanapaino, liudennus, diftongit ja äännepoikkeamat (l, s). Molemmilla puhujilla oli kuitenkin myös yksittäisiä sanoja, jotka he lausivat erityisen poikkeavasti ja joihin moni arvioija oli kiinnittänyt huomiota. Rusf4 lausui *musiikki* voimakkaasti soinnillisella sibilantilla ja Rusf5 *ystävä* venäjälle tyypillisellä j-mäisellä alulla. Kuitenkin heillä oli myös ääntämisen osaamista, sillä muista puhujista poiketen heillä ilmeni esimerkiksi suomen kielelle tyypillistä narinaa ilmauksen lopussa.

9 11 % arvioijista oli tunnistanut hänet ensikieleltään venäjänkieliseksi.

10 34 % arvioijista oli tunnistanut hänet ensikieleltään venäjänkieliseksi.

8 Päätäntö

Tutkimus osoitti, että kymmenen ensikieleltään venäjänkielisen suomen puhujan tunnistaminen vaihteli suuresti puhujittain. Tunnistaminen vaihteli 10 prosentista 90 prosenttiin. Näyttäisi siltä, että tunnistaminen liittyy näytteelle annettuun yleistaitotasoarvioon. Arvioijat nimesivät ääntämiseen liittyvät tekijät tärkeimmäksi (60 %) tunnistamisen perusteeksi. Yksittäisistä ääntämisperiteistä eniten mainintoja saivat intonaatio ja liudennus, mutta myös muita prosodisen ja äännetason piirteitä nimettiin.

Tulokset vahvistavat myös aiemman tutkimuksen (esim. Aho ym. 2016) tuloksia siitä, että haasteita tuottavat konsonanteista /l/ ja /h/ ja liudentumisen ja intonaation siirtyminen venäjästä suomeen. Saadut tulokset ovat linjassa sen suhteen, mitä tiedetään ensikieleltään venäjänkielisistä vironoppijoista, joilla niin ikään äännekestot, palatalisaatio, sanapainon tuottaminen ja vokaalien laatu ovat haasteellisia (Meister & Meister 2007) ja osin samankaltaisia (palatalisaatio ja painollisten vokaalien liian pitkä kesto) haasteita kuin puolankielisillä suomenoppijoilla (Silfverberg 1982). Tulos vokaalikestojen tuottamisen vaikeudesta ei ole aiemman kirjallisuuden valossa yllätys (Aho & Toivola 2008; Toivola 2011; Ullakonoja & Kuronen 2015). Se, mikä tutkimuksessani on kuitenkin uutta, on se, että äännekestot nimettiin yhdeksi syyksi tunnistaa ensikieleltään venäjänkielinen puhuja.

Aiemmassa tutkimuksessa (Aho ym. 2016) raportoitu painottomien vokaalien redusoituminen laadullisesti tunnistamattomaksi ei sen sijaan noussut omassa tutkimuksessani venäläisen aksentin piirteeksi. Tutkimuksessani uutta oli ”terävän ässän” nimeäminen venäläisen aksentin piirteeksi. Suomessa kuitenkin /s/:n kategoria on verrattain laaja (ja sen toteuma myös vaihtelee hieman vokaaliympäristön mukaan) (ks. esim. Raimo & Ojala 2009), joten on hieman yllättävää, että arvioijat kiinnittävät siihen huomiota. Toisaalta tiedetään, että ”ässän” havaitsemisessa suomalaiset ovat aika ”herkkiä” ja poikkeamat siinä tuntuvat tarjoutuvan suomalaisille helposti vierauden merkiksi (ks. Halonen, Nyström ym. 2020). Uutta tässä tutkimuksessa oli myös joidenkin arvioijien mainitsema /r/ venäläisen aksentin piirteenä. Venäjässä /r/ on samaan tapaan täryäänne kuin suomen kielessäkin, mutta ääntöpaikaltaan se on venäjässä dentaali ja suomessa alveolaari (Bondarko 2009; Iivonen 2009a), joten on mahdollista, että ensikieleltään venäjänkieliset ääntävät sen suomessa eri tavalla. Nämä kuulohavainnot ovat kuitenkin vain suuntaa antavia, eikä niiden perusteella voi siten ilman laajempia akustisia mittauksia tehdä päätelmää kyseisten äänneiden tyypillisyydestä venäläisen aksentin piirteenä.

Kiinnostavaa oli myös, että suomen kielen kaltaista lausuman loppuista narinaa esiintyi joillain puhujilla – piirre, jonka puuttuminen on aiemmissa tutkimuksissa yhdistetty vahvaan vieraaseen aksenttiin. Narinan puuttumista ei kukaan arvioijista kuitenkaan ollut nimennyt venäläisen aksentin piirteeksi. On mahdollista, että se on nimenomaan vieraan aksentin piirre, ei erityisesti venäläisen ja että narinan esiinty-

minen kertoo pikemminkin kielitaidon tasosta kuin vieraasta aksentista. Kallion ym. (2022) tutkimuksessa narina oli merkittävä suullisen kielitaidon tason ennustaja, kun aineistona oli suuri määrä (n=160) eri ensikielisiä puhujia Yleisten kielitutkintojen keskitason ja ylimmän tason suomen testistä. Vastaavaa tendenssiä ei tämän tutkimuksen suppeammassa aineistossa havaittu.

Se, miten hyvin kuulijat tunnistavat venäläisellä aksentilla puhutun suomen riippuu luonnollisesti tutkimusasetelmasta. Aiemmassa tutkimuksessa jopa suomea osaamattomat ensikieleltään venäjänkieliset kuulijat erottivat venäjää ensikielenään puhuvat suomen puhujat ensikielisistä puhujista varsin lyhyissä puhenäytteissä, kun vastaajille annettiin kaksi ensikielivaihtoehtoa: suomi ja venäjä (Toivola & Ullakonoja 2017). Tässä tutkimuksessa tehtävä oli kuitenkin haastavampi: puhujan ensikieltä kysyttiin avokysymyksenä, jolloin vaihtoehtoja oli, jos ei nyt rajattomasti, niin kuitenkin ainakin satoja. Lisäksi päätehtävä oli arvioida suullista kielitaitoa ja siksi aksentin alkuperän määrittäminen ei ollut kaikille arvioijille helppo eikä mielekäskään. On mielenkiintoista pohtia, miten paljon venäjän piirteitä puhenäytteessä tarvitsee olla, jotta se tunnistetaan ensikieleltään venäjänkielisen puhujan näytteeksi. Riittääkö yksittäinen venäläisellä aksentilla äännetty sana, esimerkiksi [jʉstava] tai [muzikas]? Mahdollisesti. Ainakin nämä kyseiset sanat esiintyivät puhujilla Rusf4 ja Rusf5, jotka tunnistettiin ensikieleltään venäjänkieliseksi yli 80 %:ssa arvioista. Kyseisillä puhujilla oli kuitenkin myös muita venäjän kielelle tyypillisiä piirteitä puheessaan, joten ei voida varmuudella sanoa, että tunnistaminen johtuisi pelkästään näistä sanoista.

Arvioijien antamat sanalliset perusteet ensikieleltään venäjänkielisten tunnistamisesta keskittyivät ääntämispiirteisiin, mutta myös muita puheen piirteitä, kuten sanastollisia ja rakenteellisia seikkoja mainittiin. Näiden tutkiminen olisi eittämättä kiinnostava jatkotutkimuksen aihe. Arvioijat myös eroavat siinä, kuinka hyvin he erittelevät ensikieleltään venäjänkieliseksi tunnistamisen perusteluitaan tarkemmin ääntämispiirteisiin, ja siinä, mihin äänneisiin he kiinnittävät huomiota. Jatkossa olisi tärkeä huomioida arvioijan taustojen merkitys, koska aiempi tutkimus (esim. Isaacs & Thomson 2013; Winke ym. 2012; Kahng 2023) on osoittanut, että arvioijan henkilökohtaiset ominaisuudet (esim. kielikyky) ja taustat (kuten opetus- ja arviointikokemus, kielitaito ja kyseisen aksentin tuttuus) voivat vaikuttaa arviointiin ja esimerkiksi hänen kykynsä eritellä ääntämisvirheitä. Myös puhujien taustojen merkitys olisi kiinnostava jatkotutkimuksen aihe. Onko siinä, että naiset tunnistettiin helpommin kuin miehet, kyse siitä, että arvioijilla on mielessään vahvemmin stereotypia ensikieleltään venäjänkielisestä nais- kuin miespuhujasta? Entä onko puhujan iällä tai koulutustaustalla merkitystä aksentin tunnistamisen kannalta?

9 Lopuksi

Kuten jo alussa mainittiin, toisen ja vieraan kielen oppimisen tutkimuskenttä on kansainvälisestäkin keskittynyt enimmäkseen englantiin opittavana kielenä, mihin myös ääntämisen oppimisen teorit pohjautuvat. Siispä tutkimukset, jotka keskittyvät muihin kuin englannin kieleen opittavana kielenä ovat ensiarvoisen tärkeitä. Kielitaidon arvioinnin näkökulmasta tutkimus arviointikriteerien ulkopuolisista tekijöistä on tärkeää arviointijärjestelmän kehittämisen kannalta. Arvioijakoulutuksessa on syytä kiinnittää huomioita arviointiin vaikuttaviin subjektiivisiin tekijöihin, jotta arviointi olisi entistä objektiivisempää.

Tutkimukseni tuloksia voidaan hyödyntää paitsi venäjää ensikielensä puhuvien suomen kielen opetuksessa, myös ukrainankielisten suomen opetuksessa, koska ukrainan ja venäjän kielten fonologiset järjestelmät ovat hyvin samankaltaisia ja siten ukrainalaisen aksentin piirteet suomessa ovat todennäköisesti osittain samanlaiset kuin venäläisen.

Kiitokset

Kiitän kaikkia tutkimukseen osallistuneita puhujia ja arvioijia. Kiitän myös artikkelin anonyymeja arvioijia arvokkaista kommentteista käsikirjoitukseni ensimmäisen version parantamiseksi. Lisäksi kiitän Mia Halosta artikkelin kommentoinnista sekä Mikko Kurosta ja Reeta Neittaanmäkeä tuesta analyysissä. Arvokasta oli myös *Rikkinäistä suomea: aksenttien arviointi yhteiskunnallisena portinvartijana* -hankkeen rahoittajan Suomen Akatemian (hankenumero 21000042461) antama mahdollisuus työstää tätä artikkelia tutkijavierailulla University College Londonissa. Lopuksi haluan kiittää Viola de Silvaa opeista ja inspiraatiosta venäjän ja suomen vertailevan foneettisen tutkimuksen maailmassa.

Kirjallisuus

- Aho, E., M. Nikonen & M. Toivola 2019. Suomi toisena kielenä. Teoksessa E. Tergujeff & M. Kautonen (toim.) *Suullinen kielitaito: Opi, opeta, arvioi*. Helsinki: Otava. 182–183.
- Aho, E. & M. Toivola 2008. Venäläisten maahanmuuttajien suomen prosodiasta. *Virittäjä*, 112 (1), 3–23. <https://journal.fi/virittaja/article/view/40638>
- Aho, E., M. Toivola, F. Karlsson & M. Lennes 2016. Aikuisten maahanmuuttajien suomen ääntämisestä. *Puhe ja kieli*, 36 (2), 77–96.
- Ahola, S. 2016. Puhetta arvioinnista: yleisten kielitutkintojen arvioijien käsityksiä arvioinnista. Teoksessa A. Huhta & R. Hildén (toim.) *Kielitaidon arviointitutkimus 2000-luvun Suomessa*. AFinLA-e. Soveltavan kielitieteen tutkimuksia 9. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 89–109. <https://journal.fi/afinla/article/view/60848>
- Ahola, S. 2022. *Rimaa hipoen selviää tilanteesta. Yleisten kielitutkintojen suomen kielen arvioijien käsityksiä kielitaidon arvioinnista ja suullisesta kielitaidosta*. JYU dissertations 489. Jyväskylä: Jyväskylän yliopisto. <http://urn.fi/URN:ISBN:978-951-39-9005-3>
- Ahola, S. & M. Halonen 2021. 'Broken Finnish': speaker L1 and its recognition affecting rating in National Certificates of Language Proficiency test in Finnish. Teoksessa G. Seed (toim.) *Collated Papers for the ALTE 7th International Conference, Madrid*. Madrid: Association of Language Testers in Europe, 53–57. <https://www.alte.org/resources/Documents/ALTE%207th%20International%20Conference%20Madrid%20June%202021.pdf> [luettu 26.3.2024]
- Ahola, S. & T. Hirvelä 2024. Venäjänkielisten suomen puhumisen taito taustamuuttajien valossa. *Idäntutkimus* 31 (1), 4–21. <https://doi.org/10.33345/idantutkimus.138163>
- Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proceedings*, 17, 97–110. https://www.fon.hum.uva.nl/paul/papers/Proceedings_1993.pdf
- Boersma, P. 2013. Acoustic analysis. Teoksessa R. Podesva & D. Sharma (toim.) *Research methods in linguistics*. Cambridge: Cambridge University Press, 375–396.
- Boersma, P. & D. Weenink 2023. Praat: doing phonetics by computer [Tietokoneohjelma]. Versio 6.3.08. <http://www.praat.org> [ladattu 10.2.2023]
- Bondarko, L. V. 1998. *Fonetika sovremennogo russkogo jazyka*. Sankt-Peterburg: Izdatel'stvo Sankt-Peterburgskogo gosudarstvennogo universiteta.
- Bondarko, L. V. 2009. Short description of Russian sound system. Teoksessa V. de Silva & R. Ullakonoja (toim.) *Phonetics of Russian and Finnish. General description of phonetic systems: experimental studies on spontaneous and read-aloud speech*. Frankfurt: Peter Lang, 23–35.
- Council of Europe 2020. *Common European Framework of Reference for Languages: learning, teaching, assessment*. Strasbourg: Council of Europe Publishing. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4> [luettu 5.4.2024]
- Dufva, H. & M. Halonen 2016. Svetlana Rönkkö ja Tove Hansson: vieraan aksentin performanssi mediahuumorin resurssina. Teoksessa A. Solin, J. Vaattovaara, N. Hynninen, U. Tiililä, & T. Nordlund (toim.) *Kielenkäyttäjät muuttuvissa instituutioissa*. AFinLAN vuosikirja 2016. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 145–162. <https://journal.fi/afinlavk/article/view/59724>
- Field, A. 2009. *Discovering statistics using SPSS*. London: Sage Publications.
- Flege J. E. 1984. The detection of French accent by American listeners. *The Journal of the Acoustical Society of America*, 76 (3), 692–707. <https://doi.org/10.1121/1.391256>

- Flege, J. E. 1988. The production and perception of foreign language speech sounds. Teoksessa H. Winitz (toim.) *Human communication and its disorders*. Norwood: Ablex, 224–401.
- Grosjean, F. 2010. *Bilingual, Life and Reality*. Cambridge: Harvard University Press.
- Halonen, M., S. Ahola, T. Hirvelä, R. Neittaanmäki, S. Ohranen & R. Ullakonoja 2024. Kielitaidon arviointi kansalaisuuden portinvartijana. Teoksessa T.A. Renvik & M. Säävälä (toim.) *Kotoutumisen kokonaiskatsaus 2023: Näkökulmana väestösuhteet*. TEM oppaat ja muut julkaisut 2024:1 FI. Helsinki: Työ- ja elinkeinoministeriö, 77–86.
- Halonen, M., A. Huhta, S. Ahola, T. Hirvelä, R. Neittaanmäki, S. Ohranen & R. Ullakonoja 2020. Ensikielen tunnistamisen merkityksestä suullisen kielitaidon arvioinnissa Yleisissä kielitutkinnoissa. Teoksessa S. Grasz, T. Keisanen, F. Oloff, M. Rauniomaa, I. Rautiainen & M. Siromaa (toim.) *Menetelmällisiä käännteitä soveltavassa kielentutkimuksessa – Methodological turns in applied language studies*. AFinLAN vuosikirja 2020. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 56–70. <https://doi.org/10.30661/afinlavk.89453>
- Halonen, M., S. Nyström, H. Paunonen & J. Vaattovaara 2020. *Stadin syntinen s*. Helsinki: Art House.
- Hirvelä, T., R. Neittaanmäki, S. Ahola & M. Halonen 2022. Statistical analyses of project Broken Finnish (Rikkinäistä suomea). V. 30.8.2022. <https://doi.org/10.17011/jyx/dataset/85234>
- Hirvonen, P. 1970. *Finnish and English communicative intonation*. Turun yliopiston fonetiikan laitoksen julkaisuja 8. Turku: Turun yliopisto.
- Iivonen, A. 2009a. Major features of standard Finnish phonetics. Teoksessa V. de Silva & R. Ullakonoja (toim.), *Phonetics of Russian and Finnish. General description of phonetic systems: experimental studies on spontaneous and read-aloud speech*. Frankfurt: Peter Lang. 47–65.
- Iivonen, A. 2009b. Finnish sentence accent and intonation. Teoksessa V. de Silva & R. Ullakonoja (toim.) *Phonetics of Russian and Finnish. General description of phonetic systems: experimental studies on spontaneous and read-aloud speech*. Frankfurt: Peter Lang. 67–73.
- Isaacs, T. & R. I. Thomson 2013. Rater experience, rating scale length, and judgments of L2 pronunciation: revisiting research conventions. *Language Assessment Quarterly*, 10 (2), 135–159. <https://doi.org/10.1080/15434303.2013.769545>
- Johnson, K. 2011. *Acoustic and auditory phonetics*. Hoboken: Wiley-Blackwell.
- Kahng, J. 2023. Exploring individual differences in rating second language speech: rater's language aptitude, major, accent familiarity, and attitudes. *TESOL Quarterly*, 57 (4), 1545–1557. <https://doi.org/10.1002/tesq.3217>
- Kallio, H., R. Suviranta, M. Kuronen & A. von Zansen 2022. Creaky voice and utterance fluency measures in predicting perceived fluency and oral proficiency of spontaneous L2 Finnish. Teoksessa S. Frota, M. Cruz & M. Vigário (toim.) *Proceedings of the 11th International Conference on Speech prosody 2022*. Lisbon: International Speech Communication Association (ISCA), 777–781. <https://doi.org/10.21437/SpeechProsody.2022-158>
- Kim, J., M. Toivola, L. M. Heikkola & A. Saloranta 2022. Saisinko puhua suomea? *Kielikello* 3/2022. <https://www.kielikello.fi/-/saisinko-puhua-suomea>
- Kokkonen, M. 2007. Vaatimuksena sujuva suomi. *Virittäjä*, 111 (2), 253–261. <https://journal.fi/virittaja/article/view/40570>
- Kuronen, M. & M. Kautonen 2018. Foneettisten piirteiden ja vieraan aksentin yhteydestä suomen kielessä. *Lähivõrdlusi. Lähivertailuja*, 28, 207–241. <http://dx.doi.org/10.5128/LV28.06>
- Leinonen, A. 2015. "Riittää kun saa selvää": vieraalla aksentilla tuotettu suomi nuorten arvioimana. *Jyväskylä studies in humanities* 275. Jyväskylä: Jyväskylän yliopisto.
- Lennes, M. 2009. Segmental features in spontaneous and read-aloud Finnish. Teoksessa V. de Silva & R. Ullakonoja (toim.) *Phonetics of Russian and Finnish. General description of phonetic systems: experimental studies on spontaneous and read-aloud speech*. Frankfurt: Peter Lang. 145–166.

- Lennes, M., D. Aalto & P. Palo 2009. Puheen perustaajuusjakaumat: alustavia tuloksia. Teoksessa M. O'Dell & T. Nieminen (toim.) *Fonetiikan päivät 2008*. Tampere Studies in Language, Translation and Culture. Series B 3. Tampere: Tampere University Press, 147–155.
- Lennes, M. & M. Toivola 2023. Pitch distributions in a very large corpus of spontaneous Finnish speech. *Proceedings of Interspeech 2023*. Dublin: International Speech Communication Association (ISCA), 4778–4782. <https://doi.org/10.21437/Interspeech.2023-1822>
- Meister, L. & E. Meister 2007. Perceptual assessment of Russian-accented Estonian. Teoksessa J. Trouvain & W. Barry. (toim.) *Proceedings of the 16th International congress of phonetic sciences, 6–10 August 2007, Saarbrücken Germany*. Saarbrücken: Universität des Saarlandes, 1717–1720. <http://www.icphs2007.de/conference/Papers/1262/1262.pdf>
- Mennen, I. & E. de Leeuw 2014. Beyond segments: prosody in SLA. *Studies in Second Language Acquisition*, 36 (2), 183–194. <https://www.jstor.org/stable/26328937>
- Metsämuuronen, J. 2005. *Tutkimuksen tekemisen perusteet ihmistieteissä*. Helsinki: International Methelp Ky.
- Moyer, A. 2013. *Foreign accent: the phenomenon of non-native speech*. Cambridge: Cambridge University Press.
- Munro, M. J. 1995. Nonsegmental factors in foreign accent. Ratings of filtered speech. *Studies in Second Language Acquisition*, 17 (1), 17–34. <https://doi.org/10.1017/S0272263100013735>
- Munro, M. J. & T. M. Derwing 1995. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38 (3), 289–306. <https://doi.org/10.1177/002383099503800305>
- Munro, M. J., T. M. Derwing & C. S. Burgess 2010. Detection of nonnative speaker status from content-masked speech. *Speech Communication*, 52 (7–8), 626–637. <https://doi.org/10.1016/j.specom.2010.02.013>
- Raimo, I. & S. Ojala 2009. Akustiikkaa ja artikulaatiota eli miltä puhe näyttää. Teoksessa O. Aaltonen, R. Aulanko, A. Iivonen, A. Klippi & M. Vainio (toim.) *Puhuva ihminen: puhetieteiden perusteet*. Helsinki: Otava, 174–182.
- Ruuska, K. 2020. *At the nexus of language, identity and ideology. Becoming and being a highly proficient second language speaker of Finnish*. Jyväskylä: Jyväskylän yliopisto. JYU dissertations 309. <http://urn.fi/URN:ISBN:978-951-39-8366-6>
- Saito, K., P. Trofimovich & T. Isaacs 2016. Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics* 37 (2), 217–240. <https://doi.org/10.1017/S0142716414000502>
- Silfverberg, L. 1982. Puolalaisten suomen opiskelijoiden ääntämisvaikeuksista. *Virittäjä*, 86 (1), 64–69. <https://journal.fi/virittaja/article/view/37667>
- de Silva, V. 1999. *Quantity and quality as universal and specific features of sound systems: Experimental phonetic research on interaction of Russian and Finnish sound systems*. *Studia philologica Jyväskyläensia* 48. Jyväskylä: Jyväskylän yliopisto.
- de Silva, V., A. Iivonen, L. V. Bondarko & L. C. W. Pols 2003. Common and language dependent phonetic differences between read and spontaneous speech in Russian, Finnish and Dutch. Teoksessa M. J. Solé, D. Recasens & J. Romero (toim.) *Proceedings of the 15th ICPhS*. Barcelona: Universitat Autònoma de Barcelona, 2977–2980. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_2977.pdf

- de Silva, V., K. Mäkilä, H. Kärkkäinen & R. Ullakonoja 2010. *Uusi venäjän ääntämisopas*. Helsinki: Finn Lectura.
- de Silva, V. & R. Ullakonoja 2009. Introduction: Russian and Finnish in contact. Teoksessa V. de Silva & R. Ullakonoja (toim.) *Phonetics of Russian and Finnish. General description of phonetic systems. Experimental studies on spontaneous and read-aloud speech*. Frankfurt: Peter Lang, 15–20.
- Teló, C., P. Trofimovich & M.c.G. O'Brien 2022. Disentangling professional competence and foreign accent. *Journal of Second Language Pronunciation*, 8 (3), 413–443. <https://doi.org/10.1075/jslp.22013.tel>
- Tergujeff, E. 2022. Ymmärrettävyytutkimuksen antia kielenopetukselle: esimerkkinä englanti vieraana kielenä. Teoksessa R. Kantelinen, M. Kautonen & Z. Elgundi (toim.) *Linguapeda 2021. Ainedidaktisia tutkimuksia 21*. Joensuu: Itä-Suomen yliopisto, Suomen ainedidaktinen tutkimusseura ry, 121–148. <http://hdl.handle.net/10138/352128>
- Toivola, M. 2011. *Vieraan aksentin arviointi ja mittaaminen suomessa*. Helsinki: Helsingin yliopisto.
- Toivola, M. & R. Ullakonoja 2017. Identification of Russian accented Finnish by native and non-native listeners with and without Finnish proficiency. Teoksessa M. Kuronen, P. Lintunen & T. Nieminen (toim.) *Näkökulmia toisen kielen puheeseen. Insights into second language speech*. AFInLA-e. Soveltavan kielitieteen tutkimuksia 10. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFInLA, 258–276. <https://doi.org/10.30660/afinla.73141>
- Tsunemoto, A., M. McAndrews, P. Trofimovich & E. Friginal 2022. Listener perceptions of customer service agents' performance. *Journal of Second Language Pronunciation*, 9 (2), 234–262. <https://doi.org/10.1075/jslp.21027.tsu>
- Ullakonoja, R. 2007. Comparison of pitch range in Finnish (L1) and Russian (L2). Teoksessa J. Trouvain & W. Barry. (toim.) *Proceedings of the 16th International congress of phonetic sciences, 6-10 August 2007, Saarbrücken, Germany*. Saarbrücken: Universität des Saarlandes, 1701–1704. <http://www.icphs2007.de/conference/Papers/1498/1498.pdf>
- Ullakonoja, R. & M. Kuronen 2015. Young Russian immigrants' segmental duration and length in Finnish. Teoksessa The Scottish Consortium for ICPhS 2015 (toim.), *Proceedings of the 18th ICPhS*. Glasgow: University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0347.pdf>
- Vainio, M., P. Palo, D. Aalto & U. K. Laine 2009. Lähde ja suodin – puheentuoton akustiikasta ja sen mallintamisesta. Teoksessa O. Aaltonen, R. Aulanko, A. Iivonen, A. Klippi & M. Vainio (toim.) *Puhuva ihminen: puhetieteiden perusteet*. Helsinki: Otava, 161–173.
- Volskaya, N. 2009. Aspects of Russian intonation. Teoksessa V. de Silva & R. Ullakonoja (toim.) *Phonetics of Russian and Finnish. General description of phonetic systems: experimental studies on spontaneous and read-aloud speech*. Frankfurt: Peter Lang, 37–46.
- Winke, P., S. Gass & C. Myford 2012. Raters' L2 background as a potential source of bias in rating oral performance. *Language testing*, 30 (2), 141–147. <https://doi.org/10.1177/0265532212456968>
- Yleisten kielitutkintojen perusteet 24.11.2011. https://finlex.fi/data/normit/39607-Oph_24_011_2011_su.pdf
- Ylinen, S. 2006. *Cortical representations for phonological quantity*. Helsinki: Helsingin yliopisto.

Toivola, M., P. Lintunen & L. M. Heikkola (toim.) 2024. *Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. New directions in speech research – Freely produced speech as data.* *AFinLA-teema / n:o 17*, 77–90.

**Katja Haapanen, Antti Saloranta,
Kimmo U. Peltola, Henna Tamminen**
University of Turku

Lannie Uwu-khaeb
University of Turku in Windhoek, Namibia

Maija S. Peltola
University of Turku

The production of voicing in Namibian English stops by speakers of Khoekhoegowab

Highlights

- Word-initial stop consonant voicing in Khoekhoegowab speakers' Namibian English is examined.
- The speech data consists of interviews conducted in Namibia.
- The voice onset times are significantly longer in voiceless stops than in voiced stops.
- Stop voicing contrast in Namibian English is found congruent with other varieties of English.

Abstract

This study examines whether Namibian English spoken by L1 Khoekhoegowab speakers has word-initial stop voicing contrasts, and how they might be realized in freely produced speech. The data consists of English interviews collected from nine speakers of Namibian English. For analysis, a total of 365 words beginning with a stop-vowel sequence were extracted from the interviews, and analyzed for voice onset time (VOT). To examine the realization of voicing, the extracted words are divided into voiced and voiceless categories based on their phonological voicing in other varieties of English and their VOTs are compared statistically. The VOTs of voiceless stops are significantly longer than those of voiced stops. The results suggest a short-lag vs. aspiration contrast in word-initial stops in Namibian English and provide new information about the phonetic features of Namibian English consonants.

Keywords: English, Namibian English, speech production, stop consonants, voicing

Asiasanat: englanti, namibianenglanti, puheen tuotto, klusiilit, sointi

1 Introduction

English has been the official language of Namibia since the country gained its independence in 1990 (Norro 2022a). Since then, research has indicated that a distinctly Namibian form of English may be emerging among World Englishes, separate from South African English varieties (Buschfeld & Kautzsch 2014; Schröder et al. 2021; Stell 2022b). Most Namibians are multilingual speakers of indigenous languages and typically learn English as a second or third language in school where it is used as the mode of instruction (Norro 2022b, 2022c), which means that it is likely that the phonetic characteristics of Namibian English are influenced by the phonological systems present in Bantu and Khoe languages. This suggestion is supported by previous research that has identified Bantu influences on spoken Black South African English (see e.g. Kamwangamalu 2020; Louw & de Wet 2007; Makalela 2013). According to the 2011 census, the most commonly spoken indigenous languages are Oshiwambo, Khoekhoegowab, Kavango, and Otjiherero (Namibia Statistics Agency 2013). In addition to these, 10 % of the population are native speakers of Afrikaans, which was the language of administration and education before the independence of Namibia (Norro 2022a). Today, Afrikaans is still used as a lingua franca between speakers of different L1s.

Previous studies have indeed identified certain phonetic characteristics in Namibian English that differentiate it from other English varieties, particularly South African English (Buschfeld & Kautzsch 2014; Schröder et al. 2021). It has been suggested, however, that the pronunciation of Namibian English varies among speakers and is influenced by ethnic factors, implying the existence of multiple localized sub-varieties of Namibian English (Schröder et al. 2021). On the other hand, some argue that a single, more standardized local variation of Namibian English with a

higher social status might be emerging among speakers of different local languages, bringing the ethnically diverse Namibian English variations together (Stell 2022a, 2022b). However, compared to other varieties of English, the descriptions on the phonetic characteristics of Namibian English are still relatively limited and literature often focuses on South African Englishes or English spoken by L1 Bantu speakers in general (e.g. Kamwangamalu 2020).

Khoekhoegowab is a language belonging to the Khoekhoe branch of the Khoe language family (formerly known as Khoesan/Khoisan). It is spoken by roughly 240 000 people in Namibia, or 12 % of the population, making it the second most common language after Oshiwambo (Haacke 2018). Khoekhoegowab has a fairly complex sound system, typically described as having five short and long vowels, three nasal vowels, 12 non-click consonants and 20 click consonant phonemes. In addition, it is a tonal language, with a tone system consisting of six contrasting tones (Cruttenden 1992). The existence of stop consonant voicing in Khoekhoegowab has been described in various and conflicting ways in the existing literature. Cruttenden (1992), for example, describes the language as having three stop consonants, /b/, /d/ and /g/, that have entirely predictable and therefore non-contrastive voicing. They describe the stops as being voiceless in word-initial positions with possible slight aspiration. Voiced plosives only occur word-medially, and /b/ may also occur voiceless in a word-final position, marking masculine nouns (Cruttenden 1992: 104). Fredericks and Banda (2018) report that this supposed lack of stop voicing contrasts in Khoekhoegowab is also maintained by other phonological descriptions of the language and more official language resources, such as the official Khoekhoegowab orthography (Curriculum Committee for Khoekhoegowab 2003) and a Khoekhoegowab-English dictionary (Haacke & Eiseb 2002). In all of these sources, it is claimed that the stop consonant pairs represented in writing with the seemingly contrasting letters p-b, t-d and k-g are in fact phonetically realized with identical sounds, and the use of the distinct letters is motivated by needing to distinguish between the tonal quality of the following vowel (Fredericks & Banda 2018).

Conversely, Fredericks and Banda (2018) suggest that there are, in fact, three contrasting stop voicing pairs in Khoekhoegowab: /p-b/, /t-d/ and /k-g/. Their findings are based on acoustic analysis of recorded sentences and individual words from four Khoekhoegowab speakers, and additionally listener groups of native Khoekhoegowab speakers who discussed the recordings with the researchers. The listener groups were asked, for example, if the meaning of the word would change if the initial stop consonant was changed to its voiced or voiceless counterpart while the tone was kept the same. Both the acoustic analyses of the recorded speech and the statements from the listener groups suggested that word-initial voiced and voiceless stop sounds in Khoekhoegowab are realized with different acoustic qualities, and that they cannot in all situations be switched without changing the mean-

ings of the words (Fredericks & Banda 2018). There is little information about the acoustic nature of the possible contrast in any of the examined literature.

In English, there are three contrasting stop consonant voicing pairs, /p-b/, /t-d/ and /k-g/ (Roach 2004). In a word-initial position, the difference between English voiced and voiceless stops is mainly based on VOT, as originally described by Lisker and Abramson (1967) and confirmed, for example, by Keating et al. (1983). Voiced stops in English exhibit either prevoicing, where voicing begins before the release burst of the consonant, or more typically short-lag VOT, where the delay between the release burst and voicing is very short, if not effectively non-existent. Voiceless stops are typically aspirated (e.g. Cho et al. 2019), meaning that the VOT is longer and turbulent air escapes through the vocal folds after the release burst before voicing starts. The mean VOT durations for English word-initial short-lag voiced stops are 0–20 ms and for aspirated voiceless stops 50–80 ms. Prevoicing, when it occurs, is in the range of 40–100 ms (e.g. Keating et al. 1983; Lisker & Abramson 1967; Suomi 1980).

Models of second language acquisition, such as the Revised Speech Learning Model (SLM-r) (Flege & Bohn 2021) and the Perceptual Assimilation Model of Second Language Speech Learning (PAM-L2) (Best & Tyler 2007), suggest that a second language phonological contrast of similar sounds that are considered to be allophones or other variants of a single native language phoneme category will initially both be perceived and produced like that native phoneme. This could mean that native speakers of Khoekhoegowab, assuming the lack of native stop voicing contrasts, would perceive the English voiced and voiceless stops as members of their own single stop category and then produce them as such when speaking English. With enough experience and exposure to the new contrast, a new phoneme category could form for the more unfamiliar sound in the second language contrast and the speakers would be able to use it contrastively (Best & Tyler 2007; Flege & Bohn 2021). It has been shown, for example, that neural discrimination of an unfamiliar second language fricative voicing contrast can be acquired in just days with focused training, though this is likely not reflected in the productions of the participants (Tamminen et al. 2015). If, however, a second language contrast closely matches an existing native contrast, the speakers will face little problems perceiving and producing it. This could be the case if Khoekhoegowab does have voiced and voiceless stops. (Best & Tyler 2007; Flege & Bohn 2021.)

Studies on Namibian English have typically focused on vowels (e.g. Schröder et al. 2021; Stell 2022a, 2022b), and research on the consonants of Namibian English is particularly scarce. The purpose of the current study was therefore to provide some of the first insights into the consonants of Namibian English by examining the production of word-initial stop consonants by native speakers of Khoekhoegowab. VOT was chosen as the examined feature because it is the main acoustic feature differentiating English voiced and voiceless stops, particularly in a word-initial position (Abramson &

Whalen 2017; Lisker & Abramson 1967). Based on the previously described ambiguity in the existence of stop voicing contrasts in Khoekhoegowab and its possible influence on the L2 English of its speakers, four research questions were formed:

- 1) Does the English spoken by native speakers of Khoekhoegowab have word-initial stop voicing contrasts?
- 2) If it does, how are the contrasts realized, based on VOT?
- 3) Do the word-initial stop voicing patterns follow the short-lag vs. aspiration contrast typical for other English varieties?
- 4) Are there Afrikaans influences in L1 Khoekhoegowab speakers' production of English voicing contrasts?

It was hypothesized that if there indeed is no stop voicing contrast in Khoekhoegowab, as maintained by some of the examined literature (Cruttenden 1992; Fredericks & Banda 2018), the speakers would face difficulties consistently producing it in English, as they previously would not have had to use voicing contrastively. Based on the literature, variation in VOT in Khoekhoegowab is dependent on the position of the segment in the word, i.e. it is allophonic (Cruttenden 1992), and according to models of second language acquisition, this poses the most difficult learning situation (Best & Tyler 2007; Flege & Bohn 2021). The speakers' likelihood of producing a voicing distinction might be largely dependent on their experience and exposure to a variety of English where the distinction does exist, and there could be a great deal of individual variation. Conversely, if a stop voicing contrast does exist in Khoekhoegowab, as suggested by Fredericks and Banda (2018), the speakers would likely also be able to employ it contrastively in English. The VOT values then produced, however, might not follow typical English patterns if the native contrast is acoustically realized in a different way. Proficiency in Afrikaans, typical to many Khoekhoegowab speakers (Haacke 2018), may also affect the production of voicing contrasts. Afrikaans is considered to have /p-b/ and /t-d/ contrasts, with voiced stops realized as prevoiced and voiceless stops as short-lag (Coetzee et al. 2018). While this pattern differs from the English short-lag vs. aspiration contrast, it could also be employed by the speakers in the current study for English, especially if stop consonant voicing is not a distinctive feature in Khoekhoegowab. As Afrikaans lacks the /k-g/ contrast, however, the influence of Afrikaans might result in a smaller VOT contrast between /k/ and /g/ for the speakers.

2 Materials and methods

2.1 Speakers

Nine students (aged 20–23 years, mean age 21.4 years, six females) from the University of Namibia participated in the study. Khoekhoegowab was the self-reported L1 of all participants. The reported L2s (languages acquired after L1) were English, Afrikaans, German, French, Otjiherero and Oshiwambo, of which Afrikaans and English were the only L2s spoken by all the participants. They had first learned English at the age of 3 to 7 and spoke it daily at the time of taking part in the study. The speakers were selected because they were highly proficient speakers of Namibian English and used it as a lingua franca with speakers of other Namibian languages in their everyday lives.

2.2 Interviews

The interviews where the speech data was collected were originally conducted for folkloristic research and linguistic archiving purposes and they all followed the same structure. All participants gave written informed consent before the interview. The consent covered the archiving of the interviews as well as their use for other research and teaching purposes. The participants were spoken to in English, and they were asked to answer the interviewer's questions first in Khoekhoegowab, and then to summarize their answers in English. The questions pertained both to the participants' everyday lives and their personal experiences and opinions about the historic and current presence of Finnish people in Namibia. The interviews were held at a lecture hall at the University of Namibia campus in Windhoek, and they were recorded using either an Olympus LS-P1 or a Zoom H2n audio recorder. The length of the participants' answers was dependent on both their talkativeness and their experience with the subject matter of the interview. Therefore, the total durations of the interviews ranged between 6 and 19 minutes.

2.3 Analysis

For analysis, words beginning with the target consonant-vowel (CV) sequences (with an initial stop consonant) were manually searched for from the interview recordings. Before further analysis, the words were separated into voiced and voiceless categories based on their phonological forms, i.e. whether the initial stop consonant would typically have been phonologically voiced or not in English. The words were then annotated in three tiers: the orthographic form of the whole word, the phonological form of the initial CV sequence based on British English Received Pronunciation (RP)

as described in the Cambridge Dictionary (2024), and the VOT of the initial stop. If the voicing started before the explosion (prevoicing) the VOT was marked as negative, and if voicing started after the explosion, it was marked as positive. If voicing started immediately at the time of the explosion, the VOT was marked as zero. Every instance of the target CV sequences could not be included in the analysis due to high speech rate or other factors in the recording that made it difficult to reliably assess the boundaries of the sounds. The number of analyzed stop sounds ranged between 21 and 89 between speakers. The total number of analyzed consonants was 365. The number of analyzed tokens per consonant are listed in Table 1.

TABLE 1. The number of analyzed stop consonants extracted from the English interviews.

Phoneme	/p/	/b/	/t/	/d/	/k/	/g/	Total
N	50	78	79	52	81	25	365

To statistically analyze possible VOT distinctions in word-initial stop consonant voicing, the extracted stops were divided into three voicing pairs according to English phonology (Roach, 2004): /p-b/, /t-d/ and /k-g/. The statistical analysis of the VOT data began with Shapiro-Wilk's tests of normality, which confirmed that the data were approximately normally distributed. The voiced and voiceless VOT values were then compared within each three voicing pairs with the paired samples' t-tests using the SPSS Statistics (version 27.0.1.0) software.

3 Results

Examination of the stop consonant VOT values extracted from the speakers' productions revealed that the voiceless stops had overall longer VOTs (37–56 ms) than the voiced stops (10–22 ms). The average VOT values from the whole group for each stop consonant are displayed in Figure 1.

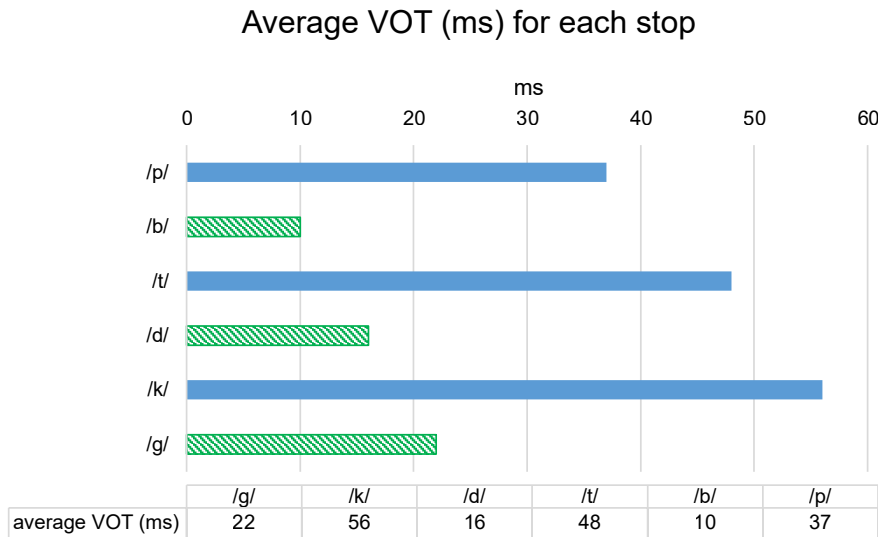


FIGURE 1. The group average of average VOTs calculated for each stop from the productions of the whole group (N=9). The bars represent the average VOT value, not the range.

The longest VOTs, on average, were produced in words beginning with /k/ (56 ms), whereas words beginning with /b/ were produced with the shortest VOT (10 ms). There appeared to be some variation in the average VOT ratios between the three consonant voicing pairs, with the VOT difference being the largest for the velar pair /k-g/ (44 ms), the smallest for the bilabial pair /p-b/ (27 ms), and intermediate for the dental-alveolar pair /t-d/ (32 ms). Examination of the individual average VOT values from each speaker revealed that the VOTs were produced fairly consistently within the group (Figure 2). Overall, there was more VOT variation between speakers in the voiceless stops /p, t, k/ than in the voiced stops /b, d, g/. There were two outliers in the speakers' average /d/ VOT values, as seen in Figures 2 and 3, but no other discrepancies emerged.

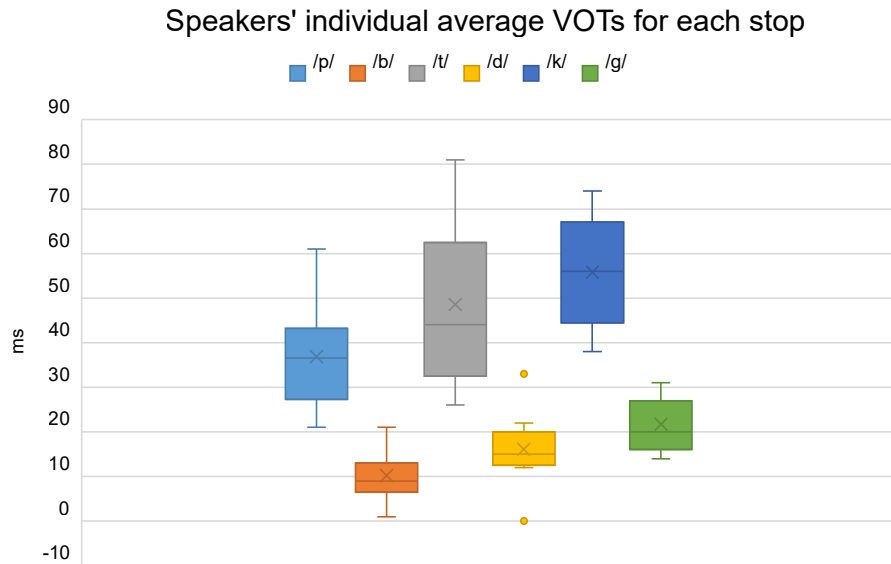


FIGURE 2. The average VOT values from each speaker for the six stops.

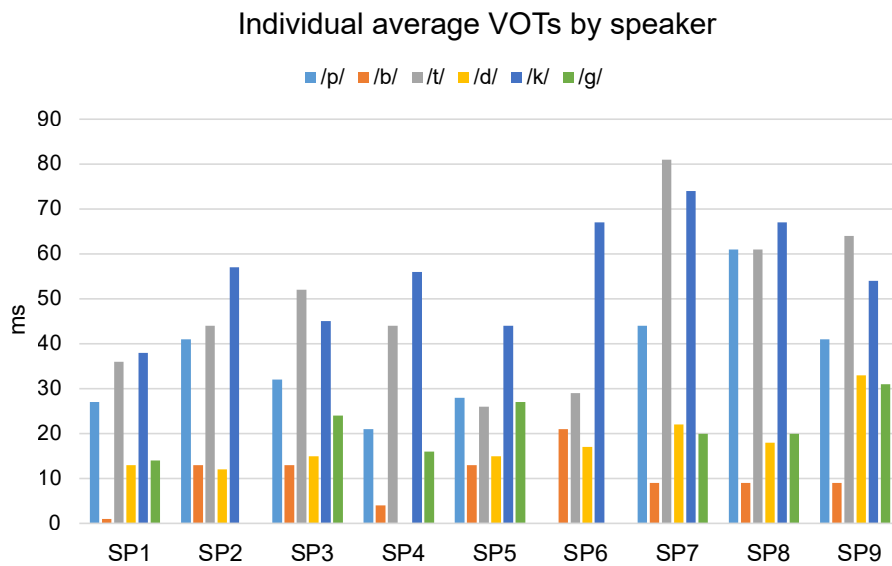


FIGURE 3. The average VOT values from each individual speaker.

The results of the paired samples' t-tests on subject averages revealed that the VOT values extracted from /p/ and /b/ differed significantly from each other ($t(8) = 6.558$, $p < .001$; Cohen's $d = 2.318$), and the same was true for the pairs /t-d/ ($t(8) = 6.155$, $p < .001$; Cohen's $d = 2.052$) and /k-g/ ($t(6) = 5.914$, $p = .001$; Cohen's $d = 2.235$). In other words, the VOTs were significantly longer in the voiceless than in the voiced stops in all the tested consonant voicing pairs.

4 Discussion

This study examined stop consonant production in Namibian English speech by native speakers of Khoekhoegowab. The research questions were:

- 1) Does the English spoken by native speakers of Khoekhoegowab have word-initial stop voicing contrasts?
- 2) If it does, how are the contrasts realized, based on VOT?
- 3) Do the word-initial stop voicing patterns follow the short-lag vs. aspiration contrast typical for other English varieties?
- 4) Are there Afrikaans influences in L1 Khoekhoegowab speakers' production of English voicing contrasts?

The hypothesis was that the speakers' production of word-initial consonant voicing might show influences of Khoekhoegowab and/or Afrikaans voicing production patterns, which differ from those of most English varieties. Possible Khoekhoegowab influences were expected to be reflected on the VOT values as a lack of clear voicing contrasts or as VOT values that are not typical for other English varieties. The second hypothesis was that possible Afrikaans influences on the speakers' production of Namibian English stop consonants would be reflected on the VOT values as a prevoiced vs. short-lag contrast and/or the lack of the /k-g/ voicing contrast.

The results show that the L1 Khoekhoegowab speakers produced word-initial stop voicing contrasts in their Namibian English speech, which was reflected in the consonant VOTs. The voiceless stop consonants /p, t, k/ were produced clearly aspirated with longer VOTs, whereas the voiced stops /b, d, g/ were produced with short-lag, i.e. very short or almost non-existent VOTs. The average VOTs for aspirated stops were slightly shorter than the VOT durations reported in some previous literature (e.g. Keating et al. 1983; Lisker & Abramson 1967), which could be at least partly explained by the high speech rate of the speakers and the free nature of the speech data, though some findings contradict the idea of an articulation rate influence (e.g. Nakai & Scobbie 2016). Though few individual words were produced with slight prevoicing, which was well under the typical 40–100 ms range reported for prevoiced stops (Keating et al. 1983; Lisker & Abramson 1967), or zero VOT, the majority of the

analyzed words with voiced initial consonants were produced with short-lag. These findings indicate that the primary cue for maintaining the stop voicing contrast in Khoekhoegowab speakers' English appears to be the aspiration of the voiceless stop rather than strong (pre)voicing of the voiced stop.

The short-lag vs. aspiration contrast found in the speakers' VOT data suggests that the speakers' production of stop voicing was similar to the voicing contrasts typically found in other English varieties (Keating et al. 1983; Lisker & Abramson 1967; Roach 2004) and was not greatly affected by the differing voicing patterns found in Khoekhoegowab or Afrikaans (Coetzee et al. 2018; Fredericks & Banda 2018). If the speakers had produced the English stops according to their L1 phonology, we would have expected there to be more variation in the stop VOTs. On the other hand, if their production had been influenced by their knowledge of Afrikaans, we would have expected the stop VOTs to show a prevoiced vs. short-lag contrast rather than a short-lag vs. aspiration contrast. In addition, the large VOT contrast found in the velar stop pair /k-g/ does not support the hypothesis that the speakers' production of English stops might have Afrikaans influences, as there is no /k-g/ voicing contrast in Afrikaans (Coetzee et al. 2018). The longer average VOT values found in the velar stop consonants /k/ and /g/ are also in keeping with previous literature on consonant voicing production in English, stating that the VOT tends to lengthen as the place of articulation moves back (Lisker & Abramson 1967). This has been discussed in detail by Cho and Ladefoged (1999) as well as by Suomi (1980), who describe how the place of articulation, length of occlusion and length of VOT are connected in English stops. As the place of constriction moves back in the mouth, the volume of the articulatory tract between the constriction and vocal folds decreases. This means that the pressure required to cause the release of the constriction can build up faster the further back the place of constriction is, resulting in shorter occlusion and a longer VOT. The exact glottal mechanisms underlying the connection between occlusion and VOT are physiologically different for voiced and voiceless English stops, but the general principle holds for both (Suomi 1980).

The study focused on L1 Khoekhoegowab speakers' production of English stop consonants but the results also offer valuable information on Namibian English phonetics in a broader perspective. There is still relatively little research on the phonetic features of Namibian English, and there is some debate on whether there is a single standardized English variety or multiple L1 dependent varieties spoken in the country (Buschfeld & Kautzsch 2014; Schröder et al. 2021; Stell 2022a, 2022b). Either way, since Khoekhoegowab is the second most common indigenous language spoken in Namibia, a notable number of Namibian English speakers are L1 Khoekhoegowab speakers. Therefore, the results of this study reveal new insights into the production of consonant voicing in the emerging Namibian English variety/varieties. Our results suggest that although Khoe and Bantu language influences and variety specific features have been found in Namibian English vowels (Buschfeld & Kautzsch

2014; Schröder et al. 2021) as well as South African English (e.g. Kamwangamalu 2020; Louw & de Wet 2007; Makalela 2013), the voicing contrasts of word-initial stop consonants in Namibian English seem to follow the common production patterns of English. These results can be used in the future to develop phonetic English training materials for L1 Khoekhoegowab speakers in Namibian schools.

More research into stop consonant voicing production in word-medial and word-final positions as well as fricative voicing production in general needs to be conducted in the future in order to gain an even better understanding of the phonetic features of Namibian English consonants. Moreover, additional data on word-initial stop consonant voicing from more L1 Khoekhoegowab speakers could be collected to verify our findings with this speaker group. Data from different L1 speakers of Namibian English could also be gathered to see if there are L1 related differences in the realization of word-initial stop consonant voicing contrasts. Due to the small number of speakers in this study, and due to their homogenous linguistic backgrounds, we need more phonetic research on Namibian English in the future to make more definite conclusions. The lack of clear Khoekhoegowab or Afrikaans influences in the current data suggests that the speakers' other languages did not affect the production of English stop consonant voicing. In other words, these results represent L1 independent stop consonant VOT production patterns in Namibian English.

5 Conclusions

The results of this study showed that L1 Khoekhoegowab speakers of Namibian English produce word-initial stop consonant voicing contrasts /p-b/, /t-d/ and /k-g/ in their speech, with voiced stops produced with significantly shorter VOTs than voiceless stops. The short-lag vs. aspiration voicing contrast found in the data is congruent with the voicing contrast production patterns found in other English varieties. The results can be used in the future to design phonetic English teaching materials and to support English learning in Namibian schools.

Acknowledgements

The authors wish to thank the Kone Foundation for funding this work, professor Erkki Sutinen and Helvi Haikokolla for their valuable help in Namibia, and all the speakers who participated in this study.

Literature

- Abramson, A. S. & D. H. Whalen 2017. Voice Onset Time (VOT) at 50: theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63, 75–86. <https://doi.org/10.1016/j.wocn.2017.05.002>
- Best, C. T. & M. D. Tyler 2007. Non-native and second-language speech perception: commonalities and complementarities. In M. J. Munro & O.-S. Bohn (eds), *Second language speech learning: the role of language experience in speech perception and production*. Amsterdam: John Benjamins Publishing Company, 13–34. <https://doi.org/10.1075/llt.17.07bes>
- Buschfeld, S. & A. Kautzsch 2014. English in Namibia: a first approach. *English World-Wide: A Journal of Varieties of English*, 35 (2), 121–160. <https://doi.org/10.1075/eww.35.2.01bus>
- Cambridge Dictionary*. Cambridge: Cambridge University Press & Assessment 2024. <https://dictionary.cambridge.org/> [accessed 17.1.2024]
- Cho, T. & L. Ladefoged 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27 (2), 207–229. <https://doi.org/10.1006/jpho.1999.0094>
- Cho, T., D. H. Whalen & G. Docherty 2019. Voice onset time and beyond: exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, 72, 52–65. <https://doi.org/10.1016/j.wocn.2018.11.002>
- Coetzee, A. W., P. S. Beddor, K. Shedden, W. Styler & D. Wissing 2018. Plosive voicing in Afrikaans: differential cue weighting and tonogenesis. *Journal of Phonetics*, 66, 185–216. <https://doi.org/10.1016/j.wocn.2017.09.009>
- Cruttenden, A. 1992. Clicks and syllables in the phonology of Dama. *Lingua*, 86 (2), 101–117. [https://doi.org/10.1016/0024-3841\(92\)90031-D](https://doi.org/10.1016/0024-3841(92)90031-D)
- Curriculum Committee for Khoekhoegowab 2003. *Khoekhoegowab: 3lli xoaigaub = orthography 3*. Windhoek: Gamsberg Macmillan.
- Flege, J. E. & O.-S. Bohn 2021. The revised speech learning model (SLM-r). In R. Wayland (ed.), *Second language speech learning: theoretical and empirical progress*. Cambridge: Cambridge University Press, 3–83. <https://doi.org/10.1017/9781108886901.002>
- Fredericks, N. & F. Banda 2018. Voicing in non-click consonants and orthographic design in Khoekhoegowab. *South African Journal of African Languages*, 38 (1), 1–11. <https://doi.org/10.1080/02572117.2018.1429852>
- Haacke, W. H. G. 2018. Khoekhoegowab (Nama/Damara). In T. Kamusella & F. Ndhlovu (eds) *The social and political history of Southern Africa's languages*. London: Palgrave Macmillan, 133–158. https://doi.org/10.1057/978-1-137-01593-8_9
- Haacke, W. H. G. & E. Eiseb 2002. *A Khoekhoegowab dictionary with and English-Khoekhoegowab index*. Windhoek: Gamsberg Macmillan. <https://doi.org/10.4314/lex.v13i1.51402>
- Kamwangamalu, N. 2020. Southern African Englishes. In C. Nelson, D. Davis & Z. Proshina (eds), *The handbook of world Englishes* (2nd ed.). Hoboken: Wiley-Blackwell, 159–172. <https://doi.org/10.1002/9781119147282>
- Keating, P., W. Linker & M. Huffman 1983. Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, 11 (3), 277–290. [https://doi.org/10.1016/S0095-4470\(19\)30827-7](https://doi.org/10.1016/S0095-4470(19)30827-7)
- Lisker, L. & A. S. Abramson 1967. Some effects of context on voice onset time in English stops. *Language and Speech*, 10 (1), 1–28. <https://doi.org/10.1177/002383096701000101>
- Louw, P. & F. de Wet 2007. The perception and identification of accent in spoken Black South African English. *Southern African Linguistics and Applied Language Studies*, 25 (1), 91–105. <https://doi.org/10.2989/16073610709486448>

- Makalela, L. 2013. Black South African English on the radio. *World Englishes*, 32 (1), 93–107. <https://doi.org/10.1111/weng.12007>
- Nakai, S. & J. M. Scobbie 2016. The VOT category boundary in word-initial stops: counter-evidence against rate normalization in English spontaneous speech. *Laboratory Phonology*, 7 (1), 131. <https://doi.org/10.5334/labphon.49>
- Namibia Statistics Agency 2013. *Namibia 2011: population & housing census main report*. Windhoek: Namibia Statistics Agency.
- Norro, S. 2022a. Factors affecting language policy choices in the multilingual context of Namibia: English as the official language and medium of instruction. *Apples - Journal of Applied Language Studies*, 16 (1), 1–20. <https://doi.org/10.47862/apples.107212>
- Norro, S. 2022b. Language practices in Namibian primary schools. *Journal of Multilingual and Multicultural Development*, 0 (0), 1–19. <https://doi.org/10.1080/01434632.2022.2121403>
- Norro, S. 2022c. Namibian teachers' practices in a multilingual context. *International Journal of Multilingualism*, 21 (1), 360–378. <https://doi.org/10.1080/14790718.2022.2065280>
- Roach, P. 2004. British English: received pronunciation. *Journal of the International Phonetic Association*, 34 (2), 239–245. <https://doi.org/10.1017/S0025100304001768>
- Schröder, A., F. Zähres & A. Kautzsch 2021. The phonetics of Namibian English. In A. Schröder (ed.) *The dynamics of English in Namibia: perspectives on an emerging variety*. Amsterdam: John Benjamins Publishing Company, 111–134. <https://doi.org/10.1075/veaw.g65.06sch>
- Stell, G. 2022a. Contact and innovation in new Englishes: ethnic neutrality in Namibian face and goat. *Journal of English Linguistics*, 50 (2), 169–200. <https://doi.org/10.1177/00754242221090522>
- Stell, G. 2022b. 'De-ethnicization' in new Englishes: perception and recognition of ethnicity in Namibian English. *Lingua*, 274. <https://doi.org/10.1016/j.lingua.2022.103355>
- Suomi, K. 1980. *Voicing in English and Finnish stops: a typological comparison with an interlanguage study of the two languages in contact*. Publications of the Department of Finnish and General Linguistics of the University of Turku. Turku: University of Turku.
- Tamminen, H., Peltola, M. S., Kujala, T., & Näätänen, R. 2015. Phonetic training and non-native speech perception—New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology*, 97 (1), 23–29. <https://doi.org/10.1016/j.ijpsycho.2015.04.020>

Toivola, M., P. Lintunen & L. M. Heikkola (toim.) 2024. Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. *New directions in speech research – Freely produced speech as data.* AFinLA-teema / n:o 17, 91–116.

Anna von Zansen
Helsingin yliopisto

Heini Kallio
Tampereen yliopisto

DigiTala – Moodle-sovellus suullisen kielitaidon automaattiseen arviointiin

Nostot

- DigiTala kehitti puheentunnistukseen ja koneoppimiseen perustuvan työkalun suomen ja ruotsin oppijoille.
- Työkalu antaa automaattista palautetta myös spontaaneista puhenäytteistä.
- Automaattinen palaute antaa tietoa puhujan kehittämiskohteista sekä konkretisoi taitotasoa.
- Artikkelissa esittelemme työkalun kehittämisvaiheet ja ominaisuudet.

Abstract

This article presents an online tool developed by the DigiTala research project in 2022. The tool is the first one to provide comprehensive automated assessment of spontaneous L2 Finnish and Swedish speech. The Moodle-based tool records and analyzes L2 learners' speech automatically and provides holistic as well as analytic feedback to learners. In addition to self-regulated learning purposes, the tool can be used to support human assessment in high-stakes tests and language courses. We discuss the benefits and weaknesses of the tool and automated speaking assessment in general from the perspectives of low-resourced languages as well as developers and users.

Keywords: automated assessment, spontaneous speech, L2 skills, agile software development

Asiasanat: automaattinen arviointi, spontaani puhe, suullinen kielitaito, ketterä ohjelmistokehitys

1 Johdanto

Suullista kielitaitoa tarvitaan yhteiskunnassa ja yhteisöissä toimimiseen. Kielenoppijat tarvitsevat paljon tilaisuuksia käyttää kieltä, jotta he saavat valmiuksia toimia kohdekielellä arkisissa kielenkäyttötilanteissa. Lisäksi oppijat saattavat tarvita harjoitusta ääntämiseen, sujuvuuteen tai ilmausten laajuuteen liittyen voidakseen itsevarmasti käyttää kieltä esimerkiksi työelämässä. Erilaisissa elämäntilanteissa kielenoppijoilla ei kuitenkaan aina ole tarvittavaa mahdollisuutta, motivaatiota tai uskallusta hakeutua aitoihin vuorovaikutustilanteisiin harjoittelemaan kohdekieltä. Lisäksi vaikka oppija hakeutuisi kielikurssille, oppituntien ja saatavilla olevan ohjauksen määrä ei välttämättä riitä suullisen kielitaidon kehittämiseen toivotussa määrin.

Puheteknologia mahdollistaa ajasta ja paikasta riippumattoman suullisen kielitaidon harjoittelun. Vuorovaikutustaitojen kehittymisen kannalta etenkin spontaanin puheen¹ harjoittelu on tärkeää. Automaattisten puheen arviointimenetelmien kehittäjät ovat kuitenkin suosineet lukupuhetta (*read-aloud*) sen ennustettavuuden kannalta (Evanini & Zechner 2020). Puheentunnistimet suoriutuvat paremmin luetun tai imitoidun kuin spontaanin puheen tunnistamisesta, ja saman sisältöiset puhenäytteet mahdollistavat esimerkiksi tarkempien ääntämismallien kehittämisen (Bernstein ym. 1990). 2000-luvulla tutkijoiden huomio laajeni spontaanin puheen automaattiseen arviointiin. Tällöin tutkimus alkoi puheen sujuvuuden mittaamisesta (Cucchiari ym. 2002) ja laajeni lopulta useiden kielitaitoa eri näkökulmista mittaavien piirteiden tarkasteluun (Xi ym. 2008). Spontaanin puheen automaattista arviointia on kehitetty kuitenkin lähinnä suurille kielille, kuten englannille, jossa on tarjolla paljon puheaineistoa niin äidinkielisiltä kuin kielenoppijoiltakin.

1 Käytämme termiä *spontaani puhe* kuvaamaan ns. vapaata puhetta, jossa puhuja tuottaa sisältöä itse ääneen lukemisen tai perässä toistamisen sijaan.

DigiTala-tutkimushankkeen (Kautonen & von Zansen 2020) Moodle-sovellus (von Zansen ym. 2022a) on ensimmäinen kotimaisten kielten automaattiseen arviointiin kehitetty työkalu, joka kattaa useita kielitaidon osa-alueita ja arvioi myös spontaaneja puhenäytteitä. Automaattiseen puheen tunnistukseen (*automatic speech recognition*), arviointiin (*automated assessment*) ja palautteeseen (*automated feedback*) perustuvan työkalun kohderyhmänä ovat suomen ja ruotsin oppijat, jotka haluavat harjoitella puheen tuottamista ja ääntämistä osana kieltenopetusta tai itsenäisesti. Suomi ja suomenruotsi ovat maailmanlaajuisesti tarkasteltuna pieniä kieliä, joissa sekä äidinkielsiä puhujia että kielten oppijoita on suhteellisen vähän. Puheteknologian näkökulmasta nämä kielet ovat vähäresurssisia: soveltuva aineistoa kuten kielenoppijoiden spontaania puhetta on rajallisesti saatavilla. Teknologisia sovelluksia kehittämällä edistämme näiden pienten kielten opiskelua.

Puhumisen harjoittelun lisäksi tutkimushankkeen kehittämää puheen automaattista arviointia voitaisiin hyödyntää yksilön tulevaisuuden kannalta tärkeissä (*high-stakes*) kielikokeissa, kuten yleisten kielitutkintojen ja ylioppilastutkinnon kielikokeissa. Yleiset kielitutkinnot toimivat portinvartijana (*gate-keeper*) esimerkiksi Suomen kansalaisuutta hakiessa. Ylioppilastutkinnon arvosanoja puolestaan hyödynnetään muun muassa jatko-opintoihin hakeutuessa. Työkalua voidaan käyttää tärkeisiin kokeisiin valmistautumisen lisäksi ihmisarviointin tukena. Yleisiin kielitutkintoihin kuuluu puhumisen osakoe, mutta ylioppilastutkinnon kielikokeista se edelleen puuttuu ratkaisemattomien käytännön kysymysten takia (Vaarala ym. 2021).

Tämän artikkelin tavoitteena on luoda katsaus puheen automaattista arviointia koskevaan aiempaan tutkimukseen, dokumentoida työkalun kehittämisvaiheet sekä tarjota läpinäkyvä ja ymmärrettävä kuvaus DigiTalan kehittämästä arviointityökalusta. Automaattisen arviointityökalun toimintaperiaatteiden kuvaaminen on välttämätöntä, jotta sen toimivuutta voidaan arvioida niin pedagogisesti kuin teknisestikin. Artikkelissa esittelemme DigiTalan Moodle-sovelluksen tutkimusperustaisen taustan, sovelluksen kehitysvaiheet sekä itse sovelluksen toiminnot. Pohdimme myös kielenoppijan puheen automaattisen arvioinnin vahvuuksia ja heikkouksia vähäresurssisten kielten sekä kehittäjien ja käyttäjien näkökulmista.

2 Tausta

Seuraavissa alaluvuissa esittelemme puheen automaattisen arviointiin liittyviä lähtökohtia (luku 2.1) sekä aiempaa tutkimusta (luku 2.2).

2.1 Puheen automaattisen arvioinnin lähtökohtia

Suullisen kielitaidon automaattisen arvioinnin tärkein komponentti on tarkka kohdekielinen puheentunnistin. Tunnistin muuntaa puheen tekstiksi, josta voidaan laskea automaattisesti erilaisia ääntämisen laatuun, kielioppiin ja sanastoon liittyviä piirteitä (Hsieh ym. 2020; Yoon ym. 2020). Lisäksi suoraan puhesignaalista voidaan mitata puheen sujuvuutta laskemalla mm. puhenopeutta ja taukojen ja epäröintien määrää tai kestoa (Hsieh ym. 2020). Kielenoppijan puheen tunnistaminen on kuitenkin hankalampaa kuin äidinkielen puhujan, koska kielenoppijan puhe on usein epäsujuvaa ja siinä voi esiintyä erilaisia kielioppi-, sanasto- ja ääntämisvirheitä. Puheentunnistinten akustiset mallit pohjautuvat äidinkielisten puhujien näytteisiin, mutta automaattista arviointia varten tunnistin täytyy mukauttaa kielenoppijan puheeseen opettamalla sitä eritasoisten oppijoiden puheella (Ylinen & Kurimo 2017).

Automaattisten arviointimallien pohjana käytetään ihmisten tekemiä arvioita eritasoisten kielenoppijoiden puheesta. Arviointialgoritmit vertaavat ihmisten arviointien puhenäytteiden laskennallisia piirteitä testattavien puhenäytteiden piirteisiin. (Loukina & Yoon 2020.) Jokaiselle arvioitavalle suullisen kielitaidon ulottuvuudelle kehitetään oma arviointimallinsa. Esimerkiksi ETS:n (English Testing Services) kehittämässä automaattisessa arviointijärjestelmässä on holistisen taitotason lisäksi kolme arvioitavaa ulottuvuutta: ulosanti (ääntäminen ja puheen sujuvuus), kielenkäyttö (kieliopin tarkkuus ja sanaston laajuus) sekä aiheenmuodostus (sisällön kattavuus ja koherenssi) (Loukina & Yoon 2020).

DigiTalan työkalussa arvioitavat puheen ulottuvuudet ovat taitotason lisäksi tehtävänannon täyttyminen, sujuvuus, ääntäminen ja ilmauksen laajuus (ks. von Zansen ym. 2022c). Tutkimukseen (ks. luku 2.2) ja koneoppimiseen pohjautuva arviointimalli huomioi sen, mitä tehtävän vastaukselta odotetaan sekä määrittää, mitä puheen piirteitä käytetään kyseisen ulottuvuuden arvioimiseen ja miten kutakin piirrettä painotetaan arvioinnissa.

Useimmiten arviointimallin kehittämiseen on käytetty ohjattua oppimista: puheaineisto, josta algoritmi oppii, on ihmisten esikäsittelmää (Evanini & Zechner 2020). Aineisto voi olla nimikoitua tai siitä on muodostettu erilaisia funktioita, joiden avulla arviointimalli päättää, miten uusia puhenäytteitä tulisi arvioida. Opetusaineistoon voidaan esimerkiksi luokitella sana- ja lausepainoja, joiden toteutumista arviointimalli vertaa arvioitavan puhenäytteen ja opetusaineiston kesken (Hsieh ym. 2020). Puheen piirteiden laskentaan käytetään kuitenkin enenevässä määrin ohjaamatonta oppimista, jolloin algoritmi oppii aineistosta itsenäisesti ilman ihmisen väliintuloa (Al-Ghezi ym. 2023).

Oikeudenmukaisuuden ja luotettavuuden kannalta automaattisen arvioinnin läpinäkyvyys (*transparency*) on tärkeää (Evanini & Zechner 2020; von Zansen & Heijala 2023). Ohjattuun oppimiseen perustuvien arviointimallien toimintaperiaatteet tiedetään, jolloin työkalun käyttäjälle voidaan tarjota tietoa siitä, mihin automaattinen

arvio perustuu. Tämä esimerkiksi mahdollistaa analyyttisen automaattisen palautteen näyttämisen oppijalle (ks. von Zansen & Heijala 2023; von Zansen & Huhta 2022).

Puheen automaattiseen arviointiin liittyy kuitenkin rajoitteita. Automaattiset arviointityökalut mahdollistavat nykyään erityyppisiä puhumisen tehtäviä, mutta tehtävät ovat yleensä rajattuja, koska työkalut on säädetty tiettyä tarkoitusta varten (Ylinen & Kurimo 2017). Rajoitetuimmat tehtävät sisältävän ääneen lukua tai lyhyitä toistoja mallin perässä, ja näistä suorituksista voidaan arvioida vain ääntämistä ja sujuvuutta (von Zansen ym. 2022c). Avoimemmista tuottamistehtävistä, kuten kuvasta kertomisesta tai annetusta aiheesta puhumisesta, voidaan arvioida ääntämisen ja sujuvuuden lisäksi myös kieliopin tarkkuutta ja ilmaisun laajuutta. Spontaanit puhetehtävät ovat lähempänä autenttisia kielenkäyttötilanteita kuin ääneen lukeminen. Kaikkia suullisen kielitaidon ulottuvuuksia ei kuitenkaan päästä vielä arvioimaan automaattisesti: nykyisistä työkaluista puuttuvat vuorovaikutteiset tehtävät (kielenoppijoiden suullisen vuorovaikutuksen automaattiseen arviointiin tähtäävästä tutkimushankkeesta ks. von Zansen 2023).

2.2 Aiempi tutkimus puheen automaattisesta arvioinnista

Luomme tässä luvussa katsauksen aiempaan sekä DigiTala-hankkeessa tehtyyn monitieteiseen tutkimukseen. Esittelemme ensin kielididaktista ja foneettista tutkimusta, jonka jälkeen esittelemme hankkeen kieliteknologista tutkimusta.

2.2.1 Kielididaktinen tutkimus

Puheen automaattiseen arviointiin liittyvä kielididaktinen tutkimus ulottuu ensinnäkin mitattavaan käsitteeseen (*construct*), tehtävänlaadintaan (*task design*) sekä tavoiteltavaan puheasuoritukseen liittyviin tarkasteluihin (Luoma 2004; Fulcher 2014). Toiseksi automaattisten arviointimallien kehitys perustuu yleensä ihmisten tekemiin arviointeihin (Evanini & Zechner 2020), joten arviointien laadun ja skaalojen toimivuuden (ks. tarkemmin luku 3.4) tutkiminen on keskeistä. Lisäksi käytettävien arviointikriteerien tulee olla linjassa puheesta mitattavien piirteiden (*speech feature*, Evanini & Zechner 2020, ks. tarkemmin foneettista tutkimusta koskeva alaluku 2.2.2) kanssa. Kolmas keskeinen aihe kielididaktisessa tutkimuksessa liittyy automaattisen arvioinnin käyttöön. Arviointityökalujen käytettävyyttä ja hyödyllisyyttä tutkitaan esimerkiksi kartoittamalla kielenoppijoiden, opettajien ja kielitaidon arvioijien näkemyksiä. Neljäntenä kielididaktisen tutkimuksen aiheena nostamme esiin puheen harjoittelua ja kielen oppimista tukevan automaattisen palautteen (Gu & Davis 2020).

Hankkeen kielididaktinen tutkimus on käsitellyt muun muassa ihmisarviointien laatua ja arviointiin osallistuvien (*stakeholders*) käsityksiä puheen automaattisesta arvioinnista. Osallistujien suhtautuminen puhetehtävien tai arvioinnin suorittamiseen tietokoneella on ollut positiivista, vaikka se on ollut heille pääosin uutta. Laitimamme

arviointikriteerit ovat osoittautuneet toimiviksi, ja kouluttamamme ihmisarvioijat ovat suorituneet arviointitehtävästä hyvin (von Zansen & Huhta 2022). Niin kouluttamiemme ihmisarvioijien (von Zansen ym. 2022c), lukiolaisten (von Zansen ym. 2022b) ja aikuisopiskelijoiden (von Zansen & Hilden 2022) kuin opettajienkin (von Zansen & Heijala 2023) tutkimuksen aineistonkeruun yhteydessä ilmaisemat näkemykset ovat tuoneet esiin alan kirjallisuudessakin (ks. esim. Evanini & Zechner 2020) esiintyviä mahdollisuuksia ja uhkia automaattiseen arviointiin liittyen.

Automaattisen arvioinnin kiistaton hyöty on mahdollisuus harjoitella puhumista ajasta ja paikasta riippumatta. Lisäksi kone arvioi kaikki puhujat väsymättä ja yhtenevillä kriteereillä. Automaattiseen arviointiin liittyvät huolet puolestaan liittyvät arvioinnin oikeudenmukaisuuteen: tunnistaako kone puheen oppijan erilaisesta aksentista tai puheeseen liittyvästä vaikeudesta huolimatta? (von Zansen ym. 2022b.) Tiettyjen ulottuvuuksien, kuten sujuvuuden ja taitotason, arvioiminen liian lyhyistä puhenäytteistä on puolestaan haastavaa niin koneelle kuin ihmisarvioijille (von Zansen ym. 2022c). Moodle on toiminut hyvin suoritusympäristönä kyseisille kohderyhmille, vaikkakin sen käyttö edellyttää jonkin verran teknisiä valmiuksia (ks. von Zansen & Hilden 2022).

Haastattelemamme opettajat käyttäisivät Moodle-työkalua esimerkiksi kotitehtävissä, suullisissa kokeissa ja osana lukiossa laadittavaa kieliprofilia. Lisäksi muotoilemamme automaattinen palaute on heidän arvionsa mukaan oppijoille ymmärrettävää ja hyödyllistä, vaikkakin opettajan apua saatetaan tarvita palautteen välittämisessä näille. Arviointikriteerien pohjalta laaditut palauteväittämät eivät kuitenkaan sisällä kaikkia puhumisen ulottuvuuksia – opettajat lisäisivät esimerkiksi kehonkielen, sisällön ja rakenteiden virheettömyyden. (von Zansen & Heijala 2023.)

Työkalun yksi merkittävä hyöty liittyy taitotasoihin, jotka pohjautuvat Eurooppalaiseen viitekehukseen ja ovat toistakymmentä vuotta kuuluneet suomalaisiin kielten opetussuunnitelmiin (Opetushallitus 2003, 2019). Opetussuunnitelmista huolimatta taitotasoa ei ole kattavasti hyödynnetty kielten opetuksessa (ks. Härmälä & Marjanen 2023: 122), joten työkalun antama holistinen palaute konkretisoisi taitotasoa niin opiskelijoille kuin opettajille. Konkretisoimalla puheesta mitattavia piirteitä työkalu voisi yhdenmukaistaa arviointia. (von Zansen & Heijala 2023.)

2.2.2 Foneettinen tutkimus

Puheen automaattiseen arviointiin liittyvä foneettinen tutkimus on usein taustatutkimusta, jossa verrataan puheesta mitattavien piirteiden ominaisuuksia ihmisten tekemiin arvioihin kielitaidosta. Tutkimuksista saatava tieto on tärkeää piirrephajaisten arviointimallien kehittämisessä. Lisäksi tutkimusten avulla voidaan päätellä, mitä piirteitä niin sanottu *black box* -malli (ks. esim. Zechner 2020) voisi käyttää puhenäytteiden arviointiin, kun halutaan tarjota kielenoppijalle sanallista palautetta numeerisen arvion lisäksi. Ihmisarvioiden ja puheen piirteiden yhteyttä on tutkittu suhteellisen paljon; etenkin englannin oppijoiden puhetta on tutkittu niin äänteiden

tuottamisen (esim. Munro & Derwing 1995; Isaacs & Trofimovich 2012) kuin puheen sujuvuudenkin näkökulmista (esim. Derwing ym. 2009; Tavakoli 2011; Kahng 2018; Kallio ym. 2022c; Peltonen & Lintunen 2022).

Sujuvuustutkimuksissa on todettu, että alkeisoppijoiden puhe on usein hitaampaa ja taukoja esiintyy useammin kuin edistyneempien kielenoppijoiden tai äidinkielisten puheessa. Ääntämisen arvioinnin tutkimuksissa korostuu virheettömyyden sijasta ymmärrettävyys, ja yksittäisten äännetason virheiden sijaan tutkijat ovat havainneet prosodisten piirteiden, kuten sana- ja lausepainotuksen, tauotuksen ja nopeuden, vaikuttavan voimakkaammin puheen ymmärrettävyyteen (Anderson-Hsieh ym. 1992; Isaacs & Trofimovich 2012; Kang 2012).

Hankkeessa kielenoppijoiden puheen tutkimus on keskittynyt etenkin spontaaneihin puhunnoksiin. Ruotsinoppijoiden monologipuhetta tutkineet Kautonen ja Kuronen (2021) havaitsivat, että sanojen ääntäminen kehittyy lineaarisesti suhteessa taitotasoon, mutta lausetason prosodiassa esiintyy enemmän virheitä B2-tasolla kuin A1–B1-tasojen puhujilla. Tämä on todennäköisesti yhteydessä muun kielitaidon kehittymiseen: B2–C1-tasoilla puhujat tuottavat enemmän puhetta ja monimutkaisempia rakenteita, jolloin lauseprosodiaa on vaikeampi hallita. Tutkimustuloksia voidaan automaattisessa arvioinnissa hyödyntää siinä, miten puheen eri piirteitä painotetaan milläkin taitotasolla: kun tiedetään, että tietyntaiset prosodiset virheet yleistyvät muun kielitaidon kasvaessa, suhteutetaan näiden piirteiden esiintyminen esimerkiksi vastauksen syntaktiseen syvyyteen ja sanavaraston laajuuteen.

Sana- ja lausepainojen vaihtelu muodostaa puheelle sen ominaisen rytmin, mutta painotuksia tuotetaan eri kielissä eri tavoin. Puherytmejä on vertailtu eri kielissä sekä äidinkielisten ja kielenoppijoiden välillä erilaisilla tavu- tai äännekeston perustuvilla parametreilla (White & Mattys 2007). Ruotsin sanapainoa tuotetaan ensisijaisesti tavukestojen avulla, ja painon tuottotavassa on enemmän haasteita ja suurempaa vaihtelua alemman taitotason kielenoppijoilla kuin edistyneemmällä kielenoppijoilla (Kallio ym. 2020, 2021). Sanapainoon liittyvistä piirteistä DigiTalan suomenruotsiin keskittyvä arviointityökalu mittaakin tavukestojen suhteita. Painotusten objektiivinen mittaaminen osana automaattista arviointia voi kuitenkin olla vaikeaa, jos kielenoppija yrittää puhua esimerkiksi riikinruotsia tai norjaa, joiden painotustavat eroavat suomenruotsista, tai jos puheen intonaatio on muuten poikkeavaa (Kallio ym. 2021).

Puhenopeudella, epäröintien määrällä ja tavukestojen hajonnalla puolestaan on tilastollinen yhteys ruotsinoppijoiden spontaanin puheen sujuvuuden arviointiin (Kallio ym. 2023). Tulokset ovat linjassa kansainvälisten tutkimusten kanssa, joissa samankaltaisten piirteiden on todettu vaikuttavan puheen sujuvuuteen (Derwing ym. 2009; Tavakoli 2011; Kahng 2018). Suomenkielisten, A2-tason ruotsinoppijoiden ja suomenruotsalaisten puheen sujuvuutta verranneet Peltonen ja Lintunen (2022) havaitsivat myös selkeän eron L1- ja L2-ryhmien välillä niin puhenopeudessa kuin useissa epäsujuvuutta kuvaavissa piirteissä. Puhenopeutta ja erilaisia epäsujuvuuksia mitataan myös DigiTalan arviointityökalussa.

Suomenoppijoillakin puhenopeus on merkittävä sujuvuuden mittari, mutta vielä paremmin ihmisten antamia arvioita ennustaa artikulaationopeus yhdistettynä erityyppisiin taukoihin, kuten puhunnosten välisten katkosten kestoon, hiljaisten taukojen esiintymistiheyteen ja epäröintien kestoon (Kallio ym. 2022b; Koivusalo 2022). Artikulaationopeus kertoo yksittäisten sanojen tuottonopeuden ilman taukoja, puhenopeus taas koko puhunnoksen tuottonopeuden taukoineen. Tulokset taukojen esiintymistiheydestä tukevat aiempaa tutkimusta, jossa suomenoppijoiden todettiin pitävän ääneen lukiessaan lyhyitä taukoja huomattavasti useammin kuin äidinkielisten puhujien (Toivola ym. 2009). Näitä piirteitä käytetään myös DigiTalan arviointityökalussa (Al-Ghezi ym. 2023). Lisäksi Kallio ym. (2022b) havaitsivat, että etenkin C1–C2-tasolle arvioiduilla suomenoppijoilla esiintyi puheessa narinaa, mutta sellainen oli yleisesti niin harvinaista, ettei sen perusteella voi luotettavasti erottaa eritasoisia puhujia toisistaan.

Kallio ym. (2022a) tutkivat taukojen sijainnin vaikutusta arvioituun kielitaitotasoon ja spontaanin puheen sujuvuuteen suomenoppijoilla. Tulosten perusteella kieliopillisten lausekkeiden sisälle jäävillä tauoilla ja epäröinneillä on merkittävä vaikutus sekä suomen taitotason että sujuvuuden arviointiin. Tulokset ovat linjassa aiempien tutkimusten kanssa (Kahng 2018; Tavakoli 2011). Taukojen sijaintiin perustuvilla parametreilla voitaisiin myös parantaa automaattisen arvioinnin tarkkuutta (Kallio & Kuronen 2023). DigiTalassa etsittiin tapoja integroida taukojen sijaintiparametreja arviointityökaluun; yksi tällainen menetelmä voisi perustua puheen automaattisen syntaktisen jäsennyksen ja akustisten taukojen mittaamisen yhdistämiseen.

2.2.3 Kieliteknologinen tutkimus

Puheen automaattisen arvioinnin teknologinen tutkimus alkoi 1990-luvulla ja keskittyi ääntämisen laadun (Bernstein ym. 1990), puheen sujuvuuden (Cucchiari ym. 1997) ja intonaation arviointiin (Eskenazi 1996). Näiden tutkimusten kohde oli vielä rajattu kielenoppijan lukemaan tai perässä toistamaan puheeseen, jota algoritmi vertasi puhujan samansisältöiseen tuotokseen. Teknologian nopea kehitys ja laskentatehon kasvu on mahdollistanut yhä monimutkaisempien algoritmien ja sitä myötä tarkempien puheentunnistimien ja monipuolisempien arviointimallien kehittämisen (ks. esim. Loukina & Yoon 2020).

Hankkeen puhe- ja kieliteknologinen tutkimus on keskittynyt automaattisen puheentunnistimen (Al-Ghezi ym. 2021; Getman 2021a) sekä automaattisen arvioinnin ja palautteen kehittämiseen (Getman 2021b; Al-Ghezi ym. 2023). Al-Ghezi ym. (2021) käyttivät Meta AI -yhtiön julkaisemaa, avoimen lähdekoodin wav2vec-algoritmia ruotsinoppijoiden puheentunnistimen kehittämiseen. Menetelmä on itseohjattu: kone oppii tunnistamaan puhetta suoraan suuresta, monikielisestä aineistosta ilman ihmisen väliintuloa. Itseohjatun alustavan opetuksen jälkeen tunnistin optimoitiin ohjatusti pienellä, ruotsinoppijoiden puheesta koostuvalla aineistolla. Getman (2021a) tutki

maisterintutkielmassaan tällaisten itseohjattujen tunnistimien soveltuvuutta suomen-ruotsin, suomen ja saksan oppijoiden puheentunnistukseen ja havaitsi niiden olevan tehokkaita työkaluja vähäresurssiseen toisen kielen oppijoiden puheentunnistukseen. DigiTalassa kehitetyt puheentunnistimet ovat tietääksemme ensimmäiset suomen ja suomenruotsin oppijoiden puheeseen adaptoidut tunnistimet.

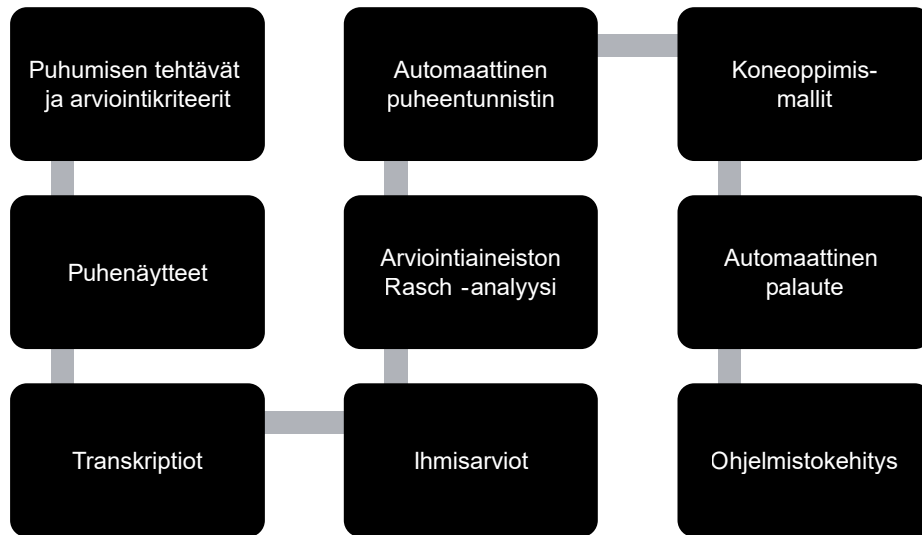
Getman on myös kehittänyt työkalua, joka antaa automaattista palautetta ruotsin kielen sanastosta ja kieliopista (Getman 2021b). Palautetyökalu paikantaa tekstistä virheitä sisältäviä sanoja tai sanajonoja ja antaa käyttäjälle korjausehdotuksia. Työkalua voi käyttää myös suullisen kielitaidon arviointiin, kun puhe on ensin muutettu tekstiksi automaattisen puheentunnistimen avulla.

Voskoboinik ym. (2023) puolestaan tutkivat eri menetelmiä automaattiseen tehtävän suorittamisen arviointiin. Testatuista menetelmistä alustavasti opetettu suuri kielimalli suoriutui sekä suomen- että ruotsinoppijoiden vastausten arvioinnista tarkemmin kuin sanojen yleisyyteen perustuva vektorimalli. Molemmat menetelmät perustuvat tuotetun ja odotetun sisällön samankaltaisuuden mittaamiselle.

Tuorein tutkimuksemme (Al-Ghezi ym. 2023) raportoi ensimmäisen, useita suullisen kielitaidon osa-alueita kattavan automaattisen arviointityökalun eri moduulien kehitysprosessit, sekä testaa ja pohtii niiden toimivuutta. Pääsääntöisesti koneen antama arvio poikkesi vähemmän ihmisen antamasta arviosta (*human-machine agreement*) kuin ihmisten arvioinnit toisistaan (*human-human agreement*), mutta sekä mallien että puheentunnistinten tarkkuus vaihteli jonkin verran. Sekä suomen- että ruotsinkielisen puheentunnistuksen tarkkuuteen vaikuttivat todennäköisesti mm. ihmisen tekemän transkription tarkkuus, puhenäytteen tekninen laatu ja puhujan puhenopeus.

3 Työkalun kehittäminen

Tässä luvussa esittelemme DigiTalan Moodle-sovelluksen (von Zansen ym. 2022a) kehittämisvaiheet. Kyseessä on ensimmäinen kotimaisten kielten automattiseen arviointiin kehitetty työkalu, joka kattaa useita kielitaidon osa-alueita. Koska vastaavaa ei ole tehty aiemmin, puuttui myös sopiva tietokanta työkalun kehittämistä varten. Projektissa on suunniteltu ja kerätty laaja tutkimus- ja kehitysaineisto, joka sisältää suomen ja ruotsin oppijoille suunnattuja suullisen kielitaidon tehtäviä (von Zansen 2022e, 2022f, 2022g), arviointikriteerit (von Zansen 2022d), puhenäytteet litteraatteineen, asiantuntija-arvioinnit (ks. von Zansen ym. 2022c), käyttäjätutkimusten kyselylomakkeet (von Zansen 2022a, 2022b, 2022c), puheentunnistimet, koneoppimismallit sekä tietovaraston Moodle-sovelluksen lataamiseen, käyttöön ja kehittämiseen (von Zansen ym. 2022a; Alanen ym. 2022). Esittelemme kuviossa 1 työkalun kehittämisvaiheet.



KUVIO 1. Työkalun kehittämisvaiheet.

Seuraavat alaluvut etenevät kuviota 1 mukailevassa järjestyksessä. Kuvaamme ensin puhumisen tehtävät (3.1), arviointikriteerit (3.2), puheaineistot (3.3) sekä ihmisarviointien keräämisen ja analyysin (3.4). Tämän jälkeen esittelemme puheentunnistimen (3.5) sekä koneoppimismallit (3.6). Lopuksi käsittelemme automaattista palautetta (3.7) sekä käyttöliittymän (3.8) ja ohjelmiston (3.9) kehitystä.

3.1 Puhumisen tehtävät

Aineistonkeruun suunnittelu käynnistyi mitattavan puhumisen käsitteen määrittelymisellä etenkin Glenn Fulcherin luokittelun mukaisesti (ks. Fulcher 2014 puhumisen käsitteen määrittelystä ja puhetehtävien ominaisuuksista). Kohderyhmän ja kielenkäyttötilanteiden lisäksi otimme huomioon puheen automaattiseen arviointiin ja tietokoneella suoritettavaan puhumisen kokeeseen liittyvät rajoitukset. Käytännössä puhumisen tehtävät suunniteltiin siten, että ne tuottaisivat mahdollisimman monipuolisia puhenäytteitä kielitaidon kattavaa arviointia varten, mutta olisivat puheentunnistimelle riittävän helppoja ja laskentatehon kannalta sopivan mittaisia. Lisäksi tehtävien suunnittelussa huomioitiin niiden pedagoginen mielekkyys muun muassa luomalla useita tehtäväkokonaisuuksia eritasoisille kielenoppijoille.

B1/B2-tasojen puhetehtävät laadittiin vastaamaan lukion opetussuunnitelman (Opetushallitus 2019) sisältöjä ja A1/A2-tasojen puhetehtävät yliopiston alkeiskurssien sisältöjä (ks. von Zansen & Hilden 2022; von Zansen ym. 2022b). Tehtäviä testattiin

opettajien ja kielenoppijoiden kanssa, ja niihin tehtiin parannuksia saadun palautteen perusteella esimerkiksi tehtävänantoja muokkaamalla.

Tehtävät sisältävät ääneen lukua, lyhyitä reagoiteja kuvan, tekstin tai äänitteen pohjalta sekä pidempiä tuottamistehtäviä, esimerkiksi kuvasta tai annetusta aiheesta kertomista (esim. kerro päivästäsi tai sinulle tärkeästä paikasta). Näistä spontaaniksi puheeksi lasketaan lyhyet reagoinnit sekä pidemmät tuottamistehtävät. Suomen oppijoiden puheen keräämisessä käytetyt tehtävät on julkaistu verkossa (von Zansen 2022e, 2022f, 2022g) lukuun ottamatta Yleisten kielitutkintojen puhekokeista saatua aineistoa. Ruotsin A1/A2-tasojen puhetehtävät on julkaistu suomen tehtävien tapaan verkossa (von Zansen & Tarvainen-Li 2024). Ruotsin B1-tason suulliset tehtävät puolestaan suunniteltiin aiemmassa projektissa (Karhila ym. 2016), eikä niitä ole julkaistu.

3.2 Arviointikriteerit

Tehtävänlaadinnan rinnalla tutkimushankkeessa laadittiin holistiset ja analyttiset arviointikriteerit (von Zansen 2022d), jotka suunniteltiin vastaamaan viimeaikaisten lukion opetussuunnitelmien kielitaitokäsitystä (puhumisen arviointikriteereistä tarkemmin Luoma 2004 ja Fulcher 2014). Kriteerien laadinnassa käytettiin pohjana ensisijaisesti vuoden 2003 opetussuunnitelman kielitaidon tasojen kuvausasteikkoja (Opetushallitus 2003), koska ne sopivat nykyisen opetussuunnitelman asteikkoa paremmin hankkeen analyttisiin ja teknisiin tavoitteisiin. Aiemmat kuvausasteikot eivät ole ristiriidassa nykyisen asteikon kanssa: Opetushallitus kannustaa hyödyntämään niitä esimerkiksi arvioinnissa. Tämä on mahdollista, koska molemmat asteikot ovat paikallisia sovellutuksia Eurooppalaisesta viitekehystä. Automaattisesta arvioinnista ks. tarkemmin alaluku 3.6.

Holistinen kriteeristö koostuu kuvauksista taitotasoille asteikolla alle A1–C2. Analyttiset kriteerit pitivät sisällään kuvaukset viidelle analyttiselle suullisen kielen taidon osa-alueelle (asteikolla 0–3/4): tehtävän suorittaminen, sujuvuus, ääntäminen, ilmaisun laajuus sekä sanaston ja kieliopin tarkkuus. Ääneenlukutehtävissä arvioitiin vain puheen sujuvuutta ja ääntämistä. Tuottamistehtävissä arvioijat määrittivät puhenäytteille taitotason ja antoivat sille arvion analyttisistä osa-alueista. Analyttiset osa-alueet arvioitiin itsenäisinä, taitotasosta riippumattomina ominaisuuksina. Kriteerejä testattiin projektitiimin kielitaidon arvioinnin asiantuntijoiden kanssa ennen niiden käyttöönottoa, ja ne havaittiin toimiviksi.

3.3 Puheaineistot

DigiTala-hanke on kerännyt kielenoppijoiden puhetta yllä kuvatuilla tehtäväkokonaisuuksilla sekä lukioista että korkeakouluista. Osallistajat suorittivat puhetehtävät Moodle-tenttinä joko luokkahuoneympäristössä tai kotonaan osana etäopetusta. Osa

osallistujista käytti ulkoista kuulokemikrofonia puhevastaustensa äänittämiseen, osa käytti kannettavan tietokoneen sisäistä mikrofonia ja osa älypuhelimien tai tabletin mikrofonia. Vähintään yksi hankkeen tutkija ohjeisti osallistujat puhetehtävien tekoon ja auttoi teknisissä ongelmissa. Puhenäytteiden keruu on dokumentoitu tarkemmin osatutkimuksissa (Al-Ghezi ym. 2023; von Zansen ym. 2022b; von Zansen & Hilden 2022). B1-tason ruotsin näytteet kerättiin aiemmassa projektissa (Karhila 2016).

Artikkelia kirjoittaessa projektissa on kerätty ja litteroitu (ks. myös luku 3.5) yhteensä 4442 puhenäytettä (noin 19 h puhetta) 325 suomenoppijalta ja 8834 näytettä (noin 12,5 h puhetta) 301 ruotsinoppijalta. Näiden lisäksi saimme puheaineistoa Yleisten Kielitutkintojen (YKI) suomen (356 litteroitua näytettä, noin 8,5 h) ja ruotsin (12 litteroitua näytettä, noin 0,3 h) puhekokeista.

3.4 Ihmisarviointien kerääminen ja analyysi

Varsinaisten puhenäytteiden ja transkriptioiden lisäksi automaattisen arvioinnin kehitys on edellyttänyt arviointien keräämistä koulutetuilta ihmisarvioijilta. Arvioijien kouluttamisella pyrimme lisäämään arvioijien ymmärrystä käytettävistä arviointikriteereistä sekä vähentämään arviointivirheitä. Arvioijat saattavat esimerkiksi tulkita annettuja kriteerejä eri tavoin (lisää arviointiin vaikuttavista tekijöistä Fulcher 2014, laajempi katsaus puhumisen arviointitutkimukseen Fan & Yan 2020).

Tutkimushanke on vuosina 2020–2023 järjestänyt yhteensä viisi arviointikierrosta, joihin on osallistunut 39 arvioijaa (ks. arviointiprosessista tarkemmin von Zansen ym. 2022c). Arvioijiksi rekrytoitiin suomen ja ruotsin kielten asiantuntijoita, joilla oli aiempaa kielitaidon arviointikokemusta yleisistä kielitutkinnoista tai ylioppilastutkinnosta. Lisäksi arvioijia rekrytoitiin aiemman tutkimushankkeen ruotsin arviointeihin osallistuneista lukio-opettajista sekä hankkeen tutkijoiden henkilökohtaisten verkostojen kautta. Ulkopuolisten arvioijien lisäksi yhteensä viisi tutkijaa projektitiimistä osallistui arviointikierruksiin.

Arvioijat koulutettiin tehtävään Zoomissa, ja projektin ulkopuolisille arvioijille maksettiin kultakin arviointikierrokselta yhtä työpäivää vastaava korvaus. Arvioinnit ja kyselyvastaukset kerättiin Moodlessa (versio 3.8.3). Koulutuksessa käytiin läpi arviointiin liittyvät ohjeet ja kriteerit (von Zansen 2022d) sekä esiteltiin taitotasoja konkretisoivat maamerkinäytteet.

Ihmisarvioijien välillä on tunnetusti tiukkuuseroja, joiden tasaamiseksi ordinaaliasteikollinen arviointiaineisto muutettiin lineaarisiksi ja alkuperäisiä pisteitä paikansapitävämmiksi arvoiksi Rasch-analyysin avulla (*Many-facet Rasch measurement*, ks. Boone ym. 2014, hankkeen aiemmista tutkimuksista von Zansen & Huhta 2022). Lisäksi puheaineistosta karsittiin tässä vaiheessa ongelmallisiksi havaitut näytteet, jotka olivat saaneet jostakin kriteeristä arvosanan 0. Näitä kullekin puhenäytteelle lasket-

tuja raakapisteitä reilumpia arvoja (*fair averages*) hyödynnettiin arviointimoduulien opettamisessa, mihin palaamme puheentunnistimen esittelyn jälkeen.

3.5 Puheentunnistin

Automaattisen puheen arvioinnin tärkein osa on tarkka puheentunnistin, joka muuntaa kielenoppijan puheen tekstiksi. DigiTalan suomen ja ruotsin puheentunnistimet ovat niin sanottuja itseohjattuja (*self-supervised*) päästä-päähän-tunnistimia (*end-to-end*, ks. esim. Baevski ym. 2020; Al-Ghezi ym. 2021; Getman 2021a). Nämä puheentunnistimet opetettiin alustavasti suurella, litteroimattomalla monikielisellä aineistolla, joka sisälsi puhetta 23 eri kielestä. Itseohjattu opetus tarkoittaa, että kone oppii suoraan puheaineistosta ilman ihmisen väliintuloa (ks. luku 2.2.3).

Ruotsin puheentunnistin opetettiin lisäksi ruotsinkielisellä, litteroimattomalla puheaineistolla. Suomen kielestä vastaavaa laajaa aineistoa ei ollut saatavilla. Alustavan opetuksen jälkeen tunnistin hienosäädettiin DigiTala-projektissa kerätyllä, litteroidulla eritasoisten suomen- ja ruotsinoppijoiden puheella (ks. luku 3.3). Litteraatteihin merkittiin muun muassa ääntövirheet ja epäröinnit helpottamaan puheentunnistimen työtä. Näin kehitetyt puheentunnistimet ovat tarkempia kuin pelkästään kohdekielisellä puheaineistolla opetetut, kun kohdekielistä aineistoa on suhteellisen vähän saatavilla. (Baevski ym. 2020; Al-Ghezi ym. 2021; Getman 2021a.)

3.6 Koneoppimismallit

Koneoppimisen menetelmillä pyritään ennustamaan ihmisten antamia arviointeja. Kutsumme näitä koneoppimismalleja arviointimalleiksi. Ihmisarvioijilta saatu arviointiaineisto kategorisoitiin *fair average* -arvoja pyöristämällä, ja jos johonkin arvosanakategoriaan jäi liian vähän näytteitä, sitä ei sisällytetty arviointimoduulien opettamiseen. Holistinen taitotason arviointimalli opetettiin ohjatusti syviä neuroverkkoja hyödyntäen. Kyseessä on päästä-päähän -malli, jossa ihmisten antamaa taitotasoarvioita ennustetaan suoraan puhesignaalista saatavien akustisten ja kielellisten piirteiden pohjalta (Chen ym. 2018; Al-Ghezi ym. 2023). Samoja mitattavia piirteitä käytetään valikoiden myös analyttisissä arviointimalleissa. Arviointimallit painottavat eri piirteitä eri tavoin riippuen paitsi arvioitavasta kielitaidon alueesta, myös kielestä ja opetusaineistosta.

Tehtävän suorittamisen arviointiin käytettiin itseohjattua koneoppimista, joka laskee semanttisen etäisyyden puhetuotoksen ja tehtävänannon välillä (ks. esim. Wang ym. 2020; Voskoboinik ym. 2023). Tämän arviointimallin tehtävänä on toimia portinvartijana, jos kokelas pelkästään lukee tehtävänannon ääneen tai ei vastaa tehtävänantoon.

Analyttiset arviointimallit ovat piirrepohjaisia (*feature-based*), eli ne mittaavat puheesta useita kymmeniä parametreja, joihin malli perustaa arvionsa (Evanini &

Zechner 2020; Al-Ghezi ym. 2023). Sujuvuuden arviointimalli perustuu akustisesta signaalista mitattuihin parametreihin, kuten puhenoiteen (montako tavua puhuja keskimäärin tuottaa tietyssä ajassa) ja hiljaisten ja täytettyjen taukojen määrään. Ääntämisen arviointimalli laskee esimerkiksi arvon, joka kuvaa puheentunnistimen tunnistusvarmuutta eli kuinka todennäköisesti sen tunnistama sana tai äänne vastaa kohdekielen akustista mallia. Lisäksi ääntämisen arviointimalli laskee puheintonaatioon ja -rytmiin liittyviä parametreja, kuten perustaajuuden hajontaa ja tavukestojen suhteita.

Ilmaisun laajuuden sekä sanaston ja kieliopin tarkkuuden arviointimallit puolestaan perustuvat puheentunnistimen tuottamasta tekstistä laskettuihin piirteisiin. Ilmaisun laajuutta kuvaavia parametreja ovat muun muassa näytteen pituus sanoissa, uniikkien sanojen osuus puhunnoksessa, vastauksen syntaktinen syvyys eli kuinka monimutkaisia rakenteita puhuja on keskimäärin käyttänyt ja verbien suhteellinen osuus (ks. esim. Yoon ym. 2020; Al-Ghezi ym. 2023). Koska kaikki edellä mainitut tekstipohjaiset parametrit kuvaavat ilmaisun laajuutta enemmän kuin kieliopillista tarkkuutta (ks. esim. Yoon ym. 2020), sanaston ja kieliopin tarkkuus jätettiin pois hankkeen lopullisesta arviointisovelluksesta.

3.7 Automaattinen palaute

Oppijoille annettava automaattinen palaute on teknis-tieteellisten STEM-alojen (*science, technology, engineering and mathematics*) lisäksi yleistynyt myös kielten opetuksessa (Deeva ym. 2021). DigiTalan Moodle-sovelluksessa oppijalle näytettävä automaattinen palaute muotoiltiin työkalun ensimmäisessä versiossa sekä ihmisarvioijien käyttämien arviointikriteerien (von Zansen 2022d, ks. myös luku 3.2) että automaattisten arviointimoduulien hyödyntämien puheen piirteiden pohjalta. Palautteessa pyrimme valitsemaan puheen ulottuvuuksia, jotka ovat koneelle laskettavissa ja ihmisille ymmärrettävissä kuten ääntäminen, sujuvuus, tehtävänantoon vastaaminen ja laajuus (ks. myös von Zansen & Heijala 2023; von Zansen & Huhta 2022). Myöhemmässä vaiheessa automaattisen arvion lisäksi oppijalle voitaisiin antaa oppimista edistäviä neuvoja, kuten Gu ja Davis (2020) kuvaavat.

3.8 Käyttöliittymä

Käyttöliittymän ohjelmoinnista vastasi neljä Helsingin yliopiston tietojenkäsittelytieteen opiskelijaa. Heidän laatimansa käyttöohjeet kuvaavat yksityiskohtaisemmin toiminnallisuuksia ja sisältävät lisää kuvia työkalusta eri käyttäjäryhmien näkökulmista (Alanen ym. 2022), joten kuvaamme käyttöliittymän seuraavaksi pääpiirteittäin. DigiTalan verkkotyökalu (von Zansen ym. 2022a) on julkaistu vapaiden ohjelmistojen lisenssillä, mikä mahdollistaa myös käyttöliittymän jatkokehityksen. Ohjelmistokehitystä kuvaamme tarkemmin luvussa 3.9.

Käyttöliittymän avulla oppija nauhoittaa puhenäytteitä, lähettää puhenäytteet arvioitavaksi sekä vastaanottaa automaattisen arvion. Opettaja puolestaan laatii puhetehtävät sekä voi halutessaan täydentää ja kommentoida koneen oppijalle antamaa arviota. Lisäksi opettaja saa ladattua oppijoidensa puhe-suoritus- ja niihin liittyvät arviot.

Työkalua käytetään verkkoselaimessa sijaitsevassa Moodlessa, johon kirjaututtuaan oppija näkee opettajan valmistelemat tehtävät, jotka voivat olla ääneenlukua tai puheen tuottamistehtäviä, kuten lyhyitä reagoiteja tai pidempiä kertomistehtäviä. Opettaja voi helposti kopioida ja muokata laatimiaan tehtäviä määrittelemällä kohdekielen (suomi/ruotsi), tehtävänannon, aikarajan, äänityskertojen määrän sekä tehtävän sisältämän aineiston. Ääneenlukutehtävien aineistona on luettava teksti tai lause tekstinä, tuottamistehtävät puolestaan voivat sisältää myös kuvia tai äänitteitä.

Testattuaan mikrofoninsa oppija äänittää vastauksensa ja lähettää sen arvioitavaksi Aalto-yliopiston palvelimille, joissa puheentunnistin ja arviointimallit prosessoivat vastauksen. Sovellus hakee palvelimelta kullekin puhenäytteelle erilliset automaattiset arviot, joten oppija voi nauhoittaa useita puhenäytteitä ja palata tarkastelemaan automaattista palautetta (kuva 1) sen valmistuttua.

1 Aloitus > 2 Tehtävä > 3 Arviointiraportti

Arviointiraportti
Lähetetty: 10.05.2022 11.21:13

Tämä palaute koskee alnoastaan nauhoittamaasi puhenäytettä, eikä se kuvaa kaikkea suullista kielitaitoasi. Automaattinen arvio on koneen tekemä. Koneita on opetettu muiden kielen oppijoiden puheella ja muulla kieliaineistolla.

Tässä tehtävässä ei ole yrityskertojen rajaa.

0:00 / 0:09

Puhenäytteesi tekstinä
öö mult oli jäi huppari teillä sinne kahvilaan eilen

Analyytinen arvio Taitasoarvio

Sujuvuus
Tämä mittari kertoo puhenäytteesi nopeudesta, taukojen määrästä ja empimisestä.

0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0

4/4

Automaattisen arvion mukaan vaikuttaa siltä, että puheesi on todella sujuvaa ja vaivatonta, ei häiritseviä taukoja, katkoksia tai empimistä.

Ääntäminen
Näet yllä, että kone muunsi puheesi tekstiksi. Voit tarkistaa tekstistä, lausuitko kaikki sanat oikein. Tämä mittari kertoo, kuinka hyvin ja varmasti kone tunnistaa puheesi. Tunnistamistarkkuuteen vaikuttavat puhenäytteet, joita kone on aiemmin opetusvaiheessa saanut.

0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0

3/4

Automaattisen arvion mukaan vaikuttaa siltä, että kone ymmärtää puhettasi, ääntämisessäsi ei vaikuta olevan suurempia ongelmia.

Tehtävänantoon vastaaminen
Tämä mittari perustuu vastauksiin, joilla kone on opetettu arvioimaan tätä tehtävää.

0 0.5 1.0 1.5 2.0 2.5 3.0

0/3

Automaattisen arvion mukaan vaikuttaa siltä, että valitettavasti kone ei ole kuullut tämänkaltaista suoritusta aiemmin, eikä siksi osaa arvioida puhettasi.

Laajuus
Tämä mittari kertoo, kuinka paljon olet puhunut sekä käyttämiesi sanojen ja lauseiden monipuolisuudesta.

0 0.5 1.0 1.5 2.0 2.5 3.0

2/3

Automaattisen arvion mukaan vaikuttaa siltä, että käytät tavallisia sanoja ja osaat tehdä niistä lauseita.

[Yritä uudestaan](#)

KUVA 1. Oppijan saama analyytinen palaute tuottamistehtävästä.

Palautesivulla oppija voi kuunnella nauhoittamansa äänitteen sekä nähdä puheensa muutettuna tekstiksi. Tuottamistehtävissä automaattinen arvio koostuu eri välilehdillä näytettävästä analyttisestä (kuva 1) ja holistisesta palautteesta (kuva 2), kun taas ää-
neenlukutehtävissä automaattinen palaute sisältää vain sujuvuuden ja ääntämisen arviot.

Arviointiraportti

Lähetetty: 10.05.2022 11.21:13

Tämä palaute koskee ainoastaan nauhoittamaasi puhenäytettä, eikä se kuvaa kaikkea suullista kielitaitoasi. Automaattinen arvio on koneen tekemä. Koneetta on opetettu muiden kielen oppijoiden puheella ja muulla kielilaineistolla.

Tässä tehtävässä ei ole yrityskertojen rajaa.

0:00 / 0:09

Puhenäytteesi tekstinä

öö mult oli jäi huppari tellä sinne kahvilaan eilen

Analyttinen arvio **Taitotasoarvio**

Taitotasoarvio



B1

Automaattisen arvion mukaan vaikuttaa siltä, että taitotasosi on B1.

Selviydyt arkielämän tilanteista kohdekielellä. Ääntämisessä on ymmärrettävää, käytät melko laajaa sanastoa ja erilaisia lauseita.

[Yritä uudelleen](#)

KUVA 2. Oppijan saama holistinen palaute tuottamistehtävästä.

Tuottamistehtävistä annettava analyttinen palaute (kuva 1) sisältää arviot sujuvuudesta, ääntämisestä, tehtävänantoon vastaamisesta ja ilmausten laajuudesta asteikolla 0–3/4. Taitotasoarvio-välilehdellä puolestaan näytetään automaattinen arvio oppijan taitotasosta asteikolla alle A1–C2 (kuva 2).

3.9 Ohjelmistotuotanto

Ohjelmisto tuotettiin yhteistyössä tietojenkäsittelytieteen opiskelijoiden kanssa tammi-toukokuun 2022 aikana. Työskentelimme ketterässä ohjelmistokehityksessä (*agile software development*) käytettävään *scrum*-periaatteen mukaisesti lyhyissä 1–2 viikon kehitysjaksoissa (*sprint*), joiden aikana työkaluun kehitettiin lisää toimintoja. Kehitysjaksojen sisältö suunniteltiin yhteisissä Zoom-videoneuvottelutyökalun avulla järjestetyissä palavereissa, joissa kävimme läpi tehdyn kehitystyön, listasimme asioita

kehitysjonoon (*backlog*) ja sovimme seuraavaksi kehitettävistä toiminnoista. (Stober & Hansmann 2010.) Lisäksi käytimme yhteydenpitoon projektin ja kehittäjien välillä Teams-ohjelmaa ja sähköpostia. Ketterän kehityksen lisäksi kehitystyötä ohjasivat suomalaisen opetussuunnitelman mukainen kielitaitokäsitys (Opetushallitus 2019), saavutettavuusvaatimukset (Etelä-Suomen aluehallintovirasto 2019) sekä saatavilla oleva aiempi tutkimus (ks. luku 2.2). Kehitystyön vaatimia päätöksiä tehtäessä priorisoitiin yleisesti ottaen oppijan etua ja työkalun käyttöä tutkimustarkoituksissa.

DigiTala-hankkeessa tutkimustuotokset jaetaan mahdollisuuksien mukaan avoimen tieteen periaatteita noudattaen. Näin ollen käyttöliittymän toteutustavaksi valittiin Moodle muun muassa avoimen lähdekoodin takia. Kehitystyössä huomioitiin Moodlen omat tarkat tyyl- ja tekniikkavaatimukset (Moodle 2023). Kehitetty työkalu on julkaistu vapaiden ohjelmistojen GNU GPL -lisenssillä, joka on niin kutsuttu *copyleft*-lisenssi: jos lisensoitua koodia muokataan, myös muokattu koodi on jaettava samalla lisenssillä. Ohjelmiston dokumentaatio on saatavilla Github-verkkosivustolla (von Zansen ym. 2022a). Vapaiden ja avointen ohjelmistojen hyödyt liittyvät avoimuuteen ja yhteistyöhön – työkalua voidaan kehittää hankkeen päättymisen jälkeenkin.

4 Pohdinta

Artikkelissa esittelimme DigiTala-tutkimushankkeessa kehitetyn automaattiseen puheentunnistukseen, arviointiin ja palautteeseen perustuvan työkalun, jonka avulla suomen ja ruotsin oppijat voivat harjoitella puheen tuottamista ja ääneen lukemista. Monitieteiseen soveltavaan tutkimukseen (luku 2.2) perustuva työkalu tarjoaa kieltenoppijoille ajasta ja paikasta riippumattomia harjoittelumahdollisuuksia sekä antaa heille sekä analyyttistä palautetta että taitotasoarvion puhe-suorituksesta.

DigiTalan työkalu on ensimmäinen, joka antaa analyyttistä palautetta useasta kielitaidon osa-alueesta nimenomaan spontaaneista suomen- ja ruotsinkielisistä puhenäytteistä. Analyyttinen palaute tukee erilaisia oppijoita, sillä kielitaidon osa-alueet kehittyvät usein eri tahtiin (ALTE 2016: 28, 40). Työkalumme antama palaute auttaa yksilöllisten kehitystarpeiden hahmottamisessa sekä puhujan yleisen taitotason määrittämisessä.

Tulevaisuudessa automaattista arviointia voidaan toivottavasti hyödyntää kielitaidon arvioinnissa laajemmin – myös yksilöä koskevissa tärkeissä (*high-stakes*) päätöksissä, kuten kansalaisuuden, työpaikan tai jatko-opiskelupaikan hakuprosesseissa (ks. ALTE 2016; Vaarala ym. 2021). Tekoälyn hyödyntämiseen kielitaidon arvioinnissa liittyy kuitenkin uhkia ja rajoitteita, joiden takia automaattista arviointia suositellaan yleensä käytettäväksi ihmisen tekemän arvioinnin rinnalla (*hybrid approach*, ks. Evanini & Zechner 2020: 11).

Koneoppimismallit ovat esimerkiksi alttiita aineistosta kumpuaville vinoumille (*bias*). Mitä rajallisempi tekoälyn opetusaineisto on, sitä tärkeämpää on huomioida siitä mahdollisesti johtuvat vinoumat. DigiTalassa kohdekielet, suomi ja suomenruotsi, ovat vähäresurssisia ja arviointityökalun kehittämiseen on käytetty huomattavasti suppeampaa opetusaineistoa kuin esimerkiksi ETS:n tuottamissa tutkimuksissa (ks. esim. Zhang ym. 2020). Haasteet kotimaisten kielten automaattisen arvioinnin käyttämisessä ja tutkimuksessa koskevatkin mahdollisten vinoumien havaitsemista malleissa ja niiden mahdollisesti aiheuttaman syrjinnän ennaltaehkäisemistä.

Tekoälyn kontekstissa vinoumalla viitataan koneoppimismalliin tai tekoälyä hyödyntävään sovellukseen, joka kohtelee tiettyä ihmisryhmää systemaattisesti epäsuotuisammin (Ojanen ym. 2022: 14). Tällainen automaattisen arviointityökalun vinouma johtuu yleensä opetusaineiston rajallisuudesta: arviointimalli oppii arvioimaan vain sellaisia puhujia (ja sellaisia tehtäviä), joiden vastauksista opetusaineisto koostuu.

Puhumisen arvioinnissa on haasteena saada kerättyä riittävän kattava aineisto, jossa huomioitaisiin kaiken tasoiset ja eri taustoista tulevat kielenoppijat. Myös kohdekielen variantit tuovat omat haasteensa. Esimerkiksi DigiTalan ruotsin arviointimallit on opetettu aineistolla, joka sisältää pääosin kieltenoppijoiden puhumaa suomenruotsia, jolloin arviointimallit eivät välttämättä toimi yhtä hyvin oppijoille, jotka puhuvat riikinruotsia (Kallio ym. 2021). Myös osa oppijoista tiedostaa puhetyylistä johtuvan vinouman mahdollisuuden (von Zansen ym. 2022b). Lisäksi DigiTalan aineistossa puhujien kielitaitotaso oli jakautunut epätasaisesti. Tämä tarkoittaa, että työkalu arvioi luotettavammin paremmin edustetuille taitotasolle osuvia puhenäytteitä kuin niitä puhenäytteitä, joille opetusaineistosta löytyy vähemmän vertailukohteita (Al-Ghezi ym. 2023).

Puhetyyliin erot korostuvat mitä pidemmistä puhunnoksista on kyse. Etenkin spontaaneissa, pitkissä puhunnoksissa voi olla paljon variaatiota, joka ei liity tai vaikuta kielitaidon arviointiin. Tällainen ”normaali” variaatio voi liittyä esimerkiksi sujuvuuspiirteisiin (ks. esim. Penttilä & Korpijaakko-Huuhka 2019; Lintunen ym. 2022). Spontaanissa puheessa esiintyy yleensä enemmän epäsujuvuuksia kuin luetussa puheessa ja niitä myös siedetään paremmin (Kallio ym. 2017). Mikäli sujuvuuden arviointimalli on opetettu suhteellisen lyhyillä ja hyvin valmistelluilla puhunnoksilla, se saattaa arvioida pidempiä puhenäytteitä turhan tiukasti. Tällaista arviointivinoumaa voidaan pienentää toisaalta aineistoa laajentamalla, toisaalta puhetehtäviä tai ohjeistuksia tarkentamalla ja siten vähentämällä tehtävän tulkinnasta johtuvaa variaatiota puhujien välillä.

Arviointivinoumat koskevat kuitenkin myös ihmisarvioijia, joiden toimintaan voidaan vaikuttaa koulutuksen ja ohjeistusten avulla sekä analysoimalla heiltä kerättyjä arviointeja (luku 3.4). Automaattinen arviointityökalu voi toimia myös koulutusvälineenä: vertaamalla omia arvioitaan koneen antamaan palautteeseen arvioija voi havaita itselleen uusia suullisen kielitaidon piirteitä tai tiedostaa arvioivansa tietynlaista puhetta tietyllä tavalla.

Yksi tekoälyn hyödyntämisen rajoite liittyy sen toimintaperiaatteiden ja käytännön soveltamisen ymmärtämiseen. Yhtäältä automaattisen arviointityökalun käyttäjien on

tärkeää ymmärtää, mihin automaattinen palaute perustuu, jotta he osaavat tulkita koneen antamaa palautetta oikein ja käyttää työkalua tarkoituksenmukaisesti. Toisaalta työkalun kehittäjillä tulee olla riittävä ymmärrys luotettavuuteen, valideiteettiin ja oikeudenmukaisuuteen liittyvistä kysymyksistä. (Evanini & Zechner 2020.)

Tämän artikkelin tavoitteena oli koostaa aiempaa puheen automaattista arviointia käsittelevää tutkimusta, dokumentoida DigiTalan arviointityökalun kehittämisvaiheet sekä kuvata työkalun toimintaperiaatteet. Työkalu toteutettiin ketterän ohjelmistokehityksen (Stober & Hansmann 2010) periaatteita noudattaen avoimeen lähdekoodiin perustuvalla Moodle-oppimisalustalle. Kokemuksen perusteella suosittelemme sekä ketterää ohjelmistokehitystä että Moodlea. Kannustammekin kielten opetuksen ja kielitaidon arvioinnin parissa työskenteleviä työkalujen kehittäjiä ja materiaalien laatijoita osallistamaan monipuolisesti eri käyttäjäryhmiä suunnitteluun varhaisesta vaiheesta lähtien. Tämän avulla pyrimme palvelemaan mahdollisimman erilaisia käyttäjiä (*design for all*, Etelä-Suomen aluehallintovirasto 2019).

Käyttäjäkokemusten tutkiminen on oleellinen osa työkalujen kehitystä, ja jatkamme sen äärellä tutkimushankkeen loppuun saakka. Oppijoille suunnattujen verkkotyökalujen käytettävyyttä voitaisiin jatkossa tutkia entistä monipuolisemmilla tutkimusmenetelmillä kuten yhdistämällä kyselyistä, havainnoinnista ja haastatteluisista saatavaa tietoa käyttäjän fysiologisiin mittauksiin. Esimerkiksi katseenseurannan avulla voitaisiin tutkia, mihin kohtiin ja kuinka kauaksi aikaa kielenoppija todella kohdistaa katseensa automaattista palautetta tarkastellessaan (ks. esim. Liu & Yu 2022). Hankkeessa saadut tulokset ja kokemukset ovat lupaavia puheen automaattisen arvioinnin ja koneoppimisen hyödyntämisen kannalta siinä. Tulevaisuudessa tarvitaan kuitenkin tutkimusta liittyen puhumisen muihin osa-alueisiin, kuten nonverbaaliseen viestintään, jota ei perinteisesti ole mitattu tai eksplisiittisesti arvioitu osana suullista kielitaitoa (akateemisten suomenoppijoiden suullisen vuorovaikutuksen automaattista arviointia koskevasta Aasis-tutkimushankkeesta ks. von Zansen 2023).

Kiitokset

Suomen Akatemia rahoittaa DigiTala-tutkimushanketta (2019–2023). Kiitämme yhteistyöstä projektin tutkijoita Raili Hildeniä Helsingin yliopistosta (rahoituspäätös 322619), Mikko Kurimoa, Ragheb Al-Gheziä, Yaroslav Getmania ja Ekaterina Voskoboinikia Aalto-yliopistosta (rahoituspäätös 322625) sekä Mikko Kurosta, Maria Kautosta ja Ari Huhtaa Jyväskylän yliopistosta (rahoituspäätös 322965). Lisäksi kiitämme Moodle-sovelluksen kehittämiseen osallistuneita Helsingin yliopiston tietojenkäsittelytieteen opiskelijoita: Tuomas Alanen, Topi Harjunpää, Joona Erkkilä, Maikki Heijala. Artikkelin oikolukemisesta kiitämme tutkimusavustaja Ilona Lähteenmäkeä. Kiitämme myös kaikkia projektin aikana aineiston keruuseen, litterointiin ja arviointiin osallistuneita avustajia, arvioijia ja opettajia.

Kirjallisuus

- Alanen, T., J. Erkkilä, T. Harjunpää & M. Heijala 2022. *Digitala Moodle plugin user manual* (1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.6535377>
- Al-Ghezi, R., Y. Getman, A. Rouhe, R. Hildén & M. Kurimo 2021. Self-supervised end-to-end ASR for low resource L2 Swedish. Teoksessa H. Heřmanský & H. Āernocký (toim.) *Proceedings of Interspeech 2021*. Brno: International Speech Communication Association (ISCA), 1429–1433. <http://dx.doi.org/10.21437/Interspeech.2021-1710>
- Al-Ghezi, R., K. Voskoboïnik, Y. Getman, A. von Zansen, H. Kallio, C. Akiki, M. Kuronen, A. Huhta & R. Hildén 2023. Automatic speaking assessment of spontaneous L2 Finnish and Swedish. *Language Assessment Quarterly*, 20 (4–5), 421–444. <https://doi.org/10.1080/15434303.2023.2292265>
- ALTE 2016. *Kielitestienv avulla osalliseksi, integroituneeksi ja kansalaiseksi. Opas päätöksenteon tueksi*. <https://alte.wildapricot.org/resources/Documents/LAMI%20Booklet%20FI.pdf> [luettu 22.11.2023]
- Anderson-Hsieh, J., R. Johnson & K. Koehler 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure. *Language Learning*, 42 (4), 529–555. <https://doi.org/10.1111/j.1467-1770.1992.tb01043.x>
- Baevski, A., Y. Zhou, A. Mohame, & M. Auli 2020. Wav2vec 2.0: a framework for self-supervised learning of speech representations. Teoksessa H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (toim.) *Proceedings of the 34th International conference on neural information processing systems (NIPS'20)*. Vancouver: Curran Associates Inc., 12449–12460. <https://proceedings.neurips.cc/paper/2020/file/92d1e1eb1cd6f9fba3227870bb6d7f07-Paper.pdf>
- Bernstein, J., M. Cohen, H. Murveit, D. Rtschev & M. Weintraub 1990. Automatic evaluation and training in English pronunciation. Teoksessa *First International conference on spoken language processing (ICSLP 1990)*. Kobe: Acoustical Society of Japan, 1185–1188. <https://doi.org/10.21437/ICSLP.1990-313>
- Boone, W. J., J. R. Staver & M. S. Yale 2014. *Rasch analysis in the human sciences*. Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-94-007-6857-4>
- Chen, L., J. Tao, S. Ghaffarzagdegan & Y. Qian 2018. End-to-end neural network based automated speech scoring. Teoksessa M. Hayes & H. Ko (toim.) *2018 IEEE International conference on acoustics, speech and signal processing (ICASSP)*. Calgary: Institute of electrical and electronics engineers, 6234–6238. <https://doi.org/10.1109/ICASSP.2018.8462562>
- Cucchiariini, C., H. Strik & L. Boves 1997. Automatic evaluation of Dutch pronunciation by using speech recognition technology. Teoksessa *1997 IEEE Workshop on automatic speech recognition and understanding proceedings*. Santa Barbara: Institute of electrical and electronics engineers, 622–629. <https://doi.org/10.1109/ASRU.1997.659144>
- Cucchiariini, C., H. Strik & L. Boves 2002. Quantitative assessment of second language learners' fluency: comparisons between read and spontaneous speech. *The Journal of the Acoustical Society of America*, 111 (6), 2862–2873. <https://doi.org/10.1121/1.1471894>
- Deeva, G., D. Bogdanova, E. Serral, M. Snoeck & J. De Weerd 2021. A review of automated feedback systems for learners: classification framework, challenges and opportunities. *Computers & Education*, 162. <https://doi.org/10.1016/j.compedu.2020.104094>
- Derwing, T. M., M. J. Munro, R. I. Thomson & M. J. Rossiter 2009. The relationship between L2 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31 (4), 533–557. <https://doi.org/10.1017/S0272263109990015>

- Eskenazi, M. 1996. Detection of foreign speakers' pronunciation errors for second language training – preliminary results. Teoksessa *Proceedings of the International congress on spoken language processing*. Philadelphia: International Speech Communication Association, 1465–1468. <https://doi.org/10.1109/ICSLP.1996.607892>
- Etelä-Suomen Aluehallintovirasto 2019. *Yleistä saavutettavuudesta*. <https://www.saavutettavuusvaatimukset.fi/yleista-saavutettavuudesta> [luettu 22.11.2023]
- Evanini, K. & K. Zechner 2020. Overview of automated speech scoring. Teoksessa K. Zechner & K. Evanini (toim.) *Automated speaking assessment: using language technologies to score spontaneous speech*. New York: Routledge, 3–20.
- Fan, J. & X. Yan 2020. Assessing speaking proficiency: a narrative review of speaking assessment research within the argument-based validation framework. *Frontiers in Psychology*, 11, 330–330. <https://doi.org/10.3389/fpsyg.2020.00330>
- Fulcher, G. 2014. *Testing second language speaking*. London: Routledge.
- Getman, Y. 2021a. *End-to-end low-resource automatic speech recognition for second language learners*. Pro gradu -tutkielma. Aalto-yliopisto, Sähkötekniikan korkeakoulu/ELEC. <https://aaltodoc.aalto.fi/handle/123456789/110588>
- Getman, Y. 2021b. Automated writing support for Swedish learners. Teoksessa P. Ljunglöf, S. Dobnik & R. Johansson (toim.) *Selected contributions from the eighth Swedish language technology conference (SLTC-2020), 25-27 November 2020*. Linköping electronic conference proceedings 184. Linköping: University of Linköping, 21–26. <https://doi.org/10.3384/ecp184171>
- Gu, L. & L. Davis 2020. Providing speech rater feature performance as feedback on spoken responses. Teoksessa K. Zechner & K. Evanini (toim.), *Automated speaking assessment: using language technologies to score spontaneous speech*. New York: Routledge, 159–175.
- Hsieh, C.-N., K. Zechner & X. Xi 2020. Features measuring fluency and pronunciation. Teoksessa K. Zechner & K. Evanini (toim.) *Automated speaking assessment: using language technologies to score spontaneous speech*. New York: Routledge, 101–122.
- Härmälä, M. & J. Marjanen 2023. *A-ruotsin oppimistulokset perusopetuksen päättövaiheessa 2022*. Helsinki: Kansallinen koulutuksen arviointikeskus. <https://www.karvi.fi/fi/julkaisut/ruotsin-oppimistulokset-perusopetuksen-paattovaiheessa-2022>
- Isaacs, T., & P. Trofimovich 2012. Deconstructing comprehensibility: identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34 (3), 475–505. <https://doi.org/10.1017/S0272263112000150>
- Kahng, J. 2018. The effect of pause location on perceived fluency. *Applied Psycholinguistics*, 39 (3), 569–591. <https://doi.org/10.1017/S0142716417000534>
- Kallio, H., M. Kautonen & M. Kuronen 2023. Prosody and fluency of Finland Swedish as a second language: investigating global parameters for automated speaking assessment. *Speech Communication*, 148, 66–80. <https://doi.org/10.1016/j.specom.2023.02.003>
- Kallio, H. & M. Kuronen 2023. Revising parameters for predicting L2 speech fluency and proficiency. Teoksessa R. Skarnitzl & J. Volín (toim.) *Proceedings of the 20th International congress of phonetic sciences 2023*. Prague: Guarant International, 2452–2456. <https://guarant.cz/icphs2023/221.pdf>
- Kallio, H., M. Kuronen & M. Kautonen 2021. Differences in acoustically determined sentence stress between native and L2 speakers of Finland Swedish. *Working papers – Lund University, Department of linguistics, General linguistics, Phonetics*, 56, 42–47. <http://urn.fi/URN:NBN:fi:ju-202111115628>

- Kallio, H., M. Kuronen & L. Koivusalo 2022a. The role of pause location in perceived fluency and proficiency in L2 Finnish. Teoksessa *Proceedings of the International symposium of applied phonetics 2022*. Lund: University of Lund, 22–27. <https://doi.org/10.21437/ISAPh.2022-5>
- Kallio, H., J. Šimko, A. Huhta, R. Karhila, M. Vainio, E. Lindroos, R. Hildén & M. Kurimo 2017. Towards the phonetic basis of spoken second language assessment: temporal features as indicators of perceived proficiency level. Teoksessa M. Kuronen, P. Lintunen & T. Nieminen (toim.) *Näkökulmia toisen kielen puheeseen. Insights into second language speech*. AFinLA-e. Soveltavan kielitieteen tutkimuksia 10. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 193–213. <https://journal.fi/afinla/article/view/73137>
- Kallio, H., A. Suni, J. Šimko & M. Vainio 2020. Analyzing second language proficiency using wavelet-based prominence estimates. *Journal of Phonetics*, 80. <https://doi.org/10.1016/j.wocn.2020.100966>
- Kallio, H., A. Suni & J. Šimko 2022c. Fluency-related temporal features and syllable prominence as prosodic proficiency predictors for learners of English with different language backgrounds. *Language and Speech*, 65 (3), 571–597. <https://doi.org/10.1177/00238309211040175>
- Kallio, H., R. Suviranta, M. Kuronen & A. von Zansen 2022b. Creaky voice and utterance fluency measures in predicting fluency and oral proficiency of spontaneous L2 Finnish. Teoksessa S. Frota, M. Cruz & M. Vigário (toim.) *Proceedings of Speech prosody 2022*. Lisbon: International Speech Communication Association (ISCA), 777–781. <https://doi.org/10.21437/SpeechProsody.2022-158>
- Kang, O. 2012. Relative impact of pronunciation features on ratings of non-native speakers' oral proficiency. Teoksessa J. Levis & K. LeVelle (toim.) *Proceedings of the 4th Pronunciation in second language learning and teaching conference*. Ames: Iowa State University, 10–15.
- Karhila, R., A. Rouhe, P. Smit, A. Mansikkaniemi, H. Kallio, E. Lindroos, R. Hildén, M. Vainio & M. Kurimo 2016. Digitala: An augmented test and review process prototype for high-stakes spoken foreign language examination. Teoksessa *Proceedings of Interspeech 2016*. San Francisco: International Speech Communication Association (ISCA), 784–785. https://www.isca-archive.org/interspeech_2016/karhila16_interspeech.pdf
- Kautonen, M. & M. Kuronen 2021. Kvantitatiivinen perspektiivi L2-talteen eri färdighetsnivåer. *Folkmarksstudier*, 59, 11–40. <https://journal.fi/folkmarksstudier/article/view/112545>
- Kautonen, M. & A. von Zansen 2020. DigiTala research project: Automatic speech recognition in assessing L2 speaking. *Kieli, koulutus ja yhteiskunta*, 11 (4). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-kesakuu-2020/digitala-research-project-automatic-speech-recognition-in-assessing-l2-speaking>.
- Koivusalo, L. 2022. *Phonetic fluency in Finnish as a second language: acoustic analysis of high school students' spontaneous speech*. Pro gradu -tutkielma, Helsingin yliopisto. <http://urn.fi/URN:NBN:fi:hulib-202206152522>
- Lintunen, P., M. Mutta, S. Olkkonen, P. Peltonen & O. Veivo 2022. Sujuvuus ja epäsujuvuus vieraan kielen oppimisen näkökulmasta: monitahoinen ilmiö edellyttää monitieteistä tutkimusta. Teoksessa S. Loukusa, T. Hautala & A.-K. Tolonen (toim.) *Sujuvaa vai sujumatonta? Puheen ja kielen sujuvuutta tutkimassa*. Helsinki: Puheen ja kielen tutkimuksen yhdistys, 75–87.
- Liu, S. & G. Yu 2022. L2 learners' engagement with automated feedback: An eye-tracking study. *Language Learning & Technology*, 26 (2), 78–105. <https://doi.org/10.1257/73480>
- Loukina, A. & S. Y. Yoon 2020. Scoring and filtering models for automated speech scoring. Teoksessa K. Zechner & K. Evanini (toim.) *Automated speaking assessment: using language technologies to score spontaneous speech*. New York: Routledge, 75–98.

- Luoma, S. 2004. *Assessing speaking*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511733017>
- Moodle 2023. *Moodle developer documentation*. https://docs.moodle.org/dev/Main_Page [luettu 22.11.2023]
- Munro, M. J. & T. M. Derwing 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45 (1), 73–97.
<https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Ojanen, A., O. Sahlgren, J. Vaiste, A. Björk, J. Mikkonen, K. Kimppa, A. Laitinen & N. Oljakka 2022. *Algoritminen syrjintä ja yhdenvertaisuuden edistäminen: arviointikehikko syrjimättömälle tekoälylle*. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 2022, 54. Helsinki: Valtioneuvosto. <http://urn.fi/URN:ISBN:978-952-383-404-0>
- Opetushallitus 2003. *Lukion opetussuunnitelman perusteet*. Helsinki: Opetushallitus.
- Opetushallitus 2019. *Lukion opetussuunnitelman perusteet*. Helsinki: Opetushallitus.
- Peltonen, P. & P. Lintunen 2022. Multilingual speakers' L1, L2, and L3 fluency across languages: a study of Finnish, Swedish, and English. *Nordand*, 17 (1), 48–63.
<https://doi.org/10.18261/nordand.17.1.4>
- Penttilä, N. & A. M. Korpijaakko-Huuhka 2019. Disfluencies in typical Finnish-speaking adults. *The Phonetician*, 27 (116), 28–41.
- Stober, T. & U. Hansmann 2010. Overview of agile software development. Teoksessa T Stober & U. Hansmann (toim.) *Agile software development*. Heidelberg: Springer Berlin, 35–59. https://doi.org/10.1007/978-3-540-70832-2_3
- Tavakoli, P. 2011. Pausing patterns: differences between L2 learners and native speakers. *ELT Journal*, 65 (1), 71–79. <https://doi.org/10.1093/elt/ccq020>
- Toivola, M., M. Lennes & E. Aho 2009. Speech rate and pauses in non-native Finnish. Teoksessa *Proceedings of Interspeech 2009*. Brighton: International Speech Communication Association (ISCA), 1707–1710. <https://doi.org/10.21437/Interspeech.2009-515>
- Vaarala, H., S. Riuttanen, E. Kyckling & S. Karppinen 2021. *Kielivaranto. Nyt!: monikielisyys vahvuudeksi -selvityksen (2017) seuranta*. Jyväskylä: Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto. <https://jyx.jyu.fi/handle/123456789/74416>
- Voskoboinik, E., Y. Getman, R. Al-Ghezi, M. Kurimo & T. Grósz 2023. Automated assessment of task completion in spontaneous speech for Finnish and Finland Swedish language learners. Teoksessa D. Alfter, E. Volodina, T. François, A. Jönsson & E. Rennes (toim.) *Proceedings of the 12th Workshop on natural language processing for computer assisted language learning (NLP4CALL 2023)*. Linköping: University of Linköping, 102–110.
<https://doi.org/10.3384/ecp197012>
- Wang, X., K. Zechner & C. Hamill 2020. Targeted content feedback in spoken language learning and assessment. Teoksessa H. Meng, B. Xu & T. Zheng (toim.) *Proceedings of Interspeech 2020*. Shanghai: International Speech Communication Association (ISCA), 3850–3854. <https://doi.org/10.21437/Interspeech.2020-1766>
- White, L. & S. L. Mattys 2007. Calibrating rhythm: first language and second language studies. *Journal of Phonetics*, 35 (4), 501–522. <https://doi.org/10.1016/j.wocn.2007.02.003>
- Xi, X., D. Higgins, K. Zechner & D. M. Williamson 2008. Automated scoring of spontaneous speech using speechratersm v1.0. *ETS Research Report Series*, 2008 (2), i–102.
<https://doi.org/10.1002/j.2333-8504.2008.tb02148.x>
- Ylinen, S. & M. Kurimo 2017. Kielenoppiminen vauhtiin puheteknologian avulla. Teoksessa H. Savolainen, R. Vilkkonen & L. Vähäkylä (toim.) *Oppimisen tulevaisuus*. Helsinki: Gaudeamus, 57–69. <http://hdl.handle.net/10138/309750>

- Yoon, S.-Y., X. Lu & K. Zechner 2020. Features measuring vocabulary and grammar. Teoksessa K. Zechner & K. Evanini (toim.) *Automated speaking assessment: using language technologies to score spontaneous speech*. New York: Routledge, 123–137.
- von Zansen, A. 2022a. DigiTala's post-rating questionnaire for human raters (Finnish, upper secondary schools, Jun2021). Zenodo. <https://doi.org/10.5281/zenodo.6477015>
- von Zansen, A. 2022b. DigiTala's post-test questionnaire for L2 Finnish learners (upper secondary schools 2021). Zenodo. <https://doi.org/10.5281/zenodo.6562884>
- von Zansen, A. 2022c. DigiTala's pre-test consent and background information form for L2 Finnish learners (upper secondary schools 2021). Zenodo. <https://doi.org/10.5281/zenodo.6562663>
- von Zansen, A. 2022d. DigiTala's rating criteria: Holistic and analytic scales for assessing L2 speaking. Zenodo. <https://doi.org/10.5281/zenodo.6477089>
- von Zansen, A. 2022e. DigiTala's speaking tasks and questionnaire for L2 Finnish learners (proficiency level A). Zenodo. <https://doi.org/10.5281/zenodo.6627533>
- von Zansen, A. 2022f. DigiTala's speaking tasks for L2 Finnish learners (proficiency level B1). Zenodo. <https://doi.org/10.5281/zenodo.6562855>
- von Zansen, A. 2022g. DigiTala's speaking tasks for L2 Finnish learners (proficiency level B2). Zenodo. <https://doi.org/10.5281/zenodo.6562865>
- von Zansen, A. 2023. The Aasis research project: automatically assessing spoken interaction in L2 Finnish. *Kieli, koulutus ja yhteiskunta*, 14 (7). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-joulukuu-2023/the-aasis-research-project-automatically-assessing-spoken-interaction-in-l2-finnish>
- von Zansen, A., T. Alanen, R. Al-Ghezi, J. Erkkilä, T. Harjunpää., M. Heijala & H. Kallio 2022a. *DigiTala Moodle plugin*. https://github.com/aalto-speech/moodle-mod_digitala
- von Zansen, A. & M. Heijala 2023. Miten suomen ja ruotsin opettajat käyttäisivät puheen automaattiseen arviointiin kehitettyä työkalua?. Teoksessa T. Mäkipää, R. Hilden & A. Huhta (toim.) *Kielenoppimista tukeva arviointi – Assessment for supporting language learning*. AFinLA-teema 15. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 124–141. <https://journal.fi/afinla/article/view/124822>
- von Zansen, A. & R. Hilden 2022. "It was cool and comfortable!" Akateemisten alkeistason S2-opiskelijoiden kokemuksia tietokoneella suoritettavasta puhumisen kokeesta. Teoksessa S. Routarinne, P. Heinonen, T. Kärki, A. Roiha. M.-L. Rönkkö & A. Korkeaniemi (toim.) *Ainedidaktikka ajassa: laajenevat oppimisympäristöt ja eri-ikäiset oppijat*. Suomen ainedidaktisen tutkimusseuran julkaisuja 22. Turku: Turun yliopisto, Suomen ainedidaktinen tutkimusseura ry, 72–90. <http://hdl.handle.net/10138/353562>
- von Zansen, A., R. Hilden & M. Sneck 2022b. Lukiolaisten käsitykset ja heidän antamansa palaute suullisen kielitaidon automaattisesta arvioinnista. Teoksessa R. Kantelinen, M. Kautonen & Z. Elgundi (toim.) *Linguapeda 2021*. Suomen ainedidaktisen tutkimusseuran julkaisuja 21. Joensuu: Itä-Suomen yliopisto, Suomen ainedidaktinen tutkimusseura ry, 176–205. <http://hdl.handle.net/10138/352128>
- von Zansen, A. & A. Huhta 2022. Developing automated feedback on spoken performance: exploring the functioning of five analytic rating scales using many-facet Rasch measurement. Teoksessa J. H. Jantunen, J. Kalja-Voima, M. Laukkarinen, A. Puupponen, M. Salonen, T. Saresma, J. Tarvainen, & S. Ylönen (toim.) *Diversity of methods and materials in digital human sciences: proceedings of the Digital research data and human sciences DRDHum conference 2022, December 1–3, Jyväskylä, Finland*. Jyväskylä: Jyväskylän yliopisto, 211–229. <http://urn.fi/URN:ISBN:978-951-39-9450-1>

- von Zansen, A., H. Kallio, M. Sneck, M. Kuronen, A. Huhta & R. Hilden 2022c. Ihmisarvioijien näkemyksiä suullisen kielitaidon automaattisesta arvioinnista, digitaalisesta arviointiprosessista sekä puhesuorituksista arvioitavista ulottuvuuksista. Teoksessa T. Seppälä, S. Lesonen, P. Ikkänen & S. D'hondt (toim.) *Kieli, muutos ja yhteiskunta – Language, change and society*. AFinLAn vuosikirja 2022. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 370–394. <https://journal.fi/afinlavk/article/view/114821>
- von Zansen, A. & L. L. Tarvainen-Li 2024. DigiTala's speaking tasks and questionnaire for L2 Swedish learners (proficiency level A1-A2). Zenodo. <https://doi.org/10.5281/zenodo.10693570>
- Zechner, K. 2020. Summary and outlook on automated speech scoring. Teoksessa K. Zechner & K. Evanini (toim.) *Automated speaking assessment: using language technologies to score spontaneous speech*. New York: Routledge, 192–204.
- Zhang, M., L. Yao, S. J. Haberman & N. J. Dorans 2020. Assessing scoring accuracy and assessment accuracy for spoken responses: using human and machine scores. Teoksessa K. Zechner & K. Evanini (toim.) *Automated speaking assessment: using language technologies to score spontaneous speech*. New York: Routledge, 32–58.

Toivola, M., P. Lintunen & L. M. Heikkola (toim.) 2024. Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. *New directions in speech research – Freely produced speech as data.* AFinLA-teema / n:o 17, 117–141.

Leena Maria Heikkola

UiT Norjan arktinen yliopisto
Åbo Akademi

Minnaleena Toivola

Helsingin yliopisto

Jeongdo Kim

Vapaa tutkija

Antti Saloranta

Turun yliopisto

Kielivalinnat ja kielenvaihtaminen ensikielisten ja ei-ensikielisten suomenpuhujien keskusteluissa – Kyselytutkimus

Nostot

- Suurin osa ensikielistä suomenpuhujista raportoi aloittavansa keskustelun oletetun ei-ensikielisen puhujan kanssa suomeksi.
- 95 % ensikielistä suomenpuhujista kertoi kuitenkin joskus vaihtavansa keskustelun kielen suomesta englanniksi.
- Kielenvaihtamisen syiksi mainitaan mm. tehokkuus, nopeus, kohteliaisuus ja halu auttaa ei-ensikielistä suomenpuhujaa.
- Ensikieliset suomenpuhujat olettavat, että ei-ensikieliset suomenpuhujat suhtautuvat kielenvaihtamiseen pääasiassa positiivisesti.



VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Abstract

Immigration into Finland has increased in recent years. Learning Finnish is essential for integrating into Finnish society. We investigated L1 Finnish speakers' language choices and language switching in everyday conversations with L2 Finnish speakers, and their reasoning for these choices. The data were gathered in an online survey with 358 respondents. The majority of L1 Finnish speakers reported starting conversations with L2 Finnish speakers in Finnish. However, 95% of respondents also reported switching the language of the conversation to English. The reported reasons were efficiency, speed, politeness and the desire to help. Most respondents also assumed that L2 Finnish speakers would react positively to the language switch. The results outline the need for mutual agreement on language choice and switching in everyday discussions. L1 Finnish speakers should understand the importance of offering possibilities to speak Finnish to facilitate language learning and to ensure equality in discussions and aid participation in Finnish society.

Keywords: language choices, language switching, L1 Finnish speakers, L2 Finnish speakers

Asiasanat: kielivalinnat, kielenvaihtaminen, ensikieliset suomenpuhujat, ei-ensikieliset suomenpuhujat

1 Johdanto

Suomessa on yhä enemmän muualta tulleita ihmisiä (Tilastokeskus 2023). Heille suomen kielen taito on yksi olennaisimpia uuteen yhteiskuntaan kotoutumisen välineitä, ja sillä on suuri merkitys niin sosiaalisten suhteiden solmimiselle kuin työllistymisellekin. Esimerkiksi osa työpaikoista edellyttää hakijalta hyvää suomen kielen taitoa, ja suomen osaaminen on tärkeää myös arjessa, vaikka hyvä kielitaito ei itsessään takaa kuulumisen kokemusta (Scotson 2018; Intke-Hernandéz 2020). Suomenoppijalle mahdollisuus käyttää suomea monipuolisesti eri tilanteissa onkin merkittävää. Käytännön esteeksi suomen käyttämiselle voi kuitenkin nousta se, että vuorovaikutustilanteissa käytetään usein englantia. Vaikka ei-ensikielisen suomenpuhujan suomen kielitaito olisi hyväkin, saattavat ensikieliset suomenpuhujat helposti vaihtaa keskustelun kielen englanniksi riippumatta siitä, millä kielellä keskustelu on aloitettu (Scotson 2020; Ruuska 2020). Tällaista kielenvaihtamista voivat motivoida monet seikat: vaihtoa voidaan pitää kohteliaisuutena kielenoppijaa kohtaan, keinona helpottaa ja nopeuttaa keskustelun tavoitteen saavuttamista tai mahdollisuutena harjoitella itse englannin käyttöä (Kim ym. 2022).

Suomea opiskelevalle vaihto englantiin voi kuitenkin olla kielenoppimisen kannalta haitaksi, sillä jos kieltä ei pääse harjoittelemaan, sen oppiminen ei edisty. Optimaalinen kielenoppiminen vaatii kielenkäyttömahdollisuuksia myös kielikurssia vapaammissa puhetilanteissa. Keskustelun kielen vaihtamisella on myös sosiaalisia vaikutuksia. Kun ensikielinen puhuja päättää yksipuolisesti vaihtaa keskustelun kieltä, hän voi asettaa ei-ensikielisen epätasa-arvoiseen rooliin, jossa tämän asema yhteisön

ulkopuolisena ihmisenä korostuu (Hippi & Lehto 2023). Ei-ensikielinen ei näin pääse osaksi suomenkielistä yhteisöä, eikä synny sosiaalista osallisuutta (Leemann ym. 2015). Yhteiskunnassa osallisuuden kokemusta rakentavat sekä kielikäytännöt että -ideologiat (Lehto 2023a), mutta ensikieliset puhujat eivät välttämättä ole tietoisia omien kielivalintojensa vaikutuksista ei-ensikielisten osallisuuteen. Ulkopuolisuus voi erityisesti korostua tilanteissa, joissa ei-ensikielinen kokee keskustelun kielen vaihtuneen esimerkiksi ulkonäön, nimen tai muun kielitaitoon liittymättömän asian takia. Tällainen ensikielisen puhujan tekemä – mahdollisesti jopa täysin tiedostamaton – jaottelu ”meihin” ja ”muihin” voi haitata Suomeen muuttaneen ihmisen integraatiota ja tuottaa negatiivisia kokemuksia.

Tämän tutkimuksen tavoitteena on tuottaa tietoa ensikielisten suomenpuhujien keskusteluista heille tuntemattomien ei-ensikielisten suomenpuhujien kanssa. Käytämme tässä tutkimuksessa nimitystä *ei-ensikielinen suomenpuhujia*, koska haluamme korostaa suomenpuhujien osallisuutta suomenkieliseen yhteisöön jo mahdollisesti vielä kehittyvän kielitaidon aikana. Tutkimuksessa tarkastellaan ensikielisten suomenpuhujien näkemyksiä arjen vuorovaikutuksellisissa puhetilanteissa tapahtuvista kielivalinnoista ja mahdollisesta kielenvaihtamisesta sekä näiden taustalla vaikuttavista syistä. Lisäksi tutkitaan ensikielisten suomenpuhujien käsityksiä ei-ensikielisten suomenpuhujien mahdollisista reaktioista vapaissa puhetilanteissa tapahtuvaan kielenvaihtamiseen.

Tutkimuksessa käytetty termi *kielenvaihtaminen* ei ole vakiintunut. Kotilainen ja Lehtimaja (2019) määrittelevät *kielenvaihdon* kielitaitoon liittyvänä kommunikaatiostrategiana. Toisin kuin Kotilainen ja Lehtimaja, korostamme kielenvaihtaminen-termillä sitä, että ensikielisen ja ei-ensikielisen välillä keskustelun etenemismvastuu ja ongelmista selviäminen näyttäytyy epätasapuolisena. Omien kokemustemme pohjalta vaikuttaa siltä, että ensikielisten suomenpuhujien yksipuolinen kielenvaihtaminen suomesta englanniksi luo epätasa-arvoa keskusteluun, etenkin jos ei-ensikielisen toiveet keskustelun kielestä sivuutetaan. Aiemmassa suomalaisessa tutkimuksessa esitetäänkin tärkeänä neuvottelut ei-ensikielisen ja ensikielisen puhujan välille siitä, millä kielellä yhteisiä keskusteluja käydään (Lehto, 2023a). Vaikka vuorovaikutustilanteita ensikielisen ja ei-ensikielisen välillä onkin tutkittu aiemmin (ks. esim. Kurhila 2001; Lilja ym. 2022; Lilja 2010; Suni 2008), ei Suomessa ole aiemmin tutkittu kielivalintoja ja kielenvaihtamista ensikielisten näkökulmasta.

Tutkimuksen tulosten perusteella pohditaan, miten ensikielisten suomenkielisten kielelliset käytännöt vaikuttavat ei-ensikielisten suomenpuhujien osallisuuteen suomenkieliseen yhteisöön ja tasa-arvoiseen keskustelukulttuuriin. Tutkimus on osa Tasku-hanketta, jonka tavoitteena on erityisesti ensikielisten suomenpuhujien herätely tähän kielelliseen tasa-arvokysymykseen aineistonkeruun ja tutkimustulosten esittelyn kautta.

Tätä tutkimusta ohjaavat seuraavat tutkimuskysymykset:

- 1) Millä kielellä ensikieliset suomenpuhujat aloittavat keskustelun ei-ensikielisen kanssa ja miksi?
- 2) Vaihtavatko ensikieliset suomenpuhujat keskustelukielen suomesta englantiin ja jos vaihtavat, miksi?
- 3) Miten ensikieliset suomenpuhujat arvioivat suomenoppijoiden reagoivan kielenvaihtamiseen suomesta englanniksi?

2 Kuka saa puhua suomea?

Suomessa, kuten monissa muissakin länsimaissa, englanninkielisen kulttuurin laaja levinneisyys ja kansalaisten keskimäärin hyvä englannin kielen taito ovat saaneet aikaan tilanteen, jossa englantia pidetään ”kaikkien” osaamana kielenä. Suomessa englantia osaavat hyvin erityisesti korkeakoulutetut nuoret aikuiset (Leppänen ym. 2011). Loppuvuodesta 2023 ilmestyneessä, Valtioneuvoston tilaamassa raportissa (Laitinen ym. 2023) englantia näyttyy suomen kielen ohella maan toiseksi käytetyimpänä kielenä. Englannin rooli on tärkeä erityisesti korkeakouluissa ja tieteessä sekä elinkeinoelämässä. Joillekin ei-ensikielisille englantia voikin esimerkiksi asiointitilanteissa olla tärkeä tukikieli, joka helpottaa arjessa pärjäämistä etenkin Suomessa asumisen alkuvaiheessa.

On kuitenkin syytä huomata, että englannin kielen taito on jakautunut maailmassa maantieteellisesti hyvin epätasaisesti. Englanninpuhujia on maailmassa noin 1,4 miljardia (Ethnologue 2023), eli vain alle viidennes maailman 8,05 miljardista asukkaasta, eikä englantia suinkaan ole lingua franca kaikissa maailman maissa. Niinpä myös Suomeen muuttavien ihmisten englannin osaaminen vaihtelee suuresti. Englannin kielitaidossa on suuria maiden välisiä eroja Afrikassa ja Lähi-idässä sekä Pohjois- ja Etelä-Euroopan välillä (Education First 2021). Keskustelun kielen vaihtaminen englantiin ei siis aina välttämättä helpota vuorovaikutusta. Harkitsematon kielenvaihtaminen voi olla jopa haitallista kielenoppimisen kannalta ja toiseuttaa ei-ensikielistä suomenpuhujaa. Tosiasia on myös, etteivät kaikki ensikielisetkään suomenpuhujat puhu englantia: esimerkiksi iäkkäämpi väestö ja maaseudun asukkaat osaavat englantia heikommin tai eivät lainkaan verrattuna valtaväestöön (Lehto 2023b). Myös se, että suomalaisessa työelämässä pärjää pelkällä englannilla, on yksinkertaistus, sillä suomea tarvitaan usein ainakin työpaikan epävirallisessa kommunikoinnissa (Lehtimaja ym. 2021; Lehtimaja 2017).

Tämän tutkimuksen kirjoittajien omien kokemusten pohjalta edes edistynyt kielitaito ei aina riitä keskustelun jatkamiseen suomeksi. Tilanne näkyy myös suomenopettajalle, jolle raportoidaan, että ensikieliset vaihtavat keskustelun kielen usein suomesta englantiin. Tällaiseen kielenvaihtamiseen on viitattu myös tutkimuksessa.

Esimerkiksi Katharina Ruuska (2020: 72) tuo väitöskirjassaan esiin ei-ensikielisten turhautumisen siihen, että ensikielinen suomenpuhuja vaihtaa kielen usein englantiin, vaikka ei-ensikielinen olisi aloittanut keskustelun tarkoituksellisesti suomen kielellä: “For a long time I would, for instance, be regularly addressed in English by native Finnish speakers, even if I had initiated the interaction in Finnish, and I found this deeply discouraging.”

Ruuska (2020) tutki väitöskirjassaan, miten kielitaidoltaan erittäin edistyneet ei-ensikieliset kohdataan arkitilanteissa ja miten he reagoivat näihin tilanteisiin. Ruuskan koehenkilöt tulkittiin usein erittäin hyvin suomea osaaviksi suomenruotsalaisiksi, virolaisiksi tai varhaisiän kaksikielisiksi. Tämä heijastelee ensikielisten suomenkielisten vaikeutta tunnustaa, että muut kuin ensikieliset voisivat puhua suomea hyvin edistyneellä kielitaitotasolla. Ruuskan tutkimustulokset osoittavat, että ei-ensikielisten mahdollisuudet osallistua ja kokea kuuluvansa yhteisöön kietoutuvat ympäristön implisiittisiin normeihin ja odotuksiin. Ei-ensikieliset kertoivat, että näissä normeissa ja odotuksissa on eroja eri kaupunkien välillä: esimerkiksi pääkaupunkiseudun monikielisemmässä ympäristössä odotukset suomen kielitaidosta ovat matalammat kuin pienemmissä, homogeenisemmissä kaupungeissa (Ruuska 2020: 242). Ruuska (2020: 249) peräänkuuluttaakin kieliyhteisön muutosta, joka tukisi ei-ensikielisten nykyistä parempaa osallisuutta ja kuuluvuutta suomalaisen yhteisöön.

Mia Scotson tutki väitöskirjassaan (2020) korkeasti koulutettuja kotoutujia suomenkäyttäjänä. Scotson (2020: 111) tuo esiin, että uuden kielen käyttäminen (ja oppiminen) edellyttää kieliympäristössä esiintyviä tarjoumia (engl. *affordance*), mutta myös ei-ensikielisen toimijuutta (engl. *agency*) (tarjoumista ks. van Lier 2000; toimijuudesta ks. Vitanova ym. 2015). Hän keskittyi tutkimuksessaan erityisesti ei-ensikielisten tunteisiin ja käsityksiin sekä sosiokulttuurisen ympäristön merkitykseen toimijuudelle. Väitöskirjansa toisessa, puolistrukturoituihin haastatteluihin perustuvassa osatutkimuksessa Scotson (2018) tarkasteli ympäristön vaikutuksia ei-ensikielisen suomen käyttöön. Suuri este ei-ensikielisen toimijuudelle suomen kielen käyttäjänä oli ensikielisten suomenpuhujien kielivalinnat tai kielenvaihtaminen englantiin, vaikka ei-ensikielinen itse halusi puhua suomea. Myös aikapaine ja vuorovaikutuksessa hoidettava asia vaikuttivat ei-ensikielisten suomen kielen käyttöön – usein negatiivisesti. Erityisesti korkeakoulutetut valitsivat englannin asiointikieleksi haastavimmiksi kokemissaan tilanteissa (Scotson 2020).

Scotsonin (2020) tutkimus osoitti, että ensikielisen puhekumppanin oikea-aikainen tuki on edellytys ei-ensikielisen toimijuudelle suomen kielellä. Onkin tärkeää, että ensikielinen suomenpuhuja antaa ei-ensikieliselle mahdollisuuden käyttää suomen kieltä silloin, kun ei-ensikielinen sitä itse toivoo. Scotson (2020: 14) nostaa myös esille, että englannin kielen jatkuva käyttäminen suomenkielisten puhekumppanien kanssa hidastaa suomen kielen oppimista. Lisäksi se asettaa ei-ensikieliset epätasa-arvoiseen asemaan, kun ensikielinen suomenpuhuja yksipuolisesti valitsee

tai vaihtaa keskustelun kielen suomesta englanniksi (vrt. Kotimaisten kielten keskus 2021a, 2021b; Kotilainen 2013).

Vaikka Suomessa monen työpaikan työntekijöillä on erilaisia kielitaustoja, työkieli voi kuitenkin usein olla englanti. Se ei kuitenkaan aina riitä yhteisen kommunikaation tarpeisiin (Kotilainen ym. 2022), vaikka se voidaan nähdä myös tasa-arvoisimpana kielivalintana, koska usein englanti on kaikille toinen tai vieras kieli. Kotilainen kollegoineen (2022) korostaa kuitenkin sitä, että kaikki eivät osaa englantia, ja sitä osaaVILLEkin sen käyttö työssä voi olla vähintäänkin kuormittavaa. Kun työpaikalla käytetään suomea, saa ei-ensikielinen tilaisuuden harjoitella kielitaitoaan. Tämäkin voi tuki aiheuttaa ongelmia, koska työntekijän ja kielenoppijan rooleja voi olla vaikea sovittaa toisiinsa. Kotilainen ym. (2022) nostavat esiin tärkeän kysymyksen: ”hyväksyykö suomenkielinen puhujayhteisömme ’vajavaisen’ suomen?” He kysyvät myös, antavatko ensikieliset suomenpuhujat ei-ensikieliselLe mahdollisuuksia harjoitella suomen käyttöä ja tukevatko he kielenoppimista kielitietoisilla käytänteillä vai ”vaihdammeko aina englantiin, kun kuulemme omasta mielestämme puutteellisesti tuotettua suomea?” (Kotilainen ym. 2022).

Lehto (2023a) tutki ei-ensikielisten kielivalintoja keskusteluissa ensikielisten suomenpuhujien kanssa. Hän keskittyi erityisesti monikielisten tutkittaviensa englannin kielen valintoihin, vaikka he itsekin huomauttavat, että kaikki Suomessa eivät osaa englantia (ks. myös Iikkanen 2017). Lehdon monikieliset puhujat tuovat esille kokemansa jännitteen siitä, että keskustelussa ensikielisten kanssa pitäisi odotusarvoisesti käyttää vain yhtä kieltä, vaikka useamman kielen käyttö olisi heille itselleen usein luonnollisempaa. Toisaalta englanti nähdään tukena tilanteessa, jossa keskustelu keskeytyisi ilman kielenvaihtamista englantiin. Usein kielenvaihtamisesta englantiin ei kuitenkaan käydä neuvottelua lainkaan keskustelukumppaneiden kesken tai suomen kielen käyttäminen johtaa ei-toivottuun lopputulokseen, esimerkiksi kielenvaihtamiseen tai keskustelun päättymiseen (Lehto 2023a).

Ensikielisten käsityksiä kielenvaihtamisesta keskusteluissa ei-ensikielisten kanssa on tutkittu vähän. Tanskassa aihetta on tutkittu hieman, esimerkiksi Benediktsson (2015) on esittänyt, että ensikieliset tanskanpuhujat saattavat vaihtaa keskustelukielen tanskasta englantiin keskustellessaan ei-ensikielisten tanskanpuhujien kanssa. Anna Bothe Jespersen ja Míša Hejná (2021) tarkastelivat tutkimuksessaan ensikielisten tanskalaisten ja ei-ensikielisten tanskanpuhujien välistä kielenvaihtamista. He olettivat, että tanskankielisten taipumus väheksyä yleistanskasta poikkeavien puhemuotojen käyttöä voi johtaa kielenvaihtamiseen, mikä voi toiseuttaa maahanmuuttotautaisia ei-ensikielisiä ja vähentää heidän osallisuuttaan tanskalaiseen yhteiskuntaan. Jespersen ja Hejná (2021) tarkastelivat kyselytutkimuksessaan sekä ei-ensikielisten ($n = 409$) että ensikielisten tanskanpuhujien ($n = 134$) käsityksiä yhteisestä vuorovaikutuksesta. Kyseinen tutkimus vastaa nyt käsillä olevaa tutkimusta erityisesti tutkimusmenetelmiensä osalta, koska tutkittavana ovat ensikieliset ja aineistonkeruuseen käytettiin

kyselyä. Vertailemmekin tanskalaisten tutkijoiden saamia tuloksia tämän tutkimuksen tuloksiin.

Jespersenin ja Hejnán kyselyyn vastanneista ei-ensikielisistä tanskanpuhujista 34,6 % ilmoitti aina aloittavansa keskustelut tanskaksi ensikielisten tanskanpuhujien kanssa, 11,5 % usein ja 18,2 % joskus. Näin ollen yhteensä 64,3 % (258 vastaajaa) ilmoitti aloittavansa keskustelut tanskaksi ainakin joskus. Kukaan tanskanoppijoista ei vastannut, ettei olisi koskaan yrittänyt aloittaa keskustelua tanskaksi ensikielisen tanskanpuhujan kanssa.

Yli puolet (54,6 %) ei-ensikielisistä tanskanpuhujista oli usein osallistunut keskusteluun, jossa ensikielinen tanskanpuhujia oli vaihtanut kielen tanskasta englanniksi. Vain 13,2 % ilmoitti, että kielenvaihtamista ei ollut koskaan tapahtunut heille. Kun ei-ensikielisiltä tanskanpuhujilta kysyttiin heidän kielellistä reaktiotaan kielenvaihtamiseen, 74,3 % raportoi keskustelun jatkuvan englanniksi ensikielisen tanskankielisen sen vaihdettua. Ei-ensikielisiltä tanskanpuhujilta kysyttiin myös, millaisia tunnereaktioita kielenvaihtaminen heissä herättää. Mitä useammin ei-ensikielinen kohtasi kielenvaihtamista, sitä negatiivisemmin hän siihen suhtautui. Yhteensä 57,9 % vastaajista ilmaisi reagoivansa kielenvaihtamiseen negatiivisesti; loppuissa vastauksissa suhtautuminen oli vaihtelevampaa. Positiivisesti siihen suhtautuivat vain harvat. Avoimissa vastauksissa ei-ensikieliset tanskanpuhujat kertoivat esimerkiksi, että kielenvaihtaminen on epäkohteliasta ja ärsyttävää tai että se tekee heidät vihaiseksi. Vaihtelevaan suhtautumiseen annettiin mm. seuraavia syitä: ”se ei haittaa minua”, ”jatkan tanskaksi” tai ”riippuu kontekstista”. Harvoja positiivisia kokemuksia toivat kohteliaisuudeksi tai auttamishaluksi tulkitut kielenvaihtamiset. (Jespersen & Hejná 2021.)

Ensikielisille tanskanpuhujille suunnatussa kyselyssä kaikki ensikieliset tanskanpuhujat vastasivat olleensa vuorovaikutuksessa tanskanoppijoiden kanssa ainakin joskus. Kun ensikielisiltä tanskanpuhujilta kysyttiin kielenvaihtamisesta englantiin, 21,2 % kertoi, että ei koskaan vaihtanut englantiin, 27,3 % vaihtoi kielen ”erittäin harvoin” ja 14,4 % ”ei usein”. Vastaajista 31,7 % vastasi vaihtavansa englantiin ”joskus” tai ”melko usein”. Kukaan vastaajista ei valinnut kielenvaihtamiseen vaihtoehtoja ”melkein aina” tai ”aina”. (Jespersen & Hejná 2021.)

Ensikielisten tanskanpuhujien kuvaamia syitä kielenvaihtamiseen oli useita. Suurin syy kielenvaihtamiseen oli vuorovaikutuksen ylläpitäminen: Kieltä vaihdettiin, koska puhujat eivät pystyneet seuraamaan keskustelua (52 %) tai ei-ensikielinen tanskanpuhujia ei pystynyt seuraamaan keskustelua (28,4 %). Vastaajista 11,8 % pyrki kieltä vaihtamalla välttämään keskustelun katkeamista kokonaan ja 43,1 % halusi ylläpitää keskustelun vauhtia. Vain pieni osa (2,9 %) raportoi eksplisiittisesti vaihtavansa kieltä ei-ensikielisen tanskanpuhujan (heikon) kielitaidon tai aksentin vuoksi. Vastaajista 18,6 % vaihtoi kieltä kohteliaisuuden tai auttamisen halun vuoksi. Kun ensikielisten tanskankielisten piti arvioida, miten kielenvaihtaminen vaikuttaa ei-ensikielisiin tanskanpuhujiin, heistä 59,9 % arvioi, että englantiin vaihtaminen ei vaikuta ei-ensikieliseen keskustelukumppaniin negatiivisesti. Vain 2 % vastasi, että tanskanoppija

saattaa tuntea itsensä syrjityksi kielitaitonsa tai ulkonäkönsä perusteella. Vastaajista 4,4 % odotti negatiivista reaktiota, koska heillä on itsellään samanlaisia kokemuksia kielenvaihtamisesta ulkomailta. 20,8 % mainitsi, että viestin konteksti ja yksilöllisyys on tärkeintä, ja muutamat (4,5 %) eivät osanneet vastata kysymykseen ollenkaan. 8,9 % vastaajista odotti ei-ensikielisten tanskanpuhujien reagoivan positiivisesti, koska englantiin vaihtaminen helpottaa keskustelua tai on heidän mielestään kohteliaampaa. (Jespersen & Hejná 2021.)

3 Menetelmät ja aineisto

Tavoitteenamme oli selvittää ensikielisten suomenpuhujien kokemuksia arkisista keskusteluista ei-ensikielisten kanssa. Erityisesti tarkastelimme ensikielisten kielivalintoja ja kielenvaihtamista ja niille annettuja syitä. Tässä luvussa esittelemme aineistonkeruussa käytetyn kyselyn (alaluku 3.1), tutkimuksen osallistujat (alaluku 3.2), aineiston analyysin (alaluku 3.3) sekä tutkimuksen eettisiä pohdintoja (alaluku 3.4).

3.1 Kysely

Keräsimme tutkimuksen aineiston Webropol-kyselyllä keväällä 2022 (ks. Liite 1). Kyselyssä oli sekä monivalintakysymyksiä että avoimia kysymyksiä. Kysymykset keskityivät vastaajien kohtaamisiin ja keskustelukäyttäytymiseen arkisissa vapaissa puhetilanteissa ei-ensikielisten suomenpuhujien kanssa. Lisäksi kartoitimme keskustelussa tehtyjen kielivalintojen ja mahdollisen kielenvaihtamisen syitä sekä ensikielisten arvioita niiden aiheuttamista tunnereaktioista ei-ensikielisissä suomenpuhujissa. Kysymyksiä oli yhteensä 25, mutta todellisuudessa vastattavien kysymysten määrä riippui vastaajan valinnoista aiemmissä kysymyksissä (ks. Liite 1). Useimmissa kysymyksissä oli sekä valmiita vastausvaihtoehtoja että avoimia tekstikenttiä.

Esitämme tuloksissa ensikielisten suomenpuhujien vastauksia kysymyksiin arkisten puhetilanteiden kielivalinnoista, kielenvaihtamisesta ja niiden syistä. Tarkastelun kohteena on ensikielisten suomenpuhujien raportoima kielivalinta (suomi, englanti, muu) keskusteluissa heille ennestään ”tuntemattoman ei-äidinkielen” kanssa (ks. tarkemmin Liite 1). Kyselyssä ei annettu vastaajalle tarkempaa kontekstia tai tietoa puhujan ominaisuuksista, jotka olisivat voineet vaikuttaa vastauksiin. Lisäksi tutkimme sitä, kertovatko ensikieliset suomenpuhujat vaihtavansa arkisen puhetilanteen kielen englanniksi ja millaisia syitä he vaihtamiseen antavat. Kolmanneksi tarkastelemme ensikielisten suomenpuhujien oletuksia ei-ensikielisten tunnereaktioihin kielenvaihtamiseen arkisissa, vapaissa puhetilanteissa.

3.2 Osallistujat

Rajasimme kyselyn vastaajajoukoksi äidinkielenään suomea puhuvat aikuiset. Keräsimme vastaajilta seuraavat taustatiedot: ikä, sukupuoli, ylin koulutustaso, ammattiala sekä asuinmaakunta. Rekrytoimme vapaaehtoiset vastaajat julkisista Facebook-ryhmistä, joiden jäsenet olivat suurelta osin kielen tutkijoita tai kielistä kiinnostuneita maallikoita. Tämä tutkimus sai alkunsa yhdessä näistä ryhmistä käydyistä keskusteluista. Jaoimme kyselyä tarkoituksellisesti Facebookissa sellaisiin ryhmiin, joissa käsitellään kielen ja kielenkäyttöön liittyviä asioita, koska oletimme ryhmän jäsenten olevan kiinnostuneita vastaamaan tällaiseen kyselyyn. Vastaajia oli yhteensä 358.

Vastaajista 98,3 % oli ensikieltään suomea puhuvia, ja 1,7 % kaksikielisiä (suomi–muu kieli). Vastaajien keski-ikä oli 47,7 vuotta, keskihajonta 13,8 vuotta. Vastaajista 77,4 % oli naisia, 19,5 % miehiä, 2 % muita ja 1,1 % ei halunnut vastata. Yli 87 %:lla vastaajista oli korkeakoulututkinto (alempi tai ylempi korkeakoulututkinto tai lisensiaatin/tohtorin tutkinto). Vastaajista 44 % oli Uudeltamaalta, 21 % Varsinais-Suomesta, 11 % Pirkanmaalta ja 8 % Keski-Suomesta. Muut alueet olivat edustettuna hyvin pienillä osallistujamäärillä. Suurimmat ammattialat olivat kasvatus ja ohjaus (27,9 %), tietotyö ja hallinto (24,9 %) tai jokin muu ala (23,7 %).

Noin joka neljännellä osallistujalla (22,6 %) ei ollut yhtään muunkielistä työtoveria. Yli puolella osallistujista (58,1 %) näitä oli muutamia, 12 %:lla noin puolet ja 4,5 %:lla suurin osa. Vastaajista 2,8 %:lla lähes kaikki työkavereista olivat muun kuin suomenkielisiä. Muunkielisten kanssa puhuttiin työpaikoilla enimmäkseen suomea (49,8 %), mutta myös englantia (27,1 %), yhtä paljon suomea ja englantia (9,7 %), suomea ja jotain muuta kieltä (9 %) tai jotain muuta kieltä (4,3 %). Muiksi käytetyiksi kieliksi listattiin venäjä, ruotsi, ranska, puola, saksa, saamen kielet, espanja, italia, viro ja suomalainen viittomakieli.

Suurin osa vastaajista oli puhunut muunkielisen kanssa viimeisen 6 kk:n aikana. Vain 5,9 % ei ollut kertaakaan puhunut muun kuin ensikielisen suomenkielisen kanssa viimeisen puolen vuoden aikana. Päivittäin muunkielisen kanssa oli puhunut 10,6 %, viikoittain 25,4 %, kuukausittain 33,2 % ja harvemmin 24,9 %. Tuloksissa tarkastelemme vastauksia niiltä ensikielisiltä suomenpuhujilta, jotka vastasivat kyselyssä puhuneensa muiden kuin ensikielisen suomenkielisen kanssa viimeisten 6 kk:n aikana ($n = 342$).

3.3 Aineiston analyysi

Kyselyaineistoa tarkastelimme määrällisesti ja laadullisesti. Ensin kävimme läpi ensikielisten suomenpuhujien vastauksia monivalintakysymyksiin vastausten määriä tarkastellen. Kysymykset käsittelivät tuntemattoman ei-ensikielisen kanssa käytävän keskustelun aloittamiskieltä (suomi, englanti tai muu) sekä kyselyssä listattuja syitä aloittamiskielen valintaan. Tarkastelimme myös osallistujien kielenvaihtamista vastausten prosenttiosuuksien kautta. Syitä kielenvaihtamiseen arvioimme Likert-asteikollisten

kysymysten pohjalta. Vastajat kuvasivat samanmielisyyttä väittämiin asteikolla: eri mieltä – hieman eri mieltä – ei samaa eikä eri mieltä – samaa mieltä – täysin samaa mieltä. Osallistajat arvioivat myös ei-ensikielisten suomenpuhujien erilaisia reaktioita mahdolliseen kielenvaihtamiseen vapaissa puhetilanteissa aika-asteikolla (ei koskaan – harvoin – joskus – usein – aina). Vastajien oli mahdollista kertoa kielivalintojen ja kielenvaihtamisen syistä myös kyselyn avoimissa kysymyksissä. Näistä nousevia tärkeimpiä teemoja analysoimme laadullisen sisällönanalyysin (Krippendorf 2019) keinoin. Luimme ensin vastaukset läpi itsenäisesti ja luokittelimme niitä alustavasti. Tämän jälkeen luokittelimme vastauksissa nousseita teemoja yhdessä keskustellen, kunnes olimme luokituksesta samaa mieltä.

3.4 Tutkimuksen eettisyys

Tutkimuksessa noudatimme hyvän eettisen tutkimustavan mukaisia periaatteita, jotka on esitetty Tutkimuseettisen neuvottelukunnan ohjeissa (Tutkimuseettinen neuvottelukunta 2023). Osallistajat antoivat kirjallisen suostumuksensa tutkimukseen täyttämällä kyselyn. Osallistujilta ei kerätty henkilötietoja, kuten nimiä tai yhteystietoja, ja raportoimme tulokset siten, että yksittäistä vastaajaa ei voida tunnistaa.

4 Tulokset

Tässä luvussa esitämme tutkimuksen tulokset tutkimuskysymysten järjestystä seuraten. Ensin tarkastelemme ensikielisen suomenpuhujan ja oletetun ei-ensikielisen keskustelun aloittamiskieltä ja siihen annettuja syitä (alaluku 4.1). Toiseksi esittelemme ensikielisten suomenpuhujien vastauksia kysymykseen kielenvaihtamisesta vapaissa puhetilanteissa ja syitä vaihtamiseen (alaluku 4.2). Kolmanneksi käymme läpi ensikielisten suomenpuhujien arvioita ei-ensikielisten reaktioista keskustelun kielenvaihtamisesta suomesta englantiin (alaluku 4.3).

4.1 Keskustelun aloittamiskieli

Suurin osa (79 %) ensikielisistä suomenpuhujista kertoi aloittavansa keskustelun itselleen tuntemattoman ei-ensikielisen puhujan kanssa suomeksi. Englanniksi keskustelun raportoi aloittavansa 26,6 % vastaajista. Vain 5,3 % kertoi aloittavansa keskustelun itselleen tuntemattoman ei-ensikielisen puhujan kanssa muulla kielellä kuin suomeksi tai englanniksi.

Valmiiksi annetuista vaihtoehdoista suomeksi keskustelun aloittamisen syiksi nousivat suomi vastaajan asuinpaikan valtakielinä (73,4 %), suomi vastaajan äidin-kielänä (15,6 %), jokin muu syy (36,9 %) sekä suomen puhuminen joka tilanteessa

(1,1 %). Mahdollisia syitä puhetilanteen aloittamiseen suomeksi tarkastelimme myös avoimista vastauksista sisällönanalyysin keinoin. Monet ensikieliset korostivat vastauksissaan, että eivät voi tietää toisen kielitaitoa ulkonäön perusteella. Vastauksissa nousi usein esiin myös, että keskustelu aloitettiin suomeksi, koska vastaaja ei halua jakaa ihmisiä ”meihin” ja ”muihin”. Usealla oli myös oletuksena, että Suomessa yhteinen kieli on suomi, jolloin keskustelut aloitetaan suomeksi. Jotkut mainitsivat myös, että on kohteliaampaa aloittaa suomeksi. Joillekin keskustelun kieleksi valikoitui suomi, koska se on heidän oma ensikielensä.

Keskustelun englanniksi aloittamiselle ensikieliset suomenpuhujat valitsivat monivalinnassa seuraavia syitä: oletan, että ei-äidinkielen puhuu parempaa englantia kuin suomea (72,6 %), on kohteliaampaa puhua englantia kuin suomea (23,2 %), haluan päästä asiassa nopeasti eteenpäin (23,2 %) ja haluan itse harjoitella englantia (7,4 %). Avoimista vastauksista nousi sisällönanalyysissa esille useita syitä keskustelun aloittamiseksi englanniksi. Useat totesivat englannin olevan lingua franca ja kansainvälisesti käytetyin vieras kieli. Esiin tuotiin myös se, että ei-ensikieliset eivät puhu tai ymmärrä suomea tai että keskustelun toinen osapuoli aloittaa keskustelun itse englanniksi. Ensikieliset kertoivat myös aloittavansa keskustelun englanniksi tottumuksesta tai koska keskustelu on sujuvampaa englanniksi. Joidenkin vastausten mukaan myös ei-suomalainen olemus sai ensikielisen aloittamaan keskustelun englanniksi. Useissa vastauksissa heijastui ajatus siitä, että ulkomainen ihminen puhuu todennäköisemmin englantia kuin suomea.

5 % vastaajista raportoi aloittavansa keskustelun tuntemattoman ei-ensikielisen ihmisen kanssa muulla kielellä kuin suomeksi tai englanniksi. Monivalintakysymyksessä vastaajat mainitsivat useimmin syiksi keskustelun aloittamiseen muulla kielellä seuraavia seikkoja: puhun henkilön oletettua äidinkieltä (78,9 %), oletan toisen henkilön puhuvan tätä kieltä (47,4 %) ja haluan harjoitella tätä kieltä (42,1 %). Kuten muissakin monivalintakysymyksissä, vastaajat saivat valita useamman kuin yhden vastauksen. Avoimista vastauksista nousi sisällönanalyysissa erilaisia syitä keskustelun avaamiseksi muulla kielellä kuin suomeksi tai englanniksi. Keskustelu avattiin muulla kielellä silloin, kun ensikielinen puhuja tiesi toisen puhuvan jotain tiettyä kieltä tai kun hän itse puhui toisen ensikieltä. Myös keskustelun tilanne vaikutti kielivalintaan: usein keskustelu aloitettiin esimerkiksi ruotsiksi ruotsinkielisellä alueella. Myös skandinaavinen nimi sai ensikieliset joskus aloittamaan keskustelun ruotsiksi. Jos ensikielinen suomenpuhuja tiesi, että toinen ei puhu suomea tai englantia, hän kertoi aloittavansa keskustelun muulla (itse osaamallaan) kielellä. Muulla kielellä aloittamiseen vaikutti myös oma mahdollisuus harjoitella tätä kieltä.

4.2 Kielenvaihtaminen keskustelussa

Kun kysyimme ensikielisiltä suomenpuhujilta, olivatko he koskaan vaihtaneet suomeksi alkaneen keskustelun kieltä englanniksi, vain 4,5 % ei ollut koskaan vaihtanut kieltä suomesta englantiin. Toisaalta vain yksi vastaaja (0,3 %) vastasi aina vaihtavansa kielen suomesta englantiin keskustelussa ei-ensikielisen kanssa. Suurin osa vastaajista vaihtoi kielen suomesta englantiin joskus (55,3 %), noin viidesosa usein (20,9 %) ja noin viidesosa harvoin (19 %). Ensikielisten suomenpuhujien vastaukset väittämiin mahdollisista syistä englanniksi vaihtamiseen on esitetty taulukossa 1.

TAULUKKO 1. Vastaukset kielenvaihtamista englantiin koskeviin väittämiin (n = 342).

Väittämä (KA=keskiarvo, KH=keskihajonta)	1 = eri mieltä	2 =hie- man eri mieltä	3 = ei samaa eikä eri mieltä	4 = samaa mieltä	5 = täysin samaa mieltä
Haluan auttaa toista puhujaa, joten vaihdan englantiin. (KA=2,9, KH=1,2)	17 %	24 %	20,5 %	33,9 %	4,7 %
Haluan päästä nopeammin asiassa eteenpäin, joten vaihdan englantiin. (KA=2,7, KH=1,2)	24,3 %	22,8 %	19 %	30,1 %	3,8 %
En ajattele asiaa sen kummemmin, vaan kieli vaihtuu itsestään suomesta englanniksi. (KA=2,0, KH=1,2)	50,9 %	15,8 %	19,6 %	10,5 %	3,2 %
Oletan, että muualta tulleet eivät halua puhua suomea, joten vaihdan englantiin. (KA=1,6, KH=0,9)	62 %	22,5 %	10,8 %	4,4 %	0,3 %
Vaihdan suomesta englantiin, koska haluan harjoitella englantia. (KA=1,3, KH=0,8)	81,3 %	7,9 %	7 %	3,5 %	0,3 %
Huomaan, että puhujan aksentissa kuuluu vieras aksentti, joten vaihdan englantiin. (KA=1,2, KH=0,6)	83 %	12,3 %	3,5 %	1,2 %	0 %

Kysyimme ensikielisiltä suomenpuhujilta avoimessa kysymyksessä syitä siihen, miksi he olivat vaihtaneet keskustelun kielen suomesta englantiin. Karkeassa laadullisessa luokittelussa nousi esiin kolme pääluokkaa: 50 % vastaajista kertoi kielenvaihtamisen olleen yhteinen tai ei-ensikielisen päätös. Noin 15 %:ssa vastauksista kielenvaihtamisen aloitteen tekijä ei käy ilmi lainkaan. Noin 35 % vastaajista kertoi kielenvaihtamisen liittyvän omaan arvioonsa ei-ensikielisen suomen kielitaidosta.

Tarkemman laadullisen sisällönanalyysin pohjalta avoimista vastauksista nousi erilaisia teemoja kielenvaihtamisen syiksi. Usein syyksi annettiin ongelma vuorovai-
kutuksessa. Tällöin kielenvaihtaminen oli ratkaisuyritys ongelman selvittämiseen ja ymmärtämisen saavuttamiseen (ks. esimerkit 1–4). Vastaajan osallistujanumero on merkitty sulkeisiin esimerkin jälkeen. Esitämme esimerkit siinä muodossa, kun vastaaja on kirjoittanut ne kyselyyn.

- (1) Keskustelussa on tullut ongelmia tai keskustelukumppani on toivonut, että puhuisimme englantia. (5)
- (2) Koska asian ymmärtäminen on ollut keskeistä. (34)
- (3) Kanssapuhuja ei ole kyennyt kommunikoimaan suomeksi ja hänelle on ollut tärkeää voida selostaa asiansa riittävässä määrin luottaen siihen, että hänet ymmärretään oikein. (41)
- (4) emme ole ymmärtäneet toisiamme ja asia ei ole edennyt (61)

Asian hoitamisen tärkeyteen liittyvät myös vastaukset, joissa ensikieliset suomenpuhujat perustelivat kielenvaihtamista nopeudella tai sujuvuudella (ks. esimerkki 5). Tämän kääntöpuolena joissain vastauksissa esiintyi eksplisiittisesti puhujan oma laiskuus tai itsekkyyt, eli implisiittinen ajatus siitä, että asian hoitaminen suomeksi olisi hidasta tai epäsujua (ks. esimerkki 6).

- (5) Hän itse vaihtaa kieltä tai keskustelu olisi nopeampaa/helpompaa meille molemmille englanniksi. (97)
- (6) Itsekkäät syyt. Jos toinen puhuu sujuvasti englantia ja haparoiden suomea, on kommunikaatio nopeampaa ja helpompaa myös minulle englanniksi. Tunnen heti vahvaa empaattista epämukavuutta, jos näen toisen takeltelevan suomen kielen kanssa, ja joudun kamppailemaan tätä tunnetta vastaan, jotta toinen saa tilaisuuden harjoitella itselleen vaikeaa asiaa. Tunne on hyvin fyysinen, kiusallinen ja inhottava, ja joissakin tilanteissa olen itse liian väsynyt tai kärsimätön sietääkseni sitä. (188)

Useista vastauksista kävi myös ilmi, että ei-ensikielinen keskustelukumppani oli pyytänyt kielenvaihtamista tai oli itse vaihtanut kielen (ks. esimerkit 7–9). Esimerkeistä heijastuu myös tahaton kielenvaihtaminen, kun keskusteluun liittyy joku toinen henkilö, joka ei puhujan tiedon mukaan puhu suomea (ks. esimerkit 8–9).

130 KIELIVALINNAT JA KIELENVAIHTAMINEN ENSIKIELISTEN JA EI-ENSIKIELISTEN SUOMENPUHUJIEEN KESKUSTELUISSA – KYSELYTUTKIMUS

- (7) Keskustelukumppani on pyytänyt sitä tai olen halunnut kokeilla ymmärtäisimmekö toisiamme paremmin englanniksi. (19)
- (8) Toinen on ehdottanut vaihtoa tai vaihtanut itse kielen. Joskus olen epähuomiossa itse vaihtanut, kun seurueeseen on liittynyt joku joka ei osaa suomea (jolloin monenkeskiset keskustelut käydään englanniksi) ja olen sen jälkeen jatkanut kahdenkeskistä keskustelua ei-äidinkielen mutta suomea osaavan kanssa. (23)
- (9) Usein puhujan pyyntö. Kysyn usein, haluaako hän käydä keskustelun suomeksi vai englanniksi. Aika usein toivotaan englantia, jos keskustelukumppani on korkeakoulutettu. Töissä kieli vaihtelee tilanteen ja paikalla olevien mukaan. (64)

Ensikielisten suomenpuhujien avoimissa vastauksissa heijastui vahvasti kielenvaihtamisen tilanteisuus. Usein kieli vaihtuu Englantiin jonkin tietyn asian selvittämiseksi tai tietyn henkilön osallistuessa keskusteluun. Kun asia on selvitetty tai tietty henkilö lähtee keskustelusta pois, voidaan taas palata suomen kieleen (ks. esimerkit 10–11).

- (10) Keskustelukumppani on pyytänyt. Kieli vaihtui takaisin suomeksi, kun tietty asia saatiin selvitettyä. (108)
- (11) Jos minä olen vaihtanut, olen halunnut selittää epäselvän asian, mutta saattanut palata takaisin suomeen. Esimerkki: pappa kysyi kaupassa nuorelta mieheltä, mikä on tölkissä eräpäivä. Ymmärsi selkeästi vähän suomea muttei kysymystä. Sanoin sille "date" ja siitä se tajusi ja sanoi sen suomeksi papalle. Sanoin jälkikäteen suomeksi että hyvin meni. (205)

4.3 Ensikielisten suomenpuhujien arviot ei-ensikielisten suomenpuhujien reaktioista kielenvaihtamiseen

Kun kysyimme ensikielisiltä suomenpuhujilta, miten ei-ensikieliset olivat reagoineet vapaan keskustelun kielenvaihtamiseen suomesta englanniksi, he antoivat seuraavia vastauksia väittämiin (ks. Taulukko 2). Useimmat vastaajat tulkitsivat ei-ensikielisten reaktiot enemmän myönteisiksi kuin kielteisiksi. Heidän näkemyksensä mukaan vaihtaminen Englantiin oli vaikuttanut usein tai aina (54,9 %) ei-ensikieliseen siten, että asia oli ymmärretty paremmin. Ei-ensikielisen tulkittiin ilahtuneen kielenvaihdosta vähintäänkin joskus (44,4 %) tai usein/aina (26,9 %). Kielteisiksi ei-ensikielisten reaktiot tulkittiin vain harvoin. Vastaajien enemmistön (75,5 %) mielestä ei-ensikielinen ei hämmentynyt kielenvaihtamisesta koskaan tai hämmentyi vain joskus. Enemmistö vastaajista (86,5 %) raportoi, että ei-ensikielinen ei harmistunut koskaan tai harmistui vain joskus kielenvaihtamisesta.

Suuri osa ensikielisistä oletti, että ei-ensikieliset osaisivat Englantia. Vastaajista suurin osa (64,9 %) oli sitä mieltä, että tilanteita, joissa ei-ensikielinen ei ollut osannut puhua Englantia, oli ollut joko harvoin tai ei koskaan. Yli puolessa (59,6 %) vastauksista

oltiin sitä mieltä, että tilanteita, joissa ei-ensikielinen ei halunnut puhua englantia vaan suomea, oli tapahtunut harvoin tai ei koskaan. Väittämään keskustelun päättymisestä kielenvaihtamisen takia vastattiin, ettei sellaista ollut tapahtunut koskaan tai että sitä tapahtui vain harvoin (91 %).

TAULUKKO 2. Vastaukset väittämiin siitä, miten ei-ensikieliset reagoivat kielenvaihtamiseen suomesta englanniksi ($n = 342$).

Väittäjä: Jos olet itse vaihtanut keskustelutilanteessa suomesta englantiin, onko keskustelussa tapahtunut seuraavaa? (KA= keskiarvo, KH= keskihajonta)	1 = Ei kos- kaan	2 = Har- voin	3 = Joskus	4 = Usein	5 = Aina
Ei-äidinkielenen* näytti ymmärtävän asian paremmin (KA=3,5, KH=0,9)	6,7 %	2 %	36,3 %	48,5 %	6,4 %
Ei-äidinkielenen ilahtui (KA=2,9, KH=1,0)	12,3 %	16,4 %	44,4 %	24,9 %	2 %
Ei-äidinkielenen ei halunnut puhua englantia vaan suomea (KA=2,2, KH=1,0)	33,6 %	25,7 %	30,7 %	9,1 %	0,9 %
Ei-äidinkielenen ei osannut puhua englantia (KA=2,1, KH=0,9)	33,9 %	31 %	29,5 %	5,6 %	0 %
Ei-äidinkielenen hämmentyi (KA=1,9, KH=0,8)	36,6 %	38,9 %	22,5 %	1,8 %	0,6 %
Ei-äidinkielenen harmistui (KA=1,5, KH=0,8)	64,3 %	22,2 %	11,7 %	1,8 %	0 %
Keskustelu päättyi (KA=1,3, KH=1,3)	77,5 %	13,5 %	7,9 %	0,6 %	0,6 %

* Tekstissä käytetään termiä ei-ensikielinen

5 Pohdinta

Tutkimuksen tavoitteena oli tutkia ensikielisten suomenpuhujien kielivalintoja ja mahdollista kielenvaihtamista arjen vapaissa puhetilanteissa ennalta tuntemattomien, oletetusti ei-ensikielisten suomenpuhujien kanssa. Lisäksi tarkastelimme ensikielisten suomenpuhujien arvioita siitä, miten ei-ensikieliset reagoivat kielenvaihtamiseen.

Tutkimuksemme vastaajista selvä enemmistö, neljä viidestä, kertoi aloittavansa keskustelun suomeksi ja vain yksi neljästä englanniksi. Syy suomen kielellä aloittamiseen oli suurella osalla se, että vastaajien asuinpaikalla puhutaan suomea. Tanskalaisessa tutkimuksessa keskustelun ei-ensikielisen kanssa aina tanskaksi aloittavien ensikielisten osuus oli huomattavasti pienempi (noin 35 %) (Jespersen & Hejná 2021). Ei-ensikielististä tanskanpuhujista taas yli 60 % kertoi aloittavansa keskustelun ensikielisen kanssa ainakin joskus tanskaksi, eli tutkimuksessa piirtyi ristiriita ensikielisten ja ei-ensikielisten kokemusten välille (Jespersen & Hejná 2021).

Tämän tutkimuksen tulokset, eli että suurin osa ensikielististä raportoi aloittavansa keskustelun ei-ensikielisen kanssa suomeksi, ovat ristiriidassa aiemman suomalaisen tutkimuksen kanssa. Aiemman tutkimuksen mukaan ei-ensikieliset suomenpuhujat kokevat, että keskustelu aloitetaan usein englanniksi (ks. esim. Hippi & Lehto 2023; Ruuska 2020; Scotson 2020). Tällöin ei-ensikieliset tulevat ensikielisen tekemän kielivalinnan kautta kategorisoiduiksi ”muiksi”, koska esimerkiksi heidän aksenttinsa tai ulkonäkönsä ei vastaa ensikielisten odotuksia tyypillisestä suomalaisesta tai suomenpuhujasta (Hippi & Lehto 2023). Havaitsemamme ristiriita voi osaltaan heijastaa sitä, että vastaajat ovat kuvitelleet tietävänsä, miten tähän kysymykseen on haluttu vastattavan (Hawthorne-effekti, ks. esim. Mackey & Gass 2005) ja ovat vastanneet sen mukaisesti oman toimintansa kuvaamisen sijaan. Toisaalta tulokset voivat viitata siihen, että vastaajajoukko on keskimääräistä kielitietoisempää ja on pohtinut kielivalintojen merkitystä esimerkiksi kielenoppimiselle.

Tässä tutkimuksessa yksi neljästä ensikielisestä suomenpuhujasta aloitti keskustelun englanniksi, koska moni heistä oletti suoraan, että ei-ensikielinen suomenpuhujia puhuu parempaa englantia kuin suomea. On kiinnostavaa, että tällainen oletus tehtiin. Avoimissa vastauksissa tähän ei kuitenkaan löytynyt tarkempaa selitystä. Kyselyn kysymyksen asettelussa viitattiin ei-ensikieliseen ”tuntemattomana ei-äidinkielenä puhujana”, eli muotoilusta ei voinut päätellä syitä keskustelukumppanin ei-ensikielisyteen. Kuitenkin jokin seikka uudessa keskustelukumppanissa on saanut vastaajan aloittamaan keskustelun jollain muulla kielellä kuin suomella.

Yli puolet ensikielististä suomenpuhujista kertoi ainakin joskus vaihtaneensa keskustelun kielen suomesta englantiin. Ensikielististä tanskanpuhujista (Jespersen & Hejná 2021) näin teki vain noin joka kolmas. Tanskalaisessa tutkimuksessa (mt.) samaa asiaa kysyttiin myös ei-ensikielisiltä: heistä yli puolet kertoi, että ensikielinen oli vaihtanut keskustelun kielen tanskasta englanniksi. Tietääksemme samankaltaista kartoitettavaa tutkimusta, jossa tarkastellaan ei-ensikielisten suomenpuhujien käsityksiä kielenvaihtamisesta keskusteluissa, ei ole. Laadullista tutkimusta aiheesta on hieman. Esimerkiksi Hipin ja Lehdon (2023) haastattelututkimuksessa ei-ensikielisillä oli kokemus siitä, että keskusteluissa ensikielisten kanssa kieli vaihtuu englanniksi. Myös muissa suomalaisissa tutkimuksissa ensikieliset raportoivat keskustelun kielen vaihtuvan usein suomesta englanniksi, vaikka heidän kielitaitonsa olisi edistyneellä tasolla (Hippi & Lehto 2023; Scotson 2020; Ruuska 2020).

Tutkimuksemme vastaajista enemmistö oli sitä mieltä, että kielenvaihtamisen syynä on halu toisen auttamiseen ja toisaalta vuorovaikutuksen sujuvuus ja halu päästä keskustelussa eteenpäin. Osalla vastaajista kieli vaihtui itsekseen ja osa piti englantiin vaihtamista kohteliaisuutena, koska oletuksena oli ei-ensikielisen parempi englannin kuin suomen taito. Lisäksi pieni joukko vastaajista oli valinnut kielenvaihtamiselle annetuista vaihtoehdoista itsestään lähtöisen syyn päästä harjoittelemaan englantia tai oletuksen siitä, että muualta tullut ei halua puhua suomea. Nämä tulokset ovat samansuuntaisia kuin Jespersenin ja Hejnán (2021) tanskalaisessa tutkimuksessa. Ensikieliset tanskanpuhujat vaihtoivat kielen englantiin kohteliaisuudesta tai auttamisen halusta tai välttääkseen keskustelun katkeamisen. Myös Lehdon (2023a) ja Scotsonin (2020) tutkimuksessa englantia käytettiin ensikielisten ja kielenoppijoiden keskusteluissa ikään kuin pakon sanelemana tilanteissa, joissa vuorovaikutus olisi muuten täysin estynyt. Jespersenin ja Hejnán (2021) sekä Scotsonin (2020) mukaan englannin käyttö voi liittyä erityisesti aikapaineeseen ja tehokkuuteen: asia on hoidettava kielellä millä hyvänsä.

Tässä tutkimuksessa vain muutama vastaaja kertoi vaihtavansa kielen muualta tulleen aksentin vuoksi eikä juuri kukaan nostanut kielenvaihtamisen syyksi eksplisiittisesti ulkonäköä. Tämä on linjassa tanskalaisen tutkimuksen kanssa: vain harva ensikielisistä nosti kielenvaihtamisen syyksi ei-ensikielisen aksentin, ulkonäköä ei juurikaan mainittu (Jespersen & Hejná 2021). Tulokset voivat heijastaa myös kyselyä aineistonkeruumenetelmänä. Kielenalinta tai -vaihtaminen ulkonäön tai aksentin vuoksi voi olla tiedostamatonta, minkä vuoksi siihen voi olla vaikea päästä käsiksi kyselyaineiston kautta. Se, että ensikieliset eivät nosta ulkonäköä tai aksenttia syyksi kielenvaihtamiseen, on kuitenkin ristiriidassa ei-ensikielisten kokemusten kanssa: toiseuttavan kielenvaihtamisen syiksi on tulkittu esimerkiksi oma ulkonäkö tai aksentti (Hippi & Lehto 2023). Kielenvaihtamisen syyt voivat toki olla myös tilannesidonnaisia: työkontekstissa aikapaine ja asian hoitaminen tehokkaasti ovat tärkeämmässä roolissa kuin vapaamuotoisemmassa kohtaamisessa (ks. tästä enemmän Hippi & Lehto 2023).

Ensikielisten antamat syyt kielivalinnoille ja kielenvaihtamiselle herättävät kysymyksiä siitä, miten he hahmottavat keskustelun merkityksiä. Varsinkin avoimissa vastauksissa kielivalintoja ja kielenvaihtamista selitettiin usein viestin ymmärtämisellä, eli keskustelu nähtiin mahdollisesti puhtaasti viestin välittämisenä ja kieli vain välineenä. Kuitenkin ihminen oppii (myös kieltä) vuorovaikutuksessa toisten kanssa (Vygotsky 1978), jolloin ensikielisillä voisi keskusteluissa ei-ensikielisten kanssa olla kielenoppimisen mahdollistajan rooli. Sen sijaan ensikieliset perustelevat englannin käyttöä keskustelussa sillä, että he voivat tällöin itse harjoitella englantia. Omaa roolia ei-ensikielisen suomenpuhujan kielenoppimisen tukena ei vastauksissa kuitenkaan ole useinkaan ymmärretty.

Kun kysyimme ensikielisiltä, miten he arvelivat ei-ensikielisten reagoivan kielenvaihtamiseen suomesta englantiin, suurin osa (yli 70 %) ensikielisistä ajatteli, että kielenvaihtaminen suomesta englantiin ilahduttaa ei-ensikielistä suomenpuhujaa.

Sen sijaan ensikieliset tulkitsivat ei-ensikielisen harmistuvan tai hämmentyvän kielen vaihtamisesta vain harvoin. Tämä tulos on lähellä tanskalaisessa tutkimuksessa esitettyä: noin 60 % ensikielisistä arvioi, että kielenvaihtaminen ei aiheuta negatiivisia reaktioita ei-ensikielisessä tanskan puhujassa (Jespersen & Hejná 2021). Sen sijaan Hipin ja Lehdon tutkimuksessa ei-ensikieliset haastateltavat nostavat esiin, että kielen vaihtaminen englanniksi kategorisoi heidät ei-suomenkielenpuhujiksi, mikä herätti heissä negatiivisia tunteita (Hippi & Lehto 2023). Myös tanskalaisten (Jespersen & Hejná 2021) tutkimuksessa noin 60 % ei-ensikielisistä ilmaisi reagoivansa kielteisesti kielenvaihtamiseen ja vain harva tulkitsi sen kohteliaisuudeksi tai auttamishaluksi. Kielteiseksi koettua kielenvaihtamista ei-ensikieliset pitivät epäkohteliaana, ja se teki heidät jopa vihaiseksi (Jespersen & Hejná 2021).

Onnistunut vuorovaikutus vaatisi ensikielisiltä ei-ensikielisen reaktioiden parempaa havaitsemista: kun ei-ensikielisen suomenpuhujan ei anneta käyttää suomea keskustelussa, hänen osallisuutensa suomenkieliseen yhteisöön vaikeutuu. Vastapuolen reaktioiden ulossulkeminen vuorovaikutuksesta voi myös johtaa jännitteisiin, jotka voivat haitata esimerkiksi asiakaspalvelutilannetta ja johtaa esimerkiksi siihen, että ei-ensikielinen ei asioi yrityksessä uudestaan (Hippi & Lehto 2023). Tarvitaankin lisää autenttisia keskustelutilanteita tarkastelevaa tutkimusta siitä, miten ensikieliset havainnoivat ei-ensikielisten reaktioita keskustelujen kielivalintoihin ja kielenvaihtamiseen.

Kun ei-ensikielinen suomenpuhija aloittaa keskustelun suomeksi, hän positioi itsensä suomenoppijaksi. Scotsonin tutkimuksessa (2018: 52) korkeakoulutetut maahanmuuttajat näkivät sisällöltään vaativamman vuorovaikutustilanteen tarjoumana, eli mahdollisuutena oppia suomea, ja positioivat itsensä suomenoppijaksi ja samalla vuorovaikutuskumppanin kielenoppimisen tukijaksi. Toisaalta asiasisällön muuttuessa monimutkaisemmaksi monet Scotsonin haastateltavat kuitenkin vaihtoivat kielen englanniksi, koska he pitivät sen käyttämistä helpompana ja nopeampana (mt.). Tässä tutkimuksessa myös ensikieliset suomenpuhujat perustelivat englantiin vaihtamista tehokkuudella ja asian hoitamisen ensisijaisuudella.

Olisi tärkeää pohtia, miten ensikieliset suomenpuhujat voisivat toimia suomenoppijoiden, tai tarkemmin sanoen suomenoppijan position ottaneiden, kielenoppimisen tukijoina silloin, kun vuorovaikutustilanteet ja niiden sisällöt monimutkaistuvat. Siitä, että kyselyymme vastanneista suuri osa tulkitsi ei-ensikielisten suomenpuhujien ilahtuneen keskustelukielen vaihtuessa englantiin, ilmenee, että kielenoppimisen tukeminen ei oikeastaan tule huomioiduksi vuorovaikutustilanteissa. Tasa-arvoisen keskustelukulttuurin kannalta olennaista olisi, että ensikieliset tiedostaisivat ei-ensikielisten tarvitsevan tukea eri vuorovaikutustilanteissa ja että ensikielisten kommunikointiin eli sanalliseen viestintään välittämiseen keskittyvä ja yksipuolinen kielenvaihtaminen ei aina vaikuta positiivisesti – vuorovaikutuksen oletetusta sujuvoitumisesta huolimatta.

Ei-ensikieliselle keskustelun käyminen joko suomeksi tai englanniksi ei välttämättä aina ole paras vaihtoehto. Lehto (2023a) tuo esille, että useat monikieliset suomenoppijat kokevat keskustelun yksikielisyyden vaatimuksen painostavana:

usein olisi helpompaa käyttää useampaa kieltä. Työtilanteissa englannin valinta voi olla perusteltua, sillä joskus työntekijän ja kielenoppijan rooleja on vaikea sovittaa toisiinsa (ks. Kotilainen ym. 2022). Se, että ensikieliset suomenpuhujat vaihtavat keskustelun kielen yksipuolisesti englanniksi, neuvottelematta ei-ensikielisen kanssa, voi kuitenkin olla ongelmallista: jos ei-ensikielisen suomeksi aloittama keskustelu vaihtuu englanninkieliseksi kerta toisensa jälkeen, hän voi pahimmillaan lakata yrittämästä kommunikoida suomeksi (Ruuska 2020).

Tässä tutkimuksessa ensikielisten suomenpuhujien vastaukset monivalintakysymyksiin ja avoimiin kysymyksiin olivat ristiriidassa keskenään. Avointen kysymysten vastausten perusteella vaikuttaa siltä, että monet ensikieliset suomenpuhujat eivät tunnista kielenvaihtamisen monimerkityksellisyyttä ja sen kautta tapahtuvaa ei-ensikielisen kategorisointia ”muiksi” (ks. Hippi & Lehto 2023). Olisikin tärkeää, että ensikieliset suomenpuhujat tulisivat tietoisiksi siitä, kuinka tärkeää suomen kielen käyttömahdollisuudet ovat suomenoppijoille – sekä kielenoppimisen että osallisuuden ja yhdenvertaisuuden näkökulmista. Myös Ruuska (2020) peräänkuuluttaa kieliyhteisöihin tietoisuutta normeista ja odotuksista, jotka voivat joko estää tai edistää ei-ensikielisten suomenpuhujien osallisuutta ja kokemusta kuuluvuudesta suomalaiseen yhteiskuntaan.

Tämän tutkimuksen vastaajajoukko oli jossain määrin valikoitunut: suurimmalla osalla vastaajista oli vähintään alempi korkeakoulututkinto. Vastaajien englannin kielitaito oli todennäköisesti hyvä, ja heille oli luontevaa käyttää englantia. Yleinen kielitietoisuus oli myös todennäköisesti korkeampi kuin väestössä yleisesti: Avoimista vastauksista käy ilmi, että näitä asioita on pohdittu jo aiemmin. Toisaalta osa vastauksista antaa ymmärtää, että kielitietoisuuden sisältöä ei välttämättä ole täydellisesti ymmärretty (ks. määritelmästä esim. Suuriniemi 2023). Avointen vastausten perusteella jotkut osallistujat todennäköisesti myös opettavat suomea. Voidaan kenties myös olettaa, että useimmat vastaajista ovat opiskelleet suomen tai ruotsin lisäksi vähintään kahta kieltä.

Tulosten tulkinnan kannalta olisi ollut hyvä, jos vastaajilta olisi kysytty, ovatko he kiinnostuneita kielenoppimisesta. Tutkimuksen tulokset ovat siten suuntaa antavia ja tarvitaankin lisää tutkimusta, jossa osallistujien taustat ovat moninaisemmat, jotta saadaan kokonaisvaltaisempi näkemys aiheesta. Lisäksi tarvitaan vertailtavaa tietoa ei-ensikielisten kokemuksista kielenkäytöstä keskusteluissa sekä tehdyistä kielivalinnoista ja kielenvaihtamisesta. Tasku-hankkeessa onkin seuraavaksi tekeillä ei-ensikielisille suomenpuhujille suunnattu kysely suomen käytöstä sekä laajempi kysely ensikielisille suomenpuhujille. Laajemman kartoituksen lisäksi tarvitaan erityisesti systemaattista etnografista näyttöä. Tulevaisuudessa tulisi tarkastella myös ulkonäön mahdollista vaikutusta ensikielisten suomenpuhujien kielivalintoihin ja kielenvaihtamiseen. Tätä voitaisiin tutkia esimerkiksi tekoälyn avulla.

Ensikielisten suomenpuhujien on hyvä ymmärtää suomen kielen käytön mahdollisuuksien merkitys ei-ensikielisten kielenoppimiselle, yhdenvertaisuudelle keskusteluissa

sekä osallisuudelle suomalaisen yhteiskuntaan. Englannin käyttö vuorovaikutuksessa ei itsestään selvästi lisää tasa-arvoa (Lehto 2023b). Ensikielisten ja ei-ensikielisten keskusteluissa tasa-arvo voi esimerkiksi toteutua niin, että ei-ensikielinen suomenpuhuja ei tule arvioiduksi tai kategorisoiduksi ulkonäkönsä tai kielitaitonsa perusteella, vaan saa mahdollisuuden olla mukana neuvottelemassa keskustelun kielestä (ks. myös Lehto 2023a; Hippi & Lehto 2023; Scotson 2020).

Tämän tutkimuksen tulosten pohjalta vahvistuu tarve tehdä molemminpuolisia sopimuksia yhteisten keskustelujen kielivalinnoista ja kielenvaihtamisesta ensikielisten ja ei-ensikielisten suomenpuhujien välillä (ks. myös Lehto 2023a; Hippi & Lehto 2023; Scotson 2020).

Kirjallisuus

- Benediktsson, A. I. 2015. *Det sproglige sammenstød. En undersøgelse af sproglige problemer i kommunikationen mellem udlændinge og danskere i Danmark og udlændinge og islændinge i Island*. Pro gradu -tutkielma. Hugvísindasviði Háskóla Íslands.
- Education First 2021. *EF English Proficiency Index 2021*. <https://www.ef.com/assetscdn/WIBlwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2021/ef-epi-2021-english.pdf> [luettu 27.2.2024]
- Ethnologue 2023. *The most spoken languages worldwide in 2023 (by speakers in millions)*. <https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/> [luettu 21.3.2024]
- Hippi, K. & L.-M. Lehto 2023. Seeking understanding: categories of linguistic (non)belonging in interviews. *Multilingua*, 42 (4), 473–498. <https://doi.org/10.1515/multi-2022-0049>
- likkanen, P. 2017. The use of language in migrant stay-at-home parents' process of integration. Experiences of inclusion and exclusion. *Apples – Journal of Applied Language Studies*, 11 (3), 121–142. <https://doi.org/10.17011/apples/urn.201712104587>
- Intke-Hernandéz, M. 2020. *Maahanmuuttajaäitien arjen kielitarinat. Etnografinen tutkimus kieliyhteyksien sosiaalistumisesta*. Helsinki: Helsingin yliopisto. <http://hdl.handle.net/10138/321636>
- Jespersen, A. B. & A. Hejná 2021. Second language learners of Danish as the linguistic other. *Otherness: Essays & Studies*, 8 (3), 129–161. https://otherness.dk/fileadmin/www.othernessandthearts.org/Publications/Journal_Otherness/Otherness_Essays_and_Studies_8.3/Second_Language_Learners_of_Danish_as_the_Linguistic_Other_-_Jespersen_and_Hejna__with_abstract_.pdf
- Kim, J., M. Toivola, L. M. Heikkola & A. Saloranta 2022. Saisinko puhua suomea? *Kielikello* 3/2022. <https://www.kielikello.fi/-/saisinko-puhua-suomea>
- Kotilainen, L. 2013. Älä englannita. *Suomen kuvalehti* 23.3.2013. <https://suomenkuvalehti.fi/jutut/kotimaa/mielipide-kotimaa/ala-englannita/> [luettu 10.11.2021]
- Kotilainen, L., S. Kurhila & I. Lehtimaja 2022. Monikielinen työpaikka – tehokkuuden, vaivattomuuden vai kielenoppimisen ehdoilla? *Kielikello* 1/2022.
- Kotilainen, L. & I. Lehtimaja 2019. Kielenvaihto monikielisissä kokouksissa: osallistujien kielitaito ja vuorovaikutuksen sujuva eteneminen. *Puhe ja kieli*, 39 (3), 221–239. <https://doi.org/10.23997/pk.76602>
- Kotimaisten kielten keskus 2021a. *Kielellisen tasa-arvon puolesta*. https://www.kotus.fi/nyt/kotus-vinkit/kiellellisen_tasa-arvon_puolesta.36726.news [luettu 10.11.2021]
- Kotimaisten kielten keskus 2021b. *Käytä näköistäsi suomea*. https://www.kotus.fi/nyt/uutistekstit/kotuksen_uutiset/kotuksen_uutiset_2021/kayta_nakoistasi_suomea.36760.news?35496_o=3 [luettu 10.11.2021]
- Krippendorff, K. 2019. *Content analysis. An introduction to its methodology*. Thousand Oaks: SAGE Publications, Inc.
- Kurhila, S. K. (2001). Correction in talk between native and non-native speaker. *Journal of Pragmatics*, 33, 1083–1110.
- Laitinen, M., S. Leppänen, P. Rautionaho & S. Backman 2023. *Englanti Suomen kansalliskielten rinnalla: kohti joustavaa monikielisyyttä*. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 2023:59. Helsinki: Valtioneuvoston kanslia. <http://urn.fi/URN:ISBN:978-952-383-055-4>

- Leemann, L., H. Kuusio & R.-M. Hämäläinen 2015. *Sosiaalinen osallisuus*. Sosiaalisen osallisuuden edistämisen koordinaatiohanke (Sokra). Helsinki: Terveiden ja hyvinvoinnin laitos. https://thl.fi/documents/155392151/190140732/Tietopaketti_Sosiaalinen_Osallisuus.pdf [luettu 14.2.2023]
- Lehtimaja, I. 2017. Korkeakoulutetun maahanmuuttajan oikeus oppia suomea. *Kieli, koulutus ja yhteiskunta*, 8 (5). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-lokakuu-2017-2/korkeakoulutetun-maahanmuuttajan-oikeus-oppia-suomea>
- Lehtimaja, I., L. Kotilainen & S. Kurhila 2021. Monikielisyyden haasteet työyhteisössä. *Työelämän tutkimus*, 19 (3), 453–463. <https://journal.fi/tyoelamantutkimus/article/view/109950/65403>
- Lehto, L.-M. 2023a. Englanti kielivalintana Suomessa asuvien ulkomaalaistaustaisten henkilöiden diskursseissa. *Virittäjä*, 127 (3). <https://doi.org/10.23982/vir.107906>
- Lehto, L.-M. 2023b. Transnationals' discourses on the English language in Finland. Teoksessa M. Frick, T. Räisänen & J. Ylikoski (toim.) *Language contacts and discourses in the Far North*. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-031-42979-8_7
- Leppänen, S., A. Pitkänen-Huhta, T. Nikula, S. Kytölä, T. Törmäkangas, K. Nissinen, L. Kääntä, T. Räisänen, M. Laitinen, H. Koskela, S. Lähdesmäki & H. Jousmäki 2011. National survey on the English language in Finland: uses, meanings and attitudes. *Studies in variation, contacts and change in English*, 5. <https://varieng.helsinki.fi/series/volumes/05/>
- Lilja, N. 2010. *Ongelmista oppimiseen: toisen aloittamat korjausjaksot kakkoskielisessä keskustelussa*. Jyväskylä: Jyväskylän yliopisto. <http://urn.fi/URN:ISBN:978-951-39-4116-1>
- Lilja, N., L. Eilola, A.-K. Jokipohja & T. Tapaninen (toim.) 2022. *Aikuiset maahanmuuttajat arjen vuorovaikutustilanteissa: suomen kielen oppimisen mahdollisuudet ja mahdottomuudet*. Tampere: Vastapaino.
- Mackey A. & S. Gass 2005. *Second language research: methodology and design*. Mahwah: Lawrence Erlbaum Associates.
- Ruuska, K. 2020. *At the nexus of language identity and ideology: becoming and being a highly proficient second language speaker of Finnish*. Jyväskylä: Jyväskylän yliopisto. <https://jyx.jyu.fi/handle/123456789/72334>
- Scotson, M. 2018. Korkeakoulutettujen maahanmuuttajien kielivalinnat vuorovaikutuksessa. *Prologi: puheviestinnän vuosikirja*, 14 (1), 44–59. <https://doi.org/10.33352/prlg.95929>
- Scotson, M. 2020. *Korkeakoulutetut kotoutujat suomen kielen käyttäjinä: toimijuus, tunteet ja käsitykset*. Jyväskylä: Jyväskylän yliopisto. <https://jyx.jyu.fi/handle/123456789/71523>
- Suni, M. 2008. *Toista kieltä vuorovaikutuksessa: kielellisten resurssien jakaminen toisen kielen omaksumisen alkuvaiheessa*. Jyväskylä: Jyväskylän yliopisto. <https://jyx.jyu.fi/handle/123456789/18556>
- Suuriniemi, S.-M. 2023. *Monikielistyvä koulu ja kielitietoisuus: perusopetuksen opetus-suunnitelmien, opettajien ja oppikirjojen kieliorientaatiot*. Helsinki: Helsingin yliopisto. <http://urn.fi/URN:ISBN:978-951-51-9412-1>
- Tilastokeskus 2023. *Väestörakenne 2022. Vieraskielisten määrä kasvoi lähes 38 000 henkilöllä*. <https://stat.fi/julkaisu/cl8lprraorrr20dut5a0tywm5> [luettu 14.2.2023]
- Tutkimuseettinen neuvottelukunta 2023. *Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa*. Tutkimuseettisen neuvottelukunnan julkaisuja 2/2023. Helsinki: Tutkimuseettinen neuvottelukunta. https://tenk.fi/sites/default/files/2023-03/HTK-ohje_2023.pdf [luettu 14.2.2023]
- van Lier, L. 2000. From input to affordance: Social-interactive learning from an ecological perspective. Teoksessa J. P. Lantolf (toim.) *Sociocultural theory and second language learning*. Oxford: Oxford University Press, 245–259.

- Vitanova, G., Miller, E. R., Gao, X. & Deters, P. 2015. Introduction to theorizing and analyzing agency in second language learning: Interdisciplinary approaches. Teoksessa P. Deters, X. Gao, E. R. Miller & G. Vitanova (toim.) *Theorizing and analyzing agency in second language learning: Interdisciplinary approaches*. Bristol, UK: Multilingual Matters, 1–13.
- Vygotsky, L. S. 1978. *Mind in society: development of higher psychological processes* (toim. M. Cole, V. Jolm-Steiner, S. Scribner, E. Souberman). Cambridge: Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>

LIITE 1: Kyselyn esittely

Valmiit vastausvaihtoehdot on esitelty kysymyksen alla. Mikäli vaihtoehtoja ei ole, on kysymykseen pyydetty avoin vastaus.

Kysely alkoi kysymyksillä vastaajan mahdollisen työpaikan kielikäytännöistä:

- *Työtovereinani on äidinkielenään muuta kuin suomea puhuvia:*
 - » *ei yhtään, muutamia, noin puolet, suurin osa, lähes kaikki*
- *Mitä kieltä/kieliä puhutte muuta kuin suomea äidinkielenään puhuvien työtoveriesi kanssa työpaikalla?*

Tämän jälkeen alkoi osuus liittyen vastaajan keskustelunavauksiin ei-äidinkielisten suomenpuhujien kanssa:

- *Kuinka usein olet keskustellut itsellesi tuntemattoman ei-äidinkielisten kanssa viimeisten 6 kk:n aikana? Keskustelulla tarkoitamme tässä lyhyttä tai pitempää vuorovaikutustilannetta. Lyhimmillään kyseessä voi olla esim. tervehdyksen vaihto, neuvominen (esim. missä on lähin rautatieasema?) tai asiakastilanne.*
- *Kuvittele mielessäsi tapaaminen itsellesi tuntemattoman ei-äidinkielisten puhujan kanssa. Millä kielellä aloitat keskustelun?*
- *Aloitatko keskustelun itsellesi tuntemattoman ei-äidinkielisten puhujan kanssa:*
 - » *suomeksi, englanniksi, muulla kielellä, millä?*

Näiden pohjustuskysymysten jälkeen siirryttiin kielikohtaisiin kysymyksiin, jotka olivat samanlaisia kielten välillä ja jotka vastaajat näkivät vain, mikäli olivat valinneet aloittavansa keskustelun kyseisellä kielellä.

- *Miksi aloitat keskustelun suomeksi/englanniksi/muulla kielellä?*
- *Jos aloitat keskustelun ei-äidinkielenään suomea puhuvan kanssa suomeksi, mitkä seuraavista vaihtoehdoista sopivat sinuun parhaiten? (voit valita yhden tai useamman):*
 - » *Aloitin keskustelun suomeksi, koska puhun vain suomea.*
 - » *Aloitin keskustelun suomeksi, koska äidinkieleni on suomi.*
 - » *Aloitin keskustelun suomeksi, koska asuinpaikkani valtakielinen on suomi.*
 - » *Aloitin keskustelun suomeksi jostain muusta syystä.*
- *Jos aloitat keskustelun ei-äidinkielenään suomea puhuvan kanssa englanniksi, mitkä seuraavista vaihtoehdoista sopivat sinuun parhaiten? (voit valita yhden tai useamman):*
 - » *Aloitin keskustelun englanniksi, koska ei-äidinkielenen puhuja ei näytä suomalaiselta.*
 - » *Aloitin keskustelun englanniksi, koska on kohteliaampaa puhua englantia kuin suomea.*
 - » *Aloitin keskustelun englanniksi, koska haluan harjoitella omaa englantiani.*
 - » *Aloitin keskustelun englanniksi, koska haluan päästä nopeasti asiassa eteenpäin.*
 - » *Aloitin keskustelun englanniksi, koska oletan että ei-äidinkielenen puhuu parempaa englantia kuin suomea.*
- *Jos aloitat keskustelun ei-äidinkielenään suomea puhuvan kanssa muulla kielellä, mitkä seuraavista vaihtoehdoista sopivat sinuun parhaiten? (voit valita yhden tai useamman)*

- » *Aloitaa keskustelun muulla kielellä, koska puhun toisen henkilön oletettua äidinkieltä.*
- » *Aloitaa keskustelun muulla kielellä, koska oletan toisen henkilön puhuvan tätä kieltä.*
- » *Aloitaa keskustelun muulla kielellä, koska haluan harjoitella tätä kieltä.*

Keskustelunavauskysymysten jälkeen siirryttiin käsittelemään kielen vaihtamista tai vaihtumista Englantiin keskustelun aikana.

- *Jos keskustelu tuntemattoman ei-äidinkielen kanssa on alkanut suomeksi, oletko vaihtanut kielen suomesta Englantiin?*
- *en koskaan/en ole keskustellut ei-äidinkielen kanssa, harvoin, joskus, usein, aina*
- *Mikä on ollut syy siihen, että olet vaihtanut kielen Englantiin?*

Mikäli vastaaja ei ollut koskaan vaihtanut Englantiin tai keskustellut ei-äidinkielen kanssa, häntä pyydettiin tämän kysymyksen kohdalla kuvittelemaan tilanne, jossa olisi kuitenkin toiminut näin. Tällöin kysymyksessä oli samat vastausvaihtoehdot ja alustus: *”Kuvittele tilanne, jossa jostain syystä vaihdat keskustelun kielen suomesta Englantiin”.*

- *Lue seuraavat väittämät. Oletko samaa vai eri mieltä (eri mieltä, hieman eri mieltä, ei eri eikä samaa mieltä, samaa mieltä, täysin samaa mieltä)?*
 - » *Mielestäni on kohteliaampaa vaihtaa kieli englanniksi kuin jatkaa suomeksi, koska toinen luultavasti osaa paremmin Englantia kuin suomea.*
 - » *Haluan auttaa toista puhujaa, joten vaihdan Englantiin.*
 - » *Oletan, että muualta tulleet eivät halua puhua suomea, joten vaihdan Englantiin.*
 - » *Haluan päästä nopeammin asiassa eteenpäin, joten vaihdan Englantiin.*
 - » *Huomaan, että puhujan suomessa kuuluu vieras aksentti, joten vaihdan Englantiin.*
 - » *Vaihdan suomesta Englantiin, koska haluan harjoitella Englantia.*
 - » *En ajattele asiaa sen kummemmin, vaan kieli vaihtuu itsestään suomesta englanniksi.*

Kyselyn lopuksi vastaaja pyydettiin arvioimaan, miten kielen vaihtaminen oli vaikuttanut keskustelutilanteeseen ja miten keskustelukumppani oli reagoanut vaihtoon.

- *Jos olet itse vaihtanut keskustelutilanteessa suomen kielen Englantiin, miten ei-äidinkielen puhuja on yleensä reagoanut kielenvaihtoon?*
- *Jos olet itse vaihtanut keskustelutilanteessa suomen kielen Englantiin, onko keskustelussa tapahtunut seuraavaa (ei koskaan, harvoin, joskus, usein, aina):*
 - » *ei-äidinkielen ilahtui*
 - » *ei-äidinkielen näytti siltä, että ymmärsi asian paremmin*
 - » *ei-äidinkielen hämmentyi*
 - » *ei-äidinkielen ei osannut puhua Englantia*
 - » *ei-äidinkielen ei halunnut puhua Englantia vaan suomea*
 - » *ei-äidinkielen harmistui*
 - » *keskustelu päättyi*
- *Kerro vapaasti omista kokemuksistasi edelliseen kysymykseen liittyen.*

Osa III

Toivola, M., P. Lintunen & L. M. Heikkola (toim.) 2024. Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. *New directions in speech research – Freely produced speech as data.* AFinLA-teema / n:o 17, 143–166.

Pauliina Peltonen & Pekka Lintunen
University of Turku

Word searches and fluency during problem-solving tasks: a mixed-methods study of L1 Finnish and L2 English interaction

Highlights

- L1 Finnish dialogues contained more solitary word searches (WSs) than L2 English dialogues.
- Correlations across L1 Finnish and L2 English WSs were weak and statistically nonsignificant.
- Collaborative WSs, in particular, contributed to maintaining interactional fluency.

Abstract

Word searches (WSs) have been widely studied in L1 and L2 interaction but have rarely been combined with fluency research. The present study bridges the gap between these research fields by examining WSs and their connections to fluency in L1 and L2 dialogue data. From a larger project, 50 participants' dialogues (25 L1 Finnish, 25 L2 English) were analyzed for WSs. The mixed-methods analyses focused on potential differences and connections in WS use across the L1 and L2 dialogues and the connections between WSs and fluency among selected participants. The results demonstrated differences especially in the use of solitary WSs across L1 and L2 and some connections between L1 and L2 WSs among individual participants, despite group-level connections not being statistically significant. The qualitative analyses illustrated the potential fluency-enhancing functions of collaborative WSs. The study has implications for L2 speech fluency research and L2 teaching.

Keywords: fluency, interaction, problem-solving, word search
Asiasanat: sujuvuus, vuorovaikutus, ongelmanratkaisu, sanahaku

1 Introduction

Fluency is a key indicator of second language (L2) oral proficiency. While fluency is a widely studied construct, it has mostly been studied quantitatively from an individual's perspective based on monologue speech data and temporal measurements related to speech rate and pausing. However, there has been recent interest to explore the construct in dialogue contexts (e.g., Sato 2014; Peltonen 2017a, 2017b, 2020). Fluency researchers have also begun to study how individual speaking style in the first language (L1) influences L2 fluency by comparing L1 and L2 speech fluency patterns (e.g., De Jong et al. 2015; Peltonen 2018). This research strand has, so far, focused solely on monologue speech, leaving the influence of an individual speaking style on L2 speech performance unexplored in interactional settings. Furthermore, while word searches (WSs) have been studied especially from a conversation analytic perspective in L1 and L2 interaction (recently, e.g., Rydell 2019; Dressel 2020; Skogmyr Marian & Pekarek Doehler 2022; Tüma & Sherman 2022), they have rarely been the focus of fluency-oriented research, despite efficient and collaborative problem-solving during WSs having the potential to contribute to maintaining fluency especially in interactional settings (see, e.g., Peltonen 2020).

This exploratory study aims to fill these gaps in fluency research by examining L1 and L2 dialogue data from the same speakers and bridging fluency-oriented research with previous studies on WSs. In particular, our aim is to identify links between efficient problem-solving and speech fluency in interactional settings. The present study thus focuses on WSs in L1 Finnish and L2 English dialogues from university students of English. In line with the focus of this thematic publication, the present study focuses on elicited but freely produced speech. The study is part of

a larger project *Fluency and Disfluency Features in L2 Speech (FDF2)*; funded by the Research Council of Finland 2020–2024; decision number 331903). From the larger data set collected for the project, L1 Finnish (n=25) and L2 English (n=25) dialogues from the same participants (N=50) were analyzed in the present study. The following research questions were addressed:

- 1) RQ1: To what extent does the participants' use of WSs differ across their L1 and L2 dialogues?
- 2) RQ2: To what extent is the use of WSs connected across L1 and L2 dialogues?
- 3) RQ3: How are WSs connected with maintaining fluency in L1 and L2 dialogues among pairs with the highest frequencies of WSs?

The quantitative RQ1 was examined based on statistical analyses (Wilcoxon signed ranks tests). RQ2 was answered with a mixed-methods approach combining quantitative analysis (Spearman's rank-order correlations) with a qualitative analysis of the connections in WS use among selected participants' dialogues. The final RQ3, focusing on two pairs, was also examined with a mixed-methods approach: qualitative analyses were complemented with quantitative fluency measurements. Thus, overall, the study adopts a mixed-methods approach for examining fluency and WSs.

2 Fluency in interaction

L2 speech fluency has traditionally been examined from an individual speaker's perspective. The bulk of L2 fluency research has thus focused on characterizing learners' utterance fluency (i.e., by measuring mostly temporal fluency-related features in a speech sample; Segalowitz 2010) based on monologue data rather than interactional data. However, the present study builds on an emerging line of research within L2 fluency studies that focuses on characterizing the nature of fluency in dialogues, i.e., between two participants (see, e.g., Sato 2014; Peltonen 2017a, 2017b, 2020; Van Os et al. 2020). While some previous L2 fluency studies have also analyzed dialogic data, they have mostly focused on comparisons across monologue and dialogue fluency (e.g., Witton-Davies 2014; Tavakoli 2016), placing less emphasis on the collaborative elements of fluency. Furthermore, while the theoretical background and most suitable measures for monologue L2 fluency have become relatively well established in the field, the definitions and measures associated with *interactional fluency* (i.e., collaborative, between-turn aspects of fluency; Peltonen 2020) are not equally well established (for discussion, see McCarthy 2010; Feng 2022).

Similarly to characterizations of monologue (utterance) fluency consisting of different but interrelated dimensions (see, e.g., Skehan's 2009 framework of speed, breakdown, and repair fluency), interactional fluency can also be characterized in terms of different components: temporal fluency (related to the frequency and dura-

tion of between-turn pauses) and cohesive devices (especially other-repetitions and collaborative completions; Peltonen 2017b; 2020). Thus, while individual fluency is mostly captured with measures related to the abovementioned dimensions of speed, breakdown (pausing), and repair (e.g., corrections) of an individual speaker in monologues (or with within-turn measures in dialogue settings), interactional fluency measures aim to capture the joint efforts of maintaining fluent speech across individual speakers' turn boundaries (Peltonen 2020, see also McCarthy 2010). Previous examinations of collaborative completions (see, e.g., Lerner 1996), in particular, have shown that they can be effectively used by participants to maintain fluency in interactive settings (Peltonen 2017a). Some of these collaborative completions occur in conjunction with WSs (ibid.), highlighting the usefulness of also examining the role of WSs in maintaining fluency in dialogues.

Overcoming lexis-related problems might, thus, contribute to maintaining individual fluency, but also interactional fluency, especially if these sequences are solved with the help of the interlocutor. This approach is connected to previous characterizations on the role of problem-solving mechanisms (see Dörnyei & Kormos 1998) or fluency resources (Peltonen 2020) in facilitating speech fluency: along with different kinds of stalling mechanisms (filled pauses [FPs], repetitions, filler words, and sound lengthenings), communication strategies (e.g., Gullberg 2011; Peltonen 2017b) used in conjunction with WSs (such as circumlocutions, approximations, all-purpose words or word coinage¹; see, e.g., Dörnyei & Scott 1997) also have the potential to contribute to fluent speech. By examining WSs both qualitatively and quantitatively, the present study aims to complement previous studies by focusing on the role of WS in contributing to fluency in dialogues, covering both L1 and L2 dialogue (on previous fluency studies examining connections between L1 and L2 fluency based on monologue data, see, e.g., De Jong et al. 2015; Peltonen 2018).

3 Word searches

WSs have been characterized in previous literature as “mostly self-initiated, forward-oriented repair in which the progressivity of the speaker’s turn is delayed because of an item (i.e., a word) is not readily available to the speaker” (Dressel 2020: 39; based on Schegloff et al. 1977). Based on this characterization, WSs can be viewed as impacting fluency negatively, causing delays and potential disruptions in the flow of speech.

1 Circumlocutions include illustrations or descriptions of target items based on their characteristics, approximations include the use of alternative lexical items (e.g., superordinate or related terms) that are related in meaning to the target item, all-purpose words include general or “empty” lexical items (e.g., thing), and word coinage involves the creating of a non-existing word to refer to the target item (Dörnyei & Scott 1997; Peltonen 2020).

Yet, as argued above, especially collaboration in solving these sequences efficiently can, potentially, help in maintaining fluency from a broader interactional perspective. Dressel (2020: 39) describes WSs as consisting of three parts: onset of the search, the search-in-progress, and the resolution (or abandonment) of the search (following Chiarenza 2010). During the onset, a WS is initiated with the use of different resources, such as cut-offs, drawls (sound lengthenings) or FPs, such as *uh* or *um* in English or *öö* in Finnish (e.g., Schegloff et al. 1977; Goodwin & Goodwin 1986; Dressel 2020; Skogmyr Marian & Pekarek Doehler 2022); that is, from a fluency perspective, by using features that have mainly been considered as disfluencies. Yet, in the context of initiating a WS, they serve as important interactional signals.

During the search itself, various verbal (lexical and non-lexical) as well as multimodal resources (e.g., gestures) are used to contribute to the marking of WS as a discrete activity (e.g., Dressel 2020: 39). Verbal resources include, for instance, the abovementioned silent pauses (SPs) and FPs as well as repetitions (e.g., Dressel 2020; Skogmyr Marian & Pekarek Doehler 2022). Lower intensity can also be used to mark the WS as distinct from the surrounding talk (e.g., Dressel 2020). Kurhila (2006: 103–104), for instance, has documented how interrogatives in conjunction with WSs are often spoken with a quieter or lower voice compared to the surrounding talk. Along with solitary searches, during which speakers display their ongoing WS (e.g., with a “thinking face”; Goodwin & Goodwin 1986) but do not invite their interlocutor to participate, WSs can also be collaborative (“joint search” in Dressel 2020), i.e., involve the interlocutor (see, e.g., Brouwer 2003; Hayashi 2003; Kurhila 2006; Svennevig 2018; Rydell 2019; Tüma & Sherman 2022). Participants can invite their interlocutors to collaborate in the WS with interrogative utterances, such as *what is it*, combined with gaze directed to the interlocutor (e.g., Kurhila 2006). A WS can also start as a solitary search and, especially if not successful, can continue as a collaborative WS engaging the interlocutor (e.g., Kurhila 2006). Collaborative WSs come close to collaborative completions (e.g., Lerner 1996), i.e., completions of another speaker’s turn, which can be used to solve problems in conjunction with WSs, especially in L2 interaction (e.g., Taguchi 2014).

The WS is completed and resolved when the word (or a paraphrase, e.g., Skogmyr Marian & Pekarek Doehler 2022) that is the object of the search is provided by the participant initiating the search or their interlocutor (e.g., Dressel 2020). The WSs can, therefore, be self-completed or other-completed and thus collaboratively solved (e.g., Rydell 2019: 62). In the case of collaborative WSs, if the interlocutor provides the solution, it can be prosodically try-marked with rising intonation (e.g., Lerner 1996). This try-marked solution can then be subsequently confirmed by the first speaker with a repetition (e.g., Peltonen 2017a). WSs can also be left unsolved and abandoned without a solution (“let it pass” strategy; see, e.g., Rydell 2019: 62).

Previous early studies on WSs in L1 interaction have been conducted especially among speakers of English (e.g., Goodwin & Goodwin 1986; Lerner 1996) as well as

other L1s, such as Japanese (e.g., Hayashi 2003; see also Greer 2013 on Japanese/English bilingual interaction). More recently, the scope of WS research has extended to other L1s and methodologically often involves the use of multimodal conversation analysis (CA). For instance, Dressel (2020) examined WSs during joint storytelling in L1 French conversations and distinguished solitary and collaborative (joint) searches. For both types, multimodal resources, gaze in particular, were important in inviting (joint) or not inviting (solitary) participation (Dressel 2020). In the present study, while the focus is on verbal resources, the videos of the interactions were also consulted in the WS analysis.

Along with studies on WSs in L1 interaction, WSs have been of interest to L2 researchers. In particular, WSs have been examined among native and non-native-speaker interactions in L2 classrooms (Brouwer 2003; Koshik & Seo 2012; Duran et al. 2022) and outside classrooms, for instance, in workplaces (Svennevig 2018) or other institutional contexts (Kurhila 2006). Recently, the scope of WS studies has expanded to cover L2 interaction in various settings and among different L2s, including L2 English/English as lingua franca conversation (Binti Abdullah 2017; Siegel 2016), L2 French conversation (Skogmyr Marian & Pekarek Doehler 2022), paired L2 Swedish speaking tests (Rydell 2019), and speaking tasks in L2 English classrooms (Tüma & Sherman 2022). Some studies have even incorporated both L1 and L2 interaction (see, e.g., Chiarenza 2010 on L1 and L2 Italian), allowing for comparisons across L1 and L2 WS patterns, but these are relatively rare. Overall, previous studies have shown that WSs are performed in a similar manner in the L1 and the L2 (Kurhila 2006), but WSs can occur more often and be more elaborate in the L2 (Gullberg 2011). Previous studies have also highlighted the role of WSs as providing learning opportunities for L2 speakers (e.g., Brouwer 2003; Svennevig 2018). The present study, incorporating L1 and L2 dialogues from the same speakers, thus offers a novel perspective to examining WSs across the participants' L1 and L2 dialogues, providing insights into potential individual communicative styles across languages (see also Gullberg 2011).

4 Data and method

The present study uses data collected for a larger project *DF2*, as mentioned in the Introduction. Of this larger data set, the dialogic problem-solving tasks were included in this study. The participants (N=50, 25 pairs) in the project were university students of English in Finland (84% major students) who completed comparable problem-solving tasks in their L1 (Finnish) and L2 (English). The mean age of the participants was 21.9 years, and 58% of them were female (34% male, 2% other, 6% did not want to report their gender). The participants had studied English for approximately 10 years at school before their university studies and represented the CEFR level C1/C2 on average

based on their LexTale scores (mean 86.4%, scores over 80% representing the C1/C2 level; Lemhöfer & Broersma 2012). The participants can, therefore, be characterized as advanced speakers of English.

During data collection, the participants completed two problem-solving tasks, one in their L1 (n=25 interactions) and another in the L2 (n=25 interactions), with the order of the tasks and the languages counterbalanced. The task was completed with the same pair in both languages to control for potential interlocutor effects. Before each task, the participants were given two minutes of individual preparation time, followed by six minutes of time for task completion. Research assistants were present to give instructions, to audio- and video-record the tasks and to inform the participants when five minutes had passed but did not otherwise intervene. The two problem-solving tasks were designed to be maximally similar to enable comparisons across the two performances: in both tasks, the participants were given sixteen items to discuss (black-and-white images printed on A4 paper) and rank according to their usefulness for either survival on a desert island (see also Peltonen 2020) or reaching a mothership after a crash landing of a spaceship on the lighted side of the moon. The data set, thus, consists of elicited but freely produced speech: while the interactions were structured around the prompt materials and task instructions, the discussions themselves were freely produced by the participants. Since the pictures in the task prompt contained some items that were potentially more difficult for the participants to name than others, especially in the L2, the task design can be viewed suitable for eliciting WSs.

For the present study, all instances of solitary and collaborative WSs were identified from the data. The data were analyzed based on the audio files and accompanying transcriptions (for transcription conventions, see Appendix 1), but the video recordings of interactions were also consulted especially to distinguish solitary WSs from collaborative ones with the help of gaze direction (see Goodwin & Goodwin 1986). Overall, WSs were identified based on participants' displays that they were engaged in a WS (cf., making a WS "public", Rydell 2019: 63). These displays could involve various resources and often combinations of them, including SPs and FPs, sound lengthenings, repetitions, and wh-questions in conjunction with a vocabulary item (e.g., Rydell 2019; see also Section 3). As many of these features are considered as indicators of (dis)fluency, a clear orientation towards a specific vocabulary item (e.g., with the use of wh-questions or all-purpose words such as *thing* in English or *juttu* in Finnish) was required in addition to hesitations for the sequence to be counted as a WS (see also Rydell 2019).

Solitary WSs were started and answered (or left unanswered altogether) by the same participant, while collaborative WSs were interlocutor-oriented and, in many cases, ended with the interlocutor completing the search (see also Section 3). Collaborative WSs could include an implicit or explicit question for the interlocutor about a vocabulary item, often one of the objects included in the task (e.g.,

P[articipant]-62: *then I suppose the ah I don't remember the name but the like äh what is this called now* [pair 15, L2 English dialogue]), but could also involve the interlocutor's completion of a search without an explicit (verbal) request to participate from the speaker (P-29: *I'm also thinking uh we can maybe: put the tent and the um what-whatever it's called um* P-38: *sleeping bag?* [pair 5, L2 English dialogue]). The WSs receiving a collaborative solution and/or involving the interlocutor in other ways were, nevertheless, assigned only to the participant initiating the WS in the frequency calculations.

To answer the quantitative RQ1, descriptive statistics were calculated and subjected to statistical analyses. Both unstandardized (raw) and standardized frequencies of WSs are reported in Section 5. The frequencies were standardized per minute of an individual participant's speaking time (ST; excluding SPs) following a common practice in fluency research (De Jong 2016). To analyze potential differences between L1 Finnish and L2 English dialogues, non-parametric Wilcoxon signed ranks tests were calculated for the standardized measures. The non-parametric tests were preferred over paired samples t-tests due to non-normality in the data based on Shapiro-Wilk's tests of normality and a visual analysis of boxplots. False discovery rate was applied to the p-values to control for type I error, i.e., potential false positives (see Larson-Hall 2010: 251–252) resulting from multiple comparisons. To answer the quantitative part of the mixed-methods RQ2, Spearman's rank-order correlations were computed. The effect sizes were interpreted based on guidelines in Plonsky and Oswald (2014): r_s 0.25 = small, r_s 0.40 = medium, and r_s 0.60 = large. Non-parametric correlations were chosen over Pearson correlations due to the abovementioned non-normality in the data. To complement this quantitative analysis, connections in WS use were also analyzed qualitatively based on those participants' productions that were chosen for a more detailed analysis (see RQ3).

Finally, to answer the mainly qualitative RQ3, selected participants' L1 and L2 dialogues were analyzed in detail with a focus on how WSs were connected with maintaining fluency from individual and interactional perspectives. The pairs with the highest raw frequencies of WSs in the L1 Finnish (pair 11, P-53 and P-55) and L2 English (pair 16, P-61 and P-63) dialogues were chosen as the focus participants for the qualitative analyses. The approach was based on "extreme case analysis" (Dörnyei 2007: 272) used in some mixed-methods designs in applied linguistic studies: by illustrating unusual cases, the most extreme forms of a particular phenomenon can be explored and extended to the average population as well, albeit perhaps in less extreme ways (Dörnyei 2007: 128; see also Peltonen 2020). A similar "critical case sampling" (Dörnyei 2007: 128) has previously been used in conjunction with L1 and L2 monologue fluency analysis regarding stalling mechanisms (Peltonen 2018), L2 monologue fluency analysis of anxious and non-anxious individuals (Szyszka & Lintunen 2023), and L2 interactional fluency analysis focusing on other-repetitions and collaborative completions (Peltonen 2017a). To contextualize the qualitative

analysis, three fluency measures were calculated for the two pairs: articulation rate (AR; syllables per minute of ST), the mean length of within-turn SPs, and the mean length of between-turn SPs. The first two measures give an impression of each participant's individual fluency from the perspectives of speed (AR) and pausing (mean length of within-turn SPs), and the final measure (mean length of between-turn SPs) characterizes their interactional fluency from the perspective of "shared" silences. In the qualitative analysis itself, along with paying particular attention to features traditionally associated with disfluency (e.g., SPs), the use of stalling mechanisms (e.g., FPs, sound lengthenings, and filler words) during WSs was also examined to characterize their role in maintaining fluency.

5 Findings

In total, 48 instances of WSs were identified from the L1 Finnish dialogues: of these WSs, 66.7% (32) were solitary WSs, while 16 were collaborative. In comparison, from the L2 English dialogues, a total of 39 WSs were identified. Of these WSs, 43.6% were solitary (17), while 22 were collaborative. Of the 25 pairs, three pairs did not produce WSs at all during their L1 Finnish or L2 English dialogues.

The descriptive statistics (means and SDs) for solitary and collaborative WSs (raw frequencies and frequencies standardized per minute of ST) are compiled in Table 1.

TABLE 1. Means (*M*) and standard deviations (*SD*) for WSs in the Finnish and English dialogues (N=50).

Measure	Finnish (L1)		English (L2)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Number of solitary WSs	0.64	0.92	0.34	0.52
Number of solitary WSs / minute of ST	0.28	0.42	0.14	0.23
Number of collaborative WSs	0.32	0.62	0.44	0.97
Number of collaborative WSs / minute of ST	0.14	0.27	0.19	0.43
Total number of WSs	0.96	1.28	0.78	1.22
Total number of WSs / minute of ST	0.42	0.60	0.33	0.54

As can be seen from Table 1, the tendencies for raw and standardized frequencies are similar: L1 Finnish dialogues, on average, contained more WSs than the L2 English

dialogues. Within the specific WS types, solitary WSs were more common than collaborative ones in the Finnish L1 dialogues, while the opposite tendency was observed for the English L2 dialogues. Based on Wilcoxon signed ranks tests for the standardized frequencies, only the difference between the number of solitary WSs per minute of ST in Finnish and English dialogues was statistically significant ($Z = -2.451, p = 0.014$), while the differences for total WSs per minute of ST ($Z = -1.131, p = 0.258$) and collaborative WSs per minute of ST ($Z = -0.120, p = 0.904$) were nonsignificant.

To answer RQ2, correlations between the two types of WSs and the total (standardized) frequencies in the two languages were calculated. Based on the results ($N=50$), none of the correlations reached statistical significance. Solitary WSs per minute of ST were weakly positively correlated across L1 and L2 dialogues ($r_s = 0.271, p = 0.057$), while no association was found between the total number of WSs per minute of ST across the dialogues in different languages ($r_s = 0.058, p = 0.690$). Interestingly, while the correlation between collaborative WSs in L1 and L2 was also weak and not statistically significant, it was negative: thus, participants producing many collaborative WSs in L2 English produced few of them in Finnish and vice versa ($r_s = -0.241, p = 0.092$).

A closer look at individual participants' performances for the selected pairs (pairs 11 and 16) suggested that, for pair 11, P-53 used two solitary and one collaborative WS, while P-55 used two solitary and three collaborative WSs in their L1 Finnish dialogue. In contrast, in their L2 English dialogue, P-53 used two solitary and one collaborative WS (exactly the same profile as in the Finnish dialogue), while P-55 did not use any WSs in the English dialogue. For pair 16, the pattern was somewhat different, as they engaged in WSs relatively rarely in their Finnish dialogue: P-61 used only one solitary WS, while P-63 used one solitary and one collaborative WS. In contrast, P-61 engaged frequently especially in collaborative WS in the English dialogue: six collaborative WSs and one solitary WS were identified from their dialogue. However, P-63 did not produce any WSs during their English dialogue. These patterns provide additional support for the lack of statistically significant correlations across L1 and L2 dialogues: as the use of WSs was, overall, relatively rare in the data, it is perhaps not surprising that clear patterns were not detected in the quantitative analysis. Furthermore, even the comparison of the two pairs showed that the patterns regarding L1–L2 connections could be relatively different, not only across individual participants, but also across pairs. While certain participants can have a particular communicative style regarding WSs across the two languages (e.g., P-53), others exhibit differing patterns in the two languages: for instance, P-61 seemed to rely on their partner more during WSs in the English dialogue than in the Finnish one (see also RQ3 below).

To answer RQ3, the two pairs' WSs were analyzed in detail, with particular attention paid to the connections between WSs and fluency. Before discussing the

qualitative analyses, information about the selected pairs' samples and quantitative fluency profiles based on selected measures is provided in Table 2.

TABLE 2. Descriptive information about the selected pairs' samples (duration, syllables) and quantitative fluency measures.

Measure	Pair 11				Pair 16			
	P-53		P-55		P-61		P-63	
	Fin	Eng	Fin	Eng	Fin	Eng	Fin	Eng
Dialogue duration	366.25	392.70	366.25	392.70	221.79	328.78	221.79	328.78
Number of syllables (individual)	295	432	1093	879	638	531	363	345
Articulation rate	275.61	227.76	311.31	246.35	342.53	225.52	336.30	256.25
Mean length of within-turn SPs	0.95	0.67	0.71	0.61	0.76	0.72	0.66	0.57
Mean length of between-turn SPs	0.81	0.66	0.81	0.66	0.85	0.99	0.85	0.99

Note: Articulation rate = syllables per minute of ST (excluding all SPs).
The dialogue durations and mean lengths of SPs are given in seconds.

As can be seen from Table 2, the durations of the dialogues in the L1 and L2 were relatively similar for pair 11, while pair 16 produced a shorter dialogue in L1 Finnish than in L2 English. P-55 was the more dominant participant in pair 11's dialogues, producing 79% of the syllables in the L1 Finnish dialogue and 67% of the syllables in the L2 English dialogue. Pair 16's dialogue was more symmetrical, with P-61 only slightly dominating the interactions (producing 64% of the syllables in the L1 Finnish and 60% in the L2 English dialogues). Regarding individual fluency (articulation rate), all participants were somewhat more fluent, i.e., faster speakers, in their L1 Finnish dialogues. Interestingly, however, regarding the other individual fluency measure (mean length of within-turn SPs), there was an opposite tendency: the SPs within turns were, on average, somewhat longer for all participants in their L1 Finnish compared to L2 English dialogues. Finally, regarding the interactional fluency measure (mean length of between-turn SPs), a difference between the two pairs was observed. Pair

11's dialogue in L1 Finnish included somewhat longer "shared" silences, on average, compared to the L2 English dialogue, whereas for pair 16, the between-turn SPs were somewhat longer in the L2 English than in the L1 Finnish dialogue. In the following, we will discuss both pairs' WS use, starting with pair 11, followed by pair 16.

As stated before, pair 11 (P-53 and P-55) produced the most WSs in the Finnish dialogues (P-53 three instances, P-55 five instances). Both of them used two solitary WSs in Finnish. P-55's WSs related to the word *hätäsoihtu* ('flare' in English, see Example 1; full English translations of the Finnish examples, translated by the first author, can be found in Appendix 2).

(1a) Flare, part 1 (pair 11, P-55)

- 1 P-55: joo (.) ku mä mietin ku (.) täs on meijä tää niinku varotus tää (0.25)
- 2 P-55: * GAZE DIRECTED TO PAPER ON THE DESK THROUGHOUT THE SEQUENCE
- 3 P-55: tai tää niinku hälytys (0.28) valo (0.67) syty[tys] juttu (0.50)
- 4 P-53: [nii]
- 5 P-55: nii mä mietin et (.) oisko kuitenkin et kiikarit?

(1b) Flare, part 2 (pair 11, P-55)

- 1 P-55: et (.) ehkä sitten kiikarit ja sit se (0.76) varoitus:: (0.80)
- 2 P-55: * GAZE DIRECTED TO PAPER THROUGHOUT THE SEQUENCE
- 3 P-55: sytyke (0.38)
- 4 P-55: °en tiedä (.) [mikä se on°]
- 5 P-53: [entä sit] toi lamppu tai mikä toi (0.55)

In Example 1a, P-55 engages in a WS searching for the word *hätäsoihtu* ('flare'). The participant produced approximations related to the target word (*varotus* and *hälytys*; English 'warning' and 'alarm' combined with the word *valo*; 'light'), followed by a further combination of an approximation (*sytytys*) and an all-purpose word (*juttu*, roughly translating to 'detonation thing'). These approximations indicate the participant's engagement in a WS, coupled with the two filler words *niinku* ('like') and SPs, with gaze directed at the paper throughout the sequence suggesting a solitary search. P-53 does not engage in the WS, providing only a minimal acknowledgment token *nii* ('yes') in line 4. While P-55's search is relatively lengthy, it only minimally interrupts the speaker's individual fluency: the sequence does not contain very long SPs and includes filler words to keep the flow of talk going. Furthermore, the multiple alternative attempts to convey the meaning of *flare* help in maintaining the flow of the interaction. Later in the dialogue (Example 1b), P-55 refers to the flares again by using a new approximation (*varoitussytyke*; 'warning kindling'), accompanied with a sound lengthening at the end of the word *varoitus*, a typical indicator for initiating a WS (e.g., Dressel 2020). This time, P-55 also explicitly marks their lack of vocabulary knowledge related to this particular item, spoken more softly than the surrounding talk (*en tiedä*

mikä se on; 'I don't know what it is' in line 4; see also Kurhila 2006). Again, P-55's gaze is directed at the paper indicating a solitary search, and P-53 does not engage in this WS but continues the discussion by shifting the focus to another item, using a solitary WS themselves (*lamppu tai mikä toi*; 'lamp or what that' in line 5). Despite the lack of P-53's engagement in P-55's WS, the dialogue is minimally interrupted. From the perspective of P-55's individual fluency, despite the WSs, their articulation rate still, overall, remained relatively high (311 syllables per minute) and the SPs were not exceptionally lengthy on average (0.71 seconds within turns).

Regarding the collaborative WSs used by the pair in their Finnish dialogue (one instance for P-53, three instances for P-55), two of them related to the same item (*onki*; 'fishing rod' in English; see Example 2).

(2a) Fishing rod, part 1 (pair 11, P-55)

- 1 P-55: ni mä mietin et tuisko sit sen (1.09)
- 2 P-55: * GAZE DIRECTED TO PAPER
- 3 P-55: tää (0.57) °kalastus::° (0.43)
- 4 P-55: * SHIFTS GAZE TOWARDS P-53
- 5 P-55: °mikä tää onkaa°
- 6 P-55: * GAZE BACK TO PAPER
- 7 (0.36)
- 8 P-53: on-
- 9 (0.26)
- 10 P-55: [nii onki nii]
- 11 P-53: [onki onki]
- 12 P-55: tulisko onki ennen kattilaa (1.30)

(2b) Fishing rod, part 2 (pair 11, P-55)

- 1 P-55: amm (0.37) >mikä tää on< (0.39)
- 2 P-55: * GAZE DIRECTED TO PAPER THROUGHOUT SEQUENCE
- 3 P-53: on[ki]
- 4 P-55: [on]ki (0.64) kattila

In Example 2a, P-55 starts their WS by shifting their gaze from the paper towards P-53, engaging their interlocutor in the search (lines 3–4). Preceded by *tää* ('this') and a SP, P-55 produces a softly spoken *kalastus* ('fishing'), with a sound lengthening at the end of the word, using these markers to distinguish the WS from the surrounding talk. Following a SP, the sequence also contains an interrogative *mikä tää onkaa* ('whatever this is'; line 5), with gaze shifting back towards the desk. This WS thus contains similar elements as Example 1: quieter speech, sound lengthening, approximations, SPs, and an interrogative expressing the participant's lack of knowledge regarding the target word. However, in contrast to Example 1, the gaze direction is different, engaging P-53 in the WS: they provide the first syllable of the target word *onki* ('fishing rod') in line

8, which is followed (after a SP) by P-55 confirming and providing the target word in line 10 (*nii onki nii*; 'yes fishing rod yes'). This is produced in overlap with P-53, who also produces and repeats the target word (*onki onki*, line 11), providing a collaborative resolution to the WS. With P-53 orienting to P-55's WSs and helping them finding the target word, the participants are able to move the discussion forward collaboratively after some minor halts in P-55's individual fluency.

Later on in the discussion they return to the same item while listing the items in the order of importance (Example 2b). P-55 engages in a brief WS in line 1, marked by an interrogative (*mikä tää on*; 'what is this') preceded by a FP (*amm*) and a SP. P-55's gaze is directed at the paper, not inviting the interlocutor to participate. Yet P-53 collaborates and provides the word *onki* (line 3), preceded only by a very brief SP of 0.39 seconds and the last syllable being produced in partial overlap with P-55's production of the target word in line 4. Compared to Example 2a, this sequence is solved more quickly and without a request for P-53 to collaborate; yet both participants provide the solution to the WS together almost simultaneously and are thus able to move the discussion forward quickly, maintaining interactional fluency and avoiding lengthy silences.

The final example from pair 11, now from their L2 English dialogue, involves a collaborative WS initiated by P-53 (Example 3).

(3a) Intercom, part 1 (pair 11, P-53)

- 1 P-53: so {*pt*_0.54} is this an intercom? (0.77)
- 2 P-55: yeah or like a: (.) [radio]phone or
- 3 P-53: [inter-]
- 4 P-53: yeah [that's]
- 5 P-55: [it-]
- 6 P-55: I think yeah that
- 7 (1.14)
- 8 P-53: I don't know the exact term but (0.83)
- 9 P-53: so they can: (0.77) I don't know who they can communicate

(3b) Intercom, part 2 (pair 11, P-53)

- 1 P-55: do you think flashlight or rope after (.) or maybe the: (.) intercom (0.69) °there°
- 2 P-53: um (.) they do need to communicate with each other

In contrast to the more indirect invitations to participate in the WS by P-55 in Example 2, in Example 3a, P-53 initiates their WS by asking directly from their interlocutor whether an item in the task is an intercom (line 1), preceded by WS markers *so* and a SP including a lip smack.² P-55 initially confirms this in line 2 (*yeah*) but continues

2 P-53's gaze direction is difficult to estimate reliably here due to their hair being in front of their eyes.

by providing a more appropriate word (*like a radiophone*), albeit with some hedging (the word *like* and sound lengthening in the indefinite article *a*). After some overlapping talk in line 6, P-55 returns to the word they previously suggested, confirming that it is likely to be the correct word (*I think yeah that*). Followed by a SP, P-53 takes the turn and, despite indicating not knowing the target word (line 8), shifts the focus from terminological negotiation towards task completion (*I don't know who they can communicate*, line 9) without adopting P-55's more target-like *radiophone*.

Interestingly, later in the discussion, the opposite happens: in Example 3b, P-55 adopts P-53's candidate word *intercom* (line 1), followed by P-53 reiterating their earlier point about the need of being able to communicate with one another (line 2). Thus, rather than using the more target-like *radiophone*, the participants both refer to the item as *intercom*. While the word *radiophone* would have depicted the item more accurately, the participants opted for a less target-like item for the purposes of maintaining interactional fluency and focusing on the completion of the task without further negotiation related to this particular vocabulary item.³

The other pair chosen for the qualitative analysis, pair 16, produced few WSs in their Finnish dialogue (P-61 one solitary WS; P-63 one solitary and one collaborative WS). Note, however, that for the two pairs, the quantitative interactional fluency (based on the mean length of between-turn SPs) was very similar in their L1 Finnish dialogues, indicating that the frequency of WSs is not reflected in (the lack of) interactional fluency in a straightforward way. Example 4 illustrates P-61's solitary WS.

(4) Star chart (pair 16, P-61)

- 1 P-61: sit mä mietin et jos on joku tämmöne (.) tähti (.) kartta-asia?
- 2 P-61: * GAZE DIRECTED TO PAPER
- 3 P-61: ni sithän jos me tietään et misä kaikki tähtikuviot on
- 4 P-61: ni sit me voidaa suunnistaa siel \$paremmi\$ *hihi*
- 5 P-61: * GAZE SHIFTED TOWARDS INTERLOCUTOR

In Example 4, P-61 is searching for the word *tähtikartta* ('star chart') in Finnish. While the example does not include as many hesitations as Example 1, only some micro-pauses, it contains markers of vagueness (*joku tämmöne*; 'something like this' in English), suggesting the participant's engagement in a solitary WS, coupled with their gaze towards the desk. The compound word *tähtikartta* itself is interrupted with a

3 Pair 18 arrived at a somewhat similar solution when discussing a magnifying glass. Neither participant used the target-like expression; instead, P-67 explicitly offered an alternative item *widening glass* (P-67: *for the purpose of this let's call it the widening glass*) and the interlocutor P-65 adopted this item. Both participants also used the expression later on in the interaction to refer to the magnifying glass. While a non-target-like expression, using this coined term helped with task accomplishment and maintaining interactional fluency, much like *intercom* in Example 3 (for a similar example involving word coinage, see Rydell 2019).

micropause and ends with an all-purpose word *asia* ('thing') attached to the end of the compound, forming a unit that contains the target word with an added element of hedging. This, coupled with the rising intonation, marks the word as potentially non-target-like, but as something close to it. P-63 does not participate in this search; instead, P-61 continues by providing a justification for why this item could be useful (lines 3–4), following the task instructions and moving the discussion forward despite the potential insecurity related to the specific vocabulary item. Furthermore, P-61's individual fluency does not seem to be disrupted during this sequence, as it does not contain any SPs.

The final example (Example 5) illustrates one of the six collaborative WSs produced by P-61 during the English dialogue.

(5) Rockets (pair 16, P-61)

- 1 P-61: um (0.37) but what if this- (0.39) these are (1.45) are those (0.69)
- 2 P-61: * GAZE DIRECTED TO THE PAPER
- 3 P-61: * POINTS AT PAPER
- 4 P-61: [rockets or the-?]
- 5 P-61: * GAZE SHIFTED TOWARDS THE INTERLOCUTOR
- 6 P-61: * WAVING GESTURE WITH RIGHT HAND
- 7 P-63: I [don't know the exact] name but like (1.37)
- 8 P-63: to let others know you're here
- 9 (0.52)
- 10 P-61: yeah okay tha[t would be] (0.33) kind of cool [*hahaha*] so
- 11 P-63: [sticks] [\$yeah\$]

In line 1, P-61 initiates the WS with a combination of a FP and SP, followed by a hesitantly delivered *wh*-question. At this point, P-61's gaze is directed to the paper (line 2). They also point at the item on the paper (line 3) to specify the referent *this* (later repaired to *these*). While producing an approximation *rockets* in line 4 (followed by an embodied completion involving a representational gesture consisting of waving of the imaginary flares in their hand, line 6), P-61 shifts their gaze towards the interlocutor, inviting them to collaborate in the search. P-63's response in line 7, stating that they do not know the target-like word, partially overlaps with P-63's previous turn, minimizing the gap between turns and enhancing interactional fluency. This is followed by a circumlocution describing the function of the object in line 8 (*to let others know you're here*). In line 10, P-61 responds after a brief between-turn SP of 0.52 seconds by stating that it would be *kind of cool*, followed by laughter. Involving P-63 in the WS ensures that there is minimal disruption to interactional fluency.

Using *wh*-questions to engage the interlocutor was a typical strategy for P-61 in the L2: in their other collaborative WSs as well, direct *wh*-questions were common (e.g., *what are these things*, *what is this called*), occasionally spoken softly, marking the activity as distinct from the surrounding talk. Compared to the examples from pair

11, P-61 thus tended to use a somewhat different strategy, more directly engaging the interlocutor in WSs. This is interesting considering that P-61 was, nevertheless, the participant producing somewhat more talk during their L1 and L2 interactions. While engaging P-63 in the WSs can be viewed as a good approach for maintaining interactional flow, from an individual fluency perspective, P-61 did not utilize many individual communication strategies to attempt to solve the WSs to maintain their individual fluency, as reflected, in particular, in their within-turn SPs in Example 5. While the local disfluencies did not seem to have a major impact on the overall mean length of within-turn SPs for this participant, the slower speech production during WSs could have been reflected in P-61's articulation rate, which was considerably slower in their L2 English dialogue (226 syllables per minute) than in the L1 Finnish dialogue (343 syllables per minute). As the interlocutor P-63 represented C1/C2 level and P-61 represented level B2 in the CEFR, P-61 potentially relied on the more proficient participant's vocabulary knowledge in their L2 English dialogue. This strategy could be useful from a language learning perspective, with the more advanced participant providing input above the interlocutor's level (scaffolding; Wood et al. 1976; see also Brouwer 2003; Peltonen 2017a).

6 Discussion and conclusion

The present study set out to examine WSs in L1 Finnish and L2 English dialogues from the same speakers using a mixed-methods approach, aiming to bridge the gap between research on WSs and fluency research. While efficient problem-solving during WSs has the potential to contribute to individual and interactional fluency (e.g., Peltonen 2020), WSs had rarely been in the focus of fluency research before the present study. Examining L1 and L2 dialogue data from the same speakers was an additional novel aspect of the present study. The findings showed differences especially in the use of solitary WSs across L1 and L2 samples (RQ1), WSs being used more often in the L2 English dialogues despite the participants' high proficiency level. Regarding RQ2, correlations between WS use in L1 Finnish and L2 English did not reach statistical significance, but based on qualitative observations, some participants demonstrated potential cross-linguistic communication styles regarding WSs (see also Gullberg 2011). However, since the qualitative analysis was based on only four participants, further research is needed to confirm this tendency discovered in the present study. Finally, based on detailed analyses of the four participants' WSs across L1 and L2 samples (RQ3), with focus on examining these sequences from the perspective of fluency-related features and linking them to the participants' quantitative fluency measures, especially collaborative WSs were found to be linked with a potential to enhance fluency from an interactional perspective (cf. Peltonen 2017a on collaborative completions).

Furthermore, while WSs contained some disfluencies, such as SPs, these were not reflected in particularly low individual fluency measures for the participants. Some differences in the participants' orientations to problem-solving were also observed: while some participants preferred solitary WSs, others relied more on their interlocutors to overcome lexis-related problems. Furthermore, the profiles were not always linked across L1 and L2 dialogues, suggesting that some participants' strategies were, at least to some extent, language-specific.

Overall, while some quantitative tendencies could be detected based on the frequencies of WSs in the L1 and L2 dialogues, the qualitative analyses suggested that individual participants and pairs display complex and somewhat different patterns regarding L1 and L2 WSs and fluency. While the present study can be viewed as exploratory in nature, being among the first to approach WSs from the perspective of fluency in an interactional setting, the study provides the basis for further research in this area. Despite WSs being marked with indicators that have traditionally been approached as disfluencies in fluency research, the present study, along with previous CA-oriented WS studies (e.g., Dressel 2020; Skogmyr Marian & Pekarek Doehler 2022), has highlighted the importance of these disfluencies in marking the WS as a separate activity from the surrounding talk and providing important interactional cues for the interlocutor. This poses implications for speech fluency research using interactional data: while on the surface level, certain features may suggest disfluency, they may also provide the speaker more time to complete their search (cf., Dörnyei & Kormos 1998 on stalling mechanisms providing planning time during problem-solving) and, as mentioned above, signal their engagement in WSs to the interlocutor. Furthermore, from a more collaborative perspective, the disfluencies may provide cues for the interlocutor to participate in the search, and as a result of the participation, contribute to maintaining the flow of the discussion, i.e., interactional fluency.

While the data set used in the present study, involving task-based dialogues, differed somewhat from previous studies on WSs in institutional (e.g., classrooms) or informal (e.g., conversation) settings, the patterns found regarding the structure of WSs were, in many respects, similar to those found in previous L1 (e.g., Dressel 2020) and L2 studies (e.g., Rydell 2019). Thus, the present study has added to the body of CA-oriented research on WSs by confirming many observations demonstrated previously in other settings and/or with other languages. Most importantly, the data set used in the present study enabled the analysis of L1 and L2 patterns from the same speakers, which has been rare in previous research. Furthermore, while WSs in L1 and L2 English interaction have been extensively studied, the present study provided insights into WS use from the perspective of a less widely studied language, Finnish (but see Kurhila 2006).

While the present study has some limitations, especially related to the preliminary nature of the qualitative findings due to the small sample size, it has provided

the starting point for examining the links between fluency and WSs with a mixed-methods approach. Importantly, WSs were employed both in the L1 Finnish and L2 English dialogues, with a statistically significant difference only being found in the use of solitary WSs. Furthermore, based on the qualitative analysis, structurally, the L1 and L2 WSs did not differ considerably from each other (see also, e.g., Kurhila 2006). This suggests that WSs are employed both in the L1 and L2 to solve especially lexis-related problems, the frequency of collaborative WSs not differing statistically significantly across L1 and L2 dialogues. Thus, from a fluency perspective, disfluencies in conjunction with WSs should not be straightforwardly categorized as impacting the individual or interactional fluency negatively; in contrast, the efficient problem-solving during these sequences may, in fact, help to maintain the overall flow of the interaction despite local disfluencies (see also Peltonen 2017b).

The findings of the present study also have implications for L2 teaching (see also Lintunen et al. 2020; Peltonen 2023). In L2 classrooms, students' attention could be drawn to the fact that WSs are common in both L1 and L2 interaction. Efficient ways of solving WS could also be practiced with various activities, such as the Alias game, in which you explain a concept without mentioning the word. In addition, different kinds of problem-solving tasks, such as the ones used for data elicitation in the present study, could be used in classrooms to practice collaborative problem-solving: since the task instructions specify a common goal and encourage collaboration and negotiation among the participants, they are likely to be useful for pedagogical purposes as well, especially combined with awareness-raising about problem-solving strategies and WSs (see also Peltonen 2020). Importantly, students should be encouraged to use their whole multimodal repertoire in these activities: by realizing that they have plenty of resources at their disposal when encountering communicative problems, the students can be supported to become efficient communicators also outside the classroom.

Literature

- Binti Abdullah, N. 2017. *A multimodal conversation analytic study of word-searches in L2 interaction*. An unpublished doctoral dissertation. Newcastle University. <https://theses.ncl.ac.uk/jspui/handle/10443/3908>
- Brouwer, C. 2003. Word searches in NS–NNS conversations: opportunities for learning? *The Modern Language Journal*, 87 (4), 534–545. <https://doi.org/10.1111/1540-4781.00206>
- Chiarenza, A. C. 2010. *Word searches in L1 and L2 Italian conversation: re-establishing intersubjectivity*. An unpublished doctoral dissertation. University of Illinois at Urbana–Champaign. <https://www.ideals.illinois.edu/items/17109>
- De Jong, N. H. 2016. Fluency in second language assessment. In D. Tsagari & J. Banerjee (eds) *Handbook of second language assessment*. Berlin: De Gruyter, 203–218.
- De Jong, N. H., R. Groenhout, R. Schoonen & J. H. Hulstijn 2015. Second language fluency: speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36 (2), 223–243. <https://doi.org/10.1017/S0142716413000210>
- Dressel, D. 2020. Multimodal word searches in collaborative storytelling: on the local mobilization and negotiation of coparticipation. *Journal of Pragmatics*, 170, 37–54. <https://doi.org/10.1016/j.pragma.2020.08.010>
- Duran, D., S. Kurhila & O. Sert 2022. Word search sequences in teacher–student interaction in an English as medium of instruction context. *International Journal of Bilingual Education and Bilingualism*, 25 (2), 502–521. <https://doi.org/10.1080/13670050.2019.1703896>
- Dörnyei, Z. 2007. *Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Dörnyei, Z. & J. Kormos 1998. Problem-solving mechanisms in L2 communication: a psycholinguistic perspective. *Studies in Second Language Acquisition*, 20 (3), 349–385. <https://doi.org/10.1017/S0272263198003039>
- Dörnyei, Z. & M. L. Scott 1997. Communication strategies in a second language: definitions and taxonomies. *Language Learning*, 47 (1), 173–210. <https://doi.org/10.1111/0023-8333.51997005>
- Feng, R. 2022. Cognitive factors influencing utterance fluency in L2 dialogues: monadic and non-monadic perspectives. *Frontiers in Psychology*, 13, article 926367. <https://doi.org/10.3389/fpsyg.2022.926367>
- Goodwin, M. H. & C. Goodwin 1986. Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62 (1–2), 51–75. <https://doi.org/10.1515/semi.1986.62.1-2.51>
- Greer, T. 2013. Word search sequences in bilingual interaction: codeswitching and embodied orientation toward shifting participant constellations. *Journal of Pragmatics*, 57, 100–117. <https://doi.org/10.1016/j.pragma.2013.08.002>
- Gullberg, M. 2011. Multilingual multimodality: communicative difficulties and their solutions in second-language use. In J. Streeck, C. Goodwin & C. LeBaron (eds) *Embodied interaction: language and body in the material world*. Cambridge: Cambridge University Press, 137–151.
- Hayashi, M. 2003. Language and the body as resources for collaborative action: a study of word searches in Japanese conversation. *Research on Language and Social Interaction*, 36 (2), 109–141. https://doi.org/10.1207/S15327973RLSI3602_2
- Koshik, I. & M.–S. Seo 2012. Word (and other) search sequences initiated by language learners. *Text & Talk*, 32 (2), 167–189. <https://doi.org/10.1515/text-2012-0009>
- Kurhila, S. 2006. *Second language interaction*. Amsterdam: John Benjamins.

- Lemhöfer, K. & M. Broersma 2012. Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44, 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Lerner, G. H. 1996. On the “semi-permeable” character of grammatical units in conversation: conditional entry into the turn space of another speaker. In E. Ochs, E. A. Schegloff & S. A. Thompson (eds) *Interaction and grammar*. Cambridge: Cambridge University Press, 238–276.
- Lintunen, P., M. Mutta & P. Peltonen 2020. Synthesising approaches to second language fluency: implications and future directions. In P. Lintunen, M. Mutta & P. Peltonen (eds) *Fluency in L2 learning and use*. Bristol: Multilingual Matters, 186–201.
- McCarthy, M. 2010. Spoken fluency revisited. *English Profile Journal*, 1 (1), 1–15. <https://doi.org/10.1017/S2041536210000012>
- Peltonen, P. 2017a. L2 fluency in spoken interaction: a case study on the use of other-repetitions and collaborative completions. In M. Kuronen, P. Lintunen & T. Nieminen (eds) *Näkökulmia toisen kielen puheeseen – Insights into second language speech*. AFinLA-e. Soveltavan kielitieteen tutkimuksia 10. Jyväskylä: Finnish Association for Applied Linguistics AFinLA, 118–138. <https://doi.org/10.30660/afinla.73130>
- Peltonen, P. 2017b. Temporal fluency and problem-solving in interaction: an exploratory study of fluency resources in L2 dialogue. *System*, 70, 1–13. <https://doi.org/10.1016/j.system.2017.08.009>
- Peltonen, P. 2018. Exploring connections between first and second language fluency: a mixed methods approach. *The Modern Language Journal*, 102 (4), 676–692. <https://doi.org/10.1111/modl.12516>
- Peltonen, P. 2020. *Individual and interactional speech fluency in L2 English from a problem-solving perspective: a mixed-methods approach*. Turku: University of Turku. <https://urn.fi/URN:ISBN:978-951-29-8137-3>
- Peltonen, P. 2023. Fluency revisited. *ELT Journal*. Epub ahead of print. <https://doi.org/10.1093/elt/ccad047>
- Plonsky, L. & F. L. Oswald 2014. How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64 (4), 878–912. <https://doi.org/10.1111/lang.12079>
- Rydell, M. 2019. Negotiating co-participation: embodied word searching sequences in paired L2 speaking tests. *Journal of Pragmatics*, 149, 60–77. <https://doi.org/10.1016/j.pragma.2019.05.027>
- Sato, M. 2014. Exploring the construct of interactional oral fluency: second language acquisition and language testing approaches. *System*, 45, 79–91. <https://doi.org/10.1016/j.system.2014.05.004>
- Schegloff, E. A., G. Jefferson & H. Sacks 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53 (2), 361–382. <https://doi.org/10.2307/413107>
- Segalowitz, N. 2010. *The cognitive bases of second language fluency*. New York: Routledge.
- Siegel, A. 2016. *Longitudinal development of word search sequences in English as a lingua franca interaction*. An unpublished doctoral dissertation. Newcastle University. <https://theses.ncl.ac.uk/jspui/handle/10443/3227>
- Skogmyr Marian, K. & S. Pekarek Doehler 2022. Multimodal word-search trajectories in L2 interaction: the use of gesture and how it changes over time. *Social Interaction: Video-Based Studies of Human Sociality*, 5 (1). <https://doi.org/10.7146/si.v5i2.130867>
- Svennevig, J. 2018. “What’s it called in Norwegian?” Acquiring L2 vocabulary items in the workplace. *Journal of Pragmatics*, 126, 68–77. <https://doi.org/10.1016/j.pragma.2017.10.017>

- Szyska, M. & P. Lintunen 2023. Zooming into the L2 speech fluency markers of anxious and non-anxious advanced L2 learners – an extreme case sampling report. *Research in Language*, 21 (4), 357–376. <https://doi.org/10.18778/1731-7533.21.4.02>
- Taguchi, N. 2014. Development of interactional competence in Japanese as a second language: use of incomplete sentences as interactional resources. *The Modern Language Journal*, 98 (2), 518–535. <https://doi.org/10.1111/modl.12087>
- Tavakoli, P. 2016. Fluency in monologic and dialogic task performance: challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, 54 (2), 133–150. <https://doi.org/10.1515/iral-2016-9994>
- Tůma, F. & T. Sherman 2022. Recruiting help in word searches in L2 peer interaction: a multimodal conversation-analytic study. *Linguistics and Education*, 67, 1–15. <https://doi.org/10.1016/j.linged.2021.100999>
- Van Os, M., N. H. de Jong & H. R. Bosker 2020. Fluency in dialogue: turn-taking behavior shapes perceived fluency in native and non-native speech. *Language Learning*, 70 (4), 1183–1217. <https://doi.org/10.1111/lang.12416>
- Witton-Davies, G. 2014. *The study of fluency and its development in monologue and dialogue*. An unpublished doctoral dissertation. Lancaster University.
- Wood, D., J. S. Bruner & G. Ross 1976. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17, 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>

Appendix 1. Transcription conventions

:	Extended or stretched sound, syllable, or word
(.)	Micropause (duration less than 0.25 seconds)
(0.65)	Timed pause (duration at least 0.25 seconds)
pt	Lip smack. Included in SP time measured and marked with curly brackets {*pt*_0.54}
um, uh	Non-lexicalized FP
heh	Laugh syllable
\$ \$	Laughing/chuckling talk between markers
° °	Passage of talk noticeably softer than surrounding talk
> <	Portions of talk delivered at a pace noticeably quicker than surrounding talk
?	Rising vocal pitch
-	Halting, abrupt cut off of a sound or word
=	Latching of contiguous utterances, with no interval or overlap
[]	Overlapping speech
* CAPS	Multimodal conduct (gestures, gaze)

Appendix 2. English translations of Examples 1–2

The translations are provided in italics below each line.

(1a) Flare, part 1 (pair 11, P-55)

- 1 P-55: joo (.) ku mä mietin ku (.) täs on meijä tää niinku varotus tää (0.25)
yeah (.) since I was thinking since (.) here is our this like warning this (0.25)
- 2 P-55: * GAZE DIRECTED TO PAPER ON THE DESK THROUGHOUT THE SEQUENCE
- 3 P-55: tai tää niinku hälytys (0.28) valo (0.67) syty[tys] juttu (0.50)
or this like alarm (0.28) light (0.67) detona[tion] thing (0.50)
- 4 P-53: [nii]
[yes]
- 5 P-55: nii mä mietin et (.) oisko kuitenkin et kiikarit?
so I was thinking that (.) should we rather take binoculars?

166 WORD SEARCHES AND FLUENCY DURING PROBLEM-SOLVING TASKS:
A MIXED-METHODS STUDY OF L1 FINNISH AND L2 ENGLISH INTERACTION

(1b) Flare, part 2 (pair 11, P-55)

- 1 P-55: et (.) ehkä sitten kiikarit ja sit se (0.76) varoitus:: (0.80)
so (.) maybe then binoculars and then the (0.76) warning:: (0.80)
- 2 P-55: * GAZE DIRECTED TO PAPER THROUGHOUT THE SEQUENCE
- 3 P-55: sytyke (0.38)
kindling (0.38)
- 4 P-55: °en tiedä (.) [mikä se on°]
°I don't know (.) [what it is°]
- 5 P-53: [entä sit] toi lamppu tai mikä toi (0.55)
[what about then] that lamp or what that (0.55)

(2a) Fishing rod, part 1 (pair 11, P-55)

- 1 P-55: ni mä mietin et tuisko sit sen (1.09)
so I was thinking that would it come after that (1.09)
- 2 P-55: * GAZE DIRECTED TO PAPER
- 3 P-55: tää (0.57) °kalastus::° (0.43)
this (0.57) °fishing::° (0.43)
- 4 P-55: * SHIFTS GAZE TOWARDS P-53
- 5 P-55: °mikä tää onkaa°
°whatever this is°
- 6 P-55: * GAZE BACK TO PAPER
- 7 (0.36)
- 8 P-53: on-
ro-
- 9 (0.26)
- 10 P-55: [nii onki nii]
[yes rod yes]
- 11 P-53: [onki onki]
[rod rod]
- 12 P-55: tulisko onki ennen kattilaa (1.30)
would the rod come before the pot (1.30)

Toivola, M., P. Lintunen & L. M. Heikkola (toim.) 2024. *Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. New directions in speech research – Freely produced speech as data.* AFinLA-teema / n:o 17, 167–190.

Sanna Olkkonen, Maarit Mutta & Pekka Lintunen
University of Turku

Fluency profiles in L2 monologue production: the role of underlying cognitive factors

Highlights

- Identifying cognitive fluency profiles helps to understand mechanisms behind speech production.
- Cluster analysis combining L1 and L2 measures is a valuable tool in the creation of fluency profiles.
- Learners with different profiles could benefit from different types of exercises in the classroom environment.

Abstract

This article examines the role of cognitive fluency and different fluency profiles that underlie L2 speech fluency, in relation to cognitive efficiency in L1 and L2, in a group ($n = 64$) of university students of English. The participants conducted, first, a freely produced monologue narration task in L1 and L2. Second, they performed two cognitive tasks in L1 and L2 measuring the efficiency of lexical access: a rapid word naming task and a Stroop task. Thirdly, L2 proficiency level was assessed with a vocabulary test. The correlations between proficiency and cognitive variables were low to moderate. Based on a cluster analysis, the participants were divided into four fluency profile groups differing in speech and disfluency rate, proficiency, and efficiency of processing. The results suggest that an awareness that learners represent various fluency profiles based on cognitive variables can aid formal teaching situations to include more versatile tasks for different kinds of learners.

Keywords: lexical access efficiency, monologue production, profiles, cluster analysis, speech fluency

Asiasanat: sanahaun tehokkuus, monologi, profiilit, klusterianalyysi, puheen sujuvuus

1 Introduction

Second language (L2) speech fluency is a core theme for both language users and those teaching or evaluating L2s for different educational or professional purposes. L2 fluency can be broadly understood as general oral proficiency or more narrowly as the speed and smoothness of production (Lennon 2000: 25). The narrow definition is common in fluency research and also followed here. Effortless and smoothly produced speech refers to speech without extensive pausing, repetitions, or self-repairs, but according to more recent studies, these so-called disfluency features are, nevertheless, an essential part of speech fluency both in first language (L1) and L2 production (see, e.g., Lintunen et al. 2020). This means that, in addition to relatively fast speech rate, even filled pauses and corrections can be features of efficient spoken production and strategically important for the flow of talk.

Language use has been studied in classroom and experimental conditions, and also in freely produced contexts. Study designs imitating authentic spoken production represent ideal contexts for collecting freely produced samples, but they induce other challenges, such as uncontrollable factors and variables. Therefore, a mixed setting including controlled tasks with controlled variables and freely produced speech can reveal new information on speakers' L2 proficiency and fluency profile most ideally, which benefits learners, teachers, and researchers. Earlier research has shown that there is a need for multiple-method studies (see Révész et al. 2022: 2) and especially for those where language users' own linguistic profiles and skills are compared (e.g., Lintunen et al. 2020).

L2 fluency is influenced by language proficiency and many individual factors relating to speech style and the efficiency of cognitive processing (e.g., De Jong 2018; Olkkonen & Mutta 2020; Peltonen 2020). It is important for both language teachers and assessors to understand that fluency features can be related to general language proficiency or may be more individual in nature. Instead of group-level analyses of fluency and proficiency, a more individual analysis can reveal the mechanisms of how speaking styles and differences in general cognitive processing have an effect on L2 fluency within speakers; especially with more proficient language users, these impacts are more highlighted. This can be seen in, for example, the frequency and duration of pauses, which are influenced by L1 speaking style, especially for proficient language users (e.g., Derwing et al. 2009; for fluency features in Finnish, see, e.g., Peltonen & Lintunen 2022; Toivola et al. 2009). In addition, repairs seem to depend on many factors, including resources for monitoring for mistakes and choosing to engage in a repair, as in whether the language user prefers speed or accuracy of the speech (see, e.g., the Complexity-Accuracy-Fluency framework, Kuiken & Vedder 2022).

In L2 research, there are studies on, for example, learner profile, aptitude profile, learning style profile, motivational profile, self-efficacy profile, and anxiety profile (see Li et al. 2022). Profiles can be used to describe the common tendencies and features of individuals or groups of individuals when learning or using a language (see, e.g., the Common European Framework of Reference, CEFR; Council of Europe 2001). Spoken L2 fluency has been studied from different perspectives (see, e.g., Lintunen et al. 2020), but studies focusing on fluency profiles are scarce. Our aim is to shed more light on this matter by examining how certain cognitive background features are related to L2 proficiency and to spoken production and fluency. The first aspect is approached with correlation tests and the second by performing a cluster analysis to examine how differences in lexical access efficiency in both L1 and L2 are related to L2 speech fluency in a group of university students of English. The participants ($n = 64$) conducted, first, a freely produced monologue task in L2 English. Second, they performed two cognitive lexical access tasks in L1 and L2, testing the automaticity of lexical access and the efficiency in its control. Furthermore, the participants' proficiency level was measured by a vocabulary test. By combining these perspectives, we can add to the understanding of different types of language users and their L2 fluency profiles that, in addition to more direct language proficiency measures, can be based on cognitive background factors behind speech production. The results can offer tools to identify fluency profiles and encourage L2 teachers to use more varied tasks that suit all kinds of learners to enhance their spoken production.

Our research questions are the following:

- 1) How is cognitive fluency related to L2 proficiency?
- 2) What kind of L2 fluency profiles can be identified based on cognitive fluency measures?

Before presenting our study, we discuss the dimensions of speech fluency and the different profile perspectives that have been studied in relation to speech fluency.

2 Literature review

2.1 Speech fluency: objective and subjective aspects

For speech fluency, the most oft-cited framework comes from Segalowitz (2010, 2016). In this framework, speech can be studied from three perspectives. Cognitive fluency is described as “the efficiency of operation of the underlying processes” of speaking (Segalowitz 2010: 165), incorporating the processes and abilities that make speech flow fluently on the utterance level and can usually be measured with multiple objective variables. Utterance fluency is the observable fluency of measurable aspects of speech, usually divided into speed, repair, and breakdown fluency (Skehan 2009, 2014; Tavakoli & Skehan 2005). Perceived fluency, the third perspective, is the subjective impression listeners have of the speaker’s production (e.g., Lintunen & Peltonen 2020). This study focuses on the link between cognitive and utterance fluency.

Utterance fluency in L2 is related to proficiency, but it is also affected by individual differences in speaking styles and cognitive processing skills, that is, cognitive fluency. There is a considerable amount of research on the features of fluent speech that may be the most reliable indicators of language proficiency. Studies have shown that speed (e.g., speech rate) and certain aspects of breakdowns (e.g., location of pauses) are most reliably linked to L2 proficiency (De Jong et al. 2013; Duran-Karaoz & Tavakoli 2020; Kahng 2020). Moreover, L2 speech rate seems to be significantly connected to L1 speech rate (Gagné et al. 2022; Peltonen 2020; Towell & Dewaele 2005), and it is possible that there is a level of proficiency in L2 after which the individual speaking style becomes more apparent (e.g., Derwing et al. 2009). Individual differences in speaking styles seem to affect especially the use of repairs in spoken production. According to research findings, for example, the number of repetitions (Olkonen 2017; Peltonen 2020) and corrections (Duran-Karaoz & Tavakoli 2020) in L2 speech are not related to proficiency in a reliable manner.

The reasons why the connection between proficiency and speech fluency might be difficult to define stem from multiple sources. Huhta et al. (2020) analysed fluency measures in language assessment and found that even if in most fluency scales certain features appear frequently (e.g., pauses and hesitation), the definition of fluency varies considerably. Individuals seem to pay attention to different subcomponents of fluency, which underlines the subjectivity of fluency perceptions. This kind of *perceived fluency* has been studied by, for example, Lintunen and Peltonen (2020), who examined university-level students’ (n = 71) perceptions and concluded that

learners had a multifaceted and rich understanding of what fluency means, but they mainly related it to spoken language, L2 proficiency, and a broad sense of fluency (Lennon 2000: 25). Even if fluency can be approached from objective or subjective perspectives, there seems to be a strong association between objective measures and subjective ratings according to Bosker et al. (2013), who reported that about 84% of the variance on ratings on fluency could be explained by objective measures of fluency, such as acoustic fluency features (i.e., speed fluency measured with the mean duration of syllables, breakdown fluency with the number of silent and filled pauses, and repair fluency with the number of repetitions and corrections). The aspects of breakdown and speed fluency seem to be the most strongly related to the subjective perception of fluency.

Besides these perceptions of fluency, differences in cognitive processing can also partly explain a lack of connection found between proficiency and fluency. Segalowitz's (2010) detailed model for *cognitive fluency* includes three aspects which determine fluent speech: the availability of *linguistic resources* (e.g., vocabulary size), *processing speed* (e.g., lexical or grammatical retrieval speed), and the efficiency of *attention control* (e.g., monitoring one's own speech for errors). Beginning language learners' cognitive resources are stretched by engaging in the multiple simultaneous processes of speech production, even with very low-level aspects of language use, which are, in general, automatically produced in L1 (such as correct pronunciation and word recognition, e.g., Levelt 1989; Towell & Dewaele 2005). Resources are freed by training these lower-level processes, but the differences between L1 and L2 performance might be quite notable even on high levels of proficiency, for instance in lexical access (Levelt 1989; Segalowitz 2010). Therefore, efficient lexical access is considered a central aspect of cognitive fluency (Segalowitz 2010: 75–76; Segalowitz & Freed 2004). A more automatic process is usually also more fluent, as it is less prone to interference from outside stimuli. Therefore, in lower proficiency levels, the lack of language skills hinders natural processing fluency. L2 cognitive fluency includes both the general cognitive processing capacity that is subject to individual variation, as well as L2-specific aspects, but their influence on L2 may, in part, be mediated by proficiency. To understand the differences in general and L2-specific processing, automaticity and control of language use differ between individuals can be studied (Segalowitz 2010; Olkkonen & Mutta 2020).

The control aspect, often studied with attention control, is argued to be one of the most important mechanisms differentiating expert language users from beginners (Segalowitz & Frenkiel-Fishman 2005). However, in contrast to automaticity, controlled processes do not seem to be as directly connected to language proficiency. In fact, Segalowitz and Freed (2004) found that efficiency of attention control was negatively related to L2 speech rate, which was explained as the better ability to monitor one's own speech. Furthermore, language users might struggle with multiple languages which they have simultaneously in their mental lexicon. It is more

difficult to inhibit the activation of the language one is more proficient in, and more cognitive resources are needed to suppress this kind of ballistic process (Meuter & Allport 1999). Therefore, the combination of automaticity and control as an indication of the efficiency of cognitive operations offers an insight into how closely L1 and L2 processing resemble each other within individuals. The hypothesis is that the more proficient one is in L2, the closer L2 processing aspects are to those of L1 (see, e.g., Marian et al. 2013), and the background speech processes in L2 approach those of L1 processing (Duran-Karaoz & Tavakoli 2020). These relationships have not yet, however, been studied together. By comparing these aspects within individuals, we also aim to obtain information on how individuals differ in their personal preferences and abilities in relation to these aspects.

2.2 Explicit and implicit profiles in spoken language studies

As mentioned above, L2 use contains individual characteristics such as individual speaking style and speech rate. L2 speech fluency is also affected by other user-internal (e.g., personality and affective variables like emotions) and user-external (e.g., formality of the situation, power relations) variables (Lintunen et al. 2020), which all contribute to creating language users' individual profiles. In research, there is a need for profile studies to understand the complex picture of individual L2 learners and users (Li et al. 2022). Next, we will discuss how profiles have been examined in relation to cognitive skills and L2 speech fluency. As the approach is relatively new, all studies below do not present learner or speaker profiles explicitly but describe individual styles or profiles implicitly.

In L2 studies, the examination of individual differences by creating profiles has been rarer than in psychological and educational research (see Sparks et al. 2012; Li et al. 2022). Sparks et al. (2012) performed a cluster analysis to study high school learners' (n=208) cognitive skills and L1 achievement skills with respect to L2 proficiency. They used a large-scale battery of cognitive tests (e.g., aptitude, intelligence) and L1 tests (e.g., the modern language aptitude test, MLAT). These measures were compared to the students' achievement after two years of high school L2 studies. Sparks et al. (2012) wanted to reveal patterns of differences, and based on their analysis, they identified three distinct profiles in the level of cognitive and L2 achievement. Their results indicate, for instance, that learners' L1 skills are strongly related to L2 aptitude and proficiency; however, they did not examine different types of cognitive profiles besides high- or low-achievement.

Cognitive fluency studies are often related to psycho- or neurolinguistically oriented frameworks, and profiles are formed primarily from a disfluency perspective, examining some speech-related disorders such as cluttering (e.g., Oliveira et al. 2010) or cerebral palsy (Hustad et al. 2010), even if an articulation difficulty does not

automatically make speech disfluent. The speech profiles are compared to fluent individuals' productions or to the comprehensibility of production, related to perceived fluency, which is not straightforwardly an objective measure because some assessment might reflect the speakers' skills, but some others, the listeners' skills. One cognitive study concerning profiles is Hustad et al. (2010), who studied speech and language problems in children with CP (with dysarthria) and examined speech and language measures (vowel area, speech rate, test of Auditory Comprehension of Language TACL-3) to create communication profiles groups. They combined the objective measures with a subjective one (parents' scaled assessment of the comprehensibility of the children's speech). Following a cluster analysis, the researchers created four profile groups based on children's speech motor involvement and possible language impairment. One of the objective measures, speech rate, was integral to the speech profiles, and this feature has also been widely used as a composite measure in L2 research as well, indicating the speed of talk and general speaking fluency (e.g., De Jong 2016; Skehan 2009, 2014; Tavakoli & Skehan 2005).

In her thesis, Dumont (2018) studied fluency and disfluency features to create non-native (French speakers of L2 English) and native (British English) speaker (dis-) fluency profiles based on utterance fluency. She used objective measures which combined product (i.e., the end product) and process level features, namely speech rate, mean length of run, phonation time ratio, unfilled pauses, filled pauses, restarts, false starts, repetitions, truncated words, foreign words, lengthenings, discourse markers, and conjunctions. Dumont (2018) used a cluster analysis to define the categories for the profile analysis. The analysis showed that temporal (dis)fluency variables had a central role in creating each profile. The most fluent non-native speaker profile had the following characteristics: high temporal fluency (few unfilled pauses, high phonation-time ratio and speech rate), many discourse markers, and few filled pauses, lengthenings and truncations; whereas fluent native speakers had a very high temporal fluency (few and short unfilled pauses, high phonation-time ratio and speech rate) and many lengthenings (Dumont 2018: 418–419). She also discovered that there was no relationship between the learners' fluency profiles and their assessed CEFR levels.

In a recent study, Mutta et al. (2023) made a comprehensive analysis of L2 fluency by examining objective variables of both written and spoken modes in a case study. They combined data from 11 university students who participated in two experiments. Students performed tasks in both L1 (Finnish) and L2 (English). The spoken tasks consisted of monologue picture description tasks, and the written tasks were short essays. The measured variables were chosen to represent central variables in speech (e.g., Peltonen & Lintunen 2022) and written fluency studies (e.g., Mutta 2020; Mutta & Laine 2022) and could be compared with each other illustrating the fluent-disfluent continuum; the used variables included, for example, speech rate, length of production, silent pauses per minute of speaking time, and

repairs per minutes of speaking time. These variables were used to create L1 and L2 fluency profiles. The results showed that four speaker profiles could be identified: fast and productive speaker, fast speaker, slow and productive speaker, and slow and reflective speaker. These same profiles were identified in the written production. Furthermore, it was found that six speakers had the same profile in spoken production in L1 and L2; on the other hand, only one participant had the same profile in L1 and L2 speaking and writing. Interestingly, these results were not related to learners' L2 proficiency level.

For their part, Saito et al. (2016) examined different lexical factors' impact on native-speaker raters' assessment of comprehensibility of L2 speech. Comprehensibility was defined as the ease of understanding learners' L2 production including how fluently they spoke without undue pauses and hesitations (i.e., the experienced perceived fluency). They collected 40 short narratives from French L2 English speakers to examine which vocabulary features influenced the evaluators' intuitive assessment of the comprehensibility of speech; they were evaluated by 10 raters. Saito et al. (2016: 686) analysed the narratives for 12 lexical variables (related to appropriateness, fluency, variability, sophistication, abstractness, and sense relations), and based on these variables, they distinguished beginner, intermediate, and advanced levels of L2 comprehensibility. Profiles were linked to these linguistic features, produced by the speakers and not to the speakers themselves in this study. Based on recent studies, De Jong (2016: 205) relates comprehensibility also to objectively measured disfluency features (e.g., filled pauses) when these are used for communicative purposes, for example, to inform the listener that the speaker is searching for an appropriate expression and the listener has to pay more attention to what is coming next. However, there is still a need for a better understanding of how the underlying mechanisms of fluency and disfluency are linked to these surface features.

The above-mentioned studies show that identifying any kind of profile is a multifaceted phenomenon and contains a lot of intraleaner variation. Studies focusing explicitly on L2 fluency profiles are scarce and, therefore, we try to shed light on this issue. Especially lacking is an understanding of the factors that underlie fluent speech, and as its relationship to cognitive fluency does not seem to be direct, it is important to strive for a more in-depth understanding of the different ways this relationship manifests itself. In our experimental study design, we approach the topic with objective measures that allow us to identify L2 fluency profiles with a cluster analysis, focusing on the efficiency of lexical access, speed and accuracy of speech, and L2 proficiency. The efficiency of lexical access was measured both by automaticity and attention control, which are thought to cover the complementary aspects of cognitive fluency (Segalowitz 2010; Olkkonen et al. in review). The aim was to study how L2 proficiency level is related to cognitive fluency and what kind of fluency profiles can be identified based on these measures.

3 Data and methods

3.1 Participants

The participants were native Finnish-speaking university students ($n = 64$; mean age (M) = 22.7, standard deviation (SD) = 5.5; 61% females) and advanced English users (number of years studying English: $M = 10.2$, $SD = 1.3$). LexTale vocabulary test (on the test and its validation, see Lemhöfer & Broersma 2012) was used to assess overall L2 proficiency level, with 70.3% of the participants on the C1/C2 level (LexTale accuracy score $M = 84.9$, $SD = 8.7$; C-level cutoff point 80%) in the CEFR scale (Council of Europe 2001). The testing was conducted in university language classes during a course as a voluntary task. The participants were asked about their vision (normal or corrected-to-normal), language-related problems, and colour vision impairments; if answers to any of these were positive, the test results were excluded from further analyses. All the tasks were conducted in the university language and computer laboratories, and the productions were recorded with microphone-headsets. Informed consent to participate in the study was obtained from the participants, and the data were processed according to the university's own ethical as well as GDPR guidelines.

3.2 Fluency tasks

The four objective measures in L1 Finnish and five in L2 English used in the current study are presented in Table 1.

TABLE 1. Profile dimensions and tasks in L1 and L2.

Profile dimensions	Task & measure
Language proficiency	LexTale (L2): lexical decision
Speech fluency	Monologue (L1 & L2): Speech rate, syllables / minute
Disfluency	Monologue (L1 & L2): false starts, reformulations, and repetitions / minute
Automaticity	Rapid word reading (L1 & L2): percentage of correct answers
Control of attention	Stroop task (L1 & L2): amount of interference (additional reaction time), number of repairs

The task to measure speech fluency was a freely produced monologue task in L1 Finnish and L2 English, based on two similar cartoon strips with six pictures. The participants

had two minutes' time to plan what to say and then another two minutes to tell a short story in their own words based on the cartoons. The cartoon strips and the order of language (L1 and L2) were counterbalanced between the participants (more on the procedure, see Peltonen 2020). The monologues were transcribed and annotated by student assistants using Praat (Boersma & Weenink 2018), and double-checked by the researchers. Based on these transcripts, for the speech fluency variable, speech rate (number of syllables, including filled pauses / total speaking time, multiplied by 60 to standardise per minute; see Kormos 2006) was counted. Speech rate, thus, includes speed and breakdown (pausing) fluency, and provides a composite measure to be used as a general fluency measure that correlates highly with, for instance, perceived fluency (see, e.g., Bosker et al. 2013). Filled pauses were included in the syllable count as they can be considered strategic ways to keep the talk flowing when encountering difficulties with the speech production (following, e.g., Kahng 2020; Peltonen 2020; Dumont 2018). To account for repairs, disfluency rate (total number of false starts, reformulations, and repetitions per minute) was calculated (Kormos 2006).

To measure cognitive fluency, following Segalowitz's (2010) division and previous research (Olkkonen 2017), we opted for one task measuring automaticity (processing of Rapid words) and one task measuring control of attention (Stroop task), as these were considered to cover both aspects of cognitive fluency. The cognitive fluency tasks were conducted in a computer laboratory using jsPsych 6.3 framework (De Leeuw 2015), counterbalancing the order of the tasks and languages between the participants. Automaticity was measured by accuracy in *Rapid words*, where the participants were shown 30 words per language, presented for 50 ms and replaced immediately by a mask. The instructions were to say the words aloud, as fast and accurately as possible, and the participants were encouraged to guess when not sure. The words ranged from 8 to 13 letters, and from 3000 to 9000+ in frequency bands of Finnish (Parole 2007) and English (Nation 2017), getting progressively longer and less frequent. The accuracy of the answers was assessed by two raters (the first author and a research assistant instructed by the first author), listening to the answers independently and scoring based on the degree of accuracy (e.g., target word *reactivate*, *ympärillään* > 1 point for correct answer, .5 point for a correct base word such as *reactive*, *ympärillä* and .25 for a correct part such as *react*, *ympäri*); interrater reliability between two raters with Cohen's Kappa was 0.87.

Control of attention was measured by a Stroop task (Stroop 1935). The participants were shown neutral and colour words, written in different colour inks, and the task was to name the colour of the ink. The task included four colour words (*punainen/red*, *sininen/blue*, *keltainen/yellow*, *vihreä/green*) and three neutral words (*ovi/door*, *ikkuna/window*, *tuoli/chair*). The words were written in red, blue, yellow, or green ink on a grey background. There were six conditions (neutral, congruent, and incongruent in L1 and L2): the congruent condition with the word printed in the corresponding colour (*punainen/red* in red); the incongruent condition with a

different colour (*punainen/red* in green); and the neutral condition, with a neutral word in any colour (*ovi/door* in red). Moreover, there were cross-language stimuli as the words shown could be either in L1 or L2, but the answering language was always in blocks. Reaction times (RTs) to name the ink were measured from the beginning of the correct answer (wrong answers and late responses of over 2100 ms, which constituted 0.6% of all answers, were excluded) using Praat (Boersma & Weenink 2018). Based on the mean RTs, *interference* effects (mean RTs in incongruent condition minus RTs in neutral condition) were calculated for both languages and cross-language conditions separately. The cross-language interference was used as an index of the strength of the involuntary activation of irrelevant language, and the ease of blocking it. Thus, a larger difference in RTs between the conditions, that is, a higher score, denotes a bigger effect. The number of repairs was also counted to account for the efficiency of the monitoring aspect of attention control. (Overall means and standard deviations in the tasks are reported in Olkkonen et al. in review).

3.3 Method

Sixty-four students' productions were analysed to study their cognitive fluency, L2 proficiency, and fluency profiles. To study the relationship between cognitive fluency and L2 proficiency, Pearson correlations were counted. To identify the profiles, we employed a hierarchical cluster analysis, the Ward method (Ward 1963). Cluster analysis is a method that groups cases together based on their similarity and thus helps to form subgroups of data sets that have high intra-group variation. In the case of language learning, for example, this can be used to form subgroups that "may utilise different pathways to language learning, including different strategies, aptitudes, motivational profiles, or different linguistic features to produce successful spoken or written language" (Staples & Biber 2015: 244; see also Dumont 2018). The hierarchical clustering method combines the closest cases together step by step and is, therefore, well-suited if the number of clusters is not decided beforehand. The distances between the clusters are computed by Euclidean distance and serve as an index of differences. All analyses were conducted on SPSS 27.

4 Results

To answer the first research question about the relationship between L2 proficiency and individual cognitive fluency variables, Pearson correlations were calculated (see Table 2). The connections were mostly low and non-significant. The correlations of accuracy in the L2 Rapid words and L1 Stroop interference to LexTale were moderate ($r = .33$). Rapid words task measures the depth and breadth of vocabulary skills, and

therefore, its connection to LexTale is not surprising. The negative connection between L1 Stroop and LexTale, on the other hand, could be attributed to good skills in control of attention (see, e.g., Ghaffarvand Mokari & Werner 2019). These connections, however, did not reach statistical significance after Bonferroni corrections.

TABLE 2. Pearson correlations between L2 proficiency (LexTale) and cognitive fluency variables.

Cognitive fluency variable	<i>r</i>
Rapid words L1	.22
Rapid words L2	.33
Stroop repairs L1	-.03
Stroop repairs L2	-.02
Stroop interference L1	-.33
Stroop interference L2	-.14
Stroop interference L1 > L2	-.16
Stroop interference L2 > L1	-.31

To answer the second research question on the different cognitive profiles in the background of L2 users' speaking profiles, we performed a cluster analysis based on the variables described in Table 1. The cluster analysis is by nature somewhat subjective, as deciding the number of meaningful clusters is often based on an examination of distribution and distances in a dendrogram, and the number of clusters varies accordingly. In the current study, based on the dendrogram (see Appendix 1; for the means and standard deviations, see Appendix 2), we opted for four clusters (henceforth, groups) (also, e.g., Mutta 2020; Mutta et al. 2023). This also aligned with the fluency dimensions used, that is, the groups based on these clusters were compared according to differences in proficiency, fluency and disfluency, automaticity, and control of attention.

Based on the means presented in Appendix 2, we examined the difference between the groups further with one-way between-group ANOVAs (as recommended, e.g., by Staples & Biber 2015) with Bonferroni corrections for multiple testing (significance level at $p < .005$). In the case of the Disfluency rate, there was no significant difference between the groups (L1: $F(3,60) = 1.97, p = .128$; and L2 $F(3,60) = .05, p = .986$), nor for L2 Stroop interference ($F(3,60) = 3.56, p = .019$). Group differences remained significant after the correction, however, between Speech rates (L1: $F(3,60) = 5.42, p = .002$; and L2 $F(3,60) = 5.22, p = .003$), L1 Stroop interference ($F(3,60) =$

21.38, $p < .001$), L1/L2 Stroop interference ($F(3,60) = 6.87, p < .001$), and L2/L1 Stroop interference ($F(3,60) = 9.63, p < .001$).

To conceptualise the groups as profiles, the results are presented in a schema table (see Table 3) by groups based on their proficiency, speed, (dis)fluency, automaticity, and interference (i.e., control of attention).

TABLE 3. Schema of fluency groups.

Profile dimension	Group 1	Group 2	Group 3	Group 4
Proficiency (LexTale)			+	-
Speed (SR)	+		-	L1 + / L2 -
Fluency (DR)	L1 + / L2 -		-	
Automaticity (RW)		-	+	L1 + / L2 -
Interference (Stroop)	+		-	L1 + / L2 -

Note. + = high score - = low score

Based on this schema, four different fluency profile groups were identified: 1) Fast and fluent, 2) Accuracy-oriented, 3) Proficient but disfluent, and 4) Less proficient but fast. These groups are further described below (see also Figures 1 and 2).

Group 1 "Fast and fluent"

This group included the fastest speakers in both L1 and L2; however, the proficiency scores, albeit high in general, were from the lower end among the participants (mean score in LexTale 80.23). Despite this, the group produced the least repairs in free speech in either language (L1 $M = 2.05$ and L2 $M = 3.27$), although the difference between the number of repairs in L1 and L2 was the highest in this group. They had the most repairs in the L2 Stroop task of the groups ($M = 1.73$), which might be related to the speed of production, as their reaction times did not suffer from this. This may reflect an emphasis on speed over accuracy. This was, moreover, the only group where L2 interfered more with L1 than vice-versa, which may indicate a fast activation of lexical material also in L2.

Group 2 "Accuracy-oriented"

The group was in the mid-range in both speech rate (L1 $M = 232.97$ and L2 $M = 153.59$) and disfluency rate (L1 $M = 2.67$ and L2 $M = 3.52$). They had the lowest accuracy in L2 rapid words ($M = 74.35$) and second lowest in the L1 version ($M = 45.49$), which indicates lower overall processing speed. In the Stroop task, this group had very few

repairs (L1 $M = 1.19$ and L2 $M = 0.63$); interestingly, they did experience interference in the task, which is seen in the longer reaction times (L1 $M = 115.56$ and L2 $M = 106.76$) compared to the Groups 1 (L1 $M = 106.40$ and L2 $M = 36.50$) and 3 (L1 $M = 57.44$ and L2 $M = 59.31$), but this did not translate into more repairs. These findings can point to either slightly lower automaticity (and therefore less need to correct) or to more emphasis on accuracy than speed, and therefore, not trying to guess the rapid words if not completely sure of the correct answer.

Group 3 "Proficient but disfluent"

This group had the highest scores in both proficiency (LexTale $M = 90.09$) and in the L2 rapid words' accuracy ($M = 62.03$), indicating that their L2 skills and automaticity were very high. Despite this, their speech rate was the lowest of all groups ($M = 141.90$). They also had the most disfluencies in L1 monologue ($M = 3.91$), and of all the groups, they were the only ones with more disfluencies in L1 than in L2 ($M = 3.43$). Furthermore, this group had very few repairs in the Stroop task in both L1 ($M = 1.06$) and L2 ($M = 0.88$), which may relate to slower speed in production. This slowness, however, seemed not to affect reception, as the rapid word scores were high (L1 $M = 83.91$ and L2 $M = 62.03$). Therefore, this group's performance in L1 and L2 resembled each other the closest of all groups, also regarding the disfluency behaviour.

Group 4 "Less proficient but fast"

This group was on the lowest level in the proficiency test along with Group 1 ($M = 80.83$), but compared to Group 1, Group 4 had the highest disfluency rate in L2; furthermore, the difference between their L1 ($M = 232.87$) and L2 speech rates ($M = 136.94$) was the largest. They had the highest accuracy scores in the L1 rapid words ($M = 89.03$), and even though they scored the second highest in L2 ($M = 55.28$), the discrepancy between the languages was, nevertheless, the largest in this measure as well. The difference between the strength of the Stroop effect between the languages (L1 $M = 320.60$ and L2 $M = 123.10$) was also highest in this group, all pointing to a less-developed and less-automatic L2 skill. However, it is important to notice that this did not translate into a slow speech rate, although with a lot of repairs. Furthermore, the relatively high score in L2 automaticity in rapid words may point to good overall automaticity which mediated even the lower language skill in the task.

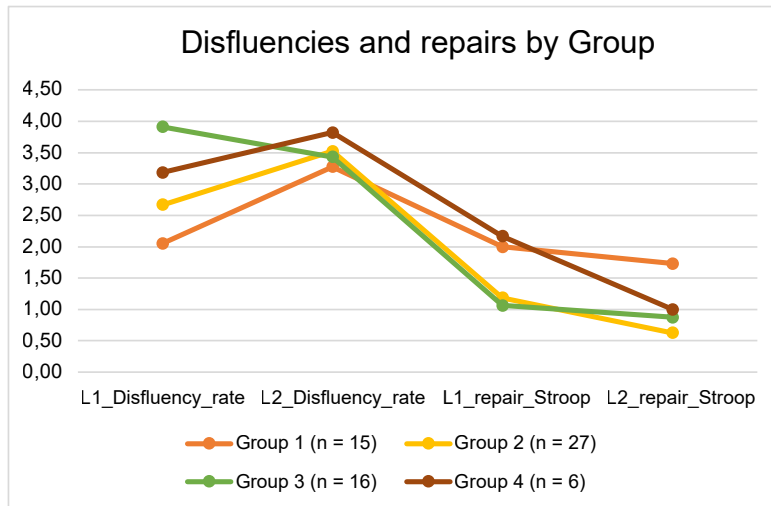


FIGURE 1. Means of disfluencies and repairs by profile groups.

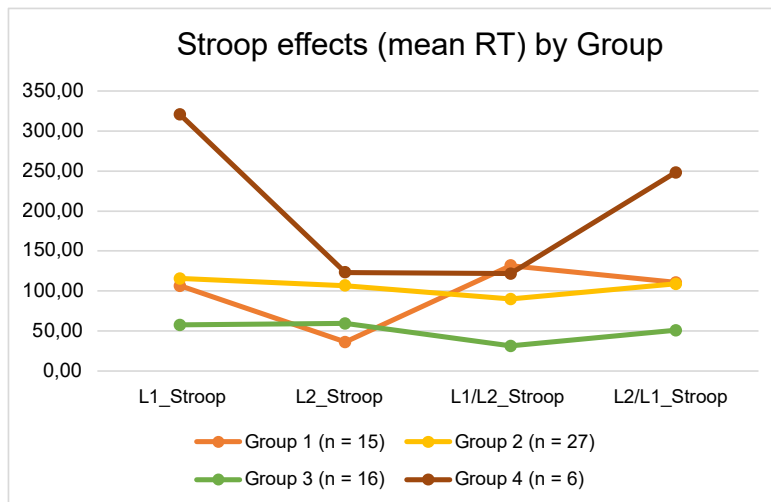


FIGURE 2. Means of different Stroop effects by profile groups.

5 Discussion and conclusion

In this article, we studied L2 learners' fluency profiles, based on cognitive fluency, L2 proficiency, and L2 speech production fluency. There is a need for multiple-method

studies where language users' individual profiles and skills are studied in their different languages to form a more comprehensive picture of L2 learners' competences. Our experimental multiple-method study design combined freely produced monologue tasks and cognitive lexical access tasks to allow us to create fluency profiles in L1 and L2 production based on utterance and cognitive fluency variables. Until today, there have been only a few studies concentrating directly on speech fluency profiles, especially from a cognitive processing viewpoint, and our study tries to fill this gap and complement other L2 fluency studies in the field.

The two research questions related to how L2 proficiency and cognitive fluency are linked together and what kind of L2 proficiency profiles can be identified based on these measures. Regarding the first research question, the correlations between the cognitive fluency variables and L2 proficiency were mostly low to moderate, suggesting that the cognitive fluency variables used here reflect more individual differences. For example, in the case of the number of repairs in the Stroop task, this is not very surprising, as the correlation between the number of repairs between languages (L1 and L2) has been shown to be high and significant (see Olkkonen et al. in review). Together these results confirm further that repair behaviour is strongly influenced by individual cognitive or speaking style differences (see, e.g., Duran-Karaoz & Tavakoli 2020; Olkkonen 2017; Peltonen 2020). Noticeably, our target group was advanced language users, and therefore, in line with previous studies (e.g., Derwing et al. 2009), it seems that the high proficiency level enables the application of personal speaking style.

To answer the second research question, we measured proficiency, speed of production, disfluency rate, automaticity, and interference in L1 and L2 to create four fluency profile groups by means of a cluster analysis. The main differences between the groups fell into two main strands: the relationship between proficiency and fluency, and the relationship between speed and accuracy.

On the connection between proficiency and fluency, earlier studies have shown that L2 speech fluency is related to proficiency level but different aspects have different impact on L2 speech production such as speech rate in L1 (Gagné et al. 2022; Peltonen 2020; Towell & Dewaele 2005) or L2 (De Jong et al. 2013; Duran-Karaoz & Tavakoli 2020; Kahng 2020), individual speaking style (Derwing et al. 2009), or use of repairs (Olkkonen 2017; Duran-Karaoz & Tavakoli 2020; Peltonen 2020). In the current study, the most proficient L2 profile (Group 3) was shown not to be the most fluent one when using L2 in freely produced speech. However, based on their L1 monologue results, we can see that their general speaking style in L1 was very similar, that is, they were quite slow and disfluent in producing a monologue in their L1 as well. Together with the observation that the difference in the speech rate between L1 and L2 for this group was the lowest, this finding supports the interpretation that the more proficient language users are, the closer their L2 speech comes to their personal speaking style in general. Group 4, with the lowest proficiency, also had

the biggest difference between the speech rates; however, it is notable that their speech rate was still slightly higher than that of Group 3. These two groups serve to illustrate the dissociation of proficiency and fluency, especially in the higher levels of proficiency, as Group 3 was characterised as proficient but disfluent, whereas Group 4 was less proficient but fast.

When speaking freely, as in the monologue task, there is the option of concentrating more on speed and unpaused speech, or ensuring as accurate production as possible. This may be, in part, related to personal differences and choices, and in the current study, we have focussed on how automaticity and interference as background factors can reveal this. In the current results, Group 1 showed fast performance in freely produced speech, and although they had a high number of repairs in the L2 Stroop task, this was not associated with slower naming of the colour. As the reaction times were measured from the beginning of the correct answer, this seems an indication of very fast performance. This was in contrast to Group 2, for whom most of the other measures were very close to Group 1, but for the low number of repairs and markedly slower reaction times in the L2 Stroop task. Fast reaction to the prompts produced more inaccurate answers, whereas taking more time assured more accurate answers in the case of interference. This is in line with some previous findings indicating that individual preferences can affect reaction times in the Stroop task (Shao et al. 2015).

It is noteworthy that, using the bilingual, mixed version of the Stroop task, we were able to find robust differences between the groups' processing abilities. The most significant group differences were found for the different Stroop interference variables. As these variables are based on reaction times in milliseconds, they are able to reveal minute differences in the processing of linguistic material. The high proficiency of Group 3 was shown in the high interference they experienced from the activating L2 words, as compared to the fluent Group 1, for whom interference was mainly from the L1 words.

There are also limitations regarding this study. Especially in cognitive fluency studies, the chosen cognitive measures have differed quite widely, and this has an impact on the results. Therefore, emphasising other cognitive processes and their relationship to the spoken outcome, the L2 fluency profiles in speech production could have also emerged quite differently. This should be considered whenever language learner/user profiles are created in later studies in order to allow comparison with other studies. The current choice of tests derived from processing-based interpretation of cognitive fluency, whereas, for example, Kahng (2020) based her cognitive fluency measures (e.g., picture naming and syntactic encoding) on Levelt's (1989) model, leading to more detailed results on different stages of speech production. As the proficiency differences between the participants were very small, future studies are required to see how the framework applies to lower levels of proficiency. In addition, although the proficiency test used here has been validated in relation to

an array of external tests, more fine-tuned and varied proficiency level assessments would provide the results with more robustness. These variables, combined with the chosen method of cluster analysis, highlight the exploratory nature of the current study. Regarding further studies, as the languages in question (English and Finnish) are not related or typologically close, an interesting follow-up would be to compare how closer languages might either help or interfere with the processing in the lexical access tests.

The importance of the identified L2 fluency profiles to language teaching and assessment is based on the fact that individual characteristics influence speech production and might intervene in assessing the production. As Huhta et al. (2020) stated, the definition of fluency in L2 assessment often varies considerably and those who assess oral production might pay attention more or less subjectively to different subcomponents of fluency. Although the correlations between individual cognitive variables and proficiency were moderate, cluster analysis revealed their important role in distinguishing fluency profiles. Cognitive background features can, therefore, provide valuable insights into how a person communicates in any language, and thus help to understand how these features interact with L2 oral production. In contrast to, for example, Sparks et al. (2012), who found three groups with quite distinct levels of proficiency, our results were more mixed. For instance, speech rate seems to be related to a learner's language proficiency, as the speaker might retrieve words more easily from their mental lexicon, but only to a degree, as it might also reflect individual characteristics in speaking style that are not related to language proficiency (see Group 1 vs. Group 4).

In educational contexts, information on different kinds of profiles might help learners and teachers to better understand the role of underlying cognitive factors in L2 fluency. Learners with different kind of speaking profiles benefit from divergent tasks to improve their speech fluency: for instance, an accuracy-oriented speaker could practise short improvised dialogues; a proficient but disfluent speaker might benefit from drill-type exercises; a less proficient but fast speaker could find support from supporting vocabulary list during exercises; and a fast and fluent speaker, being ready for the next level tasks, could benefit from more demanding exercises and language use contexts. The knowledge of various fluency profiles can further encourage teachers to create and use multiple types of exercises to motivate learners' practising their spoken language skills in a beneficial and supportive way.

Acknowledgments

We would like to thank the two reviewers for their insightful and constructive comments, which have helped to strengthen this article. The work was supported by the Research Council of Finland (project *Fluency and Disfluency Features in L2 Speech (FDF2)*, decision number 331903). Furthermore, we would like to thank MA Saara Kaltioma and the participants of the project course for their substantial help in data preparation.

Literature

- Boersma, P. & D. Weenink 2018. Praat: doing phonetics by computer [Computer program]. <http://www.praat.org/>. [accessed 17 Feb. 2021]
- Bosker, H. R., A.-F. Pinget, H. Quené, T. J. M. Sanders & N. H. De Jong 2013. What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30 (2), 159–175. <https://doi.org/10.1177/0265532212455394>
- Council of Europe 2001. *Common European framework of reference for languages: learning, teaching, assessment*. Strasbourg: Council of Europe. <https://rm.coe.int/1680459f97>
- De Jong, N. H. 2016. Fluency in second language assessment. In D. Tsagari & B. Jayanti (eds) *Handbook of second language assessment*. Berlin, Boston: De Gruyter Mouton, 203–218. <https://doi.org/10.1515/9781614513827-015>
- De Jong, N. H. 2018. Fluency in second language testing: insights from different disciplines. *Language Assessment Quarterly*, 15 (3), 237–254. <https://doi.org/10.1080/15434303.2018.1477780>
- De Jong, N. H., M. P. Steinel, A. Florijn, R. Schoonen & J. H. Hulstijn 2013. Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34 (5), 893–916. <https://doi.org/10.1017/S0142716412000069>
- De Leeuw, J. R. 2015. jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47, 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- Derwing, T. M., M. J. Munro, R. I. Thomson & M. J. Rossiter 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31 (4), 533–557. <https://doi.org/10.1017/S0272263109990015>
- Dumont, A. 2018. *Fluency and disfluency. A corpus study of non-native and native speaker (dis) fluency profiles*. An unpublished doctoral dissertation. Louvain: Université Catholique de Louvain. [accessed 17 Feb. 2021] https://dial.uclouvain.be/pr/boreal/object/boreal%3A198393/datastream/PDF_01/view
- Duran-Karaoz, Z. & P. Tavakoli 2020. Predicting L2 fluency from L1 fluency behavior: the case of L1 Turkish and L2 English speakers. *Studies in Second Language Acquisition*, 42 (4), 671–695. <https://doi.org/10.1017/S0272263119000755>
- Gagné, N., L. French & K. Hummel 2022. Investigating the contribution of L1 fluency, L2 initial fluency, working memory and phonological memory to L2 fluency development. *Language Teaching Research*. <https://doi.org/10.1177/13621688221076418>
- Ghaffarvand Mokari, P. & S. Werner 2019. On the role of cognitive abilities in second language vowel learning. *Language and Speech*, 62 (2), 260–280. <https://doi.org/10.1177/0023830918764517>

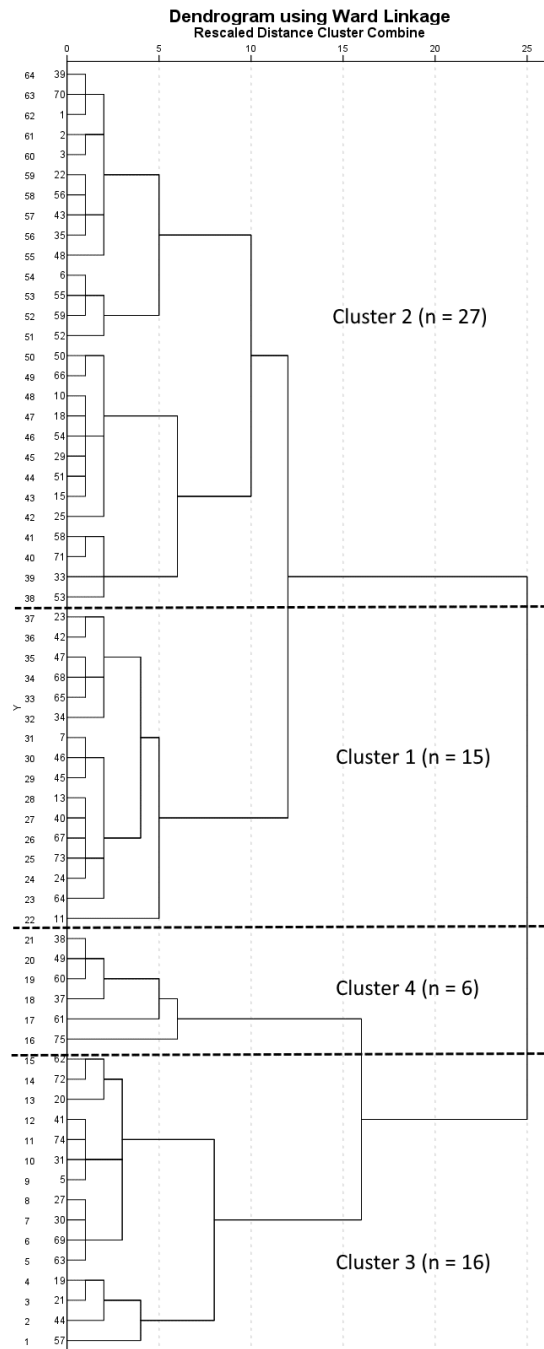
- Huhta, A., H. Kallio, S. Ohranen & R. Ullakonoja 2020. Fluency in language assessment. In P. Lintunen, M. Mutta & P. Peltonen (eds) *Fluency in L2 learning and use*. Bristol: Multilingual Matters, 129–145.
- Hustad, K. C., K. Gorton & J. Lee 2010. Classification of speech and language profiles in 4-year-old children with cerebral palsy: a prospective preliminary study. *Journal of Speech, Language, and Hearing Research*, 53 (6), 1496–1513. [https://doi.org/10.1044/1092-4388\(2010/09-0176\)](https://doi.org/10.1044/1092-4388(2010/09-0176))
- Kahng, J. 2020. Explaining second language utterance fluency: contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, 41 (2), 457–480. <https://doi.org/10.1017/S0142716420000065>
- Kormos, J. 2006. *Speech production and second language acquisition*. Mahwah: Lawrence Erlbaum Associates.
- Kuiken, F. & I. Vedder 2022. Complexity, accuracy, fluency, and functional adequacy (CAFFA). In L. Gurzynski-Weiss & Y. Kim (eds) *Instructed second language acquisition research methods*. Amsterdam: John Benjamins, 329–352.
- Lemhöfer, K. & M. Broersma 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44, 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Lennon, P. 2000. The lexical element in spoken second language fluency. In H. Riggensbach (ed.) *Perspectives on fluency*. Ann Arbor: University of Michigan Press, 25–42.
- Levelt, W. 1989. *Speaking: from intention to articulation*. Cambridge: MIT Press.
- Li, S., P. Hiver & M. Papi (eds) 2022. *The Routledge handbook of second language acquisition and individual differences*. New York & London: Routledge.
- Lintunen, P., M. Mutta & P. Peltonen 2020. Synthesising approaches to second language fluency: implications and future directions. In P. Lintunen, M. Mutta & P. Peltonen (eds) *Fluency in L2 learning and use*. Bristol: Multilingual Matters, 186–201.
- Lintunen, P. & P. Peltonen 2020. What is fluency? Learner perceptions of the concept. In P. Lintunen, M. Mutta & P. Peltonen (eds) *Fluency in L2 learning and use*. Bristol: Multilingual Matters, 16–33.
- Marian, V., H. K. Blumenfeld, E. Mizrahi, U. Kania & A.-K. Cordes 2013. Multilingual Stroop performance: effects of trilingualism and proficiency on inhibitory control. *International Journal of Multilingualism*, 10 (1), 82–104. <https://doi.org/10.1080/14790718.2012.708037>
- Meuter, R. & A. Allport 1999. Bilingual language switching in naming: asymmetrical costs of language selection. *Journal of Memory and Language*, 40 (1), 25–40.
- Mutta, M. 2020. L2 fluency and writer profiles. In P. Lintunen, M. Mutta & P. Peltonen (eds) *Fluency in L2 learning and use*. Bristol: Multilingual Matters, 63–80.
- Mutta, M. & P. Laine 2022. Tapaustutkimus monikielisten kirjoittajien kirjoitusprosessista: näppäilyntallennus apuna sanastokeskittymien tunnistamisessa [A case study of multilingual writers' writing processes: using keystroke logging to identify formulaic sequences]. *Lähivördlusi. Lähivertailuja*, 32, 120–149. <https://doi.org/10.5128/LV32.04>
- Mutta, M., P. Laine, P. Lintunen & P. Peltonen 2023. *Fluency across modes: towards a more comprehensive analysis of L2 fluency*. Presentation at the AILA 2023 conference, 17–21 July 2023. France: Lyon.
- Nation, I. S. P. 2017. *The BNC/COCA word family lists* (Version 1.0.0) [Data file]. <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists>. [accessed 17 Feb. 2021]
- Oliveira, C., A. Bernardes, G. Broglio & S. Capellini 2010. Speech fluency profile in cluttering individuals (original title: Perfil da fluência de indivíduos com taquifemia). *Pró-Fono Revista de Atualização Científica*, 22 (4), 445–450.

- Olkkonen, S. 2017. *Second and foreign language fluency from cognitive perspective: inefficiency and control of attention in lexical access*. Jyväskylä studies in humanities 314. Jyväskylä: University of Jyväskylä. <http://urn.fi/URN:ISBN:978-951-39-7062-8>
- Olkkonen, S. & M. Mutta 2020. Cognitive fluency in L2: what inaccuracies can reveal about processing and proficiency. In P. Lintunen, M. Mutta & P. Peltonen (eds) *Fluency in L2 learning and use*. Bristol: Multilingual Matters, 34–48.
- Olkkonen, S., P. Snellings, O. Veivo & P. Lintunen (in review). Cognitive fluency in L2: the effect of automatic and controlled lexical processing on speech rate.
- Parole. Kirjoitetun suomen kielen sanojen taajuuksia*. [Frequencies of written Finnish words] 2007. Kotimaisten kielten keskus [Institute for the Languages of Finland]. <https://kaino.kotus.fi/sanat/taajuuslista/parole.php>. [accessed 17 Feb. 2021]
- Peltonen, P. 2020. *Individual and interactional speech fluency in L2 English from a problem-solving perspective: a mixed-methods approach*. Turku: University of Turku. <https://urn.fi/URN:ISBN:978-951-29-8137-3>
- Peltonen, P. & P. Lintunen 2022. Multilingual speakers' L1, L2, and L3 fluency across languages: a study of Finnish, Swedish, and English. *Nordand*, 17 (1), 48–63. <https://doi.org/10.18261/nordand.17.1.4>
- Révész, A., M. Michel, X. Lu, N. Kourtali, M. Lee & L. Borges 2022. The relationship of proficiency to speed fluency, pausing, and eye-gaze behaviours in L2 writing. *Journal of Second Language Writing*, 58. <https://doi.org/10.1016/j.jslw.2022.100927>
- Saito, K., S. Webb, P. Trofimovich & T. Isaacs 2016. Lexical profiles of comprehensible second language speech. *Studies in Second Language Acquisition*, 38 (4), 677–701. <https://doi.org/10.1017/S0272263115000297>
- Segalowitz, N. 2010. *Cognitive bases of second language fluency*. New York: Routledge.
- Segalowitz, N. 2016. Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, 54 (2), 79–95. <https://doi.org/10.1515/iral-2016-9991>
- Segalowitz, N. & B. F. Freed 2004. Context, contact, and cognition in oral fluency acquisition: learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26 (2), 173–200. <https://doi.org/10.1017/S0272263104262027>
- Segalowitz, N. & S. Frenkiel-Fishman 2005. Attention control and ability level in a complex cognitive skill: attention-shifting and second language proficiency. *Memory & Cognition*, 33, 644–653. <https://doi.org/10.3758/BF03195331>
- Shao, Z., A. Roelofs, R. Martin & A. Meyer 2015. Selective inhibition and naming performance in semantic blocking, picture-word interference, and color-word Stroop tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41 (6), 1806–1820. <https://doi.org/10.1037/a0039363>
- Skehan, P. 2009. Modelling second language performance: integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30 (4), 510–532. <https://doi.org/10.1093/applin/amp047>
- Skehan, P. 2014. The context for researching a processing perspective on task performance. In P. Skehan (ed.) *Processing perspectives on task performance*. Amsterdam: John Benjamins, 1–26.
- Sparks, R. L., P. Patton & L. Ganschow 2012. Profiles of more and less successful L2 learners: a cluster analysis study. *Learning and Individual Differences*, 22 (4), 463–472. <https://doi.org/10.1016/j.lindif.2012.03.009>
- Staples, S. & D. Biber 2015. Cluster analysis. In L. Plonsky (ed.) *Advancing quantitative methods in second language research*. New York: Routledge, 243–274.

- Stroop, J. R. 1935. Studies of interference in serial verbal reaction. *Journal of Experimental Psychology*, 18 (6), 643–662. <https://doi.org/10.1037/h0054651>
- Tavakoli, P. & P. Skehan 2005. Strategic planning, task structure, and performance testing. In R. Ellis (ed.) *Planning and task performance in a second language*. Amsterdam: John Benjamins, 239–273.
- Toivola, M., M. Lennes & E. Aho 2009. Speech rate and pauses in non-native Finnish. *Proceedings of Interspeech 2009*. Brighton: International Speech Communication Association (ISCA), 1707–1710.
- Towell, R. & J.-M. Dewaele 2005. The role of psycholinguistic factors in the development of fluency amongst advanced learners of French. In J.-M. Dewaele (ed.) *Focus on French as a foreign language: multidisciplinary approaches*. Bristol: Multilingual Matters, 210–239.
- Ward Jr, J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236–244. <https://doi.org/10.2307/2282967>

APPENDIX 1.

Dendrogram obtained from Hierarchical Cluster Analysis: Speaker cluster groups across fluency and cognitive measures (Ward's method, Squared Euclidean Distance)



APPENDIX 2.

The means and standard deviations of the four cluster groups based on L2 proficiency, speech fluency, and cognitive fluency measures.

Measure	Group 1 (n = 15)	Group 2 (n = 27)	Group 3 (n = 16)	Group 4 (n = 6)
	M (SD)	M (SD)	M (SD)	M (SD)
Proficiency (LexTale accuracy %)	80.83 (6.84)	84.91 (8.38)	90.09 (6.97)	80.83 (12.29)
Speech rate L1 (syllables/min)	251.14 (33.31)	232.97 (33.65)	197.62 (44.75)	232.87 (49.04)
Speech rate L2 (syllables/min)	174.26 (25.67)	153.59 (21.60)	141.90 (27.65)	136.94 (35.91)
Disfluency rate L1 (number/min)	2.05 (3.03)	2.67 (1.83)	3.91 (2.01)	3.19 (1.95)
Disfluency rate L2 (number/min)	3.27 (2.80)	3.52 (3.22)	3.43 (3.49)	3.82 (2.19)
Rapid words L1 (accuracy %)	70.94 (18.58)	74.35 (18.47)	83.91 (14.30)	89.03 (8.34)
Rapid words L2 (accuracy %)	46.61 (17.13)	45.49 (22.27)	62.03 (17.55)	55.28 (12.52)
Stroop repairs L1 (mean number)	2.00 (2.33)	1.19 (1.69)	1.06 (1.00)	2.17 (1.94)
Stroop repairs L2 (mean number)	1.73 (2.09)	0.63 (0.88)	0.88 (1.02)	1.00 (1.10)
Stroop interference L1 (milliseconds)	106.40 (83.91)	115.56 (57.81)	57.44 (52.98)	320.60 (109.31)
Stroop interference L2 (ms)	36.50 (74.88)	106.76 (81.66)	59.31 (59.08)	123.10 (93.03)
Stroop interference L1>L2 (ms)	131.85 (68.01)	89.88 (72.94)	31.19 (45.40)	121.77 (63.85)
Stroop interference L2>L1 (ms)	110.65 (73.74)	108.87 (78.61)	50.78 (52.00)	248.03 (123.24)

Toivola, M., P. Lintunen & L. M. Heikkola (toim.) 2024. Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. *New directions in speech research – Freely produced speech as data.* AFinLA-teema / n:o 17, 191–215.

Elina Lehtilä, Pauliina Peltonen & Pekka Lintunen
University of Turku

Listener perceptions of L1 speech in L2 and L3 fluency assessment contexts: a qualitative approach

Highlights

- Listeners' target language fluency assessments are influenced by both temporal and non-temporal L1 features.
- There are individual differences across listeners in how they attend to L1 speech when assessing L2/L3 fluency.
- Half of the listeners felt that listening to L1 speech before assessing L2/L3 fluency facilitated the rating process.

Abstract

Studies on listeners' perceptions of speech fluency have typically focused on numeric L2 fluency assessments and rarely addressed individual differences in L1 speech fluency qualitatively. The objective of the present study was to examine how listeners attend to speakers' L1 speech when assessing the speakers' L2 and L3 fluency. In the study, 37 listeners were presented with freely produced L1 Finnish, L2 English (Experiment I), and L3 Swedish (Experiment II) speech by Finnish-speaking and Finnish–Swedish bilingual speakers (N=30). The listeners first listened to the speakers' L1 Finnish speech before assessing their target language fluency and then commented on how hearing the L1 speech samples affected their ratings. A qualitative analysis focusing on the listeners' comments revealed that listeners attend to both temporal and non-fluency related features in L1 speech when assessing target language fluency, but there were clear individual differences across listeners' intuitive approaches to L1 speech.

Keywords: fluency, speech perception, individual differences

Asiasanat: sujuvuus, puheen havaitseminen, yksilölliset erot

1 Introduction

Listener perceptions of a speaker's fluency play an essential role in all communicative settings. Examining subjective listener-based assessments of learners' speech fluency in different linguistic contexts and understanding the factors affecting listeners' perceptions of fluency is, therefore, crucial for the learning and teaching of second language (L2) spoken production skills. While a number of studies have investigated L2 perceived fluency from purely quantitative (e.g., Derwing et al. 2004; Bosker et al. 2013) and mixed-methods perspectives (e.g., Magne et al. 2019; Peltonen 2022), examining listeners' qualitative, freely produced comments across different target language assessment contexts can uncover important aspects of the factors underlying listeners' perceptions of fluency.

Over the last decade, connections between language learners' speech fluency in their first (L1) and target languages (L2, L3...) have received increased attention (e.g., De Jong et al. 2015; Duran-Karaoz & Tavakoli 2020; Peltonen & Lintunen 2022; Gao & Sun 2024), and the accumulating results from these studies indicate that several L2 fluency features are affected by L1 fluency features. However, to our knowledge, examining how a speaker's L1 fluency affects their target language fluency has not been studied from the listener's perspective, even though it has been suggested that individual differences in L1 speech fluency should be recognized in L2 fluency assessment contexts (De Jong 2018; Gao & Sun 2024). With a detailed qualitative analysis, the present study examines how listeners take L1 speech into account when assessing the speakers' fluency in the L2 and L3 and how helpful they perceive L1 speech to be in a fluency assessment context. Thus, the overall aim of the present

study is to gather empirical evidence on the connections between L1 and L2 fluency from the listener's perspective.

2 Approaches to listener perceptions of L2 fluency

In second language acquisition (SLA) research, fluency is traditionally defined as one of the three key components of L2 performance within the complexity, accuracy, and fluency framework (CAF; see, e.g., Housen et al. 2012), and it is typically regarded as a complex phenomenon between speaker and listener involving both speech production and perception (Lennon 2000). With reference to Lennon's (1990, 2000) distinction between broad and narrow senses of fluency, we approach fluency from the narrow sense, where speech fluency refers to the speed and efficiency of delivery, as opposed to global oral proficiency of the speaker (see also Lintunen et al. 2020). Using the term *perceived fluency* to refer to speech fluency from the listener's perspective, we subscribe to the following definition of the concept developed by Segalowitz (2010: 48): "the inferences listeners make about a speaker's cognitive fluency based on their perception of utterance fluency". Thus, although this study focuses on perceived fluency, we consider listeners' perceptions of fluency to be inherently linked to the actual temporal features of speech (i.e., *utterance fluency*) and the efficiency of the speaker's processes underlying speech production (i.e., *cognitive fluency*; Segalowitz 2010). The features of speech that are thought to reflect cognitive fluency are typically grouped into three main fluency dimensions: *speed* (speech tempo), *breakdown* (pausing), and *repair fluency* (corrections and hesitations; Tavakoli & Skehan 2005; Skehan 2009).

As stated in the introduction, L2 perceived fluency is often studied from a quantitative perspective, where numeric fluency assessments based on spoken productions are examined in terms of objective utterance fluency measures (for a recent meta-analysis, see Suzuki et al. 2021). Results from these studies indicate that listeners' perceived fluency ratings are strongly connected with measures related to speed (e.g., speech rate and articulation rate; Préfontaine et al. 2016; Kallio et al. 2023) and breakdown fluency (e.g., frequency, duration and distribution of silent pauses; Kahng 2018), and to a lesser extent with repair fluency measures (e.g., frequency of corrections; Cucchiari et al. 2002; Bosker et al. 2013). While perceived fluency has rarely been studied from a purely qualitative perspective (for an exception, see Préfontaine & Kormos 2016), a number of studies have approached L2 perceived fluency using mixed-methods designs, where quantitative analyses are typically complemented with qualitative analyses based on post-assessment interviews (Freed 1995; Magne et al. 2019; Suzuki & Kormos 2020) or raters' written comments (Kormos & Dénes 2004; Rossiter 2009; Peltonen 2022). These qualitative analyses have demonstrated that L2 fluency ratings tend to be affected by non-fluency

related factors, even when the listeners have been instructed to base their assessments on temporal and hesitation phenomena (Rossiter 2009; Magne et al. 2019). The factors that have been found to influence listeners' perceptions of L2 fluency include other linguistic aspects of language production such as pronunciation, vocabulary, and grammar, but also non-linguistic speaker attributes such as perceived confidence (Freed 1995; Rossiter 2009) and nativelikeness (Préfontaine & Kormos 2016; Magne et al. 2019).

While a consistent finding in several utterance fluency studies using correlation and regression analyses has been that there are connections between learners' L1 and L2 fluency features (e.g., De Jong et al. 2015; Duran-Karaoz & Tavakoli 2020), perceived fluency studies have rarely addressed connections between L1 fluency and listeners' assessments of L2 fluency. Some utterance fluency studies also indicate that the connections between L1 and L2/L3 fluency measures are potentially influenced by cross-linguistic differences and learners' proficiency level in the target language (Huensch & Tracy-Ventura 2017; Peltonen 2018; Peltonen & Lintunen 2022). In one of the few studies examining L1 and L2 fluency links in an assessment context, Pinget et al. (2014) examined connections between L2 Dutch fluency ratings and both traditional L2 fluency measures and L1-adjusted fluency measures (residuals obtained from correlating L1 and L2 fluency measures) for L1 Turkish and English (N=30) speakers. The L1-adjusted measures were not found to be better predictors of perceived fluency than the traditional fluency measures, which is perhaps unsurprising since, as acknowledged by the authors, their listeners only had access to the speakers' L2 Dutch speech.

In a different approach, Derwing et al. (2009) examined the relationship between L1 and L2 fluency across 16 Mandarin- and 16 Russian/Ukrainian-speaking immigrants to Canada by collecting fluency ratings based on both the learners' L1 and L2 English productions. For both learner groups, positive correlations were found between the L1 and L2 fluency ratings at the first data collection time, but not at the later stages of data collection over the course of two years. Although these results point towards a complex relationship between L1 and L2 fluency, the authors recognize that their assessment data cannot be used for a direct comparison of L1 and L2 perceived fluency, as different listeners (native speakers of the learners' L1s vs. native speakers of English) were used to rate the L1 and L2 speech samples. As neither Pinget et al. (2014) nor Derwing et al. (2009) examined connections between L1 and L2 perceived fluency from the perspective of the same listeners, how listeners attend to L1–L2 connections when assessing L2 fluency remains an unexplored area in L2 fluency research.

To address this gap in research, the present study examined how listeners intuitively attend to speakers' L1 Finnish speech when assessing their target language (L2 English/L3 Swedish) fluency. Using freely produced L1 Finnish, L2 English, and L3 Swedish speech recorded from Finnish-speaking and Finnish–Swedish bilingual

speakers, the study included two experiments: in Experiment I, we examined all speakers' (N=30) L2 English perceived fluency, and, in Experiment II, the Finnish-speaking participants' (n=20) L3 Swedish perceived fluency. In both experiments, the listeners (N=37) were first presented with short excerpts from the speakers' L1 Finnish samples and then asked to assess their target language fluency.¹

The data analyses focused on the listeners' open-ended comments on how hearing the Finnish speech samples affected their rating process. Furthermore, we examined the listeners' numeric evaluations of how helpful they felt the Finnish speech samples were in rating the target language speech samples. The study addressed the following research questions:

- 1) Which features and themes can be identified in the listeners' comments regarding the influence of L1 speech on L2 and L3 fluency assessment?
- 2) How do the themes and features identified from the comments relate to the listeners' helpfulness evaluations?

3 Methodology

3.1 Speech data

The speech data used in the present study were collected as part of a project Fluency across Multilingual Speakers (MultiFluency; funded by the Swedish Cultural Foundation in Finland) and selected based on our previous research on these speakers' utterance fluency in Finnish, Swedish, and English (see Peltonen & Lintunen 2022). The speakers were 20 Finnish-speaking and 10 Finnish–Swedish bilingual university students of language subjects from two Finnish universities, most of whom studied English as either their major (n=21) or minor (n=5) subject (see Table 1 for a summary of the speakers' background information). The target language speech samples used in Experiment I consisted of the L2 English speech samples from all 30 speakers, while the samples used in Experiment II consisted of the L3 Swedish speech samples produced by the 20 Finnish-speaking participants. All speakers volunteered to participate in the study, and informed consent was obtained from the speakers prior to participating.

To elicit spontaneous speech, all speakers were asked to perform cartoon-based narratives in Finnish, Swedish, and English. The speakers were presented with three different cartoon strips consisting of six pictures and given two minutes to plan their

1 The experiment design also included numeric fluency assessments from listeners who only heard the speakers' L2/L3 speech samples, making the total number of listeners in the two experiments 72. As the present study examines listeners' perceptions of L1 speech from a qualitative perspective, here we focus only on the listeners who had access to the speakers' L1 speech samples.

narrative before performing the task in each of the three languages. The speakers were not allowed to take notes during the planning time, but they could look at the cartoons when they were speaking. Two of the cartoon strips used in the study had been employed in previous research projects (e.g., Peltonen & Lintunen 2016), and the third cartoon strip was selected based on its comparability with the other two prompts in terms of key visual elements and narrative structure. In the first cartoon, a man sat down next to two people on a bench and lifted his hat in greeting, revealing a squirrel underneath. In the second cartoon, a man came across a group of children building a snowman, and the children asked the man to help them put a hat on the snowman’s head. The third cartoon showed two children growing a tree, reflecting their growth into adults. The first task was always done with the same cartoon in either Finnish or Swedish depending on the speaker’s L1, but the order of the additional two languages and the additional two cartoons was counterbalanced across speakers (for details, see Peltonen & Lintunen 2022). To ascertain the speakers’ overall level of oral proficiency in the target languages, four trained raters evaluated the L2 English and L3 Swedish speech samples based on CEFR criteria for qualitative aspects of spoken language use (range, accuracy, fluency, and coherence; Council of Europe 2001: 28–29). All speakers were assessed to represent level C1 on average in spoken L2 English and the Finnish-speaking participants level A2 on average in spoken L3 Swedish.

TABLE 1. Speakers’ background characteristics by language group.

Background variable	L1(s)			
	Finnish (n=20)		Finnish and Swedish (n=10)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age (years)	20.60	2.04	27.00	8.38
Duration of English studies at school (years)	10.18	1.09	10.33	2.40
Duration of Swedish studies at school (years)	6.63	1.56		
Self-rated English spoken production (1–5*)	4.0	0.65	4.1	0.74
Self-rated Swedish spoken production (1–5*)	1.85	1.04		

*Scale: 1 = weak, 2 = moderate, 3 = good, 4 = very good, 5 = excellent.

For the assessment tasks, all speech samples were first resampled to a sampling frequency of 44100 Hz and scaled to an average intensity of 70 dB (see, e.g., Pinget et al. 2014). For each L1 Finnish narrative, approximately 20-second excerpts were extracted from the beginning of the original recording, excluding any initial hesitations and ending at a natural break in speech. Based on the piloting of the two experiments, the excerpts were deemed sufficient to get an impression of the speaker's L1 speaking style. However, to increase the ecological validity of the assessment, the entire target language recordings were presented to the listeners (Préfontaine et al. 2016; Suzuki & Kormos 2020). The mean duration of the target language recordings was 64.60 s in Experiment I ($SD = 29.34$) and 56.75 s in Experiment II ($SD = 19.02$).

3.2 Listeners and assessment task

Skilled non-native speakers of the target language (see, e.g., Rossiter 2009; Magne et al. 2019) were recruited as listeners in the present study. The listeners in Experiment I ($n=15$; 11 female, 3 male, 1 preferred not to say, mean age 27.07 years) were advanced Finnish-speaking English majors at a Finnish university. The average duration of the listeners' English studies at university was 4.17 years ($SD = 0.92$), and 12 listeners had either completed or were participating in teacher training at the time of the study. The listeners in Experiment II were advanced Finnish-speaking university students of Swedish ($n=22$; all female, mean age 23.86 years) from two Finnish universities. The listeners had studied Swedish at university for 3.86 years on average ($SD = 0.79$). Sixteen listeners studied Swedish as their major subject and six as their minor subject, and all listeners had either completed or were participating in teacher training. The listeners in both experiments reported high familiarity with Finnish-accented target language speech on a scale from 1 to 6 (Experiment I: $M = 4.93$, $SD = 1.22$; Experiment II: $M = 4.91$, $SD = 1.06$). While most listeners had no prior experience in language assessment, seven listeners in Experiment I and nine listeners in Experiment II reported having some, relatively limited previous experience in assessing learners' spoken or written language skills. All listeners reported normal hearing and completed the assessment task as part of an MA-level course ($n=33$) or as volunteers ($n=4$). All listeners participated in the study voluntarily and gave informed consent to participate in the study before beginning the assessment task.

The assessment tasks were implemented online using REDCap, a secure data collection software hosted at the University of Turku (Harris et al. 2009; Harris et al. 2019). All listeners completed the assessment task unsupervised on their own computers and received all the instructions in written format (see Appendix). The listeners were first presented with a narrow definition of fluency, focusing on the temporal aspects of speech, and then instructed to rate each speaker's target lan-

guage speech rate, pausing, corrections, and overall fluency (based on their ratings of speech rate, pausing, and corrections) using 9-point scales (see, e.g., Rossiter 2009; Bosker et al. 2013; Kahng 2018; Magne et al. 2019; Suzuki & Kormos 2020). The inter-rater reliability was high for all rated aspects in both experiments, and the results regarding these numeric assessments are reported in Lehtilä et al. (in review). To mitigate the potential influence of familiarity bias, the participants were also introduced to the cartoons used in the speaking tasks before they began the assessment task (see, e.g., Derwing et al. 2004; Rossiter 2009; Kahng 2018; Suzuki & Kormos 2020). Following two practice trials, the target language speech samples were presented to the listeners in one of three pseudo-randomized orders to minimize potential order effects. Regarding the L1 Finnish speech samples, the listeners received the following instructions: "Before each [target language] sample, you will hear a short speech sample in the learners' first language (Finnish). Listen to this speech sample once before listening to the [target language] sample to get an impression of the learner's individual speaking style, but do not rate the Finnish sample." Thus, the listeners were not instructed to assess the speakers' L1 speech or to compare first and target language fluency features, as the objective was to examine how listeners intuitively attend to L1 speech when assessing target language fluency.

After completing the assessment task, the listeners were asked to evaluate how helpful the Finnish speech samples were in rating the L2/L3 speech samples using a scale with labeled endpoints (1 = not helpful at all and 6 = very helpful). After assigning the score for the overall helpfulness, the listeners were asked to comment freely on how they themselves felt hearing the Finnish speech samples had affected their rating process. These comments and the numeric helpfulness evaluations form the qualitative and quantitative data of the present study.

3.3 Analysis

To answer the first research question, the listeners' comments on how they felt hearing the speakers' L1 Finnish speech samples affected their L2/L3 fluency assessments were analyzed qualitatively using qualitative content analysis (Dörnyei 2007). Due to limited prior research in the area, the comments were coded in an inductive approach, where themes and patterns are identified from the data without the imposition of preconceived labels or categories (Miles & Huberman 1994; Dörnyei 2007). As each listener provided one comment at the end of the assessment task, altogether 37 comments were analyzed. The coding frame was developed and evaluated following O'Connor and Joffe's (2020: 20) recommendations for intercoder reliability assessment in qualitative research. Each comment was treated as a single data unit, and the initial coding frame was first developed and applied to the data by the first author, after which the second author independently coded approximately 50% of the data using the initial

coding frame. Reliability statistics (Krippendorff's alpha coefficient; Krippendorff 2004) were then calculated in R 4.0.4 (R Core Team 2021) using the irr R package (v0.84.1; Gamer et al. 2019) and interpreted individually for each code. The four codes that fell short of the acceptability threshold ($\alpha \geq .800$; Krippendorff 2004) were discussed and revised, and the first and second author independently coded the remaining data in accordance with the revised coding frame. After the revision, all codes except for one reached sufficient intercoder reliability. The code with an insufficient reliability value was excluded from the finalized coding frame.

The finalized coding frame based on the listeners' comments included five main categories: (1) Degree of effect, (2) Effect type, (3) Fluency features, (4) Speaker characteristics, and (5) Listener characteristics. Each main category had two to four subcategories, which are discussed in detail in the following section. To answer the second research question, cross-tabulations between the coded subcategories and the listeners' helpfulness evaluations were prepared to examine whether any patterns would be detected between the listeners' comments and their evaluations regarding the helpfulness of the L1 Finnish speech samples in their L2/L3 rating process.

4 Results

4.1 Listeners' perceptions of L1 speech in L2 and L3 fluency assessment

The subcategories for the five main categories of the finalized coding frame are presented in Tables 2–6, including subcategory descriptions and the percentages of comments in each subcategory in Experiments I (L2 English) and II (L3 Swedish). The first main category 'Degree of effect' reflected the magnitude of the effect that the L1 speech samples had on the L2/L3 fluency assessments, as estimated by the listeners themselves. As illustrated in Table 2, the most common degree of effect in Experiment I was mild (40.0%), whereas the most common degree of effect in Experiment II was moderate (45.5%). In addition, there were more listeners in Experiment I whose comments were classified under the subcategory 'No effect' (20.0%) compared to Experiment II (9.1%). Very few listeners in both experiments stated or implied that listening to the L1 speech samples had had a significant effect on their L2/L3 fluency assessments (Experiment I: 6.7%; Experiment II: 9.1%; note that the percentages in Experiment II do not add up to 100% due to rounding).

TABLE 2. Subcategories for the main category 'Degree of effect' and the percentages of all comments in Experiments I (L2 English) and II (L3 Swedish) for each subcategory.

Subcategory	Description	Percentage of all comments	
		Experiment I (L2 English)	Experiment II (L3 Swedish)
No effect	Listener states/implies that the L1 samples did not have any effect on their ratings	20.0%	9.1%
Mild	Listener states/implies that the L1 samples had or might have had a mild, potentially unspecified effect on their ratings	40.0%	36.4%
Moderate	Listener states/implies that the L1 samples had or might have had a moderate effect on their ratings and identifies at least one feature they attended to in the L1 samples	33.3%	45.5%
Large	Listener states/implies that they actively used the L1 sample (e.g., for all the speakers) in the rating process at least for one specified feature	6.7%	9.1%

The second main category, 'Effect type', reflected the ways in which the listeners reported the L1 speech samples had affected their rating process (note that, from Table 3 onwards, the percentages do not add up to 100% as one comment could be coded in multiple categories). As indicated in Table 3, the most common effect type was 'Facilitative' in both experiments (Experiment I: 53.3%; Experiment II: 50.0%), i.e., the listeners stated or implied that hearing the L1 speech samples made the rating process easier in some way (e.g., helped to prepare them for the L2/L3 samples). Approximately a quarter of the participants in both experiments reported consciously comparing the L1 sample to the L2/L3 sample for at least one specified feature (Experiment I: 26.7%; Experiment II: 27.3%). In addition, a few listeners in both experiments indicated that the L1 samples were not helpful in assessing L2/L3 fluency (Experiment I: 13.3%; Experiment II: 9.1%).

TABLE 3. Subcategories for the main category 'Effect type' and the percentages of all comments in Experiments I (L2 English) and II (L3 Swedish) for each subcategory.

Subcategory	Description	Percentage of all comments	
		Experiment I (L2 English)	Experiment II (L3 Swedish)
Facilitative	Listener states/implies that the L1 samples made the rating process easier in some way	53.3%	50.0%
Comparative	Listener states/implies that they consciously compared the L1 samples to the L2/L3 samples	26.7%	27.3%
Non-facilitative	Listener states/implies that the L1 samples were not helpful in the rating process for some or all of the rated variables	13.3%	9.1%

The third main category covered the three fluency features the listeners were instructed to base their L2/L3 fluency assessments on (speech rate, pausing, and corrections). As can be seen from Table 4, the most commonly mentioned fluency feature in both experiments was speech tempo (Experiment I: 60.0%; Experiment II: 45.5%), whereas pauses were less frequently mentioned (Experiment I: 20.0%; Experiment II: 13.6%). Very few listeners in Experiment II (9.1%) and none of the listeners in Experiment I mentioned attending to hesitations or corrections in the speakers' L1 speech.

TABLE 4. Subcategories for the main category 'Fluency features' and the percentages of all comments in Experiments I (L2 English) and II (L3 Swedish) for each subcategory.

Subcategory	Description	Percentage of all comments	
		Experiment I (L2 English)	Experiment II (L3 Swedish)
Speech tempo	Listener mentions speech rate as one of the features they attended to in the L1 speech samples	60.0%	45.5%
Pauses	Listener mentions pauses as one of the features they attended to in the L1 speech samples	20.0%	13.6%
Hesitations and corrections	Listener mentions corrections or hesitations as one of the features they attended to in the L1 speech samples	0.0%	9.1%

The two remaining main categories consisted of two subcategories each, covering all other speaker (Table 5) and listener (Table 6) characteristics that were identified from the comments. In addition to the three fluency features, listeners in both experiments reported attending to other, non-fluency related features (e.g., vocabulary choices or narrative strategies) in the L1 samples (Experiment I: 26.7%; Experiment II: 22.7%). Some listeners also mentioned typological or proficiency-related differences between the two languages spoken by the participants (Experiment I: 6.7%; Experiment II: 9.1%). While the frequencies of comments in the subcategories for ‘Speaker characteristics’ were very similar in both experiments, both subcategories for ‘Listener characteristics’ were more prominent in Experiment I compared to Experiment II. As summarized in Table 6, compared to Experiment II, more listeners in Experiment I reported attempting to consciously focus only on the L2 samples in the rating process (26.7%) and mentioned their personal attitudes, expectations or behaviors regarding the rating procedure (33.3%).

TABLE 5. Subcategories for the main category ‘Speaker characteristics’ and the percentages of all comments in Experiments I (L2 English) and II (L3 Swedish) for each subcategory.

Subcategory	Description	Percentage of all comments	
		Experiment I (L2 English)	Experiment II (L3 Swedish)
Personal speaking style	Listener mentions other, non-fluency related features they attended to in the L1 speech samples	26.7%	22.7%
L1–L2/L3 difference	Listener mentions either typological differences or differences in speakers’ proficiency levels between the L1 and L2/L3	6.7%	9.1%

TABLE 6. Subcategories for the main category 'Listener characteristics' and the percentages of all comments in Experiments I (L2 English) and II (L3 Swedish) for each subcategory.

Subcategory	Description	Percentage of all comments	
		Experiment I (L2 English)	Experiment II (L3 Swedish)
Deliberate disregard	Listener states/implies attempting to consciously focus only on the L2/L3 samples and/or ignore the L1 samples in the rating process	26.7%	9.1%
Personal responses	Listener mentions their personal attitudes, expectations or behaviors regarding the rating procedure	33.3%	18.2%

4.2 Relationship between the listeners' comments and their helpfulness evaluations

The distribution of the listeners' evaluations regarding the helpfulness of the L1 Finnish speech samples when assessing L2/L3 fluency in Experiments I and II are illustrated in Figure 1. As shown in the figure, the majority of the responses in Experiment I fell in the middle of the scale (3–4; $M = 3.67$, $SD = 1.23$), whereas the most common response in Experiment II was 2 ($M = 3.09$, $SD = 1.31$). Furthermore, some listeners in Experiment II evaluated the L1 samples as 'not helpful at all' (1), and none of the listeners evaluated the L1 samples as 'very helpful' (6). Overall, the distributions suggest that the listeners in Experiment II felt that the L1 Finnish samples were less helpful in assessing target language fluency than the listeners in Experiment I.

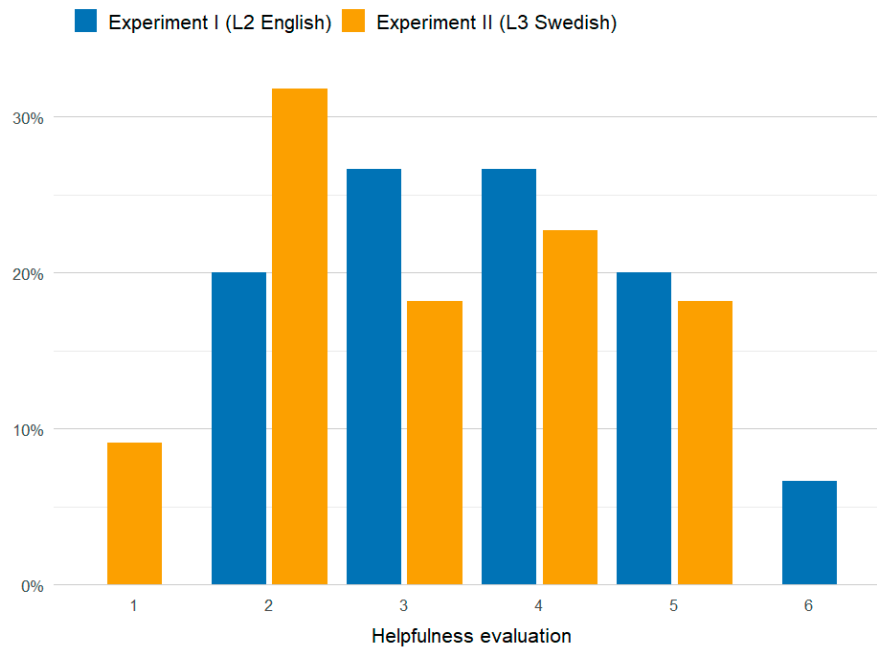


FIGURE 1. Percentage of responses regarding the perceived feeling of helpfulness of the L1 Finnish speech samples when assessing fluency (scale 1–6; 1 = not helpful at all, 6 = very helpful) in Experiments I (L2 English) and II (L3 Swedish).

The cross-tabulations between the coded subcategories and the listeners' helpfulness evaluations are summarized in Tables 7 (Experiment I) and 8 (Experiment II).

TABLE 7. Cross-tabulations of the coded subcategories according to the helpfulness evaluations in Experiment I (L2 English).

Subcategory	Helpfulness evaluation				
	2	3	4	5	6
Degree of effect					
No effect	66.7%	33.3%	–	–	–
Mild	16.7%	33.3%	33.3%	16.7%	–
Moderate	–	20.0%	40.0%	40.0%	–
Large	–	–	–	–	100.0%

Subcategory	Helpfulness evaluation				
	2	3	4	5	6
Effect type					
Facilitative	25.0%	25.0%	12.5%	25.0%	12.5%
Comparative	–	25.0%	50.0%	25.0%	–
Non-facilitative	50.0%	50.0%	–	–	–
Fluency features					
Speech tempo	11.1%	22.2%	44.4%	11.1%	11.1%
Pauses	–	–	33.3%	33.3%	33.3%
Speaker characteristics					
Personal speaking style	–	–	25.0%	50.0%	25.0%
L1–L2 difference	100.0%	–	–	–	–
Listener characteristics					
Deliberate disregard	75.0%	25.0%	–	–	–
Personal responses	20.0%	20.0%	60.0%	–	–

TABLE 8. Cross-tabulations of the coded subcategories according to the helpfulness evaluations in Experiment II (L3 Swedish).

Subcategory	Helpfulness evaluation				
	1	2	3	4	5
Degree of effect					
No effect	–	100.0%	–	–	–
Mild	25.0%	50.0%	25.0%	–	–
Moderate	–	10.0%	20.0%	50.0%	20.0%
Large	–	–	–	–	100.0%
Effect type					
Facilitative	–	18.2%	27.3%	27.3%	27.3%
Comparative	–	–	16.7%	33.3%	50.0%

Subcategory	Helpfulness evaluation				
	1	2	3	4	5
Non-facilitative	50.0%	50.0%	–	–	–
Fluency features					
Speech tempo	–	20.0%	30.0%	20.0%	30.0%
Pauses	–	–	100.0%	–	–
Hesitations and corrections	–	50.0%	–	–	50.0%
Speaker characteristics					
Personal speaking style	–	–	–	80.0%	20.0%
L1–L3 difference	50.0%	50.0%	–	–	–
Listener characteristics					
Deliberate disregard	50.0%	50.0%	–	–	–
Personal responses	–	50.0%	25.0%	25.0%	–

Regarding the degree of effect, tables 7 and 8 show that in both experiments the larger degrees of effect were associated with higher helpfulness evaluations. In particular, the subcategories of ‘No effect’ and ‘Large’ were respectively associated with lower and higher helpfulness evaluations, as demonstrated in comments 1 and 2 (the examples come directly from the responses without modifications or omissions; English translations by the first author are provided below each Swedish comment). With respect to effect type, comments 1 and 2 also illustrate that there was no clear relationship between the helpfulness evaluations and the ‘Facilitative’ effect type in either experiment. However, in both experiments the majority of the listeners who reported actively comparing the two speech samples gave the L1 samples helpfulness evaluations of 4 or 5 in the 6-point scale (see comment 2). Unsurprisingly, the listeners whose comments were coded as ‘Non-facilitative’ also evaluated the L1 samples to be less helpful in both experiments, as demonstrated by comment 3.

- (1) They didn't really affect the rating since they were not supposed to be rated. They only provided a little soundcheck before hearing the actual sample that was supposed to be rated.

(Listener 21, Experiment I; ‘No effect’, ‘Facilitative’, ‘Deliberate disregard’; helpfulness = 2)

- (2) Jag anser att de finska talproven var väldigt nyttiga eftersom jag visste en natural talhastighet för alla läsare och kunde jämföra det till de svenska talproven.

I think the Finnish speech samples were extremely helpful because I knew a natural speech rate for all the readers and could compare it to the Swedish speech samples.

(Listener 52, Experiment II; 'Large', 'Facilitative', 'Comparative', 'Speech tempo'; helpfulness = 5)

- (3) Genom finska talproven fick man bekanta sig med rösten av personen men talproven i finska och svenska var ändå så jätteolika så de hjälpte inte jättemycket med bedömning.

Through the Finnish speech samples one got to familiarize oneself with the speaker's voice but the Finnish and Swedish speech samples were still so very different that they didn't help that much with the assessment.

(Listener 46, Experiment II; 'Mild', 'Non-facilitative', 'L1–L3 difference'; helpfulness = 1)

In terms of the three fluency features, there was no clear association between 'Speech tempo' and helpfulness evaluations, as indicated by the relatively even distribution of the helpfulness evaluations for the subcategory in both experiments. In comparison, attending to pauses seemed to be associated with somewhat higher helpfulness evaluations in Experiment I, but not in Experiment II, as illustrated in comment 4.

- (4) Hearing the Finnish samples were handy because they gave perspective to certain behaviours that might not have been discernable from the English samples alone. For example, longer pauses might not be due to disfluency, but rather a linguistic strategy for the speaker. This, however, poses a problem for assessing fluency: Should L2 fluency be assessed based on the speaker's L1 fluency, or a 3rd party standard (native speakers)?

(Listener 20, Experiment I; 'Moderate', 'Facilitative', 'Pauses', 'Personal speaking style'; helpfulness = 5)

In both experiments, however, the listeners who reported attending to features related to personal speaking styles evaluated the L1 samples to be more helpful (see comments 4 and 5), whereas the listeners who mentioned differences between the L1 and L2/L3 or reported attempting to ignore the L1 samples evaluated the L1 samples to be less helpful (see comment 6).

- (5) It gave me an idea of the persons style of speaking and also often gave a "hint" whether or not the person was going to simply describe the comic or come up with a story with character names and even lines.

(Listener 17, Experiment I; 'Moderate', 'Facilitative' 'Personal speaking style'; helpfulness = 5)

- (6) Jag tycker att de finska talproven påverkade nästan inte alls mina bedömningar. Jag försökte att inte ta inte så mycket hänsyn till de finska talproven, för att jag burde ju

inte bedöma de. Men det är ju möjligt att de påverkade bedömningen, men på ett omedvetet sätt.

I feel that the Finnish speech samples hardly affected my ratings. I tried not to pay too much attention to the Finnish speech samples, since I wasn't supposed to assess them. However, it's possible that they did affect the assessment, but in an unconscious way.

(Listener 43, Experiment II; 'Mild', 'Deliberate disregard'; helpfulness = 1)

5 Discussion and conclusion

The purpose of this study was to examine how listeners intuitively attend to freely produced L1 speech when assessing L2 and L3 fluency and how helpful they perceive information about L1 speech to be in terms of fluency assessment. Based on an analysis of open-ended comments and numeric helpfulness evaluations provided by 37 listeners, the present study provided information on the thus far understudied effects of L1 speaking style on listeners' assessments of target language fluency. Furthermore, by using speech samples from the same speakers to investigate listener perceptions of L1 Finnish speech in both L2 English (Experiment I) and L3 Swedish (Experiment II) fluency assessment contexts, the study design enabled us to investigate these effects across multiple languages with differing levels of language proficiency.

The listeners' comments were categorized into five main categories and 14 sub-categories that were identified from the data based on the qualitative content analysis. The five main categories were (1) Degree of effect, (2) Effect type, (3) Fluency features, (4) Speaker characteristics and (5) Listener characteristics. With respect to the perceived degree of effect, we found that most listeners did not feel that the L1 speech samples had had a significant effect on their target language fluency assessments, with the majority of the comments falling under the subcategories of 'Mild' and 'Moderate' degree of effect in both experiments. However, the cross-tabulation analyses revealed that the perceived degree of effect was connected with how helpful the listeners felt the L1 samples were in the rating process, with higher helpfulness evaluations coinciding with larger degrees of effect in both experiments.

The different effect types revealed that approximately half of the listeners in both experiments felt that hearing the L1 speech samples somehow facilitated their rating process. However, the cross-tabulation analyses revealed no connection between the facilitative effect type and how helpful the listeners evaluated the L1 samples to be in the rating process. A possible explanation for this could be that, in the listeners' comments from both experiments, there was variation within the subcategory of facilitative effect. While some listeners felt that the L1 speech samples helped to assess target language fluency characteristics, the facilitative effect was

often more generally attributed to the opportunity to familiarize oneself with the speaker before having to assess their target language fluency. Furthermore, how helpful this familiarizing effect was perceived to be varied across listeners, as indicated by the variation found in their helpfulness evaluations. Based on our findings, it seems that giving listeners access to the speakers' L1 speech helps them to adjust and prepare for the target language speech samples, thus potentially releasing cognitive resources for language assessment, but the extent to which listeners find this helpful for L2/L3 fluency assessment specifically depends on the listener's individual characteristics.

A comparison of the three fluency features showed that speech tempo was the feature most often mentioned by the listeners in both experiments, indicating that the rate of speech is the fluency feature most consciously attended to in L1 speech. Interestingly, none of the listeners in Experiment I and very few listeners in Experiment II mentioned hesitations or corrections. Although this finding might seem surprising considering the prevalence of hesitations and self-corrections in language examinations and frameworks (Huhta et al. 2020), it echoes the results of Magne et al. (2019), who also focused on advanced L2 listeners and found that very few commented on repair fluency as influencing their fluency assessments, despite the significant correlation between their numeric ratings and the frequency of self-corrections. Thus, it is possible that the listeners in the present study attended to repair fluency features on a more subconscious level, but it is also possible that their relatively limited experience in language testing and assessment led them to focus more on speed and breakdown fluency features than repairs. Future studies could apply the assessment procedure introduced in the present study to listeners with more language testing experience (such as experienced language teachers) to corroborate the results.

In addition, a similar proportion of listeners in both experiments reported attending to other, non-fluency related features in the L1 speech samples, such as narrative strategies or vocabulary choices (see, e.g., comment 5). These results are in line with previous mixed-methods and qualitative perceived fluency studies, where listeners' fluency assessments have been found to be influenced by factors other than temporal and hesitation phenomena (e.g., Freed 1995; Rossiter 2009; Préfontaine & Kormos 2016; Magne et al. 2019). Furthermore, cross-tabulation analyses showed that the listeners who reported attending to these features also assigned higher helpfulness scores for the L1 samples, which might point towards potential connections between these non-fluency related first and target language features. As, to our knowledge, connections between L1 and L2/L3 speech have only been studied with regards to fluency measures, a possible future research direction could be to examine the relationship between first and target language speech across other domains of language production (e.g., grammatical and lexical complexity). In addition, as our study focused on perceived fluency, future studies could investigate

whether information about L1 speaking style has an effect on the assessment of other target language oral skills, such as complexity or accuracy. However, acknowledging the limitations based on the results from the two small-scale experiments and the potential effects of the populations under investigation, our findings should be confirmed with large sample sizes and different speaker and listener populations.

We also found clear differences across individual listeners in how they intuitively attended to L1 speech when assessing target language fluency: approximately a fifth of the listeners in both experiments reported actively comparing first and target language fluency features, whereas some listeners reported attempting to consciously ignore L1 fluency features and focus only on target language fluency. As the listeners in both experiments were advanced Finnish-speaking university students of the target language and the majority of the listeners had participated in teacher training, the results suggest that even listeners with similar language and educational backgrounds may employ varying approaches to how they attend to L1 speech when assessing target language fluency. Since these differences across listeners represent a potential source of rater bias especially in assessment settings where the raters are familiar with the speakers' L1 speaking styles (e.g., in classroom-based speaking tests), the findings imply that there is a need to establish a common approach as to whether and how individual differences in L1 fluency should be taken into account in these assessment contexts. Furthermore, there is a need to increase pre-service language teachers' awareness of how individual differences may affect intuitive judgements of target language fluency (see also Peltonen 2023). Regarding the implications for large-scale language testing (e.g., in high-stakes standardized language tests), if and how potential L1 speaking style effects should be taken into account is a much more complex issue due to practical considerations and potential threats to test validity. As our findings provide the first insights into how listeners attend to L1 speaking styles when assessing fluency, more research on this topic is needed to ensure fair and equal language assessment practices across different contexts (see also De Jong 2018; Gao & Sun 2024).

In conclusion, the present study has been among the first to study the connections between first and target language speech fluency from a perceived fluency perspective. The research questions addressed the themes identified from the listeners' open-ended comments and the relationship between these themes and the listeners' helpfulness evaluations. The results for the first research question revealed a wide range of approaches to L1 speech in both L2 and L3 fluency assessment contexts. In addition, while the feature most often mentioned in both experiments was L1 speech tempo, the listeners in both experiments also commented on features of L1 speech that were not related to fluency. The results for the second research question also showed that these listeners perceived information about L1 speech to be more helpful than the listeners who commented only on L1 fluency features. Overall, the study has extended previous perceived fluency research by qualitatively

examining listeners' perceptions of L1 speech when assessing fluency in different target languages and, thus, provided important insights regarding the relationship between L1 and L2/L3 perceived fluency. The findings also have implications beyond fluency research, as they indicate that the potential effect of L1 speech fluency on target language fluency assessment needs to be addressed explicitly in both teacher training and language assessment.

Literature

- Bosker, H. R., A-F. Pinget, H. Quené, T. Sanders & N. H. de Jong 2013. What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30 (2), 159–175. <https://doi.org/10.1177/0265532212455394>
- Council of Europe 2001. *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cucchiari, C., H. Strik & L. Boves 2002. Quantitative assessment of second language learners' fluency: comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, 111 (6), 2862–2873. <https://doi.org/10.1121/1.1471894>
- De Jong, N. H. 2018. Fluency in second language testing: insights from different disciplines. *Language Assessment Quarterly*, 15 (3), 237–254. <https://doi.org/10.1080/15434303.2018.1477780>
- De Jong, N. H., R. Groenhout, R. Schoonen & J. H. Hulstijn 2015. Second language fluency: speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36 (2), 223–243. <https://doi.org/10.1017/S0142716413000210>
- Derwing, T. M., M. J. Munro, R. I. Thomson & M. J. Rossiter 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31 (4), 533–557. <https://doi.org/10.1017/S0272263109990015>
- Derwing, T. M., M. J. Rossiter, M. J. Munro & R. I. Thomson 2004. Second language fluency: judgments on different tasks. *Language Learning*, 54 (4), 655–679. <https://doi.org/10.1111/j.1467-9922.2004.00282.x>
- Duran-Karaoz, Z. & P. Tavakoli 2020. Predicting L2 fluency from L1 fluency behaviour: the case of L1 Turkish and L2 English speakers. *Studies in Second Language Acquisition*, 42 (4), 671–695. <https://doi.org/10.1017/S0272263119000755>
- Dörnyei, Z. 2007. *Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Freed, B. F. 1995. What makes us think that students who study abroad become fluent? In B. F. Freed (ed.) *Second language acquisition in a study abroad context*. Amsterdam: John Benjamins, 123–148. <https://doi.org/10.1075/sibil.9.09fre>
- Gamer, M., J. Lemon & I. F. P. Singh 2019. irr: Various coefficients of interrater reliability and agreement (R package version 0.84.1). *The Comprehensive R Archive Network*. <https://CRAN.R-project.org/package=irr>
- Gao, J. & P. P. Sun 2024. How does learners' L2 utterance fluency relate to their L1? A meta-analysis. *International Journal of Applied Linguistics*, 34 (1), 276–291. <https://doi.org/10.1111/ijal.12493>
- Harris, P. A., R. Taylor, B. L. Minor, V. Elliott, M. Fernandez, L. O'Neal, L. McLeod, G. Delacqua, F. Delacqua, J. Kirby, S. N. Duda & REDCap Consortium 2019. The REDCap consortium: building an international community of software platform partners. *Journal of Biomedical Informatics*, 95, Article 103208. <https://doi.org/10.1016/j.jbi.2019.103208>
- Harris, P. A., R. Taylor, R. Thielke, J. Payne, N. Gonzalez & J. G. Conde 2009. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42 (2), 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>
- Housen, A., F. Kuiken & I. Vedder (eds) 2012. *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.32>

- Huensch, A. & N. Tracy-Ventura 2017. Understanding second language fluency behaviour: the effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, 38 (4), 755–785. <https://doi.org/10.1017/S0142716416000424>
- Huhta, A., H. Kallio, S. Ohranen & R. Ullakonoja 2020. Fluency in language assessment. In P. Lintunen, M. Mutta & P. Peltonen (eds) *Fluency in L2 learning and use*. Bristol: Multilingual Matters, 129–145. <https://doi.org/10.21832/9781788926317-011>
- Kahng, J. 2018. The effect of pause location on perceived fluency. *Applied Psycholinguistics*, 39 (3), 569–591. <https://doi.org/10.1017/S0142716417000534>
- Kallio, H., M. Kautonen & M. Kuronen 2023. Prosody and fluency of Finland Swedish as a second language: investigating global parameters for automated speaking assessment. *Speech Communication*, 148, 66–80. <https://doi.org/10.1016/j.specom.2023.02.003>
- Kormos, J. & M. Dénes 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32 (2), 145–164. <https://doi.org/10.1016/j.system.2004.01.001>
- Krippendorff, K. 2004. Reliability in content analysis: some common misconceptions and recommendations. *Human Communication Research*, 30 (3), 411–433. <https://doi.org/10.1093/hcr/30.3.411>
- Lehtilä, E., P. Peltonen & P. Lintunen in review. How does access to L1 speech affect L2 and L3 perceived fluency? Evidence from Finnish-speaking and Finnish–Swedish bilingual speakers.
- Lennon, P. 1990. Investigating fluency in EFL: a quantitative approach. *Language Learning*, 40 (3), 387–417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- Lennon, P. 2000. The lexical element in spoken second language fluency. In H. Riggenbach (ed.) *Perspectives on fluency*. Ann Arbor: University of Michigan Press, 25–42.
- Lintunen, P., M. Mutta & P. Peltonen (eds) 2020. *Fluency in L2 learning and use*. Bristol: Multilingual Matters. <https://doi.org/10.21832/9781788926317>
- Magne, V., S. Suzuki, Y. Suzukida, K. Saito, M. Ilkan & M. Tran 2019. Exploring the dynamic nature of second language listeners' perceived fluency: a mixed-methods approach. *TESOL Quarterly*, 53 (4), 1139–1150. <https://doi.org/10.1002/tesq.528>
- Miles, M. B. & A. M. Huberman 1994. *Qualitative data analysis: an expanded sourcebook* (2nd edition). Thousand Oaks: Sage.
- O'Connor, C. & H. Joffe 2020. Intercoder reliability in qualitative research: debates and practical guidelines. *International Journal of Qualitative Methods*, 19, 1–13. <https://doi.org/10.1177/1609406919899220>
- Peltonen, P. 2018. Exploring connections between first and second language fluency: a mixed methods approach. *The Modern Language Journal*, 102 (4), 676–692. <https://doi.org/10.1111/modl.12516>
- Peltonen, P. 2022. Connections between measured and assessed fluency in L2 peer interaction: a problem-solving perspective. *International Review of Applied Linguistics in Language Teaching*, 60 (4), 983–1011. <https://doi.org/10.1515/iral-2020-0030>
- Peltonen, P. 2023. Fluency revisited. *ELT Journal*. Advance online publication. <https://doi.org/10.1093/elt/ccad047>
- Peltonen, P. & P. Lintunen 2016. Integrating quantitative and qualitative approaches in L2 fluency analysis: a study of Finnish-speaking and Swedish-speaking learners of English at two school levels. *European Journal of Applied Linguistics*, 4 (2), 209–238. <https://doi.org/10.1515/eujal-2014-0018>

- Peltonen, P. & P. Lintunen 2022. Multilingual speakers' L1, L2, and L3 fluency across languages: a study of Finnish, Swedish, and English. *Nordand*, 17 (1), 48–63. <https://doi.org/10.18261/nordand.17.1.4>
- Pinget, A.-F., H. R. Bosker, H. Quené & N. H. de Jong 2014. Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, 31 (3), 349–365. <https://doi.org/10.1177/0265532214526177>
- Préfontaine, Y. & J. Kormos 2016. A qualitative analysis of perceptions of fluency in second language French. *International Review of Applied Linguistics in Language Teaching*, 54 (2), 151–169. <https://doi.org/10.1515/iral-2016-9995>
- Préfontaine, Y., J. Kormos & D. E. Johnson 2016. How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing*, 33 (1), 53–73. <https://doi.org/10.1177/0265532215579530>
- R Core Team 2021. *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/>
- Rossiter, M. J. 2009. Perceptions of L2 fluency by native and non-native speakers of English. *The Canadian Modern Language Review*, 65 (3), 395–412. <https://doi.org/10.3138/cmlr.65.3.395>
- Segalowitz, N. 2010. *The cognitive bases of second language fluency*. New York: Routledge. <https://doi.org/10.4324/9780203851357>
- Skehan, P. 2009. Modelling second language performance: integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30 (4), 510–532. <https://doi.org/10.1093/applin/amp047>
- Suzuki, S. & J. Kormos 2020. Linguistic dimensions of comprehensibility and perceived fluency: an investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, 42 (1), 143–167. <https://doi.org/10.1017/S0272263119000421>
- Suzuki, S., J. Kormos & T. Uchihara 2021. The relationship between utterance and perceived fluency: a meta-analysis of correlational studies. *The Modern Language Journal*, 105 (2), 435–463. <https://doi.org/10.1111/modl.12706>
- Tavakoli, P. & P. Skehan 2005. Strategic planning, task structure, and performance testing. In R. Ellis (ed.) *Planning and task performance in a second language*. Philadelphia: John Benjamins, 239–273. <https://doi.org/10.1075/lllt.11.15tav>

Appendix:

Rating instructions in English (a Swedish version of the instructions was used in Experiment II).

You will hear 30 speech samples in English produced by learners of English. Your task is to listen to the speech samples and rate them in terms of their **fluency**. Before each English sample, you will hear a short speech sample in the **learners' first language (Finnish)**. Listen to this speech sample once before listening to the English speech sample to get an impression of the learner's **individual speaking style**, but do not rate the Finnish sample.

In this study, fluency refers to the **ease and effortlessness** of speech and is thus separate from other aspects of oral proficiency, such as grammatical complexity and accuracy. Please give ratings for each of these three **temporal aspects of speech** using a 9-point scale:

- speech rate (**1 = very slow, 9 = very fast**)
- pausing (**1 = very many and/or very long disruptive pauses, 9 = very few and/or very short pauses**)
- corrections (**1 = very many and/or very long corrections, 9 = very few and/or very short corrections**)

Based on your ratings of speech rate, pausing and corrections, rate the overall fluency of the speaker on a scale from 1 to 9 (*1 = extremely disfluent, 9 = extremely fluent*).

You will first rate **two practice samples** in order to familiarize yourself with the rating procedure.