

OLLI KUPARINEN

Muutoksen mekanismit

Kolmen aikapisteen reaaliaikatutkimus
Helsingin puhekielestä

Tampereen yliopiston väitöskirjat 428

OLLI KUPARINEN

Muutoksen mekanismit
Kolmen aikapisteen reaaliaikatutkimus
Helsingin puhekielestä

AKATEEMINEN VÄITÖSKIRJA
Esitetään Tampereen yliopiston
Informaatioteknologian ja viestinnän tiedekunnan
suostumuksella julkisesti tarkastettavaksi
Tampereen yliopiston Pinni B -rakennuksessa
luentosalissa 1096, Kanslerinrinne 1,
Tampere, 18.6.2021, klo 12

AKATEEMINEN VÄITÖSKIRJA

Tampereen yliopisto, Informaatioteknologian ja viestinnän tiedekunta

<i>Vastuuohjaaja ja Kustos</i>	Dosentti Unni Leino Tampereen yliopisto Suomi	
<i>Ohjaajat</i>	Filosofian tohtori Liisa Mustanoja Tampereen yliopisto Suomi	Professori Jaakko Peltonen Tampereen yliopisto Suomi
<i>Esitarkastajat</i>	Professori emerita Pirkko Nuolijärvi Kotimaisten kielten keskus Suomi	Filosofian tohtori Katri Priiki Turun yliopisto Suomi
<i>Vastaväittäjä</i>	Dosentti Heikki Paunonen Helsingin yliopisto Suomi	

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

Copyright ©2021 tekijä

Kannen suunnittelu: Roihu Inc.

ISBN 978-952-03-1989-2 (painettu)

ISBN 978-952-03-1990-8 (verkkojulkaisu)

ISSN 2489-9860 (painettu)

ISSN 2490-0028 (verkkojulkaisu)

<http://urn.fi/URN:ISBN:978-952-03-1990-8>

PunaMusta Oy – Yliopistopaino
Joensuu 2021

Kerkolle

ESIPUHE

Olen akateeminen broileri. Hyppäsin suoraan lukiosta Tampereen yliopistoon suomen kielen oppiaineeseen ja valmistuin maisteriksi suurin piirtein tavoiteajassa. Valmistumisen jälkeen otin puolen vuoden syrjähyppyn yliopiston ulkopuolelle, kunnes palasin taas takaisin – ensin Tampereen yliopiston normaalikouluun ja puoli vuotta myöhemmin keskustakampukselle väitöskirjatutkijaksi. Neljä vuotta on kulunut ja tässä sitä taas ollaan. On selvää, että tämä broileri ei ole kasvanut itsekseen; seuraavassa esitän kiitoksia ihmisille matkan varrelta.

Suurin kiitos työni näinkin nopeasta valmistumisesta kuuluu Kippo-hankkeen muille jäsenille Unni Leinolle, Liisa Mustanojalle, Jaakko Peltoselle ja Jenni Santaharjulle. Unnia, Liisaa ja Jaakkoa kiitän lisäksi työni ohjauksesta. Tiedän olevani huomattavan etuoikeutettu, kun olen saanut tavata ohjaajiani käytännössä viikoittain työryhmän palaverissa. Unni on johdattanut opintojani jo pitkältä kandidaattivaiheesta asti (en ole koskaan tehnyt opinnäytettä ilman Unnin ohjausta), Liisa on toiminut sosiolingvistisenä selkänöjanani ja väsymättömänä kannustajanani, Jaakko on kärsivällisesti opastanut minua tilastollisten menetelmien maailmassa, ja Jenni on osannut tieteenalan ulkopuolisena kysyä sellaisia kysymyksiä, joissa aloitteleva kielitieteilijä on joutunut pohtimaan itsestään selvinä pitämiään linjauksia. Oman työni viime metreillä porukkaan liittyi uudeksi jäseneksi Sini Knuutila, joka saa toivon mukaan nauttia yhtä runsaasta tuesta kuin minä. Hanketta on rahoittanut koko nelivuotiskauden Koneen Säätiö, jota kiitän keskeytyksettömän tutkimustyön mahdollistamisesta.

Väitöskirjani esitarkastajia, professori Pirkko Nuolijärveä ja FT Katri Priikiä, kiitän työni tarkkanäköisestä kommentoinnista. Käsillä oleva väitöskirja on monessa suhteessa velkaa professori Heikki Paunosen urauurtavalle työlle Helsingin puhekielen parissa. Niinpä olen erityisen otettu ja kiitollinen saadessani hänet myös työni vastaväittäjäksi. Tutkimuksen aineistosta kiitän niin ikään Heikkiä sekä Hanna Lappalaista, jolta olen saanut muutenkin runsaasti tukea ja neuvoja vuosien varrella. Osatutkimusten arvioinnista kiitän kunkin julkaisun nimettömiä arvioijia.

Tampereen yliopistolla olen saanut työskennellä kahdessa eri osoitteessa. Kiitän Silja Juopperia, Emilia Luukkaa ja Anna-Kaisa Jokipohjaa, eli huonekavereitani ruokalan takaa. Aina löytyy aikaa yhdelle kahvikupilliselle! Pinnin viidennessä kerroksessa sain huonekaverikseni Mehdi Ghasemin, jota kiitän aina iloisesta mielestä ja

keskusteluhaluudesta. Toimiessani tuntiopettajana ja Kielitieteen päivien koordinaattorina olen saanut tutustua myös moniin muihin henkilökunnan jäseniin. Kiitos erityisesti Johanna Vaattovaaralle, joka otti minut mukaan järjestämään konferenssia ja hieman vahingossa tuli hankkineeksi siitä minulle vielä työpaikankin. Kielten tohtoriseminaarissa on ollut innostavaa esitellä omaa tutkimusta ja keskustella kaikenlaisista aiheista. Kiitän erityisesti Laura Eilolaa, Henrik Oksasta ja Joonas Vakkilaista, joiden kanssa olen saanut keskustella myös seminaarin ulkopuolella.

Ystäväni perusopinnoista, Elisa, Mariella, Meri, Riitu, Tanja ja Tiina, kiitos opiskeluriennoista ja yhä jatkuvasta yhteydenpidosta. Ollia ja Iikkaa kiitän pyöräily-, retkeily- ja saunaseurasta sekä elämän ilojen ja surujen jakamisesta. Aina Hollannin ja Suomen välillä matkalla olevalle Joonakselle lähetän lämpimät terveiset. Shakkipeelimme on ollut kesken jo vuosia! HMP:n keikat ovat tarjonneet mahtavan rentoutumiskeinon väitöskirjan kirjoittamisen lomassa. Kiitos siis Annina, Fanni, Hene, Henkka, Henna, Juho ja Lauri. Toivon mukaan joskus koittaa aika, jolloin pääsemme taas lavoille. Laurille lisäksi erityiskiitokset johdattamisesta säännöllisten lausekkeiden maailmaan. Kannattaa aina kysyä basistilta, jos ei itse tiedä.

Vanhemmilleni Marjalle ja Esalle lähetän kiitokset luottamuksesta kykyihini ja vapaudesta tehdä, mitä haluan. Veljeni Matti, ikuinen idolini, kiitos musiikista ja kiinnostuksesta myös minun töihini. Kiitos Anne ja Olli rentoutumislomista Akaan täyshoito-hotellissa. Mikalle kiitos visailumielestä ja kannustuksesta. Vaikka väitöskirjan valmistuminen tuntuukin hienolta, olen viimeisen neljän vuoden aikana saavuttanut paljon enemmän toisaalla. Olen saanut ihanan vaimon ja maailman viisaimman lapsen. Kerkko, kiitos naurusta, ihmettelystä ja ylitsevuotavasta rakkaudesta. Ilona, kiitos jatkuvasta tuesta ja asioiden kannattelemisesta silloin, kun olen ollut ajatuksisani tässä kirjassa. Te olette kaikki.

Nokialla vappuaattona 2021

Olli Kuparinen

TIIVISTELMÄ

Väitöskirja tarkastelee Helsingin puhekielessä tapahtuneita muutoksia 1970-luvulta 2010-luvulle sekä kielenmuutosta teoretisoivia malleja. Tutkimuksen aineistona käytetään osaa Helsingin puhekielen pitkittäiskorpuksesta, joka sisältää helsinkiläisten puhujien litteroituja haastatteluja 1970-luvulta, 1990-luvulta ja 2010-luvulta. Kielenmuutosta tarkastellaan niin piirteiden, varieteettien kuin kieliyhteisönkin tasolla. Konkreettisten kielenmuutosten lisäksi väitöskirjassa sovelletaan käytäntöön teorioita muutosmalleista sekä leksikaalisesta ja morfologisesta diffuusiosta. Tarkastelu on pitkälti tilastollista.

Tutkimus osoittaa, että tunnusmerkkisesti yleiskieliset variantit ovat väistyneet Helsingin puhekielestä tutkimuksen kattaman ajanjakson aikana. Tällaisia variantteja ovat esimerkiksi monikon ensimmäisen persoonan käyttö (*menemme*), pronominit *minä* ja *sinä* sekä sananloppuisen *z:n* säilyminen (*tulisi, viisi*). Selvimmin tämä ero näkyy 1970-luvun ja 1990-luvun aineistojen välillä. Etelähämäläiset piirteet ovat vielä 2010-luvun aineistossa yleisiä, mutta nuorten keskuudessa suositaan erityisesti valtakunnallisia puhekielen piirteitä (esim. *eA-* ja *OA-*yhtymien oikeneminen, *d:n* kato *b:n* jäljessä). Selvästi leimaavia piirteitä vältellään.

Helsingin puhekielessä yleistyy sekä *A-* että *MA-*infinitiivin tehtävissä toimiva lyhyt infinitiivi (*voin kertoa, meen kertoa*). Eteneminen tapahtuu erityisesti analogisen mallivaikutuksen voimin. 2010-luvun nuorilla puhujilla lyhyt infinitiivi on jo infinitiiveistä kaikkein yleisin. Yksittäisten piirteiden osalta tutkimus osoittaa myös, että kato muodot etenevät yleiskielen *hd-*yhtymässä diffuusiomallin mukaan hitaasti muoto-ryhmistä ja lekseemeistä toisiin, vaikka valtaosa katoesiintymistä keskittyy tiettyihin lekseemeihin (lukusanat, *lähteä, tehdä*). *A-*loppuisen vokaaliyhtymien ja yleiskielen *ts-*yhtymän vastineiden osalta muutos etenee jo aiemmin oletettuun suuntaan. *eA-* ja *OA-*yhtymät esiintyvät lähes yksinomaan oienneessa asussa, kun taas *UA-* ja *iA-*yhtymissä on enemmän vaihtelua. Yleiskielen *ts-*yhtymän sijaan Helsingissä yleistyy vaihtelu *#: t*.

William Labovin kehittämistä muutosmalleista Helsingin puhekielen aineistossa selvästi yleisin on sukupolvittaisen muutoksen malli. Tämä löydös antaa lisätukea pitkään sovelletulle näennäisaikamenetelmälle. Sen sijaan muut Labovin alkuperäisen jaottelun mallit ovat harvinaisia. Muutosmallien täydennykset elinikäinen muutos,

vetäytyvä muutos ja kaareva muutos perustuvat kaikki ikäkausivaihtelulle. Tutkimuksessa esitetään, että kankeiden muutosmallien sijaan ikäkausivaihtelua voisi tarkastella erilaisina ikäefekteinä. Lisäksi osoitetaan, että yksittäiset muutosmallit eivät sovellu erityisen hyvin kuvaamaan kolmea aikapistettä. On oletettavaa, että muutokset käyvät läpi useampia muutosmallivaiheita. Tätä havaintoa kuvaamaan tutkimuksessa esitetään myös muutosmallien ja S-käyrän yhdistelmä.

Väitöskirja tuottaa uutta tietoa niin muutoksen teorioista kuin Helsingin puhekielessä tapahtuneista muutoksista. Tutkimuksessa sovelletut tilastolliset menetelmät ovat fennistiikassa verrattain harvinaisia. Teoreettisesti väitöskirjassa tarkennetaan erityisesti käsitystä iän ja kielenmuutoksen suhteesta. Kolmen aikapisteen tarkastelu tuo uusia näkökulmia kahden aikapisteen reaaliaikatutkimusten perusteella tehtyihin tulkintoihin.

Asiasanat: Helsingin puhekieli, kielenmuutos, reaaliaika, muutosmallit, diffuusio

ABSTRACT

This dissertation examines the linguistic changes that have happened in Finnish spoken in Helsinki from the 1970s to the 2010s as well as theoretical models of language change. The data are from the longitudinal corpus of Finnish spoken in Helsinki, which consists of transcribed interviews of Helsinki natives from the 1970s, 1990s and 2010s. The dissertation monitors language change in linguistic features and varieties, as well as in the community. Theories on language change, such as patterns of change and lexical and morphological diffusion, are also applied to the data. The analysis is mostly based on statistical methods.

The study shows that there is retreat from salient features of standard Finnish in Helsinki during the time span of the data. Such features include the use of the first-person plural (*menemme*), pronouns *minä* ‘I’ and *sinä* ‘you’ and pronouncing the word final *i* (*tulisi*, *viisi*). The retreat is most visible between the 1970s’ and 1990s’ datasets. Features originating from Southern Häme are still common in the 2010s, but the youngest speakers prefer supralocal features (e.g., monophthongization of *eA* and *OA*, /d/ deletion after /h/). Enregistered features are avoided.

A shortened infinitive form that can act in the functions of both the A-infinitive and the MA-infinitive is becoming prevalent in Helsinki. It spreads mainly through analogy. The shortened infinitive is the most common infinitive of the youngest age group in 2010s. The study also shows that /d/ deletion after /h/ progresses to new lexemes and morphological groups according to a diffusion model, although it is most prevalent in certain lexemes (numerals, *lähteä* ‘to leave’, *tehdä* ‘to do’). The change is progressing as alleged in A-final vowel sequences and in standard Finnish *ts*. Monophthongization is almost exclusive in *eA* and *OA* sequences, but more variable in *UA* and *iA*. As a substitute for *ts*, the grade alternation *tt* : *t* is becoming more common.

Generational change is by far the most common pattern of change in the data. This finding supports the use of the apparent time construct. The remaining patterns of change in Labov’s original division are quite rare. The supplementary patterns life span change, retrograde change and curvilinear change are all based on age-specific changes. The study suggests that such changes could be viewed as age effects instead of rigid patterns of change. It is moreover shown that singular patterns of change do

not perform well over three time points. It is assumed that linguistic changes go through several patterns over time. To depict this, the patterns are combined with the S-curve in the study.

The dissertation presents new findings from theories of language change as well as from Finnish spoken in Helsinki. The statistical methods used are relatively rare in contemporary studies on Finnish. The dissertation deepens the understanding of the relation between age and language change. The examination of three time points questions and expands the interpretations made earlier from two points in real time.

Keywords: Helsinki Finnish, language change, real time, patterns of change, diffusion

SISÄLLYS

1	Johdanto.....	17
1.1	Tutkimuksen lähtökohdat.....	17
1.2	Väitöskirjan rakenne ja tutkimuskysymykset.....	19
2	Teoreettinen tausta.....	24
2.1	Suomalaisen variaation tutkimuksen historiaa	24
2.2	Kvantitatiivinen sociolingvistiikka.....	26
2.3	Näennäisaika- ja reaaliaikatu tutkimus	27
2.4	Kielenmuutosten teoretisointi	29
2.5	Ään suhde muutokseen	32
3	Helsingin puhekieli.....	35
3.1	Helsingin kaupungin historiaa	35
3.2	Kielelliset lähtökohdat.....	36
3.3	Slangin kehityssuuntia	39
3.4	Edeltävä tutkimus ja väitöskirjan suhde siihen	40
4	Aineisto ja menetelmät	43
4.1	Helsingin puhekielen pitkittäiskorpus	43
4.2	Aineisto.....	46
4.3	Tutkitut piirteet	48
4.4	Laskennalliset menetelmät.....	52
5	Tutkimuksen tulokset	56
5.1	Muuttuvat piirteet	56
5.1.1	Infinitiivit.....	57
5.1.2	<i>bd</i> -yhtymä.....	59
5.1.3	<i>A</i> -loppuiset vokaaliyhtymät	62
5.1.4	<i>ts</i> -yhtymä	69
5.2	Muuttuvat varieteetit	71
5.2.1	Yleiskieli.....	72
5.2.2	Hämäläismurteet.....	73
5.2.3	Valtakunnallinen puhekieli.....	74
5.3	Muutosmallit.....	75
5.3.1	Muutosmallit käytännössä.....	75

5.3.2	Muutosmallit ja S-käyrä	78
6	Koonti	82
6.1	Vastaukset tutkimuskysymyksiin	82
6.1.1	Muutos pürretasolla.....	82
6.1.2	Muutos yhteisön tasolla.....	84
6.1.3	Muutos yksilötasolla.....	85
6.1.4	Muutoksen teorit	86
6.2	Päätelmät	87
6.2.1	Muutos Helsingin puhekielessä 1970-luvulta 2010-luvulle.....	87
6.2.2	Teoreettiset päätelmät.....	89
6.3	Tutkimuksen arviointi ja jatkotutkimuksen mahdollisuudet.....	91
	Lähteet	94
	Julkaisut	103

Kuvioluettelo

Kuvio 1. OA- ja eA-yhtymien oikeneminen eri muotoryhmissä Helsingin puhekielessä.

Kuvio 2. UA-yhtymän oikeneminen partitiiveissa ja infinitiiveissä Helsingin puhekielessä.

Kuvio 3. iA-yhtymän oikeneminen partitiiveissa ja infinitiiveissä Helsingin puhekielessä.

Kuvio 4. Vahvassa asteessa esiintyvän tt-variantin edustus, heikossa asteessa esiintyvän tt-variantin edustus ja heikossa asteessa esiintyvän t-variantin edustus Helsingin puhekielessä.

Kuvio 5. Muutosmallien ja S-käyrän yhdistelmä.

Taulukkuuettelo

Taulukko 1. Labovin muutosmallit, Sankoffin ja Wagnerin täydennykset ja kaarevan muutoksen malli yhteisön ja yksilön tasolla sekä muutosprosessin kuvaus.

Taulukko 2. Aineiston jakautuminen vuosikymmenittäin, ikäryhmittäin ja puhujakohorttien mukaan.

Taulukko 3. Osatutkimuksessa KATOMUODOT Tutkitut muotoryhmät, esimerkikilekseemit ja esiintymismäärä.

ALKUPERÄISJULKAISUT

- Julkaisu I Kuparinen, Olli 2018: Infinitiivien variaatio ja muutos Helsingissä. – *Virittäjä* 122(1), 29–51.
- Julkaisu II Kuparinen, Olli – Mustanoja, Liisa – Peltonen, Jaakko – Santaharju, Jenni – Leino, Unni 2019: Muutosmallit kolmen aikapisteen pitkittäisaineiston valossa. – *Sananjalka* 61, 30–56.
- Julkaisu III Kuparinen, Olli – Peltonen, Jaakko – Mustanoja, Liisa – Leino, Unni – Santaharju, Jenni: Lects in Helsinki Finnish. A probabilistic component modeling approach. – Käsikirjoitus, hyväksytty julkaistavaksi aikakauskirjassa *Language Variation and Change*.
- Julkaisu IV Kuparinen, Olli – Santaharju, Jenni – Leino, Unni – Mustanoja, Liisa – Peltonen, Jaakko: Katomuotojen eteneminen yleiskielen *bd*-yhtymässä Helsingin puhekielessä. – Käsikirjoitus. Lähetetty arvioitavaksi.

Yhteisartikkelien vastuunjako

	II	III	IV
Alkuperäisidea	OK	OK, UL, LM, JP, JS	UL
Aineisto ^a	OK	OK	OK
Menetelmät	OK, JP	JP	JP, OK, UL
Tilastollinen analyysi	OK, JP	JP, OK, JS	OK, JP, UL, JS
Lingvistinen analyysi	OK, LM	OK, LM, UL	OK, UL, LM
Käsikirjoituksen valmistelu ^b	OK, LM, JP, JS, UL	OK, JP, LM, UL, JS	OK, JS, UL, LM, JP

OK = Olli Kuparinen, UL = Unni Leino, LM = Liisa Mustanoja, JP = Jaakko Peltonen, JS = Jenni Santaharju

^a Esiintymien poiminta Helsingin yliopiston opiskelijoiden keräämästä ja litteroimasta korpuksista.

^b Kussakin artikkelissa käytetty kirjoittajajärjestys on perusteltu osatutkimusten yhteydessä. OK on kaikissa artikkeleissa ensimmäisenä kirjoittajana ja päävastuussa.

1 JOHDANTO

Muutos on jokaisen elävän kielen ominaisuus. Siksi se on myös yksi kielentutkimuksen keskeisimmistä kohteista. Muutosta on kielentutkimuksessa tarkasteltu sekä pitkällä että lyhyellä aikavälillä. Pitkän aikavälin tarkastelussa keskiöön nousevat esimerkiksi kielten tai murteiden erkaantuminen, kun taas lyhyemmällä aikavälillä keskitytään paikallismurteiden tai kaupunkien puhekielten muutokseen. Pitkää aikaväliä tutkii erityisesti historiallinen kielitiede, ja jälkimmäiseen keskittyvät dialektologian ja sosiolingvistiikan tutkimussuunnat. Erityisen keskeiseksi tutkimuskohteeksi muutos on noussut juuri sosiolingvistiikassa, jossa on pyritty seuraamaan meneillään olevaa muutosta. William Labovin tutkimuksista 1960-luvulla lähteneen suuntauksen piirissä on pyritty selittämään yhteisöjen ja yksilöiden kielellisiä muutoksia monin eri tavoin: erilaisiin taustamuuttujiin perustuen, sosiaalisten verkostojen analyysillä, kielellisten tyylien kautta, maantieteellisesti, tilanteisesti ja niin edelleen.

Tämä väitöskirja kiinnittyy tähän perinteeseen: tutkimuksen kantavana teemana on muutos. Tämä näyttäytyy toisaalta konkreettisena, Helsingin puhekielen piirteisä ja varieteeteissa 40 vuoden ajanjaksona tapahtuneena muutoksena – toisaalta näitä konkreettisia muutoksia kuvaamaan kehitettyjen teorioiden tarkasteluna. Tutkimus antaa siis kuvan Helsingin puhekielessä 1970-luvulta 2010-luvulle tapahtuneista kielellisistä muutoksista. Tämän lisäksi tutkimus tuottaa uutta tietoa sosiolingvistiikassa perinteisesti hyödynnetyistä teorioista ja niiden puutteista. Tutkimuksen laskennalliset menetelmät ovat ainakin suomalaisessa kielitieteessä valtaosin harvinaisia. Väitöskirjan teoreettiset löydökset ja menetelmälliset valinnat tuovat näin uusia työkaluja sosiolingvistien käyttöön. Tässä luvussa kuvaan ensin tutkimuksen lähtökohtia, minkä jälkeen esittelen väitöskirjaan kuuluvat osatutkimukset sekä tutkimuskysymykset.

1.1 Tutkimuksen lähtökohdat

Tutkimuksen kohteena on Helsingin puhekieli 1970-luvulta 2010-luvulle. Väitöskirja asettuu näin osaksi yksittäisen paikkakunnan kieltä laajasti tarkastelevaa

tutkimusperinnettä (esim. Paunonen 1995 [1982]; Kurki 2005; Mustanoja 2011). Näistä tutkimuksista poiketen väitöskirja perustuu kuitenkin kolme aikapistettä kattavaan yhtenäiseen aineistohankkeeseen.¹ Haastatteluja on toteutettu samoin periaattein 1970-, 1990- ja 2010-luvulla. Kolmen aikapisteen reaaliaikatuutkimukset ovat kansainvälisestikin harvinaisia, ja tämä väitöskirja on ensimmäinen vastaavan aikahaarukan väitöskirja kaupunkikielestä suomen kielen alalla. Kolmen aikapisteen tutkiminen mahdollistaa toisaalta laajemman kuvan saamisen Helsingin puhekielessä tapahtuneista muutoksista ja toisaalta pääasiassa kahden aikapisteen varaan rakennettujen muutosteorioiden tarkastelun kolmesta aikapisteestä käsin.

Helsingin puhekielen muutoksen ohessa tarkastelen myös variaationtutkimuksessa hyödynnettyjä teoreettisia malleja. Tutkin William Labovin kehittämää ja muiden täydentämää teoriaa muutosmalleista (Labov 1994: 83–84; Paunonen 2005a; Sankoff 2005; Sankoff & Blondeau 2007; Wagner & Sankoff 2011; Wagner 2012a; Kurki 2013; Laasanen 2016), useilla aloilla hyödynnettävää S-käyrää sosiolingvistiikan kontekstissa (Wang & Cheng 1970; Bailey 1973; Kytömäki 1981) sekä kiinan kielen tutkimukseen alun perin kehitettyä, mutta muissa kielissä erilaisia sovelluksia saanutta leksikaalisen diffuusion mallia (Wang 1969; Mielikäinen 1995; Bybee 2002). Vaikka teoriat ovat niin kansainvälisessä kuin suomalaisessakin sosiolingvistiikassa hyvin vakiintuneita, niitä on käytetty lähinnä kvalitatiivisesti tapahtuneiden muutosten selityksiin. Tässä väitöskirjassa tarkastelen teorioiden toimivuutta käytännössä käyttäen apuna laskennallisia menetelmiä. Tällainen tarkastelu tuottaa tarkempaa tietoa niiden todellisesta kuvauvoimasta ja toisaalta niihin sisältyvistä teoreettisista ongelmista. Tutkimuksen osatutkimuksissa on käytetty pääosin sellaisia tilastollisia menetelmiä, joita on harvemmin hyödynnetty suomalaisessa sosiolingvistisessä tutkimuksessa. Näin väitöskirja toimii myös menetelmien esittelijänä sosiolingvistiselle yleisölle.

Labovin varhaisten sosiolingvististen tutkimusten (esim. 1966) suurimpia oivaluksia oli kielen variaation ja muutoksen selittäminen sosiaalisten taustamuuttujien avulla. Myöhemmin näiden taustamuuttujien hyödyllisyyttä on voimakkaasti kyseenalaistettu (esim. Pratt 1987; Romaine 1994: 221–225; Bucholtz 2003; Lappalainen 2004: 15–16; Vaattovaara 2009: 23). Taustamuuttujista systemaattisesti hyödynnetään yleensä vain ikää. Näin teen tässäkin tutkimuksessa. Samalla pohdin iän ja muutoksen suhteen ongelmallisuutta ja kommentoin sitä erityisesti kolmen aikapisteen reaaliaikaisiin tuloksiin perustuen. Yksittäisten piirteiden muutosta tarkastelen labovilaisen sosiolingvistiikan perinteestä poiketen pääasiassa kielen rakenteesta käsin ja

¹ Myös Tommi Kurjen väitös (2005) kattaa kolme aikapistettä, mutta näistä 1970-luvun vertailuaineisto poikkeaa selvästi tuoreemmista haastattelukierroksista.

hyödynnän selitysmalleina analogiaa ja morfologista diffuusiota. Toisaalta esimerkiksi infinitiivejä tarkastelen systeeminä, mikä paljastaa sellaista variaatiota, joka yksittäisiin morfologisiin piirteisiin keskittyessä on jäänyt varjoon (vrt. Nuolijärvi & Sorjonen 2005: 130). Uusin menetelmin pystyn tarkastelemaan esimerkiksi puhujayhteisön ryhmittelyä pelkästään kieleen perustuen, vailla sitoutumista taustamuuttujiin.

Monissa tällä vuosituhanella julkaistuissa suomalaisen sosiolingvistiikan väitöskirjoissa näkyy pyrkimys uusiin kysymyksenasetteluihin (Nuolijärvi 2000: 33–35; Mustanoja 2011: 10). Esimerkiksi sosiaalisten verkostojen (Lappalainen 2004), kansanlingvististen asenteiden (Vaattovaara 2009) ja idiolektien (Mustanoja 2011) tutkimukset kohdistavat mielenkiinnon fennistiikassa uusiin aiheisiin. Toisaalta on pyritty tarkastelemaan esimerkiksi kieliyhteisöjä etnografisesti (Lehtonen 2015) ja perinteisiä tutkimuskohteita vuorovaikutuksen näkökulmasta (Priiki 2017). Käsillä olevan väitöskirjan voi liittää tähän uudistusmieliseen joukkoon siksi, että vaikka tutkimuksen aineisto ja tutkitut piirteet ovat monella tapaa hyvin perinteisiä, olen tarkastellut niitä uusin tavoin. Pyrkimyksenä ei ole ollut sivuuttaa jo 1970-luvulla aloitettuun keruuhankeeseen väistämättä sisältyviä ongelmia, vaan etsiä niihin ratkaisuja. Pitkittäisaineistojen tutkimuksessa on välttämätöntä, että tutkimuksen tekemisen tavat uudistuvat. Näin pystytään vastaamaan yhteiskunnassa vääjäämättä tapahtuviin muutoksiin myös kielellisen analyysin tasolla (vrt. Hurttu 1999: 86–87). Tältä osin väitöskirja rinnastuu myös Kurjen (2005) väitökseen, joka esitteli fennistiikassa aiemmin harvinaisen Varbrul-menetelmän sovellusta suomalaiseen pitkittäisaineistoon.

1.2 Väitöskirjan rakenne ja tutkimuskysymykset

Väitöskirja koostuu neljästä osatutkimuksesta ja käsillä olevasta yhteenveto-osasta. Osatutkimuksista kolme on julkaistu suomenkielisissä lehdissä ja yksi englanninkielisessä lehdessä. Kolme ensimmäistä osatutkimusta on vertaisarvioitu ja neljäs on lähetetty arvioitavaksi. Osatutkimukset ja niihin tässä yhteenveto-osassa viitattaessa käytetyt nimet ovat seuraavat:

INFINITIIVIT = Kuparinen, Olli 2018: Infinitiivien variaatio ja muutos Helsingissä. – *Virttäjä* 122(1), 29–51.

MUUTOSMALLIT = Kuparinen, Olli – Mustanoja, Liisa – Peltonen, Jaakko – Santaharju, Jenni – Leino, Unni 2019: Muutosmallit kolmen aikapisteen pitkittäisaineiston valossa. – *Sananjalka* 61, 30–56.

LEKTIT = Kuparinen, Olli – Peltonen, Jaakko – Mustanoja, Liisa – Leino, Unni – Santaharju, Jenni: Lects in Helsinki Finnish. Probabilistic component modeling

approach. – Käsikirjoitus, hyväksytty julkaistavaksi aikakauskirjassa *Language Variation and Change*.

KATOMUODOT = Kuparinen, Olli – Santaharju, Jenni – Leino, Unni – Mustanoja, Liisa – Peltonen, Jaakko: Katomuotojen eteneminen yleiskielen *bd*-yhtymässä Helsingin puhekielessä. – Käsikirjoitus. Lähetetty arvioitavaksi.

Osatutkimukset voi jakaa kahtia niin, että niistä kaksi (INFINITIIVIT ja KATOMUODOT) keskittyvät yksittäisiin piirteisiin ja niiden muutoksen selittämiseen, kun taas kaksi muuta (MUUTOSMALLIT ja LEKTIT) pyrkivät tarkastelemaan laajempaa kokonaisuutta Helsingin puhekielestä. Näin variaatio näyttäytyy sekä laajana kokonaiskuvana että yksittäisten piirteiden tasolla (vrt. Nuolijärvi & Sorjonen 2005: 131). Kaikissa osatutkimuksissa käytetään aineistona Helsingin puhekielen pitkittäiskorpusta (Helpuhe; luku 4.1), mutta INFINITIIVIT-osatutkimuksen aineisto on hieman pienempi kuin kolmen muun artikkelin, koska se perustuu pitkälti aiemmin toteutettuun pro gradu -tutkimukseeni. Infinitiivejä käsittelevä artikkeli on myös osatutkimuksista ainoa oma artikkelini, kun taas muut on toteutettu tutkimusryhmän yhteisartikkeleina. Näissäkin artikkeleissa olen ensimmäisenä tekijänä ja päävastuussa.

INFINITIIVIT-osatutkimus käsittelee infinitiivien variaatiota ja muutosta Helsingin puhekielessä 1970-luvulta 2010-luvulle. Tutkimus paneutuu sinällään jo paljon fennistiikassa tutkittuihin ilmiöihin, MA-infinitiivin illatiivin ja A-infinitiivin latiivin muotoihin (esim. Saukkonen 1965; Sorsakivi 1982; Lappalainen 2004: 83–85; Nuolijärvi & Sorjonen 2005: 97–98; Mustanoja 2011: 127–150). Tutkimuksen uutuusarvo liittyy ennen kaikkea käsittelytapaan. Siinä missä näitä kahta infinitiivimuotoa on perinteisesti tutkittu toisistaan erillään, kiinnitän INFINITIIVIT-osatutkimuksessa huomiota ennen kaikkea infinitiivisysteemiin: miten nämä kaksi infinitiivimuotoa ovat langenneet yhteen, miten uusi lyhyt infinitiivi etenee ja mitä siitä mahdollisesti seuraa?

Osatutkimuksessa KATOMUODOT tarkastellaan samaan tapaan perinteistä fennistiikan tutkimuskohdetta, yleiskielen *d*:n vastineita *b*:n jäljessä. Kato on tyypillinen *t*:n heikon asteen vastineena itämurteissa, mutta jo kauan on tiedetty, että katoa esiintyy myös länsimurteissa tietyissä lekseemeissä, kuten esimerkiksi lukusanoissa (*yhbeksän*, *kabe-*) ja persoonapronominien genetiivimuodoissa (*meiän*) (Rapola 1966: 112). Osatutkimuksessa KATOMUODOT keskitytään tarkastelemaan *bd*-tapauksia Helsingin puhekielessä kolmessa aikapisteessä ja kuvaamaan niiden leviämistä leksikaalisen ja morfologisen diffuusion malleilla (Mielikäinen 1995). Näin ollen tämänkin yksittäisen piirteen muutosta tarkastellaan ennen kaikkea kielen rakenteen tasolla, eikä sosiiaalisiin syihin perustuen.

Osatutkimuksista MUUTOSMALLIT on otteeltaan teoreettisin, vaikka sen toteutus on myös hyvin laskennallinen. Osatutkimuksessa tarkastellaan tutkimuksen aineistoon valittuja kielenpiirteitä William Labovin esittämän ja muiden täydentämän muutosmallijaottelun pohjalta (Labov 1994; Sankoff 2005; Wagner & Sankoff 2011; suomeksi ks. Paunonen 2005a; Kurki 2013; Laasanen 2016). Tutkimuksessa selvitetään, ovatko usein teoreettisina malleina pidetyt muutosmallit todella sovellettavissa käytäntöön. Toisekseen tarkastellaan muutosmallien variaatiota: ovatko jotkin malleista puhtaan teoreettisia, mikä on mallien yleisyys ja miten kolmen aikapisteen tutkiminen vaikuttaa mallien toteutumiseen? Osatutkimus paljastaa sosiolingvistiikan teorian kannalta kiinnostavia seikkoja esimerkiksi näennäisaikamenetelmän toimivuudesta erilaisten ikäsidonnaisten muutosten ongelmallisuuteen.

Ainoana englanninkielisenä osatutkimuksena väitöskirjassa on LEKTIT, jonka toteuttaminen vei monista syistä suurimman osan väitöskirjan kirjoittamiseen kulu-neesta ajasta. Osatutkimus lähti kunnianhimoisesta tavoitteesta: millaisia varieteetteja Helsingin puhekielestä voidaan löytää pelkän kielellisen vaihtelun perusteella ja miten nämä varieteetit muuttuvat kolmen aikapisteen välillä? Lektejä etsittiin perinteisesti automaattisen tekstintutkimuksen parissa käytetyllä Latent Dirichlet Allocation -mallilla (LDA; luku 4.4). Koska aineistoa on paljon, vaati tämän osatutkimuksen toteuttaminen huomattavan paljon kielitieteellistä tulkintaa. Myös osatutkimuksen kirjoittaminen englanniksi ja julkaisupäätöksen saaminen arvostetusta kansainvälisestä lehdestä vaikuttivat luonnollisesti tutkimusprosessin pituuteen.

Osatutkimuksessa LEKTIT esitetään kiinnostavia, osittain lektejä aiemmin käsitel-leitä tutkimuksia (esim. Guy 2013; Gregersen & Phrao 2016) kyseenalaistavia tulok-sia. Tutkimuksen aineistosta löytyy selviä toisistaan poikkeavia lektejä. Nämä löydök-set ovat fennistiselle yleisölle kiinnostavia myös löydettyjen varieteettien kannalta: esimerkiksi voimakkaan yleiskielisen kielimuodon tai toisaalta valtakunnallisen puhe-kielen (yleispuhekielen) olemassaolosta on esitetty epäilyksiä (esim. Mantila 1997; Hurttä 1999; Anhava 2019). Osatutkimuksessa Helsingin puhekielen aineistosta löy-tyy kuitenkin vastine molemmille varieteeteille.

Kokonaisuudessaan väitöskirja tarkastelee monenlaisia kysymyksiä Helsingin puhe-kielen variaatiosta ja muutoksesta, muutoksen tavoista ja kieliyhteisöjen ja -yksilöiden variaatiomalleista. Seuraavassa esitän tutkimuskysymykset selvyiden vuoksi jaotel-tuna alaotsikoiden alle. Lisäksi mainitsen suluissa, missä tutkimuksen osassa juuri tätä kysymystä käsitellään.²

² Olen käyttänyt esittelytavan mallina Katri Priikin väitöskirjaa (2017). Samoin olen toiminut myös vastatessani kysymyksiin luvussa 6.1.

1. Muutos piirretasolla

Millaista variaatiota on havaittavissa suomen yleisimmissä infinitiivimuodoissa, MA-infinitiivin illatiivissa ja A-infinitiivin perusmuodossa? Miten niiden käyttö muuttuu 1970-luvulta 2010-luvulle? (INFINITIIVIT)

Miten yhteenlankeavat infinitiivimuodot vaikuttavat koko infinitiivijärjestelmään? Onko uusi lyhyt infinitiivi erkaantumassa lähtöinfinitiveistään? (INFINITIIVIT)

Missä lekseemeissä kato esiintyy yleiskielen *hd*-yhtymässä? Eteneekö kato ajan kuluessa uusiin lekseemeihin? (KATOMUODOT)

Mitkä piirteet yleistyvät ja mitkä vähenevät Helsingin puhekielessä kolmen aikapisteen välillä? (MUUTOSMALLIT, Yhteenvedo-osa)

2. Muutos yhteisön tasolla

Millaisia lekkejä Helsingin puhekielessä esiintyy? Mitkä variantit esiintyvät yhdessä lekkeissä? (LEKTIT)

Vastaavtko laskennallisen menetelmän avulla löydetyt lektit Helsingin puhekielen taustalla vaikuttavia varieteetteja, kuten hämäläismurteita, yleiskieltä ja slangia? (LEKTIT)

Missä piirteissä on havaittavissa yhteisötason muutosta eli sukupolvittaista tai yhteisöllistä muutosta? (MUUTOSMALLIT)

3. Muutos yksilötasolla

Missä piirteissä on havaittavissa yksilötason muutosta eli ikäsidonnaista, elinikäistä tai vetytyvää muutosta? (MUUTOSMALLIT)

Miten yksilöt käyttävät lekkejä? Käyttävätkö yksilöt vain yhtä lektiä vai yhdistelevätkö he useampia? Mitä lekkejä yhdistellään? (LEKTIT)

4. Muutoksen teoriat

Miten muutosmallit sopivat käytäntöön? Mitkä mallit ovat yleisiä ja mitkä harvinaisia? (MUUTOSMALLIT, INFINITIIVIT)

Miten muutosmallien teoria on sovitettavissa perinteiseen S-käyrään? (Yhteenvedo-osa)

Miten morfologinen ja leksikaalinen diffuusio selittävät kadon etenemistä *hd*-yhtymässä? (KATOMUODOT)

Seuraavassa luvussa esittelen tutkimuksen teoreettista taustaa eli etenkin kvantitatiivista variaationtutkimusta, kielenmuutoksia koskevia teoretisoiteja ja iän ja muutoksen suhdetta. Luvussa 3 esittelen tutkimuskohteen, Helsingin puhekielen.

Tarkastelen lyhyesti Helsingin kaupungin historiaa ja siitä seuraavia kielellisiä lähtökohtia, kunnes paneudun edeltävään tutkimukseen ja suhteutan omaa väitöskirjaani siihen. Luvussa 4 esittelen tutkimuksen aineiston, siitä tutkitut piirteet ja tutkimuksen tekoprosessin. Samassa luvussa esittelen myös tutkimuksessa käytettyjä laskennallisia menetelmiä. Luvussa 5 esittelen tutkimuksen keskeiset tulokset. Aloitan tulosten esittelyn piirretason muutoksista, jonka jälkeen tarkastelen löydettyjä varieteetteja. Viimeisessä alaluvussa tarkastelen muutosmallien sopimista käytäntöön ja niiden suhdetta S-käyrään. Luvussa 6 kokoan väitöskirjan tuloksia yhteen, pohdin niiden merkitystä ja arvioin tehtyjä valintoja. Tässä luvussa annan myös vastaukset edellä esitettyihin tutkimuskysymyksiin.

2 TEOREETTINEN TAUSTA

Tässä luvussa esittelen tutkimuksen kannalta olennaisen teoreettisen taustan. Alaluvussa 2.1 tarkastelen suomalaisen variaation tutkimuksen perinnettä. Seuraavassa alaluvussa 2.2 keskityn erityisesti kvantitatiiviseen sosiolingvistiikkaan ja sitä kohtaan esitettyyn kritiikkiin. Luvussa 2.3 tarkastelen sosiolingvistiikassa hyödynnettyjä näennäis- ja reaaliaikatuksia, jonka jälkeen luvussa 2.4 esittelen erilaisia kielenmuutosten etenemistä kuvaavia teorioita, kuten muutosmalleja ja S-käyrää. Viimeiseksi luvussa 2.5 tarkastelen ikää ja sen suhdetta muutokseen.

2.1 Suomalaisen variaation tutkimuksen historiaa

Variaation tutkimuksella on suomalaisessa kielentutkimuksessa vankka asema. Dialektologian perinteessä 1800-luvun lopulta pitkälle sotien jälkeiseen aikaan pyrittiin tallentamaan ja tutkimaan kansankieltä eli paikallismurteita. Tehtävä koettiin tärkeäksi, koska kansankielen koettiin olevan vaarassa hävitä erityisesti yleiskielen, mutta myös muiden murteiden vaikutuksesta (Lehtimäki 1983: 24; Nuolijärvi 2005: 243–254). Niinpä kielenoppaiksi tai kielimestareiksi valittiin vanhoja, koko ikänsä paikallaan pysyneitä puhujia, joiden kielen koettiin edustavan paikkakunnan aitoa murretta. Itse paikallismurteiden sisäinen variaatio jäi melko vähälle huomiolle, vaikka sitäkin toki ajoittain kommentoitiin (Nuolijärvi 1988).

Murteentutkimuksen perinne keskittyi alueellisten erojen dokumentoimiseen ja ennen kaikkea kvalitatiiviseen tutkimusotteeseen. Niinpä pyrkimyksenä oli löytää murteita toisistaan erottelevia piirteitä, eikä niinkään yksittäisen murteen sisällä piirteiden käytössä esiintyvää vaihtelua. Toisekseen keskittyminen vanhoihin, paikallaan pysyneisiin kielenoppaisiin piilotti myös murteessa mahdollisesti olemassa olevan sosiaalisen vaihtelun. On korostettava, että nämä eivät sinänsä ole tutkimuksellisia virheitä, vaan tietoisien valinnan seurausta: kun pyrkimyksenä oli ennen kaikkea kuvata ja tallentaa murteita, niihin sisältyvä kvantitatiivisen vaihtelun tarkastelu ei kuulunut tutkimustavoitteisiin (esim. Virtaranta 1966: 184–186; ks. myös Nuolijärvi 1988: 122).

Yhdysvalloissa 1960-luvulla kehittynyt sosiolingvistiikan suuntaus (esim. Labov 1966; Weinreich, Labov & Herzog 1968) levisi nopeasti myös Suomeen. Heikki Paunosen arvostelu Labovin New York -tutkimuksesta Virittäjässä (Paunonen 1971) esitteli tutkimussuuntausta, jota pian sovellettiin myös käytäntöön Helsingin puhekielen tutkimushankkeessa. Siinä missä dialektologia pyrki tallentamaan varioimattomia murteita, sosiolingvistinen metodi keskittyi ennen kaikkea variaation rakenteen selvittämiseen ja selittämiseen. Varhaisessa sosiolingvistiikassa korostuivat ennen kaikkea vaihtelun taustalla vaikuttavat sosiaaliset syyt, kuten puhujien ikä, sukupuoli ja sosiaalinen status. Suomalaisen yhteiskunnan poikkeaminen esimerkiksi englantilaisesta tai yhdysvaltalaisesta on aiheuttanut sen, että muuttujista on usein tarkasteltu lähinnä vain ikää ja sukupuolta (Nuolijärvi & Sorjonen 2005: 13).

Sosiolingvistiikan tutkimuskohteeksi vakiintui ennen kaikkea kaupunkien puhekieli. Suomessa käynnistettiin 1970-luvulla Nykysuomalaisen puhekielen muros -hanke, jonka tarkoituksena oli tarkastella suurimpien kaupunkien puhekieltä (esim. Mielikäinen 1980; Jonninen-Niilekselä 1982a; Paunonen 1995 [1982]). Myös tämän tutkimuksen aineisto on syntynyt tässä hankkeessa (ks. luku 4.1). Suomessa sosiolingvistisesti suuntautuneita tutkimuksia on tehty kuitenkin myös maaseutumurteista (esim. Mantila 1992; Kurki 2005). Sosiolingvistiikan ja dialektologian välinen raja onkin Suomessa ollut häilyvä, eikä suurta eroa eri suuntien välille ole helppo osoittaa (Siitonen 1982; Suojanen 1982c: 9). Tutkimussuunnat ovat jakaneet samantyyppiset tutkimuskäytännöt (esim. haastattelu) ja samat tutkimuskohteet (esim. fonologiset ja morfologiset piirteet). Viime vuosina murreaineistoja on tarkasteltu myös vuorovaikutuslingvistisesti (esim. Väänänen 2016; Priiki 2017; Uusitupa 2017).

Lopuksi tarkastelen vielä käsillä olevan väitöstutkimuksen suhdetta suomalaiseen variaationtutkimukseen. Tutkimusaihe on vahvan sosiolingvistinen: suuren kaupungin puhekieli. Niin ikään aineistonkeruu on toteutettu perinteisin sosiolingvistisin periaattein. Kuitenkaan taustamuuttujista en käsittele juuri muita kuin ikää. Eroaako tällainen taustamuuttujatarkastelu juurikaan dialektologian perinteestä, jossa valittiin tarkasteltavaksi vain yksi ikäryhmä (vrt. Hurttu 1999: 67)? On muistettava, että myös murteentutkimuksen parissa havaittiin ikäryhmittäisiä eroja, mutta ne eivät kuuluneet varsinaisiin tutkimusaiheisiin (Nuolijärvi 1988). Tässä väitöskirjassa dialektologian perinteen puoleen osoittaa sekin, että olen selittänyt tapahtuneita muutoksia lähinnä kielisysteemistä käsin esimerkiksi analogian ja diffuusion avulla, enkä sosiaalisin perustein. Tutkimusote on vahvan kvantitatiivinen, mutta tätä tuskin voi pitää sosiolingvistiikan yksinoikeutena. Tämä lyhyt vertailu korostaa sitä, että dialektologian ja sosiolingvistiikan tieteenalat ovat suomalaisessa variaationtutkimuksessa yhä edelleen kytkeytyneinä toisiinsa.

2.2 Kvantitatiivinen sosiolingvistiikka

Etenkin kansainvälisesti sosiolingvistinen tutkimus on ollut 2010-luvulla edelleen voimakkaan laskennallista (esim. Nevalainen, Raumolin-Brunberg & Mannila 2011; Tagliamonte & Baayen 2012; Walker, García, Cortés & Campbell-Kibler 2014; Levon & Buchstaller 2015; Hay & Foulkes 2016; Gross 2018; Nycz 2018). Joillekin tutkijoille sosiolingvistiikka onkin tarkoittanut vain laskennallista sosiolingvistiikkaa (Chambers 2009 [1995]). Myös Suomessa on tehty ja edelleen tehdään runsaasti kvantitatiivista sosiolingvistiikkaa (ks. esim. Kurki 2005; Mustanoja 2011; Priiki 2017; Lappalainen, Mustanoja & O'Dell 2019). Myös käsillä oleva väitöskirja liittyy tähän joukkoon.

Kvantitatiivisuus on saanut kritiikkiä etenkin 2000-luvun alkupuolella (Lappalainen 2004: 15–16; Nuolijärvi & Sorjonen 2005: 132; vrt. Vaattovaara 2009: 23). Vaikka kritiikki on usein osoitettu kvantitatiivisuudelle, on se lopulta kuitenkin kohdistunut ennen kaikkea ryhmän ja yksilön suhteeseen: taustamuuttujin luotuihin ryhmiin ja ryhmäkeskiarvojen käyttöön (vrt. Mustanoja 2011: 91–95). Taustamuuttujia vaivaakin essentialistinen identiteettinäkemys: kuuluminen tiettyyn yhteiskunnalliseen joukkoon (ikäryhmään, sosiaaliluokkaan tai sukupuoleen) nähdään syynä kielellisille eroille (Pratt 1987; Romaine 1994: 221–225; Bucholtz 2003; Vaattovaara 2009: 23). Tämä ei kuitenkaan johdu laskennallisuudesta. Vastaavasti näin luotujen ryhmien tarkastelu keskiarvoina ei ole mitenkään välttämätöntä kvantitatiivisessa tutkimuksessa, vaikka tähän ratkaisuun on usein Labovin (1966) asettaman esimerkin mukaan päädytty. Metodiiikan taustalla vaikuttavia taustaoletuksia on pohdittu suomalaisessa sosiolingvistiikassa melko vähän (Vaattovaara 2009: 21; ks. kuitenkin Laasanen 2016).

Vaikka sosiolingvistiikka on Suomessakin pitkään ollut kvantitatiivista, tilastolliseksi se on muuttunut vasta 2000-luvulla. Ensimmäisenä aidosti tilastollisena laajana tutkimuksena voidaan pitää Tommi Kurjen väitöskirjaa (2005), joka hyödynsi kansainvälisesti tunnettua Varbrul-menetelmää. Tämän jälkeen tilastollista tarkastelua ovat hyödyntäneet esimerkiksi Mustanoja (2011) ja Priiki (2017). Varhemman tutkimuksen kritisoiminen tilastollisten menetelmien käyttämättömyydestä ei kuitenkaan ole erityisen mielekäästä, koska menetelmien soveltaminen on ollut huomattavasti hankalampaa kuin nykyisin (ks. kuitenkin esim. Ma & Herasimchuk 1972; Horvath & Sankoff 1987). Sen sijaan nykypäivän kvantitatiivisessa sosiolingvistiikassa olisi syytä ottaa laajemmin käyttöön erilaiset tilastolliset menetelmät. Tässä tutkimuksessa on menetelmien hyödyntämisessä lähdetty liikkeelle tutkimusongelmasta: mitä halutaan selvittää? Tähän tarkoitukseen on kussakin osatutkimuksessa valittu tilastollinen

menetelmä, joka pystyy vastaamaan esitettyyn ongelmaan. Näin menetelmä palvelee tutkimustarkoitusta.

Kuten todettua, suuri osa kvantitatiivisen sosiolingvistiikan kritiikistä liittyy oikeastaan ryhmän ja yksilön suhteeseen. Kritiikin mukaan keskittyminen suuriin aineistoihin ja puhujaryhmiin häivyttää yksilöiden väliset erot. Tätä ongelmaa on pyritty kiertämään esimerkiksi puhujaverkostojen tutkimuksella (Lappalainen 2004) tai keskittymällä pelkästään idiolektitasolle (Mustanoja 2011). Tarkastelen myös tässä väitöstutkimuksessa ensisijaisesti puhujaryhmiä ja muutoksen kokonaiskuvaa. Osatutkimuksissa on kuitenkin hyödynnetty useita tapoja, joilla myös yksilöiden väliset erot ovat selvästi näkyvillä ja osa analyysia. Osatutkimuksissa MUUTOSMALLIT, LEKTIT ja KATOMUODOT variaatiota ja muutosta kuvataan ruutu- ja janakaavioin, jotka osoittavat ryhmän sisäisen vaihtelun. Niin ikään osatutkimuksissa MUUTOSMALLIT ja KATOMUODOT hyödynnetään sijalukuihin perustuvaa tilastollista testiä, joka tarkastelee yksilöiden muutosta osana puhujakohorttia (ks. tarkemmin luku 4.4). Näin ollen ryhmätason muutos perustuu kohortin yksilöiden muutokselle. Osatutkimuksessa LEKTIT taas vastataan jo Labovin (1966: 209–211) osoittamaan tarpeeseen kieliyhteisön ryhmittelystä puhtaasti kielilähtöisesti. Näin ollen väitöskirjassa keskitytään edelleen selvästi ryhmätasolle mutta yksilöt huomioon ottaen.

2.3 Näennäisaika- ja reaaliaikatutkimus

Sosiolingvistiikan keskeinen piirre on ollut meneillään olevan kielenmuutoksen tarkastelu. Tarkastelu on perustunut joko näennäisaikaan tai reaaliaikaan. Reaaliaikaiset aineistot ovat sellaisia, joissa aikaa on aineistonkeruun aikana todellisuudessa kulu- nut. Tyypillisesti puhujia on haastateltu tietyin intervalein (yleensä 10 tai 20 vuotta), ja eri haastattelukertojen välillä havaittavia eroja on tulkittu muutoksiksi reaaliajassa. Suomessa reaaliaikaisia aineistoja ovat esimerkiksi Helsingin puhekielen pitkittäiskorpus (Helpuhe) ja Tampereen puhekielen pitkittäiskorpus (TAPU). Molemmat ovat kolmen aikapisteen reaaliaikaisia aineistoja. Vastaavasti esimerkiksi Kotimaisten kiel- len tutkimuskeskuksessa 1980-luvun lopulla perustetussa murteenseuruuhankkeessa on koottu reaaliaikaisia aineistoja paikallismurteista (Kurki 2005; Nuolijärvi & Sorjo- nen 2005). Myös Nahkolan ja Saanilahden tutkimukset Virtain murteen muuttumi- sesta (2001, 2004) edustavat reaaliaikaista tutkimusta.

Reaaliaikaiset aineistot jaetaan kahtia trendi- ja paneelitutkimuksiin. Trenditutki- muksissa valitaan uusilla tutkimuskierroksilla puhujat samoin perustein kuin edelli- sellä. Näin ollen puhujat eivät ole kierroksesta toiseen samoja. Trenditutkimuksen

etuna voidaan pitää mahdollisuutta tutkimuskierrosten yhteismitallisuuteen: kullakin kierroksella on mahdollista tutkia yhtä suurta joukkoa informantteja. Paneelitutkimuksessa puolestaan seurataan samoja puhujia kierrokselta toiselle. Tämän tavan etuna on idiolektien seuraamisen mahdollisuus, mutta ongelmana se, että samoja puhujia on usein vaikeaa tavoittaa uudelleen (Paunonen 2006: 65; Sankoff 2018). Helsingin puhekielen pitkittäiskorpus on molempien tapojen yhdistelmä: keski-ikäisten ja vanhojen puhujien ryhmät 1990- ja 2010-luvulla koostuvat jo aiemmillä kierroksilla haastatelluista puhujista. Sen sijaan aina uusi nuorten ryhmä on koostettu trenditutkimuksen periaattein.

Reaaliaikaiset aineistot ovat sikäli uskottavia aineistolähteitä muutoksen tutkimukseen, että ne mallintavat aidosti etenemistä ajassa. Sen sijaan vielä melko vähän on tietoa siitä, kuinka paljon haastattelukertojen välillä on sattumanvaraista variaatiota. Olisiko esimerkiksi vain viikon välein tehtyjen haastattelujen välillä samankaltaista variaatiota kuin 20 vuoden välein tehtyjen haastattelujen välillä? Kuinka paljon haastattelujen välisestä variaatiosta voi siis asettaa reaaliaikaisen muutoksen aikaansaamaksi ja kuinka paljon satunnaisesti variaatioksi? Ja mikä lopulta on satunnaista ja mikä pysyvää? Ennen tällaisten lyhyen aikavälin tutkimusten valmistumista on ainoastaan luotettava siihen, että pitkään hyödynnetty reaaliaikamethodi pystyy tavoittamaan kielessä tapahtuneet muutokset.

Oletus kielenmuutoksesta voi perustua myös näennäisaikaan. Näennäisaikaisessa tutkimuksessa tehdään ennusteita kielitilanteesta tulevaisuudessa sen perusteella, miten eri-ikäiset puhujat käyttävät kieltä nyt. Jos siis esimerkiksi nuoret käyttävät ajankohdassa x tiettyä varianttia 80-prosenttisesti, keski-ikäiset 50-prosenttisesti ja vanhat 30-prosenttisesti, voidaan näennäisajassa ennustaa, että ajassa $x+20$ keski-ikäiset käyttävät varianttia 80-prosenttisesti ja vanhat 50-prosenttisesti. Kuten tästäkin esimerkistä huomataan, näennäisaikamenetelmä perustuu oletukselle idiolektien muuttumattomuudesta. Idiolektien muuttumattomuus on osa kriittisen iän hypoteesia, jonka mukaan yksilön kieli vakiintuu varhaisnuoruudessa (Lenneberg 1967; Meyerhoff 2011: 141; tarkemmin näennäisaikahypoteesista ks. Laasanen 2016). Reaaliaikaiset tutkimukset ovat kuitenkin osoittaneet, että idiolektit voivat muuttua läpi eliniän (esim. Nahkola & Saanilahti 2001; Kurki 2005; Sankoff & Blondeau 2007; Mustanoja 2011; Wagner & Sankoff 2011).

2.4 Kielenmuutosten teoretisointi

Variaation tutkimuksen perinteessä on teoretisoitu kielenmuutosten etenemistä usein eri tavoin. Esittelen tässä lyhyesti tutkimuksen kannalta keskeisiä teorioita ensin kieliyhteisön osalta (muutosmallit, S-käyrä) ja sen jälkeen kielen rakenteen osalta (analoginen muutos, leksikaalinen ja morfologinen diffuusio). Kaikista tässä esiteltävistä kielenmuutoksen teorioista on esitetty kuvaukset myös niissä osatutkimuksissa, joissa niitä käsitellään.

Kieliyhteisöön tietty muutos voi tulla joko ulkopuolelta eli levitä alueellisesti (horisontaalisesti) tai sisäpuolelta ja levitä yhteisön sisällä (vertikaalisesti). Muutosten etenemistä alueelta toiselle on hahmoteltu niin aaltomallilla kuin hierarkkisen diffuusion mallilla (Kurki 2005: 27–28). Aaltomallissa muutos etenee syntysijoiltaan nimensä mukaisesti aaltoina eri suuntiin (Downes 1998: 20–23), kun taas hierarkkisen diffuusion mallissa muutoksen katsotaan etenevän hyppäyksellisesti esimerkiksi suurkaupungista toiseen (Chambers & Trudgill 1998: 172–175). Tässä väitöskirjassa keskityn kuitenkin muutosten leviämiseen yhteisön sisällä, joten en tarkastele näitä malleja lähemmin. Yhteisön sisällä eteneviä muutoksia käsittelevistä malleista tarkastelen ensin William Labovin kehittämää teoriaa muutosmalleista, jonka jälkeen esittelen niin sanottua S-käyrää.

Labovin alkuperäiset muutosmallit (1994: 83–84) pyrkivät esittelemään näennäis-aikamenetelmän ongelmia. Labov jakaa muutosmallit neljään osaan yksilön ja yhteisön reaaliaikaisen muutoksen mukaan. Jos kummallakaan tasolla ei tapahdu muutosta, tilanne on stabiili. Jos vain yksilöt muuttuvat, mutta yhteisö pysyy stabiilina, kyseessä on ikäsidonlainen muutos. Jos taas yhteisö muuttuu, mutta yksilöt eivät, muutosmalli on sukupolvittainen: yhteisö muuttuu uusien sukupolvien myötä. Jos sekä yksilöt että yhteisö muuttuvat, kyseessä on yhteisöllinen muutos. Näennäisaikametodin ongelma on, että se ei voi erottaa toisistaan toisaalta sukupolvittaista ja ikäsidonnaista muutosta, toisaalta stabiilia tilannetta ja yhteisöllistä muutosta. Labovin havainnollistavasta esimerkistä on tullut teoreettinen väline, jonka kykyä kuvata aitoa aineistoa tarkasteltiin osatutkimuksessa MUUTOSMALLIT. Osatutkimuksen yhteydessä muutosmallit on myös kuvattu tarkemmin kuin tässä lyhyessä esittelyssä.

Labovin jaottelua on reaaliaikatutkimusten myötä täydennetty elinikäisen muutoksen mallilla, jossa sukupolvittaisen muutoksen lisäksi osa vanhemmista yksilöistä muuttuu meneillään olevan muutoksen suuntaan (Sankoff 2005: 1011; Sankoff & Blondeau 2007: 562), sekä vetäytyvän muutoksen mallilla, jossa vanhemmat puhujat vetäytyvät meneillään olevasta muutoksesta (Wagner & Sankoff 2011). Nämä täydennykset ovat sikäli tarpeellisia, että ne ottavat paremmin huomioon idiolekteissa

tapahtuvat muutokset, mutta toisaalta hieman redundantteja, koska ne koostuvat Labovin alkuperäisten mallien yhdistelmästä: elinikäinen muutos on sukupolvittaisen ja yhteisöllisen muutoksen yhdistelmä, kun taas vetäytyvä muutos on sukupolvittaisen ja ikäsidonnaisten muutoksen yhdistelmä. Labov (1994: 84) toteaa jo muutosmalleja esitellessään, että tavallisesti mallit esiintyvät yhdistelminä. Myös suomalaisessa tutkimuksessa on havaittu tällaisia yhdistelmämuutoksia (Kurki 2005: 240; Paunonen 2005a: 16).

Taulukko 1. Labovin muutosmallit (1–4), Sankoffin ja Wagnerin täydennykset (5–6) ja kaarevan muutoksen malli (7) yhteisön ja yksilön tasolla sekä muutosprosessin kuvaus.			
	Yhteisö	Yksilö	Muutosprosessi
1. Stabiili tilanne	vakaa	vakaa	Ei muutosta
2. Ikäsidonnaisten muutos	vakaa	muuttuva	Yksilöt poikkeavat aina tietyssä ikävaiheessa (yleensä nuoruudessa) muista. Ikävaiheen jälkeen yksilöt palaavat muun yhteisön tasolle.
3. Sukupolvittainen muutos	muuttuva	vakaa	Uusi sukupolvi poikkeaa aina muista sukupolvista. Yksilöt eivät muutu ajan myötä, joten yhteisö muuttuu.
4. Yhteisöllinen muutos	muuttuva	muuttuva	Kaikki yhteisön puhujat muuttuvat samanaikaisesti samaan suuntaan.
5. Elinikäinen muutos	muuttuva	muuttuva	Osa vanhemmista puhujista ottaa osaa meneillään olevaan sukupolvittaiseen muutokseen. Näin muutosvauhti on nopeampi kuin sukupolvittaisessa muutoksessa.
6. Vetäytyvä muutos	muuttuva	muuttuva	Yksilöt vetäytyvät meneillään olevasta sukupolvittaisesta muutoksesta vanhetessaan. Näin muutosvauhti on hitaampi kuin sukupolvittaisessa muutoksessa.
7. Kaareva muutos	mahdollisesti muuttuva	muuttuva	Yksilöt välttelevät tiettyä varianttia keski-iässä. Etenkin kansainvälisesti pidetty tilanteena, jossa yhteisö ei muutu.

Näiden kuuden muutosmallin lisäksi osatutkimuksessa MUUTOSMALLIT nostettiin esiin myös kaareva muutos (*curvilinear change*), joka etenkin kansainvälisessä tutkimuksessa on yhdistetty tasaisen variaation piirteisiin (esim. -ING; Downes 1998: 224; Tagliamonte 2012: 47; Wagner 2012a, 2012b). Kaarevan muutoksen mallissa puhujat käyttävät jotakin varianttia paljon nuoruudessaan, vähentävät sen käyttöä keski-iässä ja lisäävät jälleen sen käyttöä eläköityessään. Näin ollen variantin vähentäminen johtuisi työelämän paineista (ns. kielellisen markkinan periaate: Bourdieu & Boltanski 1975; Sankoff & Laberge 1978). Suomalaisessa tutkimuksessa kaarevaa muutosta on kutsuttu myös valemutokeksi (Nahkola & Saanilahti 2001: 123; Kurki 2005: 41). Valemutoos on kuitenkin terminä nähdäkseni ongelmallinen: esimerkiksi osatutkimuksessa MUUTOSMALLIT jotkin yhteisön muutokseen johtavat muutosprosessit

muistuttivat kaarevan muutoksen mallia. Tällainen erilaisten muutosten arvottaminen vaivaa yleisemminkin muutosmallien teoriaa (ks. myös luku 2.5). Kaikki seitsemän muutosmallia on esitetty taulukossa 1.

Muutosmallien hyödyntäminen on ollut keskeinen tapa teoretisoida jo tapahtuneita tai tapahtumassa olevia muutoksia. Huomattavasti niitä vanhempi muutoksen kuvaustapa on kuitenkin niin sanottu S-käyrä. Käyrä on alun perin lainattu kulttuurimaantieteestä kielitieteeseen (Kytömäki 1981), ja sitä sovelletaan muillakin tieteenaloilla (esim. Rogers 1962). Kielitieteessä S-käyrää on tehnyt tunnetuksi erityisesti Bailey (1973; ks. myös Wang & Cheng 1970). S-käyrällä kuvataan muutoksen etene mistä niin, että ensin on hidaskuvaihe, jossa vain harvat innovaattorit käyttävät jotakin kielellistä varianttia. Sen jälkeen muutos alkaa hiljalleen kiihtyä, ja siirrytään nopeaan keskivaiheeseen, jossa valtaosa puhujista ottaa uuden variantin käyttöön. Lopuksi muutos taas hidastuu, kun osa puhujista vielä säilyttää vanhan variantin. Väitöskirjan luvussa 5.3.2 tarkastelen muutosmallien ja S-käyrän yhteensovittamista.

Kielenmuutokset voivat edetä myös rakenteellisesti kielisysteemeissä. Tässä väitöskirjassa rakenteellisia muutoksia seurataan kahdessa fennistiikassa pitkään tutkitussa kielenpiirteessä: tulosijainfinitiiveissä ja yleiskielen *hd*-yhtymässä. Osatutkimuksessa INFINITIIVIT keskeisenä muutosta selittävänä voimana on analogia. Analogia nähtiin keskeisenä tekijänä kielenmuutoksissa jo 1800-luvun vertailevassa kielitieteessä, jossa sitä käytettiin selittämään muutoksia silloin, kun systemaattiset äänne muutokset eivät tuottaneet toivottua tulosta. Myöhemminkin analogian ja äänne muutosten suhdetta on pidetty toisiaan täydentävänä (Sturtevant 1917: 108–109; Anttila 1989: 94). Olen tässä tutkimuksessa käsitellyt analogiaa ennen kaikkea mallivaikutuksena: lyhyen infinitiivin muodot hakeutuvat kohti kaksitavuista pitkään vokaaliin päättyvää asua (*antaa, kertoo, pelaa*). Vaikka infinitiivimuodot ovat alun perin päättyneet asuunsa erilaisten äänne muutosten myötä, etenevät ne nykyään pääasiassa analogian mallivaikutuksen voimin.

Osatutkimuksessa KATOMUODOT keskiössä on diffuusio. Kansainvälisessä kielitieteessä on hyödynnetty ennen kaikkea leksikaalisen diffuusion mallia (Wang 1969; Bybee 2002), jossa äänne muutokset etenevät sanasta sanaan alkaen tyypillisesti korkean frekvenssin sanoista. Suomen kielen tutkimukseen diffuusiomallia on muokattu niin, että muutokset etenevät morfologisista muotoryhmistä toisiin. Tästä syystä mallia kutsutaan morfologiseksi diffuusioksi (Suihkonen 1992; Mielikäinen 1995). Molemmissa diffuusio teorioissa siis oletetaan äänne muutosten etenevän vähittäisesti kaikkia muotoryhmiä yhtä aikaa koskettavan äännelain sijaan. Esimerkiksi Suihkonen (1992: 286) esittää väitöskirjassaan vahva-asteisten muotojen edenneen ensiksi

essiiviin ja illatiiviin *koivikkoihin* -tyyppisissä sanoissa, ja vasta niiden jälkeen partitiiviin (*koivikkoja*). Osatutkimuksessa KATOMUODOT diffuusiota tarkasteltiin leksikaalisen ja morfologisen diffuusion hybridinä: keskenään samanlaisia leksemejä yhdisteltiin edustamaan samaa muotoryhmää, ja kadon etenemistä tarkasteltiin näiden muotoryhmien välillä sekä vielä tarkemmin yksittäisten lekseemien tasolla.

2.5 Iän suhde muutokseen

Tässä tutkimuksessa on perinteisistä sosiolingvistisistä taustamuuttujista tarkasteltu lähinnä ikää, koska iän on aiemmissa tutkimuksissa huomattu olevan taustamuuttujista keskeisin (Thelander 1979b: 93–95; Lappalainen 1996: 30; Lappalainen 2001: 77; Nuolijärvi & Sorjonen 2005: 102). Ikä ei kuitenkaan ole sekään taustamuuttujana ongelmaton: kiinnittyminen kronologiseen ikään samaistaa tiettyyn ikähaarukkaan sopivat puhujat yhdeksi kokonaisuudeksi (Lappalainen 2004: 35–36). Eckertin (1997: 154–158) mukaan olisi syytä tarkastella enemmän elämäntilanteita kuin kronologista ikää (vrt. Wagner 2012b). Esimerkiksi tämän tutkimuksen aineistossa osa 1970-luvun nuorista on jo siirtynyt työelämään, kun taas osa ei ole. Lisäksi tämä elämäntilanteiden ero kertaantuu uusiin vuosikymmeniin verratessa: 1990- ja 2010-luvulla kaikki nuoret vielä opiskelevat. Yleisemmin reaaliaikaturkimusten ongelmaksi nousee se, että aikuistumiseen liitetyt elämänvaiheet (työllistyminen, perheen perustaminen) ovat länsimaissa siirtyneet jatkuvasti eteenpäin kronologisessa iässä (Eckert 1997; Bowie & Yaeger-Dror 2015; Buchstaller 2015: 484). Näin samat ikäryhmät esimerkiksi 1970- ja 2010-luvulla ovat jo täysin erilaisissa elämäntilanteissa (vrt. Hurta 1999: 86–87). Tässä tutkimuksessa käytetty trendi- ja paneelitutkimuksen yhdistelmä auttaa osittain purkamaan ikäryhmien vertailuun liittyviä ongelmia: ikäryhmien lisäksi eri vuosikymmenten välillä tarkastellaan myös samoja puhujasukupolvia eli kohortteja.

Edellä luvussa 2.4 kommentoin jo ongelmallista termiä valemuutos. Myös muutostmallien teoriaan sisältyy taustaoletuksia iän ja muutoksen suhteesta. Ikäsidonnaisen muutokseen suhtaudutaan Labovin luokittelussa eräänlaisena ohimenevänä vaiheena, jonka jälkeen palataan normaalitilaan (ks. myös Romaine 1994: 143). Ajatus ohimenevästä muutoksesta pakottaa taustalle myös ajatuksen “oikeasta” kielestä tai normaalitilasta, jota ikäsidonnainen muutos häiritsee. Sosiolingvistisessä perinteessä onkin käytetty huomattavan paljon aikaa ja vaivaa mahdollisimman suotuisien olosuhteiden luomiselle aidon kielen tallentamiseksi ja toisaalta muutoksia tutkittaessa pyritty erottamaan aidot ja väärät muutokset. Onko kielen normaalitilaa kuitenkaan

olemassa ja miksi jotkin muutokset olisivat valheellisia? Ja miten tällainen suhtautuminen kieleen lopulta poikkeaa suomalaisessa murteentutkimuksessa korostuneesta puhtaan murteen tavoittelusta (vrt. Hurtta 1999: 55)?

Ikäsidonnaisuuden yhdistäminen erityisesti nuorten puheeseen saattaakin olla eräänlainen aineistolähtöinen harha: nuoruus on länsimaissa koulumaailman takia melko strukturoitu elämänvaihe, mikä aiheuttaa myös kielessä samankaltaisuuksia. Aikuisiällä elämänpolut erkaantuvat, eikä tällaista kaikille yhteistä vaihetta voi enää helposti löytää. Näin ollen esimerkiksi kolmea ikäryhmää tarkasteltaessa näyttää siltä, että aikuisuudessa ei ole ikäsidonnaisuutta, mutta nuoruudessa on. Myös suomalaisessa tutkimuksessa olisi hyvä tarkastella elämäntilannesidonnaisuutta, ja tutkia samanlaisia elämänvaiheita ohittavia henkilöitä: voidaanko esimerkiksi havaita, että työpaikkaa vaihtavat puhujat muuttavat samalla kieltään tiettyyn suuntaan (vrt. Wagner 2012b)? Entä parisuhteen aloittavat tai lopettavat?

Gillian Sankoffin ja Suzanne Evans Wagnerin muutosmallitädennykset elinikäinen muutos ja vetäytyvä muutos vastaavatkin tähän vaatimukseen siten, että niissä huomioidaan paremmin myös aikuisiällä tapahtuvat muutokset. Kuitenkin näissä malleissa sitoudutaan hieman liiaksi kronologiseen ikään ja toisekseen niissä jätetään irtonaiseksi ajatus siitä, että yksilöt toimivat eri tavoin. Esimerkiksi elinikäisen muutoksen kohdalla Sankoff (2005) esittää, että osa aikuisista lähtee meneillään olevaan muutokseen mukaan, kun taas osa ei. Tällaiseen toimintaan kuvittelisi olevan jonkinlaisia taustasyitä: aikuiset, jotka ovat nuorten kanssa tekemisissä (esimerkiksi kouluissa tai valmennustyössä) saattavat omaksua joitakin piirteitä nuorten puheesta. Toisaalta aikuiset, joiden työssä vaaditaan melko formaalia kieltä saattavat pysyä kielellisesti paikallaan. Näin ollen kyse ei ole niinkään iästä, vaan elämäntilanteesta. Koska aikuisten elämäntilanteet ovat huomattavan erilaisia, idiolektit saattavat kulkea hyvinkin erilaisiin suuntiin (vrt. Eckert 1997). Sosiolingvistiksi teoriassa aikuisiällä ja vanhuudessa mahdollisesti tapahtuvat kielenmuutokset ovat edelleen melkoisia arvoituksia (ks. kuitenkin esim. Paunonen 2005b; Mustanoja 2011; Lappalainen, Mustanoja & O'Dell 2019).

Toisekseen vetäytyvän muutoksen yhteydessä on selvää, että se on eräänlainen ikäsidonnaisen muutoksen vastaliike. Nuoruudessa käytetyistä piirteistä saatetaan luopua aikuisiällä elämäntilanteen muututtua ja esimerkiksi työelämän vaatimusten noustua. Vetäytyvän muutoksen erottaminen aiemmin kuvatussa kaarevasta muutoksesta saattaa siis olla usein vaikeaa. Molemmista malleista tietyt piirteet saattavat olla yleisiä nuoruudessa, koska erilaisia aikuisuuden vaatimuksia ei vielä ole. Työelämään siirryttäessä näistä piirteistä vetäydytään ja taas eläköityessä ja velvollisuuksien vähentyessä nuoruudessa käytetyt piirteet saattavat palata (kaareva muutos) tai eivät

(vetäytyvä muutos). Onko näiden kahden eri muutoksen tavan erottelu erityisen mielekästä?

3 HELSINGIN PUHEKIELI

Tässä luvussa esittelen tutkimuksen kohteen eli Helsingin puhekielen. Aloitan tarkastelun Helsingin kaupungin historian lyhyellä esitellyllä alaluvussa 3.1. Tämän jälkeen keskityn erityisesti kaupungin kielelliseen taustaan alaluvussa 3.2. Alaluvussa 3.3 tarkastelen lyhyesti slangia ja sen mahdollisia kehityssuuntia. Viimeiseksi luvussa 3.4 suhteutan käsillä olevaa väitöskirjaa Helsingin puhekielestä aiemmin tehtyyn tutkimukseen.

3.1 Helsingin kaupungin historiaa

Kustaa Vaasa perusti Helsingin kaupungin vuonna 1550 kilpailemaan Itämeren kaupasta Tallinnan kanssa (Paunonen 2006: 19). Asukkaiden kannalta kaupunkipolitiikka oli tempoilevaa: jo olemassa olevien kaupunkien kuten Rauman ja Porvoon asukkaita määrättiin muuttamaan ensin Viipuriin ja sittemmin Helsinkiin. Kovista yrityksistä huolimatta Helsinki ei juuri kasvanut, vaan esimerkiksi sen uusmaalainen naapuri Porvoo pysyi pitkään suurempana kaupunkina (Paunonen mp.) Osaltaan Helsingin kaupungin kehittämisen tarve lakkasi, kun Ruotsi valloitti kilpakumppani Tallinnan vain noin kymmenen vuotta Helsingin perustamisen jälkeen.

Alkuperäiseltä paikaltaan Vantaanjoen suulta Helsinki siirrettiin Vironniemelle eli nykyisen kantakaupungin alueelle 1600-luvulla. Kun Suomenlinnaa alettiin rakentaa Helsingin edustalle 1700-luvun puolivälissä, erityisesti ruotsinkielisten muutto alueelle lisääntyi (Paunonen 2006: 19–20). Suurin mullistus kaupungin historiassa tapahtui kuitenkin, kun Suomi siirtyi Venäjän vallan alaisuuteen, ja pääkaupunki päätettiin siirtää Turusta Helsinkiin vuonna 1812. Turun palon myötä myös yliopisto siirtyi uuteen pääkaupunkiin vuonna 1827. Nämä kaksi muutosta vaikuttivat merkittävästi Helsingin kielelliseen historiaan: niin valtion kuin yliopistonkin virkamiehet olivat äidinkieleltään ruotsinkielisiä. Pääkaupungin siirron myötä Helsingin väkiluku kasvoi 1800-luvun alkupuolella viisinkertaiseksi (Paunonen 1994: 223), ja vuosisadan lopulla teollistuminen kasvatti kaupunkia entisestään. Etenkin työläisiä muutti kaupunkiin myös suomenkielisiltä alueilta, mikä alkoi hiljalleen muuttaa kaupungin kielioloja.

Muutto oli suurinta hämäläisalueilta (etenkin Helsinkiä ympäröivästä Etelä-Hämeestä), läntiseltä Uudeltamaalta ja Pohjanmaalta (Waris 1973: 64).

1800-luvun alkupuolella kaupungin asutus sijaitsi käytännössä vain Vironniemellä (nykyisten Kruununhaan, Kluuvin ja Kaartinkaupungin alueilla). Suunnitellusti asutus laajeni pohjoiseen kohti Punavuorta ja Kamppia. Myös suunnittelematonta laajenemista tapahtui, kun köyhälistö muutti kaupungin laitamille, kuten Katajanokalle (Waris 1973: 20). Samankaltainen kehitys jatkui ajan kuluessa: työläiset muuttivat 1800-luvun lopulla Pitkäsillan pohjoispuolelle Kallioon, Sörnäisiin ja Hermanniin. Nykypäivänä nämä kaupunginosat ovat vuorostaan kasvattaneet arvoaan, ja lähiöt ovat siirtyneet jälleen kauemmaksi kaupungin keskustasta, Pohjois- ja Itä-Helsinkiin. Vuosien 2019 ja 2020 taitteessa Helsingin kaupungissa asui noin 650 000 asukasta (Vuori, Mäki & Bergroth 2020).

3.2 Kielelliset lähtökohdat

Helsingin kaupungin ominaispiirre on läpi sen historian ollut kaksikielisyys³. Kaupungin perustamisen jälkeen kaupungin vähäisistä asukkaista valtaosa oli ilmeisesti suomenkielisiä (Waris 1950: 17). Kuitenkin 1700-luvulta eteenpäin Suomenlinnan rakentajien ja ruotsinkielisten käsityöläisten muutto kaupunkiin siirsi suomenkieliset vähemmistöön (Paunonen 2006: 19). 1800-luvun alkupuolella valtaosa kaupungin asukkaista oli ruotsinkielisiä (Waris 1950: 11). Pääkaupungin ja yliopiston siirto Turusta Helsinkiin kiihdytti entisestään ruotsinkielisten virkamiesten muutttoa kaupunkiin. Vastaavasti kaupungin teollistuminen etenkin 1860-luvulta eteenpäin kasvatti työläisten muutttoa erityisesti läntiseltä Uudeltamaalta ja Etelä-Hämeestä. Työläiset eivät suinkaan olleet kaikki suomenkielisiä, vaan myös työväestön keskuudessa ruotsin asema oli vakaa.

Suomen- ja ruotsinkielisten asukkaiden suhde kääntyi vain kymmenien vuosien kuluessa 1800-luvun lopulla. Kun suomenkielisiä oli 1870-luvulla 26 % väestöstä ja ruotsinkielisiä 57 %, oli suhde vuonna 1910 jo 59 % suomenkielisiä ja 35 % ruotsinkielisiä (Paunonen 1995 [1982]: 9). Osittain tätä nopeaa käännettä selittää vuonna 1863 annettu kieliasetus, jossa määrättiin otettavaksi suomi ruotsin rinnalle kaikkeen hallinnon kielenkäyttöön kahdenkymmenen vuoden kuluessa. Toinen selittävä tekijä

³ Voitaisiin yhtä hyvin puhua myös monikielisyydestä. Esimerkiksi 1870-luvulla jopa 12 prosenttia kaupungin asukkaista puhui äidinkielenään venäjää (Paunonen 1995 [1982]: 9). Vuoden 2019 lopussa äidinkieleltään muiden kuin suomen-, saamen- ja ruotsinkielisten osuus kaupungin väestöstä oli 16,2 % (Vuori, Mäki & Bergroth 2020: 20).

oli yhä vahvistunut maassamuutto Helsinkiin myös suomenkielisiltä alueilta. Suomenkielisistä muuttajista valtaosa tuli länsimurteiden alueelta Länsi-Uudeltamaalta, Varsinais-Suomesta, Hämeestä ja Etelä-Pohjanmaalta, mutta 1900-luvun alkuvuosikymmeninä voimakkaasti myös Itä-Suomesta, erityisesti Viipurin läänistä (Paunonen 1995 [1982]: 12).

Kaupungin monikielisyys kulminoitui voimakkaimmin 1800-luvun lopulla Pitkäsillan pohjoispuolen työläiskortteleissa kehittyneeseen slangiin, joka sekoitti suomen, ruotsin ja venäjän sanastoa ja rakenteita. Slangin ominaispiirteitä olivat suomen murteille vieraat soinnilliset klusiilit (*b, d, g*) ja suuri slangisanojen varasto, josta osa on säilynyt nykypäivään asti (esim. *tsiigata, bonjata, steissi*; ks. Paunonen & Paunonen 2000). Slangi syntyi työläiskortteleissa, joissa eli niin ruotsin- kuin suomenkielisiäkin perheitä. Myös perheiden sisällä saattoi olla monikielisyyttä, jolloin esimerkiksi toinen vanhemmista puhui ruotsia ja toinen suomea. Slangin kukoistuskausi päättyi viimeistään toisen maailmansodan aiheuttamaan hajaannukseen, ja nykyään vanha slangia elää lähinnä säilyneessä sanastossa. (Paunonen 2006: 56–57.) Kommentoin slangia vielä seuraavassa luvussa 3.3.

Dokumentoitua tietoa 1900-luvun alun Helsingissä puhutusta suomen kielestä ei juurikaan ole saatavilla, mutta muuttoliikkeen perusteella voidaan päätellä puhekielen olleen sävyiltään etelähämäläistä (Paunonen 1995 [1982]: 13). Monet tuon ajan murrepiirteet ovat jo Helsingin puhekielestä hävinneet, mutta erityisesti supistumaverbien analogiset infinitiivimuodot ovat yhä yleisiä ja jopa ekspansiivisia (*pitää pelaa*). Aihetta käsitellään myös tämän väitöskirjan osatutkimuksessa INFINITIIVIT (ks. myös Itkonen 1964: 187–193; Sorsakivi 1982). Hämäläispiirteitä Helsingissä on esiintynyt ja yhä esiintyy varsin runsaasti. Näistä edelleen elinvoimaisina voidaan mainita *tt*-yleiskielen *ts:n* vastineena erityisesti vahvassa asteessa ja *A*-loppuisten vokaaliyhtymien monoftongiutumisen *eA-* ja *OA*-yhtymissä (Paunonen 2006: 39–43). Väitöskirjan osatutkimuksessa LEKTIT löydettiin myös tämän tutkimuksen aineistosta hämäläismurteinen komponentti.

Perinteisiin kansanmurteisiin perustuvan suomen kielen ohella Helsingin puhekieleen on vaikuttanut myös kirjakieleen perustuva puhuttu yleiskieli. Koska valtaosa sivistyneistöstä oli joko ruotsinkielistä tai ainakin opiskellut ruotsiksi, suomelle sivistyksen kielenä ei ollut historiallista pohjaa. Kun lisäksi Helsingissä ei ollut omaa murrepohjaa, päätyi sivistyneistö puhumaan kirjakieleen perustuvaa suomen kielen varieteettia. Kiinnostus kirjakielen normitusta kohtaan oli 1800-luvulla niin ikään voimakasta, mikä osaltaan edesauttoi tämän kielimuodon päätymistä myös puhutuksi. Normitusta ylläpiti esimerkiksi vuonna 1876 perustettu Kotikielen Seura (Paunonen 1976: 334–346). Väitöskirjan osatutkimuksen KATOMUODOT kannalta kiinnostava

on seuran alkuaikojen keskustelu *d:n* ”poisheittämisestä” yleiskielestä, jota aluksi kannatettiin. Kun auktoriteettiasemassa ollut August Ahlqvist vaati *d:n* säilyttämistä, päädyttiin lopulta tälle kannalle (Paunonen 1976: 338). Yleiskieli puhuttuna varieteettina oli erityisesti akateemisten puhujien suosiossa, ja vielä tämänkin väitöskirjan aineistosta löytyi vahvan yleiskielinen komponentti osatutkimuksessa LEKTIT.

Kommentoin tässä yhteydessä lyhyesti kielimuotoihin viittavaa terminologiaa, koska palaan siihen tutkimuksen tuloksissa luvussa 5.2. Kuten edeltä jo kävi ilmi, viittaan yleiskielellä tässä väitöskirjassa kirjakieleen perustuvaan puhetapaan (tuoreesta debatista tämän kielimuodon olemassaolosta ks. Anhava 2019; Paunonen 2019; Lyytikäinen 2019). Termiä murre käytän erityisesti silloin, kun viittaan perinteisiin kansanmurteisiin kuuluviin variantteihin. Tässä tutkimuksessa esillä ovat erityisesti hämäläismurteet. Puhekielellä viittaan toisaalta koko Helsingin alueella puhuttavaan kielimuotoon (Helsingin puhekieli), toisaalta käytän sitä vastinparina yleiskielisille varianteille. Tämä jaottelu ei ole täysin ongelmaton, koska puhekielinen ja yleiskielinen variantti voivat ajoittain langeta yhteen. Slangia käytän terminä melko vähän huolimatta sen yhdistämisestä juuri Helsinkiin (ks. luku 3.3). Koska en tarkastele väitöskirjassa juurikaan sanastoa (ks. luku 4.2), on minun vaikea ottaa kantaa nykyään lähinnä sanastossa esiintyvään slangiin. Ajoittain viittaan vanhaan slangiin, jolla tarkoitan pääasiassa 1900-luvun alkupuolella puhuttua kielimuotoa, joka poikkesi rakenteellisestikin muista puhesuomen muodoista.

Suomalaisessa variaationtutkimuksessa on ajoittain käytetty termiä yleispuhekieli kuvaamaan sellaisia variantteja, jotka ovat yleisiä tai yleistymässä laajalla alueella Suomessa (esim. Karlsson 1975; Nuolijärvi 1986: 223; kritiikistä ks. Mantila 1997; Hurtta 1999: 59–60). Tässä tutkimuksessa vältän kyseisen termin käyttöä ennen kaikkea siksi, että se sekoittaa toisiinsa yleiskielen ja puhekielen, vaikka yleispuhekielessä ei välttämättä olisi jälkeäkään yleiskielestä. Toisekseen monissa tapauksissa yleispuhekieli on viitannut lähinnä eteläisessä Suomessa käytettyihin variantteihin (Nuolijärvi & Tiittula 2000: 222–223). Tässä tutkimuksessa olen päättänyt käyttämään vastavassa tehtävässä termiä valtakunnallinen puhekieli, vaikka sekin sisältää ongelmia: ei ole ollenkaan sanottua, että kaikki piirteet olisivat valtakunnallisia. Nimityksen etuna voidaan kuitenkin pitää selkeyttä verrattuna kankeisiin käännöksiin (ylipaikallinen, engl. *supralocal*) tai useita termejä yhdisteleviin nimiin (yleispuhekieli). Lisäksi esimerkiksi Nuolijärvi ja Sorjonen (2005: 102) käyttävät joistakin varianteista nimitystä valtakunnallisesti yleistynyt variantti. On lisäksi korostettava, että valtakunnallisella puhekielellä (kuten yleensä yleispuhekielelläkin) viitataan ennen kaikkea yleistyviin variantteihin, ei välttämättä yhtenäiseen varieteettiin (vrt. Hurtta 1999: 59–60).

Osatutkimuksessa LEKTIT kuitenkin havaittiin, että kyseessä näyttää olevan ainakin Helsingissä myös konkreettinen varieteetti (ks. luku 5.2.3).

3.3 Slangin kehityssuuntia

Helsingin puhekielen kehityksessä on kolme selkeästi erottuvaa linjaa. On ruotsinkielisen eliitin kielenvaihdon yhteydessä syntynyt hyvin yleiskielinen suomi, on maakunnista muuttaneiden murteellinen suomi (erityisesti Häme) ja kolmanneksi Pitkän sillan pohjoispuolella syntynyt työläisten slangi. Koska edellä esitellyistä kielimuodoista oikeastaan vain slangi on aidosti helsinkiläinen kielimuoto, on se saanut eräänlaisen erityisaseman pääkaupungissa. Vaikka vanha slangi sellaisenaan ei enää nyky-Helsingissä ole käytössä, on se jatkanut elämäänsä markkinoinnissa ja brändäyksessä. Erityisesti slangin on omaksunut urheiluseura Helsingin IFK, joka käyttää markkinoinnissaan lähes aina slangia.

Slangipiirteet ovat siis nousseet hyvin selvästi esiin ja indeksoivat markkinoinnissa paikallisuutta. Kuitenkin helsinkiläiset puhujat suhtautuvat slangiin eri tavoin. Tämän tutkimuksen aineistossa on huomattavissa selvä juopa vanhempien ja nuorempien puhujien välillä. Vielä 2010-luvun haastatteluissa vanhemmat puhujat ovat ylpeitä Helsingin slangista ja kertovat itse osaavansa sitä tai heidän vanhempiansa puhuneen sitä. Kuitenkin 2010-luvun nuorista puhujista osa kokee slangin käytön vieraana. Tämä on kiinnostavaa siksi, että aiemmin juuri nuoret on nähty slangin suurimpana käyttäjäryhmänä (Suojanen 1982b: 157). Esimerkiksi yksi aineiston puhekielisimmistä puhujista, F81⁴, kertoo haastattelussaan, että hänen espoolaiset kaverinsa saattavat puhua Stadista, kun hän itse sanoisi olevansa vain keskustassa. Lisäksi hän jatkaa (esimerkki 1), että erityisesti pojat käyttävät slangia väkinäisesti.

(1)

tai tavallaa mä sanoisin et jätkät käyttää just enemmän
niit sillai niinku, slangisanoi jotka tuntuu väkinäisilt?
jätkät menee ain himaa dösäl, ja tälläsii juttui

Näin ollen on olemassa merkkejä siitä, että slangi on noussut niin stereotyyppiseksi osaksi Helsinkiä, että nuoret puhujat alkavat vastustaa sitä. Myös muualla maailmassa

⁴ Kullekin aineiston puhujalle on oma yksilökoodinsa Helsingin puhekielen pitkittäiskorpuksessa. Kullekin haastattelulle on lisäksi oma koodinsa. Kyseinen katkelma on haastattelusta 2010-T1N3A, joka tarkoittaa 2010-luvulla tehtyä akateemisen (T1) nuoren (3) naisen (N) haastattelua A.

on havaintoja samanlaisesta trendistä: Johnstone ja Kieslingin (2008) tutkimuksessa Pittsburghista nuoret keskiluokkaiset puhujat kyllä tunnistavat paikalliset piirteet mutta eivät itse käytä niitä (vrt. myös Buchstaller 2015: 475). Näihin piirteisiin yhdistyy jonkinlainen leima, jota puhujat haluavat vältellä. Tästä syystä stereotyyppiset piirteet usein katoavatkin kielenkäytöstä (Labov 1972: 180). Helsingissäkin tämä kehitys näkyy nuorten puheessa: 2010-luvulla nuoret käyttävät eniten valtakunnallisia puhekielen variantteja (esim. *d:n* kato ja passiivin käyttö monikossa), kun taas aiemmilla vuosikymmenillä etelähämäläinen puhekieli oli nuorten suosiossa (luku 5.2; LEKTTT). Kuten yllä mainitsin, käsillä oleva tutkimus ei tavoita slangisanaston käytössä ilmevä vaihtelua.

On vaikea tulkita, kuinka suuri merkitys slangin käyttämisellä markkinoinnissa oikeasti on sen stereotyyppikehityksessä. Näyttää kuitenkin siltä, että esimerkiksi Paunosen (1982a: 147) esittämä nuorten puhujien antinormatiivisuus voi yhtä hyvin suuntautua myös muihin kielimuotoihin kuin yleiskieleen (vrt. Hurta 1999: 78; ks. myös Kotsinas 1994: 25). Näin nuoret pyrkisivät ennen muuta välttämään leimaavia piirteitä, tulivat ne mistä lähteestä hyvänsä (vrt. Lappalainen 2001: 77). Slangipiirteet ovat nousseet niin kommentoiduiksi osiksi Helsinkiä, että niihin alkaa liittyä liian vahva paikallinen indeksi (Silverstein 2003; Eckert 2008; Johnstone & Kiesling 2008). Vastaavasti esimerkiksi Tampereen kaupungin liikennelaitoksen tunnuksena on ollut Nysse, vaikka sanan käyttö todellisuudessa lienee melko harvinaista. Myös partikkeli *nääs* yhdistetään stereotyyppisesti Tampereeseen, vaikka se käytössä on esimerkiksi partikkelia *kato* selvästi harvinaisempi (Mustanoja 2011: 208–212).

Helsingissä on tämän tutkimuksen aineiston ajanjakson aikana edetty paikallisista piirteistä kohti valtakunnallisia piirteitä. Näin myös Helsingissä havaittava kehitys vastaa jo pitkään muualla havaittuja kehityslinjoja laajalevikkisten piirteiden yleistymisestä ja leimaavien piirteiden vähenemisestä (esim. Thelander 1979a: 112–116; Mielikäinen 1982: 280–287; Paunonen 1982b: 46–52; Lappalainen 1996: 26–27; Mantila 1997; Lappalainen 2001: 77). On kiinnostavaa nähdä, johtavatko viime vuosina keskustelua aiheuttaneet tiedonvälityksen pirstaloituminen ja erilaisten kuplien syntyminen kielellisen variaation lisääntymiseen, vai jatkavatko leimalliset piirteet väistymistään (vrt. Nuolijärvi & Tiittula 2000).

3.4 Edeltävä tutkimus ja väitöskirjan suhde siihen

Helsingissä puhuttava suomi ei kuulunut 1800-luvun lopulla käynnistyneen murteen-tutkimuksen tutkimuskohteisiin. Helsingin, kuten muidenkin kaupunkien,

puhekielen katsottiin olevan monista eri lähteistä sekoittunutta kieltä. Erityisesti 1900-luvun dialektologisessa perinteessä tavoiteltiin yleiskielen ja muiden murteiden vaikutukselta säilyneiden kansankielten eli murteiden tallentamista, eikä kaupunkien sekakieli näin ollen ollut kiinnostuksen kohteena (Lehtimäki 1983: 22–25).

Sosiolingvистisen variaation tutkimuksen rantautuminen Suomeen 1960-luvulla ja erityisesti William Labovin tutkimus New Yorkin englannista (1966) herättivät kiinnostuksen myös kaupungeissa puhuttavaa kieltä kohtaan. Ensimmäisenä Suomen kaupungeista aloitettiin Helsingin puhekielen tutkimushanke vuonna 1972 Terho Itkosen aloitteesta. Tästä hankkeesta syntyi myös tämän väitöskirjan ensimmäinen aineistokierros. Vuonna 1976 Helsingin puhekielen tutkimushanke liitettiin osaksi laajempaa Nykysuomalaisen puhekielen murros -hanketta, jossa tutkittiin myös Turun, Tampereen ja Jyväskylän puhekieltä. Tämän hankkeen myötä kerättyä aineistoa ovat esitelleet muun muassa Suojanen (1982a, 1985), Mielikäinen (1980, 1981a, 1981b, 1986), Jonninen-Niilekselä (1982a) ja Mustanoja (2011). Esittelen Helsingin puhekielen pitkittäiskorpuksen tarkemmin luvussa 4.1.

Helsingin puhekielen tutkimuksen pioneeri on professori Heikki Paunonen. Paunonen johti Itkosen aloittamaa Helsingin puhekielen tutkimushanketta 1970- ja 1990-lukujen keruun aikana (ks. luku 4.1) ja on kirjoittanut aineistosta lukuisia julkaisuja. Erityisen laaja kuvaus on vuoden 1982 käsikirjoitus 1970-luvun keruun tulokista, joka julkaistiin painettuna vuonna 1995 (Paunonen 1995 [1982]). Paunonen on myöhemmin vertailut myös 1970- ja 1990-lukujen aineistoja (1995, 2005a, 2005b, 2006) ja tarkastellut myös slangia (Paunonen & Paunonen 2000, Paunonen 2016).

Paunosen vuoden 1982 käsikirjoituksen lisäksi 1980-luvulla julkaistiin myös Pirkko Nuolijärven Helsinkiin muuttaneiden pohjalaisten ja savolaisten kielellistä sopeutumista käsittelevä väitöskirja (Nuolijärvi 1986). Sen sijaan 1990-luvulla merkittävien julkaisujen suhteen oli hiljaisempi jakso, kuten sosiolingvистisten väitöskirjojen osalta ylipäänsä (Nuolijärvi & Sorjonen 2005: 13; ks. kuitenkin Paunonen 1994; 1995). 2000-luvulla tutkimuksissa alkoi painottua Helsingin monimuotoistuminen, ja toisaalta sosiolingvистisessä teoriassa puhalsivat uudet tuulet: vuonna 2004 julkaistiin Hanna Lappalaisen väitöskirja sosiaalisesta verkostosta, jonka puhujat olivat helsinkiläisiä, ja vuonna 2015 Heini Lehtosen etnografinen väitöskirja maahanmuuttajataustaisista helsinkiläisnuorista. Keskeisiä julkaisuja 2000-luvulla ovat myös kokoomateokset *Helsinki kieliyhteisönä* (Juusela & Nisula 2006) ja *Helsingissä puhuttavat suomet* (Sorjonen, Rouhikoski & Lehtonen 2015).

Perinteisen sosiolingvistiikan ohella 2000-luvulla on yleistynyt myös sosio-onomastinen tutkimus, jossa tarkastellaan pääkaupunkiseudun nimistöä (Ainiala & Vuolteenaho 2005; Paunonen, Vuolteenaho & Ainiala 2009; Ainiala & Halonen 2011;

Ainiala & Jantunen 2019; Vuolteenaho, Lappalainen & Ainiala 2019). Toisaalta on tarkasteltu niin pääkaupunkiseutulaisten kuin muidenkin asenteita Helsingin puhekieleen (Vaattovaara & Soininen-Stojanov 2006; Halonen & Vaattovaara 2017; Halonen, Nyström, Paunonen & Vaattovaara 2020).

Käsillä olevassa väitöskirjassa tarkastelen Helsingin puhekielen pitkittäiskorpusta (luku 4.1) kokonaisvaltaisesti. Näin tutkimus asettuu eräänlaiseksi jatkoksi Heikki Paunosen saman aineiston kahdesta ensimmäisestä kierroksesta tekemille tutkimuksille. Kolmen aikapisteen tarkastelu tarkoittaa kuvaa Helsingin puhekielen variaatiosta ja muutoksesta, kuten myös muutoksen teorioista. Aineistoa on koko laajuudessaan käytetty tämän väitöskirjan osatutkimusten lisäksi toistaiseksi pro gradu -tutkimuksissa (esim. Vuolle 2016) sekä Lappalaisen, Mustanojan ja O'Dellin idiolektien muutosta tarkastelevassa artikkelissa (2019).

4 AINEISTO JA MENETELMÄT

Tässä luvussa esittelen tutkimuksen aineiston ja tutkimuksessa käytetyt menetelmät. Käsittely alkaa Helsingin puhekielen pitkittäiskorpuksen esittelyllä alaluvussa 4.1. Alaluvussa 4.2 kuvaan, mitä osia korpuksesta on käytetty tässä tutkimuksessa. Toisaalta kuvaan niitä eroja, joita osatutkimuksen INFINITIIVIT ja muiden osatutkimusten aineiston välillä on. Luvussa 4.3 esittelen väitöskirjassa tutkitut piirteet ja niiden poiminnan tutkimuksen aineistosta. Viimeisessä alaluvussa 4.4 esittelen lyhyesti väitöskirjan osatutkimuksissa hyödynnetyt laskennalliset menetelmät.

4.1 Helsingin puhekielen pitkittäiskorpus

Tutkimuksen aineistona käytetään osaa Helsingin puhekielen pitkittäiskorpuksesta (Helpuhe). Korpus on syntynyt Helsingin yliopistossa toteutetussa hankkeessa, ja se sisältää haastatteluja kaikkiaan kolmelta vuosikymmeneltä: 1970-, 1990- ja 2010-luvulta. Korpuksen ajallinen laajuus on poikkeuksellinen jopa maailmanlaajuisesti. Seuraavassa esittelen tarkemmin korpuksen syntyyn johtaneita kehityskulkuja, aineiston koostamisessa tehtyjä valintoja sekä korpuksen käyttöä tässä tutkimuksessa.

Helsingin puhekielen tutkimushanke käynnistyi Helsingin yliopistossa vuonna 1972. Hankkeen johtajana toimi Terho Itkonen, tutkijana Heikki Paunonen ja tutkimusapulaisina Outi Honkasalo sekä Riitta Matilainen (Paunonen 2006: 63). Suomeen 1960-luvulla rantautuneen sosiolingvistisen metodin ohjastamana haastateltavat valittiin kielenulkoisten muuttujien perusteella. Muuttujina käytettiin perinteisiä sosiologisia taustatietoja, kuten sosiaalista luokkaa, ikää, asuinpaikkaa ja sukupuolta. Aineiston keruumenetelmänä käytettiin haastatteluja. Haastattelut olivat noin tunnin mittaisia ja käsittelivät tyypillisesti haastateltavien elämää. Useissa haastatteluissa keskustellaan myös asenteista niin Helsinkiä, kaupungin muutosta kuin kaupungissa puhuttavia kieliäkin kohtaan.

Vuonna 1976 Helsingin puhekielen tutkimushanke liitettiin osaksi valtakunnallista Nykysuomalaisen puhekielen murros -hanketta, jonka kohteena olivat myös Tampereen, Turun ja Jyväskylän puhekielet. Näistä kaupungeista Turun hanke keskittyi syntaksin tutkimukseen ja kolme muuta morfofonologisiin piirteisiin.

Hankkeista on kolmeen haastattelukierrokseen ehtinyt Helsingin ohella ainoastaan Tampereen puhekieltä käsittelevä TAPU-hanke. Aivan vastaava ei Tampereenkaan aineisto ole, sillä siinä ei ole haastateltu lainkaan nuoria puhujia 1990-luvun haastattelukierroksella. Tampereen hankkeen kahdesta ensimmäisestä haastattelukierroksesta on tarkka kuvaus Liisa Mustanojan väitöskirjassa (2011).

Helsingin puhekielen tutkimushankkeen ensimmäinen haastattelukierros toteutettiin vuosina 1972–1974. Haastateltavat valittiin kahdesta kaupunginosasta, jotka olivat akateeminen Töölö ja pääasiassa työläisten asuttama Sörnäinen. Sosiaaliluokkia oli kolme, ja luokittelu perustui Helsingin kaupungin luokitukseen. Ylimpään luokkaan luettiin vain akateemisen koulutuksen saaneet tai akateemisista perheistä tulevat lukiolaiset, kun taas alimpaan luokkaan vain kansakoulun, tai enintään ammattikoulun käyneet työläiset. Iän perusteella haastateltavat jaettiin kolmeen kategoriaan: nuoriin (15–20-vuotiaat), keski-ikäisiin (40–45-vuotiaat) ja vanhoihin (yli 65-vuotiaat). (Paunonen 1995 [1982]: 26).

Kaikkiaan 1970-luvulla tutkimusapulaiset Outi Honkasalo ja Riitta Matilainen haastattelivat 149:ää helsinkiläistä, ja 127 haastattelusta litteroitiin vähintään puolen tunnin jakso (litteroijina toimivat edellä mainittujen lisäksi myös Anna-Liisa Lepäsmä ja Merja Sorsakivi). Paunonen on 1970-luvun haastattelukierroksen tutkimuksessa käsitellyt tästä määrästä vain 96:ta haastattelua, jotta puhujaryhmät olisivat myös sosiaaliluokkien perusteella yhtä suuret (Paunonen 1995 [1982]: 27). Tässä väitöskirjassa en hyödynnä sosiaaliluokittaista vertailua, joten saatavilla ollutta litteroitua aineistoa on käytetty 1970-luvun osalta kokonaisuudessaan.

Hanketta jatkettiin kaksikymmentä vuotta myöhemmin Heikki Paunosen johdolla. Vuosina 1991–1992 Johanna Kurela haastatteli 45:tä helsinkiläistä. Tästä joukosta jo edellisellä kierroksella mukana olleita oli 29, ja uusia nuoria puhujia valittiin mukaan 16. Nuoret puhujat valittiin tasan kahdesta kaupunginosasta, Töölöstä ja Kallio-Sörnäisistä. Töölöläiset nuoret kävivät lukiota, kun taas Kallion-Sörnäisten nuoret opiskelivat joko ammattioppilaitoksessa tai kauppaoppilaitoksessa. 1990-luvun aineisto eroaa monessa suhteessa 1970-luvun aineistosta. Ensinnäkin se on huomattavasti pienempi (127 puhujaa vs. 45 puhujaa). Toisekseen uudelleen tavoitetut haastateltavat eivät enää edusta tasaisesti taustamuuttujia, vaan joukko on heterogeenisempi kuin 1970-luvulla. Erityisesti alimman sosiaaliluokan sörnäisläisiä oli vaikea löytää uudelleen haastateltavaksi, minkä takia 1990-luvun aineisto painottuu ylempiin sosiaaliluokkiin (Paunonen 2006: 65; vrt. Sankoff 2018). Tämä toisaalta kuvastanee koko Sörnäisten alueen demografista muutosta Helsingin sisällä.

Vuonna 2013 toteutettiin hankkeen kolmas aineistonkeruu. Hankkeen johtajana toimi Hanna Lappalainen ja avustajina Saira Marttila, Jenna Sorjonen, Sanni Surkka,

Suvi Syrjänen, Suvi Vierula ja Marjut Vuolle. Uusi nuorten ryhmä oli tälläkin haastattelukierroksella 16 puhujan kokoinen, ja se jakaantui tasaisesti lukiolaisiin (8) ja ammattikoululaisiin (8). 1990-luvun haastattelukierroksen puhujista tavoitettiin uudelleen 27 puhujaa. Kaikille kolmelle haastattelukierrokselle on osallistunut 13 puhujaa. Näistä 1950-luvulla syntyneiden joukkoa (9 puhujaa) ovat tarkemmin tarkastelleet Lappalainen, Mustanoja ja O'Dell (2019). Haastattelujen pohjana käytettiin samaa runkoa kuin aiemminkin vuosikymmenillä, ja vuoden 2013 haastattelijat pyrkivät tietoisesti nostamaan esiin samoja aiheita kuin aiemmin (mts. 556). Näin ollen haastatteluja voisi luonnehtia puolistrukturoiduiksi (Kurki & Mustanoja 2020). Kaikilla kierroksilla nauhoitetut haastattelut ja niihin kohdistetut litteraatit on tallennettu osaksi CSC:n Kielipankkia, jossa ne ovat luvanvaraisessa tutkimuskäytössä.

Haastattelua aineistonkeruun metodina on kommentoitu runsaasti variaationtutkimuksen perinteessä (esim. Karhu 1995; Lappalainen 2004: 15–16; Nuolijärvi 2005: 259–261; Nuolijärvi & Sorjonen 2005: 129–130; Kurki & Mustanoja 2020). Haastatteluja on suosittu etenkin tallennettavan puheen teknisen laadun ja laajuuden vuoksi. Toisekseen reaaliaikaisissa hankkeissa haastattelutilanteen toistettavuus on etu. On kuitenkin selvää, että esimerkiksi Helsingin puhekielen pitkittäiskorpuksessa suhtautuminen haastatteluun tilanteena on vuosikymmenten aikana muuttunut. Paunonen raportoi yksittäisen puhujan idiolektin muutosta käsitellessään, että puhuja oli 1970-luvulla tulkinut haastattelun selvästi formaalimmaksi tilanteeksi kuin 1990-luvulla (Paunonen 2005b: 184–186). Kyseisen puhujan vuosikymmenten välinen muutos onkin hämmästyttävän suuri. Tällaisissa tilanteissa on vaikea erottaa diakronista muutosta diatyyppisestä muutoksesta: Onko haastattelujen välinen ero lopputulos 20 vuoden kehityksestä? Vai onko ero vain suhtautumisessa haastattelutilanteeseen? Kun käsitys muutoksesta perustuu 20 vuoden välein tehtäviin haastatteluihin, on tutkijan osittain luotettava siihen, että keruumetodi tavoittaa muutokset. Olisi kuitenkin kiinnostavaa tarkastella esimerkiksi vain muutamien viikkojen aikahaarukalla toteutettujen haastattelujen välisiä eroja, jotta saataisiin käsitys siitä, kuinka suurta vaihtelu haastattelujen välillä ylipäänsä on.

Myös haastattelujen litteraatteja tutkittaessa on luotettava niiden työstäjiin, joita ei itse tunne. Kun litterointeja on tehty usealla vuosikymmenellä ja usean litteroijan voimin, on selvää, että niiden välillä on eroja. Toisekseen aineistoa litteroitaessa on tehty joitakin valintoja, jotka heijastuvat osittain myös tähän tutkimukseen. Esimerkiksi sananrajailmiöiden ja avartuneiden diftongien litteroiminen ei ole johdonmukaista. Erot ovat pääasiassa vuosikymmenten välillä (esim. *η* litteroitu vain 1990-luvulla), mutta osittain myös vuosikymmenten sisällä siksi, että osa tutkimuksessa käyttämistäni 1970- ja 1990-luvun haastattelulitteraateista on tehty 2010-luvulla. Näitä

eroja havainnollistavat näytteet 2–4. Esimerkki 2 on 1970-luvulta haastattelusta 1970-T1N3C, esimerkki 3 1990-luvulta haastattelusta 1990-T1M3C ja esimerkki 4 2010-luvulta haastattelusta 2010-S3N3A. Esimerkissä 2 on litteroitu avartuneita diftongeja ja sananrajailmiöitä ja esimerkissä 3 vielä tarkemmin esimerkiksi η-äännettä. Esimerkissä 4 sen sijaan ei ole litteroitu sananrajailmiöitä missään yhteydessä. On korostettava, että nämä eivät siis ole litterointivirheitä, vaan kullakin vuosikymmenellä on tehty erilaisia valintoja siinä, mitä ilmiöitä johdonmukaisesti litteroidaan.

(2)

niin kup pienenä vai

/OH: Niin./

no, pienenä mä pyräili aika paljon tos tota, rannassa tos merikannontiej ja siinä, sit seur-, sit siel se- niit seurasaaaren teitä ja, tai siäl, siäl päim munkkivuoressa et tota, mut sillo jo-, jos mä lähin niin ku kauemmas seurasaaarest sillom mul piti ollak kyl joku isompi ihminen mukANA AINA.

(3)

joo ky me silloḅ ku ollaḅ kotona mut, isä tekee kyllä aika paljo matkoi sillee et ei enää, tai nykyään kun se silloḅ ku me oltiin nuorempii ni kyl se oli enemmän kotona. nyt se om matkustellu aika paljon. ja mun, systeri om muuttanun nyt, se om, muuttanut tota, kihlattunsak kanssa vähäk kauemmas. et nytt o yks vähemmän perheessä

(4)

mitenköhän mä ees mä laitoin kai jotain sähköpostia tai en mä ees menny sanoo niille

mitään tai et periaattees olisin mä voinu sen opon kaa juttelee ja kattoo jotain sellasii kurssei et ois voinu pärjää mut

4.2 Aineisto

Tämän väitöskirjan osatutkimuksissa INFINITIIVEJÄ lukuun ottamatta on käytetty korpuksesta kaikkia vuoden 2017 kesällä saatavilla olleita litteraatteja. Tämä kattaa yhtä vaille kaikki haastattelut 1970-luvulta (126), 33 litteraattia 1990-luvulta ja 40 litteraattia 2010-luvulta. Tutkimuksen alkaessa erityisesti 1990-luvun ja 2010-luvun haastatteluista oli vielä osa litteroimatta tai tallentamatta osaksi Kielipankkia. Osan

juuri valmistuneista litteraateista sain suoraan Hanna Lappalaiselta, mistä osoitan suuret kiitokset. Kuten luvuista huomataan, erityisesti 1990-luvun aineisto on vajaa suhteessa tehtyihin haastatteluihin: kaikkiaan 12 haastattelua jää tämän väitöskirjan tarkastelun ulkopuolelle. 2010-luvulta pois jää vain kolme haastattelua. Väitöskirjan aineiston jakautuminen eri vuosikymmenille, kohortteihin ja ikäryhmiin on esitetty taulukossa 1.

Taulukko 2. Aineiston jakautuminen vuosikymmenittäin, ikäryhmittäin ja puhujakohorttien mukaan. Ensin mainitut luvut ovat osatutkimusten MUUTOSMALLIT, LEKTIT ja KATOMUODOT aineistona, ja sulkeissa on esitetty osatutkimuksen INFINITIIVIT aineiston jakauma.			
Kohortti	Haastattelut 1972–1974	Haastattelut 1991–1992	Haastattelut 2013
Kohortti 1 s. 1900–1907	65–74 vuotta 46 puhujaa (22)		
Kohortti 2 s. 1927–1932	40–45 vuotta 41 puhujaa (20)	59–65 vuotta 9 puhujaa (13)	81–86 vuotta 4 puhujaa (4)
Kohortti 3 s. 1952–1957	15–20 vuotta 39 puhujaa (23)	34–40 vuotta 10 puhujaa (13)	56–61 vuotta 9 puhujaa (9)
Kohortti 4 s. 1971–1975		16–21 vuotta 14 puhujaa (15)	38–42 vuotta 14 puhujaa (14)
Kohortti 5 s. 1994–1997			16–19 vuotta 13 puhujaa (17)
Yhteensä	126 puhujaa (65)	33 puhujaa (41)	40 puhujaa (44)

Päädyin jättämään litteroimattomat haastattelut pois aineistosta ensinnä siksi, että valmistuneet litteraattit on tehty tiettyjen periaatteiden mukaisesti, ja itse litteroimani aineisto olisi todennäköisesti poikennut näistä periaatteista (ks. luku 4.1). Toisekseen luvussa 4.3 esiteltävä piirteiden poiminta oli itsessään huomattavan aikaa vievää. Jos olisin tämän työn lisäksi litteroinut 15 haastattelua riittävällä tarkkuudella, olisi varsinaisen tutkimustyön aloittaminen viivästynyt huomattavasti. Lisäksi litteraattien määrä olisi tätä myötä lisääntynyt, mikä olisi vuorostaan kasvattanut työtaakkaa piirteiden poiminnassa.

Osatutkimus INFINITIIVIT poikkeaa aineiston osalta muista osatutkimuksista. Tässäkin tutkimuksessa käytetään Helsingin puhekielen pitkittäiskorpusta, mutta koska osatutkimuksen aineisto on kerätty jo ennen väitöstutkimusta, se on rakenteeltaan erilainen. Kyseisen osatutkimuksen aiheena on myös rajattu piirre – infinitiivit – minkä takia siinä on voitu hyödyntää myös litteroimattomia haastatteluja ja poimia infinitiiviesiintymät pelkältä ääninauhalta. Toisaalta edellä mainittu sananrajailmiöiden litteroinnin epätarkkuus vaati nauhojen kuuntelua infinitiivejä tutkittaessa. INFINITIIVEISSÄ käytetty aineisto on esitetty taulukossa 1 sulkeissa. Taulukon luvuista huomataan, että vuosikymmenten välinen ero on tässä aineistossa huomattavasti pienempi kuin pääasiallisessa aineistossa. Siinä missä 1970-luvulta on vähemmän haastatteluja kuin pääasiallisessa aineistossa, 1990-luvulta ja 2010-luvulta niitä on enemmän.

Vaikka tutkimuksen aineistosta on löytynyt paljon kiinnostavia tutkimuskohteita, monet asiat jäävät aineiston struktuurin takia varjoon. Esimerkiksi keskittyminen kolmeen ikäryhmään (joiden ikähaarukka vielä vaihtelee vuosikymmeneltä toiselle) ei palvele erityisen hyvin elämäntilanteiden tutkimusta. Tästäkin on kuitenkin saatavissa havaintoja, koska puhujat itse kertovat haastatteluissa elämästään. Kuitenkin joissakin haastatteluissa tällaista tietoa puhujien elämästä ei juuri saada, jolloin puhujien elämäntilanteiden vertailu on hankalaa.

Kaikissa tutkimuksissa, metodeissa ja aineistoissa on omat heikkoutensa. On kuitenkin selvää, että Helsingin puhekielen pitkittäiskorpus on kansainvälisestäkin ainutlaatuinen tutkimuskohde, joka antaa mahdollisuuden tarkkaan reaaliaikaiseen kielen tutkimukseen. Tällaista reaaliaikaista sosiolingvistiikkaa edustavat kaikki väitöskirjan osatutkimukset, vaikka niissä kussakin on omanlaisiaan metodisia ja teoreettisia erikoisuuksia. Tämän lisäksi korpukselta voidaan tehdä havaintoja elämäntilanteista ja esimerkiksi kielellisistä asenteista (ks. slangiasenteet luvussa 3.3), vaikka näiden tarkastelu ei olekaan välttämättä ollut korpuksen alkuperäinen tarkoitus.

4.3 Tutkitut piirteet

Valitsin kielelliset piirteet tutkimuksen pääasialliseen aineistoon muutamain perustein. Ensinnäkin tarkoitus oli, että piirteissä olisi variaatiota Helsingin puhekielessä. Tässä yhteydessä hyödynsin erityisesti Heikki Paunosen tutkimuksia Helsingin puhekielestä, ja tutkitut piirteet vastaavatkin melko hyvin Paunosen (2006) listausta. Lisäksi valitsin joitakin piirteitä, koska olin havainnut niissä kiinnostavaa variaatiota käytännön kielenkäytössä. Tällaisia piirteitä ovat esimerkiksi infinitiivit ja muut

kaksitavuiseen ja pitkään vokaaliin päättyvään sanahahmoon pyrkivät piirteet (esim. *mitää, sillee*). Toisekseen rajasin piirteet niin, että niitä voisi kohtuullisella varmuudella löytää litteraateista. Näin ollen keskityin paljolti morfologisiin piirteisiin, kun taas esimerkiksi tarkkoja foneettisia mittauksia en ole voinut tehdä. Valitsin piirteitä lopulta 30 ja mahdollisia variantteja piirteillä on 78.⁵ Esittelen seuraavassa piirteet ja niiden variantit lyhyesti ja kommentoin joissakin tapauksissa myös niiden poiminnassa käytettyjä periaatteita. Kunkin variantin osalta ilmoitan myös esiintymien määrän koko aineistossa sulkeissa (N). Osatutkimuksessa INFINITIIVIT tarkasteltiin listauksessa mainittujen infinitiivitapausten lisäksi A-infinitiivejä kaikissa verbityypeissä.

A-loppuiset vokaaliyhtymät

- OA- ja eA -yhtymät
 - a) partitiivissa: *autoa* (N = 1270) ~ *autoo* (2042)
 - b) nomineissa: *korkea* (832) ~ *korkee* (3949) (myös osina muita rakenteita: *oikeastaan*)
 - c) infinitiiveissä: *kertoa* (424) ~ *kertoo* (1015)
- UA-yhtymä
 - a) partitiivissa: *koulua* (1449) ~ *kouluu* (792)
 - b) infinitiiveissä: *istua* (422) ~ *istuu* (324)
- iA-yhtymä
 - a) partitiivissa: *kabvia* (9571) ~ *kabvii* (4460)
 - b) infinitiiveissä: *leikkää* (266) ~ *leikkii* (126)

Infinitiivitäydennykset

- Supistumaverbien A-infinitiivit: *pelata* (588) ~ *pela* (68)
- MA-infinitiivin illatiivi: *pelaamaan* (2062) ~ *pelaan* (272) ~ *pela* (802)
- *käydä* ja *olla* -verbien infinitiivitäydennys: *käyn pelaamassa* (362) ~ *pelaamas* / *pelaas* (226) ~ *pela* (50)

Yleiskielen *d:n* vastineet

- Monikollisissa persoonapronomeineissa ja *tietää*-verbin taivutusmuodoissa: *meidän* (1567) ~ *meirän* (7) ~ *meiän* (1356)
- *b:n* jäljessä: *tehdä* (3128) ~ *tebrä* (14) ~ *tehä* (1937)
- Lekseemeissä *edes*, *ede-* (esim. *edessä*) ja *odottaa*: *edes* (382) ~ *eres* (2) ~ *ees* (176)

⁵ Osassa piirteitä on useampia alaluokkia, joten piirteiden määrä ei välttämättä ole kovin informatiivinen.

Yleiskielen *ts:n* vastineet

- Vahvassa asteessa lekseemeissä *etsiä, katsoa, seitsemän* ja *viitsiä: seitsemän* (706) ~ *seittemän* (796)
- Heikossa asteessa lekseemeissä *etsiä, itse, katsella, katsoa* ja *viitsiä: itse* (879) ~ *itte* (309) ~ *ite* (655)

Paikallissijat (vain sanan lopussa)

- Inessiivin päättevariaatio: *talossa* (15 076) ~ *talosa* (80) ~ *talos* (5455)
- Adessiivin päättevariaatio: *talolla* (18 204) ~ *talol* (9561)
- Elatiivin päättevariaatio: *talosta* (5706) ~ *talost* (2418)

Vielä -sanan variaatio

- *vielä* (1314) ~ *viel* (473) ~ *viä / vie* (82)

*ñ*n loppuheitto

- numeraaleissa: *yksi* (200) ~ *yks* (2896)
- konditionaalissa: *tulisi* (72) ~ *tulis* (3183)
- translatiivissa: *taloksi* (464) ~ *taloks* (2847)
- supistumaverbien imperfektissä: *pelasi* (335) ~ *pelas* (1036)

i-loppuinen diftongi

- *inen*-adjektiiveissa ja konditionaalissa (*antaisi*): *punainen* (442) ~ *punanen* (2405) (ei koske sanoja *semmoinen, tommoinen, tämmöinen, sellainen, mimmoinen, tällainen, tuollainen, millainen*)
- imperfektissä: *antoi* (328) ~ *anto* (2291)

NUT-partisiippi

- *hyppinyt* (1130) ~ *hyppiny* (9536)

Pikapuhemuodot

- *ei ole* (393) ~ *ei oo* (2083) ~ *ei o* (969)
- *olen, olet* (569) ~ *oon, oot* (2408)
- *tämä, nämä* (702) ~ *tää, nää* (2765)

Yksikön 1. ja 2. persoonan pronominit

- *minä, sinä* (1435) ~ *mää, sää* (275) ~ *mä, sä* (14 427)

3. persoonan pronominit

- *hän, he* (1755) ~ *se, ne* (5331)

Monikon 1. persoona

- *menemme* (844) ~ *mennää(n)* (8169) (vain selvästi mon. 1. persoonaa vastaavissa tapauksissa)

Sananloppuinen *n*

- leksemeissä *mitään, mikään, missään, millään, ketään, kellään, kukaan, koskaan, yhtään, varmaan. mitään* (3028) ~ *mitää* (2237)

Sanahahmovaihtelu

- *sillä / tällä / tolla tavalla, silleen, tälleen, tolleen, melkein, melkeen, oikein, oikeen* (2020) ~ *sillee, tälle, tolee, melkee, oikee* (1448)

Piirteistä suuri osa kuuluu suomalaisen variaation tutkimuksen perinteisiin tutkimuskohteisiin, eli niin sanottuun piirrekaanoniin, jota on ajoittain kritisoitu (esim. Lappalainen 2004: 15–16; Nuolijärvi & Sorjonen 2005: 130). Tämän piirrekaanonin ulkopuolelta tutkimuksessa on myös tarkasteltu esimerkiksi *käydä-* ja *olla-*verbien infinitiivitäydennyksiä (Ylikoski 2005) ja joidenkin sanojen pyrkimystä kohti kaksitaivuista, pitkään vokaaliin päättyvää sanahahmoa (*mitää, melkee, sillee*). Tutkimusta aloittaessani tarkoituksenani oli tarkastella, yleistyvätkö juuri tämän mallin sanahahmot Helsingin puhemielessä, mutta tätä aihetta en lopulta ehtinyt käsitellä. Kuitenkin esimerkiksi infinitiivit näyttävät pyrkivän kohti samaa sanahahmoa. Uusissa piirteissä ongelmaksi nousee se, vastaavatko esimerkiksi *sillä tavalla* ja *sillee* aidosti toisiaan (ks. Lappalainen 2004: 111–113). Olen keskittynyt tutkimuksessa piirteisiin, jotka voidaan kohtuullisen helposti litteroida (esim. onko inessiivissä loppuheittoa vai ei). Tämä jättää luonnollisesti tutkimuksen ulkopuolelle paljon sellaista variaatiota, jossa erot ovat jatkumoluontoisia (esim. diftongien avartuminen).

Sain haastattelujen litteraatit pääasiassa CSC:n Kielipankista, jossa ne olivat valmiina txt-muodossa. Osan litteraateista sain suoraan Hanna Lappalaiselta Praat-ohjelmiston TextGrid-muodossa. TextGrid-tiedostot muunsin txt-muotoon Praatiin kirjoitetulla skriptillä (Lennes 2009). Käsittelin haastatteluja txt-muodossa Notepad++ -ohjelmistolla, koska se tuki säännöllisten lausekkeiden käyttöä. Ennen varsinaisten hakujen aloittamista poistin haastatteluista ylimääräiset huomiot (esim. haastateltava vastaa puhelimeen) sekä haastattelijoiden osuudet. Joissakin haastatteluissa on useampia puhujia (esim. haastateltavan puoliso). Jotta poiminnat tulisivat vain haastateltavien puheesta, poistin myös nämä osuudet litteraateista.

Varianttien poimimisessa käytin säännöllisiä lausekkeita (Kleene 1956; Thompson 1968). Säännölliset lausekkeet etsivät teksteistä tiettyä jaksoa joko sellaisenaan tai tietyssä kohtaa sanaa tai riviä. Esimerkiksi siis hakusana $(ssa|ssä)\backslash b$ etsii inessiivin yleiskielen mukaista päätettä sanojen lopusta. Vain sanojen loppuun keskittyvää hakua käytin paikallissijojen kohdalla, joita esiintyy aineistossa valtavasti ja tämänkaltaisen haku selkeytti niiden rajausta. Edellä esitelty hakusana on vielä melko yksinkertainen, mutta monissa tapauksissa hakusanat muuttuivat melko monimutkaisiksi. Esimerkiksi painottoman tavun *i*-loppuisten diftongien tapauksessa jätin tarkastelematta sanoja *semmoinen, tommoinen, tämmöinen, sellainen, mimmoinen, tällainen, tollainen, millainen* (vrt. Paunonen 2006: 68), mikä tarkoitti sitä, että nämä sanat oli saatava pois myös hakutuloksista. Valtaosassa tapauksista en pystynyt rajaamaan säännöllistä lauseketta niin, että se olisi löytänyt vain haluamani variantit, eikä mitään muuta. Esimerkiksi A-loppuisten vokaaliyhtymien oienneiden varianttien osalta tämä on ilmeistä: haettava variantti on käytännössä vain pitkä vokaali, jolloin sellaista hakemalla aineistosta löytyy toivottujen osumien (esim. *oikee*) lisäksi paljon tarpeetonta (esim. *hän juoksee*). Vastaavasti esimerkiksi passiivitapauksissa minun piti erottaa varsinaisista passiiveista ne tapaukset, joissa passiivia käytetään monikon 1. persoonan sijaan. Säännölliset lausekkeet siis nopeuttivat etsintätöitä, mutta eivät suinkaan automatisoineet sitä.

Osumat koodasin varianttien nimillä litteraattitiedostoihin (esim. *oapart2 = OA*-yhtymän oiennut muoto partitiivissa). Näin pystyin lopuksi laskemaan ja lisäämään varianttien määrät taulukkoon automatisoidusti, mikä nopeutti työn tekemistä ja toisaalta varmisti, että laskutoimituksissa ei tapahtunut virheitä. Varianttien laskemisen ja taulukoimisen tein R-ohjelmistolla (R Core Team 2017), kuten valtaosan tutkimuksen analyysistakin. Kokonaisuudessaan prosessissa oli siis jonkin verran työtä helpottavaa automaatiota, mutta erityisen helpoksi sitä ei silti voi luonnehtia. Suuri osa tämänkin tutkimuksen varianteista lankeaa yhteen monien muiden kielenpiirteiden kanssa (kuten yllä A-loppuisten vokaaliyhtymien oikenemisen tapauksessa), minkä takia niitä ei voi edes säännöllisin lausekkein erottaa toisistaan. Kokonaisuudessaan aineiston poiminnassa kului aikaa muun työn ohessa noin vuosi.

4.4 Laskennalliset menetelmät

Sosiolingvistiikka on perinteisesti ollut tutkimusotteeltaan laskennallista. Tutkimussuuntaa on jopa kritisoitu liiallisesta laskennallisuudesta (ks. luku 2.2). Myös suomalaisessa variaationtutkimuksessa on pitkät perinteet laskennallisten menetelmien

käytössä. Paunosen (1995 [1982]) ja Nuolijärven (1986) tutkimukset Helsingistä ovat otteeltaan hyvinkin laskennallisia, vaikka niissä ei hyödynnetäkään varsinaisesti tilastollisia menetelmiä. Uudet menetelmät näkyvät sen sijaan 2000- ja 2010-luvun väitöskirjoissa, joista Tommi Kurjen (2005) väitöskirja hyödyntää kansainvälisesti tunnettua Varbrul-analyysityökalua, Liisa Mustanojan (2011) bayesilaista tilastotiedettä ja Katri Priikin (2017) päätöspuita ja satunnaismetsää. Käsillä oleva väitöstutkimus liittyy tähän joukkoon. Kaikissa väitöskirjan osatutkimuksissa on hyödynnetty jotakin tilastollista menetelmää. Seuraavassa esittelen osatutkimuksissa hyödynnetyt tilastolliset menetelmät lyhyesti. Tarkemmat kuvaukset kustakin menetelmästä on esitetty osatutkimuksissa.

Osatutkimuksessa INFINITIIVIT hyödynnettiin khiin neliö -testejä (Pearson 1900) vuosikymmen- ja ikäryhmäkohtaisten erojen selvittämiseksi. Testin nollahypoteesina on, että vertailtavien ryhmien välillä ei ole eroa. Mikäli nollahypoteesi hylätään ($p < 0,05$), saadaan ryhmien välisille eroille tilastollista tukea. Khiin neliö -testien avulla tutkimuksessa saatiin kuva infinitiiveissä tapahtuvista muutoksista reaaliajassa, sekä ikäryhmien välisestä variaatiosta. INFINITIIVEISSÄ toteutettiin myös logistinen regressio (Cox 1958) *alkaa*-verbin täydennyksille sen selvittämiseksi, minkä tyyppiset seikat vaikuttavat infinitiivitäydennyksen valintaan (vrt. Herlin 2012). Logistisessa regressiossa selitettävä muuttuja on kategorinen (tässä tapauksessa infinitiivitäydennyksen laatu), ja selittävät tekijät voivat olla myös suhde- tai järjestysasteikollisia.

Osatutkimuksessa MUUTOSMALLIT haluttiin selvittää Labovin muutosmallien (1994: 83–84) toteutumista käytännössä. Koska muutosmallit käsittelevät yksilö- ja yhteisötason muutosta, piti myös mahdollisen tilastollisen mallin tarkastella näitä tasoja. Niinpä osatutkimuksessa muotoiltiin kaksi väitettä, joita tutkia tilastollisesti:

Väite 1: Samojen ikäryhmien kielenkäytössä ei ole eroja eri vuosikymmenien välillä (esim. 1970-luvun nuoret ja 1990-luvun nuoret eivät eroa toisistaan).

Väite 2: Samojen puhujien kielenkäytössä ei ole eroja eri vuosikymmenien välillä (esim. 1970-luvun nuoret ja 1990-luvun keski-ikäiset eivät eroa toisistaan).

Jotta näitä väitteitä voitiin tutkia tilastollisesti, oli vertailtava keskenään samoja ikäryhmiä eri vuosikymmenillä (väite 1) ja samoja puhujia eri vuosikymmenillä (väite 2). Tyypillisesti tällaisissa otoskohtaisissa vertailuissa käytettäisiin Studentin t-testiä (Student 1908), mutta koska tässä aineistossa (ja kieliaineistoissa tyypillisesti) on paljon vinoja jakaumia sekä jakauman ulkopuolisia havaintoja, tutkimuksessa hyödynnettiin ei-parametrisiä testejä. Ensimmäisen väitteen tutkimiseen käytettiin Mann-Whitneyn U-testiä (Wilcoxon 1945; Mann–Whitney 1947) ja toisen väitteen osalta Wilcoxonin merkittyjen sijalukujen testiä (Wilcoxon mt.). Molemmat testit perustuvat sijalukujen

muutokseen eli niissä muutetaan havaitut arvot järjestyslukuiksi, ja muutosta tulkitaan järjestyslukujen perusteella. Tilastollisen merkitsevyyden raja-arvona käytettiin näissäkin testeissä arvoa $p < 0,05$.

Wilcoxonin merkittyjen sijalukujen testiä käytettiin samaan tarkoitukseen myös osatutkimuksessa KATOMUODOT. Tutkimuksessa seurattiin samojen puhujien (kohortin) muutosta yleiskielen *bd*-yhtymässä esiintyvän kadon käytössä. Kyseisessä osatutkimuksessa Wilcoxonin testi oli kuitenkin melko pienessä osassa, sillä KATOMUODOISSA pääasiallinen tilastollinen menetelmä oli korrelaatioanalyysi. Korrelaatioiden avulla verrattiin toisiinsa morfologisesti ja semanttisesti perusteltuja muotoryhmiä. Mikäli korrelaatio oli jollakin vuosikymmenellä tilastollisesti merkitsevä ($p < 0,05$) kahden eri muotoryhmän välillä, pääteltiin katomuotojen edenneen kyseisiin muotoryhmiin. Korrelaatioanalyysit tehtiin vain sellaisille muotoryhmäpareille, joissa katoa esiintyi yli 10 prosentissa kaikista tapauksista ja kadon määrä kasvoi vuosikymmeneltä toiselle. Korrelaatioanalyysin ja suhteellisten osuuksien tarkastelun avulla seurattiin morfologista diffuusiota (Suihkonen 1992: 304; Mielikäinen 1995) yleiskielen *bd*-yhtymässä Helsingin puhekielessä.

Khiin neliö -testit, logistiset regressiot, korrelaatiot ja sijalukutestit ovat tilastotieteen perinteisiä työkaluja. Osatutkimuksessa LEKTIT hyödynnettiin uudempaa LDA-mallia (Latent Dirichlet Allocation; Blei, Ng & Jordan 2003). LDA-mallia on perinteisesti hyödynnetty tekstien luokitteluun niissä käytetyn sanaston avulla. Tästä syystä mallista puhutaan myös aihemallina (engl. *topic model*). Kun tietyt sanat esiintyvät teksteissä usein yhdessä, malli pääättelee niiden käsittelevän samaa aiheetta. Malli laskee todennäköisyyksiä tietyn sanan esiintymiselle tietyssä aiheessa ja toisaalta tietyn aiheen esiintymiselle tietyssä tekstissä. Sanat voivat siis kuulua useampaan aiheeseen, ja tekstit koostua useammasta aiheesta.

Osatutkimuksessa LEKTIT LDA-mallia käytettiin kuitenkin kielellisten varianttien luokitteluksi toisistaan poikkeaviksi lekteiksi. Näin ollen sanojen sijaan tutkimuksessa tarkasteltiin variantteja (esim. inessiivin loppuheitto) ja tekstien sijaan puhujien haastatteluja. Mikäli siis esimerkiksi inessiivin loppuheitto (*talos*) ja oiennut *iA*-yhtymä (*peiliä*) esiintyvät haastatteluissa usein yhdessä, malli alkaa olettaa niiden kuuluvan samaan laskennalliseen komponenttiin. Komponentit ovat siis mallin tuottamia varianttijoukkoja. Mikäli jonkin yksilön haastattelussa näyttää olevan paljon tiettyyn komponenttiin kuuluvia variantteja, malli olettaa tämän komponentin olevan merkittävä kyseisessä haastattelussa. Osatutkimuksessa eroteltiin toisistaan mallin tuottamat komponentit ja varsinaiset lekit, jotka johdettiin komponenteista kielitieteellisen analyysin avulla. Lektillä tarkoitettiin osatutkimuksessa sellaista useiden yhdessä esiintyvien varianttien joukkoa (eli laskennallista komponenttia), joka oli

selvästi muista komponenteista erillinen. Mallin etuna perinteisempiin klusterointimenetelmiin ja ulotteisuuden pudotukseen (esim. pääkomponenttianalyysi, faktori-analyysi, k-means) on sen inklusiivinen luonne: variantit voivat kuulua useampaan komponenttiin.

Väitöskirjassa on siis sovellettu useita tilastollisia menetelmiä, eikä keskitytty vain yhteen. Kunkin osatutkimuksen yhteydessä on lähdetty liikkeelle tutkimusongelmasta: mitä halutaan selvittää? Vasta tämän jälkeen on päädytty tiettyjen menetelmien valintaan. Tällainen lähestymistapa antaa sosiolingvistikalle tutkimukselle enemmän liikkumavaraa kuin keskittyminen yksittäisiin menetelmiin. Esimerkiksi Varbrul nousi kansainvälisessä sosiolingvistiikassa jo hieman liiankin korosteiseen asemaan 1980- ja 90-luvuilla (Kurki 2005: 42–43).

5 TUTKIMUKSEN TULOKSET

Tässä luvussa esitän väitöskirjan keskeiset tulokset. Tarkastelu on jaettu kolmeen osaan: alaluvussa 5.1 keskityn muutoksen kuvaamisessa piirretasolle: mitkä piirteet ovat muuttuneet Helsingin puhekielessä 1970-luvulta 2010-luvulle ja miten? Luvussa 5.2 puolestaan tarkastelen muuttuvia varieteetteja: millaisia lektejä Helsingin puhekielestä löytyy ja miten niiden käyttö on kaupungissa ajan kuluessa muuttunut? Viimeisessä alaluvussa 5.3 esittelen muutosmallien teoriasta tutkimuksessa paljastuneita seikkoja ja luon synteesin niiden kyvystä kuvata erilaisia muutoksia. Samassa yhteydessä tarkastelen myös muutosmallien ja S-käyrän yhteensovittamista.

5.1 Muuttuvat piirteet

Tässä tutkimuksen tuloksia käsittelevässä alaluvussa tarkastelen muutosta muutamissa Helsingin puhekielen piirteissä. Luvussa 5.1.1 esittelen muutosta infinitiiveissä ja luvussa 5.1.2 muutosta yleiskielen *hd*-yhtymässä. Molemmat alaluvut perustuvat osatutkimuksiin, luku 5.1.1 osatutkimukseen INFINITIIVIT ja luku 5.1.2 osatutkimukseen KATOMUODOT.

Luvuissa 5.1.3 ja 5.1.4 esittelen lyhyesti kaksi piirrekimppua, jotka olivat tarkastelun kohteena myös osatutkimuksessa MUUTOSMALLIT. Koska osatutkimuksen tavoitteena oli ennen kaikkea tarkastella muutosmallien sopivuutta käytäntöön, ei osatutkimuksessa esitelty kuin harvoja muutosmalleihin sopivia esimerkkivariantteja. Osatutkimuksesta poiketen käsittelen alaluvuissa piirrekimppuja kokonaisuuksina riippumatta siitä, sopivatko ne tilastollisesti kuvaamaan tiettyjä muutosmalleja. Tilastollinen tarkastelu on kuitenkin alaluvuissa osatutkimuksen MUUTOSMALLIT mukainen: kutakin varianttia on tarkasteltu niin yksilö- kuin yhteisötasolla sijalukutestein (ks. luku 4.4) ja tilastollisen merkitsevyyden raja-arvona on $p < 0,05$.

Käsittelen kahta fennistiikassa perinteisesti tutkittua piirrettä: A-loppuisia vokaaliyhtymiä luvussa 5.1.3 sekä yleiskielen *ts*-yhtymän vastineita luvussa 5.1.4. A-loppuisien vokaaliyhtymien oikenemista on pidetty yhtenä nopeimmin leviävistä nykysuomen piirteistä (Mantila 2004: 326–327). *ts*-yhtymän osalta Paunonen (1995 [1982]: 67–68) on puolestaan ennustanut, että Helsingissä on vakiintumassa vaihtelu $\# : t$.

Näissä alaluvuissa esittelen, vastaavatko Helsingin puhekielestä saadut tulokset olettuksia näiden varianttien yleistymisestä.

5.1.1 Infinitiivit

Suomen yleisimmät infinitiivimuodot ovat molemmat (historiallisesti) tulosijaisia. A-infinitiivin perusmuoto eli latiiivi on infinitiivimuodoista kaikkein yleisin ja MA-infinitiivin illatiivi toiseksi yleisin (Herlin, Leino & Visapää 2005: 17). Latiiivin tulosijaisuus on ajan myötä hävinnyt, ja se toimii nykyään pääasiassa merkitykseltään abstraktien verbien täydennyksenä. Tulosijainfinitiivien työnjakoa on esitellyt tarkasti Saukkonen (1965).

Osatutkimus INFINITIIVIT lähti liikkeelle havainnosta, että A-infinitiivin lyhyt muoto ja MA-infinitiivin illatiivi esiintyvät monissa tapauksissa puheessa täsmälleen samassa, pitkään vokaaliin päättyvässä asussa. Tähän muotojen yhteenlankeamiseen ovat johtaneet seuraavat äänne muutokset:

- a) A-loppuisten vokaaliyhtymien oikeneminen (*kertoa* > *kertoo*)
- b) MA-infinitiivin illatiivien MA-tunnuksettomuus (*antamaan* > *antaa*)
- c) Sananloppuisen *n*-äänneen kato (*antaa* > *antaa*)
- d) Supistumaverbien analogiset muodot (*pelata* > *pelaa*)

Sekä A-loppuisten vokaaliyhtymien oikeneminen (a) että MA-infinitiivin tunnus-
tomuus (b) ovat ekspansiivisia piirteitä puhesuomessa (Mantila 2004: 326–327; Nuolijärvi & Sorjonen 2005: 97–98). Sananloppuisen *n*:n kato on vanha ilmiö, ja kadon laajuus riippuu pitkälti äänneympäristöstä (Rapola 1966: 325–327). Sorsakivi (1982: 383–384) esittää, että sananloppuisen *n* (kuten MA-infinitiivin illatiivissa) on liittymässä sananloppuisen *k*:n sandhikaavaan (kuten A-infinitiivin perusmuodossa). Kolme ensimmäistä piirrettä ovat siis enemmän tai vähemmän tavallisia etenkin Helsingissä, mutta myös muualla Suomessa. Viimeinen kohta (d) on sen sijaan alun perin etelähämäläinen. Itkonen (1964: 187–193) esittää supistumaverbien analogisten muotojen olevan seurausta ruotsinkielisten kielenvaihdosta.

Edellä esitellyt äänne muutokset johtavat tilanteeseen, jossa sekä A-infinitiivin perusmuoto että MA-infinitiivin illatiivi esiintyvät samassa asussa: *voin antaa ~ kertoo ~ pelaa – meen antaa ~ kertoo ~ pelaa*. Tämä infinitiivimuoto nimettiin osatutkimuksessa lyhyeksi infinitiiviksi. Infinitiivimuodon tulkinta erillään A- ja MA-infinitiivistä perustuu siihen, että lyhyt infinitiivi näyttää osittain irtautuvan lähtöinfinitiiveistään ja leviävän analogisesti. Analogisena mallina on toiminut ennen kaikkea verbityyppi

antaa (vrt. Itkonen 1964: 187–193). Kyseisen infinitiivimuodon tulkinta vain esimerkiksi MA-infinitiivin illatiivin varianttina piilottaisi sen etenemisen syyt. Näin piirre toimii varoittavana esimerkkinä siitä, että keskittyminen suomalaisen variaation tutkimuksen piirrekaanoniin voi piilottaa kiinnostavia ilmiöitä (Nuolijärvi & Sorjonen 2005: 130).

Osatutkimuksessa tarkasteltiin näiden kolmen infinitiivimuodon variaatiota ja muutosta kolmella vuosikymmenellä aineistossa, jonka esittelin luvussa 4.2. Aineisto poikkesi kaikkien muiden osatutkimuksen aineistosta siitä syystä, että se on kerätty jo aiemmin ja hieman erilaisin menetelmin. Lyhyttä infinitiiviä käytetään myös *käydä*- ja *olla*-verbien täydennyksenä, mitä tarkasteltiin lyhyesti myös osatutkimuksessa. Tarkemmin näiden verbien täydennyksiä on tarkastellut Ylikoski (2005). Sananrajailmiöiden osalta osatutkimuksessa havaittiin, että lyhyt infinitiivi noudattaa vielä pitkälti lähtöinfinitiivinsä sananloppuisen *k:n* (A-infinitiivi) ja *n:n* (illatiivi) sandhikaavoja.

Muutosta tarkasteltiin tilastollisesti khiin neliö -testien avulla (ks. luku 4.4). Ikäryhmien välillä oli kaikilla vuosikymmenillä ja kaikissa ikäryhmissä tilastollisesti erittäin merkitseviä eroja ($p < 0,001$). Lyhyt infinitiivi oli yleisin infinitiivimuoto nuorten puheessa kaikilla vuosikymmenillä, ja sen määrä kasvoi ajan edetessä myös muissa ikäryhmissä. Näin ollen Helsingin puhekielessä on meneillään muutos kohti lyhyen infinitiivin käyttöä A-infinitiivin ja MA-infinitiivin kustannuksella. Osa verbityypeistä kuitenkin esiintyy lyhyessä infinitiivissä hyvin harvoin (konsonanttivartaloiset ja *dA*-verbit; *meen juoksee* – **voin saa*). Nämä verbityypit kattavat aineiston infinitiiviesiintymistä noin 30 %, jolloin muutokselle jää vähemmän tilaa. Analogisen mallivaikutuksen kannalta on todennäköistä, että lyhyt infinitiivi leviää näistä verbityypeistä ensin konsonanttivartaloisiin verbeihin, joista on mahdollista tehdä mallin mukainen kaksitavuinen ja pitkään vokaaliin päättyvä muoto (*juoksee*). Sen sijaan lyhyen infinitiivin eteneminen *dA*-verbeihin vaikuttaa vielä epätodennäköiseltä.

Infinitiivien käyttö on osittain ikäsidonnaista: nuorten ja keski-ikäisten välillä oli osatutkimuksessa aina huomattavan suuri ero. Keski-ikäisten ja vanhojen puhujien välillä ero jäi pienemmäksi. Muutos infinitiiveissä oli osittain ikäsidonnaista ja osittain sukupolvittaista, koska yhteisö kyllä muuttui, mutta lyhyen infinitiivin käyttö väheni merkittävästi nuoruudesta keski-ikään. Näin ollen muutos noudatti vetäytyvän muutoksen mallia (Wagner & Sankoff 2011), jota ei kuitenkaan itse osatutkimuksessa käsitelty. Muutosmalleja koskevat huomiot jalostuivat myöhemmin osatutkimukseksi MUUTOSMALLIT.

Lyhyen infinitiivin käyttö jakautui myös sen mukaan, kumman lähtöinfinitiivin tehtävissä se esiintyi. Nuoret käyttivät lyhyttä infinitiiviä paljon myös MA-infinitiivin illatiivin sijaan, kun taas vanhemmilla puhujilla lyhyen infinitiivin esiintymät

painottuivat pitkälti alun perin A-infinitiiviä edellyttäviin tapauksiin. Tämä on luonnollista sikäli, että A-loppuisten vokaaliyhtymien oikeeneminen on yleinen ja neutraali puhekielisyys (Mantila 2004: 326–327; ks. luku 5.1.3).

5.1.2 *hd*-yhtymä

Yksi suomalaisen murteentutkimuksen ja sociolingvistiikan keskeisimmistä tutkimuskohteista on ollut *tn* heikko aste eli yleiskielen *d*. Tämä on seurausta siitä, että mahdollisia variantteja on useita ja ne erottavat toisistaan esimerkiksi läntisiä ja itäisiä murteita. Yleiskielen *d*:n vastineina ovat länsimurteissa historiallisesti olleet *r*, *ð* ja *l*, kun taas itämurteissa suositaan katoa ja sen siirtymä-äänteitä *j*, *v* ja *b* (Rapola 1966: 99–110). Lounaismurteissa on vielä 1900-luvun alkupuolella tavattu myös näitä äännteitä edeltänyttä soinnillista dentaalispiranttia *ð* (Kettunen 1930: 104; Virtaranta 1958: 12–13). Yleiskielessä tavattava *d* on kirjakielen ja ruotsin mallista omaksuttu äänne (Lehtinen 2007: 178–181), jonka käytöstä puheessa on kiistelty vielä 1800-luvun loppun fennistien kesken (Paunonen 1976: 338). Hämmäläismurteille tyypillinen *l* on 1900-luvun saatossa kadonnut lähes kokonaan käytöstä, ja se on korvautunut yleisesti *r*:llä (Virtaranta 1958: 213–216). *ð* on murteentutkimuksessa usein esitetty yksitai heikkotäryisenä tremulanttina (Setälä 1890: 29; Itkonen 1973: 26). Tämä määrittely on kuitenkin ongelmallinen siksi, että suomen *r* on niin ikään hyvin usein yksitäryinen. Usein on siis epäselvää, mikä oikeastaan on äänneiden ero niissä tapauksissa, joissa on litteroitu sekä *r*- että *ð*-tarketta. Näitä litterointitapojen ja määritelmien ongelmia tarkastelevat lähemmin Mustanoja ja O’Dell (2007).

Vaikka *r*, *ð* ja *l* ovat läntisiä variantteja, esiintyy länsimurteissa myös katoa joissakin lekseemeissä. Katomuotoja esiintyy kaikissa äänneasemissa (pitkän vokaalin jäljessä, lyhyen vokaalin jäljessä ja *b*:n jäljessä) ja tyypillisesti taajaan esiintyvissä lekseemeissä. Näitä ovat esimerkiksi persoonapronominien genetiivimuodot (*meiän*, *teiän*), tietää-verbin taivutus (*en tiää*), lekseemi *edes* (*ees*) ja lukusanat (*yhdeksän*, *kabe-*). Kaikkein yleisin kato on juuri *b*:n jälkeisessä asemassa (Nuolijärvi 1986: 112; Paunonen 1995 [1982]: 32). Syynä tähän ovat toisaalta useat taajaan esiintyvät lekseemit (esim. *yhdeksän*, *yksi*, *lähteä*, *tehdä*) ja toisaalta se, että *b*:n jälkeisessä asemassa *d*:n kato ei aiheuta hankalia vokaali- tai konsonanttityhtymiä ja uusien siirtymä-äänneiden tarvetta.

Katomuotojen etenemistä yleiskielen *d*:n vastineeksi *b*:n jäljessä tarkasteltiin osatutkimuksessa KATOMUODOT. Katomuotoja on länsimurteiden alueella ja Helsingissäkin tutkittu jo ennestään paljon (esim. Nuolijärvi 1986: 112–114; Paunonen 1995 [1982]: 44–55; Kurki 2005: 107–119; Mustanoja 2011: 318–329). Osatutkimuksessa

keskityttiin katomuotojen etenemisen kuvaukseen morfologisen ja leksikaalisen diffuusion mallien avulla (Wang 1969; Suihkonen 1992: 286; Mielikäinen 1995; luku 2.4).

Osatutkimuksessa havaittiin, että katomuodot yleistyvät Helsingin puhekielessä vuosikymmeneltä toiselle. Muutos näyttää etenevän pääasiassa sukupolvittaisen muutoksen mallilla, mitä Paunonen (2005a: 48) epäili jo kahden ensimmäisen haastattelukierroksen perusteella. Suurimman osan katomuotojen yleistymisestä selittää pieni joukko lekseemejä, jotka ovat hyvin suuritaajuisia ja esiintyvät jo 1970-luvun aineistossa runsaasti katomuodoissa (erityisesti *yhdeksän*, *kahdeksan*, *yhde*-, *kahde*-, *lähde*-, *ehdi*-, *tehdä*, *nähdä*). Katomuodot näyttävät etenevän kuitenkin myös uusiin lekseemeihin: niin 1990- kuin 2010-luvullakin esiintyy katomuodoissa joukko sellaisia lekseemejä, joita aiemmilla vuosikymmenillä vielä ei ole katomuodoissa esiintynyt. Tätä havaintoa pyrittiin selittämään diffuusiomallilla.

Morfologisen diffuusion kuvaamisen apuna käytettiin korrelaatioanalyysia (ks. luku 4.4). Korrelaatioanalyysit tehtiin sellaisille muotoryhmille, joissa esiintyi tutkitavalla vuosikymmenellä katoa yli 10 prosentissa kaikista tapauksista ja joissa kadon määrä kasvoi lineaarisesti aineiston ajanjakson aikana. Mikäli muotoryhmien välillä oli jollakin vuosikymmenellä tilastollisesti merkitsevä ($p < 0,05$) yhteys, pääteltiin kadon edenneen näihin muotoryhmiin. Tutkitut muotoryhmät, muutama esimerkkilekseemi ja esiintymismäärä koko aineistossa on esitetty taulukossa 3. Täydellinen listaus lekseemeistä on julkaistu osatutkimuksessa.

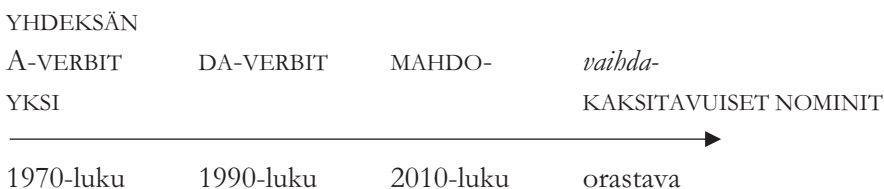
Taulukko 3. Osatutkimuksessa KATOMUODOT tutkitut muotoryhmät, esimerkkilekseemit ja esiintymismäärä koko aineistossa (N).

Muotoryhmä	Lekseemit	N
YHDEKSÄN	<i>yhdeksän</i> , <i>kahdeksan</i>	1034
YKSI	<i>yhde</i> -, <i>kahde</i> -	852
A-VERBIT	<i>lähde</i> -, <i>ehdi</i> -, <i>johdu</i> -, <i>mahdu</i> -, <i>pohdi</i> -, <i>ryhdy</i> -, <i>tahdo</i> -	723
DA-VERBIT	<i>tehdä</i> , <i>nähdä</i>	747
MAHDO-	<i>mahdoton</i> , <i>mahdollinen</i> , <i>mahdollisuus</i> , <i>mahdollistaa</i> , <i>ehdoton</i>	527
LAHTI- JA LEHTONIMET	<i>Aurinkolahde</i> -, <i>Hakalahde</i> -, <i>Haukilahde</i> -, <i>Hietalahde</i> -, <i>Sofianlehdo</i> -	240

KOHDA-	<i>kohda-, kiinnekohda-, kohokohda-, kohde, kohden, kohdata, kohdella</i>	158
PITKÄ VOKAALI	<i>hiihdä-, hiihdo-, huuhdo-, paahda-, paahdin, viihde, viihdy-, viihdyttää</i>	147
DIFTONGI	<i>kaihda-, kiehdo-, köyhdyttää, vaihde, vaihda-, vaihdo-, vaihdu-, vauhdi-</i>	155
PAINOTON TAVU	<i>erehdys, hätkähdyttää, huolehdi-, kömmähdys, kadehdi-, kiiruhda-, kolahda-</i>	100
KAKSITAVUISET NOMINIT	<i>ahdas, ehdo-, johdo-, lähdö-, lähde, lahde- (ei nimenä), lehde-, lyhdy-</i>	303
YHDESSÄ	<i>yhdessä, yhdistää, yhdistäminen, yhdistelmä, yhdistyä, yhdistys, yhdyskunta</i>	319

Osatutkimuksessa havaittiin, että katomuodot olivat jo 1970-luvun aineistossa tavallisia muotoryhmissä YHDEKSÄN, YKSI ja A-VERBIT. Nämä olivat 1970-luvun aineistossa ainoat keskenään tilastollisesti merkitsevästi korreloivat muotoryhmät. 1990-luvun aineistossa myös muotoryhmä DA-VERBIT alkoi korreloida tilastollisesti merkitsevästi kolmen ensimmäisen muotoryhmän kanssa. Tästä havainnosta pääteltiin katomuotojen edenneen muotoryhmään viimeistään 1990-luvulla. Muotoryhmä MAHDO- alkoi puolestaan korreloida tilastollisesti merkitsevästi DA-VERBIEN kanssa 2010-luvun aineistossa, mistä pääteltiin kadon edenneen muotoryhmään MAHDO- viimeistään 2010-luvulla.

Kadon osuus oli selvästi kasvussa myös muotoryhmissä DIFTONGI ja KAKSITAVUISET NOMINIT, vaikka nämä muotoryhmät eivät korreloineetkaan muiden kanssa tilastollisesti merkitsevästi. Kadon katsottiin näissä muotoryhmissä vasta orastavan. Lekseemitason tarkastelussa havaittiin lisäksi, että muotoryhmän DIFTONGI katotapauksista valtaosa esiintyi lekseemissä *vaihda-*. Tästä syystä kadon ei katsottu edenneen koko muotoryhmään vaan pelkästään kyseiseen lekseemiin. Katomuotojen suhteellisten osuuksien ja korrelaatioanalyysin perusteella hahmoteltiin katomuotojen diffuusiojärjestys Helsingin puhekielessä seuraavanlaiseksi.



Osatutkimuksen ansiona voidaan pitää tähän asti lähinnä teoreettisena mallina toimineen morfologisen diffuusion soveltamista käytännön aineistoon laskennallisesti.

Lisäksi osatutkimus osoitti, että kadon eteneminen ei suinkaan ole pysähtynyt, kuten esimerkiksi Nuolijärvi (1986: 114) esittää, vaan se etenee yhä uusiin muotoryhmiin ja lekseemeihin. On muistettava, että *bd*-yhtymän sisältäviä lekseemejä on lopulta melko vähän (tässä tutkimuksessa 191), joista vain muutama on hyvin taajaan esiintyviä. Näin ollen kadon eteneminen näiden korkean frekvenssin sanojen takana saattaa helposti jäädä huomiotta.

5.1.3 A-loppuiset vokaaliyhtymät

A-loppuisilla vokaaliyhtymillä tarkoitetaan seuraavia vokaaliyhtymiä: *OA*, *eA*, *UA* ja *iA*. Näistä kahden ensimmäisen oikeneminen on tavallinen piirre hämäläis- ja savolaismurteissa, kun taas kaksi viimeistä eivät oikene yhtä usein murteissa (piirteiden levikistä ks. Kettunen 1940: kartat 191–192; Mielikäinen 1991: 36–39). Helsingin puhekielen hämäläisasutuksen seurauksena *OA*- ja *eA*-yhtymien oikeneminen on ollut myös pääkaupungissa tavallista. Esimerkiksi Tampereen seudusta poiketen etelähämäläisissä murteissa esiintyy myös *UA*- ja *iA*-yhtymien oikenemistä (Kettunen 1940: kartta 197). Näiden Helsingin lähellä puhuttavien murteiden vaikutus on ollut luonnollisesti suuri myös pääkaupungin puhekielessä. Paunonen (1995 [1982]: 111) esittää, että yhtymien oikenemisjärjestys Helsingin puhekielessä on $eA > OA > UA > iA$.

Tässä tutkimuksessa olen tarkastellut A-loppuisia vokaaliyhtymiä kolmessa joukossa ja kolmessa muotoryhmässä. Olen aineistonkeruuvaiheessa yhdistänyt *OA*- ja *eA*-esiintymät samaan joukkoon, kun taas *UA*- ja *iA*-yhtymät olen pitänyt omina ryhminään. Jälkikäteen tarkastellen tämä ratkaisu jättää toivomisen varaa: kaikkein selvintä olisi ollut pitää kaikki ryhmät erillään, vaikka *OA*- ja *eA*-yhtymien oikeneminen eteneekin usein samassa tahdissa. Jokaisen yhtymän osalta olen tarkastellut partitiivimuotoja (*kabvia*) ja A-infinitiivejä (*sattua*), joiden lisäksi vielä *ea*- ja *oa*-nomineja (*korkea*). Aloitan tarkastelun *OA*- ja *eA*-yhtymistä. Kuviossa 1 on esitetty kaikkien muotoryhmien osalta vaihtelu ja muutos näissä vokaaliyhtymissä.

Kuvio 1 (kuten myös kuvat 2, 3 ja 4) ovat ruutu- ja janakaavioita. Samaa kaavio-tyyppiä on käytetty myös kaikissa osatutkimuksissa paitsi INFINITIIVEISSÄ. Kaavio-tyyppi kuvaa hyvin ryhmän sisäistä vaihtelua: mikäli ruudut ovat suuria, ryhmä ei ole yhtenäinen. Ruutu on piirretty tutkitun joukon alaneljänneksestä ylaneljännekseen (25 %–75 % kaikista puhujista). Koko joukon mediaaniarvo on esitetty vaakaviivana ruudun sisällä. Joissakin tapauksissa kuvioissa voi näkyä vain viiva. Tällöin kaikki joukon arvot ovat hyvin lähellä mediaania. Nämä ovat tyyppillisesti tapauksia, joissa

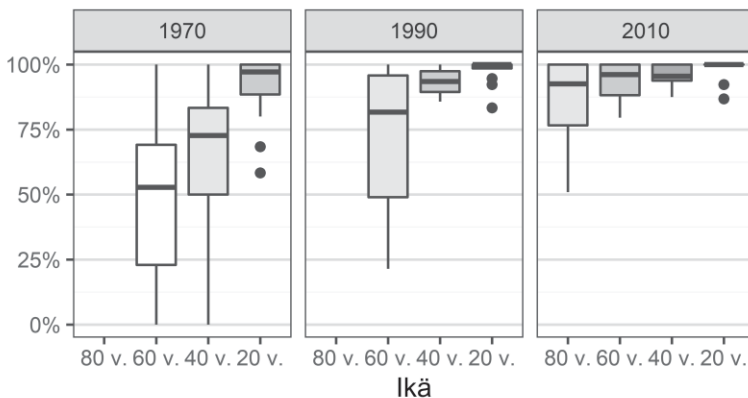
mediaani on joko 0 tai 100 prosenttia. Janat ruudun molemmissa päissä osoittavat joukon suurinta ja pienintä arvoa, ellei joukossa ole huomattavasti muista poikkeavia arvoja. Tällaiset poikkeavat arvot merkitään kaavioon pistein, mikäli ne sijoittuvat yli 1,5 ruudun mitan etäisyydelle ruudusta. Kaavioissa on lisäksi kuvattu haastattelu-aika yläotsikkona, ikäryhmä x-akselilla ja puhujaryhmien syntymäajat harmaan eri sävyin. Sama puhujaryhmä pysyy siis samanvärisenä kaikissa haastattelupisteissä (vrt. taulukko 1).

Kuviota 1 tarkastelemalla näyttää siltä, että *OA*- ja *eA*-yhtymien oikeneminen on Helsingin puhekielessä muuttumassa lähes yksinomaiseksi. Erityisesti nomineissa vanhimmatkin puhujat käyttävät oienneita muotoja lähes kaikissa tapauksissa. Osatutkimuksessa MUUTOSMALLIT nominien muutos oli osittain yhteensopiva elinikäisen muutoksen mallin kanssa: 1930-luvun taitteessa syntyneiden kohortti kasvatti oienneiden muotojen käyttöä tilastollisesti merkitsevästi 1970-luvulta 1990-luvulle eli keski-ikästä vanhuuteen ($p = 0,02$). Määrällisesti muutos oli 51 prosentista ($N = 119/234$) 73 prosenttiin ($253/347$). Vastaavalla aikavälillä myös yhteisö muuttui merkitsevästi keski-ikäisten puhujien osalta ($p < 0,001$). Tässä ryhmässä oienneiden muotojen määrä kasvoi 64 prosentista ($451/745$) 93 prosenttiin ($251/270$)⁶.

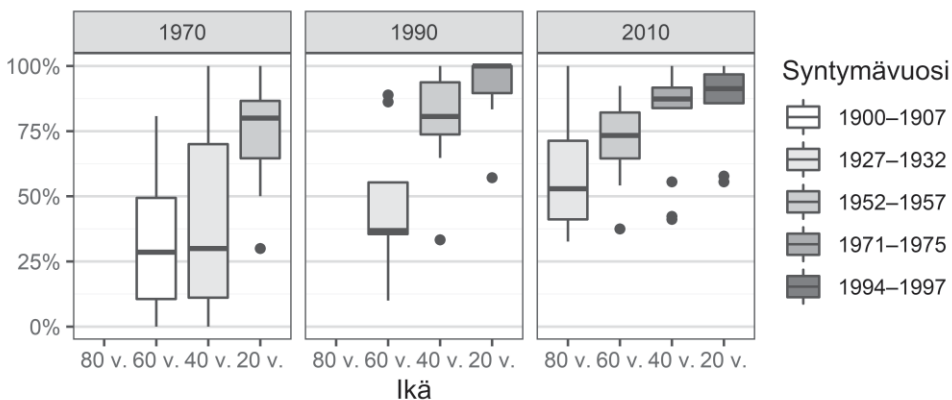
⁶ Määrälliset osuudet havainnollistavat tapahtuneen kokonaismuutoksen suuruutta. Käytetyt testit eivät kuitenkaan perustu ryhmän kokonaismuutokseen, vaan ryhmän edustajien muutokseen. Ryhmien sisäiset erot ja mediaaniarvojen muutos ovat puolestaan luettavissa kuvioista. Esitystapa on samanlainen kuin osatutkimuksessa MUUTOSMALLIT.

Kuvio 1. OA- ja eA-yhtymien oikeneminen eri muotoryhmissä Helsingin puhekielessä. Kohortit on esitetty harmaana sävyin, x-akselilla on puhujien ikäryhmä haastatteluhetkellä ja kukin vuosikymmen on asetettu omaan taulukkoonsa.

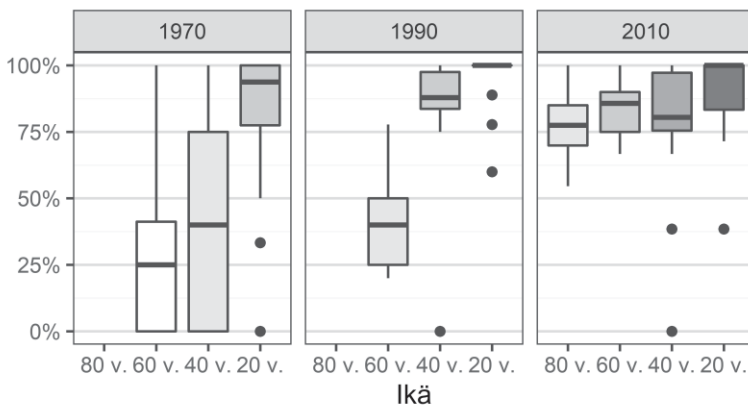
A OA- ja eA-yhtymien oikeneminen nomineissa



B partitiiveissa



C infinitiiveissä



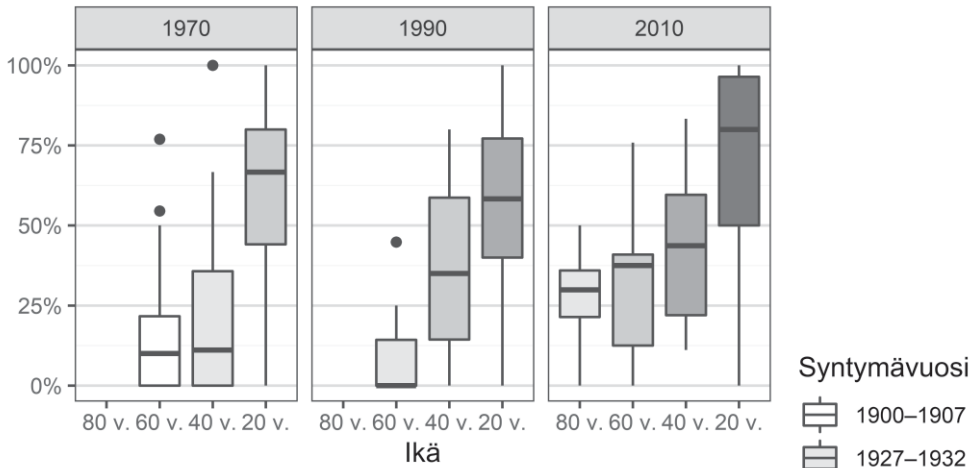
Nominien jälkeen oikeneminen näyttää olevan yleisintä infinitiiveissä (2C) ja vasta tämän jälkeen partitiivitapauksissa. Molemmissa tapauksissa muutos etenee sukupolvittaisen muutoksen mallilla. Partitiiveissa on tilastollisesti merkitsevä muutos niin keski-ikäisten kuin nuorten ryhmissä 1970-luvulta 1990-luvulle ($p < 0,001$ molemmissa ryhmissä). Keski-ikäisillä oienneiden muotojen käyttö kasvoi 40 prosentista ($N = 205/311$) 78 prosenttiin (115/148) ja nuorilla 78 prosentista (513/660) 94 prosenttiin (247/263). Infinitiivien osalta ero on 1970- ja 1990-lukujen osalta tilastollisesti merkitsevä vain keski-ikäisten ryhmässä ($p = 0,02$). Oienneiden muotojen osuus kasvoi tällä välillä 44 prosentista ($N = 104/238$) 87 prosenttiin (55/63). Lisäksi infinitiiveissä on tilastollisesti merkitsevä muutos 1990-luvulta 2010-luvulle vanhojen ryhmässä ($p = 0,02$); määrällisesti kasvua on 38 prosentista (36/94) 84 prosenttiin (65/77).

Sekä kuvio 1 että nämä tilastolliset tulokset paljastavat nopean muutoksen Helsingin puhekielessä: erityisesti 1970- ja 1990-lukujen välillä partitiivi- ja infinitiiviesiintymien oikeneminen on edennyt yhteisössä vauhdilla. Kuitenkin 2010-luvulla muutoksen eteneminen on selvästi hidastunut. 2010-luvun nuoret käyttävät oienneita muotoja hieman vähemmän kuin 1990-luvun nuoret (infinitiiveissä laskua on 96 prosentista 89 prosenttiin ja partitiiveissa 94 prosentista 85 prosenttiin). Lisäksi 1970-luvulla syntyneiden kohortti on vähentänyt oienneiden muotojen käyttöä tilastollisesti merkitsevästi niin infinitiiveissä ($p = 0,02$; 96 prosentista 85 prosenttiin) kuin partitiiveissa ($p = 0,02$; 93 prosentista 78 prosenttiin). Näin ollen 1970- ja 1990-lukujen välillä sukupolvittaisen muutoksen mallin mukaan edennyt muutos olisi 1990- ja 2010-lukujen välillä siirtynyt vetäytyvän muutoksen malliin. Palaan tähän luvussa 5.3.2 muotoillessani muutosmallien ja S-käyrän yhdistelmää.

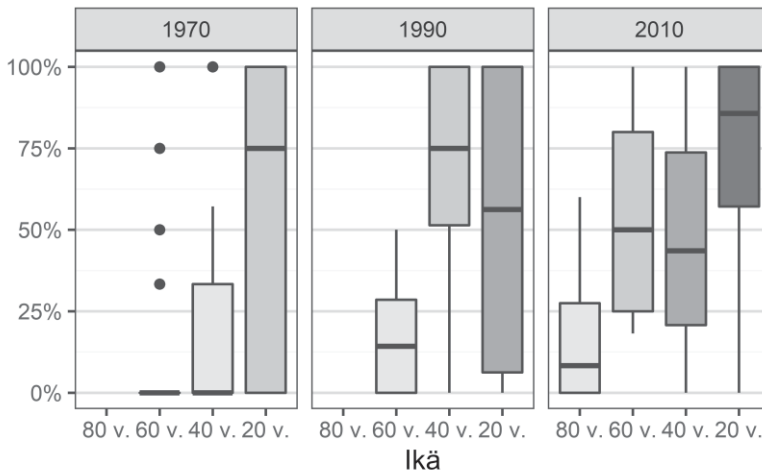
OA- ja *eA*-yhtymien oikenemisen tavallisuus oli odotettua peilaten esimerkiksi Paunosen (1995 [1982]: 106–116) tuloksiin sekä Mielikäisen (1982) ja Mantilan (2004) havaintoihin oikenemisen yleistymisestä. Seuraavassa tarkastelen *UA*-yhtymän oikenemistä partitiiveissa ja infinitiiveissä. Kuviossa 2 on esitetty tässä variantissa tapahtuneet muutokset samoin kuin edellä kuviossa 1.

Kuvio 2. *UA*-yhtymän oikeneminen partitiiveissa (A) ja infinitiiveissä (B) Helsingin puhekielessä.

A *UA*-yhtymän oikeneminen partitiiveissa



B infinitiiveissä



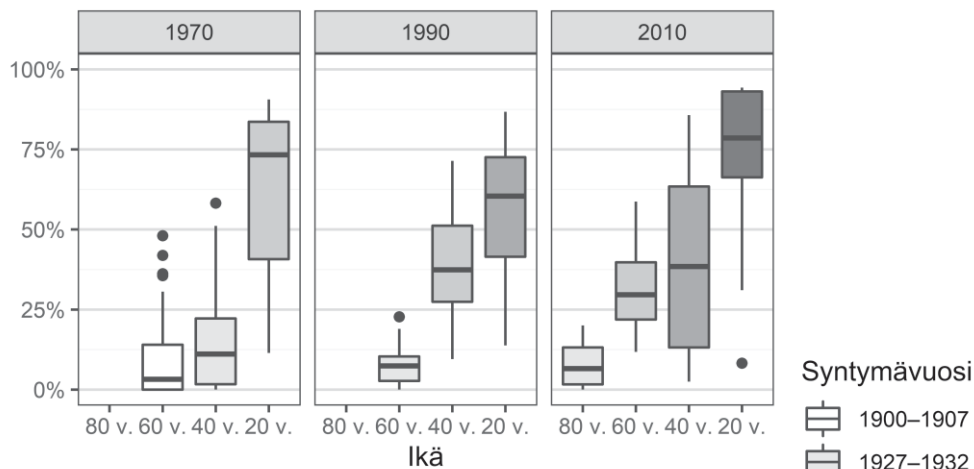
Kuvion 2 perusteella oikeneminen ei ole *UA*-yhtymässä yhtä pitkällä kuin *OA*- ja *eA*-yhtymissä. Toiseksi variaatio puhujien kesken on paljon suurempaa. Tämä johtuu osittain siitä, että esimerkiksi *UA*-infinitiivitapauksia on aineistossa melko vähän (yhteensä 724). Infinitiivien osalta ei voidakaan varmuudella sanoa paljoakaan: näyttää siltä, että oienneet muodot ovat 2010-luvulla yleistyneet, mutta vaihteluväli on huomattavan suuri. Partitiivimuodoissa oienneet muodot ovat enemmistönä nuorten

puheessa jokaisella vuosikymmenellä, mutta muiden ryhmien puheessa vähemmistöissä. Kuitenkin vanhempienkin puhujien osalta oienneiden muotojen osuus kasvaa vuosikymmeneltä toiselle. Yhteisön tasolla muutos partitiivissa on merkitsevää keski-ikäisten ryhmässä 1970- ja 1990-lukujen välillä ($p = 0,03$) ja vanhojen ryhmässä 1990- ja 2010-lukujen välillä ($p = 0,04$). Keski-ikäisillä kasvua on 20 prosentista ($N = 86/441$) 41 prosenttiin (44/108) ja vanhoilla 18 prosentista (24/137) 42 prosenttiin (59/139). Koska yksilötasolla ei samaan aikaan tapahdu tilastollisesti merkitsevää muutosta, muutos on tilastollisesti yhteensopiva sukupolvittaisen muutoksen mallin kanssa. Vain kuviota tarkastelemalla näyttäisi siltä, että *UA*-yhtymässä on huomattavasti suurempi ikäsidonlainen tendenssi kuin *OA*- ja *eA*-yhtymissä: oienneiden muotojen käyttö putoaa kohorttien sisällä nuoruudesta keski-ikään siirryttäessä. Ikäsidonlainen tendenssi sopisi yhteen sen Paunosen (1995 [1982]: 110–111) tekemän havainnon kanssa, että etelähämäläiset variantit (*UU*, *ii*) ovat leimallisempia puhekielisyyskäyttöä kuin yleisemmin kansanmurteissa käytetyt *eA*- ja *OA*-yhtymien oienneet muodot. Keski-ikässä näistä leimaavista piirteistä on osittain luovuttu. Mikäli tämä oletus pitää paikkansa, pitäisi vastaava tendenssi löytyä myös *iA*-yhtymän oikeenemisestä, jonka muutos on esitetty kuviossa 3.

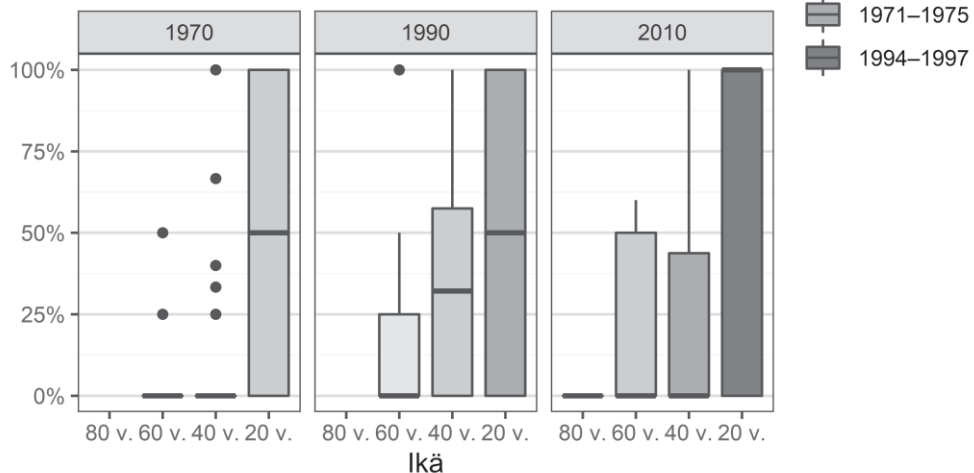
Kuviossa 3 infinitiiviesiintymissä (B) on vielä suurempi hajonta kuin edellisessä kuviossa 2. Infinitiivien osalta muutosta ei siis voi kommentoida, koska esiintymiä on liian vähän. Sen sijaan partitiivimuotoja aineistossa on runsaasti. Tässäkin vokaalilyhtymässä nuoret käyttävät oienneita muotoja suuremmassa määrin kuin kirjakielisiä muotoja. Myös *iA*-yhtymässä on havaittavissa selvä ikäefekti: puhujat vähentävät oienneiden muotojen käyttöä ikääntyessään. Tilastollisesti tämän variantin tulkinta on ongelmallista: viisi 1950-luvulla syntynyttä puhujaa kymmenestä vähentää variantin käyttöä eli vetäytyy nuoruudesta keski-ikään siirryttäessä. Kuitenkin kolme puhujaa kasvattaa osuuttaan ja kaksi pysyy jokseenkin stabiilina. Vaikka pääasiallinen trendi siis onkin vetäytyvä muutos, tälle ei löydy tilastollista tukea, koska yksilöt muuttuvat eri tavoin. Tilastollisten testien mukaan *iA*-yhtymän oikeeneminen partitiivissa noudattaakin *UA*-yhtymän tapaan sukupolvittaisen muutoksen mallia, koska yhteisö muuttuu tilastollisesti merkitsevästi keski-ikäisten osalta 1970-luvun ja 1990-luvun välillä ($p < 0,01$) ja sekä vanhojen että nuorten osalta 1990-luvun ja 2010-luvun välillä ($p < 0,01$ ja $p = 0,04$). Keski-ikäisillä osuus kasvaa 1970-luvun 15 prosentista ($N = 381/2618$) 1990-luvun 39 prosenttiin (269/686), kun taas nuorilla 1990-luvun 58 prosentista (459/790) 2010-luvun 66 prosenttiin (704/1066) ja vanhoilla 7 prosentista (66/991) 34 prosenttiin (250/735).

Kuvio 3. *iA*-yhtymän oikeeneminen partitiiveissa (A) ja infinitiiveissä (B) Helsingin puhekielessä.

A *iA*-yhtymän oikeeneminen partitiiveissa



B infinitiiveissä



Paunosen (1995 [1982]: 111) esittämä oikeenemisjärjestys osoittautuu siis kolmen aikapisteen reaaliaikaturkimuksessakin kestäväksi. Siinä missä oikeeneminen on *eA*- ja *OA*-yhtymissä jo lähes yksinomaista, on se *UA*- ja *iA*-yhtymissä edelleen selvästi varioivampaa. Myös Paunosen perustelut ovat uskottavia: pienemmän levikkialueen variantit (*UU*, *ii*) ovat leimaavampia kuin laajemman levikin piirteet (*ee*, *OO*). Vastavia havaintoja on esitetty variaation tutkimuksen perinteessä runsaasti (esim. Thelander 1979a: 112–116; Mielikäinen 1982: 280–287; Lappalainen 1996: 26–27).

5.1.4 *ts*-yhtymä

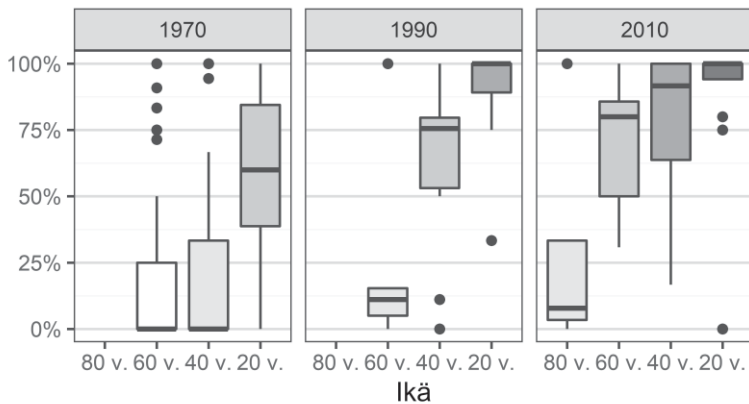
Yleiskielen *ts*-yhtymällä on suomen murteissa useita variantteja, kuten *tt*, *ht* ja *ss* (Rapola 1966: 182–195). Tyypillisesti yhtymän vaihtelua tarkastellaan erikseen vahvassa ja heikossa asteessa, koska esimerkiksi yksinäis-*t* ei esiinny lainkaan vahvassa asteessa. On myös havaittu, että yleiskieliset *ts*-variantit ovat yleisempiä vahvassa kuin heikossa asteessa (Paunonen 1995 [1982]: 62–63; Mustanoja 2011: 185). Vaikka yhtymä esiintyy monissa sanoissa, on sen tarkastelua usein rajattu tiettyihin lekseemeihin (esim. Paunonen mts. 57–58; Mustanoja mts. 176). Näin olen toiminut myös tässä tutkimuksessa, ja olen tarkastellut aineistosta ainoastaan lekseemejä *etsiä*, *itse*, *katsella*, *katsoa*, *seitsemän* ja *viitsiä*. Tyypillisesti näiden lisäksi on tarkasteltu ainakin lekseemiä *metsä*, mutta olen jostakin syystä rajannut sen poimintavaiheessa tutkittavien lekseemien ulkopuolelle. Mikäli tutkimus alkaisi uudelleen nyt, ottaisin lekseemin ehdottomasti tarkasteltavaksi.

Hämäläismurteissa *ts*:n vastineena on tyypillisesti vaihtelematon *tt* (*mettä* : *mettän*) (Rapola 1969: 118). Tämän lisäksi Helsingin puhekielessä esiintyy myös *ts*-esiintymiä sekä heikossa asteessa myös yksinäis-*t*tä. *ts*-esiintymien voidaan olettaa olevan peräisin kirjakielestä, koska variantin murre-edustus keskittyy vain kaakkoismurteisiin. Paunonen (1995 [1982]: 67–68) on huomioinut, että 1970-luvun helsinkiläisnuorilla näyttää vakiintuvan vaihtelu *tt* : *t*. On tosin huomioitava, että juuri *t*-variantti näyttää kytkeytyvän muita variantteja vahvemmin tiettyihin lekseemeihin, erityisesti pronomiiniin *itse* (*ite*) (Nuolijärvi 1986: 123–125). Kuviossa 4 on esitetty Helsingin puhekielessä yleisimpien varianttien edustus vahvassa ja heikossa asteessa.

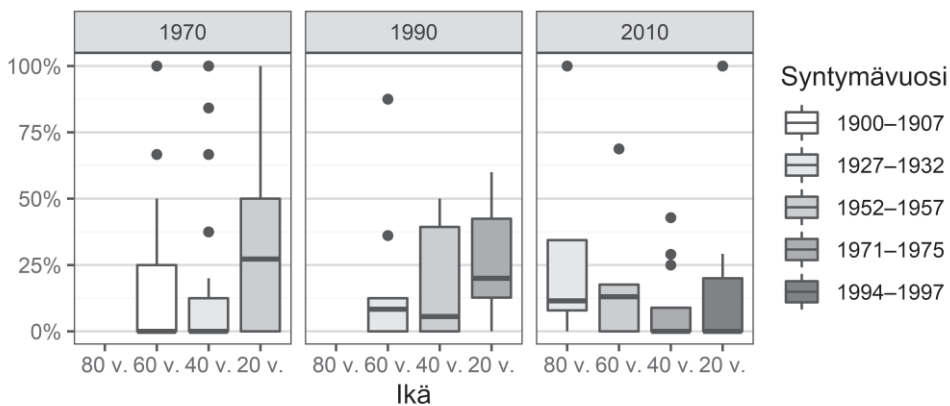
Kuviosta 5A havaitaan selvästi, että *tt* on 2010-luvulla yleisin variantti Helsingin puhekielessä vahvan asteen tapauksissa. Nuorilla puhujilla variantti on 1990- ja 2010-luvulla lähes yksinomainen, ja noin 60-vuotiaiden puhujienkin mediaani on 2010-luvulla yli 75 prosenttia. Nuorten ja keski-ikäisten ryhmissä muutos on tilastollisesti merkitsevä 1970- ja 1990-lukujen välillä ($p < 0,001$) ja vanhojen ryhmässä 1990- ja 2010-lukujen välillä ($p = 0,04$). Nuorten ryhmässä *tt*:n osuus kasvoi 1970-luvulta 1990-luvulle 62 prosentista ($N = 211/339$) 94 prosenttiin (113/120) ja keski-ikäisillä 26 prosentista (64/249) 67 prosenttiin (57/85). Vanhojen ryhmässä osuus kasvoi 1990-lukujen ja 2010-lukujen välillä 28 prosentista (29/103) 73 prosenttiin (81/111). Kuviota 5A tarkastelemalla näyttää siltä, että 1950-luvulla syntyneiden puhujien kohortti ottaa osaa meneillään olevaan muutokseen, jolloin muutos toteuttaisi elinikäisen muutoksen mallia. Koko kohortin tasolla muutos ei kuitenkaan ole tilastollisesti merkitsevä, sillä puhujat muuttuvat paljon ristiin. Tilastollisesti muutos toteuttaakin sukupolvittaisen muutoksen mallia.

Kuvio 4. Vahvassa asteessa esiintyvän *tt*-variantin edustus (A), heikossa asteessa esiintyvän *tt*-variantin edustus (B) ja heikossa asteessa esiintyvän *t*-variantin edustus (C) Helsingin puhekielessä.

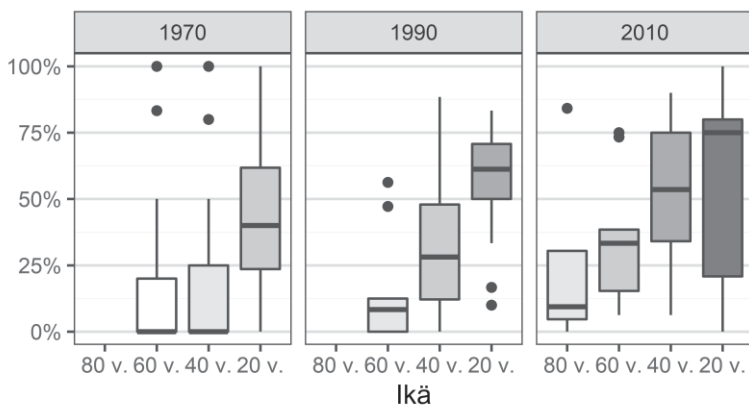
A *tt* ts:n vahvan asteen vastineena



B *tt* ts:n heikon asteen vastineena



C *t* ts:n heikon asteen vastineena



Heikossa asteessa näyttää selvältä, että yksinäis-*t* on 2010-luvulla yleisempi variantti kuin *#*. Muutos yksinäis-*t:n* käytössä on tilastollisesti merkitsevä nuorten ja keski-ikäisten ryhmissä 1970- ja 1990-lukujen välillä ($p = 0,04$ nuorilla ja $p = 0,02$ keski-ikäisillä). Nuorilla variantin osuus kasvoi 1970-luvun 38 prosentista ($N = 113/300$) 58 prosenttiin ($75/130$) ja keski-ikäisillä 18 prosentista ($38/242$) 41 prosenttiin ($45/110$). Sen sijaan variantin *#* käytössä on tilastollisesti merkitsevä ero nuorilla puhujilla 1990-luvulle 2010-luvulle ($p = 0,03$). Kuviosta 5B havaitaan, että tämä johtuu variantin *#* nopeasta vähentymisestä: variantin käyttö on nuorten ryhmässä laskenut 1990-luvun 25 prosentista ($33/130$) 6 prosenttiin ($12/216$). Heikossa asteessa *#* näyttääkin olevan jopa kokonaan häviämässä, kun taas *t*-variantin osuus nousee vuosikymmeneltä toiselle ja muutos etenee sukupolvittaisen muutoksen mallin mukaisesti.

Paunosen (1995 [1982]: 67–68) ounastelu vaihtelun *#*: *t* vakiintumisesta näyttää siis käyvän toteen. Tässä tapauksessa on kuitenkin muistettava, että tarkasteltujen lekseemien määrä on vähäinen ja niissä on saattanut tapahtua leksikaaliselta muutosta (vrt. luku 5.1.2 ja osatutkimus KATOMUODOT). Sen sijaan melko varmasti voidaan sanoa, että hämäläismurteinen vaihtelematon *#* ei ole Helsingin puhekielessä enää 2010-luvulla merkittävässä asemassa.

5.2 Muuttuvat varieteetit

Tässä alaluvussa esittelen osatutkimuksessa LEKTTI saatuja tuloksia Helsingin puhekielen varieteeteista. Tulokset perustuvat LDA-mallin (ks. luku 4.4) avulla aineistosta luotuun jakoon. LDA-malli jakoi aineiston kymmeneen laskennalliseen komponenttiin haastatteluissa yhdessä esiintyvien varianttien perusteella. Komponentteja tarkasteltiin osatutkimuksessa vielä erilaisin kielellisin perustein. Lopulta päädyttiin tilanteeseen, jossa viisi lektiä tai varieteettia olivat sekä laskennallisesti että kielellisesti mielekäs kuvaus aineistosta. Nämä lekit olivat rekisteriytynyt yleiskieli (artikkelissa *enregistered standard*), neutraali yleiskieli (*neutral standard*), hämäläismurre (*Häme dialect*), etelähämäläinen murre (*Southern Häme dialect*) ja valtakunnallinen puhekieli (*supralocal lect*). Näistä kahta ensimmäistä käsitellen alaluvussa 5.2.1, kahta seuraavaa alaluvussa 5.2.2 ja valtakunnallista puhekieltä alaluvussa 5.2.3.

Osatutkimuksessa LEKTTI LDA-mallin tuloksia tarkasteltiin todennäköisyysin $P(\text{Lekti} | \text{Variantti})$ ja $P(\text{Variantti} | \text{Lekti})$. Esitän tässä yhteenveto-osassa koonnin kunkin lekitin tunnusmerkeistä ilman tarkkoja todennäköisyyksiä. Käytän esimerkksanoja kuvaamaan lekitien välisiä eroja. On tärkeä huomata, että lekit on nimetty nimenomaan niiden tunnusmerkkisten piirteiden mukaan. Monissa lekteissä on myös

sellaisia variantteja, jotka eivät välttämättä kyseiseen kielimuotoon perinteisesti kuuluisi (esim. yleiskielen *d:n* kato hämäläismurteessa). Tästä heterogeenisyydestä huolimatta osatutkimuksessa löydettiin kuitenkin selvästi toisistaan poikkeavia lekkejä (vrt. Guy 2013; Gregersen & Pharao 2016).

Rekisteriytynyt yleiskieli: *menemme, minä, tämä, viisi, taloksi*

Neutraali yleiskieli: *kabvia, talosta, mä, tää, viis*

Hämäläismurre: *seittemän, itte, meen pelaan, se*

Etelähämäläismurre: *kabvii, voi sattuu, voi pelaa*

Valtakunnallinen puhekieli: *mennään, mä, korkee, talost, tehä*

5.2.1 Yleiskieli

Osatutkimuksen LEKTTT kymmenestä laskennallisesta komponentista kaikkein voimakkaimmin muista poikkesi rekisteriytyneeksi yleiskieleksi nimetty komponentti (rekisteriytymisestä ks. esim. Agha 2003; Lehtonen 2015: 40–41). Tämä komponentti oli yksi kolmesta komponentista, jotka pidettiin lingvistisen tarkastelun jälkeenkin ennallaan omana lektinään (muut olivat hämäläismurre ja valtakunnallinen puhekieli). Rekisteriytyneen yleiskielen suosimat variantit poikkesivat selvästi muista lekteistä. Esimerkiksi kaikki monikon ensimmäisen persoonan (*menemme*) esiintymät aineistossa kuuluvat tähän lektiin. Monikon ensimmäinen persoona näyttää tutkimuksen valossa siis olevan hyvin vahvasti rekisteriytynyt piirre, koska se kuuluu vain yhteen lektiin. Muita lektiä määrittäviä variantteja ovat pronomini muodot *minä* ja *tämä* sekä *in* säilyminen sananloppuisessa asemassa.

Rekisteriytynyt yleiskieli esiintyy hyvin runsaasti 1970-luvun vanhojen ja keskiikäisten puhujien varieteettina, mutta on kadonnut lähes kokonaan 2010-luvun aineistosta. Ero 1970-luvun vanhimpien puhujien ja aineiston muiden kohorttien välillä ilmenee monessa muussakin yhteydessä. Siinä missä rekisteriytynyt yleiskieli katoaa lähes kokonaan, neutraali yleiskieli on vielä 2010-luvun aineistossakin melko tavallinen. Neutraaliin yleiskieleen kuuluu yleiskielisten varianttien ohella myös hyvin tavallisia puhekielisyyksiä kuten *tää, mä* ja *in* loppuheitto. Lektin yleiskieliset piirteet, kuten *iA*-yhtymän säilyminen (*kabvia*) ja elatiivin pääte (*talosta*), ovat hyvin neutraaleja yleiskielisyyksiä. Osatutkimuksen valossa näyttää siltä, että yleiskielisyys on Helsingin puhekielessä vähentynyt 1970-luvulta 2010-luvulle. Monet neutraaleiksi mielletyt yleiskielisyydet ovat kuitenkin vielä 2010-luvullakin tavallisia, kun taas monet sellaiset hyvin tavalliset puhekielisyydet kuten pronominivariantit *tää* ja *mä* sekä passiivin

käyttö monikon ensimmäisen persoonan sijaan näyttävät 2010-luvulla tulevan osaksi jo kaikkein yleiskielisimpienkin puhujien kieltä. Tällaista epävirallista yleiskieltä on kommentoitu jo 1970-luvun tutkimuksissa (esim. Itkonen 1972: 379–382; Karlsson 1975: 53–64). Sellaista täysin yleiskielistä puhetta, jota esimerkiksi Paunonen käsittelee artikkelissaan (2005b: 169–170; vrt. myös Anhava 2019; Paunonen 2019) ei enää 1970-luvun jälkeen Helsingin puhekielen aineistosta löydy.

5.2.2 Hämläismurteet

Hämläismurteisia komponentteja löytyi aineistosta kaikkiaan kolme kappaletta. Näistä kuitenkin kaksi yhdistettiin etelähämläläiseksi lektiksi, ja yksi säilyi ennallaan sydänhämläläisenä lektinä. Näidenkin lektien välinen ero on osaltaan häilyvä ja perustuu ennen kaikkea muutamaiin keskeisiin variantteihin.

Hämläismurteisen lektin keskeisimpiä variantteja olivat *ts*-yhtymän *tt*-variantit vahvassa ja heikossa asteessa (*seittemän*, *itte*) sekä MA-infinitiivin illatiivin MA-tunnukseton, mutta *n*-päätteinen muoto (*meen pelaan*). Sen sijaan etelähämläläiseen lektiin kuuluivat sellaiset variantit kuin supistumaverbien analoginen A-infinitiivi (*voin pelaa*) ja *iA*- ja *UA*-yhtymien monoftongiutumisen partitiivissa (*kahvii*) ja infiniivissä (*saattaa sattuu*), joista kumpaakaan ei perinteisesti esiinny hämläläisalueen pohjoisosissa (Kettunen 1940: kartta 197; Lappalainen 1996: 25–26).

Hämläismurre näyttää aineiston valossa hiljalleen hiipuvan Helsingin puhekielessä, sillä 2010-luvulla sen variantteja käyttävät kaikkein eniten vanhat puhujat ja hekin melko vähän. Sen sijaan etelähämläläinen murre näyttää toteuttavan vetäytyvän muutoksen muutosmallia (ks. luku 2.4; Wagner & Sankoff 2011): käyttö kasvaa vuosikymmeneltä toiselle, vaikka puhujat vähentävätkin lektin käyttöä aikuistuuksaan. On kuitenkin huomattava, että 2010-luvun nuoret käyttävät lektiä huomattavasti vähemmän kuin vetäytyvän muutoksen mallin perusteella sopisi odottaa. Tämä johtuu pitkälti valtakunnallisen puhekielen vaikutuksesta (ks. luku 5.2.3).

Helsingissä puhuttavan suomen perustan muodostavat hämläismurteet ovat siis edelleen edustettuina kaupungin puhekielessä. Kuitenkin sydänhämläläiset piirteet, kuten vaihtelematon *tt ts:n* vastineena sekä tunnuseton illatiivi ovat väistymässä uusien varianttien tieltä. Tässä tapauksessa alaa valtaavat *ts:n* vastineena vaihtelu *tt : t* (*mettä : metässä*) ja illatiivin tapauksessa lyhyt infiniivi (*meen pelaa*).

5.2.3 Valtakunnallinen puhekieli

Viimeisenä tarkastelen valtakunnallista puhekieltä. Tämän lektin nimeäminen sisältää paljon enemmän problematiikkaa kuin edellisten, joiden nimet kumpusivat joko paikallisuudesta (hämäläismurteet) tai standardista (yleiskieli). Englanninkielisessä osatutkimuksessa lektiä kutsuttiin supralokaaliksi eli ylipaikalliseksi. Kumpikaan näistä nimistä ei kuitenkaan ole suomeksi erityisen tyydyttävä, sillä ne ovat etäännyttävän vierasperäisiä. Käytettyä termistöä käsittelemme luvussa 3.2.

Puhuttiin sitten valtakunnallisesta puhekielestä, yleispuhekielestä tai ylipaikallisesta puhekielestä, tyypillisesti viitataan sellaisiin variantteihin, jotka ovat yleisiä tai yleistymässä eri puolilla Suomea. Esimerkiksi Mantila (2004: 325–328) erotelee variantteja niiden neutraaliuden ja yleisyyden mukaan. Esimerkiksi passiivin käyttö monikon ensimmäisen persoonan sijaan sekä pronominit *mä* ja *sä* ovat Mantilan listauksen mukaan yleisiä ja neutraaleja puhekielisyyksiä. Molemmat variantit kuuluvat myös osatutkimuksen LEKTIT valtakunnalliseen puhekieleen. Mantilan kolmannen luokan muodostavat yleistyvät puhekielen piirteet, joista hän luettelee esimerkiksi A-loppuisten vokaaliyhtymien oikenemisen (*korkee*), A:n loppuheiton (*talosť*) ja kadon yleiskielen *d*:n vastineena. Kaikki nämä piirteet esiintyvät myös osatutkimuksen LEKTIT valtakunnallisessa puhekielessä.

Monet näistä piirteistä ovat jo leimallisesti osa Helsingin puhekieltä. Joskus niiden yleistymistä muualla Suomessa onkin tarkasteltu juuri Helsingistä lähteneinä piirteinä. Variantit ovat kuitenkin lähtöisin monesta eri suunnasta (esim. *d*:n katomuodot itämurteista ja A-loppuisten vokaaliyhtymien oienneet muodot sekä hämäläis- että savolaismurteista). Niiden on lisäksi osoitettu levinneen jo maaseutumurteiden aikana (Mantila 2004: 328). Lisääntynyt liikkuvuus ja sosiaalisen median entistä suurempi vaikutusvalta saattavat sen sijaan voimakkaasti vauhdittaa niiden yleistymistä tulevaisuudessa. Osatutkimuksessa valtakunnallinen puhekieli oli 2010-luvun nuorten keskuudessa lekteistä kaikkein yleisin. Niin valtakunnallisen puhekielen kuin neutraalin yleiskielenkin yleistymisen kertoo osaltaan Helsingin puhekielen tasoittumisesta (vrt. Milroy 2002): vahvasti leimaavia piirteitä käytetään 2010-luvulla entistä vähemmän. Vastaava tendenssi on yleinen myös maaseutumurteissa (Lappalainen 1996; Nahkola & Saanilahti 2001; Kurki 2005) ja erityisesti nuorten keskuudessa (Lappalainen 2001: 77).

Yleispuhekieli-termin käyttöä on kartettu muiden syiden ohella myös siksi, että koko varieteetin olemassaoloa on epäilty (esim. Mantila 1997: 6–14; Hurttu 1999: 56). Osatutkimuksessa LEKTIT havaittiin kuitenkin tilastollisesti, että juuri ne piirteet, joita Mantilakin pitää yleisinä tai yleistyvänä (esim. 2004: 325–328) näyttävät

esiintyvän yhdessä eli muodostavan piirrekimpun. Näin ollen valtakunnallinen puhekieli (tai yleispuhekieli) näyttää siis perustuvan kielelliseen todellisuuteen. Tämä on sinänsä luonnollinen päätelmä: jos tietyt variantit yleistyvät samaan aikaan samassa paikassa, yleistyvät tietenkin myös niiden esiintyminen yhdessä.

5.3 Muutosmallit

Tässä aluvuossa esittelen osatutkimuksessa MUUTOSMALLIT saatuja tuloksia sekä niiden vaikutusta sosiolingvistiseen teoriaan ja aineistoihin. Lisäksi pohdin lyhyesti muutosmallien ja S-käyrän yhteensovittamista, sillä vaikka molemmat edustavat sosiolingvistiikassa perinteisiä muutoksen kuvauksen tapoja, niitä on harvoin tarkasteltu kokonaisuutena. Tässä aluvuossa ei ole tarkoitus käydä tarkasti läpi osatutkimuksessa MUUTOSMALLIT esimerkkeinä esiteltyjä variantteja, vaan tarkastella muutosmallien teoriaa saatujen tulosten pohjalta. Aluvuossa 5.3.1 tarkastelen muutosmallien kykyä kuvata muutosta käytännössä. Aluvuossa 5.3.2 keskityn muutosmallien teorian ja S-käyrän yhteensovittamiseen.

5.3.1 Muutosmallit käytännössä

Muutosmalleja on esitelty tarkemmin itse osatutkimuksessa sekä tämän yhteenveto-osan luvussa 2.4. Tässä aluvuossa tarkastelen niitä käytännön kannalta: mitkä muutosmalleista näyttävät Helsingin puhekielen aineiston valossa käyttökelpoisilta ja mitkä taas jäävät lähinnä teoreettisiksi? Labovin (1994: 83–84) muutosmalleista ensimmäinen on stabiili tilanne. Stabiilissa tilanteessa ei ole sanan varsinaisessa merkityksessä kyse muutoksesta lainkaan, sillä sen paremmin yksilö- kuin yhteisötasollaakaan ei tapahdu muutosta. Stabiilia tilannetta on tyypillisesti pidetty melko teoreettisena (esim. Paunonen 2005a: 15; Kurki 2013: 206). Myös osatutkimuksessa MUUTOSMALLIT stabiili tilanne oli harvinainen. Ainoat tilanteet, joissa stabiili muutosmalli näytti jokseenkin uskottavalta, olivat sellaisia, joissa jokin variantti oli jo lähes kokonaan kadonnut (esim. *r* *d*:n vastineena) tai jokin variantti oli muuttunut lähes yksinomaiseksi (esim. *z*:n loppuheitto konditionaalissa). On kuitenkin muistettava, että variaation tutkimus keskittyy lähtökohtaisesti variaation löytämiseen; niinpä valitut piirteet ovat oletettavasti usein sellaisia, joissa variaatiota esiintyy. Sellainen tilanne, jossa jostakin piirteestä on käytössä useampia variantteja mutta näiden välinen variaatio-suhde ei ajan kuluessa millään tavalla muutu, ei vaikuta uskottavalta.

Myös ikäsidonnaista muutosta on pidetty teoreettisena (Chambers 2009 [1995]: 201). Osittain tämä johtuu siitä, että ikäsidonnaisen muutoksen määrittelyssä on ollut jonkin verran vaihtelua. Mikäli ikäsidonnainen muutos ymmärretään Labovin (1994: 84) kuvaamana syklisenä muutoksena, jossa tietyt ikäryhmät poikkeavat vuosikymmen vuosikymmeneltä muista samalla tavalla, muutosmalli on eittämättä harvinaisen. Helsingin puhekielen aineistosta löydettiin vain kaksi varianttia, joissa ikäsidonnaisen muutoksen malli näyttää toteutuvan. Muissa tutkimuksissa on kuitenkin havaittu ikäsidonnaista tendenssiä (esim. Lappalainen 2010: 287–288; Wagner 2012b). Toisaalta ikäsidonnaisuus tendenssinä ja muutosmallina saattaa olla syytä erottaa toisistaan: monissa tapauksissa näyttää todennäköiseltä, että muutoksiin liittyy ikäsidonnainen elementti. Esimerkiksi vetäytyvässä muutoksessa ja kaarevassa muutoksessa on mukana ikäefekti. Sen sijaan varsinainen syklinen ikäsidonnainen muutos, jossa yhteisö ei muutu, on huomattavasti harvinaisempi kielellinen tilanne. Wagner (2012a: 378) nostaa esiin sen tärkeän huomion, että ikäsidonnaista muutosta on pidetty toistuvana muutoskaavana, vaikka muutoksia ei ole oikeastaan tutkittu kuin kahdesta aikapisteestä käsin. Miten kahden aikapisteen välillä havaittu muutos voitaisiin olettaa toistuvaksi?

Sukupolvittaista muutosta on pidetty variaationtutkimuksessa pitkään pääasiallisena muutoksen tapana (esim. Labov 1994: 84), ja sitä se on myös Helsingin puhekielessä. Sukupolvittainen muutos toimii perustana näennäisaikaennusteille, mikä on osaltaan johtanut sen tulkitsemiseen normaaliksi muutoksen tavaksi. Helsingin puhekielen aineistossa sukupolvittainen muutos esiintyi erityisesti sellaisissa tapauksissa, joissa jokin variantti oli muuttumassa yksinomaiseksi tai katoamassa kokonaan käytöstä, eli niin sanotuissa loppuvaiheen muutoksissa. Sukupolvittaisen muutoksen kanssa tilastollisesti yhteensopivia muutoksia esiintyi myös huomattavasti varioivammissa tilanteissa. Näissä keskivaiheen muutoksissa muutos ei kuitenkaan jatkunut sukupolvittaisen muutoksen mallilla yli kolmen aikapisteen (esim. *OA-* ja *eA-*yhtymät luvussa 5.1.3).

Viimeinen Labovin alkuperäisistä muutosmalleista on yhteisöllinen muutos. Yhteisöllisessä muutoksessa sekä yksilöt että yhteisö muuttuvat. Labovin (1994: 84) tulokinnan mukaan kyseessä on yleensä uuden sanan ilmestyminen yhteisöön. Helsingin puhekielen aineistosta yhteisöllisen muutoksen kanssa yhteensopivia muutoksia löytyi vain yksi, *t:n* kato partisiipeissa (*hyppim*). Tässäkin piirteessä muutos tapahtuu vain kahden aikapisteen välillä (1970-luvulta 1990-luvulle). Sellainen tilanne, jossa yhteisöllinen muutos tapahtuisi kolmen aikapisteen yli, vaikuttaa epätodennäköiseltä. Mikäli yhteisöllistä muutosta tulkitaan hieman vapaammin, voi sitä soveltaa paremmin myös käytäntöön. Yksi yhteisöllisen kaltainen muutos Helsingin puhekielen

aineistossa on esimerkiksi tunnuksettomien illatiivien (*meen pelaan*) yhtäkkinen harvinaistuminen, kun puhujat alkavat käyttää runsaammin lyhyttä infinitiiviä (*meen pelaa*). Tätä on tarkasteltu osatutkimuksessa INFINITIIVIT. Vastaavasti *tt:n* käyttö *ts:n* vastineena heikossa asteessa vähenee nopeasti 1990-luvulta 2010-luvulle (luku 5.1.4).

Labovin neljän muutosmallin lisäksi Gillian Sankoff kollegoineen on kehittänyt vielä kaksi uutta muutosmallia: elinikäisen ja vetäytyvän muutoksen mallit. Molemmilla malleilla tuodaan Labovin malleja voimakkaammin esiin idiolektitason muutosta. Elinikäisen muutoksen malli muistuttaa muuten sukupolvittaista muutosta, mutta osa vanhemmista puhujista ikään kuin siirtyy meneillään olevan muutoksen mukaan (Sankoff 2005: 1011). Näin ollen elinikäisen muutoksen mallissa yhteisö muuttuu nopeammin kuin sukupolvittaisen muutoksen mallissa. Vetäytyvän muutoksen mallissa tilanne on päinvastainen (Wagner & Sankoff 2011): puhujat vähentävät joidenkin varianttien käyttöä ikääntyessään (tyypillisesti nuoruudesta keski-ikään). Vetäytyvä muutos on siis käytännössä ikäsidonnan muutos ja sukupolvittaisen muutoksen yhdistelmä.

Kolmen aikapisteen yli kestävästä elinikäisestä muutosta ei havaittu tilastollisesti Helsingin puhekielen aineistossa yhdessäkään variantissa. Useimmiten syynä mallin sopimattomuudelle oli se, että osa vanhemmista puhujista kyllä muuttui meneillään olevan muutoksen suuntaan, kun taas osa sitä vastaan. Sankoff ei varsinaisesti määrittele, kuinka monen puhujan täytyy muuttua muutoksen suuntaan, jotta voitaisiin puhua elinikäisestä muutoksesta. Mallin soveltaminen käytäntöön vaatisikin sen tarkempaa määrittelyä. Elinikäisen muutoksen kaltaista tendenssiä esiintyi kuitenkin joissakin piirteissä myös tässä aineistossa (esim. luvussa 5.1.3 käsitelty *eA-* ja *OA-* yhtymien oikeneminen nomineissa sekä luvussa 5.1.4 käsitelty *tt ts:n* vahvan asteen vastineena). Muutosmalli näyttää sopivan nopean keskivaiheen muutosten kuvaamiseen. Pohdin tätä tarkemmin seuraavassa alaluvussa 5.3.2.

Vetäytyvän muutoksen mallille sen sijaan löytyi tukea myös Helsingin puhekielen aineistosta. Vetäytyvän muutoksen malli näytti toteutuvan elatiivin loppuheitossa: nuoret olivat muita ikäryhmiä edellä, mutta keski-ikäistyessään he vähensivät variantin käyttöä. He eivät kuitenkaan kokonaan luopuneet variantin käytöstä, mikä johti ennen pitkää yhteisön muutokseen. Lisäksi näytti siltä, että keski-ikästä eläkeikään siirtyessään puhujat jälleen kasvattivat variantin käyttöä. Näin ollen muutoksen kuva noudattaisikin *u:n* muotoista käyrää eli kyseessä olisi kaareva muutos (Downes 1998: 224; Tagliamonte 2012: 47). Tällaista mallia on pidetty tavallisena sellaisissa tilanteissa, joissa ei ole meneillään muutosta. Esimerkiksi elatiivin loppuheiton tapauksessa näyttää kuitenkin siltä, että myös yhteisö muuttuu, vaikkakin hitaasti.

Osatutkimuksessa MUUTOSMALLIT ei tarkasteltu tilastollisesti kaarevan muutoksen mallia, joten aineistossa tällä mallilla mahdollisesti tapahtuvia muutoksia ei voitu todentaa. Toisaalta vetäytyvän muutoksen ja kaarevan muutoksen erottaminen saattaa olla oikeastaan tarpeetonta: molempien mallien peruseriaatteena on puhujien yleiskielistyminen keski-iässä. Molemmissa malleissa on oikeastaan kyse siitä, että sukupolvittainen muutos ja ikäsidonlainen muutos kohtaavat: jotakin varianttia käytetään huomattavasti enemmän nuoruudessa ja siitä irtaannutaan keski-iässä esimerkiksi työelämän tuomien kielivaatimusten johdosta (ns. kielellisen markkinan periaate; Bourdieu & Boltanski 1975; Sankoff & Laberge 1978). Erotuksena on oikeastaan vain se, että kaarevan muutoksen mallissa puhujat omaksuvat variantin uudelleen eläköityessään.

Osittain tästä huomiosta johtuen osatutkimuksessa esitettiin, että muutosmalleja voisi yksinkertaistaa tarkastelemalla niitä ikäefektin avulla: ikäsidonlainen muutos olisi näin ollen stabiili tilanne ikäefektillä ja vetäytyvä muutos puolestaan sukupolvittaisen muutoksen malli ikäefektillä. Tällainen ajattelutapa jättäisi tilaa myös erilaisille muutoksen tavoille melko jyrkästi määriteltyjen muutosmallien lisäksi. Tähän viittaa jo Labov alkuperäisessä muutosmallien esittelyssään: useimmiten muutokset ovat mallien yhdistelmiä (Labov 1994: 85).

Muutosmallien tarkastelu kolmen aikapisteen reaaliaika-aineistossa (kuten Helsingin puhekielen pitkittäiskorpuksessa) osoitti, että ne harvoin toimivat yli kahden aikapisteen. Sellaisetkin muutokset, jotka kahden aikapisteen välillä näyttivät selvästi noudattavan jotakin muutosmallia, saattoivat olla muuttuneet jo täysin toisin kolmannessa aikapisteessä. Erityisen selvästi tämä näkyi ikäsidonlainen muutoksen yhteydessä: samojen ikäryhmittäisten erojen toistuminen samanlaisina vuosikymmenestä toiseen yhteisön kuitenkin muuttumatta vaikuttaa melko epätodennäköiseltä, ja tällaisen mallin ennustaminen kahdesta aikapisteestä käsin on käytännössä mahdotonta. Niinpä muutosmalleihin kannattaa suhtautua lähinnä jo tapahtuneiden muutosten kuvaajina, eikä tulevien muutosten ennustajina.

5.3.2 Muutosmallit ja S-käyrä

Osatutkimuksessa MUUTOSMALLIT pohdittiin lyhyesti muutosmallien yhdistämistä S-käyrän kanssa. Osatutkimuksessa tämän aiheen käsittelylle ei kuitenkaan ollut tilaa. Tässä esitettävä yhdistelmä ei siis perustu osatutkimukseen, vaan on sen tuloksista ja kirjallisuudesta johtamaani päättelyä.

Mikäli muutosmallit ja S-käyrä halutaan yhdistää, on hyväksyttävä joko se, että muutokset käyvät läpi useita muutosmallivaiheita tai toisaalta luopua S-käyrään kuuluvasta muutoksen nopeuden vaihtelusta. Näistä nähdäkseni ensimmäinen oletus on mielekkäämpi. On luonnollista, että yksittäinen kielenmuutos ei etene tasaisella vauhdilla, vaan monissa tämänkin tutkimuksen piirteissä on havaittu joko sukupolvittaisen muutoksen mallia nopeampaa tai hitaampaa muutosta. Mikään muutosmallien teoriassa ei myöskään velvoita tutkijaa sitoutumaan yhteen ainoaan malliin muutoksen selittäjänä, vaikka tähän onkin ajoittain pyritty (esim. Paunonen 2005a: 48). Tämä vertautuu myös Kurjen (2005: 239) esittämään havaintoon siitä, että tietyn muutoksen diffuusiohierarkia saattaa olla erilainen muutoksen alku- tai loppuvaiheessa. Seuraavassa pyrin luomaan mallien välille synteetin tästä lähtökohdasta.

Kuten stabiilin tilanteen käsittelyssä esitin, se on tavallinen lähinnä sellaisissa tilanteissa, joissa muutos on jo tapahtunut: jokin variantti on joko kadonnut tai yksinomaisen. Niinpä stabiili tilanne voidaan asettaa molempiin päihin S-käyrää: ennen muutoksen alkamista ja muutoksen jo tapahduttua. Esimerkiksi viimeisestä käy osatutkimuksessa MUUTOSMALLIT esitetty *z*:n loppuheitto konditionaalissa, joka on aineistossa 1990-luvulta lähtien jo käytännössä yksinomaisen.

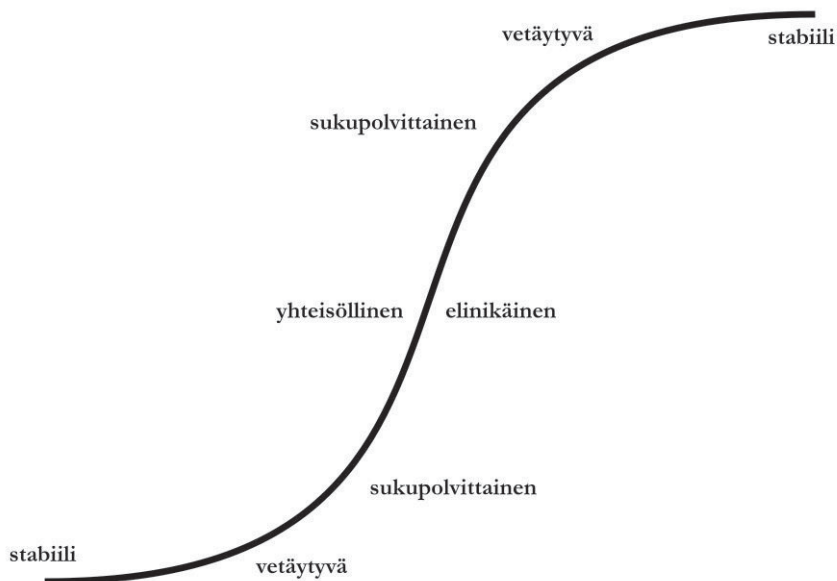
Kun muutos sitten alkaa tapahtua, on vauhti ensin hidas. Hitaassa alkuvaiheessa kyseeseen tulee lähinnä vetäytyvä muutos, joka johtaa kuitenkin yhteisön muutokseen (toisin kuin ikäsidonnainen muutos), mutta hitaammin kuin sukupolvittainen muutos. Esimerkiksi tällaisesta alkuvaiheen vetäytyvästä muutoksesta käy elatiivin loppuheitto, jota käsiteltiin osatutkimuksessa MUUTOSMALLIT. Koska sukupolvittainen muutos on nopeampi kuin vetäytyvä muutos, asettuu se seuraavaan vaiheeseen S-käyrää, ennen kaikkein nopeinta keskivaihetta. Alkuvaiheen sukupolvittaisesta muutoksesta esimerkiksi sopii MUUTOSMALLEISSA ja tämän johdanto-osan luvussa 5.1.4 esitelty yksinäis-*z*:n eteneminen *z*:n heikon asteen vastineena. Vielä 1970-luvulla variantin käyttö on muiden kuin nuorten ryhmässä vähäistä. Seuraavilla vuosikymmenillä sukupolvittain etenevä muutos kuitenkin muuttaa koko yhteisöä.

Kaikkein nopeimmillaan muutos on yhteisöllisen ja elinikäisen muutoksen mallissa. Elinikäisessä muutoksessa tapahtuu sukupolvittaisen muutoksen lisäksi myös vanhempien puhujien siirtymistä uuden variantin käyttäjiksi. Esimerkiksi *ea*- ja *oa*-nominien oikenemisessä 1930-luvun taitteessa syntyneiden kohortti ottaa osaa muutokseen, jolloin muutosvauhti on sukupolvittaisen muutoksen mallia nopeampi (ks. luku 5.1.3). Kuvaukseltaan tilanne on siis hyvin samanlainen kuin S-käyrän keskivaiheilla: suuri osa puhujista siirtyy nopeasti variantin käyttäjäksi (vrt. Kurki 2005: 240). Vastaava tilanne koskee yhteisöllistä muutosta, joka kuvaa esimerkiksi uusien sanojen ilmestymistä leksikkoon. Yhteisöllisen muutoksen mallin kanssa yhteensopiva

muutos osatutkimuksessa MUUTOSMALLIT (*zn* kato partisiipeissa) ei kuitenkaan edusta tällaista keskivaiheen muutosta, joten tämä oletus jää teoreettiseksi.

Nopean vaiheen jälkeen muutosmallit alkavat toimia jälleen käänteisessä järjestyksessä. Sukupolvittainen muutos näytti osatutkimuksessa MUUTOSMALLIT sopivan parhaiten juuri sellaisiin tilanteisiin, joissa jokin variantti oli muuttumassa yksinomaiseksi (esim. *i*-loppuisen diftongin jälkikomponentin kato imperfektissä). Tämän jälkeen vetäytyvän muutoksen voisi ajatella toimivan esimerkiksi niin, että nuoret puhujat käyttävät jotakin varianttia yksinomaisesti, mutta keski-ikässä puhujat siirtyvätkin osittain vielä toisen, harvinaisemman variantin käyttäjäksi. Tässä tutkimuksessa esimerkiksi *OA*- ja *eA*-yhtymien oikeneminen partitiiveissa ja infinitiiveissä näytti 1990- ja 2010-lukujen välillä etenevän juuri näin (ks. luku 5.1.3). Myös Wagner (2012a: 379) ja Buchstaller (2015: 483) ovat löytäneet jonkinlaisia viitteitä tämäntyyppisen muutostendenssin olemassaolosta. Tätä vaihetta seuraa jälleen stabiili tilanne, kun variaatiota ei enää esiinny. Vaikka muutosmallit on tässä kuvauksessa esitetty ikään kuin peräkkäisinä, on täysin mahdollista, että muutokset eivät käy kaikkia muutosmallivaiheita läpi. Muutosmallien sopimisesta esimerkiksi tietynlaisten piirteiden muutoksen kuvaamiseen olisikin syytä tehdä lisätutkimusta. Edellä kuvattu S-käyrän ja muutosmallien yhdistelmä on esitetty kuviossa 5.

Kuvio 5. Muutosmallien ja S-käyrän yhdistelmä.



On huomattava, että ikäsidonnainen muutos ei tässä kuvauksessa kuulu mihinkään kohtaan S-käyrää. Tämä johtuu siitä, että syklisesti tulkittuna sen ei pitäisi johtaa minkäänlaiseen muutokseen eli se asettuisi samaan paikkaan kuin stabiilikin tilanne. Kenties voitaisiin ajatella niin, että stabiilin tilanteen jälkeen siirrytään ikäsidonnaiseen vaiheeseen, joka edeltää varsinaista yhteisön muutosta. Tällainen kuvaustapa ei kuitenkaan ole yhteensopiva S-käyrän kanssa: tässä tapauksessa innovaattorit siis kehittäisivät muutoksen, mutta luopuisivat siitä. Tätä jatkuisi jonkin aikaa, kunnes uudet omaksujat jostain syystä säilyttäisivätkin osittain uuden variantin käytön ja siirryttäisiin vetäytyvän muutoksen malliin. Syklisen ikäsidonnaisen muutoksen S-käyrälle asettamisen vaikeus on toisaalta yksi osoitus siitä, että se ei välttämättä ole muutosmallina erityisen uskottava. Kuviosta puuttuu myös kaarevan muutoksen malli, koska sitä ei tarkasteltu osatutkimuksessa MUUTOSMALLIT tilastollisesti.

6 KOONTI

Tässä tutkimuksen päättävässä luvussa kokoan väitöskirjan tulokset yhteen, arvioin tutkimuksen onnistumista ja suuntaan katseen tulevaan. Vastaan tutkimuskysymyksiin luvussa 6.1 ja kokoan väitöskirjan keskeiset päätelmät toisaalta Helsingin puhekielessä tapahtuneen muutoksen ja toisaalta muutosta kuvaavien teorioiden osalta alaluvussa 6.2. Viimeisessä alaluvussa 6.3 arvioin tutkimusta ja pohdin mahdollisia jatkotutkimuksen paikkoja.

6.1 Vastaukset tutkimuskysymyksiin

Tässä alaluvussa vastaan lyhyesti luvussa 1.2 esitettyihin tutkimuskysymyksiin. Kysymysten luokittelu ja järjestys on sama kuin kyseisessä luvussa: ensin tarkastelen muutosta piirretasolla (6.2.1), sen jälkeen yhteisötasolla (6.2.2), yksilötasolla (6.2.3) ja lopuksi keskityn muutoksen teorioihin (6.2.4).

6.1.1 Muutos piirretasolla

Millaista variaatiota on havaittavissa suomen yleisimmissä infinitiivimuodoissa, MA-infinitiivin illatiivissa ja A-infinitiivin perusmuodossa? Miten niiden käyttö muuttuu 1970-luvulta 2010-luvulle?

Sekä MA-infinitiivin illatiivin että A-infinitiivin perusmuodon käyttö on vähentynyt 1970-luvulta 2010-luvulle, koska näiden yhteenlankeamisesta syntynyt lyhyt infinitiivi yleistyy. Lyhyen infinitiivin yleistyminen näyttää kuitenkin keskittyvän voimakkaammin illatiivitapauksiin, minkä takia MA-infinitiivitapausten määrä vähenee nopeammin. Variaation osalta näyttää selvältä, että erityisesti MA-infinitiivin illatiivin ja lyhyen infinitiivin välillä on ikäsidonlainen ero: nuoret käyttävät huomattavan paljon lyhyttä infinitiiviä, kun taas vanhemmat puhujat suosivat illatiivimuotoja.

Miten yhteenlankeavat infinitiivimuodot vaikuttavat koko infinitiivijärjestelmään? Onko uusi lyhyt infinitiivi erkaantumassa lähtöinfinitiiveistään?

Infinitiivimuotojen yhteenlankeamisesta syntynyt lyhyt infinitiivi vaikuttaa koko infinitiivijärjestelmään erityisesti analogisena mallina. Vaikka lyhyen infinitiivin taustalla on useita äänne muutoksia, on oletettavaa, että sen leviäminen kytkeytyy nykyään ennen kaikkea sen mallivaikutukseen: koska lyhyttä infinitiiviä käytetään jo monissa muotoryhmissä ja sekä A-infinitiivin että MA-infinitiivin tehtävissä, sitä voidaan sujuvasti käyttää yhä uusien verbien kanssa. Lyhyen infinitiivin erkaantuminen lähtöinfinitiiveistä ei vielä näytä tutkimuksen valossa todennäköiseltä, sillä se toteuttaa kahta eri sandhikaavaa. On lisäksi muistettava, että muutamat infinitiiviryhmät eivät juurikaan edustu lyhyinä infinitiiveinä (*dA*-verbit ja konsonanttivartaloiset verbit).

Missä lekseemeissä kato esiintyy yleiskielen *hd*-yhtymässä? Eteneekö kato ajan kuluessa uusiin lekseemeihin?

Kato esiintyy aineistossa yhteensä 44 lekseemissä 191:stä *hd*-lekseemistä. Kato on keskittynyt kaikkein yleisimpiin lekseemeihin, kuten *kabde-*, *kabdeksan*, *läbde-*, *mahdollinen*, *tehdä*, *yhde-*, *yhdeksän*. Kato näyttää kuitenkin etenevän uusiin lekseemeihin ajan kuluessa: 2010-luvulla kato esiintyy jo 31:ssä eri lekseemissä, kun se huomattavasti suuremmassa 1970-luvun aineistossa esiintyi 24:ssä eri lekseemissä. Näyttää myös siltä, että lekseemien morfologiset ja semanttiset ominaisuudet vaikuttavat kadon todennäköisyyteen (vrt. Kurki 2005: 117–124), mutta tätä seikkaa ei tässä tutkimuksessa ehditty tarkastella lähemmin.

Mitkä piirteet yleistyvät ja mitkä vähenevät Helsingin puhekielessä kolmen aikapisteen välillä?

Helsingin puhekielessä yleistyvät monet muuallakin Suomessa yleiset tai yleistyvät variantit, kuten A-loppuisten vokaaliyhtymien oikeeneminen erityisesti *eA*- ja *OA*-yhtymissä, *tt* : *t* -vaihtelu yleiskielen *tn*:n vastineena ja kato yleiskielen *d*:n vastineena erityisesti *b*:n jäljessä. Lisäksi monet vielä 1970-luvulla vaihtelunalaiset piirteet ovat muuttuneet käytännössä yksinomaisiksi, kuten passiivin käyttö monikon ensimmäisen persoonan sijaan, *in*:n loppuheitto konditionaalissa, numeraaleissa ja translatiivissa, *i*-loppuisen diftongin jälkikomponentin kato imperfektissä ja *inen*-adjektiiveissa sekä sananloppuisen *tn* kato partisiipeissa (*hyppiny*). Vastaavasti katoavia piirteitä ovat jo 1970-luvun aineistossa harvinaiset piirteet kuten *r* yleiskielen *d*:n vastineena. Vielä voimakkaammin vaihtelunalaisia, mutta kuitenkin väheneviä piirteitä ovat MA-

infinitiivin illatiivimuodot (*meen tekemään / tekeen*), *tt ts:n* heikon asteen vastineena ja pronominitapaukset *minä, sinä, tämä* ja *nämä* (ei kuitenkaan *hän* ja *he*, vrt. Priiki 2017).

6.1.2 Muutos yhteisön tasolla

Millaisia lekkejä Helsingin puhekielessä esiintyy? Mitkä variantit esiintyvät yhdessä lek-teissä?

Helsingin puhekielestä löytyi seuraavia lekkejä tilastollisen analyysin ja kielitieteellisen tarkastelun perusteella: rekisteriytynyt yleiskieli, neutraali yleiskieli, hämäläismurre, etelähämäläismurre ja valtakunnallinen puhekieli. On muistettava, että nämä nimet perustuvat toisaalta niihin seikkoihin, mitä tiedämme kansanmurteista, mutta toisaalta myös työryhmämme arvioihin esimerkiksi varianttien leimaavuudesta. Niinpä lektien nimet saattavat luoda vaikutelman selvärajaisemmista piirrejoukoista kuin mistä todellisuudessa on kyse. Yleiskielisten lektien varianteista valtaosa on luonnollisesti yleiskielisiä, kun taas puhekielisten lektien variantit painottuvat puhekielisiin vaihtoehtoihin. Puhekielisiä lekkejä eroteltiin toisaalta varianttien historiallisen puhuma-alueen perusteella (hämäläismurre ja etelähämäläismurre) ja toisaalta varianttien laajan levikin ja yleistymistendenssin perusteella (valtakunnallinen puhekieli). Yleiskielisten lektien välillä ero syntyi lopulta helposti: rekisteriytynyt yleiskieli sisälsi huomattavan paljon sellaisia leimallisia yleiskielisyyksiä, joita vain harvat puhujat käyttivät. Sen sijaan neutraali yleiskieli sisälsi yleiskielisten varianttien lisäksi tavallisia puhekielisyyksiä.

Vastaavatko laskennallisen menetelmän avulla löydetyt lekit Helsingin puhekielen taustalla vaikuttavia varieteetteja, kuten hämäläismurteita, yleiskieltä ja slangia?

LDA-mallin avulla löydettiin lekti, joka sisältää paljon hämäläismurteisia variantteja ja toisaalta hyvin yleiskielinen lekti, johon esimerkiksi Paunonen (2019) viittaa. Sen sijaan slangiin palautettavaa lektiä aineistosta ei löytynyt. Tämä ei ole yllättävää, sillä osatutkimuksessa LEKTIT ei juurikaan tarkasteltu sanatasoa, jossa slangi nykypäivänä näkyy selvimmän. Etelähämäläinen murre sisältää joitakin sellaisia Helsinkiin liitettäviä variantteja, jotka voitaisiin yhdistää myös slangiin (erityisesti supistumaverbien analogiset muodot).

Missä piirteissä on havaittavissa yhteisötason muutosta eli sukupolvittaista tai yhteisöllistä muutosta?

Sukupolvittainen muutos on aineistossa hyvin tavallinen muutoksen tapa. Tyypillisimmillään se koskee tilanteita, joissa jokin variantti on muuttumassa yksinomaiseksi (esim. *i*-loppuisen diftongin jälkikomponentin kato). Sukupolvittainen muutos oli tutkituista muutosmalleista selvästi yleisin Helsingin puhekielen kolmen aikapisteen aineistossa. Sen sijaan yhteisöllinen muutos vaikuttaa harvinaiselta: vain yhden variantin muutos osoittautui yhteensopivaksi yhteisöllisen muutosmallin kanssa (sananloppuisen *t:n* kato partisiipeissa). Mikäli yhteisöllistä muutosta tulkitaan hieman väljemmin, myös esimerkiksi tunnuksettoman illatiivin (*meen tekeen*) nopea väistyminen 1990- ja 2010-lukujen välillä osoittaa yhteisöllisen muutoksen piirteitä.

6.1.3 Muutos yksilötasolla

Missä piirteissä on havaittavissa yksilötason muutosta eli ikäsidonnaista, elinikäistä tai vetäytyvää muutosta?

Ikäsidonnainen muutos Labovin esittämänä syklisenä muutoksena on huomattavan harvinainen, ja sen kanssa yhteensopivia variantteja esiintyi aineistossa vain kaksi: *vielä*-sanan käyttö sekä *hän*- ja *he*-pronominien käyttö. Näistä jälkimmäisen osalta on aiemminkin löydetty ikäsidonnaista tendenssiä (Lappalainen 2010: 287–288). Lisäksi pronominvaihteluun liittyy paljon esimerkiksi vuorovaikutuksellisia seikkoja (Priiki 2017). Ikäsidonnaisen muutoksen sijaan ikäefekti vaikuttaa melko tavalliselta monissa muutoksissa. Elinikäisessä muutoksessa vanhemmat puhujat omaksuvat meilläään olevan muutoksen vanhemmalla iällä, mutta tämän muutosmallin kanssa tilastollisesti yhteensopivia variantteja ei löytynyt yhtäkään. Syynä tähän on ennen kaikkea se, että puhujat muuttuvat vanhemmalla iällä paljon ristiin: osa muutoksen suuntaan ja osa sitä vastaan. Elinikäiseen muutokseen viittavia muutoksia esiintyi aineistossa kuitenkin esimerkiksi *t:n* käytössä *t:n* vahvan asteen vastineena ja *OA*- ja *eA*-nominien oikenemisessä. Vetäytyvän muutoksen mallille löytyi myös tilastollista tukea. Tähän muutosmalliin sopiva variantti on esimerkiksi loppuheitto elatiivissa.

Miten yksilöt käyttävät lektejä? Käyttävätkö yksilöt vain yhtä lektiä vai yhdistelevätkö he useampia? Mitä lektejä yhdistellään?

Valtaosa puhujista käyttää useampaa lektiä puheessaan. Tämä löydös on yhteensopiva aikaisempien lektitutkimusten (esim. Gregersen & Pharao 2016) kanssa. Vain

hyvin harvat rekisteriytyneen yleiskielen puhujat keskittyvät vain yhteen lektiin (vrt. Paunonen 2005b). Tavallisimpia yhdistelmiä ovat toisaalta yleiskielisten lektiin yhdistelmät (rekisteriytynyt ja neutraali) ja toisaalta puhekielisten lektiin yhdistelmät (etelähämäläinen ja valtakunnallinen puhekieli).

6.1.4 Muutoksen teoriat

Miten muutosmallit sopivat käytäntöön? Mitkä mallit ovat yleisiä ja mitkä harvinaisia?

Muutosmalleista erityisesti sukupolvittainen muutos sopii hyvin käytäntöön. Tämä antaa tukea perinteiselle näennäisaikaennusteelle, joka on vuosien kuluessa saanut myös kritiikkiä. On selvää, ettei näennäisaikainen kuva muutoksesta sovi jokaisen piirteen käsittelyyn, mutta tämän tutkimuksen aineistossa sukupolvittainen muutos oli selvästi yleisin muutoksen tapa. Muut Labovin alkuperäiset muutosmallit (stabiili tilanne, ikäsidonlainen muutos ja yhteisöllinen muutos) vaikuttavat melko harvinaisilta. Täydennyksistä vetäytyvän muutoksen mallille löytyi tilastollista tukea, kun taas elinikäisen muutoksen mallille ei. Joissakin piirteissä oli kuitenkin havaittavissa elinikäistä tendenssiä. On myös korostettava, että muutosmallit eivät toimi juuri koskaan sellaisinaan yli kolmen aikapisteen, vaan selvältäkään näyttänyt muutos on saattanut myöhemmin hidastua tai pysähtyä. Tämä on sikäli luonnollista, että mikäli ajatellaan piirteiden muuttuvan esimerkiksi S-käyrän mukaisesti, niiden muutosvauhti ei tietenkään voi pysyä tasaisena.

Miten muutosmallien teoria on sovitettavissa perinteiseen S-käyrään?

Muutosmallien teoria on yhdistettävissä S-käyrän kanssa. Tämä edellyttää kuitenkin joko sen hyväksymistä, että muutokset käyvät läpi useita eri muutosmalleja tai S-käyrään sisältyvän muutosvauhdin vaihtelun oletuksen hylkäämistä. Mikäli lähdetään ensimmäisestä oletuksesta, S-käyrän hitaassa alkuvaiheessa muutoksen tapana on vetäytyvä muutos, jota seuraa sukupolvittainen muutos ja nopeassa keskivaiheessa muutoksen tapana on oltava joko elinikäinen tai yhteisöllinen muutos. Nopeaa keskivaihetta seuraavat jälleen yhteisöä hitaammin muuttavat muutosmallit eli sukupolvittainen ja vetäytyvä muutos. On todennäköistä, että kaikki muutokset eivät käy kaikkia eri muutosmalleja läpi. On myös todettava, että esimerkiksi syklinen ikäsidonlainen muutos ei näytä erityisen hyvin sopivan yhteen S-käyrän kanssa. Tämä toisaalta todistaa myös siitä, että ikäsidonlainen muutos muutosmallina ei välttämättä

ole erityisen hyödyllinen, vaan olisi kenties edullisempaa tutkia erilaisia ikäefektejä muissa muutosmalleissa.

Miten morfologinen ja leksikaalinen diffuusio selittävät kadon etenemistä *bd*-yhtymässä?

Kato etenee sekä morfologisen että leksikaalisen diffuusion mallin mukaisesti. Joissakin muotoryhmissä kato on jo 1970-luvun aineistossa hyvin vakiintunut. Kato etenee myös uusiin muotoryhmiin 1990- ja 2010-luvulla. Joissakin muotoryhmissä (esim. diftongin jälkeisessä asemassa) diffuusio on ennemminkin leksikaalista, sillä katoa esiintyy pääasiassa vain yhdessä lekseemissä (*vaihda*). Suomen kielen muutosten kuvaukseen leksikaalisen ja morfologisen diffuusion hybridimalli näyttää osatutkimuksen KATOMUODOT perusteella sopivan hyvin.

6.2 Päätelmät

Tässä alaluvussa kokoan synteessin edellä esitellyistä tuloksista. Alaluvun tarkoituksena on tarkastella, miten Helsingin puhekieli on muuttunut 1970- ja 2010-luvun välillä ja mitä tulokset tarkoittavat teoreettiselta kannalta. Alaluvussa 6.2.1 luon kokonaiskuvan Helsingin puhekielessä tapahtuneista muutoksista 1970-luvulta 2010-luvulle ja alaluvussa 6.2.2 kokoan yhteen tutkimuksen teoreettiset päätelmät.

6.2.1 Muutos Helsingin puhekielessä 1970-luvulta 2010-luvulle

Kaikkein keskeisin muutos Helsingin puhekielessä tutkimuksen ajanjaksona on yleiskielisten varianttien väistyminen. Selkeimmin ero on havaittavissa 1970-luvun vanhojen ja keski-ikäisten puhujien sekä kaikkien muiden kohorttien välillä (vrt. Paunonen 2006: 78). Erityisesti osatutkimuksessa LEKTIT havaitun rekisteriytyneen yleiskielen variantit näyttävät hävinneen 1990- ja 2010-luvulla lähes kokonaan. Tällaisia variantteja ovat esimerkiksi monikon ensimmäisen persoonan käyttö (*menemme*), persoonapronominit *minä* ja *sinä* sekä sananloppuisen *i:n* säilyminen kaikissa asemissa (*tulisi, sanoi, taloksi, viisi*). Vaikka yleiskielisten varianttien väistyminen näyttää pääasialliselta muutoksen suunnalta, monet yleiskieliset variantit ovat kuitenkin vielä 2010-luvun aineistossakin tavallisia. Esimerkiksi *iA*-yhtymän partitiiviesiintymistä yhtymän säilyttäneet muodot (*kahvia*) ovat enemmistössä kaikilla muilla ikäryhmillä paitsi nuorilla, joilla enemmistönä on oiennut variantti (*kahvi*).

Etelähämäläisyydet ovat vielä 2010-luvullakin hyvin yleisiä, ja niistä osa näyttää myös edelleen yleistyvän (esim. *iA*-yhtymien oikeeneminen). Sen sijaan perinteisemmat hämäläisyydet, kuten vaihtelematon *# t:*n vastineena näyttävät hitaasti väistyvän Helsingin puhekielestä. Osatutkimuksen LEKTIT kolmas puhekielinen varieteetti, valtakunnallinen puhekieli on 2010-luvun nuorten keskuudessa kaikkein yleisin. Kuten osatutkimuksessa MUUTOSMALLIT osoitettiin, tämä ei kuitenkaan välttämättä enusta tämän varieteetin yleistymistä Helsingin puhekielessä tulevaisuudessa. On kuitenkin huomattava, että monet piirteet valtakunnallisessa puhekielessä ovat jo joko hyvin yleisiä kaikkialla Suomessa (passiivi, *mä* ja *sä*) tai vähintään yleistymässä (*OA*- ja *eA*-vokaaliyhtymien oikeeneminen, kato *d:*n vastineena *b:*n jäljessä). Tämän varieteetin yleistymisen hidastuminen siis edellyttäisi sitä, että monet jo pitkään vallalla olleet kehityskulut yhtäkkiä muuttuisivat.

2010-luvun aineiston perusteella vaikuttaa siltä, että sekä formaalien varieteettien että puhekielisten varieteettien osalta yleisimpiä ovat sellaiset lektit, joissa on mahdollisimman vähän leimaavia piirteitä. Neutraalissa yleiskielessä käytetään yleiskielisten piirteiden lisäksi tavallisimpia puhekielisyyksiä (esim. ns. pikapuhemuodot *oon*, *tää*) ja vältellään kaikkein leimaavimpia yleiskielisyyksiä (monikon 1. persoonan käyttö). Sen sijaan valtakunnallisessa puhekielessä suositaan esimerkiksi laajan levikin piirteitä (esim. *OA*- ja *eA*-yhtymien oikeeneminen), mutta vältellään pienemmän levikin variantteja (esim. vaihtelematon *#* yleiskielen *t:*n vastineena).

Helsingin puhekieli onkin 2010-luvun aineiston valossa tasoittunut suosimaan sellaisia variantteja, joita käytetään yleisesti muuallakin. Vastaavaa tendenssiä on löydetty niin suomen kuin ruotsinkin maaseutumurteista (Thelander 1979a: 112–116; Mielikäinen 1982: 280–287; Paunonen 1982b: 46–52; Lappalainen 1996: 26–27; Mantila 1997; Mielikäinen 2003: 261–264). Toisaalta esimerkiksi yleisgemiaation nopea häviäminen Tampereen puhekielestä kertoo siitä, että tendenssi on tavallinen myös kaupungeissa (Jonninen-Niilekselä 1982b: 156–160; Tampereen lähistöltä myös Nahkola 1987: 89–92; Nahkola & Saanilahti 2001: 35–38). Suppealevikkisten piirteiden välttely on erityisesti nuorten puheelle leimallista (Lappalainen 2001: 77). Tutkimus siis antaa vahvistuksen niille havainnoille, joita on suomalaisessa sosiolingvistiikassa tehty jo lähes 40 vuoden ajan (vrt. Mantila 2006: 298–299). Väitöskirjan poikkeuksellinen aikaskaala on toisaalta tavoittanut myös sen, että vielä 1970-luvulla tämä tendenssi ei ollut yhtä voimakas kuin nykypäivänä.

Ajoittain valtakunnallisen puhekielen piirteiden yleistymistä on pidetty nimenomaan helsinkilähtöisenä (Mielikäinen 2003: 264; Nuolijärvi & Sorjonen 2005: 126). Tämä asettuu kuitenkin kyseenalaiseksi, kun tasoittumistendenssiä on pystytty suomen maaseutumurteissa seuraamaan jo vähintään 1900-luvun puoliväliin. Mikäli

kansanmurteiden sisäisestä vaihtelusta olisi kerätty tarkkoja tietoja, voitaisiin tasoittumista ajoittaa todennäköisesti vieläkin aikaisemmaksi (vrt. Nuolijärvi 1988). Näyttää siltä, että puhujat niin maaseudulla kuin kaupungissa välttelevät leimaavia piirteitä. Koska Helsingin asema valtakunnan pääkaupunkina ja ainoana metropolina on huomattavan näkyvä, tulevat nämä valtakunnalliset piirteet paremmin näkyville. On myös huomattava, että vaikka pääkaupunkiseutu on alueellisesti pieni, puhujia on kuitenkin paljon (Lappalainen 1996: 26).

6.2.2 Teoreettiset päätelmät

Tutkimus tuotti runsaasti uutta tietoa sosiolingvistiikan teoriasta. Tutkimuksen aineiston jakautuminen kolmeen aikapisteeseen mahdollisti Labovin muutosmallien tarkastelun pidemmän aikavälin aineistossa. Muutosmallien sopimattomuus näin pitkälle aikavälille sekä joidenkin muutosmallien harvinaisuus olivat tutkimuksen keskeisiä löydöksiä. Vastaavasti diffuusiomallin testaaminen käytännössä kuuluu tutkimuksen ansioihin. Tarkastelen näitä löydöksiä vielä seuraavassa.

Monien muutosmallien kanssa yhteensopivia piirteitä löytyi vain hyvin vähän tai ei lainkaan. Aineistossa selvästi yleisin on sukupolvittaisen muutoksen malli. Tämä löydös tuo lisäuskottavuutta pitkään käytetylle näennäisaikamenetelmälle, koska sukupolvittaisen muutoksen malli on juuri sen kanssa yhteensopiva. Niin sukupolvittaisessa muutoksessa kuin näennäisaikamenetelmässä puhujien kielenkäyttö ei enää vanhemmalla iällä muutu. Taustalla vaikuttaa niin sanottu kriittisen iän hypoteesi: varhaisnuorena ohitetun kriittisen iän jälkeen kieli vakiintuu (Lenneberg 1967: 178–179; Meyerhoff 2011: 141; Laasanen 2016: 65–66).

Eri muutosmallien vertailun lisäksi osatutkimuksen MUUTOSMALLIT tärkein havainto oli, että muutosmallit eivät näytä toimivan erityisen hyvin kolmen aikapisteen yli. Perinteisesti muutosmalleja on hyödynnetty kahden aikapisteen tutkimusten kuvaamiseen ja toisaalta muutoksen suunnan ja nopeuden ennustamiseen (Labov 1994: 85–98; Paunonen 2005a; Sankoff & Blondeau 2007). Tutkimus osoitti, että erityisesti ennusteiden suhteen on syytä olla varovainen: joltakin muutosmallilta kahden aikapisteen välillä näyttävät tendenssit saattavat olla kolmanteen pisteeseen tullessa pysähtyneet, harpanneet tai kääntyneet. Mikäli muutosmalleja tulkitaan yhdessä S-käyrän kanssa, tämä on luonnollista, koska muutokset käyvät läpi useampia vaiheita. Muutosmallit toimivatkin yksittäisinä lähinnä tapahtuneiden muutosten kuvaamiseen.

Muutosmallien tarkastelu osana S-käyrää tuotti havaintoja siitä, minkä tyyppiset muutokset sopivat mihinkin muutoksen vaiheeseen. Sukupolvittainen muutos näytti kiinnittyvän erityisesti tilanteisiin, joissa jokin variantti on muuttumassa yksin-omaiseksi tai häviämässä kokonaan. Vetäytyvän muutoksen malli sopi sekä orastaviin muutoksiin että loppuvaiheen muutoksiin. Elinikäisen muutoksen mallilla sen sijaan voi havainnollistaa muutoksen nopeaa keskivaihetta. Stabiilissa tilanteessa muutosta ei esiinny, joten se asettuu S-käyrän alkuun ja loppuun.

Ikäsidonnan muutos ei nähdäkseni sovi erityisen hyvin S-käyrän osaksi. Tämä ei ole muutosmallin ainoa ongelma. Ikäsidonnanaisuutta on kansainvälisessä variaati-ontutkimuksessa määritelty toisaalta syklisenä yksilöiden muutoksena, jossa yhteisö pysyy ennallaan (Labov 1994), ja toisaalta löyhemmin kuvaamaan tietyssä ikävaiheessa tapahtuvia muutostendenssejä, jotka voivat johtaa yhteisön muutokseen (esim. Sankoff 2005: 1011; Cheshire 2005: 1553). Labovin esittämä syklinen ikäsidonnan muutos on teoreettisesti selvempi, ja se mahdollisti esimerkiksi tässä tutkimuksessa muutosmallin tarkastelun tilastollisesti. Sen sijaan jälkimmäinen on todellisempi: on selvää, että monissa tämänkin tutkimuksen piirteissä on ikäkausittaista variaatiota (esim. infinitiivit, katomuodot *hd*-yhtymässä). Tätä on tässä tutkimuksessa kuvattu ikäefektiksi: ikäefektit voivat joko hidastaa (vetäytyvä muutos, kaareva muutos) tai nopeuttaa (elinikäinen muutos) yhteisön muutosta. Esimerkiksi lyhyen infinitiivin eteneminen näyttää toteuttavan vetäytyvän muutoksen mallia, kun taas *OA*- ja *eA*-yhtymien oikeminen nomineissa puolestaan näyttää olevan osittain yhteensopiva elinikäisen muutoksen mallin kanssa. Labovin ikäsidonnan muutosmalli, jossa yksilöt muuttuvat tietyssä ikävaiheessa yhä uudelleen yhteisön pysyessä ennallaan, ei sen sijaan vaikuta erityisen hyödylliseltä mallilta. Väitöskirjan keskeisenä ansiona voi pitää iän ja muutoksen suhteen tarkempaa selvittelyä.

Muutosmallien lisäksi väitöskirjassa tarkasteltiin myös diffuusiomallia. Alun perin kiinan kielen tutkimuksesta lähtenyt leksikaalisen diffuusion malli on kotimaisessa kielentutkimuksessa kääntynyt morfologiseksi diffuusioksi (Suihkonen 1992: 286; Mielikäinen 1995). Morfologisen diffuusion malli on kulkenut fennistisissä variaati-ontutkimuksissa useasti mukana (esim. Vaattovaara 2002: 513–514; Kurki 2005: 26; Mustanoja 2011: 76–77). Kuitenkin silloin kun mallia on hyödynnetty käytännössä, sitä on tarkasteltu jo tapahtuneiden muutosten levinneisyyden perusteella (Suihkonen 1992: 303; Mielikäinen 1995: 326–328). Osatutkimuksessa KATOMUODOT morfologista diffuusiota lähestyttiin tilastollisesti korrelaatioanalyysin avulla. Näin diffuusiota pystyttiin tarkastelemaan reaaliajassa tutkimuksen kolmen aikapisteen välillä. Diffuusiomalli näyttää siis muutosmallien ohella sopivan käytäntöön. Tämä havainto

toivottavasti rohkaisee myös muita tutkijoita hyödyntämään sosiolingvistiikan teoreettisia malleja myös käytännön aineistoihin.

Väitöskirjassa on keskitytty perinteisiin variaation tutkimuksen aiheisiin: kaupunkikielen reaaliaikaturkimukseen, jo dialektologiasta periytyviin tutkittaviin piirteisiin ja vuosikymmenien takaisiin teorioihin. Näitä kaikkia on kuitenkin käsitelty tuoreella tavalla: reaaliaikaturkimus kattaa kolme aikapistettä, infinitiivejä on tarkasteltu systeeminä eikä yksittäisinä piirteinä, katomuotojen etenemisestä on saatu konkreettista näyttöä, muutosmalleja on tutkittu käytännössä ja morfologista diffuusiota sekä Helsingissä puhuttavia lekkejä on mallinnettu laskennallisesti. Väitöskirja on näin ollen vankasti osa perinteistä suomalaista variaation tutkimusta sitä kuitenkin rohkeasti uudistaen.

6.3 Tutkimuksen arviointi ja jatkotutkimuksen mahdollisuudet

Tutkimusprosessi on kestänyt kokonaisuudessaan neljä vuotta. Prosessi käynnistyi varianttien poiminnalla Helsingin puhekielen aineiston litteroiduista haastatteluista. Variantit valikoitiin pääasiassa menneiden tutkimusten (erityisesti Paunonen 2006: 36–50) perusteella. Poiminnassa tehdyt valinnat ovat luonnollisesti vaikuttaneet väitöskirjan tuloksiin. Näitä valintoja ovat esimerkiksi joidenkin piirteiden jättäminen tutkimuksen ulkopuolelle (esim. monikon partitiivin päätevariaatio *tyttöjä* ~ *tyttöi*) ja toisaalta joidenkin piirteiden rajaaminen tiettyihin lekseemeihin (esim. *metsä*-sananjättäminen *ts*-yhtymän tulosten ulkopuolelle). Mikäli aloittaisin tutkimusta uudelleen nyt, valitsisin luultavasti suuremman joukon piirteitä. Toisaalta jo nykyisenkin tutkimuksen ansiona voidaan pitää sitä, että mukaan on valittu hyvin laaja joukko kielenpiirteitä.

Piirteiden poiminta oli erilaisten työkalujen, kuten säännöllisten lausekkeiden, avusta huolimatta hyvin aikaa vievää, sillä lähes kaikkien varianttien esiintymät piti vielä tarkistaa manuaalisesti. Jos aloittaisin tutkimusta nyt, käyttäisin enemmän aikaa joko piirteiden valintaan niin, että niitä olisi helpompi hakea litteroiduista haastatteluista, tai itse hakujen optimoimiseen. Piirteiden poimintaan käytetystä ajasta osan olisi näin voinut käyttää sellaisten aineistossa mukana olevien haastattelujen litteroimiseen, joista litteraattia ei ollut valmiina. Erityisesti 1990-luvun osuus jäi nykyisessä tutkimuksessa hieman liian pieneksi verrattuna kahteen muuhun vuosikymmeneen. Näin tutkimus olisi saattanut kärsiä joidenkin piirteiden puutteesta mutta olla toisaalta edustavampi vuosikymmenten välillä.

Aineiston taustamuuttujista on keskitytty tässä väitöskirjassa pääasiassa vain ikään. Näin on toimittu siitä syystä, että iän on aiemmin huomioitu vaikuttavan taustamuuttujista eniten kielelliseen variaatioon (Thelander 1979b: 93–95; Lappalainen 1996: 30; Nuolijärvi & Sorjonen 2005: 102). Vaikka taustamuuttujat ovat saaneet paljon ansaittuakin kritiikkiä osakseen (Pratt 1987; Romaine 1994: 221–225; Bucholtz 2003; Vaattovaara 2009: 23), olisi niitä voinut hyödyntää laajemmin tutkimuksen eri vaiheissa. Tämä toisaalta jättää myös mahdollisuuden Helsingin puhekielen pitkittäiskorpuksesta tehtäville jatkotutkimuksille.

Tutkimuksen ansiona pidän perinteisten variaation tutkimuksen kohteiden tarkastelua kielisysteemistä käsin. Niin infinitiiveissä tavattava vaihtelu kuin *bd*-yhtymän sijaan esiintyvä kato ovat jo vanhastaan tutkittuja kielenpiirteitä. Tässä tutkimuksessa niitä on tarkasteltu kielensisäisten muutosten kannalta. Näiden piirteiden variaatiosta ja muutoksesta on löytynyt paljon sellaista, mikä on aiemmin jäänyt huomaamatta. Siinä missä näiden kahden piirteen muutosta on tarkasteltu rakenteellisesti, on vastaavasti tarkasteltu myös kieliyhteisön muutosta. Aiemmin oletetuille varieteeteille Helsingin puhekielessä (yleiskieli, hämäläismurre) todella löytyi vastine myös laskennallisesti tarkasteltuna. Lisäksi valtakunnallisen puhekielen esiintyminen omana varieteettinaan todistaa siitä, että kyseessä ei ole pelkkä yksittäisten varianttien kokoelma, vaan todellinen kokonaisuus (vrt. Mantila 1997; Hurta 1999). William Labovin muutosmallien (1994) ja niiden täydennysten tarkastelu käytännön aineistossa yli kolmen aikapisteen on paljastanut mallien käyttökelpoisia ja ongelmallisia puolia. Vastaavasti muutosmallien ja S-käyrän yhteensovittaminen on sosiolingvistiikan teorian kannalta keskeistä.

Jatkotutkimuksen aiheita jää useita. Väitöskirjassa on tarkasteltu muutosta monesta näkökulmasta: yksittäisten piirteiden, varieteettien ja muutoksen teorioiden kannalta. Idiolektitaso on kuitenkin jäänyt vähälle huomiolle. Idiolektien muutosta ovat aineistosta tarkastelleet Lappalainen, Mustanoja ja O’Dell (2019), mutta idiolektitasolla voitaisiin tarkastella esimerkiksi vain 2010-luvun nuoria. Tässä joukossa tuntuu olevan suurta vaihtelua, ja erityisesti yksi puhuja nousee esiin hyvin yleiskielisenä. Tällainen yksittäisiin puhujiin keskittyvä tarkastelu vastaisi kansainvälisesti tehtyjä tutkimuksia (esim. Johnstone & Kiesling 2008; Buchstaller 2015).

Osatutkimuksista selvimmän jatkokysymyksen herättää KATOMUODOT: mitkä syyt vaikuttavat katomuotojen etenemiseen morfologisen diffuusion mallilla yleiskielen *bd*-yhtymässä? Esimerkiksi *bd*-yhtymän esiintyminen konsonanttivartalossa tai proprinen merkitys saattavat hidastaa kadon etenemistä, mutta näitä syitä olisi tarkasteltava lähemmin. Toisekseen osatutkimusta LEKTIT voitaisiin laajentaa koskemaan itse puhujia: millaiset puhujat käyttävät mitäkin lektiä?

Tutkimuksessa hyödynnettyjä tilastollisia menetelmiä voi hyvin hyödyntää muuhinkin puhutun kielen korpuksiin ja tutkia esimerkiksi Tampereen puhekielessä esiintyviä lekkeitä LDA-mallilla. Itse tutkimusaineiston osalta on muistettava, että nykyisellään Helsingin puhekielen korpus on melko homogeeninen, sillä puhujat on valittu tietyin periaattein. Miten hyvin tällainen korpus todella kuvaa esimerkiksi 2020-luvun Helsinkiä, jossa muiden kuin äidinkieleltään suomen-, ruotsin- ja saamenkielisten osuus väestöstä on 16 prosenttia ja valtaosa kaupungin väkiluvun kasvusta perustuu tämän väestöosan osuuden kasvuun (Vuori, Mäki & Bergroth 2020: 20)? Hankkeen seuraavilla kierroksilla onkin harkittava jälleen kerran puhujien valintaan liittyviä kysymyksiä (vrt. Nuolijärvi & Sorjonen 2005: 135). Jonkinlaista esimerkkiä voi ottaa esimerkiksi 2010-luvun lopulla kerätystä ja monikielisyyden huomioivasta Sata suomalaista kielellistä elämäkertaa -hankkeesta (Hippi, Lappalainen & Nuolijärvi 2020).

Viimeiseksi on korostettava, että tässä tutkimuksessa esitetyt tulokset koskevat vain Helsinkiä. Miten tässä kuvatut muutokset esimerkiksi infinitiiveissä, *hd*-yhtymässä, *z:n* vastineissa tai *A*-loppuisissa vokaaliyhtymissä vertautuvat muualla Suomessa tapahtuneisiin muutoksiin? Tässä yhteydessä voidaan tarkastella yhtä hyvin kaupunkien puhekieliä kuin nykyisiä paikallismurteitakin. Millaisia lekkeitä löytyy esimerkiksi Tampereen puhekielestä? Leviävätkö valtakunnallisen puhekielen variantit edelleen muualla Suomessa? Sopiiko esimerkiksi sukupolvittainen muutosmalli kuvaamaan erityisesti loppuvaiheessa olevia muutoksia myös muualla? Erityisesti muutosmallien soveltaminen ja erilaisten ikäefektien tarkastelu Helsingin ulkopuolella tuottaisi varmempaa tietoa muutosten etenemisen tavoista.

LÄHTEET

Aineistolähteet

Helpuhe = Helsingin puhekielen pitkittäiskorpus (1970, 1990, 2010) [online-puhekorpus]. - Helsinki. Helsingin yliopiston suomen kielen, suomalais-ugrialaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos, Kotimaisten kielten keskus ja Heikki Paunonen, 2014. URN: <http://urn.fi/urn:nbn:fi:lb-2014073041>.

Kirjallisuuslähteet

- AGHA, ASIF 2003. The social life of cultural value. – *Language and communication* 23 s. 231–273.
- AINIALA, TERHI – VUOLTEENAHO, JANI 2005. Urbaani muutos ja kaupunkilaiset identiteetit paikannimistön kuvaajina. – *Virittäjä* 109 s. 378–394.
- AINIALA, TERHI – HALONEN, MIA 2011. Somalialaistaustaiset nuoret nimistönkäyttäjänä Itä-Helsingissä. – *Virittäjä* 115 s. 193–220.
- AINIALA, TERHI – JANTUNEN, JARMO HARRI 2019. Korpusonomastinen tutkimus slanginimistä Hesa ja Stadi digitaalisissa diskursseissa. – *Sananjalka* 61 s. 57–79.
- ANHAVA, JAAKKO 2019. Puhuiko Suomen sivistyneistö kirjakieltä? – *Virittäjä* 123 s. 443–445.
- ANTTILA, RAIMO 1989: *Historical and Comparative Linguistics*. 2. painos. Amsterdam: Benjamin.
- BAILEY, CHARLES-JAMES 1973. *Variation and linguistic theory*. Arlington: Center for Applied Linguistics.
- BLEI, DAVID – NG, ANDREW – JORDAN, MICHAEL 2003. Latent Dirichlet Allocation. – *Journal of Machine Learning Research* 3 s. 993–1022.
- BOURDIEU, PIERRE – BOLTANSKI, LUC 1975. Le fétichisme de la langue. – *Actes de la recherche en sciences sociales* 1 s. 1–32.
- BOWIE, DAVID – YAEGER-DROR, MALCAH 2015. Phonological change in real time. – Patrick Honeybone & Joseph Salmons (toim.), *Handbook of historical phonology*. s. 603–618. Oxford: Oxford University Press.
- BUCHOLTZ, MARY 2003. Sociolinguistic nostalgia and the authentication of identity. – *Journal of Sociolinguistics* 7 s. 398–416.
- BUCHSTALLER, ISABELLE 2015. Exploring linguistic malleability across the life span. Age-specific patterns in quotative use. – *Language in Society* 44 s. 457–496.
- BYBEE, JOAN 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. – *Language Variation and Change* 14 s. 261–290.
- CHAMBERS, J. K. 2009 [1995]. *Sociolinguistic theory*. Kolmas painos. Malden: Wiley-Blackwell
- CHAMBERS, J. K – TRUDGILL, PETER 1998. *Dialectology*. Toinen painos. Cambridge: Cambridge University Press.

- CHESHIRE, JENNY 2005. Age and generation-specific use of language. – Ulrich Ammon, Norbert Dittmar, Klaus Mattheier & Peter Trudgill (toim.), *Sociolinguistics: An introductory handbook of the science of language and society* s. 1552–1563. Berlin: Mouton de Gruyter.
- COX, DAVID 1958: The regression analysis of binary sequences. – *Journal of the Royal Statistical Society B* 20 s. 215–242.
- DOWNES, WILLIAM 1998. *Language and society*. Toinen painos. Cambridge University Press, Cambridge.
- ECKERT, PENELOPE 1997. Age as a sociolinguistic variable. – Florian Coulmas (toim.), *The handbook of sociolinguistics* s. 151–167. Oxford: Blackwell.
- ECKERT, PENELOPE 2008. Variation and the indexical field. – *Journal of Sociolinguistics* 12 s. 453–476.
- GREGERSEN, FRANS – PHARAO, NICOLAI 2016. Lects are perceptually invariant, productively variable. A coherent claim about Danish lects. – *Lingua* 172–173 s. 26–44.
- GROSS, JOHAN 2018. Segregated vowels: Language variation and dialect features among Gothenburg youth. – *Language Variation and Change* 30 s. 315–336.
- GUY, GREGORY R. 2013. The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? – *Journal of Pragmatics* 52 s. 63–71.
- HALONEN, MIA – VAATTOVAARA, JOHANNA 2017. Tracing the indexicalization of the notion “Helsinki s”. – *Linguistics* 55 s. 1169–1195.
- HALONEN, MIA – NYSTRÖM, SAMU – PAUNONEN, HEIKKI – VAATTOVAARA, JOHANNA 2020. *Stadin syntinen* s. Helsinki: Art House.
- HAY, JENNIFER – FOULKES, PAUL 2016. The evolution of medial /t/ over real and remembered time. – *Language* 92 s. 298–330.
- HERLIN, ILONA 2012. Miksi alkaa tekemään? – Ilona Herlin & Lari Kotilainen (toim.), *Verbit ja konstruktiot* s. 143–174. Suomi 201. Helsinki: Suomalaisen Kirjallisuuden Seura.
- HERLIN, ILONA – LEINO, PENTTI – VISAPÄÄ, LAURA 2005. Kas siinä pulma. – Ilona Herlin & Laura Visapäätä (toim.), *Elävä kielioppi. Suomen infinittisten rakenteiden dynamiikkaa* s. 9–38. Helsinki: Suomalaisen Kirjallisuuden Seura.
- HIPPI, KAARINA – LAPPALAINEN, HANNA – NUOLIJÄRVI, PIRKKO (toim.) 2020. *Suomalaisen kielestä elämää*. SKST 1465. Helsinki: Suomalaisen Kirjallisuuden Seura.
- HORVATH, BARBARA – SANKOFF, DAVID 1987. *Delimiting the Sydney speech community*. – *Language in Society* 16 s. 179–204.
- HURTTA, HEIKKI 1999. Variaationitutkimuksen myytit ja stereotyyppit. – Urho Määttä, Pekka Pälli & Matti K. Suojanen (toim.), *Kirjoituksia sosiolingvistiikasta* s. 53–101. Tampere: Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
- ITKONEN, TERHO 1964: *Proto-Finnic final consonants. Their history in the Finnic languages with particular reference to the Finnish dialects*. Helsinki: Suomalais-Ugrilainen Seura.
- ITKONEN, TERHO 1972. Kielenhuoltomme tehtävistä. – *Viritäjä* 76 (jatkoa edelliseen vihkoon) s. 371–387.
- ITKONEN, TERHO 1973: Suomen murteiden transkription yksinkertaistamisesta. – Lauri Posti & Terho Itkonen (toim.), *FU-transkription yksinkertaistaminen* s. 19–31. Castrenianumin toimitteita 7. Helsinki: Helsingin yliopisto.
- JOHNSTONE, BARBARA – KIESLING, SCOTT F. 2008. Indexicality and experience: Exploring the meanings of /aw/-monophthongization in Pittsburgh. – *Journal of Sociolinguistics* 12 s. 5–33.
- JONNINEN-NIILEKSELÄ, KAIJA (toim.) 1982a. *Tampereen pubekieli tutkimuskohteena*. Folia Fenestica & Linguistica 6. Tampere: Tampereen yliopisto.

- JONNINEN-NIILEKSELÄ, KAIJA 1982b. Eräitä äänne- ja muoto-opillisia piirteitä. – Kaija Jonninen-Niilekselä (toim.), *Tampereen pubekieli tutkimuskohteena*. Folia Fennistica & Linguistica 6, s. 121–160. Tampere: Tampereen yliopisto.
- JUUSELA, KAISU – NISULA, KATARIINA (toim.) 2006. *Helsinki kieliyhteisönä*. Helsinki: Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.
- KARHU, ANNA 1995. Murrehaastattelu vuorovaikutustilanteena. – Marjatta Palander (toim.), *Murteiden matkassa. Jublakirja Alpo Räisäsän 60-vuotispäiväksi* s. 92–109. Studia Carelica Humanistica 6. Joensuu: Joensuun yliopiston humanistinen tiedekunta.
- KARLSSON, FRED 1975. Suomen kielen tulevaisuus. – *Sananjalka* 17 s. 51–66.
- KETTUNEN, LAURI 1930. *Suomen murteet II. Murrealueet*. SKST 188. Helsinki: Suomalaisen Kirjallisuuden Seura.
- KETTUNEN, LAURI 1940. *Suomen murteet III A. Murrekartasto*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- KLEENE, STEPHEN 1956. Representation of events in nerve nets and finite automata. – *Automata Studies (Annals of Mathematics Studies 34)* s. 3–41.
- KOTSINAS, ULLA-BRITT 1994. *Ungdomsspråk*. Uppsala: Hallgren & Fallgren.
- KURKI, TOMMI 2005. *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen. Kielenmuutosten seuraamisesta ja niiden tarkastelussa käytettävistä menetelmistä*. SKST 1036. Helsinki: SKS.
- KURKI, TOMMI 2013. Näennäisaikametodi ja muutosmallit. – *Sananjalka* 55 s. 202–212.
- KURKI, TOMMI – MUSTANOJA, LIISA 2020. Haastattelu aineistonkeruun metodina sosiolinguistikassa. – Milla Luodonpää-Manni, Markus Hamunen, Reetta Konstenius, Matti Miestamo, Urpo Nikanne & Kaius Sinnemäki (toim.), *Kielentutkimuksen menetelmiä II. Aineistonkeruusta analyysiin*. SKST 1457. Helsinki: Suomalaisen Kirjallisuuden Seura.
- KYTÖMÄKI, JORMA 1981. Maantieteellinen tutkimus lingvistin apuna. – *Sananjalka* 23 s. 25–36.
- LAASANEN, MIKKO 2016. Näennäisaikametodin taustaoletusten teoreettista tarkastelua. – *Virittäjä* 120 s. 57–83.
- LABOV, WILLIAM 1966. *The Social Stratification of English in New York City*. Washington D.C: Center for Applied Linguistics.
- LABOV, WILLIAM 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- LABOV, WILLIAM 1994. *Principles of Linguistic Change: Internal Factors*. New York: Blackwell.
- LAPPALAINEN, HANNA 1996. Variaatio ja sen selittäminen kaakkoishämäläisen murreaineiston valossa. – *Virittäjä* 100 s. 18–37.
- LAPPALAINEN, HANNA 2001. Sosiolinguvistinen katsaus suomalaisnuorten nykypuhekieleen ja sen tutkimukseen. – *Virittäjä* 105 s. 74–101.
- LAPPALAINEN, HANNA 2004. *Variaatio ja sen funktiot. Erään sosiaalisen verkoston jäsenten kielellisen variaation ja vuorovaikutuksen tarkastelua*. SKST 964. Helsinki: Suomalaisen Kirjallisuuden Seura.
- LAPPALAINEN, HANNA 2010. *Hän vai se, he vai ne? Pronominivariaatio ja normien ristiveto*. – Hanna Lappalainen, Marja-Leena Sorjonen & Maria Vilkkuna (toim.), *Kielellä on merkitystä. Näkökulmia kielipolitiikkaan* s. 279–324. SKST 1262. Helsinki: Suomalaisen Kirjallisuuden Seura.
- LAPPALAINEN, HANNA – MUSTANOJA, LIISA – O'DELL, MICHAEL 2019. Miten ja milloin yksilön kieli muuttuu? – *Virittäjä* 123 s. 550–581.
- LEHTIMÄKI, PEKKA 1983. Sekaidiolektien tutkimuksesta. – *Virittäjä* 87, s. 22–40.
- LEHTINEN, TAPANI 2007. *Kielen vuosituhannet. Suomen kielen kehitys kantauralista varhaisuomeen*. Helsinki: Suomalaisen Kirjallisuuden Seura.

- LEHTONEN, HEINI 2015. *Tyylitellen. Nuorten kielelliset resurssit ja kielen sosiaalinen indeksisyys monimielisessä Helsingissä*. Väitöskirja. Helsingin yliopisto. Verkkoaineisto. <http://urn.fi/URN:ISBN:978-951-51-1333-7>
- LENNEBERG, ERIC 1967. *Biological foundations of language*. New York: Wiley.
- LENNES, MIETTA 2009. Praat-ohjelmalla annotoidun keskustelun litteraatin vieminen tekstitiedostoon. https://lennes.github.io/spect/howto/keskustelun_litterointi.html (16.2.2021)
- LEVON, EREZ – BUCHSTALLER, ISABELLE 2015. Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing. – *Language Variation and Change* 27 s. 319–348.
- LYYTIKÄINEN, ERKKI 2019. Kyllä kirjakieltä puhuttu on. – *Virittäjä* 123 s. 588–589.
- MA, ROXANA – HERASIMCHUK, ELEANOR 1972. Speech styles in Puerto Rican bilingual speakers: a factor analysis of co-variation of phonological variables. – Joshua A. Fishman (toim.), *Advances in the Sociology of Languages, vol. II*. s. 268–295. The Hague: Mouton.
- MANN, HENRY B. – WHITNEY, DONALD R. 1947. On a test of whether one of two random variables is stochastically larger than the other. – *The Annals of Mathematical Statistics* 18 s. 50–60.
- MANTILA, HARRI 1992. *Ei tääläkän senthän jokhaisben sanhan bootakhan panna. Jälkita-vujen vokaalienvälisen b:n variaatio peräpohjalaisissa murteissa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- MANTILA, HARRI 1997. Johdanto. – Seija Makkonen & Harri Mantila (toim.), *Pohjoissuomalaisen pubekielen sosiolingvistinen variaatio*, s. 1–23. Suomen ja saamen kielen ja logopedian laitoksen julkaisuja 8. Oulun yliopisto.
- MANTILA, HARRI 2004. Murre ja identiteetti. – *Virittäjä* 108, s. 322–346.
- MANTILA, HARRI 2006. Uutta kvantitatiivista analyysia fennistisessä variaation tutkimuksessa. – *Virittäjä* 110 s. 295–299.
- MEYERHOFF, MIRIAM 2011. *Introducing Sociolinguistics*. Toinen painos. London: Routledge.
- MIELIKÄINEN, AILA (toim.) 1980, 1981a, 1981b, 1986. *Nyky-suomalaisen pubekielen murros. Jyväskylän osatutkimus. Raportit 1–4*. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 20, 24, 26, 32.
- MIELIKÄINEN, AILA 1982. Nykypuhe-suomen alueellista taustaa. – *Virittäjä* 86 s. 277–294.
- MIELIKÄINEN, AILA 1991. *Murteiden murros. Levikkeikarttoja nykypubekielen piirteistä*. Jyväskylän yliopiston suomen kielen julkaisuja 36.
- MIELIKÄINEN, AILA 1995. Morfologinen diffuusio. – *Virittäjä* 99 s. 321–336.
- MIELIKÄINEN, AILA 2003. Puitten-tyyppinen monikon genetiivi nykypubekielen tuntomerkinä. – Lea Laitinen, Hanna Lappalainen, Päivi Markkola & Johanna Vaattovaara (toim.), *Muotojen mieli. Kirjoituksia morfologiasta ja variaatiosta* s. 249–270. Kieli 15. Helsinki: Helsingin yliopiston suomen kielen laitos.
- MILROY, LESLIE 2002. Mobility, contact and language change – working with contemporary speech communities. – *Journal of Sociolinguistics* 6 s. 3–15.
- MUSTANOJA, LIISA 2011. *Idiolekti ja sen muuttuminen. Reaaliaikaturkimus Tampereen pubekielestä*. Acta Universitatis Tamperensis 1605. Tampere University Press, Tampere.
- MUSTANOJA, LIISA – O'DELL, MICHAEL 2007 Suomen *d* ja *r* sosiofonettisessä kentässä. – *Virittäjä* 111 s. 56–67.

- NAHKOLA, KARI 1987. *Yleisgemmaatio. Äänne muutoksen synty ja vaiheet kielisysteemissä erityisesti Tampereen seudun hämäläismurteiden kannalta*. SKST 457. Helsinki: Suomalaisen Kirjallisuuden Seura.
- NAHKOLA, KARI – SAANILAHTI, MARJA 2001. *Kymmenen vuotta myöhemmin. Seuruutut- kimus kielenmuutosten etenemisestä yksilön kielessä ja puhe yhteisössä*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 68.
- NAHKOLA, KARI – SAANILAHTI, MARJA 2004. Mapping language changes in real time. A panel study on Finnish. – *Language Variation and Change* 16 s. 75–92.
- NEVALAINEN, TERTTU – RAUMOLIN-BRUNBERG, HELENA – MANNILA, HEIKKI 2011. The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. – *Language Variation and Change* 23 s. 1–43.
- NUOLIJÄRVI, PIRKKO 1986. *Kolmannen sukupolven kieli. Helsinkiin muuttaneiden suurten ikäluokkien eteläpohjalaisten ja pohjoissavolaisten kielellinen sopeutuminen*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- NUOLIJÄRVI, PIRKKO 1988. Kielenulkoisten taustamuuttujien huomioon ottaminen 1800-luvun ja 1900-luvun alun murre tutkimuksissa. – Jyrki Kalliokoski, Pentti Leino & Pertti Pyhtilä (toim.), *Kieli* 3 s. 117–160. Helsingin yliopiston suomen kielen laitos.
- NUOLIJÄRVI, PIRKKO 2000. Sosiolingvistiikka kielentutkimuksen kentässä. – Kari Sajavaara & Arja Piirainen-Marsh (toim.), *Kieli, diskurssi & yhteisö* s. 13–37. Soveltavan kielentutkimuksen teoriaa ja käytäntöä 2. Soveltavan kielentutkimuksen keskus. Jyväskylän yliopisto.
- NUOLIJÄRVI, PIRKKO 2005. Haastattelu kielentutkijan aineistona. – Johanna Ruusuvuori & Liisa Tiittula (toim.), *Haastattelu. Tutkimus, tilanteet ja vuorovaikutus* s. 242–263. Tampere: Vastapaino.
- NUOLIJÄRVI, PIRKKO – TIITTULA, LIISA 2000. *Televisiokeskustelun näyttämöllä. Televisioinstituutionaalisuus suomalaisessa ja saksalaisessa keskustelukulttuurissa*. SKST 768. Helsinki: Suomalaisen Kirjallisuuden Seura.
- NUOLIJÄRVI, PIRKKO – SORJONEN, MARJA-LEENA 2005. *Miten kuvata muutosta? Pubutun kielen tutkimuksen lähtökohia murteenseuruaankkeen pohjalta*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 133. Helsinki: Kotimaisten kielten tutkimuskeskus.
- NYCZ, JENNIFER 2018. Stylistic variation among mobile speakers: Using old and new regional variables to construct complex place identity. – *Language Variation and Change* 30 s. 175–202.
- PAUNONEN, HEIKKI 1971. Sosiolingvistinen metodi ja sen sovellus. – *Viritäjä* 75 s. 401–412.
- PAUNONEN, HEIKKI 1976. Kotikielen Seura 1876–1976. – *Viritäjä* 80 s. 310–432.
- PAUNONEN, HEIKKI 1982a. Muuttuvat puhe-suomen muodot. – Maija Larmola (toim.), *Koulukäisten kieli* s. 130–152. Tietolipas 88. Helsinki: Suomalaisen Kirjallisuuden Seura.
- PAUNONEN, HEIKKI 1982b. Suomen kielen sosiolingvistinen vaihtelu tutkimuskohteena. – Matti K. Suojanen & Päivikki Suojanen (toim.), *Sosiolingvistiikan näkymiä* s. 35–110. Helsinki: Gaudeamus.
- PAUNONEN, HEIKKI 1994. The Finnish language in Helsinki. – Bengt Nordberg (toim.), *The sociolinguistics of urbanization: the case of the Nordic countries*. s. 223–245. Berlin: de Gruyter.
- PAUNONEN, HEIKKI 1995 [1982]. *Suomen kieli Helsingissä. Huomioita Helsingin puhekielen historiallisesta taustasta ja nykyvariaatiosta*. Helsinki: Helsingin yliopiston suomen kielen laitos.
- PAUNONEN, HEIKKI 1995. Puhe-suomen muuttuva omistusmuotojärjestelmä. – *Viritäjä* 99 s. 501–531.

- PAUNONEN, HEIKKI 2005a. Totta vai toiveajattelua. Näennäisaikainen muutos todellisen muutoksen kuvastajana. – Johanna Vaattovaara, Toni Suutari, Hanna Lappalainen & Riho Grünthal (toim.), *Muuttuva muoto. Kirjoituksia Tapani Lehtisen 60-vuotispäivän kunniaksi* s. 13–53. Kieli 16. Helsinki: Helsingin yliopiston suomen kielen laitos.
- PAUNONEN, HEIKKI 2005b. Helsingiläisiä puhujaprofiileja. – *Virtittäjä* 109 s. 162–200.
- PAUNONEN, HEIKKI 2006. Vähemmistökielestä varioivaksi valtakieleksi. – Kaisu Juusela & Katariina Nisula (toim.), *Helsinki kieliyhteisönä* s. 13–99. Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.
- PAUNONEN, HEIKKI 2016. *Sloboa Stadissa. Stadin slangin etymologiaa*. Jyväskylä: Docendo.
- PAUNONEN, HEIKKI 2019. Millaista kieltä suomalainen sivistyneistö puhui? – *Virtittäjä* 123 s. 584–588.
- PAUNONEN, HEIKKI – PAUNONEN, MARJATTA 2000. *Tsennaaks Stadii, bonjaaks slangii. Stadin slangin suursanakirja*. Helsinki: WSOY.
- PAUNONEN, HEIKKI – VUOLTEENAHO, JANI – AINIALA, TERHI 2009. Industrial urbanization, working-class lads and slang toponyms in early twentieth-century Helsinki. – *Urban History* 36 s. 449–472.
- PEARSON, KARL 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. – *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50 s. 157–175.
- PRATT, MARY LOUISE 1987. Linguistic utopias. – Nigel Fabb, Derek Attridge, Alan Durant & Colin MacCabe (toim.), *The linguistics of writing: arguments between language and literature* s. 48–66. New York: Methuen.
- PRIIKI, KATRI 2017. *Hän, se, tää vai toi? Vuorovaikutussosiolingvistinen tutkimus henkilöviit-tauksista Kaakkois-Satakunnan nykypubekielessä*. Annales Universitatis Turkuensis C432. Turku: Turun yliopiston julkaisuja.
- R Core Team 2017. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.
- RAPOLA, MARTTI 1966. *Suomen kielen äännehistorian luennot*. SKST 283. Helsinki: Suomalaisen Kirjallisuuden Seura.
- RAPOLA, MARTTI 1969. *Jobdatus suomen murteisiin*. Kolmas painos. Tietolipas 4. Helsinki: Suomalaisen Kirjallisuuden Seura.
- ROGERS, EVERETT M. 1962. *Diffusion of innovations*. New York: The Free Press.
- ROMAINE, SUZANNE 1994. *Language in society. An Introduction to sociolinguistics*. New York: Oxford University Press.
- SANKOFF, DAVID – LABERGE, SUZANNE 1978. The linguistic market and the statistical explanation of variability. – David Sankoff (toim.), *Linguistic variation: Models and methods* s. 239–250. New York: Academic Press.
- SANKOFF, GILLIAN 2005. Cross-sectional and longitudinal studies. – Norbert Dittmar, Klaus J. Mattheier & Ammon Ulrich (toim.), *Sociolinguistics. An international handbook of the science of language and society* s. 1003–1013. Berlin: de Gruyter.
- SANKOFF, GILLIAN 2018. Before There Were Corpora: The Evolution of the Montreal French Project as a Longitudinal Study. – Suzanne Evans Wagner & Isabelle Buchstaller (toim.), *Panel Studies of Variation and Change* s. 21–52. Oxford: Routledge.
- SANKOFF, GILLIAN & BLONDEAU, HÉLÈNE 2007. Language change across the lifespan: /r/ in Montreal French. – *Language* 83 s. 560–588.

- SAUKKONEN, PAULI 1965. *Itämerensuomalaisten kielten tulosjainfinitiivirakenteiden historiaa I*. Suomalais-Ugrilaisen Seuran Toimituksia 137. Helsinki: Suomalais-Ugrilainen Seura.
- SETÄLÄ, E. N. 1890. *Yhteissuomalainen äännehistoria I–II: Konsonantit*. Helsinki.
- SIITONEN, KIRSTI 1982. Dialektologian ja sosiolingvistiikan suhteesta. – Matti K. Suojanen (toim.), *Kirjoituksia pubekielestä* s. 43–58. Toinen painos. Turun puhekielen projektin julkaisuja 1. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 14.
- SILVERSTEIN, MICHAEL 2003. Indexical order and the dialectics of sociolinguistic life. – *Language and Communication* 23 s. 193–229.
- SKST = Suomalaisen Kirjallisuuden Seuran toimituksia
- SORJONEN, MARJA-LEENA – ROUHIKOSKI, ANU – LEHTONEN, HEINI 2015. *Helsingissä puhuttavat suomet. Kielen indeksisyys ja sosiaaliset identiteetit*. SKST 1310. Helsinki: Suomalaisen Kirjallisuuden Seura.
- SORSAKIVI, MERJA 1982. Infinitiivijärjestelmän muutoksia lasten kielessä. – *Virittäjä* 86 s. 377–391.
- STUDENT 1908. The probable error of a mean. – *Biometrika* 6 s. 1–25.
- STURTEVANT, EDGAR H. 1917. *Linguistic change. An introduction to the historical study of language*. Chicago: Chicago University Press.
- SUIHKONEN, PAAVO 1992. Klusiilien vaihtelusuhteet Kala- ja Lestijokilaakson murteissa. SKST 577. Helsinki: Suomalaisen Kirjallisuuden Seura.
- SUOJANEN, MATTI K. (toim.) 1982a. *Kirjoituksia pubekielestä*. Toinen painos. Turun puhekielen projektin julkaisuja 1. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 14.
- SUOJANEN, MATTI K. 1982b. Koululaisslangista. – Maija Larmola (toim.), *Kouluikäisten kieli* s. 153–164. Tietolipas 88. Helsinki: Suomalaisen Kirjallisuuden Seura.
- SUOJANEN, MATTI K. 1982c. Sosiolingvistiikka – synkretistinen tieteenala. – Matti K. Suojanen & Päivikki Suojanen (toim.), *Sosiolingvistiikan näkymiä* s. 7–37. Helsinki: Gaudeamus.
- SUOJANEN, MATTI K. (toim.) 1985. *Mitä Turussa puhutaan? Raportti Turun puhekielen tutkimuksesta*. Turun puhekielen projektin julkaisuja 3. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 23.
- TAGLIAMONTE, SALI 2012. *Variationist sociolinguistics. Change, observation, interpretation*. Oxford: Wiley-Blackwell.
- TAGLIAMONTE, SALI – BAAYEN, HARALD 2012. Models, forests and trees of York English. – *Language Variation and Change* 24 s. 135–178.
- THELANDER, MATS 1979a. *Språkliga variationsmodeller tillämpade på nutida Burträskedal. Del 1. Materialbeskrivning och analys påspråkdragnivå*. Studia Philologiae Scandinavicae Upsaliensia 14: 1. Uppsala.
- THELANDER, MATS 1979b. *Språkliga variationsmodeller tillämpade på nutida Burträskedal. Del 2. Teoretiska utgångspunkter och analys på språkartsnivå*. Studia Philologiae Scandinavicae Upsaliensia 14: 2. Uppsala.
- THOMPSON, KEN 1968. Regular expression search algorithm. – *Communications of the ACM* 11 s. 419–422.
- UUSITUPA, MILLA 2017. *Rajakarjalaismurteiden avoimet persoonaviittaukset*. Dissertations in Education, Humanities, and Theology 117. Joensuu: Publications of the University of Eastern Finland.
- VAATTOVAARA, JOHANNA 2002. Jälkitavujen h:n variaation kehityksestä Tornionjokilaaksoissa. – *Virittäjä* 106 s. 508–535.

- VAATTOVAARA, JOHANNA 2009. *Meän tapa puhua. Tornionlaakso pellolaisnuorten subjektiivisena paikkana ja murrealueena*. SKST 1224. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VAATTOVAARA, JOHANNA & SOININEN-STOJANOV, HENNA 2006. Pääkaupunkiseudulla kasvaneiden kotiseuturajaukset ja kielelliset asenteet. – Kaisu Juusela & Katariina Nisula (toim.), *Helsinki kieliyhteisönä* s. 223–255. Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.
- VIRTARANTA, PERTTI 1958. *Pääpainollisen tavun jälkeisen soinnillisen dentaalispirantin edustus suomen murteissa*. SKST 252. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VIRTARANTA, PERTTI 1966. Suomen kielen ja sen sukukielten kielennäisten sodanjälkeisestä tallennustyöstä Suomessa. – *Virittäjä* 70 s. 180–188.
- VUOLLE, MARJUT 2016. *”Et etutööläläisenä se Taka-Töölöki oli vähä ku ulkomaille menis”. Töölön paikannimet 1970-, 1990- ja 2010-luvuilla*. Suomen kielen pro gradu -tutkielma. Helsingin yliopisto.
- VUOLTEENAHO, JANI – LAPPALAINEN, HANNA – AINIALA, TERHI 2019. A slightly slummier area? Negotiations of place-bound identities through social spatializations and unofficial toponyms. – *Journal of Sociolinguistics* 23 s. 140–162.
- VUORI, PEKKA – MÄKI, NETTA – BERGROTH, CLAUDIA 2020. Helsingin väestö vuodenvaihteessa 2019/2020 ja väestömuutokset vuonna 2019. – *Tilastoja* 2020:10. https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/20_09_17_Tilastoja_10_Vuori_Maki_Bergroth.pdf (16.2.2021).
- VÄÄNÄNEN, MILJA 2016. *Subjektin ilmaiseminen yksikön ensimmäisessä persoonassa. Tutkimus suomen vanhoista murteista*. Annales Universitatis Turkuensis C430. Turku: Turun yliopiston julkaisuja.
- WAGNER, SUZANNE EVANS 2012a. Age grading in sociolinguistic theory. – *Language and Linguistics Compass* 6 s. 371–382.
- WAGNER, SUZANNE EVANS 2012b. Real-time evidence for age grad(ing) in late adolescence. – *Language Variation and Change* 24 s. 179–202.
- WAGNER, SUZANNE EVANS – SANKOFF, GILLIAN 2011. Age grading in the Montréal French inflected future. – *Language Variation and Change* 23 s. 275–313.
- WALKER, ABBY – GARCÍA, CHRISTINA – CORTÉS, YOMI – CAMPBELL-KIBLER, KATHRYN 2014. Comparing social meanings across listener and speaker groups: The indexical field of Spanish /s/. – *Language Variation and Change* 26 s. 169–189.
- WANG, WILLIAM S-Y. 1969. Competing Changes as a Cause of Residue. – *Language* 45 s. 9–25.
- WANG, WILLIAM S-Y. – CHENG, CHIN-CHUAN 1970. Implementation of phonological change: the Shuang-Fêng case. – *Chicago Linguistic Society* 6 s. 552–559.
- WARIS, HEIKKI 1950. Helsingiläisyhteiskunta. – *Helsingin kaupungin historia* 3. osa, 1. nide. Helsinki: Helsingin kaupunki.
- WARIS, HEIKKI 1973. *Työläisyhteiskunnan syntyminen Helsingin Pitkänsillan pohjoispuolelle*. Toinen, tarkistettu painos. Helsinki: Weilin+Göös.
- WEINREICH, URIEL – LABOV, WILLIAM – HERZOG, MARTIN 1968. Empirical foundations for a theory of language change. – Lehmann, Winfred P. & Yakov Malkiel (toim.), *Directions for Historical Linguistics: A Symposium* s. 95–188. Austin: University of Texas Press.
- WILCOXON, FRANK 1945. Individual comparisons by ranking methods. – *Biometrics Bulletin* 1 s. 80–83.

YLIKOSKI, JUSSI 2005. Puhekielen morfologisten ja semanttisten innovaatioiden tutkimusnäkymiä – esimerkkinä suomen *tekeen-* ja *tekees-*tyyppiset verbimuodot. – *Puhe ja kieli* 25 s. 187–209.

JULKAISUT

JULKAISU I: INFINITIIVIT

Infinitiivien variaatio ja muutos Helsingissä

Kuparinen, Olli

Virittäjä, 122(1), 29–51
<https://doi.org/10.23982/vir.65310>

Artikkelin käyttöön väitöskirjan osana on saatu kustantajan lupa

Infinitiivien variaatio ja muutos Helsingissä



OLLI KUPARINEN

1 Johdanto

Lähettiin pämppää / pistettiin lääväks pari kämppää lauloi helsinkiläinen Teflon Brothers -yhtye vuoden 2015 kesähitissään. Lyhyessä otteessa on useita variaationtutkijoita kiinnostavia piirteitä, mutta tässä kirjoituksessa huomioni kohteena on infinitiivi, joka näyttää menettäneen sekä tunnuksensa että sijansa. Aineistona artikkelissa on käytetty Helsingin puhekielen pitkittäiskorpusta (Helpuhe 2014), joka koostuu 1970-, 1990- ja 2010-luvuilla tehdyistä haastatteluista. Infinitiiveistä tarkastelen kahta yleisintä: A-infinitiivin historiallista latiivia ja MA-infinitiivin illatiivia. Jonkin verran kiinnitän huomiota myös *käydä-* ja *olla-*verbien inessiivitäydennyksiin.¹

Infinitiivit ovat kiinnostaneet suomen kielen tutkijoita niin dialektologian kuin sosiolingvistiikan perinteessä. Erityisesti MA-infinitiivin illatiivimuodot ovat kuuluneet sekä aluemurteiden että kaupunkikielten tutkimuskohteisiin. Tunnuksellisten ja tunnuksettomien muotojen vaihtelua on tarkasteltu etenkin hämäläismurteiden alueella (Saukkonen 1965: 50–55; Itkonen 1989; Mustanoja 2011). Myös Helsingin puhekielen tutkimuksessa on kiinnitetty jonkin verran huomiota infinitiiveihin (Sorsakivi 1982; Paunonen 1995 [1982]; Lappalainen 2004). Viime vuosikymmeninä on julkaistu infinitiivejä käsittelevää tutkimusta toisaalta kieliopin (Herlin & Visapää toim. 2005; Visapää 2008), toisaalta vanhan kirjakielen (Leino 2003) näkökulmasta. Kattavaa kuvausta infinitiivien variaatiosta puhekielessä ei kuitenkaan ole esitetty sitten Saukkosen (1965), minkä vuoksi tutkimukselleni on selvä tilaus.

Helsingin puhekielessä on viime vuosina yleistynyt lyhyt infinitiivimuoto, joka voi toimia sekä A-infinitiivin että MA-infinitiivin tehtävissä: *pitää kertoa* (A-infinitiivi) – *mennä kertoa* (MA-infinitiivi). Tässä artikkelissa tarkastelen tällaisten yhteenlankeamisten syitä, seurauksia ja infinitiivien keskinäistä variaatiota Helsingin puhekielessä neljän vuosikymmenen ajanjaksolla. Tapahtuneita muutoksia analysoin

1. Artikkelini perustuu pro gradu -tutkielmaani (Kuparinen 2015) ja on kirjoitettu Koneen Säätiön apurahalla. Kiitän Unni-Päivä Leinoa, Liisa Mustanojaa, Jaakko Peltosta ja Jenni Santaharjua käsikirjoituksen kommentoinnista. Työn varhaisessa vaiheessa olen saanut apua myös Markus Hamuselta, mistä esitän kiitokset. Lisäksi kiitän Virittäjän kahta nimetöntä arvioijaa, joiden huomiot ovat parantaneet huomattavasti kirjoituksen laatua.

muutosmallien (alaluku 2.2) avulla, mikä on suomalaisessa variaation tutkimuksessa ollut melko vähäistä (Kurki 2013: 211).

2 Ilmiön tausta ja aineisto

Tässä luvussa esittelen tutkimuskohteet ja tutkimuksen hypoteesit ja kysymykset. Jotta esittämäni tulokset asettuisivat paremmin kontekstiinsa ja olisivat lukijan helposti seurattavissa, esittelen infinitiivisysteemin taustaa laajasti alaluvussa 2.1. Vastaavasti tulosten seuraamisen helpottamiseksi kuvaan alaluvussa 2.2 lyhyesti teoriaa muutosmalleista. Alaluvussa 2.3 esittelen tutkimuksessa käytetyn aineiston ja sitä koskevat valinnat.

2.1 Tutkimuskohteet ja -kysymykset

Artikkeli keskittyy suomen kahden yleisimmän infinitiivimuodon tarkasteluun. A-infinitiivin lyhyitä muotoja (latiiveja) on Lauseopin arkiston murreaineistossa 6,72 esiintymää 1 000 tekstisanaa kohden ja MA-infinitiivin illatiiveja vastaavasti 5,73 (Herlin, Leino & Visapää 2005: 17). Nämä kaksi infinitiivimuotoa ovat tehtäviltään samantapaisia, sillä molemmat ovat olleet historiallisesti tulosjaisia. Myöhemmin A-infinitiivin latiiivisuus on hävennyt ja se on liittynyt ennen kaikkea abstraktien verbien täydennykseksi (esim. *voida*). MA-infinitiivin illatiivi toimii puolestaan tyypillisesti sijansa mukaisesti konkreettisempien, paikallissijaista täydennystä vaativien verbien kanssa (esim. *mennä*). Kahden infinitiivin² järjestelmä on yleinen itämerensuomalaisissa kielissä, ja näistä *t*-aineiden sisältänyt infinitiivi (suomessa siis A- ja E-infinitiivi) on vanhempi kuin MA-infinitiivi (Saukkonen 1965: 50). Joidenkin verbien infinitiivitäydennyksen laatu vaihtelee murretaustan mukaan (esim. *alkaa, pystyä, joutua, ehtiä*). Tunnetuin näistä verbeistä lie *alkaa*, joka länsimurteissa on saanut täydennyksekseen tyypillisesti MA-infinitiivin illatiivin (*alkaa tekemään*) ja itämurteissa puolestaan A-infinitiivin latiiivin (*alkaa tehdä*). Variaatiolle on nähty myös semanttisia syitä (Herlin 2012). (Infinitiivimuotojen historiasta ks. Saukkonen 1965: 237–247; 1966: 112.)

Vaikka kahdella tulosjainfinitiivillä onkin päällisin puolin melko selkeä työnjako, on systeemi silti osittain päällekkäinen. Esimerkiksi pohjoissaamassa käytetään tyypillisesti yhtä perusinfinitiiviä monissa eri tehtävissä eikä MA-infinitiiviä vastaava verbimuoto ole käytössä lainkaan (Jomppanen 2009: 23; Saukkonen 1965: 50). Vastaavasti monissa indoeurooppalaisissa kielissä on vain yksi tulosjainen infinitiivi. Jaakko Leino onkin esittänyt (2003: 237), että suomessa saattaisi olla meneillään tendenssi kohti yhtä yleisinfinitiiviä indoeurooppalaisten kielten tapaan. Tutkimukseni tukee tätä havaintoa ainakin infinitiivien muodon kannalta: muutamaiset yleiset ääntenmuutokset johtavat suomen kahta yleisinfinitiiviä yhteen, mitä esittelen tau-

2. Suomen kielioppiperinteen viidestä infinitiivistä on nyttemmin yleisesti luovuttu ja infinitiivit luokitellaan yleensä kolmiksi: A-, E- ja MA-infinitiiviin (VISK § 120). Näistä kuitenkin A- ja E-infinitiivillä on vanhastaan yhteinen *t*-aines (Saukkonen 1965), joten voidaan puhua kahdesta infinitiivistä. Siro (1964) on puolestaan esittänyt, että syntaktisesti suomessa on vain yksi infinitiivi. Selvää kuitenkin on, että morfologisesti infinitiiviä merkitään itämerensuomalaisissa kielissä vähintään kahdella eri tavalla.

lukossa 1. Kieliopissa mahdollisesti tapahtuvia muutoksia ja yhden infinitiivin teoriaa pohdin myöhemmin luvussa 4.

Taulukko 1.
Suomen verbityypit eri infinitiivimuodoissa.

Verbi- tyyppi	A-infinitiivin latiivi	MA-infinitiivin illatiivi	MA-tunnukseton illatiivi	Lyhyt infinitiivi
1. AA, OA, eA, UA, iA	<i>antaa</i> <i>kertoa</i>	<i>antamaan</i> <i>kertomaan</i>	<i>antaan</i> <i>kertoon</i>	<i>antaa</i> <i>kertoo</i>
2. dA	<i>saada</i> <i>imuroida</i>	<i>saamaan</i> <i>imuroimaan</i>	<i>*saan</i> <i>*imuroiin</i>	<i>*saa</i> <i>*imuroi</i>
3. tA, lA, rA, nA	<i>nousta</i> <i>tulla</i>	<i>nousemaan</i> <i>tulemaan</i>	<i>nouseen</i> <i>tuleen</i>	? <i>nousee</i> ? <i>tulee</i>
4. (t)A	<i>pelata</i>	<i>pelaamaan</i>	<i>pelaan</i>	<i>pelaa</i>
5. tarvita	<i>tarvita</i>	<i>tarvitsemaan</i>	<i>tarvitseen</i> <i>tarviin/tartteen</i>	? <i>tarvitsee</i> ? <i>tarvii/tarttee</i>
6. vanheta	<i>vanheta</i>	<i>vanhenemaan</i>	<i>vanheneen</i> ? <i>vanheen</i>	? <i>vanhenee</i> ? <i>vanhee</i>

Taulukossa 1 olen esittänyt suomen verbien muotoryhmät ja niiden variantit A-infinitiivissä, MA-infinitiivin illatiivissa, MA-tunnuksettomassa illatiivissa ja lyhyessä infinitiivissä. Infinitiivien päätyminen lyhyiksi infinitiiveiksi voidaan näin johtaa muuttaman äänne muutoksen kautta:

- a) A-loppuisten vokaaliyhtymien assimiloituminen (*kertoa* > *kertoo*)
- b) MA-infinitiivin illatiivien MA-tunnuksettomuus (*antamaan* > *antaan*)
- c) sananloppuisen *n*-äänne kato (*antaan* > *antaa*)
- d) supistumaverbien analogiset muodot (*pelata* > *pelaa*)

Näistä ensimmäinen (a) on Mantilan (2004: 326–327) havaintojen mukaan yleisty-mässä koko maassa, ja samansuuntaisia arvioita on partitiivimuotojen (esim. *palloa* > *palloo*) osalta esittänyt myös Paunonen (1995 [1982]: 110–113). Vokaaliyhtymien välillä on kuitenkin eroa: herkimmin pitkävokaalisina esiintyvät *eA* ja *OA*, kun taas *UA* ja *iA* ovat perinteisesti säilyneet (ks. esim. mts. 106–127). Kohta b puolestaan voidaan yhdis-tää hämäläismurteiden vaikutukseen, joka Helsingissä onkin ollut suuri (mts. 13–15). Nuolijärvi ja Sorjonen (2005: 97) esittävät tämän piirteen olevan myös ”hyvin ekspan-siivinen tyyppi Etelä-Suomen kaupunkien ja erityisesti nuorten kielessä”. Viimeinen (d) on vanhastaan etelähämäläisissä murteissa tavattu analoginen muoto, joka Itko-sen (1964: 187–193) mukaan on seurausta ruotsinkielisten ja suomenkielisten puhujien yhteiselosta (ks. myös Itkonen 1989: 364–365). Muoto on säilynyt Helsingissä etenkin slangissa (Paunonen 2006: 37).

Sananloppuisen *n*-äänteen kato (c) ei ole yhtä yksiselitteinen kuin muut esiteltyt ääntenmuutokset. Absoluuttisessa lopussa kato on hyvin tyypillinen (Rapola 1966: 325) ja niin ikään sellaisissa tapauksissa, joissa se on osa taivutustunnusta.³ Nasaalin ja klusililn edellä *n* kuitenkin usein assimiloituu, ja assimilaatio voi tapahtua myös likvidan, *j:n* tai *v:n* edellä (VISK § 39). Vokaalien edellä *n* on Rapolan (mts. 327) mukaan yleisesti säilynyt. Tulosijainfinitiivejä helsinkiläislasten puheesta tutkinut Sorsakivi (1982: 383–384) kuitenkin esittää, että illatiivimuodoissa sananloppuinen *n* ei katoa, vaan se liittyy sananloppuisen *k:n* sandhikaavaan. Äänteen käyttäytymiseen sananrajalla ja mahdollisiin ongelmiin palaan vielä alaluvuissa 3.1 ja 3.2.

Taulukosta 1 huomataan myös, että kaikissa verbityypeissä edellä mainitut ääntenmuutokset eivät ole mahdollisia (tyyppi 2) tai verbin vartalosta johtuen A-infinitiivi ja MA-infinitiivin illatiivi päätyvät ääntenmuutosten jälkeen eri asuihin (tyypit 3, 5 ja 6).⁴ *dA*-verbit eivät voi menettää infiniitiivin tunnusta sen paremmin A-infinitiivissä kuin MA-infinitiivissäkään, minkä takia olen merkinnyt näitä muotoja asteriskilla. Konsonanttivartaloisten ryhmä (tyyppi 3) sen sijaan päättyy MA-infinitiivin illatiivia vaativissa tapauksissa herkästi lyhyeen asuun (*meen juoksee*) mutta toimii lyhyenä infiniitiivinä harvoin A-infinitiivitäydennystä vaativissa tehtävissä (*voi juoksee*). Sentyypiset tapaukset kuin *pitää saa* ja *pitää juoksee* eivät siis vaikuta kovin todennäköisiltä. Tällaisia lipsahduksiksi tulkitsemiani esiintymiä on käyttämässäni aineistossa (Helpuhe 2014) vain muutamia. Tämä aiheuttaa haasteita aiemmin esittelemäleni ajatukselle yhden yleisinfinitiivin systeemistä, sillä *dA*-verbien ja konsonanttivartaloisten verbien osuus on tämän tutkimuksen kaikista infiniitiiviesiintymistä noin 30 %. Palaan tähän alaluvussa 4.2.

Olen edellä esiteltyt yleisinfinitiivien historiallista taustaa ja ääntenmuutoksista johtuvaa yhteenlankeamista. Tähän taustaan peilaten tutkin sitä, millaista yleisinfinitiivien sisäinen variaatio on ja miten se on muuttunut vuosikymmenten välillä. Toisekseen lyhyen infiniitiivin muoto ja merkitys kaipaavat vielä selvittämistä: kytkeytyykö lyhyt infiniitiivi A-infiniitiiviin, MA-infiniitiiviin vai yksikön kolmannen persoonan muotoon? Näitä seikkoja tarkastelen luvussa 3. Viimeisenä kysymyksenä (alaluku 4.2) pohdin koko infiniitiivijärjestelmää: onko edellä esitetyissä ääntenmuutoksissa koko järjestelmän muutoksen siemen? Seuraavaksi esittelen kuitenkin mallit, joilla muutosta on kuvattu variaationtutkimuksessa.

2.2 Muutosmallit

Sosiolingvistiikan keskeisiä tutkimuskohteita on meneillään oleva kielenmuutos. Sitä on pyritty tarkastelemaan näennäisaikametodilla ja pitkittäisaineistojen valmistumisen myötä myös reaaliajassa (ks. esim. Kurki 2005: 31–37; Laasanen 2016). Keskeinen ero menetelmien välillä on se, että jälkimmäisen avulla voidaan tarkastella muutosta

3. Lehtonen (2006: 269–270) on kuitenkin havainnut, että 2000-luvun helsinkiläisnuorilla loppu-*n* katoaa usein myös verbin persoonapäätteestä (esim. *mä sano*).

4. Tyyppien 5 ja 6 esiintymiä on aineistossani niin vähän, että jätän ne tässä kokonaan tarkastelun ulkopuolelle.

diakronisesti, kun taas näennäisaikamenetelmä tarkkailee yhtä hetkeä synkronisesti ja tekee sen perusteella ennusteita muutoksen suunnasta. Tämä perustuu pitkälti ajatuksen idiolektien muuttumattomuudesta ja niin sanottuun kriittisen iän hypoteesiin. Hypoteesin mukaan kielen oppiminen on helpointa tietyssä ikävaiheessa (lapsena tai viimeistään varhaisnuorena); tämän kriittisen iän jälkeen idiolektin ei enää katsota olenaisesti muuttuvan. (Laasanen 2016: 64–67; Meyerhoff 2011: 141.)

Toinen muutoksen tutkimuksen keskeisistä eronteista liittyy puhujayhteisön ja yksilöiden väliseen suhteeseen. Idiolektien muutos voi poiketa yhteisön muutoksen suunnasta, tai nämä muutokset voivat olla samansuuntaisia. Näitä erilaisia vaihtoehtoja on esitellyt Labov (1994: 83) nelikentässä, jonka on suomeksi esittänyt Paunonen (2005; ks. myös Kurki 2013). Reaaliaikaisen aineiston avulla saadun informaation perusteella nelikenttää on vielä täydennetty elinikäisellä muutoksella (Sankoff 2005). Erilaiset muutosmallit on koottu taulukkoon 2.

Taulukko 2.

Muutosmallit (Labov 1994; Paunonen 2005; Sankoff 2005; Kurki 2013).

	Yksilö	Yhteisö
1. Vakaus	vakaa	vakaa
2. Ikäsidonnainen muutos	muuttuva	vakaa
3. Sukupolvittainen muutos	vakaa	(asteittain) muuttuva
4. Yhteisöllinen muutos	muuttuva	muuttuva
5. Elinikäinen muutos	muuttuva	(asteittain) muuttuva

Tämän tutkimuksen tulosten osalta (luku 3) yksilön ja yhteisön suhde vaihtelee: tilanne ei ole vakaa, eikä muutos tapahdu myöskään yhteisöllisesti. Niinpä tutkimuksen kannalta keskeisiä muutosmalleja ovat ikäsidonnainen muutos, sukupolvittainen muutos ja elinikäinen muutos. Ikäsidonnaisessa muutoksessa on kyse siitä, että tietyssä elämäntavaiheessa tai iässä yksilön kieli muuttuu hetkellisesti. Kun tämä vaihe on ohitettu, vaiheelle ominaiset piirteet heikkenevät tai väistyvät (Kurki 2013: 209). Näin koko yhteisön kielitilanne pysyy vakaana. Sukupolvittaisella muutoksella tarkoitetaan puolestaan tilannetta, jossa eri sukupolvien kielenkäyttö eroaa toisistaan eikä yksilöiden kieli muutu. Näin ajan kuluessa koko yhteisön kieli muuttuu. Ikäsidonnainen muutos ja sukupolvittainen muutos näyttävät yhdestä hetkestä tarkasteltuna samalta: yksittäinen sukupolvi eroaa muista sukupolvista. Elinikäinen muutos perustuu sukupolvittaiseen muutokseen, mutta sen sijaan, että idiolektit säilyisivät ennallaan, ne muuttuvat yhteisössä tapahtuvan sukupolvittaisen muutoksen suuntaan (Sankoff 2005; Kurki 2013). Wagner ja Sankoff (2011: 303–306) ovat kuitenkin muistuttaneet myös niin sanotusta vetäytyvästä elinikäisestä muutoksesta (engl. *retrograde lifespan change*)⁵, jossa idiolektien muutos onkin yhteisössä tapahtuvaan muutokseen nähden päinvastainen. Käsitellen muutosmalleja ja niiden sopimista infinitiivien muutoksen tarkasteluun tulosten yhteydessä (luku 3).

5. Suomennos Laasaselta (2016: 66).

2.3 Aineisto ja menetelmät

Tutkimuksen aineistona käytän Helsingin puhekielen pitkittäiskorpusta (Helpuhe 2014). Aineisto on kerätty kolmella vuosikymmenellä (1972–1974, 1991–1992, 2013) ja sijoitettu myöhemmin osaksi CSC:n Kielipankkia.⁶ Informantit on ensimmäisellä kierroksella valittu kahdesta kaupunginosasta (Töölö ja Sörnäinen), kolmesta sosiaaliluokasta ja kolmesta ikäluokasta ja jaettu jotakuinkin tasan miehiin ja naisiin. Kaupunginosa- ja sosiaaliluokitus on myöhemmillä kierroksilla yhdistetty osoittamaan lähinnä informanttien koulutustaustaa.

Tässä tutkimuksessa olen tarkastellut taustamuuttujista vain ikää, koska sen on aiemmissa tutkimuksissa katsottu korreloivan taustamuuttujista voimakkaimmin kielien kanssa (esim. Lappalainen 2001; Buchstaller 2015). Kriittikinä voidaan esittää, että aineistoon luodut ikäryhmät ovat väistämättä heterogeenisiä joukkoja, joita tutkittaessa yksittäisten puhujien idiolektit jäävät varjoon (Lappalainen 2004: 15; vrt. Kurki 2007: 150–151). Tutkimuksessa pyrin kuitenkin selvittämään ennen kaikkea koko infinitiivijärjestelmän tilannetta Helsingissä. Näin ollen ryhmien koostumukset eivät ole tässä tutkimuksessa ensisijaisia, vaan tarkoituksena on tutkia yksittäisiä piirteitä ja pyrkiä havaitsemaan niissä tapahtuvia muutoksia.

Olen poiminut ja luokitellut käsin aineistosta havaitsemani infinitiiviesiintymät. Ikäryhmien ja vuosikymmenten eroja olen tarkastellut tilastollisesti khiin neliö -testin (Pearson 1900) avulla. Testin nollihypoteesissa oletetaan, että vertailtavat ryhmät eivät eroa toisistaan. Otoksesta laskettu p-arvo on todennäköisyys sille, että havaittu arvo toteutuisi ilman eroa perusjoukossa eli selittyisi pelkästään otantavirheellä. Mitä pienempi p-arvo siis on, sitä enemmän tukea saadaan nollihypoteesin hylkäämiselle. Yleisesti rajana pidetään p-arvoa 0,05. Tutkimuksen tuloksissa esitän p-arvon lisäksi myös vapausasteiden lukumäärän (sulkeissa) ja testimuuttujan arvon (yhdyserkin jälkeen) seuraavasti: $X^2(2) = 251,79, p < 0,001$. Khiin neliö -testiä hyödyntämällä aineistossa kenties muutenkin havaittavat muutokset saavat tilastollista tukea. Tätä kautta tutkimus kiinnittyy suomalaisen sosiolingvistiikan kvantitatiivisia metodeja painottaaviin tutkimuksiin (esim. Kurki 2005; Mustanoja 2011; Priiki 2017).

Eri vuosikymmenillä kerätyt aineistot eivät määrällisesti täysin vastaa toisiaan, sillä 1970-luvun osalta tutkimus on kohdistunut 96 informantista koostuvaan ydinaineistoon, 1990-luvulla haastatteluja tehtiin 45 ja 2010-luvulta käytettävissäni oli 44 informantin haastattelu. Tästä aineistosta on litteroituna julkaistu vasta osa. 1990-luvun ja 2010-luvun aineistosta litteraatteja ei tätä tutkimusta tehdessäni ollut saatavilla vastaavassa määrin kuin 1970-luvun aineistosta. Näin ollen päädyin käyttämään vain osaa 1970-luvun aineistosta ja täydentämään 1990- ja 2010-luvun informanttien määrää kuuntelemalla haastatteluja ja poimimalla niistä infinitiiviesiintymät. 1990-luvun aineistosta olen kuunnellut 14 haastattelua ja 2010-luvun aineistosta 11. Tutkimukseen eri vuosikymmeniltä

6. Hankkeen käynnistäjänä toimi Terho Itkonen, mutta pääasiallisen vastuun 1970- ja 1990-lukujen aineistonkeruusta kantoi Heikki Paunonen, jonka ideasta syntyi myös viimeisin kierros. Vuodesta 2013 projektia on johtanut Hanna Lappalainen. Haastattelijoina ovat toimineet kaikilla vuosikymmenillä Helsingin yliopiston suomen kielen opiskelijat.

valittujen haastattelujen määrä on koottu taulukkoon 3. Aineisto koostuu kolmen ikäryhmän informanteista: noin 20-vuotiaista nuorista, noin 40-vuotiaista keski-ikäisistä ja noin 60-vuotiaista vanhoista puhujista. Lisäksi 2010-luvun keruussa on haastateltu neljää noin 80-vuotiasta puhujaa, jotka ovat kuuluneet 1990-luvun keruussa 60-vuotiaiden ryhmään. Nämä neljä puhujaa olen laskenut osaksi 2010-luvun vanhojen ryhmää. Pysin valitsemaan jokaiselta vuosikymmeneltä jokseenkin saman määrän eri-ikäisiä puhujia, mutta ryhmien välillä on kuitenkin pieniä eroja. Vertailen infinitiiviesiintymiä prosentuaalisesti, joten ryhmien kokoerot eivät vaikuta tuloksiin.

Taulukko 3.
Informanttien määrä.

Vuosikymmen	Vanhat (60-v.) ⁷	Keski-ikäiset (40-v.)	Nuoret (20-v.)	Yhteensä
1970-luku	22	20	23	65
1990-luku	13	13	15	41
2010-luku	13 (9 + 4)	14	17	44
Yhteensä	48	47	55	150

Toinen huomionarvoinen seikka informanttien valinnassa on se, että 1990-luvun ja 2010-luvun vanhojen ja keski-ikäisten ryhmät koostuvat jo aiemmilla kierroksilla haastatelluista informanteista, ja ainoastaan nuorten ryhmät ovat joka kierroksella uusia. Kyseessä on siis trendi- ja paneelitutkimuksen yhdistelmä (esim. Laasanen 2016: 60). Tämä antaa erinomaiset mahdollisuudet näennäisaikahypoteesin (esim. Chambers & Trudgill 1998: 149–151; Laasanen 2016) tarkasteluun mutta antaa myös pohdittavaa, koska aiemmin haastateltuja informantteja ei ole tutkimuksessa erikseen eritelty joukosta ja tarkasteltu idiolektitasolla.

3 Infinitiivien variaatio ja muutos

Tässä luvussa esittelen tutkimuksen keskeiset tulokset. Aloitan tarkastelun infinitiivien kokonaisvariaatiosta diakronisesti (ks. alaluku 3.1). Seuraavaksi (3.2) esittelen A-loppuiseen vokaaliyhtymään päättyvien infinitiivimuotojen variaatiota ja muutosta. Lopuksi tarkastelen sananrajailmiöitä (3.3) ja *käydä*- ja *olla*-verbien täydennyksiä (3.4).

3.1 Infinitiivien variaatio ja muutos 1970-luvulta 2010-luvulle

Vertailuja varten jaoin alaluvussa 2.1 esitellyt infinitiivit kolmeen yleisinfinitiiviin: lyhyisiin infinitiiveihin, A-infinitiiveihin ja illatiivitapauksiin. Lyhyiden infinitiivien joukko koostuu pitkään vokaaliin päättyvistä, useampitavuisista muodoista: *antaa*,

7. 2010-luvun vanhojen ryhmään kuuluu yhdeksän n. 60-vuotiasta ja neljä n. 80-vuotiasta puhujaa.

kertoo, lukee, istuu, leikkii, pelaa, parantaa, luistelee. A-infinitiivien ryhmään kuuluvat standardinmukaiset A-infinitiivit (esim. *kertoa, lukea, istua, leikkiä, pelata, juosta, saada*) ja illatiivitapauksiin puolestaan kaikki muodot, joissa esiintyy joko MA-tunnus tai illatiivin sananloppuinen *n* (*antamaan, antamaa, antaan*).

Jotkin tehdyistä valinnoista ovat tulkinnanvaraisia ja vaativat siksi perusteluja. *Antaa*-tyyppi on laskettu lyhyisiin infiniitveihin, vaikka se on standardinmukainen A-infinitiivi. Se vastaa kuitenkin muodoltaan lyhyttä infiniitviä, ja sen mallivaikutus voidaan nähdä esimerkiksi Helsingin puhekielen supistumaverbeissä (Itkonen 1964); vastaava analogia koskee esimerkiksi partitiiveja (alaluku 3.2; Paunonen 1995 [1982]: 126). Toisekseen on syytä palata sananloppuisen *n*-äänteen käyttäytymiseen sananrajalla (ks. alalukua 2.1). Nyt tehdyllä jaottelulla esimerkiksi seuraavat kuvitteelliset tapaukset asetuvat kyseenalaisiksi:

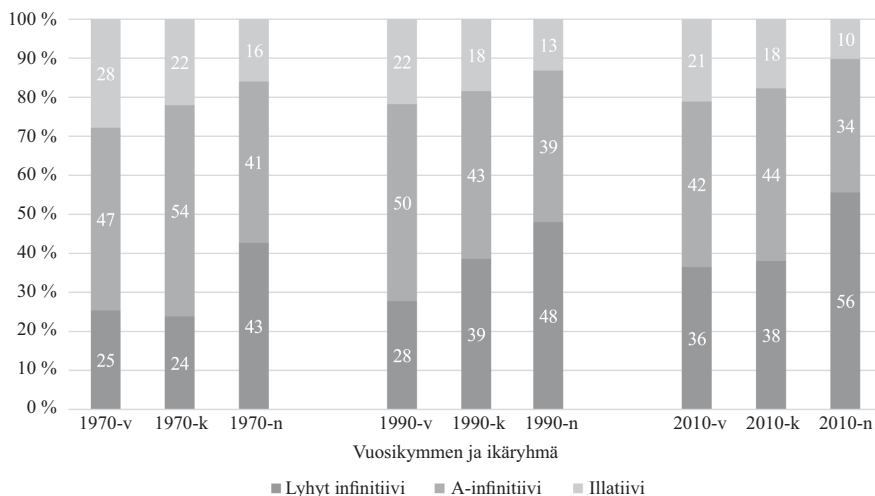
- (a) *Mennään pelaamaan niille.*
- (b) *Mennään pelaan niille.*
- (c) *Mennään pelaa.*

Näistä c-kohdan lyhyt infiniitvi voi selittyä sen asemalla absoluuttisessa lopussa, kun taas b-kohdan *pelaan*-muodossa nasaalia edeltävä *n* voidaan nähdä joko assimilaatio-ilmionä tai illatiivin päätteenä. Mikäli c-kohdan *pelaa* ei olisikaan absoluuttisessa lopussa, vaan sitä seuraisi esimerkiksi nasaali tai klusiili, se voisi käyttäytyä täysin b:n tavoin. On selvää, että näistä esimerkki b on muodoltaan läheisempi esimerkin c kuin esimerkin a kanssa, mutta edellä esitellyssä jaottelussa b päätyykin samaan ryhmään MA-tunnuksellisen illatiivin kanssa (a). Olen kuitenkin päätenyt edellä esitettyyn jaotteluun siitä syystä, että päätte *-VVn* on muodoltaan selvästi illatiivi, vaikka siinä ei olisikaan MA-tunnusta. Tästä syystä en kuitenkaan käytä nimitystä MA-infiniitviin illatiivi, vaan puhun yleisemmin illatiiveista (vaikka joukkoon mahtuu toki tapauksia, joissa on MA-tunnus eikä vuorostaan kokonaista illatiivin tunnusta). MA-tunnuksettomien illatiivimuotojen (*antaa*) osuus koko aineistosta on noin 1,8 %, joten niiden lajittelun vaikutus on lopulta vähäinen.

Infinitiivien kokonaisvariaatio esiteltyjen kolmen infiniitviin osalta kaikilla vuosikymmenillä ja kaikissa ikäryhmissä on esitetty kuviossa 1 (ks. seur. sivu). Olen poiminut kaikilta vuosikymmeniltä yhteensä 10 057 infiniitviesiintymää, joita vertailen tässä luvussa prosentuaalisesti. Kuviossa 1 voidaan havaita muutama hyvin selvä tendenssi. Lyhyt infiniitvi on jokaisella vuosikymmenellä yleisin infiniitviimuoto nuorten puheessa. Vastaavasti illatiivimuodot vähenevät jokaisella vuosikymmenellä vanhemmista nuorempiin siirryttäessä. Eri ikäryhmien välillä on kaikilla vuosikymmenillä ja kaikissa infiniitviimuodoissa tilastollisesti erittäin merkitseviä eroja ($p < 0,001$).⁸ Vaikka erot ovat merkitseviä kaikissa infiniitviimuodoissa, suurimmat ne ovat vertailtaessa lyhyttä infiniitviä muihin infiniitviimuotoihin ja toiseksi suurimmat illatiivissa. Ikäsidonnainen muutos (ks.

8. Khiin neliö -testin mukaan lyhyen infiniitviin ja muiden infiniitviimuotojen välillä on erittäin merkitseviä eroja kaikkia kolmea ikäryhmää vertailtaessa ($2(2) = 251,79, p < 0,001$). A-infiniitviin osalta lueumat ovat $2(2) = 60,20, p < 0,001$ ja illatiivin osalta $2(2) = 108,18, p < 0,001$.

alalukua 2.2) näyttääkin koskettavan juuri näitä kahta infinitiivimuotoa. A-infinitiivin osalta erot ovat pienempiä, joskin silti tilastollisesti merkitseviä. Tämä on odotuksenmukainen tulos, koska kuten alaluvussa 2.1 totesin, *dA*-verbit ja konsonanttivartaloiset verbit eivät todennäköisesti voi lyhentyä. Kun nämä verbit kuitenkin kattavat huomattavan osan kaikista infinitiiviesiintymistä (noin 30 %), jää muutokselle vähemmän tilaa. A-infinitiiveissä tapahtuvia muutoksia käsittelemme alaluvussa 3.2.



Kuvio 1.
Infinitiivien kokonaisvariaatio.

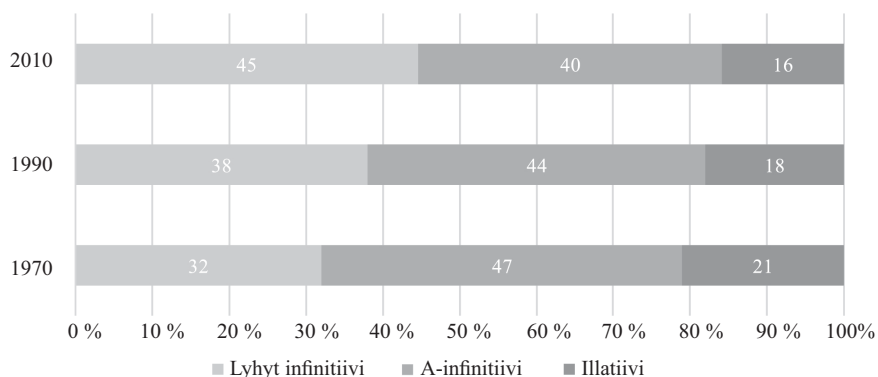
Kun vertaillaan ikäryhmiä portaittain, huomataan, että erot ovat selvimmät nuorten ja keski-ikäisten välillä. Kaikissa infinitiivimuodoissa erot nuorten ja keski-ikäisten välillä ovat khiin neliö -testin mukaan tilastollisesti erittäin merkitseviä ($p < 0,001$).⁹ Sen sijaan vanhojen ja keski-ikäisten välillä erot esimerkiksi A-infinitiiviesiintymisissä eivät ole merkitseviä ($X^2(1) = 0,19$, $p = 0,660$), vaikka muissa infinitiivimuodoissa on näidenkin kahden ikäryhmän välillä merkitseviä eroja. Tämä antaa lisätukea sille jo kuvioista 1 silmämääräisesti tehtävälle havainnolle, että infinitiivimuotojen käyttö on osittain hyvinkin ikäsidonnaista. Toisekseen tilastolliset erot kertovat siitä, että A-infinitiivi on infinitiivimuodoista vähiten altis ikäkauteen liittyvälle muutokselle.

Esittelen infinitiivien muutosmalleja seuraamalla aineiston jokaisessa vaiheessa mukana ollutta sukupolvea vuosikymmeneltä toiselle. 1970-luvun nuoret, 1990-luvun keski-ikäiset ja 2010-luvun vanhat edustavat samaa joukkoa (lukuun ottamatta neljää 2010-luvun vanhojen joukkoon laskettua jo 80-vuotiasta informanttia). Tätä joukkoa tarkasteltaessa huomataan, että sukupolvi tuntuu pitävän yllä nuorena aloitettua ten-

9. Keski-ikäisten ja nuorten väliset erot lyhyessä infinitiivissä khiin neliö -testin mukaan: $2(1) = 134,93$, $p < 0,001$. A-infinitiivissä: $2(1) = 47,59$, $p < 0,001$ ja illatiivissa: $2(1) = 38,84$, $p < 0,001$. Vanhojen ja keski-ikäisten väliset erot puolestaan ovat lyhyessä infinitiivissä merkitseviä $2(1) = 10,04$, $p = 0,002$ ja illatiivissa erittäin merkitseviä $2(1) = 16,61$, $p < 0,001$.

denssiä (sukupuolvittainen muutos), mutta lyhyen infinitiivimuodon määrät vähenevät ja illatiivimuotojen puolestaan kasvavat vuosikymmeneltä toiselle (ikäsidonnainen muutos). Voidaan siis todeta, että lyhyen infinitiivin yleisyys on osittain ikäsidonnaista mutta muutos on kuitenkin selvästi myös sukupuolvittaista (ks. alalukua 2.2; Kurki 2013: 206–210). Olennaista on myös huomata, että illatiivi ja lyhyt infinitiivi ovat A-infinitiiviä selvemmin ikäsidonnaisia. Näin ollen lyhyt infinitiivi MA-infinitiivin illatiivin tehtävissä on nuorille leimallinen piirre, kun taas lyhyt infinitiivi A-infinitiivin tehtävissä on selvemmin yleisesti hyväksytty eikä leimaudu tiettyyn ryhmään (leimautuneista piirteistä ks. Mantila 2004). Tällä voi olla huomattavia vaikutuksia muodon yleistymiseen, mihin palaan alaluvussa 4.2.

Lyhyt infinitiivi on siis leimallisesti nuorten infinitiivimuoto, mutta se säilyy heidän puheessaan jossain määrin myös myöhemmillä vuosikymmenillä eikä väisty täysin ikääntymisen myötä. Sen sijaan Sankoffin (2005) esittämää elinikäistä muutosta, jossa vanhemmat ikäryhmät ikään kuin hyppäävät muutokseen mukaan, ei infinitiivimuodoissa juuri näytä tapahtuvan, vaan vanhemmiten muutoksen suunta on ennemminkin nuorten suosimasta muodosta pois päin – erityisesti siis MA-infinitiivin illatiivin tehtävissä. Tästä huolimatta lyhyt infinitiivi yleistyy vuosikymmeneltä toiselle, mitä voidaan havainnollistaa laskemalla kaikkien ikäryhmien esiintymämäärät yhteen ja vertailemalla vuosikymmenkohtaisia määriä (kuvio 2).



Kuvio 2.
Vuosikymmenkohtainen vertailu.

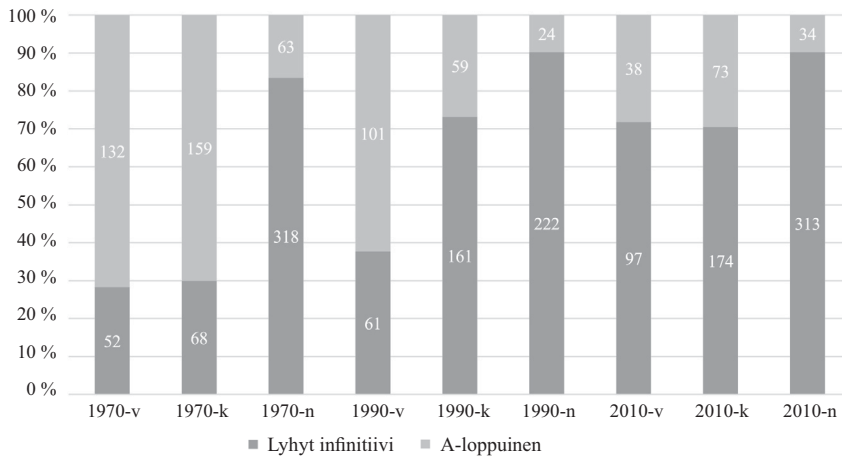
Kuviosta 2 havaitaan, että infinitiivien välillä tapahtuu muutosta myös vuosikymmenittäin. Lyhyt infinitiivi yleistyy jokaisella vuosikymmenellä kahden muun kustannuksella. Erot vuosikymmenten välillä ovat tilastollisesti erittäin merkitseviä ($p < 0,001$), ja suurimmat erot löytyvät jälleen vertailtaessa lyhyttä infinitiiviä muihin infinitiivimuotoihin.¹⁰ Näin ollen myös yhteisön kielenkäyttö muuttuu selvästi vuosikymmeneltä toiselle siirryttäessä eikä muutos jää ikäsidonnaiseksi. 2010-luvun

10. Verrattaessa lyhyttä infinitiiviä muihin infinitiivimuotoihin vuosikymmenten välillä erot ovat khiin neliö -testin mukaan erittäin merkitseviä ($2(2) = 99,14, p < 0,001$). Samoin A-infinitiivissä ($2(2) = 27,77, p < 0,001$) ja illatiivissa ($2(2) = 33,14, p < 0,001$) erot ovat tilastollisesti erittäin merkitseviä.

aineistossa lyhyt infinitiivi on kaikkein yleisin variantti, mutta jo 1970-luvun aineistossa lyhyitä infinitiivimuotoja on huomattavan suuri määrä. Tätä selittää etenkin se, että *antaa*-tyyppi on laskettu lyhyiden infinitiivien joukkoon. Vastaavasti A-loppuisten vokaaliyhtymien assimilaatio (*kertoa* > *kertoo*) on jo vanha ilmiö. Suurin osa lyhyistä infinitiiveistä on vielä 2010-luvullakin A-infinitiivilähtöisiä, ja tarkastelen niitä seuraavaksi. Itse puhetilanteen kannalta lähtöinfinitiivi on sinänsä yhdentekevä, mutta sen tarkastelu kertoo muutoksen lähtökohdista ja suunnasta.

3.2 A-loppuiseen vokaaliyhtymään päättyvien infinitiivimuotojen muutos

Edellisessä luvussa totesin, että A-infinitiiviesiintymissä erot eri ikäryhmien ja vuosikymmenten välillä eivät ole yhtä merkittäviä kuin kahdessa muussa infinitiivimuodossa. Tätä selitin sillä, että osa A-infinitiiveistä ei voi lyhentyä. Suurin osa lyhyistä infinitiiveistä toimii kuitenkin nimenomaan A-infinitiivin tehtävissä, ja ne syntyvät ennen kaikkea A-loppuiseen vokaaliyhtymään päättyvien infinitiivimuotojen assimilaation kautta. Näitä esittelen kuviossa 3.



Kuvio 3.

A-loppuisten vokaaliyhtymän esiintymät prosentuaalisesti ja määrällisesti.

Kuviosta 3 huomataan, että myös A-loppuisten vokaaliyhtymän sisältävissä tapauksissa nuoret käyttävät huomattavan paljon enemmän lyhyitä infinitiivejä kuin niiden standardinmukaisia vastineita. Toinen olennainen huomio on se, että muutos on jälleen selvästi sukupolvittaista: sama puhujajoukko (1970-luvun nuoret, 1990-luvun keskiikäiset ja 2010-luvun vanhat) pitää kiinni lyhyestä infinitiivistä. Toisaalta muutos on myös ikäsidonnaista: varianttien väliset erot kaventuvat ryhmän sisällä puhujien vanhetessa. Näin ollen voidaan tässäkin yhteydessä todeta näennäisaikahypoteesin olevan oikeansuuntainen mutta ikäsidonnaisten muutoksen hidastavan sukupolvittaista muutosta (vrt. Laasanen 2016).

A-loppuiseen vokaaliyhtymään päättyvien verbien osalta erot ovat khiin neliö -testin mukaan erittäin merkitseviä vuosikymmenten välillä ($X^2(2) = 110,26$, $p < 0,001$). Muoto-

ryhmiä erikseen tarkasteltaessa vain *iA*-loppuisissa verbeissä (esim. *leikkiä*) ei ole khiin neliö -testin mukaan tilastollisesti merkitseviä eroja. Tämä noudattelee *A*-loppuisten vokaaliyhtymien assimilaatiosta jo vanhastaan tiedettyä kaavaa: kuten olen jo edellä tuonut esiin, *eA* ja *OA* assimiloituvat herkemmin kuin *UA* ja *iA* (esim. Paunonen 1995 [1982]: 107–109). Korpuksen 1970-luvun aineistoa käsittelevässä tutkimuksessaan Paunonen toteaa, että *iA*-yhtymä assimiloituu kaikissa muotoryhmissä vain kolmasosassa tapauksista. Nuorten esiintymiin nojaten hän kuitenkin onastelee erojen kaventuvan (mts. 110–113). Tässä tutkimuksessa ero onkin infinitiivien osalta selvästi kaventunut, sillä 2010-luvun aineistossa pitkävokaaliset muodot (*pitää leikkii*) ovat niukassa enemmistössä. *iA*-loppuisten infinitiivien esiintymämäärä on kuitenkin kovin vähäinen, joten niitä ei voi suoraan vertailla Paunosen tarkastelemiin huomattavan frekventteihin partitiivimuotoihin.

Paunonen esittää (mts. 126), että partitiivimuodoissa *A*-loppuisten vokaaliyhtymän variaatio on lopulta johtanut partitiivin muodostussäännön muutokseen: alkuperäinen sääntö (vartalo + *A*) on korvautunut säännöllä vartalo + vartalovokaalin pidennys. Malliksi tälle muutokselle Paunonen esittää jo valmiiksi *A*-loppuisia sanoja: *kalaa, kesää* > *ovee, taloo, tuntii, lintuu*. Vastaavanlainen analogia vaikuttaa myös infinitiivimuotojen taustalla, ja ylipäänsä kaksitavuinen, pitkään vokaaliin päättyvä sanahahmo esiintyy Hel-singin puhekielessä monissa eri muotoryhmissä. Tähän palaan alaluvussa 4.2.

3.3 Sananrajailmiöt lyhyessä infinitiivissä

Kun tarkastellaan infinitiivien järjestelmää, on syytä huomioida myös sanojen rajoilla tapahtuvat ilmiöt. *A*-infinitiivin lativiisija esiintyy nyky Suomessa jäännösloppukkeena ja on yksi voimakkaimmin rajageminaatiota toteuttavista muotoryhmistä (VISK § 34; Karlsson & Lehtonen 1977: 10–15). *MA*-infinitiivin illatiivin loppu-*n* sen sijaan assimiloituu sitä seuraavien konsonanttien laadusta riippuen hieman eri tavoin. Samalla tavalla illatiivi ja latiivi käyttäytyvät kuitenkin *l:n*, *r:n*, *j:n* ja *v:n* edellä: *mennään pelaal lätkää* vrt. *voidaan pelaal lätkää*. Sananrajalla tapahtuvat ilmiöt kertovatkin siitä, mihin muotoihin lyhyt infinitiivi kiinnittyy: toteuttaako se *k:n* sandhikaavaa (kuten *A*-infinitiivin lativi), *n:n* sandhikaavaa (kuten illatiivi) vai ei mitään (kuten yksikön kolmannen persoonan muoto)?

Tätä selvittääkseni jaoin sananrajailmiöt kolmeen ryhmään: Ensin poistin joukosta lausumanloppuiset ja vokaalia edeltävät tapaukset, koska niissä rajailmiöitä ei esiinny.¹¹ Loput tapaukset jaottelin sen mukaan, tapahtuuko sananrajalla assimilaatioilmiöitä vai ei; rajageminaatio- ja assimilaatiotapaukset on siis yhdistetty. Ongelmia erotteluun tuottaa se, että eri vuosikymmenten haastattelut on litteroitu erilaisella tarkkuudella (Paunonen, henkilökohtainen tiedonanto 19.5.2017). 1970-luvulla ja 2010-luvulla rajageminaatioiden merkinnässä on suurta horjuntaa (tai siitä on luovuttu kokonaan), kun taas 1990-luvulla merkinnät ovat huomattavan tarkkoja. Niinpä olen tarkistanut kuuntelemalla joitakin epäselviä kohtia tämän tutkimuksen tarpeisiin.

11. Vokaalien edellä rajageminaatio voi toteutua glottaaliklusiilina, mutta sen esiintyminen on huomattavasti horjuvampaa kuin konsonanttien kahdentumisen (Karlsson & Lehtonen 1977: 16).

Tutkimus osoittaa, että A-infinitiivitapauksissa rajageminaatio toteutuu melko järjestelmällisesti kaikissa ikäryhmissä ja kaikilla vuosikymmenillä. MA-infinitiiviesiintymissä sen sijaan assimiloitumattomat tapaukset ovat niukassa enemmistössä kaikilla vuosikymmenillä. Vastaavat erot näkyvät myös lyhyen infinitiivin sananrajailmiöissä: A-infinitiivilähtöisissä lyhyissä infinitiivimuodoissa rajageminaatio toteutuu 70–80 prosentissa tapauksista (*voi pelaal lätkää*), kun MA-infinitiivilähtöisten kohdalla on enemmän horjuntaa. Nuorten puheessa MA-infinitiivilähtöisissä muodoissa on 1970-luvulla lähes saman verran assimilaatiotapauksia (*mennään pelaal lätkää*) kuin assimiloitumattomia (*mennään pelaa lätkää*). 1990- ja 2010-luvulla nuorten puheessa assimilaatiotapauksia on jo enemmistö. Ikäsidonaisuus näyttääkin vaikuttavan myös MA-infinitiivilähtöisen lyhyen infinitiivimuodon käyttäytymiseen sananrajalla, sillä keski-ikäisten ja vanhojen puheessa assimiloitumattomat muodot ovat enemmistössä. MA-infinitiivilähtöisen lyhyen infinitiivin esiintymämäärät sananrajalla ovat kuitenkin keski-ikäisten ja vanhojen ryhmässä selvästi pienemmät kuin nuorilla, mikä vaikeuttaa uskottavaa vertailua. Lisäksi on muistettava, että sananrajailmiöissä on paljon variaatiota niin idiolektitasolla kuin erilaisissa äänneympäristöissä (Karlsson & Lehtonen 1977: 15), minkä takia ryhmätason vertailut ovat lähinnä suuntaa antavia. Havaintojen perusteella voidaan todeta, että lyhyt infinitiivi ei ole ainakaan vielä sananrajalla vakiintunut tiettyyn muotoon, vaan se noudattaa lähtöinfinitiivien sandhikaavoja. Orastavaa ikäsidonaisuutta on kuitenkin havaittavissa siinä, että nuorten puheessa MA-infinitiivilähtöinen lyhyt infinitiivi vaikuttaa assimiloituvan herkemmin kuin vanhojen tai keski-ikäisten puheessa.

3.4 *käydä*- ja *olla*-verbien täydennykset

Lyhyen infinitiivin muotoja esiintyy jonkin verran myös *käydä*- ja *olla*-verbien täydennyksinä tilanteissa, joissa yleiskielessä käytettäisiin inessiiviä: *kävin pelaa* < *kävin pelaamassa*. Myös MA-tunnuksettomia illatiivi- ja inessiivimuotoja (*käydä pelaan* ~ *pelaas*) tavataan murteissa. Näiden verbien täydennysten alueellista variaatiota ja taustaa on esitellyt Ylikoski (2005).

Tämän tutkimuksen aineiston esiintymämäärät ovat näiden verbien täydennysten osalta hyvin vähäisiä, enkä siksi käsittele niitä tilastollisesti. Suurin osa aineiston täydennyksistä on kaikilla vuosikymmenillä inessiivimuotoja tai niiden loppuheittoisia variantteja (*käydä pelaamas*). Aineistosta voidaan kuitenkin havaita, että 2010-luvun nuorten puheessa lyhyet infinitiivimuodot ovat melko yleisiä myös *käydä*- ja *olla*-verbien täydennyksinä (10/45 esiintymää). Huomattavaksi määrän tekee se, että esimerkiksi 1970-luvun nuorilla esiintymiä on vain kaksi kaikkiaan hieman alle sadasta esiintymästä (2/97) ja vielä 1990-luvun nuorillakin vain vajaa kymmenesosa (5/54). Näin ollen voidaan suuntaa antavasti todeta, että lyhyt infinitiivi *käydä*- ja *olla*-verbien täydennyksenä vaikuttaa olevan vasta yleistymässä Helsingin puhekielessä.

3.5 Kokoavasti tuloksista

Tässä luvussa olen esitellyt tutkimuksen keskeiset tulokset infinitiivien kokonaisvariaation osalta ja eritelty myös muutoksen taustalla olevia syitä. Tärkeimpänä havain-

tona esitin, että lyhyt infinitiivi yleistyy vuosikymmeneltä toiselle ja se on 2010-luvulla jo kaikkein yleisin infinitiivimuoto. Lisäksi kiinnostavaa on se, että sekä lyhyt infinitiivi että illatiivimuoto vaikuttavat edustavan niin sanottua ikäsidonnaista muutosta: lyhyt infinitiivi on erityisesti nuorten suosima muoto, kun taas illatiivi vanhempien puhujien. Pelkällä ikäsidonnaisuudella eroja ei kuitenkaan voi selittää, sillä lyhyt infinitiivi yleistyy myös vuosikymmeneltä toiselle sukupolvittaisen muutoksen myötä, eli puhujat eivät täysin luovu tästä variantista ikääntyessään. Lyhyen infinitiivin yleistyminen kiinnittyy ennen kaikkea MA-infinitiivin illatiivimuotojen lyhentymiseen (*meen kertoo*) ja A-loppuiseen vokaaliyhtymään päättyvien muotojen assimiloitumiseen (*voin kertoo*). Assimilaatiotapauksissa on havaittavissa eroja eri vokaaliyhtymien välillä: *iA*-yhdistelmä assimiloituu hitaammin kuin muut yhdistelmät. Vastaava tendenssi on havaittu myös partitiivimuotojen tutkimuksessa (Paunonen 1995 [1982]: 122–132).

Muodollisesti lyhyt infinitiivi ei ole vielä vakiintunut tietynlaiseksi, vaan se kiinnittyy yhä pitkälti lähtöinfinitiiveihinsä: A-infinitiivilähtöiset lyhyet infinitiivimuodot toteuttavat loppukahdennusta, kun taas MA-infinitiivilähtöisissä varianteissa loppu-*n* ei yhtä voimakkaasti assimiloitu seuraavaan äänteeseen. Tässä suhteessa lyhyen infinitiivin sisäinen variaatio on vielä suurta, ja myöhempien tutkimusten näytettäväksi jää, kieliopillistuuko lyhyt infinitiivi tiettyyn muotoon ja merkitykseen vai jääkö se nykyiselleen – kahden tulosijaisen infinitiivimuodon kontaminaatioksi.

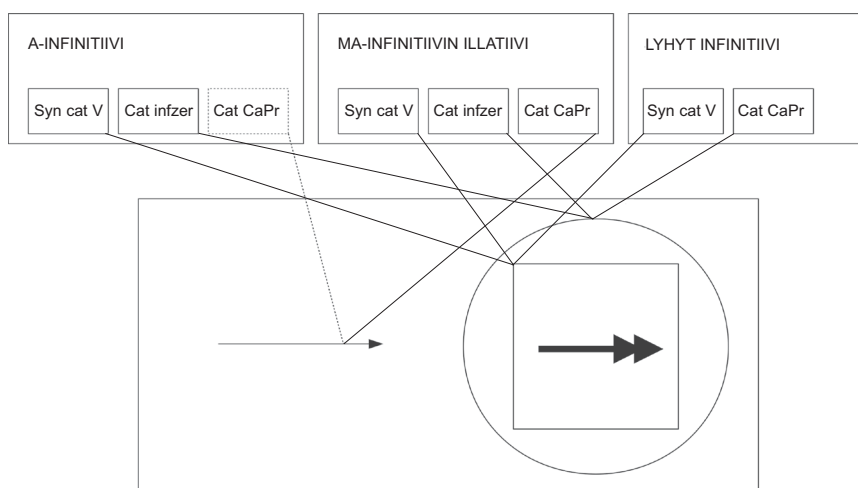
4 Infinitiivijärjestelmä muutoksessa

Tässä luvussa esittelen infinitiivijärjestelmän nykytilannetta Helsingin puhekielessä. Tarastelu on siis ennen kaikkea synkronista ja pohjautuu edellisessä luvussa esittelemiini muutoksiin. Luon myös katseen tulevaisuuteen ja pohdin muutostendenssejä.

4.1 Kolme infinitiiviä – morfologiset ja semanttiset erot

Olen edellä osoittanut aiemmissä tutkimuksissa (esim. Sorsakivi 1982; Paunonen 1995 [1982]) ounasteltujen muutosten edenneen melko pitkälle. Lyhyt infinitiivi on yleistynyt vuosikymmeneltä toiselle, ja 2010-luvun aineistossa se on jo kaikkein yleisin infinitiivimuoto. Muodon synnyn ja yleistymisen taustalla vaikuttavat ennen kaikkea muutamat äänne muutokset, jotka ovat ajaneet infinitiivimuodot yhteen. Toisaalta muotojen päädyttyä yhteiseen asuun analoginen mallivaikutus on entuudestaan lisännyt muodon yleisyyttä. Systemaattiset äänne muutokset ovat siis luoneet infinitiivijärjestelmään kaoottisuutta, jota analogia on vuorostaan systematisoinut (Sturtevant 1917: 108–109; Anttila 1989: 94). Voidaan kuitenkin puhua vain osittaisesta produktiivisuudesta, sillä analogia ei ainakaan vielä toistaiseksi yllä *dA*-verbeihin (**pitää syö*) tai konsonanttivartaloisten verbien A-infinitiivimuotoihin (**pitää tulee*). Näiden verbi-muotojen muutosmahdollisuuksia käsitellen alaluvussa 4.2.

Kolmen (yleis)infinitiivin järjestelmässä on paljon päällekkäisyyttä: eri infinitiivit eroavat toisistaan lähinnä muodoltaan, ja merkityserot ovat vähäisempiä. Infinitiivien merkityksien eroja esittelen konstruktiona kuviossa 4.



Kuvio 4.
Infinitiivien merkityksenkuvaus (Leino 2003: 105 mallina; Kuparinen 2015: 72).

Kuviossa 4 olen yhdistänyt konstruktiokuvaukset merkityksenkuvaukseen. Kuvaus on hyvin abstrakti eikä välttämättä kovin informatiivinen. Siitä voidaan kuitenkin erottaa infinitiivien välisiä eroja. MA-infinitiivin illatiivi täyttää kaikki kolme kuvauksen kohteena olevaa ominaisuutta: sillä on verbivartalo (V), infinitiivin tunnus (infzer) ja sijapäätte (CaPr). A-infinitiivin sijapäätte on hävennyt ja esiintyy nykyään jäännösloppukkeena. Näin ollen olen merkinnyt sen kuvaan katkoviivalla. Lyhyessä infinitiivissä en katso olevan sijaa lainkaan, vaan se koostuu pelkästään verbivartalosta ja infinitiivin tunnuksesta, joka lyhyessä infinitiivissä on vartalovokaalin pidennys. Lyhyt infinitiivi siis erkaantuu infinitiivien sijajärjestelmästä ja lähestyy puhdasta deverbaaliutta. Näin ollen infinitiivijärjestelmässä voisi olla tapahtumassa muutos kohti yhden yleisinfinitiivin järjestelmää indoeurooppalaisten kielten – tai pohjoissaamen – tapaan.

Myönnän, että kuvio 4 on pitkälti idealisoitu kuvaus infinitiivien järjestelmästä, joka on jatkuvasti jäsentyvää. Alaluvussa 3.3 huomautin, että lyhyt infinitiivi ei ole vielä muodollisesti erkaantunut lähtöinfinitiiveistään, vaan se toteuttaa esimerkiksi loppukahdennusta lähtöinfinitiivinsä mukaan. Voitaisiinkin esittää ajatus, että lyhyt infinitiivi on vain yksi variantti sekä A-infinitiivistä että MA-infinitiivin illatiivista.¹² Tämä variantti vain sattuu olemaan samanlainen molemmissa tapauksissa. Muotojen samuus vaikuttaa kuitenkin huomattavalla tavalla niiden tulkintaan. Onkin jok-

12. Tämä on pitkälti ollutkin tapana fennistisessä variaation tutkimuksessa: lyhyen infinitiivin muodot eivät ole mitenkään uusia puhekielessä, mutta niiden näkymättömyys alan kirjallisuudessa johtuu ennen kaikkea siitä, että ne on tulkittu eri tavoin: MA-infinitiivilähtöiset tapaukset MA-tunnuksettomiksi illatiiviksi ja A-infinitiivilähtöiset tapaukset A-infinitiiveiksi, joissa on tapahtunut erilaisia ääntenmuutoksia. Tämä on osaltaan seurausta suomalaisen puhekielen tutkimuksen kiinnittymisestä tiettyihin äänne- ja muotopiirteisiin (vrt. Nuolijärvi & Sorjonen 2005: 130–136).

seenkin selvää, että nykyisellään lyhyt infinitiivi leviää ennen kaikkea analogian voimin: infinitiiviä merkitään Helsingin puhemielessä vartalovokaalin pidennyksellä. Tätä kieliopillistumispolkua ja analogiaa käsittelem tarkemmin seuraavassa alaluvussa.

Infinitiivien merkityseroja on sanomalehtitekstien osalta tarkastellut Herlin (2012). Tutkimuksessaan hän vertailee A-infinitiiviä ja MA-infinitiivin illatiivia *alkaa*-verbin täydennyksenä ja esittää, että verbin merkityksen ohella myös kuvatun alkamisprosessin luonne ohjaa infinitiivin valintaa (ks. myös englannista Wierzbicka 1988: 85–87). Tätä edesauttaa se, että *alkaa* on merkitykseltään jokseenkin väljä (Herlin mts. 149). *Alkaa* tarjoaakin hyvän mahdollisuuden myös tämän tutkimuksen aineiston infinitiivien merkityserojen tarkasteluun, koska se voi saada täydennyksekseen minkä vain infinitiivimuodoista. Näin voidaan selvittää, onko infinitiivien merkityksessä kenties jotakin sellaista, joka edesauttaa tai estää infinitiivien sekoittumista ja leviämistä. Herlinin (mts. 162–170) mukaan MA-infinitiivin illatiivin esiintymistä *alkaa*-verbin täydennyksenä edesauttavat

- a) prosessin pitkäkestoisuus ja vähittäisyyden ilmaukset
- b) preesens aikamuotona
- c) epäspesifi tekijä ja
- d) prosessin dynaamisuus.

Lisäksi Herlin nostaa esiin erityisesti frekventatiivien johdosluokan, joka tuntuu herkästi esiintyvän MA-infinitiivin illatiivissa.

Tarkastelen tässä lyhyesti aineiston litteroiduista haastatteluista *alkaa*-verbin infinitiivitäydennyksiä suhteuttamalla niitä Herlinin tarjoamiin selittäviin tekijöihin. Käytän apuna logistisen regression mallia, jossa selittäviin tekijöihin nojaten voidaan arvioida kategoristen muuttujien todennäköisyyksiä (Cox 1958). Malli antaa siis mahdollisuuden laskea, miten eri tekijät vaikuttavat infinitiivimuotojen todennäköisyyksiin. On huomioitava, että selittävien tekijöiden arvioiminen on tässä tapauksessa jokseenkin tulkinnanvaraista: esimerkiksi prosessin pitkäkestoisuudessa on useita tasoja, joiden luokitus saattaa olla ajoittain hankalaa.

Aineistossani on kaikkiaan 158 *alkaa*-verbin täydennyksenä toimivaa infinitiiviä, ja ne jakautuvat hyvinkin tasan eri varianttien kesken: A-infinitiivitapauksia on 57, illatiivitapauksia 51 ja lyhyttä infinitiiviä 50. Taulukossa 4 (seur. sivulla) on esitetty logistisen regression tuottamat vetosuhteet (engl. *odds ratio*, 'todennäköisyyksien suhde'), kun illatiivimuotoja ja lyhyitä infinitiivejä on verrattu A-infinitiiviin. Näin ollen tilanteessa, jossa vetosuhde on suurempi kuin yksi, kyseinen infinitiivimuoto on todennäköisempi kuin A-infinitiivi vastaavassa tehtävässä. Vastaavasti arvot alle yhden tarkoittavat infinitiivimuodon olevan vähemmän todennäköinen. Esimerkiksi tässä aineistossa preesens-muotoiset *alkaa*-verbit saavat todennäköisemmin illatiivitäydennyksen kuin A-infinitiivin (vetosuhde 1,32), kun taas perfektissä tilanne on päinvastainen (0,38).

Koska aineisto on melko pieni ja joitakin selittäviä tekijöitä esiintyy aineistossa kovin vähän, käytän tulosten esittämisessä lisäksi luottamusväliä. Luottamusvälillä tarkoitetaan sellaista kahden arvon väliä, jonka voidaan valitun luottamustason (tässä tapauksessa 95 %) varmuudella sanoa sisältävän oikean arvon. Luottamusväli siis ottaa

huomioon aineiston koon ja hajonnan. Esimerkiksi vähittäisyyden ilmauksia on aineistossa niin vähän, että niiden luottamusväli kasvaa huomattavan suureksi. Mikäli vetosuhteiden luottamusväli on kokonaisuudessaan yli yhden, voidaan todeta 95 prosentin varmuudella positiivinen yhteys – mikäli alle yhden, yhteys on negatiivinen. Jos luku yksi (1) sisältyy luottamusväliin, yhteyttä ei voida tällä varmuudella todentaa.

Taulukko 4.

alkaa-verbin täydennysten selittävien tekijöiden vetosuhteet ja luottamusvälit.

Muuttuja	Illatiivi Vetosuhde	Illatiivi Luottamusväli 95 %	Lyhyt infinitiivi Vetosuhde	Lyhyt infinitiivi Luottamusväli 95 %
Preesens	1,32	0,58–3,04	1,16	0,48–2,78
Perfekti	0,38	0,06–2,33	1,80	0,43–7,46
Pitkäkestoinen	1,92	0,82–4,49	2,24	0,96–5,20
Frekventatiivi	1,96	0,48–7,97	0,14	0,01–1,46
Epäspesifi	0,54	0,23–1,27	0,44	0,18–1,08
Dynaaminen	1,19	0,52–2,73	0,48	0,19–1,24
Vähittäinen	1,68	0,32–8,82	3,61	0,79–16,58

Jos tarkastellaan taulukosta 4 vain aineistossa esiintyviä vetosuhteita, Herlinin esittämät selittävät tekijät näyttävät uskottavilta: illatiivissa preesens, pitkäkestoisuus, frekventatiivisuus, dynaamisuus ja vähittäisyys saavat kaikki yli yhden meneviä arvoja, eli A-infinitiiviin verrattuna illatiivin todennäköisyys on suurempi. Vain epäspesifin tekijän vetosuhte jää alle yhden, mikä on vastoin aiempaa hypoteesia. Lyhyt infinitiivi saa illatiivin tavoin suuremman todennäköisyyden A-infinitiiviin verrattuna preesensissä, pitkäkestoisuudessa ja vähittäisyydessä, mutta poikkeaa illatiivista frekventatiivien ja dynaamisuuden alhaisissa todennäköisyyksissä. Erityisesti frekventatiivien osalta ero on merkittävä: frekventatiiviverbi esiintyy lyhyenä infinitiivinä aineistossa vain kerran. Tähän osasy lienee tosin niiden muodossa: suuri osa frekventatiiviverbeistä kuuluu konsonanttivartaloisiin verbeihin (esim. *etsiskellä*), joiden olen todennut jo aiemmin esiintyvän harvoin lyhyinä infinitiivimuotoina (ks. alalukua 2.1).

Lisäksi vetosuhteista voidaan havaita, että lyhyt infinitiivi on sekä A-infinitiiviä että illatiivia todennäköisempi perfektissä, pitkäkestoisten prosessien esittämisessä ja vähittäisyyden ilmausten yhteydessä. Tämän aineiston nojalla lyhyt infinitiivi toimii siis monenlaisissa sekä A-infinitiivin että illatiivin tehtävissä, eikä ole syytä olettaa, että ainakaan nämä merkitystekijät estäisivät lyhyen infinitiivin leviämistä. Kuitenkin tapaukset, joissa täydennys on frekventatiivi, tekijä ei ole spesifi tai toiminta on dynaamista, näyttävät suosivan enemmän A-infinitiiviä tai illatiivia kuin lyhyttä infinitiivimuotoa. Aineiston pieni koko ja suuri hajonta aiheuttavat kuitenkin sen, että kovin suurella varmuudella edellä esitettyjä tendenssejä ei voida todistaa: kun tarkastellaan luottamusvälejä, yksikään selittävä tekijä ei saa kokonaisuudessaan alle yhden tai yli yhden meneviä arvoja. Infinitiivimuotojen merkitysten lähempi tarkastelu jääkin jatkotutkimuksen varaan.

4.2 Kohti yhtä infinitiiviä?

Kieliopillistumispolku tulosijaisesta muodosta yleisinfinitiiviksi on maailman kielissä yleisesti havaittu (Haspelmath 1989), ja osaltaan myös A-infinitiivin lyhyt muoto on noudattanut tätä kaavaa, kun se on siirtynyt alun perin tulosijaisesta latiivista abstraktimpiin tehtäviin (Saukkonen 1965; Sorsakivi 1982). Onkin syytä vertailla A-infinitiivin latiivin historiallista kehitystä lyhyen infinitiivin kehitykseen: molemmissa tapauksissa on edetty kohti yhtä – sijansa ainakin osittain menettänyttä – yleisinfinitiiviä, joka toimii monissa infinitiivisissä tehtävissä. Saukkonen esittää, että latiivin lisäksi itämerensuomalaisissa kielissä olisi ollut muitakin A-infinitiivin tulosijaisia vastineita ikään kuin selventämässä infinitiivin tulosijaisuutta (Saukkonen 1965: 20–21). Itämerensuomen kantakielivaiheessa A-infinitiivin latiivin rinnalle kehittynyt MA-infinitiivin illatiivi on lopullisesti täydentänyt tämän aukon. Ainakin Helsingin puhekielessä tendenssi tuntuu jälleen olevan kohti yhtä infinitiiviä, mutta pohdittavaksi jää, ilmaantuuko sen rinnalle jälleen selvästi tulosijaisempi muoto.

Toinen – infinitiivijärjestelmän ulkopuolinen – selitys lyhyen infinitiivin yleistymiselle on sanahahmojen mallivaikutus. Infinitiivien mallina vaikuttaa selvästi olevan kaksitavuinen ja pitkään vokaaliin päättyvä muoto, kuten *antaa*. Analogian vaikutuksesta *antaa*-hahmo on levinnyt myös muihin verbinmuotoihin (esim. *kertoo*, *istuu*). Tässä yhteydessä pitää muistaa myös Itkosen (1964) esittämä supistumaverbien analogiapolku. Nämä kolme muotoa voidaan esittää rinnakkain:

$$\frac{\text{hän antaa}}{\text{pitää antaa}} = \frac{\text{hän kertoo}}{\text{pitää kertoo}} = \frac{\text{hän pelaa}}{\text{pitää pelaa}}$$

Ajatusta sanahahmon mallivaikutuksesta tukee se havainto, että sanahahmon yleisyys Helsingin puhekielessä ei rajoitu vain infinitiiveihin, vaan vastaavaa hahmoa edustavat esimerkiksi passiivimuodot (*mentii*, *mennää*), A-loppuiseen vokaaliyhtymään päättyvät muodot (*kelloo*, *korkee*), verbien kolmannen persoonan muodot (*hän pelaa*, *ne pelaa*), eräät pronomini muodot (*mikää*, *mitää*, *kukaa*, *ketää*) ja lisäksi eräät yksittäiset sanat (*sillee*, *tällee* < *sillä tavalla*, *tällä tavalla*; *oikee*, *melkee* < *oikein*, *melkein*; *nopee* < *nopeasti*). Sanahahmo vaikuttaa osaltaan myös siihen, miksi *dA*-verbit eivät muutu lyhyiksi infinitiiveiksi: ne eivät voi asettua samaan muottiin vartalonsa takia. Tästä syystä vaikuttaa todennäköisemmältä, että lyhyt infinitiivi leviää ennemmin konsonanttivartaloisiin verbeihin (*juosta*) kuin *dA*-verbeihin (*saada*), koska konsonanttivartaloisista verbeistä voidaan tehdä myös kaksitavuinen vokaalivartalo, jonka loppuvokaali pidentyy: *pitää juoksee*. Suomessa on jo pitkään havaittu konsonanttivartaloitten muutoksia vokaalivartaloiksi, joskin etenemisen on todettu tapahtuvan jopa yksittäinen lekseemi kerrallaan (Paunonen 2003: 201–209; diffuusioilmioista suomen kielessä ks. Mielikäinen 1995).

Tässä tutkimuksessa olen osoittanut, että lyhyt infinitiivi on yleistynyt jokaiselta vuosikymmeneltä toiselle siirryttäessä. Osaltaan tärkeä havainto on se, että lyhyt infinitiivi ei ole 1970-luvunkaan aineistossa harvinainen, vaan se kattaa jopa 32 % kaikista esiintymistä (kuvio 2 s. 38). Näin ollen variaatiotilanne ja muutoksen siemen on jo paljon vanhempi kuin mihin tämän aineiston avulla on mahdollista päästä käsiksi, vaikka lyhyen infinitiivi-

muodon kasvu kaikkein yleisimmäksi variantiksi onkin tapahtunut vasta 1990-luvun alun jälkeen. Tulevaisuuden tilanteen ennustaminen käykin sikäli vaikeaksi, että selvin muutos on jo tapahtunut. Lyhyen infinitiivin leviäminen konsonanttivartaloisiin verbeihin (tai jopa *dA*-verbeihin) vaatii jo huomattavaa infinitiivien uudelleentulkintaa. Näin ollen voitaisiin ennustaa, että lyhyt infinitiivi kyllä yhä yleistyy niissä muotoryhmissä, joissa sen yleistyminen on mahdollista. Sen sijaan sen koko systeemin läpäisevä produktiivisuus näyttää tämän tutkimuksen valossa vielä epätodennäköiseltä. Yhden infinitiivin systeemiin on olemassa analogian kautta syntynyt mahdollisuus, mutta sen toteutuminen lienee useampien sukupolvien takana. Näin pitkälle menevän muutoksen ennustaminen on käytännössä mahdotonta, koska jo yhteiskunnan muutos tällaisella ajanjaksolla on valtava (Hurta 1999: 86–87; Laasanen 2016: 67–68).

Toinen lyhyen infinitiivin etenemistä jarruttava tekijä on sen osittainen ikäsidonaisuus: nuorten ryhmä käyttää lyhyttä infinitiiviä suhteellisesti huomattavan paljon, mutta puhujien ikääntyessä variantin suosio laskee. Jotta tähän asti havaittu muutos on voinut tapahtua, on sen täytynyt kuitenkin olla myös sukupolvittaista (tai elinikäistä), eli vaikka lyhyen infinitiivimuodon käyttö on vähentynyt sukupolven ikäännyttä, se ei kuitenkaan ole loppunut kokonaan. Näin jokainen sukupolvi on käyttänyt lyhyttä infinitiivimuotoa aina hieman enemmän kuin edellinen sukupolvi, jolloin koko yhteisö on muuttunut ajassa. Vaikuttaa siltä, että ikäsidonaisuus koskettaa nimenomaan MA-infinitiivin illatiivin tehtävissä toimivia lyhyitä infinitiivejä (ks. kuviota 1 s. 37). Näin voitaisiin todeta, että lyhyen infinitiivin käyttö MA-infinitiivin illatiivin tehtävissä on selvästi nuorille leimallista, kun taas A-infinitiivilähtöisissä tapauksissa tällaista leimautumista ei ole. Jotta lyhyt infinitiivi voisi aidosti vakiintua systeemiin, täytyisi muodon kenties muuttua puhujien silmissä neutraalimmaksi (vrt. Mantila 2004).

Tässä luvussa olen eritellyt lyhyen infinitiivin taustaa ja jäsentymistä osaksi infinitiivijärjestelmää. Leviämismalleiksi olen esittänyt toisaalta järjestelmällisten äänteenmuutosten vakiintumista infinitiiveihin analogian vaikutuksesta ja toisaalta näiden uusien infinitiivimuotojen mallivaikutusta. Olen todennut myös sanahahmojen vaikuttavan infinitiivin yleistymiseen. On luonnollista tulkita lyhyen infinitiivin irtautuvan ajan myötä lähtöinfinitiiveistään ja vakiintuvan omaksi yleisinfinitiivikseen. Sen sijaan sen vakiintumisen vaikutus kahteen muuhun infinitiiviin jää tulevien tutkimusten näyttäväksi. Tässä tutkimuksessa olen osoittanut, että infinitiivijärjestelmässä on selviä muutostendenssejä, mutta muutoksen etenemisen ennustaminen on vaikeaa. Systeemin nykyistä, kolmen infinitiivin tilannetta pitävät yllä muutamit verbinmuodot, jotka eivät toistaiseksi juuri edustu lyhyinä infinitiivimuotoina. Vastaavasti lyhyt infinitiivi vaikuttaa tämän tutkimuksen reaaliaikaisen aineiston valossa olevan osittain ikäsidonainen piirre, jonka käyttö vähenee puhujien ikääntyessä mutta ei siinä määrin, ettei tapahtuisi myös sukupolvittaista muutosta.

5 Lopuksi

Olen tarkastellut tässä artikkelissa kolmen yleisinfinitiivin variaatiota Helsingin puhekielessä 1970-luvulta 2010-luvulle. Analyysini osoittaa, että muutamit tunnetut äänteen-

muutokset ajavat A-infinitiivin lyhyen muodon (latiivin) ja MA-infinitiivin illatiivin samanmuotoisiksi. Tämän systeemiin syntyneen infinitiivimuodon olen nimennyt lyhyeksi infinitiiviksi ja tulkinnut sen eriävän analogian vaikutuksesta kahdesta muodosta omilleen. Tutkimukseni keskeinen tulos on lyhyen infinitiivin yleistyminen niin, että se on 2010-luvulla jo kaikkein yleisin infinitiivimuoto. Muutos on ollut sukupolvittaista ja alkanut jo ennen tämän aineiston ensimmäistä haastattelukierrosta. Erityisesti MA-infinitiivin illatiivin tehtävissä toimivat lyhyet infinitiivimuodot (*meen kertoo*) ovat ikäsidonaisia ja nuorten puheelle leimallisia. Puhujien ikääntyessä illatiivimuodot (*ker-tomaan*) lisääntyvät ja nuorille leimalliset piirteet jäävät vähemmälle kuin pelkällä näennäisaikaotetuksella voitaisiin päätellä. A-infinitiivilähtöisissä tapauksissa (*pitää kertoo*) tällaista leimallisuutta ei ole.

Lyhyt infinitiivi kiinnittyy edelleen lähtöinfinitiiveihinsä, mikä voidaan huomata etenkin sananrajalla tapahtuvista ilmiöistä: A-infinitiivilähtöiset lyhyet infinitiivit toteuttavat rajageminaatiota, kun taas MA-infinitiivilähtöisten tapausten loppu-*n:n* assimilaatiossa on horjuntaa. Myöskään merkityksen tasolla ei ole havaittavissa suuria eroja infinitiivien välillä. Näin ollen voidaan tulkita, että lyhyen infinitiivin irtaantuminen lähtöinfinitiiveistään ja matka kohti paljasta deverbaaliutta on vielä kesken, vaikka merkkejä siitä on olemassa. Eteneminen kohti yhtä yleisinfinitiiviä saamen tai indoeurooppalaisten kielten tapaan ei siis vielä näytä selvältä, vaan sen toteutuminen jää tulevien tutkimusten tarkasteltavaksi. Muutoksen etenemistä hidastavat erityisesti muutamit verbityypit (konsonanttivartaloiset ja *dA*-verbit), jotka eivät vielä juuri edustu lyhyinä infinitiiveinä, ja toisaalta lyhyen infinitiivin osittainen leimallisuus nuorten puhekieleen kuuluvana piirteenä.

Lähteet

Aineistolähteet

Helpuue 2014 = Helsingin yliopisto, suomen kielen, suomalais-ugrialaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos, Kotimaisten kielten keskus & Paunonen, Heikki (2014). Helsingin puhekielen pitkittäiskorpus (1970, 1990, 2010) [puhekorpus]. Kielipankki. <http://urn.fi/urn:nbn:fi:lb-2014073041>.

Kirjallisuuslähteet

- ANTTILA, RAIMO 1989: *Historical and comparative linguistics*. Toinen painos. Amsterdam: Benjamins.
- BUCHSTALLER, ISABELLE 2015: Exploring linguistic malleability across the life span. Age-specific patterns in quotative use. – *Language in Society* 44 s. 457–496. <https://doi.org/10.1017/S0047404515000391>.
- CHAMBERS, J. K. – TRUDGILL, PETER 1998: *Dialectology*. Toinen painos. Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9780511805103>.
- COX, DAVID 1958: The regression analysis of binary sequences. – *Journal of the Royal Statistical*

- Society B* 20 s. 215–242.
- HASPELMATH, MARTIN 1989: From purposive to infinitive – a universal path of grammaticization. – *Folia Linguistica Historica* 10 s. 287–310. <https://doi.org/10.1515/flih.1989.10.1-2.287>.
- HERLIN, ILONA 2012: Miksi alkaa tekemään? – Ilona Herlin & Lari Kotilainen (toim.), *Verbit ja konstruktiot* s. 143–174. Suomi 201. Helsinki: Suomalaisen Kirjallisuuden Seura.
- HERLIN, ILONA – LEINO, PENTTI – VISAPÄÄ, LAURA 2005: Kas siinä pulma. – Ilona Herlin & Laura Visapää (toim.), *Elävä kielioppi. Suomen infiniittisten rakenteiden dynamiikkaa* s. 9–38. Helsinki: Suomalaisen Kirjallisuuden Seura.
- HERLIN, ILONA – VISAPÄÄ, LAURA (toim.) 2005: *Elävä kielioppi. Suomen infiniittisten rakenteiden dynamiikkaa*. SKST 1021. Helsinki: Suomalaisen Kirjallisuuden Seura.
- HURTTA, HEIKKI 1999: Variaation tutkimuksen myytit ja stereotyyptit. – Urho Määttä, Pekka Pälli & Matti K. Suojanen (toim.), *Kirjoituksia sosiolingvistiikasta* s. 53–101. Tampere: Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
- ITKONEN, TERHO 1964: *Proto-Finnic final consonants. Their history in the Finnic languages with particular reference to the Finnish dialects*. Helsinki: Suomalais-Ugrilainen Seura.
- 1989: *Nurmijärven murrekirja*. SKST 498. Helsinki: Suomalaisen Kirjallisuuden Seura.
- JOMPPANEN, MARJATTA 2009: *Pohjoissaamen ja suomen perusinfinitiivi vertailussa leat ja olla -verbien yhteydessä*. Acta Universitatis Ouluensis B 92. Oulu: Oulu University Press.
- KARLSSON, FRED – LEHTONEN, JAAKKO 1977: *Alkukahdennus. Näkökohtia eräistä suomen kielen sandhi-ilmiöistä*. Turku: Turun yliopiston suomalaisen ja yleisen kielitieteen laitos.
- KUPARINEN, OLLI 2015: *Kohti ääretöntä. Jäsentyvä infiniitivikonstruktiojoukko Helsingin puhekielessä 1970-luvulta 2010-luvulle*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto. <http://urn.fi/URN:NBN:fi:uta-201512112521>.
- KURKI, TOMMI 2005: *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen*. SKST 1036. Helsinki: Suomalaisen Kirjallisuuden Seura.
- 2007: Variaation tutkimuksen nykynäkymiä. – *Sananjalka* 49 s. 143–162.
- 2013: Näennäisaikametodi ja muutosmallit. – *Sananjalka* 55 s. 202–212.
- LAASANEN, MIKKO 2016: Näennäisaikametodin taustaoletusten teoreettista tarkastelua. – *Virittäjä* 120 s. 57–83.
- LABOV, WILLIAM 1994: *Principles of linguistic change vol. 1. Internal factors*. Oxford: Blackwell.
- LAPPALAINEN, HANNA 2001: Sosiolingvistinen katsaus suomalaisnuorten nykypuhekieleen ja sen tutkimukseen. – *Virittäjä* 105 s. 74–101.
- 2004: *Variaatio ja sen funktiot. Erään sosiaalisen verkoston jäsenten kielellisen variaation ja vuorovaikutuksen tarkastelua*. SKST 964. Helsinki: Suomalaisen Kirjallisuuden Seura.
- LEHTONEN, HEINI 2006: Morfologinen variaatio maahanmuuttajataustaisten helsinkiläisnuorten puheessa. – Kaisu Juusela & Katariina Nisula (toim.), *Helsinki kieliyhteisönä* s. 255–274. Helsinki: Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.
- LEINO, JAAKKO 2003: *Antaa sen muuttua. Suomen kielen permissiivirakenne ja sen kehitys*. SKST 900. Helsinki: Suomalaisen Kirjallisuuden Seura.
- MANTILA, HARRI 2004: Murre ja identiteetti. – *Virittäjä* 108 s. 322–346.
- MEYERHOFF, MIRIAM 2011: *Introducing sociolinguistics*. Toinen painos. London: Routledge.
- MIELIKÄINEN, AILA 1995: Morfologinen diffuusio. – *Virittäjä* 99 s. 321–334.
- MUSTANOJA, LIISA 2011: *Idiolekti ja sen muuttuminen. Reaaliaikatutkimus Tampereen puhekielestä*. Acta Universitatis Tamperensis 1605. Tampere: Tampere University Press.
- NUOLIJÄRVI, PIRKKO – SORJONEN, MARJA-LEENA 2005: *Miten kuvata muutosta? Puhutun kielen tutkimuksen lähtökohtia murteenseuruuhankeeseen pohjalta*. Helsinki: Kotimaisten kielten tutkimuskeskus.

- PAUNONEN, HEIKKI 1995 [1982]: *Suomen kieli Helsingissä*. Helsinki: Helsingin yliopiston suomen kielen laitos.
- 2003: Suomen kielen morfologisista muutosmekanismeista. – Lea Laitinen, Hanna Lappalainen, Päivi Markkola & Johanna Vaattovaara (toim.), *Muotojen mieli. Kirjoituksia morfologiasta ja variaatiosta* s. 187–248. Kieli 15. Helsinki: Helsingin yliopiston suomen kielen laitos.
- 2005: Totta vai toiveajattelua. Näennäisaikainen muutos todellisen muutoksen kuvastajana. – Johanna Vaattovaara, Toni Suutari, Hanna Lappalainen & Riho Grünthal (toim.), *Muuttuva muoto. Kirjoituksia Tapani Lehtisen 60-vuotispäivän kunniaksi* s. 13–53. Kieli 16. Helsinki: Helsingin yliopiston suomen kielen laitos.
- 2006: Vähemmistökielestä varioivaksi valtakieleksi. – Kaisu Juusela & Katariina Nisula (toim.), *Helsinki kieliyhdistyksenä* s. 13–99. Helsinki: Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.
- PEARSON, KARL 1900: On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. – *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50 s. 157–175.
- PRIIKI, KATRI 2017: Hän, se, tää vai toi? Vuorovaikutussociolingvistinen tutkimus henkilöviitauksista Kaakkois-Satakunnan nykypuhekielessä. *Annales Universitatis Turkuensis C432*. Turku: Turun yliopiston julkaisuja.
- RAPOLA, MARTTI 1966: *Suomen kielen äännehistorian luennot*. SKST 283. Helsinki: Suomalaisen Kirjallisuuden Seura.
- SANKOFF, GILLIAN 2005: Cross-sectional and longitudinal studies. – Norbert Dittmar, Klaus J. Mattheier & Ammon Ulrich (toim.), *Sociolinguistics. An international handbook of the science of language and society* s. 1003–1013. Berlin: de Gruyter.
- SAUKKONEN, PAULI 1965: *Itämerensuomalaisten kielten tulosijainfinitiivirakenteiden historiaa I*. Suomalais-Ugrilaisen Seuran Toimituksia 137. Helsinki: Suomalais-Ugrilainen Seura.
- 1966: *Itämerensuomalaisten kielten tulosijainfinitiivirakenteiden historiaa II*. Suomalais-Ugrilaisen Seuran Toimituksia 140. Helsinki: Suomalais-Ugrilainen Seura.
- SIRO, PAAVO 1964: Infinitiivin asema suomen kieliopissa. – *Sananjalka* 6 s. 20–26.
- SKST = Suomalaisen Kirjallisuuden Seuran Toimituksia.
- SORSAKIVI, MERJA 1982: Infinitiivijärjestelmän muutoksia lasten kielessä. – *Virittäjä* 86 s. 377–391.
- STURTEVANT, EDGAR H. 1917: *Linguistic change. An introduction to the historical study of language*. Chicago: Chicago University Press.
- VISAPÄÄ, LAURA 2008: *Infinitiivi ja sen infiniittisyys*. SKST 1181. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VISK = Hakulinen, Auli – Vilkuna, Maria – Korhonen, Riitta – Koivisto, Vesa – Heinonen, Tarja Riitta – Alho, Irja 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkkoersio. <http://scripta.kotus.fi/visk> URN:ISBN:978-952-5446-35-7 (24.1.2018).
- WAGNER, SUZANNE – SANKOFF, GILLIAN 2011: Age grading in the Montréal French inflected future. – *Language Variation and Change* 23 s. 275–313. <https://doi.org/10.1017/S0954394511000111>.
- WIERZBICKA, ANNA 1988: *The semantics of grammar*. Amsterdam: John Benjamins.
- YLIKOSKI, JUSSI 2005: Puhekielen morfologisten ja semanttisten innovaatioiden tutkimusnäkömiä – esimerkkinä suomen *tekeen-* ja *tekees-*tyyppiset verbimuodot. – *Puhe ja kieli* 25 s. 187–209.

Variation and change in the use of infinitives in Helsinki Finnish

The article examines the two most frequent infinitives in Finnish – the lative of the A-infinitive (*pelata*) and illative of the MA-infinitive (*pelaamaan*). It also presents a contamination form of the two, termed the ‘shortened infinitive’ (*pelaa*). All three forms roughly mean ‘to play’. The analysis is based on the longitudinal corpus of the Finnish variety spoken in Helsinki. The corpus contains interviews recorded in the 1970s, 1990s and 2010s. The aim of the article is to examine variation and change in the use of these three different infinitive forms over a period of time in three different age groups. In addition, the article seeks to examine the underlying causes and patterns of such change.

The article shows that change in the use of infinitives in Helsinki is both generational and an example of age-grading. The use of the shortened infinitive is typical of the youngest age group, whereas older speakers prefer other infinitive forms. Although there is slight idiolectic change (when focusing on the informants interviewed in the earlier rounds of the project), each subsequent generation uses the shortened infinitive more frequently than the preceding generation. This leads to change within the community over time.

The shortened infinitive is the most frequent infinitive form in the 2010s. It is used to represent the functions both of the A-infinitive and the MA-infinitive. Some earlier studies have suggested that there may be a tendency towards using only one infinitive in Finnish, as is the case in many Indo-European languages. This study also indicates that such on-going change may eventually lead Finnish in this direction. However, this is only a case of partial productivity: some verb forms have not yet shortened in Helsinki at all. Thus, it must be left for future studies to evaluate whether the shortened infinitive will ever fully replace the two other infinitives. The article nevertheless presents a thorough overview of the incipient changes in the infinitive system.

Infinitiivien variaatio ja muutos Helsingissä

Artikkeli tarkastelee suomen yleisimpiä infinitiivimuotoja, A-infinitiivin lativia (*pelata*) ja MA-infinitiivin illatiivia (*pelaamaan*), sekä näiden yhteenlankeamisesta syntyneitä infinitiivimuotoja, lyhyttä infinitiiviä (*pela*). Aineistona käytetään Helsingin puhekielen pitkittäiskorpusta, joka sisältää haastatteluja 1970-, 1990- ja 2010-luvuilta. Artikkelin tarkoituksena on selvittää kolmen infinitiivimuodon variaatiota ja muutosta kolmella vuosikymmenellä ja kolmessa ikäryhmässä variaationanalyysin keinoin. Tutkimuksen kohteena ovat myös muutoksen etenemisen mallit ja muutoksen taustalla vaikuttavat syyt.

Tulokset osoittavat, että Helsingin puhekielen infinitiiveissä tapahtuu kahdenlaista muutosta: ikäsidonnaista ja sukupolvittaista. Lyhyt infinitiivi on erityisesti nuorten puhekielelle tyypillinen piirre, jonka käyttö on sitä vähäisempää, mitä vanhemmista puhujista on kyse. Toisaalta jo aiemmilla kierroksilla haastatellut puhujat ovat säilyttäneet melko hyvin infinitiivivarianttinsa. Näin koko yhteisö muuttuu, kun muutos etenee sukupolvittain.

Lyhyt infinitiivi on 2010-luvulla jo kaikkein yleisin infinitiivimuoto. Sitä käytetään sekä A-infinitiivin että MA-infinitiivin tehtävissä. Aiemmin on ounasteltu, että suomessa saattaisi olla meneillään tendenssi kohti yhtä yleisinfinitiiviä esimerkiksi indoeurooppalaisten kielten tapaan. Artikkelissa osoitetaan muutos tähän suuntaan, mutta kyse on kuitenkin osittaisesta produktiivisuudesta: *dA*-verbit (esim. *saada*) ja konsonanttivartaloiset verbit (esim. *juosta*) eivät vielä juuri esiinny lyhyinä infinitiiveinä. Näin lyhyen infinitiivin mahdollinen muutos yhdeksi yleisinfinitiiviksi kahden muun tilalle jää myöhempien tutkimusten arvioitavaksi, vaikka tässä artikkelissa luodaankin jo kattava katsaus infinitiivijärjestelmässä orastaviin muutoksiin.

Kirjoittajan yhteystiedot (address):
etunimi.sukunimi@uta.fi

Kirjoittaja on väitöskirjatutkija Tampereen yliopistossa.

JULKAISU

II: MUUTOSMALLIT

Muutosmallit kolmen aikapisteen pitkittäisaineiston valossa

Kuparinen, Olli – Mustanoja, Liisa – Peltonen, Jaakko – Santaharju, Jenni – Leino,
Unni

Sananjalka 2019, 61, 30–56
<https://doi.org/10.30673/sja.80056>

Artikkelin käyttöön väitöskirjan osana on saatu kustantajan lupa



OLLI KUPARINEN, LIISA MUSTANOJA, JAAKKO PELTONEN,
JENNI SANTAHARJU ja UNNI LEINO

Muutosmallit kolmen aikapisteen pitkittäisaineiston valossa

1 Johdanto

Yksi variationistisen sosiolingvistiikan keskeisistä tutkimuskohteista on meneillään olevan muutoksen tarkastelu. Muutoksia on tarkasteltu sekä yhdestä aikapisteestä käsin (näennäisajassa) että pitkittäisaineistojen valmistuttua monen eri aikapisteen välillä (reaaliajassa). Näennäisaikamenetelmää käytettäessä tarkastellaan usean eri ikäryhmän kieltä ja tehdään synkronisesta tilanteesta päätelmiä muutoksen suunnasta. Jos siis esimerkiksi nuorten ikäryhmä poikkeaa muista ikäryhmistä yhdessä aikapisteessä, voidaan näennäisaikamenetelmää hyödyntäen ennustaa yhteisön kielen muuttuvan kohti nuorten kielenkäyttöä.¹

Näennäisaikamenetelmä perustuu oletukselle idiolektien stabiiliudesta aikuisiässä. Tämän niin sanotun kriittisen iän hypoteesin mukaan yksilön kieli ei juuri muutu varhaisnuorena ohitetun kriittisen iän rajapyykin jälkeen (Lenneberg 1967; Meyerhoff 2011, 141). Reaaliaikatutkimusten valmistumisen myötä on kuitenkin saatu yhä vahvempaa todistusaineistoa siitä, että idiolektit voivat muuttua myös kriittisen iän jälkeen (Palander 2005; Kurki 2005; Sankoff–Blondeau 2007; Mustanoja 2011; Wagner–Sankoff 2011; Buchstaller 2015).

Keskeiseen asemaan muutoksen tarkastelussa on noussut William Labovin jaottelu muutosmalleista (1994, 83–85), jota on myöhemmin täydennetty (Sankoff 2005; Wagner–Sankoff 2011). Suomeksi muutosmalleja ovat esitelleet esimerkiksi Paunonen (2005a), Kurki (2013) ja Laasanen (2016). Labov pyrki muutosmallijaottelullaan osoittamaan, että yhdestä aikapisteestä ei voida ennustaa muutoksen etenemistä näennäisajassa, vaan tarvitaan reaaliaikatutkimuksia (Labov 1994, 84). Muutosmallijaottelua soveltavissa

1 Artikkelin on kirjoitettu osana Koneen Säätiön rahoittamaa hanketta. Tekijöistä toisen ja kolmannen osuudet ovat samanarvoiset, ja nimet ovat aakkosjärjestyksessä. Kiitämme *Sananjalan* nimettömiä arvioijia arvokkaista kommentista.

tutkimuksissa onkin usein tarkasteltu kahta aikapistettä reaaliajassa (esim. Sankoff–Blondeau 2007; Wagner–Sankoff 2011).

Tässä tutkimuksessa tarkastelemme kielenmuutosta reaaliaikaisesti kolmessa aikapisteessä käyttäen aineistona Helsingin puhekielen pitkittäiskorpusta (Helpuhe 2014). Osoitamme, että kolmannen aikapisteen lisääminen monimutkaistaa huomattavasti muutosmallien soveltamista: kahden aikapisteen välillä tietyltä muutosmallilta näyttäneet muutokset saattavat pysähtyä, kääntyä tai harpata eteenpäin toisen ja kolmannen aikapisteen välillä.

Esittelemme käytännön esimerkkejä muutosmalleista Helsingin puhekielessä ja pyrimme luomaan kokonaiskuvaa muutoksen etenemisestä. Tutkimuksen keskiössä on fonologinen ja morfologinen variaatio. Tarkastelumme lähtee liikkeelle perinteisestä muutosmallijaosta, mutta etenee pohtimaan sen taustaoletuksia myös kriittisesti. Vaikka tutkimuksemme tavoite onkin ennen kaikkea metodologinen, käsittelemme samalla myös Helsingin puhekielen nykytilannetta ja muutoksen suuntaa. Tutkimus vastaa muun muassa Kurjen (2013, 211) esittämään tarpeeseen suomen muutosten tarkastelusta muutosmallien avulla.

2 Muutosmallit

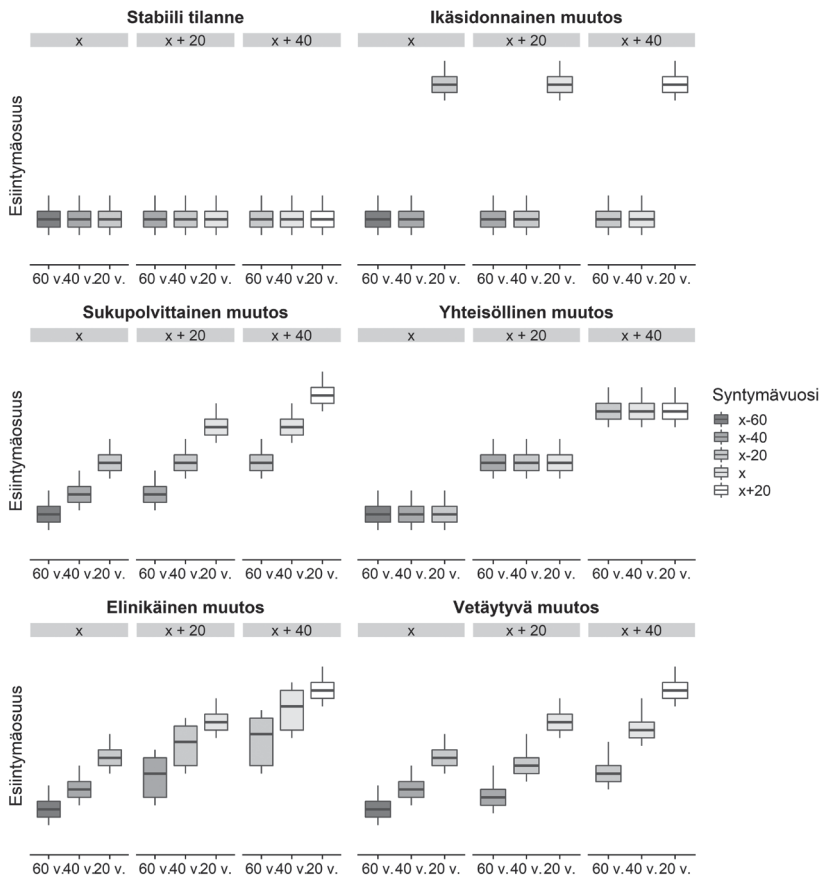
Labov (1994, 83) on hahmotellut teoreettisen kehityksen, joka pyrkii tavoittamaan sekä yhteisön että yksilön kielenmuutoksen etenemisen. Nämä muutoksen etenemisen tavat on yksinkertaistettu paljon siteerattuun nelikenttään (1–4 taulukossa 1; suomennokset Paunonen 2005a; Kurki 2013). Labovin alkuperäisiä malleja on myöhemmin täydennetty kahdella (taulukossa kursivoituna 5–6; Sankoff 2005; Wagner–Sankoff 2011; suomennokset Kurki 2013; Laasanen 2016).

Taulukko 1. Muutosmallit (Labov 1994; Sankoff 2005; Wagner–Sankoff 2011)

	Yksilö	Yhteisö
1. Stabiili tilanne	vakaa	vakaa
2. Ikäsidonnainen muutos	muuttuva	vakaa
3. Sukupolvittainen muutos	vakaa	(asteittain) muuttuva
4. Yhteisöllinen muutos	muuttuva	muuttuva
5. <i>Elinikäinen muutos</i>	<i>muuttuva</i>	<i>(asteittain) muuttuva</i>
6. <i>Vetäytyvä muutos</i>	<i>muuttuva</i>	<i>(asteittain) muuttuva</i>

Kuviossa 1 havainnollistamme kolmessa aikapisteessä ja kolmessa ikäryhmässä ilmeneviä muutosmalleja ruutu- ja janakaavioin (engl. *boxplot*). Kuvion 1 esimerkit ovat

teoreettisia, eivätkä perustu aitoon aineistoon. Kaaviotyyppejä käytetään kuitenkin myös analyysiluvussa 4, joten esittelemme lyhyesti kaavion ominaisuudet. Nimensä mukaisesti kaaviossa on ruutuja ja janoja. Ruudun yläreuna vastaa tutkitun joukon yläneljännestä ja alareuna alaneljännestä. Ruudun keskellä oleva vaakaviiva vastaa joukon mediaaniarvoa. Ruudun sisään mahtuvat siis joukon arvot 25 prosentista 75 prosenttiin, ja koko joukon keskimääräinen arvo on merkitty vaakaviivalla. Mikäli varsinaisen ruudun sijaan näkyy vain vaakaviiva, kaikki joukon arvot ovat hyvin pienellä alueella mediaanin ympärillä. Ruudusta lähtevien janojen päät vievät joukon suurimpaan ja pienimpään arvoon. Jos joukossa on huomattavasti muista poikkeavia arvoja, nämä merkitään kuvioon pistein. Pisteet merkitään, jos arvo on 1,5 ruudun mitan etäisyydellä ruudun reunasta. Kaavioiden yläreunaan on merkitty haastattelu-aika, x-akselille ikäryhmä ja harmaan eri sävyissä puhujien syntymäaika. Sama puhujakohortti pysyy siis väriltään samana kaikissa teoreettisissa haastattelupisteissä. Kuvion 1 jälkeen esittelemme muutosmallit lyhyesti.



Kuvio 1. Ruutu- ja janakaaviot muutosmalleista kolmessa teoreettisessa aikapisteessä ja kolmessa ikäryhmässä.

Stabiili tilanne

Stabiilissa tilanteessa yksilöt säilyttävät idiolekteissa havaittavat variaatiosuhteensa läpi eliniän, eikä yhteisössäkään tapahdu muutosta. Täysin stabiilia tilannetta on pidetty pitkälti teoreettisena (Paunonen 2005a, 15). Tämä johtunee kuitenkin myös sosiolingvistisen tutkimuksen omista lähtökohdista. Kun tutkimuksiin pyritään valitsemaan varioivia piirteitä, stabiilien piirteiden esiintyminen on luonnollisesti harvinaista.

Ikäsidonnainen muutos

Ikäsidonnaisessa muutoksessa yksilöiden kieli muuttuu heidän elinikänsä aikana, mutta yhteisö pysyy vakaana. Idiolekit siis muuttuvat tiettyä ajanjaksona, jonka jälkeen ne palaavat ennalleen. Kuviossa 1 huomataan tyypillinen ikäsidonnainen malli: nuorten ikäryhmä eroaa muista jokaisessa aikapisteessä, mutta koska sama kaava jatkuu jokaisella vuosikymmenellä, yhteisö itsessään ei muutu.

Yksi ongelma ikäsidonnaisen muutosmallin sovelluksessa on sen vaihteleva käyttö (Wagner 2012a, 378–379). Edellä olemme esitelleet syklisen mallin, jossa jokin ikäryhmä aina uudelleen poikkeaa muista (Labov 1994, 46; Sankoff 2005, 1011). Joskus ikäsidonnaisuudella on tarkoitettu vain tilannetta, jossa ikä korreloi kielellisten erojen kanssa eli ikäryhmät eroavat toisistaan (esim. Downes 1998, 225–226, 238–239). Downesin näkemys ikäsidonnaisuudesta onkin oikeastaan vain sukupolvittaisen muutoksen (ks. alla) pysäytyskuva. Samassa yhteydessä Downes (mts. 239) puhuu kuitenkin ikäefekteistä (*age-effects*), joilla hän viittaa eri ikäkausien normien tuomiin paineisiin (vrt. kielellisen markkinan periaate: Bourdieu–Boltanski 1975; Sankoff–Laberge 1978). Tagliamonte (2012, 46–48) puolestaan ajattelee tavallisen ikäsidonnaisen muutoksen noudattavan u:n muotoista käyrää: nuoruudessa ja vanhuudessa puhkelellisia muotoja käytetään enemmän kuin keski-ässä.

Vastaava käsitteiden huojunta näkyy myös suomenkielisessä kirjallisuudessa, kun Paunonen (2005a, 15) kääntää Labovin ikäsidonnaisen mallin (*age-grading*) ikäkausivaihteluksi, jota taas Kurki (2013, 205) tarjoaa sekä ikäsidonnaisen muutoksen että elinikäisen muutoksen yläkäsitteeksi. Kun tässä tutkimuksessa puhumme ikäsidonnaisesta muutoksesta, viittaamme sillä nimenomaan Labovin esittämään sykliseen malliin.

Syklisen ikäsidonnaisen muutoksen taustalla vaikuttaa oletus lyhytkestoisesta poikkeavuudesta, jonka jälkeen idiolekti palautuu normaaliksi. Tällainen malli kuulostaa sikäli vieraalta, että se olettaa jonkinlaisen ”aidon idiolektin” olemassaolon, josta vain hetkeksi harhaidutaan (Laasanen 2016, 73). Esimerkiksi Sankoff (2005, 1011) erottaa elinikäisen muutoksen ikäsidonnaisesta muutoksesta siitä syystä, että ikäsidonnaisuus on ohimenevää ja sukupolvittain tyypillisesti nuorten puheesta toistuvaa, kun taas elinikäinen muutos on pysyvää idiolektin muutosta aikuisiällä. Tällainen erottelu on kuitenkin hieman vaarallinen: on helppo huomata muutos nuoruudesta aikuisuuteen ja tulkita edeltänyt vaihe nuoruudesta johtuneeksi erikoisuudeksi. Sen sijaan aikuisuus on käsitteellisesti epäselvä elämänvaihe: mistä se alkaa, mihin se päättyy ja voiko sitä pitää yhtenäisenä ajanjaksona?

Tämä voidaan kytkeä myös sosiaalitieteitä yleisemmin vaivaavaan ikäharhaan: ainoastaan keski-ikä on vakaa elämänvaihe, johon joko pyritään mukautumaan nuoruudesta aikuisuuteen siirryttäessä tai siitä joudutaan päästämään irti vanhetessa (Eckert 1997, 157–158; vrt. Labov 1994, 46). Voisiko ikäsidonnaisen muutoksen nuorisokeskeisyydessä olla siis kyse siitä, että nuoruus on länsimaissa kohtuullisen standardi elämänvaihe kouluelämän takia? Näin ollen ikäsidonnaisuus olisikin elämäntilannesidonnaisuutta, ja nuorten elämäntilanteet vain ovat huomattavasti yhtenäisempiä kuin aikuisten (vrt. Eckert 1997). Sosiolingvistiikan perinteessä iän ja muutoksen suhteen tutkimus keskittyy vahvasti nuoruuteen, ja tätä näkökulmaa olisi syytä laajentaa.

Sukupolvittainen muutos

Ikäsidonnainen ja sukupolvittainen muutos näyttävät yhdestä hetkestä tarkasteltuna samalta: ikäryhmät eroavat toisistaan. Edellä esitellyn ikäsidonnaisen muutoksen sijaan sukupolvittainen muutos johtaa yhteisön muutokseen, koska ikäryhmien välillä on eroja ja idiolektit säilyvät läpi eliniän.

Sukupolvittaista muutosta on pidetty tyypillisenä muutoksen tapana. Esimerkiksi Labov toteaa sukupolvittaisen muutoksen olevan normaalia muutosta (Labov 1994, 84). Reaaliaikaisissa tutkimuksissa on kuitenkin huomattu idiolektien muuttuvan läpi eliniän (esim. Mustanoja 2011; Buchstaller 2015), jolloin sukupolvittaisen muutoksen normaaliutta on alettu entisestään kyseenalaistaa. Selvää kuitenkin on, että näennäisaikannusteet pitävät monissa tapauksissa paikkansa, jolloin sukupolvittaista muutosta on tapahtunut. Tärkeää onkin pyrkiä erottelemaan, minkä tyyppiset ilmiöt muuttuvat sukupolvittain ja mitkä eivät.

Yhteisöllinen muutos

Yhteisöllisessä muutoksessa yhteisön jäsenet iästä riippumatta muuttuvat samanaikaisesti samaan suuntaan. Perinteisesti tulkiten yhteisöllisessä muutoksessa kaikki ikäryhmät ovat keskenään jo lähtökohtaisesti samassa tasossa ja muuttuvat ajassa saman verran (ks. kuvio 1). Tällainen samantasoisuuden vaatimus supistaa kuitenkin yhteisöllisen muutoksen käyttöalan esimerkiksi uusien sanojen ilmaantumiseen yhteisön kieleen. Yhteisöllinen muutos olisikin mielekkäämpää tulkita niin, että se kattaa minkä tahansa muutoksen, jossa kaikki ikäryhmät siirtyvät samanaikaisesti samaan suuntaan, oli ikäryhmien välillä alkutilanteessa millaisia eroja tahansa. Labov (1994, 84) itse asiassa kirjoittaa yhteisöllisen muutoksen määritelmässä tähän tapaan. Myöhemmin hän kuitenkin esittää, että stabiilia tilannetta ja yhteisöllistä muutosta ei voida erottaa toisistaan näennäisaikamenetelmällä, mikä voidaan tulkita vain niin, että molempien on näytettävä yhdestä aikapisteestä samalta eli tasaiselta suoralta. Tällaiseksi yhteisöllinen muutos on myös tulkittu (esim. Kurki 2013, 208).

Elinikäinen muutos

Gillian Sankoff on täydentänyt Labovin alkuperäistä jaottelua elinikäisellä muutoksella (2005; Sankoff–Blondeau 2007). Elinikäinen muutos ja seuraavassa esiteltävä vetäytyvä muutos ovat syntyneet reaaliaika-aineistojen havainnoista. Elinikäinen muutos perustuu Labovin jaottelun sukupolvittaiselle muutokselle siten, että siinäkin muutos etenee ennen kaikkea sukupolvittain. Erona sukupolvittaiseen muutokseen on kuitenkin se, että elinikäisessä muutoksessa osa vanhemmista puhujista tempautuu meneillään olevan muutoksen mukaan, jolloin muutosnopeus on sukupolvittaista muutosta nopeampi.

Vetäytyvä muutos

Viimeinen muutosmalli on rinnakkainen elinikäisen muutoksen kanssa siten, että myös se on syntynyt käytännön havainnoista (Wagner–Sankoff 2011). Vetäytyvässä muutoksessa tietty variantti näyttää yhdestä havaintohetkestä katsottuna yleistävän voimakkaasti, koska se on hyvin yleinen nuorten ryhmällä. Kun tämä nuorten ryhmä siirtyy keski-ikään, puhujat vetäytyvätkin osittain variantin käytöstä. Muutos on siis hitaampaa kuin sukupolvittaisen muutoksen mallissa, mutta nopeampaa kuin ikäsidonnaisessa mallissa, jonka mukaan yhteisö ei muutu lainkaan.

Labov pyrki malleillaan osoittamaan, että näennäisaikametodilla tarkasteltuna muutokset näyttävät samanlaisilta: stabiili ja yhteisöllinen muutos tasaiselta suoralta, kun taas sukupolvittainen muutos ja ikäsidonnaisten muutosten nousuvalta tai laskevalta käyrältä (Labov 1994, 84). Elinikäinen muutos ja vetäytyvä muutos ovat puolestaan käytännönläheisempiä malleja, jotka ovat kuitenkin käytännössä Labovin alkuperäisten mallien yhdistelmiä. Esimerkiksi elinikäinen muutos on sukupolvittaisen muutoksen ja yhteisöllisen muutoksen yhdistelmä: muutos on pääpiirteittäin sukupolvittaista, mutta puhujat muuttuvat aikuisiälläkin meneillään olevan muutoksen suuntaan. Vastaavasti vetäytyvä muutos on osittain ikäsidonnaista ja osittain sukupolvittaista muutosta, koska nuoruudessa käytetyt variantit eivät kokonaan vetäydy puhujien idiolekteista.

Vaikka Sankoff (2005), Blondeau (Sankoff–Blondeau 2007) ja Wagner (Wagner–Sankoff 2011) ovatkin onnistuneesti käsitteistäneet sekä elinikäisen muutoksen että vetäytyvän muutoksen, näihin malleihin viitataan toistuvasti jo aiemmissa muutosmalleja käsittelevissä tutkimuksissa niitä varsinaisesti nimeämättä. Esimerkiksi Labov toteaa jo nelikenttensä esitetyään, että todellisuudessa mallit toiminevat ennemmin useiden mallien yhdistelminä kuin yksittäisinä (Labov 1994, 85).² Vastaavasti Paunonen (2005a, 16) esittelee sekä elinikäisen muutoksen että vetäytyvän muutoksen yksinkertaisesti suhteessa sukupolvittaiseen muutokseen: muutos on näennäisaikaista kuvaa nopeampi, jos yksilöt muuttuvat meneillään olevan muutoksen suuntaan ja näennäisaikaa hitaampi, jos he

2 Tosin käsitellessään Cedergrenin (1987) tuloksia Labov pohtii, sopiiko niihin paremmin sukupolvittainen vai ikäsidonnaisten malli, kun tulokset viittaavat näiden yhdistelmään eli Wagnerin ja Sankoffin termin vetäytyvään muutokseen (Labov 1994, 94–97).

muuttuvat sitä vastaan. Elinikäisen muutoksen kaltaisen mallin on omassa aineistossaan havainnut myös Kurki (2005, 240), mutta hän nimittää sitä muutoksen haitarimaiseksi etenemiseksi. Nämä muutosmallit on siis jo pitkään huomioitu, mutta vasta myöhemmin käsitteellistetty.

3 *Aineisto ja menetelmät*

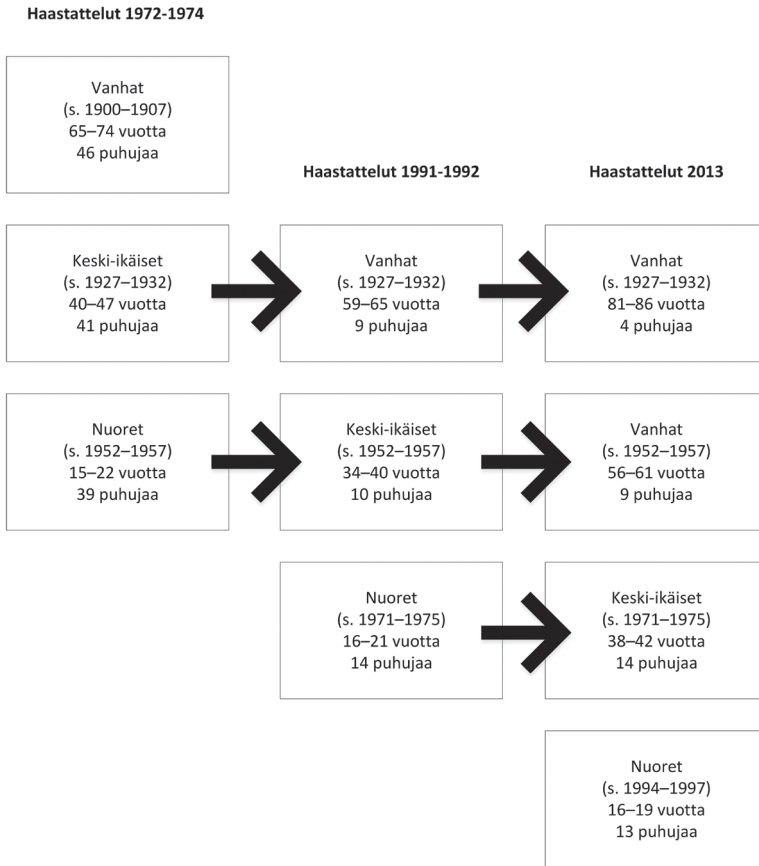
Tutkimuksen aineistona käytämme Helsingin puhekielen pitkittäiskorpusta (Helpuhe 2014), jonka haastattelut on toteutettu kolmella vuosikymmenellä (vuosina 1972–1974, 1991–1992, 2013). Hankkeen aloitti professori Terho Itkonen, kaksi ensimmäistä haastattelukierrosta toteutti professori Heikki Paunonen ja 2010-luvun kolmannesta aineistonkeruusta lähtien hankkeen johtajana on toiminut dosentti Hanna Lappalainen. Kaikilla vuosikymmenillä haastattelijoina ovat toimineet Helsingin yliopiston opiskelijat. Aineisto on osittain digitoitu ja sijoitettu osaksi CSC:n kielipankkia. Informantit valittiin alun perin Töölöstä ja Sörnäisistä, kolmesta sosiaaliluokasta ja kolmesta ikäryhmästä. 1990-luvulta lähtien kaupunginosadikotomiasta on demografian muutoksen takia luovuttu, ja sosiaaliluokkien luokittelua on niin ikään vähennetty kahteen. 2010-luvulla on haastateltu neljää ikäryhmää, sillä myös neljä 80-vuotiasta puhujaa on tavoitettu haastatteluun.

Korpus on toteutettu paneeli- ja trenditutkimuksen yhdistelmänä siten, että edellisillä kierroksilla haastatelluista nuorten ja keski-ikäisten ryhmistä on koottu seuraavien kierrosten keski-ikäisten ja vanhojen puhujien ryhmät, joiden lisäksi kullakin vuosikymmenellä mukaan on otettu uusi nuorten ryhmä. Koska kaikkia puhujia ei ole tavoitettu uudelleen haastateltaviksi, aineisto on pienentynyt 1990-luvun ja 2010-luvun haastattelukierroksilla. Olemme käyttäneet tässä tutkimuksessa 1990-luvun litteroitua haastattelua, vaikka korpus itsessään on vieläkin suurempi. Suurin osa litteraateista on tehty noin puolen tunnin haastattelukatkelmasta, kun taas jotkin erityisesti 1990-luvun ja 2010-luvun litteraateista kattavat noin tunnin mittaisen jakson. Esitämme aineiston jakautumisen vuosikymmenittäin ja ikäryhmittäin kuviossa 2.

Olemme tarkastelleet aineistosta 34:ää kielenpiirrettä. Näistä suurin osa on jo aiemmin tutkittuja Helsingin puhekielen piirteitä (esim. Paunonen 1995; 2006). Tutkimuksessa tarkastellut piirteet on esitelty taulukossa 2. Tässä yhteydessä käsittelemme piirteitä kuitenkin valikoiden sen mukaan, miten ne sopivat muutosmallien esittelyyn.

Olemme poimineet esiintymät litteraateista käyttäen säännöllisiä lausekkeita (Kleene 1956; Thompson 1968). Tämän jälkeen olemme vielä tarkastaneet löydetty esiintymät ja koostaneet ne taulukkoon automaattisesti käyttäen R-ohjelmistoa (R Core Team 2017), jolla olemme tehneet myös tutkimuksen tilastolliset analyysit ja kuviot.

Muutoksen tarkastelussa lähdemme liikkeelle tilastollisesta testaamisesta: tulkitsemme muutokseksi sellaiset osa-aineistojen väliset erot varianttien edustuksessa, jotka ovat tilastollisesti merkitseviä (vrt. Mustanoja 2011, 79). Tilastollisesti merkitsevä muutos ei välttämättä tarkoita, että variaatiosuhde olisi muuttunut dramaattisesti, vaan mahdollistaa myös pienempien, orastavien muutosten havaitsemisen, kunhan aineistoa on



Kuvio 2. Aineiston jakautuminen vuosikymmenittäin ja ikäryhmittäin (mallina Pau-
nonen 2006, 66).

Taulukko 2. Tutkitut piirteet.

Piirre	Variantti 1	Variantti 2	Variantti 3
OA, eA -loppuiset vokaaliyhtymät partitiivissa nomineissa infinitiiveissä	a) <i>autoa</i> b) <i>korkea</i> , c) <i>kertoa</i>	<i>autoo</i> , <i>korkee</i> , <i>kertoo</i>	
iA -loppuiset vokaaliyhtymät partitiivissa infinitiiveissä	a) <i>kahvia</i> , b) <i>leikkiä</i>	<i>kahvii</i> , <i>leikkii</i>	

Piirre	Variantti 1	Variantti 2	Variantti 3
UA -loppuiset vokaaliyhtymät partitiivissa, infinitiiveissä	a) <i>koulua</i> , b) <i>istua</i>	<i>kouluu</i> , <i>istuu</i>	
Supistumaverbien A-infinitiivi	<i>pelata</i>	<i>pelaa</i>	
MA-infinitiivin illatiivi	<i>pelaamaan</i>	<i>pelaan</i>	<i>pelaa</i>
<i>käydä</i> + infinitiivi	<i>käydä</i> <i>pelaamassa</i>	<i>käydä pelaamas</i> / <i>pelaas</i>	<i>käydä pelaa</i>
VVdV (määrälekseemeissä)	<i>meidän</i>	<i>meiän</i> , <i>meijän</i>	<i>meirän</i>
<i>hd</i>	<i>tehdä</i>	<i>tehä</i>	<i>tehrä</i>
VdV	<i>edes</i>	<i>ees</i>	<i>eres</i>
<i>vielä</i>	<i>vielä</i>	<i>viel</i>	<i>vie</i> , <i>viä</i>
<i>ts</i> (määrälekseemeissä) vahvassa asteessa heikossa asteessa	<i>seitsemän</i> (a) <i>itse</i> (b)	<i>seittämän</i> (a) <i>itte</i> (b)	<i>ite</i> (b)
Inessiivi	<i>talossa</i>	<i>talosa</i>	<i>talos</i>
Adessiivi	<i>talolla</i>	<i>talol</i>	
Elatiivi	<i>talosta</i>	<i>talost</i>	
<i>i:n</i> loppuheitto (tapaukset analysoitu erikseen)	<i>yksi</i> , <i>tulisi</i> , <i>taloksi</i> , <i>pelasi</i>	<i>yks</i> , <i>tulis</i> , <i>taloks</i> , <i>pelas</i>	
Painottoman tavun <i>i</i> -loppuiset diftongit	a) <i>punainen</i> , <i>antaisit</i> b) <i>puhui</i>	a) <i>punanen</i> , <i>antasisit</i> b) <i>puhu</i>	
NUT-partisiippi	<i>purrut</i> , <i>hyppinyt</i>	<i>purru</i> , <i>hyppiny</i>	
<i>ei ole</i>	<i>ei ole</i>	<i>ei oo</i>	<i>ei o</i>
<i>olla</i> , <i>tulla</i> (1. ja 2. persoona)	<i>olen</i> , <i>olet</i>	<i>oon</i> , <i>oot</i>	
<i>tämä</i> , <i>nämä</i>	<i>tämä</i> , <i>nämä</i>	<i>tää</i> , <i>nää</i>	
1. ja 2. persoonan pronominit	<i>minä</i> , <i>sinä</i>	<i>mää</i> , <i>sää</i>	<i>mä</i> , <i>sä</i>
3. persoonan pronomini (yksikkö ja monikko)	<i>hän</i> , <i>he</i>	<i>se</i> , <i>ne</i>	
Monikon 1. persoona	<i>menemme</i>	<i>mennään</i>	
<i>n:n</i> loppuheitto (määrälekseemeissä)	<i>mitään</i> , <i>kukaan</i>	<i>mitää</i> , <i>kukaa</i>	
<i>sillee</i> , <i>oikee</i>	<i>sillä tavalla</i> , <i>oikein</i>	<i>sillee</i> , <i>oikee</i>	

riittävästi. Vastaavasti pieneen aineistomäärään voi mahtua määrällisesti suuriakin muutoksia, jotka eivät ole tilastollisesti merkitseviä, koska niistä ei saada riittävän luotettavaa tietoa. Tällaisia muutoksia on hyvä tarkastella kvalitatiivisesti, mutta niistä ei voi aineiston määrän perusteella luotettavasti sanoa muutoksen todellisuudesta. Tilastollisten merkitsevyyksien lisäksi tutkimme jokaisen esimerkin osalta myös käytännön merkittävyyttä. Tätä tarkastelemme pääosin kohorttien suhteellisten ja absoluuttisten arvojen vertailun sekä laadullisen pohdinnan avulla.

Testaamme kahta väitettä, joista toinen koskee yhteisön muutosta ja toinen yksilöiden muutosta. Kukin tarkasteltu muutosmalli voi vaatia jommankumman tai kummankin väitteistä pitävän paikkansa.

Väite 1: Samojen ikäryhmien kielenkäytössä ei ole eroja eri vuosikymmenien välillä (esim. 1970-luvun nuoret ja 1990-luvun nuoret eivät eroa toisistaan).

Väite 2: Samojen puhujien kielenkäytössä ei ole eroja eri vuosikymmenien välillä (esim. 1970-luvun nuoret ja 1990-luvun keski-ikäiset eivät eroa toisistaan).

Väitteitä tilastollisesti testaamalla voimme löytää eri mallien kanssa yhteensopivat muutokset seuraavasti: Jos molemmat väitteet jäävät voimaan, muutosta ei ole havaittavissa ja tilanne vaikuttaa stabiililta. Jos ensimmäinen väite jää voimaan, mutta toinen väite hylätään, muutos on yhteensopiva syklisen ikäsidonnan muutosmallin kanssa: yhteisö ei muutu, mutta yksilöt muuttuvat. Tämän vastakohta, ensimmäisen väitteen hylkäys ja toisen jääminen voimaan, puolestaan viittaa sukupolvittaiseen muutokseen, koska idiolekit eivät muutu, mutta yhteisö muuttuu ajassa.

Jos molemmat väitteet hylätään, vaihtoehtoiksi jäävät yhteisöllinen, elinikäinen tai vetäytyvä muutos. Näiden erottaminen toisistaan vaatii muutoksen suunnan ja yksilöiden tarkempaa tarkastelua. Yhteisöllinen muutos vaatii kaikkien ikäryhmien muutosta samaan suuntaan samanaikaisesti. Elinikäisen muutoksen ehtona on toisen väitteen hylkäys siten, että keski-ikäiset ja vanhat puhujat kasvattavat variantin käyttöä edellisestä haastattelukierroksesta. Vetäytyvän muutoksen ehto on päinvastainen: vanhemmat puhujat vähentävät variantin käyttöä. Kaikissa kolmessa mallissa on kuitenkin ehtona myös ensimmäisen väitteen hylkäys eli yhteisön muutos.

Haluamme korostaa, että tässä esiteltävällä tilastollisella menetelmällä voimme tutkia ainoastaan sitä, onko jokin aineistossa havaittu variaatiotilanne ristiriidassa eri muutosmalleihin kuuluvien oletusten kanssa. Jos jonkin muutosmallin vaatimat ehdot (esimerkiksi stabiilissa tilanteessa molempien väitteiden jääminen voimaan) eivät tilastollisten testien mukaan pidä paikkaansa jonkin piirteen osalta, tämä muutosmalli voidaan sulkea pois selityksenä kyseisen piirteen variaatiolle – tässä esimerkissä siis tilanne ei voi olla ainakaan stabiili. Jos taas tilastollinen testaus ei hylkää mallin vaatimia ehtoja, voimme sanoa, että muutosmalli on yhteensopiva kyseisen piirteen variaation kanssa.

Testaamme väitteitä vertailemalla otoksia. Koska ensimmäisen väitteen tapauksessa puhujat eivät ole samoja, käytämme riippumattoman otoksen testiä. Vastaavasti toisessa

väitteessä puhujat ovat samoja, jolloin käytämme riippuvan otoksen testiä. Käsillä olevassa puhutun kielen aineistossa on selvästi vinoja jakaumia ja paljon jakauman ulkopuolisia havaintoja. Tästä syystä käytämme ei-parametrisiä testejä: Mann–Whitneyn U-testiä (Wilcoxon 1945; Mann–Whitney 1947) riippumattomille otoksille (väite 1) ja Wilcoxonin merkittyjen sijalukujen testiä (Wilcoxon 1945) riippuville otoksille (väite 2). Teemme kaikki testit suhteellisille osuuksille litteraattien pituuden ja piirteiden yleisyyden vaikutuksen pienentämiseksi. Tämä ei ole sijalukuihin perustuvissa testeissä ongelma, mutta saattaa korostaa joidenkin pienifrekvenssisten piirteiden muutoksen merkitsevyyttä. Tällaisen tilanteen sattuessa kuitenkin kommentoimme mahdollisia ongelmia.

Yksilöiden muutosta testaamme riippuvalla otoksella, jossa puhujat ovat samoja eri vuosikymmenien välillä. Taulukossa 3 olemme esittäneet tutkittavat otosjoukot. On huomattava, että kohortit ovat melko pieniä ja monissa tapauksissa painottuvat ylempään sosiaaliluokkaan. Tämä on tavallinen ongelma paneelitutkimuksissa (Paunonen 2006; Sankoff–Blondeau 2007; Sankoff 2018), koska alemman sosiaaliluokan puhujia on tyyppillisesti vaikeampi tavoittaa uuteen haastatteluun. Paneelitutkimuksissa yhteisön muutosnopeus saattaa näkyä hitaammalta kuin trenditutkimuksissa (Sankoff–Blondeau 2007, 581).

Taulukko 3. Usealla vuosikymmenellä haastatellut puhujat.

Syntymäaika	Haastattelut	Ikäryhmät	Puhujia yhteensä	Miehet	Naiset	Ylempi sosiaaliluokka	Alempi sosiaaliluokka
1927–1932	1970 ja 1990	40–60	9	4	5	7	2
1927–1932	1990 ja 2010	60–80	4	2	2	2	2
1952–1957	1970 ja 1990	20–40	10	6	4	8	2
1952–1957	1990 ja 2010	40–60	9	5	4	7	2
1971–1975	1990 ja 2010	20–40	13	6	7	7	6

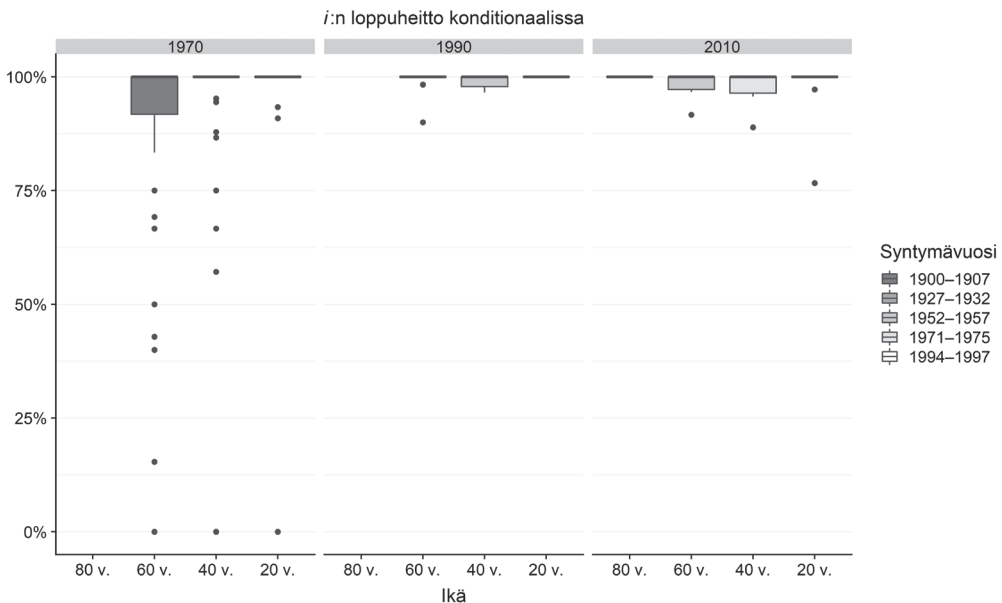
4 Muutosmallit Helsingin puhekielen aineistossa

Tässä luvussa esittelemme tutkimuksen keskeiset tulokset. Olemme tehneet kaikille tutkituille piirteille Mann–Whitneyn U-testin ja Wilcoxonin merkittyjen sijalukujen testin. Testien tulosten perusteella olemme valikoineet kutakin muutosmallia havainnollistamaan yhden tai useampia esimerkkipiirteitä. Koska piirteitä on paljon ja tarkastelumme on menetelmäkeskeinen, emme esittele piirteiden historiallista taustaa. Jokaisesta piirteestä annetaan kuitenkin lähdeviitteitä tutkimuksiin, joissa itse ilmiötä on tarkasteltu syvemmin.

Tilastollisen merkitsevyyden rajana pidämme p-arvoa 0,05 mutta ilmoitamme kuitenkin myös tarkat lukemat kahden desimaalin tarkkuudella, paitsi jos p-arvo on hyvin pieni (alle 0,01). Kuviolina käytämme samanlaisia ruutu- ja janakaavioita kuin luvussa 2. Samaa kohorttia edustavat puhujat on merkitty kuvioihin samalla värillä haastattelujasta riippumatta. Haluamme korostaa, että kyseessä ei aina ole täsmälleen sama joukko: esimerkiksi 1950-luvulla syntyneitä puhujia on 1970-luvun aineistossa 39, kun heitä 1990-luvulla on enää 10. Juoksevassa tekstissä vertailemme kuitenkin kohortin edustuksia niin, että kyseessä on nimenomaan sama puhujajoukko – tässä esimerkissä siis samat 10 puhujaa 1970-luvulta ja 1990-luvulta.

4.1 Stabiili tilanne

Stabiilia tilannetta on pidetty teoreettisena mahdollisuutena (esim. Paunonen 2005a, 15; Kurki 2013, 206), mutta tämä on osittain aineistolähtöistä (ks. luku 2). Omasta aineistostamme löytyy muutamia piirteitä, joissa ei tapahdu tilastollisesti merkitsevää muutosta sen paremmin yhteisön kuin yksilöidenkään kielessä. Voimme kuitenkin luokitella nämä piirteet kolmeen eri kategoriaan: hyvin pienifrekvenssisiin piirteisiin (esim. supistuma-verbit), jo lähestulkoon yksinomaisiin variantteihin (esim. *i*:n loppuheitto konditionaalissa) tai näiden vastakohtaan eli lähes täysin esiintymättömiin variantteihin (*r* yleiskielen *d*:n vastineena). Näistä esimerkkinä esitämme *i*:n loppuheiton konditionaalissa 3 (piirteestä Rapola 1966, 493).



Kuvio 3. *i*:n loppuheitto konditionaalissa.

Kuviosta 3 nähdään, että *i:n* loppuheiton variaatiotilanne on Helsingin puhekielen aineistossa jo hyvin vakiintunut. Ainoastaan 1970-luvun vanhoilla ja keski-ikäisillä puhujilla esiintyy variantin käytössä vielä hyvin alhaisiakin suhteellisia osuuksia. Nämä kaksi ryhmää ovat muutenkin aineistossa selvästi muita yleiskielisempiä (Paunonen 2005a, 37). Osuuksia tarkastelemalla huomataan, että ryhmien välillä on vielä pieniä eroja: esimerkiksi 1970-luvun vanhoilla puhujilla loppuheiton osuus on 92 prosenttia ($n = 350/381$), kun se 1990-luvun vanhoilla on 99 prosenttia ($n = 256/258$). Erot ryhmien välillä eivät kuitenkaan ole tilastollisesti merkitseviä ($p = 0,51$). Tämä ei kuitenkaan suoraan todista tilanteen olevan stabiili, vaan vain osoittaa, että aineiston perusteella stabiilia muutosmalleja ei voitu sulkea pois (ks. luku 3).

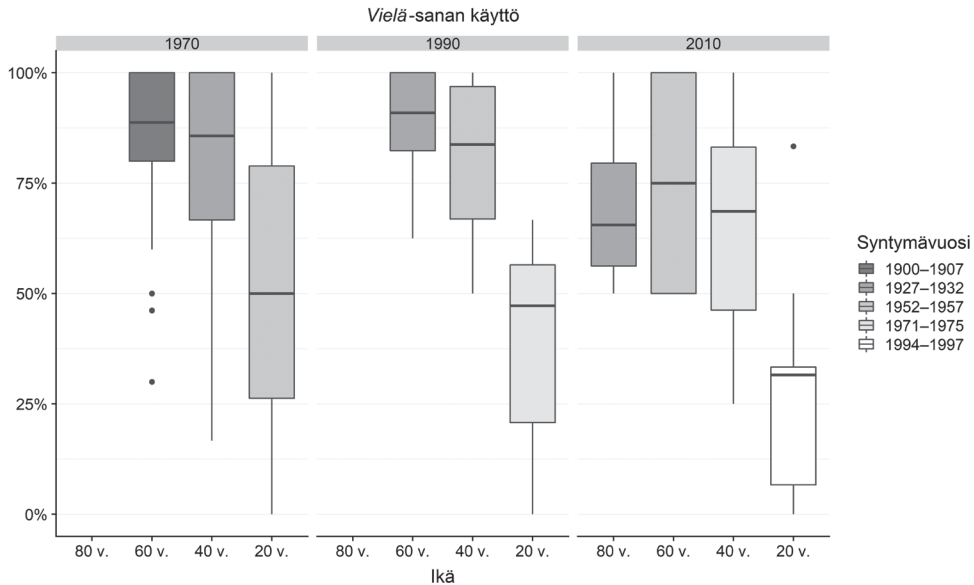
Koska stabiilin muutosmallin kanssa yhteensopivia variaatiotilanteita oli aineistossa vain muutamissa piirteissä, osoittautuu malli myös Helsingin puhekielen aineistossa harvinaiseksi. Muutosmallin tulkitseminen täysin teoreettiseksi riippuu pitkälti siitä, miten jo yksinomaisiin variantteihin suhtaudutaan. Sellaisia tapauksia, joissa varianttien välillä vielä olisi vaihtelua, mutta nämä vaihtelusuhteet pysyisivät kolmen (tai edes kahden) aikapisteen välillä tasaisina, ei ainakaan tästä aineistosta löydy. Konditionaalinen *i:n* loppuheitonkin tapauksessa näyttää siltä, että vielä esimerkiksi 1950-luvulta kerätyssä aineistossa variaatiota olisi saattanut olla huomattavasti enemmän.

4.2 Ikäsidonnan muutos

Myös toinen Labovin muutosmalleista, ikäsidonnan muutos, on tulkittu harvinaiseksi (Chambers 2009, 201). Tämä harvinaisuuden oletus perustuu ikäsidonnan mallin sykliseen luonteeseen: tietyt ikäryhmät poikkeavat muista aina uudelleen ja uudelleen. Myös tämä oletus käy toteen Helsingin puhekielen aineistossa. Ikäsidonnan muutoselle annetut ehdot toteutuvat vain kahden variantin osalta. Piirteistä toinen, *vielä*-sanana käyttö (pro *viel* ja *viä*), on esitetty kuviossa 4. Toinen ikäsidonnan mallin kanssa yhteensopiva variantti aineistossa on demonstratiivipronominien käyttö henkilöviitteissä (Vilppula 1989; Priiki 2017), jossa on aiemminkin havaittu ikäsidonnan tendenssiä (Lappalainen 2010, 287–288).

Tässä tutkimuksessa *vielä*-sanaa on tarkasteltu erillään siitä syystä, että siihen liittyy useita erilaisia kehityssuuntia: *A:n* loppuheitto (*viel*), diftongien avartuminen (*viäl*) ja eräänlainen välikonsonantin kato (*viä*) (Virtaranta 1946, 296–298). Vastaavia ilmiöitä esiintyy myös sanoissa *tuolla*, *siellä* ja *tiedä*, mutta näitä olemme tarkastelleet adessiivin yhteydessä (*tuolla* ja *siellä*) ja yleiskielen *d:n* vastineiden yhteydessä (*tiedä*). Diftongien avartumista ei ole Helsingin puhekielen aineistossa litteroitu johdonmukaisesti, joten sen tarkastelu jää tutkimuksen ulkopuolelle.

Puhujat kasvattavat *vielä*-sanana käyttöä aikuistuessaan. Erityisen selvästi muutos näkyy 1970-luvulla syntyneiden kohortissa: 1990-luvulla ryhmän puhujat käyttävät *vielä*-sanaa 42 tapauksessa 117:stä (36 %). Keski-ikässä 2010-luvulla ryhmän puhujat käyttävät yleiskielistä varianttia 94 tapauksessa 141:stä (67 %). Wilcoxonin testin mukaan *p*-arvo onkin tämän kohortin osalta tilastollisesti merkitsevä, 0,02. Myös 1950-luvulla



Kuvio 4. Vielä-sanan käyttö.

syntyneiden puhujien kohortissa ero on tilastollisesti merkitsevä ($p = 0,04$), mutta erot eivät ole ryhmän kokonaisuutta tarkasteltaessa numeerisesti yhtä suuria (1970-luvulla *vielä*-sanan osuus on 60 % ja 1990-luvulla 78 %). Tämä muutos näkyy selvästi myös kuviossa 4: nuorten ja keski-ikäisten välillä on aina selvä ero. Myös toinen ikäsidonaisuuden malliin kuuluva vaatimus, oletus yhteisön stabiiliudesta ja syklistyydestä, toteutuu tässä variantissa, sillä ryhmien välillä ei ole tilastollisesti merkitseviä eroja. Kuviossa 4 1970- ja 1990-lukujen variaatiotilanne näyttääkin lähes kivettyneeltä, mikä sopisi täydellisesti ikäsidonaisen muutoksen malliin. Kuitenkin 2010-luvulle tultaessa huomataan myös toisenlaista, yhteisöllistä muutosta: kaikki ryhmät ovat vähentäneet variantin käyttöä. Suurin pudotus tapahtuu keski-ikäisten ryhmässä, jonka edustus putoaa 11 prosenttiyksikköä (1990-luvulla 78 %, $n = 53/68$, 2010-luvulla 67 %, $n = 102/152$). Tämä pudotus ei muuta varianttien edustussuhteita, eikä U-testin mukaan kyseessä ole tilastollisesti merkitsevä muutos ($p = 0,13$), mutta täysin staattisesta tilanteesta ei voida puhua.

Tässä piileekin eräs muutosmallien soveltamisen ongelmakohta: tähän asti muutosmalleille on sovellettu lähinnä kahden aikapisteen reaaliaikakatutkimuksiin (esim. Sankoff-Blondeau 2007; Sankoff-Wagner 2011). Mikäli tilanne näyttää kahden aikapisteen välillä yhdeltä muutosmallilta, tehdään helposti johtopäätös, että muutos etenee kyseisen mallin mukaisesti. Kuitenkin kolmannen aikapisteen mukaan tuominen saattaa muuttaa kahden aikapisteen välillä havaitun muutostendenssin todennäköisyyttä radikaalistikin.

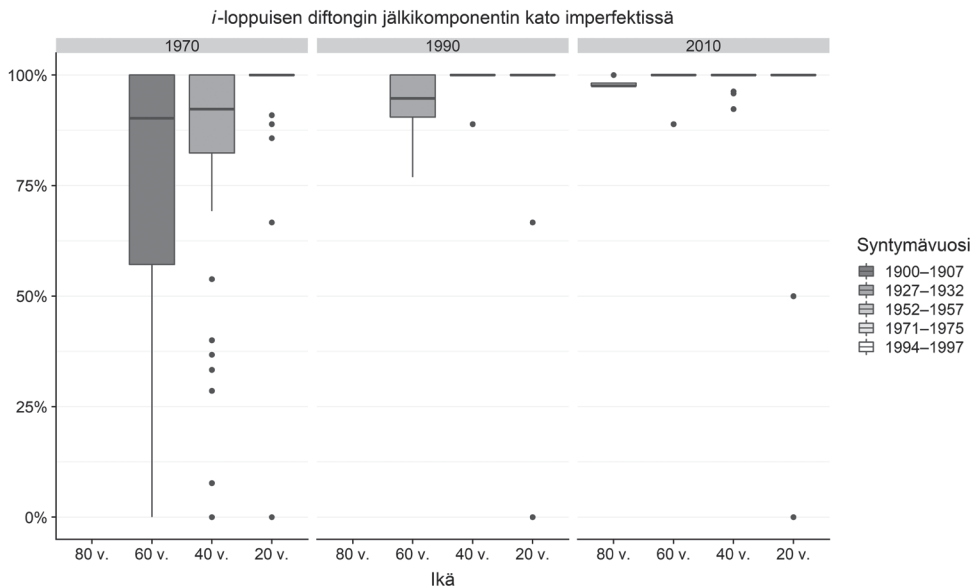
Vielä-sanan variaatiossa havaittu ikäsidonaisuus näyttää liittyvän muihinkin loppuheittoisiin variantteihin esimerkiksi adessiivissa ja elatiivissa, mikä tukisi tulkintaa, että kyse on oikeastaan samasta piirteestä (vrt. Mustanoja 2011, 287–289). Elatiiviin palataan

luvussa 4.6. Niin *vielä*-sanan kuin paikallissijojenkin variaatio voidaan tulkita tasaiseksi variaatioksi, joka ei ole välttämättä johtamassa minkäänlaiseen muutokseen. Tällaisissa piirteissä onkin kansainvälisissä tutkimuksissa löydetty ikäsidonnaista tendenssiä (esim. englannin *-ing* -päätevariaatio; Wagner 2012b).

4.3 Sukupolvittainen muutos

Siinä missä ikäsidonnaista muutosta on tavattu pitää harvinaisena muutosmallina, sukupolvittaista muutosta on yleensä pidetty normaalina muutoksen tapana (esim. Labov 1994, 84). Tämä näkemys perustuu oletukselle idiolektien pysyvyydestä eli niin sanottuun kriittisen iän hypoteesiin. Toisaalta sukupolvittaisen muutoksen korostuminen sociolingvistiksessä tutkimuksessa johtuu toisesta kriittisen iän hypoteesin seurauksesta eli näennäisaikamenetelmästä. Jotta näennäisaikamenetelmään voidaan luottaa, on muutoksen oltava ainakin osittain sukupolvittaista (tai stabiilia). Viimeisenä on tietenkin otettava huomioon, että myös reaaliaikatutkimuksissa on havaittu sukupolvittaisesti eteneviä muutoksia (ks. esim. Paunonen 2005a; Sankoff–Blondeau 2007).

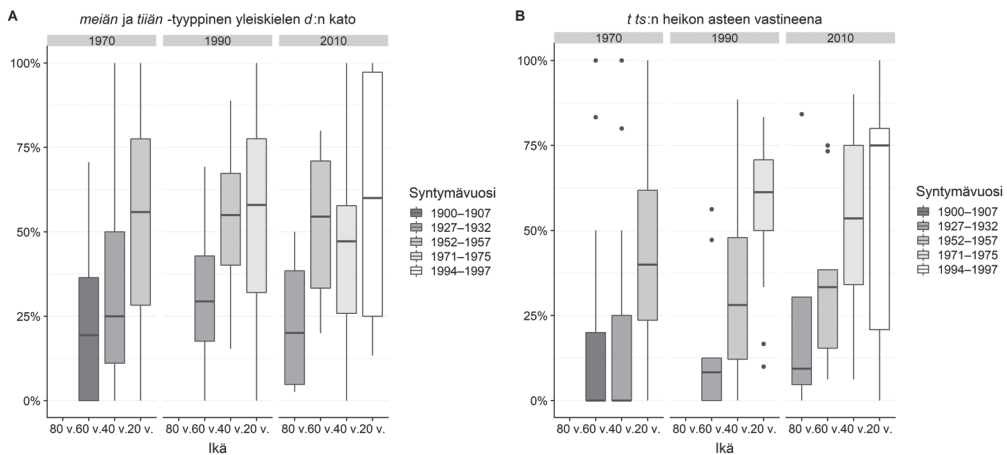
Myös käsillä olevassa tutkimuksessa sukupolvittainen muutos vaikuttaa huomattavan tavalliselta muutoksen tavalta. Sukupolvittainen muutos näyttää tässä aineistossa kiinnittyvän erityisesti muutoksen loppuvaiheessa oleviin tilanteisiin, joissa variantista tulee joko yksinomainen tai se poistuu käytöstä kokonaan. Tällaisia tapauksia on aineistossa kaikkiaan neljä kappaletta, joista esimerkkinä on kuviossa 5 esitetty *i*-loppuisen diftongin jälkikomponentin kato imperfektissä (*hän puhu eilen*) (Rapola 1966, 467–485).



Kuvio 5. *i*-loppuisen diftongin jälkikomponentin kato imperfektissä.

Kuviosta huomataan sukupolvittainen tendenssi, joka johtaa yhteisön muutokseen ajassa. Tilastollisten testien mukaan muutos on tosin merkitsevä vain keski-ikäisten osalta 1970- ja 1990-lukujen välillä ($p = 0,01$). 1970-luvulla keski-ikäiset käyttävät katoedustusta 80 prosentissa tapauksista ($n = 509/635$) ja 1990-luvulla 99 prosentissa ($n = 86/87$). Koska suhteelliset osuudet ovat 1990-luvulta eteenpäin jo hyvin lähellä sataa prosenttia, ryhmien vertailu ei tuota muissa tapauksissa tilastollisesti merkitseviä eroja. Esimerkiksi nuoret käyttävät katoedustusta 99-prosenttisesti niin 1970- kuin 1990-luvullakin ($n = 339/343$ ja $69/70$).

Sukupolvittaisen mallin kanssa yhteensopivia tapauksia on aineistossa kuitenkin myös selvästi varioivammassa tilanteissa. Kuviossa 6 esittelemme kaksi hieman erilaista tapausta, joissa molemmissa on kuitenkin pääasiallisena muutosmallina sukupolvittainen muutos.



Kuvio 6. Meiän ja tiiän -tyyppinen yleiskielen *d*:n kato (A) ja yksinäis-*t* *ts*:n heikon asteen vastineena (B).

Molemmat esimerkit poikkeavat kuvion 5 tapauksista siinä, että ne näyttävät olevan selvästi suuremman variaation alaisia. Aloitamme yleiskielen *d*:n kadosta määrälakseemisessä. Kyseessä on niin sanottu länsimurteen kato (Virtaranta 1946, 117–122; Paunonen 1995, 37–44), joka esiintyy monikollisissa persoonapronomineissa (*meiän*) ja *tietää*-verbin taivutuksessa (*tiiän*). 1970-luvun tilanne näyttää selvästi siltä, että yhteisössä on meilläään muutos: ikäryhmät eroavat toisistaan, ja nuoret ovat muutoksen kärjessä (ks. myös Paunonen 1995, 37). Seuraavassa aikapisteessä 1990-luvulla muutos näyttää edenneen odotuksenmukaisesti: puhujien idiolektit ovat pysyneet kohtuullisen vakaina. Esimerkiksi 1950-luvulla syntyneiden kohortissa on katoedustusta 1970-luvun aineistossa 56 prosenttia ($n = 161/288$) ja 1990-luvulla 55 prosenttia ($67/123$). Yhteisö puolestaan on muuttunut: 1970-luvun keski-ikäisillä katoedustusta on 31 prosenttia ($123/400$) ja 1990-luvun keski-ikäisillä 55 prosenttia ($67/123$). Näiden ryhmien välillä on tilastollisesti

merkitsevä ero ($p = 0,01$). Kuitenkin jo 1990-luvun pysäytyskuvassa on tapahtunut jotakin odottamatonta: nuoret eivät ole kasvattaneet katoesiintymäänsä juuri lainkaan, vaan keski-ikäiset ovat ottaneet nuorten ryhmän kiinni.

Sukupolvittaisen mallin uskottavuus tässä variantissa heikkenee entisestään 2010-luvulle tultaessa. Siinä missä 1950-luvulla syntyneet ovat edelleen säilyttäneet idiolektinsa vakaina (58 %, $n = 127/218$), 1970-luvulla syntyneiden osalta varianttien yleisyys-suhteet ovat muuttuneet. 1990-luvulla nuoruudessaan kohortin puhujat käyttivät katoedustusta 70-prosenttisesti ($n = 155/221$) ja 2010-luvulla vain 41-prosenttisesti (85/210). Wilcoxonin testin mukaan muutos ei kuitenkaan ole tilastollisesti merkitsevä ($p = 0,29$). On muistettava, että Wilcoxonin sijalukutesti vertaa keskenään yksilöiden edustuksia. Tässä tapauksessa yhdeksän kohortin 13 puhujasta vähentää katoedustusta nuoruudesta keski-ikään. Kolme puhujaa puolestaan kasvattaa osuuttaan reilusti. Näin ollen testi ei pidä eroa tilastollisesti merkitsevänä, koska ryhmän sisäinen varianssi on suurta. Kvalitatiivisesti tarkasteltuna näyttää kuitenkin siltä, että sukupolvittaisen muutoksen malli ei enää tämän kohortin osalta ole uskottava, vaan puhujien kielenkäyttö on muuttunut nuoruudesta keski-ikään siirryttäessä. Tässäkin tapauksessa siis kolmannen aikapisteen tuominen mukaan sotkee kahdessa aikapisteessä selvältä näyttäneen muutosmallin.

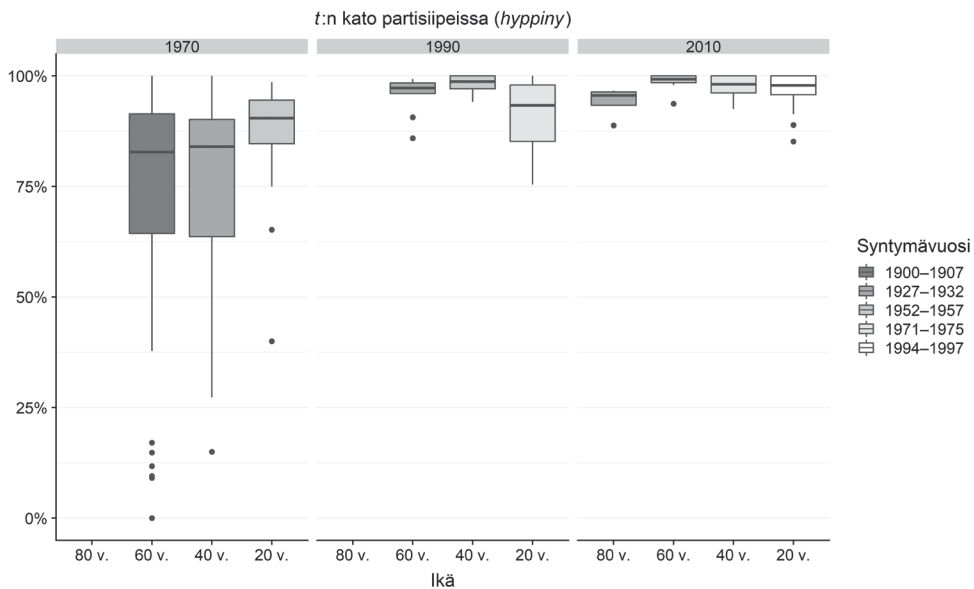
Toisessa esimerkkitapauksessa (kuviot 6B) tarkastelemme *t:n* käyttöä *ts:n* heikon asteen vastineena lekseemeissä *katsolkatse-*, *itse*, *viitsi-*, *etsi-* (Rapola 1966, 182–195; Paunonen 1995, 58–68). Tässä tapauksessa huomaamme muutamia sukupolvittaisen muutoksen mallista poikkeavia ilmiöitä: puhujat vähentävät variantin käyttöä nuoruudesta keski-ikään siirryttäessä ja kasvattavat sitä taas keski-ikästä vanhuuteen. Esimerkiksi 1970-luvulla syntyneillä muutos nuoruudesta keski-ikään on 57 prosentista ($n = 70/122$) 44 prosenttiin (107/243). Keski-ikästä vanhuuteen voidaan niin ikään havaita muutos esimerkiksi 1930-luvulla syntyneillä puhujilla: 1970-luvulla yksinäis-*t:n* osuus on kohortissa 11 prosenttia ($n = 7/66$) ja 1990-luvulla 22 prosenttia (35/156). Mikään näistä muutoksista ei kuitenkaan ole tilastollisesti merkitsevä, eli muutos on tilastollisesti tulkiten yhteensopiva sukupolvittaisen muutoksen mallin kanssa. Tämä johtuu pitkälti samasta syystä kuin yllä: puhujat muuttuvat ristiin. Voimme kuitenkin lingvistisesti tulkiten havaita idiolekteissa muutostendenssin.

Tällaista variaatiokaavaa, jossa seurauksena on *u:n* muotoinen käyrä (variantin käyttö vähenee keski-ikässä ja kasvaa eläköityessä) on pidetty tavallisena tapauksissa, joissa ei ole meneillään muutosta (Downes 1998, 224; Tagliamonte 2012, 47), ja sen on nähty olevan seurausta aikuisuuden paineista (ks. luku 2). Yksinäis-*t:n* käytössä *ts:n* heikon asteen vastineena on kuitenkin selvää, että meneillään on koko yhteisöä koskeva muutos. Tämän variantin osalta yksittäinen muutosmalli ei mitenkään riitäkään kuvaamaan itse tilannetta. Pääasiallinen muutoksen tapa on edelleen sukupolvittainen muutos, mutta mukana on selvää ikäsidonaisuutta, koska keski-ikässä käyttö vähenee ja vanhetessa käyttö jälleen kasvaa. Nuoruuden korkea tasoa ei kuitenkaan enää vanhana saavuteta. Kun vielä aina uusi nuorten ryhmä kasvattaa variantin käyttöä, voidaan todeta, että yhteisö muuttuu, mutta ei niin nopeasti kuin sukupolvittainen muutos antaa aiheen olettaa. Toisaalta muutos on nopeampaa kuin vetäytyvän muutoksen mallissa, joka ei oleta käytön kasvua enää

myöhemmissä ikävaiheissa. Tämä muotoilu on jo melko monimutkainen, mutta pohjimmiltaan kyse on siitä, että sukupolvittainen muutos kohtaa kielellisten markkinoiden periaatteen (Bourdieu–Boltanski 1975; Sankoff–Laberge 1978).

4.4 Yhteisöllinen muutos

Yhteisöllisen muutoksen vaatimuksena on sekä yksilöiden että yhteisön muutos samaan suuntaan. Tapauksia, joissa tällainen yhdensuuntainen muutos tapahtuu tilastollisesti merkittävästi, löytyy aineistosta vain yksi, NUT-partisiippien sananloppuisen *t:n* kato (kuvio 7). Piirteestä ovat kirjoittaneet esimerkiksi Rapola (1966, 314) ja Mielikäinen (1978, 108–112).



Kuvio 7. *t:n* kato partisiipeissa.

Partisiippien katomuodoissa muutos on jo 1970-luvulla melko pitkällä. 1990-luvulle tultaessa on kuitenkin tapahtunut selvä hyppäys: sekä vanhat että keski-ikäiset puhujat käyttävät katomuotoja lähes yksinomaisesti. Siinä missä 1970-luvun vanhojen katoedustus oli 72 prosenttia ($n = 1109/1538$) on se 1990-luvulla 96 prosenttia ($1067/1106$). Vastaavasti keski-ikäisillä muutos on 76 prosentista ($n = 1116/1462$) 98 prosenttiin ($664/676$). Tilastollisesti tarkasteltuna sekä vanhoilla että keski-ikäisillä 1970-luvulta 1990-luvulle siirtäessä *p*-arvo on erittäin merkittävä eli pienempi kuin 0,001. Erikoista on kuitenkin nuorten ryhmän staattisuus: kato-osuus on sekä 1970-luvulla että 1990-luvulla 92 prosenttia ($n = 1689/1842$ ja $533/577$, $p = 0,32$). Näin ollen täysin yhteisöllisestä muutoksesta ei voida tämänkään variantin kohdalla puhua.

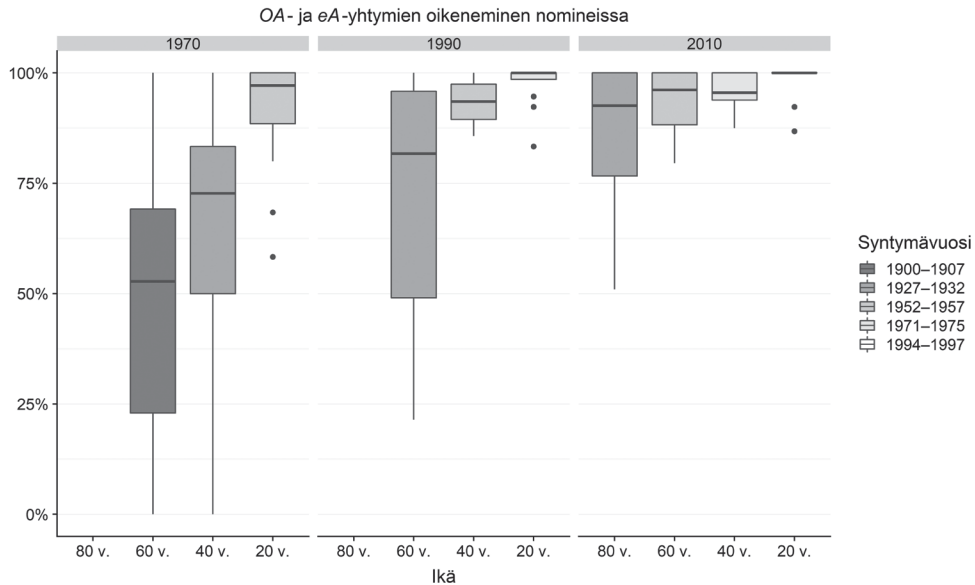
Myös idiolekteissa tapahtuu muutos 1970-luvun ja 1990-luvun välillä: 1930-luvulla syntyneiden osalta 74 prosentista ($n = 378/509$) 96 prosenttiin ($n = 1067/1106$, $p < 0,01$) ja 1950-luvulla syntyneiden osalta 94 prosentista ($n = 909/972$) 98 prosenttiin ($664/676$, $p = 0,02$). Koska katomuotojen käyttö lähentelee sataa prosenttia kaikissa ikäryhmissä 1990-luvulla ja 2010-luvulla, ei merkitsevää muutosta enää näiden vuosikymmenten välillä tapahdu. Onkin vaikea kuvitella tilannetta, jossa tapahtuisi yhteisöllinen muutos samassa piirteessä useamman kuin kahden aikapisteen välillä, koska tyypillisesti yhteisöllinen muutos on nopea, koko yhteisön siirtymä käyttämään tiettyä varianttia.

Yhteisöllinen muutos vaikuttaa tämän tutkimuksen aineistossa harvinaiselta muutoksen tavalta. Osittain tämä liittyy siihen, että yhteisöllinen muutos on yleensä yhdistetty leksikaalisiin piirteisiin, joita emme ole tässä tutkimuksessa juurikaan tarkastelleet. Toiseksi yhteisöllisen muutoksen ehdot ovat melko tarkat: kaikki yhteisön jäsenet muuttuvat samanaikaisesti samaan suuntaan ja mielellään vielä samantasoisesti. Jos emme seuraa vaatimuksia näin tarkasti, voimme löytää tämänkin tutkimuksen aineistosta tapauksia, joissa tapahtuu selvä yhteisöllisen muutoksen kaltainen käänne, kun jokin variantti pysäyttää toisen etenemisen. Kaikkein selvimmin muutos näkyy MA-infinitiivin variaatiossa. Tunnuksettomat illatiivimuodot (tapaukset, joissa MA-tunnus on kadonnut, mutta illatiivin sananloppuinen n on vielä säilynyt; *meen pelaan*) häviävät 1990-luvun ja 2010-luvun välillä lähes täysin käytöstä, kun puhujat käyttävät enenevässä määrin lyhyttä infinitiiviä (*meen pelaa*). Piirrettä on tarkasteltu samasta aineistosta jo aiemmin (Kuparinen 2018), joten emme tässä paneudu siihen syvemmin.

4.5 Elinikäinen muutos

Elinikäisen muutoksen testaaminen tilastollisesti on sikäli hankalaa, että mallin määritelmät ovat epäselviä. Esimerkiksi Sankoff (2005, 1011) määrittelee elinikäisen muutoksen niin, että osa vanhemmista puhujista muuttaa puhettaan meneillään olevan muutoksen suuntaan. Mikä sitten on riittävä osa, jotta voimme puhua elinikäisestä muutoksesta? Entä jos osa puhujista muuttaakin puhettaan päinvastaiseen suuntaan? Toisaalta elinikäistä muutosta on määritelty niin, että näennäisaikainen muutoksen kuva aliarvioi muutoksen nopeutta (Buchstaller 2015, 459). Tämä tarkoittaisi sitä, että vanhempien puhujien pitäisi muuttaa puhettaan suurimmaksi osaksi meneillään olevan muutoksen suuntaan eikä sitä vastaan – muutenhan muutoksen nopeus ei olisi suurempi kuin näennäisaikaisessa enusteessa (siis muutosmallien terminä sukupolvittaisessa muutoksessa).

Näistä lähtökohdista tarkastelemme yhtä esimerkkipiirrettä, jossa on ainakin jonkin verran elinikäisen muutoksen piirteitä. Kuitenkaan tilanteita, joissa elinikäinen muutos olisi selvä kolmen aikapisteen välillä, ei aineistosta löydy ainuttakaan. Monissa tapauksissa puhujat muuttuvat ristiin vanhemmalla iällä: osa puhujista muuttuu meneillään olevan muutoksen suuntaan ja osa sitä vastaan. Mikäli yhteisö ei kokonaisuutena muutu riittävästi, on vaikea perustella elinikäisen muutoksen olemassaoloa. Kuviossa 8 on esitelty OA- ja eA-yhtymien oikeneminen nomineissa (*korkee*) (Rapola 1966, 413–427).



Kuvio 8. OA- ja eA-yhtymien oikeneminen nomineissa.

Kuviosta 8 havaitaan elinikäisen muutoksen mukainen tendenssi 1930-luvun taitteessa syntyneiden puhujien edustuksessa: puhujat kasvattavat oienneiden muotojen käyttöä vuosikymmeneltä toiselle.³ Tälle kuvasta tehtävälle havainnolle löytyy myös tilastollista tukea: Wilcoxonin merkittyjen sijalukujen testin mukaan 1930-luvulla syntyneiden puhujien kielenkäyttö muuttuu tilastollisesti merkitsevästi 1970-luvulta 1990-luvulle eli keski-ikästä vanhuuteen ($p = 0,02$). Keski-ikässä kohortin puhujat käyttävät oienneita muotoja 51 prosentissa tapauksista ($n = 119/234$) ja 60-vuotiaina 73 prosentissa ($253/347$). Muissa kohorteissa ei kuitenkaan ole tilastollisesti merkitseviä eroja, ja kaikkien aikapisteen välillä esiintymien määrät hieman vähenevät nuoruudesta keski-ikään siirryttäessä. Esimerkiksi 1970-luvulla syntyneillä puhujilla tämä muutos on 98 prosentista ($n = 375/383$) 95 prosenttiin ($408/431$).

Yhteisö muuttuu eA- ja OA-yhtymien oikenemisen suuntaan tilastollisesti merkitsevästi vain 1970-luvulta 1990-luvulle siirryttäessä keski-ikäisten ryhmässä ($p < 0,001$). 1970-luvulla keski-ikäiset käyttävät oienneita muotoja 64 prosentissa piirteen esiintymistä ($n = 451/705$) ja 1990-luvulla jo 93-prosenttisesti ($251/270$). Samalla aikavälillä on myös vanhojen ryhmässä määrällisesti selvä muutos, vaikka testin tulos ei aivan ylläkään tilastolliseen merkitsevyyteen ($p = 0,06$). Ryhmän edustus muuttuu 1970-luvun 49 prosentista ($n = 306/622$) 1990-luvun 73 prosenttiin ($253/347$). 1990-luvun ja 2010-luvun

3 On huomioitava, että 2010-luvulla kohortissa on vain 4 puhujaa, eikä joukon tarkastelu ryhmänä ole erityisen mielekästä.

välillä ikäryhmiä vertailtaessa ei ole enää merkittäviä eroja, mikä johtuu osaltaan siitä, että esimerkiksi 2010-luvulla kaikkien ryhmien edustus on jo yli 90 prosentin.

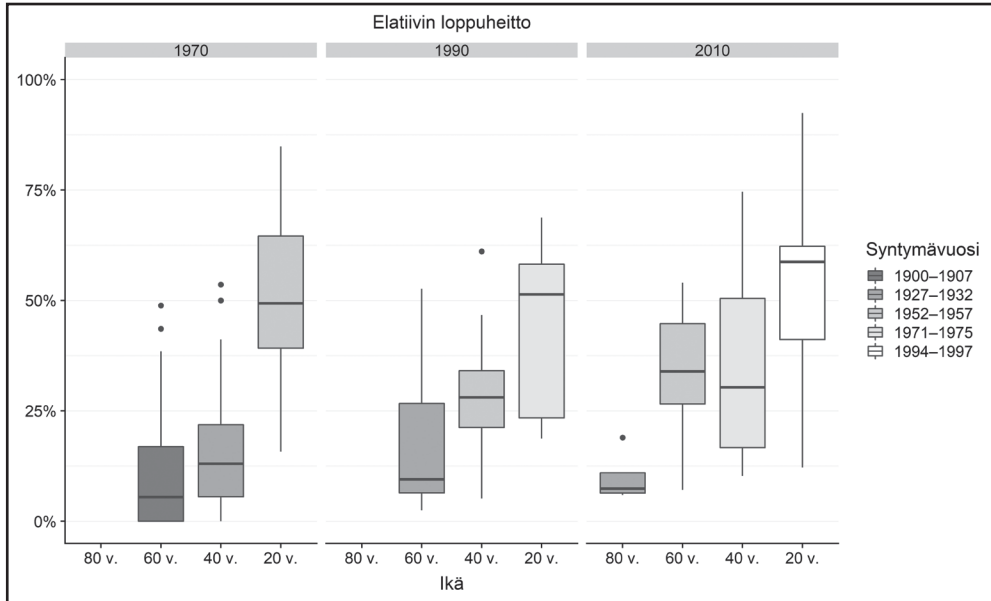
Myös *tt:n* käytössä *ts:n* vahvan asteen vastineena (*seitsemän*) toteutuu samantapainen, yhden puhujakohortin toteuttama elinikäinen muutos. Tässä variantissa 1950-luvulla syntyneiden puhujien kielenkäyttö muuttuu elinikäisen muutoksen mallin tavoin, mutta muissa kohorteissa vastaavaa tendenssiä ei ole. Elinikäinen muutos näyttää siis toteutuvan tilanteissa, joissa jokin puhujakohortti ottaa osaa meneillään olevaan nopeaan muutokseen. Esimerkiksi juuri *OA-* ja *eA-*yhtymien oikeneeseen malli sopii hyvin, sillä monofongivariantti on koko Suomessa edelleen voimakkaasti yleistyvää (Mantila 2004, 326–327). Sosiolingvistiikassa paljon hyödynnettyyn *S-käyrään* (Bailey 1973) verraten voisi tulkita niin, että *S-käyrän* nopeassa keskivaiheessa olevat muutokset hakeutuvat elinikäisen muutoksen malliin. Tämä on kuitenkin eräänlainen tautologia: jotta muutos voisi olla nopeassa vaiheessa, suurimman osan yhteisön jäsenistä on osallistuttava siihen eli muutoksen on oltava yhteisöllistä tai elinikäistä. *S-käyrän* ja muutosmallien suhde on kiinnostava teoreettisen pohdiskelun aihe, johon emme kuitenkaan tässä paneudu tarkemmin.

4.6 Vetäytyvä muutos

Vetäytyvässä muutoksessa puhujat vähentävät variantin käyttöä vanhemmalla iällä, mutta eivät niin paljoa kuin ikäsidonnaisessa mallissa. Tämä johtaa ennen pitkää myös yhteisön muutokseen, mutta muutoksen nopeus on hitaampi kuin sukupolvittaisen muutoksen mallissa. Tämän mallin osalta tilastollisen testaamisen ongelmat ovat pitkälti samanlaisia kuin elinikäisessä muutoksessa: puhujat muuttuvat paljon ristiin. Muutaman variantin käytössä tilastollista tukea vetäytyvälle muutokselle kuitenkin löytyy. Esimerkkinä tällaisesta variantista esittelemme elatiivin loppuheiton variaatiota kuviossa 9 (piirteestä Rapola 1966, 496–497).

On selvää, että elatiivin loppuheitossa on ikäefekti: nuoret ovat selvästi muita edellä. Koska tämän lisäksi yhteisökin muuttuu, muutos on vetäytyvää eikä pelkästään ikäsidonnaisista. Yhteisön muutos saa tilastollista tukea 1970-luvun ja 1990-luvun keski-ikäisten välillä ($p = 0,01$). 1970-luvulla keski-ikäisten esiintymistä loppuheittoa on 15 prosenttia ($n = 207/1425$) ja 1990-luvulla 29 prosenttia ($108/373$). Määrällisesti suuri muutos tapahtuu myös 1990-luvun ja 2010-luvun vanhojen puheessa, kun loppuheittoa edustus muuttuu 13 prosentista ($n = 81/628$) 33 prosenttiin ($156/471$). Muutos on tilastollisen merkitsevyyden rajalla ($p = 0,05$), mutta lingvistisesti voidaan puhua melko selvästä erosta.

Yksilötasolla suurin muutos tapahtuu 1950-luvulla syntyneiden puheessa 1970-luvun ja 1990-luvun välillä. Nuoruudessaan ryhmän puhujat käyttävät loppuheittoa muotoa 46 prosentissa tapauksista ($n = 285/625$) ja keski-ikässä 1990-luvulla vain 29 prosentissa ($108/373$). Tässäkään tapauksessa liikutaan aivan tilastollisen merkitsevyyden raja-arvolla ($p = 0,05$). Vastaavasti 1970-luvulla syntyneiden kohortin loppuheittoaisten varianttien määrä laskee 47 prosentista 33 prosenttiin. Puhujat siis vähentävät vanhetessaan variantin käyttöä, mutta eivät kokonaan luovu siitä. Tämän lisäksi on huomattava, että esimerkiksi 1950-luvulla syntyneiden kohortti kasvattaa loppuheittoaisten varianttien osuuttaan



Kuvio 9. Elätiivin loppuheitto.

keski-ikästä vanhuuteen, joskin vain 4 prosenttiyksikköä. Näin ollen muutos näyttää noudattavan u:n muotoista käyrää (ks. luvut 2.2 ja 4.3).

Ainakin tämän aineiston osalta näyttää siltä, että tällainen u:n muotoinen kaarevan muutoksen malli on huomattavasti tavallisempi kuin tasaisesti laskeva malli (vetäytyvä muutos). Kahden aikapisteen tutkimuksessa nuoruudesta keski-ikään selvästi laskeva variantin käyttö saattaa johdattaa tulkitsemaan muutoksen vetäytyväksi, kun taas kolmannessa aikapisteessä saman joukon muutos näyttääkin noudattavan u:n muotoista käyrää.

5 Lopuksi

Vaikka muutosmallien teoriaa voidaan pitää myös käytäntöön sopivana, näyttää siltä, että osa malleista toimii vain hyvin harvoissa tilanteissa. Esimerkiksi stabiilia tilannetta voidaan löytää vain varianteista, jotka ovat joko yksinomaisia tai kokonaan esiintymättömiä. Yhteisöllisen muutoksen ehtoihin sopiva muutos esiintyy aineistossa vain yhdessä piirteessä ja ikäsidonainainkin vain kahdessa.

Kaikkein yleisimmältä muutosmallilta vaikuttaa odotuksenmukaisesti sukupolvittainen muutos. Vaikka kriittisen iän hypoteesi ja näennäisaikamenetelmä ovat saaneet viime vuosina arvosteluakin osakseen, niiden kanssa yhteensopiva sukupolvittaisen muutoksen malli on kuitenkin edelleen hyvin uskottava. Tällainen muutos näyttää esiintyvän erityisesti tilanteissa, joissa muutos on lähestymässä loppuaan. On myös muistettava, että elinikäisen ja vetäytyvän muutoksen mallit perustuvat juuri sukupolvittaiselle muutokselle.

Vaikka muutosmalleja on täydennykset mukaan luettuina jo kuusi kappaletta, olemme törmänneet toistuvasti malliin, jota ei ole laskettu näiden kuuden mallin joukkoon. U:n muotoisen käyrän muodostava variaatiokaava nousi esiin sekä sukupolvittaisen muutoksen (luku 4.3) että vetäytyvän muutoksen (4.6) yhteydessä. Vanhastaan tällainen käyrä taas on liitetty ikäsidonaisuuteen (ks. luku 2.2). Kaarevan muutoksen yhdistäminen vain ikäsidonaisuudesta muuttuviin tasaisen variaation piirteisiin näyttää tämän tutkimuksen valossa liian kapealta käyttöalalta: esimerkiksi *t:n* käyttö *ts:n* heikon asteen vastineena (kuvio 6B) toteuttaa toisaalta kaarevaa mallia, mutta yhteisö kuitenkin selvästi myös muuttuu ajassa.

Kun kaareva muutos lasketaan mukaan, neljässä muutoksen tavassa on mukana jonkinlainen ikäefekti. Niin ikäsidonaisuudessa muutoksessa, elinikäisessä muutoksessa, vetäytyvässä muutoksessa kuin kaarevassakin muutoksessa idiolekteissa tapahtuu muutoksia ikäsidonaisuudesta. Onko näiden erottelu neljäksi eri muutosmalliksi mielekästä? Olisiko sittenkin syytä tulkita ikäsidonaisuus laavammin ja käsitellä sitä nimenomaan efektinä? Ikäefekteistä puhuu muun muassa Downes (1998, 239), ja näkemys lähestyy jo Hockettin (1950, 453) alkuperäistä ikäsidonaisuuden määritelmää, jonka mukaan samanikäiset puhujat muodostavat ryhmiä ja puhuvat samalla tavoin. Näin tulkittuna esimerkiksi ikäsidonainen muutos olisi stabiili tilanne ikäefektillä, kun taas elinikäinen muutos ja vetäytyvä muutos vastaisivat sukupolvittaisen muutoksen mallia ikäefektillä. Tällainen ajattelutapa toisaalta vähentäisi erilaisten mallien määrää ja toisaalta avaisi selvästi, mistä erilaiset variaatiokaavat ovat peräisin. Sen sijaan, että yritettäisiin päättää, mikä malli kuvaa tilannetta parhaiten (Labov 1994, 94–97; Paunonen 2005a, 48), voitaisiin tarkastella muutoksia erilaisten efektien yhdistelminä. Tämä saattaisi myös avata uudella tavalla poikkeavia variaatiokaavoja, sillä nykyiseen jaotteluun kiinnittyminen jättää jotkin muutoksen tavat väistämättä sivuun.

Muutosmallien teoriaan sisältyvä ikäkäsitys vaatii muutenkin tarkempaa tarkastelua. Esimerkiksi Sankoff (2005, 1011) määrittelee ikäsidonaisuuden hetkellisenä poikkeamana ja katsoo elinikäisen muutoksen edustavan lopullista muutosta. Onko tämä kuitenkin vain ikäkeskeistä harhaa? On helppo tulkita muutos nuoruuden ja keski-ikä välillä ikäsidonaisuudeksi, kun taas muutos keski-ikästä vanhuuteen jää tutkimuksen näkökulmasta lopulliseksi, koska puhujaa ei enää haastatella myöhemmin. Näin jälkimmäinen muutos näyttää todellisemmalta kuin ensimmäinen, vaikka molemmat olisivatkin yhtä lailla ikään tai elämäntilanteeseen sidottuja. Osaltaan tämä liittyy sosiaalitieteitä yleisemmin vaivaavaan ikäharhaan: nuoruuden ja vanhuuden muutoksia tarkastellaan ikään liittyvinä, kun taas keski-ikässä tapahtuvia idiolektin muutoksia lähestytään yksilön oman elämän verkostojen kautta (vrt. Eckert 1997, 157–158).

Muutosmallien kuvaama kielellinen todellisuus perustuu oletukselle yhteiskunnan muuttumattomuudesta. Kieli muuttuu tietyn mallin mukaan ikään kuin tyhjiössä, ulkoisten vaikuttimien ulottumattomissa. Vastaavanlainen oletus on sisäänrakennettu jo näennäisaikahypoteesiin (Laasanen 2016, 67–68). Tämä ei välttämättä ole ongelma sinänsä, jos muutosmalleja käytetään ennen kaikkea selittämään jo tapahtuneita muutoksia, jolloin

yhteiskunnassa tapahtuneet muutokset voidaan huomioida selityksen yhteydessä. Hankaluuksiin saatetaan sen sijaan päätyä, jos muutosmallien perusteella pyritään luomaan ennusteita tulevasta. Käsillä olevassa aineistossa selvä yhteiskunnallinen muutos voidaan havaita 1970-luvun ja 1990-luvun välillä. Paunonen (2005b, 195) toteaa, että 1970-luvun aineistoon saatiin vielä 'työläiskaupungin' ja 'virkamieskaupungin' edustajia, mutta tämän jälkeen kaupunginosadikotomiasta on täytynyt luopua. Tämä muutos näkyy aineiston puhujissa: siinä missä 1970-luvun vanhat puhuvat hyvin yleiskielisesti, 1990-luvun vanhat ovat huomattavasti puhekielisempiä (Paunonen 2006, 47). Muiden vuosikymmenten ja ikäryhmien välillä yhtä suuria eroja ei ole.

Labov pyrki muutosmallien teoriallaan osoittamaan, että näennäisaikametodi ei voi erottaa toisistaan stabiilia tilannetta ja yhteisöllistä muutosta eikä toisaalta ikäsidonnaista ja sukupolvittaista muutosta (Labov 1994, 84–85). Tähän hän tarjosi ratkaisuksi kahden aikapisteen reaaliaikatutkimuksia. Reaaliaikatutkimukset ovat näyttäneet toteen toisaalta Labovin teorian, mutta tuoneet myös paljon uutta tietoa esimerkiksi idiolektien muuttumisesta aikuisiällä (esim. Kurki 2005; Sankoff–Blondeau 2007; Mustanoja 2011; Buchstaller 2015). Tutkittuamme kieltä kolmessa aikapisteessä olemme huomanneet, että kahdenkin aikapisteen tutkimus saattaa olla vajavainen: yksittäinen malli toimii muutoksen kuvana harvoin yli kolmen aikapisteen (ks. myös Wagner 2012a, 378). Jokin muutos näyttää kahden aikapisteen välillä selvästi yhdeltä mallilta, mutta kolmas aikapiste ei enää toteutakaan yhtä selvästi tätä mallia. Muutos on saattanut esimerkiksi pysähtyä (esim. kuvio 6A) tai ottaa harppauksen eteenpäin (kuvio 4). Tämä tutkimus osoittaa, että kahden aikapisteen – ja luultavasti kolmenkin – reaaliaikatutkimusten perusteella tehtyihin ennusteisiin kannattaa suhtautua varoen. Tämä on luontevaa jatkoa Labovin alkuperäiselle ajatukselle: reaaliaikaiset tutkimukset toivat esiin näennäisaikametodin mahdolliset puutteet, mutta eivät tuominneet sitä kokonaan käyttökelttomaksi. Vastaavasti tässä olemme nostaneet esiin kahden aikapisteen reaaliaikatutkimusten mahdollisia puutteita, jotta ne voidaan ottaa tulevaisuudessa paremmin huomioon.

Lähteet

Aineistolähteet

Helpuhe 2014 = Helsingin puhekielen pitkittäiskorpus (1970, 1990, 2010) [online-puhekorpus]. - Helsinki. Helsingin yliopiston suomen kielen, suomalais-ugrilaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos, Kotimaisten kielten keskus ja Heikki Paunonen, 2014. URN: <http://urn.fi/urn:nbn:fi:lb-2014073041>.

Kirjallisuuskäsitteet

BAILEY, CHARLES-JAMES 1973: *Variation and linguistic theory*. Center for Applied Linguistics, Arlington.
 BOURDIEU, PIERRE - BOLTANSKI, LUC 1975 : Le fétichisme de la langue. *Actes de la recherche en sciences sociales* 1, 1–32. <https://doi.org/10.3406/arss.1975.3417>
 BUCHSTALLER, ISABELLE 2015: Exploring linguistic malleability across the life span. Age-specific patterns in quotative use. *Language in Society* 44, 457–496. <https://doi.org/10.1017/s0047404515000391>.

- CEDERGREN, HENRIETTA 1987: The spread of language change: Verifying inferences of linguistic diffusion. PETER LOWENBERG (toim.): *Language spread and language policy: Issues, implications, and case studies*, 45–60. Georgetown University Press, Washington, DC.
- CHAMBERS, J. K. 2009: *Sociolinguistic theory*. Kolmas painos. Wiley-Blackwell, Malden.
- DOWNES, WILLIAM 1998: *Language and society*. Toinen painos. Cambridge University Press, Cambridge.
- ECKERT, PENELOPE 1997: Age as a sociolinguistic variable. FLORIAN COULMAS (toim.): *The handbook of sociolinguistics*, 151–167. Blackwell, Oxford.
- HOCKETT, CHARLES 1950: Age-grading and linguistic continuity. *Language* 26, 449–457. <https://doi.org/10.2307/410396>
- KLEENE, STEPHEN 1956: Representation of events in nerve nets and finite automata. *Automata Studies (Annals of Mathematics Studies 34)*, 3–41. <https://doi.org/10.1515/9781400882618-002>
- KUPARINEN, OLLI 2018: Infinitiivien variaatio ja muutos Helsingissä. *Virittäjä* 122, 29–51. <https://doi.org/10.23982/vir.65310>
- KURKI, TOMMI 2005: *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen*. SKST 1036.
- 2013: Näennäisaikametodi ja muutosmallit. *Sananjalka* 55, 202–212.
- LAASANEN, MIKKO 2016: Näennäisaikametodin taustaoletusten teoreettista tarkastelua. *Virittäjä* 120, 57–83.
- LABOV, WILLIAM 1994: *Principles of linguistic change vol 1. Internal factors*. Blackwell, Oxford.
- LAPPALAINEN, HANNA 2010: Hän vai se, he vai ne? Pronominivariaatio ja normien ristiveto. HANNA LAPPALAINEN, MARJA-LEENA SORJONEN ja MARIA VILKUNA (toim.): *Kielellä on merkitystä. Näkökulmia kielipolitiikkaan*, 279–324. SKST 1262.
- LENNEBERG, ERIC 1967: *Biological foundations of language*. Wiley, New York. <https://doi.org/10.1080/21548331.1967.11707799>
- MANN, HENRY B. - WHITNEY, DONALD R. 1947: On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* 18, 50–60. <https://doi.org/10.1214/aoms/1177730491>
- MANTILA, HARRI 2004: Murre ja identiteetti. *Virittäjä* 108, 322–346.
- MEYERHOFF, MIRIAM 2011: *Introducing Sociolinguistics*. Toinen painos. Routledge, London.
- MIELIKÄINEN, AILA 1978: Aktiivin II partisiippi Etelä-Savon murteissa. *Virittäjä* 82, 101–121.
- MUSTANOJA, LIISA 2011: *Idiolekti ja sen muuttuminen. Reaaliaikatuutkimus Tampereen puhekielestä*. Acta Universitatis Tampereensis 1605. Tampere University Press, Tampere.
- PALANDER, MARJATTA 2005: *Lapsuudesta keski-ikään. Seuruututkimus itäsavolaisen yksilömurteen kehityksestä*. Suomi 191. Suomalaisen Kirjallisuuden Seura, Helsinki.
- PAUNONEN, HEIKKI 1995 [1982]: *Suomen kieli Helsingissä*. Helsingin yliopiston suomen kielen laitos, Helsinki.
- 2005a: Totta vai toiveajattelua. Näennäisaikainen muutos todellisen muutoksen kuvastajana. JOHANNA VAATTOVAARA, TONI SUUTARI, HANNA LAPPALAINEN ja RIHO GRÜNTAL (toim.): *Muuttuva muoto. Kirjoituksia Tapani Lehtisen 60-vuotispäivän kunniaksi*, 13–53. Kieli 16. Helsingin yliopiston suomen kielen laitos, Helsinki.
- 2005b: Helsingiläisiä puhujaprofileja. *Virittäjä* 109, 162–200.
- 2006: Vähemmistökielestä varioivaksi valtakieleksi. KAISU JUUSELA ja KATARIINA NISULA (toim.): *Helsinki kieliyhteisönä*, 13–99. Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.
- PRIIKI KATRI 2017: Hän, se, tää vai toi? *Vuorovaikutussociolinguvistinen tutkimus henkilöviittauksista Kaakkois-Satakunnan nykypuhekielessä*. Annales Universitatis Turkuensis C432. Turun yliopiston julkaisuja, Turku.
- RAPOLA, MARTTI 1966: *Suomen kielen äännehistorian luennot*. SKST 283.
- R Core Team 2017: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Saatavilla <https://www.R-project.org/>.
- SANKOFF, DAVID - LABERGE, SUZANNE 1978: The linguistic market and the statistical explanation of variability. DAVID SANKOFF (toim.): *Linguistic variation: Models and methods*, 239–250. Academic Press, New York.

- SANKOFF, GILLIAN 2005: Cross-sectional and longitudinal studies. NORBERT DITTMAR, KLAUS J. MATTHEIER ja AMMON ULRICH (toim.): *Sociolinguistics. An international handbook of the science of language and society*, 1003–1013. de Gruyter, Berlin.
- 2018: Before there were corpora: The evolution of the Montreal French project as a longitudinal study. SUZANNE EVANS WAGNER ja ISABELLE BUCHSTALLER (toim.): *Panel Studies of Variation and Change*, 21–52. Routledge, Oxford.
- SANKOFF, GILLIAN - BLONDEAU, HÉLÈNE 2007: Language change across the lifespan: /r/ in Montreal French. *Language* 83, 560–588. <https://doi.org/10.1353/lan.2007.0106>
- SKST = Suomalaisen Kirjallisuuden Seuran Toimituksia. Suomalaisen Kirjallisuuden Seura, Helsinki.
- TAGLIAMONTE, SALI 2012: *Variationist sociolinguistics. Change, observation, interpretation*. Wiley-Blackwell, Oxford.
- THOMPSON, KEN 1968: Regular expression search algorithm. *Communications of the ACM* 11: 419–422. <https://doi.org/10.1145/363347.363387>
- VILPPULA, MATTI 1989: Havainnot *hän-* ja *he-*pronominiin käytöstä suomen murteissa. *Virittäjä* 93, 389–400.
- VIRTARANTA, PERTTI 1946: *Länsiyhästakuntalaisten murteiden äännehistoria I. Konsonantit*. SKST 230.
- WAGNER, SUZANNE - SANKOFF, GILLIAN 2011: Age grading in the Montréal French inflected future. *Language Variation and Change* 23, 275–313. <https://doi.org/10.1017/S0954394511000111>
- WAGNER, SUZANNE EVANS 2012a: Age grading in sociolinguistic theory. *Language and Linguistics Compass* 6, 371–382. <https://doi.org/10.1002/lnc3.343>
- 2012b: Real-time evidence for age grad(ing) in late adolescence. *Language Variation and Change* 24 s. 179–202. <https://doi.org/10.1017/S0954394512000099>
- WILCOXON, FRANK 1945: Individual comparisons by ranking methods. *Biometrics Bulletin* 1, 80–83. <https://doi.org/10.2307/3001968>

OLLI KUPARINEN, LIISA MUSTANOJA, JAAKKO PELTONEN, JENNI SANTAHARJU, and UNNI LEINO: *Patterns of change in a longitudinal corpus of three points in time*

This article applies William Labov's theory of patterns of change to the *Longitudinal Corpus of Finnish Spoken in Helsinki*. The corpus consists of interviews recorded in three points in time: the 1970s, 1990s and 2010s. This article sets off to examine whether the studied features change in accordance with the patterns over these three points in time. The study also analyses underlying presumptions and possible problems in the theory. In addition to these theoretically important objectives, this article provides information on the variation and change in Finnish spoken in Helsinki.

There is at least one linguistic feature compatible with each pattern of change in the data. However, some of the patterns seem to appear only in specific situations. Generational change is the most common of the patterns, which reaffirms the findings of previous studies. This study also identifies a connection between the stages of change and patterns of change: a generational pattern, for example, seems to be common in changes nearing completion.

One crucial finding in this study is that a singular pattern of change rarely occurs over three points in time. A change that follows a certain pattern between two points in time might stop or advance on to a third point. Labov's original theory set out to prove that studies in apparent time could not distinguish the patterns of change that appear to be similar synchronically. This article indicates that a real time study of two points in time might likewise be falsely interpreted.

Olli Kuparinen
olli.kuparinen@tuni.fi
Informaatioteknologian ja viestinnän tiedekunta
Kanslerinrinne 1
PL 300, 33014 Tampereen yliopisto

Liisa Mustanoja
liisa.mustanoja@tuni.fi
Informaatioteknologian ja viestinnän tiedekunta
Kanslerinrinne 1
PL 300, 33014 Tampereen yliopisto

Jaakko Peltonen
jaakko.peltonen@tuni.fi
Informaatioteknologian ja viestinnän tiedekunta
Kanslerinrinne 1
PL 300, 33014 Tampereen yliopisto

Jenni Santaharju
jenni.leppanen@helsinki.fi
Bio- ja ympäristötieteellinen tiedekunta
Viikinkaari 1
PL 65, 00014 Helsingin yliopisto

Unni Leino
unni.leino@tuni.fi
Informaatioteknologian ja viestinnän tiedekunta
Kanslerinrinne 1
PL 300, 33014 Tampereen yliopisto

JULKAISU

III: LEKTIT

Lects in Helsinki Finnish – a probabilistic component modeling approach

Kuparinen, Olli – Peltonen, Jaakko – Mustanoja, Liisa – Leino, Unni – Santaharju,
Jenni

Käsikirjoitus, hyväksytty julkaistavaksi julkaisussa Language Variation and Change

Artikkelin käyttöön väitöskirjan osana on saatu kustantajan lupa

Lects in Helsinki Finnish - a probabilistic component modeling approach

Olli Kuparinen, Tampere University

Jaakko Peltonen, Tampere University

Liisa Mustanoja, Tampere University

Unni Leino, Tampere University

Jenni Santaharju, University of Helsinki

Olli Kuparinen

Faculty of Information Technology and Communication Sciences

Kanslerinrinne 1 (Pinni B)

FI-33100 Tampere

olli.kuparinen@tuni.fi

Short title:

Lects in Helsinki Finnish

Accepted for publication in Language Variation and Change.

Lects in Helsinki Finnish - a probabilistic component modeling approach

Abstract

This article examines Finnish lects spoken in Helsinki from the 1970s to the 2010s with a probabilistic model called Latent Dirichlet Allocation. The model searches for underlying components based on the linguistic features used in the interviews. Several coherent lects were discovered as components in the data, which counters the results of previous studies that report only weak co-variation between features that are assumed to present the same lect. The speakers, however, are not categorical in their linguistic behavior and tend to use more than one lect in their speech. This implies that the lects should not be considered in parallel with seemingly uniform linguistic systems such as languages, but as partial systems that constitute a network.

Keywords: lect, coherence, real time change, Finnish, Latent Dirichlet Allocation

The research project was funded by the Kone Foundation.

The coherence of linguistic varieties has been increasingly questioned in recent years, for example in studies on ethnolects (Boyd & Fraurud, 2010; Wolfram, 2007), dialects (Gregersen & Phrao, 2016), sociolects (Guy, 2013) and registers (Geeraerts, 2010). Results indicate that the lects labeled by linguists and laymen seem not to be realized in actual language use, and they should thus be reconsidered. The issue of lectal coherence (Guy, 2013) has been examined by focusing on patterns of covariation instead of singular variables (e.g., Gross, 2018; Oushiro, 2016).

The sociolinguistic research on covariation patterns was introduced in Labov's study of English in New York City (Labov, 1966: 209–211). The study presented two approaches to stratification of linguistic data: social grouping and linguistic grouping. Social grouping (the assignment of speakers to sociologically defined groups) gained considerable popularity among the field after Labov's initial study, whereas linguistic grouping (search for patterns of linguistic behavior) has surfaced more rarely until recently (e.g., Horvath & Sankoff, 1987; Ma & Herasimchuk, 1972; Thelander, 1979).

Social grouping does not scrutinize linguistic systems per se but seeks to distinguish how (perceived) social systems affect linguistic features. If multiple features seem to behave uniformly in regard to social factors, the collection of these forms is considered a variety or a lect – a linguistic system under certain circumstances. Therefore, the varieties discovered in studies built on social grouping (i.e., sociolects) are reified by social factors such as class ('middle class speech'), ethnicity ('African-American English') or locale ('Rinkeby Swedish'). This sort of essentialist approach has several problems that have been discussed at length (e.g., Bucholtz, 2003; Pratt, 1987).

Linguistic grouping starts from language use and searches for patterns of covariation. The issue with linguistic grouping is that the discovered patterns are often difficult to interpret. For instance, Ma and Herasimchuk (1972: 271) discard four of ten patterns found in their study in part because "co-occurring items made little linguistic or sociolinguistic 'sense' as a unity". The discovery, although disregarded in the study, is of utmost importance. The items were found to co-occur, but not in a

way the researchers expected: the lects were not distinct and identifiable but rather obscure and overlapping.

The article at hand contributes to the field of lectal coherence by searching for lects of Finnish spoken in Helsinki. Based on real time interview data from the 1970s to the 2010s, we examine how phonological, morphological, and lexical features co-vary and thus constitute lects in Helsinki. We define a lect as a pattern of several frequently co-occurring linguistic features. This definition is entirely usage-based and does not require social similarity of the speakers (cf. sociolect). Whereas many of the studies mentioned focus on a small number of features (e.g., Guy, 2013; Oushiro, 2016), we scrutinize 34 linguistic alternation variables with 78 possible variants. We do not restrict to mutually exclusive patterns; each variant can appear in several lects and each speaker can exhibit features of multiple lects in their speech. To uncover the lects, a probabilistic component model called Latent Dirichlet Allocation (Blei, Ng & Jordan, 2003) is used. The present article addresses the following questions:

1. Which linguistic variants co-vary and thus constitute the lects?
2. Can the discovered lects be traced to previously proposed lects of Finnish spoken in Helsinki, such as the standard language, the Häme dialect or the Helsinki slang (Paunonen, 1994)?
3. Are the speakers likely to use single lects or combine them? And if so, which lects combine?

The growing body of work on lectal coherence suggests that the co-variation among features is likely to be weak (cf. Guy, 2013), and that there will be overlap and uncertainty between the lects (cf. Ma & Herasimchuk 1972). Thus, we expect that although we might discover lects that can be traceable to previously proposed lects of Finnish spoken in Helsinki, they will not be easily identifiable. Similarly, we expect idiolects to be combinations of several underlying lects (cf. Geeraerts, 2010).

We commence with a brief history of the linguistic landscape of Helsinki, which clarifies why the Finnish capital is of particular interest when studying lectal coherence. We then introduce our data and the model used in more detail, before proceeding on to the results.

Helsinki as a linguistic mosaic

Helsinki was founded in 1550 but remained a minor town until the late 18th century. Sweden lost the territories of Finland to Russia in the Finnish war, and the new Russian regime moved the capital from Turku to Helsinki in 1812. At the time, the population of Helsinki was approximately 4000 and almost entirely Swedish speaking (Paunonen, 1994; Waris, 1951). Considering that the study at hand scrutinizes Finnish varieties, this is highly important. As the settlement's history is relatively short and the population was mostly Swedish speaking, the city lacks a Finnish dialect basis. The linguistic situation of the early 20th century is illustrated in Figure 1.

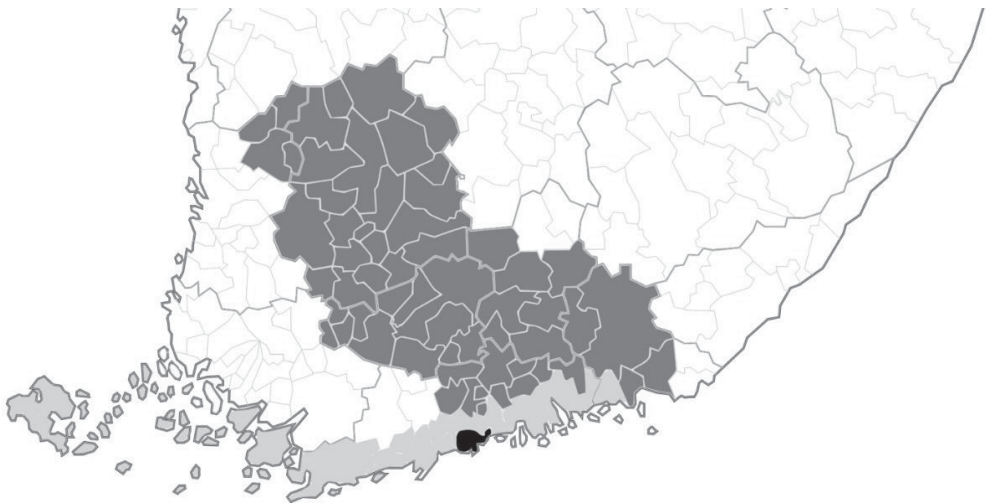


Figure 1. Map of southern Finland. Present-day Helsinki in black, Swedish-speaking areas in light gray and the areas of the Häme dialect in dark gray (based on Itkonen, 1989). The map represents the linguistic situation in the early 20th century.

The capital status and industrialization rapidly increased the number of residents and spurred migration from Finnish speaking areas. The migrants were mostly from neighboring regions Uusimaa and Häme. The Finnish variety spoken for the most part in these areas was the Häme dialect, that is of interest also in this study (dark

gray in figure 1). The dialect itself is divided into several subdialects (Itkonen, 1989), from which the Southern Häme dialect is of particular interest in our study. Although there was migration from the Eastern parts of Finland as well, the features of the Eastern dialects never gained much popularity in Helsinki (Paunonen, 2006). The language situation changed speedily: from 10 percent of Finnish speaking residents in 1850 to 45.5 percent in 1890 (Åström, 1956). In the beginning of the 20th century Finnish was already the majority language in Helsinki.

In the wake of the 19th century nationalist movement several members of the Swedish-speaking elite decided to start speaking Finnish between each other and in their homes (Paunonen, 1994: 229). As stated earlier, there was no base dialect of Finnish in Helsinki to learn, which led the elite to speak a standardized version of the language. The Finnish standard language is an artificial creation that combines features of the Western and Eastern dialects and is thus not based on any single natural dialect (Lehikoinen & Kiuru, 1989). It is also relatively new, as the standardization was executed as late as the 19th century. The standard language as a spoken variety is highly salient, as it differentiates from all the dialects. We argue that we are almost certain to discover a lect traceable to standardized Finnish in the results of this study.

In the working-class suburbs Finnish and Swedish co-existed, generating the old Helsinki slang around the turn of the 20th century. The slang combined Swedish, Finnish, and Russian lexis, utilized Swedish phonemes (e.g., /b/) and was spoken by speakers of Finnish- and Swedish-speaking descent (Paunonen, 1994). The old Helsinki slang has faded since the second World War and the new Helsinki vernacular is generally Finnish based with new loan words coming mostly from English.

The amount of people speaking other languages than Finnish or Swedish as their native language has risen rapidly in the last 30 years in Helsinki. In 1990 only 1.3% of residents were speakers of other languages, whereas in 2019 the proportion was 15.7% (Mäki & Vuori, 2019). Unfortunately, speakers of other languages cannot be reached in the study at hand, as the data is exclusive to first language speakers of Finnish. The increase of other languages is bound to affect the lects of Finnish spoken in Helsinki in the future (cf. Gross, Boyd, Leinonen & Walker, 2016).

Helsinki has seen several language and dialect contacts in the past 200 years: between Swedish and Finnish and Russian, between Finnish dialects, between dialects and the standard language and between the increasing number of new languages in recent decades. Mufwene (2001: 4–6) describes such a situation as producing a feature pool, from which new varieties are constructed and modified (cf. Cheshire, Kerswill, Fox & Torgersen, 2011). The old Helsinki slang is an example of an output variety that was composed of features of the pool selectively and constituted a unique combination of the input varieties.

In a similar vein but for stylistic practices, Eckert (2008) describes the process of bricolage (see also Hebdige, 1979). In bricolage linguistic variables (and other social resources) are distinct features of a style. Once a feature is perceived, the observer can select to use it or not to use it in their own speech. The use of a feature in a new setting changes both the meaning of the feature and the originally perceived style. Thus, bricolage might disrupt the coherence of lects as a speaker decides to use a feature not originally belonging to a certain lect or, condense the lects in even more coherent bundles as speakers accommodate to lectal norms in their selections. For a contrastive study of a variety approach and a stylistic approach to lects, see Quist (2008).

To conclude, the input varieties of Helsinki Finnish can be simplified to three separate lines: the highly standardized Finnish introduced by the elite, the rural dialects (especially Häme) of the migrants and the slang of the working-class neighborhoods. Residents of the capital area themselves define the lects spoken in Helsinki mostly in relation to the old slang or the standard language, using characterizations such as ‘modern Helsinki slang’, ‘neither book language nor slang’ and ‘standard with a few slang words’ (Vaattovaara & Soininen-Stojanov, 2006). In the same study, most cited features of Helsinki Finnish are slang words, fast tempo, and the *Helsinki-s* (see Halonen & Vaattovaara 2017). Thus, the dialectal features, which in reality form the basis of the spoken language in the area (Paunonen, 1995) are not perceived or identified by the speakers.

Data

Longitudinal Corpus of Finnish Spoken in Helsinki

The data analyzed in this article is a subset of the Longitudinal Corpus of Finnish Spoken in Helsinki (Helpuhe). The corpus project started in the beginning of the 1970s as a part of a sociolinguistic enterprise to study the spoken language of Finland's biggest cities. In the guidelines of the sociolinguistic studies of the time, the method was to conduct interviews of the city's native residents with varying social backgrounds. The original endeavor involved three age groups, three social classes, two socially different districts and two genders. The neighborhoods chosen in the original study were academically inclined Töölö and a working-class district Kallio-Sörnäinen, which was also the birthplace of the old Helsinki slang (Paunonen, 2006). In total, 149 speakers were interviewed between 1972–1974, from which a corpus of 96 speakers was created.

The project continued in 1991–1992, when the two youngest age groups of the 1970s study were pursued again. Nearly half of them were reached (29/64) and interviewed again, supplemented with a new group of youngsters. In total the number of interviewees is significantly lower in the 1990s subset than in the 1970s (45 and 96 respectively). The youngest group was chosen using the same guidelines as in the 1970s, but the distinction between three social classes was simplified to present high schoolers and vocational school attendees. Moreover, the demographic conditions of the Kallio-Sörnäinen district had changed in the 20 years between the interviews and it did not anymore represent a working-class neighborhood; thus, the district restrictions were loosened. The third cycle of the project was carried out in 2013 in a similar fashion. A good proportion of the interviewees in the last cycle was reached (27/45) and augmented again with a group of adolescents, divided equally in high school and vocational school. All in all, 13 speakers have been interviewed in all three time points.

Three age groups were studied in all three decades. These represent young speakers, middle-aged speakers, and old speakers, but vary in the actual ages of the interviewees. In the 2010s subset there is also an additional group of four 81 to 86-year-old informants. The ages, birth years and number of speakers in each group is presented with the temporal structure of the data in Figure 2. The corpus is a combination of panel and trend data, as the 1990s and 2010s subcorpora consist of middle-aged and old speakers already interviewed in earlier stages, bolstered with the young group. This leads to imbalance in scale, as the 1970s subset is much larger than the following ones. The data analyzed here consists of 126, 33 and 40 informants in the 1970s, 1990s and 2010s, respectively. The panel data also tends to emphasize the academic speakers, as it has been harder to reach the speakers of lower social classes (Paunonen, 2006; cf. Sankoff, 2018).

Through all three time points, the interviews were conducted by the University of Helsinki students. A recurrent pattern of questions was provided for the interviews in every cycle, but it has not been followed equivalently in every interview. The questions enquired about the informants' and their families' lives, traditions, perceptions of Helsinki and the Helsinki Finnish along with other dialects and languages. Every interview lasted for about an hour, and at least 30 minutes of every interview has been transcribed. The transcribed excerpts are not consistent (e.g., first 30 minutes) but vary between the interviews. It has been noted that some speakers in the 1970s regarded the interviews as formal situations (Paunonen 2005: 186), and such attitudes were not present in the 1990s or 2010s. This might affect some of the changes we observe.

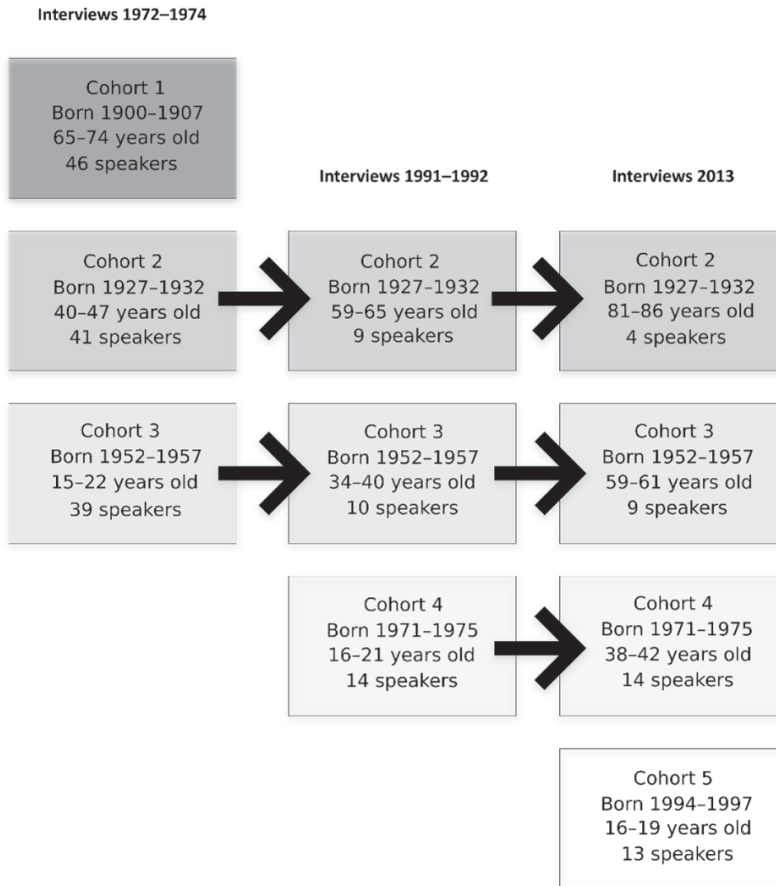


Figure 2. Structure of the data (based on Paunonen, 2006). Time of the interviews is presented among the x-axis and the cohorts are visualized in shades of gray. The number of speakers in each cohort is reported in the rectangles along with the ages at the time of the interview.

Analyzed features

The features analyzed in this study are mostly based on previous studies of Helsinki Finnish (e.g., Paunonen, 1995, 2006). In addition, some features were chosen because of their prevalence in a pilot study. The features are mostly morpho-phonological, but

some syntactic and lexical items are studied as well. We present the analyzed linguistic features and their respective variants in Table 1. Each feature is given a short name to refer to in the text. The frequency of each variant is also presented in the table. To give the reader an understanding of the variants and their relations to previously proposed lects, we give a brief explanation. However, we do not test if the previously proposed lects occur in the study (a hypothesis model), but our exploratory model finds collections of variants that best explain the variance in the data. For an international audience however, some backdrop is needed.

The first variant in Table 1 represents the pronunciation of the feature in standard Finnish. The standard is a combination of Western and Eastern dialect features, which leads the standard and the Häme dialect, for instance, to coincide in some cases. Using the standard variant therefore does not always represent deviation from dialectal features. Some standard variants have a social index (Silverstein, 2003), however. Using the first-person plural form /'menemme/ (1PL1) for instance comes across as very formal.

The differing social meanings present problems for the idea of coherent lects. Consider a case in which a speaker uses a single standard language variant that comes across as highly formal amid other varying features. Another speaker consistently uses several standard language features that are not as salient markers. The one variant can provoke a second-order indexical link (Eckert, 2008; Silverstein, 2003), whereas the use of several first-order features might not be noticeable to laymen. Thus, the first speaker might be perceived as more correct and formal (to use folk linguistic terms), although in the light of lectal coherence this would not be the case. Earlier studies (e.g., Campbell-Kibler, 2011; Johnstone & Kiesling, 2008) have reported on how a single variant affects perceptions on matters such as intelligence of the speaker, formality of the situation, or locality. A similar perception study on lects would shed light to the ways in which laymen label different ways of speaking.

Continuing with Table 1, the second variants represent the use of the variable in the Häme dialect. This is not the case in the variables INF-COMP and INE (marked with asterisk) however, in which the first variant is used in the Häme dialect. In these two cases, the second variant represents the feature in other Western dialects. In all

the cases where the second variant is empty, the Häme dialect coincides with the standard. In some cases, the second variant is not exclusive to the Häme dialect but more generally used in Finnish dialects (e.g., DIPH, PCP, COP, 3PN, 1PL). The last column (Variant 3) collects variants of differing origins: for instance, from Southern Häme, which differs slightly from the main Häme dialect (A-INF), Eastern dialects (HD) and widespread vernacular features (1PN).

Table 1. Studied features. Variant 1 represents pronunciation in standard Finnish and Variant 2 in the Häme dialect. Variant 3 represents variants from several differing origins, most of which can be associated with Southern Häme.

Feature		Variant 1	Variant 2	Variant 3
Monophthongization of vowel pairs (in partitive case, infinitives, and nouns separately)	/oɑ/, /øæ/, /ea/ and /eæ/ (OA-PTV, OA-INF, OA-N)	/'autoɑ/ 'car-PTV' OA-PTV1: 1270 OA-INF1: 424 OA-N1: 832	/'auto:/ OA-PTV2: 2042 OA-INF2: 1015 OA-N2: 3949	
	/ia/, /iæ/, /ua/ and /yæ/ (IA-PTV, IA-INF, UA-PTV, UA-INF)	'leikkiæ/ 'to play (with toys)' IA-PTV1: 9571 IA-INF1: 266 UA-PTV1: 1449 UA-INF1: 422		/'leikki:/ IA-PTV3: 4460 IA-INF3: 126 UA-PTV3: 792

				UA-INF3: 324
Infinitives	Contracted verb A-infinitive (A-INF)	/ˈpelata/ ‘to play (a game)’ 588		/ˈpela:/ 68
	MA-infinitive illative (MA-INF)	/ˈpela:ma:n/ ‘to play’ 2062	/ˈpela:n/ 272	/ˈpela:/ 802
	Infinitive complementizer to käydä ‘to go, to visit’ (INF-COMP)*	/ˈkæydæ ˈpela:mas:a/ ‘to go play’ 362	/ˈkæydæ ˈpela:mas/ or /ˈpela:s/ 226	/ˈkæydæ ˈpela:/ 50
Weak grade alternants of /t/ (d, r, ø)	After a long vowel (VVDV)	/ˈmeidæn/ ‘our’ 1567	/ˈmeiræn/ 7	/ˈmeiæn/ 1356
	After /h/ (HD)	/ˈtehdæ/ ‘to do’ 3128	/ˈtehræ/ 14	/ˈtehæ/ 1937
	After a short vowel (VDV)	/ˈedes/ ‘even’ 382	/ˈeres/ 2	/ˈe:s/ 176

Alternants of /ts/ in strong and weak consonant grades (TS-S and TS-W) (in certain lexemes)		/’itse/ ‘one-self’ TS-S1: 706 TS-W1: 879	/’it:e/ TS-S2: 796 TS-W2: 309	/’ite/ TS-W3: 655
Apocope	In inessive case (INE)*	/’talossa/ ‘in a house’ 15 076	/’talosa/ 80	/’talos/ 5455
	In adessive case (ADE)	/’talolla/ ‘at a house’ 18 204		/’talol/ 9561
	In elative case (ELA)	/’talosta/ ‘from a house’ 5706		/’talost/ 2418
	of /i/ (in numerals, conditional, translative case and past tense)	/’yksi/ ‘one’ 1399 (all cases combined)	/’yks/ 12 253 (all cases combined)	
	of /n/ (in certain lexemes) (N-APO)	/’mitæ:n/ ‘nothing’ 3028	/’mitæ:/ 2237	

Shortening of an unstressed diphthong (DIPH)		/ˈpʊnainen/ 'red' 442	/ˈpʊnanen/ 2405	
Participle (PCP)		/ˈpʊr:ut/ 'bit- ten' 1130	/ˈpʊr:u/ 9536	
Shortening of the copula verb olla 'to be'	1st and 2nd pers. (COP-1P)	/ˈolɛn/ 'I am', /ˈolet/ 'you are' 569	/ˈo:n/, /ˈo:t/ 2408	
	3rd pers. neg. (COP-NEG)	/ˈei ˈole/ 'is not' 393	/ˈei ˈo/ 969	/ˈei ˈo:/ 2083
Pronouns	'This', 'these' (THIS-PN)	/ˈtæmæ/, /ˈnæmæ/ 702	/ˈtæ:/, /ˈnæ:/ 2765	
	1st and 2nd pers. pronouns (1PN)	/ˈminæ/ 'I', /ˈsinæ/ 'you' 1435	/ˈmæ:/, /ˈsæ:/ 275	/ˈmæ/, /ˈsæ/ 14 427

	3rd pers. pronouns (3PN)	/ˈhæɪn/ ‘s/he’, /ˈhe/ ‘they’ 1755	/ˈse/ ‘it’, /ˈne/ ‘them’ 5331	
1st pers. pl. (present and past tense) (1PL)		/ˈmenemme/ go-1PL ‘we go’ 844	/ˈmen:æ:n/ go- PASS 8169	
Lexical items		/ˈsil:æ ‘tavalla/ ‘like that’ (SIL- LEE) 2020		/ˈsille:/ 1448
		/ˈvielæ/ ‘still, yet’ (VIEL) 1314	/ˈviæ/ 82	/ˈviel/ 473

Before continuing to the method of the study, a word of caution is presented. The interviewees selected to the corpus are native speakers of Finnish born in Helsinki. Although some of them have since moved or lived elsewhere in between the interviews, the corpus is designed to be quite homogenic, continuing the traditions of dialectology. This of course influences the variance we encounter and the different lects the method is able to discover. Ethnolects, for instance, are not reachable with the corpus at hand.

The features selected for the study also need to be scrutinized. Most features are morpho-phonological, only few lexical, and none prosodic. It has been stated that

the surface of the language (i.e., lexical, and phonetic features) shows greater social stratification than grammatical features (Meyerhoff & Walker, 2013). However, Levon and Buchstaller (2015) found this does not hold for listener perceptions: listeners attended a morphosyntactic variant differently than a phonetic one, but it was subject to evaluation, nonetheless. Even in studies of production results are conflicting (e.g., Cheshire, Kerswill & Williams, 2005). We believe emphasis on the morpho-phonological domain in our study does not compromise the results. It must also be noted that Finnish is an agglutinative language, and results of studies on English are not therefore directly applicable.

Latent Dirichlet Allocation

The aim of the current study is to discover which variants co-occur in the data and thus create a linguistic pattern, a *lect*. Several methods for identifying co-variation have been used in sociolinguistic settings before, such as factor analysis (Spearman, 1904), principal component analysis (Pearson, 1901) and k-means clustering (Forgy, 1965). In the study at hand a Latent Dirichlet Allocation (LDA) model is used (Blei et al., 2003).

LDA is a probabilistic model of the occurrence of countable features in each document. Typically, these countable features have been the number of occurrences of different words, but in our work, they are the number of occurrences of different linguistic variants listed in Table 1. In LDA, these counts in each document are assumed to arise from a mixture of underlying components. In machine learning literature these components have traditionally been called *topics*. We abstain from using this term, as we do not search for topics in a linguistic sense. We use the term *component* when looking at the data at the level of the computational model and *lect* when looking at the linguistic interpretation of the computational analysis.

Each of the components $k=1, \dots, K$ has a probability distribution over the features $v=1, \dots, V$, denoted as $p(v | k)$, and each document d has a probability distribution (mixture) over the components, denoted as $p(k | d)$. The probability for a particular

feature to occur in the document is then a sum over the components, $p(v | d) = \sum_{k=1, \dots, K} p(v | k) p(k | d)$. The contents of each distribution $p(v | k)$ and each distribution $p(k | d)$ are found by fitting (optimizing) the model for the observed data, according to what best describes it. In practice, features that often co-occur will form one of the components and have high probabilities in it, but the same feature can also become associated with multiple components if, for example, it co-occurs with one group of features in one context and with another group in another context.

The decision to use LDA for our corpus of spoken Finnish is based on an analogy. LDA has already been used in text classification and considers documents and words. Based on the co-occurrence of words in the documents, LDA infers hidden topics for the corpus. For instance, if the words *dog*, *bone* and *fetch* appear in several documents together, whereas *cat*, *meow* and *kitten* appear in others, the model infers that these two collections of words represent different topics (dogs and cats, although the model does not know the labels). If we understand lects as collections of frequently co-occurring linguistic features, the analogy is simple enough. The interview transcriptions we use correspond to the text documents and the linguistic features we have picked correspond to the words in the documents. Instead of finding latent topics as in text classification, we attempt to find underlying lects: if the features IPL1, INE1 and ADE1 co-occur in several interviews the model starts to infer that they represent a lect. However, the model does not restrict the number of co-occurring variables that can define a component, which means that it could also discover collections of only a few features. In these cases, we must use linguistic reasoning to decide which components represent lects. This is another reason to use the term component when discussing the results of the model.

Although the analogy is for the most part simple, there are some differences between these two data types. First, when the model is used for text documents, all the words (apart from very common words such as *and*) are used to fit the model. In our data we have focused on a chosen set of linguistic features and collected them from the interviews. Thus, our decisions already force the model in a certain direction. Another important difference is the fact that text corpora are generally sparse. This means that the words might appear frequently in some documents but not at all in

others; in terms of matrices, there are a lot of zeroes. There is not a lot of sports' vocabulary in articles about quantum physics, and vice versa. This is not the case in our data. We have only 78 possible variants in our data, several of which appear in most of the interviews.

One of the benefits of using LDA instead of other clustering methods like k-means is the fact that it is a mixed-membership model. This means that each of the variants under study can be a part of any component and each interview can be a mixture of unlimited components. It therefore supports a fluid understanding of language use, as the variants and components can be combined in an unrestricted manner. This is an important attribute: if a computational model is asked to do strict categorization, the results may appear more coherent than the linguistic reality is.

The variants were searched for in interview transcriptions using regular expressions (Thompson, 1968). The search results were manually verified and collected to a data frame automatically, using the statistical software R. The LDA model was then run on the resulting data frame using the R package `topicmodels` (Grün & Hornik, 2011).

Results

We set out to examine which features co-vary in Finnish spoken in Helsinki and if the discovered components were intelligible and interpretable in terms of previously proposed lects. We hypothesized that there was likely to be weak co-variation of features in the data and that although we might discover components traceable to different historical processes, we would also find plenty of obscurity in the data.

We begin by fitting the model to find the statistically most coherent components. After this initial stage, we start analyzing the components in regard to their distinctiveness, which is defined as the variants' exclusiveness to components. In this section we will also comment on the components that can be traced to previously proposed lects. To take a closer look at the lects and their utilization, we examine the primary components in the interviews and scrutinize the volume in which they are

used. We furthermore analyze which components are likely to combine in use. Finally, we take a short look at the variation and change of the lects in real time. Based on these analyses of the inferred components, we define the linguistically realistic lects, and their divergence from the mathematical components.

Coherence and distinctiveness

When using the LDA model, the number of components must be defined beforehand. There are several mathematical methods to find the best fit. In this study we used two separate tests (Cao, Xia, Li, Zhang & Tang, 2008; Deveaud, SanJuan & Bellot, 2014). We performed a hundred iterations of model training for each fit, where the number of components ranged from two to fifteen. Both tests found a model with ten components to present the best fit and thus the most coherent components. We further checked the three best models of each of the preceding tests for the Akaike Information Criterion (Akaike, 1974) and Bayesian Information Criterion (Schwarz, 1978) values. These values also signaled that the model with ten components presented the best fit for the data. We thus ran the model for the whole dataset with ten components. We ran the model for the three decades' (1970s, 1990s, 2010s) separately as well and compared the results with the whole data set. The division to three subsets did not significantly alter the discovered components which led us to focus on the complete set in the analysis.

The LDA model returns two objects: probabilities for variants per component and probabilities for components per interview speech. The probability of a variants' appearance in a component emphasizes the frequency of the variants in the data and not their exclusiveness to one component. Some variants appear in an interview hundreds of times, whereas others occur only a few times (see Table 1). This is problematic for linguistic interpretation, as the components are often headlined by the same features. To counter the issue and discover which variants are mostly associated with which components, we calculated conditional probabilities using Bayes' theorem ($P(\text{Component}_k | \text{Variant}_v)$, probability of Component_k given Variant_v). This process

brings the variants on the same scale: because there are ten components and the probabilities always sum to 1, the average probability per component (given a variant) is 0.1. Thus, conditional probabilities above 0.1 index a higher association with a component than in average. If a Component_k has a conditional probability of 1 given Variant_v, all utterances of Variant_v are associated with Component_k.

We scrutinize the variants that produce highest conditional probabilities in the components in Figure 3. We use the feature labels presented in Table 1, followed by the number of the variant. For instance, 3PN2 represents the second variant of the 3rd person pronoun variable. In the figure the complete probability scale is shown. The more exclusive the variants are to the component in question, the longer the bars. We want to emphasize that exclusiveness to a component does not necessarily indicate that the variant is highly important to it: for example, if a variant appears only 7 times in the whole corpus, it does not carry much weight in the fitted model, but might still be (and most likely is) highly exclusive to a single component. Thus, the conditional probabilities shown in Figure 3 might emphasize variants that are very rare in the corpus. These cases are commented when necessary. The probabilities of variants per components (which in turn emphasize very frequent variants) are presented as a supplementary figure in the Appendix.

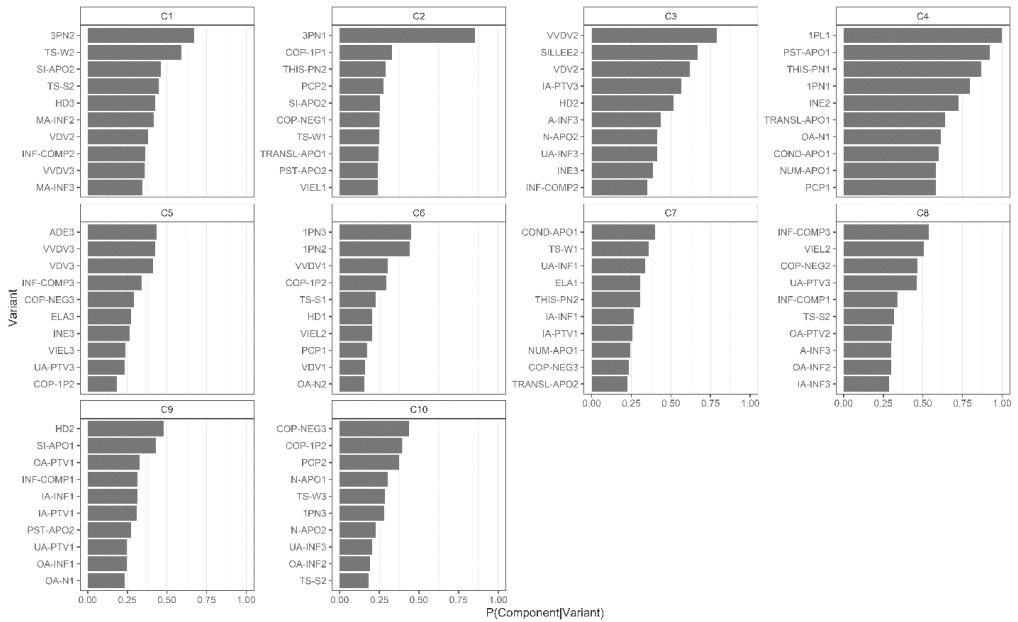


Figure 3. Highest conditional probabilities of components given variants. Components are labeled as abbreviations (C1 = Component 1). The variants are labeled as in Table 1.

The first observation of the figure is that C2 is headlined by a single variant. C6 is also unevenly distributed. C5, C7, C8, C9 and C10 are more coherent internally but the probabilities are not high, which indicates that these components consist of features frequently used in other components as well. Finally, the three components that appear the most distinct are C1, C3 and C4. We will begin by analyzing these three components before moving on to the more uneven ones.

C4 appears to be the most distinct of the discovered patterns as some variants are almost exclusive to it. The ten highest ranking variants instantly reveal the nature of this component, as they are almost entirely from standard Finnish. The only exception is INE2, an infrequent variant originally from the Western dialects. The variant producing the highest conditional probability (1PL1) was discussed earlier in regard to second order indexicality. It is completely exclusive to C4. Other variants in C4 are also quite distinct and index formality or correctness: the preservation of the word-final /i/ in PST-APO1, TRANSL-APO1 and NUM-APO1 as well as the use

of standard language pronouns /'minæ/ 'I', /'sinæ/ 'you' (1PN1) and /'tæmæ/ 'this' (THIS-PN1) are all relatively rare in Helsinki (Paunonen, 1995). Thus, our hypothesis of discovering a standard Finnish variety in the corpus is realized in C4. It is in fact not a surprise that this lect is the most distinct of all, given its enregistered nature (Agha, 2003). Similarly to RP in Agha's study, standard Finnish is also a variety whose correctness is guaranteed by someone else. Hence there are clearer rules to follow to use this lect. This opens possibilities for ideological work: a speaker may refer to the variants in standard Finnish to sound upper class or more educated.

Unlike C4, C1 and C3 do not have any standard Finnish variants among their most distinct ones. Both have several variants of the Häme dialect but C3 has more features associated with Southern Häme, which is situated closer to the capital region. IA-PTV3, A-INF3 and INE3 are all common in the Helsinki region, but not in the heartlands of Häme. Interestingly though, C1 uses zero as the alternative to standard Finnish /d/ (HD3, VVDV3), which is a variant originating in the Eastern dialects but very common in spoken Finnish all over the country at present. The alternative to standard /d/ in C3 is /r/ (VVDV2, VDV2, HD2), historically used in the Western dialects (including Häme) but quite rare in contemporary Finnish. It is also very rare in the data at hand, which means that it cannot be given too much weight in the analysis. Concluding from these differences between the two components, C1 can be traced to the Häme dialect, whereas C3 can be traced to Southern Häme.

These three components are thus interpretable in terms of previously proposed lects and can be labeled accordingly to ease reading of the article. We shall call C4 the enregistered standard (cf. RP in Agha 2003), C1 the Häme dialect and C3 the Southern Häme dialect. To reflect on our hypotheses, the results are quite surprising. The three most distinct components are easily traceable to previously proposed lects. Thus, the coherence of these lects is also visible in the data at hand. Especially the enregistered standard (C4) shows significant coherence and distinctiveness. However, we need to take a further look at the seven remaining components before making overall interpretations of our results.

C6 and C10 share features, as there are both 1PN (1st and 2nd person pronouns) and COP-1P (1st and 2nd person present tense of the verb *olla* 'to be') variants

among them. They are thus constituted by phrases such as /'mæ 'oon/ 'I am', which are naturally quite likely to occur in interview speech. Both components are mostly focused on these two variables. C6 hardly gets even a 0.25 conditional probability for any other variant. Moreover, in C10, the highest conditional probability belongs to the negation of the copula (COP-NEG3 /'ei 'oo/) which also appears with the 1st and 2nd person pronouns. As there are no restrictions on the number of variants needed to compile a component, only a few frequently co-occurring linguistic items suffice. Such a possibility was discussed in the previous section, and these two components demonstrate it in practice. Thus, although components 6 and 10 are collections of co-occurring features, they must be interpreted as repeating phrases rather than lects.

C2 is similar to components 6 and 10, as the conditional probability given 3PN1 (/ 'hæn/ 's/he') is 0.85 and the next highest probability is 0.33 (COP-1P1). This might indicate that the 3rd person pronoun appears in multiple contexts and is not tied to any of the other components. The variant is discursive in nature and interpreted as age-graded (Lappalainen, 2010), which could explain its somewhat general behavior in regard to the lects. In conclusion, components 2, 6 and 10 are not lects in the sense of this study. The model discovers the underlying structure in the data and thus does not judge if the components represent lects or phrases. A possible solution to this would be to eliminate lexical features from the input of the model, as the model would then not trace phrases. It would however be quite unsatisfactory to neglect a complete level of language use to fit a model. We have decided to keep our data intact but exclude these components from the lectal analysis.

The ten highest conditional probabilities for C5 all appear given highly colloquial variants. For instance, apocope in the grammatical cases (ADE3, ELA3, INE3) and the substitution of standard Finnish /d/ with zero in certain lexemes are supralocal colloquial features in Finland (Mantila, 2004). Given that young speakers generally prefer to use supralocal variants (Lappalainen, 2001) it is most likely that C5 is used by the youngest age group. It must also be noted that this component does not have a very high conditional probability on any of the variants, which indicates that the features are also used in other components. The lack of high probabilities also supports the idea of C5 as supralocal as it combines variants of several domains. We

label C5 as a supralocal lect. C8 is also colloquial, but slightly more local than C5. For instance, variant VIEL2 /'viæ/ is a salient marker of the Häme dialect and A-INF3 is a specialty of the Southern Häme region. C8 is thus more reminiscent of C3 discussed earlier than the supralocal lect (C5).

Finally, C7 and C9 appear to use mostly standard Finnish variants. The highest conditional probability for C9 is produced given HD2 /'tehræ/ which is very rare in the corpus and is thus overemphasized when looking at the probabilities of the components given variants. This means that HD2 is the most exclusive of the variants in C9, but, because there are only 14 occurrences of this variant in total, it does not carry much weight. Looking beyond that variant, C9 consists of standard Finnish variants that are very common in spoken Finnish. This fact differentiates C9 from the enregistered standard (C4), which uses almost exclusively formally marked features. Based on this difference, we expect that C9 is used by many of the same speakers as the enregistered standard (C4) but is generally more likely to appear. C7 is similar to C9, but also produces some very common colloquial features such as THIS-PN2 /'tæ:/. The three components that utilize mostly standard Finnish variants thus create a formality continuum. The variants mostly associated with C4 are salient markers, while the variants associated with C9 are more common. Finally, C7 consists of some common colloquialisms on top of the mostly standard Finnish basis. To define the lects from this continuum is not straightforward, and we shall analyze their relations to each other in more detail in the next section.

The inclusion of the seven new components in addition to the three most distinct ones obscures the picture slightly. Components 2, 6 and 10 were all analyzed as consisting of only one or two lexical features and are thus not lects at all. C8 is quite close to the Southern Häme dialect (C3), whereas C9 and C7 are close to the enregistered standard (C4). The supralocal lect (C5) however appears to differ from others in the unique combination of the features, even though they produce quite low conditional probabilities.

The results show the capability of the LDA model as part of a lectal analysis process. The discovery of components traceable to historical lects (the enregistered standard, the Häme dialect, the Southern Häme dialect) and the supralocal lect is very

promising and counters the findings of earlier studies on lectal coherence (Gregersen & Phrao, 2016; Guy 2013). The overlap between some components (C3 and C8; C7 and C9) on the other hand concurs with the results of these studies, and it is not directly clear if they present one or several lects. Finally, although the components based on repeating words and phrases (C2, C6, C10) are not of use in this study, they likewise show that the model can find latent structure in the data. In this study, we have used LDA as the starting point for sociolinguistic analysis, but the tool could be equally useful for other branches of linguistics such as lexicology, phraseology, or even forensic linguistics.

Primary components and a network analysis

The LDA model does not force the speakers into users of exclusive components (hard clustering) but provides a distribution of components for each speaker. By analyzing the probabilities of each component given the speaker, it is possible to examine how much of a speaker's linguistic behavior can be attributed to a single underlying component. Furthermore, by analyzing the co-occurring components we can uncover links between them. We should remark that the components discovered by the model are technical collections of co-occurring features over the whole dataset and thus abstract. We do not claim that the speakers combine clear-cut lects in their speech, but that we can discover underlying components in the whole data.

We commence with an analysis of primary components, i.e., the components that explain the largest proportion of the speakers' linguistic behavior. Counts of the primary components are, from the largest value to the smallest, 66 (C9), 31 (the enregistered standard C4), 22 (C8), 19 (C6), 15 (the supralocal lect C5), 15 (C7), 14 (the Southern Häme dialect C3), 8 (C10), 6 (the Häme dialect C1), and 3 (C2). The hypotheses we proposed in the last section turn out to be true: C9 appears more than the enregistered standard (C4), although both are very frequent. The primarily used components differ from the most distinct ones apart from the enregistered standard (C4). Both the Häme dialect (C1) and the Southern Häme dialect (C3) are quite rare as primary components. It seems that the other colloquial components (C8 and C5)

are more frequent than the most distinct ones. It is also visible that the colloquial speakers are more spread out between different lects than the speakers that use lects associated with standard Finnish. This is a natural conclusion, given there are more colloquial domains than there are standardized ones.

To get a deeper understanding of the links between the components, we will scrutinize them as a network, which is presented in Figure 4. The network is drawn based on a Fruchterman-Reingold layout (Fruchterman & Reingold, 1991) using the *igraph* package in R (Csárdi 2019). The gray lines are drawn from each speaker to a component, if the probability of the component (given speaker) is more than 0.16, which was the smallest probability of a primary component in the data. Thus, the speakers might have a different number of lines drawn depending on the distribution of components. The width of the gray line represents the weight of the link: the thicker the line is between the speaker symbol and the component, the bigger the probability of the component within the interview speech. The speaker links act as springs: if a speaker uses two components, these are pulled closer together. Hence, components that are positioned close to each other tend to share a lot of speaker links. We want to emphasize that the network presentation is an approximation and optimized for readability, and different runs on the same data might result in different layouts. The non-lectal components (C2, C6, C10) are presented in the graph to preserve the original distribution between the components, but they will not be discussed in the analysis.

The components that utilize more standard variants appear on the left-hand side of the figure, while the more colloquial ones are focused on the right-hand side. This divide is also apparent in the decades: most of the speakers connected to the enregistered standard (C4) and C9 were interviewed in the 1970s. Thus, the Finnish spoken in Helsinki is becoming more colloquial over the time span of this study. The age of the speakers is not visible in Figure 4, but there are no young speakers utilizing the enregistered standard, and only 11 with C9. Conversely in C8, C3 and especially C5, most of the speakers represent the youngest age group.

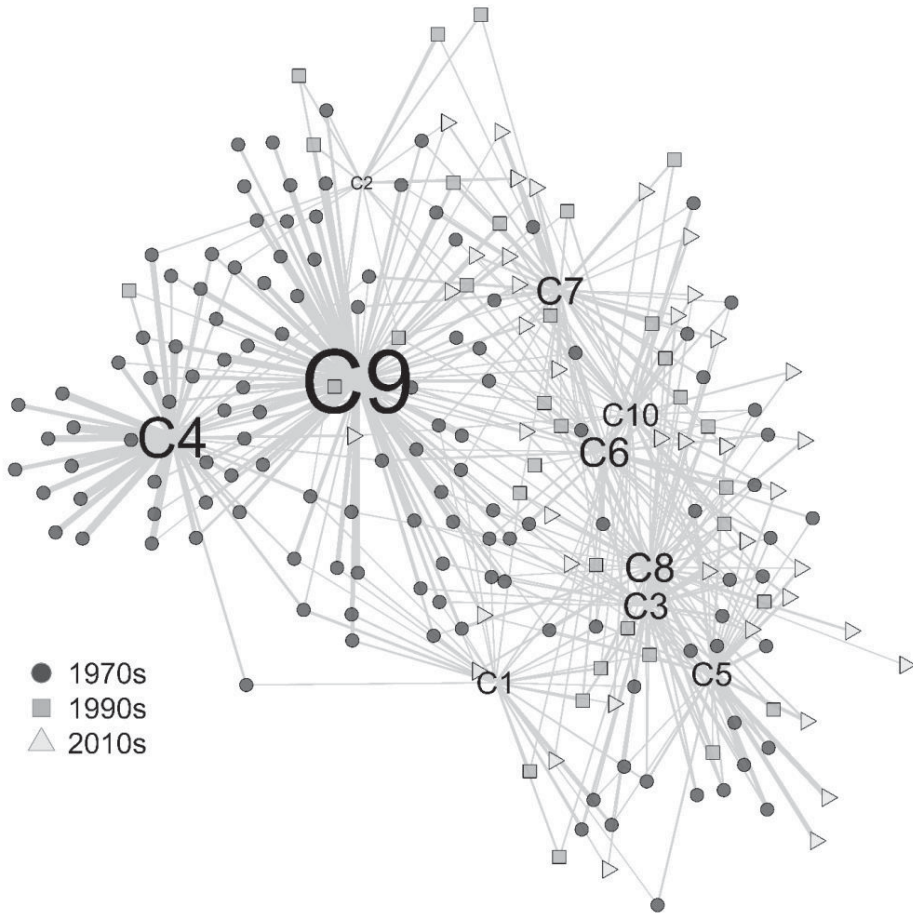


Figure 4. Network of components and speakers. Times of the interviews are presented in different shapes and shades. The width of the line presents a stronger connection. The size of the component labels corresponds to the total weight of the components' usage in the data. The components close to each other share a lot of the same speakers.

The width of the line between the speaker symbol and the component represents the proportion of the component in the speech of the informant. From the figure we can thus interpret that C9 and especially C4 (the enregistered standard) generally explain a lot of the linguistic behavior in the interviews in which they appear. This is understandable given the standardized nature of these components. One important observation from the figure is the fact that most of the speakers tend to have more than one

link. Thus, the components are used only partially. This is important, as lects have been seen as analogous to uniform systems such as languages and dialects. If, however, the speakers combine several lects in their speech, is this a wrong model?

In the previous section we interpreted that C3 (labeled as the Southern Häme dialect) and C8 are quite similar. This is also evident in Figure 4, as they are almost overlapping. Given their similar distribution of variants and their proximity in the figure, we combine these two components to present the Southern Häme dialect. We also noted that C4, C9 and C7 form a continuum in regard to standard Finnish. This is also visible in Figure 4, and supports the division created by the LDA model: these three components are distinct enough that they should not be immediately combined. Thus, in addition to the enregistered standard (C4), we have a component that could be defined as a neutral standard (C9) and another that utilizes common colloquialisms, which we shall describe as a lax standard (C7). This raises a question whether they represent lects or situational registers. Another issue with these three components is their use over time: in the 1970s C4 and C9 are very frequent, but seem to disappear almost completely later, whereas C7 is utilized mostly in the 1990s and 2010s. This leads us to the question of the next section: are we observing three lects or one lect changing over time?

Change in real time

In this section we scrutinize the components in regard to the interview rounds and the age groups. We commence with the three standard Finnish components discussed earlier, the enregistered standard (C4), the neutral standard (C9) and the lax standard (C7). The change over time in these three components is presented as a boxplot in Figure 5.

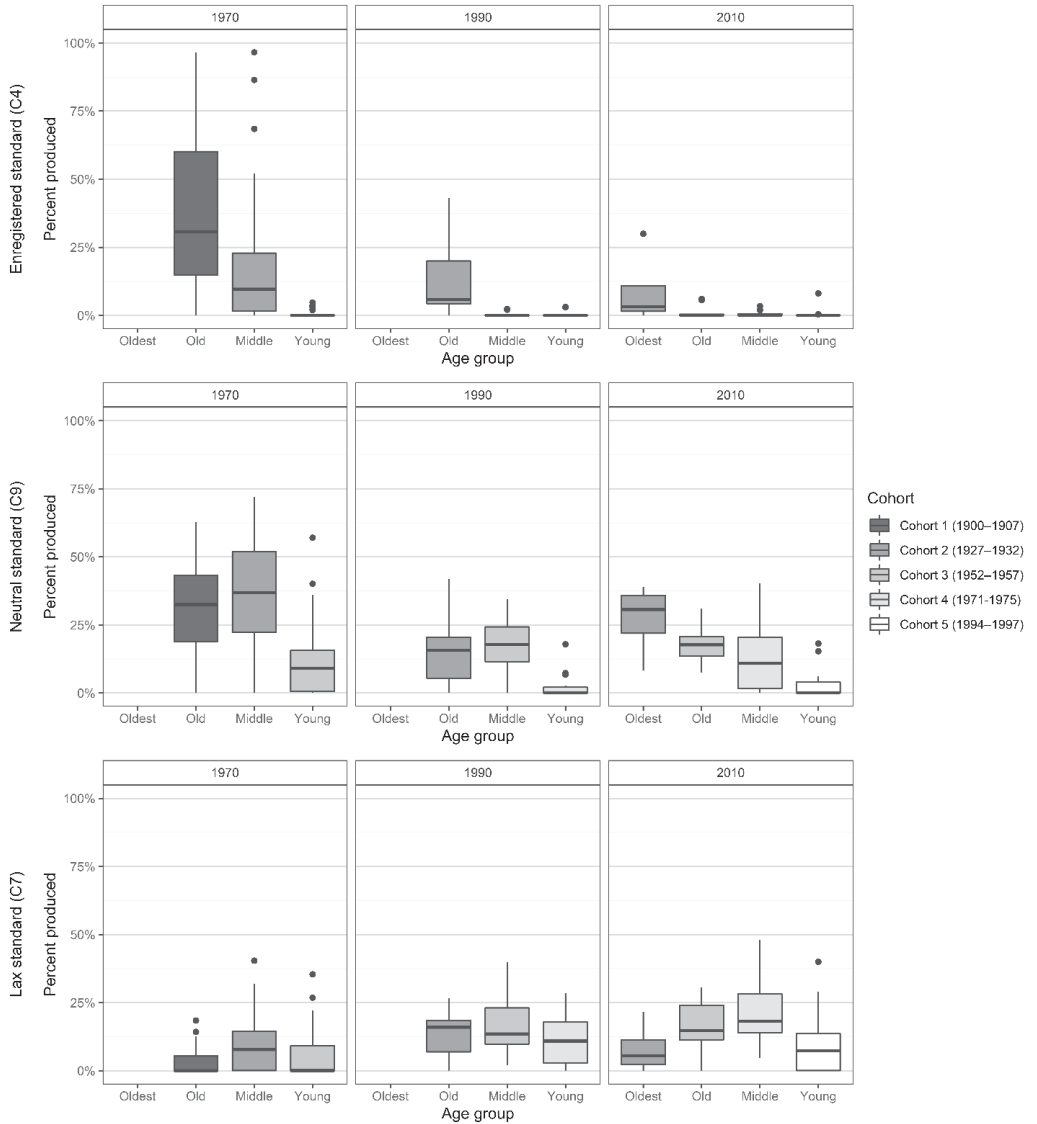


Figure 5. Change over time in the standard Finnish components. The cohorts and their shades correspond with Figure 2.

The enregistered standard (C4) is in strong decline, with the change happening mostly generationally. The youngest speakers in the 1970s do not use the lect or take it on later either. The middle-aged group in the 1970s preserves the usage of the lect throughout their lifespan. Among the old speakers in the 1970s, the lect is the

most popular one. This was to be expected as the usage of standard Finnish as a spoken variety started in the late 19th century, while this group was born in the first decade of the 20th century. The enregistered standard is thus one of the most used lects in the data, but its usage is heavily focused on the 1970s.

The neutral standard (C9) is actualized in a different pattern. Although the use of this component also decreases over time, it seems to be still actively used in the 2010s. For both the 1970s and the 1990s the middle-aged group uses C9 the most. This would suggest an age-graded pattern: a shift towards the standard language in middle age due to pressures from the linguistic marketplace (Bourdieu & Boltanski, 1975; Buchstaller, 2015). This pattern does not hold in the 2010s, but even then, the maximum values of C9 are presented in the middle-aged group. A similar pattern is visible in the lax standard (C7). The middle-aged speakers use it the most in all the decades, although in the 1990s the median of the old speakers is higher. Compared to C4 and C9, the lax standard is not as popular in the 1970s. However, in the 2010s it is the most frequent standard Finnish component.

In conclusion it is not easy to decide whether we should view these components as separate lects or as changing iterations of the same lect. The unique change pattern and the collection of highly salient features of the enregistered standard (C4), however, leads us to label it as a lect in its own right. C7 and C9 however share a lot of features and a similar change pattern. There is not a clear division between these two components, and we have decided to view them as the same lect, which is based on standard Finnish but garners common colloquialisms over time.

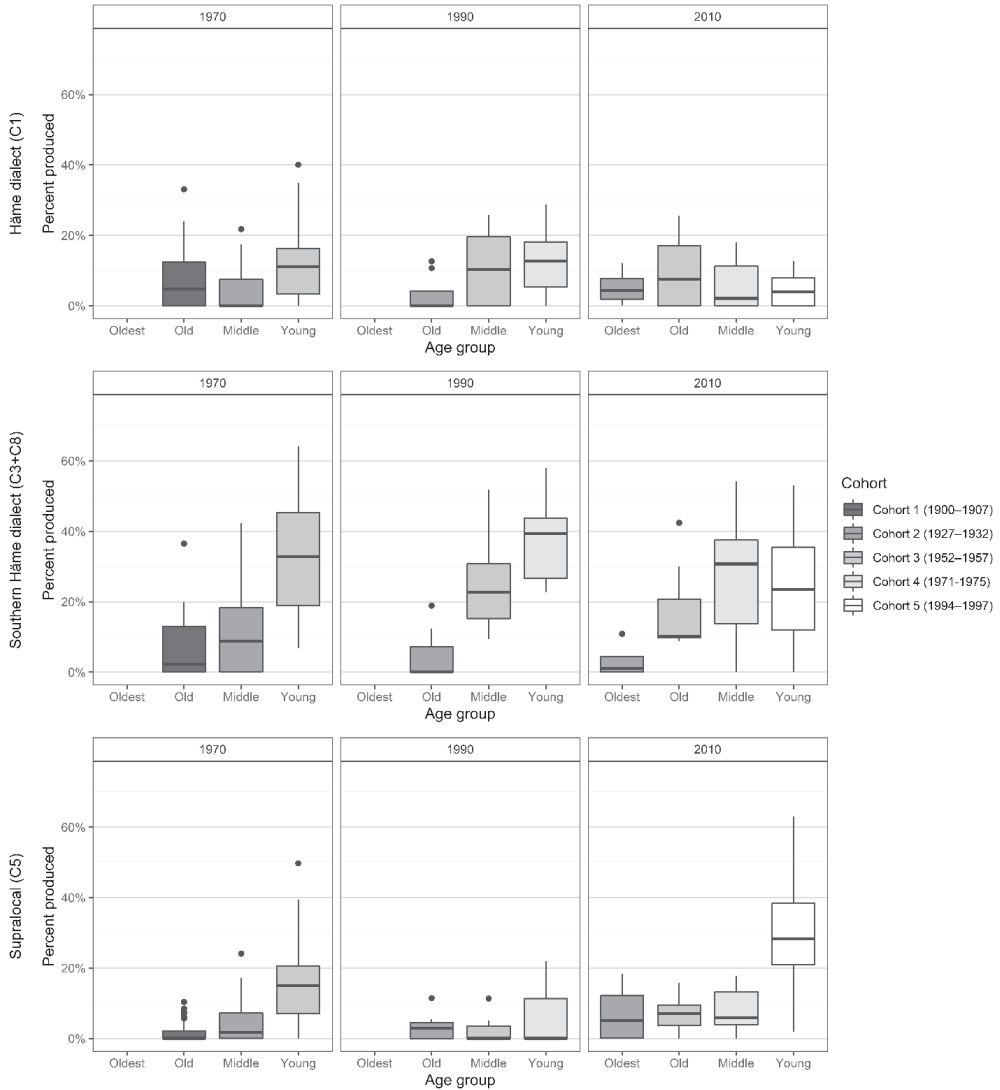


Figure 6. Change over time in the colloquial lects.

The change over time in the colloquial lects is presented in Figure 6. The Häme dialect (C1) seems to be on a generational change pattern, as the cohort born in the 1950s keeps their usage quite steady. It is however not popular among the young speakers in the 2010s and is thus possibly fading in Helsinki. The Southern Häme dialect (C3+C8) is very frequent among young speakers in the 1970s and 1990s, but not so in the 2010s. This lect presents a retrograde change (Wagner & Sankoff, 2011) as the

usage rate drops as the speakers age. The overall use however stays relatively steady which means that the Southern Häme dialect is still regularly utilized in Helsinki. However, the young speakers in the 2010s do not use the Southern Häme variety as much as in the 1990s. The reason behind this is the rise of the supralocal lect (C5). The situation in the 1970s in regard to this lect is reminiscent of the Southern Häme dialect: the older the speakers, the less they use this lect. Interestingly though, the supralocal lect is used less in the 1990s than in the 1970s and even the youngest group seems to diverge from it. Conversely in the 2010s it is the most popular lect among the youngest group and the older groups use it more as well. This might be due to increased mobility and the rise of social media. The popularity of the supralocal lect also suggests that some dialect levelling (Milroy 2002) is happening in the community.

The change in real time has provided more evidence on the lects in Helsinki. Although the enregistered standard is one of the most used lects in the data, its usage is already in decline. The same is true for the Häme dialect (C1). The Southern Häme dialect however seems to remain constant throughout the decades. There is dialect levelling happening in the data, as the supralocal lect (C5) is the most popular among the young speakers in the 2010s, and the neutral standard (C7+C9) utilizes some non-salient colloquial variants as well.

Before moving on to the conclusions, we briefly comment on the sociolinguistic background variables of the speakers besides age. It is not surprising that the enregistered standard and the neutral standard are preferred by the more academic speakers, whereas the Southern Häme dialect and supralocal lect are used more by the working class. A similar difference is evident between the two neighborhoods. Interestingly men use the Southern Häme dialect more than women, who in turn prefer the supralocal lect. Thus, the fact that the supralocal lect is gaining popularity is supported by the fact that women typically lead language change (Labov 1990).

Conclusions

The results presented suggest that there are coherent and distinct lects in the Finnish spoken in Helsinki. The findings contradict our initial hypothesis that the variants would show only weak co-variation and that the lects would be obscure. Moreover, the lects that seemed to be the most distinct in our study were the ones that have been discussed before both in linguistics and by laymen.

Three of the components discovered by the LDA model were directly deemed as lects (C1 as the Häme dialect, C4 as the enregistered standard and C5 as the supralocal lect). Three components were discarded as they were more reminiscent of repeating phrases or words rather than lects. After a linguistic analysis, four of the components were combined into two: C7 and C9 to the neutral standard, and C3 and C8 to the Southern Häme dialect. Thus, ten components initially discovered by the LDA model finally led to five linguistically reasoned lects.

Earlier studies on lectal coherence have often focused on individuals (e.g., Gregersen & Phrao, 2016; Guy, 2013). It has been noted that very few speakers are consistent in their usage of lects, and our results approve: apart from the most extreme cases of the enregistered standard, hardly any speaker uses one lect exclusively or even for most of the time. This observation raises an important theoretical question: if the lects are coherent, but not used in a coherent fashion, what causes the divergence?

Following Geeraerts (2010), we argue that we should not consider the lects in parallel with discrete language systems, but as a network of partial systems (cf. Figure 4). Thus, the analogy between languages, dialects and lects in general does not seem to hold true in actual usage. Whether the model of a homogeneous linguistic system is correct for languages and dialects either is a question on its own, and shall not be considered here. Geeraerts also argues that lects have prototype structure: each feature may be more or less typical of a certain lect. Our model has provided evidence of this: many of the variants appear in several lects but are more typical for some of them. Moreover, if we consider for instance the supralocal lect, none of the variants are highly typical of the lect but it is still useful and quite popular. It must be noted

that the studied features are predominantly morpho-phonological. This results from the agglutinative structure of Finnish, but also means that the results might not be directly applicable to, for instance, the languages of the Indo-European family.

We have shown that the lects in Helsinki Finnish are for the most part constructed by coherent bundles of features. Thus, the perceptions of lects may be grounded in the linguistic reality after all. However, as hardly any speaker uses one lect invariably, it is difficult to argue that a listener could perceive a coherent lect from a single speaker's linguistic behaviour. In this regard we approve of Gregersen and Phrao's (2016) view of lects being guided by salient features. Similarly, Boyd and Fraurud (2010) argue that a single salient feature may suffice to label a speaker a user of Rinkeby Swedish. This is essentially what Irvine (2001) defines as erasure: as a distinction is noted, other sociolinguistic evidence is ignored, and a homogeneous variety is imagined. A possible future study should focus on lectal perceptions based on single salient variants (cf. Campbell-Kibler, 2011).

Our findings suggest that we should not consider lects and their coherence in regard to single speakers, but as underlying collections of features that the speakers can incorporate in their speech. This sort of view is compatible with the idea of bricolage, as put forth by Eckert (2008). Thus, considering lects as objects in their own right might be a false presumption as they are hardly ever used in a complete manner. However, we do not encourage a model in which each variant is evaluated separately, as proposed by Guy (2013). We have shown that the variants indeed still co-vary and cluster when focusing on a larger community of speakers, and that these underlying lects still offer interesting information about the relations between the variants and the people that use them. Therefore, we should not deem lects non-existent or incoherent, but treat them as parts of an intricate system, exploited by speakers in a complex, yet meaningful way.

References

- Agha, Asif. (2003). The social life of cultural value. *Language and communication* 23:231–73.
- Akaike, Hirotugu. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19:716–23.
- Blei, David, Andrew Ng, & Michael Jordan. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3:993–1022.
- Bourdieu, Pierre, & Luc Boltanski. (1975). Le fétichisme de la langue. *Actes de la recherche en sciences sociales* 1:1–32.
- Boyd, Sally, & Fraurud, Kari. (2010). Challenging the homogeneity assumption in language variation analysis: Findings from a study of multilingual urban spaces. In P. Auer & J. E. Schmidt (eds.), *Language and space: An international handbook of linguistic variation*. Berlin: Mouton De Gruyter. 686–706.
- Bucholtz, Mary. (2003). Sociolinguistic nostalgia and the authentication of identity. *Journal of Sociolinguistics* 7:398–416.
- Buchstaller, Isabelle. (2015). Exploring linguistic malleability across the life span: Age-specific patterns in quotative use. *Language in Society* 44:457–96.
- Campbell-Kibler, Kathryn. (2011). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change* 22:423–41.
- Cao, Juan, Xia, Tian, Li, Jintao, Zhang, Yongdong, & Tang, Sheng. (2008). A density-based method for adaptive IDA model selection. *Neurocomputing – 16th European Symposium on Artificial Neural Networks* 72:1775–1781.
- Cheshire, Jenny, Kerswill, Paul, & Williams, Ann. (2005). Phonology, grammar and discourse in dialect convergence. In P. Auer, F. Hinskens, & P. Kerswill (eds.), *Dialect change: Convergence and divergence of dialects in contemporary societies*. Cambridge: Cambridge University Press. 135–167.
- Cheshire, Jenny, Kerswill, Paul, Fox, Sue, & Torgersen, Eivind. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15:151–196.

- Csárdi, Gábor. (2019). igraph R package. Available at <https://igraph.org/r/>.
- Deveaud, Romain, Sanjuan, Éric, & Bellot, Patrice. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique* 17:61–84.
- Eckert, Penelope. (2008). Variation and the indexical field. *Journal of Sociolinguistics* 12:453–76.
- Forgy, Edward W. (1965). "Cluster analysis of multivariate data: efficiency versus interpretability of classifications". *Biometrics*. 21:768–69.
- Fruchterman, Thomas M.J. and Reingold, Edward M. (1991). Graph Drawing by Force-directed Placement. *Software - Practice and Experience*. 21:1129–64.
- Geeraerts, Dirk. (2010). Schmidt redux: how systematic is the linguistic system if variation is rampant? In K. Boye & E. Engberg-Pedersen (eds.), *Language usage and language structure*. Berlin: Mouton de Gruyter. 237–62.
- Gregersen, Frans, & Phraao, Nicolai. (2016). Lects are perceptually invariant, productively variable: A coherent claim about Danish lects. *Lingua* 172–173:26–44.
- Gross, Johan, Boyd, Sally, Leinonen, Therese, & Walker, James A. (2016). A tale of two cities (and one vowel): Sociolinguistic variation in Swedish. *Language Variation and Change* 28:225–47.
- Gross, Johan. (2018). Segregated vowels: Language variation and dialect features among Gothenburg youth. *Language Variation and Change* 30:315–36.
- Grün, Bettina, & Hornik, Kurt. (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software* 40:1–30.
- Guy, Gregory R. (2013). The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics* 52:63–71.
- Halonen, Mia, & Vaattovaara, Johanna. (2017). Tracing the indexicalisation of the notion “Helsinki s”. *Linguistics* 55:1169–95.
- Hebdige, Dick. (1979). *Subculture: The Meaning of Style*. New York: Methuen.
- Helpuhe = The Longitudinal Corpus of Finnish Spoken in Helsinki (1970s, 1990s and 2010s). University of Helsinki, Institute for the Languages of Finland and

- Heikki Paunonen. URN: <http://urn.fi/urn:nbn:fi:lb-100110016072> Accessed September 15, 2018.
- Horvath, Barbara, & Sankoff, David. (1987). Delimiting the Sydney speech community. *Language in Society* 16:179–204.
- Irvine, Judith T. (2001). Style as distinctiveness: The culture and ideology of linguistic differentiation. In P. Eckert and J. Rickford (eds.), *Style and sociolinguistic variation*. Cambridge: Cambridge University Press. 21–43.
- Itkonen, Terho. (1989). *Nurmijärven murrekirja*. [Dialect book of Nurmijärvi]. Helsinki: The Finnish Literature Society.
- Johnstone, Barbara, & Kiesling, Scott F. (2008). Indexicality and experience: Exploring the meanings of /aw/-monophthongization in Pittsburgh. *Journal of Sociolinguistics* 12:5–33.
- Labov, William. (1966). *The social stratification of English in New York City*. Washington D.C: Center for Applied Linguistics.
- (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2:204–54.
- Lappalainen, Hanna. (2001). Sosiolingvistinen katsaus suomalaisnuorten nykypuhekieleen ja sen tutkimukseen. [A sociolinguistic overview of the contemporary spoken Finnish of young speakers and its research.] *Virittäjä* 105:74–101.
- (2010). Hän vai se, he vai ne? Pronominivariaatio ja normien ristiveto. [She or it, they or those? Variation in pronouns and the tug of war between norms.] In H. Lappalainen, M-L. Sorjonen & M. Vilkuna (eds.), *Kielellä on merkitystä. Näkökulmia kielipolitiikkaan*. Helsinki: The Finnish Literature Society. 279–324.
- Lehikoinen, Laila, & Kiuru, Silva. (1989). *Kirjasuomen kehitys*. [The development of written Finnish.] Helsinki: University of Helsinki.
- Ma, Roxana, & Herasimchuk, Eleanor. (1972). Speech styles in Puerto Rican bilingual speakers: a factor analysis of co-variation of phonological variables. In J.A. Fishman (ed.), *Advances in the Sociology of Languages, vol. II*. The Hague: Mouton. 268–95.

- Mantila, Harri. (2004). Murre ja identiteetti. [Dialect and identity.] *Virittäjä* 108:322–46.
- Meyerhoff, Miriam, & Walker, James A. (2013). An existential problem: The sociolinguistic monitor and variation in existential constructions on Bequia. *Language in Society* 42:407–28.
- Milroy, Leslie. (2002). Mobility, contact and language change – working with contemporary speech communities. *Journal of Sociolinguistics* 6:3–15.
- Mufwene, Salikoko S. (2001). *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.
- Mäki, Netta, & Vuori, Pekka. (2019). Helsingin väestö vuodenvaihteessa 2018 / 2019 ja väestönmuutokset vuonna 2018. [The population of Helsinki in 2018/19 and changes in the population in 2018.] *Tilastoja* 2019:9.
- Oushiro, Livia. (2016). Social and structural constraints in lectal cohesion. *Lingua* 172–173:116–30.
- Paunonen, Heikki. (1994). The Finnish language in Helsinki. In B. Nordberg (ed.), *The sociolinguistics of urbanization: the case of the Nordic countries*. Berlin: de Gruyter. 223–45.
- (1995). Suomen kieli Helsingissä. [Finnish in Helsinki]. Helsinki: University of Helsinki.
- (2006). Helsingiläisiä puhujaprofiileja. [Speaker profiles in Helsinki.] *Virittäjä* 109:162–200.
- (2006). Vähemmistökielestä varioivaksi valtakieleksi. [From a minority to a varying majority language.] In K. Juusela & K. Nisula (eds.), *Helsinki kieliyhteisönä*. Helsinki: University of Helsinki. 13–99.
- Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine* 2 (11): 559–572.
- Pratt, Mary Louise. (1987). Linguistic utopias. In N. Fabb, D. Attridge, A. Durant & C. MacCabe (eds.), *The linguistics of writing: arguments between language and literature*. New York: Methuen. 48–66.
- Quist, Pia. (2008). Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *International Journal of Bilingualism* 12:43–61.

- R = R Core Team. (2017). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.
- Sankoff, Gillian. (2018). Before there were corpora: The evolution of the Montreal French project as a longitudinal study. In S. Wagner & I. Buchstaller (eds.), *Panel Studies of Variation and Change*. Oxford: Routledge. 21–52.
- Schwarz, Gideon E. (1978). Estimating the dimension of a model. *Annals of Statistics* 6:461–464.
- Silverstein, Michael. (2003). Indexical order and the dialectics of sociolinguistic life. *Language and Communication* 23:193–229.
- Spearman, Charles (1904). General intelligence objectively determined and measured. *American Journal of Psychology* 15: 201–93.
- Thelander, Mats. (1979). *A Qualitative Approach to the Quantitative Data of Speech Variation*. Uppsala: University of Uppsala.
- Thompson, Ken. (1968). Regular expression search algorithm. *Communications of the ACM* 11:419–22.
- Vaattovaara, Johanna, & Soininen-Stojanov, Henna. (2006). Pääkaupunkiseudulla kasvaneiden kotiseuturajaukset ja kielelliset asenteet. [Regional identity and linguistic attitudes among the capital region residents.] In K. Juusela & K. Nisula (ed.), *Helsinki kieliyhteisönä*. Helsinki: University of Helsinki. 223–55.
- Wagner, Suzanne Evans, & Sankoff, Gillian. (2011). Age grading in the Montréal French inflected future. *Language Variation and Change* 23:275–313.
- Waris, Heikki (1951). Helsinkiläisyhteiskunta. [The Helsinki society.] In *Helsingin kaupungin historia 3 (2): ajanjakso 1809–1875*. Helsinki: City of Helsinki. 89–211.
- Wolfram, Walt. (2007). Sociolinguistic folklore in the study of African American English. *Language and Linguistics Compass* 1:292–313.
- Åström, S.-E. (1956). Kaupunkiyhteiskunta murrosvaiheessa. [Urban society in a turning point.] In *Helsingin kaupungin historia 4 (2)*. Helsinki: City of Helsinki. 9–333.

Appendix

Figure S1 shows the probabilities of variants given components produced by the LDA model. It corresponds to Figure 3 presented in the article, which represents probabilities of components given variants. Figure S1 emphasizes the most frequent variants in the data, whereas Figure 3 emphasizes the most exclusive variants in the components. Please note that the probability scales of the figures do not match (0–1 in Figure 3, and 0–0.4 in Figure S1).

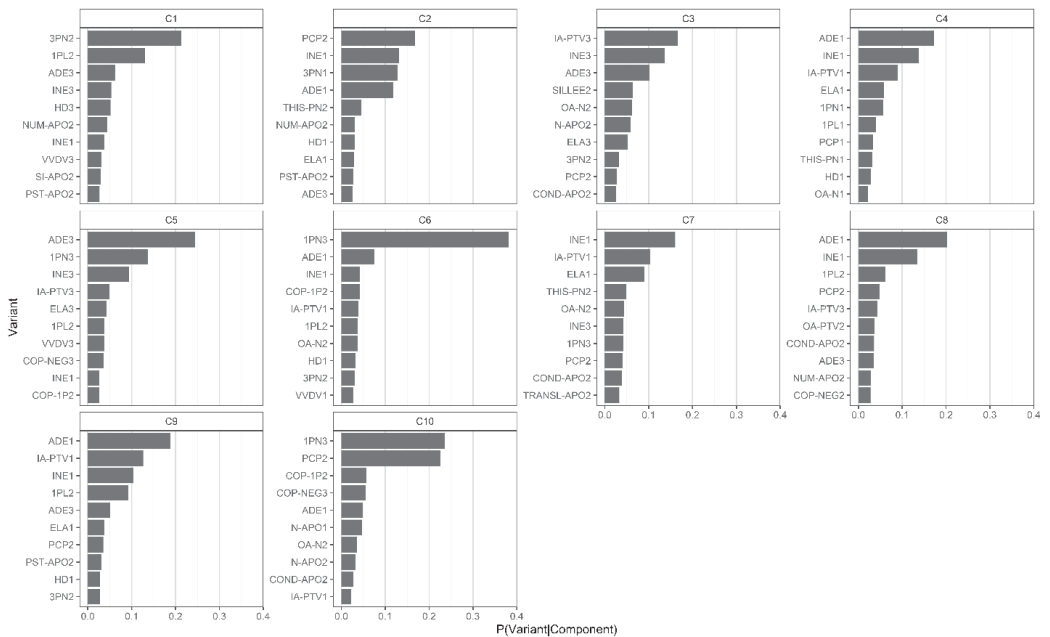


Figure S1. Highest conditional probabilities of variants given components.

JULKAISU

IV: KATOMUODOT

Katomuotojen eteneminen yleiskielen *hd*-yhtymässä Helsingin puhekielessä

Kuparinen, Olli – Santaharju, Jenni – Leino, Unni – Mustanoja, Liisa – Peltonen,
Jaakko

Käsikirjoitus. Lähetetty arvioitavaksi.

