



# Katomuotojen eteneminen yleiskielen *hd*-yhtymässä Helsingin puhekielessä

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

## 1 Johdanto

Suomen kielen variaation tutkimuksessa yksi keskeisimmistä tarkasteltavista piirteistä on perinteisesti ollut *t:n* heikon asteen eli yleiskielen *d:n* edustuksen tarkastelu. Tämän piirteen varianteina on murteissamme kuultu niin *r*, *ð*, *l*, *j*, *v*, *h* kuin katokin. Näiden lisäksi myöhäiskantasuomesta periytynyttä soinnillista dentaalispiranttia *ð* on kirjattu vielä 1800- ja 1900-lukujen vaihteessa syntyneillä lounaismurteisilla puhujilla. (Kettunen 1930: 104; Virtaranta 1958: 12–13.)

Katomuodot (ja siirtymä-äänteet *j*, *v* ja *h*) ovat tavallisia itämurteissa, kun taas länsimurteissa on suosittu tremulanttiäänteitä (*r* ja *ð*). Nykyisin länsimurteiden alueella tyypillisin variantti on kuitenkin yleiskielinen *d*. Tyypillisesti yleiskielen *d:n* vastineita on tarkasteltu lyhyen vokaalin jälkeen (VdV), pitkän vokaaliaineksen jälkeen (VVdV) ja *h:n* jälkeen (hd; luokittelusta esim. Paunonen 1995 [1982]: 31–33; Mustanoja 2011: 307–311).

Vaikka katoedustus on tulkittu itämurteiseksi, on länsimurteissa kaikissa äänneympäristöissä taajaan esiintyviä lekseemejä, joissa katoedustus on tyypillinen. Esimerkiksi *ees* (VdV), *meiän* (VVdV) ja *kaheksan* (hd) ovat kaikki yleisiä länsimurteiden alueella. Mantila (2004) on luokitellut katoedustuksen yleistäväksi puhekielisyydeksi erityisesti *h:n* jäljessä. Tutkimuksissa on havaittu useita lekseemejä, joissa kato on tässä äänneasemassa tavallinen (esim. Nuolijärvi 1986: 112; Mustanoja 2011: 322, 335). Kato on ollut länsimurteissa tyypillinen lukusanoissa (*kaheksan*, *yheksän*, *kahe-*, *yhe-*) jo 1900-luvun alkupuolen murreaineistoissa (Rapola 1966: 112). Helsingissä on yleisesti katoa lisäksi keskeisissä verbeissä (*tehä*, *nähä*, *lähe-*, *ehi*, *taho*) sekä sanoissa *mahollinen*, *mahoton*, *mahollisuus* ja *kohalla* (Paunonen 1995 [1982]: 45). Helsinkiin muuttaneiden pohjalaisten katoedustus ei juuri poikkea Paunosen esittämästä listasta, mihin perustuen Nuolijärvi (1986: 112, 114) esittääkin länsimurteisen kadon leviämisen jähmettyneen näihin yleisiin sanoihin.

Käsillä oleva tutkimus lähtee liikkeelle seuraavasta kysymyksestä: onko *hd*-yhtymässä tavattava kato aidosti jähmettynyt vai eteneekö se yhä? Muutosta tarkastellaan Helsingin puhekielestä kolmen aikapisteen reaaliaikaisen aineiston avulla (Helpuhe 2014 -aineisto). Tutkimushypoteesimme on, että kato ei ole pysähtynyt tiettyihin lekseemeihin, vaan se etenee diffuusiomallin (Wang 1969; Mielikäinen 1995) mukaisesti

hitaasti leksikon ja morfologisten muotoryhmien sisällä. Tätä kehitystä mallinimme tilastollisesti korrelaatioanalyysin avulla. Lisäksi oletamme, että niissä lekseemeissä, joissa katoa on jo aiemmin esiintynyt, katoesiintymien osuus on entisestään kasvanut.<sup>1</sup>

## 2 Taustaa

Tässä luvussa esittelemme tutkimuksen taustan. Alaluvussa 2.1 tarkastelemme yleiskielen *d*:n vastineita erityisesti Helsingin puhekielessä. Alaluvussa 2.2 esittelemme tutkimuksen teoreettisen kehityksen eli diffuusiomallin.

### 2.1 Yleiskielen *d*:n vastineet

Yleiskielen *d*:n vastineiden runsaus murteissa perustuu historiallisiin kehityskulkuihin *t*:n astevaihtelussa. Soinnillinen dentaalispirantti  $\delta$ , jonka jatkajia kaikki nykyiset variantit ovat, kehittyi keskikantasuomalaisella kaudella *t*:n heikon asteen allofoniksi vokaalien, likvidojen ja *h*:n jäljessä (Lehtinen 2007: 144–145). Astevaihtelu perustui siis aluksi puhtaasti äänne- ja painoasemaan, mutta varhaisuuden kaudella  $\delta$  kehittyi omaksi foneemikseen. Spirantti on kuitenkin ajan myötä saanut murteissa uusia vastineita tai kadonnut kokonaan. Itämurteissa tyypillinen kato ajoitetaan jo muinaiskarjalan kaudelle, kun taas läntisistä varianteista varhaisimpina pidetään hämäläismurteiden jo hiipunutta *l*:ää (Rapola 1966: 99–110; ks. myös Virtaranta 1958: 90, 214–216; kritiikistä Kallio 2013). Länsimurteissa tavallinen *r* (ja yksitäreiseksi tulkittu vastine  $\delta$ ; näiden problematiikasta ks. Mustanoja & O’Dell 2007) on kehittynyt vasta myöhemmin, ja sen katsotaan levinneen Länsi-Suomen rannikoilta laajemmalle sisämaahan (Virtaranta 1958: 213–216). Nykyinen yleiskielen *d* ei perustu murteisiin, vaan se on syntynyt alun perin kirjakielessä ruotsin ortografiaan sopeutumisesta ja siirtynyt myöhemmin puhuttuun kieleen (Lehtinen 2007: 178–181).

Helsingin puhekielessä on tavattu läntisiä tremulantti-variantteja (*r*,  $\delta$ ) ja katoa määrällekseemeissä. Koska yleiskielen vaikutus Helsingin puhekieleen on ollut muutenkin suuri (Paunonen 1995 [1982], 2006), myös yleiskielinen *d* on saanut jo 1800-luvun lopulla merkittävän aseman. Täysin kivutonta *d*:n vakiintuminen ei ole ollut edes yleiskieliseen puheeseen, vaan sen ”poisheittäminen” on ehdotettu jo Kotikielen Seuran alkuaikoina (Paunonen 1976: 338). Osaltaan *d*:n vakiintumiseen osaksi Helsingin puhekieltä on vaikuttanut sen tavallisuus slangisanoissa (esim. *stadi*). Paunonen (1995 [1982]: 33) toteaa, että Helsingin puhekielessä vallitsee yleiskielisen *d*:n ja (itämurteisen) kadon välinen variaatiotilanne.

*hd*-yhtymän toinen elementti, *h*, on suomen kielessä verraten nuori konsonantti. Varhaiskantasuomessa *h*:ta ei käytetty, vaan se on kehittynyt nykyisiin asemiinsa eri-

---

1. Ensimmäistä kirjoittajaa lukuun ottamatta kirjoittajien nimien järjestys on arvottu. Tutkimuksen alkuperäisidea on Leinin, aineiston muuttamisesta analysoitavaan muotoon on vastannut Kuparinen, tilastollisista menetelmistä Peltonen, Kuparinen, Leino ja Santaharju ja lingvistisestä analyysistä Kuparinen, Mustanoja ja Leino. Käsikirjoitus on valmistettu yhdessä. Artikkelin kirjoittaminen osana Koneen Säätiön rahoittamaa hanketta.

laisten äänne muutosten seurauksena keskikantasuomalaisena aikana. Tässä tutkimuksessa tarkasteltava yleiskielen *hd*-yhtymä on pääosin syntynyt *kt* > *ht* -muutoksen seurauksena (esim. *kahtena* < \**kaktena*), mutta jotkin tapaukset ovat syntyneet myös *pt* > *ht* -muutoksessa (*vyyhti* << \**vipte*). Myöhemmin *ht* on mukautunut astevaihteluun (\**kahden* < \**kahten*). (Rapola 1966: 245.)

Nyky-suomessa yleiskielen *d*:n katoesiintymät vaikuttavat tarttuvan herkimmin juuri *h*:n jälkeiseen asemaan. Paunonen (1995 [1982]: 32) raportoi Helsingin puhekielen 1970-luvun aineiston *hd*-yhtymistä katoprosentiksi 29,1, kun se oli pitkän vokaali-aineuksen jäljessä 19,2 ja lyhyen vokaalin jäljessä vain 1,4. Vastaavasti Nuolijärven tutkimuksessa (1986: 112) savolais- ja pohjalaismuuttajien puheessa katoedustus oli selvästi suurempi *hd*-yhtymässä kuin muissa asemissa. Niin ikään itämurteissa katomuodot näyttävät säilyvän parhaiten juuri *h*:n jälkeisessä asemassa (esim. Ikäheimonen 2012: 66). Kadon esiintyminen länsimurteissa ylläpitää siis osaltaan kadon säilymistä itämurteissa (Räsänen & Palander 2015: 16).

Yksi mahdollinen syy katoedustuksen hakeutumiselle juuri *h*:n jälkeiseen asemaan on se, että tässä asemassa *d*:n kato ei aiheuta hankalia vokaaliyhtymiä, joita joudutaan korjaamaan uusilla siirtymä-äänneillä (esim. *meijän*, *saaha*). Helsingin puhekielessä (ja länsimurteissa yleisesti) kato on keskittynyt määrälekkeseemihin. Sen sijaan tremulanttiääntämys voi esiintyä missä tahansa leksemissä *h*:n jälkeen niin vanhoilla työläispuhujilla (Paunonen 1995 [1982]: 45) kuin Helsinkiin muuttaneilla pohjalaisilla (Nuolijärvi 1986: 112).

## 2.2 Diffuusiomallit

Sanoista tai muotoryhmistä toisiin hitaasti etenevää muutosta on kuvattu diffuusion käsitteellä. Alun perin kiinan kielen tutkimukseen kehitetty leksikaalisen diffuusion malli (Wang 1969) esitti äänne muutosten etenevän sanasta sanaan kaikissa sanoissa yhtä aikaa tapahtuvan äänneläillisen muutoksen sijaan. Myöhemmin on esitetty kaikkien äänne muutosten etenevän diffuusiomallin mukaisesti, vaikka äänne muutos lopulta päätyisikin koskemaan koko leksikkoo (Oliveira 1991).

Koska leksikaalinen diffuusio on typologialtaan analyttiseen kiinan kieleen kehitetty malli, sen sovelluksissa muihin kieliluokkiin on tehty muutoksia. Esimerkiksi flekteeraavan englannin kielen tutkimuksissa on voitu ottaa mukaan myös morfeemien tarkastelua (Bybee 2002). Vielä enemmän diffuusiomallia pitää soveltaa suomen kaltaiseen agglutinoivaan kieleen. Fennistisessä tutkimuksessa käyttöön onkin vakiintunut termi morfologinen diffuusio (Suihkonen 1992: 304; Mielikäinen 1995).

Morfologisella diffuusiolla tarkoitetaan sellaisia muutoksia, jotka eivät etene samanaikaisesti esimerkiksi kaikkiin sijamuotoihin, vaan muutos etenee hitaasti ensin osaan muodoista ja vasta myöhemmin toisiin. On selvää, että esimerkiksi *a*:n ja *ä*:n loppuheitossa paikallissijoissa (*talos*, *talost*, *talol*, *talolt*) on eroa sijojen välillä loppuheiton toteutumisessa. Äänne muutos on siis morfologisesti ehdollista. Erityisen hyvin diffuusio näkyy suomen kielen hitaassa etenemisessä kohti vokaalivartalaisuutta (Paunonen 2003: 201–207). Konsonantivartaloista on ensin luovuttu essiivissä (*piennä* > *pienenä*, *lassa* > *lapsena*) ja vasta myöhemmin edetty kohti partitiivia (*pienä* > \**pieneä*,

*lasta* > \**lapsea*). Muutoksessa havaitaan muotoryhmittäisen morfologisen diffuusion ohella myös sanakohtaista, leksikaalista diffuusiota. Vokaalivartalaisuus on edennyt partitiiviin asti esimerkiksi sanoissa *tuomea* (< *tuonta*) ja *niemeä* (< *nientä*), mutta ei vielä sanassa \**sutea* (*sutta*).

Myös tässä tutkimuksessa liikutaan leksikaalisen ja morfologisen diffuusion raja- mailla. Päähuomio on leksikaalisessa muutoksessa (eli muutoksen etenemisessä esimerkiksi lukusanoista paikannimiin), mutta koska käsittelemme *t:n* heikon asteen vastineita, on selvää, että diffuusiosta on myös morfologinen elementti. Esimerkiksi sanojen välillä on eroa siinä, ovatko ne yksi- vai kaksivartaloisia, eli missä asemassa *t:n* heikko aste esiintyy. Tämän aineiston frekventeimmistä verbeistä on näin ollen selvä ero: *lähtää* esiintyy vain vokaalivartaloisena (*lähdetään*), kun taas verbillä *tehdä* kaikki heikon asteen tapaukset esiintyvät konsonanttivartalossa (*tehdään*). On mahdollista, että katotapausten leviämisen taustalla on myös taivutus- tai persoonamuodosta toiseen etenevää morfologista diffuusiota, mutta sen tarkastelu jää tämän tutkimuksen ulkopuolelle.

Suihkonen (1992) on tarkastellut klusiilien kadossa tapahtuvaa diffuusiota Kala- ja Lestijokilaaksossa. Hän perustaa päätelmänsä diffuusiojärjestyksestä varianttien levinneisyydelle ja yleisyysuhteille (mts. 303), kuten tekee myös Mielikäinen (1995: 326–328). Pitkittäisaineistojen avulla diffuusiota on kuitenkin mahdollista seurata myös ajassa. Tässä tutkimuksessa tarkastelemme yleiskielen *hd*-yhtymän katomuotojen etenemistä Helsingin puhekielessä 40 vuoden ajanjaksolla, kolmessa eri aikapisteessä. On korostettava, että diffuusio on ennen kaikkea teoreettinen malli ja kuvaustapa. Se ei siis anna lainkaan selitystä sille, miksi esimerkiksi yleiskielen *d* katoaa Helsingin puhekielessä tietyissä lekseemeissä ennen toisia (tai ylipäänsä). Näitä seikkoja on pohdittava muista lähtökohdista; esimerkiksi lekseemien tai muotoryhmien korkean frekvenssin on havaittu laukaisevan uudennoksia (esim. Lehtimäki 1983: 38; Nahkola 1987: 43), mutta toisaalta analogisesti etenevissä tapauksissa ne myös ylläpitävät vanhoja rakenteita kuten konsonanttivartaloita (Mielikäinen 1995: 324–325). Vastaavasti jotkin seikat voivat estää muutoksen etenemistä diffuusiomallin mukaisesti: muutos voi esimerkiksi aiheuttaa hankalia vokaali- tai konsonanttiyhtymiä.

Myös länsimurteissa tapahtuvaa yleiskielen *d:n* katoa *h:n* jäljessä on selitetty keskeisten sanojen korkealla esiintymistiheydellä, joka pätee esimerkiksi lukusanoihin ja *lähteä*-verbin taivutusmuotoihin. Nuolijärvi (1986: 112) esittää lisäksi, että lukusanoissa – joissa kato on kaikkein yleisin – selityksenä voisi olla se, että esimerkiksi *kahdeksan* ja *yhdeksän* ovat jo astevaihtelun ulkopuolella. Tämä selitys ei sellaisenaan päde läheisiin lukusanoihin *kahde-* ja *yhde-* eikä keskeisiin verbeihin (*ehdi-*, *lähde-*, *tehdä*), jotka ovat yhä astevaihtelussa. Myös sanojen merkityksellä voidaan olettaa olevan vaikutusta katotaajuuteen. Esimerkiksi erisnimien voidaan katsoa vastustavan katoa, vaikka sille olisi olemassa äänteelliset edellytykset (vrt. Lappalainen, Mustanoja & O’Dell 2019: 559). Helsingin puhekielen haastatteluissa toistuvien *lahti*-päätteisten paikannimien (esim. *Ruoholahti*, *Taivallahti*, *Hietalahti*) avulla on mahdollista tarkastella tätä ilmiötä. On myös luultavaa, että leksikaalistuneen sanan *yhdessä* katotaajuus on pienempi kuin kantasanalla *yhde-*, vaikka äänteellisesti niiden asema on samanlainen (vrt. Jonninen-Niilekselä 1982: 124).

### 3 Aineisto

Tutkimuksen aineistona käytetään Helsingin puhekielen pitkittäiskorpusta (Helpuhe 2014). Aineistonkeruu aloitettiin 1970-luvulla Helsingin yliopistossa ja liitettiin myöhemmin osaksi Nykysuomalaisen puhekielen murros -hanketta, jonka tavoitteena oli kerätä puhekielistä aineistoa Suomen suurimmista kaupungeista. Helsingin puhekielen pitkittäiskorpus oli näistä aineistoista ensimmäinen, johon kerättiin kolmas haastattelukierros. Hankkeen käynnisti 1970-luvulla Terho Itkonen ja sitä johti 1970- ja 1990-luvuilla Heikki Paunonen. Hanna Lappalainen on johtanut hanketta 2010-luvun keruusta lähtien. Aineisto on pääosin litteroitu ja digitoitu osaksi CSC:n Kielipankkia. Perinteiset taustamuuttujat kuten ikä, sukupuoli, asuinpaikka ja sosiaaliluokka on esitetty aineiston metatiedoissa.

Korpus mahdollistaa paneeli- ja trenditutkimusten tekemisen ja yhdistelemisen. Aineisto on paneelimuotoinen siinä mielessä, että kunkin uuden haastattelukierroksen keski-ikäiset ja vanhat puhujat on valittu jo edellisellä kierroksella haastateltujen joukosta. Tämä helpottaa idiolektitason seurua (ks. esim. Lappalainen ym. 2019), mutta väistämättä pienentää myöhempien haastattelukierrosten laajuutta, koska kaikkia haastateltavia ei ole voitu tavoittaa. Trendimuotoinen aineisto on puolestaan siksi, että uusi nuorten ryhmä on valittu aina samanlaisin periaattein. Aineiston kaikista kolmesta kierroksesta on toteutettu vielä niukalti tutkimuksia (ks. kuitenkin Kuparinen 2018; Kuparinen, Mustanoja, Peltonen, Santaharju & Leino 2019; Lappalainen ym. 2019; Kuparinen, Peltonen, Mustanoja, Leino & Santaharju 2021).

Tässä tutkimuksessa aineistona käytetään 1990-luvun haastattelua korpuksesta, joka on itsessään vielä jonkin verran suurempi. Korpuksesta on käytetty vain niitä valmiiksi litteroituja haastatteluita, joita tutkimusta aloitettaessa oli saatavilla. Tästä syystä esimerkiksi 1990-luvun haastattelujen osuus koko aineistosta jää melko pieneksi. Aineiston jakauma vuosikymmenten ja ikäryhmien kesken on esitetty taulukossa 1.

**Taulukko 1.**

**Aineiston jakautuminen kohortteihin ja haastattelukierroksiin.**

Kohortti	Haastattelut 1972–1974	Haastattelut 1991–1992	Haastattelut 2013
Kohortti 1 s. 1900–1907	65–74 vuotta 46 puhujaa		
Kohortti 2 s. 1927–1932	40–47 vuotta 41 puhujaa	59–65 vuotta 9 puhujaa	81–86 vuotta 4 puhujaa
Kohortti 3 s. 1952–1957	15–22 vuotta 39 puhujaa	34–40 vuotta 10 puhujaa	56–61 vuotta 9 puhujaa
Kohortti 4 s. 1971–1975		16–21 vuotta 14 puhujaa	38–42 vuotta 14 puhujaa
Kohortti 5 s. 1994–1997			16–19 vuotta 13 puhujaa

Olemme poimineet aineiston litteraateista *hd*-lekseemien esiintymät käyttäen säännöllisiä lausekkeita (Thompson 1968) ja tilastoineet ne lekseemeittään. Näin ollen kai-

kista haastatteluista on tilastoitu kaikki *hd*-tapaukset ( $N = 5\,536$ ), esiintymien laatu ( $d$ ,  $N = 3\,315$ ; kato,  $N = 2\,221$ ) ja esiintymislekseemit (yhteensä 191 lekseemiä). Aineistossa esiintyvät *hr*-tapaukset olemme tässä tutkimuksessa jättäneet huomiotta, koska niiden määrä on hyvin vähäinen ( $N = 14$ ).

## 4 *hd*-yhtymän variaatio ja muutos

Tässä luvussa tarkastelemme *hd*-yhtymän variaatiota ja muutosta tutkimuksen kolmen aikapisteen välillä. Alaluvussa 4.1 pyrimme luomaan yleiskuvan *hd*-yhtymän muutoksesta tutkimuksen ajanjaksona. Alaluvussa 4.2 tarkastelemme muutosta morfologisen diffuusion näkökulmasta ja keskitymme muotoryhmien välisiin eroihin kadon esiintymisessä. Koska *hd*-yhtymässä esiintyvän kadon on aiemmin huomattu kiinnittyvän voimakkaasti yksittäisiin sanoihin (esim. Nuolijärvi 1986: 112; Paunonen 1995 [1982]: 45; Nahkola & Saanilahti 2001: 51–52), tarkastelemme muotoryhmien välisen vertailun jälkeen vielä erikseen lekseemitasoa alaluvussa 4.3. On mahdollista, että jossakin muotoryhmässä yksittäinen lekseemi alkaa esiintyä huomattavasti muita saman ryhmän lekseemejä enemmän katomuodossa, jolloin diffuusio on leksikaalisesti ehdollista. Näin tarkastelumme on morfologisen ja leksikaalisen diffuusion hybridi.

### 4.1 Muutoksen yleiskuva

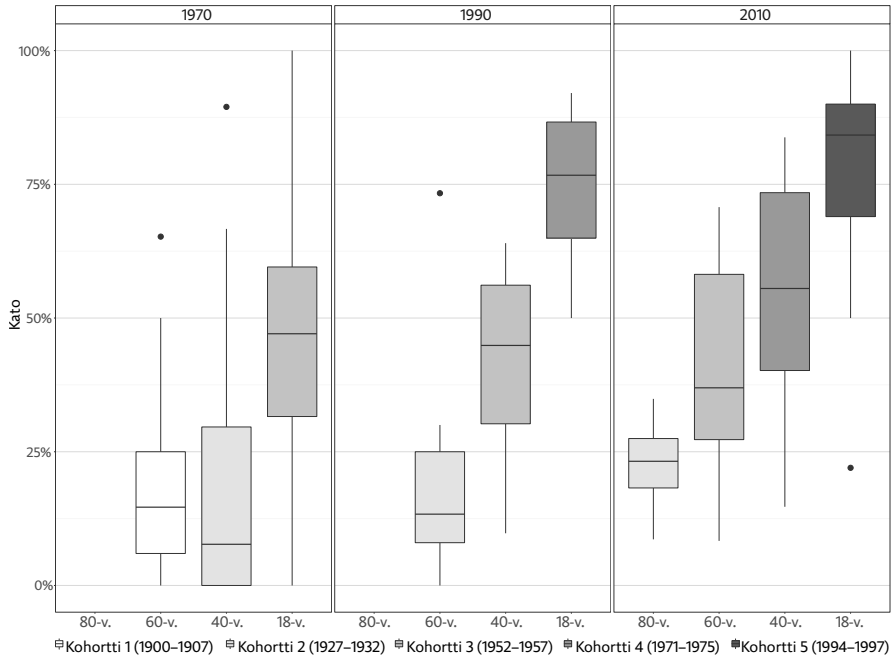
Yleiskuvan saamiseksi *hd*-yhtymän muutoksesta hyödynnämme ennen kaikkea yksinkertaisia frekvenssien ja suhteellisten osuuksien vertailuja. Tämän lisäksi tarkastelemme myös idiolektien muutosta kohorttien sisällä, koska aineisto on osittain paneeli-muotoinen. Tähän tarkoitukseen käytämme Wilcoxonin merkittyjen sijalukujen testiä (Wilcoxon 1945).<sup>2</sup> Idiolektitasoa on tutkinut tarkemmin ja useamman piirteen osalta Lappalainen kollegoineen (2019).

*hd*-yhtymässä havaittava muutos tutkimuksen aineistossa on esitetty ruutu- ja jana-kaaviona kuviossa 1 (seur. sivulla). Ruutu- ja jana-kaaviot edustavat samanaikaisesti ryhmiä ja yksilöitä niiden sisällä. Ruudun reunat kattavat ryhmän arvojen kaksi keskimmäistä neljänestä, ja ruudusta lähtevät janat esittävät ryhmän arvojen alinta ja ylintä neljänestä. Keskiviiva on ryhmän mediaaniarvo. Yksittäiset pisteet ovat muista huomattavasti eroavia tapauksia. Ikäryhmät on asetettu x-akselille, ja y-akselilla on esitetty katotapauksien osuus kaikista *hd*-tapauksista.

Kuviosta 1 havaitaan selvästi, että katotapaukset ovat yleistyneet Helsingin puhekielessä voimakkaasti 1970-luvulta 2010-luvulle. Siinä missä 1970-luvun nuorten

2. Wilcoxonin testi on tilastollinen hypoteesitesti, joka tarkastelee kahta toisistaan riippuvaista näytejoukkoa. Tässä tutkimuksessa verrataan keskenään samojen yksilöiden arvoja kahdessa eri aikapisteessä. Testi tarkastelee, onko arvon muutos ensimmäisestä aikapisteestä toiseen (pseudo-)mediaaniltaan nollassa poikkeava. Testin nollahypoteesina on, että joukkojen välillä ei ole eroa. Jos nollahypoteesi hylätään, voidaan tulkita aineistossa tapahtuneen muutosta ajanjaksojen välillä. Toinen vaihtoehto riippuvaisten otosten vertailuun olisi Studentin t-testi, joka kuitenkin olettaa aineiston olevan normaalisti jakautunut, mitä käsillä oleva aineisto ei ole.

mediaaniarvo oli alle 50 prosentin (47 %), on se 2010-luvun nuorilla jo 84 %, ja muidenkin ikäryhmien osuudet ovat kasvaneet. On siis selvää, että muutoksen suunta on kohti katomuotoja. Tämä havainto itsessään on kuitenkin varsin vajavainen, kun tunnetaan ilmiön luonne: kato on voimakkaasti keskittynyt tiettyihin lekseemeihin. Niinpä on tarkasteltava, onko kato yleistynyt samoissa lekseemeissä, joissa se on ennenkin ollut tavallinen, vai onko mukaan tullut uusia lekseemejä. Tätä tarkastelemme luvussa 4.3.



**Kuvio 1.**  
*hd*-yhtymän katomuotojen suhteellisten osuuksien muutos vuosikymmenittäin ja ikäryhmittäin.

Paunonen arvioi vuoden 2005 artikkelissaan, että *hd*-yhtymän muutos etenee Helsingin puhekielessä sukupolvittaisen muutoksen mallin mukaisesti (Paunonen 2005: 48; muutosmalleista ks. Labov 1994; Laasanen 2016; Kuparinen ym. 2019). Sukupolvittaisessa muutoksessa aina uusi sukupolvi on joko kasvattanut tietyn variantin käyttöä tai vähentänyt sitä, mikä johtaa ajan kuluessa yhteisön muutokseen. Sen sijaan idiolektitasolla muutosta ei tapahdu, vaan puhujat säilyttävät variantin puheessaan jotakuinkin samalla tasolla kuin nuoruudessaan. Paunosen tutkimus perustuu Helsingin puhekielen pitkittäiskorpuksen kahteen ensimmäiseen haastattelukierrokseen (1972–1974, 1991–1992), kun taas tässä tutkimuksessa tarkastellaan myös kolmatta aikapistettä (vuotta 2013).

Kuviota 1 tarkastelemalla Paunosen arvio on hyvinkin uskottava. Erityisesti 1970- ja 1990-lukujen välillä kohorttien katoesiintymien määrä ei näytä juuri muuttuvan. Li-

säksi aina uusi nuorten ryhmä käyttää entistä enemmän katomuotoja. 2010-luvulle tultaessa arvio ei näytä enää yhtä uskottavalta (vrt. Kuparinen ym. 2019). Kohortti 4 on vähentänyt selvästi katomuotojen käyttöä keski-ikään tultaessa, eikä 2010-luvun nuorten ryhmäkään näytä juuri eroavan 1990-luvun vastaavasta. On kuitenkin huomattava, että kuviossa näkyvät ruudut sisältävät myös sellaisia puhujia, jotka ovat osallistuneet vain yhteen haastattelukierrokseen. Näin ollen kuviossa esitetyt kohortit eivät ole vuosikymmenten välillä täysin vertailukelpoisia. Tästä syystä tarkastelemme useammalle haastattelukierrokselle osallistuneita puhujia vielä tilastollisesti käyttäen Wilcoxonin (1945) merkittyjen sijalukujen testiä. Wilcoxonin testi vertaa keskenään riippuvaisia otoksia eli tässä tapauksessa samoja yksilöitä eri vuosikymmeninä. Aineiston paneelimuodon ansiosta voimme verrata toisiinsa tietyn kohortin puhujia vuosikymmenestä toiseen ja tutkia, onko kohortissa tapahtunut tilastollisesti merkitsevää muutosta. Olemme soveltaneet samaa menetelmää koko aineistoon jo aiemmin (Kuparinen ym. 2019).

Vertailu koskee kohortteja 2, 3 ja 4, joista kahta ensimmäistä on haastateltu kaikilla vuosikymmenillä (joskin kohortista 2 vain neljää puhujaa 2010-luvulla) ja kohorttia 4 1990- ja 2010-luvuilla. Näin erilaisia vuosikymmenten välisiä vertailuja tulee yhteensä 5. Tilastollisen merkitsevyyden raja-arvona käytetään arvoa  $p < 0,05$ , mutta koska kyseessä on monivertailu (rinnakkaisia testejä on 5), lasketaan raja-arvoa Holm–Bonferroni-korjauksella (Holm 1979). Korjauksen tavoitteena on estää nollahypoteesin hylkäysvirhe eli väärät positiiviset tulokset (Neyman & Pearson 1933). Teknisesti korjaus tapahtuu järjestämällä kaikkien testien  $p$ -arvot pienimmästä suurimpaan ja jakamalla haluttu merkitsevyytaso (tässä tapauksessa  $0,05$ ) testien lukumäärällä (tässä tapauksessa 5). Mikäli pienin  $p$ -arvo on pienempi kuin näin saatu merkitsevyytaso (tässä tapauksessa  $0,05/5 = 0,01$ ), nollahypoteesi hylätään tämän testin osalta. Näin edetään järjestyksessä seuraavaksi pienimmästä  $p$ -arvosta suurimpaan, kunnes jonkin testin  $p$ -arvo ei enää alita merkitsevyytaso. Tämän jälkeen kaikki jäljelle jääneet nollahypoteesit hyväksytään. Jokaisen  $p$ -arvon kohdalla testien määrä vähenee yhdellä, jolloin viiden testin esimerkkitapauksessamme toiseksi pienimmän  $p$ -arvon kohdalla merkitsevyytaso on  $0,05/4 = 0,0125$ , kolmanneksi pienimmän kohdalla  $0,05/3 = 0,0167$  ja niin edelleen.

Wilcoxonin testin perusteella kohorttien sisällä ei vuosikymmenten välillä ole tilastollisesti merkitsevää muutosta muissa tapauksissa kuin jo kuviostakin huomattussa 1970-luvulla syntyneiden kohortissa (4) 1990-luvulta 2010-luvulle tultaessa. Tässä tapauksessa tilastollisen merkitsevyyden raja-arvo ( $p < 0,01$ ) alittuu selvästi ( $p = 0,004$ ). On vaikea sanoa, kertooko tämä muutos ilmiön hidastumisesta vai onko taustalla muita syitä. Ainakin on huomattava, että 1990-luvulla nuoruudessaan tämän kohortin puhujat käyttivät huomattavan paljon katomuotoja, jolloin myös ero seuraavaan vuosikymmeneen näyttäytyy helposti suurena. Kuviosta 1 käy esimerkiksi ilmi, että huolimatta tästä pudotuksesta 2010-luvun keski-ikäisten ryhmä käyttää enemmän katomuotoja kuin 1990-luvun keski-ikäisten ryhmä. Yhteisön muutos on siis edennyt yksilöiden muutoksesta huolimatta. Tässä tapauksessa voitaisiin puhua vetäytyvästä muutoksesta (Wagner & Sankoff 2011). Mikäli vetäytyvän muutoksen tendenssi jatkuu myös tulevaisuudessa, katomuotojen määrän kasvu hidastuu. Tässä tapauksessa on kuitenkin edelleen muistettava lekseemitaso: on luonnollista, että kasvu hidastuu,



jos uusia lekseemejä katomuodoille ei synny. Näin päädytään tilanteeseen, jossa tietyissä lekseemeissä kato on hyvin todennäköinen (esim. lukusanoissa) ja toisaalta suurimmassa osassa leksikköä kato ei ole todennäköinen. Mikäli uusia lekseemejä kuitenkin alkaa esiintyä katomuodoissa, katomuotojen määrä voi kasvaa nopeastikin. Seuraavissa kahdessa alaluvussa pyrimme selvittämään, miten muutos etenee muoto-ryhmittäin ja lekseemeittäin.

## 4.2 Morfologinen diffuusio

Yleiskielen *hd*-yhtymää käsittelevissä tutkimuksissa lekseemejä on ajoittain jaettu erilaisiin tarkasteltaviin ryhmiin. Esimerkiksi Nahkola ja Saanilahti (2001: 51) jakavat lekseemit kahteen lukusanojen ryhmään (*yhdeksän*, *kahdeksan* sekä *yksi*, *kaksi*) ja muihin, kun taas Kurjen (2005: 109) lukusanojen luokittelussa myös *yksi* ja *kaksi* on erotettu toisistaan. Lappalainen kollegoineen (2019: 559) sen sijaan jakaa aineiston esiintymät lukusanoihin, nimiin ja muihin. Luokittelemme *hd*-yhtymän sisältäviä lekseemejä toisistaan poikkeaviin muotoryhmiin myös tässä tutkimuksessa, mutta ryhmiä on huomattavasti enemmän kuin edellä mainituissa lähteissä.

Morfologisen diffuusion mallin mukaan kielenmuutokset eivät etene kaikkiin muotoryhmiin samanaikaisesti. Tässä tapauksessa on tärkeää selvittää, missä vaiheessa kato etenee mihinkin muotoryhmään eli missä järjestyksessä diffuusio tapahtuu. Tarkastelemme lekseemijoukot ovat ensisijaisesti morfologisesti perusteltuja, mutta huomioimme myös merkitysseikkoja ja tutkimme esimerkiksi lahti- ja lehtonimiä (esim. *Ruoholahti*, *Sofianlehto*). Diffuusiomallimme siis poikkeaa sellaisesta morfologisesta diffuusiosta, jossa tutkitaan muutoksen etenemistä esimerkiksi sijamuodosta toiseen.

Vertailtavia muotoryhmiä luodessa olemme pyrkineet löytämään kompromissin riittävän erottelun ja toisaalta riittävän aineiston välillä. Näin ollen esimerkiksi lekseemit, joissa *hd*-yhtymä esiintyy pitkän vokaalin jäljessä, on yhdistetty omaksi muotoryhmäkseen, vaikka se kattaa eri sanaluokkien sanoja. Vastaavasti samankin sanaluokan sanoja on voitu erotella toisistaan muotopiirteiden takia. Esimerkiksi keskeisissä verbeissä *lähteä*-tyypin verbit (A-VERBIT) on eroteltu *tehdä*-tyypin verbeistä (DA-VERBIT), koska *tehdä*-tyypin verbeissä *hd*-yhtymä esiintyy konsonanttivartalossa. Tarkasteltavat muotoryhmät ja niihin kuuluvat lekseemit on esitelty taulukossa 2. Kullekin muotoryhmälle on lisäksi annettu nimi, johon viitata tekstissä. Juoksevassa tekstissä muotoryhmistä käytetään selvyuden vuoksi kapiteelikirjaimia.

On huomattava, että taulukossa 2 esitellyt muotoryhmät eivät kata kaikkia aineistossa esiintyviä lekseemejä. Esimerkiksi muut nimet kuin LAHTI- JA LEHTONIMET (kuten *Vihti*) jäävät tarkastelun ulkopuolelle, koska niiden määrä on vähäinen. Taulukossa esitelty jako on sikäli perusteltu, että kukin muotoryhmä rakentuu vähintään yhdelle nimittäväälle tekijälle ja kunkin ryhmän esiintymämäärät ovat vertailukelpoisia keskenään.

Tarkastelemme muotoryhmissä tapahtuvaa muutosta ensin absoluuttisissa esiintymissä ja suhteellisissa osuuksissa tapahtuvana kokonaismuutoksena. Tämän tarkastelun jälkeen vertailemme korrelaatioanalyysin avulla vielä sellaisia muotoryhmiä, joissa katoa esiintyy runsaasti tai joihin se näyttää etenevän tutkimuksen ajanjakson aikana. Näin diffuusiota tarkastellaan myös tilastollisesti.

Korrelaatioanalyysissä tarkastellaan kerrallaan kahta muotoryhmää yhdessä aikapisteessä niin, että jokaisen yksilön tuottamien katomuotojen suhteellista osuutta muotoryhmän kaikista esiintymistä verrataan toisen muotoryhmän vastaavaan osuuteen samalta yksilöltä. Korrelaatioanalyysi kertoo siis muotoryhmien välisistä yhteyksistä kussakin aikapisteessä. Mikäli esimerkiksi muotoryhmät YHDEKSÄN ja YKSI korreloivat 1970-luvulla, voidaan todeta kadon esiintyvän niissä haastateltavien puheessa samansuuntaisesti tässä aikapisteessä. Jos vastaavasti muotoryhmät YHDEKSÄN ja DIFTONGI eivät korreloivat 1970-luvulla, muotoryhmien käytössä kyseisessä aikapisteessä ei löydetä yhteyttä yksilöiden puheessa.

### Taulukko 2.

Muotoryhmät, niihin kuuluvat lekseemit ja esiintymien määrä.

Muotoryhmä	Lekseemit	N
YHDEKSÄN	<i>yhdeksän, kahdeksan</i>	1034
YKSI	<i>yhde-, kahde<sup>3</sup></i>	852
A-VERBIT	<i>lähde-, ehdi-, johdu-, mahdu-, pohdi-, ryhdy-, tahdo-</i>	723
DA-VERBIT	<i>tehdä, nähdä</i>	747
YHDESSÄ	<i>yhdessä, yhdistää, yhdistäminen, yhdistelmä, yhdistyä, yhdistys, yhdyskunta</i>	319
MAHDO-	<i>mahdoton, mahdollinen, mahdollisuus, mahdollistaa, ehdoton</i>	527
LAHTI- JA LEHTONIMET	<i>Aurinkolahde-, Espoonlahde-, Hakalahde-, Haukilahde-, Hietalahde-, Humallahde-, Huopalahde-, Kaisaniemenlahde-, Kauklahde-, Kilpilahde-, Kivenlahde-, Korpilahde-, Lapinlahde-, Meilahde-, Pääkslahde-, Ruoholahde-, Sepetlahde-, Sofianlahde-, Suolahde-, Suomenlahde-, Suvilahde-, Taivallahde-, Töölönlahde-, Vanhankaupunginlahde-, Vehkalahde-, Vesilahde-, Virolahde-</i>	240
KOHD-	<i>kohda-, kiinnekohda-, kohokohda-, kohde, kohden, kohdata, kohdella, kohdistaa, kohdistua</i>	158
PITKÄ VOKAALI	<i>hiihdä-, hiihdellä-, hiihdo-, huuhdo-, paahda-, paahdin-, viihde-, viihdy-, viihdyttää</i>	147
DIFTONGI	<i>kaihda-, kiehdo-, köyhdyttää-, vaihde-, vaihdella-, vaihda-, vaihdo-, vaihdos, vaihdu-, vaihdunta, vauhdi-, vauhdikas, vauhditon</i>	155
KAKSITAVUISET NOMINIT	<i>ahdas, ehdo-, johdo-, lähdö-, lähde, lahde- (ei nimenä), lehde-, lyhdy-, puhdas, vahdi-, rohdos, suhde, tähde-, tahdi-, tehdas</i>	303
PAINOTON TAVU	<i>erehdys, hätkähdyttää-, huolehdi-, kömmähdys, kadehdi-, kiiruhda-, kolahda-, lipsahda-, lipsahdus, luonnehdi-, möksähdä-, mätkähdä-, molskahda-, murehdi-, nukahda-, pälkähdys, paukahdus, perehdy-, perehdyttää-, piipahda-, pujahda-, pulpahda-, purjehdi-, purjehdus, putkahda-, pyörähdä-, pysähdy-, pysähdys, räjähdys, romahda-, romahdus, sykähdyttää-, tärähdys, tapahdu-, tervehdi-, tervehdys, tipahda-, tulehdus, unohda-, unohdus</i>	100

Koska tutkimuksessa on tarkasteltu useampaa aikatasoa, voidaan korrelaation avulla seurata myös katomuotojen etenemistä: jos huomataan, että 1990-luvulla YHDEKSÄN ja DIFTONGI korreloivatkin, voidaan päätellä katomuotojen joko eden-

3. Lekseemeistä käytetään tässä tutkimuksessa perusmuotojen sijaan sellaista kirjoitusasua, jossa *hd*-yhtymä näkyy.

neen muotoryhmään DIFTONGI tai vähentyneen muotoryhmässä YHDEKSÄN. Kadon esiintymäämääriä ja suhteellisten osuuksien muutosta tarkastelemalla voidaan päätellä, kumpi vaihtoehtoista on toteutunut. Näin kokonaismäärien tarkastelu ja korrelaatioanalyysi täydentävät toisiaan.

Koska tarkasteltavia muotoryhmiä on useita ja tutkimuksessa on kolme aikapistettä, kaikkien korrelaatioparien tutkiminen kaikissa aikapisteissä lisäisi todennäköisyyttä saada tilastollisissa testeissä tuloksia, jotka näyttävät merkitseviltä mutta johtuvat sattumasta (vrt. Wilcoxonin testit luvussa 4.1). Tästä syystä jätämme korrelaatioanalyysin ulkopuolelle sellaiset muotoryhmät, joissa ei ole millään vuosikymmenellä yli 10 prosentin katoedustusta, sekä sellaiset muotoryhmät, joissa kadon osuus ei kasva ajan kuluessa. Myös korrelaatiotestien merkitsevyytensä lasketaan Holm–Bonferroni-korjauksella luvussa 4.1 esitettyyn tapaan.

Korrelaatioista raportoimme p-arvon sekä korrelaatiokertoimen (r). Korrelaatiokerroin r saa arvon väliltä [-1, 1]. Mitä kauempana r on arvosta 0, sitä vahvempi on muuttujien välinen lineaarinen riippuvuus (negatiivisen riippuvuuden tapauksessa arvo on negatiivinen ja positiivisen riippuvuuden tapauksessa positiivinen). On korostettava, että korrelaatio ei kerro kausaalisuhteesta, vaan ainoastaan siitä, että katomuodot esiintyvät vertailtavissa muotoryhmissä yksilöiden puheessa samansuuntaisesti. Tästä syystä korrelaatioanalyysi sopii hyvin yhteen morfologisen diffuusion kanssa, sillä diffuusiomalli ei myöskään selitä muutoksia, vaan pelkästään kuvaa niiden etenemistä.

Aloitamme muotoryhmittäisen tarkastelun vuosikymmenittäisten kokonaismäärien vertailulla. Taulukossa 3 on esitetty kustakin muotoryhmästä katomuotojen määrä, esiintymien kokonaismäärä ja kadon prosentuaalinen osuus kullakin vuosikymmenellä. Taulukosta voidaan havaita jo kuviosta 1 todettu tendenssi katomuotojen yleistymiselle koko aineistossa ajan kuluessa. On kuitenkin selvää, että katomuotojen lisääntyminen ei ole jakautunut muotoryhmien kesken tasaisesti, vaan joissakin muotoryhmissä kadon määrä on hyvin pieni vielä 2010-luvullakin.

Taulukossa 3 esitettyjen katomuotojen osuuksien lisäksi tarkastelemme muotoryhmiä korrelaatioiden avulla. Korrelaatiokertoimien ja korrelaatioparien p-arvojen avulla voimme selvittää, missä muotoryhmissä kato etenee samassa tahdissa. Näin voimme tarkentaa kokonaismäärien perusteella luotavaa kuvaa diffuusiojärjestyksestä. Jätämme korrelaatioanalyysin ulkopuolelle sellaiset muotoryhmät, joissa katoa ei juurikaan esiinny (PITKÄ VOKAALI, PAINOTON TAVU) sekä sellaiset muotoryhmät, joissa kadon osuus ei kasva lineaarisesti vuosikymmeneltä toiselle (YHDESSÄ, KOHDA-). Teemme korrelaatiotestit kullakin vuosikymmenellä vain niille muotoryhmille, joissa kadon osuus on yli 10 prosenttia. Näin esimerkiksi muotoryhmää DIFTONGI tarkastellaan vain 2010-luvulla. Ehdot täyttäviä muotoryhmiä on 1970-luvulla viisi, 1990-luvulla kuusi ja 2010-luvulla kahdeksan. Korrelaatiotestejä tehdään yhteensä 53 kappaletta. Tilastollisen merkitsevyyden taso on  $p < 0,05$ , jota lasketaan Holm–Bonferroni-korjauksella.

Taulukosta 3 huomataan selviä eroja muotoryhmien välillä katomuotojen esiintymisessä. Kato on 1970-luvulla kaikkein yleisin muotoryhmissä A-VERBIT, YHDEKSÄN ja YKSI, joissa katoa esiintyy noin puolessa kaikista tapauksista. DA-VERBEISSÄ kato on tavallinen, mutta ei yhtä yleinen kuin kolmessa edellä mainitussa. Lisäksi muoto-

ryhmissä MAHDO- ja KOHDA- katoa esiintyy yli kymmenessä prosentissa kaikista tapauksista. Muissa muotoryhmissä kato on joko hyvin harvinainen tai sitä ei esiinny lainkaan. Tilastollisesti merkitseviä korrelaatiopareja 1970-luvulla muodostavat YHDEKSÄN ja YKSI ( $r = 0,59$ ,  $p < 0,0001$ ), YHDEKSÄN ja A-VERBIT ( $r = 0,54$ ,  $p < 0,0001$ ) sekä YKSI ja A-VERBIT ( $r = 0,56$ ,  $p < 0,0001$ ). Näissä kolmessa muotoryhmässä kato näyttää siis etenevän samansuuntaisesti.

### Taulukko 3.

Katomuotojen määrä, esiintymien kokonaismäärä ja kadon osuus kullakin vuosikymmenellä.

Muotoryhmä	1970-luku	1990-luku	2010-luku
YHDEKSÄN	302/555 (54 %)	178/239 (74 %)	219/240 (91 %)
YKSI	219/494 (44 %)	136/191 (71 %)	120/167 (72 %)
A-VERBIT	171/311 (55 %)	152/214 (71 %)	149/198 (75 %)
DA-VERBIT	72/300 (24 %)	104/236 (44 %)	117/211 (55 %)
YHDESSÄ	9/152 (6 %)	19/100 (19 %)	11/67 (16 %)
MAHDO-	33/252 (13 %)	78/183 (43 %)	39/92 (42 %)
LAHTI- JA LEHTONIMET	5/143 (3 %)	1/30 (3 %)	7/67 (10 %)
KOHDA-	11/102 (11 %)	2/24 (8 %)	5/32 (16 %)
PITKÄ VOKAALI	0/74 (0 %)	0/37 (0 %)	1/36 (3 %)
DIFTONGI	1/68 (1 %)	1/33 (3 %)	15/54 (28 %)
KAKSITAVUISET NOMINIT	3/181 (2 %)	7/59 (12 %)	12/63 (19 %)
PAINOTON TAVU	1/37 (3 %)	0/24 (0 %)	2/39 (5 %)

1990-luvulle tullessa muotoryhmissä YHDEKSÄN, A-VERBIT ja DA-VERBIT kadon määrä on lisääntynyt noin 20 prosenttiyksikköä. Eniten kadon osuus on kasvanut muotoryhmissä YKSI ja MAHDO-, joissa kadon osuus on kasvanut 27 ja 30 prosenttiyksikköä. Sen sijaan 1970-luvulla muotoryhmän MAHDO- kanssa samalla tasolla olleessa muotoryhmässä KOHDA- kato on pysynyt suurin piirtein ennallaan. Näyttää siis siltä, että kato ei välttämättä etene systemaattisesti sellaisissakaan muotoryhmissä, joissa sitä jo ennestään esiintyy. 1990-luvulla katoa esiintyy yli 10 prosentissa kaikista tapauksista myös muotoryhmissä YHDESSÄ ja KAKSITAVUISET NOMINIT. Korrelaatiotesteissä tilastollisesti merkitseviä pareja on tullut kolme lisää 1970-luvun tilanteeseen verrattuna. Edelleen korreloivat YHDEKSÄN ja YKSI ( $r = 0,78$ ,  $p < 0,0001$ ), YHDEKSÄN ja A-VERBIT ( $r = 0,73$ ,  $p < 0,0001$ ) sekä YKSI ja A-VERBIT ( $r = 0,83$ ,  $p < 0,0001$ ). Korrelaatiokertoimet ovat näissä pareissa entisestään kohentuneet ja osoittavat hyvin vahvaa yhteyttä muotoryhmien välillä. 1990-luvulla myös DA-VERBIEN muotoryhmä korreloi kolmen edellä mainitun kanssa: YHDEKSÄN ja DA-VERBIT ( $r = 0,60$ ,  $p = 0,0003$ ), YKSI ja DA-VERBIT ( $r = 0,66$ ,  $p < 0,0001$ ) sekä A-VERBIT ja DA-VERBIT ( $r = 0,78$ ,  $p < 0,0001$ ). Kato on siis selvästi edennyt myös DA-VERBEIHIN 1990-luvun aineistossa.

Siinä missä kadon määrä kasvoi muotoryhmissä YHDEKSÄN, YKSI, A-VERBIT ja DA-VERBIT 1970-luvulta 1990-luvulle suurin piirtein saman verran, on 2010-luvulle

tultaessa muotoryhmien YKSI ja A-VERBIT katotilanne pysynyt lähes 1990-luvun tasolla. Myös 1970- ja 1990-lukujen välillä selvästi katomuodoissa yleistyneen muotoryhmän MAHDO- katoedustus on jäänyt 1990-luvun tasolle. Sen sijaan muotoryhmässä YHDEKSÄN kadon osuus on edelleen kasvanut 17 prosenttiyksikköä ja DA-VERBEISSÄKIN 11 prosenttiyksikköä. Kaikkein eniten kadon osuus on kuitenkin kasvanut muotoryhmässä DIFTONGI (25 prosenttiyksikköä). Lisäksi kato on lisääntynyt kaikissa muissa muotoryhmissä paitsi muotoryhmässä YHDESSÄ. On tosin huomattava, että muotoryhmien KOHDA-, PITKÄ VOKAALI ja PAINOTON TAVU absoluuttinen määrä on niin vähäinen, että niiden tarkastelu suhteellisin osuuksin voi olla harhaanjohtavaa. Kato on 2010-luvulla selvästi yleisin muotoryhmässä YHDEKSÄN.

Taulukosta 3 havaittava katomuotojen etenemisen hidastuminen muotoryhmissä YKSI ja A-VERBIT heijastuu myös korrelaatiotesteihin. Muotoryhmät YHDEKSÄN ja YKSI korreloivat vielä 2010-luvullakin ( $r = 0,57$ ,  $p = 0,001$ ), mutta korrelaatiokerroin on pienempi kuin 1970- ja 1990-luvuilla. Muotoryhmät YHDEKSÄN ja A-VERBIT eivät enää 2010-luvulla korreloi tilastollisesti merkitsevästi. Muotoryhmien YKSI ja A-VERBIT välinen korrelaatiokerroin on vielä 2010-luvullakin suuri ( $r = 0,77$ ,  $p < 0,0001$ ), kuten jo osuuksien perusteella saattoi olettaa. Kolmas tilastollisesti merkitsevä korrelaatio 2010-luvulla on DA-VERBIEN ja muotoryhmän MAHDO- välillä ( $r = 0,67$ ,  $p = 0,0002$ ). Kadon eteneminen muotoryhmään MAHDO- saa siis myös tilastollista tukea 2010-luvun aineistosta.

Muotoryhmissä DIFTONGI, KAKSITAVUISET NOMINIT ja LAHTI- JA LEHTONIMET kadon osuus on vielä melko pieni, mutta kaikkien kolmen muotoryhmän suhteelliset osuudet osoittavat kasvua. LAHTI- JA LEHTONIMIEN muotoryhmässä katoa esiintyy 2010-luvulla kaikkiaan seitsemän kertaa. Näistä esiintymistä kuitenkin yksittäisen nuoren naispuhujan esiintymiä on viisi, jolloin tulosta on vaikea laajentaa koskemaan koko puhujayhteisöä. Sen sijaan muotoryhmässä DIFTONGI katoa esiintyy 2010-luvulla seitsemällä puhujalla ja KAKSITAVUISISSA NOMINEISSA kahdeksalla puhujalla. KAKSITAVUISISSA NOMINEISSA katoa esiintyy jo 1990-luvulla kaikkiaan viidellä puhujalla. Koska LAHTI- JA LEHTONIMET ovat perusmerkitykseltään nimenomaan kaksitavuisia nomineja, näyttää siltä, että sanan esiintyminen nimessä jossain määrin hidastaa kadon etenemistä tässä aineistossa (vrt. Lappalainen ym. 2019: 559, 561).

Korreloivia muotoryhmiä ovat siis 1970- ja 1990-luvuilla YHDEKSÄN, YKSI ja A-VERBIT. Näissä muotoryhmissä kato on jo 1970-luvun aineistossa hyvin vakiintunut. 2010-luvun aineistossa muotoryhmien asetelma kuitenkin rikkoutuu: vahvasti korreloivat edelleen YKSI ja A-VERBIT, kun taas muotoryhmien YKSI ja YHDEKSÄN korrelaatio on heikentynyt ja A-VERBIEN ja muotoryhmän YHDEKSÄN välinen korrelaatio ei ole enää tilastollisesti merkitsevä. Syynä on se, että muotoryhmän YHDEKSÄN kato-osuus on kasvanut hyvin suureksi. 1990-luvulla myös muotoryhmä DA-VERBIT korreloi kolmen ensin mainitun kanssa. Kato näyttää vakiintuneen muotoryhmään siis viimeistään 1990-luvulla, vaikka kadon osuus ei missään tutkimuksen vaiheessa yllä ilman muotoryhmien YHDEKSÄN, YKSI ja A-VERBIT tasolle.

2010-luvun aineistossa DA-VERBIT ei enää korreloi kolmen muun kanssa, vaan muotoryhmän MAHDO- kanssa. Kadon voidaan korrelaatioiden perusteella katsoa vakiintuneen muotoryhmään MAHDO- siis viimeistään 2010-luvulla. DA-VERBIEN

tilanne on sikäli kiinnostava, että *tehdä* ja *nähdä* nostetaan usein esimerkeiksi tajaan katomuodoissa esiintyvistä verbeistä (Nuolijärvi 1986: 112; Paunonen 1995 [1982]: 45; Nahkola & Saanilahti 2001: 52; Kurki 2005: 109). Helsingin puhekielen aineiston perusteella näyttää kuitenkin siltä, että katoa ei esiinny niissä yhtä runsaasti kuin A-VERBEISSÄ (esim. *lähde*-). Koska verbiryhmien frekvenssi on lähes yhtä suuri (ks. taulukkoa 2), syy katomuotojen pienemmälle osuudelle DA-VERBEISSÄ löytynee morfologisista seikoista. Kurki (2005: 122–123) on Varbrul-analyysin<sup>4</sup> avulla osoittanut, että lekseemin kuuluminen astevaihteluun vähentää kadon esiintymistä (vrt. Nuolijärvi 1986: 114). Näiden verbien perusteella vaikuttaa kuitenkin siltä, että myös vartalovaihtelulla voi olla merkitystä kadon etenemiselle. On nimittäin huomattava, että A-VERBEISSÄ *hd*-muodot esiintyvät aina vokaalivartalossa (*lähde*-), kun taas DA-VERBEISSÄ yhtymä esiintyy konsonanttivartalossa. Aiheen syvällisempi tarkastelu jää kuitenkin tämän tutkimuksen ulkopuolelle.

Taulukossa 3 esitettyjen määrien ja korrelaatioanalyysien avulla voidaan hahmotella muotoryhmien välistä diffuusiojärjestystä. Jo 1970-luvulla katoa esiintyy selvästi muotoryhmissä YHDEKSÄN, YKSI ja A-VERBIT. Nämä kolme muotoryhmää myös korreloivat 1970- ja 1990-luvuilla. Hieman näitä vähemmän katoa esiintyy DA-VERBEISSÄ, jotka korreloivat kolmen muun kanssa 1990-luvun aineistossa. Näin kadon eteneminen DA-VERBEIHIN voidaan tämän tutkimuksen kontekstissa ajoittaa viimeistään 1990-luvulle. Muotoryhmässä MAHDO- kato on vielä 1970-luvulla orastavaa, mutta 1990-luvulla jo melko vahvaa. Viimeistään 2010-luvun aineistossa kadon tulkitaan edenneen myös tähän muotoryhmään, sillä se korreloi muotoryhmän DA-VERBIT kanssa. 2010-luvulla kato vaikuttaa orastavan muotoryhmissä DIFTONGI ja KAKSITAVUISET NOMINIT, mutta tälle havainnolle emme saa tilastollista tukea. Asetelmassa 1 katomuotojen morfologinen diffuusio on hahmoteltu janalle.

#### **Asetelma 1.**

##### **Katomuotojen morfologinen diffuusio.**

1970-luku	1990-luku	2010-luku	orastava
YHDEKSÄN	DA-VERBIT	MAHDO-	DIFTONGI
A-VERBIT			KAKSITAVUISET NOMINIT
YKSI			

Muotoryhmäkohtaisen tarkastelun perusteella olemme saaneet kadon etenemisestä jo melko selvän kuvan. Sellaisissa muotoryhmissä, joissa katoa esiintyy melko vähän (KAKSITAVUISET NOMINIT, DIFTONGI, LAHTI- JA LEHTONIMET), yksittäisten lekseemien vaikutus voi kuitenkin olla huomattavan suuri, minkä vuoksi tarkastelemme katomuotojen etenemistä vielä leksikaalisen diffuusion tasolla. Lisäksi katomuodot eivät vastoin odotuksia kasvaneet lineaarisesti kahdessa muotoryhmässä (YHDESSÄ ja KOHDA-). Lekseemikohtainen tarkastelu voi valaista näidenkin muotoryhmien sisäistä dynamiikkaa paremmin.

4. Varbrul on sosiolinguistiikassa perinteisesti käytetty logistiseen regressioon perustuva tilastollinen menetelmä (Cedergren & Sankoff 1974).

### 4.3 Leksikaalinen diffuusio

Tutkimus lähti liikkeelle Nuolijärven (1986) jo 35 vuotta sitten esittämästä olettamuksesta, että katomuotojen leviäminen on jähmettynyt tiettyihin lekseemeihin. Tässä alaluvussa tarkastelemme, missä lekseemeissä kato tutkimuksen aineistossa esiintyy ja kasvatattaako kato alaansa uusiin lekseemeihin ajan kuluessa. Mikäli kadon leviäminen on aidosti pysähtynyt tiettyihin lekseemeihin, täytyy tätä ilmiötä pyrkiä selittämään. Mikäli taas kato on hiljalleen edennyt uusiin lekseemeihin, on syytä tarkastella diffuusiojärjestystä: mihin lekseemeihin kato näyttää etenevän? Lisäksi lekseemitason tarkastelu auttaa selittämään alaluvussa 4.2 havaittuja muotoryhmäkohtaisia muutoksia.

Kaikkiaan *hd*-yhtymä esiintyy tutkimuksen aineistossa 191 lekseemissä. Lekseemi on tässä tutkimuksessa ymmärretty niin, että esimerkiksi saman kantasanan eri johdokset edustavat eri lekseemejä. Näin ollen esimerkiksi sanat *mahdollinen*, *mahdoton*, *mahdollistaa* ja *mahdollisuus* edustavat kaikki eri lekseemejä. Näistä 191:stä *hd*-lekseemistä kato esiintyy 44:ssä. Lisäksi tässä lähes kahdensadan haastattelun korpuksessa esiintyy yli 5 kertaa vain 61 lekseemiä 191:stä. Näin ollen niistä lekseemeistä, joissa *hd*-yhtymä esiintyy, vain harva on frekventtejä, mikä osaltaan selittää katomuotojen keskittymistä tiettyihin lekseemeihin ja muotoryhmiin.

Seuraavassa tarkastelemme katoesiintymiä sisältävien lekseemien muutosta kolmen aikapisteen välillä kahdella eri tasolla. Toisaalta kiinnitetään huomiota itse lekseemeihin: tuleeko vuosikymmenten kuluessa uusia lekseemejä, joissa kato esiintyy? Toisaalta tarkastellaan sitä, kasvaako katoesiintymien osuus niissä lekseemeissä, joissa kato on jo aiemminkin esiintynyt.

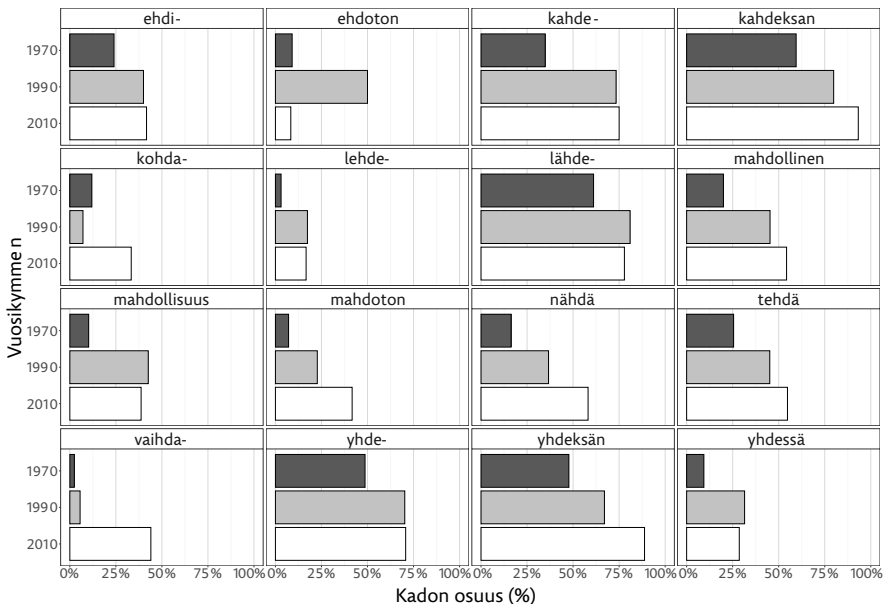
Kaikilla vuosikymmenillä kato esiintyy seuraavissa lekseemeissä: *yhde-*, *kahde-*, *kahdeksan*, *yhdeksän*, *tehdä*, *nähdä*, *lähde-*, *ehdi-*, *mahdollinen*, *mahdoton*, *mahdollisuus*, *yhdessä*, *kohda-*, *lehde-*, *vaihda-* ja *ehdoton*. Näiden 16:n lekseemin lisäksi 1970-luvulla kato esiintyy seuraavissa 8 lekseemissä (katomuotojen määrä sulkeissa): *tahdo-* (11), *kahdeksikko* (8), *yhdeksikkö* (4), *Lapinlahde-* (3), *Hietalahde-* (2), *(jonkin) tähden* (2), *puhdas* (1) ja *pyörähdä-* (1). Yhteensä kato esiintyy siis 24 lekseemissä 148:sta mahdollisesta 1970-luvun aineistossa. Kiinnostava tapaus on *pyörähdä-*, jossa *hd*-yhtymä esiintyy painottoman tavun jäljessä, kun kaikissa muissa tapauksissa yhtymä sijaitsee painollisen (yleensä ensimmäisen) tavun jäljessä.

1990-luvun aineistossa esiintyy katoa 25 lekseemissä. Sellaisia lekseemejä, joita ei esiinny katomuotoisena 1970-luvun aineistossa, ovat *mahdu-* (4), *lähdö-* (2), *yhdistys* (2), *kohde* (1), *ahdistaa* (1), *Töölönlahde-* (1) ja *ehdotus* (1). Katomäärät ovat hyvin pieniä, mutta on muistettava, että 1990-luvun aineisto on huomattavasti pienempi kuin 1970-luvun aineisto (33 puhujaa ja 126 puhujaa). Myös *hd*-yhtymän sisältävien lekseemien määrä on 1990-luvulla pienempi (94 lekseemiä). Kaikki yllä mainitut lekseemit esiintyvät myös 1970-luvun aineistossa, mutta *hd*-muotoisina. Tällä pienellä otannalla näyttäisi siltä, että kato siirtyy myös uusiin lekseemeihin. Vielä selvemmin tästä muutoksesta todistaa 2010-luvun aineiston tarkastelu.

Kun 1990-luvulla katoa esiintyi 25:ssä lekseemissä ja 1970-luvulla 24:ssä, 2010-luvun aineistossa kato esiintyy 31:ssä eri lekseemissä. Kokonaan uusia katomuotoisia lekseemejä 2010-luvun aineistossa ovat *Ruoholahde-* (5), *vaihdella* (2), *ehdottaa* (2), *mah-*

*dollistaa* (2), *viihdy-* (1), *tapahdu-* (1), *yhdistää* (1), *unohda-* (1), *vaihtoehto-* (1), *vauhdikas* (1), *Aurinkolahde-* (1) ja *Haukilahde-* (1). 2010-luvun uusissa lekseemeissä huomio kiinnittyy muotoseikkoihin. Kolmessa uudessa lekseemissä *hd*-yhtymä esiintyy pitkän vokaalin tai diftongin jäljessä (*vaihdella*, *viihdy-*, *vauhdikas*), kun tähän asti mainituista lekseemeistä kaikissa muissa paitsi lekseemissä *vaihda-* *hd*-yhtymä on esiintynyt lyhyen vokaaliaineksen jäljessä. Lisäksi kahdessa lekseemissä yhtymä esiintyy painottoman tavun jäljessä (*tapahdu-*, *unohda-*). Nämä tapaukset ovat edelleen harvinaisia, mutta osoittavat, että kato pyrkii myös uusiin, morfologisesti toisistaan poikkeaviin lekseemeihin.

Katomuodot näyttävät etenevän myös uusiin lekseemeihin Helsingin puhekielessä. Esiintymämäärät ovat kuitenkin hyvin pieniä, jolloin katomuotojen määrän kasvu syntyy pääosin hyvin frekventtien lekseemien entistä suuremmasta katotaajuudesta. Tätä tarkastelemme seuraamalla kaikilla vuosikymmenillä esiintyvien lekseemien katomääriä kuviossa 2. Kuviossa kukin lekseemeistä on esitetty omana kaavionaan. Eri vuosikymmenten osuus on kussakin kaaviossa omalla rivillään ja esitetty eri sävyin.



**Kuvio 2.**

**Kaikilla vuosikymmenillä katomuodossa esiintyvien lekseemien suhteelliset osuudet.**

Kuviosta 2 huomataan yleisenä tendenssinä se, että 1970-luvun ja 1990-luvun aineistojen välillä näyttää olevan monissa tapauksissa suuri ero. Sen sijaan 1990-luvun ja 2010-luvun välillä erot jäävät pienemmiksi. Valtaosassa lekseemeistä katomuotojen osuus kasvaa vuosikymmeneltä toiselle. Selvän poikkeuksen tähän tekee lekseemi *ehdoton*, joka esiintyy huomattavan paljon katomuodoissa 1990-luvulla. Lekseemi on



melko harvinainen (koko aineistossa 76 esiintymää), ja 1990-luvulla kolme puhujaa käyttää tätä lekseemiä huomattavan monta kertaa katomuodoissa. Näin ollen lekseemin poikkeava variaatiokaava selittyy pääosin muutamien yksilöiden perusteella. Selvimmin vuosikymmeneltä toiselle katomuodot kasvavat lukusanoissa *kahdeksan* ja *yhdeksän* sekä verbeissä *tehdä* ja *nähdä*. Tämä oli havaittavissa jo luvun 4.2 muotoryhmittäisen tarkastelun perusteella.

Kun tarkastellaan lekseemejä vuosikymmenittäin, 1970-luvulla katomuotoja esiintyy eniten lekseemeissä *lähde-* (N = 154/252, 61 %), *kahdeksan* (188/316, 59 %), *yhde-* (165/339, 49 %) ja *yhdeksän* (114/239, 48 %). Näissä huomio kiinnittyy kahteen seikkaan. Ensinnäkin tulokset näyttävät vastaavan esimerkiksi Tampereen puhekielen vastaavia. Jonninen-Niilekselän (1982: 124) Tampereen puhekieltä käsittelevässä tutkimuksessa yleisimmät lekseemit katomuodoissa ovat juuri lukusanat ja verbi *lähteä*. Toisekseen sanojen *yhde-* ja *yhdeksän* kato-osuus on lähestulkoon yhtä suuri, mikä on osittain odotuksenvastaista, sillä *yksi* osallistuu edelleen astevaihteluun (vrt. Nuolijärvi 1986: 114). Jos verrataan esimerkiksi vastaavassa suhteessa olevia lekseemejä *kahde-* (N = 54/155, 35 %) ja *kahdeksan* (59 %), on ero paljon selvempi. Tätä voi osittain selittää sanan *yhdeksän* toistuvuus haastatteluaineistoissa vuosiluvuista puhuttaessa (*tuhat yhdeksänsataa*).<sup>5</sup>

1990-luvun aineistossa yleisimpiä lekseemejä katomuodoissa ovat edelleen *lähde-* (140/173, 81 %) ja *kahdeksan* (111/139, 80 %). Huomattavasti enemmän katomuotoja esiintyy kuitenkin 1990-luvulla lekseemissä *kahde-* (44/60, 73 %). Näin ollen edellä käsitelty lekseemien *kahde-* ja *kahdeksan* ero on pienentynyt. Vastaavasti *yhde-* (92/131, 70 %) ja *yhdeksän* (67/100, 67 %) esiintyvät 1990-luvullakin kutakuinkin yhtä paljon katomuodoissa. Muita 1990-luvulla huomattavasti 1970-lukua enemmän katomuodoissa esiintyviä lekseemejä ovat *mahdollinen* (29/64, 45 %), *mahdollisuus* (23/54, 43 %), *yhdessä* (17/54, 31 %) sekä edellä käsitelty *ehdoton*.

Siinä missä 1990-luvulla *kahdeksan* ja *yhdeksän* olivat jokseenkin yhtä yleisiä katomuodoissa kuin niiden kantasanat *kahde-* ja *yhde-*, on tilanne 2010-luvulla muuttunut. Selvästi eniten katomuotoja esiintyy lekseemeissä *kahdeksan* (124/133, 93 %) ja *yhdeksän* (95/107, 89 %). Näiden jälkeen yleisin on edelleen verbi *lähde-* (137/176, 78 %), ja vasta sen jälkeen tulevat lukusanat *kahde-* (33/44, 75 %) ja *yhde-* (87/123, 71 %). Muotoryhmien YHDEKSÄN ja YKSI sekä A-VERBIT erkaantuminen 2010-luvulla havaittiin jo alaluvussa 4.2. 1990-luvulta 2010-luvulle katomuodot ovat prosenttiyksiköiden valossa yleistyneet huomattavasti lekseemeissä *kohda-* (5/15, 33 %), *vaihda-* (11/25, 44 %) ja *nähdä* (25/43, 58 %).

Kuvio 2 valaisee myös muutamien luvussa 4.2 käsiteltyjen muotoryhmien muutosta. Lekseemi *vaihda-* aiheuttaa suurimman osan muotoryhmän DIFTONGI katoesiintymistä, joten kadon määrän kasvu tässä lekseemissä aiheuttaa pääosin myös taulukossa 3 havaitun kadon määrän kasvun muotoryhmässä DIFTONGI. 2010-luvulla katoa esiintyy kuitenkin myös kahdessa muussa muotoryhmän DIFTONGI lekseemissä (*vaihdella* ja *vauhdikas*). Sen sijaan muotoryhmän KOHDA- kaikki katoesiintymät yhtä lukuun ottamatta esiintyvät lekseemissä *kohda-* koko aineistossa. Vastaavasti muotoryhmän YHDESSÄ

5. Nämä tapaukset on tilastoitu samaan lekseemiin kuin muutkin *yhdeksän*-sanon esiintymät.

katoesiintymistä valtaosa esiintyy nimenomaan lekseemissä *yhdessä* (36 esiintymää 39:stä). Niinpä näiden kolmen muotoryhmän sijaan olisi kenties mielekkäämpää tarkastella yksittäisiä lekseemejä (*vaihda-*, *kohda-*, *yhdessä*). Luvun 4.2 lopussa hahmoteltua diffuusiojärjestystä onkin syytä tarkentaa niin, että muotoryhmän DIFTONGI sijaan esitetään vain lekseemi *vaihda-*. Asetelmassa 2 on esitetty tarkennettu diffuusiojärjestys.

**Asetelma 2.**

Katomuotojen tarkennettu morfologinen diffuusio.

1970-luku	1990-luku	2010-luku	orastava
YHDEKSÄN	DA-VERBIT	MAHDO-	<i>vaihda-</i>
A-VERBIT			KAKSITAVUISET NOMINIT
YKSI			

On korostettava, että kyseinen diffuusiojärjestys on aineistosta tehty tulkinta, joka perustuu muotoryhmien ja lekseemien kokonaismääriin sekä muotoryhmien välisiin korrelaatioihin. Esimerkiksi lekseemit *kohda-* ja *yhdessä* esiintyvät aineistossa melko paljon katomuodoissa, mutta koska kato ei etene niissä lineaarisesti, niitä ei ole esitetty diffuusiojärjestyksessä. Kuvion 2 perusteella lekseemi *yhdessä* yleistyy katomuodoissa 1990-luvulla ja *kohda-* puolestaan 2010-luvulla.

## 5 Lopuksi

Tässä tutkimuksessa olemme tarkastelleet katomuotojen esiintymistä yleiskielen *hd*-yhtymässä Helsingin puhekielessä ja mallintaneet kadon etenemistä muotoryhmien ja lekseemien välillä kolmen aikapisteen reaaliaika-aineistossa. On selvää, että kato on yleistynyt huomattavasti Helsingin puhekielessä 1970-luvulta 2010-luvulle. 1990-luvulla syntyneiden puhujien kohortin (kohortti 5) mediaaniarvo kadon osuudelle on jopa 84 prosenttia. Pelkästään tätä yleiskuvaa tarkastelemalla voitaisiin tulkita, että kato on levinnyt huomattavan pitkälle Helsingin puhekielessä ja näennäisaikannusteen mukaan lopulta korvaa *hd*-yhtymän säilyttäneet muodot. Tässä tutkimuksessa olemme kuitenkin tarkastelleet myös katomuotojen etenemistä. Katomuodot olivat jo 1970-luvulla hyvin yleisiä tietyissä hyvin frekventeissä lekseemeissä (*kahdeksan*, *yhdeksän*, *kahde-*, *yhde-*, *lähde-*, *tehdä*, *ehdi-*), ja katomuotojen osuus on näissä lekseemeissä entisestään kasvanut. Tämä selittää valtaosan katomuotojen yleistymisestä. Havaitimme kuitenkin, että katomuodossa esiintyviä lekseemejä tulee lisää niin 1990- kuin 2010-luvuillakin. Näin kato siis etenee myös uusiin lekseemeihin, eli leksikaalista diffuusiota tapahtuu.

Lekseemejä muotoryhmiksi yhdistelemällä havaitsimme myös morfologista diffuusiota. Muotoryhmien kokonaismäärien muutoksen ja korrelaatioanalyysin perusteella hahmottelimme diffuusiojärjestyksen katomuotojen etenemiselle. Diffuusiojärjestyksessä otimme huomioon niin leksikaalisen kuin morfologisen diffuusion tasot.

Kuten Mielikäinen (1995: 331) toteaa, diffuusiomalli on ennen kaikkea kuvaus-tapa. Se ei pyri eikä pysty selittämään muutoksia, vain kuvaamaan niitä. Koska tässä

tutkimuksessa on tietoisesti keskitytty diffuusion kuvaamiseen, itse muutoksen syyt ovat jääneet käsittelemättä. Myös tämän tutkimuksen tuloksista voidaan kuitenkin saada jonkinlainen kuva siitä, mitkä seikat saattavat vaikuttaa katomuotojen etenemiseen. Kuten lukuisissa tutkimuksissa ennen tätä on havaittu, frekvenssillä näyttää olevan roolinsa myös tämän muutoksen etenemisessä (esim. Lehtimäki 1983: 38; Suihkonen 1992: 286; Bybee 2002), sillä kaikkein eniten katomuodoissa esiintyvät lekseemit ovat myös hyvin taajaan esiintyviä. Jotkin seikat näyttävät kuitenkin hidastavan katomuotojen etenemistä. Esimerkiksi *hd*-yhtymän sijoittuminen konsonantivartaloon (*tehdä*) vokaalivartalon (*lähde*-) sijaan tai esiintyminen painottoman tavun (*pyörähdä*-) tai pitkän vokaalin (*hiihdä*-) jäljessä näyttävät vähintään hidastavan katoa. Selvin esimerkki tästä on juuri verbi *tehdä*, joka on koko aineiston yleisin lekseemi, mutta kadon osuus on siinä kuitenkin pienempi kuin lekseemeissä *yhdeksän*, *kahdeksan*, *yhde*-, *kahde*- ja *lähde*-. Oletimme myös, että sanojen merkityksellä olisi vaikutusta katomuotojen esiintymiseen. Esimerkiksi LAHTI- JA LEHTONIMET esiintyvät katomuodoissa vähemmän kuin KAKSITAVUISET NOMINIT, joihin myös *lahti* ja *lehto* kuuluisivat yleisniminä. Näiden sanojen esiintyminen erisnimissä näyttää siis vaikuttavan kadon todennäköisyyteen. Tarkempi katomuotojen etenemisen syiden tarkastelu jää kuitenkin jatkotutkimuksen varaan.

Niin ikään jatkotutkimukselle mahdollinen kohde olisi diffuusion tarkastelu esimerkiksi sijamuotojen osalta (vrt. esim. Suihkonen 1992). Tässä tutkimuksessa sijamuotojen (tai verbeissä aikamuotojen) välille ei ole tehty eroa, vaan lekseemejä on tarkasteltu sellaisenaan sekä osana muotoryhmiä. On hyvin mahdollista, että katomuotojen etenemisessä voi olla eroa myös sijamuotojen välillä. Viimeiseksi on muistettava, että tässä tutkimuksessa on tarkasteltu vain Helsingin puhekieltä. Olisi kiinnostavaa tarkastella katomuotojen etenemistä myös muualla länsimurteiden alueella, jotta voidaan selvittää, koskevatko tässä esitellyt diffuusiojärjestykset vain Helsinkiä vai ovatko ne havaittavissa yleisemminkin.

## Lähteet

### Aineistolähteet

Helpuhe 2014 = Helsingin puhekielen pitkittäiskorpus (1970, 1990, 2010). 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

BYBEE, JOAN 2002: Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. – *Language Variation and Change* 14 (3) s. 261–290. <https://doi.org/10.1017/S0954394502143018>.

CEDERGREN, HENRIETTA – SANKOFF, DAVID 1974: Variable rules. Performance as a statisti-

- cal reflection of competence. – *Language* 50 (2) s. 333–355. <https://doi.org/10.2307/412441>.
- HOLM, STURE 1979: A simple sequentially rejective multiple test procedure. – *Scandinavian Journal of Statistics* 6 (2) s. 65–70.
- IKÄHEIMONEN, SUVI 2012: *Liperin murre vuosina 1990 ja 2000. Idiolekteittaisten muutosten tarkastelua reaaliajassa*. Suomen kielen pro gradu -tutkielma. Itä-Suomen yliopisto.
- JONNINEN-NIILEKSELÄ, KAIJA 1982: *Tampereen puhekieli tutkimuskohteena*. Folia Fennistica & Linguistica. Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitoksen julkaisu 6. Tampere: Tampereen yliopisto.
- KALLIO, PETRI 2013: Hämäläisen *l*:n ajoituksesta. – *Virittäjä* 117 (2) s. 266–271.
- KETTUNEN, LAURI 1930: *Suomen murteet II. Murrealueet*. SKST 188. Helsinki: SKS.
- KUPARINEN, OLLI 2018: Infinitiivien variaatio ja muutos Helsingissä. – *Virittäjä* 122 (1) s. 29–51. <https://doi.org/10.23982/vir.65310>.
- KUPARINEN, OLLI – MUSTANOJA, LIISA – PELTONEN, JAAKKO – SANTAHARJU, JENNI – LEINO, UNNI 2019: Muutosmallit kolmen aikapisteen pitkittäisaineiston valossa. – *Sananjalka* 61 s. 30–56. <https://doi.org/10.30673/sja.80056>.
- KUPARINEN, OLLI – PELTONEN, JAAKKO – MUSTANOJA, LIISA – LEINO, UNNI – SANTAHARJU, JENNI 2021: Lects in Helsinki Finnish. A probabilistic component modeling approach. – *Language Variation and Change* 33 (1) s. 1–26. <https://doi.org/10.1017/S0954394521000041>.
- KURKI, TOMMI 2005: *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen*. SKST 1036. Helsinki: SKS.
- LAASANEN, MIKKO 2016: Näennäisaikametodin taustaoletusten teoreettista tarkastelua. – *Virittäjä* 120 (1) s. 57–83.
- LABOV, WILLIAM 1994: *Principles of linguistic change. Volume 1. Internal factors*. Oxford: Blackwell.
- LAPPALAINEN, HANNA – MUSTANOJA, LIISA – O'DELL, MICHAEL 2019: Miten ja milloin yksilön kieli muuttuu? – *Virittäjä* 123 (4) s. 550–581. <https://doi.org/10.23982/vir.67808>.
- LEHTIMÄKI, PEKKA 1983: Sekaidiolektien tutkimuksesta. – *Virittäjä* 87 (1) s. 22–40.
- LEHTINEN, TAPANI 2007: *Kielen vuosituhanneet. Suomen kielen kehitys kantaauralista varhais-suomeen*. Tietolipas 215. Helsinki: SKS.
- NAHKOLA, KARI 1987: *Yleisgeminaatio. Ääntenmuutoksen synty ja vaiheet kielisysteemissä erityisesti Tampereen seudun hämäläismurteiden kannalta*. SKST 457. Helsinki: SKS.
- NAHKOLA, KARI – SAANILAHTI, MARJA 2001: *Kymmenen vuotta myöhemmin. Seuruututkimus kielenmuutosten etenemisestä yksilön kielessä ja puheyhteisössä*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 68.
- NEYMAN, JERZY – PEARSON, EGON S. 1933: The testing of statistical hypotheses in relation to probabilities a priori. – *Mathematical Proceedings of the Cambridge Philosophical Society* 29 s. 492–510.
- NUOLIJÄRVI, PIRKKO 1986: *Kolmannen sukupolven kieli. Helsinkiin muuttaneiden suurten ikäluokkien eteläpohjalaisten ja pohjoissavolaisten kielellinen sopeutuminen*. Tietolipas 100. Helsinki: SKS.
- MANTILA, HARRI 2004: Murre ja identiteetti. – *Virittäjä* 108 (3) s. 322–346.
- MIELIKÄINEN, AILA 1995: Morfologinen diffuusio. – *Virittäjä* 99 (3) s. 321–334.
- MUSTANOJA, LIISA 2011. *Idiolekti ja sen muuttuminen. Reaaliaikatutkimus Tampereen puhekielestä*. Acta Universitatis Tamperensis 1605. Tampere: Tampereen yliopisto. <https://urn.fi/urn:isbn:978-951-44-8417-9>.
- MUSTANOJA, LIISA – O'DELL, MICHAEL 2007: Suomen *d* ja *r* sosiofoneettisessa kentässä.

– *Virittäjä* 111 (1) s. 56–67.

- OLIVEIRA, MARCO ANTONIA DE 1991: The neogrammarian controversy revisited. – *International Journal of the Sociology of Language* 89 s. 93–105.
- PAUNONEN, HEIKKI 1976: Kotikielen Seura 1876–1976. – *Virittäjä* 80 (3–4) s. 310–432.
- 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 kieliyhteisönä* s. 13–99. Helsinki: Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.
- RAPOLA, MARTTI 1966: *Suomen kielen äännehistorian luennot*. SKST 283. Helsinki: SKS.
- RÄSÄNEN, MAARET – PALANDER, MARJATTA 2015: Kansandialektologinen testi murrepiirteiden keskinäisestä murteellisuusjärjestyksestä. – *Virittäjä* 119 (1) s. 4–33.
- SKS = Suomalaisen Kirjallisuuden Seura
- SKST = Suomalaisen Kirjallisuuden Seuran toimituksia
- SUIHKONEN, PAAVO 1992: *Klusiilien vaihtelusuhteet Kala- ja Lestijokilaakson murteissa*. SKST 577. Helsinki: SKS.
- THOMPSON, KEN 1968: Regular expression search algorithm. – *Communications of the ACM* 11 (6) s. 419–422. <https://doi.org/10.1145/363347.363387>.
- VIRTARANTA, PERTTI 1958: *Pääpainollisen tavun jälkeisen soinnillisen dentaalispirantin edustus suomen murteissa*. SKST 252. Helsinki: SKS.
- WAGNER, SUZANNE – SANKOFF, GILLIAN 2011: Age grading in the Montréal French inflected future. – *Language Variation and Change* 23 (3) s. 275–313. <https://doi.org/10.1017/S0954394511000111>.
- WANG, WILLIAM S-Y. 1969: Competing changes as a cause of residue. – *Language* 45 (1) s. 9–25.
- WILCOXON, FRANK 1945: Individual comparisons by ranking methods. – *Biometrics Bulletin* 1 s. 80–83. <https://doi.org/10.2307/3001968>.

## The advancement of /d/ deletion after /h/ in Helsinki Finnish

The article examines the linguistic variable /hd/ (*kahdeksan*) in Helsinki Finnish and models the advancement of /d/ deletion in the variable (*kaheksan*) in real time. The article uses a subset of the longitudinal corpus of Finnish spoken in Helsinki, which includes recordings dating from three points in time: the 1970s, the 1990s and the 2010s. The article presents an overview of the variation and change in the variable between these three time points, and simultaneously examines differences between lexemes and lexeme groups. Deletion is assumed to follow a diffusion pattern, slowly advancing from one morphological group to another.

The article shows that such deletion becomes more common over time. The observed change is mostly due to a few high-frequency lexemes, but deletion advances to new lexemes as well. The diffusion is followed both in morphological groups and in lexemes. Methodologically the study uses statistical correlation tests and Wilcoxon signed-rank tests. The results indicate that deletion advances according to the diffusion pattern in Helsinki Finnish even though deletion is most common in very frequent words.

## Katomuotojen eteneminen yleiskielen *hd*-yhtymässä Helsingin puhekielessä

Artikkeli tarkastelee yleiskielen *hd*-yhtymän (*kahdeksan*) esiintymiä Helsingin puhekielessä ja mallintaa katomuotojen (*kaheksan*) etenemistä reaaliajassa. Artikkelin aineistona käytetään osaa Helsingin puhekielen pitkittäiskorpuksesta, joka kattaa haastatteluja kolmesta eri aikapisteestä: 1970-, 1990- ja 2010-luvuilta. Artikkelin esittää kokonaiskuvan *hd*-yhtymän variaatiosta ja muutoksesta näiden kolmen aikapisteen välillä ja tarkastelee myös lekseemikohtaisia eroja. Katomuotojen oletetaan etenevän diffuusiomallin mukaisesti hitaasti muotoryhmästä toiseen.

Artikkeli osoittaa, että katomuodot yleistyvät Helsingin puhekielessä haastattelu-kierrokselta toiselle. Valtaosa havaittavasta muutoksesta on peräisin muutamista keskeisistä lekseemeistä, mutta kato etenee myös uusiin lekseemeihin. Kadon etenemistä seurataan niin morfologisten muotoryhmien kuin yksittäisten lekseemien tasolla. Tilastollisina menetelminä hyödynnetään korrelaatiotestejä ja Wilcoxonin testejä. Tulokset osoittavat kadon etenevän Helsingin puhekielessä diffuusiomallin mukaisesti muotoryhmästä toiseen, mutta kato on kaikkein yleisintä hyvin yleisissä sanoissa.

Kirjoittajien yhteystiedot:

etunimi.sukunimi@helsinki.fi (Olli Kuparinen)

etunimi.leppanen@helsinki.fi (Jenni Santaharju)

etunimi.sukunimi@tuni.fi (Unni Leino, Liisa Mustanoja, Jaakko Peltonen)

Olli Kuparinen on kieliteknologian tutkijatohtori Helsingin yliopistossa. Artikkelin kirjoittamisen aikaan hän työskenteli väitöskirjatutkijana Tampereen yliopistossa. Jenni Santaharju on evoluutiobiologian apurahatutkija Helsingin yliopistossa. Unni Leino on suomen kielen yliopistonlehtori Tampereen yliopistossa. Liisa Mustanoja on suomen kielen yliopistonlehtori Tampereen yliopistossa. Jaakko Peltonen on tilastotieteen ja data-analyysin professori Tampereen yliopistossa.