



KATRI PRIIKI

Puhutun suomen kielioppia ja yksilöllistä vaihtelua

Kvantitatiivinen tutkimus *hän*-pronominista Kaakkois-Satakunnan nykypuhekielessä

1. Johdanto

Tässä artikkelissa tarkastelen pronominia *hän* henkilöviittauksissa lounaisten välimurteiden¹ alueelta äänitetyssä keskusteluaineistossa. Tavoitteena on asettaa *hän*- ja *se*-pronominien vaihteluun vaikuttavia tekijöitä tärkeysjärjestykseen. Pronominien vaihtelusta on kirjoitettu paljon aiemmin, mutta kattavaa sekä kielen sisäiset että ulkoiset tekijät huomioivaa tutkimusta ei ole tehty. Uutta tässä artikkelissa on myös suomen kielen tutkimuksessa vasta vähän käytetty tilastollinen monimuuttujamenetelmä, satunnaismetsä (*random forest*). Vaikka artikkelin painotus on kvantitatiivisessa tutkimuksessa, asemoin tutkimukseni vuorovaikutussosiolingvistiikan (ks. esim. Gumperz 2001; Rampton 2006) kenttään. Pyrin tarkastelemaan kriittisesti sitä, mitä vaihtelun ilmiöitä tilastollinen analyysi tavoittaa ja kuinka paljon vaihtelusta on mahdollista selittää laskettavissa olevien tekijöiden avulla. Tarkennan määrällistä analyysiä laadullisella, vuorovaikutustekijät huomioivalla tarkastelulla. Artikkelin on osa väitöskirjahanketta (Priiki tekeillä), jossa tarkastellaan pronomineja *hän*, *se*, *tää* ja *toi*.

Aiemmin pronominin *se* ja *hän* vaihtelussa olennaisiksi tekijöiksi on havaittu toisaalta kielenulkoisia sosiaalisia muuttujia, toisaalta kielensisäiseen kontekstiin liittyviä tekijöitä. Monissa tutkimuksissa on todettu, että erilaiset henkilöt käyttävät *hän*-pronominia eri määrin (Paunonen 1995, 165–177), ja myös murteiden välillä on eroja (esim. Siitonen 2008, 88–91; Vilppula 1989, 389). *Hän*-pronominin frekvenssi on kytköksissä myös tilanteeseen vaihteluun (Lappalainen 2010a, 2010b; Mielikäinen 1982, 282; Nuolijärvi 1986, 290–300). Alueellisten, sosiaalisten ja tilanteisten tekijöiden lisäksi *hän*-pronominin

1 Rapola (1990) kutsuu ryhmää *lounais- ja hämäläismurteiden väliseksi siirtymämurteiksi*. Tiiviimmän ilmauksen vuoksi suosin Virtarannan ja Yli-Paavolan (1980) käyttämää nimitystä.

käyttöä puhekielessä säätelee sen tehtävä logoforisena pronomininä: se viittaa referaatissa lausuman alkuperäiseen tuottajaan, esim. *se sano että hänelle voi soittaa* (VISK, § 1469). Olen väitöskirjan aiemmassa osatutkimuksessa todennut samasta aineistosta, että logoforisuus laajenee eläytymiseen ja tulkintaan siten, että pronomini *hän* toimii usein ajattelua tai tunnetta ilmaisevan verbin subjektina (Priiki 2014b).

Tässä artikkelissa kokoan aiemmin *hän* ~ *se* -vaihtelun kannalta tärkeiksi havaitut muuttujat samaan analyysiin. Kummilla on suurempi vaikutus, kielenulkoisilla vai -sisäisillä tekijöillä? Ovatko sosiaaliset muuttujat, joiden suhteen vaihtelua on tarkasteltu paljon, tärkeämpiä kuin kielensisäiset muuttujat? Entä miten voidaan päästä käsiksi puhujakohtaisiin eroihin, joita aiemmissa tutkimuksissa (Lappalainen 2010a; Nuolijärvi 1986, 187; Paunonen 1995, 172) on luonnehdittu suuriksi?

Satakuntalaisuus puheessa -hankkeen (ks. Kurki & Siitonen 2009) aineistosta on valittu tarkastelun kohteeksi äänitteitä viidestä kaakkoissatakuntalaisesta kunnasta. Niissä puhuttava lounainen välimurre on kiinnostava kohde kolmaspersoonaisen pronominiin tarkastelulle, koska siinä pronominia *hän* on käytetty verrattain paljon myös muuten kuin logoforisesti. Lounaiset välimurteet sijoittuvat lounaismurteiden ja hämäläismurteiden väliin: lounaismurteissa *hän*-pronominia on vanhastaan käytetty runsaasti viittaamassa niin ihmisiin kuin eläimiinkin, kun taas hämäläismurteissa käyttö on rajoittunut referaatteihin. (Vilppula 1989, 389; SMS, s.v. *hän*.) Tällä vuosituhanella äänitetyssä satakuntalaisaineistossa kenelläkään *hän* ei ole yksinomainen henkilöön viittaava pronomini, mutta nuoristakin ainakin osa käyttää sitä silloin tällöin (Kurki, Siitonen, Väänänen, Ivaska & Ekberg 2011, 98; Priiki 2011). Aineistosta, sen rajauksesta ja käsittelystä kerron enemmän luvussa 2. Luvussa 3 esittelen käytetyn tilastollisen menetelmän, satunnaismetsän. Tulokset esitetään koko aineiston osalta luvussa 4 ja puhujaryhmiin jaoteltuna luvussa 5. Lopuksi kokoan havainnot ja pohdin niiden merkitystä luvussa 6.

2. Aineisto ja sen luokittelu

Tutkimuksen pääaineistona on 24 tuntia monenkeskisiä keskusteluja Kokemäeltä, Harjavallassa, Huittisista, Säskylästä ja Köyliöstä. Ääniteaineisto on kerätty Satakuntalaisuus puheessa -tutkimushankkeessa 2009–2011 (hankkeesta Kurki & Siitonen 2009). Tässä tarkasteltavan ääniteaineiston muodostavat 19 keskustelua, joissa on mukana 33 henkilöä, joiden puhetta tarkastellaan. Keskusteluihin osallistuvat myös aineistoa keränneet äänittäjät, mutta heidän tuottamiaan pronomini viittauksia ei ole poimittu aineistosta tehtyyn tietokantaan. Aineiston kerääjät ovat Satakunnasta kotoisin olevia suomen kielen opiskelijoita. He ovat äänittäneet kotiseudullaan ihmisiä, jotka ovat asuneet pääosan elämästään tutkimuksen paikkakunnilla. Olen kerännyt äänitteiden litteraateista 2 833 yksinään nominilausekkeen muodostavaa yksilöllistä *se*- tai *hän*-pronominia, jotka viittaavat henkilöön, joka ei ole läsnä keskustelutilanteessa.

Keskityn tutkimuksessa yksiköllisiin muotoihin, sillä aiemmin on todettu yksiköllisen *hän*-pronominin poikkeavan frekvenssiltään monikollisesta *he*-muodosta (Kurki ym.

2011, 95; Lappalainen 2010a, 287–288; Nuolijärvi 1986, 180–181; Paunonen 1995, 167). Tästä syystä yksiköllisiä ja monikollisia viittauksia on syytä tarkastella erillään. Aineiston ulkopuolelle on rajattu keskustelutilanteessa läsnä olevaan henkilöön viittaaminen, sillä olen todennut, että aineistossa pronominin jakauma on erilainen viitattaessa läsnäolijaan kuin poissaolijaan (Priiki 2012, 2015a).

Aineistossa 97 % yksiselitteisessä puheen referoinnissa logoforisesti käytetyistä pronomineista on *hän*-pronomineja. Logoforista käyttöä voidaankin pitää parhaiten *hän*-pronominin esiintymistä puhekielessä ennustavana yksittäisenä kontekstitekijänä (Priiki 2014b). Kvantitatiivisessa osuudessa analyysi on rajattu prototyypisiin johtolauseesta ja referaattiosasta koostuviin referointeihin, koska tällaiset rakenteet on helppo tunnistaa ja laskea yksiselitteisesti. Dramaattisessa kerronnassa referoinnin johtolauseet eivät kuitenkaan ole pakollisia ja dialogia voidaan esittää ilmaisemalla pelkästään persoonan vaihtelulla, ketä referoidaan (esim. Laitinen 2005, 86; ks. myös Priiki 2014a). Tällaisten tapaus-ten luokittelu logoforisiksi on kuitenkin joskus tulkinnanvaraista ja *hän*-esiintymien vertailu *se*-pronominin käyttöön ongelmallista laajan aineiston kvantitatiivisessa tarkastelussa, sillä joskus nimenomaan *hän* luo vaikutelman referoinnista. Pohdin logoforisuuden rajoja ja laajentumia lyhyesti luvussa 6 ja tarkemmin toisaalla (Priiki tulossa).

Puheen välittämisen lisäksi referointiin kuuluu ajattelun ja muiden mielenliikkeiden kerrontaa, joka tehdään johtolauseesta ja referaattiosasta koostuvalla rakenteella² (1). Kiinnostava havainto jo luokitteluvaiheessa on, että ajattelun referointi prototyypistä puheen referointia muistuttavalla rakenteella on aineistossa selvästi harvinaisempaa kuin puheen referointi.³ Johtolauseesta ja referaattiosasta koostuvissa ajattelun referoinneissa esiintyviä logoforisia pronomineja on aineistossa vain 9 esiintymää: 7 *hän*-pronominia ja 2 *se*-pronominia. Esiintymien vähäisyyden vuoksi olen yhdistänyt ajattelun referoinnit samaan luokkaan puheen referoinnin kanssa (asetelmassa 1 logoforinen käyttö).

(1) ja kun se meinaa että **hän** pääsee (Sapu-188)

Asetelmassa 1 on listattu tässä tutkimuksessa huomioidut muuttujat ja niiden luokat. Kielensisäisen kontekstin muuttujista olen saman aineiston aiemmassa tarkastelussa havainnut merkitseväksi logoforisen käytön lisäksi pronominisubjektin verbin merkityksen, kun verbit luokitellaan puhumista tai mielenliikkeitä ilmaiseviin verbeihin (Priiki 2014b). Nämä verbit on otettu tarkasteluun, koska aiemmin on esitetty logoforisen *hän*-pronominin leviävän johtolauseisiin (Ylikahri 1996, 187) sekä referentin ajatusten ja tunteiden tulkintaan (Siitonen 2008). Tässä aineistossa predikaattiverbin vaikutus on sellainen, että pronomini *hän* esiintyy suhteellisesti muita kolmaspersoonaisia pronomineja

2 Esimerkeissä johtolauseet on alleviivattu.

3 Kun toisen henkilön tunteita kuvaillaan, aineistossa esiintyy *että*-sivulauseiden lisäksi tunteen aiheuttajaa kuvaavia *kun*-sivulauseita. Tässä aineistossa päälauseessa mainittuun kokijaan viittaava *hän*-pronomini vaikuttaa esiintyvän melko tavallisesti myös *kun*-lauseissa, esim. *se oli vihanen ku häntä mentiin* (.) *herätettiin* (Sapu-188). *Hän*-pronominin tavallisuudesta sivulauseissa ja etenkin *kun*-lauseissa on havaintoja vanhoista Satakunnan murteista (SMS, s.v. *hän*; Toivola-Junttila 1966, 16; Vilppula 1989, 394), ja ilmiön tarkastelu nykykielessä vaatisi oman tutkimuksensa.

Asetelma 1. Tarkasteltavat muuttujat ja niiden luokat.

Kielensisäinen konteksti

Verbin merkitys	puhumisverbi / kognitio- tai emootioverbi / muu
Virittyneisyys (<i>priming</i>)	edellä <i>hän</i> / <i>se</i> / muu viittaus
Logoforinen käyttö	kyllä / ei

Kielenulkoiset kontekstit

Puhujan ikä	vanha (72–84 v.) / keski-ikäinen (49–53 v.) / nuori (12–31 v.)
Puhujan sukupuoli	nainen / mies
Referentin ikä	vanha / keski-ikäinen / nuori
Referentin tuttuus	vieras / tuttava / läheinen
Referentin institutionaalinen rooli	kyllä / ei
Yksilöllinen vaihtelu	yksilöt pseudonyymeittäin (33 nimeä)

useammin ajattelua tai tunnetta kuvaavan verbin subjektina mutta harvemmin puhumista ilmaisevan verbin subjektina (Priiki 2014b). Samassa yhteydessä olen todennut, että aineistossa pronomien *se* ja *hän* välillä ei ole tilastollisesti merkitsevää eroa siinä, toimivatko ne lauseessa subjektina vai muussa funktiossa (Priiki 2014b).

Merkitseväksi tekijäksi on aiemmassa tutkimuksessa (Priiki 2014b) osoittautunut myös *se*, millaista viittausta samasta tarkoitteesta edellä käytetään. Käytän ilmiöstä (engl. *priming*) nimitystä *virittyneisyys*⁴. Virittyminen on assosiaation mekanismi: samanlaiset tai yhteen liittyvät asiat tulevat mieleen helpommin kuin erilaiset. Puhekielen kannalta *se* tarkoittaa sitä, että on todennäköistä toistaa keskustelussa aiemmin esiintyneitä sanoja, muotoja tai rakenteita. (Bock & Griffin 2000; Kahneman 2012, 66–73.) Ilmiön on todettu olevan kytköksissä pronominin ja nollan vaihteluun ensimmäisessä persoonassa englannissa ja espanjassa (Torres Cacoullos & Travis 2014; Travis 2007). Tässä aineistossa virittyneisyys ilmenee viittausten ketjuuntumisena siten, että *se*-pronominia edeltää usein samaviitteinen *se*, *hän*-pronominia *hän* (Priiki 2014b).

Kielenulkoisen kontekstin muuttujat on valittu aiempien tutkimusten perusteella. Perinteisten sosiolingvististen muuttujien, iän ja sukupuolen, on aiemmin havaittu vaikuttavan pronomien *se* ja *hän* vaihteluun siten, että naiset, vanhat ja korkeasti koulutetut käyttävät *hän*-pronominia miehiä, nuoria ja matalasti koulutettuja enemmän (Mielikäinen 1988, 107; Nuolijärvi 1986, 187; Paunonen 1995, 171; Ylikahri 1996, 185–186).

4 Sosiaalipsykologiassa *priming* käännetään termillä *pohjustus* (Tieteen kansallinen termipankki, s.v. *priming*). Kahnemanin (2012, 66–73) teoksen käännöksessä *se* on kuitenkin suomennettu *viritysvaikutukseksi*. Tietyn sanan tai muodon ollessa kyseessä *virittyminen* on mielestäni *pohjustusta* kuvaavampi ilmaus.

Tässä tutkimuksessa näistä tarkastellaan vain ikää ja sukupuolta. Puhujien koulutustausta on tiedossa, mutta aineistosta puuttuvat ääripäät: kaikilla äänitetyillä on jonkinlainen perusasteen jälkeinen koulutus, kenelläkään ei akateemista koulutusta. Tämän vuoksi koulutustaustaa ei ole otettu muuttujaksi.

Sukupuolenkaan vaikutuksen tarkastelun suhteen aineisto ei ole ideaali, sillä puhujajoukko on naisvaltainen. Lisäksi aineiston miehet tuottavat naisiin verrattuna vähän henkilöviitteisiä kolmaspersoonaisia pronomineja. Aineiston epäsuhta on samansuuntainen Ylikahrin (1996, 185–186) ja Lappalaisen (2014) havaintojen kanssa: miehet puhuvat vähemmän ihmisistä tai ainakin viittaavat heihin harvemmin kolmaspersoonaisella pronomiinilla. Lappalainen (2014) arvelee haastatteluaineistonsa perusteella ilmiön liittyvän siihen, että haastattelijana on nuori nainen. Myös tässä tarkasteltavien äänitteiden kerääjät ovat parikymppisiä naisia. Tällaisen äänittäjän seurassa miehet eivät ehkä koe yhtä luontevaksi puhua perheenjäsenistään ja ystävistään kuin naiset.

Hän-pronomien referenteistä on aiemmissa tutkimuksissa tehty ristiriitaisia havainnoja. Tässä tutkimuksessa mahdollisesti vaikuttavina tekijöinä tarkastellaan referentin ikää, tuttuutta ja institutionaalista roolia. Lappalaisen (2010a) aineistossa *hän* viittaa yleisimmin läheiseen henkilöön, esimerkiksi puhujaa vanhempaan perheenjäseneseen. Satakuntalaisuus puheessa -hankkeen äänitteistä taas on laadullisin havainnoin arveltu, että tyypillinen *hän*-pronominin referentti olisi puhujalle etäinen (Kurki ym. 2011, 97). Mielikäinen (1982, 282) mainitsee Jyväskylän puhekielen perusteella *hän*-pronominin viittaussuhteina etäiset, kunnioitetut ja puhujalle tärkeät henkilöt. Kun referentti on keskustelussa läsnä, kaakkoissatakuntalaisessa aineistossa *hän* viittaa suhteellisesti useammin kuin muut kolmaspersoonaiset pronominit siihen keskustelun osallistajaan, joka on puhujalle vierain (Priiki 2012). Tämän tutkimuksen määrällistä analyysiä varten tehdyssä koodauksessa läheisiksi luokitellut henkilöt ovat puhujan perheenjäseniä tai hyviä ystäviä. Tuttaviksi on koodattu naapureita, koulu- ja työkavereita sekä ystävien perheenjäseniä. Vierain on määritelty esimerkiksi julkisuuden henkilöt, jotka puhuja on tavannut vain ohimennen tai ei koskaan.

Ikäluokan suhteen referentit on jaettu kolmeen ryhmään kuten puhujatkin. Referentin ikäryhmä on määritelty sillä perusteella, millaisena hänet kuvataan. Esimerkiksi kun puhuja kertoo poikansa lapsuudesta, on referentti koodattu nuoreksi, vaikka tämä äänityshetkellä on keski-ikäinen. Referenteistä on merkitty tietokantaan myös, esitetäänkö heidät institutionaalisessa roolissa, esimerkiksi lääkärinä tai opettajana⁵. Nimittäin Lappalaisen havaintojen mukaan *hän*-pronominia käytetään paitsi institutionaalisissa tilanteissa (Lappalainen 2010b), myös viittaamassa henkilöön, johon puhujalla on institutionaalinen suhde (Lappalainen 2010a, 291–292).

Ennalta tärkeiksi havaittujen sosiaalisten tekijöiden lisäksi tässä tutkimuksessa pyritään kartoittamaan, kuinka tärkeitä yksilöiden väliset erot vaihtelun kannalta ovat. Kaikki aineiston 33 puhujaa on koodattu tietokantaan pseudonyymeittäin (asetelmassa 1

5 Referentin ominaisuuksia koskevia tietoja ei aina ole mahdollista määritellä. Epäselvät tapaukset on jätetty tyhjiksi, jolloin ne eivät vaikuta muuttujan selittävyteen.

yksilöllinen vaihtelu). Yksilöllinen vaihtelu poikkeaa muista muuttujista siksi, että se on satunnaisvaikutusmuuttuja, kun taas muut tarkasteltavat muuttujat ovat kiinteän vaikutuksen muuttujia. Kiinteän vaikutuksen muuttujilla arvot ovat pysyviä ja niitä on rajallinen määrä: esimerkiksi missä tahansa puhujien joukossa kaikki voidaan luokitella miehiksi tai naisiksi. Yksilöiden kohdalla kuitenkin erilainen otos samasta populaatiosta sisältäisi erilaisia luokkia – eri kokoelmassa olisi eri puhujia. Koska perinteisissä monimuuttujamalleissa kaikkia muuttujia on pidettävä kiinteinä, yksilöllisen vaihtelun merkitystä ei pystytä yleistämään siihen joukkoon, jota otoksen on tarkoitus edustaa. Tässä tutkimuksessa käytettävissä menetelmissä analyysiin on kuitenkin mahdollista sisällyttää satunnaisvaikutusmuuttujia. (Tagliamonte & Baayen 2012.) Esittelen menetelmät seuraavassa luvussa.

Päähuomio tarkastelussa on ääniteaineistosta tehtävissä havainnoissa. Vertaan äänitteiden pronomineista löytyviä tendenssejä kuitenkin myös kielenkäyttäjien käsityksiin, joita olen kerännyt pienimuotoisella lomakekyselyllä. Kysely on kerätty osana aiempaa tutkimusta (Priiki 2011). Kyselyyn on vastannut 39 henkilöä: 27 eri-ikäistä satakuntalaista sekä 12 henkilön verrokkiryhmä eri puolilta Suomea.⁶ Satakuntalaisista vastaajista 8:n puhetta on tallennettu myös äänitteisiin. Kyselyvastaukset antavat laadullisen vertailukohdan ääniteaineiston havainnoille ja näkökulmaa sen arvioimiseen, mitä vaihtelun piirteitä puhujat tiedostavat ja mitä eivät.

3. Menetelmät

Tarkastelen pronomiinien vaihteluun vaikuttavia tekijöitä satunnaismetsämenetelmää (*random forest*) käyttäen. Olen aiemmin (Priiki 2014b, 2015b, 2016) käyttänyt saman aineiston tarkasteluun päätöspuuta (CHAID ja ECHAID⁷). Satunnaismetsä on päätöspuiden pohjalta kehitetty työkalu (Breiman 2001). Sen perustana on binäärinen päätöspuu-algoritmi, mutta yhden koko aineistoa kuvaavan sijaan päätöspuuta tehdään satoja tai tuhansia – näin muodostuu metsä.

Päätöspuu on luokitteleva monimuuttujamenetelmä, joka jakaa aineistoa rekursiivisesti joko kahteen tai useampaan osaan siten, että aineiston osien välille muodostuu mahdollisimman suuria eroja. Se palvelee myös tämän artikkelin analyysin yhdessä vaiheessa (luku 5). Päätöspuuanalyysi sopii erityisen hyvin aineistoon, jossa selittävien muuttujien välillä on keskinäisiä riippuvuuksia, koska se on taustaoletuksiltaan hyvin salliva. Aiemmat saman aineiston tarkastelut päätöspuulla (Priiki 2014b, 2015b, 2016) ovatkin paljastaneet kiinnostavia muuttujien välisiä kytköksiä.

Tämän artikkelin tavoitteena on koota aiemmin merkitseviksi havaitut tekijät samaan analyysiin ja selvittää niiden tärkeysjärjestys. Kun merkitseviä muuttujia on runsaasti, päätöspuu kohtaa ongelmia. Ensinnäkin suuresta aineistosta syntyvä monihaaranainen

6 Kysymykset ja lisätietoja vastaajista Priiki 2011, liitteet.

7 Toteutettu IBM SPSS Statistics -ohjelmalla (versio 22).

kuvio ei enää ole havainnollinen. Toisekseen puun ongelmaksi voi koitua sen hierarkisuus: jos selittävästä muuttujista kaksi on miltei yhtä merkitseviä, analyysi näyttää kuitenkin vain merkitsevemmän. Tällöin alaryhmissä ei ehkä enää olekaan tarpeeksi esiintymiä, jotta toinen, miltei yhtä merkitsevä muuttuja tulisi näkyviin. (Tagliamonte & Baayen 2012, 24.) Tässä artikkelissa keskityn muuttujien tärkeysjärjestykseen ja siksi olen valinnut menetelmäksi satunnaismetsän. Siinä muodostuvia puita testataan arvioimalla, kuinka hyvin ne selittävät tarkasteltavaa ilmiötä. Näin menetelmä asettaa selittävät tekijät paremmuusjärjestykseen. Kuten päätöspuu, myös satunnaismetsä on menetelmä, joka sopii myös pienehkölle ja epätasaiselle aineistolle, jossa selittävien tekijöiden välillä on keskinäisiä riippuvuuksia (Tagliamonte & Baayen 2012, 24). Satunnaismetsää ovat fennistiikassa aiemmin käyttäneet Ilmari Ivaska (2014, 2015) ja Milja Väänänen (2015).

Olen toteuttanut analyysit R-ohjelmointiympäristössä *party*-paketin *cforest*-funktiolla. Satunnaismetsäanalyyssissä aineistosta otetaan useita satunnaisotoksia, joista jokaisesta muodostetaan päätöspuu. Jokaisen satunnaisotospuun kuvaamaa mallia verrataan jäljelle jääneen aineiston tilanteeseen. (Tagliamonte & Baayen 2012, 22.) Satunnaismetsä tuottaa kullekin selittävälle tekijälle sen tärkeyttä ilmaisevan tunnusluvun (*variable importance*⁸). Analyysi vertaa todellisia selittäviä tekijöitä täysin satunnaiseen tekijään. Mitä suurempi ero satunnaistekijään on, sitä tärkeämpi selittäjä on kyseessä. (Tagliamonte & Baayen 2012, 23.) Koska satunnaistekijöiden arvot ovat satunnaisia, eri analyysikertojen tulokset voivat poiketa toisistaan. Tulokset on siis varmistettava useilla analyysikerroilla. Saadut kertoimet ilmaisevat vain tarkasteltavien tekijöiden järjestystä ja suhdetta toisiinsa. (Baayen, Janda, Nessel, Dickey, Endresen & Makarova 2013; Strobl, Malley & Tutz 2009, 342.)

Kuten satunnaistekijöiden, myös epärelevanttien tekijöiden arvot vaihtelevat sattumanvaraisesti nollan ympärillä. Jos rajakohdan, jossa muuttujan tärkeys on olennaisesti suurempi kuin nolla, haluaa määritellä, voi muuttujaksi lisätä binäärisen satunnaisluvun. Tällainen muuttuja on varmuudella täysin merkityksetön selitettävän muuttujan kannalta, ja sen itseisarvo kertoo, kuinka paljon muuttujan selitysarvo voi pelkän sattuman takia poiketa nolasta (Lohmann 2013, 11–12).

Satunnaismetsä on sitä luotettavampi, mitä enemmän puita siinä on.⁹ Tässä tutkimuksessa puita on kussakin satunnaismetsäanalyyssissä 1 000. Toinen analyysiin vaikuttava luku on satunnaistekijöiden määrä, johon suositus on muuttujien määrän neliöjuuri (Cutler, Edwards, Thomas, Beard, Cutler, Hess, Gibson & Lawler 2007). Olen toistanut jokaisen analyysin kolmella eri jyvityksellä (*seeding*, suomennos Ivaska 2014, 170). Jyvitys voi olla mikä tahansa luku, ja sen avulla algoritmi muodostaa käytettävät satunnaistekijät. Luvuissa 4 ja 5 esitettävät tulokset perustuvat aina kolmen analyysin keskiarvoon (keskiarvoja hyödyntävään ratkaisuun on päätynyt myös Ivaska 2015, 220). Lisäksi tuloksissa ilmoitetaan esitetyn mallin konkordanssiluku (C). Jotta mallin voidaan ajatella selittävän vaihtelua hyvin, C-arvon tulee olla vähintään 0,8 (Baayen ym. 2013, 8).

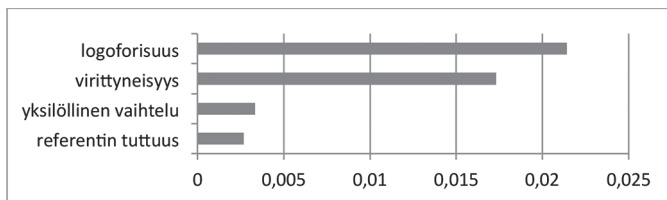
8 Väänänen (2015) käyttää *variable importance* -käsitteestä suomennosta *muuttujan selittävyys*.

9 Mitä suurempi puiden, tarkasteltavien muuttujien ja aineiston esiintymien määrä on, sitä enemmän laskentatehoa menetelmä vaatii.

4. Pronominien *hän* ja *se* vaihteluun vaikuttavien tekijöiden tärkeys

Olen poiminut aineiston äänitteistä tarkasteluun 549 henkilöviitteistä *hän*-pronominia ja 2 285 *se*-pronominia. Aineistosta on tehty satunnaismetsä, jossa selitettävänä muuttujana on pronomini (*se* tai *hän*) ja selittäjinä edellä asetelmassa 1 esitetyt muuttujat. Päätöspuiden kokoa on rajattu siten, että kuhunkin voi syntyä vain 20 haaraa. Satunnaismetsä on toistettu kolmella eri jyvityksellä. Niitä muuttujia, joiden arvo on kaikissa analyyseissä parempi kuin satunnaistekijän itseisarvo, löytyy vain neljä: logoforisuus, virittyneisyys, yksilöllinen vaihtelu ja referentin tuttuus. Näiden neljän tekijän selittävyys esitetään kuviossa 1. C-arvo on 0,94 eli malli selittää vaihtelua varsin hyvin.

Kuvio 1. Pronominien *hän* ja *se* vaihteluun vaikuttavien tekijöiden selittävyys (C = 0,94).



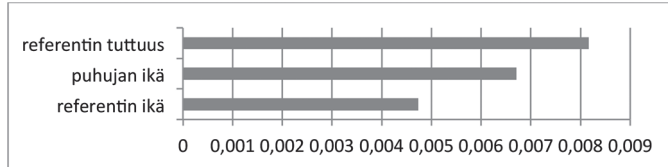
Muuttujista tärkeimmiksi osoittautuvat pronominin logoforisuus sekä virittyneisyys. Logoforisuus on odotuksenmukaisesti tärkeä tekijä, sillä useimmilla aineiston puhujilla referaatissa viittaamassa johtolauseessa mainittuun puhujaan esiintyy vaihtelematon *hän*. Yllättävää on, että pronominin virittyneisyys on tässä aineistossa miltei yhtä tärkeä tekijä. Virittyneisyyden merkitystä selittää se, että hyvin suurelle osalle aineiston pronomini-esiintymistä on mahdollista määritellä edeltävä viittaus, sillä pronominiensimaininnat ja muuten epäselvät tapaukset ovat harvinaisia. Koko aineistossa 60 %:a *se*-pronomineista edeltää samaviitteinen *se* ja 44 %:a *hän*-pronomineista edeltää *hän*. Selvännän logoforisuuden ja virittyneisyyden ilmenemistä tarkemmin erilaisten puhujaryhmien tarkastelun myötä luvussa 5.

Kielenulkoisista muuttujista tärkein on yksilöllinen vaihtelu. Tulos ei yllätä, koska pronominien vaihtelussa yksilöllisten erojen on jo aiemmin havaittu olevan suuria (Lappalainen 2010a; Nuolijärvi 1986, 187; Paunonen 1995, 172; Priiki 2011). Yksilöllinen vaihtelu on todettu tärkeimmäksi selittäjäksi monien muidenkin kielenpiirteiden kohdalla, joiden vaihtelua on tarkasteltu tilastollisin menetelmin (ks. esim. Baayen 2008; Klavan, Pilvik & Uiboed 2015, 213; Tagliamonte & Baayen 2012, 24). Muista kielenulkoisista tekijöistä vain referentin tuttuus osoittautuu merkitseväksi tässä analyysissä.

Jos logoforisuuden ja virittyneisyyden vaikutus eliminoidaan eli tarkastellaan vain sellaisia pronomini-esiintymiä, jotka eivät ole logoforisia ja joita edeltää jokin muu viittaus kuin sama pronomini, ja yksilöllinen vaihtelu -muuttuja jätetään pois, aineistosta löytyy muitakin kielenulkoisia merkitseviä muuttujia (kuvio 2). Tällöin referentin tuttuuden ohella merkitseviksi nousevat puhujan ikä ja referentin ikä. Nämä kaikki ovat tekijöitä, jotka on nostettu esiin aikaisemmissa *hän*- ja *se*-pronominien vaihtelua kuvaavissa

tutkimuksissa (Lappalainen 2010a; Paunonen 1995, 171; Mielikäinen 1982, 282; Nuolijärvi 1986, 290–300). Tällainen perinteisille kielenulkoisille muuttujille perustuva malli esitetään kuviossa 2.

Kuvio 2. Pronominien *hän* ja *se* vaihteluun vaikuttavien tekijöiden selittävyys ei-logoforisissa ja virittymättömissä tapauksissa ilman yksilöllinen vaihtelu -muuttujaa ($C = 0,76$).



Kuvion 2 mallin selitysteho (C) ei ole hyvä: C -arvo 0,76 on alle sen rajan, että mallia voidaan Baayenin ym. (2013, 8) mukaan pitää hyväksyttävänä. Jos kuvion 2 ei-logoforisten ja virittymättömien pronominien vaihtelun analyysiin lisätään muuttujaksi yksilöllinen vaihtelu, C -arvo nousee 0,84:een. Tämäkään ei ole kovin hyvä luku, mikä kertoo, että ei-logoforisten ja virittymättömien *se*- ja *hän*-pronominien vaihteluun vaikuttavat muutkin tekijät kuin tässä tarkastellut. Pohdin seuraavissa luvuissa, millaiset tekijät voivat olla kyseessä.

Kuten edellä olen todennut, aiemmassa tutkimuksessa *hän*-pronominin referenttejä on kuvattu toisinaan etäisiksi, toisinaan läheisiksi puhujalle. Tässä tarkasteltavissa keskusteluissa *hän* viittaa yleensä puhujalle läheisiin ystäviin ja perheenjäseniin. Kun aineistoa katsotaan tarkemmin, tuttuus tai vieraus ei kuitenkaan sinänsä vaikuta vaihtelun perimmäiseltä syytä. Tuttuus nimittäin kytkeytyy logoforisuuteen ja siihen, millaisia kertomuksia erilaisista referenteistä kerrotaan. Selvennän kytköstä jäljempänä. Näyttää siltä, että erilaisia referenttejä esitellään ja kuljetetaan keskustelussa erilaisin keinoin: näiden kattava tarkastelu vaatisi oman tutkimuksensa, jossa otettaisiin huomioon pronominien ohella myös henkilöviittaukset nimillä ja muilla substantiivilausekkeilla.

Aiempien havaintojen (Lappalainen 2010a; Paunonen 1995, 171) perusteella on odotuksenmukaista, että vanhat puhujat käyttävät *hän*-pronominia nuoria enemmän. Ilman yksilöllinen vaihtelu -muuttujaa tällainen yhteys aineistosta löytyykin (kuvio 2). Tässä 33 puhujan aineistossa yksilöiden väliset erot osoittautuvat kuitenkin suuriksi, ja yksilöllinen vaihtelu selittää pronominien käyttöä selvästi paremmin kuin puhujan ikä kolmeen ikäryhmään jaoteltuna. Referentin ikä taas on kytköksissä puhujan ikään: vanhat puhuvat useammin vanhoista henkilöistä, nuoret nuorista. Seuraavassa luvussa perehdyn yksilölliseen vaihteluun tarkemmin jakamalla puhujat ryhmiin sillä perusteella, kuinka paljon he käyttävät *hän*-pronominia. Näitä ryhmiä tarkastelemalla kuva puhujan iän, referentin iän ja tuttuuden sekä *hän*-pronominin käytön korrelaatiosta selventyy.

Satunnaismetsäanalyysin tuloksissa on kyse kunkin tekijän vaikutuksesta vaihteluun yleisesti koko aineiston tasolla. Tästä syystä jotkin tekijät voivat saada matalan arvon, vaikka ne tietyssä aineiston osassa olisivat hyvinkin tärkeitä vaihtelun selittämiseksi.

(Lohmann 2013, 13.) Esimerkiksi kognitio- ja emotioverbiin määrä on koko aineistoon suhteutettuna vähäinen. Kyseisten verbiin subjektien joukossa *hän*-pronominien suhteellinen osuus on suurempi kuin muissa yhteyksissä (Priiki 2014b), mutta ilmiö ei harvinaisuutensa vuoksi selitä kovin suurta osaa koko aineiston *hän*-esiintymistä. Tästä syystä verbiin merkitys ei nouse tärkeäksi selittäjäksi satunnaismetsäanalyysissä. Aineiston tarkastelu puhujaryhmittäin seuraavassa luvussa paljastaa, että jotkin tekijöistä – edellä mainitun verbiin merkityksen lisäksi referentin institutionaalinen rooli – ovat varteenotettavia selittäjiä tietyssä puhujaryhmässä, vaikkei niillä ole selitysvoimaa koko aineiston tasolla.

5. Puhujaprofileja

Tässä luvussa perehdyn tarkemmin siihen, millaisia puhujien väliset erot *hän* ~ *se* -vaihteluun vaikuttavissa tekijöissä ovat. Kun koko aineiston *se*- ja *hän*-pronominista tehdään päätöspuu, jossa yksilöllinen vaihtelu on selittävänä tekijänä, algoritmi jakaa puhujat viiteen ryhmään. Seuraavaksi esittelen, millaisia eroja ryhmien välillä on siinä, mitkä tekijät pronominien vaihteluun vaikuttavat. Ensimmäisessä ryhmässä vaihtelua ei ole juuri lainkaan. Muista neljästä ryhmästä olen kustakin tehnyt kolme satunnaismetsää, ja esitän analyysien keskiarvot kuvioina.

Kuten edellä totesin, osa äänitetyistä henkilöistä on vastannut myös ääniteaineistoa tukevaan lomakekyselyyn. Kunkin ryhmän kohdalla luon katsauksen myös siihen, kuinka eri puhujaryhmien äänitteistä löytyvät *hän*-pronominin esiintymistä selittävät tekijät ovat yhteydessä kyselyyn vastanneiden puhujien omiin arvioihin.

5.1. Ryhmä 1. Lähes vaihtelematon *se* (yli 99 %)

Ensimmäisessä ryhmässä on yhdeksän puhujaa, jotka käyttävät *hän*-pronominia hyvin vähän tai eivät ollenkaan. Yhteensä nämä yhdeksän puhujaa tuottavat 584 kolmaspersoonaista henkilöön viittaavaa pronominia, mutta *hän*-pronominia esiintyy vain yhdellä ryhmän puhujista. Tämä puhuja on 26-vuotias Karri¹⁰ (Sapu-tunnukset 188b ja 190d), joka puhuu ihmisistä enemmän kuin muut ryhmään kuuluvat. Karrin tuottamista 282 pronominiesiintymästä *hän*-pronominin (3 esiintymää) muodostavat noin yhden prosentin: määrä ei eroa tilastollisesti merkitsevästi vaihtelemattomista puhujista. Kaikki Karrin *hän*-pronominin ovat selvästi logoforisia siten, että referaattia edeltää puhumista ilmaiseva johtolause, kuten esimerkissä 2.

(2) Ei *se* tykkää ajjaap pikatiätä (Sapu-190)

01 Karri: mt kais **se** katos se ei- ei **se** tykkää ajjaap pikatiätä.

02 [(**se**) sannoo aina

03 Mika: [siitt_om pa:ljom parempi ajjaan niinkun [tämmösel (tiällä).

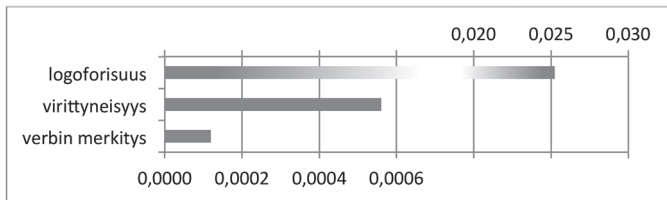
04 Karri: [↑juu

10 Puhujien nimet ja esimerkeissä esiintyvät yksityishenkilöiden nimet on muutettu.

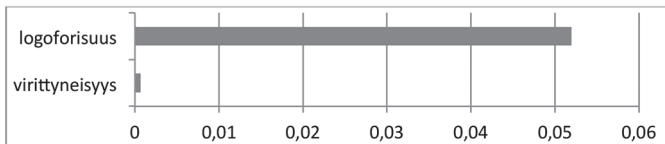
aineistossa yhteensä 413. Yhtä lukuun ottamatta aineiston nuoret puhujat kuuluvat siis kahteen vähiten *hän*-pronominia käyttävään ryhmään. Ryhmässä 3 on neljä naista: nuori, kaksi keski-ikäistä ja vanha. Heiltä on tallentunut aineiston keskusteluihin tarkasteltavia pronomineja 843, ja niistä on *hän*-pronomineja keskimäärin 14 %. Ryhmä 3 kuvaa hyvin sitä, miten jako korreloi puhujien iän kanssa – siihen kuuluu puhujia kaikista ikäryhmistä, kun taas kolmatta ryhmää vähemmän *hän*-pronominia käyttäviin kuuluu pääosin nuoria. Jäljempänä esittelen eniten aineistossa *hän*-pronominia käyttävät ryhmät 4 ja 5, joihin puolestaan kuuluu pääosa vanhoista puhujista.

Käsittelen sekä vähän että jonkin verran *hän*-pronominia käyttävät puhujat tässä luvussa, koska näissä ryhmissä tärkeimmät pronominiin vaihtelua selittävät tekijät ovat samoja. Esitän molemmista ryhmistä tehtyjen satunnaismetsäanalyysien tulosten keskiarvot kuvioissa 3 ja 4. Kuten edellä, näissäkin kuvioissa esitän vain ne tekijät, joiden selittävyys ylittää kaikissa analyyseissä satunnaisen tekijän itseisarvon.

Kuvio 3. Vähän *hän*-pronominia käyttävien puhujien *hän* ~ *se* -vaihteluun vaikuttavien tekijöiden selittävyys ($C = 0,96$).



Kuvio 4. Jonkin verran *hän*-pronominia käyttävien puhujien *hän* ~ *se* -vaihteluun vaikuttavien tekijöiden selittävyys ($C = 0,84$).



Kuviossa 3 esitetyt tulokset osoittavat vähän *hän*-pronominia käyttävien puhujien pronominivaihtelulle kolme merkitsevää selittäjää: logoforisuuden (selittävyys 0,026), virittyneisyyden (0,00051) ja subjektipronominin predikaattiverbin merkityksen (0,000095). Korkea C -arvo (0,96) kertoo, että näiden tekijöiden avulla on mahdollista selittää lähes kaikki ryhmän *hän*-pronominit. Myös jonkin verran *hän*-pronominia käyttävien ryhmässä (kuvio 4) merkitsevät selittäjät ovat logoforisuus (0,033) ja virittyneisyys (0,00054). Logoforisuuden selittävyys on kummassakin ryhmässä yli 50-kertainen virittyneisyyden selittävyyteen verrattuna. Kuviossa 3 se esitetään eri asteikolla (yläakseli).

Vähän *hän*-pronominia käyttävien tuottamista 22 *hän*-pronominista 16 (73 %) on selvästi logoforisia. Vastaavasti jonkin verran *hän*-pronominia käyttävillä joko puheen tai ajattelun referoinneissa esiintyy ryhmän 116 *hän*-pronominista 71 esiintymää (61 %).

Kummassakin näistä ryhmistä *hän*-pronominin logoforisuus kuitenkin horjuu – johtolauseella alkavissa referaateissa tavataan viisi puhujaa viittaavaa *se*-pronominia.

Näiden kahden puhujaryhmän osa-aineistoissa *se*-pronomineista 61 % seuraa toista samaviitteistä *se*-pronominia. Myös *hän*-pronominin edellä on melko usein samaviitteinen *hän* (28 %). *Hän*-pronominit virittyvät näillä puhujilla kuitenkin heikommin kuin koko aineiston tasolla, jossa *hän*-viittauksista 44 % seuraa toista *hän*-viittausta. Tämä johtuu yksinkertaisesti siitä, että *hän*-viittaukset ketjuuntuvat vahvemmin niillä puhujilla, joilla niitä esiintyy paljon. Tästä kerron lisää jäljempänä ryhmien 4 ja 5 esittelyn yhteydessä.

Ryhmän 2 osa-aineistossa *se*- tai *hän*-subjektin predikaattina esiintyvistä 23 kognitio- ja emootioverbistä joka viidennen (6 esiintymää; 21 %) subjekti on *hän*, vaikka muuten ryhmään kuuluvien henkilöiden puheessa *hän*-pronominin osuus on keskimäärin vain 5 %. Kuitenkin viisi näistä kuudesta kognitio- tai emootioverbin *hän*-subjektista on logoforisia, kuten esimerkissä 4.

(4) Harjavallas aina asunu (Sapu-161)

01 Anna: se oj just aika kyl (.) paha sillai ku (.) mt (.) mut niin^{ko}
 02 ↑se kaveri just kuka Turus nyt (.) on opiskelemas ni
 sanoki et
 03 tota et .hhkyl **hän** kuitenki haluis ko Harjavallas on niin^{ku}
 04 aina asunu et
 05 N: .mhh
 06 Anna: nytt=on sit tosissaav vaan nää vuadet opiskellus siäl ni

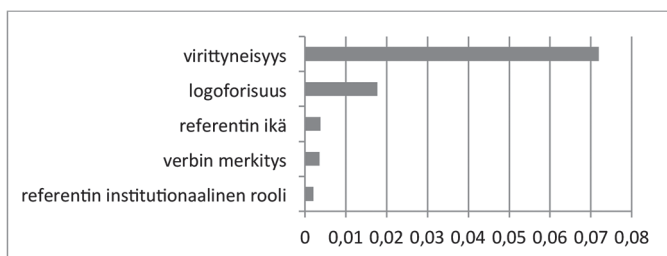
Vähän *hän*-pronominia käyttävien henkilöiden *hän*-esiintymät siis selittyvät parhaiten kielensisäisten tekijöiden eli virittyneisyyden ja logoforisuuden kautta. Sama pätee jonkin verran *hän*-pronominia käyttäviin. Jonkin verran *hän*-pronominia käyttävien puhujien ryhmä on kuitenkin hajanaisempi kuin vähän *hän*-pronominia käyttävien ryhmä. Sen lisäksi, että siihen kuuluu kaikenikäisiä puhujia, myös pronominiin käyttötavat vaikuttavat sekalaisilta. Ryhmän 3 satunnaismetsäanalyysin C-arvo (0,84) on heikompi kuin ryhmän 2 analyysissä (0,96). Tämä kertoo, että osa-aineistossa on enemmän sellaisia tapauksia, joita ei selitä virittyneisyys eikä logoforisuus, vaan vaihtelun taustalla on tässä tehdyn määrällisen analyysin tavoittamattomissa olevia syitä.

Neljä vähän tai jonkin verran *hän*-pronominia käyttävää puhujaa on vastannut lomakekyselyyn. Kaikki tiedostavat käyttävänsä arkipuheessa henkilöihin viitattaessaan pääosin *se*-pronominia. He mieltävät *hän*-pronominin kuuluvan asiointitilanteisiin sekä kohteli-aaksi viittaukseksi vanhoihin tai vieraisiin ihmisiin. Kaksi neljästä tunnistaa *hän*-pronominin epäkohteliaan, ironiseksi tai ivalliseksi käsitettävän käytön. On mahdollista, että he käyttävät sitä ironisessa kaiuttamisessa, vaikka sopivia konteksteja ei äänitteisiin ole tallentunut (*hän*-pronominin ironisesta käytöstä ks. Priiki 2014a). Logoforista käyttöä tai pronomini viittausten taipumusta ketjuuntua kukaan ei kyselyssä mainitse.

5.3. Ryhmä 4. Melko paljon hän-pronominia (36 %)

Neljännän ryhmän puhujat käyttävät *hän*-pronominia melko paljon: heillä sitä on keskimäärin 36 % *hän*- ja *se*-pronominien yhteenlasketusta määrästä (729 esiintymää). Ryhmään kuuluu yhdeksän henkilöä, jotka yhtä 51-vuotiasta miestä lukuun ottamatta on luokiteltu iältään vanhoiksi (72–84-vuotiaat). Tässä ryhmässä huomiota herättää, että siihen kuuluvat aineiston ainoan keski-ikäisen miehen lisäksi myös kaikki aineiston kolme vanhaa miestä. Kuten edellisistä ryhmistä, olen jälleen tehnyt kolme satunnaismetsäanalyysiä. Esitän tulosten keskiarvon merkitseviksi osoittautuvien tekijöiden osalta kuviossa 5.

Kuvio 5. Melko paljon *hän*-pronominia käyttävien puhujien *hän* ~ *se* -vaihteluun vaikuttavien tekijöiden tärkeysjärjestys (C = 0,86).



Kuvio 5 osoittaa, että ryhmässä 4 merkitsevien tekijöiden joukko on erilainen kuin vähemmän *hän*-pronominia käyttävillä. Logoforisuuden selittävyys ei ole yhtä hallitseva – se on selittävyydeltään vasta toiseksi tärkein ja mahtuu nyt samalle asteikolle muiden kanssa. Tärkeimmäksi selittäjäksi nousee virittyneisyys. Virittyneisyyden suurempi merkitys tässä ryhmässä onkin loogista: kun *hän*-pronomeineja on puhujilla enemmän, ne myös esiintyvät peräkkäin useammin, ja kun käyttö ei rajoitu prototyyppiseen referointiin, mahdollisuuksia *hän*-pronominin esiintymiseen on enemmän. Tässä ryhmässä selittäjänä on myös verbin merkitys (ks. Priiki 2014b), joka on tärkeä myös ryhmässä 2 (kuvio 2 edellä). Tässä osa-aineistossa *hän*-pronomeineista 15 % toimii ajattelua tai tunnetta ilmaisevan verbin subjektina, henkilöviitteisistä *se*-pronomeineista vain 3 %. Tällaisen verbien käyttöä *hän*-subjektin kanssa kuvastaa esimerkki 5. Ennen katkelmaa Raisa on kertonut löytäneensä lapsena veljeltään saamansa kirjeen, josta veli oli nolostunut.

(5)

```
01 Raisa: ja sit mää snoi et ku mull_on se talles et mää voin sen sul
02      vaik näyttää. <@jaa no mamma on käskeny@>. ettei hän
03      [↑keh(d)annu- aikunem ↑miäs sit- ööh öö ninku: (0.5)
04 Esko:  [hö hö
05 Raisa: myäntää >et (tai) < tunnustaa et hänki on_ollul ↑laps
06      ja kirjottanut ↑lapsellisesti.
07 Esko:  joo [o.
08 Raisa:  [minust se oli nin täyttä asiaa et-
09 Esko:  oli se >ihan ihan ihan i- ihan< semmost asiaa mitä hä- mitä
10      sillon hän koki.
```

(Sapu-165)

Esimerkissä 5 kyse on selvästi logoforisen käytön laajenemisesta prototyypin referoinnin ulkopuolelle. Katkelman alussa Raisa kuvaa keskustelua veljensä kanssa ja referoi veljensä reaktiota suoralla sitaatilla, joka merkitään ääntä muuttamalla. Tätä seuraa tulkinta veljen reaktion syistä tämän mielenliikkeiden referoinnilla. Johtolauseessa Raisa käyttää *kehdata*-verbiä, jonka subjektina on *hän* (r. 2–3). Katkelman lopussa Esko tuottaa tulkinnan kirjeen sisällön aitoudesta riveillä 9–10: verbinä on *kokea* ja subjektina säilyy edelleen *hän*.

Toisin kuin vähemmän *hän*-pronominia käyttävillä puhujilla, ryhmässä 4 referentin ikä ja institutionaalisuus ovat merkitseviä. Niiden selittävyys on kuitenkin heikompa kuin virittyneisyyden tai logoforisuuden. Ryhmän puhujien *hän*-pronomeista 44 % viittaa vanhaan henkilöön, *se*-viittauksista 31 %. Referentin ikä on yhteydessä referentin tuttuuteen: ryhmän viittauksissa vanhaan henkilöön (260 esiintymää) valtaosassa puhuja tuntee referentin joko hyvin tai ainakin jotenkin – vieraaseen vanhaan henkilöön on vain kolme viittausta.

Lomakekyselyn vastauksissa kielenkäyttäjät mieltävät vanhoihin henkilöihin *hän*-pronominilla viittaamisen syyksi kohteliaisuuden, mutta aineiston keskustelukatkelmien laadullisessa analyysissä kertomisen tapa vaikuttaa osuvammalta selitykseltä. Puhujalle tutut vanhat ihmiset esiintyvät usein kertomuksissa, jotka kertoja on alun perin todennäköisesti kuullut päähenkilöltä itseltään. Esimerkki 6 on kolmen vanhan naisen ja nuoren äänittäjän keskustelusta, ja siinä Maija kertoo tarinaa, jossa pääosassa on hänen puolisonsa Martti. Martti on aineistossa luokiteltu vanhaksi ja puhujalle läheiseksi referentiksi.

(6) Varalasiat (Sapu-210)

- 01 Maija: Martil [oli: sit s- koko soran_ajam puseron taskus ku
 02 Erja: [mm m
 03 Maija: sillon_oli sotilaspuseros [semmonen tasku
 04 Kaija: [mm.
 05 Maija: **hänell**_oli varalasiat [siinä.
 06 Kaija: [jaha. joo.
 07 Maija: ei **hän** tarvinnun niit kertaakam mut nep puto:s kerran Gaevooh kh
 08 [hh **häm** ph- ne varalasiat h h h h [h h h
 09 N: [hh h
 10 Erja: [hh h
 11 Kaija: niin [(mul vissiin kans).
 12 Erja: [saatiiks ne siält [pois.
 13 Maija: [eei niit siält saatu **hän** [(.)
 14 N: [hh
 15 Maija: **hänel** oli tiätyst niit sittek kottoo lähettivät toiset et
 16 **hänel** oli varalasiat sitä varten taskus ain et jos jos
 17 mennee muttei selvis siäl mut (.) ei kotonaka mennyk ku
 18 sen_yhren_ainoon kerran ja mää löysin nek kyl iha
 19 hetii eikä ne rikki ollum menny. mut **hänn**_ei uskaltanu
 20 itten niit ruvet hakemaa ku sano et kyl **hän** sen siit
 21 tiäsi ettei nek kaukan oom mut ko,
 22 Kaija: niij joo mut voi astuup pääl

Hän-pronominilla voidaan varsinaisen logoforisen käytön ohella tuoda kertomukseen toisen käden tiedon sävyä (ks. Lappalainen 2010a, 302–304). Tässä kertomuksessa

ainakin osasyynä *hän*-pronominin käytölle voidaan nähdä, että kertomus Martin silmälasien putoamisesta kaivoon sodan aikana todennäköisesti on Martin itsensä Maijalle kertoma ja siinä mielessä toisen käden tietoa. Esimerkissä näkyy hyvin myös se, mitä virittyneisyys aineistossa tarkoittaa: erisnimen (*Martil*, rivi 1) jälkeen Maijan viittauksena Marttiin toistuu läpi koko kertomuksen *hän*.

Institutionaalisessa roolissa, siis esimerkiksi opettajana tai lääkärinä, esitettävään referenttiin puolestaan viitataan ääniteaineistossa muita henkilöreferenttejä harvemmin pronomiinilla *hän*, mikä on vastoin useiden puhujien lomakekyselyssä ilmaisemia käsityksiä. Tämän ryhmän puhujat keskustelevat äänitetyissä keskusteluissa melko paljon institutionaalisista referenteistä (172 esiintymää). Viittauksista näihin henkilöihin 80 % tehdään *se*-pronominilla, kun keskimäärin *se*-pronominin osuus on 64 %. Vaikuttaa kuitenkin siltä, että referentin esiintyminen institutionaalisessa roolissa ei ole pronominin valinnan kannalta niin olennaista kuin se, miten referenttiin kyseisessä vuorovaikutustilanteessa suhtaudutaan. Lappalaisen (2010a) aineistossa keskustellaan esimerkiksi tutusta pastorista positiiviseen sävyyn ja tähän viitataan usein *hän*-pronominilla. Myös Siitosen (2008) aineistossa *hän*-viittauksiin liittyy referenttiä arvostava sävy. Tämän aineiston keskusteluissa lääkäreistä, opettajista ja pankkivirkailijoista puhutaan harvoin positiiviseen sävyyn, mikä voi selittää suurta *se*-pronominin osuutta. Toinen selitys on, että institutionaaliset referentit esiintyvät erityyppisissä kertomuksissa kuin vaikkapa puhujan perheenjäsenet tai että heidät esitellään kertomuksissa eri tavalla (esim. erisnimi vs. yleisnimi). Aineiston kertomusten tarkka analysointi vaatisi kuitenkin oman laadullisen tutkimuksensa.

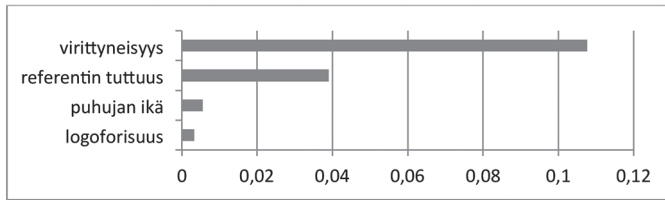
5.4. Ryhmä 5. Yli puolet *hän*-pronominia (56 %)

Viimeisen ryhmän muodostavat neljä puhujaa – kaksi keski-ikäistä ja kaksi vanhaa naista, jotka käyttävät *hän*-pronominia tässä aineistossa eniten. Heillä kaikilla yli puolet henkilöihin kohdistuvista *hän*- ja *se*-viittauksista on *hän*-pronomeineja. Yhteensä tässä ryhmässä tarkasteltavia pronominesiintymiä on 264.

Aiemmissa helsinkiläisiä ja Helsinkiin muuttaneita koskevissa tutkimuksissa (Lappalainen 2010a, 297; Nuolijärvi 1986, 182; Paunonen 1995, 173) mukana on sellaisia puhujia, joilla *hän* on yksinomainen viittaus henkilöreferentteihin yksikössä. Verrattuna näihin puhujiin kaakkoissatakuntalaisista nekin, joilla *hän*-pronominia esiintyy eniten, käyttävät sitä melko vähän. Tarkasteltavassa aineistossa eniten *hän*-pronominia käyttävän puheessa sitä on 61 % *hän*- ja *se*-viittausten yhteenlasketusta määrästä. Tällä henkilöllä on sikäli poikkeuksellinen tausta, että hän ei ole syntyperäinen satakuntalainen vaan on muuttanut Harjavaltaan alle kouluikäisenä evakkona. On siis mahdollista, että *hän*-pronominin suosimisella on yhteys hänen taustaansa, sillä kaakkoismurteissa on vanhastaan käytetty *hän*-pronominia jopa enemmän kuin lounaismurteissa (Vilppula 1989, 389). Paljon *hän*-pronominia käyttävien puhujien pronominvaihtelua selittävät tekijät on esitetty kuviossa 6.

Kuviosta 6 nähdään, että näillä puhujilla vaihteluun vaikuttavat tekijät ovat jälleen erilaisia kuin aiemmissa ryhmissä. Vähemmän *hän*-pronominia käyttävissä ryhmissä tärkein tekijä, logoforisuus, on tässä osa-aineistossa vain juuri ja juuri merkitsevä. Näillä

Kuvio 6. Paljon *hän*-pronominia käyttävien puhujien *hän* ~ *se* -vaihteluun vaikuttavien tekijöiden tärkeysjärjestys (C = 0,85).



puhujilla tärkein selittäjä on virittyneisyys, siis sama kuin melko paljon *hän*-pronominia käyttävillä (ryhmä 4). Vaikka *hän*- ja *se*-pronomeja on puhujilla suunnilleen yhtä paljon, ne eivät esiinny keskustelussa tasaisesti vuorotellen, vaan kumpikin pronomini ryhmittyy tiettyihin konteksteihin. Kielenulkoisista tekijöistä tärkein on referentin tuttuus: *hän*-pronomeista 49 % viittaa puhujalle läheiseen henkilöön, *se*-pronomeista vain 23 %. Paljon *hän*-pronominia käyttävillä vanhoilla *hän*-viittauksia on vielä enemmän (57 %) kuin saman ryhmän keski-ikäisillä (52 %).

Paunosen (1995) ja Nuolijärven (1986, 187) havainnoissa helsinkiläisten ja Helsinkiin muuttaneiden kielestä *hän*-pronominin runsas käyttö kytkeytyy yleiskielisyyteen. Tässä kaakkoissatakuntalaisessa aineistossa *hän*-pronominin suosijat eivät kuitenkaan vaikuta muilta kielenpiirteiltään yleiskielisiltä. Heillä esiintyy runsaasti alueen murteelle tyypillisiä äänne- ja muotopiirteitä, kuten loppuheittoa (*ain, häne*), inessiivin päätteitä *-s* ja *sA* (*tosa Peräkallios*), yleiskielen *ts:n* vastinetta *tt* (*mettä, mettän ~ metän*) diftongien avartumisia (*viäl, tual*), puolipitkiä toisen tavun vokaaleja (*varo:tteli, tosa:*) ja yleisgminaatiota (*mennee, pimmees*) (piirteistä ks. Kettunen 1940; Rapola 1990). Erään paljon *hän*-pronominia käyttävän henkilön puhetta kuvastava esimerkki 7 on osa kertomussikermää, jonka aiheena ovat kohtaamiset hirvien ja peurojen kanssa.

(7) Siäll_oli hirvii (Sapu-211)

```

01 Maire: määki ain Riinaa varo:ttelij ja pelo:ttelin ko (.) häne
02      (.)likkakaveris asu: tosa Peräkallios.
03 N:    .(mhh)
04      (0.9)
05 Maire: ja sillon ku här rupes_ajamaa skootteril ja (1.7) no sen
06      nyv viäl saa pysäytettyyki. mut sit ku auton kans ni (.)
07      pimmees liikuttiin [ni
08 N:    [.jhaa
09 Maire: mää sit <@koit nyt> käsittää@ ku hänki ajaa >sillaika< HIUKAN
10      SILLAI R:EIPPAASTI,
11 N:    nii::
12 Maire: .hh et siin_ei ihan ain välttämät sata:ner riitä'. (1.5)
13      et koit ↑nyt käsi:ttää. et siäl ihan_oikeest voi ol niit
14      hirvii.
15 N:    jooh
16 Maire: sit hän_oli kerran nähny'.

```

Katkelman 7 kertomuksessa köyliöläinen Maire viittaa tyttärensä Riinaan toistuvasti *hän*-pronominilla, mutta katkelma on muuten hyvin puhekielinen. Katkelmassa riveillä

1, 12 ja 13 Mairella esiintyy lounaimurteisia puolipitkiä vokaaleita, rivillä 7 yleisgemiinaatio ja riveillä 6 ja 13 avartuneet diftongit. Rivillä 2 hän käyttää Köyliön seudun murteeseen vanhastaan kuulunutta (ks. Kettunen 1940, 119) kolmannen persoonan possessiivisuffiksia (*liikkakaveris*). Lisäksi läpi katkelman esiintyy loppuheittoisia muotoja. Rungas *hän*-pronominin käyttö ei siis tarkoita, että puhujat samalla käyttäisivät muita yleiskielen normien mukaisia muotoja tai pelkästään laajalevikkisiä puhekielisyyskäytöksiä. Voidaankin ajatella, että tämän aineiston lounaisissa välimurteissa, joissa *hän* on vanhastaan ollut melko yleinen (esim. Vilppula 1989, 389), se on yhä arkikieleen sopiva viittauskeino ainakin vanhoilla ja osalla keski-ikäisistä.

Ryhmän puhujista lomakekyselyyn ovat vastanneet edellä mainittu Aila sekä hänen tyttärensä, keski-ikäinen Arja. Heidän käsityksensä *hän*-pronominin käyttökonteksteista ovat samanlaiset kuin muillakin vastaajilla: *hän* viittaa heidän mielestään vieraisiin, vanhempiin ja arvostettuihin henkilöihin sekä pikkulapsiin. Arja mainitsee institutionaaliset henkilöt, Aila taas ”asialliset” tilanteet mutta toisaalta myös läheiset ystävät. Kiinnostavaa on, että kumpikaan tämän ryhmän vastaajista ei tunnista mahdollisuutta käyttää *hän*-pronominia negatiivisissa yhteyksissä ironisesti, vaikka muista vanhoista ja keski-ikäisistä vastaajista usea sellaisen käytön tunnistaa.

6. Päätelmät

Olen tässä artikkelissa tarkastellut kielensisäisiä ja ulkoisia tekijöitä, joilla on yhteys siihen, milloin puheessa käytetään pronominia *hän* viittaamassa henkilöön.¹¹ Aiemmin henkilöviitteisen *hän* ~ *se* -vaihtelun määrällinen tarkastelu on keskittynyt pitkälti kielenulkoisten muuttujien kuten aluemurteen (esim. Siitonen 2008, 88–91; Vilppula 1989, 389), puhujan iän ja sukupuolen (esim. Paunonen 1995, 165–177) sekä puhetilanteen (esim. Lappalainen 2010a, 2010b; Mielikäinen 1982, 282; Nuolijärvi 1986, 290–300) tarkasteluun. Referentin ja puheen sävyn merkityksestä on esitetty laadullisia havaintoja (Kurki ym. 2011; Lappalainen 2010a; Mielikäinen 1982, 282; Siitonen 2008). Lisäksi on todettu puhutun kielen *hän*-pronominilla olevan erityistehtävä logoforisena pronomininä (Laitinen 2005; VISK, § 1469).

Olen pyrkinyt kuvaamaan tilastollista monimuuttujamenetelmää käyttäen, kuinka tärkeitä eri tekijät vaihtelun selittämisen kannalta ovat kaakkoisen Satakunnan alueelta kerätyssä keskusteluaineistossa. Tärkeimmiksi tekijöiksi nousevat puhutun kielen kielioppiin ja vuorovaikutukseen liittyvät ilmiöt: logoforisuus ja virittyneisyys. Pronominin *hän* esiintymistä ennustaa parhaiten se, että pronomini sijaitsee referoidussa lausumassa viittaamassa johtolauseessa mainittuun puhujaan. Prototyypin logoforisuus kytkeytyy

11 Tätä tutkimusta tehdessäni tein vastaavia satunnaismetsäanalyyskejä myös aineiston *tää*- ja *toi*-pronominista. Näin suppeasta aineistosta ei kuitenkaan löytynyt luotettavia todisteita siihen suuntaan, että näiden pronominin vaihtelu olisi kytköksissä kielenulkoisiin muuttujiin, vaan tulokset ainoastaan vahvistavat aiemmin esittämäni kytkökset ensimmäintoihin ja lohkeamarakenteisiin (Priiki 2015b).

vahvasti nimenomaan puheen referointiin; toisen henkilön ajatusten esittäminen vastavalla rakenteella on aineistossa vähäistä.

Logoforisen tehtävän tärkeys korostuu niillä puhujilla, jotka käyttävät *hän*-pronomiinia hyvin vähän. Heillä ainoat *hän*-esiintymät ovat prototyypillisissä johtolauseellisissa puhereferaateissa. Tämä puoltaa johtolauseella tehtyjen referointien erottamista muusta, tulkinnanvaraisemmasta referoinnista: *hän*-pronominin asema on kaikkein vankin juuri prototyypillisissä referaateissa. Kuten aiemminkin on havaittu (Kurki ym. 2011, 97), *hän*-pronomiinia pelkästään prototyypillisissä referoinneissa käyttävät satakuntalaispuhujat ovat pääosin nuoria. Muunlaisissa referoinneissa *hän* voi nuorten kielessä nykyään saada ironisen tulkinnan (ks. Priiki 2014a). Osalla puhujista prototyypillisissäkin referoinneissa esiintyy toisinaan pronomini *se*.

Toisaalta logoforisuus menettää merkitystään *hän*-pronomiinia enemmän suosivilla puhujilla. Eniten *hän*-pronomiinia käyttävillä prototyypinen logoforisuus ei selitä pronominin käyttöä kovin hyvin, sillä näillä puhujilla suuri osa *hän*-pronomineista esiintyy muissa konteksteissa kuin johtolauseellisissa referoinneissa. Moni päälause-esiintymä kuitenkin selittyy logoforisen käytön laajentumisena referentin ajatusten ja tunteiden tulkintaan ja referentin näkökulman välittämiseen (ks. myös Lappalainen 2010b, 302–304). Logoforisuuden laajentuminen näkyy määrällisesti siinä, että aineistossa *hän*-pronominit ovat suhteellisesti yleisempiä ajattelua ja tunnetta ilmaisevien kuin muiden verbien subjekteina (Priiki 2014b). Kognitio- ja emootioverbit ovat kuitenkin niin harvinaisia, ettei ilmiö auta selittämään *hän* ~ *se* -vaihtelua koko aineiston tasolla. Referoinnin laajentumiseen onkin vaikea päästä käsiksi määrällisen analyysin keinoin. Lisää määrällistä tietoa voisi tuottaa, jos vastaavanlaisiin analyyseihin lisäisi muuttujaksi myös lauseen aikamuodon sekä pronominin sijainnin sivu- tai päälauseessa.

Toiseksi tärkeimmäksi tekijäksi aineiston *hän* ~ *se* -vaihtelun selittäjänä osoittautuu virittyneisyys. Virittyneisyys kuvaa siis sitä, että pronomineilla *se* ja *hän* on kummallakin taipumus ketjuuntua: jos henkilöön viitataan kerran pronominilla *hän*, on tavallista, että myös seuraava viittaus samaan henkilöön tehdään samalla pronominilla. Suomen henkilöviitteiset kolmaspersoonaiset pronominit toimivat siis tässä suhteessa samalla tavoin kuin vaikkapa espanjan ja englannin ensimmäisessä persoonassa vaihtelevat pronomini ja nolla (ks. esim. Torres-Cacoullous & Travis 2014; Travis 2007). Virittyneisyyttä ilmiönä on aiemmissa tutkimuksissa kuvattu tiedostamattomaksi ja automaattiseksi tavaksi (Bock & Griffin 2000, 177). Näin ollen ei yllätä, että kielenkäyttäjät eivät kuvaa sitä ääniteaineistoa täydentävän kyselyn vastauksissa. Voidaankin ajatella, että puhutun kielen kielioppiin liittyvät ilmiöt edustavat intuitiivista, nopeaa ja tiedostamatonta ajattelua, kun taas kielenkäyttäjien tiedostamat kohteliaisuuteen ja yleiskielen käytön vaatimukseen liittyvät ilmiöt edustavat keskittyneitä ajattelua (psykologiassa järjestelmät 1 ja 2, esim. Kahneman 2012, 30–34). Päinvastoin kuin logoforisuuden, virittyneisyyden merkitys kasvaa mitä enemmän puhujat käyttävät *hän*-pronomiinia.

Kielenulkoisista muuttujista parhaiten pronominin *hän* esiintymistä selittää yksilöllinen vaihtelu. Toiset tarkasteltavista puhujista siis suosivat pronomiinia *hän*, toiset *se*, eikä ero selity kovin hyvin puhujan ominaisuuksilla. Kun puhujat jaetaan ryhmiin sen

perusteella, kuinka paljon he käyttävät *hän*-pronominia, ikäryhmäjaon ääripäät, nuoret ja vanhat, kuitenkin eroavat. Puhujan sukupuoli ei tilastollisessa analyysissä ole merkitsevä tekijä, mutta ryhmäjaossa nuoret miehet sijoittuvat pääosin vähiten *hän*-pronominia käyttäviin (ryhmä 1), vanhat miehet taas melko paljon *hän*-pronominia käyttäviin (ryhmä 4). Miehillä pronominiin käytössä näyttäisi siis olevan vähemmän vaihtelua kuin naisilla.

Keski-ikäisiksi luokiteltuja puhujia kuuluu kaikkiin viiteen ryhmään. On mahdollista, että tarkasteltavassa kielimuodossa on käynnissä muutos, ja keski-ikäisistä osa on vanhojen puhujien tavoin lounaismurteisen *hän*-pronominia suosivan systeemin kannalla, osa taas nuorten tavoin rajoittaa *hän*-pronominin ainoastaan prototyyppiseen referointiin, kuten esimerkiksi hämäläismurteissa on vanhastaan tehty. Toisaalta Helsingin puhekielistä on aiemmin esitetty, että henkilöviitteinen *se* on nuorisokielen piirre ja että nuoret alkavat myöhemmin käyttää *hän*-pronominia (Lappalainen 2014). Muutoksen toteaminen joko idiolekteissa tai alueen murteessa vaatisi kuitenkin erilaisen aineiston.

Olen verrannut ääniteaineiston pronominivaihtelua puhujien omiin käsityksiin *hän* ~ *se* -vaihtelun syistä. Lomakekyselyn vastauksista nousee esiin käsitys *hän*-pronominista piirteenä, joka kuuluu asiallisiin tilanteisiin ja institutionaaliseen kielenkäyttöön. Puhujat ovat siis tietoisia normitetun kirjakielen ja puhutun kielen erosta. Kielensisäiset tekijät logoforisuus ja virittyneisyys taas ovat tiedostamattomia. Keskustelun rakentumiseen ja vuorovaikutukseen liittyvistä ilmiöistä tunnistetaan kuitenkin *hän*-pronominin ironinen käyttö, jonka moni vastaaja mainitsee (ks. Priiki 2014a). Vaikka puhujien käsityksissä *hän* on normitetun yleiskielen piirre, ääniteaineiston perusteella sitä voidaan vielä nykypäivänäkkin pitää lounaiseen välimurteeseen kuuluvana. Nimittäin puhujat, jotka käyttävät paljon *hän*-pronominia, eivät tässä aineistossa ole muilta kielenpiirteiltään yleiskielisiä. Nuorten *se*-pronominin suosiminen sen sijaan voi olla pyrkimystä muissa murteissa ja epävirallisessa yleispuhekielessä tavallisemman järjestelmän suuntaan.

Myös sen vaikutusta, millaisesta referentistä puhutaan, pohtivat lomakekyselyssä sekä nuoret että vanhat vastaajat. Ääniteaineistossa referentin ominaisuuksien vaikutus tulee näkyviin vain niillä keski-ikäisillä tai vanhoilla puhujilla, jotka käyttävät *hän*-pronominia paljon. Kielenkäyttäjien mielestä pronominia *hän* käytettäisiin vieraista ja institutionaalisisessa roolissa esitettävistä henkilöistä. Tämän aineiston keskusteluissa kumpiinkin viittaa kuitenkin muita henkilöitä useammin pronomini *se*. Puheenaiheena olevan henkilön ominaispiirteitä enemmän siihen, mitä pronominia hänestä käytetään, vaikuttavat puhujille tiedostamattomat, keskustelun rakenteeseen liittyvät säännöt.

Puhujalle tutut henkilöt esiintyvät kertomuksissa, joissa tapahtumia kerrotaan päähenkilön näkökulmaan eläytyen ja joissa tästä syystä esiintyy sekä prototyyppisesti että laajemmin käsitettynä logoforisia *hän*-pronomineja. *Hän*-pronomini kytkeytyy myös kerronnan strategioihin. Läheisistä kerrotaan keskustelussa asioita, jotka referentti on mitä ilmeisimmin kertonut puhujalle joskus aiemmin, jolloin *hän*-viittaus voi tuoda kertomukseen toisen käden tiedon sävyä. Tämä on päinvastainen havainto Vilppulan (1989, 392) esittämään verrattuna, sillä Vilppulan mukaan *hän*-viittauksia esiintyisi nimenomaan puhujan kertoessa itse todistamistaan tapahtumista. Monet niin tämän aineiston kuin Vilppulankin (1989) esimerkeistä selittyvät nähdäkseni näkökulman vaihtelun ajatuksella:

kun pronomininä on *hän*, näkökulma on referentin, ei puhujan itsensä. Aiemmissä tutkimuksissa tehdyt osin ristiriitaiset havainnot *hän*-pronominin referentin etäisyydestä tai läheisyydestä voisivat selittyä tätä kautta. Esimerkiksi Kurjen ym. (2011, 98) ”ajallisesti ja joskus myös sosiaalisesti etäisiin” henkilöihin viittaavat *hän*-pronominit esiintyvät ehkä sellaisten asioiden yhteydessä, joista puhujalla ei ole omakohtaisia havaintoja.

Tulokset täydentävät aiemmin tehtyjä havaintoja *hän*-pronominin esiintymisestä. Kielensisäisten muuttujien tärkeys korostaa sitä, että myös sosiolingvistiksestä näkökulmasta tehtävässä tutkimuksessa on olennaista ottaa huomioon puhutun kielen kielioppiin ja vuorovaikutuksen rakentumiseen liittyvät ilmiöt. Kansanlingvistisen kyselyaineiston vertailu äänitettyihin keskusteluihin osoittaa, että kielenkäyttäjien havainnot vastaavat aineiston todellista tilannetta heikosti.

Taustaoletuksiltaan salliva monimuuttujamenetelmä satunnaismetsä palvelee hyvin *hän* ~ *se* -vaihtelun kaltaisen ilmiön tarkastelussa, kun mahdollisia tärkeitä selittäjiä on useita. Yksilöllisen vaihtelun tarkastelu muuttujana paljastaa, miten puhujat jakautuvat, kun ryhmät muodostetaan todellisen käytön perusteella eikä tarkastella ennalta asetettuja luokkia. Satunnaismetsämenetelmän piirre on, että se kuvaa tekijöiden tärkeyttä koko aineiston tasolla. Haittapuolena on, että harvinaiset mutta tietyssä kontekstissa merkitsevät ilmiöt eivät satunnaismetsän tuloksissa näy tärkeinä. Tämän artikkelin tuloksissa tämä ilmenee siinä, että aiemmin samassa aineistossa merkitseväksi osoittautunut mutta frekvenssiltään harvinainen ilmiö, *hän*-pronominin esiintyminen mielenliikkeitä ilmaisevan verbin subjektina (ks. Priiki 2014b), näyttäytyy tärkeänä selittäjänä vain yhdessä puhujaryhmässä. Tämä ei kuitenkaan kumoa aiempia tuloksia (Priiki 2014b), vaan kuvastaa vain sitä, että vaihtelun kokonaiskuvassa ilmiö jää sellaisten tekijöiden varjoon, joilla pystytään selittämään mahdollisimman suuri osa aineiston pronomini esiintymistä.

Satunnaismetsämallien selitysvaimoista kertovat C-arvot vaihtelevat tutkimuksessa esitellyissä analyyseissä. Vähän *hän*-pronominia käyttävien puhujien tarkastelussa korkeat C-arvot ilmaisevat, että heidän puheessaan tavattavasta *hän* ~ *se* -vaihtelusta pystytään tilastolliseen analyysiin soveltuvilla tekijöillä selittämään hyvin suuri osa. Kun *hän*-pronominin määrä kasvaa, lisääntyy myös sellaisten esiintymien määrä, joita tässä tarkasteltujen tekijöiden avulla ei voi selittää. Kyseessä ovat vuorovaikutukseen ja kerronnan dynamiikkaan kytkeytyvät ilmiöt, joiden luokittelu tilastollista analyysiä varten on vaikeaa. Tilastollinen tarkastelu kaipaakin tuekseen laadullista analyysiä.

Lähteet

Aineslähteet

Lomakekyselyn vastaukset, tekijän hallussa

Satakuntalaisuus puheessa (Sapu) hankkeen äänitteet 154, 155, 157, 159, 161, 163, 164, 165, 167, 187, 188, 190, 202, 209, 210, 211, 222, 223 ja 224. Lauseopin arkiston kokoelmat. Turun yliopiston kieli- ja käänntieteen laitos, suomen kieli ja suomalais-ugrilainen kielentutkimus.

Tutkimuskirjallisuus

- BAAYEN, R. HARALD 2008: *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge University Press.
- BAAYEN, R. H. – JANDA, L. A. – NESSET, T. – DICKEY, S. – ENDRESEN, A. – MAKAROVA, A. 2013: Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37, 253–291.
- BOCK, KATHRYN – GRIFFIN, ZENZI 2000: The persistence of structural priming: transient activation or implicit learning. *Journal of Experimental Psychology: General* 129 (2), 177–192.
- BREIMAN, LEO 2001. Random forests. *Machine Learning* 45, 5–32.
- CUTLER, RICHARD D. – EDWARDS JR. – THOMAS C. – BEARD, KAREN H. – CUTLER, ADELE – HESS, KYLE T. – GIBSON, JACOB – LAWLER, JOSHUA J. 2007: Random forests for classification in ecology. *Ecology* 88, 2783–2792.
- GUMPERZ, JOHN 2001: Interactional sociolinguistics. A personal perspective. DEBORAH SCHIFFRIN, DEBORAH TANNEN ja HEIDI HAMILTON (toim.): *The handbook of discourse analysis*, 215–228. Blackwell, Oxford.
- IVASKA, ILMARI 2014: Edistyneen oppijansuomen avainrakenteita. Korpusnäkökulma kahden kielimuodon tyypillisiin rakenteellisiin eroihin. *Virittäjä* 118, 161–193.
- 2015: Longitudinal changes in academic Learner Finnish. A key structure analysis. *International Journal of Learner Corpus Research* 1(2), 210–241.
- KAHNEMAN, DANIEL 2012: *Ajattelu, nopeasti ja hitaasti*. Suom. Kimmo Pietiläinen. Terra Cognita, Helsinki.
- KALLIO, JUSSI 1978: Kolmannen persoonan dilemmat. *Sananjalka* 20, 52–74.
- KETTUNEN, LAURI 1940: *Suomen murteet III A. Murrekartasto*. SKST 188. SKS, Helsinki.
- KLAVAN, JANE – PILVIK, MAARJA-LIISA – UIBOAED, KRISTEL 2015: The Use of Multivariate Statistical Classification Models for Predicting Constructional Choice in Spoken, Non-Standard Varieties of Estonian. *SKY Journal of Linguistics* 28, 187–224.
- KURKI, TOMMI – SIITONEN, KIRSTI 2009: Satakuntalaisuus puheessa. *Sananjalka* 51, 158–166.
- KURKI, TOMMI – SIITONEN, KIRSTI – VÄÄNÄNEN, MIJJA – IVASKA, ILMARI – EKBERG, JARI 2011: Ensi havainnot Satakuntalaisuus puheessa hankkeesta. *Sananjalka* 53, 83–108.
- LAITINEN, LEA 2005: *Hän*, the third speech act pronoun in Finnish. RITVA LAURY (toim.): *Minimal reference: the use of pronouns in Finnish and Estonian discourse*, 75–106. Studia Fennica Linguistica. SKS, Helsinki.
- LAPPALAINEN, HANNA 2010a: *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. SKS, Helsinki.
- 2010b: *Se ja hän* puhutussa kielessä. *Kielikello* 4/2010, 4–7. Kotimaisten kielten tutkimuskeskus, Helsinki.
- 2014: Idiomatic and generational changes in the light of real time data. The case study of 3rd person pronouns. Esitelmä konferenssissa Sociolinguistic Symposium 20, Jyväskylä 15.–18.6.2014.
- LOHMANN, ARNE 2013. Is tree hugging the way to go? Classification trees and random forests in linguistic study. *Vienna English Working Papers* 22.
- NUOLIJÄRVI, PIIRKKO 1986: *Kolmannen sukupolven kieli. Helsinkiin muuttaneiden suurten ikäluokkien eteläpohjalaisten ja pohjoissavolaisten kielellinen sopeutuminen*. SKS, Helsinki.
- MIELIKÄINEN, AILA 1982: Nykypuhe-suomen alueellista taustaa. *Virittäjä* 86, 277–294.
- 1988: Naiset puhekielen säilyttäjinä ja uudistajina. LEA LAITINEN (toim.): *Isosuinen nainen. Tutkielmia naisesta ja kielestä*, 82–111. Yliopistopaino, Helsinki.
- PAUNONEN, HEIKKI 1995: *Suomen kieli Helsingissä. Huomioita Helsingin puhekielen historiallisesta taustasta ja nykyvariaatiosta*. Helsingin yliopiston suomen kielen laitos.
- PRIIKI, KATRI 2011: *Läsnäolo, tuttuus, tulkinta ja sävy. Ihmisviitteinen hän ja sen tehtävät Harjavallan, Kokemäen ja Huittisten nykypuhekielessä*. Pro gradu -tutkielma. Turun yliopiston kieli- ja kääns-

nöstieteiliden laitos.

- PRIIKI, KATRI 2012: Läsnaölijaan viittaava *hän* Harjavallan, Kokemäen ja Huittisten nykypuhekielessä. *Sananjalka* 54, 30–49.
- 2014a: Antagonistin äänellä. *Hän*-pronominin käyttö valituskertomuksessa. *Virittäjä* 118, 194–220.
- 2014b: Kaakkois-Satakunnan henkilöviitteiset *hän, se, tää* ja *toi* subjekteina. *Sananjalka* 56, 86–107.
- 2015a: Korjauksia ja kiusoitteluja. Läsnaölijaan viittaavat *hän, tää* ja *toi* puhumista ilmaisevien verbien subjekteina. *Sananjalka* 57, 29–53.
- 2015b: *Se oli iha hullu se hammaslääkäri*. Kaakkois-Satakunnan henkilöviitteiset *se, hän, tää* ja *toi* eteenpäin lohkeavan konstruktion osina ja ensimmäintoina. *Puhe ja kieli* 35:2, 47–71.
- 2016: Henkilöviitteisten kolmaspersoonaisen pronomien vaihtelu keskustelukumppanin mukaan. *Puhe ja kieli* 36:2, 123–144.
- tulossa: The Finnish logophoric pronoun *hän* – a quantitative approach. *Journal of Estonian and Finno-Ugric Linguistics*.
- tekeillä: *Hän, se, tää* vai *toi*? Vuorovaikutussosiolinguvistinen tutkimus henkilöviittauksista Kaakkois-Satakunnan nykypuhekielessä. Väitöskirja. Turun yliopiston kieli- ja käännöstieteiliden laitos.
- RAPOLA, MARTTI 1994: *Johdatus suomen murteisiin*. Tietolipas 4. SKS, Helsinki.
- RAMPTON, BEN 2006: *Language in late modernity. Interaction in an urban school*. Cambridge University Press.
- SKS = Suomalaisen Kirjallisuuden Seura.
- SKST = Suomalaisen Kirjallisuuden Seuran Toimituksia.
- SMS 2012 = *Suomen murteiden sanakirja*. Kotimaisten kielten keskuksen verkkojulkaisuja 30. Kotimaisten kielten keskus. <http://kaino.kotus.fi/sms>. Päivitetty 29.2.2016.
- SIITONEN, KIRSTI 2008: Sävypronomini *hän*. *Sananjalka* 50, 87–109.
- STROBL, C. – MALLEY, J. – TUTZ, G. 2009: An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods* 14(4), 323–348.
- TAGLIAMONTE, SALI A. – BAAYEN, R. HARALD 2012: Models, forests and trees of York English. *Language variation and change* 24(2), 135–178.
- Tieteen kansallinen termipankki. Helsingin yliopiston suomen kielen, suomalais-ugrialaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos. <http://tieteentermipankki.fi/>
- TOIVOLA-JUNTILA, LEILA 1966: Havainoja *hän*-pronominin käytöstä suomen läsimurteissa. Pro gradu työ. Helsingin yliopiston suomen kielen laitos.
- TORRES CACULLOS, RENA – TRAVIS, CATHERINE 2014: Prosody, priming and particular constructions. The patterning of English first-person singular subject expression in conversation. *Journal of Pragmatics* 63, 19–34.
- TRAVIS, CATHERINE 2007: Genre effects on subject expression in Spanish: priming in narrative and conversation. *Language Variation and Change* 19 (2), 101–135.
- VILPPULA, MATTI 1989: Havainoja *hän-* ja *he*-pronominien käytöstä suomen murteissa. *Virittäjä* 93, 389–400.
- VIRTARANTA, PERTTI – JAAKKO YLI-PAAVOLA (toim.) 1980: *Suomen murteet. Näytteitä uralilaisista kielistä I*. Suomi 119: 1. SKS, Helsinki.
- VISK 2005 = *Iso suomen kielioppi*. AULI HAKULINEN, MARIA VILKUNA, RIITTA KORHONEN, VESA KOIVISTO, TARJA RIITTA HEINONEN ja IRJA ALHO (toim.). SKS, Helsinki. <http://scripta.kotus.fi/visk>
- VÄÄNÄNEN, MILJA 2015: Subjektin ilmaiseminen persoonapäätteellä ja pronomilla. Yksikön 1. persoonan muotojen käyttö lounaimurteissa, lounaisissa siirtymämurteissa ja kaakkoismurteissa. *Sananjalka* 57, 7–28.
- YLIKAHRI, KRISTIINA 1996: Referoinnin laajentumat *hän-*, *he-* ja *se-*, *ne*-pronominien käytössä Siikaisten murteissa. *Virittäjä* 100, 182–203.

KATRI PRIIKI: Grammatical rules and individual variation: A quantitative study of the pronoun hän in contemporary colloquial speech in Satakunta

This article examines variation in Finnish third-person pronouns referring to a person. The personal pronoun *hän* is used in written, normative language, but in colloquial speech, speakers most commonly use the demonstrative pronoun *se*. In spoken language, *hän* has a logophoric function: it occurs in indirect quotes to refer to the original speaker.

The variation between *hän* and *se* has been studied quite widely from the viewpoint of sociolinguistics. Previous studies have shown that older people use *hän* more than younger people, women more than men, and the highly educated more than the less educated. *Hän* is more likely to be used to refer to old, respected, and distantly known people, and perhaps institutional actors.

The data of this study are contemporary conversations from a dialect in which *hän* has been moderately frequent. The study considers whether social, grammatical or interactional factors are best for explaining the variation between *hän* and *se*. The data is analyzed using random forests, a relatively new multivariate statistical method. The results show that the most important factors are the logophoric function and priming, i.e. which variant precedes the reference. When considering social factors, the differences between individual speakers explain variation better than the factors reported in previous sociolinguistic studies.

The variation between individual speakers is examined by dividing the speakers into five groups based on the frequency of the use of the pronoun *hän*; the differences between these five groups are statistically significant. Those who use *hän* very little (mostly young people), use it only in logophoric contexts – some not even there. In the speech of those who use *hän* a lot (mostly older people), the importance of other factors increases. The speakers themselves believe that their own usage of *hän* follows guidelines inspired by the rules of normative language: they do not recognize the effect of logophora or priming.

Katri Priiki
katri.priiki@utu.fi
Suomen kieli ja suomalais-ugrilainen kiellentutkimus
20014 Turun yliopisto