

# Suomen kielen hallinta ja sen kehitys

*Peruskoululaiset ja nuoret aikuiset*

*Toimittaneet*

ANNELI PAJUNEN

MARI HONKO



Suomen kielen hallinta  
ja sen kehitys



# Suomen kielen hallinta ja sen kehitys

*Peruskoululaiset ja nuoret aikuiset*

*Toimittaneet*

ANNELI PAJUNEN JA MARI HONKO



## SUOMALAISEN KIRJALLISUUDEN SEURAN TOIMITUKSIA 1472

Teos on Suomalaisen Kirjallisuuden Seuran nimeämien asiantuntijoiden tarkastama.



VERTAISARVIOITU  
KOLLEGIALT GRANSKAD  
PEER-REVIEWED  
[www.tsv.fi/tunnus](http://www.tsv.fi/tunnus)

© 2021 Anneli Pajunen, Mari Honko ja SKS

Lisenssi CC BY-NC-ND 4.0 International

Kannen suunnittelu: Timo Numminen

Taitto: Maija Räisänen

EPUB: Tero Salmén

ISBN 978-951-858-407-3 (nid.)

ISBN 978-951-858-408-0 (EPUB)

ISBN 978-951-858-409-7 (PDF)

ISSN 0355-1768 (nid.)

ISSN 2670-2401 (verkkojulkaisut)

DOI <https://doi.org/10.21435/skst.1472>

Teos on lisensoitu Creative Commons CC BY-NC-ND 4.0 International lisenssillä. Tutustu lisenssiin englanniksi osoitteessa <https://creativecommons.org/licenses/by-nc-nd/4.0/> tai suomeksi osoitteessa <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.fi>.



Teos on avoimesti saatavissa osoitteessa <https://doi.org/10.21435/skst.1472> tai lukemalla tämä QR-koodi mobiililaitteella.



Hansaprint Oy, Turenki 2021

# Sisällys

Esipuhe 7

MYÖHEMPI KIELENKEHITYS: SANATIEDOSTA RAKENNE- JA  
TEKSTITIETOON 8

*Anneli Pajunen ja Mari Honko*

## I Sanat ja merkitykset

UNELMIENI PÄIVÄ -SANASTO KEHITYKSELLISESTÄ NÄKÖKULMASTA 61

*Anneli Pajunen ja Mari Honko*

LIIKETILANTEEN VÄYLÄN JA SUUNNAN KIELENTÄMINEN  
SUOMESSA 108

*Emilia Tuuri*

REFERENTIAALISET MÄÄRÄISET ILMAUKSET JA NIIDEN SUHDE  
KERTOMUSRAKENTEeseen ALAKOULUIÄSSÄ 155

*Anna-Kaisa Kouvo, Mari Honko ja Anneli Pajunen*

MONIÄÄNISYYDEN ILMAISUN KEHITTYMINEN PERUSKOULUN  
PÄÄTTÖVAIHEIDEN KIRJOITELMISSA 199

*Anneli Niinimäki*

## II Muodot ja rakenteet

MENEILLÄÄN OLEVA KIELENMUUTOS PUHUTUSTA KIELESTÄ  
KIRJOITETTUUN JA SUOMEN POSSESSIIVISUFFIKSI 249

*Mikko Laasanen ja Anneli Pajunen*

KIRJOITTAMISEN TARKKUUDEN KEHITTYMINEN JA YHTEYS  
KIRJOITTAJAN KOKONAISKIELITAITOON ALAKOULUIÄSSÄ 302

*Mari Honko*

KIELEN RAKENTEEN HALLINTA PERUSKOULULAISILLA JA NUORILLA  
AIKUISILLA 367

*Anneli Pajunen ja Seppo Vainio*

SYNTAKTISTA RAKENNETTA TIIVISTÄVÄT JA KOMPLEKSISTAVAT TEKIJÄT  
PERUSKOULULAISILLA JA NUORILLA AIKUISILLA 421

*Anneli Pajunen ja Seppo Vainio*

Kirjoittajat 480

Abstract 481

Hakemisto 482

## Esipuhe

Tutkimus peruskoululaisten ja nuorten aikuisten kielenhallinnasta ja hallinnan kehityksestä keskittyy siihen varhaislapsuutta myöhempään aikaan, jolloin kielelliset tiedot ja taidot monipuolistuvat ja rikastuvat. Kielenhallinnan kehitys on elinikäistä, mutta aluksi kyse on kasvamisesta kirjoitetun eli niin sanotun akateemisen kielen hallintaan tavoitteena sellaiset kielelliset valmiudet, joita tarvitaan ammatillisessa koulutuksessa ja yhteiskunnassa toimimisessa. Keskitymme tässä julkaisussa peruskoululaisiin ja nuoriin aikuisiin. Julkaisu kuuluu tutkimushankkeeseen ”Suomen kielen myöhempi kehitys: kouluikäiset ja nuoret aikuiset”. Hanketta on rahoittanut Koneen Säätiö vuosina 2013–2015 sekä Suomen Kulttuurirahasto lukuvuonna 2013–2014. Hankkeen toimintaan on osallistunut joukko tutkijoita Tampereen, Turun ja Jyväskylän yliopistoista, ja hyvin monet opiskelijat ovat tehneet opinnäytteensä hankkeen yhteydessä. Kiitämme kaikkia hankkeen toimintaan ja tämän julkaisun ilmestymiseen osallistuneita tahoja, mutta ennen kaikkea haluamme kiittää kaikkia niitä oppilaita, jotka ovat luovuttaneet hankkeelle aineistoa, kaikkia niitä opettajia, jotka ovat päästäneet meidät tunneilleen, ja kaikkia niitä rehtoreita ja muita kouluviranomaisia, jotka ovat antaneet tälle toiminnalle luvan. Alaikäisten osalta kiitämme myös lapsensa osallistumiseen luvan antaneita vanhempia. Suurimman kiitoksen ansaitsevat aineistoa keränneet ja sitä loputtoman tarkkaavaisesti analysoineet Hanna-Riikka Hodges ja Soile Itkonen, Maija Muhonen, Kristiina Viitala ja Annukka Yli-Paavola.

Heinäkuussa 2021

Anneli Pajunen ja Mari Honko



# Myöhempi kielenkehitys: sanatiedosta rakenne- ja tekstitietoon

Anneli Pajunen

 <https://orcid.org/0000-0001-8597-7491>

Mari Honko

 <https://orcid.org/0000-0002-3240-622X>

## Johdanto

Tämän teoksen aiheena on kielenhallinnan kehittyminen eli kieliyhteisölle yhteisen kielellisen tiedon vähittäinen lisääntyminen varhaislapsuuden jälkeen niin sanotun myöhemmän kielenkehityksen aikana (*later language development*, ks. esimerkiksi Berman 2004; Nippold 2006). Lukujen päähuomio on suomea ensikielenä oppineissa, äidin-kielisissä suomenpuhujissa.

Kielenhallinta edellyttää tiedostamatonta tai tietoista tietoa kielestä: intuitiivista tietoa kielen fonologisesta, morfologisesta ja syntaktisesta rakenteesta, näiden tasojen käyttökonteksteista ja käytön pragmaattisista ehdoista sekä lisäksi sanastosta. Riittävä kielenhallinta on edellytys kielenkäyttötaidoille, kommunikaatiolle sekä lukemaan ja kirjoittamaan oppimiselle. Kielenhallinnan kehitys alkaa puhutusta kielestä, mutta se ei ole riippuvainen puhutun ja kirjoitetun kielen modaaliteettieroista eikä varieteettieroista, vaan lapsi ja nuori oppii niitä äidinkieltensä merkityksiä,

muotoja ja käyttötapoja, joita hänellä on edellytys oppia ja joille hän ympäristössään altistuu. Kielenhallinta kattaa siis yksilön kaiken kielitiedon, se ei rajaudu kirjoitettuun yleiskieleen, vaikka hallinnan kehitystä on ehkä helpoin tutkia kirjoitetun kielen kontekstissa ja vaikka juuri myöhemmän kielenkehityksen vaiheissa hallinta laajenee puhutusta kielen kirjalliseen käyttöön. Kielenhallinta ja sen kehitys ei edellytä preskriptiivisen kieliopin normiston tai termien tuntemusta vaan kielitieto on puheyhteisölle yhteistä tietoa. Tähän kielen ydinalueeseen soveltuu yksimieli-syyperiaate (*consensus principle*): yleisimmät sanat ja rakenteet hallitaan varmasti mutta muutoksenalaisten ja harvinaisten ilmiöiden osalta epävarmuus kasvaa (ks. Labov 1975: 8–9, 30–32; Itkonen 1978: 151–154; Hulstijn 2015; 2019). Kielenhallinta kehittyy vielä aikuisiässä mutta lapsilla ja nuorilla aikuistasoon pääsy kestää pitkään. Oman tutkimuksemme tavoitteena on ollut tehdä näkyväksi se, mikä on keskeistä yhteistä kielellistä tietoa peruskoulun eri vaiheissa oleville ja miten se iän mukaan kasvaa ja kehittyy erityisesti kielessä harvinaisempien ilmiöiden osalta. Oletus on, että harvinaisten ilmiöiden hallinnassa hajonta on suurta.

Pieni lapsi oppii kielenkehityksen alkuvaiheissa paitsi äidinkielen (ensikielensä tai -kieltensä) äännerakenteen myös arkipäivän perussanas-toa, sanontoja ja fraaseja sekä morfologian ja syntaksin yksinkertaisia rakenteita. Kehitys toisin sanoen alkaa varmasta tiedosta (tai siitä tiedon osasta, joka ensimmäisenä osoittautuu varmaksi). Puhe kohdistuu aluksi pääasiassa konkreettisiin, läsnä oleviin asioihin, ja yksittäisten ilmausten viittausala voi olla huomattavasti suppeampi tai laveampi kuin aikuisten kielessä (Hoff 2005). Kolmen neljän ikävuoden vaiheilla lapsi oppii systemaattisemmin irrottamaan puheensa sisällön puhehet-kestä eli kertomaan myös asioista, jotka eivät ole läsnä tai totta (Doherty 2009; Suomessa esimerkiksi Heimo 2003). Myöhemmän kielenkehityk-sen alkuvaiheissa eli noin viisivuotiaana lapsen elämässä alkaa voimakas kielellisen kehityksen vaihe. Hän oppii kertomaan suullisia tarinoita ja huomioimaan kuulijoita, ja kaikkiaan kielelliset tiedot ja taidot moni-puolistuvat ja rikastuvat monella tavalla. Lapsi oppii harvinaisempaa sanastoa, useampisanaisia ilmauksia sekä monimutkaisia rakenteita ja oppii hallitsemaan kielenkäyttöä erilaisissa tilanteissa. Tarkastelemme tässä teoksessa sekä sanaston että kompleksisten semanttisten ja raken-

teellisten ilmiöiden kielentämisen kehitystä eri-ikäisillä peruskoululaisilla ja nuorilla aikuisilla. Luku- ja kirjoitustaidon saavuttaminen vauhdittaa lasten kielenkehitystä mutta ne eivät ole kehityksen edellytyksiä.

Myöhemmän kielenkehityksen vaihe on todellisuudessa jatkumo (*life span development*) mutta sitä tarkastellaan yleensä vaiheittain siten, että erotellaan esikoulun ja ensiluokkien vaihe, alakoulun keski- ja loppuvaihe, yläkouluvaihe ja sen päätyminen sekä varhaisaikuisuus, josta edetään elinkaaren mukaisesti vanhuuteen asti (ks. esimerkiksi Nippold 2006). Aiemmin ajateltiin, että kyse on ikään kuin kehityskaaresta, jonka alku ja loppu ovat heikomman osaamisen vaihetta (ks. esimerkiksi Jakobson 1968). Ajateltiin toisin sanoen, että vanheneminen vaikuttaisi kielitietoa ja kielenkäyttötaitoja heikentävästi. Nykyisin ajatellaan ennemmin, että suotuisasti vanhentuneen eli esimerkiksi kielelliseen toimintaan vaikuttavilta sairauksilta säästyneen vanhan ihmisen kielenhallinta on nuoremman veroinen tai häntä parempikin, vaikka sanat eivät aina löydy yhtä nopeasti kuin ennen ja vaikka tarinasta tulee välillä syrjähdelyä. Samoin jos aiemmin ajateltiin, että nuori aikuinen on kielellisesti kuin valmis paketti, nykyisin tiedetään, että kielenkehitys ja -hallinta varioi huomattavasti jokaisessa ikävaiheessa. Esimerkiksi nuoret aikuiset ovat vielä vaiheessa, jossa kompleksisten rakenteiden hallinnassa on puutteita, rekisterit ovat melko kapeat ja sanavarasto parhaimmillaankin 20–30 prosenttiyksikköä pienempi kuin vanhimmilla ihmisillä (ks. esimerkiksi Kemper & Sumner 2001; Kavé & Halamish 2015; Keulers, Stevens, Mandela & Brysbaert 2015).

Tässä kirjassa keskitymme kielellisen kehityskaaren keskivaiheille, erityisesti 12-vuotiaista nuoriin aikuisiin, ja pyrimme esittelemään peruskouluikäisten Suomessa syntyneiden ja suomea äidinkielenään puhuvien lasten ja nuorten keskimääräistä kielenhallintaa ja toisaalta näyttämään ikäkausittaisen osaamisen vaihtelun. Nuoret aikuiset ovat sekä vertailukohtana että erillistutkimuksen kohteena. Oletus on, että maahanmuuttajataustaiset peruskoulua Suomessa käyneet lapset ja nuoret tavoittelevat samankaltaista osaamista. Osaamisen erot suhteutuvat Suomessa asumisen pituuteen sekä kielellisten kontaktien määrään ja laatuun mutta myös lukuisiin yksilöllisiin ja sosiaalisiin tekijöihin (ks. tarkemmin Honko 2013; Ruuska 2020).

Lähtökohtanamme on, että myöhempi kielenkehitys alkaa puhutun kielen pohjalta ja että puhutun ja kirjoitetun kielen ydin on sama mutta suhde muuttuu vähitellen vuorovaikutukselliseksi. Oletamme myös, että kielenhallinnassa ei ole suuria sukupuolieroja, vaikka niitä voi olla kielenkäyttötaidoissa, kuten lukemisessa ja kirjoittamisessa. Tutkimus perustuu tutkimushankkeeseen nimeltä ”Suomen kielen myöhempi kehitys: lapset, nuoret ja nuoret aikuiset”, jonka tuloksia on kirjallisuudessa esitelty jo jonkin aikaa. Hankkeen julkaisut etenevät ikään kuin elinkaaren mukaisesti: Alkuvaiheessa esillä oli alakouluvaihe (ks. Pajunen 2012; Honko 2013). Peruskoululaisista ja nuorista aikuisista on myöhemmältä 2010-luvulta useita julkaisuja (Pajunen, Itkonen & Vainio 2015; 2016; Vainio, Pajunen & Häikiö 2019; Pajunen & Itkonen 2019; Laasanen, Pajunen & Häikiö, arvioitavana) ja käsillä oleva kokoomateos keskittyy kokonaan tähän vaiheeseen. Vireillä hankkeessa on vanhempien ihmisten kielenhallinnan tutkimus.

Tässä yhteydessä keskitymme toisaalta semanttisen tason hallintaan (I osio), toisaalta muotojen hallintaan (II osio) peruskoululaisten ja nuorten aikuisten näkökulmasta. Sanatieto muodostaa kielenhallinnan perustan, josta kehitys alkaa ja josta se myöhemmän kielenkehityksen aikana huomattavasti täsmentyy ja tarkentuu määrällisesti ja laadullisesti myös vaikeiden semanttisten ilmiöiden kielentämiseen. Muodot tarjoavat ne ainekset, joilla merkityksiä kielennetään muiden ymmärrettäviksi ilmauksiksi ja teksteiksi. Näkökulmat luonnollisesti limittyvät, vaikka lähtökohta on – merkityksestä muotoon ja muodosta merkitykseen – vastakkainen. Metodisesti osiot I ja II eroavat selvästi: I osiossa analyysi alkaa semanttisten ilmiöiden kartoittamisesta, II osiossa lähtökohtana ovat muodot.

Sekä sanatieto ja kyky kielentää merkityksiä että tieto muodoista kehittyvät vähittäin. Tässä teoksessa pyrimme isoilla koehenkilömäärillä osoittamaan erilaisia kielellisen osaamisen ikäkausimediaaneja ja osaamisen vaihtelua. Tutkimus on toteutettu monimetodisesti käyttäen sekä korpustutkimuksen että kokeellisen tutkimuksen metodeja. Tutkimusaineisto on kerätty kynä-paperitestein, kokeellisesti sekä kirjoittamistehävillä. Perustelemme seuraavaksi tutkimuksen perusoletuksia, nimitäin sanatiedon keskeisyyttä, puhutun kielen mallia sekä sukupuolten

kielenhallinnan samanlaisuutta ja toisaalta ikäkausivaihtelun osuutta. Sen jälkeen esittelemme tutkimuksen aineistot ja koehenkilöt sekä hankkeen tutkijoiden työnjakoa ja kirjan lukujen näkökulmia. Osiot I ja II koostuvat kumpikin neljästä eri luvusta.

Tutkimuksemme kohdistuu suomen kieleen. Teoksessa käytetty käsitteistö on kuitenkin enemmän lingvististä kuin fennististä, koska vertailututkimus on pitkälti kansainvälistä. Myös tavoiteltu lukijakunta ulottuu fennistiipiirin ulkopuolelle. Toivomme, että kirja tarjoaa uusia näkökulmia kenelle tahansa lingvistisestä tutkimuksesta sekä suomen kielen myöhemmän kehitysvaiheen tutkimisesta, oppimisesta ja opettamisesta kiinnostuneelle. Kirjan luvut sopivat hyvin yksittäinkin esimerkiksi yliopisto-opintojen kurssikirjallisuudeksi.

#### PERUSOLETUKSET KIELENHALLINNAN TUTKIMISEEN

Kielenhallinnan tutkimuksessamme perusoletuksia on ollut neljä: Kielenkehityksessä sanatiedon laatu ja määrä on keskeistä ja alussa kehityksen mallina on puhuttu kieli. Kielenhallinta ja sen kehitys ei eriydy sukupuolen mukaan, mutta ajallisesti se voi edetä eritahtisesti. Väestötasolla kielenhallinta on normaaliajakautunutta mutta varioi yksilötasolla ja kehityksellisesti. Varioinnin keskeinen taustatekijä on kielellinen älykyys, mutta siihen vaikuttaa myös muita kielenensisäisiä ja kielenulkoisia tekijöitä. Käsillä olevan teoksen luvut ja teemat ankkuroituvat näihin perusoletuksiin, ja niiden avulla voidaan hahmottaa lasten ja nuorten odotuksenmukaista kielenkehitystä ja poikkeamia siitä.

#### SANATIEDON KESKEISYYS

Kielen oppiminen on elinikäinen prosessi, joka alkaa sanojen oppimisesta ja jatkuu vanhuuteen (ks. esimerkiksi Ramscar, Hendrix, Love & Baayen 2013). Yksilön sanastollinen tieto ja taito onkin parhaimmillaan vasta iäkkäillä. Esimerkiksi laajoissa Alankomaissa toteutetuissa väestöpohjaisissa nettitesteissä yli 70-vuotiaat tunsivat vajaasta 30 000 testisanasta yli 80 prosenttia mutta 13-vuotiaat vain vähän yli puolet ja nuoret aikuiset noin 60–65 prosenttia (Keulers ja muut 2015). Suomen osalta tilanne näyttää samansuuntaiselta (Pajunen, tekeillä). Kuudesluokkalaiset tunsivat testisanoina olleista melko yleisistä liikeverbeistä

hyvin noin puolet ja yhdeksäsluokkalaiset noin 60 prosenttia. Nuorten aikuisten ryhmässä liikeverbien tuttuus varioi erityisen paljon, jos sana oli harvinainen eli sana, jonka esiintymiä on miljoonassa tekstisanassa keskimäärin alle yksi (Pajunen & Itkonen 2019). Yksilön sanavaraston dynaamisuuden vuoksi voitaisiinkin oikeastaan puhua sanavarannosta, joka karttuu ja muovautuu jatkuvasti kielenkäyttökokemuksen myötä.

Ensimmäinen perusoletuksemme on, että kielen hallinnassa ja sen kehittymisessä juuri sanatiedon laatu ja määrä on keskeistä. Siksi sanatiedon arviointi on ollut käsillä olevassa varhaislapsuuden jälkeisessä myöhemmän kielenkehityksen tutkimushankkeessa keskiössä. Sanatiedon hallintaa on hankkeessa tutkittu monella tapaa, ja nämä tutkimustulokset ovat taustoittaneet hankkeen muuta tutkimusta. Sanatiedon hallintaa on arvioitu sekä morfologisten ja semanttisten testien avulla että kielenkäytöstä (ks. Honko 2013; Pajunen, Itkonen & Vainio 2015; 2016; Vainio, Pajunen & Häikiö 2019; tekeillä); toisin sanoen termiä *sanatieto* ei pidä ymmärtää liian kapeasti. Tämän teoksen ensimmäiset neljä lukua käsittelevät sanatiedon ja merkitysten hallintaa joko sana-analyysin tai semanttis-pragmaattisten ilmiöiden kielentämisen näkökulmasta. Kyse on siis itse asiassa kyvystä ilmaista olio- ja tapahtumamerkityksiä ja laajemmin tekstuaalisia merkityksiä.

Sanatason keskeisyyden kielenhallinnassa näkee hyvin esimerkiksi suomen johdoksista, joiden käytön määrä ja leksikaalinen monimuotoisuus lisääntyvät iän mukaan lineaarisesti. Tämä on sanaston kasvun kannalta merkitsevää, koska suomessa suuri osa sanastosta on joko johdoksia tai yhdyssanoja. Johtaminen on keskeisempi ja vaikeampi ilmiö kuin yhdistäminen, koska kaikkien avointen sanaluokkien sanoja johdetaan ja koska johtaminen on rekursiivista eli toistuvaa ja osin sisäkkäistä. Lisäksi sanavartalot varioivat johtamisen yhteydessä ja johtimia, derivaatiotapoja ja derivaatioketjuja on paljon. Alakoulun päättövaiheessa nuori osaa johdoksia sanakohtaisesti mutta ei yleensä sääntöperäisesti. Hän saattaa siis tuntea sanan *käveleskellä* mutta ei yleensä osaa yhdistää muita samalla tavalla johdettuja sanoja ryhmäksi tai ei tiedä, mistä kaikesta muista sanoista samalla johtimella voidaan tuottaa samantapaista lisämerkitystä ilmaisevia sanoja. Suomen tapaisessa synteettisessä kielessä vasta nuorilla aikuisilla johtaminen on sääntöperäistä siten, että

ymmärretään kielessä esiintymättömät sanat mahdollisiksi sanoiksi, jos ne on muodostettu sääntöjä (voimakkaasti) rikkomatta (Koivisto 2013; Vainio ja muut 2019; ks. myös Berman 2007). Sääntöperäisyyden ymmärtäminen on ikään kuin avain systemaattiseen analogiseen mallintamiseen (ks. Itkonen 2005), mutta peruskoululaisia näyttää vielä sitovan enemmän ”joko oikea sana tai sitten ei” -ajattelu. Pieni lapsi osaa toki hyödyntää analogioita mutta ei kovin systemaattisesti. Johdosten käytön lisääntyminen edellyttääkin suomen monimutkaisen sanajohdon sisäistämistä ja produktiivistumista (Bertram, Laine & Virkkala 2000; Kusnetsoff 2017; Vainio ja muut 2019). Sama pätee muista derivaattorikkaista kielistä (ks. Ravid & Avidor 1998; Carlisle 2000; Štekauer, Valeira & Körtvélyessy 2012; Clark 2014). Sanojen johtamisen hallinta eroaa vielä nuorten aikuisten ja peruskoulun päättövaiheessa olevien välillä, toisin sanoen nuoret eivät hallitse derivaatiota vielä 15-vuotiaanaakaan täysin sääntötasoisesti tai laadullisesti nuorten aikuisten tapaan (Vainio ja muut 2019). Vanhempien kielenkäyttäjien osalta tutkimustieto puuttuu.

Sanatiedon hyvä hallinta ja laaja sanasto on vahvan luku- ja kirjoitustaidon edellytys (ks. esimerkiksi Carlisle 2000; Qian 2002; Nippold 2006, 25–29, 89; Milton 2009, 170–192; Schmitt 2010, 3–12; Lee 2011; suomen osalta Saarela 1997; Honko 2013; Kusnetsoff 2017). On luonnollista, että lähinnä ikäluokalle tuttua (tai kontekstista pääteltävää) sanastoa ymmärretään luettaessa ja käytetään kirjoituksessa. Luku- ja kirjoitustaito ovat myöhemmässä kielenkehityksessä keskeisiä saavutuksia, mutta ikäluokkatasoinen kielenhallinta on näiden taitojen hallintaan oton perusedellytys. Siten luku- tai kirjoitustaito eivät ole tässä tutkimuksessa varsinainen tutkimuskohde, mutta jos halutaan tukea kehittyvää luku- ja kirjoitustaitoa, on tarkoituksenmukaista seurata ja tukea myös kielenhallintaa.<sup>1</sup>

1 Lukutaitoa on Suomessa tutkittu melko paljon – kirjoitustaitoa jonkin verran – mutta pääasiassa esikoulusta koulun alkuun, ei niinkään muiden peruskouluikäisten osalta (ks. Kulju, Kauppinen, Hankala, Harjunen, Pentikäinen & Routarinne 2017; ks. myös Tuovila, Kairaluoma & Mustonen, toim. 2020). Eniten hieman vanhempien peruskoululaisten luku- ja kirjoitustaitoa koskevaa tutkimusta on tehty Opetushallituksen toimeksiannoista. Nämä tutkimukset osoittavat samoja kehityslinjoja, joita tässäkin tutkimuksessa tulee esille. Lappalaisen monet raportit (ks. esimerkiksi 2003) olivat aikanaan yhtenä tämän tutkimuksen innoittajana.

Tässä yhteydessä katsomme eri näkökulmista nimenomaan sitä, mitä lapset ja nuoret keskimäärin osaavat tietyllä ikätasolla ja miten tämä osaaminen muuttuu ja oletettavasti kehittyy ikätasolta toiselle. Lasten ja nuorten osaamista verrataan nuoriin suomenkielisiin (ylioppilastutkinnon suorittaneihin) aikuisiin, joiden jo voi olettaa hallitsevan äidinkieltään riittävällä tasolla, vaikka niin kielenkäyttötottumuksissa kuin kieli-taidossakin tiedetään esiintyvän vaihtelua (ks. esimerkiksi Malin, Sulku-nen & Laine 2013). Tutkimuksemme nuorten aikuisten ryhmä ei edusta koko ikäluokkaansa, koska se ei ole tavoitettavissa yhtä helposti kuin peruskoululaiset. Lisäksi aikuisten osallistumiseen vaikuttaa tehtävän luonne ja sen vaatima aika ja vaivannäkö. Kirjoittaminen on tehtävä-tyyppinä vaativa ja karkotti tutkimukseemme jo muuten suostuneita. Aikuisryhmä edustaa tutkimuksessa lasten ja nuorten tavoitetasoa eli sitä, mihin he voivat pyrkiä ollakseen akateemisilta taidoiltaan itsenäisiä toimijoita.

Käsillä olevassa teoksessa sanatietoa ja merkitysten kiellentämistä käsittelee I osion neljä lukua. Niistä ensimmäinen tarkastelee sanaston diversiteettiä, yleisyytasoja ja semanttisia luokkia samasta aiheesta kirjoitetuissa kirjoitelmissa, toinen ja kolmas analysoivat semanttisesti sanatasoa kompleksisempien ilmiöiden kiellentämistä ja neljäs moniää-nisyyden ilmaisun kehittymistä tekstitasolla. Lisäksi sanatietoa käsitel-lään sanahahmon ja sanarakenteen näkökulmista II osiossa. Hankkeen sanatietoa koskevia aiempia julkaisuja ovat muun muassa Honko 2013; Pajunen ja muut 2015; 2016; Honko, Jarvis & Vainio 2019; Vainio ja muut 2019; Laasanen, Pajunen & Häikiö, arvioitavana; sekä Vainio, Pajunen & Häikiö, tekeillä.

#### PUHUTUN KIELEN MALLI KIRJOITETULLE

Toinen perusoletuksemme on, että lapsen kielenkehitys perustuu alku-vaiheessaan puhuttuun kieleen ja suulliseen syötökseen. Puhuttu kieli on lauserakenteeltaan yleensä melko yksinkertaista ja virkkeet koostuvat useammin ketjulauseista kuin alisteisista ja hierarkkisista rakenteista (ks. Pajunen & Palomäki 1984). Lapsi käyttää vielä myöhemmän kielen-kehityksen alkuvaiheissa puhutun kielen mallin mukaisia perusraken-



teita – teksti etenee peruslauseilla<sup>2</sup> – ja yhdistää niitä konnektiiveilla (ks. esimerkiksi Tolchinsky 2004, 236; Berman 2007, 357; Hulstijn 2015). Lapsi osaa jo kertoa tarinoita (ks. Suvanto 2012), mutta vain harva kirjoittaa niitä. Lapsen sanasto on yleistä ja konkreettista, ja lapsi saattaa jo ymmärtää konventionaalisia metaforia mutta alkaa tuottaa niitä itse vasta teini-iässä (vrt. Tolchinsky 2004, 238). Koulua aloittava lapsi on aika tyypillisesti kirjaimellisen merkityksen vaiheessa ja kaikki kuvainnollinen kielenkäyttö on vähissä. Tässä vaiheessa lasten väliset kielelliset kehityserot ovat hyvin suuret.

Myöhemmän kielenkehityksen aikana sekä puhutun että kirjoitetun kielen hallinta kehittyy. Lapsi oppii ymmärtämään puhutun ja kirjoitetun kielen modaliteettieroja ja oppii ottamaan erot huomioon tekstejä tuottaessaan. Lisäksi opitaan muut tekstityyppi- ja rekisterierot. Sekä sanaston että syntaksin kehitys on huomattavaa (ks. esimerkiksi Berman 2004; 2007; Nippold 2006). Syntaktisesti kehitys näkyy harvinaisempien rakenteiden haltuunotosta sekä lauseiden, lauserakenteiden ja tekstijaksojen kompleksistumisesta. Rakenteiden muutosta voi mitata erilaisilla tiheysmittareilla, koska muutokset ovat usein myös määrällisiä. Tekstitasolla kehitys näkyy esimerkiksi suunnittelun lisääntymisenä, lukijan huomioimisena, ajan ja paikan aiempaa parempana ankkuroimisena sekä viittausketjujen pitenemisenä. Myöhemmän kielenkehityksen aikana opitaan tekstityypit ja opitaan myös yhdistelemään niitä. Lapsi ainoastaan linkittää tapahtumia, mutta nuorella kertovaan tekstiin alkaa jo ilmestyä kuvailevia ja arvioivia jaksoja (ks. Tolchinsky 2004, 240).

Tässä tutkimuksessa kielellisen rakenteen kehittymistä yksinkertaisesta monimutkaisempaan tarkastellaan kirjoitelma-aineistolla. Aineistonkeruun yhteydessä kirjoitelmatehtävä ohjeistettiin tekstityyppejä tai -lajia määrittelemättä ”kirjoita päivästä – ” (eikä esimerkiksi ”kerro päivästä”). Tuloksena oli iän mukaan pitenevä kirjoitelma yhdestä sanasta parhaillaan kertomusrakenteen (Labov & Waletzky 1967) mukaiseen kirjoitelmaan ja edelleen kertoviin moniäänisiin kirjoitelmiin, joissa

2 Kielitieteessä peruslause (*simple basic sentence*) on yhden predikaation kielennys ja sekä rakenteeltaan että semantiikaltaan yksinkertainen. Peruslause on kaikissa kieliopeissa syntaksin kuvauksen lähtökohhta ja muodoltaan joko intransitiivi-, transitiivi- tai kopulalause (ks. esimerkiksi Itkonen 2008; 2009; Pajunen, Vainio, Hyönä & Bertram, tulossa).

oli kuvailevia ja arvioivia jaksoja. Tarinoiden suullinen kertominen voi olla hyvää jo kielenkehityksen varhaisvaiheessa, mutta niiden kirjallinen tuottaminen yleensä noudattaa kertomusrakenteen skeemaa – eli sisältö koostuu koherentisti – vasta noin 10-vuotiaana (ks. Berman 2007, 348). Omassa hankkeessamme ei ole erityisesti keskitytty kertomuksen kehittämiseen, mutta aineistossamme on havaittavissa selviä viivästyksiä tästä oletusiästä.

Tutkimuksemme varsinainen tavoite on tutkia merkityksien kielentämisen ja muotojen hallinnan kehitystä, ei kertovan tekstin tai tekstilajien hallinnan kehittymistä. Aineistoa analysoitiin kyllä myös tekstitasolla, vaikka se ei ollut ongelmatonta. Ensinnäkin aineiston laadullinen holistinen arvio (onnistuneisuus, tekstityypin tai -lajin toteutumisen aste) osoittautui alakouluikäisten tekstejä lukuun ottamatta vaikeaksi toteuttaa, ja arvioijareliabiliteetti vaihteli liikaa (50–80 %). Sisällön laadullinen arviointi onkin aina jossain määrin subjektiivista (ks. Kuuskorpi 2012). Omat arvioijamme tuntuivat painottavan eri tekijöitä, jotka voi niputtaa joko rakenne- tai sisältökriteereihin. Osa luokitteli tekstejä rakenteen ja koherenssin mukaan (perusasiat kunnossa tai sitten ei), ja osa taas painotti lisäksi tekstin synnyttämää vaikutelmaa (persoonallinen ja omaperäinen tai arkipäiväinen ja lattea). Holistisempi analyysi olisi edellyttänyt kriteerityypistä arviointia, joka olisi perustunut kirjoitelmien kokonaisrakenne- tai jaksoanalyysiin, mutta jo sadankin tekstin, sadoista puhumattakaan, kertomusskeema- ja jaksoanalyysi on valtava urakka (ks. tarkemmin esimerkiksi Juvonen 2004; 2010; Kalliokoski 2006; 2008; Karjalainen 2008; Juvonen, Virtanen & Voutilainen 2012). Toiseksi, vaikka hankkeen kirjoitelma-aineisto on luokiteltu jokseenkin systemaattisesti, kaikki tekstitason muuttujat eivät ole luokituksessa vielä mukana. Sisällöllisesti tekstitason hallinnan kehitys luonnollisesti näkyy koherenssin kautta. Koherenssin lisääntyminen on esillä I osion luvuissa, jotka osoittavat kertomuksellisuuden ja koherenssin yhteyden.

Valinnat ovat perusteltavissa myös siten, että kirjoittaminen on hyvin laaja-alainen ilmiö. Se edellyttää kognitiivista ja neuropsykologista kypsyysyttä, hienomotoriikan riittävää kehittyneisyyttä, kykyä suunnata tarkkaavaisuutta moniin asioihin yhdellä kertaa, toimivaa episodi- ja työmuistia sekä muun muassa kykyä suunnitella (ks. esimerkiksi Hooper

ja muut 2011). Kirjoittamiseen vaikuttaa voimakkaasti myös kirjoittajan motivaatio, käsitys itsestään kirjoittajana ja usko omiin saavutuksiin (ks. esimerkiksi Pajares & Valiante 2001; Routarinne & Absetz 2013). Lisäksi tiedetään, että taustatekijöinä on myös yhteiskunnallisia ja ympäristötekijöitä. Kehityksellisessä kontekstissa lukemisen ja kirjoittamisen tutkiminen edellyttää kognitiivisen profiilin (kuten muistin, motoriikan) kontrollointia. Suomessa tällaista tutkimusta ovat tehneet erityisesti psykologit ja logopedit (ks. esimerkiksi Niemi 2007; Torppa, Niemi, Vasalampi, Lerkkanen, Tolvanen & Poikkeus 2019). Omassa hankkeessamme emme lingvisteinä voineet kontrolloida neuropsykologisia ominaisuuksia ja kirjoittajien pystyvyyduskosta emme keränneet tietoa. Motivaation rajaamme tässä yhteydessä niin sanottuun ulkoiseen motivaatioon, esimerkiksi tilannekohtaiseen haluun yrittää tai tehdä jotain.

Puhutun kielen yksinkertaisempi rakenne kompleksistuu myöhemmän kielenkehityksen vaiheessa, kun sekä kielentämisen spesifisyys kasvaa että tekstin rakenne tiivistyy. Toisaalta kielennettävät ilmiöt ja niitä ilmaisevat rakenteet erilaistuvat (Berman 2007). Lapsen ja nuoren pitää kouluikäisenä saada hallintaansa kielellisiä ilmiöitä, jotka eivät ole puhutussa kielessä (yhtä) tärkeitä tai jotka eivät vaikuta puheessa. Usein kyseessä on heijastuma siitä, että puhetilanteessa osallistujat ovat läsnä ja tilanne on yhteinen, joten lisäinformaatiota ei tarvita. Kirjoituksessa lukija ei ole tilanteessa periaatteessa läsnä (vrt. kuitenkin Kallionpää 2014), joten hänen on saatava tilanteinen informaatio haltuunsa muulla tavoin kuin havainnoimalla. Siten puhutun kielen malli ei enää yksin riitä myöhemmän kielenkehityksen aikana.

Myöhemmän kielenkehityksen tutkimushankkeessa on osoitettu muun muassa, että kompleksisista kielenilmiöistä ei voi rakentaa muutoksen ennusteita puhutun kielen muutoksien varaan ja että hyvin kompleksisten ilmiöiden kielentäminen ei onnistu puhutun kielen hallinnan varassa. Pikkulapsilla puhutun kielen malli on ensikielen osalta yleensä melko samantapainen olosuhteista riippumatta – tavalliset rutinit ja lähiympäristö – vaikka ei vaihtelematon. Lasten kasvaessa mallina on enemmän kirjoitettu kieli, ja tämä malli erilaistuu puhuttua kieltä paljon enemmän yksilöllisesti koulutuksen ja harrastusten mukaan eli yhä enemmän kulttuuristen, sosiaalisten ja poliittistenkin syiden varas-

sa (Tolchinsky 2004, 235–236). Kompleksisten ilmiöiden kielentämistä kuten kehystämistä, referentiaalisen koherenssin ja moniäänisyyden ilmaisua, tutkitaan erikseen kolmessa I osion luvussa. Ne osoittavat, että semanttis-pragmaattisten ilmiöiden kielentämisen hallintaan saanti on jokseenkin hidasta.

Kielen rakenteen hallinnan osalta keskityimme analyyttisiin, kehityksellisiin muuttujiin kuten muotojen kompleksistumiseen. Sanaston ja morfologian kompleksistumisen lisäksi myös syntaktisen kompleksisuuden eli esityksen tiivyyden kasvu osoittautuu kehitykselliseksi. Sanaston tai morfologisen ja syntaktisen tiivyyden lisääntymisessä ei ole mainittavia sukupuolieroja, joskin miespuoliset kirjoittajat suosivat hieman kompleksisempaa ilmaisua. Muotomuuttujia käsitellään erityisesti teoksen II osiossa, jossa esillä on puhutun ja kirjoitetun kielen vuorovaikutuksellisuus, jota arvioidaan muun muassa possessiivisuffiksien käyttöön liittyen. Lisäksi osiossa analysoidaan muun muassa äännekirjain-vastaavuuden kehittymistä sekä morfologisten ja syntaktisten muotojen kompleksistumista.

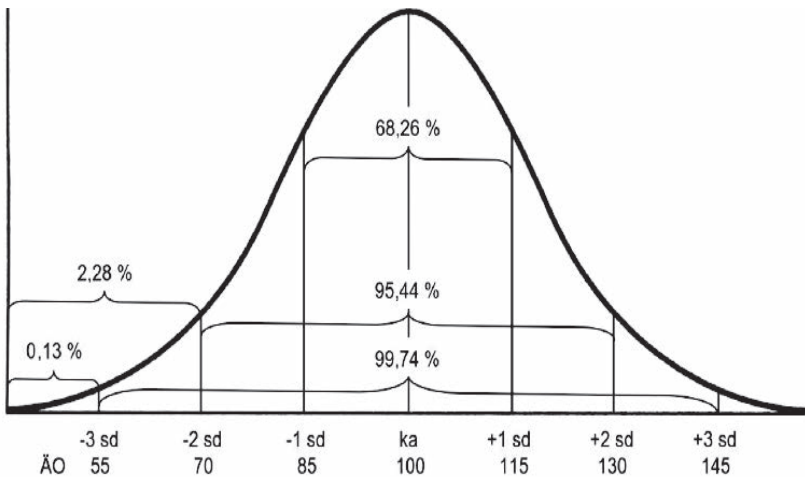
#### SUKUPUOLTEN SAMANLAISUUS

Kolmas keskeinen perusoletus hankkeessa on sukupuolten kognitiivisten kykyjen keskimääräinen samanlaisuus (*gender similarity hypothesis*, Hyde & Linn 1988; Hyde 2005). Yleisessä älykkyydessä ei katsota olevan sukupuolieroja, vaikka eroja voi olla joissain kognitiivisissa kyvyissä, jotka osaltaan vaikuttavat yleiseen älykkyyteen (Jensen 1998; Hyde 2005; Camarata & Woodcock 2006). Esimerkiksi Keith, Reynolds, Patel ja Ridley (2008) testasivat yleistä älykkyyttä ja kahdeksaa laaja-alaista kognitiivista kykyä ikäluokkiin 6–59 kuuluneella 7 000 koehenkilöllä käyttäen Woodcock-Johnson III -älykkyytestejä. Tuloksena löytyi vain kaksi johdonmukaisesti iästä riippumatta ilmenevää sukupuolieroja. Toinen koski prosessointinopeutta (*processing speed*), joka oli naisilla selvästi parempi, ja toinen verbaalista kykyä (*verbal ability*), joka oli miehillä hieman naisia parempi. Samaa älykkyydestä käyttäneet Camarata & Woodcock (2006) osoittivat, että lasten ja nuorten osalta on joitain sukupuolittuneita eroja mutta että ne ovat jo aikuisilla tasaantuneet. Kun omissa tutkimuksissamme voimme osoittaa, että derivaatiomorfologian tai semanttisen

tiedon keskimääräisessä hallinnassa ei ole mainittavia sukupuolittuneita eroja, tilanne on odotuksenmukainen. Pojilla oli reaktioaikoja ja virheitä mittaavassa derivaatiomorfologian testissä hieman tyttöjä enemmän virhevalintoja, mutta vastaavassa semanttisen tiedon testissä pojat olivat tyttöjä selvästi nopeampia. Reynoldsin, Schreiberin, Hojovskyn, Schwartzin ja Kaufmanin (2015) mukaan oikeinkirjoituksessa ja kirjoittamistaidoissa olisi suurimmat sukupuolierot tyttöjen hyväksi. Samat erot näkyvät aineistostamme ja saavat paljon huomiota samanlaisuusoletuksen vastaisina (ks. Pajunen 2016a).

Sukupuolierot voidaan jakaa dimorfisiin eli kaksijakoisiin ja allomorfiisiin eli tilastollisiin (ks. Heilig 2019). Dimorfiset erot liittyvät yksinomaan lisääntymiseen eivät ominaisuuksien jakautumiseen. Allomorfiset sukupuolierot jakautuvat Gaussin käyrän mukaisesti siten, että enemmistö ihmisistä sijoittuu jakauman keskialueelle, jossa on paljon monien ominaisuuksien päällekkäisyyttä (ks. esimerkiksi kaavio 1). Pienillä jakaumeroilla ei ole päällekkäisyysalueella merkitystä, mutta ne muuttuvat sitä tärkeämmiksi, mikä kauemmas keskiarvosta edetään. On kuitenkin muistettava, että mikään yksittäinen testi ei määritä yksilön asemaa asteikolla; se voi poiketa suuntaan tai toiseen muun muassa vireystilan tai tehtävyyppien tuttuuden vaihdellessa. Mainittakoon vielä, että sukupuolten välisiä kielelliseen älykkyyteenkin kytkeytyviä eroja on pyritty tutkimaan myös aivokuvantamisen avulla. Niin sanotussa Philadelphia-tutkimuksessa noin tuhannen koehenkilön aivokuvantamistulokset osoittivat, että nais- ja miesaivoissa hermosoluyhteydet olisivat erilaisia siten, että miesaivoissa olisi vahvemmat yhteydet aivopuoliskon sisällä ja naisaivoissa aivopuoliskojen välillä (ks. Ingahalikar ja muut 2014). Tuloksia ei ole yleisesti hyväksytty, koska alkuperäistutkimuksen efekti-koot olivat pienet ja koska toisaalta tuloksia on ylitulkittu jopa niin, että on aihetta puhua neuroseksismistä (ks. esimerkiksi Ribbon 2016).

Pitäydymme siis samanlaisuusoletuksessa. Tässä yhteydessä allomorfisten tilastoerojen voi ymmärtää tarkoittavan, että kielenhallinta on väestötasolla jatkuva ja normaali-jakautunut muuttuja ja että väestötasolla ei ole systemaattisia tai merkityksellisiä kielenhallinnan sukupuolieroja. Voimme olettaa, että vähintään 70 prosenttia suomenkielisestä väestöstä hallitsee äidinkielenä oppimansa kielen ikätasoonsa nähden vähintään



Kaavio 1. Keskihajonnat ja älykkyydosamäärät, alueiden prosentiosuudet (Psyko­logiliitto).

keskimääräisesti ja että pienellä osalla väestöstä (2–3 % asteikon kum­massakin päässä) on kielen hallinnassa joko erityisiä vahvuuksia tai suuria ongelmia. Hallinta kehitty­ iän mukaan tai voi muuttua patolo­ gisesti. Hallinnan kehitys voi sukupuolittain olla eriaikaista, mutta syyt eriaikaisuuteen voivat vaihdella. Oletuksemme on, että hallinnan erot näkyvät parhaimmin harvinaisten, vaikeiden tai kompleksisten ilmiöi­ den kielentämisestä ja että jakaumaeroja selittää kielellinen älykkyys, vaikka ei yksin. Yksilötasolla osaamista kuitenkin täytyy arvioida moni­ puolisemmin kuin väestötasolla, jotta satunnaiset kielenulkoiset tekijät tai mittauksen kapeus eivät vaikuta arviointiin.

Hankkeessa on muun muassa sanaston ja semanttisen tiedon osalta saatu melko odotuksenmukaisia tuloksia, kuten jo todettiin, mutta tu­ lokset myös osoittavat, että peruskoululaisilla osaaminen kehitty­ y hie­ man eriaikaisesti. Vaikka keskimääräinen kehitys on monen mitatun kielellisen ominaisuuden suhteen molemmilla lineaarista, tyttöjen kehi­ tys alkaa aikaisemmin kuin poikien. Toisaalta pojat saavuttavat tyttöjä koko peruskoulun ajan ja sama linja jatkuu lukiossa (ks. Harjunen, Mar­ janen & Karlsson 2019, 37–38). Ajallinen kehitysero voi olla luonteeltaan biologinen, sosiaalinen tai kulttuurisiin syihin kietoutuva, ja syihin ei

sinänsä ole tässä yhteydessä tarve ottaa kantaa. Hankkeessa ei myöskään lähtökohtaisesti ole ollut tarvetta eritellä koehenkilöitä sukupuolen mukaan, eikä koehenkilöiltä ole suoraan kysytty heidän sukupuoltaan. Suomessa viimeisen kahdenkymmenen vuoden aikana käyty laaja keskustelu poikien heikosta äidinkielen taidosta ja erityisesti kirjoittamistaitojen heikkenemisestä (ks. esimerkiksi Lappalainen 2003; 2011) on kuitenkin nostanut sukupuolieron yhdeksi tutkimuksen muuttujaksi.

Uusimman PISA-tutkimuksen mukaan erot näkyvät nykyisin myös lukutaidossa ja muissakin oppiaineissa ja lisäksi tyttöjen ja poikien erot ovat OECD-maiden suurimmat (ks. Leino ja muut 2019; Saarinen 2020). Taitoeroja on havaittu tyttöjen ja poikien välillä niin suomen- ja ruotsinkielisillä kuin suomea toisena kielenä oppivillakin (Silverström & Rautopuro 2015; Kuukka & Metsämuuronen 2016; Kauppinen & Marjanen 2020). Tässä yhteydessä viittaamme äidinkielen hallinnan käsitteellä kuitenkin aineistomme laadun vuoksi ensisijaisesti suomea ensikielellään oppineisiin ja koulussa S1-oppimäärää (suomen kieli ja kirjallisuus) opiskelleisiin lapsiin, nuoriin ja aikuisiin. Tarpeen vaatiessa tarkennamme määrittelyä lukukohtaisesti. On huomattava, että PISA-raportin mukaan (mts. 124) myös oppilaiden välinen sosioekonomisesti selittyvä vaihtelu on kasvanut ja vastaavasti koulujen sijainnin ja oppimistulosten väliset yhteydet ovat voimistuneet (samoin Harjunen ja muut 2019; Saarinen 2020). Sama sosioekonominen vaihtelu näkyy myös omassa aineistossamme, vaikka se ei ole muuttujana mukana. PISA-raportissa kiinnitetään erityisesti huomiota heikkojen (suomenkielisten) osaajien määrän kasvuun, eli toisin sanoen keskimääräisesti osaavien joukko on pienentynyt ja heikkojen kasvanut yli ääriarvojen (vrt. myös Kusnetsoff 2017). Tilanne ei siten enää olisi normaalijakauman mukainen. Tällainen poikkeava muutos pitäisi kyetä selittämään.

Ensimmäiseksi on kuitenkin tärkeintä arvioida, minkälaisia kielellisiä sukupuolieroja on ja miten ne voisi selittää. Erot esimerkiksi morfologian tai semanttisen tiedon hallinnassa eivät selittäneet saman koehenkilöryhmän kirjoittamistaitojen eroja.<sup>3</sup> Siten sukupuolierojen tausta-

3 Kuuselan (2011) mukaan heikoimmiksi arvioitujen 9.-luokan kirjoittajien lukutaito ja kielentuntemus on myös heikkoa, ja taso on näkyvissä samoilla oppilailla jo 7.-luokalla. Kielentuntemus määritellään tutkimuksessa melko yleisesti: kieli, sanasto ja peruskäsitteet.

tekijä ei ilmeisesti ole (pelkästään) kielitieto eikä kielellinen älykkyys vaan kyse on todennäköisesti enemmän jostain yhteiskunnallisesta tai opetuksellisesta tekijästä. Ulkoisen motivaation vähäisyys, kielteinen asenne ja pystyvyysuskon puutteet (*self-efficacy*, ks. Bandura 1997; Kuusela 2011; Harjunen ja muut 2019) ovat yhteiskunnallisesti latautuneita tekijöitä, joten vastuu ei voi olla yksin yksilötasolla. Testaamme tässä yhteydessä lähes kaikissa osatutkimuksissa eroja sukupuolen mukaan ja tuloksena on, että kirjoituksen kautta mitattavissa muuttujissa jakaumat eivät todellakaan aina ole odotuksenmukaisia vaan painottuvat liikaa pojilla alakvartiliin. On monia muuttujia, joissa suunta on alaspäin kaikilla peruskouluoppilailla. Kielitiedossa ero ei ole yhtäläinen, mikä vaikuttaa merkitykselliseltä, koska kielenhallinta on luku- ja kirjoitustaidon kehittymisen lähtökohta ja taustatekijä.

#### KEHITYKSELLISET MUUTTUJAT: HAJONTA VAI VARIAATIO

Neljäs taustaoletus on ollut, että vaikka väestötasolla kielenhallinta on normaalijakautunutta, yksilötasolla, kehityksellisesti ja suhteessa kielen resursseihin vaihtelu on suurta. Hankkeessa on tutkittu vaihtelun määrää ja laatua monella tapaa ja pyritty arvioimaan, liittyykö vaihtelu lähinnä tiettyyn ikätasoon vai liittyykö se myös muihin tekijöihin kuten yksilöihin tai kielennettävään ilmiöön. Edellisen tapauksen vaihtelua nimitämme (ryhmätason) hajonnaksi, jälkimmäistä (yksilötason) variaatioksi. Edellinen oletuksen mukaan vähenee iän mukaan, joten kyse on kehityksellisistä hajontamuuttujista. Jälkimmäinen voi oletuksen mukaan myös lisääntyä iän mukaan, joten kyse on myös ikäluokkatasolla monesta eri syystä varioivista muuttujista.

Pyrimme löytämään ne muuttujat, joissa vaihtelu liittyy ensi sijassa kielenhallinnan suuriin eroihin ikätasolla, ja ne muuttujat, joissa variaatio syntyy myös kielennyksen vaihtoehtojen – resurssien – suuresta määrästä. Edellisessä tapauksessa kyse on siis hajonnasta ja oletus on, että se tasaantuu ikäryhmittäin. Nuoremmilla ikäryhmäkohtainen hajonta voi olla suurempaa kuin normaali jakauma antaisi odottaa, mutta tyypillisessä kielenkehityksessä sen pitäisi vähetä selvästi iän mukaan (ks. esimerkiksi Metsämuuronen 2006; Harjunen, Marjanen & Karlsson 2019). Selvitämme, mikä on peruskoululaisilla se keskimääräisen hal-



linnan alue, jonka alittaminen on hyvä huomata ja johon opetuksessa voi reagoida. Vastaavasti pyrimme osoittamaan, missä määrin on kyse myös resursseihin liittyvästä variaatiosta. Luonnollisesti sekä hajonta että variaatio linkittyvät samoihinkin muuttujiin, mikä tekee arvioinnista hankalaa. Esimerkkinä voi mainita sanaluokkaindeksit, joissa sama arvo voi viitata kielitaidon kehittyneisyyteen tai kehittymättömyyteen. Esimerkiksi korkea substantiivitiheys voi toisaalta tarkoittaa tiivistä ja argumendoivaa tyyliä, toisaalta sanaluettelomaisuutta. Päätelmiä ei siis pidä tehdä yhden muuttujan jakaumista.

Oletus on, että hajonnan vähenemistä ilmaisevia kehityksellisiä muuttujia löytyy enemmän teksti- ja muototasolla, kehityksellisiä ja varioivia enemmän merkityksen kielentämisessä. Tämä selittyy yksinkertaisimmillaan siten, että yksikköjä on muototasolla suhteellisen vähän ja merkitystasolla paljon. Kehityksellisiä muuttujia tutkimme koululaisryhmässä pääasiassa näennäis- ja reaaliaikametodilla vertailukohtanamme nuoret aikuiset. Kielentämiserojen kautta syntyvä variaatio näkyy luotettavimmin aikuisikäluokkatasolla, jossa kielenhallinta on jo vähintään perustasolla eli jossa kielenhallinnassa ei ole oleellisia puutteita. Peruskouluikäisillä hajonta- ja varioivia muuttujia on paikoin hyvin vaikea pitää erillään ilman yksityiskohtaista laadullista analyysiä.

Varioivia muuttujia eli kielentämisen erojen lähteitä on viime aikoina tutkittu erityisesti tapahtumien kielentämisen yhteydessä ja on voitu osoittaa, että lähteitä on yksilöstä, stimuluksesta ja kielellisestä toimintaympäristöstä, havainnosta ja ympäristötekijöistä koulutuksen määrään (ks. Bohnemeyer ja muut 2014; Palmer, Lum, Schlossberg & Gaby 2017). Yksilöllistä variaatiota voi syntyä esimerkiksi sen mukaan, että perkeptiivinen ja kognitiivinen prosessi, jossa huomio (*attention*) kiinnittyy yksittäisiin ärsykkeisiin, on hyvin valikoiva ja yksilöittäin vaihteleva (Baker & Levin 2015). Koehenkilö voi tilanteittain yksinkertaisesti olla näkemättä jotain tai tulkita näkemänsä väärin (*change blindness*, ks. Gibbs, Davies & Chou 2016). Lisäksi siinä missä yksi näkee yksityiskohtia ja kielentää ne, toinen kielentää yleisemmällä tasolla. Muisti, näkökyky ja stressi vaikuttavat huomiokykyyn.

Kielentämisstrategiat voivat vaihdella yksilöllisesti mutta myös kehityksellisesti. Pieni lapsi voi kertoa lukijaa huomioimatta pitäen itseään

keskipisteenä, nuori voi kehystää tarinaansa enenevästi, mutta yleensä vasta aikuinen tiivistää ajallisesti ja vaihtelee tarinan tapahtumajärjestystä ja näkökulmia. Iän myötä kyky suunnitella kasvaa. Variaatiota syntyy myös kielennykseen käytettävissä olevista kielen resurssien määrästä ja niiden vaihtelevasta hallinnasta. Jos käytettävänä on vain yleisiä sanoja tai muotoja, kerronta on monen mitattavissa olevan muuttujan kannalta erilaista kuin jos tuntee ja käyttää yleisten sanojen tai muotojen spesifimpiä vastineita. Myös tekstityyppien iän mukana kasvava monipuolisuus vaikuttaa variaatioon ja toisaalta monet mittarit reagoivat tekstin ominaisuuksiin. Jos kirjoitetussa kielessä on paljon kehityksellistä variaatiota, variaatio voi perustua myös kulttuuriseen tai muuhun yksilölliseen erilaistumiseen.

Kehityksellinen hajontavaihtelu on esillä pääosassa teoksen luvuista, mutta myös kielentämisen (kehitykselliseen) yksilövariaatioon kiinnitetään kaikissa teoksen luvuissa huomiota. Erikseen sitä tarkastellaan spatiaalisen kehystämisen yhteydessä, muun muassa koska siihen suomen kielessä on tarjolla paljon kielen resursseja ja toisaalta yksilöllisiä kielentämisstrategioita. Tekstitasoista variaatiota tarkastellaan muun muassa possessiivisuffiksien käyttöä tutkimalla. Molemmista luvuista koehenkilöt ovat pääasiassa nuoria aikuisia. Kehityksellinen vaihtelu tulee esiin myös tekstin koherenssin – referentiaalisen, spatiaalisen, temaattisen ja temporaalisen – kehittymisen yhteydessä (ks. esimerkiksi Givón 1993; 2017).

Painopiste on kehityksellisissä varioivissa muuttujissa I osion luvuissa ja kehityksellisissä hajonnan pienenemistä ilmaisevissa muuttujissa II osion luvuissa. Edellisessä lähtökohtana on kielennettävän ilmiön semanttinen ja laadullinen analyysi, jälkimmäisessä muotojen käyttö. Edellisessä variaatio voi lisääntyä iän myötä, jälkimmäisessä vaihtelu voi vähetä, kun saavutetaan ikäluokkatasoinen keskimääräinen taso. Molemmista tapauksista taustalla on luonnollisesti myös kielelliset älykkyyserot.

#### TUTKIMUSKYSYMYKSET JA -METODIT

Edellä käsiteltyjen taustaoletusten kautta muotoutuvat myös tämän teoksen keskeisimmät tutkimuskysymykset:

- 1) Miten sanatiedon ja semanttisen tiedon hallinta kehittyi?
- 2) Miten kielen hallinta muotojen tasolla kehittyi?
- 3) Miten sukupuolierot näkyvät kielenhallinnassa?
- 4) Milloin vaihtelu tasaantuu ja milloin ei, eli milloin vaihtelun väheneminen on kehityksellistä ja milloin kyse on myös kielellisestä ja yksilöllisestä variaatiosta, jonka määrä ei ole sidoksissa pelkästään kehitykseen?

Kielenhallintaa on hankkeessa tutkittu monimetodisesti ajatuksena, että käyttämällä sekä korpus- että kokeellisia metodeja voidaan tutkittavasta ilmiöstä saada kokonaisvaltaisempi kuva. Korpusaineistoa on usein vaikea arvioida ilman tietoa koehenkilöreaktioista tai ilman laadullista etukäteistietoa, joten tietoa on kerätty myös kyselyin ja elisitoimalla (ks. Itkonen & Pajunen 2010). Aineiston ensivalinta ja analyysi on kuitenkin aina ollut luonteeltaan laadullista käsiteanalyysiä.

Sekä korpus- että testiaineisto analysoidaan laadullisesti ja tilastollisesti. Laadullinen analyysi nojaa aineistojen sana-, lause- tai tekstityyppikoodaukseen ja tilastollinen analyysi on tässä yhteydessä ennen kaikkea kuvailevaa, koska tavoitteena on osoittaa osaamisen vaihtelu. Teoksen tehtävä ei ole osoittaa erojen tilastollista merkitsevyyttä, mutta isot koehenkilömäärät takaavat sen, että eri mittarien kautta arvioituina erot ikäryhmittäin ovat tilastollisesti pääasiassa merkitseviä. Muissa yhteyksissä julkaistuissa artikkeleissa tuloksia on analysoitu myös monimuuttujamenetelmin.

Tutkimushankkeen ja tämän teoksen keskeisin ja tärkein tutkimustulos on, että peruskoululaisilla ikäkausiosaamisen hajonta on hyvin suurta. Lisäksi tendenssinä on, että koko populaatioissa alakvartiilitason osajien määrä kasvaa. Myös muissa yhteyksissä on todettu, että äidinkielen hallinnan erot alkavat olla tasavertaisten opiskelumahdollisuuksien takaamisen kannalta liian suuria (vrt. myös Pöysä & Kupiainen 2018). Äidinkielen hallinta peruskoulun loppuvaiheessa näyttää ennakoivan melko voimakkaasti esimerkiksi menestymistä myöhemmin ylioppilaskokeissa (Harjunen ja muut 2019). Tämä on nähtävissä sekä äidinkielen ja kirjallisuuden kouluarvosanan ja tutkintomenestyksen välisenä korrelaationa että oppiainearvioinnista riippumattomalla tekstitaitomittarilla

arvioituna (mts. 11–12). Viimeisimmän lukutaitoa mittaavan PISA-kartoituksen perusteella heikkojen lukijoiden osuus on kasvanut Suomessa selvästi eli 8,1 prosentista 13,5 prosenttiin (OKM 2019). Tämä merkitsee, että Suomessa on yhä enemmän nuoria ja nuoria aikuisia, joiden kielilliset taidot voivat rajoittaa heidän itsenäistä toimintaansa tekstittyneessä yhteiskunnassa. Tilannetta ei voi selittää kielellisen älykkyyden väestötason jakaumalla, ja suurimmat ongelmat koskevat poikia ja heidän jatko-opiskelumahdollisuuksiaan.

## AINEISTO

Hankkeen tutkimukset perustuvat empiiriseen aineistoon. Aineisto on kerätty useammassa vaiheessa, ja sisällöllisesti se on jaoteltavissa sähköiseen korpusaineistoon, kyselyin ja testein tuotettuun aineistoon sekä kirjoitustehtävillä koottuun aineistoon. Useimmiten yksi ja sama koehenkilö on tuottanut useammanlaista aineistoa samalla kertaa tai eriaikaisesti, mikä mahdollistaa vertailun kielen eri osa-alueilla tai kehityksellisesti. Pääosin kehitystä kuitenkin tutkitaan näennäisaikametodin avulla.

## SÄHKÖISET KORPUKSET

Korpuksat koostuvat sanomalehtiaineuksesta, fiktiosta, alakoulun oppikirjoista ja nuorten aikuisten ylioppilaskirjoitelmista, ja ne ovat kaikki sähköisessä muodossa ja automaattisesti koodattu. Sanaesiintymämäärä on yhteensä noin 45 miljoonaa. Lehtiaines on *Helsingin Sanomista* (Pajunen & Virtanen 2002a), *Savon Sanomista* (Pajunen & Virtanen 2002b) ja *Turun Sanomista* (Laine & Virtanen 1999b). Fiktioaines on 2000-luvulla ilmestyneistä, lähinnä jännitys- tai salapoliisiromaaneiksi luokiteltavista teoksista ja oppikirjat edustavat 2000-luvun alakoulun oppikirjoja (ks. Jaakola 2004). Kirjoitelmakorpus on koostettu ylioppilaskirjoitelmia esittelevistä kirjoista ja sen sanamäärä on yhteensä miljoona.<sup>4</sup> Kun muut korpusaineistot edustavat 2000-luvun kieltä, kirjoitelmakorpuksat

4 Kirjoitelmakorpus on koostettu Suomalaisen Kirjallisuuden Seuran ja Äidinkielen opettajain liiton julkaisemista ainekokoelmista, jotka aluksi keskittyivät niin sanottuihin valioaineisiin eli erityisen hyviksi arvioituihin kirjoitelmiin. Myöhemmin valikoima on monipuolistunut eri arvosanoja saaneisiin kirjoitelmiin. Laadulliset sisällölliset erot eivät ole tutkimuksessamme mukana. Korpus ei ole julkinen.

sessä on tekstejä vuosilta 1948–2003. Lisäksi aineistossa on pieni määrä 1970-luvun Lauseopin arkiston materiaalia ja siihen pohjavia kirjoitettua ja puhutun kielen tietoja (ks. Pajunen & Palomäki 1984). Lauseopin arkiston (= LA) aineisto on 1970-luvun kirjoitettua yleiskieltä Oulun korpuksen jakauman mukaan koostettuna (sanomalehdet, aikakauslehdet sekä ei-fiktiivinen kirjallisuus, ks. Saukkonen 1982) tai se edustaa aluemurteita.

Korpuksset – LA:ta lukuun ottamatta – on koodattu morfosyntaktisesti FDG-jäsentimellä (*functional dependency grammar*, ks. Voutilainen 2002) ja ne on koottu konkordanssityyppisiksi tietokannoiksi (Laine & Virtanen 1999a; Virtanen & Pajunen 2001). Korpuksista voi tehdä hakuja lekseemin tai sanamuodon, sanaluokan, funktion, muodon ja frekvenssin avulla. Tarpeen vaatiessa hakusanan esiintymäympäristöä voi rajata paitsi edellä mainituin muuttujin myös etäisyyden tai pääsanan suhteen. Sähköisiä korpuksia on käytetty frekvenssien ja vertailulukujen laskemiseen ja testisanojen vertaistamiseen. Possessiivisuffiksien käytön tutkimus perustuu sähköiseen korpusaineistoon.

#### TESTIT JA KIELENTÄMINEN

Testien avulla on elisitoitu eli kirjoiteltu suunnitelluilla asetelmilla monenlaista kielellistä ainesta. Testien yhteydessä on tuotettu myös kielentämis- ja kirjoittamistehtäviä. Kirjoitustehtävänä on ollut joko kirjoitelma unelmapäivästä, kuvan tai tapahtuman kielentäminen tai kuvasarjatarinan kertominen. Aineistot on kerätty vuosina 2007–2020. Testauskokoontapanoja on useammanlaisia:

- a) Määritelmätesti, sana-assosiaatiotesti, derivaatiotesti (kaikki paperi- ja kynätestejä), kirjoitelma unelmapäivästä, tietoa morfologian hallinnasta ja muun muassa ylioppilastutkintotodistuksesta. Koehenkilöinä on nuoria aikuisia yhteensä 240 (ks. myös Pajunen ja muut 2015; 2016; Laasanen ja muut, arvioitavana).
- b) Semanttisen ja morfologisen leksikaalisen päätöksen kaksi testiä, molemmat elektronisia online-testejä reaktioaikoinen, kirjoitelma unelmapäivästä. Koehenkilöinä on kuudes- ja yhdeksäsluokkalaisia sekä nuoria aikuisia yhteensä 300 (myös Vainio ja muut 2019; Vainio, Pajunen & Häikiö tekeillä).

- c) Määritelmätesti, sana-assosiaatiotesti, vapaa assosiaatiotesti, kirjoitustehtävänä kuvan kielentäminen, tuttuustesti, reitin kielentämistesti. Koehenkilöinä on 157 kuudes- ja yhdeksäsluokkalaista (Pajunen, tekeillä). Testit antavat tietoa peruskoululaisten sanastonhallinnasta ja erityisesti spatiaalisten ilmiöiden kielentämisestä.
- d) Sanatiedon hallintaa mittaavat testit ja kirjoitelma, taustatietolomake, tieto arvosanoista ja opettaja-arvio. Pitkittäistutkimus 3.–4.- ja 5.–6.-luokkalaista, kaikkiaan noin 600 S1- ja S2-koehenkilöä. (Tuloksista ks. Honko 2013). Luku norminmukaisuuden hallinnan kehityksestä perustuu tähän aineistoon, mutta Hongon laaja aineisto (2013) taustoittaa koko hanketta monella tavalla. Tämä on hankkeen ainoa osio, jossa S2-oppijoita on edustavasti mukana.
- e) Suullinen videoitujen liike- ja tilatapahtumien kielennys *Trajectoire*-videosarjan (ks. Ishibashi, Kopecka & Vuillermet 2006) ja fiktiivisen liikkeen (ks. Blomberg 2014) kuvasarjan avulla. Koehenkilöinä on 50 nuorta aikuista. Testit antavat tietoa nuorten aikuisten spatiaalisten ilmiöiden kielentämisestä. *Trajectoire* koostuu 76 videoklipistä, joista 55 esittää varsinaisen liiketilanteen, kaksi on tehtävänannon yhteydessä katsottuja harjoitusvideoita ja loput ovat ”fillereitä” eli mukana tutkimuksen tarkoituksen hämärtämiseksi. Videot ovat kymmenisen sekuntia pitkiä, ja niissä on näyteltyjä liiketilanteita vaihtelevissa luonnonympäristöissä. Kuvasarjassa on 36 tilaa ilmaisevaa piirrettyä kuvaa.
- f) Suullisia nimeämis- ja kielennystehtäviä; Olion- ja tapahtuman nimeäminen 800 kuvan avulla (vrt. Szekely, D’Amico, Devescovi, Federmeier, Herron, Iyer, Jacobsen, Arévalo, Vargha & Bates 2005), koehenkilöinä 60 aikuista. Nimeäminen 100 kuvan avulla, koehenkilöinä 50 nuorta aikuista. Kuvakirjakerronta Mayerin (1969) *Frog, where are you?* -kuvakirjasta (ks. Strömqvist ja Verhoeven 2004), koehenkilöinä 35 nuorta aikuista. Testeistä osa on toteutettu elektronisina. Testit antavat tietoa nuorten aikuisten sanastonhallinnasta.
- g) Joukko semanttisia subjektiivisia testejä, joissa on koehenkilöarvioita tiettyjen sanojen tuttuudesta, konkreettisuudesta, kuvitteellisuudesta, abstraktiudesta tai sanan herättämästä tunteesta tai sen

merkityksestä asteikolla positiivinen–negatiivinen (ks. esimerkiksi Söderholm, Häyry, Laine & Karrasch 2013; Citron, Cacciari, Kucharski, Back, Conrad & Jacobs 2016). Koehenkilöinä on yhteensä noin 200 peruskoululaista ja nuorta aikuista. Subjektiiivisten testien avulla on vertaistettu varsinaisia kokeita, mutta niiden avulla voi selittää myös lasten ja nuorten tekemiä valintoja ja arvioida sanaston kehitystendenssejä.

Testit ovat antaneet runsaasti tietoa sanatiedon laadusta ja määrästä kouluikäisillä ja nuorilla aikuisilla. Testeillä tuotettu aineisto on analysoitu soveltuvin kielensisäisin ja -ulkoisin muuttujin Excel-tietokantoihin siten, että samojen koehenkilöiden eri testitulokset voi yhdistää. Lisäksi on laskettu erilaisia frekvenssivertailulukuja. Elektroniset testit on toteutettu ohjelmalla Pohjankukka & Vainio (2013) tai ohjelmalla E-Prime-3 joko ryhmittäin samassa tilassa tai yksilötestauksin. Kaikki käytetyt testityypit, myös kynä-paperitestit, noudattavat yleisesti kielitietoa testaamaan käytettyjä testiformaatteja, jotka tutkimusryhmä on soveltanut suomen kieleen. Koska niitä on käytetty onnistuneesti samantapaisissa konteksteissa aiemmin, niitä voi pitää valideina kielenhallinnan osa-alueiden testaukseen.

#### UNELMAKIRJOITELMAT

Kirjoitelma-aineisto on kerätty vuosina 2006–2010 ja 2012–2014 ja se koostuu alakoululaisten, yläkoululaisten sekä nuorten aikuisten kirjoitelmista (ks. taulukko 1). Vanhimman aineiston ytimenä on Maija Purasen ja Annukka Yli-Paavolan keräämä alakoululaisaineisto, jota on muuttaman koulun erillisaineistolla täydennetty parina alkuperäiskeruuta seuranneena vuonna (= niin kutsuttu Koulukorpus, ks. Pajunen, Puranen ja Yli-Paavola 2010). Alakouluaineistossa on kaupunki-, taajama- ja haja-asutusalueen kouluja sekä kielikoulu- ja kielikylpyluokka. Koko alakoulun aineisto on kerätty maaliskuussa, jotta mahdollisimman moni ensimmäisen luokan oppilaista jo osaisi lukea ja kirjoittaa. Maaliskuu on myös alakoulun lukutestin vertailukuukausi (Lindeman 2000, 3). Vuosina 2013 ja 2014 kerättyyn kouluaineistoon kuuluu alakoulun kuudes luokka ja yläkoulun aineistoa. Keruu on toteutettu vastaavilta alueilta ja

myös siihen on valikoitunut yksi kieliluokka. Vanhempien koululais-  
ten osalta keruuajalla ei katsottu olevan enää merkitystä, koska kielelli-  
set valmiudet ovat paremmat. Tähän aineistoon eli niin kutsuttuun  
Unelmakorpuksen (ks. Pajunen 2016b) on yhdistetty myös aiempi  
kuudensien luokkien aineisto. Näin mahdollistuu eri aikoina kerätyn  
kuudensien kirjoitelmien vertailu. Unelmakorpuksen kuuluu paitsi pe-  
ruskoululaisten kirjoitelmia myös nuorten aikuisten kirjoitelmat, jotka  
on kerätty suomen kielen pääsykokeen yhteydessä tai ammattikorkea-  
koululaisilta vuosina 2012–2014. Koulukorpuksen rinnakkaisaineisto  
(= ns. S1–S2-korpus), jossa on yhteensä 550 S1- ja S2-kirjoittajien kir-  
joitelmaa, on kerätty vuonna 2007 (toinen ja kolmas luokka) ja 2010 (vii-  
des ja kuudes luokka). Unelma-aineistoihin viitataan joko korpuksen  
nimellä tai lyhyemmin esimerkiksi sanalla Unelmakirjoitelmat.

Taulukko 1. Koulu- ja Unelmakorpuksen kirjoitelmien määrä.<sup>5</sup>

Ikäluokka	F	M	Kaikki
Alakoulu 1.–5.	481	466	947
Kuudennet	162	158	320
Yläkoulu	137	133	270
Aikuiset	244	98	342
Yhteensä	1024	855	1879

Unelmakirjoitelmat on kirjoitettu samasta aiheesta eli unelmien päivästä  
samalla tehtävänannolla: ”kirjoita unelmiesi päivästä, jonka olet kokenut,  
haluaisit kokea tai voisit kokea maailmassa, missä mikä tahansa olisi  
mahdollista” (ks. Pajunen 2012). Aikaa kirjoittamiseen on ollut yksi oppi-  
tunti ja etukäteissuunnittelua ei ole ollut.<sup>6</sup> Kirjoitelmien kokonaissana-  
määrä kasvaa lineaarisesti peruskoulun ikäluokasta toiseen yhdestä sa-  
nasta muutamaan sataan, aikuisia puolestaan on pyydetty kirjoittamaan  
noin 300 sanan kirjoitelma (ks. taulukko 2). Sanaesiintymien määrä  
koko aineistossa on noin 234 000. S1–S2-korpus on kirjoitettu aiheesta  
Unelmien päivä (ks. Honko 2013), ja sen sanamäärä on noin 60 000.

5 F = tytöt ja naiset, M = pojat ja miehet.

6 Sintonen (2011) vertaili suunnittelemattomia ja suunniteltuja Unelmakirjoituksia pro gradu -tutkiel-  
massaan. Tuloksena oli, että suunnittelu toisaalta nostaa tekstin tasoa, toisaalta yhdenmukaistaa  
kirjoitelmia.



Tehtävänanto on ollut Unelmakirjoitelmia täsmällisempi ja kirjoittajia on opastettu kirjoittamaan satu tai tarina. Alussa on keskusteltu unelman käsitteestä. Aiheen ja ohjeistuksen ero näkyy aineistossa jonkin verran.

Taulukko 2. Sanaesiintymien määrä Unelmakirjoitelmissa.

Ikäluokka / Sanaesiintymä	n	Keskimäärin kirjoittajittain
Alakoulu 1.–5.	63 641	67,2
Kuudennet	42 605	133,1
Yläkoulu	38 365	142,1
Aikuiset	89 192	260,8
Yhteensä	233 803	

Kirjoittajien motivaatio (halu suostua kirjoittamaan tai viitseliäisyys) näyttää vaihdelleen yksilöittäin ja myös ryhmittäin.<sup>7</sup> Uusimmissa keruissa motivaatiota on pyritty tasaamaan arpomalla kirjoittajien kesken pieniä palkintoja. Motivaation nostatus näyttää kasvattaneen sanamäärää mutta ei ole niinkään lisännyt rakenteen suunnittelua – ja kirjoitelman tasoa –, mikä näkyy, jos verrataan vuosien 2006 ja 2014 kuudensien aineistoja keskenään (ohjatun suunnittelun vaikutuksesta ks. Sintonen 2011, motivoinnista ks. myös Kyösti 2018; motivaation osatekijöistä ja vaikutuksesta ks. Troia, Harbaugh, Shankland, Wolbers & Lawrence 2013). Aikuisista ammattikorkeakoululaisille on maksettu rahapalkkio; pääsykokeeseen osallistuville itse koetilanne on ollut motivoiva.

Koko Unelma-aineisto (eli Koulu-, Unelma- ja S1–S2-korpus) on ensin kirjoitettu sähköiseen muotoon teksteiksi ja koodattu myöhemmin manuaalisesti sana kerrallaan Excel-tietokantaan. Jokainen sanaesiintymä on luokiteltu sanarakenteen ja sanaluokan mukaan sekä syntaktisesti että semanttisesti. Kaikki sanaesiintymät on myös kantasanaistettu eli muutettu lekseemeiksi. Aineisto on koodattu myös sanatasoa ja lausetta laajempiin yksiköihin, joista voidaan arvioida muun muassa kirjoitel-

7 Aineiston kerääjät välittivät usein koehenkilökommentteja ja kuvailivat keruutilanteen tunnelmaa tai häiriötekijöiden määrää; nämä antoivat tukea päätelmiin kirjoittajien (ulkoisesta) motivaatiosta. Koehenkilöt saattoivat myös kysyä, vaikuttiko suoritus oppiainearviointiin; vastaus luonnollisesti oli kielteinen. Sisäinen motivaatio ei ole tällaisilla havainnoilla arvioitavissa, joten viittaamme termillä *motivaatio* juuri ulkoiseen motivaatioon.

mien syntaktista hierarkkisuutta ja erilaisia leksikaalisia suhteita. Lisäksi esitetään tekstityyppiä ja koko kirjoitelmaa koskevia holistisia arvioita, mutta on syytä muistaa, että ei ole olemassa sellaisia holistisia hyvän kirjoitelman kriteerejä, joista vallitsisi yksimielisyys (ks. esimerkiksi Koutsoftas & Gray 2012, 407; Bouwer, Béguin, Sanders & van der Bergh 2015, 95). Tasomuuttujaa käytettiin vain alakouluaineistossa, koska siinä tasoerot olivat selvän kehitykselliset. Aineiston koodaajina on ollut useita henkilöitä, ja koodausta on testattu ja yhtenäistetty määrittelemällä tarkat sana-, sanaluokka- ja sanarakennekriteerit, semanttisen analyysin kriteerit sekä lause- ja tekstitason analysointikriteerit perus- ja rajatapauksineen. Muuttujien määrittelyä on usein jouduttu tekemään sekä sana- että tekstikontekstissa. Rajanveto on usein vaikea vaikkapa substantiivien ja adverbien välillä. Aineiston koodisuunnitelmassa on yli 60 sivua, ja juuri koodaus on ollut tutkimuksen työläin vaihe. Tulokset luonnollisesti perustuvat aineiston yksityiskohtaiseen koodaukseen, joka on sama kaikissa osatutkimuksissa. Korpuksia on käytetty myös kymmenissä opinnäytetöissä, ja opiskelijat ovat saaneet aineistonsa valmiiksi koodattuna. Siten nämä opinnäytetöiden laadulliset analyysit täydentävät hankkeen tutkimusta luotettavasti.

Koulu- ja Unelmakorpuksen tehtävänannossa ei viitata tekstityyppiin mutta otsikko on ohjannut kertovaan tekstiin, jossa tapahtumia kuvataan kronologisesti aamusta iltaan. S1–S2-kirjoitelmissa ohjeistettiin kirjoittamaan satu tai tarina, mutta lopputulos ei ratkaisevasti eroa muista. Aineistosta muodostuu kehityksellisesti kuvaava kaari kertomusrakenteen tavoittelusta aina enemmän kuvailevaa ja erittelevää ainesta sisältävään kerrontaan. Alakoulun yläluokilla kirjoitelmat alkavat tavoittaa kertomusrakenteen skeemaa (Pajunen 2012) ja kaikkein kehittyneimmät yksittäiset kirjoittajat arvioivat samalla unelmapäivän käsitettä jo kuudennella luokalla. Aikuisista suurin osa ainakin jossain määrin erittelee tai kuvailee. Vastaavasti irrottaudutaan tapahtumia pinnallisesti luettelevasta päiväskeemasta (*aamulla heräsin – illalla menin nukkumaan*). Petälän (2018) tutkimuksen mukaan keski-ikäiset kirjoittavat vastaavasta teemasta siten, että irtautuvat sekä tapahtumajärjestyksestä että kertomusrakenteesta. Teksti säilyy silti pääpiirteissään kertovana, vaikka evaluaation määrä nousee. Heinosen (2018) mukaan viisi yle-

sintä aihetta kuudes- ja yhdeksäsluokkalaisilla sekä nuorilla aikuisilla on harrasteet, ravinto, skeemaa jäsentävät triviaalirutiinit, ystävät sekä hyödykkeet. Nuorimmat kertovat juhlista, ystäväistä, lemmikeistä ja huippulomapäivistä; rahalle annetaan suuri merkitys. Vanhemmilla omaperäisyys lisääntyy, mutta toisaalta painotetaan aika samansuuntaisesti ystävyysuhteita ja arjen pienten tapahtumien onnea. Monet kirjoittelmissa ovat koskettavia, monet hauskoja, monet surullisia. Jo peruskoululaisilla mutta ennen kaikkea nuorilla aikuisilla kirjoitelma on elämäntarinatyyppinen (vrt. Habermas & de Silveira 2008; McLean & Breen 2009): se päivä elämässä, minä jotain käänteentekevää, merkityksellistä tapahtui tai odotetaan tapahtuvan. Kyse voi olla häistä, hautajaisista, opiskelemaan pääsystä, lapsen syntymästä tai vaikkapa unelmien matkasta. On tyypillistä, että tämäntyyppisissä kirjoitelmissa omaelämäkerrallisuus iän myötä lisääntyy (ks. Pasupathai & Mansour 2006).

Kirjoitelmien tekstityyppi on siis kertova ja kirjoitelmien tavoitemuoto on tyypillisesti kertomus tai omaelämäkerta. Kertomus on määritelmän mukaan tapahtumien representaatio ja tyypillisesti tapahtumat kerrotaan tapahtumajärjestyksessä. Kerrontaa sitoo sisäinen aikarakenne (ks. esimerkiksi Berman & Slobin 1994). Kirjoitelmissa suurimpia verbi-luokkia ovat odotuksenmukaisesti tapahtumia ja liikkeitä ilmaisevat tapahtuma- ja liikeverbit, ja suurin adverbiryhmä koostuu aikaa ilmaisevista. Käännekohdat kuljettavat kertomusta eteenpäin. Alakoulussa kertomustavoitemuodon saavuttamista mitattiin kertomusskeema-analyysillä (Pajunen 2012; Honko 2013) ja narratologisilla kriteereillä (Vesala 2012). Jos narratiivisuus määritellään hyvin tiukasti, suurin osa vielä 3.-luokkalaisistakaan ei tuota kertomuksia. Noin 60 prosenttia kuudesluokkalaisten kirjoitelmista on heikkoja tai vahvoja kertomuksia, mutta sana- tai lauseluettelaita tai muulla tavoin keskeneräisiä kirjoitelmia löytyy myös.

Kertomuksissa kertovan tekstin oheen syntyy arvioivia ja kuvailevia osuuksia; tämä on selvä kehityksellinen piirre. Örnmarkin (2011) mukaan argumentoivassa tekstissä monen tekstityypin yhdistelmät ovat ennemmin heikon kirjoittajan ominaisuus: erityisesti tukeutuminen (minäkeskeiseen) kertovaan kieleen argumentoivan sijasta on heikkouden signaali. Siten kehitykselliset tekstiominaisuudet näyttävät vaihte-

levan tekstityypeittäin. Esimerkiksi nuorilla aikuisilla yksinomaisesti kertova teksti vaikuttaa lapsekkaalta, koska lukija odottaa aikuiselta jo perusteltuja siitä, miksi jokin asia on tärkeä, tai edes tärkeyden kuvailemalla näyttämistä. Aineiston arvioivat kirjoitelmat, joissa ei ole lainkaan muiden tekstityyppien ominaisuuksia, rajoittuvat pariin tekstiin. Jos kirjoitelma keskittyy arviointiin, arvioinnin kohteeksi näyttää nousevan unelma, ei unelmien päivä. Kertova teksti on myös moniäänistä ja moniäänisyys lisääntyy kehityksellisesti, samoin lukijan huomioiminen ja vuorovaikutuksellisuus. Kertovalla tekstillä on toisin sanoen oikeastaan aika moninaisia ominaisuuksia, jotka saadaan hallintaan hitaan kehityksellisesti. Onnistunut kertova teksti on siten vaativa kirjoitustehtävä, vaikka se usein mielletään koulunkäynnin alkutaipaleille kuuluvaksi.

Alakoululaisia käsittelevissä luvuissa aineistona on joko Koulukorpus tai S1–S2-korpus. Yläkoululaisia ja nuoria aikuisia analysoidaan sekä Koulukorpuksen että Unelmakorpuksen aineistolla, mutta rakenteellinen kompleksisuus on tarkimmin koodattu Unelmakorpuksen. Osaksi käytössä on ainakin paikoin koko aineistoa, osaksi on tehty rajauksia. Norminmukaisuutta on arvioitu S1–S2-korpuksesta, jonka 650 kirjoittajalta on kerätty tietoa sekä eriaikaisesti että kyselytестein.

## KOEHENKILÖT

Koehenkilöt ovat joko yleisopetuksen peruskoululaisia tai nuoria aikuisia. Peruskoululaiset edustavat tasaisesti koko ikäluokkaansa – lukuun ottamatta kahden kieliluokan oppilaita<sup>8</sup> – mutta mukana ei ole erityisluokkien oppilaita. Aikuisilla on lähes kaikilla ylioppilastausta. S2-oppilaita on kerätty edustavasti vain osaan aineistoa. Kaikki suomi toisena kielenä -oppilaat ovat Suomessa syntyneitä tai ennen esikouluikää Suo-

8 Kolmasluokkalaiset kieliluokan oppilaat on valittu kielellistä ja matemaattista älykkyyttä mittaavan testin avulla peruskoulun 2. luokan keväällä. Nämä oppilaat kuuluvat monen muuttujan suhteen yläkvartiiliin ja ero valikoimattomaan kolmannen luokan aineistoon on keskimäärin 1,5–2 kehitysvuotta. Sekä ala- että yläkoulun aineistossa on yksi kieliluokka, alakoulun aineistossa on lisäksi kielikylpyluokkia. Alakoulun kieliluokasta *Helsingin Sanomien* antamien tietojen mukaan 60 % kirjoitti äidinkielestä laudaturin. Luokan kielellinen taso jo alakoulussa tietyksi osoittaa, että kyse ei ole yksin koulun vaikutuksesta (ks. myös Harjunen ja muut 2019). Hyvä osaamistaso ei kuitenkaan rajoitu kielikouluihin, vaan aineistossa on esimerkiksi eräs 6. luokka, joka suoriutui myöhemmin ylioppilaskirjoituksista lehtitietojen mukaan erinomaisesti. Vastaava osaaminen käy ilmi omasta tutkimuksestamme.

meen muuttaneiden maahanmuuttajataustaisten vanhempien lapsia, jotka ovat tutkimuksen alkuvaiheessa asuneet Suomessa vähintään noin 4 vuotta ja sen loppuvaiheessa vähintään noin 7 vuotta (Honko 2013, d)-aineisto). Koehenkilöiden kokonaismäärää on vaikea laskea tarkkaan, mutta heitä on vähintään 2 600. Luku nousee, jos mukaan lasketaan hankkeen opiskelijatöihin erikseen kerätty aineisto. Testeihin on osallistunut noin 1 400 koehenkilöä; ja Unelmakirjoitelmia on yhteensä runsas 2 400. Osa koehenkilöistä on osallistunut vain yhteen tehtävään, osa moneen. Sähköisistä korpuksista koehenkilöiden määrää ei voi arvioida, mutta koska mukana on useita sanomalehtiä, kirjailijoita ja ylioppilaskokeeseen osallistuneita, määrä on vähintään sadoissa. Sähköisten korpusaineistojen tuottajien ikä vaihtelee, mutta kyse on työikäisistä ihmisistä tai nuorista aikuisista.

Peruskoululaisten aineistot on kerätty eri puolilta Suomea, mutta keuruu painottuu eteläiseen ja läntiseen Suomeen. Kouluja on mukana keuruussa yli 20 ja luokkia kymmeniä, mikä nostaa aineiston edustavuutta. Peruskoulun aineistossa on mukana kaikenlaisia koehenkilöitä, ja heitä ei ole esikarsittu esimerkiksi (lievän) kielihäiriön perusteella.<sup>9</sup> Otos edustaa siten koko yleisopetuksessa mukana olevaa suomenkielistä ikäluokkaa, samoin tulokset. Yleisopetuksen luokkien taso vaihtelee melko sattumanvaraisesti. Suorituksista ei voi aina päätellä, johtuvatko niiden puutteet ulkoisen motivaation, pystyvyyden tai osaamisen puutteesta vai onko kyse kielihäiriöstä. Vaikka esimerkiksi suomalaiskoululaisten lukutaito on aikaisempien kansainvälisten PIRLS- ja PISA-kielitaloituskartoitusten perusteella huippuluokkaa, lukumotivaatio ja lukemiseen sitoutuminen ovat joillakin mittareilla mitattuna vertailumaiden heikoimpia (KTL 2017; Leino ja muut 2019). Poikien suhtautuminen sekä lukemiseen että kirjoittamiseen ja omiin tekstintaitoihin on tyttöjä kielteisempää (Harjunen & Rautopuro 2015; Leino ja muut 2019).

Asennetekijöiden merkitys oppimiseen voi olla suuri sekä suoraan että välillisesti: on vaikea syventyä asioihin, jotka eivät tunnu merkityk-

9 Karsinta ei olisi ollut mahdollistakaan, koska lieviä häiriöitä ei aina diagnosoida (vrt. Kuusela 2011). Lievät kielihäiriöt näkyvät tutkimuksen aineistossa muun muassa kirjoitusvirhetyyppien perusteella sekä toisaalta leksikaalisissa testeissä sattumanvaraisena vastaustapana (ks. Ahvenainen & Holopainen 2000; Kulju & Mäkinen 2009; Lehtinen 2014; vrt. Honko, tässä teoksessa).

sellisiltä, ja oppia sitä, millä ei ole yksilöllistä merkitystä tai minkä oppiminen ei tunnu mahdolliselta. Jos äidinkielen kouluopetus ja käytetyt materiaalit ovat erityisesti poikien näkökulmasta omaan elämänpiiriin ja tavoitteisiin nähden etäisiä (Sulkunen & Kauppinen 2018), koulu ei pysty myöskään parhaalla mahdollisella tavalla tukemaan positiivisten asenteiden tai henkilökohtaisen suhteen syntymistä opittaviin sisältöihin. Ulkoisen motivaation puute näkyy erityisesti haluttomuutena ylipäätään tarttua kirjoittamistehtävään mutta kietoutuu luonnollisesti myös osaamiseen, koska kirjoittamistaito ei kehity ilman harjoittelua. Motivaatiota pyrittiin nostamaan paitsi pienillä palkkioilla myös testien vaihtelevilla toteuttamistavoilla. Osaksi tässä onnistuttiin, ja erityisesti sähköiset testit kiinnostivat. Toisaalta kokeen toteuttajat raportoivat häiriökäyttäytymisestä myös sähköisissä testeissä.

Peruskoululaisia tarkastellaan osaksi luokkakohtaisesti jakaumalla 1.–9. luokkaan. Osaksi keskitytään peruskoulun päättövaiheisiin eli alakoulun kuudenteen ja yläkoulun yhdeksänteen luokkaan, ja tällöin ajatus on tutkia peruskoululaisia muutosvaiheessa. Lisäksi tutkimuksen ensimmäisiä tavoitteita oli tutkia juuri kuudesluokkalaisia, jotta saataisiin tietoa yläkoulun alkuun eli vaiheeseen, josta aineenopettajat aloittavat. Monissa osatutkimuksissa aineisto on valittu kahden luokkavuoden välein. Tämä siksi, että yhden ja saman ikäluokan sisällä kielellinen kehitysero voi olla jopa parisen vuotta, jolloin seurauksena on, että ikäluokkakehitys peittyy liian tiheällä koehenkilövalinnalla. Aineistossa on yleistä, että kolmannen ja neljännen, viidennen ja kuudennen sekä seitsemännen ja kahdeksannen luokan välillä muuttujissa on jonkinlaista häiriövaihtelua. Ero saattaa olla yhteydessä paitsi yksilölliseen kehitykseen myös murrosikään ja siihen liittyviin muutoksiin kirjoittamismotivaatiossa, keskittymiskyvyssä tai tehtäväorientaatioissa. Vuosien 2013–2014 keruusiin osallistuneiden motivaatiota pyrittiin nostamaan arpomalla luokittain pari elokuvalippua osallistumispalkkioksi. Tulosten perusteella näyttää siltä, että arviointitilanteessa motivaatioon voidaan vaikuttaa palkkioin vain hyvin rajallisesti. Palkkio nostaa vastaamishalukkuutta mutta ei niinkään vastaamisen laatua. Syvemmälle juurtuneet negatiiviset asenteet esimerkiksi oppiainetta, tekstitaitoja tai omaa osaamista kohtaan jäivät tämän vaikutuksen tavoittamattomiin, eikä palkkio

luonnollisestikaan voi testitilanteessa tasoittaa aikaisemmin syntyneitä taitoeroja.

Aikuisten testi- ja kirjoitelma-aineisto on kerätty ensin kolmessa Etelä-Suomen kaupungissa vuosina 2012, 2013 ja 2014. Nämä koehenkilöt eivät kaikki asu kyseisissä kaupungeissa vaan tulevat ympäri Suomea. Osa on ammattikorkeakouluopiskelijoita, osa on osallistunut suomen kielen pääsykokeeseen. Myöhempiin testeihin ja kielentämistehtäviin osallistuneet aikuiset on rekrytoitu pääosin korkeakouluopiskelijoista tai korkeakouluun pyrkijöistä, mutta jonkin verran mukana on myös muita vapaaehtoisia. Koehenkilöistä ammattikorkeakoululaisille maksettiin pieni osallistumispalkkio. Myös aikuisaineistossa taso vaihtelee, mutta otos ei edusta koko nuorten aikuisten ikäluokkaa, koska koehenkilöt ovat pääosin suorittaneet ylioppilastutkinnon ja he ovat vapaaehtoisia. Pääsykoeryhmä on nuorten aikuisten ryhmistä monimuotoisin, koska ryhmä ei keruun hetkellä koostunut korkeakouluopiskelijoista. Pääsykokeeseen osallistuneista kymmenisen prosenttia läpäisi pääsykokeen, muiden osalta ei ole tietoa myöhemmistä valinnoista. Ammattikorkeakoululaiset edustavat hyvin monia eri koulutusaloja humanistisista aloista terveydenhoitoon ja tekniikkaan, mikä lisää tulosten edustavuutta.

Yksittäisten testien koehenkilömäärät ovat vaihdelleet 20–600. Ryhmäkohtaisesti on yleensä pyritty 100 koehenkilöön. Peruskoulun kuudennen ja yhdeksännen luokan joukko, joka osallistui neljän testin sarjaan, käsitti yli 150 oppilasta. Yksi testeistä oli kirjoitettava kielentämistehtävä. S1- ja S2-koehenkilöiden kielitietotesteihin osallistui noin 650 alakouluikäistä, heistä 550 kirjoitti myös Unelmakirjoitelman. Lisäksi aineistossa on noin 50 kirjoitelmaa muilta samojen luokkien oppilailta. Sähköisiin semanttisiin ja morfologisiin testeihin osallistui yhteensä 300 koehenkilöä. He jakautuivat kuudes- ja yhdeksäsluokkalaisiin sekä nuoriin aikuisiin. Myös nämä koehenkilöt kirjoittivat Unelmakirjoitelman. Osa-aineistot pääsykoeaineistoa lukuun ottamatta sisältävät suunnilleen saman verran mies- ja naiskirjoittajia.

Tutkimukseen on hankittu tarvittavat luvat, jotka pyydettiin yksittäin usealta eri taholta. Alle 15-vuotiaiden vanhemmilta on pyydetty lupa lasten osallistumiseen, ja kaikkien koululaisten osalta on pyydetty lupa alueviranomaisilta, koulujen rehtoreilta ja luokkien opettajilta. Jokainen

koululainen on osallistunut oppitunnilla tehtyyn testiin muiden tavoin, mutta hän on voinut kieltää aineistonsa käytön merkitsemällä kiellon koepaperiin. Kieltäytyjiä oli muutamia ja heidän aineistoaan ei ole sisällytetty tutkimukseen, vaan se on hävitetty. Aikuisten tuottamien aineistojen keruuseen on pyydetty suostumus paitsi koehenkilöiltä itseltään myös kouluviranomaisilta, alueviranomaisilta ja koulujen rehtoreilta, jos keruu on toteutettu koulun tiloissa.

Koehenkilöiltä on pyydetty paperitesteihin nimi. Osa testeistä on toteutettu sähköisesti, ja tällöin testiohjelma on antanut koehenkilölle yksilöllisen nimikoodin, jonka avaimella testauksen eri osat pystyttiin yhdistämään. Koehenkilöiden nimiä ei ole liitetty aineistotietokantoihin eikä niitä ole säilytetty. Osa oppilaista on alkuaankin identifioinut itsensä vain etunimellä. Koulujen nimiä tai koulupaikkakuntia ei mainita tietokannoissa. Aineisto on myös muuten anonymisoitu. Kaikki henkilönimet on kirjoitelmiin muutettu mutta yleisimmät paikannimet, huiipaikkojen nimet ja vastaavat ovat muuttamattomia, jos ne eivät viittaa koulujen sijaintikuntiin tai lähikuntiin.

Aineiston keruun aikana vuoden 2015 tasa-arvolain uudistus, joka laajensi sukupuoleen kohdistuvan syrjinnän estämisen myös identiteetin ja sukupuolen ilmaisuun, ei ollut vielä voimassa. Koehenkilöt eivät siten voineet tehdä kyselyn sukupuolta koskeviin taustatietoihin mahdollisesti haluamiaan valintoja. Tämä ei ole sikäli ongelma, että keneltäkään koehenkilöltä ei kysyty tietoa sukupuolesta, vaan aineisto on jaoteltu myöhemmin etunimien perusteella nais- ja mieskoehenkilöiden tuottamaksi, jotta väitetyt kirjoittamistaitoeroja voidaan tutkia. Aineistoon ei osunut sukupuolineutraaleja nimiä, joten jaottelu ei tuonut ongelmia. S2-aineiston keruun yhteydessä oppilaan sukupuoli on koodattu oppilaan tapaamisen tai ilmoitetun tiedon perusteella. Monet aikuiskirjoittajat pohtivat kirjoitelmassaan identiteettiään, mutta näitä pohdintoja ei käsitellä erikseen kuten ei muutenkaan kirjoitelmien yksityisempiä sisältöjä. Tutkimuksen alkuperäistä paperiaineistoa tai sähköisiä testejä ei ole säilytetty. Aineistosta kootut tietokannat ovat sen tutkijan hallussa, jolla niihin on tekijänoikeus. Vanhimpaan aineistoon ei ole keruun yhteydessä pyydetty lupaa aineiston jatkokäyttöön, joten tämä osa aineistosta on jatkossakin yksityistä.



## LUVUT

Teoksen luvut jakautuvat kahteen temaattiseen osioon, joista ensimmäisessä lukujen analyysit lähtevät merkityksestä ja toisessa muodosta. Perusteena on kielellisen eron lisäksi erilainen kielenhallinnan ja kielenkehityksen kulku. Esittelemme lyhyesti teoksen sisältöä kyseisen kahtiajaon ja itse lukujen argumentaation kannalta.

## MERKITYKSESTÄ MUOTOON JA MUODOSTA MERKITYKSEEN

Teoksen luvut on jaettu kahteen temaattiseen osioon. Ensimmäisessä osiossa näkökulma ja analyysin lähtökohta on semanttinen. Kielenhallintaa käsitellään näissä luvuissa sen mukaan, miten peruskoululaiset ja nuoret aikuiset ilmaisevat sanamerkityksiä ja miten he kielentävät semanttis-pragmaattisia merkityksiä. Näitä merkityksiä luonnollisesti ilmaistaan muodoilla, mutta muodot saavat oikeutuksensa vasta merkityksien määrittelyn kautta. Luvut on järjestetty kehityksellisesti vaikeutuvaan järjestykseen siten, että I osio alkaa sanamerkityksistä (Pajunen & Honko) ja laajenee sitten spatiaalisen merkityksen ilmaisuun (Tuuri) ja edelleen referentiaalisten viittauksien (Kouvo, Honko & Pajunen) ja moniäänisyyden ilmaisuun (Niinimäki). Sanamerkityksissä kehityksellinen linja kulkee asteikoilla yleisyys, konkreettisuus ja spesifisyys. Toisin sanoen iän mukaan käytössä olevien sanojen määrä kasvaa ja niiden yleisyys laskee, merkitykset muuttuvat abstraktimmiksi (samasta teemasta kirjoitettaessa) ja sanamerkitykset spesifisoituvat, niiden granulariteetti eli merkityksen täsmällisyys kasvaa. Nämä yleistykset koskevat myös spatiaalisen merkityksen ilmaisun hallintaa, mutta spatiaalisilla merkityksillä luodaan myös koherenssia ja ilmaisu laajenee sosiokulttuuriseen ja yksilöpsykologiseen suuntaan. Referentiaaliset ilmaukset lisääntyvät kertomuksellisuuden kehittyessä; vasta samojen diskurssireferenttien riittävän moni uudelleenmaininta lukijan kannalta identifioitavasti on (yksi) merkki kertomusrakenteen hallinnasta. Jos yksiäänisen diskurssin viittaussuhteiden ilmaisu on hallinnassa vasta alakoulun ylemmillä luokilla, moniäänisen kerronnan kerrostuneiden viittaussuhteiden hallinta kehittyy vielä hitaammin.

Semanttisten ilmiöiden kielentäminen edellyttää niiden semanttisen luonteen ja erilaisten funktioiden ymmärtämistä ja monien hierarkkis-

ten tai vaihtoehtoisten kielentämisen keinojen hallintaa. Niiden kielentämisessä kuvausta ei voi perustaa yhden ryhmän analyysien varaan, koska polku aikuisilmaisuuon on niin mutkikas: usein ilmaisu, joka pie-nellä alakoululaisella on kehityksellinen saavutus, on yläkoululaisella heikko suoritus ja aikuiselta jo kadonnut.

Toisessa osiossa lukujen näkökulma ja analyysin lähtökohta on kirjoitel-mien muodollisissa yksiköissä. Näillä muodoilla luonnollisesti ilmaistaan merkityksiä, mutta muotolähtökohta rajaa analyysin näkökulmaa esimerkiksi kompleksistumisasteikon mukaisesti. Kielenhallintaa käsi-tellään sen mukaan, miten lapset, nuoret ja nuoret aikuiset hallitsevat morfologisia ja syntaktisia muodollisia yksikköjä ja miten hallinta kehi-tyksellisesti muuttuu. Osio on järjestetty muotojen yksiselitteisyyden mukaiseen järjestykseen siten, että ensimmäinen luku (Laasanen & Pajunen) käsittelee muodollisesti hyvin yksiselitteisiä possessiivisuffik-seja ja pyrkii selvittämään paitsi käytön määrää, myös sen suhdetta pu-hutusta kielestä esitettyihin suffiksien katoamisennusteisiin. Toisessa luvussa (Honko) analysoidaan alakouluikäisten kirjoitelmien lähinnä sanatasolla esiintyviä normipoikkeamia, ja niiden iän mukaista vähene-mistä sekä vähenemisen suhdetta kielenhallintaan suomea toisena kiele-nään puhuvilla. Kolmas ja neljäs luku (Pajunen & Vainio) käsittelevät kompleksistumisasteikolla morfologisia ja syntaktisia yksikköjä – sana-rakennetta, sanaluokkajakaumiin perustuvia indeksejä sekä lauseraken-etta. Tavoitteena on selvittää, miten sanatasolla ja kertovan rakenteen tasolla mitattavat yksiköt hallitaan ja mitä muutosta niissä iän mukaan tapahtuu. Muotojen kompleksistuminen tarkoittaa käytännössä niillä ilmaistujen merkityksien tiiviimpää – ja abstraktimpaa – esittämistä. Tiivistyminen puolestaan nostaa kertovan rakenteen globaalikoherenssia.

Jos I osiossa kehitys on hyvin hidasta, II osiossa se on nopeampaa. Keskeinen ero on myös, että I osiossa analyysi on enemmän laadullista, II osion muotoja voidaan helpommin poimia ja mitata ja mittauksen avulla osoittaa ikäkausittaisia osaamisen mediaaneja, koska tiivistyvien muotojen hallinta näkyy kertovan tekstin samanlaistumisena toisin kuin merkitysten ilmaisujen hallinta, joka näkyy vaihtoehtojen monipuolis-tumisena. Ero johtuu jo yksinkertaisesti siitä, että muototasolla on yksik-köjä suhteellisen vähän, merkitystasolla niitä on paljon. Ajatus merkityk-

sen ja muodon epäsymmetriasta ei ole tietenkään uusi, mutta tässä se on pyritty esittämään täsmällisesti.

#### LUKUJEN ABSTRAKTIT

Myöhempi kielenkehitys: sanatiedosta rakenne- ja tekstitietoon  
Anneli Pajunen ja Mari Honko

Johdantona toimiva ja tutkimusta taustoittava luku esittelee Myöhemmän kielenkehityksen tutkimushankkeen, jonka yhteydessä koko teos – sekä lukuisat opinnäytteet – on kirjoitettu. Pajunen ja Honko määrittelevät teoksen keskeisen käsitteistön, yhteiset tutkimuskysymykset, aineiston ja koehenkilöt, teoksen rakenteen ja yksittäisten lukujen teeman. Teoksen keskeisenä teemana on suomen kielen hallinta lapsilla, nuorilla ja nuorilla aikuisilla ja sen kehityksellinen eteneminen peruskouluikäisillä verrattuna nuoriin aikuisiin. Kielenhallinnan kehittymistä arvioidaan merkityksien kielentämisestä ikäkausittain ja toisaalta muotojen kielentämisestä. Merkitystason hallintaa tarkastellaan sanastotiedon ja semanttis-pragmaattisten ilmiöiden kielentämisen kannalta ja muotojen hallintaa morfologian ja syntaksin kautta.

#### I Sanat ja merkitykset

Osion I analyysiluvut vastaavat etenkin tutkimuskysymyksiin 1 (sanasto ja semanttinen tieto) ja 4 (kehityksellinen, kielellinen ja yksilöllinen variaatio).

Unelmieni päivä -sanasto kehityksellisestä näkökulmasta  
Anneli Pajunen ja Mari Honko

Pajunen ja Honko tarkastelevat Koulu- ja Unelmakorpuksen kirjoittajien käyttämää sanastoa laadullisesta ja määrällisestä näkökulmasta. Koehenkilöiden ja kirjoitelmien määrä vaihtelee tarkastelussa vajaasta tuhannes-

ta lähes pariin tuhanteen. Tarkimmin on analysoitu niiden koehenkilöiden kirjoitelmia, jotka osallistuivat myös semanttisen ja morfologisen prosessoinnin testaukseen. Sanatiedon osalta sukupuolierot eivät ole osoittautuneet kovin suuriksi, ja sukupuoli ei siksi ole kaikissa sanasto-analyyseissä mukana. Frekvenssilaskelmat osoittavat, että koululaiset käyttävät paljon kielen yleisintä sanastoa ja että ero on suurempi peruskoululaisten ja aikuisten välillä kuin peruskouluryhmien. Tämä voi johtua vahvasta koulukirjoittamisen konventiosta mutta heijastaa tietysti myös aikuisten laajempaa elämänpiiriä ja kokemusta. Diversiteettierot kasvavat aineistossa kehityksellisesti, eli käytetty sanasto monimuotoistuu karttuvan iän ja kielenkäyttökokemuksen myötä. Sukupuolen mukainen ero on melko pieni – mikäli sitä ylipäättään voidaan testata – ja aikuisilla toisensuuntainen kuin peruskoululaisilla. Semanttinen analyysi osoittaa, että eri ikäluokat kirjoittavat samasta aiheesta ensisilmäyksellä samoilla aineksilla vaikka täsmällisemmin. Tarkempi analyysi osoittaa, että yleistasolla on jatkuva siirtymä konkreettisesta abstraktiin, joten elollisista ja konkreettisista elottomista kirjoittaminen vähenee iän mukaan, mutta että lisäksi abstraktistumiskehitys kulkee konkreettisiin tarkoitteisiin viittaavien ryhmässä siten, että peruskoululainen kirjoittaa tietyn luokan kuten syötävän aineen esiintymästä (*karkki, limsa, sipsi, pizza*), aikuinen enemmän itse luokan yleistasolla (*ruoka, juoma, vesi*). Semanttinen analyysi osoittaa kirjoitelmat myös selvän kertoviksi: tapahtumaverbit ja ajan adverbit vallitsevat.

### Liiketilanteen väylän ja suunnan kielentäminen suomessa Emilia Tuuri

Tuuri keskittyy spatiaalisuuden alaan kuuluvien liikkeen väylän ja suunnan kielentämiseen nuorilla aikuisilla. Jos ryhmätason hajonta yleensä kehityksellisesti vähenee, yksilötason variaatio toisaalta kasvaa, ja se kasvaa erityisesti spatiaalisessa kontekstissa. Siinä vaikuttavia tekijöitä ovat tila ja liike, liikkeen rajattuus, kielentämisen resurssit, tapahtumien havaitsemiseen liittyvät tekijät ja lisäksi yksilölliset kielentämisstrategiat. Aineistona on videolle kuvattuja lyhyitä liiketapahtumia eli videostimulusia ja niiden kielentäjinä 50 nuorta aikuista. Väylä ja suunta eroavat

toisistaan juuri rajattuuden kannalta: väylän ilmauksissa liikeradalle ilmaistaan alkupiste (esimerkiksi *ulos talosta*), päätepiste (esimerkiksi *sisään taloon*) tai keskivaihe (esimerkiksi *pellon poikki*), joka rajaa muuttujan liikkeen suhteessa kiintopisteeseen. Suunnan ilmauksissa (esimerkiksi *taloa kohti*, *poispäin*) liikettä ei rajata kielellisesti. Luku keskittyy erityisesti väylän keskivaiheeseen ja suuntaan, jotka asettuvat monien videostimulusten kohdalla juuri rajattuuden kannalta vaihtoehtoisksi tavoiksi jäsentää liiketilannetta: käveleekö poika *sillan yli* vai *siltaa pitkin*, nouseeko nainen *portaat* vai *portaita ylös*? Koehenkilöt ilmaisevat videostimuluksissa näkyviä liiketilanteita joko melko yhdenmukaisesti tai hyvin varioiden, ja Tuuri analysoi myös tähän variaatioon liittyviä taustatekijöitä. Käy myös ilmi, että yhdeksäsluokkaisilla on pääosin samat resurssit jo käytössään kuin nuorilla aikuisilla mutta kuudesluokkaisilla ei vielä ole. Kielitypologiselta kannalta Tuuri osoittaa, että suomessa kielennetään väylää ja suuntaa rikkaasti moniin muihin kieliin verraten. On odotuksenmukaista, että tällainen moninaisuus kehittyi verrattain hitaasti.

Referentiaaliset määräiset ilmaukset ja niiden suhde kertomusrakenteeseen alakouluiässä

Anna-Kaisa Kouvo, Mari Honko & Anneli Pajunen

Kouvo, Honko ja Pajunen tarkastelevat kertovan tekstin koherenssin eli tekstuaalisen yhtenäisyyden syntyä. Luku keskittyy referentiaaliseen koherenssiin ja tutkii kertomuksellisuuden kehittymistä tarkastelemalla tarkoitteiden määräisyyden ilmaisemisen muutoksia alakouluikäisten teksteissä. Nuorimmat alakouluikäiset kirjoittavat atomistisia tekstejä, mikä näkyy muun muassa siten, että tekstiin tuodaan koko ajan uusia tarkoitteita, referenttejä, mutta niistä ei ole enempää sanottavaa. Iän myötä samoista tarkoitteista kerrotaan pidempään eli kerrotaan useammista toisiinsa linkittyvistä tapahtumista samojen henkilöiden tai asioiden näkökulmasta. Samalla kun samoihin tarkoitteisiin viittaaminen lisääntyy, lisääntyvät myös keinot osoittaa kielellisin keinoin, että kyse on jo tekstiin esitellyistä tarkoitteista eli tutusta tiedosta. Yksijäseniset ilmaukset korvautuvat toisin sanoen erityyppisillä määritteillä tarkenne-

tuilla monijäsenisillä ilmauksilla. Luku osoittaa myös, että puhutun kiel-  
len keinot ilmaista spesifisyyttä artikkelinomaisella *se*-pronominilla eivät  
kuulu alakouluikäisten kirjoitettuun kieleen. Jo nuorimmilla koululaisil-  
la on siis tässä suhteessa taito erottaa puhuttu ja kirjoitettu kielimuoto  
omassa kirjoittamisessaan, jos kirjoittaminen ylipäätään vielä onnistuu.

Moniäänisyyden ilmaisun kehittyminen peruskoulun päättövaiheiden  
kirjoitelmissa

Anneli Niinimäki

Niinimäellä tarkastelun kohteena on moniäänisyyden tuottaminen kir-  
joitettuun kieleen. Moniäänisyys on semanttis-pragmaattisesti komplek-  
sinen ilmiö, jonka avulla ilmaistaan kerronnan kielellistä moniaineksi-  
suutta. Moniäänisyyttä on tekstiin rakennettu moninäkökulmaisuus ja  
tulkintojen vaihtoehtoisuus. Muutokset kertojan tai puhujan osallistuja-  
roolissa rakentavat moniäänisyyttä, samoin etääntyminen omasta ker-  
ronnasta ja viittaukset tekstin ulkopuolelle. Moniäänisyyden kielentä-  
mistä opitaan puhutussa kielessä melko varhain, mutta sen hallinta  
kirjoitetussa kielessä etenee hitaasti. Tämä prosessi voi olla kesken vielä  
peruskoulun päättövaiheissa. Kehitys suhteutuu kielellisen koherenssin  
luomiseen ja kertomusrakenteen tai yleensä tekstin rakentamisen oppi-  
miseen. Moniäänisyyden tyypeiksi on alakoululaisaineiston analyysin  
avulla määritelty intertekstuaalisuus, metakieliset viittaukset tapahtuma-  
kontekstin ulkopuolelle, suora ja epäsuora esitys sekä vapaa epäsuora  
esitys muunneltuihin. Varsinaisena aineistona on kuudes- ja yhdeksäs-  
luokkalaisten Unelmakirjoitelmia. Kouluikäisillä tarkastelun keskiöön  
nousee referointi, jonka kautta kirjoittajat tuovat kertovaan tekstiinsä  
toisen äänen. Kuudesluokkalaisilla referoinnin funktio on tyypillisesti  
kerrontaa eteenpäin vievä, ei pohdintaan liittyvä, kerronnan pysäyttävä.  
Yhdeksäsluokkalaisilla taas funktiona on episteeminen modaalisuus,  
minäkertojan ajatuksien ja tuntemuksien kuvaaminen. Samalla suoran  
esityksen määrä vähenee ja epäsuoran kasvaa. Toinen tyypillinen moni-  
äänisyyttä luova ilmiö on kertomuksen metatasoon kuuluva vuorovaiku-  
tuksen luominen lukijaan. Referoinnin taidoissa ja vuorovaikutuksen  
rakentamisessa on puutteita vielä peruskoulun päättövaiheissa. Osalla

kirjoittajista teksti kuudennella vuosiluokalla ja vielä peruskoulun päätösvaiheessa on täysin yksiäänistä. Referoitua puhetta ei eroteta muusta kerronnasta, ja vuorovaikutuksen taidot puuttuvat. Toisaalta on myös joukko kirjoittajia, jotka osaavat lähestyä tarkoituksenmukaisesti lukijaa ja luoda kokemuksiinsa perustuvan kuvan omasta sisäisestä ajatusmaailmastaan.

## II Muodot ja rakenteet

Toisessa osiossa vastataan erityisesti tutkimuskysymyksiin 2 (muotojen hallinnan kehitys) ja 3 (sukupuolierot) ja osin 4 (kehityksellinen variaatio).

Meneillään oleva kielenmuutos puhutusta kielestä kirjoitettuun ja suomen possessiivisuffiksi  
Mikko Laasanen & Anneli Pajunen

Laasanen ja Pajunen tarkastelevat puhutun ja kirjoitetun kielen modaaliteettieroja ja toisaalta niiden samanlaisuutta ja erilaisuutta ja toteavat, että vaikka kirjoitetun kielen kehitys on alisteinen puhutulle myöhemmän kielenkehityksen alkuvaiheissa, myöhemmin suhde kehittyy vuorovaikutukselliseksi. Argumentaatio rakentuu erityisesti puhutun kielen perusteella meneillään olevan muutoksen – possessiivisuffiksien käytön vähenemisen ja mahdollisen kadon – varaan ja keskeinen tarkasteltava luvussa on, siirtykö puhutussa kielessä meneillään oleva muutos suoraan tai viipeellä kirjoitettuun kieleen. Keskeistä on siis, miten pitkälle kompleksisten ilmiöiden kielentämisen yhteydessä oleva kirjoitetun kielen muotoyksikkö saa vaikutuksia puhutusta kielestä. Tarkastelu ei luvussa ole kehityksellistä, jotta kehityksellinen vaihtelu ja muutos eivät sekoittuisi, vaan perustuu nuorten aikuisten eriaikaisiin kirjoitelmiin. Joitain kehityksellisiäkin havaintoja kyllä esitetään. Ajallinen ulottuvuus kattaa noin 70 vuotta 1940-luvulta 2000-luvulle. Vertailuaineistona on lehtikieltä, fiktiota ja oppikirjatekstejä. Possessiivisuffiksien käyttöä arvioidaan määrällisesti ja sanaluokkakohtaisesti. Lisäksi tarkastellaan suffiksien käyttöä niin sanottujen indeksisanojen eli eroavaa tai eroamatonta

ta omistusta ilmaisevien sanojen yhteydessä. Koska kyse on inherentisti tai juridisesti omistettavia objekteja ilmaisevista sanoista, voi olettaa, että possessiivisuffiksien käyttö juuri niiden yhteydessä olisi välttämättöminä ja vähäinen tai entistä vähäisempi käyttö ilmaisisi helposti muutoksen. Tutkimus osoittaa, että kyseisenä ajanjaksona possessiivisuffiksien käytössä ei ole selvää ajallista muutosta, eikä käyttö myöskään aineiston eri osissa ole vähentynyt. Aineisto osoittaa sen sijaan kiistatta, että kirjoitetun kielen eri osa-aineistoissa possessiivisuffikseja käytetään eri määrin ja eri tavoin. Indeksisanoissa käyttö on yleisintä, muuten kyse on tekstilaji- ja tyyppieroista.

Kirjoittamisen tarkkuuden kehittyminen ja yhteys kirjoittajan kokonaiskielitaitoon alakouluiässä  
Mari Honko

Honko keskittyy tarkastelemaan norminmukaisen kirjoitetun kielen haltuunottoa eli kirjoittamisen tarkkuuden kehittymistä alakouluikäisillä koehenkilöillä ja vertaa suoriutumista kielenhallintaa mittaaviin testiartoihin, todistusnumeroihin ja opettaja-arviointeihin. Koehenkilöinä on yhteensä 600 suomea äidinkielenä tai toisena kielenä puhuvaa. Kaikki suomea toisena kielenä puhuvat ovat Suomessa syntyneitä tai pitkään ennen kouluikää Suomessa asuneita. Hongon tutkimus on pitkittäistutkimus, ja samoja koehenkilöitä on testattu kirjoituttamalla Unelmakirjoitus kahtena eri ajankohtana, ensin alakoulun alkuvaiheessa ja myöhemmin loppuvaiheessa. Muuttajat ovat äänne- ja muotorakenteesta, mutta tarkastelussa on mukana myös sana- ja lausetason piirteitä. Äännetasolla kirjoittamisen kehittyminen edellyttää äänne–kirjainvastaavuuden eli auditiivisen diskriminaation kehittymistä, toisin sanoen kykyä analysoida puhutun kielen yksiköitä. Alakouluikäisillä norminmukaisuus kehittyi selvästi sekä puhutun kielen että kirjoitetun kielen vaikutuksessa; kirjoittamisessa edellinen heikkenee iän myötä, jälkimmäinen voimistuu. Aineistosta kuitenkin nousee oletuksen vastaisesti pääosin pojista koostuva ryhmä, jonka tulokset viittaavat lieviin kielihäiriöihin, vaikka kielihäiriödiagnosi puuttuu. Honko osoittaa, että norminmukaisuuden kehittyminen on sidoksissa kielenhallintaan. Norminmukaisuus-



den astetta voisi siten pitää alustavana kielenhallinnan kriteerinäkin, jos kirjoittamiskonteksti, tekstin luonne ja muut tekijät otetaan huomioon. Tulosten pohjalta voi myös arvella, että aina heikot kirjoittamistaidot eivät riipu pelkästään motivaation puutteesta.

Kielen rakenteen hallinta peruskoululaisilla ja nuorilla aikuisilla  
Anneli Pajunen ja Seppo Vainio

Pajunen ja Vainio keskittyvät teoksen kahdessa loppuluvussa kielenhallinnan kehitystrendeihin, joita voi kuvata asteikolla yksinkertaisesta – eli puhutun kielen kaltaisesta – rakenteesta kompleksiseen rakenteeseen. Ensimmäisessä kuvaus etenee sanarakenteesta lauserakenteeseen ja osoittaa, että sanojen pidentyessä ja lauserakenteiden kompleksistuessa tekstit tiivistyvät. Samalla niiden rakennetta kuvaavissa indekseissä tulee näkyviin ikäkausikohtaisia tavanomaisen suoriutumistason mediaaneja, joiden avulla voi arvioida ikäluokkatasoista osaamista ja kielenhallintaa. Nuorimmilla alakouluikäisillä osaamisen hajonta on hyvin suurta, mutta jo kuudesluokkalaisilla muuttujille on laskettavissa osaamisen tavanomaista tasoa ilmaisevia mediaaneja ja myös kehitys seuraavaan ikäluokkaan on ennustettavissa. Aineisto tosin osoittaa, että monissa muuttujissa alakvartiiliosuudet ovat normaalijakauman odotusarvoihin nähden liian suuret ja että monilla peruskoulun pojilla muuttujat sijoittuvat suoralle viivalle nousevan sijaan. Tämä on huolestuttava tulos, vaikka syitä ei saa aineistolla esiin. Kaikkiaan muotoon keskittyvät kirjan luvut osoittavat, että kielen hallinta etenee ikäluokasta toiseen melko tasaisesti paljastaen ikäkausiosaamisen keskitason. Puhutun kielen rakenne on pitkälti kehityksen lähtökohtana, mutta sen vaikutus ei ole kehitystä rajaava tekijä. Voi tosin arvella, että jos jo puhutun kielen hallinnassa on ongelmia, ne heijastuvat myös kirjoitetun kielen hallintaan. Pajunen ja Vainio arvioivat kirjoitelmiä muodollisten ja rakenteellisten mittarien avulla.

Syntaktista rakennetta tiivistävät ja kompleksistavat tekijät  
peruskoululaisilla ja nuorilla aikuisilla  
Anneli Pajunen ja Seppo Vainio

Pajunen ja Vainio osoittivat edellisessä luvussa, että kirjoitetun kielen syntaktinen rakenne tiivistyy iän myötä. Nyt he erittelevät tarkemmin, mikä ominaisuus milloinkin vaikuttaa syntaktisen rakenteen kompleksisuuteen, joka selvästi kasvaa ikäryhmittäin. Toisin sanottuna, mikä nostaa eri kompleksisuusmuuttujien arvoja ja mitä muodon hallinnan kannalta näistä ominaisuuksista voidaan päätellä. Sukupuolen mukaan arvioiden kompleksisuuserot ovat vähäisempiä kuin monet muut määrälliset erot. Keskeisin ero näyttää olevan se, että mieskirjoittajat tuottavat hieman kompleksisempää tekstiä jo kuudennelta luokalta lähtien. Aikuisryhmässä erityisesti miehille ominaista on tiivistää kerrontaa lauseenvastikkeita käyttäen. Kertova teksti on miehillä usein tiivistahtinen tapahtumasarja, jossa erityisesti temporaalirakenteilla taustoitetaan ja taikaistetaan tapahtumia. Myös naiskirjoittajilla on paljon tapahtumasarjakirjoitelmia, mutta tekstin kulku on toinen. Jos sukupuolieroja mitataan sanamäärinä, erot ovat naisten hyväksi, mutta jos muotoeroja katsotaan tarkemmin, tilanne ei olekaan aivan yksinkertainen. Aineistosta nousee näkyviin myös yhdeksännen luokan poikien kirjoitelmia, joissa muuten nousevana linjana näkyvä kehitys on monen mittarin kautta mitattuna suoralla viivalla: arvot ovat usein samat kuin kuudennen luokan pojilla. Äkkipitäen näyttää, että mitään kehitystä ei olisi kielenhallinnassa tapahtunut, mutta tämä ei voi olla koko totuus, koska elektronisissa testeissä tulokset ovat linjassa ikäluokkaan ja osin parempia kuin tytöillä. Ero voi liittyä myös ymmärtämisen ja tuottamisen väliseen eroon, sillä nämä kehittyvät eriaikaisesti. Motivaation puute on yksi taustatekijä, mutta se vaikuttaa usein enemmän tekstin koherenssiin kuin muotoyksiköihin. Väestötasolla täytyy tietysti huomata, että noin 2–3 prosentin alisuoriutuminen on odotuksenmukaista.

## Lähteet

- Ahvenainen, O. & Holopainen, E. 2000. *Lukemis- ja kirjoittamisvaikeudet: Teoreettista taustaa ja opetuksen perusteita*. Jyväskylä: Special Data.
- Babayigit, S. 2015. The dimensions of written expression: Language groups and gender differences. *Learning and Instruction* 35, 33–41.
- Baker, L. & Levin D. 2015. The role of relational triggers in event perception. *Cognition* 136, 14–29.
- Bandura, A. 1997. *Self-efficacy: The exercise of control*. New York: Freeman.
- Berman, R. A. 2004. Between emergence and mastery: The long developmental route of language acquisition. Julkaisussa: R. A. Berman (toim.) *Language development across childhood and adolescence*. Amsterdam: John Benjamins, 9–34.
- Berman, R. A. 2007. Developing linguistic knowledge and language use across adolescence. Julkaisussa: E. Hoff & M. Shatz (toim.) *Blackwell handbook of language development*. Oxford: Blackwell Publishing, 347–367.
- Berman, R. A. & Slobin, D. 1994. Narrative structure. Julkaisussa: R. A. Berman & D. Slobin (toim.) *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale, NJ.: Lawrence Erlbaum Ass., 39–84.
- Bertram, R., Laine, M. & Virkkala, M. 2000. The role of derivational morphology in vocabulary acquisition: Get by with a little help from my morpheme friends. *Scandinavian Journal of Psychology* 41, 287–296.
- Blomberg, J. 2014. *Motion in language and experience: Actual and non-actual motion in Swedish, French and Thai*. Lund: Lundin yliopisto.
- Bohnemeyer, J., Donelson, K. T., Tucker, R. E., Benedicto, E., Capistrán Garza, A., Eggleston, A., Hernández Green, N., de Jesús Selene Hernández Gómez, M., Herrera Castro, S., O’Meara, C. K., Palancar, E., Pérez Báez, G., Polian, G. & Romero Méndez, R. 2014. The cultural transmission of spatial cognition: Evidence from a large-scale study. *Proceedings of the 36th annual meeting of the cognitive science society*, 212–217.
- Bouwer, R., Béguin, A., Sanders, T. & van den Bergh, H. 2015. Effects of genre on the generalizability of writing scores. *Language Testing* 32: 1, 83–100.
- Camarata, S. & Woodcock, R. 2006. Sex differences in processing speed: Developmental effects in males and females. *Intelligence* 34, 231–252.
- Carlisle, J. F. 2000. Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing* 12, 169–190.
- Citron, F., Cacciari, C., Kucharski, M., Back, L., Conrad, M. & Jacobs, A. 2016. When emotions are expressed figuratively: Psycholinguistic and affective norms for 619 idioms for German. *Behavior Research Methods* 48, 91–111.
- Clark, E. 2014. Acquisition of derivational morphology. Julkaisussa: R. Lieber & P. Štekauer (toim.) *The Oxford handbook of derivational morphology*. Oxford: Oxford University Press, 424–439.
- De Zeum, E. L., van Beijsterveldt, C. E. M., Glasner, T. J., de Geus, E. J. C. & Boomsma, D. I. 2016. Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children. *Learning and Individual Differences* 47, 156–166.

- E-Prime-3. Psychology Software Tools. <https://pstnet.com>. Viitattu 31.1.2020.
- Doherty, M. J. 2009. *Theory of mind: How children understand others' thoughts and feelings*. Hove: Psychology Press.
- Gibbs, R., Davies, G. M. & Chou, S. 2016. A systematic review of factors affecting the likelihood of change blindness. *Crime Psychology Review* 2: 1, 1–21.
- Givón, T. 1993. *Coherence in text, coherence in mind*. Amsterdam: John Benjamins.
- Givón, T. 2017. *The story of zero*. Amsterdam: John Benjamins.
- Habermas, T. & de Silveira, C. 2008. The development of global coherence in life narratives across adolescence: Temporal, causal and thematic aspects. *Developmental Psychology* 44: 3, 707–721.
- Harjunen, E. & Rautopuro, J. 2015. *Kielenkäytön ajattelua ja ajattelun kielentämistä. Äidinkielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2014: keskiössä kielentuntemus ja kirjoittaminen*. Helsinki: Kansallinen koulutuksen arviointikeskus.
- Harjunen, E., Marjanen, J. & Karlsson, J. 2019. *Äidinkielen pieni pitkittäisarviointi 2014–2017: Perusopetuksen päätöstä lukion päättöön*. Helsinki: Kansallisen koulutuksen arviointikeskus.
- Heilig, M. 2019 [2018]. *Naisen aivot, miehen aivot: Onko aivoilla sukupuolta?* Kääntäjä: S.-A. Ahvonen [Alkuteos *Hon, han och hjärnan*]. Jyväskylä: Atena.
- Heimo, H. 2003. ”Kaikki on enemmän kuin joka toinen”: *Oppimisvaikeudet ja suhdekäsitteet esiopetuksessa*. Helsinki: Helsingin yliopisto.
- Heinonen, S. 2018. *Narratiivisuuden sisällöllinen kehittyminen aiheanalyysin kautta tarkasteltuna: Kvantitatiivinen tutkimus aiheiden esiintymisen vaihtelusta iän ja sukupuolen mukaan ”Unelmieni päivä” -kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Hoff, E. 2005. *Language development*. Belmont: Thomson Wadsworth.
- Honko, M. 2013. *Alakouluikäisen leksikaalinen tieto ja taito: Toisen sukupolven suomi ja S1-verrokki*. Tampere: Tampereen yliopisto.
- Honko, M., Jarvis, S. & Vainio, S. 2019. Lukijat sanaston monimuotoisuutta määrittämässä: Leksikaalisen diversiteetin tarkastelua määrällisen ja laadullisen tutkimuksen rajapinnassa. *Virittäjä* 123: 1, 44–78.
- Hooper, S., R., Costa, L.-B., McBee, M., Anderson, K. L., Yerby, D. C., Knuth, S. B. & Childress A. 2011. Concurrent and longitudinal neuropsychological contributors to written language expression in first and second grade students. *Reading and Writing* 24, 221–242.
- Hulstijn, J. 2015. *Language proficiency in native and non-native speakers: theory and research*. Amsterdam: Benjamins.
- Hulstijn, J. 2019. An individual-differences framework for comparing nonnative with native speakers: perspectives from BLC theory. *Language Learning* 69: 1, 157–183.
- Hyde, J. S. 2005. The gender similarity hypothesis. *American Psychologist* 60, 581–592.
- Hyde, J. S. & Linn, M. C. 1988. Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin* 104: 1, 53–69.
- Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Hakonarson, H., Gur, R. E. & Verma, R. 2014. Sex differences in the structural connectome of the human brain. *PNAS* 111, 823–828.
- Ishibashi, M., Kopecka, A. & Vuillermet, M. 2006. *Trajectoire: Matériel visuel pour élicitation*

- des données linguistiques*. Laboratoire Dynamique du Langage (CRNS/Université Lyon 2) – Fédération de Recherche en Typologie et Universaux Linguistiques, CRNS, France.
- Itkonen, E. 1978. *Grammatical theory and metascience*. Amsterdam: John Benjamins.
- Itkonen, E. 2005. *Analogy as structure and process: Approaches in linguistics, cognitive psychology and philosophy as science*. Amsterdam: John Benjamins.
- Itkonen, E. 2008. *Maailman kielten erilaisuus ja samuus. Osa I*. Turku: Turun yliopisto.
- Itkonen, E. 2009. *Maailman kielten erilaisuus ja samuus. Osa II*. Turku: Turun yliopisto.
- Itkonen, E. & Pajunen, A. 2010. *Empiirisen kielitieteen metodologia*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Jaakola, M. 2004. *Peruskoulun 2. ja 3. luokan oppikirjojen sanasto – dynamiikasta luettavuuteen*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Jakobson, R. 1968 [1940–1942]. *Child language, aphasia and phonological universals*. The Hague: Mouton.
- Jensen, A. R. 1998. *The g factor: The science of mental ability*. Westport: Praeger Publisher.
- Juvonen, R. 2004. *Kertomus ylioppilasaineesta: Kevään 1999 tehtävistä 23 kirjoitettujen aineiden tekstilajianalyysejä*. Suomen kielen pro gradu -tutkielma. Helsingin yliopisto.
- Juvonen, R. 2010. *Evaluoiva että-yhdyslause ja retoriset rakenteet suomenkielisissä ylioppilasaineissa. Virittäjä* 114, 39–70.
- Juvonen, R., Virtanen, M. & Voutilainen, E. 2012. Fennistisen tekstilajitutkimuksen suuntia. Julkaisussa: V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiililä & M. Virtanen (toim.) *Genre-analyysi: Tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 453–469.
- Kalliokoski, J. 2006. *Tekstilajin taju ja toisella kielellä kirjoittaminen*. Julkaisussa: A. Mäntynen, S. Shore & A. Solin (toim.) *Genre – tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura, 240–265.
- Kalliokoski, J. 2008. *Suomi toisena kielenä ja koulussa kirjoittaminen*. Julkaisussa: S. Routarinne & T. Uusi-Hallila (toim.) *Nuoret kielikuvassa: Kouluikäisten kieli 2000-luvulla*. Helsinki: Suomalaisen Kirjallisuuden Seura, 349–370.
- Karjalainen, M. 2008. *Yläkoululaiset tekstin rakentajina*. Julkaisussa: S. Routarinne & T. Uusi-Hallila (toim.) *Nuoret kielikuvassa: Kouluikäisten kieli 2000-luvulla*. Helsinki: Suomalaisen Kirjallisuuden Seura, 290–306.
- Kauppinen, M. & Marjanen, J. 2020. *Millaista on yhdeksäsluokkalaisten kielellinen osaaminen? Suomen kielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2019*. [https://karvi.fi/app/uploads/2020/08/KARVI\\_1320.pdf](https://karvi.fi/app/uploads/2020/08/KARVI_1320.pdf). Viitattu 20.11.2020.
- Kavé, G. & Halamish, V. 2015. Doubly blessed: Older adults know more vocabulary and know better what they know. *Psychology and Aging* 30, 68–73.
- Keith, T. Z., Reynolds, M. R., Patel, P. G., Ridley, K. P. 2008. Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock-Johnson III tests of cognitive abilities. *Intelligence* 36, 502–525.
- Kemper, S. & Sumner, A. 2001. The structure of verbal abilities in young and older adults. *Psychology and Aging* 16, 312–322.
- Keuleers, E., Stevens, M., Mandela, P. & Brysbaert, M. 2015. Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology* 68: 8, 1465–1492.
- Koivisto, V. 2013. *Suomen sanojen rakenne*. Helsinki: Suomalaisen Kirjallisuuden Seura.

- Koutsoftas, A. & Gray, S. 2012. Comparison of narrative and expository writing in students with and without language-learning disabilities. *Language, Speech, and Hearing Services in Schools* 43, 395–409.
- KTL 2017. Suomalaislapset edelleen parhaiden lukijoiden joukossa. Koulutuksen tutkimuslaitoksen tiedote 5.12.2017. <https://www.jyu.fi/ajankohtaista/arkisto/2017/12/tiedote-2017-12-05-12-10-32-101275>. Viitattu 6.8.2020.
- Kulju, P. & Mäkinen, M. 2009. Sanelukirjoituksen arviointi – tyyppivirheistä sanarakenteiden fonologiseen analyysiin. *Virittäjä* 113: 4, 508–532.
- Kulju, P., Kauppinen, M., Hankala, M., Harjunen, E., Pentikäinen J. & Routarinne, S. 2017. Reviewing Finnish studies on writing in basic education. Towards a pedagogy for diversity. Julkaisussa: N. Pyry, L. Tainio, K. Juuti, R. Vasquez & M. Paananen (toim.) *Changing subjects, changing pedagogies: Diversities in school and education*. Helsinki: Helsingin yliopisto, 110–127.
- Kusnetsoff, T. 2017. *Suomenkielisten nuorten morfologinen tietoisuus ja sen yhteys luettun ymmärtämiseen*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Kuukka, K. & Metsämuuronen, J. 2016. *Perusopetuksen päättövaiheen suomi toisena kielenä (S2) -oppimäärän oppimistulosten arviointi 2015*. Helsinki: Kansallinen koulutuksen arviointikeskus.
- Kuusela, J. 2011. Kun kirjoittaminen ei suju. Julkaisussa: E. Harjunen, R. Juvonen, J. Kuusela, B. Silén, M. Sääkslahti & M. Örnmark (toim.) *Miten peruskoululaiset kirjoittavat? Näkökulmia ja kysymyksiä. Perusopetuksen 9. luokan äidinkielen ja kirjallisuuden oppimistulosten seuranta-arvioinnin aineistoa 2010*. Helsinki: Kansallinen koulutuksen arviointikeskus.
- Kuuskorpi, T. 2012. *Psykologisten testien käyttö Suomessa*. Turku: Turun yliopisto.
- Kyösti, J. 2018. *Motivoinnin vaikutuksia kuudesluokkalaisten kirjoittamiin kertomuksiin*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Laasanen, M., Pajunen, A. & Häikiö, T. Arvioitavana. Kielen natiivihallinnan variaatio ja sosiolingvistinen meneillään olevan kielenmuutoksen tutkimus.
- Labov, W. 1975. *What is a linguistic fact?* Lisse: The Peter de Ridder Press.
- Labov, W. & Waletzky, J. 1967. *Narrative analysis: Oral versions of personal experiences*. Julkaisussa: J. Helm (toim.) *Essays on the verbal and the visual arts*. Seattle: American Ethnological Society, 12–44.
- Laine, M. & Virtanen, P. 1999a. WordMill. Lexical Search Program. Turku: University of Turku.
- Laine, M. & Virtanen, P. 1999b. Turun Sanomat 1993–1994 [= TS-korpus]. Morfosyntaktisesti analysoitu tietokanta, 23 miljoonaa sanamuotoa.
- Lappalainen, H.-P. 2003. *Osaat lukea – osaatko kirjoittaa? Perusopetuksen 6. vuosiluokan suorittaneiden äidinkielen ja kirjallisuuden oppimistulosten arviointi 2002*. Helsinki: Opetushallitus.
- Lappalainen, H.-P. 2011. *Sen edestään löytää. Äidinkielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2010*. Helsinki: Opetushallitus.
- Lee, J. 2011. Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics* 32: 1, 69–92.
- Lehtinen, R. 2013. *Kielen norminmukaisuus alakoululaisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.

- Leino, K., Ahonen, A., Hienonen, N., Hiltunen, J., Lintuvuori, M., Lähteinen, S., Lämsä, J., Nissinen, K., Nissinen, V., Puhakka, E., Pulkkinen, J., Puro, J., Sirén, M., Vainikainen, M.-P., Vettenranta, J. 2019. *PISA18 Ensituloksia*. Opetus- ja kulttuuriministeriön julkaisuja 2019: 40.
- Lindeman, J. 2000. *Ala-asteen lukutesti: Käyttäjän käsikirja*. Jyväskylä: Gummerus.
- Malin, A., Sulkunen, S. & Laine, K. 2013. *PIAAC 2012 Kansainvälisen aikuistutkimuksen ensituloksia*. Helsinki: Opetus- ja kulttuuriministeriö.
- Mayer, M. 1969. *Frog, where are you?* New York: Dial Press.
- McLean, K. C. & Breen, A. V. 2009. Processes and content of narrative identity development in adolescence: Gender and well-being. *Developmental Psychology* 45: 3, 702–710.
- Metsämuuronen, J. 2006. Äidinkielen ja kirjallisuuden oppimistulosten ja asenteiden muuttuminen perusopetuksen ylempien luokkien aikana. Oppimistulosten arviointi 3/2006. Helsinki: Opetushallitus.
- Milton, J. 2009. *Measuring second language vocabulary acquisition*. Bristol: The Multilingual Matters.
- Nippold, M. A. 2006 [3. painos]. *Later language development: School-age children, adolescents and young adults*. Austin: Pro-Ed.
- Nippold, M. A., Hesketh, L. J., Duthie, J. K. & Mansfield, T. C. 2005. Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research* 48, 1048–1064.
- Niemi, P. 2007. Lukemistutkimuksen arvoitus: opetusta hylkivät oppilaat. *NMI-bulletin* 17: 3, 8–12.
- OKM 2019. PISA 2018: Suomi lukutaidossa parhaiden joukossa. Opetus- ja kulttuuriministeriön tiedote 3.12.2019. <https://minedu.fi/-/pisa-2018-suomi-lukutaidossa-parhaiden-joukossa>. Viitattu 17.11.2020.
- Pajares, F. & Valiante, G. 2001. Gender differences in writing motivation and achievement of middle school students: A function of gender orientation. *Contemporary Educational Psychology* 26, 366–381.
- Pajunen, A. 2012. Kirjoittamistaitojen kehitys 8–12-vuotiailla: Alakoululaisten unelmakirjoittelmat. *Virittäjä* 116: 1, 3–34.
- Pajunen, A. 2016a. Sukupuolierot kielen hallinnassa (etenkin kirjoittamisessa). Esitelmä. Språkets funktion 6.–7.6. 2016. Åbo Akademi.
- Pajunen, A. 2016b. Unelmakorpus. Peruskoululaisten ja nuorten aikuisten kirjoitelmia aiheesta Unelmieni päivä. Excel-tietokanta.
- Pajunen, A. 2019. Kielenhallinnan odotuksenmukaisuus peruskoululaisilla kielen eri osa-alueilla. Esitelmä. Språkets funktion 4.6.2019. Åbo Akademi.
- Pajunen, A. Tekeillä. 6.- ja 9. luokkalaisten sanaston hallinta sana- ja kielentämistehtävin arvioituna.
- Pajunen, A. & Palomäki, U. 1984. *Tilastotietoja suomen kielen rakenteesta 1*. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Pajunen, A. & Virtanen, P. 2002a. Helsingin Sanomat 2000–2001 [= HS2000-korpus]. Morfosyntaktisesti analysoitu tietokanta, 24 miljoonaa sananmuotoa.
- Pajunen, A. & Virtanen, P. 2002b. Savon Sanomat 2001 [SS2000-korpus]. Morfosyntaktisesti analysoitu tietokanta, 8 miljoonaa sananmuotoa.

- Pajunen, A., Puranen, M. & Yli-Paavola, A. 2010. Koulukorpus. Alakoululaisten kirjoitelmia aiheesta Unelmieni päivä. Excel-tietokanta.
- Pajunen, A., Itkonen, E. & Vainio, S. 2015. Sanamerkityksen hallinta nuorilla aikuisilla. *Virittäjä* 119: 2, 160–187.
- Pajunen, A., Itkonen, E. & Vainio, S. 2016. Nuorten aikuisten kyky määrittellä sanojen merkityksiä. *Virittäjä* 120: 4, 477–515.
- Pajunen, A. & Itkonen, E. 2019. Intuition and beyond: A hierarchy of descriptive methods. Julkaisussa: A. Mäkilähde, V. Leppänen, E. Itkonen (toim.) *Norms and normativity in language and in linguistics*. Amsterdam: Benjamins, 213–234.
- Pajunen, A., Vainio, S., Hyönä, J. & Bertram, R. Tulossa. The role of derivative morphemes and lexical frequency in accessing transitivity information in reading: An eye tracking study in Finnish. Julkaisussa: A. Mäkilähde, T. Alho, V. Leppänen & A. Pajunen (toim.) *Unity and diversity of linguistics*. Routledge.
- Palmer, B., Lum, J., Schlossberg, J. & Gaby, A. 2017. How does the environment shape spatial language? Evidence for sociotopography. *Linguistic Typology* 21: 3, 457–491.
- Pasupathai, M. & Mansour, E. 2006. Adult age differences in autobiographical reasoning in narratives. *Developmental Psychology* 42: 5, 798–808.
- Petälä, I. 2018. *Kertomuksen rakenne, arviointi ja kompleksisuus aikuisten käännekohtakertomuksissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Pohjankukka, J. & Vainio, S. 2013. The internet based language testing program. Department of Psychology, University of Turku. <http://kielitestit.utu.fi/>. Viitattu 31.5.2020.
- Psykologiliitto. Kognitiivisten testien pistemäärien sanalliset kuvaukset. [https://www.psyli.fi/files/836/sanalliset\\_kuvaukset.pdf](https://www.psyli.fi/files/836/sanalliset_kuvaukset.pdf). Viitattu 31.5.2020.
- Pöytä, S. & Kupiainen, S. (toim.) 2018. Tytöt ja pojat koulussa: Miten selättää poikien heikko suoriutuminen peruskoulussa? Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 36/2018.
- Qian, D. D. 2002. Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning* 52, 513–536.
- Ramscar, M., Hendrix, P., Love, B. & Baayen, R. H. 2013. Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon* 8: 3, 450–481.
- Ravid, D. & Avidor, A. 1998. Acquisition of derived nominals in Hebrew: Developmental and linguistic principles. *Journal of Child Language* 25: 229–266.
- Reynolds, M., Schreiber, C., Hajovsky, D. B., Schwartz, B. & Kaufman, A. S. 2015. Gender differences in academic achievements: Is writing an exception to the gender similarity hypothesis? *The Journal of Genetic Psychology* 176: 4, 211–234.
- Ribbon, G. 2016. The trouble with girls? *The Psychologist* 29: 12, 918–922.
- Routarinne, S. & Absetz, P. 2013. Writing performance and writing self-efficacy: The case of Finnish fifth graders. Julkaisussa: S. Tavares (toim.) *Why do we write as we write?* Oxford: Inter-Disciplinary Press, 51–66.
- Ruuska, K. 2020. *At the nexus of language, identity and ideology: Becoming and being a highly proficient second language speaker of Finnish*. Jyväskylä: Jyväskylän yliopisto.



- Saarela, L. 1997. *Peruskoululaisten kirjoitelmien kehittyminen sanastotutkimuksen valossa*. Oulu: Oulun yliopisto.
- Saarinen, A. 2020. *Equality in cognitive learning outcomes: The roles of educational practice*. Helsinki: Helsingin yliopisto.
- Saukkonen, P. 1982. *Oulun korpus. 1960-luvun suomen yleiskielen tutkimusmateriaali*. Oulu: Oulun yliopisto.
- Schmitt, N. 2010. *Researching vocabulary: A vocabulary research manual*. London: Palgrave Macmillan.
- Silverström, C. & Rautopuro, J. 2015. *Språk och skrivande i årskurs 9. En utvärdering av läsresultat i modersmål och litteratur våren 2014*. [https://karvi.fi/app/uploads/2015/09/KAR-VI\\_1815.pdf](https://karvi.fi/app/uploads/2015/09/KAR-VI_1815.pdf). Viitattu 1.6.2020.
- Sintonen, M. 2011. *Suunnittelemisen vaikutus kertomuksen rakenteeseen viides- ja seitsemäsluokkalaissa kirjoittajilla*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Arévalo, A., Vargha, A. & Bates, E. 2005. Timed action and object naming. *Cortex* 41, 7–25.
- Štekauer, P., Valera, S. & Körtvélyessy, L. (toim.) 2012. *Word-formation in the world's languages: A typological survey*. Cambridge: Cambridge University Press.
- Suvanto, A. 2012. *Lapsi tarinaa rakentamassa: Kielihäiriöisten lasten kerrontataidot ja niiden kuntouttaminen*. Oulu: Oulun yliopisto.
- Strömqvist, S. & Verhoeven, L. (toim.) 2004. *Relating events in narrative*. Volume 2. Typological and contextual perspectives. New Jersey: Lawrence Erlbaum Ass.
- Sulkunen, S. & Kauppinen, M. 2018. ”Ainakin muutaman pojan luokallani tiedän lukevan” – Poikien (osin piiloiset) luku- ja kirjoitustaidot. Julkaisussa: A. Kivijärvi, T. Hyyki, H. Lunabba (toim.) *Poikatutkimus*. Tampere: Vastapaino, 146–173.
- Söderholm, C., Häyry, E., Laine, M. & Karrasch, M. 2013. Valence and arousal ratings for 420 Finnish nouns by age and gender. *Plos One* 8: e72859.
- Tolchinsky, L. 2004. The nature and scope of later language development. Julkaisussa: R. A. Berman (toim.) *Language development across childhood and adolescence*. Amsterdam: John Benjamins, 233–247.
- Torppa, M., Niemi, P., Vasalampi, K., Lerkkanen, M.-K., Tolvanen A. & Poikkeus, A.-M. 2019. Leisure reading (but not any kind) and reading comprehension support each other: A longitudinal study across grades 1 to 9. *Child Development* 91: 3, 876–900.
- Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A. & Lawrence, A. M. 2013. Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing* 26, 17–44.
- Tuovila, S., Kairaluoma, L. & Mustonen V. (toim.) 2020. *Luku- ja kirjoitustaidon pedagogiikka yläkouluun*. Rovaniemi: Lapin yliopisto.
- Vainio, S., Pajunen, A. & Häikiö, T. 2019. Acquisition of derivational morphology: School-aged children and young adults. *First language* 39: 2, 139–157.
- Vainio, S., Pajunen, A. & Häikiö, T. Tekeillä. Acquisition of semantic knowledge: School-aged children and young adults.
- Vesala, S. 2012. *Kerronnan keinot alakoululaisten teksteissä*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.

- Virtanen, P. & Pajunen, A. 2002. ContextMill. The easy to use concordancer. Aurit Software, Turku.
- Voutilainen, A. 2002. Conexor FDG [= Functional dependency grammar] for Finnish. Linguistic description. Conexor – Natural knowledge. Moniste.
- Örnmark, M. 2011. Tyylitaju ja genretietoisuus 9. luokalla. Julkaisussa: E. Harjunen, R. Juvonen, J. Kuusela, B. Silén, M. Sääkslahti & M. Örnmark (toim.) *Miten peruskoulu-laiset kirjoittavat? Näkökulmia ja kysymyksiä. Perusopetuksen 9. luokan äidinkielen ja kirjallisuuden oppimistulosten seuranta-arvioinnin aineistoa 2010*. Raportit ja selvitykset 2011: 2, 40–51. [www.oph.fi/julkaisut](http://www.oph.fi/julkaisut). Viitattu 22.4.2020.



# I

## Sanat ja merkitykset

Sanatiedon hallinnan kehittyminen ja sanaston kasvu sekä erilaisten alati kompleksistuvien semanttis–pragmaattisten ilmiöiden kielentäminen on myöhemmän kielenkehityksen vaiheissa keskiössä. Sanatiedon hallinta on kehityksellistä ja sitä voidaan tutkia arvioimalla leksikon kokoa ja sen käyttöä, mutta yhtä hyvin voidaan tarkastella, miten abstraktimmalla tasolla olevia merkityksiä, kuten tekstimerkityksiä, kielennettään.

Sanatasolla objekteilla ja tapahtumilla on sanamerkityksien kannalta katsoen muutamia vastineita, jotka ovat esimerkiksi hyponyymi–hyperonyymisuhteisia tai toisensa synonyymeja tai antonyymejä. Yleistäminen ja spesifiointi lisäävät vaihtoehtoja, mutta silti kielellisiä resursseja on jossain määrin hallittavissa oleva määrä, kun nimetään objekteja ja yksittäisiä tapahtumia. Tekstitasolla tilanne on monimutkaisempi, etenkin jos on kirjoittamalla kielennettävä ilmiöitä, joita ei puhutussa kielessä ilmaista samalla tavoin tai samalla tarkkuudella ja tietoisesti. Tekstitasolla merkitysten kielentäminen edellyttää kirjoitettuun kielimuotoon hiljalleen kasvamista. On ensinnäkin tiedostettava ja hahmotettava ilmiön luonne – kuten tekstin moniäänisyys tai koherenssi – ja ymmärrettävä, että se pitää kielentää tietyllä tavalla, jotta lukija ymmärtää tekstiä. Lisäksi on tunnettava kielentämisessä käytettävissä olevat keinot ja hahmo-

tettava kulloisenkin tilanteen edellyttämät vaatimukset. Siten sanatasolla kehitys on – ainakin aluksi – nopeampaa kuin abstraktimmalla tekstitasolla.

Sanatiedon ja tekstitason merkitysten kielentämistä selvitetään I osiossa neljästä näkökulmasta. Sanatiedon hallintaa käsitellään osion ensimmäisessä luvussa. Seuraavissa luvuissa käsitellään tekstitason merkityksistä spatiaalisuuden ja referentiaalisen koherenssin ja moniäänisyyden ilmaisun kehittymistä sekä vaihtoehtoisten strategioiden että kielellisten resurssien kannalta. Sanatasolla voidaan osoittaa kehityksellisiä trendejä, vaikka yksilön sanatiedon määrä aina vaihtelee. Tekstitasolla trendit ovat suhteellisempia, koska vaikka kielentämisen keinot monipuolistuvat yksinkertaista moninaisempiin iän mukaan, samalla ilmiöiden vaihtoehtoiset kielentämistavat, näkökulmien moninaisuus ja vaikkapa kielenkäyttäjien kielentämiseen vaikuttavat yksilölliset havaintopsykologiset erot kasvavat. Hitaasti kehittyy myös kyky ymmärtää kirjoittajan ja lukijan perspektiivien eroja ja deiktisiä ilmiöitä. Tekstitasolla yksilöllinen ja kielellinen variaatio enemmän kasvaa kuin kanavoituu kehityksellisesti määriteltävään uomaansa. Analyysi lähtee I osiossa merkityksistä.

Hankkeessa ei ole erikseen tutkittu luku- ja kirjoitustaidon tai yleensä äidinkielen opettamista. Jos tehdyn tutkimuksen perusteella kuitenkin voisi kiinnittää huomiota opettamiseen, on selvä, että erityisesti monimutkaisten tekstitason merkitysten kielentämiseen kannattaisi kiinnittää huomiota. Tekstimerkitysten kielentämisen oppiminen käy hitaasti, keinoja ei yleensä saada hallintaan vielä alakouluiässä eikä aina yläkouluiässäkään ja oppilaiden väliset erot kasvavat juuri tekstitason merkityksien kielentämisen hallinnassa. Tutkimukset myös osoittavat, että tekstitasolla kielellisellä syötöksellä kuten luetun kuulemisella ja opetuksella on suuri merkitys. Opettaminen ei tosin ole helppoa, eikä tutkimustietoakaan ole tarpeeksi.

# Unelmieni päivä -sanasto kehityksellisestä näkökulmasta

*Anneli Pajunen*

 <https://orcid.org/0000-0001-8597-7491>

*Mari Honko*

 <https://orcid.org/0000-0002-3240-622X>

## Johdanto

Sanaston hyvä hallinta on lukuisissa tutkimuksissa osoitettu hyvän luku- ja kirjoitustaidon keskeiseksi taustatekijäksi (ks. esimerkiksi Carlisle 2000; Qian 2002; Nippold 2006, 25–29, 89; Wagner, Muse & Tannenbaum 2007; Milton 2009, 170–192; Schmitt 2010, 3–12; suomen osalta Saarela 1997; Honko 2013; Kusnetsoff 2017). Hart ja Risley (1995, 2003) ovat osoittaneet, että Yhdysvalloissa heikko sanatietao päiväkoti-ikäisenä selittää heikosta lukutaidosta yhdeksänvuotiaana jopa 90 prosenttia. Nuorimpien osalta on myös tuloksia, joiden mukaan sanaston monimuotoisuus ennustaa taitoa kirjoittaa kertomuksia (Olinghouse & Wilson 2013). Varhaislapsuudessa syntyneet erot sanastonhallinnassa voivat siis olla kauaskantoisia ja joko haitata tai edistää toiminnallisen kielitaidon kehittymistä. Suomesta vastaavat tutkimukset puuttuvat,

tosin äidinkielen arviointitutkimuksissa on joitain viitteitä heikon osaamisen ja sanaston yhteydestä (ks. esimerkiksi Kuusela 2011).<sup>10</sup>

Aikaisemmin on havaittu, että kielen, sanaston ja peruskäsitteiden osaaminen kehittyy ikäluokkatasolla yläkoulun aikana lukemisen ja kirjoittamisen taitoa vähemmän ja joskus jopa taantuu (Harjunen & Rautopuro 2015, 24–25). Nykyisissä opetussuunnitelman perusteissa tarkoitus on kuitenkin tukea oppilaita tarkastelemaan kieltä vuorovaikutuksessa ja myös sanastoa sen keskeisenä osana. Tekstien tulkinna ja tuottamisen kuvauksen yhteydessä oppilaita kannustetaan erityisesti laajentamaan sana- ja käsitevarastoaan. Sanasto mainitaan erikseen kunkin ikätason äidinkielen ja kirjallisuuden oppimäärien kuvauksissa, ja sanasto on nimettynä sisältönä mukana myös opetussuunnitelman perusteiden erityisen tuen osiossa osana kommunikaatiotaitoja, joihin sanojen tunnistaminen ja käyttökin kuuluvat (OPH 2014, 72).

Kieli läpäisee kaikki oppiaineet, mutta erityisesti sanasto huomioidaan opetussuunnitelman perusteissa kielten oppimäärien kuvauksissa sekä suomen kielen ja kirjallisuuden kuvauksissa. Alkuopetuksessa korostetaan sanojen havainnointia, sanojen ja sanontojen merkitysten sekä tekstien sanavalintojen pohtimista ja oman käsitevarannon laajentamista. Vuosiluokilla 3.–6. rohkaistaan sana- ja käsitevarannon laajentamiseen sekä analyyttiseen tarkasteluun kuten ”selittämään, vertailemaan ja pohtimaan sanojen, niiden synonyymien, kielikuvien, sanontojen ja käsitteiden merkityksiä ja niiden hierarkioita” (mts. 162–164). Oman tuottamisen yhteydessä mainitaan tekstien elävöittäminen ja sanavalintojen tarkastelu osana tekstin merkityksiä. Vuosiluokilla 7.–9. oppilaita kannustetaan vakiinnuttamaan ja laajentamaan sana- ja käsitevarantoaan sekä pohtimaan kielellisten valintojen merkityksiä ja seurauksia. Eriksen kannustetaan sanastoon liittyvien rekisteri- ja tyylipiirteiden tutkimiseen sekä eri kielten sanaston vertailuun ja kuhunkin tekstiin sopivien ilmaisu-

10 Hartin & Risley (1995) tutkimuksen mukaan lapsen leksikon koko on Yhdysvalloissa riippuvainen vanhempien sosioekonomisesta asemasta. Michaels (2013) kritisoi tätä väitettä jyrkästi. Hänen mukaansa tutkimuksen asetelma on rakennettu tulosta ennustavaksi eikä sana-aineistoa ole koodattu esimerkiksi kehitykselliseltä kannalta kriittistä abstraktisanastoa erotellen. Kolmas ja yhteiskunnalliselta kannalta vakavin kritiikin aihe on Michaelsin mukaan se, että huono koulusuoriutuminen langetetaan tutkimuksessa lasten ja heidän vanhempiensa syyksi.

tapojen valikointiin. (Mts. 291–291.) Valtakunnalliset opetussuunnitelman perusteet eivät kuitenkaan anna konkreettisia neuvoja siihen, millä tavalla sanatietoa tulisi opettaa tai tuottamisen ja tulkinnan taitoja harjoitella. Paikallisella opetussuunnitelmatyöllä, opettajien omalla valvutuneisuudella ja aktiivisuudella sekä oppikirjantekijöillä onkin merkittävä rooli siinä, miten esimerkiksi sanamerkitysten havainnointiin ohjataan.

Oma kokemuksemme opettajina ja opettajankouluttajina sekä aikaisempi tutkimus antavat viitteitä siitä, että sanastoa käsitellään opetuksessa usein yhä melko yksipuolisesti ja mekaanisesti niin, että painopiste on irrallisten sanojen nimeämisessä sekä luokittelussa. Kauppisen, Tarnasen ja Aallon (2014) tutkimuksen mukaan luokanopettajaopiskelijat eivät myöskään hahmota sanastoa osaksi äidinkielen opetuksen sekä reaaliaineiden opetuksen oppiaineintegraatiota, vaan yhteyksiä nähdään enemmän tekstilajituntemukseen ja jopa kielitiedon opettamiseen. Silloin, kun sanastoa ja sanatietoa opetuksessa tarkastellaan, huomio tuntuu kiinnittyvän kykyyn nimetä ja luokitella sanoja annettujen kategorioiden mukaan, ei itse prosessiin tai luokitteluperusteisiin kuten sanojen käytön, merkityksen ja muodon analyttiseen havainnointiin. Kognitiiviseen päättelyyn ja esimerkiksi luokitteluperusteiden kielentämiseen ohjaamisen kulttuuri tarvitsisi vahvistusta, jotta oppiminen – ajattelun ja kielellisten taitojen tukeminen – nousisi keskiöön kategorioiden läpi käymisen sijaan. Oletus tuntuu olevan, että sanasto opitaan riittävän hyvin käyttöyhteyksistä ja erityisesti omaehtoisen lukemisharrastuksen avulla. Tämä näkyy paitsi oppimateriaaleista ja opetussuunnitelmista myös reaktioista: kun lukutaidon PISA-tulokset ovat laskeneet, ratkaisuksi on kehitelty tuettuja lukukampanjoita.<sup>11</sup> Tutkimuksen valossa tämä on tärkeää mutta tuntuu riittämättömältä tilanteessa, jossa suomalaislasten ja nuorten lukemisaktiivisuus vähenee ja luku- ja kirjoitustaito heikkenee.

Koulutuksen pitäisi tasata ja sen oletetaan tasaavan osaamistasoa (ks. esimerkiksi Hadley & Rispoli 2012), mutta oman tutkimuksemme

11 Esimerkiksi Opetushallituksen Lukuliike ja siihen nidotut kampanjat koettavat lisätä lukemisharrastusta (ks. <https://lukuliike.fi>). Tavoite on hyvä, mutta ei välttämättä kohdistu siihen kasvavaan joukkoon, jonka ongelma on kielenhallinnassa tai ainakin kirjoitetun kielimuodon vieraudessa.



mukaan sanamerkityksen hallinnan taso, kyky analyttisesti määritellä sanojen merkityksiä ja taito ymmärtää tai tuottaa johdoksia vaihtelee paljon myös ylioppilastutkinnon suorittaneilla nuorilla aikuisilla, joiden ikäluokastaan koulutetuimpina pitäisi omata melko yhtenevä taitotaso (ks. Pajunen, Itkonen & Vainio 2015; 2016; Vainio, Pajunen & Häikiö 2019; Laasanen, Pajunen & Häikiö, arvioitavana). Sanaston hallinnan tutkimus on Suomessa ollut vähäistä, mutta kansainvälinen tutkimus on kasvavaa ja tiedetään, että vahvat sanastotaidot ovat yhteydessä paitsi hyviin tekstitaitoihin ja niiden edelleen kehittymiseen – myös hyvään koulumenestykseen ja opintopolulla etenemiseen (ks. esimerkiksi Hart & Risley 2003; Treffers-Daller & Milton 2013; Nation & Coxhead 2021). Sanaston hyvä hallinta on myös itseisarvo, joka säilyttää niin kulttuuria kuin edistää yksilön kykyä ilmaista itseään täsmällisesti ja osuvasti.

Kartoitamme tässä yhteydessä koululaisten sanastotaitojen kehittymistä tutkimalla, miten eri-ikäisten kirjoittajien sanaston käyttö muuttuu samasta aiheesta eli unelmien päivästä laadituissa Unelmakirjoitelmissa (alakouluikäisten unelmasanastosta tarkemmin ks. Laine-Leinonen 2013). Ikäluokat ovat alakoululaiset ja yläkoululaiset, joita tarkastellaan myös koululaisten ryhmänä, ja osaksi ryhminä ovat kolmas- ja kuudesluokkalaiset sekä yhdeksäsluokkalaiset. Vertailukohtana ovat nuoret aikuiset. Aineiston sanaesiintymien määrä on yli 200 000 ja ryhmittäin korkeintaan 10 000 eri sanaa (lekseemiä). Oletus aiemman tutkimuksen perusteella on, että käytetyn sanaston yleisyys laskee samalla, kun sanaston diversiteetti eli eri lekseemien määrä ja myös semanttinen monipuolisuus lisääntyy. Käsittelemme aineistoa ryhmätasolla, ja tavoitteena on paitsi osoittaa sanaston monipuolistuminen myös luoda tietoa ikäryhmittäisestä sanaston hallinnan keskimääräisestä tasosta. Jotta yksilöiden väliset erot eivät peittyisi, kuvaamme analyysien yhteydessä myös yksilöllistä vaihtelua. Käsittelemme ensin tutkimuksen teoreettista taustaa ja keskeisiä käsitteitä sekä aineistoa, ja keskitymme sitten kirjoitelmasanaston yleisyyteen, diversiteettiin ja lopuksi semanttiseen luonteeseen. Loppulukuun kootaan tulokset.

## Sanaston frekvenssitekijät, rikkausluvut sekä semanttinen luokitus

Sanaston hallinnan tutkimuksessa tavoitteena on yleensä saada tietoa käytössä olevan sanaston yleisyydestä, sen moninaisuudesta ja semanttisesta luonteesta. Yleisyyttä voidaan tarkastella lekseemi- ja esiintymätasolla sekä hyödyntämällä kattavuuslukuja, yleisyystasoja ja niin sanottuja perhefrekvenssejä. Näin voidaan tehdä ennusteita yleisimmin hallitun ja käytetyn sanaston luonteesta, sanojen oppimisistä, prosessin luonteesta ja niin edelleen. Puhutaan niin sanotuista frekvenssitekijöistä (ks. Bybee & Hopper 2001; Baayen 2014): yleisyys vaikuttaa suuresti siihen, minkälaiselle kielelle kielenkäyttäjät altistuvat ja on siksi yhteydessä myös sanaston tyypilliseen oppimisjärjestykseen (ks. esimerkiksi Treffers-Daller & Milton 2013). Sanaston moninaisuutta teksteissä arvioidaan erilaisilla variaatio- ja rikkausluvuilla (ks. esimerkiksi Jarvis 2013), ja semanttinen analyysi puolestaan edellyttää aineiston alaluokitusta ensin sanaluokkakohtaisesti ja edelleen merkityksen mukaan (ks. esimerkiksi Miller & Fellbaum 1991).

Sanojen yleisyyden arviointiin tarvitaan massakorpus, jonka koon olisi luotettavuuden vuoksi hyvä olla vähintään kymmenissä miljoonissa. Massakorpuksessa koko on yleensä tärkeämpi kuin sen koostumus, koska mikään korpus ei ole harvinaisten sanojen osalta edustava ja koska yleisimmät sanat selviävät luotettavasti pienemmästäkin massakorpuksesta. Itse käytämme HS2000-korpusta (n. 24 milj. sanaesiintymää),<sup>12</sup> jota täydennämme Tieteellisen laskennan antamin yleisyyksin (CSC korpus, 44 milj. sanaesiintymää). Korpusfrekvenssit annetaan yleisyytenä miljoonassa sanassa.

Yleisyytasoja lasketaan jakamalla lekseemit yleisyyden mukaan tuhannen siivuihin: Ensimmäinen tuhat sisältää kielen yleisimmät sanat. Yleensä riittää, jos erotellaan kymmenen yleisintä tasoa (ks. Laufer & Nation 1999, 39; Milton 2009, 22–43), ja tällöin harvinaiset sanat voidaan määrittellä niiksi, jotka kuuluvat yleisyytasoille 9 tai sen yli ja

<sup>12</sup> HS2000-korpuksen kokonaissanamäärä on noin 31 miljoonaa; laskelmat perustuvat sana-ainesmerkkimäärään.

joiden esiintymää miljoonassa sanassa on yksi tai vähemmän. Mitä harvinaisempi sana on, sen epätodennäköisempää myös on, että se osuu laskelmiin käytettävän korpuksen sanastoon. Siten harvinaiset sanat ovat myös niitä, joita ei korpuksista löydy. Koska harvinaisista sanoista on vaikea saada luotettavaa yleisyystietoa (vrt. esimerkiksi Keuleers, Stevens, Mandera & Brysbaert 2015), niiden tuttuus tietyssä ikäluokassa on parempi testata erikseen käyttämällä subjektiivisia testejä (vrt. Pajunen & Itkonen 2019). Kehityksellisen tutkimuksen tarpeisiin riittää usein karkea yleisyystasojatelu hyvin yleisestä tavalliseen ja harvinaiseen, koska aikuisenkin puhujan sanatieto alkaa nopeasti heiketä jo yleisyystasoilla 9–10.

Perhefrekvenssejä voidaan laskea suomen tapaisessa kielessä muun muassa tutkimalla sanueitten jäsenkokoja (kantasana ja sen johdokset ja yhdyssanat): mitä suurempi sanue, sen todennäköisemmin sana tunnetaan hyvin ja prosessoidaan nopeasti (ks. Pajunen, Vainio & Itkonen 2015; Vainio, Pajunen & Häikiö 2019). Suomen kielessä substantiiviperheet ovat paljon suurempia kuin verbiperheet, koska substantiivilekseemien ja verbilekseemien suhde on korkeintaan 8:2 ja koska substantiiveja muodostetaan myös yhdistämällä, verbejä vain harvoin. Substantiiviperhe on suuri, jos jäseniä on vähintään parisenkymmentä, mutta verbiperhe on suuri, jos niitä on vain muutamiakin. Perhefrekvensseihin kuuluu myös niin sanottu taivutusperhefrekvenssi, joka saadaan yksinkertaisesti laskemalla massakorpuksesta sanan kaikki taivutusmuototyypit ja esiintymät tyypeittäin. Kaikki korpuksessa esiintyvät muodot, kuten kunkin substantiivisanan yksikkö ja monikko eri sijoissa tai liitepartikkeleissa taivutettuina, ovat laskennassa eri tyyppisiä. Esimerkiksi *hykerrellä*-verbin taivutusperhekoko on HS2000-korpuksessa 8 ja yleisyys 21 (ks. taulukko 3). Nyt verbien taivutusperheet ovat paljon suurempia kuin substantiivien. Verbilekseemien toistuvuus on korkeampi kuin substantiivilekseemien, ja se nostaa eri konteksteissa ja siten useammissa eri taivutusmuodoissa esiintymistä.

Taivutusperheluvut korreloivat melko hyvin esiintymäyleisyyteen korkeimpien lukujen osalta, mutta pienet taivutusperheet viittaavat sanan käytön vähäisyyteen tai rajoitukseen ja merkityksen vähittäiseen idiomaattisuuteen. Taivutusperhekoko vaihtelee yhdestä kahteen kol-

meen sataan. Koko osoittaa, kuinka monipuolinen malli kielenkäyttäjälä sanasta on. Iso taivutusperhe tukee esimerkiksi vartalonvaihteluiden hallintaa, pieni vastaavasti heikentää.<sup>13</sup> Kehityksellisestä näkökulmasta katsoen yleisen sanan vähäinen taivutus voisi viitata rajoittuneeseen käyttökontekstiin, mutta melko yleisen sanan maksimaalinen taivutus viittaisi monipuoliseen käyttökontekstiin. Jälkimmäinen tilanne edistää sanan opittavuutta.

Taulukko 3. Esimerkki taivutusmuotoperheestä.

Lekseemi	Yleisyys	Taivutusperhe	Sanamuodon esiintymä	Sanamuodon yleisyys
<i>hykerrellä</i>	21	8	<i>hykerrellen</i>	1
			<i>hykerrellä</i>	2
			<i>hykertelee</i>	7
			<i>hykertelemään</i>	1
			<i>hykerteleväksi</i>	1
			<i>hykertelevät</i>	3
			<i>hykerteli</i>	4
			<i>hykertelivät</i>	2

Englannissa samaan perhefrekvenssiin lasketaan sekä sanue- eli derivaatioperhe että taivutus(muoto)perhe (ks. Bauer & Nation 1993). Vastaava laskelma sisältäisi suomessa esimerkiksi paitsi kaikki *käsi*-sanamuodot myös sen johdokset ja kaikkien näiden johdosten kaikki eri taivutusmuototyypit. Tällainen perhefrekvenssi on suomesta mahdoton laskea, koska mikään ohjelma ei yhdistä kannan kaikkia johdoksia yhteen – yhdyssanoista puhumattakaan – eikä laske lisäksi kaikkia niiden taivutusmuototyyppejä. Koska suomessa johtamattomien kantasanojen määrä on pieni ja johtaminen keskeinen sananmuodostuskeino, voi myös epäillä, olisiko tällaisista perheluvuista hyötyä. Nuorimmat kouluikäiset eivät myöskään vielä hallitse suomen sananmuodostuksen pe-

13 Sanat *keriminen*, *keritseminen* ja *kerkeäminen* voivat muodosta riippuen saada kantasanakseen joko *keriä* tai *keritä*-muodon ja toisinaan jokin taivutusmuoto laukaisee myös *kerjetä*-verbin. HS2000-korpuksessa taivutusmuotoja on *keritä*-verbistä 10 (joista 1 liittyy keritsemiseen, 9 kerkeämiseen) ja *keriä*-verbillä 14 (joista esimerkiksi passiivimuoto voi liittyä sekä keritsemiseen että kerimiseen). Lampaan keritsemiskuvan kielennystehtävässä suurin osa koehenkilöinä toimineista nuorista aikuisista nimesi kuvan kerimistapahtumaksi (Pajunen 2017). Toisin sanoen *keritsemisen* pieni taivutusperhe korpuksessa ennusti virhetulosta, vaikuttipa siihen mikä tahansa tekijä.

riaatteita vaan joko tuntevat tai eivät tunne yksittäisiä johdoksia näkemättä samaan perheeseen kuuluvien johdosten välisiä yhteyksiä (ks. Kusnetsoff 2017).

Laskemiseen liittyy synteettisessä kielessä monia periaatteellisia ongelmia, joita toistaiseksi ei ole ratkaistu, vaikka hyviä laskemistyökaluja löytyisikin. Miten esimerkiksi määritellään sanueen lähtökohtana oleva (kanta)sana, tai miten otetaan produktiivisuus huomioon.<sup>14</sup> Paras tuntemamme väline taivutusmuotojen laskemiseen on WordMill (ks. Laine & Virtanen 1999), joka antaa tiedot suoraan hakutuloksena mutta määrittelee kantasanana (tai lemman) paikoin epätoivottavasti (esimerkiksi *sammakoiden* < \*sammakoida). Ohjelma myös sisällyttää taivutusmuotoihin muutamia johdoksia, jotka halutessaan voi erotella vain manuaalisesti. Sanueiden kokoa joutuu toistaiseksi laskemaan sanakirjojen avulla. Synteettisestä kielestä saadaan myös sanarakennetta ja sanapituuksia arvioimalla kuva frekvenssitekijöistä ja sanaperheistä: johtamattomat sanat ovat yleensä johdoksia ja yhdyssanoja yleisempiä ja lyhyet sanat pitkiä yleisempiä.

Käytämme sanaston kehittymisen tutkimuksen tukena yleisyyslukuja (/miljoona sanaesiintymää), yleisyystasotietoja ja taivutusperhefrekvenssejä. Sanaperhelukuja emme käytä mutta niiden vaikutus peruskouluikäisten kielenhallintaan tuli esiin hankkeen derivaatiotestissä (ks. Vainio ja muut 2019). Lisäksi tarkastelemme sanaston monimuotoisuutta (diversiteettitunnusluku *sum of probabilities*) sekä sen jakautumista sisältö- ja kieliopillisiin sanoihin sekä eri sisältösanaloukkien sanoihin. Nämä tarkastelutavat mahdollistavat hyvin erilaajuisten sekä kokonaisrakenteeltaan että tyyliltään erilaisten tekstien vertailun, ja ne ovat perusteltuja myös aikaisemman oppijankielen tutkimuksen perusteella. Varsinaisia oppimisikäitietoja ei ole suomen kielestä kerätty.

Analysoimme kirjoitelmien sanaston myös semanttisesti. Semanttinen luokitus voi perustua yksin käsiteanalyysiin, jolloin ei tarvitse tehdä

14 Hakulinen (1979, 478) toteaa sekä yksilön sanaston että murreosanosten kokoa arvioidessaan, että suomessa lukuihin on laskettava mukaan paitsi johtamaton sana myös jokainen johdos. Sanakirjoissahan hyvin produktiiviset johtimet, kuten teettokausatiivijohtimet ja *minen*-johdin/ muoto jätetään pääasiassa huomiotta. Kummankin oppiminen on kehityksellisesti mielenkiintoista. Lopputulema kuitenkin on, että derivaatioperheen laskeminen on suomessa erityisen vaikeaa.

yrkkää sanaluokkaeroa. Käytännössä on helpompi lähteä leksikaalisista luokista, jotka perustuvat ontologis-semanttiseen kolmijakoon 'oliot', 'tapahtumat' ja 'ominaisuudet'. Nämä luokat kattavat leksikaalisesta massasta valtaosan (91 % omassa aineistossamme) ja niiden jäsenet vaikuttavat siihen, millä täsmällisyydellä ja abstraktiuden tasolla maailmassa esiintyviä todellisia ja kuviteltuja olioita ja tapahtumia pystytään nimeämään. Ne myös ilmaisevat kulttuuris-yhteiskunnalliset sävyt ja tyylilliset erot, ja muun muassa frekvenssitekijöiden suhteen ne eroavat toisistaan selvästi. Peruskolmijakoa tarkastellaan alaluokittelemalla ne useihin merkityksen alaluokkiin.

Myös syntaktiset, kielikohtaiset sanaluokat, kuten konjunktiot ja postpositiot, on koodattu aineistoon, mutta ne sivuutetaan tässä yhteydessä. Sanaluokkajako on luonnollisesti jatkumonluonteinen, on selvät substantiivit, adjektiivit ja verbit ja sitten luokkien väliin jäävät erikoisryhmät, kuten itsenäisesti käytettävät adjektiivit tai substantiivien luokasta irronneet adverbit, joten luokat jossain määrin risteävät. Verbi- ja adjektiiviluokituksissa voi tukeutua melko runsaaseen kirjallisuuteen, sen sijaan toimivia substantiiviluokituksia ei ole olemassa. Paras malli on ehkä Millerin (1998) käsitteellinen luokitus, joka listaa niin johtamattomat kuin johdetut substantiiviluokat. Sama malli on lukemattomissa eri kielten WordNet-sovelluksissa käytössä. Semanttinen analyysi on tehtävä manuaalisesti, mutta sen avulla saadaan tietoa sekä teemakeskeisestä sanastosta että kirjoitelmien sanaston abstraktiusasteesta. Yleisyystekijät, diversiteetti ja sanaston abstraktistuminen ovat sanatasolla keskeisiä kehityksellisiä muuttujia.

## Aineistona Unelmasanasto

Unelmasanasto on koostettu Koulu- ja Unelmakorpuksen aineistoista, joissa on yhteensä lähes 1900 kirjoitelmaa ja noin 230 000 sanaesiintymää (ks. tarkemmin johdantoluvusta). Sana-analyysiä varten aineisto valikoitiin siten, että kaikki erikoisaines poistettiin, koska tavoitteena oli saada näkyviin leksikaalinen erilaisuus lekseemimäärän sijaan. Sanaesiintymät muutettiin lekseemeiksi noudattamalla seuraavia periaatteita:

- 1) Aineistosta poistettiin:
  - a) propriit (henkilön- ja paikannimet sekä tavaranimet), sana-liitot, numerot ja merkit,
  - b) epäselvyydet (ei analysoitavissa, rauniosanat tms.),
  - c) vieraskieliset sanat sekä
  - d) interjektiot ja numeroalkuiset sanat (9-v).
  
- 2) Osa johdetuista sanaesiintymistä on laskettu edellisen derivaatio-vaiheen sanaan:
  - a) Esimerkiksi *minen*-johdoksen esiintymät on laskettu vastaavan verbin esiintymiin, mikäli se löytyy aineistosta. Jos verbiä ei ole aineistossa, *minen*-johdoksen esiintymä on luettu omaksi lekseemikseen.
  - b) Kaikki partisiippien (aktiivin ja passiivin partisiipit sekä *ma*-partisiipit) esiintymät on laskettu vastaaviin verbilekseemiin, mutta jos kyseinen verbilekseemi puuttuu aineistosta, partisiippi on oma lekseeminsä.

Näillä kriteereillä esiintymämäärät vähenivät kokonaisesiintymämääristä kaikissa aineiston osissa muutaman prosenttiyksikön (ks. taulukko 4). Nuorimmilla vähennys johtui runsaista nimi- ja numeroesiintymistä, vanhimmillä oli paljon syntaktisia johdoksia. Vaikka esimerkiksi *minen*-johdokset lisääntyvät kehityksellisesti, niiden katsottiin ensi sijassa osoittavan syntaktista taitoa, jolla sanottavaansa saattaa tiivistää. Lekseemimäärät taulukossa 4 ovat ryhmäkohtaisia, eli ne on laskettu ryhmittäin vain kerran, mutta eri ryhmissä samat lekseemit voivat toistua.

Taulukko 4. Unelmasanaston frekvenssit.

Ikäluokka	Esiintymät	Esiintymät karsinnan jälkeen	Ero %	Lekseemit
Alakoulu 1.–5.	64 914	60 860	6,2	4 854
Kuudennet	42 282	39 422	6,8	4 396
Yläkoulu	36 189	34 209	5,5	4 507
Aikuiset	89 192	87 126	2,3	10 069
Yhteensä	232 577	221 617	4,7	

Leksikaalinen sana-aines on jatkoanalysoitu kolmella eri tavalla, jotta saadaan selville iän mukainen muutos ja kehitys sanaston käytössä. Ensimmäkin koko aineistolle laskettiin yleisyystiedot isoista sanomalehtikorpuksista (HS2000-korpus, ks. Pajunen & Virtanen 2002). Tulokset tarkistettiin sekä sanakohtaisesta (ks. Laine & Virtanen 1999) että kontekstia sisältävästä korpuksista. Yleisyystiedot laskettiin sitten ryhmittein siten, että alakoulun 6. luokka laskettiin erikseen, koska siitä oli aineistoa paitsi vuodelta 2006 myös vuodelta 2014. Yläkoulun 7.–9. luokkien aineisto laskettiin pienuutensa vuoksi yhtenä ryhmänä. Aikuisryhmät analysoitiin erikseen, mutta erojen pienuuden vuoksi niitä käsitellään yhtenä. Toisekseen eri ikätasojen vertailua varten kolmannen, kuudennen ja yhdeksännen luokan peruskoululaisten sekä ammattikorkeakoululaisten aineistosta tehtiin sanaston diversiteetilaskelmat koehenkilökohtaisesti (ks. taulukko 5). Kaikki tekstit, joissa sanamäärä oli pienempi kuin 42, poistettiin analyysistä, koska lyhyet tekstit vinouttavat tuloksia ja koska käytetty diversiteettitunnusluvun menetelmä (*sum of probabilities*) edellyttää tiettyä vähimmäissanamäärää (McCarthy & Jarvis 2007, 472). Tässä menettelyssä tietyn tekstin sanastollista monimuotoisuutta kuvataan tunnusluvulla, joka lasketaan kunkin tekstissä esiintyvän sanan esiintymistodennäköisyyksien summana. Laskennassa käytetään apuna etukäteen asetettua parametria (tässä tapauksessa 42), joka kuvaa todennäköisyyksien laskennassa käytettävän tekstikatkelman pituutta. Kolmanneksi koko ensimmäisen kohdan lekseemiaineisto analysoitiin myös semanttisesti jakamalla aineisto ensin sanaluokkiin ja sitten luokittelemalla se sanaluokkakohtaisesti.

## Unelmasanaston yleisyys

Frekvenssitekijät vaikuttavat sanaston hallintaan ja käyttöön voimakkaasti. Voidaankin ennustaa, että yleisin sanasto hallitsee nuorimpien peruskoululaisten kirjoittamista, mutta iän myötä sanasto harvinaistuu ja monipuolistuu. Tarkastelemme tässä yhteydessä niin sanottuja kattavuuslukuja, lekseemien ja esiintymien yleisyyksiä, käytetyn sanaston

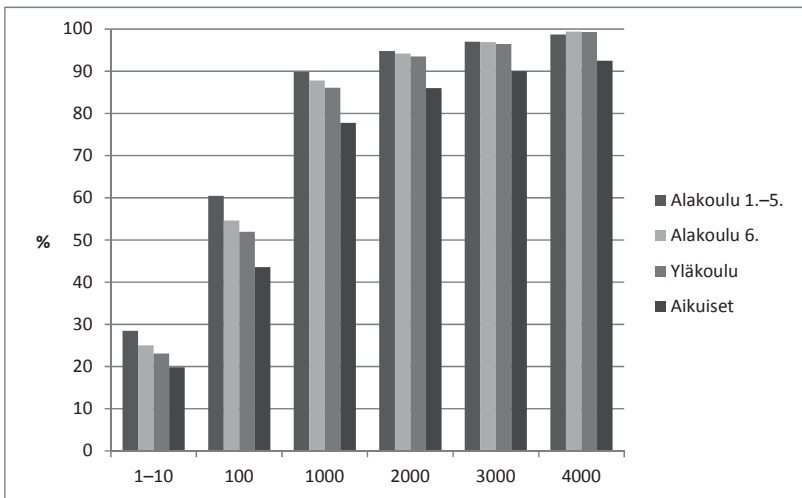


yleisyytasoja sekä taivutusperheitä ja suhteutamme eri frekvenssitekijöitä myös esimerkiksi sanojen rakenteeseen.

Yleisyystiedot laskettiin HS2000-korpuksen avulla. Haut tehtiin sekä WordMill- että ContextMill-ohjelmilla. Ensimmäinen antaa tiedoksi sana- ja taivutusmuotofrekvenssit ja erittelee sanan kaikki ne esiintymämuodot, mitkä lukuihin on sisällytetty, ja jälkimmäinen antaa sanojen käyttökontekstista halutulla tavalla rajatun tiedon. Esiintymämuotoihin sisältyvät muun muassa *minen*-johdokset ja eri funktioissa käytetyt partitiippimuodot. Saatuja frekvenssitietoja verrattiin Tieteellisen laskennan (CSC) antamiin yleisyyksiin. Frekvenssitiedot on siten koostettu yhteensä lähes 70 miljoonan sanaesiintymän aineistosta. Koska tiedot ovat peräisin sanomalehtikielestä, niiden perusteella saadaan luotettavaa tietoa suomen yleisestä sanastosta mutta ei aina tavoiteta arkikielen sanastoa: vaikka frekvenssilistojen kärki on tekstilajista riippumatta hyvin samankaltainen, ei asiategististä koostettuun frekvenssilistaan luonnollisesti juurikaan nouse puhutun kielen rekistereille tyypillistä sanastoa. Sellaisenaan tiedot eivät myöskään anna luotettavaa tietoa koko kielen harvinaisesta sanastosta. Tässä tutkimuksessa kuitenkin riittää, että saadaan tehtyä ero yleisen ja harvinaisen välillä.

Käsitlemme tuloksia luokittelemalla esiintymäyleisyyttä pääasiassa ryhmiin hyvin yleiset ja yleiset (yleisyytaso 1, yleisyytasot 2–4), tavalliset (yleisyytasot 5–9) ja harvinaiset (yleisyytaso >10). Lisäksi on puuttuvien joukko, johon kuuluvat ne, jotka eivät satu esiintymään HS2000-korpuksessa, ja ne, jotka eivät ole suomen vakiintuneita sanoja vaan esimerkiksi satunnaismuodosteita. Ikäryhmiä käsittelemme hieman eri tarkkuudella. Tarkin jako erottaa alakoulun luokat 1.–5. ja 6., yläkoulun ja aikuiset. Monet erot näkyvät jo jaottelulla ala- ja yläkouluun, mutta kuudes luokka on analyysissämme ikään kuin siirtymätyyppiä. Koululaisten kirjoitelmissa kertovan tekstin luonne ja kerrontastrategia ovat melko samanlaisia etenkin aikuisteksteihin verrattuna, joten monet peruserot ovat osoitettavissa sanastotasolla jo vertaamalla koululaisia (melko hyviin) aikuiskirjoittajiin.

Hieman yleistäen voidaan sanoa, että yleisimmät kielen lekseemit opitaan ensimmäiseksi, ja toisaalta, että yleisimmät lekseemit ovat monikäyttöisyytensä vuoksi kommunikoinnissa ne hyödyllisimmät (ks. sanas-



Kaavio 2. Suomen kielen yleisimpien sanojen kattavuus kaikista sanaesiintymistä eri-ikäisten Unelmakirjoitelmissa: 10–4 000 yleisintä sanaa (= lekseemiä).

ton yleisyyden ja toiston vaikutuksesta oppimiseen esimerkiksi Honko 2013, 29–31). Lisäksi tarvitaan toki myös harvinaisempaa sanastoa, jonka osaamisen tarpeet ja siksi myös hyödyt määrittyvät tiiviimmin kielenkäytön yksilöllisten ja tilanteisten tarpeiden mukaan. Lekseemejä, joiden esiintymiä miljoonassa sanassa on tuhansia tai edes tuhat, on yleensä hyvin vähän; suomessa muun muassa *olla* ja *ja* ovat yleisyyssä.

Yleisimpien lekseemien merkitys on tärkeä myös tekstin kattavuuslukujen kannalta. Kattavuus tarkoittaa sitä prosenttiosuutta, minkä tietty määrä lekseemien esiintymiä kattaa tekstin kokonaissanamassasta (ks. Laufer & Ravenhorst-Kalovski 2010). Miltonin (2009, 46) mukaan jo runsas kymmenkunta lekseemiä voi englannissa kattaa neljänneksen korpuksen sanaesiintymistä, sata noin puolet ja tuhat kolme neljännes. Englanti on analyttinen kieli, joten kattavuuteen vaikuttavat myös kielipolliset sanat. Unelmasanastossa kymmenen yleisintä lekseemiä kattaa 20–30 prosenttia tekstimassasta, sata noin puolet ja tuhat yleisintä jo 80–90 prosenttia (ks. kaavio 2). Nuorimpien ja vanhimpien kirjoittajien välinen ero on yleisyysryhmittäin kymmenisen prosenttiyksikköä, ja kattavuus laskee lineaarisesti iän myötä. Yleisimpien sanojen osuus aikuisaineiston koko tekstimassasta muistuttaa englannin suhteita,

mutta tähän vaikuttaa myös tekstin luonne ja teeman yhteneväisyys. Moninaisemmassa aineistossa peittävyys olisi suomessa englantia alempi, koska suomessa ei ole englannin tapaisia prepositioita eikä esimerkiksi artikkeleita.

Kaikissa ikäryhmissä kymmenen yleisimmän lekseemin joukkoon kuuluvat *ja, olla, se, minä ja päivä*. Ne kuvastavat toisaalta tekstityyppiä ja toisaalta kirjoitelman aihetta. Sen sijaan muun muassa relatiivipronomini *joka* nousee yleisimpien listalle vasta kuudennella luokalla, jolloin puolestaan sanat *sitten* ja *kun* eivät enää esiinny listalla. Yleiskielessä yleinen *ei* taas ilmestyy listalle vasta kuudennen jälkeen; unelmista ei niinkään kirjoiteta kielteisessä mielessä. Kuudennella *mennä*-verbin rinnalle ilmestyy *lähteä*-verbi, mutta aikuisten aineistossa yleisimpien sanojen listalla ei enää ole verbien *olla* ja *ei* lisäksi muita verbejä. Yleisimpien lekseemien esiintymäjärjestys toisin sanoen tuntuu aika hyvin kuvaavan syntaksin taitojen kehittymistä: esimerkiksi relatiivilauseet lisääntyvät kehityksellisesti, kuten kirjan loppuluvusta käy ilmi. Nuorimmilla yleisimmät kymmenen lekseemiä kattavat lähes kolmanneksen tekstimasasta, mikä osoittaa, että näiden lekseemien toistuvuus on erittäin korkea. Aikuisilla yleisimmät sanat ovat joko kieliopillisessa funktiossa käytettäviä konjunktioita, verbejä tai teemaan liittyviä (*unelma, päivä*).<sup>15</sup>

Kaikissa ikäryhmissä sadan yleisimmän lekseemin joukosta noin kolmannes on verbejä. Tekstikattavuuden kannalta siirtymä yleisimmistä iän myötä harvinaisempiin käy jokseenkin lineaarisesti. Yleisyyslähteinä toimivien sanomalehtikorpuksen sana-aineksista puuttuvat lekseemit ovat nuorimmilla sadunomaisia satunnaismuodosteita tai arkikielisiä (*suklaatulivuori, älämölänpoistoaine, mörköpeli, porkkanaylläri, rentoilu*), aikuisilla valtaosaan nousevat spesifiä merkitystä ilmaisevat satunnaismuodosteet tai tilanteiset yhdyssanamuodosteet (*aamurapsutus, helikopterinraato*). Lisäksi kirjoitelmissa on joitain arkikielen sanoja (*mussutella, pörräily*) tai uudissanoja, joita 2000-luvun alun sanomalehdissä ei vielä esiintynyt (*peukutus*).

Unelma-aineiston sanasto on luonnollisesti paitsi teemaan ja tekstityyppiin sidoksissa myös pieni: kussakin koululaisryhmässä eri leksee-

15 Kirjoitelmien otsikointia (Unelmieni päivä) ei ole laskettu sanamassaan mukaan.

mien määrä jää alle 5 000:n (kokonaissanamäärä 34 209–60 860), vain aikuisaineistossa lekseemejä on enemmän. Siten sanasto ei kuvasta suomen yleiskielen yleisyysuhteita riittävästi muiden kuin kaikkein yleisimpien sanojen osalta. Varsinainen yleisyystasoluokitus suomen kielestä vielä puuttuu. Lekseemien jakauma yleisyystasoihin on ryhmitäin melko tasainen (ks. taulukko 5). Erot ovat suurimmat koululaisten ja aikuisten välillä. Koululaisilla kirjoitelmissa käytetyistä lekseemeistä on hyvin yleisiä (yleisyystaso 1) tai yleisiä (yleisyystasot 2–4) lähes 30 prosenttia ja tavallisia (yleisyystasot 5–9) noin kuudennes. Koululaisryhmissä yleisimpien määrä hieman kasvaa eikä laske iän mukaan, kuten odottaisi, mutta on mahdollista, että ero kuvastaa syntaksin taitojen kehitystä. Aikuisilla hyvin yleisiä tai yleisiä lekseemejä on selvästi vähemmän kuin koululaisilla eli noin viidennes. Harvinaisia lekseemejä on kaikilla määrällisesti paljon, vaikka esiintymiä harvinaisella lekseemillä on tyypillisesti vain yksi. Tässä koululaisten ja aikuisten ero on suurin eli yli kymmenen prosenttiyksikköä. Massakorpuksista eri syistä puuttuvia lekseemejä on kymmenisen prosenttia. Aineistossa on luonnollisesti vähän suljettujen sanaluokkien jäseniä ja niiden joukossa on paljon hyvin yleisiä lekseemejä, avoimien luokkien tilanne on taas päinvastainen. Harvinaisista lekseemeistä suurin osa kuuluu substantiiveihin, verbeissä on enemmän toistuvasti käytettyjä, yleisiä lekseemejä. Käsittelemme sanaluokkien suhdetta yleisyyteen semanttisen analyysin yhteydessä tarkemmin.

Taulukko 5. Unelmakirjoitelmasanaston jako yleisyystasoihin eri ikävaiheissa: eri yleisyysluokkien lekseemien osuus kaikista eri lekseemeistä.

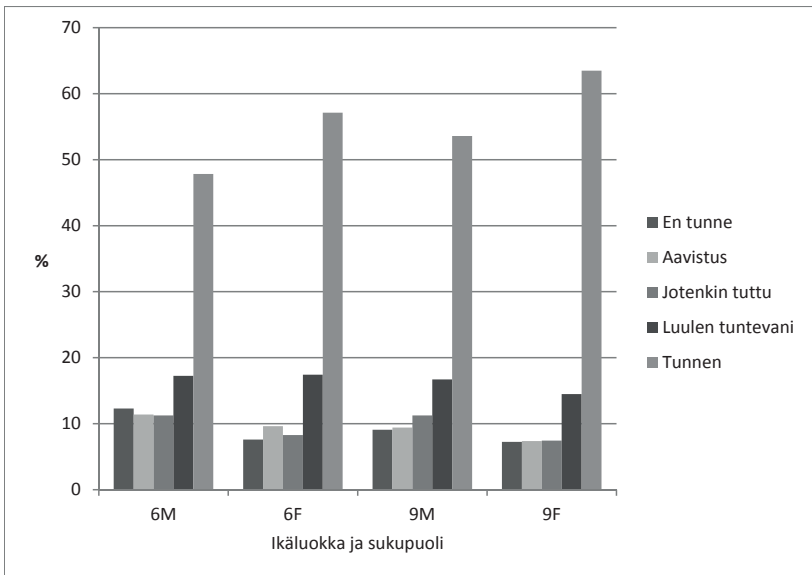
Lekseemin yleisyys / Ikäluokka %	Alakoulu 1.–5.	Alakoulu 6.	Yläkoulu	Aikuiset
Hyvin yleinen	12,2	13,8	14,2	7,2
Yleinen	14,2	15,0	15,8	10,6
Tavallinen	18,9	18,4	20,6	17,2
Harvinainen	44,3	44,6	43,0	56,9
Puuttuu korpuksesta	10,5	8,2	6,5	8,1

Unelmasanaston sanaesiintymistä kaksi kolmannesta kuuluu suomen tuhannen yleisimmän lekseemin joukkoon eli ne ovat hyvin yleisiä (ks. taulukko 6). Luku on peruskoululaisilla lähes 20 prosenttiyksikköä matalampi kuin kattavuuslaskelmissa (vrt. kaavio 2) ja tämä ero liittyy lekseemien toisteisuuteen. Esiintymistä noin 15 prosenttia kuuluu yleiseen sanastoon (yleisyystasot 2–4), kymmenisen prosenttia tavalliseen (yleisyystasot 5–9) tai harvinaiseen sanastoon. Peruskoululaisten välillä yleisyysuhteissa ei ole suuria eroja, mutta trendi on lievästi laskeva. Nuoret aikuiset eroavat selvemmin peruskoululaisista kuin nämä toisistaan.

Taulukko 6. Unelmasanaesiintymien jako yleisyystasoittain.

Sanaesiintymien yleisyys / Ikäluokka %	Koululaiset	Aikuiset
Hyvin yleiset	67,7	60,9
Yleiset	14,3	15,9
Tavalliset	8,2	10,0
Harvinaiset	8,7	12,2
Puuttuu korpuksesta	1,1	1,0

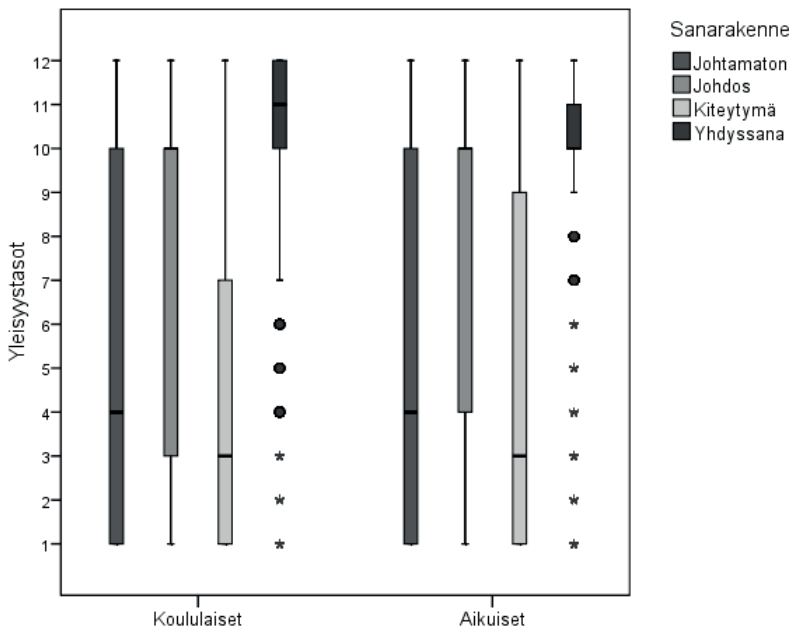
Peruskoululaisilla sanasto on melko tavanomaista ja tuttua, ja ero aikuisiin on lekseemi- ja esiintymätasolla hyvin selvä. Eroa selittää ainakin osaksi se, että koululaisilla on vielä suuria puutteita sanastossaan. Peruskoululaisille tehdyssä tuttuustestissä, jossa aineistona oli 40 melko harvinaista liikeverbiä, sanoista tunnettiin hyvin noin puolet (ks. kaavio 3). Loppuja tunnettiin heikosti tai ei ollenkaan. Esimerkiksi verbit *kuhista*, *parveilla* ja *rämpä* olivat monille tuntemattomia. Aineisto on testattu myös nuorilla aikuisilla ja heillä tuntemus on ollut yli 90 prosenttia. Luetun ymmärtämisen kannalta katsotaan, että noin 95 prosenttia sanaesiintymistä pitäisi olla tuttuja, ennen kuin luettu oikeasti ymmärretään (Milton 2009, 52). Prosentti ei ehkä suoraan sovellu suomeen, jossa ymmärtämisongelmia voivat tuoda myös harvinaisemmat taivutus- ja etenkin johdinainekset sekä näiden käyttöön liittyvä morfofonologinen sanavartaloiden vaihtelu, mutta lienee silti viitteellinen.



Kaavio 3. Melko harvinaisten liikeverbien tuttuudesta kuudes- ja yhdeksäsluokkalaisilla (N = 157).

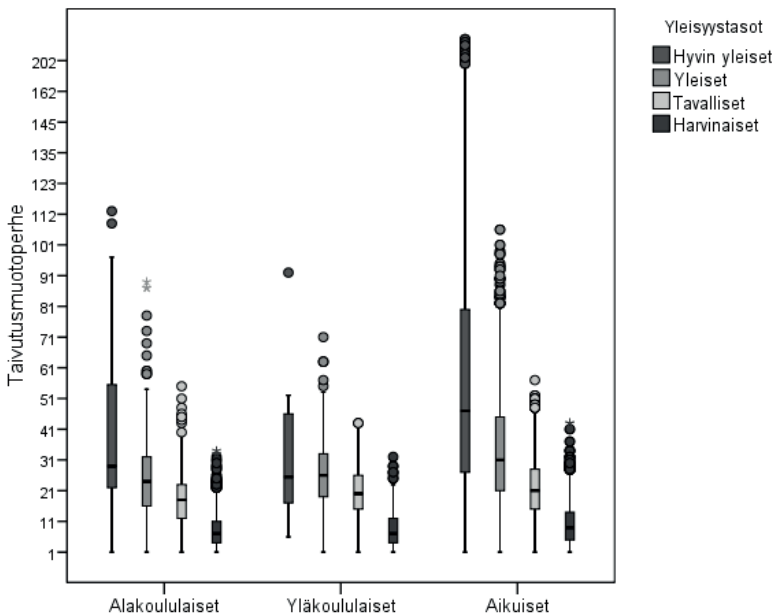
Yleisyystasoluokitukseen heijastuu suomen tapaisessa kielessä myös lekseemien sanarakenne (ks. kaavio 4).<sup>16</sup> Johtamattomien sanojen yleisyystaso on keskimäärin matalampi kuin johdettujen ja johdettujen matalampi kuin yhdyssanojen. Kiteytymät eli lähinnä substantiivien tai adjektiivien paradigmasta irronneet sanat ovat tyypillisesti adverbeja, ja ne sijoittuvat alimpiin yleisyystasoluokkiin. Yhdyssanoissa mediaani on korkein. Yleisyystaso nousee kehityksellisesti jokaisessa sanarakennetyypissä eli sanasto harvinaistuu. Esimerkiksi johdetuista lekseemeistä keskeisin osa sijoittuu koululaisilla yleisyystasoille 3–10, aikuisilla tasoille 4–10. Koululaisilla on enemmän vertailukorpuksesta puuttuvia yhdyssanoja, jotka aineistossa ovat satunnaismuodosteita.

16 Niin sanotussa boxplot-kaaviossa pylvään keskiviiva ilmaisee keskimmäisen arvon eli mediaanin, leveä osa sen ympärille sijoittuvat havainnot (yhteensä 50 % kaikista havainnoista) ja viivat ala- tai yläkvartiilin havainnot. Pallot ja tähdet viittaavat poikkeaviin havaintoihin. Kaaviossa 4 yleisyystaso 11 ja 12 viittaavat korpuksesta puuttuviin sanoihin ja satunnaismuodosteihin, epäsanat on laskelmista poistettu.



Kaavio 4. Sanan yleisyystason ja sanarakenteen riippuvuus ikäluokittain.

Taivutusperheiden ja yleisyystasojen ristiintaulukointi osoittaa, että yleisyys ja taivutusmuotojen määrä korreloivat yleisimmissä lekseemeissä hyvin vahvasti (ks. kaavio 5): kaikki lekseemit, joille massakorpuksista löytyy yli sata eri taivutusmuotoa tai niiden kombinaatiota kuuluvat suomen yleisimpien lekseemien joukkoon. Tuhannen yleisimmän joukkoon kuuluvilla taivutusmuotoja voi olla kolmekin sataa, mutta mediaani on huomattavasti matalampi, noin 50. Suurin osa paljon taivutusmuotoja saaneista lekseemeistä on verbejä, ja yksittäisellä verbillä on sanomalehtitekstissä aina enemmän taivutusmuotoja kuin yksittäisellä substantiivilla toisteisuuserojen takia. Erosta voi päätellä, että verbisyn-taksin opetus voisi hyödyttää, jos kirjoittamistaitoja halutaan edistää. Monella koululaisella ongelma on lauseenmuodostuksessa, ja juuri verbit ovat lauseenmuodostuksen ydin. Eri verbimuodoilla tuotetaan paitsi leksikaaliset myös temporaaliset ja modaaliset merkitykset sekä muute-taan lauseen argumenttien asemaa.



Kaavio 5. Massakorpuksen taivutusperheet ja Unelmasanaston lekseemit yleisyystasojakauman mukaan ikäluokittain.

Lehtikielessä lekseemien taivutusperheiden koko laskee yleisyystasoinen siten, että harvinaisempia taivutetaan vähemmän kuin yleisempiä. Nuorempien ja vanhempien unelmasanasto näyttää olevan hiukan erilaista taivutusperheiden suhteen, koska aikuisilla käytössä olevia sanoja taivutetaan lehtikielessä enemmän kuin niitä sanoja, joita koululaiset käyttävät (ks. kaavio 5). Ero näkyy monista yleisyysluokista ja syntyy pitkälti aikuisten melko abstraktimerkityksisistä ja pääosin johdetuista verbeistä. Yleisimmistä verbeistä voi esimerkkeinä mainita verbit *käsitellä*, *korostaa*, *arvostaa*, *sallia* ja *poistaa* ja harvinaisista muun muassa verbit *sovitella*, *turhautua*, *kontrolloida*, *aloitella*, *puhutella*; peruskoululaiset eivät näitä verbejä käytä lainkaan. Ero viittaa siihen, että aikuisten sanasto vastaa paremmin lehtikielen sanastoa kuin koululaisten, ja siihen, että aikuisilla sekä leksikaalinen että syntaktinen variaatio on suurempi kuin koululaisilla. Peruskoululaisten ryhmissä taivutusperheiden koko kasvaa iän mukaan tasaisesti ja osoittaa muutoksia sanastossa. Yläkoululaissanastossa



on muun muassa verbit *arvioida*, *uudistaa*, *hävittää*, *lakkauttaa*, *miellyttää*, *teloittaa* ja *sulattaa*, jotka kokonaan puuttuvat alakoululaisilta.

Massakorpuksesta laskettujen taivutusperheiden koko ilmaisee myös sen, miten moninaisesti kielenkäyttäjät altistuvat tietyille lekseemille. Tällainen luku sopii käyttöpohjaisen kielenoppimisteorian (ks. esimerkiksi Tomasello 2003; myös MacWhinney 2001, 464) mukaisiin näkemyksiin, vaikka ei kerrokaan koko totuutta siitä, mitä yksittäisiä sanoja eri yksilöt kohtaavat tai merkityksellistävät ja todennäköisimmin oppivat. Tulokset osoittavat selvästi, että Unelmasanastossa on eniten esiintymiä sanoista, joiden taivutus on massakorpuksissa moninaista. Vähemmän taivutettujen lekseemien esiintymät kasvavat kuitenkin lineaarisesti. Leksikon ja sen käyttökontekstien jatkuva laajeneminen ja harvinaisempiin lekseemeihin ja muotoihin kehittyminen on vastoin radikaaleimpia – ja yksinkertaisimpia – frekvenssitekijöiden roolia korostavia näkemyksiä, joiden mukaan ilmaukset automatisoituvat yli perinteisten morfeemi- ja sanarajojen ja muuttuvat leksikaalisiksi kollokaatioiksi (vrt. esimerkiksi Bybee & Hopper 2001). Tällainen näkemys ei ylipäätään sovellu suomen tapaiseen taivutuskieleen.

## Unelmasanaston diversiteetti

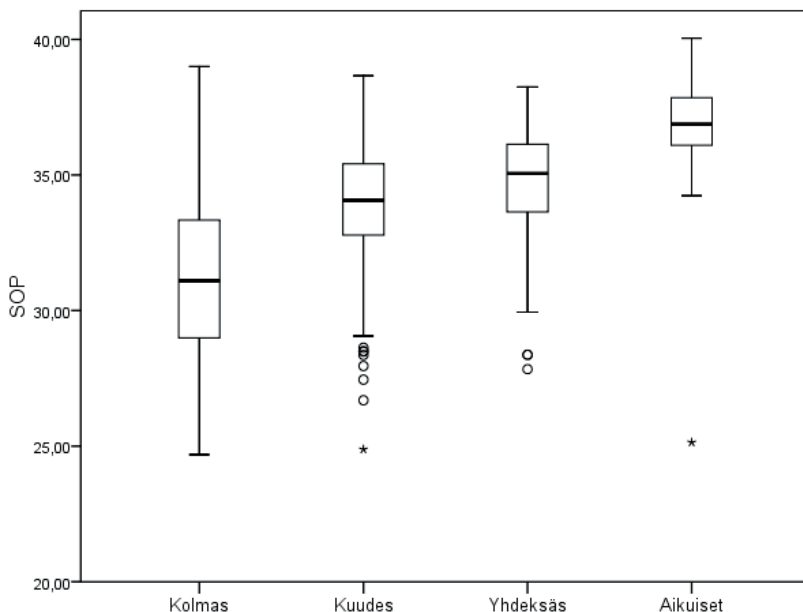
Sanaston diversiteettiä koskevaan tarkasteluun poimittiin otos eri-ikäisten aineistoa. Alkuvaiheessa otettiin mukaan kaikki kolmannen, kuudennen ja yhdeksännen vuosiluokan koululaisten sekä ammattikorkeakoululaisten kirjoitelmat eli yhteensä 729 tekstiä, joista 67 poistettiin seuraavassa vaiheessa liian lyhyinä (tekstin sanaesiintymiä oli alle 42). Lopullisen aineiston kokonaissanamäärä oli yhteensä 105 942 sanetta ja 662 kirjoitelmaa (ks. taulukko 7).

Unelmakirjoitelmien sanaston diversiteettiä on verrattu kirjoittajien ikään (kaaviot 6–7). Kaaviot on piirretty koko siitä aineistosta, josta käytetyn diversiteettitunnusluvun (*sum of probabilities*, SOP) laskeminen on mahdollista (kirjoitelman sanamäärä on vähintään 42).<sup>17</sup> Käytetty

17 Koska kaaviot on pyritty pitämään luettavassa muodossa, yksi yhdeksäsluokkalaisen kirjoitelman poikkeavan korkea arvo (lähes 50) jää näkymättömiin.

Taulukko 7. Unelma-aineiston määrä diversiteetilaskelmissa.

Ikäluokka	Kirjoitelmat	Kirjoittajan sukupuoli	Poistot	Sanamäärä
Kolmas	154	69 F, 85 M	30	10 422
Kuudes	313	160 F, 153 M	27	41 543
Yhdeksäs	161	88 F, 72 M	10	24 983
AMK	101	43 F, 58 M	Ei	28 994

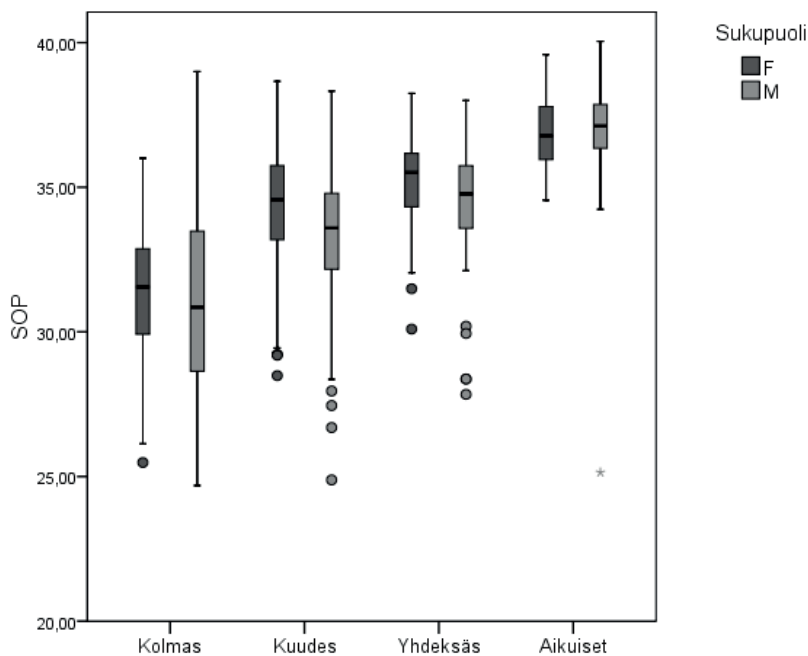


Kaavio 6. Kirjoittajan iän yhteys kirjoitelmien sanaston monimuotoisuuteen (SOP) ikäluokittain.

mittaustapa on arvioitu tekstipituuden suhteen melko neutraaliksi ja siten turvalliseksi hyvinkin eripituisten tekstien arvioinnissa (McCarthy & Jarvis 2007, 460; ks. myös Malvern, Richards, Chipere & Durán 2009), mutta kaikkein lyhimmissä teksteissä yleisimmät sanat eivät ehdi toistua yhtä todennäköisesti kuin hieman pidemmissä. Tarkastellussa aineistossa sanaston keskimääräinen diversiteetti kasvaa lähes lineaarisesti ikäluokasta seuraavaan (ks. kaavio 6). Samalla yksilöiden välinen

hajonta pienenee eli vanhimpien kirjoittajien ryhmissä yksittäisten kirjoitelmien sanaston diversiteetti on lähempänä ikäluokan keskimääräistä tasoa. Kehitys on selkeä ikäryhmien ylintä neljänestä lukuun ottamatta: vaikka sanastollisesti köyhiä kirjoitelmia esiintyy nuorimmilla kirjoittajilla selvästi enemmän kuin vanhemmilla kirjoittajilla, jo kaikkein nuorimpien aineistossa on myös sanastoltaan vaihtelevia tekstejä. Tätä selittää osittain leksikaalisten sidoskeinojen vähäinen käyttö ja nuorimpien koululaisten kirjoitelmien lyhyys, mutta osittain myös nuorimmilla on sanastollisesti moninaisia ja muutenkin tasokkaita kirjoitelmia.

Sukupuolen ja sanaston diversiteetin yhteys ei ole yksiselitteinen, sillä vaikka peruskoululaisten kirjoitelmissa sanasto näyttää olevan vaihtelevampaa tyttöjen kuin poikien teksteissä (ks. myös Honko 2013), nuorten aikuisten kirjoitelmissa tilanne on päinvastainen (kaavio 7). Poikien ryhmissä yksilölliset erot sanaston diversiteetissa ovat hieman suurem-



Kaavio 7. Kirjoittajan sukupuolen yhteys kirjoitelmien sanaston monimuotoisuuteen (SOP) ikäluokittain.

mat kuin tyttöjen ryhmissä, mutta sekä pojilla että tytöillä yleistaso on muutamaa poikkeusta lukuun ottamatta jo peruskoulun yhdeksännellä luokalla vähintään kohtalaista tasoa (SOP >30, englannin kielestä McCarthy & Jarvis 2007, 464).

Samaan aikaan kun kirjoitelmien sanaston diversiteetti kirjoittajan iän mukaan lisääntyy, myös sanojen keskimääräinen pituus kasvaa. Molempia piirteitä voidaan pitää kehityksellisinä siinä mielessä, että ryhmätasolla ne kuvaavat kehittyvää kieli- ja kirjoitustaitoa, joka tarkastellussa aineistossa on kiinteässä suhteessa kirjoittajan ikään. Myös tietyn ikäluokan sisällä tekstipituuden ja sanaston diversiteetin välillä on tarkastellussa aineistossa löyhä yhteys: pisimmät kirjoitelmat ovat sanastoltaan moninaisempia hieman todennäköisemmin kuin lyhyet. Sanaston diversiteetti ei yksilötasolla kuitenkaan ole voimakkaasti riippuvainen tekstipituudesta, mikä on odotuksenmukaista myös aiemman tutkimuksen valossa (Honko 2013). Esimerkiksi yhdeksäsluokkalaisten kirjoitelmissa korrelatiivinen yhteys tekstipituuden ja sanastollisen diversiteetin välillä on heikko, eikä sitä voi pitää kirjoittamisen ennustettavuuden kannalta merkityksellisenä. Siitä huolimatta heikkokin (tilastollisesti merkitsevä) positiivinen yhteys vahvistaa näkemystä leksikaalisesta diversiteetistä kehityksellisenä piirteenä.

On huomionarvoista, että vanhempien kirjoittajien tekstit ovat paitsi sanastoltaan vaihtelevampia myös pisimpiä, vaikka tekstipituuden kasvaessa korkean sanastollisen diversiteetin ylläpitäminen ja kasvattaminen käy yhä haastavammaksi. Sanaston vaihtelevuuteen vaikuttaa luonnollisesti myös aiheiden määrä. Heinosen (2018) mukaan kuudes- ja yhdeksäsluokkalaisten sekä ammattikorkeakoululaisten Unelmakirjoitelmissa on käsittelyssä keskimäärin 3–4 aihetta, mutta kaikkein nuorimmilla aiheita on enemmän kuin aikuisilla. Yleisimmät aiheet ovat harrasteet, ravinto, ystävät ja hyödykkeet. Nuorimmat kirjoittavat mielellään perheestä ja kuulumisuudesta, vanhimmat urasta, arvoista ja parisuhteesta. Kun aikuiset kirjoittavat pidempiä kirjoitelmia kuin kouluikäiset, he samalla käsittelevät harvempia aiheita mutta kutakin aihetta monipuolisemmin. Toisin sanoen nuorimmilla sanaston monipuolisuus syntyy useammin pinnallisista luettelonomaisista kirjoitelmista, aikuisilla monipuolisesta saman aiheen käsittelystä.

Aineistosta on lisäksi laskettu leksikaalinen tiheys (sisältösanojen osuus sanamassassa) ja substantiivi–verbi-indeksi eli substantiivien osuus substantiivien ja verbien summasta (NTVR eli *noun-to-verb ratio*). Näiden lukujen on ajateltu kertovan tekstien informaatiotiheydestä ja tiivyydestä, ja niiden suhde sanastolliseen diversiteettiin on siksi kiinnostava. Leksikaalinen tiheys on Unelmakirjoitelmissa 75–80 prosenttia kokonaissanamäärästä ja nousee hieman iän myötä, kuten Pajunen ja Vainio tässä kirjassa osoittavat. Leksikaalisen tiheyden ja sanaston diversiteetin välinen yhteys koko tekstiaineistossa on heikko mutta positiivinen. Sanastoltaan vaihtelevissa kirjoitelmissa vaihtelu saavutetaan todennäköisimmin käyttämällä runsaasti avointen eli sisältösaneluokkien sanoja, ja kaikkein toisteisimmissa kirjoitelmissa puolestaan kieliopillisten sanojen osuus on hieman keskimääräistä suurempi. Mitä enemmän leksikaalinen tiheys jää alle kirjoitelmista laskettujen mediaaniarvojen, sitä varmemmin myös sanaston diversiteetti jää alle ikämediaanin. Tämä tulos sopii jo aiemmin esiin tulleeseen tendenssiin, jonka mukaan ikäryhmittäisissä mediaaneissa pyörivät arvot viittaavat ikäluokan osaamistasolla onnistuneeseen tekstiin.

Substantiivi–verbi-indeksi on kirjoitelmissa ja oletettavasti kertovassa tekstissä ylipäätään keskimäärin 1,1:1 ja nousee iän mukaan muutaman prosenttiyksikön, tavanomainen vaihteluväli on Pajusen ja Vainion mukaan 50–57 prosenttia eli keskimäärin hieman enemmän substantiiveja kuin verbejä (ks. myös Biber, Conrad & Reppen 1998). Jos arvo on matala, teksti koostuu peruslauseista ja on lähellä puhutun kielen rakennetta, mutta jos arvo on korkea, teksti on rakenteeltaan kompleksista (paljon alisteisia verbirakenteita) ja/tai abstraktia (paljon substantiiveja). Siten oletus on, että substantiivi–verbi-mediaanissa olevissa kirjoitelmissa sanasto olisi melko vaihtelevaa. Toisaalta oletus myös on, että mitä korkeampi substantiivi–verbi-indeksi on, sitä vaihtelevampaa sanasto on. Nämä oletukset näyttäisivät pitävän paikkansa yhdeksäsluokkalaisilla ja etenkin aikuisilla. Kuudesluokkalaisilla hajonta kummassakin arvossa on suurta ja viittaa toisaalta puhutun kielen kaltaiseen yksinkertaiseen rakenteeseen, toisaalta lauseenmuodostamisongelmiin.

## Unelmasanaston semanttinen analyysi

Unelmasanaston semanttinen analyysi alkaa sanaluokkajaolla, joka on perustaltaan ontologis-semanttinen: substantiiveilla nimetään olioita, adjektiiveilla niitä kuvaillaan ja verbeillä predikoidaan substantiiveilla nimettävistä olioista paitsi tapahtumia myös tiloja. Tämä kuvaus edustaa tietysti sanaluokkia prototyyppeinä (Croft 1991), ja se korreloi kielittäin muun muassa merkinnän, kuten taivutuksen ja johtamisen, luonteeseen ja määrään (suomesta ks. Pajunen 1994; 1998; 2010). Substantiivit, verbit ja adjektiivit edustavat niin sanottuja pääsanaluokkia, ja ne voidaan määritellä paitsi ontologis-semanttisesti myös fonologisin, morfologisin, syntaktisin ja pragmaattisin kriteerein. Keskityimme tässä yhteydessä analysoimaan semanttisesti erityisesti substantiiveja ja vain sivuamme muita sanaluokkia, koska juuri substantiivien semanttinen luokitus paljastaa parhaiten kirjoitelma-aiheen käsittelyn iänmukaiset muutokset. Substantiivien luokituksesta ja niiden roolista Unelmakirjoittelmissa ei myöskään ole aiempaa tietoa (Koulukorpuksen adjektiiveista, ks. Kuisma 2011; adverbeista, ks. Koivunen 2012; verbeistä ks. Pajunen 2012).<sup>18</sup> Kieliopillisia sanaluokkia ei tässä yhteydessä analysoida määriä pidemmälle (Koulukorpuksen konjunktioista ks. Nahkola 2012; modaalisuuden ilmaisuista ks. Keskinen 2012; pronomineista Palonen 2013).

Substantiivien analyysi nojaa paitsi semanttiseen kirjallisuuteen (Miller 1998; Frawley 1992; Ravid 2006; Gärdenfors 2014) myös subjektiivisia semanttisia muuttujia arvioivien testien tuloksiin (vrt. Stavroula-Thaleia, Vigliocco, Del Campo, Vinson & Andrews 2011; Söderholm, Häyry, Laine & Karrasch 2013; Yap, Lim & Pexman 2015, Pollock 2018). Adjektiivien luokitus noudattaa Dixonin (1977, 2004) tunnettua luokitusta (ks. myös Pajunen 1994), verbien luokitus perustuu suomesta tehtyihin analyyseihin (Pajunen 1999; 2001; 2006), ja adverbit on luokiteltu lähinnä perinteisen kieliopin mukaisesti. Subjektiivisilla testeillä

18 Koulukäisten sanastonhallintaa hankkeessa ovat erilaisin testein tutkineet muun muassa Härkönen (2012, verbi-idiomit), Jääskeläinen (2008, verbisanasto), Rantala (2012, olotilanilmaukset), Routama (2008, spatiaalinen sanasto) ja Turunen (2012, abstraktisanat).

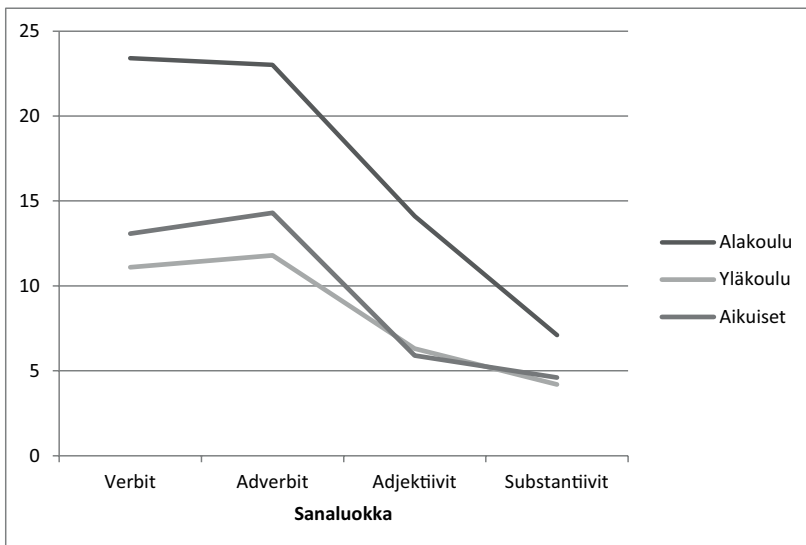
saadaan esiin kielenkäyttäjän näkemykset esimerkiksi sanan konkreettisuudesta, tuttuudesta tai vaikkapa sen herättämän tunteen voimakkuudesta tai merkityksen kielteisyydestä. Lisäksi luokittelussa välttämättä sivutaan Popperin esittämää todellisuuden kolmijakoa fysikaalisten ja mentaalisten olioiden ja tapahtumien maailmaan (maailmat 1 ja 2) sekä abstraktien entiteettien maailmaan (maailma 3), koska olioista voidaan puhua monilla eri todellisuustasoilla. Esimerkiksi sana *koulu* voi viitata tiettyyn koulurakennukseen, tietyn asteen koulutusta antavaan kouluun sekä instituutioon. Tämän kanssa risteää käsitteen ja sen toteutuman tai esiintymän suhde, joka voi todellisuudessa niin ikään olla monikerroksinen (ks. Wetzel 2009). Esimerkiksi aineesta voidaan puhua abstraktimmalla (*hyvä ruoka*) tai sen toteutuman konkreettisemmalla tasolla (*jäätelötuutti*). Unelmasanaston vertailussa olisi hyväksi, jos aineisto olisi kauttaaltaan voitu luokitella tyyppi-esiintymä-tasoisesti, mutta käytännössä tällainen tarkkuus on isossa aineistossa mahdoton. Substantiiviaineisto on luokiteltu myös suhteessa hierarkiatasoihin, joita tässä yhteydessä käsitellään vain oletuksella, että perustason jäsenet ovat johtamattomia sanoja (*kissa, koira*), geneerisen tason jäsenet voivat olla sekä johtamattomia että johdettuja (*eläin, elolliset*), ja spesifin tason jäsenten sanarakenne voi olla mikä tahansa mutta tyypillisimmin johdos tai yhdyssana (vrt. Taylor 1995, 46–51). Lisäksi aineistoon on koodattu erikseen lekseemit, joiden tarkoite on jossain määrin opaakki (*puuha, kampeet, vehje, kamat*); tähän luokkaan kuuluvat myös relationaaliset sanat (*siivu, viipale, lohko*).

Kielenhallinnan ja kirjoittamisen kehitys näkyy jo sanaesiintymien sisäisestä jakaumasta sanaluokkiin (ks. taulukko 8). Adjektiivi- ja substantiiviesiintymien määrä nousee lineaarisesti, verbien, konjunktioiden ja pronomiinien määrä laskee yhtä lailla lineaarisesti. Erot selittyvät joko kerronnan monipuolistumisen kautta tai sitten niitä selittää tendenssi yksinkertaisesta kompleksiin. Substantiivien osuus kasvaa määrällisesti eniten.

Lekseemimäärät ovat substantiiveissa suurimmat ja muissa leksikaalisissa luokissa – verbeissä, adjektiiveissa ja adverbeissa – pienempiä mutta muissa luokissa yksittäisillä lekseemeillä on enemmän esiintymiä. Korkein toisteisuus on verbi- ja adverbilekseemeillä (ks. kaavio 8). Toistei-

Taulukko 8. Unelmakirjoitelmien sanaesiintymät sanaluokittain ja ikäluokittain.

Sanaesiintymät sanaluokittain / Ikäluokka %	Alakoulu	Yläkoulu	Aikuiset
Adjektiivit	5,8	7,0	9,4
Adverbit	13,4	13,7	12,2
Konjunktiot	11,0	10,2	7,0
Numeraalit	1,3	1,1	0,8
Pronominit	10,4	9,3	7,4
Substantiivit	27,5	28,7	36,8
Verbit	30,6	29,8	26,2



Kaavio 8. Saman lekseemin toisteisuus (kpl keskimäärin) ikäluokittain ja sanaluokittain.

suus laskee kehityksellisesti, ja ero on suurin alakouluikäisten ja muiden välillä.<sup>19</sup> Substantiivit voivat olla sanarakenteeltaan johtamattomia, johdettuja tai yhdyssanoja, muut ovat pääasiassa joko johtamattomia tai johdettuja.

19 Murrosikä tai kehitystaantuma saattaa näkyä U-muotoisena kehityksenä lineaarisen sijaan (vrt. esimerkiksi McNeil 2007), ulkoisen motivaation puute pelkäästään poikkeuksena. Mainittakoon, että ulkoisen motivaation puute heijastuu eniten tuotettujen sanojen määrään, ei niinkään osaamistasoon (vrt. Kyösti 2018; Pajunen & Vainio, tämä teos).



Myös yhden esiintymän lekseemien eli HL-lekseemien (*hapaks legomenon*) määrä muuttuu kehityksellisesti siten, että niiden määrä on aikuisilla aina suurempi kuin koululaisilla. Kehityksellinen muutos on suurin adjektiivilekseemeissä eli 13 prosenttiyksikköä. HL-lekseemien määrä on adjektiivi- ja substantiiviesiintymistä 6–13 prosenttia ja verbi- ja adverbiesiintymistä vain 1,3–3 prosenttia; ero liittyy sanaluokkien lekseemimääräeroihin.

#### SUBSTANTIIVIEN SEMANTTINEN LUOKITUS

Substantiiveja voi semanttisesti jaotella monella tavalla ja joka luokituksessa on rajatapauksia. Lähtökohta voi olla käsitteellinen WordNet-ajattelun tapaan (ks. Miller 1998, 29; Vossen 2002, 5). WordNet-luokitus koostuu niin sanotuista synonyymijoukoista (*synset*), joissa sanat ovat oletuksen mukaan keskenään vaihdettavissa. Luokituksessa on 25 niin sanottua aloittajakategoriaa, ja entiteetit on jaoteltu kategorioittain toisaalta ominaisuuksia periyttäväksi hierarkiaksi, toisaalta yksittäisiksi. Hierarkiassa alaluokkia ovat organismit (eläimet, inhimilliset, kasvit), objektit (artefaktit, luonnonilmiöt, aine), abstraktiot (ominaisuus, määrä, suhde, aika) sekä psykologiset ominaisuudet (kognitio, emootio ja motivaatio). Alaluokittamattomia entiteettejä ovat aktiviteetti ja tapahtuma, paikka ja tila, ryhmä, omistus ja muoto. Millerin luokitusta on sovellettu lukuisissa WordNetin kielisovelluksissa (ks. esimerkiksi Fellbaum 1998; Vossen 2002). Substantiivit on Unelmaluokituksessa jaoteltu Millerin luokitusta soveltaen neljään pääryhmään a)–d) mutta pääasiassa käsittelemme substantiiveja näiden ryhmien alaryhminä, joissa on jonkin verran otettu tyyppi–toteutuma-jakoa huomioon:

- a) Organismit: inhimilliset (yleiset ja roolinimet), elolliset (eläimet, kasvit; elollisista predikoitavat tilat ja tapahtumat; ruumiinosa-nimet)
- b) Objektit ja muut: elottomat (konkreettiset ja laskettavat esineet; konkreettiset spatiaaliset rakennelmat) sekä ainesanat (aine; ruokanimi ja ruoka-annos), luonnontapahtumat, kollektiivit ja institutiot; spatiaaliset alueet ja luonnolliset muodostelmat; tapahtumanimet

- c) Abstraktiot: abstraktioasteeltaan vaihtelevat asianimet, temporaaliset nimet
- d) Luokittelun ulkopuoliset

Substantiivien luokittelu heijastaa paitsi Millerin luokitusta myös elollisuus- ja konkreettisuusasteikkoa (ks. myös Frawley 1992; Ravid 2006; Gärdenfors 2014, 115–134). Osaksi luokittelu on riippuvainen aineistosta. Elollisuusasteikolla oliot jakautuvat inhimillisiin, elollisiin, elottomiin konkreettisiin sekä abstrakteihin. Elollisuus voidaan määritellä joko biologian (kädelliset ja muut eläimet, kasvi, bakteeri ja niin edelleen) tai kielen kriteerein (ihminen, korkeaelollinen eli inhimillistettävä, mielikuvan luova eläin, muut eläimet ja niin edelleen) (ks. esimerkiksi Frawley 1992, 89). Konkreettisuusasteikolla konkreettisuus vähenee suunnilleen suhteessa elollisuusasteikkoon mutta alaluokissa vaihtelevammin. Konkreettisimpia ovat niin elolliset, havaittavat, kosketeltavat kuin laskevat ja manipuloitavat materiaaliset objektit. Konkreettisuus laskee, jos elottomalla oliolla on suuri ulottuvuus, koko tai vastaava (esimerkiksi vuori vs. käteen mahtuva kivi), mikä estää sen käsiteltävyyttä, kosketeltavuutta ja kerralla havaitsemista (ks. Wetzel 2009; Pajunen, Itkonen & Vainio 2016; Rosen 2017). Oliot – ja niiden nimet – eivät kuitenkaan sijoitu suoraan kummallekaan asteikolle, joten aineisto on ensin luokiteltu pienempiin semanttisiin luokkiin, joiden elollisuus- ja konkreettisuusaste on arvioitu myöhemmin erikseen. Niin sanottujen subjektiivisten semanttisten ominaisuuksien – kuten konkreettisuus, kuviteltavuus, polaarisuus ja niin edelleen – arviointi perustuu koehenkilötesteihin, ja ne arvioidaan asteikollisina (ks. esimerkiksi Pajunen & Itkonen 2019).

Elollisuusasteikon elolliset lukeutuvat organismeihin, ja ne on melko helppo yksilöidä. Organismeihin sijoittuvat tässä yhteydessä myös elollisiin osa-kokonaisuussuhteessa olevat ruumiinosanimet ja edelleen elollisesta predikoitavat fysiologiset ja psykologiset tilat ja tapahtumat (*uni, stressi; alitajunta, innostus*), jotka elollisuusasteikolla edustavat abstraktiota. Elottomia objekteja ovat kaikki (melko) konkreettiset oliot, joita ei ole luettu elollisiin. Kaavioissa elottomia eritellään tarpeen mukaan eri alaluokkiin. Rajatuimmin elottomilla viitataan esineisiin (*kynä, pöytä, ovi*), rakennelmiin (*laituri, kivitalo*) sekä ontologialtaan useammantasoiseihin

instituutioihin (*koulu, kirkko, häät*), jotka kirjoitelmissä edustavat sekä olioita (esimerkiksi koulurakennus) että instituutioita. Nämä elottomat eivät ole merkittäviä alueellisilta ulottuvuuksiltaan toisin kuin varsinaiset spatiaaliset alueet ja niihin viittaavat sanat (*suu, viljelmä*). Unelmasanaston ainesanoista suuri osa liittyy ruokaan, ja erityisesti nimetään ruokien nimiä ja ruoka-annoksia, koska syöminen kuuluu oleellisesti päiväskriptiin. Ruoka edustaa käsitteen tasoa (*type*) ja ruoka-annos sen toteutumaa (*token*). Näitä ruoan toteutumia on erityisen paljon koulu-laiskirjoittajilla, mutta eroa ei ole erikseen taulukoitu.

Konkreettisuusasteikon jaossa on oleellista erottaa havaittavat (*pöytä, juosta, vieri*) ja materiaaliset (*pöytä*) havainnontakaisista (*ajatella, ajatus, päätös*) ja immateriaalisista (*ajatella, juosta, vieressä*). Havaittavia ja materiaalisia voi todennäköisesti koskettaa ja liikuttaa, ne sijaitsevat jossakin ja niillä on muoto, hahmo tai muu sellainen. Periaatteessa kaikkein konkreettisimpia ovat ne, jotka voidaan havaita kaikilla aisteilla (näkö, kuulo, haju, maku, tunto ja tuntoaistin alatyypit kosketus, asentoaisti ja niin edelleen) eli esimerkiksi laskettavat ja kosketeltavat esineet. Havaittavia ja immateriaalisia ovat konkreettiset teot ja esimerkiksi läheisyyden ja etäisyyden relaatiot. Brysbaertin, Warrinerin ja Kupermanin (2014) 40 000 sanan konkreettisuusarvioinneissa koehenkilöt nojautuivat useimmiten visuaaliseen ja haptiseen aistiin. Sellaisten substantiivien konkreettisuus onkin yleensä korkea, joiden kuviteltavuuskin on korkea, joista on melko helppo saada mielikuva.

Suomalaiskoehenkilöt (yhteensä 28) arvioivat eri semanttisiin luokkiin kuuluvia sanoja sekä kuviteltavuuden että konkreettisuuden suhteen. Arvontiasteikko oli 1–5. Tulosten mukaan objekti- eli esinesanojen sekä luonto- ja ainesanojen kuviteltavuus on hyvin korkea ja niiden konkreettisuus on joko hyvin korkea (esineet) tai melko korkea (luontosanat). Relationaaliset sanat jakaantuivat sekä kuviteltavuuden että konkreettisuuden suhteen pääasiassa niiden moniontologisuuden mukaan: esimerkiksi sanoilla *raja, kolka* tai *pieli* voi olla sekä spesifi merkitys (*mökki lähellä Venäjän rajaa*) että yleinen, abstraktimpi merkitys (*rajana vain mielikuva*). Abstraktisubstantiivien kuviteltavuus on matala, mutta konkreettisuusasteikoille ne sijoittuvat koehenkilönäkemyksen mukaan joko melko abstrakteihin (*haju, tauko, huhu*), abstrakteihin (*huoli,*

*harmi*) tai hyvin abstrakteihin (*usko*). Tämä ero luonnollisesti kuvastaa taas kerran eroa tyyppin ja sen toteutuman välillä (vertaa *hieno tuoksu* ja *laittaa tuoksu*).<sup>20</sup>

On syytä huomata, että vaikka konkreettisuus voidaan perustaa ontologiin kriteereihin, abstraktius liittyy sekä kieleen että mieleen. Lisäksi abstraktit merkitykset voivat olla kulttuurisia ja osin yksityisiäkin. Oppimisen kannalta ero on huomattava ja vaikuttaa oppimisen lisäksi käyttöön. Konkreettisuusasteikko on kehitykselliseltä kannalta mielenkiintoinen ja tiedetään, että erityisesti substantiiviluokissa kielen ja kirjoittamisen kehitys näkyy tekstin abstraktistumisena (Ravid 2006). Alakouluikäisillä onkin vielä paljon puutteita abstraktisanaston hallinnassa (ks. esimerkiksi Turunen 2012). Puutteet heijastuvat hyvin todennäköisesti myös abstraktisanaston käyttöön, koska konkreettisten sanojen motivoitua syntaksi hallitaan paremmin kuin motivoitumattoman abstraktisanaston. Ero näkyy niin kokeellisesti (ks. esimerkiksi Barsalou 2008) kuin erityisryhmilläkin (Kukkonen & Pajunen 1985).

Koko Unelma-aineistossa on analysoitavia substantiivileksemejä 8 863 ja niiden esiintymiä yhteensä 63 651.<sup>21</sup> Substantiivien semanttisten luokkien jakaumat ovat eri ikäryhmissä aika samanlaisia, mutta selvä siirtymä konkreettisista abstrakteihin on näkyvissä. Vastaava kehityksellinen trendi on vielä selvempi substantiiviesiintymissä (ks. taulukko 9). Keskeinen on 5–10 prosenttiyksikön siirtymä inhimillisistä ja elollisista sekä konkreettisista elottomista abstraktimpiin luokkiin. Toisin sanoen aivan odotuksenmukaisesti iän myötä kirjoitetaan samasta aiheesta selvästi abstraktimmin. Ainesanojen osuus pienenee; ruoasta puhutaan hieman vähemmän. Inhimillisten roolinimistä (*opettaja, kirjailija*) ja ryhmistä (*porukka, ryhmä, pelijoukkue*) kirjoitetaan samoin tai hieman enemmän vanhemmissa ikäryhmissä, samoin aluesanoista ja spatiaalisista ulottuvuuksista (*metsä, kulkureitti, matka*), mutta elollisista – kuten lemmikeistä – kirjoitetaan vähemmän. Ajan ilmaukset (*ikä, kuukausi, tulevaisuus, vapaapäivä*) lisääntyvät. Suurimmat prosentuaaliset muutokset

20 Testin jälkeen koehenkilöt kommentoivat *raja*-tyypin sanojen luokittelua sanomalla, että oli vaikea tietää, pitikö luokitella yleistasolla vai ei.

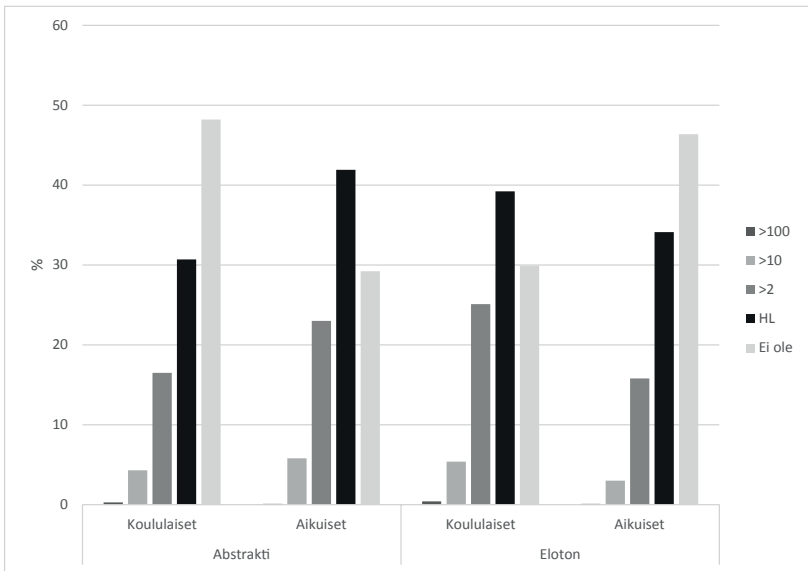
21 Substantiiviluokituksen on tehnyt kaksi henkilöä toisistaan riippumatta ja tulokset on myöhemmin yhtenäistetty.

Taulukko 9. Substantiiviesiintymien jakauma semanttisiin luokkiin ikäluokittain.

Substantiiviesiintymien semanttinen luokitus / Ikäluokka %	Alakoulu n = 22 560	Yläkoulu n = 9 779	Aikuiset n = 26 477
Inhimillinen	10,4	7,5	6,6
Inhimillisten rooli	2,3	2,5	2,3
Elolliset	6,5	4,6	2,5
Inhimillisen tai elollisen tila tai osa	3,2	4,4	8,3
Ryhmä	1,2	1,5	1,5
Aine	8,1	9,0	7,0
Eloton	30,8	28,0	22,5
Alue	6,5	7,8	6,7
Ulottuvuus	2,2	2,7	2,5
Relaatio	0,8	1,2	1,1
Aika	14,5	15,5	16,0
Asia	13,5	15,3	22,9

koskevat elottomia konkreettisia objektisanoja (31 % > 22 %) ja elottomia abstraktisanoja (13 % > 23 %). Elottomien alaluokissa muutos näkyy parhaiten esinimien vähenemisenä ja teon nimitysten lisääntymisenä. Esinimet (*kumi, kynä*) ovat konkreettisuusasteikolla korkeammalla kuin teonnimet (*vihellys, potku*), joten elottomiin viittaavan substantiiviluokan sisälläkin näkyy iänmukainen abstraktistumismuutos, vaikka luokan koko sinänsä pienenee iän mukaan. Jos luokituksessa olisi otettu huomioon myös se, viittaako kirjoittaja käsitteeseen tai instituutioon niiden toteutumien sijaan, kehityksellinen ero näkyisi selvemmin.

Suurimmat kehitykselliset muutokset esiintymätasolla näkyvät siis objektisanoissa ja abstraktisanoissa. Jos eroa tarkastellaan lekseemitasolla ja esiintymäluokkatasolla (kaavio 9), nähdään, että koululaisilla ja aikuisilla näiden sanojen luokat koostuvat osin samoista lekseemeistä, mutta erot kasvavat siten, että koululaisilta puolet koko aineiston yhteenlasketuista abstraktisanoista puuttuu, aikuisilta vain noin 30 prosenttia. Lisäksi aikuisilla on abstrakteja HL-lekseemejä enemmän kuin koululaisilla. Objektisanoissa tilanne on päinvastainen. Erot näkyvät vielä sellaisissa yleisyysluokkien lekseemeissä, joissa esiintymiä on muutama (>2 kaaviossa 9). Hyvin yleisiä luokan jäseniä on molemmilla vähän ja eroja ei niissä sanottavasti ole.



Kaavio 9. Elottomiin ja abstrakteihin viittaava sanasto koululaisilla ja aikuisilla jaoteltuna yhteiseen sanastoon (sanan esiintymiä luokittain yli sata, yli kymmenen, yli kaksi, vain yksi) ja erilliseen sanastoon (sana ei esiinny molemmilla ryhmillä eli ei ole yhteinen).

Substantiivit viittaavat olioihin, jotka voi luokitella konkreettisuusasteikolle hieman eri kohtiin riippumatta siitä, mihin semanttisiin luokkiin ne kuuluvat. Esimerkiksi konkreettisen, ihmisen tekemän, laskettavan ja manipuloitavan kokaisen esineen (*kynä, kuppi*) konkreettisuus on korkeampi kuin spatiaalisen alueen (*piha, metsä, saari*), kuten jo todettiin ja kuten koehenkilöiden testaus myös osoitti. Vastaavasti fysiologinen, elollisesta predikoitava tapahtuma (*sydänkohtaus, haukotus*) on konkreettisempi kuin vastaava mentaalinen tila tai tapahtuma (*ajatuksenjuoksu, uteliaisuus*). Lisäksi vaikka sanan tarkoite olisi konkreettinen, sana ei välttämättä kielennä tarkoitetta kovin konkreettisena. Esimerkiksi sanat, jotka viittaavat yksilöitäviin olioihin (*mies*) ovat asteikolla konkreettisempia kuin yksilöistä koostuviin kollektiiveihin viittaavat sanat (*miesryhmä, miesjoukko*), vaikka inhimilliset sinänsä kuuluvat asteikolla ykkössijalle. Vastaavasti sanan merkityksen granulariteetti eli yksityiskohtaisuus voi

Taulukko 10. Substantiiviesiintymien jakauma konkreettisuusasteikolle ikäryhmittäin.

Konkreettisuusasteikko / Ikäluokka %	Alakoulu	Yläkoulu	Aikuiset
1, konkreettinen	37,8	30,3	25,0
2	29,8	32,1	26,8
3	21,5	24,1	24,9
4	10,1	11,5	18,5
5, abstrakti	0,9	2,0	4,8

vaikuttaa konkreettisuuteen. Esimerkiksi hyponyymi–hyperonyymisuhteisessa hierarkiassa hyperonyymi (*laji, eläin*) luokituu vähemmän konkreettiseksi kuin perustason (*koira*) tai spesifitason (*susikoira*) sana. Meronyymi–holonyymisuhteiset sanat voivat olla vähemmän konkreettisiä kuin hyponyymi–hyperonyymisuhteiset, mutta erityisesti relationaaliset, elottoman osan ja kokonaisuuden suhdetta ilmaisevat sanat ovat tilanteisia tarkoitteitaan vähemmän konkreettisiä. Esimerkkinä voi mainita suomen lukuisat erilaisia reunaosia kielentävät sanat (*sivu, laita, syrjä, pääty* jne.).

Sekä substantiivilekseemien että -esiintymien (ks. taulukko 10) yhteydessä konkreettisuus vähenee kehityksellisesti lähes lineaarisesti. Erot alakoululaisten ja aikuisten välillä ovat asteikon ääripäissä yli 10 prosenttiyksikön luokkaa eli myös tilastollisesti merkitseviä. Konkreettisuusluokituksessa inhimilliset ja elolliset sekä etenkin konkreettiset esinesanat edustavat konkreettisimpia, kollektiivisanat, ainesanat ja aluesanat pääosin luokkaa 2. Luokkiin 3 ja 4 sijoittuu vielä kuviteltavissa oleviin tarkoitteisiin viittaavia jo jonkin verran abstrakteja sanoja ja erilaisia verbaalisiin tarkoitteisiin viittaavia sanoja. Kumpaankin ryhmään kuuluvia sanoja voidaan ajatella joko yksilöivästi tai yleistasolla (esimerkiksi *allekirjoitus* todistuksessa tai todisteena), tai niihin voi liittyä voimakkaita negatiivisia tunnereaktioita (*sää, kirous*) ja siten konnotaatioita käytönkin tasolla. Puhtaana abstrakteja ovat sanat, jotka eivät herätä mielikuvia tai (yleensä) kielteisiä tunnereaktioita (esimerkiksi *hyvyys, aitous*) (ks. Söderholm ja muut 2013).

Substantiivit eroavat muiden pääsanaluokkien sanoista sikäli, että lekseemien määrä on suuri ja yksittäisen lekseemin yleisyys pieni. Suurin osa Unelmanaston substantiiveista kuuluukin yleisyystasolle 10 (mediaani), ja aika monissa semanttisissa luokissa neljännes substantiivi-

lekseemeistä on harvinaisia tai puuttuu lehtikorpukselta. Kirjoitelman näkökulma aiheeseen ja kulloinenkin teema vaikuttavat substantiivisanastoon enemmän kuin muiden sanaluokkien valintoihin. Ero etenkin verbeihin on huomattava. Yleisimmät substantiivilekseemit löytyvät luokista inhimillisen rooli, kollektiivi, aika ja abstraktisana. Substantiivisanastossa näkyy yleisyysmuutoksia semanttisen luokan ja iän mukaan, muun muassa inhimillisiin viittaavat lekseemit sekä aika ja abstrakti-lekseemit harvinaistuvat iän mukaan. Koululaisilla roolinimet ovat harvinaisempia, ja toisaalta sekä elollisiin että elottomiin viittaavat sanat ovat yleisempiä kuin aikuisilla. Kehitykselliset erot eivät tosin ole kovin suuret. Aika usein ero näkyy siten, että kaikkein yleisin tietyn luokan sanasto puuttuu aikuisilta.

Koululaisilla johtamattomat sanat hallitsevat kaikissa semanttisissa substantiiviluokissa. Tämä on odotuksenmukaista, koska vielä yhdeksäsluokkalaisillakin on puutteita derivaation hallinnassa (ks. Vainio ja muut 2019). Aikuisilla johtamattomia on paljon luokissa inhimillinen, eloton ja asiasana sekä myös elollisen tilaa tai osaa nimeävissä. Kummallakin ikäluokalla yhdyssanoja on eniten elottomiin ja asioihin viittaavissa sanoissa. Nämä ovatkin aineiston suurimmat semanttiset luokat. Myös saman lekseemin toisteisuus siirtyy abstraktimpiin sanoihin iän myötä. Leksikaalisessa hierarkiassa johtamattomat sanat ilmaisevat tyypillisesti perustasoa (*kissa, koira*) ja johdetut ja yhdyssanat spesifiä tasoa (*siamilainen, kisuliini*). Siten spesifin tason sanoilla kielennettäessä granulariteettiaste on korkeampi kuin perustason sanoilla. Koululaisten ja aikuisten suurimmissa substantiiviluokissa eli objekti- ja asiasanoissa granulariteettiero tulee näkyviin erityisen hyvin, jos verrataan sanarakennemuutosta kunkin semanttisen luokan sisällä: aikuisilla johtamattomat sanaesiintymät laskevat koululaisiin verrattuna elottomiin viittaavissa ja asiasanoissa yli 10 prosenttiyksikköä ja inhimillisiin viittaavissa yli 30 prosenttiyksikköä johdettujen hyväksi. Aikuisilla tosin sanottuna substantiivisanasto harvinaistuu sekä abstraktistuu ja ilmaus samalla täsmentyy.



MUIDEN LEKSIKAALISTEN SANALUOKKIEN SEMANTTINEN LUOKITUS  
Verbit on luokiteltu toisaalta aikapaikkaisia asiantiloja kielentäviin ja toisaalta niihin, joilla koodataan asiointiloihin suhtautumista tai vastaavaa. Adjektiivit puolestaan on luokiteltu niiden kielentämän ominaisuuden mukaan. Adverbit jakautuvat ajan, tavan, paikan ja määrän, intensiteetin sekä kommenttien ilmaisuihin. Vertailemme tuloksia ensin semanttisten luokkien perusteella ja siirrymme sitten yleisyystasoittaiseen analyysiin.

Asiantiloja kielentävissä verbeissä on teko- ja tapahtumaverbejä, tilaverbejä ja liikeverbejä, suhtautumista tai vastaavaa ilmaisevissa mentaaliverbejä eli aistihavainto-, emootio-, kognitio- ja kommunikaatioverbejä. Lisäksi on erotettu muutama pienryhmä. Luokkien erottelukriteerit ovat paitsi semanttisia myös syntaktisia (ks. Pajunen 2001). Merkitykseltään melko konkreettisilla verbeillä (havaittavat immateriaaliset) ilmaistaan tekoja, tapahtumia ja tiloja mutta iän myötä samoja verbejä aletaan käyttää abstraktimmin. Esimerkiksi yleiset liikeverbit, kuten *mennä* ja *tulla*, voivat liikkeen sijaan ilmaista aineiston vanhemmilla koehenkilöillä muutosta. Mentaaliverbeillä ilmaistaan asiointiloihin kohdistuvia propositionaalisia asenteita ja referoidaan. Niiden käyttö lisääntyy kehityksellisesti (ks. myös Pajunen 2012).

Verbilekseemeissä tapahtumaverbien määrä nousee kehityksellisesti ja liikeverbien laskee. Mentaaliverbeissä alakoululaisilla on enemmän kommunikaatioverbejä ja aikuisilla kognitioverbejä. Esiintymien osalta kehityksellinen muutos on selvempi: mentaaliverbien esiintymämäärät – erityisesti kognitioverbien – lisääntyvät, vastaavasti liike- ja tilaverbien esiintymämäärät vähenevät (ks. taulukko 11). Teko- ja tapahtumaverbien esiintymät sen sijaan nousevat. Muutokset ovat toisin sanottuna ennako-odotuksen mukaisia, ja niitä selittää muun muassa se, että skriptimäinen, siirtymiä ja rutiineja korostava päiväskeemakirjoittaminen vähenee.

Taulukko 11. Verbiesiintymien jakauma semanttisiin luokkiin ikäryhmittäin.

Verbien semanttiset luokat / Ikäluokka %	Alakoulu n = 30664	Yläkoulu n = 10184	Aikuiset n = 25048
Tapahtuma	26,4	28,8	29,4
Liike	25,7	22,7	17,6
Tila	23,9	21,8	19,9
Mentaali	19,0	20,4	26,0
Kieliopillinen	5,0	6,4	7,1

Lekseemitasolla johdettuja verbejä on enemmän kuin johtamattomia muissa luokissa paitsi kieliopillisissa. Verbiyhdyssanat ovat suomen kielessä harvinaisia ja niitä esiintyy harvoissa verbiluokissa (ks. myös Pajunen 2006). Esiintymätasolla johtamattomat vallitsevat kaikissa verbiluokissa. Johdettujen verbien määrä kasvaa ikäryhmittäin kaikissa muissa verbiluokissa paitsi tilaverbeissä ja kieliopillisissa luokissa. Eniten ne lisääntyvät tapahtuma- ja tekoverbeissä, joissa johtaminen muuttaa argumenttirakennetta eli on syntaktista (*kulua > kuluttaa; aueta > aukaista*) tai spesifioi merkitystä (*aloittaa > aloitella*). Verbimerkityksetkin siis spesifioituvat iän myötä. Kognitioverbeissä esiintymät kasvavat sekä johtamattomissa että johdetuissa. Verbijohdosten määrän pysyminen kuitenkin suhteellisen vähäisenä lienee tekstityyppi- ja aiheominaisuus.

Unelmasanaston adjektiivit on luokiteltu Dixonin (1977, 2004) perusluokituksen mukaisesti seuraavasti: arvoa tai arvottamista ilmaisevat (*hyvä, huono, huikea, ihana, merkityksellinen*), dimensiota ilmaisevat (ulottuvuus ja muoto: *pitkä, pyöreä*), fyysistä ominaisuutta ilmaisevat (*kova, pehmeä*), inhimillistä taipumusta ilmaisevat (*avulias, ilkeä, rehti*), ikää tai aikaa ilmaisevat (*nuori, viisivuotias; keväinen, eilisiltainen*), materiaalista ominaisuutta ilmaisevat (*arkullinen, kurainen, pilvetön*) sekä väriä ja nopeutta ilmaisevat. Lisäksi on eroteltu suomessa adjektiiviluokkaan kuuluvat järjestys (*viimeinen, seuraava*), kansalaisuus (*ulkomaalainen, ruotsalainen; ruotsinkielinen*), pronominaalit ja muut (satunnaismuodosteet).

Adjektiivilekseemit jakautuvat näihin semanttisiin luokkiin odotuksen mukaisesti siten, että fyysistä ja materiaalista ominaisuutta ilmaisevia on paljon, inhimillistä taipumusta ilmaisevia suhteellisen paljon. Sen sijaan arvoa ja arvottamista ilmaisevia on paljon odotusarvoon nähden, mutta ero johtuu osaksi siitä, että aineisto on koodattu ennemmin käytötappaa kuin yksittäistä lekseemiä painottaen. Arvoa ilmaisevia lekseemejä on suomen kielessä melko vähän, mutta arvottamista voi ilmaista subjektiivisin painotuksin. Osaksi arvolekseemien yleisyys johtuu myös kirjoitelmien aiheesta, sillä monet kirjoittajat ovat pyrkineet evaluoimaan unelmapäivän käsitettä. Evaluointipyrkimys nostaa myös värilekseemien määrää. Evaluointi nousee vielä merkittävämmäksi adjektiiviesiintymien osalta (ks. taulukko 12): arvoa ja arvottamista ilmaisevia

adjektiiveja käytetään eniten sekä koululaisilla että aikuisilla (yli kolmannes esiintymistä). Dimensiota ja fyysistä ominaisuutta ilmaistaan lähes yhtä usein, muissa luokissa esiintymämäärät jäävät melko vähiin.

Koululaisten ja aikuisten välillä on havaittavissa tässäkin yhteydessä kehityksellinen ero siten, että koululaiset arvottavat adjektiiveilla useammin kuin aikuiset ja ilmaisevat myös ulottuvuuksia useammin. Nuorimmilla arvottaminen voi tosin olla kiva/hyvä-toteamusta (vrt. Kuisma 2011) sen sijaan, että teksti rakennettaisiin arvoja perustellen. Aikuiset ilmaisevat koululaisia useammin fyysisiä ja materiaalisia ominaisuuksia sekä inhimillisiä taipumuksia. Aikuisilla on lähes jokaisessa semanttisessa luokassa enemmän lekseemejä kuin koululaisilla, toisin sanottuna diversiteetti on korkeampi. Koululaisilla on enemmän värileksemejä ja -esiintymiä.

Taulukko 12. Adjektiiviesiintymien jakauma semanttisiin luokkiin ikäluokittain.

Adjektiivien luokka / Ikäluokka %	Koululaiset n = 7 680	Aikuiset n = 8 197
Arvo	39,7	36,1
Dimensio	17,3	13,5
Fyysinen	13,3	16,7
Inhimillinen	5,6	9,0
Järjestys	4,7	3,9
Materiaalinen	2,7	4,7
Temporaalinen	6,6	7,3
Väri	4,5	3,0
Muu	5,7	5,8

Adverbit on luokiteltu aikaa (*jo, päivittäin, kauan*), paikkaa (*kotiin, loitolla, taakse, ääreen*), tapaa (*alasti, hellästi, hissukseen, jalan*), määrää (*kerrallaan, lähes, osaksi, vailla*) ja suhtautumista ilmaiseviin (*muka, kenties, varmasti, valitettavasti*) adverbeihin sekä intensiteetin ilmaisuihin (*sangen, perin, kunnon*) ja pronominaaleihin (*jotenkin, miltei, silloin*), jotka voivat ilmaista kaikkia adverbien tavallisia funktioita ja lisäksi toimia lauseita yhdistävinä elementteinä. Koululaiset käyttävät aikuisia enemmän kaikkien muiden adverbiluokkien lekseemejä paitsi kommentti- ja tapa-adverbeja. Erot ovat melko suuria lähinnä paikan adverbeissa koululaisten hyväksi ja tavan adverbeissa aikuisten hyväksi (noin 5 prosenttiyksikköä ja yli

10 prosenttiyksikköä). Esiintymien osalta aikuiset käyttävät enemmän ajan, määrän ja tavan ilmaisuja ja myös kommenttiadverbeja (taulukko 13). Aineisto siten osoittaa, että aikuiset paitsi kehystävät kerrontaansa enemmän kuin kouluikäiset, he myös evaluoivat sitä melko monipuolisesti. Koululaisilla on enemmän proadverbeja (yli 10 prosenttiyksikköä), joista huomattava osa syntyy lauseketjutyyppisestä *sitten*-sanan käytöstä, myös paikan adverbeja on koululaisilla enemmän. Intensiteettiadverbeja on molemmilla ryhmillä yhtä paljon.

Taulukko 13. Adverbiesiintymien jakauma semanttisiin luokkiin ikäluokittain.

Adverbien luokka / Ikäluokka %	Koululaiset n = 17 844	Aikuiset n = 11 372
Intensiteetti	7,4	7,5
Kommentti	3,0	6,9
Määrä	9,7	10,3
Paikka	20,4	18,5
Proadverbi	25,0	14,9
Temporaalinen	24,2	27,1
Tapa	10,3	14,9

Verbien, adjektiivien ja adverbien semanttinen analyysi osoittaa selvää siirtymää arvioivampaan suuntaan. Asenteiden ilmaisun ja kommenttien osuus lisääntyy ja toisaalta subjektiivisen arvottamisen osuus vähenee. Kehityksellinen ero näkyy sekä yleisyystasojen että ilmaisujen granulariteetin nousussa. Tapahtumien, arvojen ja ajan runsas ja lisääntyvä kielentäminen puolestaan liittyy kertovaan tekstiin ja kertomusrakenteen kehittymiseen.

## Lopuksi

Unelmasanaston analyysi osoittaa sanojen valintaan liittyvän useanlaisia tekijöitä. Osa on tekstin ominaisuuksia, osa liittyy kirjoittajan tyylivalintoihin ja osa on selvästi kehityksellistä. Aineiston hallitseva tekstityyppi on kertomus ja kasvava osa niistä on omaelämäkertatyypisiä. Kertomuksille ominaisessa sanastossa on paljon tapahtumaverbejä sekä kehystäviä ajan ja paikan adverbeja. Kerronta on usein minämuotoista tai re-

feroivaa; tapahtumia koetaan usein yhdessä perheen tai ystävien kanssa ja kyseiset tapahtumat pyritään evaluoimaan positiivisiksi. Kirjoitelmien aihe Unelmieni päivä tuo paljon päiväskriptisanastoa, ja valitut teemat nostavat tiettyjä tapahtumasarjoja esiin: esimerkiksi häöpäivänä toteutuva unelma koostuu hänille tyypillisestä tapahtumasarjasta pukeutumisesta vihkimiseen, onnitteluihin, hääkakun leikkuuseen, tanssimiseen, kukkapuketin heittoon ja häämatkaan. Näihin eri tapahtumiin liittyvä sanasto ei vielä yksittäisiä sananvalintoja tarkastelemalla tuo esiin sinänsä juuri kertomukselle tyypillisiä piirteitä eikä myöskään paljasta kehityksellisiä trendejä. Kehityksellisiä piirteitä tulee näkyviin erityisesti frekvenssitekijöiden, sanaston diversiteetin ja sen granulariteettiasteen kautta sekä elollisuus- ja konkreettisuusasteikon muutoksista.

Myös kertomusrakenteen hallinta on kehityksellistä. Suomalaislapset oppivat hallitsemaan kertomusskeemaa keskimäärin ala- ja yläkoulun vaihteessa, nuoruudessa kertomusskeema monipuolistuu kuvailuun ja arviointiin. Nuorilla aikuisilla kirjoitelma saattaa olla tyylipuhdas kertomus unohtumattomasta kokemuksesta tai nautinnollisesta päivästä, mutta se voi myös olla tunnelmia kuvaileva ja maalaileva tai unelmia arvioiva – tai vähän kaikkea. Ehkä vähäinen suunnittelu pitää kertomukset edelleen vaikka ajallisesti aiempaa tiiviimpinä kuitenkin melko kronologisesti etenevinä. Vasta keski-ikäisillä kirjoittajilla tämä kronologia näyttää rikkoutuvan, ja kertomus voidaan laatia hyvin monella tapaa, kuten Petälä (2018) osoittaa. Kertovaan tekstiin liittynee myös ainakin osaksi sanaston suhteellinen yksinkertaisuus, suhteellisen suuri johtamattomien ja melko yleisten sanojen määrä.

Tyylilliset valinnat tai strategiat heijastuvat myös sanastoon. Kirjoitelmista voi erottaa kaksi vallitsevaa tyylistrategiaa, joista toista voisi nimittää kuvailemaan pyrkiväksi ja toista pelkistyneeksi. Edellinen tuottaa lapsilla värikkäitä adjektiiveja ja erikoisia juonen käännteitä, vanhemmilla kirjoittajilla melko yksityiskohtaista kuvailua, jälkimmäinen enemmän toteavaa kerrontaa, jossa päättely jää lukijan kontolle. On myös rönsyilevää ja tiivistä kerrontaa. Tyylilliset valinnat peittyvät usein tilastojen ulottumattomiin mutta on syytä muistaa, että osa koehenkilö- ja ikäluokka-kohtaisesta variaatiosta johtuu tyylillisistä valinnoista. Yhtä lailla on kuitenkin syytä panna merkille, että vaikka tyylilliset valinnat vaikuttavat

enemmän aikuis- kuin lapsikerrontaan, tutkimus silti paljastaa selviä iänmukaisia sanastollisia kehitystrendejä.

Koululaiset ovat kirjoitelma-aineiston sana-analyysin perusteella keskimäärin vielä peruskoulun päättyessäkin aika kaukana nuorten aikuisen tasosta. Tämä näkyy erityisen hyvin niin sanotuista peittoluvuista: 2 000–4 000 lekseemiä kattaa koululaisten sanamassasta lähes kaiken mutta aikuisilta vain 90 prosenttia. Koululaiskirjoitelmissa on hyvin yleisiä ja tavallisia lekseemejä merkitsevästi enemmän kuin aikuiskirjoitelmissa; aikuisilla taas harvinaisia lekseemejä on vastaavasti huomattavasti enemmän. Ero tarkoittaa, että aikuiset ilmaisevat täsmällisempiä merkityksiä kuin koululaiset. Sanaesiintymien tasolla massa koostuu molemmilla hyvin yleisistä sanoista, mutta yleisyysero näkyy silti. Aikuisilla myös johdetut sanat ja kiteytymät ovat yleisyydeltään harvinaisempia kuin koululaisilla mutta koululaisilla on enemmän harvinaisia yhdyssanoja. Ero kuitenkin syntyy koululaisten suosimista satunnaisuudosteista, kyse ei ole merkityksen ilmaisun täsmällisyydestä. Taivutusmuotoanalyysi perustui lekseemeille ominaiseen taivutusmuotoprofiiliin sanomalehtikielessä. Kun tätä profiilia verrattiin unelmasanaston yleisyystasoihin, kävi ilmi, että aikuissanasto liittyy vaihtelevampaan syntaksiin kuin koululaissanasto. Tämän voinee tulkita niin, että aikuiset kirjoittavat yleiskielen tapaan ja koululaiset poikkeavat siitä enemmän. Vastaava kehitys tulee esiin sanaston diversiteettianalyysistä. Nuorilla aikuisilla leksikaalinen erilaisuus – eri lekseemien määrä – on selvästi korkeampi kuin yhdeksäsluokkalaisten, jotka taas peittoavat kuudesluokkalaisten. Suurin ero on kolmasluokkalaisten ja kuudesluokkalaisten välillä.

Diversiteettiä mitattiin myös sukupuolten välillä – ainoana tässä tutkimuksessa. Tulos oli, että ainoastaan kolmasluokkalaisten tyttöjen ja poikien diversiteetti eroaa: mediaani on tytöillä korkeampi mutta pojilla keskeisten havaintojen hajonta on suurempi. Eron vähäisyys on odotuksen mukaista, koska muun muassa älykkyyttutkimukset eivät osoita semanttisen tiedon sukupuolittuneisuutta. Se on linjassa myös sanastotestiemme tulosten kanssa. Jos sitten sanastotaidot katsotaan luku- ja kirjoitustaitojen edellytykseksi, tulos on vähintään mielenkiintoinen. Peruskoululaisten aineisto edustaa koko ikäluokkaa, nuorten aikuisten

vain ikäluokan koulutetumpaa osaa. Koululaisten tuloksissa on paljon poikkeavia havaintoja sekä tytöillä että pojilla mutta aikuisryhmästä nämä puuttuvat. Koululaisten aineistosta nouseva suuri poikkeavien havaintojen määrä viittaa erityistuen tarpeeseen. Kaikkein poikkeavin osa koululaisista eli ne, jotka eivät ole saaneet tekstiä sanottavasti aikaiseksi, nostaa tuettavien joukkoa.

Semanttisesti sanastosta näkyy hyvin aiheen käsittelyn yleinen abstraktistuminen, joka tulee esiin erityisesti substantiivisanastosta mutta myös verbisanastosta. Inhimillisistä, elollisista ja konkreettisista elottomista kirjoitetaan ikäryhmittäin aina vähemmän mutta asioista aina enemmän. Tämä tarkoittaa, että vaikka kirjoitelman aihe on kaikilla sama, aihe käsitteellistetään korkeammalle abstraktiotasolle ja siitä kirjoitetaan tiivistetympin. Kun suunnitteluaikaa ei ollut millään kirjoittajaryhmällä, ero on kehityksellinen. Substantiivien semanttisista luokista näkyy selvä siirtyminen konkreettisista objektisanoista abstrakteihin asiasanoihin, mutta muutos tapahtuu myös luokittain siten, että muun muassa objektisanoista kirjoitetaan abstraktiotasoltaan korkeammalla tyyppien tasolla niiden yksittäisten esiintymäesimerkkien sijaan. Sanaston harvinaistuminen ja nimeämisen diversiteetin nousu taas osoittavat granulariteettiasteen nousun kehitykselliseksi. Yhteenvetona voi siis todeta, että sanamerkitysten abstraktistuminen, harvinaistuminen ja täsmentyminen ovat kehityksellisiä.

Lopuksi voi vielä pohtia, miksi abstraktistuminen on kehityksellistä. Eihän abstrakti teksti aina vaikuta edes tavoiteltavalta. Kehityksellisessä kontekstissa eroa voi luonnehtia seuraavasti: Konkreettisten sanojen oppiminen ja muistaminen on helpompaa kuin abstraktisanojen, konkreettismerkityksistä sanoista keksii paremmin assosiaatioita ja tarkoitteet, joihin ne viittaavat, ovat kuviteltavissa ja niistä voi muodostaa mielikuvia. Tämä kaikki tukee näiden sanojen käyttöä ja kontekstin hallintaa. Abstraktit sanat viittaavat kulttuurisiin ja yksilöllisiin konstruktiioihin ja niiden syntaksi opitaan kontekstista – tai analogiasuhteesta konkreettisten sanojen käyttöön –, joten niiden syntaksi ei ole motivoitua (ks. esimerkiksi Barsalou 2008). Abstraktin sanaston oppiminen ja käyttöön ottaminen vaatii siis aikaa. Taitava tekstin rakentaminen edellyttää paitsi yleisyystason nostoa myös kykyä täsmentää ilmaisua.

## Lähteet

- Baayen, H. 2014. Experimental and psycholinguistic approaches. Julkaisussa: R. Lieber & P. Štekauer (toim.) *The Oxford handbook of derivational morphology*. Oxford: Oxford University Press, 95–117.
- Barsalou, L. W. 2008. Grounded cognition. *Annual Review of Psychology* 59, 617–645.
- Bauer, L. & Nation, P. 1993. Word families. *International Journal of Lexicography* 6: 4, 253–279.
- Biber, D., Conrad, S. & Reppen, R. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Brysbaert, M., Warriner, A. B., Kuperman, V. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46, 904–911.
- Bybee, J. & Hopper, P. 2001. Introduction to frequency and the emergence of linguistic structure. Julkaisussa: J. Bybee & P. Hopper (toim.) *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 1–24.
- Carlisle, J. F. 2000. Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing* 12, 169–190.
- Croft, W. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: Chicago University Press.
- CSC = Tieteellinen laskenta, yleisyyskorpus (44 miljoonaa sanaesiintymää). <https://korp.csc.fi/suomen-sanomalehtikielen-taajuussanasto-B9996.txt>. Viitattu 12.12.2014.
- Dixon, R. M. W. 1977. Where have all the adjectives gone? *Studies in Language* 1, 19–80.
- Dixon, R. M. W. 2004. Adjective classes in typological perspective. Julkaisussa: R. M. W. Dixon & A. Y. Aihkenvald (toim.) *Adjective classes: A cross-linguistic typology*. Oxford: Oxford University Press, 1–49.
- Fellbaum, C. (toim.) 1998. WordNet. An electronic lexical database. Cambridge: The MIT Press.
- Frawley, W. 1992. *Linguistic semantics*. Hillsdale: Lawrence Erlbaum Ass.
- Gärdenfors, P. 2014. *The geometry of meaning. Semantics based on conceptual spaces*. Cambridge: The MIT Press.
- Hadley, P. A. & Rispoli, M. 2012. Lifespan perspective on individual differences in grammatical abilities. *Linguistic Approaches to Bilingualism* 2: 3, 269–272.
- Hakulinen, L. 1979. *Suomen kielen rakenne ja kehitys*. Neljäs, korjattu ja lisätty painos. Helsinki: Otava.
- Harjunen, E. & Rautopuro J. 2015. Kielenkäytön ajattelua ja ajattelun kiellentämistä. Äidin-kielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2014: keskiössä kielentuntemus ja kirjoittaminen. Helsinki: Kansallinen koulutuksen arviointikeskus.
- Hart, B. & Risley, T. R. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore: Brooks.
- Hart, B. & Risley, T. R. 2003. The early catastrophe: The 30 million word gap by age 3. *American Educator* 27: 1, 4–9.
- Heinonen, S. 2018. *Narratiivisuuden sisällöllinen kehittyminen aiheanalyysin kautta tarkasteltuna: Kvantitatiivinen tutkimus aiheiden esiintymisen vaihtelusta iän ja sukupuolen mu-*



- kaan "Unelmiäni päivä" -kirjoitelmissa. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Honko, M. 2013. *Alakouluikäisen leksikaalinen tieto ja taito: Toisen sukupolven suomi ja 51-verrokki*. Tampere: Tampereen yliopisto.
- Härkönen, P. 2012. "Jopas taas melkoisen sopan keitin." Verbi-idioomit *Aku Ankassa ja kuinka nuoret niitä ymmärtävät*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Jarvis, S. 2013. Capturing the diversity in lexical diversity. *Language Learning* 63, 87–106.
- Jääskeläinen, Heikki. 2008. *Puhekuplat täyteen. Peruskoululaisen luoman kertomuksen loogisuus ja mielikuvituksellisuus sekä kertomuksessa käytetyn verbisanaston monipuolisuus*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Kauppinen, M., Tarnanen, M. & Aalto, E. 2014. "Voisin pyytää oppilaita alleviivaamaan kaikki adjektiivit" – luokanopettajaopiskelija kielitietoisien aineenoppimisen ohjaajana. Julkaisussa: M. Mutta, P. Lintunen, I. Ivaska & P. Peltonen (toim.) *AFINLA-e*. Soveltavan kielitieteen tutkimuksia 7, 81–100.
- Keskinen, J. 2012. "Jos joskus pitäisi keksiä joku mielikuvitusmaailma, niin aivan varmasti jokaisella ihmisellä olisi ainakin omat unelmat: siitähän jokainen unelmoi." *Modaalisuus alakouluikäisten lasten unelmakirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Keuleers, E., Stevens, M., Mander, P. & Brysbaert, M. 2015. Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology* 8, 1665–1692.
- Koivunen, J. 2012. *Adverbien käyttö alakoululaisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Kuisma, J. 2011. *Ominaisuuden ilmausten käytön kehittyminen alakouluikäisten kirjoitelmissa*. Pro gradu -tutkielma. Tampereen yliopisto: Suomen kieli.
- Kukkonen, P. & Pajunen, A. 1986. Rektio ja agrammatismi. *Viritäjä* 90: 22–45.
- Kusnetsoff, T. 2017. *Suomenkielisten nuorten morfologinen tietoisuus ja sen yhteys luetun ymmärtämiseen*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Kuusela, J. 2011. Kun kirjoittaminen ei suju. Julkaisussa: E. Harjunen, R. Juvonen, J. Kuusela, B. Silén, M. Sääkslahti & M. Örnmark (toim.) *Miten peruskoululaiset kirjoittavat? Näkökulmia ja kysymyksiä. Perusopetuksen 9. luokan äidinkielen ja kirjallisuuden oppimistulosten seuranta-arvioinnin aineistoa 2010*. Raportit ja selvitykset 2011: 2, 13–22. [www.oph.fi/julkaisut](http://www.oph.fi/julkaisut). Viitattu 22.4.2020.
- Kyösti, J. 2018. *Motivoinnin vaikutuksia kuudesluokkalaisten kirjoittamiin kertomuksiin*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Laasanen, M., Pajunen, A. & Häikiö, T. Arvioitavana. Kielen natiivihallinnan variaatio ja sosiolingvistinen meneillään olevan kielenmuutoksen tutkimus.
- Laine, M. & Virtanen, P. 1999. *WordMill. Lexical Search Program*. Center for Cognitive Neuroscience. Turku: University of Turku.
- Laine-Leinonen, J. 2013. *Koulukorpuksen leksikko: 1.–6.-luokkalaisten aktiivinen sanavarasto*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Laufer, B. & Nation, P. 1999. A vocabulary size test of controlled productive ability. *Language Testing* 16: 1, 33–51.

- Laufer, B. & Ravenhorst-Kalovski, G. C. 2010. Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language* 22: 1, 15–30.
- MacWhinney, B. 2001. Emergentist approaches to language. Julkaisussa: J. Bybee & P. Hopper (toim.) *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 449–470.
- Malvern, D. D., Richards, B. J., Chipere, N. & Durán, P. (toim.) 2009 [2004]. *Lexical diversity and language development. Quantification and assessment*. New York: Palgrave Macmillan.
- McCarthy, M. & Scott, J. 2007. Vocd: A theoretical and empirical evaluation. *Language Testing* 24: 4, 459–488.
- McNeil, N. 2007. U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology* 43, 687–695.
- Michaels, S. 2013. Déjà Vu all over again: What's wrong with Hart & Risley and a "linguistic deficit" framework in early childhood education? *Learning Landscapes* 1, 23–41.
- Miller, G. A. 1998. Nouns in WordNet. Julkaisussa: C. Fellbaum (toim.) *WordNet. An electronic lexical database*. Cambridge: The MIT Press, 23–46.
- Miller, G. A. & Fellbaum, C. 1991. Semantic networks in English. *Cognition* 41, 198–229.
- Milton, J. 2009. *Measuring second language vocabulary acquisition*. Bristol: The Multilingual Matters.
- Nahkola, T. 2012. *Yhdyslauseiden kehitys kouluikäisen kielessä: Tutkimus 1.–6.-luokkalaisten kirjoitelmista*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Nation, I. S. P. & Coxhead, A. 2021. *Measuring native-speaker vocabulary size*. Amsterdam: John Benjamins.
- Nippold, M. A. 2006 [3. painos]. *Later language development: School-age children, adolescents and young adults*. Austin: Pro-Ed.
- Olinghouse, N. G. & Wilson, J. 2013. The relationship between vocabulary and writing quality in three genres. *Reading and Writing* 26, 45–65.
- OPH 2014. Perusopetuksen opetussuunnitelman perusteet. Helsinki: Opetushallitus.
- Qian, D. D. 2002. Investigating the relationship between vocabulary knowledge and academic reading comprehension: An assessment perspective. *Language Learning* 52: 3, 513–536.
- Pajunen, A. 1994. Adjektiivikategorian universaaliudesta. *Virittäjä* 98: 4, 513–542.
- Pajunen, A. 1998. Adjectives in spoken language discourse. *Word* 49: 3, 341–368.
- Pajunen, A. 1999. *Suomen verbirektiosta*. Turku: Turun yliopisto.
- Pajunen, A. 2001. *Argumenttirakenne. Asiantilojen luokitus ja verbien käyttäytyminen suomen kielessä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Pajunen, A. 2006. Verbisanasto uusiutuu. *Puhe ja kieli* 26: 4, 205–219.
- Pajunen, A. 2010. Sanojen synteettisyysasteesta suomen kielessä. *Virittäjä* 114: 4, 481–501.
- Pajunen, A. 2012. Kirjoittamistaitojen kehitys 8–12-vuotiailla: Alakoululaisten unelmakirjoitelmät. *Virittäjä* 116: 1, 3–34.
- Pajunen, A. 2016. Sukupuolierot kielen hallinnassa (etenkin kirjoittamisessa). Esitelmä. Språkets funktion 6.–7.6.2016, Åbo Akademi.
- Pajunen, A. 2017. Tapahtumien nimeämisen normitus ja variaatio. Esitelmä. MYK 15.12.2017. Turun yliopisto.

- Pajunen, A. & Virtanen, P. 2002. Helsingin Sanomat 2000–2001 [= HS2000-korpus]. Morfosyntaktisesti analysoitu tietokanta, 24 miljoonaa sananmuotoa.
- Pajunen, A. Puranen, M. & Yli-Paavola, A. 2010. Koulukorpus. Alakoululaisten kirjoitelmia aiheesta Unelmieni päivä. Excel-tietokanta. Tekijöiden hallussa.
- Pajunen, A. Itkonen, E., Vainio, S. 2015. Sanamerkityksen hallinta nuorilla aikuisilla. *Virittäjä* 119: 2, 160–187.
- Pajunen, A. Itkonen, E. & Vainio, S. 2016. Nuorten aikuisten kyky määritellä sanoja. *Virittäjä* 120: 4, 477–515.
- Pajunen, A. & Itkonen, E. 2019. Intuition and beyond: A hierarchy of descriptive methods. Julkaisussa: A. Mäkilähde, V. Leppänen & E. Itkonen (toim.) *Norms and normativity in language and in linguistics*. Amsterdam: John Benjamins, 213–234.
- Palonen, O. 2013. *Pronominit 8–12-vuotiaiden kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Pollock, L. 2018. Statistical and methodological problems with concreteness and other semantic variables. A list of memory experiment case study. *Behavioral Research Methods* 50: 3, 1198–2016.
- Rantala, J. 2012. *Ihmisen olotilaa kuvaavien ilmausten sanastonhallintataidot yläkoulussa ja lukiossa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Ravid, D. 2006. Semantic development in textual contexts during the school years: Noun scale analyses. *Journal of Child Language* 33, 791–821.
- Rosen, G. 2017. *Abstract objects*. Stanford: Center for the study of language and information. <https://plato.stanford.edu/entries/abstract-objects>. Viitattu 1.3.2018.
- Routama, K. 2008. *Peruskoulun 2. ja 3.-luokkalaisten spatiaalisen sanaston hallinta. Konkreetit ja metaforiset ilmaukset*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Saarela, L. 1997. *Peruskoululaisten kirjoitelmien kehittyminen sanastotutkimuksen valossa*. Oulu: Oulun yliopisto.
- Schmitt, N. 2010. *Researching vocabulary. A vocabulary research manual*. London: Palgrave Macmillan.
- Stavroula-Thaleia, K., Vigliocco, G., Del Campo, E., Vinson, D. P. & Andrews, M. 2011. The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General* 140: 1, 14–34.
- Söderholm, C., Häyry, E., Laine, M. & Karrasch, M. 2013. Valence and arousal ratings for 420 Finnish nouns by age and gender. *Plos One* 8: 8, e72859.
- Taylor, J. R. 1995 [1989]. *Linguistic categorization. Prototypes in linguistic theory*. Oxford: Oxford University Press.
- Tomasello, M. 2003. *Constructing a language. A usage-based theory of language acquisition*. Cambridge: Harvard University Press.
- Treffers-Daller, J. & Milton, J. 2013. Vocabulary size revisited: The link between vocabulary size and academic achievement. *Applied Linguistic Review* 4: 1, 151–172.
- Turunen, R. 2012. *Viides- ja seitsemäsluokkalaisten abstraktien substantiivien ja verbien hallinta*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Vainio, S., Pajunen, A. & Häikiö, T. 2019. Acquisition of Finnish derivational morphology: School-age children and young adults. *First language* 39: 2, 139–157.

- Wagner, R. K., Muse, A. E. & Tannenbaum, K. R. (toim.) 2007. *Vocabulary acquisition: Implications for reading comprehension*. New York: The Guilford Press.
- Wetzel, L. 2009. *Types and tokens: On abstract objects*. Cambridge: The MIT Press.
- Vossen, P. (toim.) 2002. EuroWordNet. General document. <http://www.hum.uva.nl/~ewn>. Viitattu 1.3.2018.
- Yap, M. J., Lim, G. Y. & Pexman, P. M. 2015. Semantic richness effects in lexical decision: The role of feedback. *Memory & Cognition* 43: 1148–1167.

# Liiketilan väylän ja suunnan kielentäminen suomessa

*Emilia Tuuri*

 <https://orcid.org/0000-0002-8298-6875>

## Johdanto

Kielenhallinnassa voidaan hahmottaa kahtalaisia muuttujia: kehityksellisesti iän mukana muuttuvia ja toisaalta sellaisia, jotka muuttuvat kehityksellisesti mutta myös varioivat ikäluokkatasolla jatkuvasti (tarkemmin Pajunen & Honko, johdantoluku). Jälkimmäisen tyyppin muuttujissa variaatioon voi liittyä kompleksisuutta kielennettävän ilmiön ja kielen resurssien tasolla sekä eroja esimerkiksi yksilöpsykologisissa tekijöissä. Yksi merkityskenttä, jolla ilmaisu on luonnostaan varioivaa, on liikkeen ilmaisu. Liike on keskeinen osa ihmisen kokemusmaailmaa, ja ihmiselle intentionaalisen olentona erityisen tärkeänä näyttäytyy kokemus liikkeestä suuntaisena, etenevänä, paikasta toiseen vievänä. Paikanmuutokseen johtavan liikkeen ilmaisun on esitetty jäsentyvän kielissä kahtalaisesti sen mukaan, millä elementeillä niissä tyyppillisimmin ilmaistaan liikkumisen väylä (esimerkiksi Talmy 2000b). Tätä ensisijaisiin ilmaisu-tapoihin perustuvaa, vahvasti yleistävää typologiaa on sittemmin haastettu kuvaamalla liikkeen kielentämiseen liittyvää laajaa kieltenvälistä variaatiota (esimerkiksi Ibarretxe-Antuñano 2009; Goschler & Stefano-

witsch 2013) ja osoittamalla yksittäisiä kieliä typologiaan sopimattomiksi (esimerkiksi thai, ks. Slobin 2004; Zlatev & Yangklang 2004).

Suhteellisen vähän on sen sijaan pohdittu sitä, missä määrin yksittäis-kielen sisällä on variaatiota liiketilanteiden<sup>22</sup> ilmaisussa (ks. kuitenkin esimerkiksi Croftin ja muiden [2010] konstruktioihin perustuva typologia). Yhtä vähän on tutkittu sitä, miten tämä kompleksisen tilanteen kielentäminen kehityksellisesti opitaan ja hallitaan. Koska edellä esitetty kehityksellisten muuttujien kahtiajako ei ole yksiselitteinen vaan muut-tajat voivat jossakin määrin edustaa molempia luokkia, ennen kehityksellistä tarkastelua muuttujaa on tutkittava sellaisella ikäluokkatasolla, jolla hallinnan voidaan olettaa olevan melko ongelmatonta eli aikuis-tasolla, ja näin kartoitettava kielennettävään ilmiöön liittyvää variaatiota.

Spatiaalisten merkitysten jäsentyminen kielelliseksi ilmaisuksi on viimeaikaisten tutkimusten valossa monen tekijän summa. Esimerkiksi spatiaalisten viitekehysten tutkimuksessa on todettu, että siihen vaikuttavat ainakin kieli ilmaisuvarantoineen, fyysinen ympäristö sekä sosio-kulttuuriset tekijät (ks. esimerkiksi Dasen & Mishra 2010; Bohnemeyer ja muut 2014; Palmer, Lum, Schlossberg & Gaby 2017). Kielen vaihtelevat resurssit mahdollistavat yksilöiden erilaiset valinnat ja niistä syntyvän variaation, mutta yksilöiden välillä on myös huomattavia eroja spatiaalisen kognition tarkkuudessa (ks. esimerkiksi Wolbers & Hegarty 2010; Meneghetti, Pazzaglia & De Beni 2011), mikä voi näkyä spatiaali-ssa kielenkäytössä.

Suomen kielen liikkeen ilmaisua on tarkasteltu aiemmin lähinnä yksit-täisistä muotokeinoista käsin (esimerkiksi Leino, Helasvuo, Lauerma, Nikanne & Onikki 1990; Sivonen 2005; Ojutkangas 2003; 2012; Huumo 2010; Lehismets 2014).<sup>23</sup> Tässä tutkimuksessa näkökulma on ilmiö-lähtöinen: aineisto on rajattu visuaalisella elisitaatiokehyksellä ja analy-soitu Zlatevin (esimerkiksi 2003; 2007) *holistic spatial semantics* -teoria-

22 Käytän tässä yhteydessä *liiketapahtuman* (*motion event*, esimerkiksi Talmy 1985; 2000) sijaan käsitettä *liiketilanne* (*motion situation*, ks. Zlatev, David & Blomberg 2010, 393), koska se viittaa luontevammin sekä rajattuihin tapahtumiin että rajaamattomiin aktiviteetteihin (rajattuudesta tarkemmin väylän ja suunnan yhteydessä).

23 Spatiaalisten suhteiden ilmaisuun liittyvä tutkimus on ollut etenkin kognitiivisesti suuntautuneessa fennistiikassa runsasta, joten en pyri tässä kattavaan esitykseen vaan mainitsemaan joitakin esimer-kejä.

kehyyksen eli kokonaisvaltaisten semanttisten kategorioiden avulla, joista keskitytään erityisesti liiketilanteen väylään ja suuntaan.<sup>24</sup> Lähtökohtaisena ajatuksena on, että kielissä on keinot tiettyjen perustavanlaatuisien spatiaalisten kategorioiden kielentämiseen mutta ilmaisutavat ja ilmaisuuden laajuus vaihtelevat kielten resurssien mukaan.

Tässä yhteydessä tarkasteltavat kielennökset ovat liiketilanteiden kuvauksia, joihin sisältyy translokaatio eli muuttujan paikanmuutos suhteessa spatiaaliseen viitekehyykseen, kuten kiintopisteeseen tai puhujan näkökulmaan (Zlatev ja muut 2010; tässä seuraava luku). Muuttuja on entiteetti, jonka sijainti tai liike on spatiaalisessa ilmauksessa huomion kohteena, ja kiintopiste on entiteetti, johon muuttujan sijainti tai liike suhteutuu (esimerkiksi Talmy 1975).<sup>25</sup> Muuttujan ja kiintopisteen suhde on liikeilmauksessa dynaaminen, ja jos liikettä ei ilmaista, kyseessä on staattinen lokaatio. Tässä tarkasteltavana on omavoimainen liike, joka tapahtuu ilman ulkopuolista aiheuttajaa (esimerkiksi *mennä* vs. *viedä*).

Tavoitteena on tarkastella spatiaalisesti rajatun väylän ja spatiaalisesti rajaamattoman suunnan (Zlatev 2007; ks. myös Miller & Johnson-Laird 1976) ilmaisukeinoja ja niiden käyttöä liikeilmauksissa. Väylän osalta tarkastelen ensisijaisesti sitä, millaisilla kielenyksiköillä suomen puhujat ilmaisevat väylän KESKIVAIHETTA (esimerkiksi *metsän läpi*, *tien poikki*). Lisäksi tarkastelen, millaisiin elisitaatiokehyksessä esitettyihin liiketilanteisiin erityisesti liittyy kielenkäyttäjien välistä variaatiota rajattuuden suhteen ja millaisista kielenkäyttäjät ovat yksimielisiä. Kielensisäisen variaation näkökulmien rinnalla kulkee typologinen näkökulma: suhteutan tulosten perusteella suomea tarkasteltujen merkityskategorioiden osalta Talmy'n (2000b) typologiaan sekä muihin liikkeen ilmaisun tyologisessa tarkastelussa esitettyihin yleistyksiin.

24 Zlatev (2007) käyttää semanttisten kategorioiden merkintätapana isoa alkukirjainta. Tässä yhteydessä käytetään kuitenkin suomen konventioiden mukaisesti pientä alkukirjainta kategorioiden nimissä: muuttuja (*Figure*), kiintopiste (*Landmark*), liike (*Motion*), väylä (*Path*), suunta (*Direction*), viitekehys (*Frame of Reference*), ala (*Region*) ja tapa (*Manner*). Kategorioiden arvot, kuten väylän ALKU, KESKIVAIHE ja LOPPU merkitään kapiteeleilla.

25 Hahmopsykologiaan pohjaavat käsitteet *Figure* ja *Ground* on kielitieteessä ottanut käyttöön Talmy (1975). Vastaavista merkityskategorioista on käytetty myös käsitteitä *referent* ja *relatum* (Miller & Johnson-Laird 1976) sekä *trajectory* ja *landmark* (Langacker 1987). Suomessa vakiintuneimmat termit ovat *muuttuja* ja *kiintopiste*.

Elisitaatiokehystä käytettäessä kiinnostuksen kohteena ovat tyypillisesti sellaiset kielen piirteet, joista ei voi saada tietoa suoraan kieliopeista tai sanakirjoista, ja pyrkimyksenä on yleensä edustavan aineiston saaminen kieltenväliseen vertailuun. On kuitenkin myös tarpeen tarkastella, kuinka paljon variaatiota yksittäisen kielen aineisto kätkee sisäänsä ja millaiset tekijät sitä aiheuttavat. Riittävän laajalla koehenkilöjoukolla elisitaatiokehys tuo esiin myös variaation siinä, millaisia resursseja yksittäiskielen puhujayksilöt valitsevat samojen tilanteiden kuvaamiseen. Tämän tutkimuksen aineisto on kerätty 50:ltä suomea äidinkielenään puhuvalta koehenkilöltä, joista valtaosa on nuoria aikuisia. Kansainvälisessä liikkeen typologiaa kartoittavassa hankkeessa<sup>26</sup> on kerätty myös kehityksellistä aineistoa, ja vastaavasti sitä kerätään myöhemmän kielenkehityksen hankkeessa. Siten tässä esitellyt tulokset nuorten aikuisten käyttämistä liikkeen kielentämisen keinoista mahdollistavat myöhemmän kehityksellisen näkökulman. Alustava vertailu Unelmakirjoitelma-aineistoon (ks. aineistosta Pajusen ja Hongon johdantolukua) osoittaa, että vaihtelu lasten ja nuorten aikuisten välillä on odotettavaa myös spatiaalisten kielenainesten, kuten väylän ja suunnan ilmaisuun käytettyjen grammien<sup>27</sup> osalta.

Tarkastelu etenee seuraavasti: Seuraavassa luvussa avaan väylän ja suunnan käsitteitä ja niiden suhdetta spatiaaliseen rajattuuteen sekä esittelen rajanylityksen (*boundary-crossing*) käsitteen ja siihen liittyvät tyypologiset yleistyksiset. Kolmannessa luvussa esittelen käytetyn elisitaatiokehityksen ja kuvailen sillä koottua suomen kielen aineistoa. Neljäs luku keskittyy väylän KESKIVAIHEEN ja viides suunnan ilmaisukeinoihin aineistossa. Kuudennessa luvussa tarkastelen tapaustutkimuksen omaisesti väylän KESKIVAIHEESEEN ja suuntaan liittyvää variaatiota sellaisten stimulusten kielennöksissä, joissa kiintopiste on ihmiselle tyypillinen kulkureitti. Loppuluku kokoaa tarkastelun keskeiset tulokset.

26 Zlatev, Jordan ja muut: Phenomenology and Typology of Motion (PATOM).

27 Spatiaaliset kielenainekset ovat tyypillisesti syntaktisesti monifunktioisia: sama yksikkö saattaa toimia kontekstista riippuen esimerkiksi adpositiiona tai adverbina (*luolan sisään* vs. *sisään luolaan*), pre- tai postpositiiona (*pitkin polkua* vs. *polkua pitkin*). Myöskään rajanveto adpositioiden ja relationaalisten substantiivien välillä ei ole aina mielekästä. Käytänkin spatiaalisista kieliopillisista sanoista yleisnimitystä *grammi* (ks. esimerkiksi Svorou 1994; fennistiikassa esimerkiksi Ojutangas 2003; 2012).



## Väylä ja suunta

Liikkeen kielentämisen kuvauksessa väylän (*Path*) käsite on ollut tärkeä lähtökohta. Väylä ilmaisee liikerataa, jolla liike etenee (Talmy 2000b). Etenevä liike suuntautuu aina jonnekin, ja yleensä on oletettu, että suunta sisältyy väylään. Holistinen spatiaalinen semantiikka kuitenkin erottaa väylän ja suunnan (Zlatev 2003) ja perustaa eron liikeradan rajattuuteen. Käsittelen tätä erottelua yksityiskohtaisemmin.

### VÄYLÄ JA SUUNTA RAJATUN JA RAJAAMATTOMAN LIIKERADAN ILMENTYMINÄ

Talmy (esimerkiksi 1985; 2000b) mukaan keskeinen eronteko maailman kielten välillä on se, millä kielenyksiköllä väylä niissä tyypillisimmin ilmaistaan. Talmy (2000b) nimittää verbikehyskieliksi (*verb-framed languages*) niitä, jotka tyypillisesti ilmaisevat väylän pääverbissä (esimerkiksi romaaniset ja seemiläiset kielet) ja satelliittikehyskieliksi (*satellite-framed languages*) niitä, joissa väylän ilmaisemiseen käytetään useimmiten verbiin yhteydessä olevaa niin sanottua satelliittia, kuten verbiin liittyvää partikkelia tai verbiprefiksiä (esimerkiksi germaaniset ja slaavilaiset kielet). Satelliittikehyskielissä liikeverbi ilmaisee tyypillisesti tavun (esimerkiksi *kävellä*, *juosta*). Suomalais-ugrilaiset kielet edustavat Talmy (2000b, 60) mukaan satelliittikehyskieliä, mutta kattavaa empiiristä kartoitusta suomen ilmaisukeinoista ei ole toistaiseksi tehty. Lisäksi satelliitin käsitteen soveltuvuutta sijakieliin on kyseenalaistettu (ks. esimerkiksi Imbert, Grinewald & Sörös 2011; Naidu, Zlatev, Duggirala, Van De Weijer, Devylder & Blomberg 2018).

Talmy (2000b) teoriassa väylä (*Path*) on jokaisen spatiaalista siirtymää ilmaisevan lauseen ydinskeema (*core schema*). Talmy väylän käsitettä on kritisoiu siitä, että se on liian laaja ja häivyttää liiketilanteiden keskeisiä erontekoja, kuten liikkeen rajattuuden (*boundedness*) ja rajaamattomuuden (*unboundedness*) (ks. esimerkiksi Zlatev ja muut 2010; Blomberg 2014, 67–68, 74–76). Zlatev (esimerkiksi 2003; Zlatev ja muut 2010) onkin erottanut väylän (*Path*) ja suunnan (*Direction*) käsitteet jäsentääkseen rajattuuden ja rajaamattomuuden vaihtelua. Väylä ja suunta ovat liikeradan ilmentymiä, jotka eroavat toisistaan rajattuu-

den kannalta.<sup>28</sup> Skemaattisesti määritelty väylä koostuu ALUSTA, KESKIVAIHEESTA ja LOPUSTA (esimerkki 1), jotka kielissä tyypillisesti erotetaan omilla keinoillaan ilmaistaviksi vaiheiksi ja joista voidaan ilmaista kerrallaan yksi tai useampia. Rajatussa liiketilanteessa muuttuja käy läpi tilanmuutoksen suhteessa yhteen tai useampaan väylän vaiheeseen. (Zlatev ja muut 2010, 395–396.) Rajattu liiketilanne vastaa Vendlerin (1967) teelistä eli päätepisteeseen johtavaa suoritusta (*accomplishment*). Sen sijaan rajaamattomassa liiketilanteessa ei implikoida tilanmuutosta vaan liike voisi periaatteessa jatkua loputtomiin (vaikka käytäntö lopulta asettaakin rajansa), ja tällöin on pikemminkin kyse liikkeen suunnasta (esimerkki 2) (Zlatev ja muut 2010, 395–396).

- 1) Väylän vaiheet: *Tyttö kävelee puistosta* [ALKU] *tien yli* [KESKIVAIHE] *kirjastolle* [LOPPU].<sup>29</sup>
- 2) Suunta: *Tyttö kävelee tieltä päin pihaa pitkin kohti kirjastoa.*

Esimerkkien 1 ja 2 liiketilanteet ovat translokatiivisia, eli ne ilmaisevat siirtymän suhteessa spatiaaliseen viitekehykseen (*Frame of Reference*): muuttuja liikkuu suhteessa erilaisiin kiintopisteisiin (objektikeskeinen, ks. esimerkiksi 1 ja 2), tilanteen havainnoijan tai muun osallistujan näkökulmaan (näkökulmakeskeinen, esimerkiksi *Tyttö tulee puistosta*) ja/tai elinympäristön määrittämiin kiinteisiin suuntiin (geosentrinen, esimerkiksi *Tyttö kävelee pohjoiseen/ylöspäin*) (Zlatev ja muut 2010, 394–395; Tuuri, tulossa). Esimerkin 3 kaltaiset lokaation ilmaukset eivät liikeverbistä huolimatta konstruoinnin kannalta sisällä translokaatiota, eli paikanmuutosta ei kiennetä. Vendlerin (1967) luokituksessa tällaiset ilmaukset edustavat rajaamattomien liiketilanteiden tapaan aktiviteettia (*activity*). Erona rajaamattomiin liiketilanteisiin on se, että suuntaa ei ilmaista (jatkuvuus on kuitenkin merkityksestä pääteltävissä). Zlatev, David & Blomberg (2010) katsovat lokaation ilmausten edustavan väylän nolla-arvoa, PAIKKAA.

28 Liikeilmausten rajattuuden analyysissä keskitytään spatiaaliseen ulottuvuuteen, mutta sitä on vaikea täysin erottaa liikkeeseen väistämättä liittyvästä ajallisesta ulottuvuudesta ja aspektista, jonka kuvauksessa rajattuuden käsitettä on pääasiassa käytetty (ks. esimerkiksi Heinämäki 1984; Nurminen 2015).

29 Esimerkit, joista puuttuu aineistoviittaus, ovat itse koostettuja.

3) *Tyttö kävelee puistossa.*

Liikkeen tiettyyn lokaatioon sijoittavat PAIKKA-ilmaukset muodostavat rajattuuden kannalta erikoistapauksen: liikkeen voi nähdä rajattuna siinä mielessä, että se tapahtuu kokonaan kiintopisteen muodostaman alueen sisällä. Kiintopisteen sisällä kuljettu liikerata jää kuitenkin rajoiltaan avoimeksi. Esimerkin 4 lauseista 4a jättää liikeradan avoimimmaksi siinä mielessä, että periaatteessa se ei kerro mitään miehen etenemistavasta sillan muodostaman kiintopisteen rajojen sisäpuolella, vaikka maailmantiedon perusteella oletus onkin, että mies kulkee sillan kulkusuunnan mukaisesti eikä esimerkiksi askella edestakaisin sillan kaiteelta toiselle. Lause 4b ilmaisee rajatun väylän ja määrittelee sen ulottuvan sillan alusta loppuun, kun taas 4c jättää liikeradan rajaamattomaksi mutta eksplikoii kuitenkin liikkeen sillan kulkusuunnan mukaiseksi eli ilmaisee suunnan.

4a) PAIKKA: *Mies kävelee sillalla.*

4b) Rajattu: *Mies kävelee sillan yli.*

4c) Rajaamaton: *Mies kävelee siltaa pitkin.*

Keskeinen teoreettinen lähtökohta tässä luvussa esittämässäni rajattuuden analyysissä on spatiaalisen merkityksen holistisuus (esimerkiksi Zlatev 2003): spatiaalisen siirtymän ilmaisu ei asemoidu lauseissa useinkaan yksittäisen kielenyksikön ilmaistavaksi vaan jakaantuu eri kielenyksiköiden, niin kieliopillisten kuin leksikaalistenkin, ilmaistavaksi. Tähän liittyy läheisesti semanttisen distribuution (Sinha & Kuteva 1995) käsite. Sillä viitataan yhden merkityksen toistumiseen useassa muodossa (esimerkiksi *sisään taloon*) ja spatiaalisen merkityksen jakautumiseen kieliopillisten ja leksikaalisten elementtien kesken (esimerkiksi spatiaalisen relaation erilaisuus saman sijamuodon sisältävissä lauseissa *Taulu on seinällä* ja *Taulu on pöydällä*, siis kiintopisteen vaikutus merkityksen muodostumisessa).

Hyödynnän semanttisessa analyysissä myös semanttisen granulariteetin eli yksityiskohtaisuuden käsitettä. Sovellan van Stadenin ja Narasimhanin (2012) erottamia granulariteetin tasoja ja tarkastelen ensinnäkin,

miten videolla näkyvä tilanne segmentoidaan (mitä videolla esiintyvistä elementeistä kielennetään) ja toisekseen, millaisin keinoin ja kuinka spesifisti spatiaalisia merkityksiä ilmaistaan. Kehityksellisesti sekä holistisuus että granulariteetti ovat oletettavasti haasteellisia, koska ne edellyttävät objektien ominaisuuksien ja sanojen merkityksen tarkkaa hallintaa.

#### RAJANYLITYS

Rajatuista liiketilanteista voidaan erottaa alatyypiksi ne, joissa ylitetään spatiaalinen raja. Liikkeen kielentämistapojen kieltenvälisen variaation tarkastelun kannalta keskeinen on ollut Asken (1989) havainto, että Talmyn leksikalisointityyppeihin erottamat kielet eroavat toisistaan ensisijaisesti teelisten eli päätepisteen tai tuloksen ilmaisevien lauseiden osalta. Slobin ja Hoiting (1994) muotoilevat saman periaatteen siten, että niin sanottu väyläverbi tarvitaan verbikehyskielissä nimenomaan ilmauksiin, joissa ylitetään spatiaalinen raja. Raja ymmärretään laajasti: se kattaa sekä siirtymät ulos rajatusta tilasta (väylän ALUSSA, esimerkki 5a) tai rajatun tilan sisään (väylän LOPUSSA, esimerkki 5b) että linjamaisen rajan ylityksen tai laajemman alueen poikki kulkemisen (väylän KESKIVAIHEESSA, esimerkki 5c).<sup>30</sup>

5a) Hän astuu **ulos talosta**.

5b) Bussi ajaa **tunneliin**.

5c) Koira loikkaa **aidan yli/juoksee pihan poikki**.

Esimerkiksi verbikehyskieleksi määritellyssä ranskassa taloon juokseminen ilmaistaan pääsääntöisesti väyläverbillä, kuten *entrer* 'mennä sisään', ja tapa voidaan ilmaista infiniittisellä rakenteella, jos se katsotaan tarpeelliseksi (*L'homme est entré dans la maison [en courant]*. 'Mies meni [juosten] sisään taloon.'). kun taas talolle asti juoksemisen, johon ei sisälly rajanylitystä, voi ilmaista tapaverbillä (*L'homme a couru jusqu'à la maison*. 'Mies juoksi talolle saakka.'). (Slobin & Hoiting 1994, 496).

30 Grammit *pois* ja *ulos* ovat muodoltaan tulosijaisia, mutta niiden merkityksessä keskeistä ei ole päätepisteen ilmaisu vaan katsotaan niiden korostavan siirtymää väylän ALUSSA. Päätepiste eli väylän LOPPU jää avoimeksi, ja se voidaan ilmaista erikseen, kuten lauseessa *Poika tulee ulos luolasta merenrantaan*, jossa *ulos luolasta* on väylän ALUN ilmaus ja *merenrantaan* väylän LOPUN.

Tämä verbikehyskieliin liittyvä ilmiö tunnetaan kirjallisuudessa nykyisin nimellä *boundary-crossing constraint*, jonka suomennan tässä *rajanylitysrajoitteeksi*. Slobinin (2004, 226–227) mukaan verbikehyskielissä on taipumus ilmaista nimenomaan verbillä teelistä tilanmuutosta, jollainen rajanylitys keskeisesti on. Tapaverbien ilmaisema liike edustaa pikemminkin aktiviteettia (Vendler 1967, 107), eivätkä tapaverbit siten kovin luontevasti sovellu näissä kielissä rajanylityksen ilmaisuun. Satelliittikehyskielissä, joissa rajanylitys tulee tyypillisesti ilmaistuksi muualla kuin pääverbissä, mainitut tekijät yhdistyvät ongelmattomasti.

Viimeaikaiset tutkimukset viittaavat kuitenkin siihen, että ilmiössä on kyse pikemminkin tendensseistä kuin varsinaisesta rajoitteesta. Tyypillisissä verbikehyskielissä tapaverbien käyttöä on havaittu liittyvän erityisesti nopeisiin, yhtäkkiisiin siirtymiin, kuten veteen hyppäämiseen, ja toisaalta väylän KESKIVAIHEESSA tapahtuviin ylityksiin (ks. esimerkiksi Özçalışkan 2015 turkista). Kopeckan (2009) mukaan ranskassa tapaverbien ja lokatiivisten adpositioiden yhdistelmiä esiintyy rajanylityskonteksteissa enemmän kuin on oletettu ja ilmiöön liittyy monitahoista tilanteista vaihtelua. Rajanylitysilmausten tarkempi analyysi yksittäis-kielissä on siis tarpeen.

Suomen kielessä spatiaalisten rajojen ylittämisen ilmauksiin ei liity rajoitteita verbikehyskielten tapaan, mutta suomessa on käytössä sekä satelliitti- että verbikehyskielille tyypillisiä strategioita rajanylitysten ilmaisuun ja rajanylitysten ilmaisu tuottaa aineistoon variaatiota, kuten myöhemmin tulee tarkemmin ilmi.

## Elisitaatiokehysellä koottu aineisto

Strukturoidun elisitaation perustana on elisitaatiokehys (*elicitation grid*), sarja ei-kielellisiä stimuluksia, joilla tähdätään tietyn aihealueen ilmaisu-tapojen ja niissä esiintyvien hienovaraistenkin erojen tavoittamiseen (ks. esimerkiksi Levinson ja muut 2003). Metodin ilmeisin hyöty on se, että samojen stimulusten käyttö mahdollistaa variaation systemaattisen tarkastelun niin kielten kuin yksilöidenkin välillä. Stimulusten avulla rajamalla pyritään saamaan mahdollisimman yhteismitallista ja vertailu-

kelpoista aineistoa. Luonnollisesti elisitaatiokehysten on oltava huolellisesti suunniteltuja, koska metodilla saadaan kielestä esiin juuri niitä piirteitä, joihin kehys ohjaa. On varmistuttava rajausten mielekkyydestä ja mahdollisesti täydennettävä tarkastelua esimerkiksi korpusaineistoilla. Huomioitava on myös koehenkilömäärän vaikutus: kielelle tyypillistä ilmaisutapaa voi tuskin päätellä muutaman koehenkilön aineistosta. Elisitaatio vertautuu yksilöllisiä reaktioita tuottavaan projektiiviseen testiin, joka mittaa normittamatonta, tyypillistä käyttäytymistä laatuasteikolla ja jossa valmista vertailukohtaa ei ole olemassa tai maksimaalista suoritusta määritelty.

Liikkeen leksikalisoinnin kieltenvälisessä tutkimuksessa on laajasti käytetty kuvien selostamiseen perustuvaa elisitaatiota; tunnetuimpana esimerkkinä ovat *Frog, where are you?* -kuvakertomuksella (Mayer 1969) koottuihin aineistoihin perustuvat tutkimukset (ks. esimerkiksi Berman & Slobin 1994; Strömqvist & Verhoeven 2004). Liikeilmausten aineistonkeruuseen video on kuitenkin havainnollisuudessaan luontevampi formaatti. Tässä tutkimuksessa liikeilmausten aineistonkeruuseen käytettiin *Trajectoire*-videosarjaa (Ishibashi, Kopecka & Vuillermet 2006), jolla on kerätty aineistoa kymmenistä eri kielistä. *Trajectoire* koostuu 76 videoklipistä, joista 55 esittää varsinaisen liiketilanteen, kaksi on tehtävänannon yhteydessä katsottuja harjoitusvideoita ja loput ovat ”fillereitä” eli mukana tutkimuksen tarkoituksen hämärtämiseksi. Videot ovat 8–14 sekuntia pitkiä, ja niissä on näyteltyjä liiketilanteita vaihtelevissa luonnonympäristöissä.

Videoissa vaihtelevat perusmuuttujat ovat muuttuja (näyttelijät ovat naisia, miehiä ja lapsia, jotka esiintyvät joko yksittäin tai ryhmänä), erilaiset kiintopisteet (esimerkiksi tie, metsä, luola tai meri) ja muuttujien erityyppiset liikeradat suhteessa kiintopisteisiin (esimerkiksi sisään tai ulos jostakin siirtyminen tai läpi kulkeminen). *Trajectoire*-kehys on kehitetty ensisijaisesti rajattujen liiketilanteiden tarkasteluun. Mukana on myös rajaamattomia liiketilanteita esittäviä videoita sekä rajattuuden suhteen monitulkintaisia videoita. Liikeradat vaihtelevat myös kompleksisuudeltaan hyvin yksinkertaisesta muutaman erilaisen vaiheen sisältävään (esimerkiksi polkua pitkin kävely ja astuminen sisään polun päässä olevaan luolaan). Videoissa liikutaan pääasiassa kolmella tavalla: kävel-

len, juosten ja hypäten. Elisitaatiokehityksen sisältöä erittelevät tarkemmin Kopecka ja Ishibashi (2011).

Visuaalisessa virikemateriaalissa on väistämättä muuttujana myös kuvakulma, ja on huomioitava myös sen mahdolliset vaikutukset kielenlennöksiin. *Trajectoire*-kehyksessä pääkuvakulmat ovat sivusta (muuttuja liikkuu kameran ohitse oikealle tai vasemmalle,  $n = 20$ ), edestä (muuttuja liikkuu kameraa kohti,  $n = 12$ ) ja takaa (muuttuja liikkuu kamerasta pois päin,  $n = 15$ ). Osassa videoista kuvakulma ei ole aivan yksiselitteinen, vaan ne on kuvattu joko hiukan viistosta tai niillä näkyvään liikkeeseen sisältyy muuttujan käänös. Nämä videot olen määritellyt yhdistelmiksi sivusta/edestä ( $n = 4$ ), sivusta/takaa ( $n = 3$ ) ja edestä/takaa ( $n = 1$ ).

Suomen aineisto on kerätty 50 suomenkieliseltä aikuiselta vuosina 2013–2015. Kukin koehenkilö katsoi *Trajectoire*-videosarjan yksi video kerrallaan, ja aina yhden videon katsottuaan koehenkilö kielensi lyhyesti näkemänsä eli vastasi kysymykseen ”Mitä videossa tapahtui?”. Videoista käytettiin kolmea erilaista esitysjärjestystä, jotta videoiden järjestyksen mahdollisia vaikutuksia kielenlennöksiin saataisiin vähennettyä. Keskimääräinen yhden koehenkilön tuottama sanamäärä on noin 453, ja koehenkilöiden tuotteliaisuus vaihteli suuresti: pienin sanamäärä on 254 ja suurin 1283.

Koehenkilöille annettiin ohje lyhyistä, yksilauseisista kielenlennöksistä, ja 90,3 prosenttia kielenlennöksistä sisältääkin vain yhden liiketilannetta kuvaavan lauseen (ks. taulukko 14). Aineistossa on myös 7,4 prosenttia kahden liikelauseen ilmauksia (esimerkki 6) ja 1,1 prosenttia kolmen liikelauseen ilmauksia (esimerkki 7).

6) Nainen **tuli** luolasta ja **käveli** polkua pitkin. (trFi#17-023)<sup>31</sup>

7) Poika **juoksee** hiekalla ja **hyppää** ison kiven yli ja **tulee** kohti kameraa. (trFi#18-063)

31 Suluisissa olevat aineistoviittaukset esitetään kansainvälisen Trajectoire-projektin yhtenäistä mallia mukaillen sisällyttäen seuraavat tiedot: viittaus aineistonkeruutyökaluun (tr = *Trajectoire*), kielen koodi (Fi = suomi), koehenkilön tunnistenumero ja viimeisenä kielenlennöksessä kuvaillun videon numero. Esimerkeistä on lihavoitu kulloinkin tarkastellut tai tulkinnan kannalta erityisen keskeiset kielenyksiköt.

Aineistossa ovat mukana ne liikevideoiden kielennökset, joissa on mukana translokaatiota eli muuttujan liike suhteutuu johonkin spatiaaliseen viitekehyykseen, tyypillisimmin kiintopisteeseen tai katsojan näkökulmaan. Tarkastelun ulkopuolelle rajattuja tapauksia ovat paitsi täysin muuhun kuin liikkeeseen keskittyvät kielennökset (esimerkiksi *Sukukokous lammen rannalla*) myös kielennökset, joissa ilmaistaan pelkästään muuttujan liike (ja korkeintaan tapa) ilman spatiaalista kehystämistä (esimerkiksi *Nainen kävelee*). Näitä tapauksia edustaa 1,2 prosenttia kielennöksistä.

Taulukko 14. Liikelauseiden määrä aineiston kielennöksissä (n = 2 740).

Liikelauseita / kielennös	Kielennökset %
0	1,2
1	90,3
2	7,4
3	1,1

Koehenkilöiden mediaani-ikä aineistonkeruuhetkellä oli 26, ja valtaosa heistä on nuoria aikuisia, pääasiassa eri alojen korkeakouluopiskelijoita. Koehenkilöistä 33 on naisia ja 17 miehiä. Aineistonkeruutilanteet videoitiin, litteroitiin standardiortografialla ja annotoitiin ELAN-ohjelmalla (Sloetjes & Wittenburg 2008). Koko aineisto koostuu 2 740 kielennöksestä, joissa on yhteensä 16 889 sanaa (filleri- ja harjoitusvideot on poistettu tästä aineistosta). Aineisto on analysoitu kahtena eri versiona: morfologisesti ja semanttisesti sana kerrallaan sekä semanttisesti kokonaisina kielennöksinä. Molempia aineistoversioita on käytetty tarkastelussa tutkimuskysymysten mukaan. Seuraavissa luvuissa tarkastellaan koko aineistosta poimittuja väylän KESKIVAIHEEN ja suunnan ilmauksia ja ilmaisukeinoja ja lopuksi tarkastelua rajataan tiettyjen videoiden kielennöksiin, toisinaan koko translokatiivisten liikeilmausten aineistoon tai muuntyyppisiin videoihin suhteuttaen.



## Väylän KESKIVAIHEEN kielentämisen keinot

Väylän KESKIVAIHEEN ilmaisukeinojen variaatio on suomessa runsasta, ja tässä luvussa esittelen aineistossa esiintyvät suomen kielen resurssit, joilla voidaan ilmaista väylän KESKIVAIHETTA. Keskeisimmäksi ilmaisukeinoksi erottuu laaja joukko grammeja. Suomen väylän KESKIVAIHEEN (reitit) grammeja on aiemmassa tutkimuksessa tarkasteltu muun muassa yksittäisten adpositioiden merkityksiin keskittyen (esimerkiksi Salmi 1994; Seppälä 2010), muutamaa grammaa vertaillen (esimerkiksi Lehismets 2014) sekä grammien syntaktiseen asemaan liittyviä merkityseroja tarkastellen (esimerkiksi Huumo 2010). Taulukossa 15 on esitetty tämän tutkimuksen aineistossa esiintyvät väylän KESKIVAIHEEN grammit frekvensseineen. Vasemmassa sarakkeessa ovat grammit, joilla väylän KESKIVAIHE kuvataan joko kiintopisteen laidalta toiselle ulottumiseksi tai kiintopisteen läpäisemiseksi, jotka edustavat tyypillisesti rajanylityksiksi<sup>32</sup> tulkittavia merkityksiä. Oikeassa sarakkeessa ovat grammit, joilla kuvataan kiintopistettä hetkellisesti sivuavaa väylää, siis vähemmän prototyyppisesti rajanylityksiksi miellettyjä ohituksia. Grammit *yli* ( $n = 305$ )<sup>33</sup> ja *ylitse* ( $n = 15$ ) voivat ilmaista kumpaakin merkitystyyppiä riippuen siitä, millaisen kiintopisteen yhteydessä ne esiintyvät: esimerkissä 8 *yli* kuvaa kiintopisteen päästä päähän ulottuvaa väylää ja esimerkissä 9 pikeminkin pistemäisesti kiintopistettä sivuavaa väylää.

8) *Nainen kävelee nurmikentän yli.* (trFi#23-051)

9) *Poika hyppää hietikolla kiven yli.* (trFi#4-063)

32 Rajanylityksen määrittely kirjallisuudessa on vaihtelevaa. Esimerkiksi Egan ja Graedler (2015) käsittelevät rajanylityksenä vain kolmiulotteisia ulos- ja sisään-tyyppejä. Arkikielessä raja taas mielletty pikeminkin kaksiulotteiseksi, ja myös oletus rajan lineaarisuudesta nousee keskeiseksi (ks. esimerkiksi *Kielitoimiston sanakirja* (2018), jossa *rajan* määrittelyssä käytetään muun muassa sanoja *linja*, *ääriiviiva* ja *reuna*). Esimerkiksi Özçalışkanilla (2015) väylän KESKIVAIHEEN rajanylitys on mukana analyysissä, mutta esimerkit edustavat vain melko prototyyppisiä tapauksia eli lineaaristen objektien ylittämistä (esimerkiksi aidan yli hyppääminen) tai alueen päästä päähän kulkemista (esimerkiksi maton yli ryömiminen). On avoin kysymys, missä määrin esimerkiksi ohitukset ovat samaistettavissa ylityksiin.

33 Jatkoissa kielenyksikön frekvenssi aineistossa ilmaistaan tällä tavalla ensimmäinnän yhteydessä.

Taulukko 15. Aineistossa esiintyvät väylän KESKIVAIHEEN grammit frekvensseineen.

Poikki-tyyppi	Ohi-tyyppi
<i>poikki</i> (139)	<i>ohi</i> (302)
<i>läpi</i> (13)	<i>takaa</i> (66)
<i>halki</i> (12)	<i>ohitse</i> (32)
<i>välistä</i> (6)	<i>editse</i> (20)
<i>kautta</i> (5)	<i>edestä</i> (19)
<i>keskeltä</i> (1)	<i>taitse</i> (3)
<i>lomitse</i> (1)	<i>vierestä</i> (1)
	<i>yli</i> (305)
	<i>ylitse</i> (15)

Grammin *yli* esimerkin 8 mukaista 'päästä päähän' -merkitystä edustavat myös *poikki* (n = 139) (esimerkki 10) ja harvakäyttöisempi *halki* (n = 12) (esimerkki 11). Yksittäisissä kielennöksissä vastaavaa merkitystä ilmaistaan myös *läpi*-grammilla (n = 13) (esimerkki 12).

- 10) Nainen kävelee varjoisan **pihan poikki**. (trFi#20-051)  
 11) Nainen kävelee **nurmikon halki** kohti puita. (trFi#9-039)  
 12) Nainen kävelee **puiston läpi** metsään. (trFi#13-057)

Grammi *läpi* on kuitenkin tyypillisempi voimadynaamisessa esteen läpäisemisen merkityksessään (esimerkki 13) (ks. Lehismets 2013).

- 13) Mies kävelee **puskien läpi** metsään. (trFi#33-026)

Aineisto ei huomattavien frekvenssierojen vuoksi mahdollista väylän KESKIVAIHEEN grammien laajempaa semanttista vertailua, mutta jo pinta-puolinen katsaus osoittaa, etteivät merkitykset ole selvärajaisia.

Grammien semanttinen erityispiirre on alan (*Region*) tarkka spesifointi väylän ilmaisun ohella (ks. esimerkiksi Zlatev 2007; Svorou 1994). Ala on topologista relaatiota tarkentava kategoria muuttujan ja kiintopisteen välillä: se ilmaisee, mihin osaan kiintopistettä tai sitä ympäröivää tilaa muuttujan liike tarkalleen kohdistuu. Esimerkiksi ohituksen kuvaamiseen käytetään paljon grammeja *ohi* (n = 302) ja *ohitse* (n = 32), mutta lisäksi ohituksen merkitys yhdistyy merkityksiin 'edessä' ja 'takana' grammeissa *edestä* (n = 19), *editse* (n = 20), *takaa* (n = 66) ja *taitse* (n = 3). Esimerkin 14 *takaa* ilmaisee väylän KESKIVAIHEEN ja

tarkentaa muuttujan mies ja kiintopisteen puu relaatiota kahdella alan arvolla, jotka ilmaisevat lateraalista vierekkäisyyttä ja takana olemista.

14) *Mies kävelee suuren puun takaa.* (trFi#8-039)

väylä: KESKIVAIHE

ala: SIVU+TAKAOSA

Edellisten lisäksi harvakseltaan käytettyjä, merkitykseltään spesifejä väylän KESKIVAIHEEN grammeja aineistossa ovat *välistä* ( $n = 6$ ), *kautta* ( $n = 5$ ), *keskeltä* ( $n = 1$ ), *vierestä* ( $n = 1$ ) ja *lomitse* ( $n = 1$ ).

Adnominaaleihin liittyvää väylän ilmaisua useista typologisesti erilaisista kielistä vertaava Papahagi (2011, 124) toteaa, että väylän KESKIVAIHEEN ilmaisu adnominaalein on semanttiselta granulariteetiltaan tyypillisesti niukempaa kuin väylän ALUN ja LOPUN: jos kielessä on yksinkertaisia adnominaaleja KESKIVAIHEEN ilmaisuun, ne ovat yleensä hyvin yleismerkityksisiä. Esimerkiksi semanttisesti väljästä väylän KESKIVAIHEEN adnominaalista käy ranskan adpositio *par*, joka voi kontekstista riippuen vastata useita suomen grammeja, kuten *kautta*, *läpi*, *halki*, *pitkin* ja jopa *kohti*. Suhteellisen runsasta väylän KESKIVAIHEEN ilmaisu on Papahagin (2011, 123 ja 125) mukaan romaniassa, mutta siinäkin osa muodoista lankeaa yhteen lokaation ilmaisun kanssa. Suomi muodostaa siis poikkeuksen Papahagin yleistykseen: väylän KESKIVAIHE eroaa kyllä suomessa väylän ALUSTA ja LOPUSTA, mutta erona ei ole ilmaisun pieni vaan pikemminkin suuri granulariteetti.

Koska väylän KESKIVAIHE on granulariteetiltaan korkea ja tulee yleensä ottaen ilmaistuksi harvemmin kuin väylän ALKU ja LOPPU (ks. esimerkiksi Talmy 2000a, 270),<sup>34</sup> on oletettavaa, että se tarkentuu lasten kielenkäytössä verrattain hitaasti. Tähän viittaa myös elisitoidun aineiston vertailu Koulu- ja Unelmakorpuksen (ks. Pajunen, Puranen & Yli-Paavola 2010; Pajunen 2016; ks. myös Pajunen ja Hongon johdannon aineistoesittelyä). Elisitoidussa aineistossa käytetyistä 16 väylägrammista

34 Talmyn mukaan liikettä ilmaistaessa tyypillisesti ikkunoidaan eli nostetaan huomion kohteeksi ja siten myös kielennetään alku- ja päätepiste, kun taas niiden välille jäävät liikeradan vaiheet jäävät helpommin spesifioimatta (*medial gapping*). Suomessa Talmy (2000a) huomion ikkunoinnin (*windowing of attention*) käsitettä on soveltanut väylän analyysiin esimerkiksi Ojutkangas (2003; 2012).

kuusi on käytössä alakoulun luokilla 1.–5. (*halki, kautta, läpi, ohi, poikki, yli*), 13 on käytössä myös nuorten aikuisten kirjoitelma-aineistossa, ja yhdeksän väylägrammin (*halki, läpi, kautta, keskeltä, takaa, vierestä, lomitse, ohitse* sekä *ylitse*) käyttö lisääntyy iän myötä eli on runsaampaa nuorilla aikuisilla kuin alakoululaisilla tai yhdeksäsluokkalaisilla.<sup>35</sup> Grammien käyttö kasvaa siitä huolimatta, että juuri nuoremmat kirjoittajat kuvaavat teksteissään usein hyvin konkreettisesti päivän etenemistä aamusta iltaan ja liikkumista eri paikkojen välillä. Ylipäätään aikuisten aineistossa väylägrammeja on käytössä monipuolisemmin kuin kouluikäisten aineistossa; samasta ilmiöstä kertoo Pajusen ja Hongon tässä kirjassa osoittama yleinen leksikaalisen diversiteetin kasvu iän myötä.

Väylän KESKIVAIHEEN sijailmaisuus on suomessa suhteellisen vähäistä. Aineistossa esiintyy harvakseltaan väylän KESKIVAIHEEN ilmaisuna elatiivi ( $n = 6$ ), jota voidaan käyttää läpi kuljettavan kiintopisteen yhteydessä, kuten esimerkissä 15. Tämä käyttö on rajoittunutta, koska semanttisesti hyväksyttävien kiintopisteiden joukko on pieni (Lauerma 1990, 134). Aineistossa elatiivilla ilmaistun väylän KESKIVAIHEEN kiintopisteinä esiintyvät *reikä*, (*luolan*) *suuaukko* ja *kallionhalkeama*. Elatiivi näkyy väylän KESKIVAIHEEN ilmaisussa myös grammien (esimerkiksi *välistä*) muodustuskeinona.

15) *Nainen meni aidassa olevasta reiästä.* (trFi#16-057)

Toinen KESKIVAIHEEN ilmaisuun liittyvä sija, prolatiivi, sen sijaan tulee aineistossa esiin ainoastaan grammien (esimerkiksi *ylitse, taitse, lomitse*) osana. Sillä ilmaistaankin tyypillisesti välillisempää väylää kuin elisitaatiomateriaalin omavoimaista liikettä edustavissa liiketilanteissa on: esimerkiksi sanomalehtikorpuksessa (Pajunen & Virtanen 2002) prolatiivi esiintyy 1400 kertaa ja 82 eri lekseimin yhteydessä, mutta joukossa ei väylägrammien lisäksi juuri ole inhimillisen muuttujan omavoimaiseen

35 Vain grammit *läpi, ohi ja yli* ovat nuorimmilla yksittäisiä tapauksia yleisempiä. Monia käytetään vielä hieman poikkeuksellisesti (esimerkiksi *Ja [haluaisin] matkustaa maapalloon yli, 3M70*) tai pelkkää suunnan ilmaisten. On ylipäätään yleistä, että nuorimmat kirjoittajat ilmaisevat spataalisessa kontekstissa vain suunnan tai väylän LOPUN. Viljasen (2012) mukaan 8-vuotiaat rajaavat väylän LOPUN suhteen tai jättävät liikeradan rajaamatta, 9-vuotiailla mukaan tulee myös ALUN suhteen rajattuja väyliä.

liikkeeseen soveltuvia väyliä. Sen sijaan prolatiivi on korpuksessa frekventein välineen ilmaisimena (esimerkiksi *puhelimitse, postitse*). Toinen keskeinen käyttöalue on laajojen, tyypillisesti kulkuvälinein kuljettavien reittien ilmaisu (esimerkiksi *rautateitse, ilmaitse*).

Väylän KESKIVAIHETTA voi suomessa ilmaista myös verbeillä. Aineistossa esiintyvät *ylittää* (n = 62), *ohittaa* (n = 38) ja *alittaa* (n = 1) ovat suomen selkeimpiä väyläleksikalisoivia verbejä, koska ne voivat ilmaista yksinään koko väylänformaation eikä ilmaisu jakaudu välttämättä verbin lisäksi sijaan tai adpositioon, kuten väylän ALKUA ja LOPPUA ilmaisevien verbien kohdalla tapahtuu (*ylittää tie* vs. *poistua luolasta, saapua luolaan*) (Pajunen 2001, 192). Typologiselta kannalta on huomionarvoista, että suomen kielessä ylipäätään on lainaamattomia väyläverbejä – esimerkiksi germaanisissa kielissä ne ovat tyypillisesti lainaa romaanisista kielistä, kuten englannin *cross, enter* ja *exit*. Frekvenssiltään väyläverbit ovat kuitenkin tässä aineistossa verrattain pieniä, kuten jäljempänä kaa-viosta 10 sekä taulukosta 16 käy ilmi.

Väylän KESKIVAIHEEN ilmaisu voi suomessa yhdistyä myös kiintopisteen ilmaisuun. Näin käy esimerkkien 16 ja 17 kaltaisissa tilanteissa, joissa kiintopisteenä on muuttujan kulkema reitti. Suomen kielen typologinen erityispiirre on rajattuuden ilmaisu objektin sijanmerkinnällä: totaaliobjekti (*portaat*) määrittää liiketilanteen rajatuksi, kun taas partitiiviobjekti (*portaita*) tuo mukanaan rajattomuustulkinnan (ks. esimerkiksi Heinämäki 1984).<sup>36</sup> Näissä tapauksissa rajaa aspektin ja spatiaalisen rajattuuden välille on jokseenkin keinotekoista vetää.

16) *Nainen nousi portaat.* (trFi#16-029)

17) *Nainen nousi portaita.* (trFi#43-074)

Partitiiviobjektilla ilmaistaan rajaamattomaan liikerataan liittyy suomen aineistossa myös grammeja: *pitkin* (n = 286) ja sen murteellinen variantti *pisin* (n = 8) sekä *myöten* (n = 6) ja *myöden* (n = 1), joiden täydennyksenä

<sup>36</sup> Kyseisten muotojen asettumisesta objektin ja adverbiaalin välimaastoon ks. esimerkiksi Leino (1991, 63–65).

partitiivi toimii.<sup>37</sup> Zlatev, David ja Blomberg (2010, 396) määrittelevät *along*-tyyppiset rajaamattomat liikeradat osaksi suunnan ilmaisua, koska väylän ja suunnan keskeinen ero on juuri rajattuudessa. Vektoriksi kuvattun suunnan käsite ei vaikuta aivan riittävältä kuvaamaan *pitkin*-tyyppisten<sup>38</sup> grammien merkitystä, koska niiden merkityksessä on keskeistä myös kiinteä yhteys kiintopisteen muotoon. Rajattuuden kannalta ero on kuitenkin selvä: siinä missä *poikki*-tyyppiset grammit liittyvät aineistossa aina rajattuihin liiketilanteisiin, *pitkin*-tyyppiset grammit jättävät liikeradan avoimeksi. Rajaamattomaksi *pitkin*-grammin kuvaa myös Huumo (2010, 542). Lehimetsä (2013, 349) erottaa sen tarkastelemistaan *läpi*-, *yli*- ja *kautta*-grammeista voimadynamiikan (Talmy 1985; 2000b) perusteella: *pitkin*-grammilla ei ole hänen mukaansa mahdollista kuvata esteen läpäisemistä, mikä luonnollisesti sopii yhteen yleisemmän rajaamattomuustulkinnan kanssa.

Rajattuuutta käsitellään tässä tutkimuksessa kuitenkin enemmän ilmaus- tai lausetason piirteenä kuin yksittäisten kielenainesten ominaisuutena, ja tämä merkityksen holistisuus huomioiden aineisto on luokiteltu arvioimalla jokaista kielennöstä erikseen. Vaikka tietty kielenyksikkö olisikin inherentisti rajaamaton, se voi esiintyä rajatussa kontekstissa (ks. esimerkiksi Huumo 2010, 543). Suomen aineistossa *pitkin*-ilmaukset esiintyvät rajatuissa konteksteissa (esimerkki 18) jonkin verran tyyppillisemmin kuin partitiiviobjekti (esimerkki 19). *Pitkin*-ilmaus on osana rajattua kielennöstä 40,4 prosentissa käyttöyhteyksistään ja pelkkä partitiiviobjekti 26,6 prosentissa.

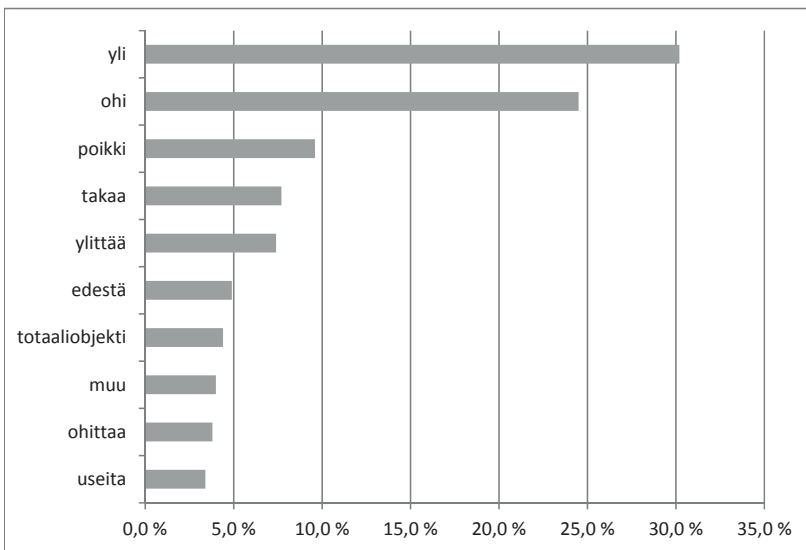
18) *Nainen kävelee luolasta metsään polkua pitkin.* (trFi#29-025)

19) *Nainen kävelee polkua luolaan.* (trFi#33-053)

37 Larjavaaran (2019, 110, 372) mukaan partitiivin käyttö reitin ilmaisussa näyttää olevan suomessa vanhempaa perua kuin *pitkin*-tyyppiset adpositiorakenteet eikä muodoilla ole suoranaista kehityksellistä yhteyttä.

38 Tarkoitan jatkossa *pitkin*-tyyppisillä grammeilla ja *pitkin*-ilmauksilla grammien *pitkin*, *pisin*, *myöten* ja *myöden* käyttöä. *Poikki*-tyyppisillä grammeilla viitataan grammeihin *poikki*, *halki* ja *läpi* sekä grammin *yli* niiden tapaiseen käyttöön.

Kun väylän KESKIVAIHEEN ilmaisua tarkastellaan kokonaisuutena, spatioaliset grammit nousevat selvästi tyypillisimmäksi ilmaisukeinoksi (ks. kaavio 10). Väyläverbit *ohittaa* ja *ylittää* sekä väylän ilmaisun yhdistyminen totaaliobjektilla ilmaistuun kiintopisteeseen ovat harvemmin käytettyjä keinoja. Kuten kaaviosta 10 käy ilmi, yleisin KESKIVAIHEEN grammi aineistossa on laajakäyttöinen *yli* (30,2 %). Lähes yhtä paljon käytetään grammia *ohi* (24,5 %). Yksittäisten grammien osuuksiin ei sinällään pidä kiinnittää liikaa huomiota, koska niiden määrä on olennaisesti sidoksissa stimuluksiin. Kaavion luokkiin on yhdistelty samankaltaisia grammeja seuraavasti: edestä = *edestä/editse*, ohi = *ohi/ohitse*, poikki = *poikki/halki*, takaa = *takaa/taitse*, yli = *yli/ylitse*. Luokassa muu on frekvenssiltään edellä mainittuja pienempiä grammeja, ja luokkaan useita on luokiteltu kielennökset, joissa väylän KESKIVAIHETTA kielennetään ainakin kahdella eri keinolla.



Kaavio 10. Väylän KESKIVAIHEEN ilmaisukeinot aineistossa.

Grammien laaja käyttö näkyy myös väylän KESKIVAIHEEN yhteydessä käytetyissä verbeissä (ks. taulukko 16): grammit esiintyvät tyypillisimmin tapaverbien kanssa (esimerkki 20), ja tämä verbiluokka onkin väylän

Taulukko 16. Väylän KESKIVAIHEEN ilmaisussa käytetyt verbit. Selvyyden vuoksi tarkasteltavana ovat kielennökset, joissa on ilmaistu pelkästään väylän KESKIVAIHE eikä eri vaiheiden yhdistelmiä.

Ilmaustyyppi %	Tapa <i>kävellä</i>	Väylä <i>ylittää</i>	Suunta: deiksis <i>mennä</i>	Suunta: muu <i>nousta</i>	Geneerinen <i>kulkea</i>	Yhteensä n
Väylän KESKIVAIHE	80,6	12,9	1,4	1,7	3,4	589

KESKIVAIHEEN ilmaisussa selvästi suurin (80,6 %). Vaihtoehtona grammirakenteille ovat väyläverbit eli lähinnä *ylittää* ja *ohittaa* (esimerkki 21). Niitä käytetään 12,9 prosentissa kielennöksistä. Geneerisiä liikeverbejä, joita edustaa tässä aineistossa ensisijaisesti tavan ja liikkeen suuntaisuuden suhteen neutraali *kulkea* (esimerkki 22), käytetään 3,4 prosentissa kielennöksistä.

20) *Kolme naista kävelee sillan yli.* (trFi#28-045)

21) *Poika ylittää pienen kivisillan.* (trFi#14-050)

22) *Mies kulkee polun yli.* (trFi#48-049)

Suuntaverbit on tässä katsauksessa jaettu kahtia: deiktisiin ja muihin suuntaverbeihin. Deiktiset verbit *mennä* (n = 131) ja *tulla* (n = 283) ovat väylän KESKIVAIHEESSA vain marginaalisesti edustettuina (1,4 %). Tämä on kiinnostavaa suhteessa väylän ALKUUN ja LOPPUUN, joiden ilmaisussa ne nousevat keskeisiksi erityisesti rajanylitysilmausten yhteydessä (Tuuri & Belliard, tulossa). Esimerkin 23 kaltaiset kielennökset, joissa deiktinen verbi yhdistyy rajanylitykseen väylän KESKIVAIHEESSA, ovat siis aineistossa yksittäistapauksia. Tässä kohdin nousee esiin kysymys elisitaatiotilanteen luomasta mahdollisesta rekisterierosta: arkisessa kielenkäyttötilanteessa *mennä X:n yli* saattaisi olla käytössä enemmän kuin elisitaatioaineistossa.

23) *Naiset menivät sillan yli.* (trFi#25-045)

Muita suuntaverbejä aineistossa ovat vertikaaliset *nousta* (n = 73) ja *laskeutua* (n = 39) ja niiden lisäksi *lähestyä* (n = 6). Väylän KESKIVAIHEESSA niiden pieni osuus (1,7 %) koostuu esimerkin 24 kaltaisista kielennöksistä, joissa verbi ilmaisee vertikaalisen suunnan.



24) *Nainen laskeutui portaat.* (trFi#16-022)

Väylän KESKIVAIHEEN ilmaisu on suomessa kielitypologisesti katsoen rikasta. Käytetyin ilmaisukeino on spatiaalinen grammi, ja väylä- ja suunta-verbit, suuntasija ja objektin sija ovat selvästi harvinaisempia.

## Suunnan kielentämisen keinot

Suunta on tässä tutkimuksessa merkityskategoria, joka kattaa joukon erilaisia rajaamattoman liikkeen spatiaalisten ulottuvuuksien ilmaisukeinoja. Suunnan käsitettä sivuttiin edellisessä jaksossa partitiiviobjektilla ilmaistun liikeradan sekä *pitkin*-tyyppisten grammien yhteydessä. Vaikka ne eivät aivan ongelmattomasti mielly suunnan ilmaisuksi, ne edustavat selvästi rajaamatonta liikerataa suhteessa rajatun liikeradan ilmaisukeinoihin, kuten totaaliobjektiin ja *poikki*-tyyppisiin grammeihin. Tässä luvussa esittelen muut aineistossa esiintyvät suunnan ilmaisukeinot. Väylän ja suunnan rajanveto ei kaikkien kielenyksiköiden kohdalla ole yksiselitteistä, vaan merkitykset määrittyvät osin holistisesti kontekstin mukaan.

Yksi keskeinen aineistossa ilmaistava suunta on kiintopisteestä tai katsojasta loittoneminen. Sitä ilmaistaan erityisesti grammeilla *pois* (n = 119) ja *poispäin* (n = 51). Myös *kauas*-grammin komparatiivi *kauemmas* (n = 6) ilmaisee samantyyppistä merkitystä kuin *poispäin*. Näistä grammeista aineistossa tyypillisin *pois* on kuitenkin yksi väylän ja suunnan rajanvedon kannalta haastavista kielenyksiköistä. Osa sen aineistoesiintymistä on samantyyppisissä konteksteissa kuin väylän ALKUA ilmaiseva grammi *ulos* (n = 123), osa edustaa selvästi rajaamatonta suuntaa grammin *poispäin* tapaan ja osa jää monitulkintaiseksi näiden välimaastoon. Seuraavaksi esittelen tarkemmin näiden kielenyksiköiden käyttöä aineistossa.

Grammien *pois* ja *ulos* käyttöä liittyy usein samoihin videoihin. Esimerkit 25 ja 26 ovat saman videon kielennöksiä, ja havaitaan, ettei grammin valinta luo olennaista merkityseroa vaan molemmat grammit ilmaisevat yhdessä elatiivisijan kanssa rajanylytystä väylän ALUSSA, siirtymää luola-kiintopisteen ulkopuolelle.

25) *Nainen kävelee pois luolasta.* (trFi#36-023)

26) *Nainen kävelee ulos luolasta.* (trFi#39-023)

Grammien merkityseroa voi hahmotella niiden kanssa esiintyvien kiintopisteiden kautta. Elisitaatiokehysten liikevideoissa selvärajaisin, säiliömäisin kiintopiste on useissa videoissa esiintyvä luola. Grammin *ulos* kaikista käyttökonteksteista 75,4 prosenttia on juuri kiintopisteen luola yhteydessä, mikä tukee Ojutkankaan (2012, 174) tuloksia, joiden mukaan Lauseopin arkiston murreaineistossa *ulko*-kantaisten grammien prototyypinen kiintopiste spatiaalista suhdetta ilmaistaessa on rakennus. Sen sijaan *pois*-grammi esiintyy kiintopisteen luola kanssa vain 30,1 prosentissa käytöistään, eli sitä käytetään vaihtelevammin eri kiintopisteiden kanssa. Lisäksi sillä on myös käyttöä, johon *ulos* ei sovi. Siihen lukeutuvat kiintopisteet, joiden kanssa tarvitaan spatiaalisen relaation tarkentava grammi, kuten esimerkin 27 *luota*. Väylän ALKU on esimerkiksi ilmaistu rakenteella *puun luota*, mutta on tulkinnanvaraista, onko *pois* rakenteessa osana väylän ilmaisua vai edustaako se objektikeskeisen viitekehysten kautta ilmaistua suuntaa – vai kantaako se itse asiassa molempia merkityksiä.

27) *Nainen hyppelehtii pois puun luota.* (trFi#26-033)

*Kielitoimiston sanakirjassa* (2018) grammin s.v. *ulos* spatiaalinen perusmerkitys määritellään seuraavasti: ”*pois* jnk sisästä, jstak t. jnk ulkopuolelle”, ja grammin s.v. *pois* vastaava määritelmä on laajempi: ”toisaalle, muualle, erilleen, eroon, sivuun, irti. i. ilmauksissa jotka merkitsevät toiseen paikkaan t. jnk ulkopuolelle t. jstak etäälle siirtymistä, siirtämistä t. joutumista. – –” Toisin sanoen *ulos* ilmaisee välttämättä rajanylityksen väylän ALUSSA, kun taas *pois* on merkitykseltään väljempi ja voi ilmaista sekä rajanylityksen (esimerkki 25) että etääntymisen kiintopisteestä (esimerkki 27). Eroa ilmentävät luonnollisesti myös grammien (substantiivi<sup>39</sup>)täydennysten sijat: *ulos* on mahdollinen vain elatiivisijaisen

39 Tässä aineistossa *ulos*-grammin täydennyksenä toimivat sanat ovat pääosin erilaisia substantiiveja, mutta sivuhuomautuksena voidaan todeta, että ulkopaikallissijaisina täydennyksinä voivat esiintyä pronominikantaiset adverbit: *ulos täältä/sieltä*.

täydennyksen kanssa, kun taas *pois* esiintyy elatiivin lisäksi ablatiivin kanssa. Suomen *pois* näyttäytyykin merkitykseltään laajempana kuin esimerkiksi germaanisten kielten vastineensa. Frawley (1992, 261, alaviite 4) toteaa, että englannissa *in* 'sisään' ja *out of* 'ulos' muodostavat keskenään konverssisuhteen, mutta *away* 'pois, kauemmas' on vain assosiaation (*coincidence*) negaatio. Vastaavasti esimerkiksi ruotsissa ja saksassa on omat kielenyksikkönsä, jotka ilmaisevat samoja merkityksiä kuin *away* ja *out of*, mutta suomeen Frawleyn analyysi ei päde, koska *pois* voi ilmaista kumpaakin merkitystä.

Grammin *pois* merkityksen moninaisuus hahmottuu paremmin, kun mukaan vertailuun otetaan grammi *poispäin*<sup>40</sup>. Se ei esiinny aineistossa *pois*- ja *ulos*-grammien kaltaisesti elatiivisijan ja säiliömäisen kiintopisteen kanssa. Esimerkissä 25 ilmauksella *pois luolasta* kielennetään ensisijaisesti naisen siirtymä luolan ulkopuolelle. Ainoa tulkintavaihtoehto on, että nainen on ensin ollut luolassa mutta siirtyy nyt sen ulkopuolelle. Siirtymän jälkeen nainen ei välttämättä etäänny kauas luolasta vaan saattaa jäädä myös sen välittömään läheisyyteen. Kun tilannetta verrataan sinällään mahdolliseen keksittyyn esimerkkiin 28, käy selväksi, että grammin *poispäin* käyttö samassa kontekstissa edellyttäisi naisen selvää etääntymistä luolan sijainnista. Sen sijaan esimerkki jättää avoimeksi, onko nainen lainkaan ollut sisällä luolassa. Grammin *poispäin* osana on partikkeli *päin*, jonka merkitys sellaisenaan on 'kohti' tai 'vastaan' ja joka olo- tai suuntasijaisen spatiaalisen ilmauksen yhteydessä ilmaisee suuntaisuutta tai summittaisuutta (Hakulinen, Vilkuna, Korhonen, Koivisto, Heinonen & Alho 2004, 697). Voidaan myös sanoa, että *päin*-partikkeli on nimenomaan rajaamaton, suunnan vektoria ilmaiseva elementti (ks. myös Nikanne 1990, 44).

### 28) *Nainen kävelee poispäin luolasta.*

Edeltävistä esimerkeistä käy ilmi, miten spatiaalinen informaatio rakentuu eri osien kautta: siihen vaikuttaa paitsi grammi ja sija myös maailmantieto kiintopistettä edustavan entiteetin ominaisuuksista. Spatiaali-

<sup>40</sup> Vastaavasti muodostettu grammi *ulospäin* esiintyy aineistossa 2 kertaa, ja *sisäänpäin* ei lainkaan.

sen merkityksen distribuutiota (ks. Sinha & Kuteva 1995) voidaan havainnollistaa vielä toisen *poispäin*-grammia koskevan esimerkin kautta: Esimerkissä 29 kiintopisteenä on puu, joka ei hahmotu luolan tapaan säiliöksi. On siis selvää, ettei lauseen merkitykseen sisälly siirtymää pois puun sisältä tai edes sen oksilta vaan pelkästään etäännyminen puun läheisyydestä. Esimerkin 29 tulkinta kääntyy jo rajaamattomuuden puolelle, koska aivan tarkka lähtöpiste jää avoimeksi ja sijoittuu vain jonnekin suhteellisen lähelle puuta. Se vertautuu esimerkin 30 ilmaukseen *kauemmas puusta*, jossa lähtötilanne jää samaan tapaan epätarkaksi.

29) *Nainen kävelee puusta poispäin.* (trFi#10-032)

30) *Nainen käveli kauemmas puusta.* (trFi#5-052)

Jos *pois*-grammin käyttöä jäsennetään väylän ja suunnan väliseksi jatku-moksi, *pois* näyttää siis ilmaisevan selvimmin väylää sentyyppisissä konteksteissa, joissa voidaan käyttää myös *ulos*-grammia. Toisessa ääri-päässä *pois* ilmaisee suuntaa samankaltaisesti kuin *poispäin*. Esimerkit 31–33 ovat saman videon kielennöksiä, joissa *pois* ja *poispäin* ovat jokseenkin samanlaisessa käytössä. Videolla nainen kävelee puistoaukiolla metsänreunaa kohti. Video on kuvattu melko kaukaa takaapäin. Naisen saapumista metsään ei näytetä, eikä liikkeelle näy myöskään selvää alkupistettä, vaan videon alkaessa nainen on jo liikkeellä. Esimerkeissä *pois* ja *poispäin* edustavat siis näkökulmakeskeistä suunnan ilmaisua: niillä ilmaistaan naisen etäännyvän katsojasta itsestään (esimerkki 31 ja 32) tai eksplisiittisesti mainitusta kamerasta (esimerkki 33), jonka sijainti luonnollisesti vastaa katsojan näkökulmaa. Tällaisissa konteksteissa *pois* on harvinainen mutta *poispäin* hyvin tavallinen.

31) *Nainen kävelee pois.* (trFi#6-052)

32) *Nainen käveli poispäin puistikossa.* (trFi#17-052)

33) *Nainen kävelee nurmikkoa pitkin poispäin kamerasta.* (trFi#8-052)

Yksittäisissä kielennöksissä *pois* ja *poispäin* esiintyvät myös yhdessä. Kielennöksessä 34 siirtymän luolasta ilmaisee lauseke *pois sieltä luolasta*,

ja *poispäin* tarkentaa oletettavasti suuntaa näkökulmakeskeisesti, koska nainen liikkuu videolla kauemmas katsojasta.

- 34) *Siinä oli luola ja sit nainen kävelee pois sieltä luolasta poispäin.* (trFi#31-025)
- 35) *Siinä oli- nainen makas nurmikolla ja mies käveli pois sen luota poispäin.* (trFi#22-035)

Esimerkissä 35 aineistonkeruutilanne näkyy ehkä hiukan epäluontevana tarkkuutena, mutta kiinnostavaa on liikkeen suunnan tähdentäminen *poispäin*-grammilla. Lauseke *pois sen luota* näyttäisi siis ilmaisevan väylän ALUN, siirtymän erilleen kiintopisteestä nainen ja *poispäin* ilmaisevan suuntaa. Videolla mies kulkee suoraan katsojaa kohti, joten *poispäin* ei ilmaise miehen liikettä suhteessa katsojan sijaintiin vaan edelleen objektikeskeisesti suhteessa maassa makaavaan naiseen.

Ilmauksille *pois* ja *poispäin* vastakkaista suuntaa edustavat erilaiset tavat kielentää lähestymistä, 'kohti'-merkitystä. Aineistossa tämänkaltaista merkitystä ilmaistaan runsaasti grammilla *kohti* (n = 191), yksittäisissä tapauksissa myös grammeilla *kohden* (n = 1) ja *päin* (n = 3).<sup>41</sup> Papahagin (2011, 125) mukaan on typologisesti tavallista, ettei 'kohti'-merkitystä ilmaisevaan adnominaaliin sisälly kiintopisteen muotoon liittyvää tietoa eli tässä käytetyin termein alan kielentämistä, joka suomessa tyypillisesti kytkeytyy väylän ilmaisuun. *Kohti*-ilmauksiin ei liity samaa tilannekohtaista vaihtelua väylän ja suunnan välillä kuin *pois*-ilmauksiin, vaan ne edustavat yksiselitteisesti suuntaa.

Valtaosa *kohti*-tyyppisistä ilmauksista edustaa objektikeskeistä viitekehystä, eli niissä on ilmaistu kiintopisteobjekti, johon eteneminen suhteutetaan (esimerkki 36).

- 36) *Nainen kulkee kohti puuta.* (trFi#48-061)

41 Grammilla *päin* voi ilmaista myös KONTAKTIA (esimerkiksi *Nainen kävelee melkein puuta päin puistossa* [trFi#40-061]), mutta tätä merkitystä edustavia tapauksia ei ole tässä laskettu mukaan suunnan ilmaisuun.

Viitekehys voi kuitenkin olla myös näkökulmakeskeinen, kuten esimerkiksi 37, jossa kielentäjä suhteuttaa liikkeen omaan näkökulmaansa *kohti*-grammilla. Lisäksi 'kohti'-merkityksistä suunnan ilmaisua sisältyy esimerkiksi 37 *tulla*-verbiin. Se liittyy aineistossa kuitenkin ensisijaisesti väylän ALUN ilmaisuun (esimerkki *Nainen tulee luolasta*) ja jää siksi tässä yhteydessä suuresta frekvenssistään huolimatta vähälle huomiolle (ks. tarkemmin Tuuri & Belliard, tulossa).<sup>42</sup> Esimerkissä 38 käytetty verbi *lähestyä* (n = 6) ilmaisee selvemmin rajaamatonta kohti-tyyppistä merkitystä mutta on puolestaan aineistossa harvinainen. Niin ikään pienifrekvenssinen *lähelle*-grammin komparatiivi *lähemmäs* (n = 5) edustaa aineistossa käytöltään kohti-tyyppistä merkitystä ja on käytössä nimenomaan näkökulmakeskeisesti, katsoja itse kiintopisteenään (ks. esimerkiksi 39).

37) *Nainen tulee kohti luolasta.* (trFi#50-023)

38) *Mies lähestyy nurmikolla makaavaa naista.* (trFi#19-036)

39) *Poika kävelee lähemmäs.* (trFi#14-067)

Vertikaalisen suunnan ilmaisua sisältyy grammeihin *ylös* (n = 116) ja *ylöspäin* (n = 34) sekä *alas* (n = 82) ja *alaspäin* (n = 12) sekä verbeihin *nousta* (n = 73) ja *laskeutua* (n = 39). Suuntaisuutta myös yhdistyy tavan ilmaisuun verbeissä *kiivetä* (n = 23) ja *kivuta* (n = 3).<sup>43</sup> Vertikaalista liike-rataa ilmaisevissa grammeissa suunta on aina keskeisenä merkityksenä, mutta niiden käytössä voi myös hahmottaa väylän ja suunnan jatku-moa samankaltaisesti kuin edellä on esitetty *pois*- ja *poispäin*-grammien käytön jäsentyvän. Esimerkiksi grammin *ylös* merkityksen nyansseja havainnollistaa hyvin video 029, jossa nainen nousee sammaloituneita kiviportaita. Liikkeen lähtöpisteenä näkyy videolla luolaksi tulkittava tila, josta käsin video on kuvattu. Nainen siis etäännytty katsojasta. Esimerkki 40 havainnollistaa paitsi väylän ja suunnan rajanvedon häilyvyyttä ver-

42 Vastaavasti *mennä*-verbin voi katsoa kantavan merkitystä 'pois/poispäin', mutta se liittyy aineistossa ensisijaisesti väylän LOPUN ilmaisuun ja jää siis tässä tarkastelussa sivuun (ks. kuitenkin Tuuri & Belliard, tulossa).

43 Tässä aineistossa verbit *kiivetä* ja *kivuta* edustavat ylöspäin suuntautuvaa liikettä, mutta marginaalisesi niillä on myös mahdollista ilmaista alaspäin suuntautuvaa liikettä, kuten *Kielitoimiston sanakirjan* (2018) esimerkkilauseessa *Kiipeä heti alas pöydältä!*

tikaalisten grammien kohdalla, myös jälleen spatiaalisen merkityksen distribuutiota (Sinha & Kuteva 1995): Elatiivi ilmaisee väylän ALKUA, mutta sen rinnalla onkin tyypillisempien väylän ALUN grammien (*pois* tai *ulos*) sijaan vertikaalinen *ylös*, joka hahmottuu tässä sekä suunnan ilmaisimeksi että osaksi väylän ilmaisua. Kiintopisteestä kielentäjä käyttää lekseemiä *onkalo*, joka tarkoittaa syvännettä ja sisältää siten kiintopisteen muodon kautta rakentuvaa tietoa poistumisen vertikaalisesta suunnasta. Ilmaus *onkalosta ylös* näyttää siis ilmaisevan toisteisesti sekä väylän ALKUA että suuntaa.

40) *Ja siinä nainen käveli tommosia portaita pitkin jostain pimeästä onkalosta ylös.* (trFi#22-029)

Toisessa saman videon kielennöksessä (esimerkki 41) väylän ALKU on sen sijaan ilmaistu omassa *luolasta ulos* -lausekkeessaan, ja *ylös*-grammi onkin selvemmin osana suunnan ilmaisua.

41) *Nainen kävelee luolasta ulos kivirappusia pitkin ylös.* (trFi#20-029)

Grammit *ylös* ja *alas* ovat mahdollisia sekä väylän että suunnan ilmaisussa, kun taas *päin*-partikkelilla muodostetut *ylöspäin* ja *alaspäin* liittyvät odotuksenmukaisesti aina suunnan ilmaisuun (esimerkki 42).

42) *Poika kävelee hiekkaisista polkua alaspäin.* (trFi#35-067)

Vertikaalisten ilmausten rajatuinta ääripäätä edustavat esimerkin 43 kaltaiset ilmaukset: kiveltä maahan hyppääminen on luonteeltaan niin selvärajainen ja punktuaalinen tapahtuma, että grammin *alaspäin* rajamattomuus olisi ristiriidassa tapahtuman rajattuuden kanssa.

43) *Poika hyppää alas kiveltä.* (trFi#26-034)

Jonkin verran aineistossa käytetään myös lateraalisen akselin suunnan ilmauksia *oikea* ( $n = 50$ ) ja *vasen* ( $n = 57$ ). Näitä näkökulmakeskeisiä ilmauksia käytetään sekä kielennöksen ainoana spatiaalisena informaati-

tiona että muun väylän tai suunnan ilmaisun ohella, kuten käy ilmi esimerkeistä 44 ja 45, jotka ovat saman videon granulariteetiltaan erilaisia kielennöksiä.

44) *Nainen käveli vasemmalta oikealle.* (trFi#43-048)

45) *Nainen tuli metsästä vasemmalta ja meni polun poikki oikealle.* (trFi#25-048)

Harvakseltaan käytettyä suunnan ilmaisua edustaa *eteenpäin* (n = 7). Se suhteuttaa muuttujan liikkeen muuttujan aiempaan sijaintiin eikä ole riippuvainen havainnoijan näkökulmasta (ks. myös Blomberg 2014, 64). Tämä havainnollistuu ilmauksen käyttöä tarkastelemalla, vaikka otos toki jääkin pieneksi: ilmausta käytetään kaikista kuvakulmista kuvattujen videoiden kielennöksissä. Esimerkissä 46 kielennetyissä videossa nainen kävelee puun luota suoraan kameraa kohti, eli kielentäjä kuvaa frontaalista akselia muuttujan, ei itsensä näkökulmasta.

46) *Nainen kävelee puun edestä eteenpäin.* (trFi#7-032)

Kouluikäisten ja nuorten aikuisten kirjoitelmissa suuntaa ilmaisevat grammit ovat yleisemmin käytössä kuin väylän *KESKIVAIHETTA* ilmaisevat grammit. Nuorimmat käyttävät runsaasti grammeja *ulos* ja *pois*, mutta myös *kohti* esiintyy muutaman kerran, satunnaisemmin käytössä on muita. Grammit *ulos* ja *pois* voivat olla nuorimmilla myös toistensa vaihtoehtoja mutta eivät niinkään enää vanhimmilla, joilla esimerkiksi (*läheteä*) *pois* korvautuu usein verbillä *poistua*. Ehkä huomiota herättävintä on *kohti*-grammin runsastuminen yläkoululaisista nuoriin aikuisiin. Sen tyypillinen käyttökonteksti on kerronnan vaihdokset tai siirtymät, joissa otetaan suunta kohti jotain uutta paikkaa, lähdetään kohti jotakin. Leksikaalinen diversiteetti nousee iän myötä samoin kuin väylän *KESKIVAIHEEN* kielennöksissä.

Verbeistä tapaverbit hallitsevat sekä suunnan ilmaisua (esimerkki 47) että myös tässä tutkimuksessa rajaamattomaan liikerataan rinnastettuja *PAIKKA*-ilmauksia (esimerkki 48) (ks. taulukko 17). Suunnan ilmauksista tapaverbejä käytetään 78,8 prosentissa ja *PAIKKA*-ilmauksista 96,2



Taulukko 17. Suunta- ja PAIKKA-ilmauksissa käytetyt verbit.

Ilmaustyyppi %	Tapa kävellä	Suunta: muu nousta	Geneerinen kulkea	Suunta: deiksis mennä	Muu	Yhteensä n
Suunta	78,8	13,6	5,5	0,5	1,1	419
PAIKKA	96,2	–	3,4	0,5	–	208

prosentissa. Muista verbeistä keskeisiksi nousevat vain muuta kuin deiktistä suuntaa ilmaisevat suuntaverbit. Niiden 13,6 prosentin osuus suunnan ilmauksissa koostuu lähinnä vertikaalisten *nousta-* ja *laskeutua-*verbien käytöstä (esimerkki 49).

- 47) *Ihmiset kävelevät kohti rantaa.* (trFi#2-066)  
 48) *Poika juoksee rantahietikolla.* (trFi#11-042)  
 49) *Nainen nousee metsäpolkua.* (trFi#14-070)

Geneeristen verbien osuus on suunnan ilmauksissa hiukan suurempi kuin PAIKAN ja väylän KESKIVAIHEEN ilmauksissa (ks. taulukko 16). Osuus koostuu kokonaan *kulkea-*verbin käytöstä (esimerkki 50).

- 50) *Nainen kulkee polkua pitkin ylöspäin.* [trFi#32-070]

## Väylän KESKIVAIHEEN ja suunnan variaatio kulkureitin ilmaisussa

Erilaiset kulkureitit ovat liiketilanteissa luontevia kiintopisteitä ja myös *Trajectoire-*elisitaatiokehityksessä keskeisiä: niitä edustavat videoissa paitsi erilaiset tiet ja polut myös sillat sekä portaat. Tässä luvussa tarkastelen rinnakkain väylän KESKIVAIHEEN ja suunnan sekä niiden ohessa myös PAIKAN ilmaisuja, koska nämä näyttäytyvät erityisesti kulkureittejä kuvaavien videoiden kohdalla vaihtoehtoisina käsitteistämisen tapoina: kokenkilöt tulkitsevat samaa videota esimerkiksi sillan ylitykseksi (väylän KESKIVAIHE), siltaa pitkin kulkemiseksi (suunta) ja sillalla kulkemiseksi (PAIKKA). Väylän KESKIVAIHEEN ilmaisu esittää tilanteen spatiaalisesti rajattuna, suunnan ja PAIKAN ilmaisu rajaamattomana.

Tarkastelun taustaksi on hyvä huomioida, että kun otetaan huomioon koko liikeilmausten aineisto, siis myös väylän ALUN ja LOPUN ilmaukset, yksimielisimmin rajattuina kielennetyt videot ovat sellaisia, joissa näkyy selkeä rajanylitys (ks. myös Tuuri & Belliard, tulossa). Kaikkien tai lähes kaikkien koehenkilöiden rajattuina kielentämiin videoihin kuuluu paitsi siirtymiä pois rajatusta tilasta (esimerkiksi *Nainen kävelee pois metsästä*) tai rajattuun tilaan (esimerkiksi *Poika kävelee luolaan*) myös väylän KESKIVAIHEEN rajanylityksiä, joissa ylitetään este (esimerkiksi *Mies hyppää tukin yli*). Rajanylitys näyttää pääsääntöisesti tulevan kielennetyksi, jos sellainen videossa on. Kulkureittiä kuvaavien videoiden tapauksessa rajanylityksen ilmaisu ei kuitenkaan näyttäydy aivan yhtä selkeänä tendenssinä, kuten tämän luvun analyysissä käy ilmi.

Kulkureittien lisäksi aineistossa esiintyviä väylän KESKIVAIHEESEEN ja suuntaan liittyviä kiintopisteitä ovat erilaiset objektit ja henkilöt sekä laajemmat alueet, kuten puistot ja metsät. Näitä muita tyyppisiä otan tarpeen mukaan vertailukohdiksi. Keskityn analyysissä kahteen granulariteetin tasoon van Stadenin ja Narasimhanin (2012) esitystä soveltaen: siihen, mitä liiketilanteen elementtejä ilmaistaan, ja siihen, millaisia kielellisiä keinoja ilmaisussa käytetään. Yleisluontoisemman analyysin lomassa tarkastelen keskeisiin ilmaisukeinoihin liittyviä merkityspiirteitä yksityiskohtaisemmin. Seuraavassa tarkastelen ensin ilmauksia, joissa kiintopisteenä on tie tai polku, sitten ilmauksia, joissa kiintopisteenä on silta, ja lopuksi ilmauksia, joissa kiintopisteenä on portaat.

#### KIINTOPISTEENÄ TIE TAI POLKU

Videoista, joiden kiintopisteenä on tie tai polku (ks. taulukko 18), kaksi on ilmaistu pääosin rajattuna väylänä, mutta ei kuitenkaan aivan yksimielisesti. Videoissa 049 (40/50 rajattu) ja 048 (39/50 rajattu) katsojan eteen aukeaa polku metsäisessä maisemassa. Videossa 049 polun yli kulkee mies oikealta vasemmalle ja videossa 048 nainen vasemmalta oikealle. Muuttuja ilmestyy tielle oksien katveesta ja polun ylityksen jälkeen katoaa jälleen polun toisella puolen olevaan metsikköön. Valtaosa koehenkilöistä kuitenkin keskittyy kielennöksissään väylän KESKIVAIHEESEEN eli kielentää vain polun ylityksen. Liiketilanteet kuvataan pääosin rajanylityksenä joko grammilla tai *ylittää*-verbillä. Grammeista käy-

tetyin on *yli*, mutta kielennöksissä esiintyy monipuolisesti myös muita grammeja: *poikki*, *läpi* ja jopa rajaamaton *pitkin* (esimerkki 51), vaikka liike ei etene polun suuntaisesti, mikä jälleen viittaa grammien merkityksen joustavuuteen ja mahdollisesti subjektiivisiin eroihin niiden hahmotuksessa. Rajaamattomana kyseiset videot on kielennetty tyypillisimmin suhteellisen epätarkasti PAIKKA-ilmauksella (esimerkki 52). Luokassa n/a ovat kielennökset, jotka eivät sisällä minkäänlaista spatiaalista informaatiota (esimerkiksi *Mies lenkkeilee*).

Taulukko 18. Väylän KESKIVAIHETTA kuvaavat videot<sup>44</sup>, joissa kiintopisteenä on tie tai polku.<sup>45</sup>

Video	Kuvakulma	Väylä	Suunta	PAIKKA	n/a
049: Mies kävelee polun yli	sivusta	40/50	2/50	8/50	–
048: Nainen kävelee polun yli	sivusta	39/50	4/50	7/50	–
037: Mies juoksee polkua ylös	sivusta/edestä	9/50	34/50	5/50	2/50
041: Ryhmä kävelee metsätietä	sivusta/edestä	4/50	28/50	18/50	–
067: Poika kävelee kivikkoista polkua	edestä	3/50	35/50	11/50	1/50
065: Poika kävelee vuoristopolkua	sivusta	–	34/50	14/50	2/50

51) *Mies kulkee polkua pitkin metsään.* (trFi#32-049)

52) *Nainen kävelee puistossa.* (trFi#4-048)

Muut videot, joissa muuttuja liikkuu suhteessa polkuun tai tiehen, edustavat pääosin rajaamatonta liikerataa. Video 065 on useimmin rajaamattomana kielennetty video: suunta- ja PAIKKA-ilmauksia siitä käyttää 48 koehenkilöä. Se on myös ainoa video, jonka liikerataa kukaan koehenkilöistä ei ole kielentänyt rajatuksi. Videolla poika kävelee korkealla vuorenrinteellä hyvin kapeaa polkua vasemmalta oikealle. Video on kuvattu sivusta ja melko kaukaa, joten pojan etenemistä näkyy suhteellisen pitkä matka. Poika on liikkeellä videon alkaessa ja jatkaa liikettään videon

44 Videoiden kielelliset kuvaukset edustavat tässä ja muissa taulukoissa luonnollisesti vain yhtä tapaa kielentää tilanne. Selkeyden vuoksi kuvauksissa on käytetty koko aineistossa yleisimpiä tapaverbejä, vaikka joidenkin videoiden tapauksessa muuntyyppinen verbi voi olla tavallisempi.

45 Tässä tarkasteltujen videoiden kielennöksistä 16:ssa rajattuustulkinta syntyy muuten kuin väylän KESKIVAIHEEN ilmaisen kautta. Esimerkiksi videon 048 kielennöksessä *Nainen tulee jostaki siimeksestä ja jatkaa kävelyä* [trFi#44-048] koehenkilö keskittyi tyypillisemmin huomioidun polun ylityksen sijaan väylän ALKUUN, jossa nainen poistuu metsästä. Nämä tapaukset muodostavat kuitenkin vain 0,5 prosenttia tässä tarkasteltujen videoiden rajatuista kielennöksistä.

päätyessä, minkä vuoksi onkin odotuksenmukaista, ettei tilannetta tul-  
kita spatiaalisesti rajatuksi.

Videon 065 kuvauksissa on eniten *pitkin*-tyyppisiä grammeja (eli *pitkin-* ja *myöten*-grammeja varianteineen): niitä käyttää 21 koehenkilöä. Grammin *pitkin* merkitys kuvataan *Kielitoimiston sanakirjassa* (2018) paitsi *myöten*-synonyymilla myös 'seuraamisen' tai 'noudattamisen' ilmaisuksi. Kapean polun seurailu vuorenrinteen mukaisesti näyttäytyykin otollisena kontekstina *pitkin*-grammille. Lähes yhtä moni, 20 koehenkilöä, käyttää *pitkin*-tyyppisiä grammeja videosta 041, jossa muuttaman henkilön joukko kulkee metsän siimeksessä mutkittelevaa polkua. Tässäkin videossa kiintopiste on suhteellisen pitkä ja kapea polunpätkä, jonka muotoa muuttajat seurailevat.

Ylipäätään *pitkin*-tyyppisten grammien käyttö liittyy aineistossa selkeästi lineaarista kulkureittiä edustaviin kiintopisteisiin, ja sama pätee myös pelkällä partitiiviobjektilla kielentämiseen (ks. esimerkki 53).

53) *Mies juoksee polkua.* (trFi#9-037)

Laajempien alueiden yhteydessä kummankintyyppiset ilmaukset ovat epätyypillisiä. Poikkeuksen muodostaa video 076, jonka liikerataa 19 koehenkilöä ilmaisee partitiiviobjektilla ja kolme *pitkin*-rakenteella. Videolla pikkupoika kävelee rantakalliolla kohti kameraa, ja kallio viettää selvästi alaspäin. Partitiiviobjekti esiintyykin videon kielennöksissä nimenomaan vertikaalista suuntaa ilmaisevien elementtien kanssa (esimerkki 54 ja 55), jotka yhdessä kiintopisteen kanssa ilmaisevat muuttujan liiketilan.

54) *Poika kävelee kalliota alas.* (trFi#4-076)

55) *Pikkupoika laskeutui varovasti kalliorinnettä.* (trFi#18-076)

Pelkkä partitiiviobjekti ei tyyppisesti näytäkään riittävän ainoaksi spatiaaliseksi informaatioksi laajempaa aluetta kuvaavien kiintopisteiden kanssa (ks. keksitty esimerkki 56). Lineaariseen kiintopisteeseen partitiiviobjekti yhdistyy luontevasti sellaisenaankin (ks. esimerkki 53).

56) *Mies juoksee ?puistoa/aukiota/rantahiekkaa.*

Epätavallisia ovat myös *pitkin*-ilmaukset aluemaisten kiintopisteiden yhteydessä. Niitä esiintyy ainoastaan kolme nurmikko-kiintopisteen (esimerkki 57), kolme kallio-kiintopisteen ja kaksi (ranta)hiekka-kiintopisteen yhteydessä.

57) *Nainen kävelee nurmikkoa pitkin pois päin kamerasta.* (trFi#8-052)

Laajaa aluetta kuvaavien kiintopisteiden kanssa tavanomaisempaa ilmaisuvoisikin olla *pitkin*-adposition prepositiokäyttö, joka Huumon (2010, 536) mukaan kantaa hajanaisuuden ja epämääräisyyden merkityksiä, kun taas postpositiokäyttö liittyy enemmän varsinaiseen väylän tai suunnan ilmaisuun. Elisitoidun aineiston liikeradat ovat kuitenkin pääosin melko suoraviivaisia, joten *pitkin*-grammin prepositiokäyttöä (esimerkki 58) on koko aineistossa vain kuusi esiintymää.

58) *Poika juoksee pitkin rantahiekkaa.* (trFi#20-042)

PAIKKA-ilmausten osuus on koko aineistossa 7,6 prosenttia. Kolmea kaikista elisitaatiokehysten videoista eniten PAIKKA-ilmauksia tuottanutta videota yhdistää se, että kiintopiste on laaja alue, ei tarkkarajainen kulkureitti. PAIKKA-ilmauksia käyttää enemmistö, 37 koehenkilöä, vain videosta 042, jossa sivusta kuvattu poika juoksee hiekkarannalla vasemmalta oikealle korkean ruovikon reunaa (esimerkki 59). Videosta 051, jossa nainen kävelee puistossa, niitä käyttää 25 koehenkilöä. Video 076, jossa poika kävelee rantakallioilla, on kielennetty PAIKKA-ilmauksella 21 kielennöksessä.

59) *Poika juoksee rantahietikolla.* (trFi#45-042)

Niistä videoista, joissa kiintopisteenä on kulkureitti, PAIKKA-ilmauksia käyttää 0–18 koehenkilöä. Osa tapauksista perustuu kuitenkin siihen, että kiintopisteeksi on itse asiassa kielennetty liiketilanteen laajempi ympäristö, esimerkiksi *polun* sijaan *metsä*. Videota 041 on kuvattu kul-

kureittivideoista eniten PAIKKA-ilmauksilla, mutta lähempi tarkastelu osoittaa, että 18 PAIKKA-ilmauksesta seitsemässä kiintopisteenä on alue (esimerkki 60) kulkureitin (esimerkki 61) sijaan.

60) *Joukko ihmisiä kulkee metsässä.* (trFi#30-04I)

61) *Ryhmä ihmisiä kävelee metsäpolulla.* (trFi#34-04I)

#### KIINTOPISTEENÄ SILTA

Niihin videoihin, joissa kiintopisteenä on silta, lukeutuvat eniten rajattuna kielennetyt kulkureittivideot (ks. taulukko 19). Videossa 046 (49/50 rajattua) mies ylittää siltaa, jolla seisoskelee muutaman naisen joukko keskustelemassa. Videossa 045 (46/50 rajattua) naisjoukko ylittää siltaa. Pääosin rajattuna on kuvattu myös video 047 (42/50 rajattua), jossa nainen ylittää siltaa, jolla mies seisoo katselemassa veteen. Näitä enemmän suunnan ja PAIKAN ilmaisua liittyy videoon 050 (22/50 rajattua), jossa poika kävelee siltaa pitkin.

Taulukko 19. Videot, joiden kiintopisteenä on silta.

Video	Kuva- kulma	Väylä	Suunta	PAIKKA	n/a
046: Mies kävelee sillan yli naisten ohi	takaa	49/50	–	1/50	–
045: Naiset kävelevät sillan yli	takaa	46/50	2/50	2/50	–
047: Nainen kävelee sillan yli miehen ohi	sivusta	42/50	2/50	1/50	5/50
050: Poika kävelee siltaa pitkin/sillan yli	sivusta/ takaa	22/50	14/50	14/50	–

Videoissa 046 ja 047 väylän KESKIVAIHETTA ilmaistaan suhteessa kahteen eri kiintopisteeseen: molemmissa muuttuja ylittää sillan ja lisäksi kulkee elollisen kiintopisteen ohi. Itse asiassa videoita 046 ja 047 on tyyppisemmin kielennetty rajattuna juuri ohituksen, ei niinkään ylityksen kannalta: videosta 046 koehenkilöistä 35 kielentää pelkän ohituksen, videosta 047 puolestaan 39 koehenkilöä. Silta on tyyppinen kiintopiste siinä mielessä, että se on muuttujaa pysyvämpi ja suurempi taustaelementti, mutta on myös odotuksenmukaista, että koehenkilöiden huomio kiinnittyy videolla näkyviin ihmisiin, jotka tulevat myös kielennösten kiintopisteiksi. Kahden erillisen väylän KESKIVAIHEEN rajanylityksen yhdistäminen samaan kielennökseen on näissä tapauksissa jokseenkin

harvinaista (n = 9), vaikka monet koehenkilöt tuottavat ylipäätään pitkiä ja kompleksisia kuvauksia. Tällaisista kielennöksistä useimmissa (n = 7) rajanylitykset on ilmaistu erillisissä lauseissa (esimerkki 62). Vain kaksi kielennöstä sisältää kaksi eri kiintopisteisiin liittyvää väylän ilmausta, joista toisessa on väyläverbi ja toisessa väylägrammi (esimerkki 63).

62) *Mies ylittää sillan ja ohittaa keskustelevat naiset.* (trFi#33-046)

63) *Nainen ylittää sillan miehen ohi.* (trFi#15-047)

Video 050 on siltavideoista ainoa, jossa rajaamattoman liikeradan ilmaisu (suunta ja PAIKKA-ilmaukset) on rajattua yleisempää. Video onkin rajattuuden suhteen erityisen monitulkintainen: siinä poika kävelee ranta-maisemassa kivisellä polkumuodostelmalla, joka koostuu kahdesta peräkkäisestä pienestä sillasta. Videolla näkyy vain yhden sillan ylitys, mutta toinenkin silta näkyy videolla pojan etenemissuunnassa. Video havainnollistaa hyvin eroja yksittäisten koehenkilöiden kielentämisstrategioissa. Esimerkissä 64 koehenkilö poimii videosta pelkän ylityksen, kun taas esimerkissä 65 ylityksen lisäksi huomioidaan liikkeen jatkuminen ja pyritään näin kattamaan koko tilanne.

64) *Poika ylittää sillan.* (trFi#15-050)

65) *Poika kävelee kivisillan yli ja jatkaa muuria myöten.* (trFi#10-050)

Esimerkissä 66 koehenkilö kuvaa liikeradan rajaamattomaksi pitkin-tyyppisellä grammilla ja tarkentaa *polun* sijaintia PAIKKA-ilmauksella *kivimuodostelmien päällä*. Lisäksi hän täydentää videolla nähdyn suhteellisen monimutkaisen liiketilanteen kuvausta toteamuksella *alhaalla virtaa vesi*. Esimerkissä 67 koehenkilö käyttää pelkästään PAIKKA-ilmauksia, jotka kuitenkin kuvaavat tilanteen melko yksityiskohtaisesti, kaikki keskeiset elementit huomioiden.

66) *Poika kävelee semmosta polkua pisin joka on kivimuodostelmien päällä ja alhaalla virtaa vesi.* (trFi#24-050)

67) *Poika kävelee veden ääressä tällaisen pienen kivistä rakennetun sillan päällä tai polun- kivipolun päällä.* (trFi#1-050)

Erityisesti videon 050 yhteydessä nähtävä vaihtelu *sillan yli-* ja *siltaa pitkin* -ilmauksissa nostaa esiin myös eroja *yli*-grammin käsitteistytksessä eri kiintopisteiden kanssa, mikä on jälleen esimerkki semanttisesta distribuutiosta. Lekseemin *silta* merkitys määrittää *yli*-grammin merkityksen erilaiseksi kuin muiden aineistossa esiintyvien lineaarista kulkureittiä kuvaavien lekseemien kanssa. Esimerkiksi ilmauksessa *tien yli* grammin *yli* merkitys hahmotetaan *poikki*-grammia vastaavaksi: muutuja kulkee tien kulkusuunnan vastaisesti, tien reunasta toiseen. Silta-kiintopiste kuitenkin muodostaa poikkeuksen, koska ilmauksella *sillan yli* kielennetään sillan kulkusuunnan mukainen sillan alusta loppuun kulkeminen. Ilmauksilla *sillan yli* ja *siltaa pitkin* ilmaistaan siis nimenomaan ero liiketilanteiden rajattuudessa.

Sama kiintopisteen vaikutus spatiaalisen lekseemin merkitykseen on luonnollisesti nähtävillä myös *ylittää*-verbin käytössä. Ilmaus *ylittää silta* käsitetään suhteessa kiintopisteeseen toisensuuntaiseksi liikkumiseksi kuin *ylittää tie*. Verbi *ylittää* esiintyy aineistossa valtaosin juuri näiden kahden kiintopisteen kanssa. Niiden lisäksi yhdessä kielennöksessä miehen loikka maassa olevan paksun tukin yli kielennetään *puun ylittämiseksi*. Selvästi eniten verbiä *ylittää* käytetään videosta 045, jossa naisjoukko kävelee sillan yli ja heidän etenemistään kuvataan takaapäin niin, että silta näkyy suoraan niin naisten kuin videon katsojankin edessä (esimerkki 68).

68) *Ystävykset ylittävät sillan.* (trFi#49-045)

Videon 045 kielensi *ylittää*-verbillä 19 koehenkilöä, kun muiden silta-videoiden kohdalla verbin osuus vaihtelee 0–5 esiintymän välillä. Vertailukohtien vähyys mahdollistaa kuitenkin vain alustavia arvioita siitä, miksi *ylittää*-verbi on niin yleinen juuri tässä videossa. Video 045 poikkeaa muista siltavideoista kuvakulmaltaan, ja on mahdollista, että suoraan eteen aukeava silta tuottaa enemmän *ylittää*-verbiä. Toisaalta muihin videoihin liittyy erityyppistä liiketilanteen kompleksisuutta: videoissa 046 ja 047 on toinen, ohitettava kiintopiste, ja videon 050 liiketilanteen rajattuus on tulkinnanvaraisempaa.



## KIIINTOPISTEENÄ PORTAAT

Elisitaatiokehyksessä on kuusi videota, joissa kulkureitti-kiintopistettä edustavat portaat. Kaikissa niistä näkyy portaiden kiipeäminen joko kokonaan tai lähes kokonaan (ks. taulukko 20). Neljässä tarkastelluista videoista on portaiden lisäksi selvästi näkyvillä toinen kiintopiste: Videoissa 022 ja 029 on mukana siirtymä luolaan portaiden lopussa tai ulos luolasta portaiden alussa, ja videoissa 073 ja 074 poistuminen veden ääreltä tai saapuminen veden äärelle. Samansuuntaista liikettä erilaisista kuvakulmista kuvaavat videot 070 ja 071 keskittyvät selvemmin portaissa kulkemiseen, mutta niissäkin on videon lopussa nähtävissä, että nainen saapuu jonnekin, mahdollisesti luolan päälle.

Taulukko 20. Väylän ja suunnan ilmaisukeinojen osuus videoista, joiden kiintopisteenä ovat portaat.

Video	Kuva- kulma	Väylä: totaali- objekti	Suunta: partitiivi	Suunta: <i>pitkin</i> - ilmaus	Suunta: partitiivi + <i>pitkin</i> - ilmaus	PAIKKA	Väylän ALKU tai LOPPU
022: Nainen kävelee portaita alas luolaan	edestä	7/50	28/50	11/50	–	–	4/50
029: Nainen kävelee portaita ylös luolasta	takaa	5/50	21/50	19/50	–	–	5/50
070: Nainen kävelee portaat ylös (luolan päälle)	edestä	3/50	31/50	13/50	2/50	–	1/50
071: Nainen kävelee portaat ylös (luolan päälle)	takaa	3/50	36/50	8/50	2/50	–	1/50
073: Nainen kävelee portaat alas veden äärelle	takaa	9/50	24/50	8/50	–	–	9/50
074: Nainen kävelee portaat ylös veden ääreltä	sivusta	5/50	29/50	6/50	1/50	–	9/50

Portaat mainitaan valtaosassa kielennöksiä, vaikka osassa videoista on mukana myös väylän ALUN tai LOPUN rajanylitys, joka aineistossa yle-

sesti tulee laajasti kielenneyiksi.<sup>46</sup> Taulukon 20 oikeanpuoleisessa sarakkeessa on niiden kielennösten osuus, joissa ei ole kielennetty väylän KESKIVAIHETTA, SUUNTA tai PAIKKAA vaan niiden sijaan väylän ALKU tai LOPPU, kuten luolaan saapuminen tai sieltä poistuminen. On todennäköistä, että portaat on kiintopisteenä niin silmiinpistävä eli salienti (*landmark salience*, esimerkiksi Caduff & Timpf 2008), että koehenkilöt huomioivat sen helposti myös kielennöksen tasolla. Portaat-kiintopiste siis erottuu aineiston muista kiintopisteistä siinä mielessä, että se tulee osassa kielennöksiä ilmaistuksi ainoana kiintopisteenä, vaikka videolla näkyy toiseen kiintopisteeseen liittyvä rajanylitys. Ero käy ilmi, kun porrasvideoita verrataan esimerkiksi videoon 027, jossa nainen poistuu metsästä ja lisäksi kuvataan pitkään naisen kävelyä nurmikolla. Kaikki koehenkilöt huomioivat naisen poistumisen metsästä. Osa huomioi lisäksi kielennöksessään nurmikolla kävelyn, mutta se ei yhdessäkään kielennöksessä syrjäytä rajanylityksen eli metsästä poistumisen ilmaisu. Porrasvideoista sen sijaan on tyypillistä esimerkin 69 tapaan mainita ainoastaan portaita pitkin eteneminen ja jättää luolasta poistuminen mainitsematta.

69) *Nainen astelee portaita pitkin.* (trFi#26-029)

Rajatun väylän KESKIVAIHEEN ilmaisukeinoa, totaaliobjektia (*Nainen nousee portaat*), käyttää pieni osa (3–9) koehenkilöistä kustakin videosta. Tyypillisimmin portaissa eteneminen ilmaistaan rajaamattomana suuntana. Kaikista videoista eniten käytetty (21–36 koehenkilöä) ilmaisukeino on *portaita*-tyyppinen partitiiviobjekti, vaikka useimmissa videoissa tilanne on selkeästi rajattu joko väylän ALUN tai LOPUN suhteen. Melko paljon on käytetty myös *pitkin*-ilmauksia (6–19 koehenkilöä). Muutamassa kielennöksessä on sekä *pitkin*-ilmaus että toista kiintopistettä käyttävä partitiiviobjekti, kuten esimerkissä 70. Sen sijaan *PAIKKA*-ilmauksia (*portaissa/portailla*) ei käytetä portaat-kiintopisteestä lainkaan.

46 Tarkastelua ei ole rajattu lekseemiin *portaat*, vaan mukaan on laskettu kaikki lekseemit, joilla on kielennetty videoilla näkyvä kulkureitti. Lekseemin *portaat* lisäksi mukana on siis sen lähisynonymejä, kuten *raput* ja *rappuset*, sekä yksittäisiä tapauksia, joissa luonnontilaiset, sammaloituneet portaat on tulkittu (*kivi*)poluksi.

70) *Mies juoksee polkua myöten mäkeä ylös.* (trFi#10-037)

Kiintopiste portaat on vertikaalinen, mikä tuo sen ilmaisun tarkasteluun mukaan vertikaalista suuntaa ilmaisevat kielenaineokset. Vertikaalisia kielenyksiköitä eli grammeja *ylös*, *alas*, *ylöspäin* ja *alaspäin* tai verbejä *nousta*, *laskeutua*, *kiivetä* ja *kivuta* käytetään 78,5 prosentissa porrasvideoiden kielennöksistä. Portaat ja muut siirtymät erityisesti rakennusten sisällä erottuvat myös Koulu- ja Unelmakorpuksessa keskeisenä kontekstina vertikaalisille grammeille, joita esiintyy kaikenikäisten teksteissä. Lap-silla tyypillinen konteksti näille grammeille on myös mäenlasku, johon kuuluu laskeminen mäkeä alas ja kiipeäminen takaisin ylös.

Seuraavaksi tarkastelen kiintopisteen portaat eri ilmaisukeinojen yhdistymistä vertikaalista suuntaa ilmaiseviin kielenaineksiin porrasvideoissa (ks. taulukko 21). Totaaliobjektilla ilmaistun väylän yhteydessä on vertikaalisen suunnan ilmaus yhtä lukuun ottamatta kaikissa kielennöksissä, ja monissa ilmaisu on jakaantunut verbin ja grammin kesken (esimerkki 71). Siinä yhdessä kielennöksessä (esimerkki 72), jossa vertikaalisuutta ei ilmaista, mukana on kuitenkin päätepisteen ilmaisu *ihan veden rajaan asti*. Totaaliobjektia ei siis käytetä ainoana liikeradan spesifikaationa, vaan sitä täydentää jokin muu väylän tai suunnan ilmaisu. Keksitty esimerkki 73 onkin intuitiivisesti hyväksyttävyyden rajamailla.<sup>47</sup>

71) *Nainen laskeutui portaat alas.* (trFi#27-022)

72) *Nainen kävelee kiviset raput ihan veden rajaan asti ja jää katselemaan veteen.* (trFi#18-073)

73) ?*Nainen kävelee portaat.*

Taulukko 21. Vertikaalisen suunnan ilmaisu videoissa, joissa kiintopisteenä on portaat.

Ilmaisukeino %	Vertikaalisen suunnan ilmaisu	Ei vertikaalisen suunnan ilmaisu
Totaaliobjekti (n = 32)	96,9	3,1
Partitiiviobjekti (n = 169)	92,3	7,7
<i>pitkin</i> -ilmaisu (n = 65)	55,3	44,6
Partitiiviobjekti + <i>pitkin</i> -ilmaisu (n = 5)	100	–

47 Totaaliobjektin *portaat* voisi ajatella muodostavan luontevan lauseen lähinnä erityisessä kontekstissa, kuten pitkistä portaista koostuvan liikuntasuorituksen yhteydessä (*Kävin juoksemassa Pispalan portaat*).

Totaali- ja partitiiviobjektin välillä ei ole suurta eroa yhdistymisessä vertikaalisiin kielenaineisiin. Partitiiviobjekti yhdistyy vertikaalisen suunnan ilmaisuun 92,3 prosentissa käytöistään. Edempänä pohdittiin grammien *ylös* ja *alas* asettumista väylän ja suunnan rajamaille, tapauskohtaisesti määrittyviksi. Grammien *ylöspäin* ja *alaspäin* todettiin edustavan *päin*-partikkelin ansiosta selvemmin suuntaa. Objektin sijan ja vertikaalisen grammin yhdistelmät asettuvat rajattuuden suhteen jatkumolle: Rajatuinta päätä edustaa totaaliobjektin ja lyhyen grammin yhdistelmä, kuten *portaas alas*, jossa totaaliobjekti on rajattu elementti ja sen ohella *alas*-grammi voi myös saada rajatun tulkinnan. Välvaiheille sijoittuu partitiiviobjekti lyhyen grammin kanssa, kuten ilmaus *portaita alas*, jossa *portaita* on rajaamaton ilmaus mutta *alas* on mahdollista tulkita joko rajatuksi tai rajaamattomaksi. Selvimmin rajaamattoman tulkinnan saa *portaita alaspäin* -tyyppinen rakenne, jossa sekä partitiiviobjekti että *päin*-päätteinen grammi ovat rajaamattomia elementtejä.

Kielennöksissä, joissa liikerataa ilmaisee *pitkin*-ilmaus, vertikaalisen suunnan ilmaisu ei ole yhtä yleistä kuin pelkän partitiivi- tai totaaliobjektin kanssa: vertikaalinen suunta on ilmaistu 55,3 prosentissa kielennöksistä (esimerkki 74). Lähes yhtä yleistä on siis jättää vertikaalinen suunta spesifioimatta esimerkin 69 tapaan.

74) *Nainen nousi pientä polkua pitkin ylös.* (trFi#5-070)

Syy vertikaalisten ilmausten suhteelliseen vähyyteen *pitkin*-ilmausten yhteydessä ei käy suoraan ilmi aineistosta, mutta on mahdollista, että kyse on kompleksisten rakenteiden välttämisestä: esimerkiksi *polkua ylös* on yksinkertaisempi ilmaisu kuin kahden grammin ketjuttaminen *polkua pitkin ylös* -rakenteeseen.

## Lopuksi

Tämän luvun tavoitteena on ollut luoda katsaus tiettyjen keskeisten spatioaalisten merkitysten, rajattua liikerataa edustavan väylän ja rajaamattomaa liikerataa edustavan suunnan, ilmaisuun suomen kielessä. Kehityksel-

linen variaatio on kontrolloitu rajaamalla koehenkilöjoukko nuoriin tai nuorehkoihin aikuisiin. Väylän KESKIVAIHEEN ja suunnan ilmaisua on tarkasteltu esittelemällä niiden ilmaisukeinoja laajassa elisitaatiokehyskellä kootussa liikeilmausten aineistossa. Ilmaisukeinoja on analysoitu semanttisesti ja niiden käyttöä on tarkasteltu erilaisia liiketilanteita kuvaavien videostimulusten yhteydessä. Aineiston erityyppisten videoiden tuottamaa ilmaisua tarkastelemalla on tarkennettu kuvaa siitä, miten väylän ja suunnan merkitykset rakentuvat suomessa ja minkätyyppiisiin liiketilanteisiin liittyy variaatiota suomen kielen puhujien kesken.

Spatiaalinen merkitys on tässä luvussa nähty holistisena: se jakaantuu eri kielenyksiköiden kesken ja rakentuu sekä kieliopillisten että leksikaalisten elementtien kautta. Tutkimuksessa tunnistettiin suomen liikkeen ilmaisussa käytetyistä kielenyksiköistä myös tilanteista variaatiota siinä, millaista merkitystä ne missäkin kontekstissa ilmaisevat. Esimerkiksi grammi *pois* ja vertikaaliset grammit *ylös* ja *alas* voivat liittyä vaihtelevasti sekä väylän että suunnan ilmaisuun, eikä raja merkitysten välillä ole aina ilmiselvä. Lisäksi kiintopisteenä toimiva lekseemi vaikuttaa keskeisesti väylägrammien ja -verbien merkitykseen spatiaalisen merkityksen distribuution (Sinha & Kuteva 1995) hengessä: esimerkiksi *sillan yli* ja *ylittää silta* kuvaavat erilaista liikerataa kuin *tien yli* ja *ylittää tie*.

Erityisesti väylän KESKIVAIHEESEEN keskittyvistä videoista nousee esiin suomen kielen laajojen, vaihtoehtoisten resurssien luoma variaatio. Suomessa KESKIVAIHEEN grammeja on useita, ja usein yhden liiketilanteen kuvaamiseen on monia eri vaihtoehtoja (esimerkiksi *poikki*, *halki*, *läpi* ja *yli*). Grammien variaatio näyttää olevan suomessa suurta myös typologisten yleistysten valossa (vrt. Papahagi 2011), ja lisäksi KESKIVAIHE voi tulla ilmaistuksi verbeissä tai objektin sijassa. Jatkotutkimuksen tehtäväksi jää verrata tätä ilmaisun aluetta yksityiskohtaisemmin suomen ja muiden kielten välillä. Myös suunnan ilmaisun keinoja on aineistossa runsaasti. Niiden tapauksessa moninaisuutta selittää osaltaan myös tilanteiden laaja kirjo: rajaamattomiin liiketilanteisiin kuuluu paitsi rajaamattomalla liikeradalla liikkumista (esimerkiksi *tietä pitkin*, *portaita*) myös loittonemista (esimerkiksi *luolasta pois päin*), lähestymistä (esimerkiksi *luolaa kohti*) sekä liikettä vertikaalisella (esimerkiksi *ylöspäin*), lateraalisella (esimerkiksi *oikealle*) ja frontaalisella (esimerkiksi *eteen-*

*päin*) akselilla. Myöhemmin on analysoitava, miten näin kompleksisessa ilmiössä osaaminen yhdenmukaistuu ja kielellisten resurssien käyttö monipuolistuu ja miten kehitys suhteutuu muiden spatiaalisten merkitysten, kuten väylän ALUN ja LOPUN ilmaisuun. Väylän KESKIVAIHEEN ja suunnan ilmaisua tarkasteltiin paitsi koko aineistosta myös tapaus- tutkimuksen omaisesti niistä videoista, joissa on kiintopisteenä ihmiselle tyypillinen kulkureitti. Niissä rajattu väylän KESKIVAIHE sekä rajaamattomat suunta- tai PAIKKA-ilmaukset hahmottuvat usein vaihtoehtoisiksi tavoiksi käsitteistää liiketilanne. Samoista videoista on käytetty esimerkiksi ilmauksia *sillan yli* ja *siltaa pitkin*, joista ensimmäinen ilmaisee tilanteen rajattuna ja jälkimmäinen rajaamattomana.

Variaatiota keskittyy tiettyihin videoihin, ja joissakin tapauksissa se liittyy selvästi monitulkintaisuuteen, kuten videossa 050, jossa kahden tai mahdollisesti useammankin sillan muodostama polkurakennelma on jokseenkin poikkeuksellinen ja vaikeasti kuvailtavissa. Eroja hahmottuu kuitenkin myös suhteessa kiintopisteen tyyppeihin. Esimerkiksi *pitkin-* tyyppisillä grammeilla ilmaisua liittyy odotuksenmukaisesti eniten videoihin, joiden kiintopisteenä on pitkä ja kapea polku. Kiintopisteen portaat yhteydessä taas esiintyy lähes poikkeuksetta vertikaalista suuntaa ilmaiseva elementti ja ylipäättään vahvaa suuntaisuutta sikäli, ettei PAIKKA-ilmauksia käytetä lainkaan videoista, joissa kulkureittinä käytetään portaita. Teiden, polkujen ja siltojen kanssa PAIKKA-ilmauksia esiintyy jonkin verran, vaikka ne pääasiassa liittyvätkin aineistossa laajempaa aluetta kuvaaviin kiintopisteisiin. Portaat näyttävät aineistossa myös niin huomion kiinnittävänä kiintopisteenä, että ne mainitaan lähes aina ja usein myös ainoana kiintopisteenä, vaikka videolla näkyisi tyypillisesti huomiota herättävä rajanylitystilanne. Kiintopisteiden salienssiin liittyvä variaatio jää odottamaan myöhempää, perusteellisempaa kartoitusta.

Tässä luvussa esitetty analyysi ei pyri yksiselitteisesti vastaamaan siihen, voiko suomen luokitella Talmyn (2000b) tapaan satelliittikehyskieleksi vai pitäisikö sen katsoa edustavan muuta tyyppiä, mahdollisesti yhdessä muiden sijakielten kanssa (ks. Naidu ja muut 2018). Tämän kysymyksen tarkastelu edellyttää myös liikkeen ilmaisun muiden osalueiden, kuten väylän ALUN ja LOPUN analyysiä (ks. Tuuri & Belliard, tulossa). Tästä tarkastelusta kuitenkin selviää esimerkiksi se, että väylän

KESKIVAIHEEN ja suunnan ilmaisussa tapaverbit ovat selvästi keskeisin verbiluokka, vaikka suomessa on myös tavan suhteen neutraaleja väyläverbejä, kuten *ylittää* ja *ohittaa*, sekä suuntaverbejä, kuten *nousta* ja *laskeutua*.

Tässä tutkimuksessa nouseva kuva liikeradan kielentämisestä suomessa tukee myös osaltaan liikeilmausten kieltenväliseen vertailuun perustuvia käsityksiä satelliittikehyskieliksi määriteltyjen kielten ilmaisun moninaisuudesta (esimerkiksi Slobin 2004; Ibarretxe-Antuñano 2009). Sen lisäksi, että satelliittikehystystä suosivat kielet ilmaisevat tapaa verbikehyskieliä monipuolisemmin (esimerkiksi Slobin 2006), ne sallivat tyypillisesti myös laajempaa väyläinformaation ilmaisua (ks. esimerkiksi Berman & Slobin 1994, 118–119; Pajunen 2001, 210). Ibarretxe-Antuñano (2009, 410) pitää laajoja leksikaalisia ja morfologisia resursseja yhtenä väylän korosteisuutta lisäävänä ja selittävänä tekijänä: esimerkiksi baskin yhdistelmä väyläverbejä, paikallissijoja, lokatiivisia adpositioita ja relationaalisia substantiiveja johtaa laajempaan väylän ilmaisuun kuin muissa perinteisesti verbikehyskieliksi määritellyissä kielissä. Tämän tarkastelun perusteella huomattavaksi nousee se, että suomen ilmaisuvarannot näyttäytyvät laajoina myös väylän KESKIVAIHEEN merkityksessä, jonka on esitetty olevan semanttiselta granulariteetiltaan kielissä tyypillisesti niukka (Papahagi 2011).

Aineiston rajaamisessa tiettyjen stimulusten avulla on sekä etunsa että haasteensa. Metodi mahdollistaa systemaattisen vertailun eri näkökulmista, esimerkiksi kielten, puhujajaksilöiden ja erilaisten stimulus-tyyppien välillä. Elisitoimalla saadaan kielennöksiä *juuri* niistä tilanteista, joita halutaan tarkastella. Samalla kuitenkin saadaan kielennöksiä *vain* niistä tilanteista, joita stimuluksiin on sisällytetty. Virikemateriaalin ominaisuudet saattavat vaikuttaa kielennöksiin osin ennakoimattomilla tavoilla; esimerkiksi Özçalışkan (2015) epäilee, että videoformaatti saattaa lisätä liikkumisen tavan kielentämistä. Elisitaatiotilanne on myös kielenkäyttötilanteena erikoislaatuinen, mikä saattaa joiltakin osin vaikuttaa kielennöksiin. Esimerkiksi *mennä*-verbin käyttö väylän KESKIVAIHEESSA (*mennä ohi* -tyyppiset rakenteet) on aineistossa vähäistä, mikä viittaa osaltaan siihen, että monella kielentäjällä elisitoitu aineisto edustaa arkipuhekieltä muodollisempaa rekisteriä. Kokonaiskuvan saavut-

tamiseksi tarvitaan siis myös muunlaista aineistoa. Toisaalta alustava vertailu Unelmakirjoitelmiin viittaa siihen, että elisitaatiotehtävä tuottaa suhteellisen todenmukaista kuvaa suomenkielisten nuorten aikuisten spatiaalisen sanaston käytöstä: lähes kaikki elisitoidussa aineistossa käytetyt väylän KESKIVAIHEEN grammit ovat mukana myös nuorten aikuisien kirjoitelma-aineistossa ja suunnan ilmaisut ovat niin ikään samoja. Alakouluikäisillä vastaavia grammeja on vähemmän, niitä käytetään harvemmin ja usein käyttö vaikuttaa vakiintumattomalta, joten väylän ja suunnan kielentämistä kannattaa tarkastella myös kehityksellisestä näkökulmasta.

## LÄHTEET

- Aske, J. 1989. Path predicates in English and Spanish: A closer look. *Proceedings of the 15th annual meeting of the Berkeley Linguistics Society*, 1–14.
- Berman, R. A. & Slobin, D. I. (toim.) 1994. *Relating events in narrative: A cross-linguistic developmental study*, Volume 1. Hillsdale: Lawrence Erlbaum Ass.
- Blomberg, J. 2014. *Motion in language and experience: Actual and non-actual motion in Swedish, French and Thai*. Lund: Lundin yliopisto.
- Bohnenmeyer, J., Donelson, K. T., Tucker, R. E., Benedicto, E., Capistrán Garza, A., Eggleston, A., Hernández Green, N., Hernández Gómez, M. de J. S., Herrera Castro, S., O'Meara, C. K., Palancar, E., Pérez Báez, G., Polian, G. & Romero Méndez, R. 2014. The cultural transmission of spatial cognition: Evidence from a large-scale study. *Proceedings of the 36th annual meeting of the Cognitive Science Society*, 212–217.
- Caduff, D. & Timpf, S. 2008. On the assessment of landmark salience for human navigation. *Cognitive Processing* 9, 249–267.
- Croft, W., Barðdal, J., Hollmann, W., Sotirova, V. & Taoka, C. 2010. Revising Talmy's typological classification of complex events. Julkaisussa: H. C. Boas (toim.) *Contrastive studies in construction grammar*. Amsterdam: John Benjamins, 201–235.
- Dasen, P. R. & Mishra, R. C. 2010. *Development of geocentric spatial language and cognition: An eco-cultural perspective*. Cambridge: Cambridge University Press.
- Egan, T. & Graedler, A.-L. 2015. Motion into and out of in English, French and Norwegian. *Nordic Journal of English Studies* 14: 1, 9–33.
- Frawley, W. 1992. *Linguistic Semantics*. Hillsdale: Lawrence Erlbaum Ass.
- Goschler, J. & Stefanowitsch, A. (toim.) 2013. *Variation and change in the coding of motion events*. Amsterdam: John Benjamins.
- Hakulinen, A., Vilkkuna, M., Korhonen, R., Koivisto, V., Heinonen, T. R. & Alho, I. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Heinämäki, O. 1984. Aspect in Finnish. Julkaisussa: C. de Groot & H. Tommola (toim.)



- Aspect bound: A voyage into the realm of Germanic, Slavonic and Finno-Ugrian aspectology.* Dordrecht: Foris Publications, 153–177.
- Huumo, T. 2010. Suomen väyläadpositioiden prepositio- ja postpositiokäyttöjen merkityseroista. *Viritäjä* 114: 4, 531–561.
- Ibarretxe-Antuñano, I. 2009. Path salience in motion events. Julkaisussa: G. Jiansheng, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura & Ş. Özçalışkan (toim.) *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin*. New York: Psychology Press, 403–414.
- Imbert, C., Grinewald, C. & Sörös, A. 2011. Pour une catégorie de "satellite" de Trajectoire dans une approche fonctionnelle-typologique. *Faits de Langues. Les Cahiers: Revue de Linguistique* 3, 99–116.
- Ishibashi, M., Kopecka, A. & Vuillemet, M. 2006. *Trajectoire: Matériel visuel pour élicitation des données linguistiques*. Fédération de Recherche en Typologie et Universaux Linguistiques, CRNS.
- Kielitoimiston sanakirja*. 2018. Helsinki: Kotimaisten kielten keskuksen verkkojulkaisu 35. Päivitetty julkaisu. Päivitetty 6.6.2018. URN:NBN:fi:kotus-201433. Viitattu 30.3.2020.
- Kopecka, A. 2009. L'expression du déplacement en français: l'interaction de facteurs sémantiques, aspectuels et pragmatiques dans la construction du sens spatial. *Langages* 173, 54–77.
- Kopecka, A. & Ishibashi, M. 2011. L'(a-)symétrie dans l'expression de la Source et du But: Perspective translinguistique. *Faits de Langues. Les Cahiers: Revue de Linguistique* 3, 131–149.
- Langacker, R. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford: Stanford University Press.
- Larjavaara, M. 2019. *Partitiivin valinta*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Lauerma, P. 1990. Spatiaalinen kenttä. Julkaisussa: P. Leino, M.-L. Helasvuo, P. Lauerma, U. Nikanne & T. Onikki (toim.) *Suomen kielen paikallissijat konseptuaalisessa semantiikassa*. Helsinki: Helsingin yliopisto, 108–145.
- Lehismets, K. 2013. *Yli, läpi ja kautta*. Suomen grammirakenteiden voimadynamiikkaa. *Viritäjä* 117: 3, 346–371.
- Lehismets, K. 2014. Suomen kielen väylää ilmaisevien adpositioiden yli, läpi, kautta ja pitkin semantiikkaa. Tartto: Tartu Ülikooli Kirjastus.
- Leino, P. 1991. *Lauseet ja tilanteet: Suomen objektin ongelmia*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Leino, P., Helasvuo, M.-L., Lauerma, P., Nikanne, U. & Onikki, T. 1990. *Suomen kielen paikallissijat konseptuaalisessa semantiikassa*. Helsinki: Helsingin yliopisto.
- Levinson, S. C., Meira, S. & The Language and Cognition Group. 2003. 'Natural concepts' in the spatial topological domain – adpositional meanings in crosslinguistic perspective. An exercise in semantic typology. *Language* 79, 485–516.
- Mayer, M. 1969. *Frog, where are you?* New York: Dial Press.
- Meneghetti, C., Pazzaglia, F. & De Beni, R. 2011. Spatial mental representations derived from survey and route descriptions: When individuals prefer extrinsic frame of reference. *Learning and Individual Differences* 21, 150–157.

- Miller, G. A. & Johnson-Laird, P. N. 1976. *Language and perception*. Cambridge: Harvard University Press.
- Naidu, V., Zlatev, J., Duggirala, V., Van De Weijer, J., Devylder, S. & Blomberg, J. 2018. Holistic spatial semantics and post-Talmian motion event typology: A case study of Thai and Telugu. *Cognitive Semiotics* 11: 2, 1–27.
- Nikanne, U. 1990. *Zones and tiers: A study of thematics structure*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Nurminen, S. 2015. *Lauseen aspektin ilmaiseminen suomen kielessä: Tutkimus olla-, tehdä- ja tulla-verbillisistä tekstilauseista käyttökonteksteissaan*. Turku: Turun yliopisto.
- Ojutkangas, K. 2003. Suuntaa ilmaisevat adverbiaalit, suhdesääntö ja käsin: ”ikkunointiin” perustuva analyysi. *Virtittäjä* 107: 2, 162–184.
- Ojutkangas, K. 2012. Ikkuna ulos. Suomen ulko-grammit ja huomion ikkunoiminen. *Virtittäjä* 116: 2, 169–197.
- Pajunen, A. 2001. *Argumenttirakenne: Asiantilojen luokitus ja verbien käyttäytyminen suomen kielessä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Pajunen, A., Puranen, M. & Yli-Paavola, A. 2010. *Koulukorpus. Alakoululaisten kirjoitelmia aiheesta Unelmieni päivä*. Excel-tietokanta.
- Pajunen, A. 2016. *Unelmakorpus. Peruskoululaisten ja nuorten aikuisten kirjoitelmia aiheesta Unelmieni päivä*. Excel-tietokanta.
- Pajunen, A. & Virtanen, P. 2002. Helsingin Sanomat 2000–2001. Morfosyntaktisesti analysoitu tietokanta, 24 miljoonaa sananmuotoa.
- Palmer, B., Lum, J., Schlossberg, J. & Gaby, A. 2017. How does the environment shape spatial language? Evidence for sociotopography. *Linguistic Typology* 21: 3, 457–491.
- Papahagi, C. 2011. Pour une typologie des systèmes adnominaux. *Faits de Langues. Les Cahiers: Revue de Linguistique* 3, 117–130.
- Salmi, T. 1994. Spatiaalinen ja metaforinen yli. Julkaisussa: P. Leino & T. Onikki (toim.) *Näkökulmia polysemiaan. Suomen kielen kognitiivista kielioppia* 2. Helsinki: Helsingin yliopisto, 164–188.
- Seppälä, P. 2010. *Kautta*-postposition polysemiaa. *Sananjalka* 52, 46–66.
- Sinha, C. & Kuteva, T. 1995. Distributed spatial semantics. *Nordic Journal of Linguistics* 18: 2, 167–199.
- Sivonen, J. 2005. *Mutkia matkassa: Nykysuomen epäsuoraa reittiä ilmaisevien verbien kognitiivista semantiikkaa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Slobin, D. I. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. Julkaisussa: S. Strömquist & L. Verhoeven (toim.) *Relating events in narrative. Volume 2: Typological and contextual perspectives*. Mahwah: Lawrence Erlbaum Ass., 219–257.
- Slobin, D. I. 2006. What makes manner of motion salient? Julkaisussa: M. Hickmann & S. Robert (toim.) *Space in languages: Linguistic systems and cognitive categories*. Amsterdam: John Benjamins, 61–81.
- Slobin, D. I. & Hoiting, N. 1994. Reference to movement in spoken and signed languages: Typological considerations. *Proceedings of the twentieth annual meeting of the Berkeley Linguistics Society: General session dedicated to the contributions of Charles J. Fillmore*, 487–505.

- Sloetjes, H. & Wittenburg, P. 2008. Annotation by category – ELAN and ISO DCR. *Proceedings of the 6th international conference on language resources and evaluation*.
- Strömquist, S. & Verhoeven, L. (toim.) 2004. *Relating events in narrative: Typological and contextual perspectives*. Mahwah: Lawrence Erlbaum Ass.
- Svorou, S. 1994. *The grammar of space*. Amsterdam: John Benjamins.
- Talmy, L. 1975. Figure and ground in complex sentences. *Proceedings of the first annual meeting of the Berkeley Linguistics Society*, 419–430.
- Talmy, L. 1985. Lexicalization patterns: Semantic structure in lexical forms. Julkaisussa: T. Shopen (toim.) *Language typology and syntactic description. Volume 3: Grammatical categories and the lexicon*. Cambridge: Cambridge University Press, 57–149.
- Talmy, L. 2000a. *Toward a cognitive semantics Volume 1: Concept structuring systems*. Cambridge: The MIT Press.
- Talmy, L. 2000b. *Toward a cognitive semantics. Volume 2: Typology and process in concept structuring*. Cambridge: The MIT Press.
- Tuuri, E. Tulossa. Concerning variation in encoding spatial motion. Evidence from Finnish. *Nordic Journal of Linguistics*.
- Tuuri, E. & Belliard, M. Tulossa. Liikkeen kielentämisen variaatio lähisukielissä: suomen ja viron väljän ilmaisu. *Finnish Journal of Linguistics*.
- van Staden, M. & Narasimhan, B. 2012. Granularity in the cross-linguistic encoding of motion and location. Julkaisussa: M. Vulchanova & E. van der Zee (toim.) *Motion encoding in language and space*. Oxford: Oxford University Press, 134–148.
- Vendler, Z. 1967. *Linguistics in philosophy*. Ithaca: Cornell University Press.
- Viljanen, M. 2012. 8–9-vuotiaiden S1- ja S2-oppilaiden syntaktisten valmiuksien kehittyminen. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Wolbers, T. & Hegarty, M. 2010. What determines our navigational abilities? *Trends in Cognitive Science* 14, 138–146.
- Zlatev, J. 2003. Holistic spatial semantics of Thai. Julkaisussa: G. B. Palmer & E. H. Casad (toim.) *Linguistics and non-Indo-European languages*. Berlin: Mouton de Gruyter, 305–336.
- Zlatev, J. 2007. Spatial semantics. Julkaisussa: H. Cuyckens & D. Geeraerts (toim.) *The Oxford handbook of cognitive linguistics*. Oxford: Oxford University Press, 318–350.
- Zlatev, J. & Yangklang, P. 2004. A third way to travel: The place of Thai in motion-event typology. Julkaisussa: S. Strömquist & L. Verhoeven (toim.) *Relating events in narrative. Volume 2: Typological and contextual perspectives*. Mahwah: Lawrence Erlbaum Ass., 159–190.
- Zlatev, J., David, C. & Blomberg, J. 2010. Translocation, language and the categorization of experience. Julkaisussa: V. Evans & P. Chilton (toim.) *Language, cognition and space: The state of the art and new directions*. Sheffield: Equinox, 389–418.
- Özçalışkan, Ş. 2015. Ways of crossing a spatial boundary in typologically distinct languages. *Applied Psycholinguistics* 36, 485–508.

# Referentiaaliset määräiset ilmaukset ja niiden suhde kertomusrakenteeseen alakouluiässä

*Anna-Kaisa Kouvo*

*Mari Honko*

 <https://orcid.org/0000-0002-3240-622X>

*Anneli Pajunen*

 <https://orcid.org/0000-0001-8597-7491>

## Johdanto

Kertominen on ihmismielelle ominainen tapa kuvata tapahtumia. Kerroksen tuloksena syntyy teksti, kertomus, joka esittää kokemamme tai kuvittelemamme tapahtumat toisiinsa liittyvinä tapahtumaketjun osina. Juuri kertomusten avulla ihminen voi ymmärtää, muistaa ja jakaa niin kokemuksia kuin tapahtumiakin (Pietikäinen & Mäntynen 2009, 105). Samaistumisen edellytyksenä on kuitenkin, että teksti on myös sen vastaanottajalle ymmärrettävä eli koherentti. Koherentti, yhtenäinen kertova teksti syntyy monen tekijän summana, mutta perusedellytys on, että tilanne, tapahtumat ja henkilöt ankkuroidaan diskurssiin. Tämä tarkoittaa, että tapahtumia kielentävät lauseet on suhteutettava temaattisesti, ajallisesti ja paikallisesti sekä toiminnallisesti – tekstillä on toisin sanoen oltava tietty teema, ja se tulee sijoittaa aikaan ja paikkaan.

Koherentissa kertomuksessa tapahtumat kerrotaan suhteessa valittuihin henkilöihin. Myös henkilöt on aluksi esiteltävä diskurssiin ja sen jälkeen merkittävä tekstissä tunnistettavasti. Koherentissa kertomuksessa tapahtumat eivät etene sattumanvaraisesti vaan kielentyvät yleensä sarjana joko kertomusrakenteen tai juonirungon eli skriptin mukaisesti (Hickmann 2004; Givón 2017, 57–64). Tapahtumat asetetaan tärkeysjärjestykseen, mitä voidaan havainnoida esimerkiksi tutkimalla virkkeiden tai T-yksiköiden rakennetta eli sitä, miten yhteen rakenneyksikköön kielennetään päätapahtuman lisäksi sitä taustoittavat, perustelevat tai muut vastaavat tapahtumat (ks. T-yksiköistä esim. Schneider, Vis Dubé & Hayward 2005). Koko tapahtumasarja voidaan puolestaan esittää joko tapahtumajärjestyksen mukaisesti, takaumin, tiivistetysti tai yksityiskohteisesti. Kertomuksen kirjoittaja voi siis pakata kokonaisuuden siten, että informaatio välittyy vastaanottajalle halutuin painoituksin ja että tuttu ja uusi informaatio erottuvat toisistaan.

Yksi oleellinen koherenssia luova tekijä on juuri referentiaalisten eli viittaavien ilmauksien käyttö. Niiden avulla viestin vastaanottaja pystyy identifioimaan puheena olevat tarkoitteet kaikkien vaihtoehtojen joukosta oikein: hän tietää tai päättelee, kuka tekee mitään ja kenelle jotain tapahtuu. Tarkoite, diskurssireferenti, tuodaan tekstiin esimerkiksi esitelyrakennetta käyttäen ja siihen viitataan jatkossa käyttäen pronomineja tai referentiaalisia määräisiä ilmauksia (ks. Halliday & Hasan 1976; Bamberg 1983; Hickmann 2004; Habermas & de Silveira 2008; Givón 2017). Tämän luvun teemana on analysoida, miten koherenssi syntyy kehityksellisessä kontekstissa juuri referentiaalisten ilmauksien ja kertonvan tekstin yhteistyönä. Tarkastelu rajautuu erityisesti kontekstissaan määräisiin ilmauksiin. Aineistona on Koulukorpuksen Unelmakirjoitelmia (ks. Pajunen, Puranen & Yli-Paavola 2010; ks. myös tämän teoksen johdantoa).

Semanttis-pragmaattisena ilmiönä koherenssia ilmaistaan viittaus-suhteiden lisäksi monilla muillakin keinoilla (Halliday & Hasan 1976), ja sitä voidaan tutkia niin kielellisen muodon ja merkityksen kannalta kuin kehityksellisestä näkökulmasta eli keskittyen koherenssin esiintyvyyteen autenttisissa, eri-ikäisten kirjoittajien kirjoittamissa teksteissä. Jo se, miten esimerkiksi referentiaalinen koherenssi määritellään – yh-

den tekstiyksikön tulkinta riippuu jostakin toisesta yksiköstä – osoittaa, että koherenssi joka tapauksessa on tekstitason eli tapahtumien kielen-tämisen tason ilmiö.

Informaation kulkua tekstissä voidaan osoittaa tekstireferenttien tois-tuvuudella eli muun muassa topiikin jatkuvuuden avulla. Topiikin jatku-vuus kertovassa tekstissä osoitetaan tyypillisesti kertojaan tai henkilöön viittaavilla persoonapronomineilla (Genette 1980). Viittausketjujen pi-tuutta rajaa ymmärrettävyys ja toistoa tarvitaan erityisesti jaksojen raja-kohdissa. Pitkät viittausketjut ovat ominaisia kertovalle tekstille, mutta erittelevässä tekstissä koherenssi nojaa enemmän leksikaalisiin keinoi-hin. Informaation kulku käy ilmi myös diskurssireferenttien merkin-nästä, jota tässä yhteydessä tutkitaan. Ilmiön moninaisuus antaa olettaa, että koherenttien ilmausten ja tekstien tuottaminen ja ymmärtäminen on kehityksellisesti hidasta; tosin varsinainen tutkimustieto suomen osalta vielä pääosin puuttuu.

Koherentin tekstin tuottaminen edellyttää vastaanottajan huomioon ottamista. Tekstin tuottajalla ja vastaanottajalla voi olla aiempaa tai muu-ten yhteistä tietoa tai sitten tuottaja joutuu päättämään, mitä vastaan-ottaja tietää tai voi päätellä kontekstin, tilanteen tai sosiaalis-kulttuurisen tiedon perusteella. Jos vastaanottaja on tekstin tuottajalle läheinen, teh-tävä on helpompi kuin jos hän on etäinen tai jos vastaanottajana on heterogeeninen ryhmä. Kehityksellisesti tilanne on vielä hankalampi. Pikkulapset pystyvät suullisesti kertomaan koherentteja tarinoita jo reip-paasti alle kouluikäisinä (Suvanto 2012), mutta ne kerrotaan yleensä tutuille kuulijoille ja sanoman välittämiseen voidaan käyttää myös ei-ver-baalaisia keinoja kuten ilmeitä ja osoituksia. Lisäksi osaamiseen vaikut-taa opittu malli eli esimerkiksi kuultujen satujen ja tarinoiden määrä. Tarinoiden siirtäminen kirjoitettuun muotoon tuo monia lisävaikeuksia ja -vaatimuksia alkaen motorisista valmiuksista aina kognitiiviseen kyp-syyteen (Hoff 2009). Erityisesti tekstin tuottajalla pitää olla kyky kuvi-tella, mitä vastapuoli tietää ja ymmärtää, näkee tai kokee. Hänellä toisin sanoen pitää jo olla kehittynyt mielen teoria (ks. esimerkiksi Astington & Baird 2005).

Unelmakirjoitelmat edustavat annetusta teemasta vapaasti kertomis-ta, vaikka aiheen raja-kaus yhteen päivään (Unelmieni päivä) antaa kirjoitta-

jalle mahdollisuuden tukeutua päivä-skriptiin. Aineistosta on tutkimusta varten poimittu alakoulun jokaiselta luokalta 8000 tekstisanaa. Tutkimus rajautuu alakouluikäisiin, koska voi olettaa, että koherentin tekstin tuottaminen on juuri heillä vielä vaiheessa (Pajunen 2012). Aineiston kokonaissanamäärä on 40000 sanaa (eli 60 % ko. luokkien teksteistä), mutta analyysi keskittyy erityisesti toisen, neljännen ja kuudennen luokan aineistoon, jotta kehitykselliset erot näkyvät selvästi.<sup>48</sup> Teksteistä on analysoitu funktioltaan referentiaalisiksi arvioitavat ilmaukset.

Tutkimus etenee seuraavasti. Ensin luodaan tutkimuksen teoreettinen konteksti ja määritellään keskeiset käsitteet. Pohdimme diskurssi-referentteinä toimivien tarkoitteiden jakoa tuttua ja uutta informaatiota ilmaiseviin. Lisäksi tarkastelemme referentiaalisen koherenssin ymmärtämis- ja tuottamistutkimusta kehityskontekstissa sekä kertovan tekstin luonnetta ja kehitystä. Luvun empiirisessä osiossa käsittelemme alakouluikäisten referentiaalisen koherenssin ilmaisun kehittymistä määraisten ilmauksien näkökulmasta. Oheisjuonteena kulkee puhutussa suomessa melko tavallisen artikkelinomaisen *se*-tarkenteella merkityn lausekkeen esiintyminen lasten kirjoitetussa kielessä. Oletuksena on, että alakouluikäisillä on vaikeuksia referentiaalisen koherenssin ilmaisussa, erityisesti päätellyn tai oletetun tuttuuden osalta, ja että kehitys suhteutuu kertomusrakenteen hallintaan saamiseen. Puheessa alakouluikäiset lapset sekä ymmärtävät että tuottavat korrekkeja viittaussuhteita mutta etenkin luovassa kirjoittamisessa kehitys on oletettavasti vielä kesken.

## Referentiaaliset ilmaukset ja kertova teksti

Kehityksellisessä kontekstissa referentiaaliset, tunnistettaviin tarkoitteisiin viittaavat ilmaukset synnyttävät koherenttia tekstiä vaiheittain. Tekstit ovat nuorimmilla kirjoittajilla aluksi vailla koherenssia, sitten osittain koherentteja ja vähitellen syntyy koherentti kertomus. Vastaavasti refe-

<sup>48</sup> Luokkakohtaisen otoksen koko määräytyi 2. luokan aineiston sanamäärästä. Ylemmiltä luokilta valittiin juoksevassa järjestyksessä yhtä monta tyttöjen ja poikien kirjoitelmaa. 1. luokan aineistossa oli paljon yhden sanan kirjoitelmia, joten sitä ei kannattanut sisällyttää analyysiin.

referentiaalisten ilmausten tuttuus on moninaista ja aluksi monilla ilmaisu-keinoilla toteutettavissa. Lapsen kyky tulkita ilmauksien koherenssia on vaiheittaista aivan samoin kuin sen tuottaminenkin. Oletuksemme onkin, että vaikka kertova teksti edellyttää koherenssia luovia referentiaalisia ilmauksia, kehitys on hidasta ja alakouluikäisillä vielä kesken-eräinen. Määrittelemme tässä luvussa analyysissä tarvittavat keskeiset käsitteet tarkemmin.

## TUTTU JA UUSI INFORMAATIO

Referentiaalisten ilmauksien eli tekstin diskurssireferentteihin viittaavien ilmauksien ymmärtäminen edellyttää viestin vastaanottajalta maailmantietoa, kontekstin ymmärrystä ja tekstin tulkintaa. Jos aihepiiri ja sisältö ovat vastaanottajalle kokonaan vieraita, hän ei näe toisen kannalta täysin koherenttia tekstiä koherentiksi, ei irrota siitä samaa merkitystä eikä huomaa sen sävyjä. Esimerkiksi 2020-luvun nuorten toisin kuin vanhempien ihmisten voi olla vaikea ymmärtää, mitä eroa on viittauksella vaikkapa 1950-luvun tai 1970-luvun Kekkos-aikaan, koska ymmärtäminen edellyttää paitsi kyseisen henkilön myös näiden vuosikymmenten yhteiskunnallis-poliittisten tapahtumien ja ennen kaikkea ilmapiirin tuntemusta. Vähimmällä ne ymmärtää ajanjaksot kokenut.

Koherenttiuden aste voi vaihdella myös muista syistä kuin vastaanottajan tiedon vaihtelun takia. Tekstin tuottaja voi esimerkiksi olla vielä sellaisessa kehityksellisessä vaiheessa, että hän ei hallitse koherenssin tuottamisen kielellisiä keinoja. Bamberg (1983) asettaa tekstikoherenssin asteikolle täydestä koherenssista osittaiseen, epäkoherenttiin, ei-ymmärrettävään ja edelleen aina tekstittömyyteen (*unscorable*) asti. Täysin koherentti teksti tunnistaa topiikin, pysyy aiheessa, orientoi lukijaa ja kertoo johdonmukaisessa järjestyksessä. Epäkoherentti teksti listaa aiheita nostamatta mitään niistä keskeiseksi, ei anna lukijalle orientoivaa tietoa jne. Yhden lauseen tai virkkeen tuotos ei muodosta tekstiä lainkaan. Gundel, Hedberg & Zacharski (1993) asettavat referentiaaliset ilmaukset annettuusasteikolle (*givenness scale*), jossa tietty kognitiivinen status merkitään tietyllä kielellisellä keinolla, esimerkiksi tuttuun referenttiin viittaava ilmaus merkitään artikkelikielissä määräisellä artikkelilla. Alakoululaiset todennäköisesti vasta tavoittelevat koherenttien viittausten tuottamista



ja niiden ilmaisemisen korrektia kielellistä keinoa. Voi olettaa, että nuorimpien alakoululaisten tekstit ovat koherenssiasteikon alapäässä ja että ne vähitellen kehittyvät iän mukaan asteikon yläpäästä tavoitteleviksi.

Referentiaalisen koherenssin kannalta oletus on, että koherenssiasteikon yläpäässä tekstin osallistujat esitellään ja heihin viitataan tunnistetavasti, epäkoherentissa tekstissä saatetaan esitellä useita osallistujia, mutta heistä ei ole mainintaa enempää sanottavaa. Myös koherenssin ja kertomusrakenteen kannalta oleellinen vastaanottajan orientointi jää heikosti koherenteissa teksteissä vähiin. Epäkoherentti kertova teksti voi silti olla vaihtelevasti ymmärrettävä. Ensinnäkin kertova teksti on ylipäättään helpompi tulkita kuin erittelevä, ja tapahtumien osittainenkin ankkurointi aikaan ja paikkaan auttaa hahmottamaan tekstiä. Unelmakirjoitelmien kontekstissa sekä kirjoittajan ikä että kirjoitelman aihe antavat lukijalle monia tulkintavihjeitä. On esimerkiksi odotuksenmukaista, että nuoren kirjoittajan maailmassa perheenjäsenillä ja koululla on suuri rooli, joten lukija todennäköisesti pyrkii identifioimaan diskurssin henkilöitä perheenjäseniin tai luokkatovereihin, vaikka heitä ei erikseen esiteltäisi tai viittaussuhteita merkittäisi kognitiivista statusta vastaavalla tavalla.

Referentiaaliset ilmaukset viittaavat siis joko tuttuun tai uuteen tietoon. Edellinen ilmaistaan esimerkiksi artikkelikielissä määräisillä muodoilla tai pronomineilla, jälkimmäinen epämääräisillä artikkeleilla. Lähtökohtaisesti perusjako tuttu ja uusi vaikuttaa helpolta ymmärtää, mutta tosiasiaassa erityisesti tuttuuden käsite on kirjallisuudessa saanut monia eri nimityksiä ja vaihtelevia tulkintoja. Ero selitti vanhemmassa alan kirjallisuudessa lähinnä syntaktisten rakenteiden käyttöeroja ja sanajärjestysvaihteluita, mutta tekstin pragmaattiseen tulkintaan ei kiinnitetty huomiota (ks. esimerkiksi Chafe 1976; 1987; 1997). Prince (1981, 225–231; 1992) oli ensimmäisiä, joka kiinnitti huomiota siihen, että tuttuuden ja uutuuden vastakohtaisuutta (*old–new, known–new, presupposition–focus, given–new*) ei ollut riittävästi määritelty ja että tuttuustermillä (*given*) saattoi olla ainakin kolme eri merkitystä. Ensinnä kysymys oli diskurssireferentin tunnistamisesta edeltävän kontekstin perusteella (*predictability/recoverability*), toiseksi salienssista eli jossain suhteessa huomion herättämisestä tai ilmeisyydestä ja kolmanneksi yhteisestä tie-

dosta (*shared knowledge*) tai oletetusta tuttuudesta (*assumed familiarity*). Prince osoitti myös, että nämä tekijät eivät ole itsenäisiä vaan toisistaan riippuvia ja monella tavoin tekstin tuottajan olettamuksiin liittyviä.

Pääteltävä tuttuus on tuttuuden tyypeistä hankalin, ja Prince pyrki antamaan sille useita tulkintoja, joiden mukaan puhuja ensinnäkin olettaa, että kuulija voi tai voisi ennustaa tietyn kielellisen yksikön esiintyvän tietyssä ilmauksessa ja tietyssä kohtaa lausetta. Toiseksi puhuja olettaa, että kuulijalle tulisi mieleen jokin tietty entiteetti ilmaisun kuullessaan, ja kolmanneksi, että kuulija tietää, olettaa tai voi päätellä tietyn asian (sitä välttämättä tietoisesti ajattelematta). Prince (1981, 237) on luonut tutun ja uuden laajalti tunnetun hierarkian (ks. taulukko 22), jossa osa tarkoitteista on täysin uusia, joko kontekstiin ankkuroimattomia tai tavalla tai toisella ankkuroituja. Osa on pääteltäviä siten, että ne joko sisältävät laajempaa tietoa tai eivät sisällä. Osa on esimerkiksi teksti- tai tilanneassosiaatioiden kautta aktivoituvia ja siten mahdollisesti tunnistettavia.

Taulukko 22. Princen tuttuustaksonomia (1981, 237).

Oletettu tuttuus	Alaluokka	Kiinnitys
Uusi	Täysin uusi	Ankkuroimaton
	Täysin uusi	Ankkuroitu
Pääteltävä	Sisällymätön	
	Sisältyvä	
Aktivoitu ( <i>evoked</i> )	Tekstuaalisesti	
	Tilanteisesti	

Vaikeasti arvioitavaa pääteltävyyttä on myöhemmässä kirjallisuudessa pyritty täsmentämään (ks. esimerkiksi Birner 2006; Loock 2013). Kuten luvun alussa jo todettiin esimerkin avulla, pääteltävyys on ongelmallista muun muassa, koska sen sisältyvyys voi olla rajaamatonta. Siten voi olla vaikea nähdä, miten laajoja yleistyksiä milloinkin voisi tai pitäisi tehdä. Abstraktikäsitteiden osalta pääteltävyys on erityisen hankalaa, koska jo niiden merkitys on kulttuurinen ja ehkä vain rajatusti yhteinen (Loock 2013). Tekstin tuottajalla ja sen lukijalla voi myös olla hyvin erilainen käsitys entiteetin tuttuudesta, ja tämä ongelma on tavallaan ratkaisematon: tuottajan ja vastaanottajan teksti voi aina poiketa toisistaan (Sperber &

Wilson 1986). Birner (2006) on ehdottanut pääteltävien jaottelua vastaanottajan kannalta tuttuihin ja uusiin. Loock (2013) on ehdottanut, että ongelmaa voi kuroa umpeen nojaamalla ns. kohteliaisuusoletukseen (vrt. Brown & Levinson 1987), jolloin viestin vastaanottaja tekisi eräänlaisia hyväntahtoisia tulkintoja, yrittäisi seurata tekstin tuottajaa parhaansa mukaan ja ymmärtämään pyrkivästi. Toisaalta tekstin tuottaja voi tosin olla tarkoituksella epäkohtelias tehdessään vastaanottajan tiedosta oletuksia ja vastaanottaja vastaavasti vastaanottamaton. Yksi tapa olisi Loockin (2013) mukaan lisätä Princen luokitukseen vastaanottajan määrittelemättömät tai ratkaisemattomat tapaukset eli kaikki sellainen, minkä sisältöä ei voi varmasti päätellä. Loockin mukaan nämä määrittelemättömät tapaukset pyritään tekstiin merkitsemään tietyillä rakenteilla lisäämällä esimerkiksi määritteiden tai parenteesien määrää. Ne olisivat siten kielellisesti tunnistettavissa. Gundelin ja muiden (1993) annettuusasteikossa (*givenness scale*) on kuusi vastaanottajan mieleen oletettuun statukseen perustuvaa tilaa, jotka ovat implikatiivisessa suhteessa: fokuksessa, aktivoitunut, tuttu (*familiar*), täysin identifioitavissa (*uniquely identifiable*), referentiaalinen ja tyypiltään tunnistettavissa (*type identifiable*). Tilat loisivat Gundelin ja muiden mukaan riittävät edellytykset spesifin referoivan ilmauksen tunnistamiseen tai saavutettavuuteen. Kehittyvässä kirjoittamisessa hierarkia näyttäytyy kuitenkin vielä selvästi osittaisena.

Lähtökohtaisesti onkin ajateltavissa, että erityisesti lapsilla olisi ongelmia yhteiseksi oletetussa tuttuudessa, että heillä olisi yllädotuksia vastaanottajan tiedosta tai päättelykyvystä ja että he eivät hallitsisi kaikkia tarvittavia kielellisiä keinoja. Toinen vaihtoehto on luonnollisesti, että nuorimmat kirjoittamista vasta opettelevat eivät tule edes ajatelleeksi, että heidän pitäisi kirjoittaessaan ottaa joku lukijakin huomioon. Kertomusrakennemalleissa oletetaan yhdeksi mallin osioksi tekstin henkilöiden esittely tai yleensä orientaatio. Unelmakirjoitelmissa orientaatiojako on vielä viidesluokkalaisillakin melko harvoin mukana, mutta tekstin etukäteissuunnittelu nostaa sen esiintyvyyttä siten, että suunnitelluissa viidesluokkalaisten teksteissä noin 60 prosentissa on esittelevä osio (ks. Sintonen 2011). Toisin sanoen alakouluikäinen voi hyvin tuoda lukijalle uutta tietoa diskurssiin ilman esittelyä tai aikuisten käyttämiä merkintä-

keinoja. Hän voi myös olettaa itselleen tutun yhteiseksi tiedoksi tai muuten tutuksi. Perheenjäsenten lisäksi luokkatoverit ja ystävät ilmestyvät tekstiin usein ilman esittelyä. Oletus siis on, että nuorimmat kirjoittajat ilmaisevat oletettua tuttuutta ankkuroimalla sitä tilannekontekstiin ja kielenulkoiseen kontekstiin. Lukijan pitäisi siten vastaavasti rakentaa ymmärtämisensä kielenulkoiseen ja kulttuuriseen tietoon.

Jossain määrin on avoin kysymys, kirjoittaako tekstin tuottaja juuri ja vain lukijaa huomioiden erottaessaan tuttua ja uutta informaatiota kielellisin keinoin, vai vaikuttaako hänen tuotukseensa myös muita tekijöitä. Arnoldin (2008) mukaan tekstin tuottajat pystyvät hyvin ottamaan huomioon vastaanottajan tarpeen ymmärtää referentiaalisia ilmauksia. He tekevät ensinnäkin eräänlaisia globaaliolletuksia vastaanottajasta ja ottavat huomioon paitsi ymmärtämistarpeen, myös aiemman kokemuksen ja mahdollisen yhteisöllisyyden. Lisäksi he pystyvät ottamaan huomioon yksilöiden eroa viestin vastaanottamisessa. Tekstin tuottajat tekevät myös eräänlaisia lokaaliolletuksia siitä, mikä on huomion kohteena (aktivoituneena) eli ottavat referentin kognitiivisen statuksen huomioon. He vihjaavat statuksesta muun muassa syntaksin avulla: subjektin positio on prominentein eli etualaisin, objektin seuraava ja niin edelleen. Positioiden prominenttijärjestykseen suhteutuu myös referentiaalisten ilmausten saavutettavuushierarkia (*accessibility hierarchy*). Osa päätöksistä perustuu Arnoldin mukaan kuitenkin selvästi muuhun kuin vastaanottajan huomioimiseen. Tuottajilla on taipumus huomioida esimerkiksi tekstin rakenteellinen tai temporaalinen raja, tekstin jaksotus, mikä voi perustua tarpeeseen selvittää omasta kognitiivisesta taakasta. Tekstin tuottaja voi myös olla egosentrinen kirjoittaja, joko kehityksellisistä syistä tai muuten.

Yhteinen tausta, tieto ja oletettu tuttuus vaikuttavat referentiaalisiin ilmauksiin, niiden tulkintaan ja tuottoon, mutta kirjallisuudessa terminologia, tulkinnat ja teoreettiset valinnat vaihtelevat. Serratrice ja Allen (2015) koostavat tätä keskustelua myös tutkimuksen kehittymisen ja tarpeellisen lisätutkimuksen perspektiivistä. Tässä yhteydessä leksikaalisesti ilmaistu niin sanottu tuttu tieto jaotellaan kolmijaolla uudelleenmaininnat (konteksti, ilmeinen, päätelty), se-tarkenteella ilmaistut ja (kirjoittajalle) muuten tutut (eli uuteen tietoon ankkuroidut). Lisäksi on

pari pikkuryhmää, joihin palataan myöhemmin. Jaottelu pohjaa Princen taksonomian mukaiseen aineiston analyysiin ja osoittaa jo lähtökohtaisesti, että alakoululaisaineistossa aikuistyyppisen oletetun tuttuuden luokitus on liian täydellinen. Usein esimerkiksi se, mikä on kirjoittajan mieleen aktivoitunut tai mikä hänen mielestään on tuttua, on lukijan pääteltävä yleistiedon varassa, so. ilman kielellisiä vihjeitä.

Tutun ja uuden informaation erottaminen ja kielellinen merkintä on referentiaalisen koherenssin kannalta keskeistä. Henkilö- ja asiaviitteiset diskurssireferentit esitellään kielissä prototyyppisesti substantiivien luokkaan kuuluvilla tarkoitteilla, ja uusi tekstiin tuotu informaatio merkitään eri tavoin kuin tuttu: tämä on luonnollisesti analyysin perusta. Tapahtumiin viittaavia ilmauksia ei käsitellä tässä yhteydessä.

#### REFERENTIAALISEN KOHERENSSIN YMMÄRTÄMINEN JA TUOTTAMINEN

Referentiaalisen koherenssin prosessointitutkimusta on tehty paljon ja hyvin monenlaisissa asetelmissa ja monenlaisten puheyhteisöjen yhteydessä. Tutkimus nojaa niin kielelliseen kuin visuaaliseen materiaaliin ja sekä ymmärtämis- että tuottotehtäviin. Prosessointitutkimus on yleensä kiinnostunut paitsi kompleksin kielen prosessoinnista etenkin sosio-pragmaattisen tiedon ja ei-kielellisen kontekstin ymmärtämisestä, toisin sanoen pitkälti juuri yhteiseksi oletetun ja pääteltävän tiedon ymmärtämisestä. Kehityksellisen tutkimuksen rinnalla kulkee kielellisten erityisryhmien testaus (ks. John, Rowe & Mervis 2009; Loukusa, Mäkinen, Gabbatorea, Laukkanen-Nevala & Leinonen 2017). Yleiskäsitys on, että tekstien koherenssi ja koheesio tyyppillisesti lisääntyvät kielissä kehityksellisesti (Sonnenschein & Whitehurst 1983; 1984; Berman & Slobin 1994, 40). Sekä aikuiset että lapset nojaavat referentin saavutettavuushierarkiaan siten, että korkea informaatio ja matala saavutettavuus kohdistavat huomion leksikaalisiin aineksiin ja matala informaatio ja korkea saavutettavuus pronomineihin ja pro-muotoihin (Serratrice & Allen 2015).

Ymmärtämistutkimusten perusteella yleinen näkemys on myös, että lapset pystyvät noin 8–9 vuoden iässä toimimaan referentiaalisen kommunikaation testeissä lähes aikuisen tavoin (ks. esimerkiksi Glucksberg,

Krauss & Higgins 1975; Ackerman 1979; 1981; Beal & Belgrad 1990). He toisin sanoen erottavat viittaussuhteiltaan epäselvät ja toisaalta informatiiviset ilmaukset ja ymmärtävät, mitä tuottaja on pyrkinyt sanomaan. Viiden kuuden vuoden iässä lapsilla on vielä suuria ongelmia tehdä eroa kirjaimellisen ilmaisuuden ja toisaalta tavoitellun merkityksen välillä, paitsi jos heille vihjataan eron mahdollisuudesta. Eroa on myös testityyppien tuloksissa: offline-testit tuottavat paljon variaatiota, jossa on mahdollisesti testihälyä, online-testit osoittavat, että 3–4-vuotiaiden katse kohdistuu kuuntelun yhteydessä tarinan kuviin – ja pronomiinien korrelaateihin – samantapaisesti kuin aikuisten mutta paljon hitaammin (Sekerina 2015). Prosessoinnin nopeutuminen on sekini hidasta ja jatkuu pitkälle keskilapsuuteen.

Kehityksellinen ymmärtämistutkimus myös osoittaa, että referentiaalinen kommunikaatio ei nojaa yksin kielenhallintaan ja kielen keinoihin, vaan merkitseviä tekijöitä ovat niin lasten yleiset kognitiiviset kyvyt, toiminnan ohjauksen taidot, heidän (reseptiivisen) leksikkonsa koko, kronologinen ikä kuin mielen teoria ja oikeastaan lukuisa joukko muitakin tekijöitä. Erityisesti referentiaalisessa kommunikaatiossa lasten kronologinen ikä ja myös mielen teoria selittävät tuloksia merkitsevästi (Ackerman 1981; Beal & Belgrad 1990; John ja muut 2009). Suomen osalta tutkimusta on vähän ja referentiaalisen kommunikaation ymmärtämistä ei ole tietääksemme erikseen kokeellisesti tutkittu. Loukusan ja muiden (2017) pragmaattisen ymmärtämistutkimuksen tulosten mukaan 4–5-vuotiaat suomenkieliset lapset vastasivat referentiaalisia viittauksia laajemmalle ulottuviin mutta päättelemistä edellyttäviin sosio-pragmaattisiin kysymyksiin heikosti mutta 8-vuotiaat jo huomattavasti paremmin. Erityinen kehitysvaihe osui aikavälille 4–7 ikävuotta. – Myös ymmärtämistutkimuksen tulokset antavat olettaa, että ei ole odotettavissa, että esimerkiksi referentiaalisen koherenssin luominen onnistuisi vielä kaikilta alakoulun alkuvaiheissa olevilta. Ymmärtäminenhan tavallisesti edeltää (hallittuja) tuottamistaitoja, vaikka ilmeisesti juuri pronominisyntaksissa ja perspektiivoinnissa järjestys voi olla toinenkin (ks. Hendriks 2014).

Tuottamista on tutkittu testein ja kuvasarjakielennöksin, mutta erityisesti online-tuottamisesta on vähän tutkimusta (Serratrice & Allen 2015).

Useimmiten tutkimuksissa on kyse suullisesta kerronnasta tai tuottamistehtävistä, ei luovasta kirjoittamisesta. Serratricen ja De Catin (2020) mukaan aikuisilla on taipumus käyttää pronomineja eniten subjektiasmassa ja lauseen alussa sekä myös topiikkina. Aikuiset myös käyttävät informatiivisempia ilmauksia, kuten erisnimiä ja indefiniittisiä ilmauksia, jos referentti viittaa uuteen tietoon tai jos konteksti voisi jäädä epäselväksi. Jo esikouluikäisillä lapsilla on vastaavia taitoja mutta ongelmia lähinnä vastaanottajan näkökulman huomioimisessa. Yksilöllisen vaihtelun määrä on kuitenkin suuri. Serratricen ja Allenin 5–7-vuotiaille teettämässä referentiaalisen kommunikaation tuottotehtävässä variaation mahdollisina taustavaikuttajina tutkittiin kognitiivisen kontrollin taitoja, verbaalista työmuistia, kielenhallintaa, kielelle altistumista ja sen käytön määrää sekä niin sanottuja sosioekonomisia muuttujia (esimerkiksi vanhempien koulutusta). Tuloksien mukaan merkittäviä vaikuttavia tekijöitä sopivan referentiaalisen ilmauksen valintaan ovat erityisesti kognitiivinen kontrolli (kuten kyky miettiä valintaa riittävästi), verbaalinen työmuisti ja kielenhallinta: ne selittävät yksiköllistä variaatiota hyvin. Näin on todennäköisesti myös Unelmakirjoitelmien kontekstissa, vaikka koehenkilöitä ei ole erikseen näiden muuttujien osalta testattu.

Mäkinen ja muut (2014) tutkivat kissa-kuvasarjan avulla suullisen kerronnan kehittymistä 4–8-vuotiailla suomalaislapsilla. He mittasivat kehittymistä tavanomaisin mittarein. Mukana oli kokonaissanamäärä, virkemäisen yksikön sanamäärä ja rakenteellinen kompleksisuus, sisällön yksiköiden määrä sekä kyky tuottaa referentiaalista koheesiota (referentiaaliset viittaukset). Koehenkilöjoukkoon kuului yli 170 lasta. Tulosten mukaan tekstin tuottamisen kehitys näkyi kautta linjan eri muuttujissa, mutta parhaiten kertomuksen kehittymistä erottelevina tekijöinä toimivat referentiaaliset ilmaukset. Niiden avulla saatiin näkyviin merkitsevä ero alle kouluikäisten ja kouluikäisten lasten välillä. 7–8-vuotiailla referentiaalinen täsmällisyys nousi jopa yli 80 prosentin, tosin ei aivan kaikilla. Samanlaisia tuloksia on saatu muistakin kielistä kuten kantonin kiinasta, espanjasta ja englannista. Mäkinen ja muut (2014, 37) arvelevat, että koska referentiaaliset viittaukset ovat vaativia tuottaa, ne täydellistyvät vasta siinä vaiheessa, kun kertomusrakenne hallitaan. Kahdeksanvuotiailla esimerkiksi tekstissä taaksepäin viittaavien anafora-

suhteiden hallinta ei vielä ole aikuistasolla. Mäkinen ja muut katsovat myös, että kehitystrendi sekä täsmällisten viittaussuhteiden että sisällön yksiköiden määrien lisääntymisen taustalla on paralleelinen.

Nuorimmillakin alakouluikäisillä on siis jo melko hyvät edellytykset ymmärtää referentiaalisia ilmauksia, ja he pystyvät niitä myös tuottamaan etenkin, jos tuottoa tukee tehtävän rakenne ja jos tuottamista ei häiritse muu liian vaativa prosessointi. Tosin lukijan tietoinen huomiointi ja kielellisen ilmaisun keinot opitaan myöhään (Tomasello 2003; Hickmann 2004) ja aikuistason semanttis-pragmaattinen hallinta alkaa kehittyä vasta noin kymmenvuotiailla (Hickmann 2004, 285; Berman 2007). Onnistumiseen vaikuttaa kielenhallinnan taso, kyky keskittyä tehtävään, hyvä toiminnanohjaus ja kyky ottaa vastaanottaja huomioon. Luova kirjoittaminen luo haastavamman kontekstin kuin ohjeistettu ja rakenteellinen tehtäväkertominen. Lisäksi motoriset valmiudet, muistin toiminnot sekä tekstityyppien ja tekstilajien tuntemus vaikuttavat suoriutumiseen. On odotuksenmukaista, että suurin osa toisluokkalaisista ei vielä täysin onnistu referentiaalisen koherenssin luomisessa vapaassa tuottotehtävässä ja että kuudesluokkalaisilla tavoite olisi jo saavutettavissa, jos tekstiä ylipäätään syntyisi. Tarkastelu keskittyy tässä yhteydessä määräisten ilmausten tuottoon kirjoitetussa kielessä. Määräisyyden tarkastelun kehittymistä tutkitaan suhteessa kertovan tekstin kehittymiseen.

#### KERTOMUS JA KERTOVA TEKSTI

Tekstilajeja on tyypillisesti lähestytty sosiaalisen toiminnan näkökulmasta: samaan tekstilajiin kuuluvilla teksteillä on vuorovaikutustapahtuman näkökulmasta samankaltaiset tavoitteet ja rakenne (Virtanen 2010; Heikkinen & Voutilainen 2012). Tekstilajit eivät kuitenkaan ole tarkkarajaisia, vaan ne voivat kerrostua ja limittyä monin tavoin (Shore & Mäntynen 2006, 11). Vaikka kertomus tekstilajinomaisena käsitteenä on laajalti käytössä eri kielissä ja kulttuureissa, se voi itse asiassa todentua hyvin monenlaisina teksteinä. Mikäli kertomusta tarkastellaan laajemmin tekstilajien supertyyppinä tai tekstityyppinä, sen voidaan katsoa pitävän sisällään lukuisia toisilleen läheisiä tekstilajin kaltaisia alatyyppejä, kuten jännityskertomus, satu, tieteiskertomus, vitsi, juoru tai tunnustus.



Kertomus ei siis ole tekstilaji samassa mielessä kuin vaikkapa päiväkirja tai raportti, vaan se toimii ennen kaikkea yläkategoriana, joka erottelee toisistaan kertovat ja ei-kertovat tekstit tai diskurssit. Lisäksi kertova teksti tekstityyppinä tarjoaa keinon tarkastella eri tekstilajia edustavien tekstien sisäistä rakentumista. (Tekstityypeistä ja -lajeista tarkemmin ks. Bamberg (toim.) 1997; Virtanen 2010; Heikkinen & Voutilainen 2012.)

Tekstityyppinä esitetään yleensä rajallinen määrä (ks. Werlich 1974). Prototyypisenä ja muista tyypeistä erottuvana tekstityyppinä on pidetty juuri kertovaa tekstiä (*narrative*). Kertova tekstityyppi kuvaa ja luo mielikuvia todellisista ja kuvitelluista maailmoista. Kertovien tekstien tehtävä on selkeästi herättää mielikuvia, ei niinkään saada aikaan toimintaa (Virtanen 2010, 54–58, 62). Kertova tekstityyppi siis nimensä mukaisesti kertoo toiminnasta ja juonen etenemisestä, minkä takia se usein rakentuu toimintaa kuvaavista verbeistä ja on melko usein menneessä aikamuodossa. Lisäksi kertovassa tekstityypissä on aktiivisia toimijoita, jotka tekevät jotakin tai joille tapahtuu jotakin (Pietikäinen & Mäntynen 2009, 101). Teksti- ja diskurssityypit ja tekstilajit kuitenkin risteävät, eikä selkeää rajanvetoa voi tehdä. Niinpä kertomuksetkaan eivät ole vaihteluttomia tai yhteneviä, vaan niissä voi Virtasen (2010, 55) mukaan toteutua monenlaisia strategioita, kuten listaamista ja luettelointia tai suostuttelua. Kertovan tekstityypin lisäksi kertomukset sisältävät yleensä myös deskriptiivisiä tekstityyppejä, sillä niissä paitsi kerrotaan tapahtumia, myös kuvataan paikkoja ja ihmisiä. Kertomuksessa voi siis olla monenlaista kielellistä toimintaa, sillä tapahtumien kuvaaminen vaatii monesti myös paikkojen, henkilöiden ja tunnelmien kuvaamista sekä tapahtumien merkityksen arvioimista (ks. esimerkiksi Pietikäinen & Mäntynen 2009, 100).

Tekstilajien ja teksti- ja diskurssityyppien rakentuminen on opittava taito, ja toisaalta niitä voidaan käyttää myös luovasti. Pietikäisen ja Mäntynsen (2009, 104–109) mukaan narratiivit ovatkin sekä kulttuurisesti ja historiallisesti kiteytyneitä tapoja kertoa tapahtuneesta että ainutlaatuisia, yksilön omia tapoja merkityksellistää kokemuksia ja tuntemuksia. Kertomukset siis ilmaisevat kertojan yksilöllisiä kokemuksia, mutta samalla ne hyödyntävät kulttuurisia ja yhteisöllisiä kertomusvarantoja ja kertomuksellisia resursseja. Niiden kulttuurinen puoli näkyy etenkin

siinä, että ne ovat ymmärrettävissä vain, jos vastaanottaja jakaa saman käsityksen esimerkiksi ajasta. Narratiivisuuteen, kuten muuhunkin sosiaaliseen ja kielelliseen toimintaan, kasvetaan.

Muihin tekstityyppeihin verrattuna kertovilla teksteillä on kuitenkin kaksi vahvaa ominaispiirrettä: kertovissa teksteissä tapahtumien aikajäsennys (usein myös kausaalisuuden ilmaisu) ja osallistujien kuvaaminen ovat keskeisiä strategioita (Virtanen 2010, 54–56). Kertomukselle on siis tyypillistä, että tapahtumasarja noudattaa ajallisesti loogista rakennetta, joka voi kaunokirjallisissa tuotoksissa olla varsin monimutkainenkin kokonaisuus. Ajallinen loogisuus pohjautuu kuitenkin ajatukseen toisiaan seuraavista tapahtumista sekä niiden syistä että seurauksista. Onnistunut kertomus on siis juonellisesti eheä ja etenee loogisesti, mikä näkyy esimerkiksi ajallisena ja toiminnallisena yhtenäisyytenä (Givón 1983; 2017; Hickmann 2004; Pietikäinen & Mäntynen 2009). Lisäksi kokonaisuus on tekstuaalisesti yhdenmukainen. Analysoimme Unelmakirjoitelmien kertomusrakennetta tekstityyppiluokituksen mukaan, joka ottaa huomioon kehittyvän kirjoittamisen ja johon määräisten ilmausten tyypit ja esiintymämäärät voi suhteuttaa.

Matilaisen (1989) mukaan kertojатаival alkaa alakoulun alussa tyypillisesti yksittäislauseesta tai -lauseista koostuvana tekstinä, jossa ajallinen ja paikallinen tausta on usein ilmaisematta, tapahtumia ei ole liitetty tapahtumien ketjuun ja ennen kaikkea tekstin yhteen liittävä toisteisuus puuttuu. Toisteisuus erityisesti kertovassa tekstissä syntyy uuden tarkoitteen esittelystä ja tähän samaan tarkoitteeseen viittaamisesta jatkossa niin kauan, että päädytään tietyn tapahtumasarjan loppuun. Matilaisen mukaan nuorimmat noudattavat aluksi niin sanottua atomistista hahmottamisstrategiaa, joka tuottaa yksittäisiä havaintoja jostain teemasta tai useammista teemoista. Seuraavassa vaiheessa tekstiä tuotetaan eräänlaisen serialistisen hahmottamisstrategian mukaan, joka tuottaa tapahtumaketjuja usein ajallisessa järjestyksessä. Vasta holistinen tai analyttinen strategia tuottaa Matilaisen mukaan tekstejä, joissa kerrota kohdistetaan lukijalle halutussa tärkeysjärjestyksessä oleellista painottaen. Teksteinä ensimmäisessä vaiheessa tuotetaan listoja, toisessa skriptejä ja kolmannessa kertovia tekstejä (ks. Pajunen 2012), jos kertomus on tavoiteltu tekstityyppi. Listat ovat Bambergin tekstikohe-

renssiasteikkoja soveltaen tekstittömiä tai epäymmärrettäviä, skriptit ovat todennäköisimmin osittain tai epätäydellisesti koherentteja ja vasta ker-  
tovat tekstit pyrkivät olemaan koherentteja.

## Määräisyyden ilmaiseminen kehittyvässä kertovassa tekstissä

Tarkastelemme määräisyyden ilmaisun kehittymistä alakouluikäisten kirjoitelmissa suhteessa kertomuksena eriasteisesti onnistuneisiin teksteihin. Määrittelemme ensin ne ilmaisukeinot, joita lapsilla on käytettävissään, ja sitten käsittelemme niiden ilmenemistä Unelmakirjoitelmissa.

### MÄÄRÄISYYDEN LAJIT JA TEKSTIEN TYYPPILUOKITUS

Määräisyyden ilmaisemisen tarkastelussa lähtökohdaksi voidaan ottaa esimerkiksi kielten luokittelu määräisen artikkelin perusteella, kuten Dryer (2013) tekee jakaessaan maailman kielet viiteen ryhmään sen perusteella, onko kielissä määräistä ja/tai epämääräistä artikkelia vai ei ja missä muodossa ilmaisu milloinkin on. Määräisellä artikkelilla erotetaan määräiset ilmaukset epämääräisistä, joilla saattaa olla omat merkintäkeinonsa (Larjavaara 2007, 294). Suomen kieli kuuluu Dryerin (2013) jaottelun viidenteen ryhmään, jonka muodostavat kielet, joissa ei ole määräistä eikä epämääräistä artikkelia. Suomen tapaisessa artikkelittomassa kielessä on useampia muita keinoja tehdä syntaksissa ero tutun ja uuden informaation – määräisen ja epämääräisen – välillä. Näitä ovat muun muassa sijanvalinta, lauserakenteen tasolla tehdyt valinnat tai leksikaaliset keinot. Uusi informaatio tuodaan tekstiin joko esitelyrakenteella (e-subjekti ja käänteinen sanajärjestys) tai jonkinlaisilla tarkenteilla (*eräs, muuan, joku, yksi*). Tuttu tarkoite osoitetaan uudelleen maininnalla ja tähdentämällä sitä mahdollisesti määräisillä tarkenteilla (*tämä/se X*) tai esimerkiksi anaforisilla pronomineilla (*hän, se*). Pelkistään uusi informaatio ilmaistaan lyhyemmin – tai esimerkiksi epämääräisellä artikkelilla, jos kieli sen tuntee – mutta tuttua informaatiota voidaan merkitä monella tavalla myös siten, että puheena olevan tar-

koitteen (tai siihen viittaavan substantiivin) sisältävä rakenne kompleksistuu.

Suomen kielessä tarkoitteet voivat tekstissä olla merkityksensä puolesta määräisiä tai epämääräisiä (Penttilä 1963, 532). Lisäksi ne voivat olla geneerisiä eli yleisiä tai puolidefiniittisiä eli puolimääräisiä. Penttilä (mp.) kuitenkin toteaa, ettei suomen kielessä ole mitään mainittavia muotokeinoja näiden eri merkityslajien – määräisen, epämääräisen, geneerisen ja puolimääräisen – ilmaisemiseksi, vaan tarkoitteen määräisyyslaji jätetään yleensä erikseen ilmaisematta. Muotokeinojen puutteen syynä lienee se, että suomen kaltaisessa ei-artikkelikielessä puhuja ja kirjoittaja saavat itse päättää, esittävätkö he tekstinsä tarkoitteet merkitsemättöminä vai määräisiksi merkittyinä (Juvonen 2000, 29). Koska tarkoitteiden määräiseksi merkitseminen on vapaaehtoista, määräisyyden keinovalikoima on suomen kielessä melko laaja ja vakiintumaton.

Larjavaara (2007, 300) kuvaa suomen kaltaisten artikkelittomien kielten definiittisyyttä kuulijanäkökulmaiseksi spesifiksi representaatioksi. Hänen mukaansa (2007, 301) suomen tyyppisen kielen yleisessä definiittisyydessä on kyse universaalista kommunikatiivisesta merkityksestä, joka on mukana kaikissa kielissä mutta jolla on erilaisia ilmaisimia eri kielissä. Yleisimpiä ilmaisimia kaikissa kielissä ovat demonstratiivit, ja niiden määräiseksi tunnistaminen voi perustua muun muassa kontekstiin ja sanajärjestykseen: lauseenalkuinen sana tulkitaan yleensä määräiseksi ja lauseenloppuinen epämääräiseksi. Toisin sanoen tunnistaminen noudattaa saavutettavuushierarkiaa (Arnold 2008). Semanttinen definiittisyys on Larjavaaran (2007, 301) mukaan suomen kielessä relatiivista: kaikki kielen viittaavat ilmaukset eivät ole joko definiittisiä tai indefiniittisiä. Tekstin tuottaja ilmaisee ennen kaikkea kokonaisuuksia, jolloin ei ole olennaista, että vastaanottaja tunnistaa jokaisen kokonaisuuden osan. Ymmärtämistutkimus ei tosin täysin tue Larjavaaran näkemystä, koska lukija pyrkii ratkaisemaan epäselvätkin viittaussuhteet lukiessaan (vrt. esimerkiksi Glucksberg ja muut 1975; Beal & Belgrad 1990), mikäli lukee ymmärtääkseen. Lapset pystyvät tähän jo noin kahdeksanvuotiaasta.

Rajaamme tässä yhteydessä tarkastelun sellaisiin määräisiin ilmauksiin, joissa määräinen tarkoite kielennetään substantiivilla. Jos substantiivilla on tarkenne, kokonaisuutta voidaan tarkastella substantiivilau-

sekkeena. Substantiivilausekkeen alatyypit eli ne, joissa substantiivin edustajana on pronomini tai numeraali (vrt. esimerkiksi Hakulinen, Vilkuna, Korhonen, Koivisto, Heinonen, Alho 2004, 431), jäävät aineiston ulkopuolelle. Aineistoon kuuluu sekä yleis- että erisnimiä, ovathan tekstien henkilöhahmoihin viittaavat nimitykset olennainen osa kertomusta. Lisäksi jotta *se*-tarkenteen oletettua artikkelistumista voidaan mitata suhteessa mahdollisten ilmauksien kokonaisuusmääriin, aineiston kaikki määräiset ilmaukset ovat sellaisia, jotka periaatteessa voivat saada pronominitarkenteen. Aineiston ulkopuolelle jäävät siis ne määräiset ilmaukset, jotka ilmaisevat temporaalisia eli ajallisia tai spatiaalisia eli paikkaan liittyviä merkityksiä ja joita tarkennetaan proadverbeillä. Lisäksi tutkimusaineiston ulkopuolelle jäävät luonnostaan määräiset persoonapronominit (ks. Hakulinen ja muut 2004, 1352), pelkkien pronomien muodostamat lausekkeet, adjektiiviedussanaiset lausekkeet sekä ellipsin takia edussanattomat lausekkeet. – Määräisten ilmauksien tarkastelu rajautuu noin 24 000 tekstisanan aineistoon, joka on poimittu toisen, neljännen ja kuudennen luokka-asteen Unelma-aineistosta. *Se*- tai *ne*-tarkenteella merkityjä ilmauksia tarkastellaan laajemmasta aineistosta eli 2.–6. luokan teksteistä (n. 40 000 tekstisanaa), jotta esimerkkejä löytyy riittävä määrä.

Määräinen ilmaus on määräinen, kun sen tarkoite on tilannekontekstissaan sillä tavoin ainoa, että tekstin tuottaja voi olettaa vastaanottajan tietävän, mistä tarkoitteesta on kyse. Tarkoitteen uudelleenmaininta on määräisyyden tyyppitapaus (Hakulinen ja muut 2004, 1350). Näin ollen tekstissä jo aikaisemmin mainittuun, tekstistä tai kontekstista pääteltyyn tai muuten ilmeiseen tarkoitteeseen viittaavat substantiivilausekkeet ovat määräisiä (ks. Prince 1981; 1992; Vilkuna 2000; Hakulinen ja muut 2004). Määräisiä ovat paitsi kontekstista päätelty myös aktivoidut ilmaukset, jotka ovat tyyppillisesti suhteessa vastaanottajaan ja liittyvät siihen kehukseen, jonka konteksti luo hänen mieleensä (ks. Prince 1981; Vilkuna 1992). Kehittyvässä kirjoituksessa myös ensimmäinnat voivat olla määräisiä, etenkin jos niiden tarkoite on tavalla tai toisella ankkuroitu kontekstiin (vrt. taulukko 22, Prince 1981). Määräisyys suhteutuu myös implikaatiiviseen tunnettuus- ja saavutettavuushierarkiaan (Gundel ja muut 1993), joten joidenkin tarkoitteiden – erityisesti uudelleenmai-

nintojen – määräisyys on moninkertaista. Kun analysoinnissa on kuitenkin kyse kehittyvästä kielestä, on selvä, että esimerkit ovat välillä tulkinnanvaraisia. Uudelleenmainintojen vastakohtana toimivat ensimmäinnat saavat epämääräisen tulkinnan, mikäli tarkoite ei ole tuttuudeltaan ilmeinen tai lauseke ei sisällä määräisyyttä ilmaisevia määrittettä, joita ovat muun muassa pronominitarkenteet, genetiivimääritteet ja possessiivisuffiksit (Hakulinen ja muut 2004, 1350).

Määräisyyden lajit muodostavat tässä tutkimuksessa viisi ryhmää (ks. taulukko 23), jotka toimivat aineiston käsittelyn perustana. Ensimmäinnaltaan määräinen ilmaus voi kuulua taulukon 23 ryhmiin 2–5, eli määräisyys voi perustua tarkenteeseen, tuttuun tarkoitteeseen, universaali-pronominiin tai kvantitatiivisuuteen. Uudelleenmainintoihin on puolestaan viitattu jo aiemmin tekstissä tai tarkoitteet ovat kontekstinsa perusteella niin ilmeiset, että ilmauksen voi katsoa olevan lukijalle tuttu eli kontekstissaan määräinen.

Taulukko 23. Määräisyyden lajit.

Ryhmän nimi	Sisältö
1. Uudelleenmaininta	Aikaisemmin mainittu Tekstistä tai kontekstista päätelty Muuten ilmeinen
2. Tarkenteet	Demonstratiivipronomini Genetiivimäärite Possessiivisuffiksi
3. Tutut tarkoitteet	Erisnimet Tutut henkilöt ja paikat Perheenjäsenet Tarkoitteet, joita on olemassa vain yksi kerrallaan
4. Universaalipronominit	Kaikki-merkityksiset pronominit
5. Kvantitatiivisuus	Nominatiivi- tai genetiivimuotoinen määräite

Määräisten ilmauksien lajit suhteutetaan niin sanottuihin kertomustyypppeihin, jotka kuvaavat kehittyvää kirjoittamista kertovaan tekstiin johdattelevasta aiheesta Unelmieni päivä (aineistoa esitellään tarkemmin kirjjan johdantoluvussa). Oppilaskohtaisten tasoerojen ja luokka-asteiden välisten kehityserojen takia aineisto sisältää hyvin heterogeenisen tekstijoukon (vrt. Matilainen 1989). Osa teksteistä on helposti luokiteltavissa kertomuksiksi, mutta osa muistuttaa enemmän luetteloa tai kuvausta

kuin kertomusta (vrt. Pajunen 2012). Tekstien ominaisuuksiin perustuva luokittelu muodostaa jatkumon, jonka toisessa päässä on kertomusmaiseksi luokiteltavat tekstit ja toisessa päässä luetteloksi laskettavat tekstit (ks. taulukko 24). Näiden ääripäiden välillä on kolme tekstiryhmää: luettelomaiset kertomukset eli skriptit, keskeneräiset kertomukset ja lyhyempään hetkeen keskittyvät kuvaukset. Luokittelu poikkeaa Pajusen (2012), Matilaisen (1989) ja Vesalan (2012) tekemistä holistisista arvioinneista jonkin verran, koska tavoitteena on saada määräisten ilmauksien käytön kehittyminen mahdollisimman selvästi näkyviin.

Taulukko 24. Tekstien luokittelu sisällön ja rakenteen perusteella määriteltäviin tekstityyppeihin.

Tekstin tyyppi	Piirreluettelo
Kertomusmainen teksti	Tekstistä erottuu selkeä juoni, jonka etenemistä kertoja kuvaa Kertoja perustelee tapahtumien väliset syy- ja seuraussuhteet Sisällön kannalta olennaiset asiat pysyvät mukana koko kertomuksen ajan
Luettelomainen teksti	Tapahtumien väliin jää selittämättömiä aukkoja Kerronta etenee pintapuolisesti ja kuvauksen määrä on vähäistä Tapahtumasta toiseen siirrytään usein, minkä takia teksti vaikuttaa juonirungolta
Keskeneräinen kertomus	Teksti alkaa kertomusmaisesti mutta loppuu yllättäen
Kuvaus	Teksti keskittyy kuvaamaan lyhyttä hetkeä Teksti ei juurikaan sisällä tapahtumia
Luettelo	Teksti listaa tapahtumat peräkkäin Kokonaisuus voi muodostua yhden verbin ympärille Siirtymät ja tapahtumien kuvaus jäävät huomiotta

Kertomusmaisten tekstien ryhmään kuuluvat kirjoitelmat, joista erottuu selkeä juoni, jonka etenemistä tekstin kertoja kuvaa. Teksti muodostaa koherentin kokonaisuuden, jossa juonen kannalta olennaiset asiat pysyvät mukana koko kertomuksen ajan. Kertomusmaisen tekstin kertoja perustelee tapahtumien väliset syy- ja seuraussuhteet sekä sitoo kertomuksensa aikaan ja paikkaan. Kertomus on siis rakenteeltaan johdonmukainen ja tekstuaalisesti yhdenmukainen, ja se herättää mielikuvia luomistaan todellisista tai kuvitelluista maailmoista. Koulukorpuksen teksteissä nämä mielikuvat liittyvät unelmien päivään ja sen sisältöön.

Luettelomaisten kertomusten ryhmä sijoittuu tekstityyppijatkumolla melko lähelle kertomusmaisten tekstien ääripäätä. Ryhmän tekstit muistuttavat kertomuksia mutta ovat niin pelkistettyjä, että teksti muodostaa pikemminkin juonirungon kuin varsinaisen kertomuksen. Yksi luettelomaisen kertomuksen tunnuspiirre on peräkkäisyyttä ilmaisevan temporaalisen *sitten*-adverbin käyttö virkkeen alussa. Tapahtumasta toiseen siirrytään nopeasti ja tapahtumien välille saattaa jäädä pitkiäkin selittämättömiä aukkoja. Nämä tyhjät kohdat eivät kuitenkaan vaikuta harkituilta tyylikeinoilta, minkä takia tekstiin syntyy luettelomaisuuden vaikutelma. Luettelomaisen kertomuksen kerronta pysyttelee hyvin yleisellä tasolla ja kuvauksen määrä on varsin vähäistä. Osallistujien kuvaaminen jää vähemmälle, mutta tekstin tapahtumat liittyvät toisiinsa, mikä erottaa luettelomaisen kertomuksen puhtaasti luetteloiksi nimetyistä teksteistä.

Jatkumon seuraavan ryhmän muodostavat keskeneräiset kertomukset. Ne alkavat varsin kertomusmaisesti mutta loppuvat yllättäen. Kertoja voi kuvailla tapahtumia hyvinkin tarkasti tekstin alkupuolella, mutta kertomus ei saavuta loppuratkaisuaan tekstin aikana. Kirjoittaja ei siis ole saanut tekstiään valmiiksi aikarajojen puitteissa tai on väsynyt kirjoittamiseen ja lopettanut tekstin kesken. Keskeneräisten kertomusten tärkein piirre onkin juuri tarinan yllättävä katkeaminen: teksti etenee loogisesti, kirjoittaja osoittaa tapahtumien aikajäsennyksen ja kuvaa osallistujia mutta ei vie juonta loppuun asti. Lisäksi aineistossa on tekstejä, jotka alkavat varsin kertomusmaisesti mutta joissa kerronta muuttuu alun jälkeen luettelomaisempaan suuntaan. Tällaiset tekstit eivät kuitenkaan kuulu keskeneräisten kertomusten ryhmään, sillä niissä tapahtumat kerrotaan loppuun saakka. Nämä tekstit on luokiteltu luettelomaisiin kertomuksiin.

Neljännän ryhmän muodostavat kuvauksiksi luokiteltavat tekstit. Kuvaukset keskittyvät päivää lyhyempään hetkeen ja kuvaavat yksittäisen asian tai tapahtuman melko tarkasti. Kuvauksiksi luokiteltavissa teksteissä ei ole juonirunkoa eikä teksti pidä sisällään runsasta tapahtumien joukkoa. Osallistujia kuvataan mutta esimerkiksi tapahtumien aikajäsennys puuttuu. Ryhmään kuuluvat myös aineiston lyhimmat tekstit, jotka eivät lyhyytensä ja tapahtumaköyhyytensä perusteella kuulu mihinkään muuhun jatkumon tekstiryhmään.



Kertomusmaisten tekstien vastakohtana tekstityyppijatkumolla on luetteloiksi nimetyt tekstit. Niissä unelmien päivän tapahtumat on listattu peräjälkeen eikä yksittäistä tapahtumaa kuvata mainintaa tarkemmin. Luettelo voi koostua yksittäisistä sanoista, lauseista tai parin virkkeen kokonaisuuksista. Usein luetteloksi luokiteltava teksti rakentuu yhden verbin, kuten haluamista tai saamista ilmaisevan, ympärille. Teksti ei sisällä perusteltuja syy- ja seuraussuhteita eikä siirtymiä tai itse tekemistä kuvailla tarkemmin. Tekstin sisältämät tapahtumat eivät välttämättä liity toisiinsa, ja niiden ajallinen etäisyys toisistaan voi olla hyvinkin pitkä. Juuri tapahtumien kausaalinen ja ajallinen irrallisuus erottaa luettelot luettelomaisista kertomuksista.

Kun aineiston tekstejä on jaettu tekstin tyyppijatkumon ryhmiin, ei ole juurikaan huomioitu sitä, pysyykö oppilaan kirjoittama teksti annetussa aiheessa. Koska tutkimuksen kohteena ovat määräisten ilmauksien yhteydet tekstin luokittukseen, aiheessa pysyminen ei ole olennainen luokittelukriteeri. Huomio keskitetään siis tekstin sisällön ja rakenteen muodostamaan kokonaisuuteen ja siihen, millaisia yhteyksiä määräisten ilmauksien tekstikohtaisella esiintymämäärällä on tähän kokonaisuuteen (vrt. Labov & Waletzky 1967).

#### MÄÄRÄISET ILMAUKSET UNELMAKIRJOITELMISSA

Määräisiä ilmauksia eli lausekkeita, joiden pääsanana on tavalla tai toisella tuttuun tarkoitteeseen viittaava substantiivi, on aineistossa parisen tuhatta, määrällisesti lähes yhtä paljon aineiston joka luokka-asteella, mutta kirjoittajakohdainen ja suhteellinen jakauma on vino (ks. taulukko 25). Hieman useampi kuin joka kymmenes (14,8 %) teksti sisältää vain epämääräisiä ja geneerisiä ilmauksia ja luokka-asteittain tarkasteltuna tällaisten tekstien osuus on odotuksenmukaisesti suurin toisen luokan aineistossa (22,7 %). Neljännellä ja kuudennella määräisiä ilmauksia jätetään käyttämättä enää alle kymmenesosassa tekstejä. Sukupuolittain tarkasteltuna vain epämääräisten ja geneeristen ilmauksien käyttäminen näyttäisi olevan leimallisempaa pojille, sillä 73,5 prosenttia tällaisten tekstien kirjoittajista on poikia. Nämä tekstit ovat useimmiten luettelomaisia kokonaisuuksia, jotka rakentuvat haluamista tai saamista tarkoittavien verbien ympärille (ks. esimerkki 1). Myös melko yleisellä tasolla

Taulukko 25. Määräiset ilmaukset ikäluokittain Unelmakirjoitelmissa.

Unelmakirjoittelien otos	Toinen	Neljäs	Kuudes
Tekstejä otoksessa	n = 163	n = 96	n = 71
Määräisiä ilmauksia	n = 643	n = 655	n = 678
Tekstit, joissa on määräisiä ilmauksia	77,3 %	92,7 %	92,9 %
Tekstit, joissa ei ole määräisiä ilmauksia	22,7 %	7,3 %	7,0 %
Määräisiä ilmauksia keskimäärin / teksti	5,1 kpl	7,4 kpl	10,3 kpl

pysyttelevät kuvaukset unelmien päivän ohjelmasta sisältävät vain epämääräisiä ja geneerisiä ilmauksia (esimerkki 2). Vaikka geneeriset ilmaukset toistuvat tekstissä, tarkoitteen uudelleenmainintoja ei lasketa määräisiksi, koska niillä ei viitata tiettyyn tarkoitteeseen vaan ylipäätään tarkoitteiden luokkaan.

1) *Haluaisin mönkiän ja paystation2. Haluaisin hiihtää paljon ja lasketella paljon. Haluaisin ajaa paljon rallia ja jokkista. Haluaisin rahaa paljon ja ostaisin lemikiä. Haluaisin autoja. Haluaisin isoon huvilan ja ruokaa paljon ja syödä paljon ja lumilaudan ja laskettelusukset ja oman metsän ja traktorin.* (2M189)<sup>49</sup>

2) *unelmieni päivä on olla viistoista-vuotias koska saan ajaa mopolla kouluun ja kotiin. Se on hauskaa kun saa ajaa mopolla ees taas. kivaa on myös vapaa ajalla ajella mopolla joka paikassa. Mopolla pääsee nopeempaa, kuin pyörällä tai kävellen esim. koulusta kotiin. Voisin ottaa mopoauton, koska sillä pystyy harjoitteleen autolla ajamista ja pystyis tuunaamaan sitä. Mopoautolla olisi kivaa talvella johonkin mutkaan heittää auton takaperää ja laittaa kunnan poppis-stereot ja isot subba-rit taakse. Niin sois kunnolla musiikki autossa. Loppu* (6M87)

Kirjoittajakohtaisesti tarkasteltuna määräisten ilmauksien käyttäminen lisääntyy selvästi iän myötä. Kun toisluokkalaisten ylipäätään määräisiä ilmauksia sisältävissä kirjoitelmissa niitä on keskimäärin 5,1, luku on neljäsluokkaisilla 7,4 ja kuudesluokkalaisilla jo 10,3 (ks. taulukko 25). Tiheys siis kasvaa kehityksellisesti, mikä on odotuksenmukaista.

49 Koulukorpuksen kirjoittajat on identifioitu seuraavasti: kaksi ensimmäistä merkkiä paljastavat luokkasteen ja sukupuolen ja loput tekstin tunnisteet itse korpuksessa.

Mäkisen ja muiden (2014) arvelun mukaan referentiaalisten ilmausten onnistunut tuottaminen edellyttää kertomusrakenteen hallintaa. Kun määräisten ilmausten määriä vertaa Pajusen (2012) antamiin lukuihin samojen kirjoitelmien – joskin hieman laajemman joukon – kertomusrakenteen jonkinasteisesta noudattamisesta ja kun ottaa huomioon, että laskelmissamme eivät ole aivan kaikki referentiaaliset ilmaukset, näyttää siltä, että arvelu on tyttöjen osalta oikeansuuntainen: toisluokkalaisten tyttöjen kirjoitelmista kertomusrakennetta noudattaa 60 prosenttia, neljäs- ja kuudesluokkalaisten noin 80 prosenttia. Poikien osalta yhteys ei näy (30 %, 50 % ja 60 %), mikä voi johtua esimerkiksi siitä, että pojille oman tarinan kirjoittaminen on haasteellisempaa kuin vaikkapa kuvasarjasta kertominen. Määräisten ilmausten keskimääräinen esiintymistiheys ei muutenkaan kerro koko totuutta niiden käytöstä, sillä kirjoittajakohdainen vaihtelu on niin suurta.

Kirjoitelmien määräiset ilmaukset ( $n = 1976$ ) ovat definiittisiä joko uudelleenmaininnan, tarkenteen, tutun tarkoitteen, universaalipronominin tai kvantitatiivisuuden perusteella (tarkemmin ks. taulukko 23). Kun nämä ilmaukset jaetaan kahteen ryhmään sen perusteella, onko lauseke määräinen jo ensimmäinnaltaan vai vasta uudelleenmaininnan myötä, huomataan, että kunkin luokka-asteen määräisistä ilmauksista noin puolet on ensimmäintöjä ja puolet uudelleenmainintöjä. Jo tämä jakauma on paljonpuhuva ottaen huomioon, että uudelleenmaininta on kuitenkin eräänlainen määräisyyden prototyyppi. Määräisten ensimmäintöjen lähempi tarkastelu osoittaa, että valtaosa jo ensimmäinnaltaan määräisistä ilmauksista saa definiittisyytensä joko tutun tarkoitteen tai tarkenteen myötä. Toisen luokan aineistossa tuttuihin tarkoitteisiin viittaavat ensimmäinnat muodostavat yli puolet kaikista määräisistä ensimmäinnoista, kun neljännellä luokka-asteella niiden määrä jää alle puoleen ja kuudennella luokka-asteella vajaan kahteen viidesosaan (ks. taulukko 26). Väheneminen on asteittaista ja toisen ja kuudennen välillä ero on jo noin 16 prosenttiyksikköä. Kehityssuunta on päinvastainen tarkenteen perusteella määräisissä ensimmäinnoissa, sillä nyt määrä kasvaa (28,2 %, 34,9 % ja 37,9 %). Muutos on jälleen asteittainen, ja toisen ja kuudennen välinen ero on kasvanut jo lähes kymmeneen prosenttiyksikköön. Ero on selvästi kehityksellinen ja tarkoittaa, että nuorimmat

kirjoittajat puhuvat paljon paitsi perheenjäsenistä (*äiti, isä*) myös ystävis-tään (*Liisa, Pekka*) heitä lukijalle esittelemättä. Vanhemmilla kirjoittajilla tällaisiin ilmauksiin sisältyy jokin luonnehdinta (*äitini, ystäväni, luokka-toverini*) myös ensimmäinnassa. Kuudesluokkalaisilla on eniten myös määräisten ilmauksien pikkutyyppejä, mikä voi kertoa määräisyyden ilmaisemisen keinojen monipuolisemmasta hallinnasta.

Taulukko 26. Määräisten ilmausten ensimmäintöjen lajit.

Määräiset ensimmäinnat / Ikäluokka %	Toinen	Neljäs	Kuudes
Tarkenteellinen määräisyys	28,2	34,9	37,9
Tuttu tarkoite	53,4	49,3	37,2
Tarkenteellinen tuttu tarkoite	8,2	6,3	8,4
Universaalipronominin sisältävä	4,9	4,6	5,3
Kvantitatiivisesti määräinen	3,6	3,6	10,5
Muu	1,6	1,3	0,7

Määräisten ensimmäintöjen tarkastelusta tuttuja kehityssuuntia on ha-vaittavissa myös uudelleenmainintöjen osuuksissa (ks. taulukko 27), jos-kaan kehitys ei ole yhtä lineaarista kuin ensimmäinnoissa. Yksinkertaisten uudelleenmainintöjen määrä kasvaa selvästi iän mukaan, ja toisen ja kuu-dennen luokan välinen ero on jo 10 prosenttiyksikköä. Tarkenteellisten uudelleenmainintöjen prosentuaalinen osuus niin ikään kasvaa, ja toisaal-ta tuttuihin tarkoitteisiin viittaavien uudelleenmainintöjen osuus piene-nee merkittävästi eli noin 20 prosenttiyksikköä. Uudelleenmaininnoissa kirjoittajat hyödyntävät pelkän toistamisen (esimerkki 3) lisäksi kehyksen aktivaatioon perustuvia määräisiä ilmauksia (esimerkit 4–5) ja voivat merkitä tällaiset uudelleenmaininnat myös tarkenteella (esimerkki 6).

Taulukko 27. Määräisten ilmauksien uudelleenmainintöjen lajit.

Määräiset uudelleenmaininnat / Ikäluokka %	Toinen	Neljäs	Kuudes
Uudelleenmaininta	41,7	38,6	51,4
Uudelleenmaininta ja määräinen tarkenne	11,2	39,3	20,4
Uudelleenmaininta ja tuttu tarkoite	41,4	17,9	22,1
Yhdistelmätyyppi	3,6	3,4	3,8
Uudelleenmaininta ja universaalipronomini	1,5	0,6	0,3
Uudelleenmaininta ja kvantitatiivisuus	0,6	0,0	0,3
Muut	0,0	0,3	1,8

- 3) *Pusukala tuli ja pussasi minu. Se oli todella hieno. Keräsin sukellusreppuun simpukoit, korallia ja otin vesikamerallani kuvia **pusukalasta**. (4F23)*
- 4) *Illalla menen saunaan ja sieltä monta kertaa uimaan. Vesi ei ole kylmää vaan sopivasti viileää – **Järvi** on ihan tyyni. (6F1)*
- 5) *Kotona huudan äitiä ja kysyn häneltä innokkaana risteilystä. **Lähtö** olisi klo 17. (6F4)*
- 6) *Siskoni meni allergiatesteille. Kysyin illalla äitiltä oliko siskoni allerginen. Olin hirveän iloinen koska hn ei ollut allerginen. Seuraavana päivänä olin kaverini kanssa. Äiti soitti minulle ja sanoi että voitaisiin ostaa tänään **se hamsteri**. (4F3)*

Muut määräisen uudelleenmaininnan lajit ovat melko vähämerkityksisiä. Tämä on odotuksenmukaista, koska määräisyyden ilmaisussa uudelleenmaininta ja tarkenne ovat hierarkian huipulla. Kaikista määräisistä ilmauksista jo ensimmäinnaltaan määräisten määrä siis laskee ja vastaavasti vasta uudelleenmaininnan jälkeen määräisten määrä kasvaa, mikä osoittaa, että referentiaalinen koherenssi kasvaa kehityksellisesti. On kuitenkin muistettava, että aineistoon ei ole sisällytetty anaforisia viittauksia, jotka ovat yksi uudelleenmaininnan perustyyppi. Anaforiset pronominit (*hän, se*) ovat kertovassa tekstissä usein subjektipositiossa (ks. esimerkiksi Arnold 2008) ja samasta aineistosta tehdyissä Pajusen (2012) laskelmissa deiktisten pronominisubjektien määrä laskee yli 10 prosenttiyksikköä ja anaforisten määrä taas nousee toisen ja viiden luokan välillä yli 10 prosenttiyksikköä, kuudennella vähän vähemmän. On huomattava, että deiktisetkin pronominit ovat luontaisesti määräisiä mutta tilanneviitteisinä ne ovat kehityksellisesti vähemmän merkitseviä. Koherenssin kasvuun viittaa myös Heinosen (2018) tulos, jonka mukaan aiheiden määrä laskee Unelmakirjoitelmissa kehityksellisesti, kun taas sanamäärä kasvaa: samasta aiheesta kerrotaan iän myötä aiempaa enemmän.

#### MÄÄRÄISTEN ILMAUKSIEN LAJIT JA TEKSTIEN TYYPIT

Kyky ymmärtää ja tuottaa referentiaalisia ilmauksia kehittyä hitaasti ja edellyttää monenlaisia taustatekijöitä, joista semanttis-pragmaattinen

hallinta lienee nykykontekstissa tärkein. Tekstin tuottajan on myös hallittava ne kielen keinot, joilla määräiset ilmaukset merkitään lukijallekin ymmärrettäviksi. Referentiaalisten ilmausten on arveltu yhdistyvän niin ikään hitaasti kehittyvään taitoon tuottaa tekstilajeittain tyyppillisiä rakenteita. Analysoimme tässä alajaksossa, miten paljon määräisiä ilmauksia on teksteissä toisella, neljännellä ja kuudennella luokalla ja miten nämä määrät suhteutuvat Unelmakirjoitelmien tekstityyppiluokitukseen (ks. taulukko 24). Näin toivottavasti saadaan arvelua tarkempi vastaus referentiaalisen koherenssin ja kertomusrakenteen hallinnan yhteydestä (Berman 2007; vrt. Mäkinen ja muut 2014).

Määräiset ilmaukset on uudelleenluokiteltu esiintymämäärien perusteella muutamaaan luokkaan, jotta mahdolliset kehityserot saataisiin paremmin näkyviin. Luokista oleellisimmat ovat kolme ensimmäistä: tekstissä ei ole lainkaan määräisiä ilmauksia tai tekstissä on joko 1–5 tai 6–10 määräistä ilmausta (taulukko 28). Jos määräisiä ilmauksia ei ole lainkaan, tekstissä viitataan pelkästään epämääräisiin tai geneerisiin tarkoitteisiin. On myös mahdollista, että teksti etenee *minä*-subjektin varassa. Joka toisessa toisluokkalaisen tekstissä määräisiä ilmauksia on 1–5, mutta neljäsluokkalaisilla niiden määrä jää näin pieneksi enää alle puolessa ja kuudesluokkalaisilla enää joka kolmannessa tekstissä. Yleisyysluokassa 6–10 suhde on jo kääntynyt päinvastaiseksi: kuudesluokkalaisten tekstejä on yleisyysluokassa eniten ja toisluokkalaisten vähiten. Myös hajonta eri yleisyysluokkiin on kuudesluokkalaisilla suurinta.

Taulukko 28. Määräisten ilmauksien määrä teksteistä.

Määräisiä ilmauksia teksteittäin / Ikäluokka %	Toinen n = 163	Neljäs n = 91	Kuudes n = 71
0	22,7	7,3	7,0
1–5	52,1	45,8	31,0
6–10	16,6	24,0	28,2
11–15	4,9	13,5	12,7
16–20	1,8	5,2	7,0
21–25	1,8	2,1	7,0
26–30	0,0	1,0	2,8
Yli 30	0,0	1,0	4,2

Vain yhden määräisen ilmauksen sisältävät tekstit pysyttelevät melko yleisellä kertomisen tasolla. Toisluokkalaisten aineistossa tällaiset tekstit ovat luettelomaisia (esimerkki 7), kertovat tapahtumista vain suurpiirteisimmät tapahtumalinjat (esimerkki 8, luettelomainen kertomus) tai ovat huomattavan lyhyitä. Samankaltaisia piirteitä on myös neljännen ja kuudennen teksteissä, joissa esiintyy vain yksi määräinen ilmaus. Neljäs- luokkalaisilla kyse on luettelevista tapahtumalistoista, päivää lyhyempään ajanjaksoon keskittyvistä pikku kuvauksista tai luettelomaisista kertomuksista (esimerkki 9). Kuudesluokkalaisilla taas yhden määräisen ilmauksen tekstit ovat luettelomaisia kertomuksia, kuvauksia tai peräti luetteloita. Luettelomaiset kertomukset kertovat vain pääkohdat päivän tapahtumista (esimerkki 10). Kuvauksiksi luokiteltavat tekstit taas keskittyvät jonkin asian havainnollistamiseen, eivätkä päivän tapahtumat ehdi alkaa tai niistä kerrotaan hyvin vähän.

Luettelo:

7) *No mennä helsinkiin ja Olla 2 viikkoa ja mennä lentokoneella afrikan ja syödä karkkia ja uida ja pelata pleikkaria. ja leikkia kave-reitten kanssa ja lukea Harry Potteria ja mennä särkän niemeen. Ja seikkailia afrikassa ja suunnistaa ja ajaa pyörällä.* (2M180)

Luettelomaisia kertomuksia:

8) *Silloin oli kaunis päivä ja aurinko paistoi. Me olimme esbaniassa. Siellä oli hieno uima allas. Sieltä meni rengas josta piti sukeltaa pore-altaaseen. Kun me mentiin takaisin lentokoneella. Kun me noustiin ilmaan niin alkoi tuntua omituiselta. Kun me oltiin perillä oli ihanaa tulla siistii kotiin kun ennen lähtöä me siivottiin.* (2F44)

9) *Olipa kerran talvinen päivä. Olimme menossa hiihtämään koulun kanssa. Silloin kun olimme päässeet Niihamaan, silloin sai olla lämpimässä ja kioskillä yms. Sellainen oli unelmieni päivä.* (4M103)

10) *Olin autiosaarella ja kello oli 19.30. Naisia ja juomaa riitti vaikka muille jakaa. Tanssit alkoivat juuri ja kivaa on. Kello 21.00 alkoi limbo kisa, jonka voitin ja sain valita yhden tytön ja me pääsimme sviittiin yöksi. Se oli mahtavaa, kello 23.15 olimme Maijan (jonka valitsin limbokisan voitosta) kanssa sviitissämme ja pelasimme monopolia kello 1.30 asti, sitten menimme nukkumaan.* (6M113)

Muutaman määräisen ilmauksen sisältävissä teksteissä esiintyvät jo kaikki tekstityypit. Ryhmään kuuluviissa toisluokkalaisten teksteissä on sekä luettelomaisia, yleisellä tasolla pysytteleviä kirjoituksia (esimerkki 11) että kertomusmaisempaa tekstiä (esimerkki 12). Luettelomaisten kertomusten ja kertomusmaisten tekstien määrä on aiempaa suurempi, mutta luettelot ja kuvaukset ovat yhä varsin yleisiä. Neljäsluokkalaistil- taas suosituimmaksi ryhmäksi nousevat luettelomaiset kertomukset ja kertomusmaisuus valtaa alaa, kun määräisten ilmausten tekstikohtainen määrä kasvaa (esimerkki 13). Samankaltaista trendiä noudattavat myös kuudesluokkalaisten tekstit: muutaman määräisen ilmauksen tekstit ovat pääsääntöisesti luettelomaisia kertomuksia (esimerkki 14) mutta painotus siirtyy kertomusmaisempaan suuntaan (esimerkki 15).

Luettelo:

11) *Unelmieni päivä on että meidän koko perhe menisi lomalle. Kahdeksaksi viikoksi ja että mikään ei menisi pieleen. Eikä kukaan riitelisi eikä huutaisi ja että kaikki lähtisi hyvillä mielin. tai iskä ostaisi kaikkea kiivaa Essille ja mulle. se olisi kivaa. Että olisi kuuma. ja se oli minun unelmieni päivä.* (2F54)

Luettelomainen kertomus:

12) *Olipa kerran tyttö joka oli aina melkein iloinen. ja sen tytön nimi oli ilona. mutta kerran hanelle tapahtui ihan outoa. kun hän meni kellarin hän näki siellä peikon. mutta se olikin vain vaatekasa. mutta kuitenkin hän säikähti. aika paljon. mutta sitten hän lopuksi nauro. kuinka hassu hän oikein oli pelästy vaate kasaa. sitten häntä alkoi väsyttään ja meni nukkumaan.* (2F36)

Kertomus:

13) *Oli rauhallinen ja lyhyt yö, minä nukuin. Puut rapisivat ja tuuli puhalsi lujaa. minä heräsin ja menin ikkunaan ja huomasin kuinka kaunista ulkona oli. Ulkona aurinko nousi ja valaisi taivaan. Minä menin ottamaan viltin ja juoksin takapihalle. Minä en ollut ikinä eläisäni nähnyt mitään yhtä kaunista. Levitin viltin nurmikolle ja menin makaamaan. Yhtäkkiä sorsia tuli viltilleni ja pyysivät leipää. Syötin sorsia ja he nukahtivat viltilleni. Jätin linnut rauhaan. Minä menin sisälle ja olin tosi onnellinen. Tämä päivä oli unelmieni päivä. Loppu. Tämä tapahtui minulle viime kesänä.* (4M94)



Luettelomainen kertomus:

14) *Nukun 10.00, syön kunnan aamupalan pakkaan tavarat. Mua tullaan limusiinilla Hakeen. Sitten meen lentokoneella Kanadaan. Sie mulle on varattu jäähalli ja sie on kaikki mailman jääkiekko mailat ja saan testailla niitä. Sit valkkaan siitä parhaat ja mee NHL-peliin illalla. Teen 6 maalia ja 3 syöttö pistettä. ja mut valitaa NHL-all staariin. Sain siitä pelistä 10 miljoonaa ja sitten ostin tommy Hilfigerin kaikki vaatteet!! (6M80)*

Kertomus:

15) *Heräsin aamulla yhdeksän aikoihin lämpöiseen thaimaan aurinkoon. Menimme perheeni ja isovanhempieni kanssa nauttimaan erinomaisesta aamupalasta. Aamiaisen jälkeen lähdimme pitkähäntä veneillä kohti korallisaarta. Kun nousimme maihin, lähdimme heti snorklaamaan ja uimaan. Söimme eväämme saarella, jonka jälkeen leivän murusilla ruokimme kaloja kädestä. – – Menimme hotellille kahdeksan aikoihin, missä söimme ja joimme, katsoimme esityksiä ja meille ojennettiin stipendi ja mitali siitä, kun voitimme lusikoiden sukellus kilpailun. Kävimme vielä kaiken tämän jälkeen iltauinnilla, jonka jälkeen luonnon ääniä kuunnellen menimme nukkumaan. (6M97)*

Kun määräisten ilmausten tekstikohtainen määrä kasvaa yli kuuteen, kertomusmaiset tekstit kasvattavat osuuttaan. Toisen luokan aineistossa eniten (6–25) määräisiä ilmauksia sisältäviä tekstejä on noin neljännes. Kertomusmaisten tekstien määrä on tässä tekstiryhmässä suurempi kuin aiemmin, vaikka yleisimpiä ovat luettelomaiset kertomukset. Ne siis sisältävät runsaasti tapahtumia, joista kertoja kuitenkin puhuu melko yleisellä tasolla (esimerkki 16). Kertomusmaisista teksteistä taas erottuu selkeä juoni, jonka etenemistä kertoja kuvailee luettelomaisuutta tarkemmin (esimerkki 17). Tekstiin on kirjattu myös siirtymät tapahtumasta toiseen ja tietyt asiat, kuten keskeisiin henkilöihin viittaaminen, säilyvät mukana koko tekstin ajan.

Luettelomainen kertomus:

16) *Unelmieni päivä kun minä sain hamsterin minä ruokin sitä sen mimi oli tipsulla oli oma juoksupyörä ja se sai olla meillä 2 vuotta*

*sitten tulin kouluun minä jätin tipsun kotiin tipsu loysi kaksikymmentä senttiä se rupesi syömäänsitä se järsi kullat siitä pois sitten siitä tuli hopea. loppu rahat mutta minulla oli rahaa minä pistin rahat purkkiin niin me päästii ruotsista suomeen kulutettiin rahat. (2M188)*

Kertomus:

17) *Kun Tiina meni kerran kouluun niin hän näki että Emmalla oli uusi rannekello. Sitten kun Tiina tuli kotiin koulusta niin hän sanoi vihaisesti äidilleen: minä haluan uuden pyörän. Kun kerran Emmakin on saanut rannekellon ja vaikka mitä kivaa. Aina ei saa mitä haluaa sanoi äiti töihin keskittyneenä. Sitä paitsi huomenna on synttärini jatkoi Tiina. Sitten huomisaamuna Tiina meni kouluun ja tuli koulusta. Mutta kun hän avasi kotioven niin siellä odottivat kaikki hänen kaverinsa ja äiti ja isä. Muistitte siis syntymäpäiväni sittenkin tietysti sanoivat äiti ja isä yhteen ääneen. Arvaapa mitä tässä paketissa on. Voisiko se olla kyllä on se on uusi polkupyörä kiitos äiti ja isä tämä on unelmieni syntymäpäivä kirjoitan sen päiväkirjaani. (2F15)*

Neljäs- ja kuudesluokkalaisten tekstit, joissa on runsaasti määräisiä ilmauksia, ovat suurimmaksi osaksi kertomusmaisia kirjoituksia unelmapäivän sisällöstä (esimerkki 18), ja niissä on selkeä juoni alkutilanteineen, tapahtumien kuvailuineen ja loppuratkaisuineen. Vieläkään kaikki kirjoitukset eivät ole kertomusmaisia, mutta luettelot ja kuvaukset ovat jo harvinaisia. Runsaasti määräisiä ilmauksia sisältävät kertomusmaiset tekstit ovat melko pitkiä ja niissä kerrotaan tapahtumista paikoitellen hyvinkin tarkasti. Kerrontaan kuuluu lisäksi vuoropuhelua henkilöhahmojen välillä. Runsas määräisten ilmauksien käyttö selittyikin usein tiettyjen asioiden tai henkilöhahmojen toistumisella.

Kertomus:

18) *– Herätys Helka! isä huudahti. – Nyt lähdetään sukeltamaan merelle, isä lisäsi. – Mutta kellohan on vasta kuusi aamulla, Helka sanoi väsyneenä. – Jos tahdot nähdä kaloja, niin mennään! isä huudahti. Aamu Thaimaan saarella oli hyvin lämmin. Eipä aikaakaan, kun minä ja isi olimme sukellusvehkeet niskassa. Hyppäsimme veteen. Meidän mukana oli thaimaalainen joka opasti meidät koralliriutoille.*

*Koralleja oli erivärisiä. Simpukat olivat todella hienoja. Mutta sitten tuli odottamani kohta. Pusukala tuli ja pussasi minua! Se oli todella hienoa. Keräsin sukellus reppuun simpukoita, korallia ja otin vesikamerallani kuvia pusukalasta. Nousimme pintaan ja sanoin: – Kiitos isi! Isikin piti sukeltamisesta. Loppu. (4F23)*

Sukupuolittain tarkasteltuna erot koko aineistossa ovat selvät. Ylipäätään mää räisiä ilmauksia käyttävien toisluokkalaisten tyttöjen ja poikien lukumäärä on lähes sama, mutta suurin osa pojista käyttää vain yhtä tai kahta mää räistä ilmausta kun taas tytöillä suurin ryhmä on kolmesta neljään mää räistä ilmausta keräävät tekstit. Suurin tiheys on pojilla 11 mää räistä ilmausta tekstissä, tytöillä 25. Suuret tiheydet ovat kuitenkin yksittäistapauksia. Neljäsluokkalaisten poikien aineistossa vain muutaman mää räisen ilmauksen sisältävät tekstit kattavat lähes puolet teksteistä. Tyttöjen teksteissä esiintymämäärä vaihtelee enemmän. Kuudennen teksteissä erot ovat sukupuolittain tarkasteltuna vielä alempia luokkia selkeämpiä. Tyttöjen teksteissä käytetään runsaasti mää räisiä ilmauksia, ja vain vajaassa viidenneksessä on alle seitsemän mää räistä ilmausta. Poikien teksteille taas on tyypillistä, että mää räisten ilmauksien käyttömäärä jää asteikon alapäähän: alle kymmenen mää räistä ilmausta on neljässä viidesosassa poikien teksteistä.

Sukupuolierot näkyvät selvästi myös kaikkien luokka-asteiden tekstityyppijakaumista (ks. taulukko 29). Yleisimmät mää räisiä ilmauksia sisältävät tekstityypit ovat toisella luokalla luettelomaiset kertomukset ja luettelot. Selvästi yli puolet toisluokkalaisten luettelomaisista kertomuksista on tyttöjen kirjoittamia ja luetteloista runsas puolet vastaavasti poikien. Selkein ero sukupuolten välillä on tyttöjen kertomusmaisten tekstien suuri määrä. Neljännellä ja kuudennella luokalla tekstityyppien painopiste siirtyy kertomusmaisiin teksteihin ja luettelomaisiin kertomuksiin. Kummallakin luokka-asteella yli puolet tyttöjen teksteistä on kertomusmaisia ja poikien puolella yleisyyttään kasvattavat luettelomaiset kertomukset. Samalla kuvausten ja luetteloiden määrä vähenee molempien sukupuolten aineistossa.

Taulukko 29. Määräisiä ilmauksia sisältävät tekstit tekstityyppijatkumon mukaan alakoulun luokilla.

Määräiset ilmaukset ja tekstin tyyppi / Ikäluokka ja sukupuoli	Toinen Tytöt	Toinen Pojat	Neljäs Tytöt	Neljäs Pojat	Kuudes Tytöt	Kuudes Pojat
Kertomusmainen	15	5	23	5	15	5
Luettelomainen	29	20	8	23	11	23
Keskeneräinen	2	0	3	0	0	1
Kuvaus	3	10	4	10	0	1
Luettelo	17	25	2	11	2	8

## MÄÄRÄISTEN ILMAUSTEN KÄYTÖN KEHITTYMINEN JA KERTOMUKSEN SYNTY

Määräisten ilmauksien määrä kasvaa alakoululaisten teksteissä kehityksellisesti ja samalla tekstin kertomusmaisuus kasvaa. Määräisten ilmauksien ja tekstityyppien suhde näkyy selvimmin, jos lasketaan jokaiselle tekstityypille luokittain sille ominainen määräisten ilmauksien mediaani. Eri tekstityypeistä kertomusmaiset tekstit ja keskeneräiset kertomukset saavat korkeimmat määräisten ilmauksien mediaanit. Toisella luokalla kertomusmaisten tekstien mediaani on 9,5, neljäsluokkalaisilla 10 ja kuudesluokkalaisilla peräti 16. Myös keskeneräisten kertomusten mediaanit liikkuvat melko samoissa lukemissa. Kertomusmaisuuuden edellytyksenä vaikuttaisi siis olevan melko runsas määräisten ilmauksien määrä. Luonnollisesti se on seurausta juonen kannalta olennaisten asioiden, kuten henkilöhahmojen, toistumisesta ja niihin viittaamisen runsaudesta. Kuudennen luokan korkeat mediaanit selittyvät osin myös tekstien piteneemisellä (ks. Koulukorpuksen sanamääristä teoksen johdantolukua): pidempään tekstiin mahtuu enemmän uudelleenmainintoja. Tosin teksti voi alakoululaisella olla pitkä myös luettelevuuden ja irrallisten lisien takia.

Kertomusmaisten tekstien ja keskeneräisten kertomusten väliin sijoittuvat luettelomaiset kertomukset sisältävät selvästi vähemmän määräisiä ilmauksia (mediaani 4–6) kuin kertomusmaisuuuteen pohjaavat tekstit. Luettelomaiset kertomukset esittelevät tapahtumista vain pääkohdat, minkä takia kokonaisuus muistuttaa juonirunkoa. Tapahtumasta toiseen siirrytään nopeasti, jolloin toistuvien asioiden määrä vähenee. Toistoa – ja sitä kautta määräisten ilmauksien määrää – vähentää myös

kerronnan pintapuolisuus, sillä luettelomaisissa kertomuksissa asioita ja tapahtumia ei esitellä tai kuvailla tarkasti. Koska teksti pitää sisällään paljon tapahtumia, jotka vaihtuvat tiheään tahtiin, lukijalle esitellään paljon uusia tarkoituksia sen sijaan, että viitattaisiin toistuvasti jo tuttuun tarkoituksiin ja rakennettaisiin teksti niiden varaan.

Kaikkein matalimmat määräisten ilmausten mediaanit (1–3,5) löytyvät odotuksenmukaisesti tekstityyppijatkumon luettelomaisimmista teksteistä eli kuvauksista ja luetteloista. Niiden matalat mediaanit selittynevät toiston vähäisyydellä sekä tekstin sisältämien asioiden ja tapahtumien irrallisuudella. Luettelot sisältävätkin runsaasti epämääräisiä ja geneerisiä ilmauksia, jotka viittaavat epätarkkaan tai yleiseen joukkoon. Kuvauksiksi luokiteltavat tekstit taas ovat monesti niin lyhyitä, ettei niissä edes ehditä viitata kovinkaan moneen tarkoitteeseen saati toistaa viittauksia tekstin aikana.

Referentiaalinen koherenssi kasvaa siten selvästi kertomusmaisuu-den – ja kertomusrakenteen – myötä. Tavallaan tämä on kehäpäätelmä, koska kertomus edellyttää koherenssia, eli kertomus määrittellään ainakin osaksi referentiaalisten viittausten avulla. Silti on vaikea sanoa, onko yhteys kausaalinen (vrt. Mäkinen ja muut 2014). Referentiaalisia ilmauksia voi nimittäin olla melko runsaasti myös luettelomaisissa kertomuksissa. Lisäksi jo alle kouluikäiset lapset ymmärtävät ja tuottavat referentiaalisia viittauksia puhutussa kielessä, vaikka he eivät osaisi lukea ja kirjoittaa. Kirjoittamisen vaikeus ja siinä onnistuminen on siis referentiaalisista viittauksista riippumaton ongelma.

## Alakoululaisten *se*-tarkenteelliset määräiset ilmaukset

Määräisten ilmauksien lisääntyminen kertomuksissa on paitsi kehityksellistä myös hidasta. Tarkastelemme lopuksi, ovatko artikkelinomaiset *se*-tarkenteet, jotka ovat tuttuja puhutussa kielessä, löytäneet tiensä kehittyvään kirjoitukseen helpottamaan määräisyyden ilmaisua. Yleensä oletetaan, että lapset ensin kirjoittavat kuten puhuvat (ks. esimerkiksi Berman 2007). Puhutun kielen vaikutus on myös Myöhemmän kielenkehityksen tutkimushankkeen yksi taustahypoteesi, jonka mukaan aloit-

televa kirjoittaja nojaa puhuttuun kieleen, josta sitten niin sanottu akateeminen kielenkäyttö irtautuu vähitellen (ks. teoksen johdantoluvun perusoletuksia). Tarkastelumme poikkeaa aiemmasta artikkelistumistutkimuksesta sekä kehitysnäkökulman että mahdollisiin esiintymisiin vertaamisen perusteella. Näin saadaan näkyviin tendenssin ja satunnaisuuden mahdollinen ero.

Puhutussa suomessa käytetään erilaisia tarkenteita määräisyyttä ilmaisemaan (Hakulinen ja muut 2004, 1353–1356). Tarkenteena voivat toimia esimerkiksi demonstratiivipronominin ja demonstratiiviset proadverbit, kuten esimerkiksi viittausketjussa (*yks mies* – *se mies*). Näistä erityisesti artikkelimaisen *se*-pronominin tarkennekäyttöä on tutkittu suomen osalta pitkään, sillä puhutussa kielessä painoton *se* on niin tavallinen, että sen on katsottu tehtävänsä osalta lähestyvän määräistä artikkelia. Tämä ei ole yllättävä kehitys: esimerkiksi Larjavaaran (2007, 294–295) mukaan kaikissa – artikkelittomissakin – kielissä käytetään demonstratiiveja ilmaisemaan substantiivin tarkoitteen mainittuutta tai muuta tuttuutta. Lauryn (1996, 162) mukaan jo Setälä (1883) oli havainnut, että *se*-pronominia käytetään puhemielessä määräisen artikkelin tapaan. Setälän jälkeen *se*- ja *ne*-pronominin tarkenteiden käyttöön ja artikkelimaistumiseen ovat ottaneet kantaa muun muassa Larjavaara (1990), Laury (1996) sekä Juvonen (2000). Artikkelistumiskeskustelu ei kuitenkaan ole johtanut yksimielisyyteen, ja joka tutkijalla tuntuu olevan oma näkemyksensä tarkenteen käyttötiheydestä ja tehtävästä.

*Se*-tarkenteellisia ilmauksia tarkastellaan kaikilta alakoulun luokilta, jotta esimerkkejä löytyisi riittävä määrä. Aineistossa on nyt yhteensä 40 000 sanaa ja tässä sanamäärässä *se*- tai *ne*-demonstratiivipronominia käytetään substantiivia määrittävänä tarkenteena vain 78 kertaa. Läheskään kaikki kirjoittajat eivät ylipäättään käytä tarkenneilmauksia, vaan niitä esiintyy vain noin joka kymmenennessä (11,5 %) tekstissä. Sukupuolittain tarkasteltuna tilanne on tasainen. Tarkenteellisten ilmausten käyttäminen on harvinaista myös siinä mielessä, että suurin osa käyttää *se*- tai *ne*-tarkennetta tekstissään vain kerran (ks. taulukko 30). Siten myös tekstikohtaisten esiintymämäärien perusteella käyttö näyttää harvinaiselta. Toisen, neljännen ja kuudennen luokan aineistojen kaikkiin määräisiin ilmauksiin suhteutettuna tarkenneilmauksien osuus jää kul-

lakin vuosiluokalla alle kolmeen prosenttiin. Prosenttiosuus näyttäisi vielä pienenevän iän mukaan, mutta aineisto on varsinaiseen kvantitatiiviseen tarkasteluun hieman pieni. Kokonaisuudessaan voi todeta, että tarkenneilmauksien käyttö on sattumanvaraista eikä osoita *se-* ja *ne-*artikkelin syntyä tai yleistymistä kehittyvään kirjoitettuun kieleen.

Taulukko 30. *Se/ne*-tarkenteiden määrä teksteittäin eri ikäluokilla.

Tarkenteiden määrä tekstissä / Ikäluokka	Toinen	Kolmas	Neljäs	Viides	Kuudes
Yksi tarkenne	10	14	9	11	7
Kaksi tarkennetta	2	4	1	1	1
Kolme tarkennetta	1			2	

Suurin osa (79,5 %) aineiston *se*-tarkenteellisista määräisistä ilmauksista on luokiteltavissa määräisyyden tyyppitapauksiksi eli uudelleenmaininnoiksi, joiden avulla viitataan johonkin tekstissä jo aiemmin mainittuun tarkoitteeseen. Uudelleenmaininnalla voidaan viitata tuttuun tarkoitteeseen joko aiemmin mainitun substantiivin avulla (esimerkki 19) tai esimerkiksi synonyymisellä ilmauksella (esimerkki 20). Uudelleenmaininnan kriteeri ei siis ole saman sanan tai lausekkeen esiintyminen uudelleen tekstissä, vaan kaikki samaan tarkoitteeseen viittaavat sanat lasketaan uudelleenmainintojen ryhmään. Kokonaisuudessaan uudelleenmainintojen ryhmä on kuitenkin varsin heterogeeninen, ja valtaosa ryhmän tarkenteellisista ilmauksista on yksittäistapauksia, jolloin yksi määräinen ilmaus merkitään tarkenteella vain yhden kerran tekstin aikana. Uudelleenmainintojen systemaattinen merkitseminen pronomini-tarkenteella ei siis ole tavanomaista oppilaiden tuottamissa koulukirjoitelmissa.

19) *Olen ratsastus leirillä siellä oli paras hoitohevonen Sillä minä hyppin esteitä ja laukkasin ja kävin maastossa ja nukuin tallin ylä kerrassa valvoin yöllä ja pesin varusteita mutta sitten aamulla ne varusteet oli märät ja mun pitilainata tunnille toisen hevosen satulaa. (3F1)*

20) *Olipa kerran pieni makkaran näköinen koira. Se meni lottoamaan. Ja 2 tunnin päästä se sai Lotto voiton. Mutta koira rupesi katu-  
maan mitä oli oikein tehnyt. Koska hän ei tiennyt mitä tekisi niillä rahoilla. (2F41)*

Muita aineistossa esiintyviä *se*-tarkenteellisen määräisen ilmauksen viittaussuhteita ovat kehyksen kautta tuttuun tarkoitteeseen viittaaminen (11,5 %) tai ensimmäinnän käyttäminen (9 %). Kehyksen kautta tunnetun tarkoitteen tuttuus perustuu toiston sijaan jonkin edellä esitetyn ilmauksen aktivoimaan kehykseen (esimerkki 21). Jo ensimmäisellä mainintakerrallaan tarkenteella merkityn ilmauksen avulla taas tuodaan keskusteluun uusi tarkoite, jonka tekstin tuottaja olettaa olevan jossain mielessä tuttu vastaanottajalle. Tarkenteellisen ensimmäinnän tuttuus voi perustua esimerkiksi tekstin henkilöhahmojen yhteiseen tietoon tekstin kuvaamasta maailmasta tai siihen, että tarkoite saa selityksensä tai tulkintakontekstinsa pian ensimmäinnän jälkeen (esimerkki 22).<sup>50</sup> Sekä kehyksen kautta tuttuun tarkoitteeseen viittaavat että jo ensimmäinnassa esiintyvät pronominitarkenteet ovat kuitenkin hyvin harvinaisia tässä aineistossa.

21) *Tai sitten voisin mennä väikka pariisiin. Ja pariisissa voisi katsella sitä ihanaa tornia joka pariisissa on.* (3F39)

22) *Pääsen idolini Rihannan konserttiin, sillä hän on saapunut Suomeen, joka on yksi niistä maista, joissa hän kiertueellansa käy.* (6F19)

## Lopuksi

Referentiaalinen koherenssi eli tunnistettava diskurssireferentteihin viittaaminen ja ylipäättään koherenssi eli tekstien yhtenäisyys, loogisuus, tapahtumajärjestyksen huomioonottaminen, ajallinen ja paikallinen sijoittuminen jne. lisääntyvät kehityksellisesti. Kehitys alkaa puhutun kielen yhteydessä jo esikouluiässä, mutta kirjoittamisen osalta kehitys on paljon hitaampaa. Osa alakouluikäisistä hallitsee viittaussuhteiden ilmaisun jo alakoulun alussa, mutta keskimäärin kehitys nopeutuu alakoulun keskivaiheilla ja jatkuu koko alakoulun ajan. Vaikka lapsi hallit-

<sup>50</sup> Perinteisen kieliopin mukaan esimerkissä (22) on kyse kataforisesta, tekstissä eteenpäin viittaavasta pronominita (ks. Penttilä 1963, 514) eikä varsinaisesta artikkelimaisesta käytöstä.



sisi monia viittaussuhteiden kielellisen merkinnän keinoja jo varhain, aikuistason semanttis-pragmaattinen hallinta saavutetaan (keskimäärin) alakoulun loppupuolella. Kehityksen taustalla on monia kielenulkoisia tekijöitä. Kronologisen iän lisäksi mielen teorian on oltava riittävän kehittynyt, jotta lapsi osaa ottaa huomioon viestin vastaanottajaa. Vastavasti lapsen on kyettävä riittävästi kontrolloimaan toimintaansa, jotta hän pystyy suunnittelemaan tekstin tuottoa. Olemme tässä luvussa tutkineet referentiaalisten ilmauksien käyttöä rajaamalla tarkastelun määräisiin ilmauksiin eli sellaisiin substantiivilausekkeisiin, joita voidaan tarvittaessa rajata pronomineilla. Tutkimuksen aineistona on ollut alakouluikäisten lasten vapaasti tuottamia kirjoitelmia annetusta aiheesta eli niin kutsuttuja Unelmakirjoitelmia.

Kirjoitelmien koherenssi lisääntyy asteittain siten, että sekä määräisten ilmauksien määrä että kirjoitelmien kerronnallisuus kertomusrakenteen kautta määriteltynä lisääntyvät. Nuorimpien tekstit ovat usein yhden lauseen epätekstejä tai eriasteisesti epäkoherentteja. Tekstikoherenssi kasvaa kirjoitelmien rakenteen alkaessa noudattaa kertomusrakennetta ja viittaussuhteiden kehittyessä täsmällisiksi. Suurin osa aineiston kirjoittajista on tässä vaiheessa vasta alakoulun lopussa. Tytöillä tekstikoherenssi kehittyi eritahtisesti poikien kanssa.

Määräisillä ilmauksilla ilmaistaan niin sanottua tuttua informaatiota eli tekstiin tavalla tai toisella jo tuotua informaatiota, jota jatkossa täydennetään. Tutun informaation referentin täytyy olla palautettavissa kontekstista, sen status voi olla havainnon kannalta salientti tai sitten tuttuus perustuu oletettuun tuttuuteen. Nuorimmilla alakouluikäisillä oletettu tuttuus on keskeinen tuttuuden muoto. Alakouluikäinen olettaa elinympäristönsä kuulija-lukijan kannalta tutuksi ja tuo perheeseen, kouluun ja ystäviin viittaavia ilmauksia esittelemättä tekstiin. Osaksi tuttuus on ymmärrettävissä tilanteisen ja kulttuurisen aktiivoinnin kautta. Jos teksti on hyvin epäkoherentti, määräisiä ilmauksia on vähän tai ei ollenkaan: teksti luettelee unelmakohteita eikä kerro unelmapäivän sisällöstä.

Alakouluikäisten määräisiä ilmauksia ei voi arvioida esimerkiksi Gundelin ja muiden (1993) tapaan implikaatiivisella annettuusasteikolla, koska huomion kohteena ja aktivoituneena voi olla myös ilmaus, joka ei ole identifioitavissa tekstin tasolla. Siten kehittyvää kieltä voidaan mää-

räisten ilmauksien kannalta arvioida myös suhteessa siihen, miten ne vielä rikkovat erilaisia asteikkoja – kuten tekstikoherenssin asteikkoa, annettuusasteikkoa, saavutettavuusasteikkoa ja niin edelleen. Alakouluikäinen ja häntä nuorempikin kyllä ymmärtää referentiaalisia ilmauksia ja niiden puutteita ja kykenee myös tuottamaan niitä mutta parhaiten ohjeistetussa kontekstissa. Erityisesti luova kirjoittaminen on lapsen kannalta haastavin konteksti. Tämä johtuu suorituksen moniulottuvuudesta ja haastavuudesta ja tietoisien referentiaalisten ilmauksien tuottamisen kuormittavuudesta.

Määräisten ilmausten käytöllä on yhteys tekstin kokonaisrakenteen edustamaan tekstityyppiin: sekä tekstien kertomusmaisuus että määräisten ilmausten tekstikohtainen esiintymismäärä kasvavat alemmilta luokilta ylemmille siirryttäessä. Määräisten ilmausten käytön lisääntyminen koulunkäynnin edetessä näyttää siis liittyvän kertomusmaisesta tekstilajin hallintaan, kun luettelomaisuus tekstissä samalla vähenee. Määräiset ilmaukset eivät kuitenkaan lisäänty tasaisesti, vaan määräisten ensimmäisten osuus vähenee ja määräisten uudelleenmainintojen osuus kasvaa, kun siirrytään alemmilta luokilta ylemmille. Tämä voi olla merkki esimerkiksi siitä, että tiettyihin tarkoitteisiin viitataan toistuvasti tekstin aikana eli ne pysyvät mukana tekstissä aiempaa kauemmin. Määräisten uudelleenmainintojen suhteellisen osuuden kasvu voi liittyä myös tekstin sisäisen koherenssin kasvuun, sillä yksi yhtenäisen tekstin piirre on samasta tai samoista tarkoitteista puhuminen läpi koko tekstin. On kuitenkin muistettava, että tekstissä toistuviin tarkoitteisiin voi viitata myös pronominin avulla, mihin tämä tutkimus ei kiinnitä huomiota. Tarkoitteiden toistuvuudesta ja tekstin koherenssista saisikin kattavamman kuvan tutkimalla myös pronominiäityksiä ja selvittämällä, kuinka usein yhden tekstin määräiset ilmaukset liittyvät samaan tarkoitteeseen.

Kun määräisiä ilmauksia tarkastellaan lajeittaisten käyttömäärien näkökulmasta, havaitaan sekä ensi- että uudelleenmainintojen kohdalla, että tarkenteella merkittyjen substantiivilausekkeiden osuus kasvaa ylemmille luokille siirryttäessä. Tämä voi viitata esimerkiksi tekstien tarkoitteiden aiempaa huolellisempaan esittelyyn ja määrittelyyn. Tarkoitteisiin joko esitellään määräisenä heti alusta alkaen tai sidotaan myöhemmässä vaiheessa esimerkiksi jonkin henkilöihahmon ominaisuudeksi.

Tarkoitteiden huolellisempaan määrittelyyn liittyy myös kvantitatiivisesti määräisten ensimmäintöjen osuuden kasvu. Pelkän mainittuuden perusteella määräisten uudelleenmainintojen lisääntyminen taas viitanee tarkoitteiden pysyvyyteen ja toistuvuuteen tekstissä. Tämä voi esimerkiksi pienentää tekstin sisältämien tarkoitteiden kokonaismäärää, mihin viittaa sekä ensi- että uudelleenmainintojen kohdalla havaittava tuttuihin tarkoitteisiin viittaavien ilmauksien osuuden väheneminen. Tällöin tekstit sisältävät esimerkiksi aiempaa pienemmän henkilöhaamojoukon.

Vaikka tekstien kokonaistarkoittemäärä vähenisikin ylemmille luokka-asteille siirryttäessä, kasvavat tekstien sisältämät määräisten ilmauksien esiintymämäärät. Täytyy kuitenkin muistaa, että samalla kasvaa myös tekstikohtainen kokonaissanamäärä, joten määräisten ilmausten esiintymistiheys voi olla vanhempien informanttien teksteissä lähes samaa luokkaa kuin nuorimmilla kirjoittajilla. Koska määrällistä kehitystä kuitenkin tapahtuu juuri määräisten ilmauksien suurimpien tekstikohtaisen esiintymämäärien kohdalla, niiden käyttö lisääntyy kirjoitustaidon kehittyessä. Tähän viittaa myös se, että vain muutamia määräisiä ilmauksia sisältävien tekstien osuus kaikista luokka-asteen teksteistä pienenee ylemmille luokka-asteille siirryttäessä.

Samalla kun määräisten ilmauksien esiintymämäärät kasvavat ja ensi- ja uudelleenmainintojen lajeittaiset osuudet muuttuvat, lisääntyy myös tekstien kertomusmaisuus: kuudesluokkalaisten teksteistä nimitäin jo reilusti yli puolet on luokiteltavissa kertomusmaisiksi teksteiksi tai luettelomaisiksi kertomuksiksi. Mediaanien valossa näyttäisikin siltä, että kertomusmaisuus vaatii muita tekstityyppisiä enemmän määräisiä ilmauksia. Tämä selittyy sillä, että hyvässä kertomusmaisessa tekstissä tietyt asiat toistuvat usein ja pysyvät mukana kertomuksessa, jolloin myös määräisten ilmauksien kokonaismäärä väistämättä kasvaa.

Määräisten ilmauksien esiintymämäärä ei kuitenkaan näytä vaikuttavan *se-* ja *ne-*demonstratiivitarkenteella merkittyjen ilmausten esiintymiseen. Ne ovat harvinaisia joka luokka-asteella, ja pääasiassa niitä käytetään vain kerran tekstin aikana. Aineiston valossa näyttäisikin siltä, ettei suomen kielen *se-*pronomini ole ainakaan laajamittaisesti kehittymässä kirjoitettuun kieleen määräiseksi artikkeliksi.

## LÄHTEET

- Ackerman, B. 1979. Children's understanding of definite descriptions: Pragmatic inferences to the speaker's intent. *Journal of Experimental Child Psychology* 28, 1–15.
- Ackerman, B. 1981. Encoding specificity in the recall of pictures and words in children and adults. *Journal of Experimental Child Psychology* 31, 193–211.
- Arnold, J. E. 2008. Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes* 23: 4, 495–527.
- Astington, J. W. & Baird, J. A. (toim.) 2005. *Why language matters for theory of mind*. Oxford: Oxford University Press.
- Bamberg, B. 1983. What makes a text coherent? *College composition and Communication* 34: 4, 417–427.
- Bamberg, M. (toim.) 1997. *Narrative development: Six approaches*. Mahwah: Lawrence Erlbaum Ass.
- Beal, C. R. & Belgrad, S. L. 1990. The development of message evaluation skills in young children. *Child Development* 61, 705–712.
- Berman, R. A. & Slobin, D. 1994. Narrative structure. Julkaisussa: R. A. Berman & D. Slobin (toim.) *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale: Lawrence Erlbaum Ass., 39–84.
- Berman, R. A. 2007. Developing linguistic knowledge and language use across adolescence. Julkaisussa: E. Hoff & M. Shatz (toim.) *Blackwell handbook of language development*. Malden: Blackwell Publishing, 347–367.
- Birner, B. 2006. Inferential relations and non-canonical word order. Julkaisussa: B. Birner & G. Ward (toim.) *Drawing the boundaries of meaning*. Amsterdam: John Benjamins, 31–51.
- Brown, P. & Levinson, S. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Chafe, W. L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. Julkaisussa: C. Li (toim.) *Subject and topic*. New York: Academic Press, 25–55.
- Chafe, W. L. 1987. Cognitive constraints on information flow. Julkaisussa: R. Tomlin (toim.) *Coherence and grounding in discourse*. Amsterdam: John Benjamins, 21–51.
- Chafe, W. L. 1997. Polyphonic topic development. Julkaisussa: T. Givón (toim.) *Conversation: Cognitive, communicative and social perspectives*. Amsterdam: John Benjamins, 41–55.
- Dryer, M. S. 2013. Definite articles. Julkaisussa: M. S. Dryer, M. Haspelmath (toim.) *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/37>. Viitattu 31.5.2020.
- Genette, G. 1980. *Narrative discourse: An essay in method*. Oxford: Basil Blackwell.
- Givón, T. 1983. Topic continuity in discourse: An introduction. Julkaisussa: T. Givón (toim.) *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins, 1–42.
- Givón, T. 2017. *The story of zero*. Amsterdam: John Benjamins.
- Glucksberg, S., Krauss, R. M. & Higgins, E. T. 1975. The development of referential communication skills. Julkaisussa: F. D. Horowitz (toim.) *Review of child development research*. Volume 4. Chicago: University of Chicago Press, 305–345.

- Gundel, J. K., Hedberg, N. & Zacharski, R. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69: 2, 274–307.
- Habermas, T. & de Silveira, C. 2008. The development of global coherence in life narratives across adolescence: Temporal, causal and thematic aspects. *Developmental Psychology* 44: 3, 707–721.
- Hakulinen, A., Vilkkuna, M., Korhonen, R., Koivisto, V., Heinonen, T.-R., Alho, I. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Halliday, M. A. K. & Hasan, R. 1976. *Cohesion in English*. New York: Longman.
- Heikkinen, V. 2012. Teksti. Julkaisussa: V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiilikä & M. Lounela (toim.) *Genreanalyysi – tekstilajintutkimuksen käsikirja*. Helsinki: Suomalaisen Kirjallisuuden Seura, 59–66.
- Heikkinen, V. & Voutilainen, E. 2012. Genre – monitieteinen näkökulma. Julkaisussa: V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiilikä & M. Lounela (toim.) *Genreanalyysi – tekstilajintutkimuksen käsikirja*. Helsinki: Suomalaisen Kirjallisuuden Seura, 17–47.
- Heinonen, S. 2018. *Narratiivisuuden sisällöllinen kehittyminen aiheanalyysin kautta tarkasteltuna: kvantitatiivinen tutkimus aiheiden esiintymisen vaihtelusta iän ja sukupuolen mukaan "Unelmieni päivä" -kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Hendriks, P. 2014. *Asymmetries between language production and comprehension*. Dordrecht: Springer.
- Hickman, M. 2004. Coherence, cohesion, and context: Some comparative perspectives in narrative development. Julkaisussa: S. Strömqvist & L. Verhoeven (toim.) *Relating events in narrative: Typological and contextual perspectives*. Amhwhah: Lawrence Erlbaum Ass., 281–306.
- Hoff, E. 2009 [2005]: *Language development*. Belmont: Wadsworth.
- John, A. E., Rowe, M. L. & Mervis, C. B. 2009. Referential communication skills of children with Williams syndrome: Understanding when messages are not adequate. *AJIDD*, 114: 2, 85–99.
- Juvonen, P. 2000. *Grammaticalizing the definite article. A study of definite adnominal determiners in a genre of spoken Finnish*. Doctoral dissertation. Stockholm: Stockholm University.
- Labov, W. & Waletzky, J. 1967. Narrative analysis: Oral versions of personal experiences. Julkaisussa: J. Helm (toim.) *Essays on the verbal and the visual arts*. Seattle: American Ethnological Society, 12–44.
- Larjavaara, M. 1990. *Suomen deiksis*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Larjavaara, M. 2007. *Pragmasemantiikka*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Lauri, R. 1996. Sen kategoriasta – onko suomessa jo artikkeli? *Virtittäjä* 100: 2, 162–181.
- Loock, R. 2013. Extending further and refining Prince's taxonomy of given/new information. A case study of non-restrictive, relevance oriented structures. *Pragmatics* 23: 1, 69–91.
- Loukusa, S., Mäkinen, L., Gabbatore, I., Laukkanen-Nevala, P., Leinonen, E. 2017. Understanding contextual and social meaning in typically developing Finnish-speaking four- to eight-year-old children. *Psychology of Language and Communication* 21: 1, 408–428.
- Maier, E. 2009. Japanese reported speech: Against a direct–indirect distinction. Julkaisussa: H. Hattori, T. Kawamura, T. Idé, M. Yokoo & Y. Murakami (toim.) *New Frontiers in Artificial Intelligence*. Berlin: Springer, 133–145.

- Matilainen, K. 1989. *Kirjoitustaidon kehittyminen neljän ensimmäisen kouluvuoden aikana*. Joensuu: Joensuun yliopisto.
- Mäkinen, L., Loukusa, S., Nieminen, L., Leinonen, E. & Kunnari, S. 2014. The development of narrative productivity, syntactic complexity, referential cohesion and event content in four- to eight-year-old Finnish children. *First Language* 34: 1, 24–42.
- Pajunen, A. 2012. Kirjoittamistaitojen kehitys 8–12-vuotiailla: Alakoululaisten unelmakirjoitelmät. *Virtittäjä* 116: 1, 3–34.
- Pajunen, A., Puranen, M. & Yli-Paavola, A. 2010. Koulukorpus. Alakoululaisten kirjoitelmia aiheesta Unelmieni päivä. Excel-tietokanta.
- Penttilä, A. 1963. *Suomen kielioppi*. Toinen, tarkistettu painos. Porvoo: WSOY.
- Pietikäinen, S. & Mäntynen, A. 2009. *Kurssi kohti diskursssia*. Tampere: Vastapaino.
- Prince, E. 1981. Toward a taxonomy of given/new information. Julkaisussa: P. Cole (toim.) *Radical pragmatics*. New York: Academic Press, 223–254.
- Prince, E. 1992. The ZPG letter: Subjects, definiteness, and information-status. Julkaisussa: S. Thompson & T. Mann (toim.) *Discourse description: Diverse analyses of a fundraising text*. Amsterdam: John Benjamins, 295–325.
- Schneider, P, Vis Dubé, R., & Hayward, D. 2005. The Edmonton narrative norms instrument. <http://www.rehabresearch.ualberta.ca/enni>. Viitattu 15.3.2020.
- Schonnenschein, S. & Whitehurst, G. J. 1983. Training referential communication: The limits of success. *Journal of Experimental Child Psychology* 35, 426–436.
- Schonnenschein, S. & Whitehurst, G. J. 1984. Developing referential communication: A hierarchy of skills. *Child Development* 55, 1936–1945.
- Sekerina, I. A. 2015. Online evidence for children’s interpretation of personal pronouns. Julkaisussa: L. Serratrice & S. E. M. Allen (toim.) *The acquisition of reference*. Amsterdam: John Benjamins, 213–239.
- Serratrice, L. & Allen, S. E. M. 2015. Introduction: An overview of the acquisition of reference. Julkaisussa: L. Serratrice & S. E. M. Allen (toim.) *The acquisition of reference*. Amsterdam: John Benjamins, 1–24.
- Serratrice, L. & De Cat, C. 2020. Individual differences in the production of referential expressions: The effect of language proficiency, language exposure and executive function in bilingual and monolingual children. *Bilingual Language and Cognition* 23: 2, 371–386.
- Setälä, E. N. 1883. *Lauseopillinen tutkimus Koillis-Satakunnan kansankielestä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Shore, S. & Mäntynen, A. 2006. Johdanto. Julkaisussa: A. Mäntynen, S. Shore & A. Solin (toim.) *Genre – tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura, 9–41.
- Sintonen, M. 2011. *Suunnittelemisen vaikutus kertomuksen rakenteeseen viides- ja seitsemäsluokkaisilla kirjoittajilla*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Sperber, D. & Wilson, D. 1986. *Relevance: Communication and cognition*. Oxford: Blackwell.
- Suvanto, A. 2012. *Lapsi tarinaa rakentamassa. Kielihäiriöisten lasten kerrontataidot ja niiden kuntouttaminen*. Oulu: Oulun yliopisto.
- Tomasello, M. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Vesala, S. 2012. *Kerronnan keinot alakoululaisten teksteissä*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.

- Vilkuna, M. 1992. *Referenssi ja määräisyys suomenkielisten tekstien tulkinnassa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Vilkuna, M. 2000. *Suomen lauseopin perusteet*. 2., korjattu painos. Helsinki: Edita.
- Virtanen, T. 2010. Variation across texts and discourses: Theoretical and methodological perspectives on text type and genre. Julkaisussa: H. Dorgeloh & A. Wanner (toim.) *Syntactic variation and genres*. Berlin: Walter de Gruyter, 53–84.
- Werlich, E. 1974. *Typologie der Texte. Entwurf eines Textlinguistischen Modells zur Grundlegung einer Textgrammatik*. Berlin: Quelle & Meyer.

# Moniäänisyyden ilmaisun kehittyminen peruskoulun päättövaiheiden kirjoitelmissa

*Anneli Niinimäki*

ORCID <https://orcid.org/0000-0003-0050-10>

## Johdanto

Moniäänisyys tarkoittaa Bahtinin (1991 [1963]) mukaan kerronnan kielellistä moniaineisuusia, joka tulee esiin paitsi kertomuksen rakenteessa myös sen välittämässä ideologiassa. Moniäänisyyttä on tekstiin rakennettu moninäkökulmaisuus ja tulkintojen vaihtoehtoisuus. Muutokset kertojan tai puhujan osallistujaroolissa rakentavat moniäänisyyttä. Moniäänisyys näkyy, kun jokin kertomuksen aines aktivoituu uudessa kehyksessä. Martin ja White (2005, 102–104) katsovat moniäänisiksi sellaiset ilmaukset, joissa aiemmat näkemykset ja aiheesta aiemmin sanottu tuodaan tekstissä jollakin tavalla näkyviin. Yksiäänisiä puolestaan ovat ilmaukset, joissa ei näy, että asiasta on tai voisi olla useampia tulkintoja (Martin & White 2005, 98–102). Useissa viimeaikaisissa tutkimuksissa moniäänisyyttä tarkastellaan lisäksi vuorovaikutuksen näkökulmasta (ks. esimerkiksi Juvonen 2014; Virtanen, Rahtu & Shore 2018). Moniäänisyyden ilmenemismuotoja ovat muun muassa intertekstuaalisuus, ironia, metatekstuaalisuus, retoriset kysymykset sekä avoin toiseen puhe-tilanteeseen viittaaminen eli referointi. Moniäänisyyttä tuotetaan eri



tekstilajeissa erilaisin kielellisin valinnoin, ja ne synnyttävät erilaisia merkityksiä ja tulkintoja. Kertovassa tekstissä moniäänisyys on aikuisten kohdalla tyypillisesti tekstien välistä intertekstuaalisuutta sekä saman tekstin sisäistä toiseen puhetilanteeseen viittaamista, toisen tai oman puheen referointia (ks. esimerkiksi Kalliokoski 2005a).

Kehittyvässä (kertovassa) kielessä moniäänisyyden ilmaisu on vielä vakiintumatonta ja aluksi hyvin yksinkertaista. Pääosin ilmaisu on yksiaänistä. Kaikki moniäänisyyden lajit tuovat teksteihin kerroksellisuutta ja kompleksistavat myös tekstin syntaktista rakennetta, ja siksi on jo lähtökohtaisesti oletettavaa, että moniäänisyyttä rakentavien ilmaisukeinojen tavoittaminen vaatii aikaa. Unelmakirjoitelmissa ehkä yksinkertaisinta moniäänisyyttä edustavat alakoululaisten kirjoitelmien yksittäiset vieraskieliset sanat, erisnimet, kuten pelien nimet sekä auktoriteettien – kuten isän ja äidin – mielipiteitä kaiuttavat asenteiden ilmaisut ja lukijan puoleen kurkottaminen (Unelmakirjoitelmista ks. tämän teoksen johdantoa). Aloituksissa ja lopetuksissa voidaan irrottautua kerronnasta metakielisin aineksin tai esittelemällä kertoja. Kaikkein nuorimpien eli ensimmäisen luokan oppilaiden Unelmakirjoitelmissa ei ole lainkaan suoraa eikä epäsuoraa esitystä ja muutkin moniäänisyyden ilmaisumuodot ovat lähes olemattomissa. Varsinainen referointi, puhujan tai kertojan puheen ilmaisukeinojen hallintaan otto on monen tutkimuksen mukaan kesken vielä 12-vuotiailla ja jopa 15-vuotiailla eli ikäryhmissä, joihin tässä luvussa keskitytään. Kehitys ei myöskään näytä täysin lineaariselta eli ikäkaudesta toiseen tasaisesti kehittyvältä (ks. esimerkiksi Nordqvist 2001; Köder 2016).

Moniäänisyyden ilmenemismuodoista referointi on ehkä tutuin ja tutkituin. Referoinnin oppiminen kestää kauan muun muassa siitä syystä, että se edellyttää kykyä vaihtaa perspektiiviä puhujasta kertojaan ja kykyä kuvitella, mitä vastaanottaja tietää, eli mielen teorian kehittyneisyyttä (Astington & Baird 2005; Doherty 2009). Vaikka oppiminen on hidasta, referoinnin esittämistavat ovat kuitenkin melko helposti mallinnettavissa toisin kuin vaikka ironian ja intertekstuaalisuuden. Referointi ilmenee kielellisesti muun muassa spatiaalisen, referentiaalisen ja temporaalisen deiksiksen muutoksina ja kehyslauserakenteina. Nuorimpien unelmakerronnassa erityisesti tempuksen esittäminen huojuu ja

kertomus noudattaa tarinan aikaa useimmiten kronologisesti ja ilman tiivistyksiä (ks. Vesala 2012).

Puhutussa ja kirjoitetussa kielessä moniäänisyyden ilmaisu eroaa huomattavasti sekä ilmaisumuodoiltaan että funktioiltaan – vaikka yhtymäkohtiakin on – mutta oppimisen ja käytön lähtökohtana on puhutun kielen suora esitys (Nordqvist 2001). Puhutussa kielessä moniäänisyyden luonnollinen esittämistapa on multimodaalinen: esityksessä voidaan käyttää hyväksi paitsi kieltä myös kasvojen ilmeitä, eleitä, äänensävyä ja muita prosodian keinoja sekä toiminnan esittämistä (*re-enactment*) (ks. Strömqvist, Nordqvist & Wengelin 2004; Good 2015; Hodge & Gormier 2019). On selvä, että puhujan äänen näin kokonaisvaltainen esittäminen antaa paljon mahdollisuuksia myös sisällön ja esittäjän arvottamiseen. Kertoja ei yleensä toista puhutun eikä kirjoitetun kielen suorassa esityksessä koko sanallista esitystä sellaisenaan vaan valitsee kontekstin edellyttämän osan esityksestä. Aiemmin katsottiin, että suora esitys olisi sanatarkkaa – operoitiin niin sanotulla sananmukaisuuskriteerillä – ja että epäsuora esitys olisi enemmän tiivistelmä esityksen tietosisällöstä. Todellisuudessa ero on paljon suhteellisempi.

Suoran ja epäsuoran esityksen sisällöllistä eroa on filosofiassa kuvattu termeillä *de dicto* ja *de re*; suunnilleen samaa ilmaistaan termeillä *de-piktiivinen* eli esittävä ja *deskriptiivinen* eli kuvaileva (Nordqvist 2001). Puhutussa kielessä tämä eronteko vastaa suoraa ja epäsuoraa esitystä, mutta kirjoitetussa kielessä suoraa esitystä voidaan tiivistää puhuttua enemmän ja painottaa kertojan haluamalla tavalla. Tiivistys on välttämätöntäkin, koska muisti ja ajallinen etäisyys asettavat rajansa. Kirjallisuudessa tehdään ero myös raportoivan ja projektoivan esityksen välille (Nordqvist 2001). Ensimmäisessä on kyse jonkin todellisuudessa sanotun referoinnista (*speech reporting*) ja jälkimmäisessä fiktiivisestä aineksesta (*speech projecting*). – Projektio-termi on käytössä myös hieman laajemmassa mielessä (ks. esimerkiksi Shore 2005).

Tämän tutkimuksen aineistossa eli Unelmakirjoitelmissa on tässä mielessä aina kyse fiktiivisen maailman projektioista, koska ei voi tietää, referoivatko kirjoittajat jotain oikeasti tapahtunutta tai vain kuviteltua. Tehtävänannossa kirjoittajia ohjeistettiin kirjoittamaan unelmapäivästä, joka on tapahtunut, jonka toivoisi tapahtuvan tai joka voisi tapahtua, jos

mikä tahansa olisi mahdollista (Pajunen 2012). Nuorimpien unelmat on kirjoitettu näiden vaihtoehtojen mukaan, ja kaksi viimeistä vaihtoehtoa tuottavat aina fiktiivisen kontekstin. Vanhemmilla kirjoittajilla täysin fiktiiviset kirjoitelmat ovat vähissä, mutta heillä kerrontaa täydentää arviointi ja kuvailu. Myös tekstityyppi- ja lajivaihtelut hankaloittavat suoran esityksen todenperäisyyden arviointia, jos se muuten olisikin mahdollista. Ongelma koskee erityisesti luovaa kirjoittamista. Raportoivan ja projektoivan kerronnan ero on sen sijaan helpompi – ehkä helpoin – tehdä kontekstissa, jossa kielentäminen on perustunut kuva-aineistoon eli tilanteessa, jossa arviointiin on ulkoinen kriteeri.

Puhuttu kieli rakentuu paitsi peräkkäisesti myös samanaikaisesti, kirjoitettu ainoastaan peräkkäisesti. Kirjoitetussa kielessä toisen puhe pitää esittää yksin kielen keinoin, mikä antaa vain vähän välineitä ilmaista puhujan ja puheen ei-sanallisia ominaisuuksia. Tyypilliset vaihtoehdot ovat illokutiivisia sävyjä ilmaisevat kehys- tai johtolauseen puheaktiverbit (*sanoa* > *vaikertaa*, *valittaa*, *uhkailla*; *kilahtaa*), tavan adverbis tai arvottavat adjektiivit (*sanoa rumalla äänellä*, *huutaa rumasti*), imitatiivit (*olin että auts*) sekä välimerkit (!!?) ja erilaiset tekstin muotoon ja aseteluun liittyvät typografiset keinot. Nuorimmilla peruskoululaisilla imitatiivit ja arvottavat välimerkit ovat Unelma-aineistossa jo monipuolisesti käytössä, mutta esimerkiksi puheaktiverbien, tavan adverbien ja adjektiivien kirjo kasvaa aikuisikäen asti muun leksikaalisen diversiteetin kasvun myötä (ks. Pajunen 2012). Verbeillä voi hyvin ilmaista illokutiivista sävyä – esimerkiksi *väittää*, *kysyä*, *todeta* ja *valittaa* – tai henkilön puhumistapaa ja olemusta tarkemmin kuin yleisverbillä *sanoa*; alakoululaisten kirjoitelmissa *sanoa*-verbi on silti vielä lähes yksinomainen.

Referointi liitetään tyypillisesti kertovaan tekstiin, jossa sen keskeisenä funktiona on kuljettaa juonta kronologisesti eteen päin (Vincent & Perrin 1999, 292; Nordqvist 2001). Vincent ja Perrin määrittivät laajan ranskankielisen haastatteluaineiston avulla kertovan tekstin referoidut lauseet Labovin (1972; 1982) jaotteluun tukeutumalla niihin, joiden funktio oli narratiivinen, ja niihin, joilla oli muita funktioita. Nämä ei-narratiiviset referoivat lauseet he edelleen jaottelivat evaluoiviin (*appreciative: judgment or opinion*), tuen (*support*) tai auktoriteetin ilmaisiin. Tuloksena oli, että yli puolet suorasta esityksestä omaa aineistos-

sa narratiivisen funktion mutta epäsuorasta vain 34 prosenttia. Epäsuorista esityksistä suurin osa on auktoriteetin ilmaisuja, tietoa siitä, kehen tekstin argumentaatioissa vedotaan.

Referoidun puheen multifunktionaalisuudesta puhuu myös Berman (1996): funktio voi olla paitsi puhujan ilmaiseminen, myös evaluointi, depiktiivinen tai deskriptiivinen esittäminen, esityksen elävöittäminen, etäisyyden luominen tai juonen edistäminen. Referoidun puheen funktiot erilaistuvat myös tekstityypin ja -lajin mukaan; erittelevässä tekstissä referoitu puhe on epäsuoraa esitystä ja ei-narratiivista funktioltaan (Köder 2016). Funktio vaikuttaa muun muassa siihen, miten referoinnin sisältö pitää kielentää ja mitä edellytyksiä vastaanottajalla on sisältöjä ymmärtää. Nykynäkemyksen mukaan niin sanottu pragmaattisen tason ilmiökielentäminen liittyykin vahvasti kielentämistilanteen fyysiseen ympäristöön, erilaisiin sosiokulttuurisiin tekijöihin sekä ikäkaudelle ominaisiin kielellisiin resursseihin. Lisäksi taustalla vaikuttavat iän lisäksi kielentäjien yksilöpsykologiset ominaisuudet (Bohnemeyer ja muut 2014; Mar 2004). Jos narratiivin ja erittelevän tekstin referoitujen lauseiden funktiot ovat erilaisia, vastaavasti niiden ymmärtäminen on erilaista, koska edelliseen liittyy referoitujen tapahtumien osallistujia, joiden intentiot, tavoitteet ja tunteet pitää ja voi tavoittaa, mutta jälkimmäisessä ei ole näin selviä tekijöitä, joihin tulkintaa voi perustaa (ks. Mar 2004). Lisäksi on hallittava modaliteettierot eli osattava referoida sekä puhutun että kirjoitetun kielen konventioiden mukaan.

Ymmärrys moniäänisyyden ja etenkin referoinnin moninaisuudesta on kasvanut 2000-luvulla tehdyn tutkimuksen myötä paljon, mutta samalla myös jo tehty tutkimus on käynyt vaikeammiksi arvioitavaksi, koska on ymmärretty saatujen tulosten vaihtelevan metodien, koehenkilöiden iän ja tekstityyppien mukaan. Unelmakirjoitelmissa tavoitemuoto on kertova ja kirjoitelmat lähestyvät tätä muotoa vähittäin. Ala- ja yläkouluun päättövaiheessa kirjoitelmat ovat jo pääosin kertovia. Kirjoitelman teema eli unelmien päivä kuvataan usein hyvin toiminnallisesti tai vanhimmilla arvottavasti; erityisesti arvottava käsittelytapa näyttää vierovan suoraa esitystä. Aikuiskirjoitelmissa myös omaelämäkerrallisuus kasvaa ja lisää toisaalta yksiäänistä kerrontaa, toisaalta kokijanäkökulmaa ja vapaata epäsuoraa esitystä tai eläytymisesitystä.

On hyvin tunnettua, että suoran ja epäsuoran esittämisen kielellisiä eroja ei voi esittää jyrkällä kahtiajaolla piirre kerrallaan määritellen vaan ennemmin jatkumoluonteisena (Maier 2009; Evans 2012; suomesta mm. Ikola 1961, Hakulinen, Vilkuna, Korhonen, Koivisto, Heinonen & Alho 2004; Niinimäki 2015). Erityisesti kehityksellisessä kontekstissa eronteko ei voi olla kovin jyrkkä. Typologisessa kontekstissa ajatus on, että kielten vertailu referoinnin ilmaisussa on valtavan merkinnän kirjjon vuoksi lähes mahdotonta, vaikka voidaan esittää joitakin kanonisen suoran ja epäsuoran esityksen ominaisuuksia. Suoralla esityksellä on kielissä Evansin (2012) mukaan selvimmät määriteltävät ominaisuudet ja ilmeisesti vain suora esitys esiintyy kaikissa kielissä – epäsuorassa esityksessä on sen sijaan vaihtelua sen mukaan, miten moni tekijä on muun muassa deiktisissä muutoksissa mukana. Vaihtelu koskee erityisesti tempusta ja toisaalta logoforisia johtolauseen puhujaan viittavia referaatin pronomineja. Suomessakin tulee referoinnin yhteydessä esiin *hän*- ja *se*-pronominien logoforinen vaihtelu (ks. esimerkiksi Kallio 1978). Kielikohtainen vaihtelu voi myös näkyä sääntöjen tiukkuuden eroina. Esimerkiksi saksa ja hollanti eroavat Köderin (2016) mukaan siinä, miten tiukasti alisteisen lauseen käänteistä sanajärjestyksestä on noudatettava – ja vastaavasti siinä, miten selvästi suora ja epäsuora esitys eroavat kielellisesti.

Kehityksellisessä kontekstissa kirjoitetun kielen suorassa ja epäsuorassa esityksessä näkyy samantapainen monensuuntainen huojunta, ja ongelmat ovat selvimmät juuri luovassa kirjoittamisessa, josta kuva-elisitoinnin tapainen kielentämis- ja vertailumalli puuttuu. Keskityn tässä yhteydessä tarkastelemaan, miten kuudes- ja yhdeksäsluokkalaiset ilmaisevat kirjoitelmissaan moniäänisyyteen liittyvää intertekstuaalisuutta, metatekstiä, miten he esimerkiksi puhuttelevat lukijaa eli ovat tietoisia toisesta näkökulmasta ja miten he ilmaisevat suoraa ja epäsuoraa esitystä ja niiden välimuotoja. Joitain havaintoja esitän koko Unelma-aineistosta. Tarkastelen vielä ennen varsinaista analyysiä, miten referointitaitojen kehitystä on aiemmin tutkittu ja mitä on saatu selville.

## Kehityksellinen näkökulma

Pienet lapset oppivat ilmaisemaan puhutussa kielessä moniäänisyyttä – etenkin suoraa esitystä – hyvin nuorina. Jo noin kaksivuotiaat matkivat toista ääntä puheessaan ja noin kolmivuotiaina he käyttävät myös epäsuoraa esitystä (esimerkiksi Nordqvist 2001). Kansainvälistä kehityksellistä tutkimusta laajemmin määriteltävän moniäänisyyden kielentämisestä ei juuri näytä olevan ja referoinnistakin sitä on tehty suhteellisen vähän. Tutkimusta on myös paljon enemmän puhutusta kuin kirjoitetusta kielestä. Uusin tutkimus näyttää kohdistuvan puhutun kielen multimodaaliseen suoraan esitykseen (ks. esimerkiksi Hodge & Gormier 2019) tai erityisryhmien tutkimukseen (ks. esimerkiksi Engberg-Pedersen & Vang Christensen 2017). Puhutun ja kirjoitetun kielen kontekstissa laajoja kehityksellisiä tutkimuksia on tehty ainakin ruotsin (Nordqvist 2001) sekä saksan ja hollannin (Köder 2016) kielistä. Kummassakin on mukana useita ikäryhmiä 2–5-vuotiaista lapsista koululaisryhmiin (9-, 12- ja 15-vuotiaat) ja aina nuoriin aikuisiin. Esimerkiksi englantia on tutkittu pienimuotoisemmin (Goodell & Sacks 1992; Hickmann 1993). Suomessa peruskouluikäisten referointitaidoista ei ole paljoa tutkimusta ikäryhmätasollakaan, mutta Kalliokoski (2005b) on tutkinut suomea toisena kielenä kirjoittavien abiturienttien ja yhdeksäsluokkalaisten tekstien moniäänisyyttä ja koherenssia sekä Juvonen (2014) epäsuoraa esitystä ylioppilasaineaineistolla. Ylioppilasaineiden referointikeinojen toteutumista ja merkitystä tekstissä on kartoittanut myös Berg (2011). Metakielellistä ainesta ilmaisevia aloituksia ja lopetuksia on myös tutkittu jonkin verran (ks. esimerkiksi Komppa 2012). Tämä tutkimus pyrkii vastaamaan erityisesti kouluikäisten referointitaitoja kartoittavan tutkimuksen tarpeeseen.

Moniäänisyyden ilmaisun kehittymistä on siis tutkittu eniten puhutussa kielessä ja metodina ovat useimmiten olleet roolileikkien leikittäminen tai kuvakirjajelisitaatio. Etenkin *Frog where are you* -tarinan (Mayer 1969; Berman & Slobin 1994) avulla on tuotettu suullista kerrontaa ja uudelleenkerrontaa, lisäksi uudemmissa tutkimuksissa on mukana multimodaalinen kerronta. Myös kirjoitustutkimus perustuu ennen kaikkea kuvajelisitaatioon, mutta luovan kirjoittamisen yhteydessä moni-

äänisyyden ja referoinnin ilmaisun kehittymistä ei ole tietääkseni laajemmin tutkittu. Kuvaelisitaatio toisaalta tukee tutkimusasetelman luomista, toisaalta nostaa vaatimustasoa, koska se edellyttää visuaalista lukutaitoa (ks. Harris Wright, Capilouto, Srinivasan & Fergadiotis 2011). Tutkimusten vertailtavuutta heikentää se, että varhempi tutkimus on usein perustunut melko pieniin koehenkilöryhmiin. Toisinaan tuloksia on myös vaikea tulkita, koska aiemmin ei niinkään kiinnitetty huomiota siihen, että tulokset voivat riippua käytetyistä metodeista ja tekstityypeistä.

Moniäänisyys näyttääkin olevan ilmiönä suhteellisen altis tutkimusmetodille. Se millä menetelmällä aineisto tuotetaan, vaikuttaa muun muassa siihen, missä funktiossa kielellinen aines esiintyy ja miten yleistä se on. Narratiivien juonta edistävä referointi on koetilanteissa helpompi tuottaa kuin tietoa tai kokemusta välittävä. Psykologin tutkimus osoittaa, että moniäänisyyden ilmaisumuoto vaikuttaa myös ymmärtämiseen ja taustalla oleviin muistiedustumiin (ks. Eerland & Zwaan 2018). Suora esitys koetaan eloisammaksi ja havaittavammaksi, mutta epäsuoran esityksen kautta kertomuksen päähenkilö muistetaan paremmin. Eerlandin ja Zwaanin (2018) koesarjassa esityksen muodon vaikutusta sisällön muistamiseen ei löytynyt, vaikka koehenkilöillä oli ongelmia viittaussuhteiden ymmärtämisessä. Köderin ja Maierin (2016) mukaan lapsilla on ongelmia viittaussuhteiden tulkinnassa suorassa mutta ei epäsuorassa esityksessä. Tilanne kuitenkin vaihtelee asetelman mukaan: suora esitys on lapsille vaikeampi tulkita silloin, kun asetelma on rakennettu tiedonvälittämiseen, mutta ei silloin, kun se on narratiiville tyypilliseen tapaan rakennettu juonen eteen päin kuljettamiseen (Köder & Maier 2018). Referoinnin ja deiktisten muutosten yhteydessä kompleksistuvien kerroksellisten viittaussuhteiden ymmärtämistä onkin tutkittu eniten (ks. esimerkiksi Köder 2016) ja tällainen asetelma on helpoin rakentaa elisitaation pohjalle. Yleensä ajatellaan, että ymmärtäminen edeltää tuottamista – eli kun ymmärretään esimerkiksi kompleksisia viittaussuhteita, niitä pystytään tuottamaan – mutta toisinaan suhde voi olla myös asymmetrinen (Hendriks 2014). Kehityksellinen konteksti vaikeuttaa arviointia entisestään.

Yleisiä aiheita referointitutkimuksessa ovat olleet myös käytettyjen muotojen ja funktioiden kehittyminen, erilaisten sekaantumamuotojen

esiintyminen sekä erilaiset kehystykset, kuten johtolauseen esiintyminen referoinnissa ja johtolauseessa käytetyt verbit sekä moniäänisyyden ilmaisun rajojen merkintä välimerkeillä (ks. esimerkiksi Goodell & Sacks 1992). Nordqvistin (2001) mukaan vielä kaksitoistavuotiaista noin 20 prosenttia ei anna lukijalle vihjettä siitä, kenen puhetta referoidaan tai millä ilmaisun tasolla liikutaan. Unelma-aineistossa peruskoululaiset eivät käytä suoran esityksen konventionaalisia välimerkkejä kovinkaan ahkerasti. Muutama viidesluokkalainen käyttää lainausmerkkejä ja hie-man useampi kuudesluokkalainen myös kaksoispistettä. Deiktisistä muutoksista parhaiten huomioidaan henkilöviittaukset.

Kansainvälisen tutkimuksen mukaan moniäänisyyden ja etenkin referoinnin kehitys alkaa aina puhutusta kielestä ja suorasta lainauksesta, joka on aluksi tyypillisesti sanatasoista. Kyse on oikeastaan sanatason intertekstuaalisuudesta. Monista lähinnä kuvaelisaatioon perustuvista tutkimuksista käy ilmi, että kehitys on hyvin hidasta. Eri ikäryhmillä voi myös olla erilaisia preferenssejä moniäänisyyden ilmaisun keinoja valitessaan. Goodell ja Sacks (1992, 414–416) osoittivat, että yhdeksänvuotiailla ja aikuisilla oli enemmän epäsuoraa (60 %) kuin suoraa esitystä, mutta kuusivuotiaat suosivat samassa elisitointitehtävässä suoraa esitystä. Nuorimmilla on vaikeuksia ottaa kuulija huomioon. Suorassa esityksessä nuorimmat koehenkilöt antavat yleensä äänen tarinan (kuva-kirjan tai vastaavan) päähenkilölle mutta vanhimmat kaikille henkilöille. Ero näkyy hyvin myös sammakkotarinan kirjallisesta kielentämistehtävästä. Nordqvistin (2001) aineistossa kouluikäiset koehenkilöt referoivat pääosin päähenkilön eli pienen pojan tekemisiä ja jättivät tarinan muut henkilöt eli sammakon, koiran ja etsintäreitillä kohdatut muut eläimet huomiotta tai ilmaisivat näitä esimerkiksi imitatiivein. Teksteissä oli paljon ambivalentteja viittaussuhteita ja informaatiota oli selvästi vähemmän kuin vanhemmilla kirjoittajilla. Viisitoistavuotiailla suora esitys väheni ja kaikista ikäryhmistä juuri viisitoistavuotiaat käyttivät epäsuoraa esitystä eniten. Heidän tuotoksissaan ei-verbaliiset ja paralingvistiset piirteet olivat näkymättömissä ja marginalisoituneet. Referoinnin sisäl-lön illokutiivisten sävyjen ilmaisu puheaktiverbien variaationa oli edel-leen vähäistä. Aikuiset tuottivat sammakkotarinoita monessa mielessä yhdeksänvuotiaiden tavoin, mutta erona yhdeksänvuotiaisiin aikuiset



projisoivat puhetta laajempaan henkilöjoukkoon ja käyttivät samalla useammanlaisia syntaktisia rakenteita helpottaakseen puhujan identifiointia. Heidän leksikkonsa oli muita rikkaampi ja he yhdistivät puheaktiverbejä ja tavan adverbijä ilmaistakseen epäsuorassa esityksessä esimerkiksi puhumistavan ilmaisuja (*frågade B. förvirrad* 'kysyi B. hämentyneenä'). Monista tutkimuksista käy ilmi paitsi moniäänisyyden ilmaisukeinojen hidas omaksuminen myös kehityksen ei-lineaarisuus ja eri-ikäisten erilaiset preferenssit. Semanttis-pragmaattisella tasolla kehitys on ylipäättään hidasta (Berman 2004; 2007).

Tarkastelen tässä yhteydessä peruskoululaisten Unelmakirjoitelmien avulla, miten moniäänisyyden ilmaiseminen kehittyy alakoulusta yläkouluun. Analysoin systemaattisesti toistasataa Myöhemmän kielenkehityksen tutkimushankkeen kuudes- ja yhdeksäsluokkalaisten Unelmakirjoitelmia, jotka on kerätty vuonna 2014, mutta perustan analyysini moniäänisyyden ilmaisukeinojen kartoitukseen laajemmasta aineistosta. Kartoitus on tarpeen, koska luovan kirjoittamisen osalta kehityksellinen tutkimus on niin vähäistä. Aloitan tarkasteluni kokoamalla aikaisemman tutkimuksen havainnointia siitä, miksi moniäänisyyttä voidaan pitää hyvän kertomuksen ominaisuutena ja yhtenä rakennetekijänä. Intertekstuaalisuus on koululaisten kirjoitelmissa melko vähäistä – usein sanatasaista – ja tulkinnallista, se on mielenkiintoista muun muassa, koska se kaiuttaa nuorten kulttuuritaustaa. Korpus itsessään tarjoaa runsaasti materiaalia sen havainnoimiseen, kuinka hyvän kertomuksen kirjoittaja ottaa lukijansa huomioon ja kuinka hän nousee oman tekstinsä yläpuolelle metatekstin kautta. Huomattava osa aineistoni tarkastelusta liittyy kuitenkin referointiin, jonka avulla kirjoittajat tuovat tekstiinsä oman tai jonkun muun aiemman äänen tai oman ajatuksensa. Erityisesti tarkastelen, kuinka kokemuksellisuuden ilmaisu kehittyy ja kuinka tekstiä etäännytetään kertojasta: millä tavoin tuodaan kokeva minä kertovan minän rinnalle? Elisitoidulla aineistolla ei ole suoraan ollut mahdollista tutkia kokemuksellisuuden ilmaisua, joten kirjoitetussa kehityksellisessä kontekstissa näkökulma on uusi.

## Kertovan tekstin moniäänisyys

Labovin (1972, 361–262; 1982) mukaan kertomuksen runko koostuu narratiivisista, temporaalisesti orientoituneista lauseista ja niin sanotuisista vapaista lauseista, joista puuttuu temporaalisesti orientoiva aines. Vapaat lauseet eivät liity ajallisesti kerronnan tapahtumiin eivätkä vie sitä eteenpäin; niiden tehtävä on ennen kaikkea evaluoiva, kuten jo mainittiin. Fludernik (1966, 43–46) puhuu kertomuksen kokemuksellisuudesta (*experientiality*), sen kyvystä heijastaa inhimillistä kokemusta. Kertomus merkitsee kertojalle jotain tunteen tasolla ja on sen vuoksi kertomisen arvoinen. (Vrt. Hyvärinen 2012, 403.) Unelmakirjoitelmien tehtävänanto ja otsikko suorastaan johdattelee kokemuksellisuuden ilmi tuomiseen. Melkein kaikissa aineistoni teksteissä on tarinaan osallistuva minäkertoja – tai ainakin sen käyttöön siirrytään aloituksen jälkeen. Koko alakoulun aineistossa hallitsee minämuotoinen kertojanääni (Vesala 2012, 77). Minäkerronnan rakenteista voidaan erottaa kertova minä ja kokeva minä (Stanzel 1985 [1979], 135). Rakenteet eivät kuitenkaan ole täysin vastakohtaisia vaan kohtaavat tapauksissa, joissa kertojan ääni yhtyy henkilön ääneen (vapaa epäsuora esitys). Kertova minä kuljettaa kerrontaa eteenpäin, kääntyy lukijan puoleen ja kommentoi. Reflektioiva, kokeva minä taas tuo esiin kertojan tunteita, ajatuksia ja havaintoja. (Stanzel 1985 [1979], 222; Cohn 2006 [1999], 120–121.) Voi jopa kysyä, onko kertomus tekstilajinsa mukainen, ellei se ole moniääninen. Hyvä kertomus ei etene tapahtumasta toiseen, vaan kertoja voi keskeyttää sen eri tavoin, muun muassa puheen tai ajatusten esittämisellä, referoinnilla. Kerronta keskeytyy silloinkin, kun kertoja kääntyy kuulijan tai lukijan puoleen ja kertoo tälle, mistä on kysymys, tai esittää muita kommenttejaan. Keskeytys on merkki siitä, että tekstissä on mukana evaluoiva näkökulma. (Vrt. Labov 1972, 370–373.)

Esimerkissä 1 kuudesluokkalaisten kertomus etenee narratiivisin lausein, joiden predikaatit ovat toimintaa ilmaisevia (*heräsin, kävelin, lähdin, menttiin, käytiin* ja niin edelleen). Havaintoa, ajattelua ja asenteita ilmaisevat verbit puuttuvat. Lukijalle ei välitetä kokemusta siitä, millaista on elää kuvattujen tapahtumien läpi. Kirjoitus on yksiääninen lukuun

ottamatta ensimmäisen kappaleen viimeistä virkettä, jossa referaatissa tuodaan esiin äidin ääni:

1) Eräänä aamuna, kun **heräsin kävelin** alakertaan syömään aamupalaa, kun **olin syönyt lähdin** kavereitten kanssa ulos. **me mentiin uimaan**. Uimisen jälkeen **me lähdettiin** kaupungille. Kavereitten kanssa **me käytiin** koskikeskuksessa jäätelöllä. Minä **lähdin** kotiin, mutta kun olin kotona äiti sanoi, että nyt **lähdetään** Eedenii (kylpylään).

Kello oli 6 päivällä kun **tultiin** kylpylästä minä **lähdin** taas kavereitten kanssa ulos. Illalla, kun **tulin** kotiin **menin nukkumaan**. (6F9)<sup>51</sup>

Kalliokoski (2005b, 234) toteaa suomea toisena kielenä opiskelevien yhdeksäsluokkalaisten kirjoittajien tekstien kertomuksellisuuden olevan paradoksaalisesti kyseenalainen tapauksissa, joissa kertomus koostuu pelkistä kertovista lauseista. Asiantilojen muodostama kokonaisuus jää saavuttamatta. Referoinnin puuttuminen korostaa ”epäkertomuksen” vaikutelmaa: *Olipa kerran auto varas. Se varasti autoja joka päivä. Se asui Itäkeskuksessa kun saapui yö hän lähti keikkaa autoja se lähti Itäkeskuksen ostarille siellä oli sellainen auto kuin MAZDA 626, jonka rekisteri oli LHG-223 se meni – –*. Seuraavassa esimerkissä 2 on oman aineistoni yhdeksäsluokkalaisten moniääninen teksti:

2) Tulen kotiin noin yhdeltä – **kuusi tuntia peliin** – kun puhelin soi. **Vastaan, ja siellä on Rasse, joka pyytää tulemaan** noin 45 min aikaisemmin hallille hiomaan vielä ylivoimaa. **Vastaan tulevani** ja suljen puhelimen. **Pohdin, mitä enää on tehtävissä**. menen yläkertaan ottamaan päiväunia. Katson jälleen kelloa ”**miksi olen näin hermostunut**”. – – **En ole enää varma**, mitä teen väliajalla, kun ovikello soi odottamatta. Kävelen rappuset alas ja avaan oven – **se on Joonas, paras kaverini** jo valmiina kooveen tuulipuvussa ”**Huh, ulkona on kylmä!**” **hän heittää heti kävellessä ja yhdyn** – **ulkona todella on kylmä**. Alan pakkaamaan ja **kysyn, soittiko Rasse Joonaksellekin, joka vastaa kyllä**. (9M1)

51 Aineistoviitteestä käy ilmi luokka-aste, kirjoittajan sukupuoli (F, M) ja kirjoitelman numero.

Esimerkissä 2 moniäänisyys luodaan eri keinoja käyttämällä. Kertoja etäännyttää kerronnan kerrontahetkestä ja ottaa erilaiset äänet mukaan. Katkelmissa kuljetetaan toimintaa eteenpäin: tullaan kotiin, vastataan puhelimeen, valmistaudutaan illan peliin ja tavataan kaveri, mutta kertoja keskeyttää tapahtumien kuvailun toistuvasti. Kertojan puhetta referoidaan puheaktina (*vastaan*) ja epäsuorana esityksenä (*vastaan tulevani; kysyn, soittiko Rasse Joonaksellekin*). Kertojan ajatuksia referoidaan vapaana epäsuorana esityksenä (*kuusi tuntia peliin; ”miksi olen näin hermostunut”; ulkona todella on kylmä*) ja epäsuorana esityksenä (*Pohdin, mitä enää on tehtävissä*). Kahden muun henkilön puhetta esitetään puheaktina (*pyytää tulemaan*) ja suorana esityksenä (*”Huh, ulkona on kylmä!”; kyllä*). Kirjoittaja siis muuttaa esitystavasta toiseen tiheästi, ja kertova minä sekä kokeva minä vaihtuvat. Rakenteena on johtoilmaus sekä siihen liitetty referaatti, mutta kirjoittaja luottaa paikoin lukijan kykyyn erottaa kokeva minä myös ilman johtoilmausta. Esimerkit 1 ja 2 heijastavat arvioinnin kehityksellistä lisääntymistä aineistossa.

Moniäänisyyden tuottamisesta kehityksellisessä kontekstissa on suomen osalta vähän tietoa, ja etenkin luovassa kirjoittamisessa kansainvälinenkin tutkimus on olematonta. Perustan siksi tämän tutkimukseni varsinaisen kehityksellisen analyysin alakoulun kirjoitelmien avulla tehtyyn kartoitukseen,<sup>52</sup> josta selviää, miten moniäänisyys syntyy ja miten se lisääntyy alakouluikäisten kirjoitelmissa. Jokainen kirjoitelma on arvioitua sen mukaan, esiintyykö jokin moniäänisyyden ilmaisukeino vähintään kerran vai ei lainkaan. Moniäänisyyden tyypeiksi on katsottu intertekstuaalisuus, metakieliset tapahtumakontekstin ulkopuolelle viittaukset, suora ja epäsuora esitys sekä vapaa epäsuora esitys muunnelmiseen. Taulukkoon 31 on yksinkertaisuuden vuoksi merkitty vain, monessako kirjoitelmassa esiintyy suoria viittauksia lukijaan (mutta ei muuta meta tekstiä) sekä suoraa tai epäsuoraa esitystä, koska nämä oli laskennallisesti helpoin tavoittaa. Ne ovat myös kouluikäisillä moniäänisyyden kehityksellisesti tärkeimmät ilmenemismuodot. Suora ja epäsuora esitys erottuvat kehittyvässä kirjoituksessa toisistaan selvimmin referentiaali-

52 Jaottelu on tehty vuoden 2006 aineiston avulla; varsinainen analyysini perustuu vuoden 2014 aineistoon, koska vuoden 2006 aineistossa ei ole yläkoululaisten kirjoitelmia.

sen deiksiksen perusteella (*X kysyi menettekö te vs. menemmekö me*), joten esitykset on laskelmissa eroteltu juuri persoonadeiksiksen avulla. Epäsuora esitys rajautuu lähinnä johtolauseen avulla erottuviin, joten periaatteessa taulukkoon ei ole merkitty vapaata epäsuoraa esitystä tai ajatusten referointia; ne eivät kuitenkaan ole alaluokille ominaisia moniäänisyyden muotoja.

Ensimmäisen luokan aineistossa ei ole varsinaisesti moniäänisiä tekstejä, mutta intertekstuaalisuutta on jonkin verran: kirjoittajat viittaavat esimerkiksi peleihin ja kirjoihin. Jo toisluokkalaisten kirjoitelmissa moniäänisyyttä ilmaistaan useammalla keinolla mutta yhä vähän: viidesosassa kirjoitelmista käytetään suoraa esitystä vähintään kerran, epäsuoraa esitystä on yhteensä vain neljässä kirjoitelmassa ja suora viittaus lukijaan yhdessä. Koko alakoulun aineistossa suoran esityksen määrä vaihtelee hieman työillä mutta kasvaa pojilla iän mukaan. Epäsuoran esityksen määrä kasvaa alakoulussa lähes lineaarisesti mutta sukupuolittuneesti omaa tahtiaan. Alakoulun aineistossa moniäänisyyttä ilmaistaan monipuolisimmin viidennellä luokalla, joten kasvu on selvän kehityksellistä, kuten oletus oli. Monet kirjoittajat, jotka moniäänisyyttä ylipäättään ilmaisevat, ilmaisevat sitä useammin kuin kerran. Lukijan suora puhuttelu lisääntyy muun moniäänisyyden lisääntyessä ja se tulee tyypillisesti esiin samoissa kirjoitelmissa, joissa on myös referointia. Se selvästi edellyttää kirjoittajalta kykyä ottaa lukija huomioon, ja kirjoitelmat, joissa lukijaa puhutellaan, ovat myös kertomusskeeman mukaisia (Labov & Waletzky 1967). Lukijaa huomioidaan myös epäsuoremmin erilaisin metakielisin keinoin. Esimerkiksi Vesalan (2012) mukaan aloituksissa ja lopetuksissa on iän mukaan enenevästi metatekstiiä. Metakielisiä aloituksia ja lopetuksia ei ole merkitty taulukkoon 31.

Käsittelen jatkossa moniäänisyyden ilmaisua kuudennen ja yhdeksännen luokan oppilailla Unelmakorpuksessa eli vuoden 2014 aineistossa sen kehityksellisessä ilmenemisjärjestyksessä eli etenen intertekstuaalisuudesta metakieleen ja lukijan puhutteluun sekä lopulta varsinaiseen referointiin.

Taulukko 31. Moniäänisyyden ilmaisukeinojen osuus alakoululaisten kirjoitelmista.

Ikäluokka / Moniäänisyys	Kirjoitelmia	Lukijan puhuttelu %	Suora esitys %	Epäsuora esitys %
1. F	57	0	0	0
1. M	55	0	0	0
2. F	80	1,3	22,5	3,7
2. M	83	0	7,3	1,2
3. F	53	3,7	13,2	7,5
3. M	72	2,7	12,5	2,8
4. F	93	5,4	30,1	12,9
4. M	93	2,2	12,9	11,8
5. F	75	13,3	36,0	33,3
5. M	79	12,6	22,8	25,3

## Intertekstuaalisuus ja metateksti

Intertekstuaalisuus ja metateksti edellyttävät lapselta pragmaattisen kompetenssin kehittyneisyyttä, jotta hän pystyy viittamaan tekstinsä ulkopuolelle tai kommentoimaan omaa tekstiään. Moniäänisyyden ilmaisukeinot voivat vielä olla hakusessa, mutta kirjoitelmista välittyvä pyrkimys päästä vuorovaikutukseen lukijan kanssa. Käsittelen ensin kuudes- ja yhdeksäsluokkalaisten Unelmakirjoitelmien intertekstuaalisuutta ja sitten niissä ilmenevää metakieltä.

### INTERTEKSTUAALISUUTTA ETSIMÄSSÄ

Laajimman määritelmän mukaan (Barthes 1979 [1971], 169) kaikki kirjalliset tekstit ovat intertekstuaalisia, koska kaikkiin teksteihin on kuttunut sitaatteja, viitteitä tai kaikuja toisista teksteistä. Nämä piirteet ovat usein jo tunnistamattomia, niin että niitä on mahdoton jäljittää. Suppeamman määrittelyn mukaan (Fairclough 1992, 84–85, 101–105; Solin 2006, 83) intertekstuaalisuudeksi luetaan lukijan jäljitettävissä oleva suora viittaus, teoksen nimi, sitaatti, lainausmerkein osoitettu toinen teksti mutta myös implisiittisemmät suhteet kuten alluusiot. Tällöin lainattu aines tekee uuden tekstin moniääniseksi ja antaa mahdollisuuden erilaisiin tulkintoihin, dialogiin kirjoittajan ja lukijan välillä.

Lapsi kertoo ja kuvailee teksteissään sitä todellisuutta, jossa hän elää. Jo hyvin nuorena (esikouluikäisenä) lapsi saattaa tuoda tekstiinsä intertekstuaalisia piirteitä ja hänen tekstinsä voi teknisistä puutteista huolimatta olla intertekstuaalista ja moniäänistä (Kauppinen 2012, 592–593; vrt. Nordqvist 2001; Köder 2016). Moniäänisyyttä kannattelevaksi intertekstuaalisuudeksi (vrt. Seppänen 2015, 40) riittänee yleisluontoinenkin viittaus esimerkiksi muualta luettuun tai kuultuun tietoon. Moniäänisyyden määrittely kaipaa kehittyvien kirjoittajien kohdalla kuitenkin rajauksia ja tarkennusta, sillä vielä peruskoulun päättövaiheessakin intertekstuaalisuus on tulkinnallista.

Selkeästi tunnistettavat intertekstuaaliset viittaukset edellyttävät luke-neisuutta tai ylipäättään elämäkokemusta. Siksi ei ole yllättävää, että aineistoni kertomuksissa niitä on vain vähän lukuun ottamatta esimerkiksi teosten ja pelien nimiä. Kuitenkin jo kuudesluokkalainen osaa luoda kerrotun maailman, jossa on viittauksia hänen lukemiensa teosten tai tunteimiensa pelien maailmoihin: *Unelmani on se että pokémonit olisivat totta. Tiedän ettei se ole mahdollista mutta on kuin olisin nähnyt Pidoven.* (6M13). Kirjoittajan intertekstuaaliseen varantoon kuuluvat *Pokémon*-pelisarjan lajit. Viittauksia elokuvaan ja muihin populaarikulttuurin muotoihin on sekä kuudes- että yhdeksäsluokkalaisten kertomuksissa: unelmien päivänä haluttaisiin nähdä muun muassa *Spider-Man 4* ja *X-Men: Days of Future Past*. Kuvaavasti erisnimiä on Unelma-aineiston nuorimmilla kirjoittajilla huomattavasti enemmän kuin vanhemmilla.

Kertovan fiktiivisen tekstin moniäänisyyttä syntyy myös silloin, kun siirrytään tarinamaailmasta toiseen todellisuuteen. Kouluteksteissäkkin näin tapahtuu. Toinen todellisuus ja moniäänisyys luodaan paikan määrittein ja intertekstuaalisiin viittauksiin esimerkiksi suosikkikyhtyeisiin ja suosikkipeleihin. Konditionaalini käytöllä luodaan illuusio kuvitellusta todellisuudesta. Esimerkissä 3 kertoja siirtyy unelmassaan Las Vegasiin ja toteuttaa siellä urheilijan unelmaansa. Esimerkissä 4 kertoja puolestaan johdattaa lukijan vaihtoehtoiseen todellisuuteen suosikkibändiä tapaamaan:

3) *Minun unelmieni päivänä pääsisin Las Vegasiin katsomaan UFC:tä. Haluaisin nähdä Jon Jonesin. – – Katsoisin elokuvia kavereiden kans-*

sa elokuva teatterissa. **Hyppäisin** lasku varjo- ja benji hypyn. **valmistuisin** yliopistosta ja **juhლისin** koko yön. **voittaisin** olympia kultaa. (9M6)  
4) Syömässä ollessamme **saisimme** viestin **One Directionin** managerilta. Hän **kertoisi** että pääsemme tapaamaan heitä! Kun **olisimme** syöneet limusiini **veisi** meidät **toiseen, hienompaan hotelliin, jossa One Direction odottaisi meitä.** (9F26)

Esimerkissä 5 konditionaalin avulla ei luoda toista todellisuutta, mutta kirjoittajalla on alluusioita toisiin kulttuureihin (jalkapuuhan kytkeminen; hiljainen hetki ja kertojalle kumartaminen) sekä yhteiskunnallista tietämystä (kansanedustajien lainsäädäntövalta):

5) Unelmieni päivä **olisi** sellainen, että **ei olisi** koulussa sääntöjä koko päivänä ja että kaupassa **saisi** hakea kaikkea **ilmaseks**. Sitten, Saksa **pyyhittäis** maailmankartalta ja että kaikki **Aasialaiset teloitettaisiin** (ei sillä että mulla ois jotain aasialaisia vastaan mutta se on ärsyttävä maan-osa). Lisäksi kaikki kansanedustajat laitettaisiin jalkapuuhan, kunnes ne lopettaa pakollisten verojen antamisen.

*Sekä jokaisen suomalaisen täytyy pitää päivän aikana hiljainen hetki ja kumartaa mulle.*

*PS. en ole sadisti* (9M3)

Esimerkin 5 teksti on aineistoni ainoa ironisesti kaksiaäninen teksti. Kirjoittaja osoittaa kriittisyyttä tehtävänantoa tai aihetta kohtaan valitsemalla unelmien päivän toiveiksi pahennusta herättäviä asioita ja esittämällä aggressionsa aasialaisten teloittamisesta sekä kansanedustajien jalkapuuhan laittamisesta. Hän kuitenkin selittää puhettaan ja lähestyy lukijaa kahdella huomautuksella: (ei sillä että mulla ois jotain aasialaisia vastaan mutta se on ärsyttävä maan-osa); *PS. en ole sadisti*. Kyseessä on selvästi kaksiaäninen puhe. Kriittistä asennetta tehtävään osoittaa paitsi lähestymistapa myös harkitut puhekielisyudet *ilmaseks, pyyhittäis, mulla ois, mulle*. Kysyä voi, kuinka suuri on ollut kirjoittajan motivaatio.

Koululaisten hyvin lyhyet, muutaman virkkeen kirjoitukset viitannevat nimenomaan motivaation puuttumiseen, kirjoittamishaluttomuuteen, eivät pelkästään kirjoittamisen taitotasoon tai kirjoittajan kielelli-



siin kykyihin. Alakoulun Unelmakirjoitelmia tutkinut Vesala (2012, 54) toteaa, että suuri joukko kuudesluokkalaisia poikia on laatinut erittäin suppean tekstin, ja kysyy myös, onko kyse kirjoitustaidosta vai kielteisestä asennoitumisesta oppiainetta tai tehtävänantoa kohtaan – vai molemmista. Niin ikään yhdeksäsluokkalaisten kohdalla alisuoriutuminen on pojille tyypillisempää kuin tytöille. Kirjoitelman ironinen, halveksuva tai aggressiivinen sävy on merkki oppilaan kielteisestä suhtautumisesta. Sen sijaan viittaukset omaan kirjoittajaidentiteettiin antavat kuvan positiivisesta kirjoittajakuvasta (Seppänen 2015, 48).

Esimerkin 5 kirjoittaja ehkä alisuorittaa, mutta kirjoittamisen perustaidot tuntuvat olevan hallinnassa. Molemmilla päättötason luokkasteilla aineiston konditionaalissa kirjoitetut kertomukset ovat valtaosin lyhyitä ja luettelomaisia sekä yksiaänisiä. Konditionaalissa kirjoitetut tekstit ovat varsinkin yhdeksäsluokkalaisten melkein kaikki poikien kirjoittamia, ja myös kuudennella luokalla valtaosa lyhyistä konditionaaliteksteistä on poikien tekemiä.<sup>53</sup> Esimerkissä 6 moduksen käyttö on lukinnut kirjoittajan kerrontatilanteeseen ja vaikeuttanut tekstin laventamista moniääniseksi. Teksti etenee pelkin narratiivisin lausein (predikaatteina *alkaisi, tulisi, lähtisin, kävisimme, syötäisiin*):

6) *Unelmien päivä alkaisi Rauhallisella herätyksellä. Aamupala tulisi vuoteeseen. Aamupalan jälkeen lähtisin perheen ja kavereiden kanssa huvipuistoon. Kävisimme kaikissa laitteissa. Lounas syötäisiin ravintolassa.* (6M3)

Vähäinenkin intertekstuaalisuus osoittaa, että kirjoittajalla on jo kykyä laventaa näkökulmaansa tietoisesti yli ulkoisen tapahtumasarjan. Ensimmäisen luokan aineisto myös osoitti, että moniäänisyyden luonti kirjoitettuun kieleen alkaa yksinkertaisilla viittauksilla luettuun ja koettuun.

53 Unelmakirjoitelmien sanamäärä kasvaa koko peruskoulun ajan. Keskimäärin poikien sanamäärä on 60 % tyttöjen kirjoitelmien sanamäärästä. Sanamäärä osoittaa ikäluokan keskimääräistä kehitystä etenkin alakouluikäisillä (Pajunen 2012, 13).

## METATEKSTI – LUKIJAN HUOMIOON OTTAMISTA

Myöhemmän kielenkehityksen vaiheessa lapset kehittyvät narratiivisen puheen tuottajista narratiivisen tekstin tuottajiksi ja tekstien suunnitelmallisuus lisääntyy (Tolchinsky 2004, 240). Lapsen pragmaattinen kompetenssi kehittyy (vrt. Blum-Kulka 2004, 192–193, 201; Nippold 2006; Makkonen-Craig 2011), mikä on edellytys varsinaisen moniäänisyyden syntymiselle. Puheen muuttaminen kirjalliseksi kerronnaksi tuottaa useita tekstuaalisia muutoksia ja kehitys näkyy muun muassa siinä, että lapsi oppii ottamaan lukijan huomioon tekstin tuottamisessa eikä kerro asioita pelkästään oma itsensä keskipisteenä.

Aineistoni kuudes- ja yhdeksäsluokkalaisten teksteissä löytyy kertomuksen metatasoon kuuluvaa vuorovaikutukseen kategorisoitavaa moniäänisyyttä, oman tekstin tai tekstin tuottamisen asettamista kommentoinnin kohteeksi ja tietoisien vuorovaikutuksen luomista lukijaan – tai ainakin sellaisen tavoittelua. Pysin seuraavaksi selvittämään, millä kielen keinoilla kuudes- ja yhdeksäsluokkalaisten luovat moniäänisyyttä, kun he kääntyvät lukijan puoleen. Miten nämä evaluoivat kohdat keskeyttävät kerronnan ja mitä funktioita niillä on? Millaisia eroja on kuudesluokkalaisten ja yhdeksäsluokkalaisten keinoissa?

Käsitelen tekstejä fiktiivisinä teksteinä, vaikka niissä on kirjoittajan oman elämän yksityiskohtia mukana. Kirjoituksissa luodaan hypoteettinen maailma, joka tosin rikkoutuu, kun kertoja kääntyy lukijan puoleen ja viittaa kirjoitustilanteeseen. Otan huomioon kirjoittajan ja kertojan eron ja tarkastelen kirjoittajan ja lukijan valintoja moniäänisyyden tulkinnoissa. Koska pohjaan tarkasteluni tekstiin, en näe tarkoituksenmukaiseksi käyttää fiktiivisten tekstien lukemiseen käytettävää sisäislukijan käsitettä (vrt. Virtanen, Rahtu, Shore 2018, 19–20).

Perinteinen koululaisdiskurssi eroaa muista diskursseista.<sup>54</sup> Sillä on oma kommunikaatiotilanteensa, koulukirjoittamiskontekstinsa, ja tekstit kirjoitetaan usein koulussa tietyinä rajattuna aikana yhdelle lukijalle,

54 Opetusmaailmassa kirjoittamisen diskurssit ovat muuttamassa, ja mukaan on tullut genrediskurssin, prosessidiskurssin ja luovuusdiskurssin myötä sosiaalisen diskurssin piirteitä (Kallionpää 2017, 84). Laajan kirjoittamistaidon opettaminen tähtää luku- ja kirjoittamiskäytänteiden muuttamiseen, jolloin painotetaan, että koululaiset kirjoittavat eri välinein erilaisissa tekstiympäristöissä tekstejä, joissa kommunikatiiviset tavoitteet ovat perinteistä koulukirjoittamista keskeisempiä.

opettajalle. Tällä seikalla on merkitystä nimenomaan sen suhteen, miten kirjoittaja ottaa lukijan huomioon. Myös Unelmakirjoitelmien alakouluikäisten kommunikaation tyypillinen piirre on suora kontakti vastaanottajaan (Vesala 2012, 76–77). Unelmakirjoitelmat tehtiin opettajan kontrollista irrallaan tutkimuskäyttöön mutta kuitenkin aikuislukijoille. Mielenkiintoinen kysymys on, kenen puoleen kuudesluokkalainen kirjoittaja kääntyy, kun hän aloittaa kirjoituksensa *Moi! Minä olen Teppo* tai *Hei! Olen Vilma* tai lopettaa sanoilla *Heh* tai *No, nyt tiedätte ainakin osan "Unelmieni päivän" tapahtumista?* Tuttavalliset puhuttelut voivat olla opettajan tai jonkun tuntemattoman puoleen kääntymistä. Kirjoittaja siis osannee ottaa huomioon kuvitteellisen lukijakunnan myös perinteisessä koulukirjoittamisen kontekstissa.

Vuoropuhelua lukijan kanssa rakennetaan paikoin suorasukaisesti: *nyt asiaan; seuraavaksi tulee mieleen; aion nyt kirjoittaa siitä*. Jo kuudesluokkalaisilla on tekstin kulkua kommentoivaa, kertomustekstille tyypillistä vuorovaikutuksen rakentamista: *Mutta se siitä aiheesta, nyt asiaan!* Keinoerot kuudes- ja yhdeksäsluokkalaisten välillä ovat kuitenkin havaittavissa. Peruskoulun päättävä kommentoi tekstin rakennetta kokonaisvaltaisemmin kuin kuudesluokkalainen, ja kommentointiin liittyy evaluoivaa suhtautumista. Lukijaa johdatellaan huomaamaan kertova minä (esimerkki 7).

7) *Toisena* tulee mieleen ystävät ja perhe hyvinvoivina tulevaisuudessa-kin. – *Seuraavaksi* tulee unelmista mieleen urheilu. (9F16)

Kertomuksen genressä lukijan huomioon ottava moniäänisyys liittyy etenkin orientaatioon, evaluointiin ja loppupäätelmiin (Labov 1972, 363). Aikuiskirjoittajilla kertomuksen orientaatio on tärkeä osa tekstiä. Sen tehtävänä on saada lukija tietoiseksi siitä, minkä vuoksi tarinan kertominen on perusteltua. Myös suomi toisena kielenä -ylioppilaskokeen kirjoitelmia tutkinut Komppa (2012, 58) toteaa, että orientointi on kirjoitelmien kokonaisrakenteen kannalta keskeinen tekstin osien välinen suhde. Aloitukset ja lopetukset ovat useissa teksteissä konkreettisesti erotettavissa. Pääideana on tässä opiskelun vaiheessa tuoda esiin, mistä aiheesta ja mistä näkökulmasta kirjoitelmassa aihetta käsitellään.

Unelmakirjoitelmien aloituksissa lukijan huomioon ottaminen on useissa tapauksissa melko yksinkertaista tehtävänantoon viittaamista. Ei sanota, miksi kertoja haluaa kertoa tarinansa. Luodaan tarinamaailma, joka usein on minäkertojan kotiympäristö tai tuttu matkakohde tuttuine henkilöineen. Useassa kuudesluokkalaisten kirjoituksessa (8a–c) on kaavamainen *Unelmieni päivänä heräsin* -aloitus. Myös suuri osa yhdeksäsluokkalaisten kirjoituksista (8d–e) alkaa samansisältöisellä toteamuksella:

- 8a) *Unelmien päivä alkaisi Rauhallisella herätyksellä.* (6M3)
- 8b) *Unelmieni päivä alkaisi alkaisi aamulla kun herään.* (6M11)
- 8c) *Minun unelmienipäivä alkaisi aikaisella herätyksellä.* (6F5)
- 8d) *Minun unelmieni päivä alkaisi heräämällä Havajilaisesta hotellista kaverini Paavon kanssa.* (9M7)
- 8e) *Unelmieni päivä olisi sellainen että herään aamulla ja minua ei väsyttä.* (9M19)

Tehtävänannon *Unelmieni päivä* mainitseminen tekstin alussa vaikuttaa kuitenkin olevan yritys asettaa oma teksti objektiksi ja päästä kirjoittamisen alkuun. Aloitukset ovat kuitenkin yksiäänisiä, sillä niissä ei varsinaisesti käännytä lukijan puoleen eikä tulla ulos kertomuksen maailmasta. Niissä ei siis tapahdu etäännyttämistä. Myös viides- ja seitsemäsluokkalaisilla suosituin tapa päästä tekstin alkuun on viitata annettuun aiheeseen: *Nyt kerron unelmapäivästäni*. Sen sijaan esimerkiksi seitsemäsluokkalaisten tekstin keskeiseen sisältöön viittaava aloitusvirke *Haluaisin yhden päivän ajaksi siivet* edustaa harvinaisempaa aloitustapaa. (Sintonen 2011, 37.)

Yhdeksäsluokkalaiset viittaavat kuudesluokkalaisia selkeämmin orientaatioissa siihen, mikä on tehnyt päivästä unelmien arvoisen, ja jatkavat teeman syventämistä. Kuudesluokkalaisilla lukijan huomiointi on melko yksinkertaista (9), yhdeksäsluokkalaisilla on moninaisempia keinoja ottaa lukija huomioon. He eivät toista tehtävänannon sanoja *Unelmieni päivä* vaan lähestyvät kauempaa kertomuksen maailmaa (10–11). Kokeva minä tuodaan kertovan minän rinnalle ja pysäytetään kerronta aurinkoiseen aamuun, ja samalla tuodaan tekstin teema sisäisen minän

kautta lukijan tietoisuuteen (*Tänään se tapahtuu; En voi uskoa tätä todeksi; Avaan verhot surullisina*).

9) *Unelmieni päivä jonka toivoisin toteutuvan olisi se, että näkisin isäni jota en ole koskaan tavannut.* (6F15).

10) *On aamu hotelli Presidentissä. Auringon säteet pyrkivät huoneeseeni sälekaihtimien väleistä. Olen juuri herännyt suureen päivään. Huonekaverini ovat jo ylhäällä. **Tänään se tapahtuu**, mistä olen aina haaveillut. Se on salibandyn MMfinaali. **En voi uskoa tätä todeksi**, mutta totta se on. Kotiyleisön edessä täydellä Hartwall areenalla, ja Ruotsi vastassa. Huippua!* (9F12)

11) *Auringon säteet pääsevät läpi vaaleiden verhojen raosta ja raotan unenpöpperöisiä silmiäni, jotka juuri ja juuri aukeavat. Pyyhin naamaani käsillä ja nousen hitaasti istumaan. Kävelen löntystellen ikkunan ääreen. **Avaan verhot surullisina**. Unelmani on Amerikassa eikä kerrostalossa Suomessa.* (9F23)

Aloituksen lisäksi kertomuksen lopetus on kohta, jossa tyypillisesti käännetään lukijan puoleen ja asetetaan oma teksti objektiksi. Kuudesluokkalaisten kirjoitelmissa lukee usein yksinkertaisesti *Loppu* tai nukkumaan menon toteamus *Illalla tulin kotiin menin nukkumaan*. Joissain kohdin kuitenkin palataan kertomuksen teemaan ja painotetaan, miksi kertomuksen esittäminen oli tärkeää (12–13), tai korostetaan omaa kokemusta (14):

12) *Unelmat olivat totta, Ja se johtui siitä että olin uskonut tosiunelmaan!* (6F12)

13) *Ai niin, vielä yksi unelma jäi kertomatta/ oppia unelmoimaan lentämisestä ja nälänhädän parantamisesta. Heh.* (9F16)

14) *Olisi vain minä ja jää.* (9F17)

Yksi osa oman tekstin objektiksi asettamista on huomion kiinnittäminen omiin kielellisiin valintoihin. Kielen näkyviin asettaminen edellyttää kirjoittajalta tietynasteista metakielistä tietoisuutta (Herlin 2005; Niinimäki 2015, 81). Unelmakirjoitelmista ei juurikaan löydy kuudesluokkalaisilta

tai yhdeksäsluokkalaisilta sellaisia konventionaalistuneita ilmauksia kuin *nimitetään joksikin, voidaan sanoa, suoraan sanoen, toisin sanoen tai totta puhuen* (ks. kuitenkin esimerkki 16: *tarkemmin ajateltuna*). Koulu-laisten kirjoituksissa metakielisyyttä on vain rajoitetusti ja rajoittunein keinoin toteutettuna. Useissa tapauksissa metakieleksi määrittäminen on tulkinnanvaraista, sillä nominaalilauseessa määritetään yhtä tarkoitetta, tehtävänannon lauseketta *unelmieni päivää* (15a–d)):

- 15a) *Minun unelmieni päivä olisi että saisin tehdä mitä haluan.* (9F26)  
15b) *Minun unelmien päivä on se, että pääsen kuuluisaksi.* (6M2)  
15c) *Unelmienipäivä olisi että voitaisin lotossa 10 miljoonaa euroa.* (6M25)  
15d) *Unelmieni päivä olisi sellainen että herään aamulla ja minua ei väsyttä.* (9F20)

Esimerkeissä 15a–d ei ole tietoista kielen ilmauksen asettamista lukijan huomion kohteeksi, mutta jos tarkoitteen määrittelyä pidetään metakielisyytenä, näitä määrittelyitä voi pitää moniäänisinä. Tarkoitteiden määrittelyissä ei ole aikuisten teksteissäkään selvää rajaa sen suhteen, milloin on kyse kielellisestä selityksestä, milloin tarkoitteen kuvauksesta (Niinimäki 2015, 82). Jonkinlaista lausekkeen merkityksen esiin nostamista kuitenkin tehdään, kun käytetään ilmausta *olisi että, on se että tai sellainen että*.

Myös kuudesluokkalaisten teksteistä löytyy ikäkaudelle ominaista sanojen merkitysten tarkastelua, joka parhaimmillaan on todellisuuteen viittaavaa pohdintaa (*tarkemmin ajateltuna*, 16). Nimittämiseen saattaa liittyä myös evaluointia (17). Voimakas nimitys – *karjuja* – yhdistetään äitiin ja lause rakennetaan niin, että huomio kiinnittyy juuri nimityksen valintaan. Tähdentäminen viittaa kertojan tietoisuuteen kielellisestä valinnastaan. Kuudesluokkalaisten metakielisyys on tyypillisimmillään huomion kiinnittämistä kertomuksen maailman ihmisten ja eläinten nimiin. Lukija joutuu ulos tarinamaailmasta ainakin kohdissa *ihmettelette varmaankin kaverini nimeä* (18) sekä esimerkissä, jossa kertoja jää pohtimaan Lola-nimen antamista koiralle (19). Esimerkeissä liikutaan

kahdella tasolla, osittain lauseen merkitykseen liittyvällä, osittain merkitystä pohtivalla.

16) *Tarkemmin ajateltuna päivä on kovin lyhyt aika elää unelmaansa.* (6F27)

17) sittä *se toinen karjuja, eli äiti, oli heittänyt itse kastellusta vessa-paperista tekemääni upeaa rekvisiittaa roskiin!* (6F27)

18) *Oli sovittu, että tulisin mökille bussilla, sillä nuunuun\* mökin vieressä oli bussipysäkki. ihmettelette varmaankin kaverini nimeä. Ei se ollut hänen nimensä mutta kaikki kutsuvat häntä sillä nimellä.* (6F31)

19) *Sillä aikaa kun koira nukkuu mietin, että minkä nimen annan sille. Päädyn kuitenkin antamaan sille nimeksi Lola. Mielestäni Lola sopisi täydellisesti nimeksi kauniille mustalle karvakasalle.* (6F11)

Moniäänistä metatason pohdintaa on erityisesti niissä yhdeksäsluokkalaisten kertomuksissa, joissa kirjoittaja pohtii oman unelmapäivänsä merkitystä ja todellisuus pohjaa. Silloin puhuja kommentoi omaa puhettaan ja suhteuttaa sanomaansa kulttuuria tai tekstuaalisia odotuksia vasten, minkä vuoksi ilmaukset voidaan nähdä moniäänisinä. Kehittyneimpiä pohdintoja voi verrata aikuisten teksteissä esiintyvään meta-kielisyyteen. Sanataiteessa kieli nostetaan usein tarkkailun kohteeksi ja esimerkiksi Lauri Viidan *Moreenissa* kulkee näkyvästi taso, jolla kiinnitetään huomio itse kieleen ja arvioidaan sen kykyä viitata todellisuuteen: *Tuli kovasti yrittää, kilvoitella, kuten sanottiin, hyväksyä ehdot mitkä tahansa ja hävitä reilusti, kunniallisesti.* (Viita 1973 [1950], 20). Esimerkeissä 20–21 pysähdytään pohtimaan, mitä unelmat kertojalle merkitsevät ja millainen niiden todellisuusviittaavuus on. Käytetään ilmausta *vaikea käsite*, mikä viittaa selvästi kielelliseen pohdintaan. Moniäänisyys saadaan molemmissa esimerkeissä aikaan myös viittaamalla muiden puheeseen (*jotkut osaavat; hän kysyi minulta*). Esimerkeissä kielen nostaminen metatasolle näyttää suunnitellulta ja lukijan huomioon ottavalta:

20) *Unelmat ovat minulle hirvittävän vaikea käsite. Jotkut osaavat tarinoida sivutolkulla, kuinka haluaisivat osata lentää tai parantaa nälänhädän.* (9F16)

21) Kesällä ollessani kaverini Karoliinan kanssa saarella, *hän kysyi minulta yhden kysymyksen: millainen sinun unelmiesi päivä olisi? Totta puhuen, en tiennyt silloin yhtään minkälainen päivä se voisi olla. Kysyin saman kysymyksen häneltä ja vastaus jollakin lailla yllätti, koska en ollut itse tullut ajatelleeksi asiaa siltä kannalta.* (9M5)

Tämän teoksen johdannossa viitataan paitsi oppilaiden erilaisiin kielellisiin kykyihin ja valmiuksiin myös heidän erilaiseen motivaatioonsa tuottaa tekstiä ja uskoon omiin taitoihinsa eli niin sanottuun pystyvyysuskoon. Kouluympäristössä oppilaat ovat kirjoitustilanteessa mukana sosiaalisessa prosessissa ja heillä on omassa yhteisössään erilaisia rooleja, joiden taustalla on erilaisia arvoja, ideologioita ja identiteettejä (vrt. Heikkinen & Voutilainen 2012, 17). Myös opetukselliset seikat vaikuttavat tulokseen. Koululla onkin vastuu tasoittaa paitsi kielellisten valmiuksien eroja myös niitä eroja, joiden taustalla ovat edellä mainitut arvoin ja identiteettiin liittyvät ennakkoluulot. Seuraava esimerkki 22 on osoitus kirjoittajansa taidoista ja korkeasta motivaatiosta. Siinä on niin kehittynyttä metakielen käyttöä, että syntyy vaikutelma suunnittelusta retorisesta kuviosta. Kirjoittaja käyttää ajanmääritteiden toistoa kohosteisesti, ja lukija havaitsee nämä merkitykset. Yhdeksäsluokkalaisten tekstissä voidaan puhua myös metafiktiosta, sillä siinä tietoinen toisto viittaa kirjoittamisen hetkeen mutta on osa fiktion maailman tapahtumia, eikä lukija joudu ulos tästä maailmasta:<sup>55</sup>

22) *Herään, ja kerrankin väsymys on poissa. Ei sitä uuvuttavaa tahmeutta, kaikkennieleväää toivottomuutta, yllä leijailevaa ukkospilveä. Kerrankin voin vain huoletta nousta, leijailla ajatuksiini uppoutuneena ympäri asuntoa. Ei kiirettä, turhautumista, ei ketään tyräämässä sanoillaan. – Sinä päivänä annoin itseni uskaltaa. Ehkä huomenna*

55 Lyriikkaan liitettävällä *metafiktio*n käsitteellä (Hallila 2006) viitataan kirjoittamisen hetkeen, kirjoittamiseen ja kertomisen aktiin. Käsitteen avulla voidaan valottaa myös koululaiskertomuksen kirjoittamisen aktia. Metafiktio on metatasoisuudestaan huolimatta fiktion sisässä esitettyä, osa fiktiivistä maailmaa. Tässä määritelmässä vertailtavuus koululaisteksteihin liittyy kirjoitushetkeen ja kirjoittamisen aktiin viittamiseen.



*kaikki on taas loisin, ehkä huomenna ahdistus palaa. Ehkä huomenna hallitsee taas pelko ja linnut putoavat taivaalta ja skeittini jää huoneen nurkkaan, ehkä huomenna kasvoilta paistaa pettymys. Ehkä huomenna, muttei tänään.* (9M6)

## Puheen ja ajatusten referointi

Aikaisempaan puheeseen tai ajatteluun viittaaminen, referoinnin eritavat, tuovat tekstiin useissa tapauksissa selkeästi erotettavan moniäänisyyden tason. Referointikeinot muodostavat jatkumon suorasta esityksestä epäsuoraan esitykseen, puheaktin ilmaisemiseen ja kerrontaan. Jatkumoa on käsitelty niin suomen kielen tutkimuksessa (esimerkiksi Penttilä 1957; Ikola 1961; Kuiri 1984; Hakulinen ja muut 2004; Kalliokoski 2005a.) kuin kansainvälisessä kirjallisuudessa (esimerkiksi Nordqvist 2001, Köder 2016). Erityisesti kehityksellisessä kontekstissa on tyypillistä, että erot eivät ole kaikilta osin selvät.

Unelma-aineisto osoittaa, että niin alakoululaisten kuin kuudesluokkalaisten ja yhdeksäsluokkalaisten referoinnin tavoissa, määrässä sekä muusta tekstistä erottamisesta on selviä eroja ikäryhmien välillä. Tarkastelen referointia niin kieliopin kuin tekstilajin sekä tekstin sisältämien erilaisten näkökulmien ja vaihtoehtoisten tulkintojen kannalta (vrt. Kalliokoski 2005a, 36.) Tulkinnat liittyvät nimenomaan siihen, mitä merkityksiä lukija liittää referoinnin kautta tuotuihin toisiin ääniin, eli miten referointi ohjaa lukijan tulkintaa (vrt. Virtanen, Rahtu & Shore 2018, 19–20).

### SUORA ESITYS – PUHEEN REFEROINTIA

Suorilla sitaateilla on eri tekstilajeissa erilaisia funktioita, mutta yhteistä on, että niiden kautta voidaan välittää lukijalle sellaisia henkilön puheen piirteitä, joita eivät epäsuora esitys tai vapaa epäsuora esitys välitä, esimerkiksi imperatiivimuotoisuus, puhuttelut ja partikkelien käyttö (Hakulinen ja muut 2004, 1401–1402; Kuiri 1984). Samoin voidaan suoraan tuoda ilmi tilannesidonnaisia ja affektiivisiä piirteitä sekä murteellisuuksia. Suorassa esityksessä referoija periaatteessa toistaa sen, mitä toinen

henkilö sanoi, ja myös sen, miten hän sanoi tai ajatteli. Yhdeksi suoran esityksen keskeiseksi funktioksi kertovassa tekstissä nousee henkilöiden idiolektin esittäminen ja sen myötä henkilökuvien rakentaminen.

Ensimmäisen luokan aineistossa ei ole vielä yhtään suoraa esitystä, mutta käyttö kasvaa kehityksellisesti. Kuudesluokkalaisilla<sup>56</sup> suora esitys on melko yleisesti käytössä (32 prosentilla aineistoni kirjoittajista, tyttöistä jonkin verran useammalla kuin pojista), mutta on vaikea tietää, kuinka moni kuudesluokkalainen kirjoittaja välttää referointia harjaantumattomuuttaan. Kuitenkin suuri osa niistä, jotka referointia käyttävät, on jollain tavoin tietoinen puheen erottamisen keinoista. Seuraavassa esimerkissä 23 kuudesluokkalaisen kirjoitelmassa käytetään dialogia kerronnan keinona niin, että dialogilla on myös kerronnan pysäyttävä, puhujia luonnehtiva funktio. Tyttöjen epäröinti ja herra Koskisen ystävällisyys paljastuvat repliikeistä, jotka eivät ole pelkästään tapahtumia eteenpäin vievässä tehtävässä. Kirjoittaja osaa erottaa vieraan puheen kerronnasta ja käyttää välimerkkejä normien mukaisella tavalla:

23) *Avasimme oven numero 1. Sisällä oli kokonainen kylpylä, ja kavერიni kysyi minulta: "Onkohan järkevää mennä sisään? Emmehän tiedä edes kuka on kuski!" "No, selvitetään asia.", sanoin. Huusimme kuskia, ja ovesta 3 tuli ulos kiltin näköinen mies. Hän sanoi: "Herra Koskinen, palveluksessanne neidit. Saatte tehdä täällä ihan mitä haluatte. Tämä limusiini on nyt omistuksessanne." "Voi kiitos!" sanoimme yhteen äänen. Meistä ja Herra Koskisesta tuli hyviä ystävyksiä. (6F4)*

Toisessa kuudesluokkalaisen kirjoitelmassa (esimerkki 24) rajataan referaatti johtoilmauksen avulla, mutta suoran esityksen välimerkit puuttuvat eikä virkerajoja ole merkitty. Oma ja lainattu puhe eivät erotu. Kirjoittaja luonnehtii henkilöitä puhekielisin repliikein: huolehtivainen äiti ja innokas pikkupoika.

56 Muhsen (2019, 27–29) mukaan vuosien 2006 ja 2014 kuudesluokkalaisilla suora esitys kasvaa kertomusrakenteen hallinnan myötä. Kertovia kirjoitelmia on aineistossa 187 ja niissä on yhteensä 120 suoraa esitystä, epäsuoran esityksen määrä on selvästi vähäisempi. Yhteensä kuudennen kirjoitelmia on 320.

24) *Silloin kun Pikku-Anssi oli syönyt jäätelön hän sanoi äidilleen Haluan särkänniemeen. Äiti vastasi okei kulta mutta ensin haemme siskosi ja isoveljesi mukaan. ja isäsi silloin Anssi Hyppäsi ilmaan ja huusi jippiiiii! (6M21)*

Pajunen (2012, 26) toteaa, että monet alaluokkien kirjoittajat vuoden 2006 Unelmakirjoitelmissa käyttävät vuoropuhelua vain kerrontaa eteen päin vievänä keinona. Muhosen (2019) mukaan tämä on vielä yleisempää vuoden 2014 aineistossa. Tällöin suora esitys harvoin kuvailee referoitavaa. Vuoden 2014 Unelmakirjoitelmissa kerronta etenee lyhyin vuoroin: *Hän kysyi äidiltä tuletko kanssani tee kuppeihin? Äiti vastasi okei! Sitten veli kysyi Hei Anssi tuletko kummitusjunaan Anssi ei uskaltanut sillä siellä oli pimeää* (6M21). Referoinnin keskeinen funktio narratiivissa onkin luonnollisesti juonen eteen päin kuljettaminen (ks. esimerkiksi Berman 2004) mutta suoran esityksen dialogimainen ketjutus tekee tekstistä raskasta. Kirjoitetun kielen osalta ei kuitenkaan ole selviä sääntöjä siitä, milloin käyttää suoraa milloin epäsuoraa esitystä ja kuinka paljon niitä kulloinkin kannattaa käyttää; kehityksellisesti suoran esityksen käyttäminen kirjoitetussa kielessä kuitenkin vähenee. Kuudennen luokan pojilla epäsuora esitystapa on jo hieman yleisempi kuin suora, työillä on päinvastoin. Ammattikorkeakoululaisten Unelmakirjoitelmissa on virkkeenkaltaisia T-yksikköjä kaikkiaan lähes 4 000, niistä 1,6 prosentissa on suoraa esitystä ja 6,1 prosentissa epäsuoraa (Tiiro 2021).<sup>57</sup> Puhutussa kielessä suora esitys voi lisääntyäkin ja silloin sen kuvailevuutta väritetään multimodaalisin keinoin.

Useissa kuudesluokkalaisten teksteissä suoraa esitystä ei rajata selkeästi kertomuksessa omaksi moniäänisyyttä tuottavaksi jaksokseen. Esimerkissä 25 kuudesluokkalainen tyttö perustelee kirjoitelmansa orientaatioissa miksi esittää kertomuksensa ja kuvailee lauantaiaamun väsyneen tunnelman ja syntymäpäiväkoristelun yksityiskohtia mutta ei

57 Epäsuoran esityksen määrää on aikuiskirjoittajien omaelämäkerrallisissa tarinoissa aika vaikea arvioida, koska ne on usein kirjoitettu kokonaan kokemuksellisesta näkökulmasta. Prosentti voisi olla huomattavastikin korkeampi, jos mukaan laskettaisiin erilaiset tuntemuksien tms. ilmaisut. Annettu prosentti on laskettu epäsuoran esityksen kehyslauseiden avulla.

erota orientaatiota muusta tekstistä eikä muitakaan kerronnan uusia näkökulmia. Hän kirjoittaa moniäänistä tekstiä, jossa tulevat esiin kertojan omat tuntemukset sekä perheenjäsenten repliikit, mutta hän ei rajaa puheenvuoroja eikä luo tekstiinsä koherenssia (vrt. Kalliokoski 2005b, 225). Kirjoittaja ei siis paljoakaan opasta lukijaa moniäänisyyden tulkinna.

25) *Heräsin varhain lauantai aamuna. Nousin sängystä väsyneesti. Laahustin keittiöön. Wow! Lamppuihin oli ripustettu serpenttiiniä, pöydällä oli valtava kakku jossa oli kaksitoista kynttilää, sekä pöydällä oli paljon muitakin herkkuja. Tuoleista riippui ilmapalloja. Mutta miksi tämä kaikki? Hyvää syntymäpäivää! Havahduin. Aivan, tänään oli minun syntymäpäiväni! Paikalla oli veljeni, vanhempani, sukulaiseni ja kaverini kaikkilla leveä hymy kasvoillaan. Istu pöytää! äiti sanoi. Puhalsin kynttilät, söin kakkua, sekä mutustin karkkia. Aika kiva aamupala, ajattelin. Paljon onneaaaa vaan, paljon onneaaaa vaan raikui. "Aamiaisen" jälkeen puin vaatteet päälle ja tein muut aamupuuhat. Äiti sanoi: viemme sinut syntymäpäiväsi kunniaksi Fun Parkkiin. Jeeeeeeeeeee! huusin. (6F15)*

Kuudesluokkalaisilla leksikaalinen variaatio on kehystävissä johtolauseissa usein vielä vähäistä. Koko alakoulun aineistossa on yleistä käyttää *sanoa*-verbiä puheenvuoron merkinä tai puheenomaisesti, vaikka *kysyä*, *vastata* tai jokin merkitykseltään spesifisempi verbi olisi odotuksenmukaisempi (vrt. Pajunen 2012, 25.) Kehityksellinen ero näkyy hyvin, jos verrataan yksittäisten puheaktiverbien keskimääräisiä esiintymiä Unelmakirjoitelmissa. Alakoulun luokilla 1.–5. esiintymiä on yhdellä puheaktilekseemillä keskimäärin lähes kymmenen ja kuudennella luokalla yhdeksän mutta yläkoulussa ja aikuisilla määrä putoaa kuuden paikkeille: leksikaalinen variaatio siis lisääntyy. Vastaavasti kaikista verbiesiintymistä puheaktiverbien osuus lähes kaksinkertaistuu alakoulusta aikuisiin. Toisin sanoen illokutiivisten sävyjen ilmaisumahdollisuudet kasvavat selvästi iän mukaan. Samanlaisen kehityksen toteavat myös Nordqvist (2001) ja Köder (2016). Esimerkissä 26 kuudesluokkalaisten pojan tekstissä *sanoa* toistuu johtolauseen verbinä. Kerronta etenee vuoropuhelu-

na. Kirjoittaja rajaa referaatin johtoilmauksen sekä kaksoispisteen avulla mutta muuten välimerkkejä ja isoja kirjaimia käytetään säästeliäästi.

26) *Hän luuli, että olisin karannut tai minut olisi siepattu. **sanoin** äidille: tule hakemaan mut täältä K-City-marketi pihasta. äiti **sanoi**: tulen ja sitten saat siivota. nähdään kohta. Viiden minuutin kuluttua tuttu Nissan tuli. Menin kyytiin. Äitini **sanoi** älä sitten ikinä tee tuolaista. Mutta **sanoin**: ME VOITIMME 50 MILJOONAA EUROA. Äiti **sanoi**, että höpöhöpö Näytin Sekkiä äidille. äiti meinasi ajaa ojaan, kun näki sekini. **sanoin**, että nyt olemme rikkaita, mutta pyöräni varastettiin. Äidin hymy ei kuitenkaan hyytynyt. Hän **sanoi**: Entä sitten olemme rikkaita. (6M17)*

Esimerkissä 27 kuudesluokkalainen tyttö sen sijaan taitaa välimerkit suoraa esitystä käyttäessään. Suora esitys keskeyttää kerronnan ja johtolauseen verbi *huikata* kuvaa hyvin kohteena olevaa henkilöä. Lisäksi kirjoittaja luottaa lukijan kykyyn tehdä päätelmiä. Tässä tapauksessa kirjoittaja pystyy etäännyttämään tekstin itsestään ja teksti on aidosti moniäänistä.

27) *Kopissa istuvat kaikki vanhat kaverit omilla vakipaikoillaan. Essi katsahtaa jälle ja **huikkaa**: ”Meillä on sitten Susanna. Vauhtia!” Kaikki kaivavat nenäliinat äkkiä treenilaukuistaan ja kiiruhtavat yhdessä jälle. (6F13)*

Suoran esityksen käyttö vähenee huomattavasti peruskoulun päättövaiheeseen edettäessä. Yhdeksäsluokkalaisten kertomuksista suora esitys on lähestulkoon kadonnut, sillä vain vajaa kymmenes osa aineistoni kirjoittajista referoi puhetta suorana esityksenä. Keskusteluun viitataan useasti vain puheaktin tiivistelmällä kuten esimerkeissä 28a ja b.

28a) *Puhun mahdollisimman hyvää Amerikkaa ja **kerron** osoitteen ja lähdemme matkanpäähämme. (9F23)*

28b) ***Kertoisimme** itsestämme ja Suomesta jotain. (9F26)*

Ne harvat yhdeksäsluokkalaiset, jotka käyttävät suoraa esitystä, yleensä osaavat tuoda toisen äänen tekstiin, arvottaa omaa puhettaan ja referoida myös ajatuksiaan. He osaavat ottaa aikaisemmat puheenvuorot tarkoituksenmukaisesti oman kertomuksensa osaksi. Kirjoitelmassa 29 toinen ääni tuo vieraan henkilön näkökulman, jonka avulla kertoja arvottaa asioita. Esimerkin 30 kirjoittaja pysäyttää kerronnan dialogin avulla. Hän varioi myös puheaktiverbiä tarkoituksenmukaisesti ja käyttää puhetta erottavia välimerkkejä melko virheettömästi. Kolmannen esimerkin 31 kirjoittaja tietää, että suorana esityksenä voi referoida myös omia ajatuksiaan:

29) *"Unelmeni päivään kuuluisi se, että minulla olisi niin paljon rahaa, että voisin maksaa kaikki perheeni velat pois," Karoliina kertoi. Sen jälkeen omat unelmieni päivän aiheet tuntuivat todella itsekkäiltä.*

(9M5)

30) *"Oletko pakannut matkalaukkusi" huutaa äiti samalla kun kävelee vessan ohi missä olen pesemässä hampaitani. "Joo joo" yritän sanoa selvästi mutta suuni oli täynnä hammastahnaa. – – "Olemme jonottaneet jo yli puoli tuntia" huokaisin. Kunnes jono alkoi liikkua, huokailuni loppui.* (9F11)

31) *En saanut muita sanoja suustani kuin "kiitos" "Mutta kiitoshan oli juuri sopiva sana", mietin.* (9F11)

Suoran esityksen käytön väheneminen vastaa tarkoin monista muista kielistä esitettyä kehitystä (vrt. Goodell & Sachs 1992; Nordqvist 2001; Köder 2016). Ilmiö selitetään usein viittaamalla referoinnin keinojen epälineaariseen kehitykseen: alakouluikäiset ja nuoret aikuiset käyttävät paljon suoraa esitystä – vaikka eri tavoin –, murrosikäiset tukeutuvat epäsuoraan esitykseen. Väite ei toteudu aikuisten Unelmakirjoitelmista, mutta ero voi johtua aiheesta ja siitä, että kerronta painottuu kokemuksellisesti. Nordqvistin aineisto on kuvaelisitaation avulla tuotettua sammakkotarinaa, jossa kerronta etenee etsintäteemana ja etsinnän aikana kohdataan useita eläimiä: ruotsalaiset aikuiskoehenkilöt ilmaisevat näitä kohtaamisia sekä pojan että eläinten fiktiivisten repliikkien suorana esityksenä. Suomen aineistossa sammakkotarina kirjoitti vain puhutussa

versiossa suoraa esitystä mutta ei lainkaan kirjoittamalla tuotetussa (sammakkotarinoista ks. johdantolukua). On vaikea sanoa, mistä ero johtuu. Myös aikuisten Unelmakirjoitelmissa suora esitys on lähes olematonta, kuten jo todettiin.<sup>58</sup>

#### EPÄSUORA ESITYS – PUHEEN JA AJATUSTEN REFEROINTIA

Suora ja epäsuora esitys ovat keskeisiä referoinnin pääluokkia (*oratio directa* ja *oratio obliqua*) (ks. Evans 2012). Suorassa esityksessä (Hakulinen ja muut 2004, 1406–1407) referaattiosan keskus on representoidussa tilanteessa, epäsuorassa esityksessä puhehetken aikaisessa tilanteessa, jolloin asiaan otetaan etäisyyttä kerrontahetken perspektiivistä. Puhuja ja puhetilanne ratkaisevat, millaisen muodon lause saa. Referoinnin keinojen esittäminen jatkumona (esimerkiksi Hakulinen ja muut 2004, 1400; Leech & Short 1981, 324; Evans 2012) perustuu kertojan kielellisen osuuden määrään, joka lisääntyy asteittain suorasta esityksestä epäsuoraan ja referoinnin muihin muotoihin.

Syntaktisesti suoran ja epäsuoran esityksen eroa luonnehditaan niin, että suorassa esityksessä lauseiden suhde on rinnastussuhde, epäsuorassa esityksessä alistussuhde. Varsinkaan aikuisten sanataiteen kertomuksissa referointikeinojen rajat eivät ole kategorisia, vaan ne on häivytetty ilmeisen tarkoituksellisesti. Samoin raja rinnasteisuuden ja alistaisuuden välillä on hämärtynyt. Tapauksia voi verrata puhutun kielen referointiin, jossa ero suoran ja epäsuoran esityksen välillä katoaa. (Vrt. Niinimäki 2015, 142–148.) Neutraaleiksi luonnehdituissa tapauksissa muodon ja sisällön distinktio katoaa. Tällöin *että*-konjunktio ei ilmaise alistaisuutta vaan tekstuaalista sidoksisuutta johtolauseen ja referaatin välillä ja vastaa *näin*-sanana käyttöä: *Hän sanoi NÄIN, pula on aina läsnä* –. (Kuiiri 1984, 98.)

Peruskoulun ylemmillä luokilla epäsuora esitys lisääntyy suhteessa suoraan esitykseen. Samoin merkitykseltään kompleksisten puheaktiverbien käyttö lisääntyy, kuten todettiin edellä. Kuudesluokkalaisilla *että*-lauseella alistettua epäsuoraa esitystä on noin puolella kirjoittajista, tosin koko kuudennen aineistossa jonkin verran vähemmän. Yhdeksäs-

<sup>58</sup> Aikuiskoehenkilöiden määrä Unelma- ja sammakkoaineistossa on yhteensä lähes 400.

luokkalaisten kertomuksissa epäsuoraa esitystä on yli puolella kirjoittajista, siis selkeästi useammalla kuin suoraa esitystä.<sup>59</sup>

Tarkastelen epäsuoraa esitystä siltä kannalta, miten sen avulla tuodaan tekstiin muita ääniä ja samalla tarkastelen myös, missä funktioissa epäsuora esitys koululaisten teksteissä on: referatiivisena, kirjoituksen etenemiseen vaikuttavana, vai episteemisenä, tekstin dialogisen tilan (= mukana vähintään kaksi ääntä tai näkökulmaa) säätelyyn liittyvänä (vrt. Virtanen & Juvonen 2018, 191 blogikirjoittamisen vuorovaikutuksesta). Epäsuorassa esityksessä äänet ovat joko kertojan oman puheen tai muun henkilön puheen tuottamia toisia ääniä tai minäkertojan omien aiempien ajatusten esille tuomista (Virtanen & Juvonen 2018, 185). Myös koululaisten teksteissä epäsuora esitys todentuu referaatteina, joissa puhuja referoi joko puhetta tai omia ajatuksiaan, kokemuksiään ja tuntemuksiaan. Koska koululaisten tekstit ovat minämuotoisia, niissä esitetty ajatusten referointi on pelkästään omien ajatusten esittämistä (muutamassa tapauksessa on me-muoto).

Tapaukset poikkeavat toisistaan kerronstrategiassa niin, että puheaktiverbiä käytettäessä referointi on tyypillisesti tarinaa kuljettavassa tehtävässä, mutta ajatuksia ja kokemuksia referoitaessa puhuja pysäyttää kerronnan ja tuo esiin kokevan minän. Kuudesluokkalaisten epäsuora esitys (esimerkit 32–33) on huomattavasti enemmän puheen referointia kuin yhdeksäsluokkalaisten, joilla epäsuora referointi painottuu ajatusten, kokemusten ja tuntemusten esittämiseen. Kuudes- ja yhdeksäsluokkalaisten Unelmakirjoitelmissa epäsuora esitys noudattaa vain harvoin niitä normeja, joilla kirjoitetun tekstin referaattia ohjeistetaan muuttamaan meneillään olevan puhetilanteen mukaiseksi. Koko alakoulun aikana *että*-lause toteutuu usein sekaesityksenä, jota on käytetty aiemmin myös vapaan epäsuoran esityksen nimityksenä (Hakulinen ja muut 2004, 1405–1406, 1413–1414).

Tarkastelen ensin sellaista epäsuoraa esitystä, jossa alisteisessa lauseessa referoidaan puhetta. Esimerkissä 32 kertoja tuo tekstiin toisen

59 En käsittele erikseen tapauksia, joissa kirjoittaja käyttää lauseenvastikerakennetta, koska niitä esiintyy vain satunnaisesti: *Nousin sängystä ja puin vaatteet päälleni kävelin kohti keittiötä jossa oli lappu jossa äiti kertoi lähenteensä ylitöihin ja että pöydällä olevan pitsan sai syödä* (6M18).



äänen, ja esimerkissä 33 kertoja referoi omaa aikaisempaa puhettaan. Kumpikaan kirjoittaja ei muuta repliikin aikamuotoa tai modusta (oikeammin *lähdettäisiin*) eikä ajanmääritteitä (*nyt; tänään*) kerrontahetken mukaiseksi. On aika yleistä, että deiktisistä kategorioista juuri tempus jää muuntamatta (Evans 2012). Myös yhdeksäsluokkalaisilla epäsuoran esityksen aikamuotojen ja muiden deiktisten elementtien muuttaminen puhehetken mukaisiksi vaihtelee. Esimerkissä 34 kirjoittaja tuo tarinaan uusia ääniä kertojan äänen rinnalle, mutta referointi on funktioltaan pääosin tarinan kuljettamista eikä kerronnan pysäyttämistä tapahdu. Kirjoittaja ei etäännyttä kerrontaa itsestään, ei siirrä deiktistä keskusta tapahtumahetkestä puhehetkeen eikä muuta persoonapronomineja tai deiktisiä elementtejä puhetilanteen mukaisiksi (*täällä; menettä*). Lisäksi puhekielinen *että*-konjunktio on kysymyslauseen alussa.

32) *Minä lähdin kotiin, mutta kun olin kotona äiti sanoi, että nyt lähdetään Eedenii (kylpylään).* (6F9)

33) *Myöhemmin aamupäivällä soitin kaverilleni, ja kyselin, että voiskohan hän tulla tänään kanssani kaupunkiin.* (6F4)

34) *Mies kysyi, että mitä me täällä teemme Sanoimme, että tulemme pelastamaan eläimet Mutta sitten mies sillä hetkellä soitti poliisit. kun poliisit tuli paikalle he kysyivät, että mikä nyt on ongelmana. Sanoimme, että menettä katsomaan tuonne sisälle miten eläimet kärsivät. – – Poliisi sanoivat meille, että saatte viedä kaikki eläimet ja hoitaa ne kuntoon.* (9F25)

Edelliset sekaesityksenä referoidut kohdat voi rinnastaa puhutun kielien referointiin, jossa ero suoran ja epäsuoran esityksen välillä katoaa. Vaikka koululaiskirjoittaja ei ottaisikaan referaatin funktiota huomioon, hän kuitenkin toteuttaa edellisenlaisissa tapauksissa suoran esityksen funktiota luonnehtia repliikillä puhujaa, kuten poliisia (*että mikä nyt on ongelmana*) tai nuoria tyttöjä (*menettä katsomaan tuonne sisälle*). Sekaesityksen näkemystä tukee myös *että*-konjunktin olemassaolo melkein poikkeuksetta epäsuoran kysymyksen alussa. Samoin tapahtumahetken tuominen puhehetken osaksi sekaesityksenä on luontevaa, vaikka se on oikeinkirjoituskäytänteiden vastaista. Voikin todeta, että vaikka kirjoitta-

jat eivät tukeudu normeihin, he osaavat tuoda toisen äänen tekstiinsä. Myös aikuisten sanataiteilijoiden kertomuksissa on usein tarkoituksellisesti sekaesityksen piirteitä. Varsinkin vapaan epäsuoran esityksen kautta saadaan henkilön ääni kuuluviin kertojan äänen rinnalla.

Kirjoittajat tuovat toisinaan esiin myös tiedollisen arvionsa referoidusta asiantilasta. Kertoja *tietää* asiantilan (37), *luulee* asiantilasta jotain (35) tai vain *ilmaisee ajattelevansa* asiantilaa (36). Esimerkkitapauksissa kirjoittaja jäsentää kehyslauseen avulla tekstiä ja ottaa lukijan huomioon. Ajatuksen samoin kuin puheen referointi tapahtuu sekaesityksenä, ja tapahtumahetki on läsnä esimerkiksi ilmauksissa *tästä päivästä tulee* (36) sekä *tästä* (37):

35) *aluksi luulin että se oli isoveljeni tekemä pila.* (6M5)

36) *Heräsin aamulla ja ajattelin, että tästä päivästä tulee kaikkein paras päivä.* (6F4)

37) *Heräsin virkeämpänä ja tiesin että tästä tulisi hyvä päivä.* (9M12)

Juvonen (2014, 59) toteaa lukioikäisten kirjoittajien yhdyslauseiden *tietää*-rakenteista, että niiden käyttämisessä on kyse lukijan odotusten herättämisestä tekstin etenemisen suhteen. Tällöin lukijan rooliksi hahmottuu havainnoida tekijän ajattelua eikä osallistua siihen. Juvosen mukaan kyse on usein toistuvasta ja kiteytyneestä tavasta osoittaa asennoituminen puheenalaiseen asiaan, jäsentää vuorovaikutusta ja ylläpitää suhdetta vastaanottajaan tai puhekuppaniin. Peruskoululaisten teksteissä rakenteita *tiesin että*, *luulin että* ja *huomasin että* käytetään vastavissa funktioissa. Kyse on paitsi tekstinsisäisestä viittaamisesta myös oman tekstin jäsentämisestä ja lukijan huomioon ottamisesta eli tekstinulkoisesta moniäänisyydestä.

Joskus kirjoittaja juuttuu korostamaan kerrontaansa ja käyttää ajatusta, havaintoa tai kokemusta ilmaisevaa verbiä toistuvasti. Esimerkeissä 38a ja 38b kirjoittajat pitävät puhehetkeä ja kertovaa minää korostetusti esillä toistamalla *huomata*-verbiä. Toisto voi olla joko taitamattomuutta tai tietoista retorisen keinon käyttöä.

38a) *Mutta hetken päästä huomasi etten ollut kotona. Huomasi minulla olevan mustan hatun, sinisen viitan ja valkoiset vaatteet.* (6M11)  
38b) *huomaan, että Äitini on käynyt kaupassa – ja huomaa, että kaverini ei ole vielä herännyt – huomaa, että kylässäni (siinä pelissä) on ruoka loppu – joten päätän lopettaa pelin ottaeni huomioon .että metripitsä näyttääkin olevan valmis äitini mukaan.* (9M13)

Kuudesluokkalaisilla tämäntapainen referointi liittyy tilanteeseen, ei esimerkiksi yleiseen pohdintaan, yleisiin normeihin tai aikaisempiin tapahumiiniin. Funktio on tyypillisesti kerronnallinen. Yhdeksäsluokkalaisilla taas funktiona on episteemisen modaalisuuden mukaan tuominen, kuten esimerkissä 38b, jossa kirjoittaja välittää minäkertojan ajatuksia ja tuntemuksia ilmauksin *tietää että, huomata että, ajatella että ja toivoo että*.

#### VAPAATA EPÄSUORAA ESITYSTÄ JA TAJUNNANVIRTAA

Vapaa epäsuora esitys tuo tekstiin moniäänisyyden, sillä siinä yhdistyvät sekä kertojan että referoitavan henkilön näkökulmat ja sanamuodot. Deiktiset elementit voivat olla kerrontatilanteen mukaiset mutta sanavalinnat aikaisemman puhetilanteen ja alkuperäisen puhujan valintoja (ks. Hakulinen ja muut 2004, 198–200.) Kielellisistä seikoista voi päätellä, kuinka kokonaisvaltaisesti kertoja ottaa toisen äänen mukaan. Vaikka esitystapaa käyttävät pääosin osaavat aikuiskirjoittajat, saman tapaista referointikeinoa käyttävät jonkin verran myös varsinkin kehittyneet yhdeksäsluokkalaiset. Myös kuudesluokkalaisten kirjoitelmissa kuvaillaan omia ajatuksia, ja yhteys vapaan epäsuoran esityksen kaksiaäänisyyteen löytyy usein siitä, että kertovan minän ääni yhdistyy kerrottuun minään.

Esimerkin 39 kuudesluokkalaisten teksti on monin tavoin moniääninen. Aluksi tuodaan esiin vieras ääni epäsuorana esityksenä (*sain kuulla että*), minkä jälkeen kertovan minän ääni yhtyy kerrottuun minään. Näin kertoja tuo ilmi tietonsa tulevasta pelistä.

39) *Samalla sain kuulla että Valmentajat valitsivat minut maajoukkueeseen. Ensimmäinen peli on kuukauden päästä Espanjaa vastaan.* (6M19)

Esimerkissä 40 kertoja toteaa kertovana minänä aikaisen aamuherätyksensä mutta tekee sen jälkeen näkökulman vaihdoksen kokevaan minään: *Tänään on minun syntymäpäiväni*. Tulkitsen mietteen vapaaksi epäsuoraksi esitykseksi. Sen jälkeen kerronta jatkuu kellonajan toteamisella.

40) *Heräsin aikaisin aamulla. Tänään on minun syntymäpäiväni. Katsoin kelloa ja se oli 7.05. (6M22)*

Minäkertoja esimerkissä 41 tuo kaverin äidin äänen mukaan vapaana epäsuorana esityksenä referaatilla *Vasikka oli karannut* ja siirtyy sen jälkeen toiminnan kuvaukseen. Pluskvamperfekti *oli karannut* viittaa toiselta saatuun tietoon, johon kertoja tukeutuu. Esimerkissä 42 vaarin ääni tulee yhdeksäsluokkalaisen kertomuksessa esiin minäkertojan puheen lomassa. Konditionaali on osoitus kertojan ja vaarin näkökulmien limittymisestä. Kertoja tuo konditionaalina kautta esiin epäroivän suhtautumisensa vaarin aikeisiin (Hakulinen ja muut 2004, 1412–1413).

41) *Keskellä yötä Tomin äiti huutaa meitä. Vasikka oli karannut, etsimme kaikkialta, mutta vasikkaa ei löytynyt. (9M8)*

42) *Vastasin ja vaari pyysi minut laivan ravintolaan jossa minua odotaisi yllätys. (9F11)*

Vapaata epäsuoraa esitystä edustavat myös esimerkit 43–44, joita täydentävät parenteettiset jälkilauseet *ajattelin innoissani* ja *ajattelin*:

43) *Tätä minä olen odottanut, tästä päivästä minä olen haaveillut, ajattelin innoissani. (9F11)*

44) *Aika kiva aamupala, ajattelin. (6F15)*

Vaikka nimitys *vapaa* kertoo, että esitysmuodosta puuttuu johtoilmaus, pidän moniäänisyyttä, sitä että referoitavan ääni kuuluu kertojan äänen rinnalla, vapaan epäsuoran esityksen oleellisimpana piirteenä. Sen sijaan irrallinen jälkilause ei ole yhtä merkittävä kuin referaattia edeltävä, alistava lause. Sen funktiona on lähinnä täydentää vapaan epäsuoran

esityksen referaattia. (Hakulinen ja muut 2004, 1418–1419; Niinimäki 2015, 201–205).

Erilaiset puheen ja ajatusten esitysmuodot limittyvät, kuten edelliset koululaisten referaatitkin osoittavat. Koska moniääniseksi määrittelyssä on paljon rajatapauksia, ei ole tarpeen määritellä eroja tarkasti. Berg (2011, 342) toteaa abiturienttien aineistoaineiden referoinnista, että juuri vapaa esitys ja toisaalta monivirkeinen referaatti ovat ongelmallisia kuvattavia, ja referoinnin tunnistaminen ja kategorian määrittäminen voivat vaatia koko tekstin tarkastelua. Edelliset esimerkit osoittavat, että moniäänisyyden luominen onnistuu monin keinoin, vaikka varsinkin vapaaksi epäsuoraksi esitykseksi nimeäminen saattaa olla tulkinnanvaraista.

Aikaisemmin vapaata epäsuoraa esitystä nimitettiin eläytymisesitykseksi (ks. Hakulinen ja muut 2004, 1405), jossa viitataan referoitavan tuntemuksiin. Tuonkin esiin eläytymisesityksen rinnalla toisen fiktiivisille teksteille tyypillisen kertojan mielenliikkeitä kuvaavan ilmaisumuodon, tajunnanvirran kuvauksen, jossa välitön ja karsimaton esitystapa ovat oleellisia.<sup>60</sup> Tajunnanvirrassa kertoja muuttaa omaa kertojan asemaansa: kertova minä muuttuu kokevaksi minäksi, jolloin tekstiin tulee moniäänisyyden taso. Myös kehittyneissä koululaisten kirjoituksissa löytyy tajunnanvirran kuvaukseksi luonnehdittavaa omien tuntojen kuvausta. On huomattava, että Unelmakirjoitelmien tehtävänanto ohjaa minäkertojaa tarkastelemaan asioita sisäisestä näkökulmasta, jolloin ulkoinen kerronta luontevasti muuttuu sisäiseksi ja moniäänisyyden ilmiö toteutuu.

Esimerkissä 45 kertojan tajunnan kuvauksessa on kysymys moniäänisestä puheesta. Minäkertoja tuo kokevan minän mukana vieraat äänet, *sanoilla viiltämisen* ja *pois häätämisen*: Tässä tapauksessa moniäänisyys syntyy sitäkin kautta, että tajunnanvirrassa käytetään viittauksia ulkopuolisten toteutumattomiin puheakteihin. Referointitapa on puheaktin kuvaus, jossa referoitavan ääni on mahdollisimman vähän esillä ja jota ei selkeästi eroteta muusta kerronnasta.

60 Kerrontateknikka, jossa henkilön mielenliikkeet, aistimukset, satunnaiset mielteet, muistumat, jopa unet pyritään esittämään mahdollisimman välittömässä ja karsimattomassa muodossa, jatkuvana miellevirtana. Samalla rikotaan usein kieliopin, syntaksin ja logiikan sääntöjä. (Hosiaisuusluoma 2003, 903.)

45) *Kun lähden ulos, kukaan ei katso minuun arvostelevasti, kukaan ei viillä minua sanoillaan, kukaan ei häädiä minua pois.* (9M6)

Koululaiskirjoittajilla on myös sellaista fiktiiviselle kerronnalle ominaista tekstityypiltään kuvaavaa tekstiä, jossa ympäristön ja tajunnan kuvaus sekoittuvat ja jossa kertovaa ja kokevaa minää ei voi erottaa toisistaan. Yhdeksäsluokkalaisen tekstin katkelma (esimerkki 46) on moniääninen, sillä kertoja vaihtaa kerronnan tasoa omasta tajunnastaan (*viikon paras päivä*) ulkoisen kuvaukseksi (*aurinko paistoi; tuuli henkäili hiljaa*):

46) *Oli viikon paras päivä, PERJANTAI! Ja päivää vielä paremmaksi teki aivan käsittämättömän upea keli. Aurinko paistoi lämpimästi lähes pilvettömältä taivaalta, tuuli henkäili hiljaa kahisuttaen puiden vihreitä lehtiä.* (9F13)

Esimerkissä 47 on lukijalle välittyvää kertojan tajunnan kuvausta. Minäkertoja, luistelijatyttö, jää pohtimaan asioita ja lähestyy lukijaa retorisin kysymyksiin, jotka myös tuovat tekstiin moniäänistä vuorovaikutteisuutta. Kertojan asema muuttuu toiminnan kuvauksesta (hermostuminen ja kaatuilu) oman toiminnan arvioimiseen ja siirtymiseen kuvitteelliseen maailmaan, josta kertoja kertoo epäsuoraan esitykseen rinnastettavin *että*-lausein. Moniäänisyys syntyy, kun asemointi muuttuu kertovasta minästä kokevaksi minäksi.

47) *Saatan hermostua oman äitinikin katseesta. Tämä johtaa yleensä kompasteluun tai jopa kaatumiseen. Ei kaatumisessa mitään pahaa ole, kaikkihan sitä tekevät. Mutta joskus se pistää arsyttämään. Jos kaikki katselijat näkevät, kun kaadun, mitä he ajattelevat? Tuomitsevatko he minut sen perusteella huonoksi luistelijaksi tai epäonnistujaksi? – – Mutta salaa minä unelmoin, ihan melkein joka kerta, että kymmenen hengen sijaan jäällä olisin vain minä. Että kaikkialla vallitsisi hiljaisuus, jonka rikkoisi vain minun terieni ääni ja vaimea tömähdys vain minun kaatuessani.* (9F17)

Esimerkeistä 45–47 näkyy sekä iän että kirjoittajakohtaisen kehityksen tuloksena lisääntyvä kuvailu ja evaluointi. Nimenomaan vapaan epäsuoran esityksen käyttö on kehittyneiden kirjoittajien hallussa. Kehittyntä kirjoitustaitoa näyttää myös osoittavan, että lukijaa johdatteleva johtoilmaus on uskallettu jättää pois. Tällöin kokeminen ja kertominen limittyvät ja kerronta tiivistyy. Berg (2011, 342) toteaa, että referointikeinojen käyttö paljastaa abiturienttien oman syntaksin mahdollisuudet ja rajoitukset. Samoin peruskoulun kuudes- ja yhdeksäsluokkalaisilla kyse on ilmeisesti myös oman syntaksin kehittyneisyydestä referointitapojen käytössä.

Oppikirjoissa vapaata epäsuoraa esitystä ei käsitellä peruskoulun aikana, mutta silti jotkut kirjoittajat ottavat esitystavan käyttöönsä tietyssä kehityksensä vaiheessa. Tällöin kyse on ilmeisesti kielellisestä harrastuneisuudesta kuten kirjoittamisesta ja lukemisesta, jotka tukevat kielen oppimista. Myös S1- ja S2 -ryhmien leksikon kehittymiseen vaikuttaa monipuolinen kielellinen harrastuneisuus ja positiivinen suhtautuminen kieleen (Kalliokoski 2005b).

## Lopuksi

Moniäänisyys on semanttis-pragmaattisesti kompleksinen ilmiö, jonka kielentäminen opitaan puhutussa kielessä melko varhain mutta jonka siirtyminen kirjoitettuun kieleen on erittäin hidas prosessi. Tämä prosessi voi olla kesken vielä peruskoulun päättövaiheissa. Kehitys suhteutuu kielellisen koherenssin luomiseen ja kertomusrakenteen tai yleensä tekstin rakentamisen oppimiseen (vrt. Mäkinen, Loukusa, Laukkanen, Leinonen & Kunnari 2014). Suomalaislapsilla kertomusskeema alkaa olla hallinnassa alakoulun lopulla – ei tosin kaikilla – ja referentiaalisten viittausten määrä kasvaa nopeasti niillä lapsilla, joilla kerronta muutenkin onnistuu (ks. Kouvon ja muiden analyysiä tässä teoksessa). Yksiäänisessä kerronnassa referentiaaliset viittaussuhteet ovat yksiselitteisempiä kuin moniäänisessä, joten on odotuksenmukaista, että referentiaalinen koherenssi täydentyy nopeammin yksiäänisessä kuin moniäänisessä kerronnassa. Ongelmat näkyvät muun muassa viittaussuhteiden epäselvyyksinä.

Moniäänisyyden kehityksellistä analyysiä on kirjoitetussa kielessä tehty pääasiassa kuvaelisitaatioiden avulla. Muu tutkimus on yllättävän vähäistä eikä luovasta kirjoittamisesta löydy pitkää kehityskaarta paljastavaa tutkimusta. Moniäänisyys näyttää kuitenkin olevan ilmiö, jonka ymmärtäminen ja tuottaminen vaihtelee herkästi aineiston laadun ja tutkimusasetelman mukaan. Kuvaelisitaation avulla tuotettu kirjoitetun tekstin moniäänisyys näyttää toteutuvan kehityksellisesti vaihdellen eikä muutos osoittaudu iän mukaan lineaariseksi. Esimerkiksi sammakkotarinan elisitaatiossa pienet lapset ja toisaalta aikuiset käyttävät paljon suoraa esitystä ja viisitoistavuotiaat taas epäsuoraa (Nordqvist 2001). Näin voisi tehdä sen johtopäätöksen, että moniäänisyys kehittyy U-käyrän tapaisesti. Tilanne ei kuitenkaan vastaa Unelmakirjoitelmien suoran ja epäsuoran kielentämisen suhdetta. Unelmakirjoitelmissa suora esitys on alaluokilla yleisempää ja sen esiintyvyys laskee tasaisen lineaarisesti peruskoulun loppuun, ja aikuisilla epäsuora esitys on huomattavasti suoraa yleisempää. Samalla kertova minä antaa tilaa kokevalle. Koska moniäänisyyden analyysi on herkkä tutkimusmetodille, on hyvin mahdollista, että tämä tulos riippuu unelmateemasta: lapset kirjoittavat toiminnallisia Unelmakirjoitelmia, nuoret aikuiset kokemuksellisia.

Tässä tutkimuksessa aineisto on kerätty kirjoitustehtävän avulla samasta teemasta, ja kirjoitelmia on peruskoulun kaikilta vuosiluokilta sekä nuorilta aikuisilta. Analyysi keskittyi peruskoulun päättövaiheisiin eli kuudes- ja yhdeksäsluokkalaisiin, mutta jotta moniäänisyyttä voitiin suhteellisen yhtenäisessä aineistossa tutkia kehityksellisen kaaren mukaan, alustava kehityskartoitus tehtiin alakoulun luokkien 1.–5. aineistosta. Aineisto analysoitiin sen mukaan, miten moniäänisyyden kielentäminen alkaa ja miten se jatkuu. Kirjoitelmat arvioitiin ensin laskeamalla, monessako niistä ylipäätään ilmenee moniäänisyyttä, ja sitten sen mukaan, minkätyyppistä moniäänisyyden lajia niissä kielennettiin.

Kaikissa alaluokkalaisten kirjoitelmissa esiintyy intertekstuaalisuutta. Se edustuu alakoulun alussa nuorimpien kirjoittajien kirjoitelmissa lähinnä pelien ja kirjojen sekä muiden populaarikulttuurin tuotteiden niminä. Voisi myös kuvitella, että unelmana hyvin tyypillinen lottovoitto ja sillä hankittavat auto, talo ja maailmanympärysmatka kaiuttavat 8-vuotiailla ennemmin kulttuurista kuin omaa unelmaa. Kuudesluokkalaisilla



intertekstuaalisuus on jo muualta saatua tietoa, liittyy lapsen omaan todellisuuteen ja lavenee iän myötä toisen todellisuuden luomiseen. Kyse voi olla esimerkiksi unelmapeleistä tai suosikkiyhtyeistä, joiden avulla rakennetaan kohtaamisia, joissa kirjoittajalla on epätavallinen ja todellisuudessa epätodennäköinen rooli. Kielellisesti toista todellisuutta rakennetaan usein konditionaalimuodolla, ja ongelma voi tulla, että kirjoittaja ei osaa niitä kielellisiä ajan ilmaisun keinoja, joilla hän siirtyisi reaali maailman ja kuvitellun maailman välillä. Ongelma on yleinen vielä yhdeksäsluokkalaisten ja erityisesti pojilla. Tempus- ja modusongelmat näkyvät moniäänisessä tekstissä myös nykyhetken ja menneen ajan ilmaisuissa.

Metakieltä esiintyy sekä aloituksissa että lopetuksissa iän mukaan lisääntyvästi, mutta lukijan suoraa – hallittua – puhuttelua esiintyi vain vanhemmilla kirjoittajilla. Aineistoni kuudes- ja yhdeksäsluokkalaisten teksteistä löytyy kertomuksen metatasoon kuuluvaa vuorovaikutukseen kategorisoitavaa moniäänisyyttä, oman tekstin tai tekstin tuottamisen asettamista kommentoinnin kohteeksi ja tietoisien vuorovaikutuksen luomista lukijaan – tai ainakin sellaisen tavoittelua. Metakielisyyttä näkyy lukijalle esittäytymisessä, tekstin kulun kommentoimisessa ja vuoropuhelussa. Kun kuudesluokkalaisten metakielelliset keinot ovat yksinkertaiset ja suorat, yhdeksäsluokkalaisten suoriutuu samasta kokonaisvaltaisemmin jaksottamalla esimerkiksi koko kirjoitelmansa meta-kielisiin raameihin. Kuudesluokkalaisten toistaa kirjoitelmansa aloituksessa tyypillisesti tehtävänannon, yhdeksäsluokkalaisten sen sijaan pohtii ja arvioi sanojen merkityksiä, esimerkiksi unelman käsitettä. Hän käyttää myös toistoa tai muita kohosteisia keinoja, joilla voi painottaa tekstiä kokonaisvaltaisemmin.

Suoran esityksen kaksi keskeistä funktiota Unelmakirjoitelmissa ovat juonen kuljetus ja puhujien luonnehdinta. Juonenkuljetusfunktio on hallitsevana ja toteutuu usein raskaana dialogimuotona. Puhujien luonnehdinnassa on sen sijaan käytössä moninaisempia ja osuvampia keinoja. Esimerkiksi eräs kuudesluokkalaisten poika sijoittaa itsensä näppäristi ja ironisestikin omalle murrealueelleen parilla sanalla: *Häh! Mitä nyt? Kui näi aikasi?* (6M19). Kehyslauseita ei joka yhteydessä käytetä mutta muuten illokutiiviset sävyt monipuolistuvat. Muodollisesti refe-

rointi on 2.–5.-luokkalaisilla ikään kuin referointijatkumon asteikolla, eli deiktisiä muutoksia ei tehdä systemaattisesti ja sekä persoona-, tempus- että spatiaalinen deiksis huojuvat (ks. myös Vesala 2012). Referoinnin merkintäkonventiot ovat alakoululaisilla useimmiten tuntemattomia. Yhdeksäluokkalaisilla on suoraa esitystä niin vähän, että välimerkkien käyttö ei erotu. Alakoululaisten aineistossa lähinnä persoonadeiksis soveltui suoran ja epäsuoran esityksen erottelukriteeriksi ja osittain piti tukeutua laajempaan kontekstiin; kuudesluokkalaiset jakautuvat tässä kahteen eritasoiseen ryhmään.

Ensimmäisen luokan aineistosta ei löydy lainkaan referointia. Sekä suora että epäsuora esitys lisääntyvät toiselta luokalta kuudennelle, mutta jo neljännellä luokalla alkaa epäsuoran esityksen suhteellinen määrä nousta. Viidennen ja kuudennen luokan pojat käyttävät epäsuoraa esitystä suoraa enemmän. Yhdeksäluokkalaiset käyttävät suoraa esitystä erittäin vähän; tämä on vertailututkimusten mukaista. Esimerkiksi Nordqvistin (2001) aineistossa juuri viisitoistavuotiaat käyttävät epäsuoraa esitystä eniten koko eri-ikäisten koehenkilöiden joukosta. Koululaiskirjoittaja ottaa epäsuoralla esityksellä etäisyyttä referoitavaan tapahtumaan kerrontahetken perspektiivistä. Epäsuoran esityksen funktio on vielä kuudesluokkalaisilla edistää juonta, mutta yhdeksäluokkalaisilla funktio on enenevästi episteeminen ja epäsuoralla esityksellä kuvataan enemmän kertojan tai puhujan ajatuksia kuin tapahtumia sinänsä. Muodollisesti epäsuoran esityksen merkintätavat horjuvat samoin kuin suoran esityksen. Kerronta saa myös vapaan epäsuoran esityksen piirteitä, muun muassa koska kehystäminen usein puuttuu. Deiktiset elementit voivat silloin olla kerrontatilanteen mukaiset mutta sanavalinnat ovat aikaisemman puhetilanteen ja alkuperäisen puhujan valintoja. Vapaa epäsuora esitys on ominaista yhdeksäluokkalaisille kirjoittajille ja on osaksi tietoista tajunnanvirtaa, osaksi vaikuttaa sattumanvaraiselta.

Puhutussa kielessä jo kaksivuotiaat lapset käyttävät moniäänistä ilmaisuja. Puhutussa kielessä moniäänisyyttä voidaan esittää esimerkiksi multimodaalisilla keinoilla, mutta kirjoitetussa kielessä tilanne on hankalampi. Yleistä näyttää olevan, että erilaiset viittaussuhteet ja deiktiset muunnokset, referoinnin eri funktiot ja kielellisen esittämisen tavat sekä hallittu lukijaan viittaaminen tulevat hallintaan samalla, kun kerto-

van tekstin kertomusrakenne tai -skeema aletaan hallita. Kehitys on pitkä ja oppiminen hidasta, ja kaikki yläkoululaiseltaan eivät näytä pääsevän tähän vaiheeseen. Serratrice, Hesketh ja Ashworth (2015) osoittavat priming- eli toistoasetelmalla, että moniäänisyyden ilmaisun keinojen oppimista auttaa aikuisen antama syötös. Jos kouluikäisille lapsille esimerkiksi luetaan moniäänisiä tekstejä, vaikutus heijastuu moniäänisyyden esittämiseen myös kirjoitetussa kielessä esimerkiksi siten, että johtolauseissa käytetyn sanaston diversiteetti kasvaa ja rakenteet monipuolistuvat. Myöhemmin toteutettu seurantatutkimus myös osoitti, että vaikutus on pitkäaikainen. Toisin sanoen kyky ilmaista moniäänisyyttä hyötyy mallista ja opettamisesta.

Unelmakirjoitelmissa huomio kiinnittyy kirjoitelmien suuriin eroihin. Toisessa päässä ovat kirjoittajat, joiden kertomuksen genreen kuuluva teksti on yksiäänistä kuudennella vuosiluokalla ja vielä peruskoulun päättövaiheessa. Vuorovaikutuksen taidot puuttuvat, eikä referoitua puhetta tai ajattelua osata erottaa muusta tekstistä. Toisessa päässä ovat kirjoittajat, jotka osaavat lähestyä lukijaa tarkoituksenmukaisesti ja luoda kokemuksiinsa perustuvan kuvan omasta sisäisestä ajatusmaailmastaan. Kielenhallinta ja kirjoitustaidot luonnollisesti vaihtelevat, ja jokaisen ei ehkä tarvitse osata kirjoittaa samalla tavalla tai tasolla. Voi kuitenkin kysyä, vaihteleeko osaaminen odotettua enemmän ottaen huomioon, että opetuksen pitäisi vähentää taitoeroja. Voi myös kysyä, edistäisikö oppimista nykyistä systemaattisempi kielioopin ja semanttis-pragmattisten ilmiöiden kielentämisen opettaminen (vrt. esimerkiksi Seppänen 2015; Rättyä 2017).

## LÄHTEET

- Astington, J. W. & Baird, J. A. (toim.) 2005. *Why language matters for theory of mind*. Oxford: Oxford University Press.
- Bahtin, M. 1991 [1963]. *Dostojevskin poetiikan ongelmia*. Kääntäjät: P. Nieminen & T. Laine. [Alkuteos *Problemy poetiki Dostojevskogo*.] Helsinki: Orient Express.
- Barthes, R. 1979 [1971]. From work to text. Julkaisussa: J. V. Harari (toim.) *Textual strategies: Perspectives in post-structuralist criticism*. Ithaca, New York: Cornell University Press, 73–81.
- Berg, M. 2011. Referointi, erittelymuotti ja tekstitaidon koe. *Virittäjä* 115, 317–348.
- Berman, R. A. & Slobin, D. 1994. Narrative structure. Julkaisussa: R. A. Berman &

- D. Slobin (toim.) *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale: Lawrence Erlbaum Ass., 39–84.
- Berman, R. A. 1996. Form and function in developing narrative abilities. *Julkaisussa: D. I. Slobin, J. Gerhardt, A. Kyratzis & J. Guo (toim.) Social interaction, social context and language: Essays in honor of Susan Ervin-Tripp*. Hillsdale: Lawrence Erlbaum Ass., 343–367.
- Berman, R. A. 2004. Between emergence and mastery: The long developmental route of language acquisition. *Julkaisussa: R. A. Berman (toim.) Language development across childhood and adolescence*. Amsterdam: John Benjamins, 9–34.
- Berman, R. A. 2007. Developing linguistic knowledge and language use across adolescence. *Julkaisussa: E. Hoff & M. Shatz (toim.) Blackwell handbook of language development*. Malden: Blackwell Publishing, 347–367.
- Blum-Kulka, S. 2004. The role of peer interaction in later pragmatic development: The case of speech representation. *Julkaisussa: R. A. Berman (toim.) Language development across childhood and adolescence*. Amsterdam: John Benjamins, 191–210.
- Bohnenmeyer, J., Donelson, K. T., Tucker, R. E., Benedicto, E., Capistrán Garza, A., Eggleston, A., Hernández Green, N., de Jesús Selene Hernández Gómez, M., Herrera Castro, S., O'Meara, C. K., Palancar, E., Pérez Báez, G., Polian, G. & Romero Méndez, R. 2014. The cultural transmission of spatial cognition: Evidence from a large-scale study. *Proceedings of the 36th annual meeting of the Cognitive Science Society*, 212–217.
- Cohn, D. 2006. *Fiktio mieli*. Kääntäjät: P. Korhonen, M. Lehtimäki, K. Mikkonen & S. Palomäki. [Alkuteos *The Distinction of Fiction* 1999.] Helsinki: Gaudeamus.
- Doherty, M. J. 2009. *Theory of mind: How children understand others' thoughts and feelings*. Hove: Psychology Press.
- Eerland, A. & Swaan, R. A. 2018. The influence of direct and indirect speech on source memory. *Psychology* 4: 1, 5 (1–12).
- Engberg-Pedersen, E. & Vang Christensen, R. 2017. Mental states and activities in Danish narratives: Children with autism and children with language impairment. *Journal of Child Language* 44, 1192–1217.
- Evans, N. 2012. Some problems in the typology of quotation: A canonical approach. *Julkaisussa: D. Brown, M. Chumakina & G. G. Corbett (toim.) Canonical morphology and syntax*. Oxford: Oxford University Press, 66–98.
- Fairclough, N. 1992: *Discourse and social change*. Cambridge: Polity Press.
- Fludernik, M. 1996. *Towards a 'natural' narratology*. London: Routledge.
- Good, J. S. 2015. Reported and enacted actions: Moving beyond reported speech and related concepts. *Discourse Studies* 17: 6, 663–681.
- Goodell, E. W. & Sachs, J. 1992: Direct and indirect speech in English-speaking children's retold narratives. *Discourse Processes* 15, 395–422.
- Hakulinen, A., Vilkkuna, M., Korhonen, R., Koivisto, V., Heinonen, T.-R. & Alho, I. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Hallila, M. 2006. *Metafiktio käsité. Teoreettinen, kontekstuaalinen ja historiallinen tutkimus*. Joensuu: Joensuun yliopisto.
- Harris Wright, H., Capilouto, G. J., Srinivasan, C. & Fergadiotis, G. 2011. Story processing ability in cognitively healthy younger and older adults. *Journal of Speech, Language, and Hearing Research* 54, 900–917.

- Heikkinen, V. & Voutilainen, E. 2012. Genre – monitieteinen näkökulma. Julkaisussa: V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiililä & M. Lounela (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 17–47.
- Hendriks, P. 2014. *Asymmetries between language production and comprehension*. Dordrecht: Springer.
- Herlin, I. 2005. Metakielen funktioita *x sanoen* -konstruktion valossa. Referointi ja moniäänisyys kielenkäytön ilmiöinä. Julkaisussa: M. Haakana ja J. Kalliokoski (toim.) *Referointi ja moniäänisyys*. Helsinki: Suomalaisen Kirjallisuuden Seura, 360–405.
- Hickmann, M. 1993. The boundaries of reported speech in discourse: Some developmental aspects. Julkaisussa: J. Ludy (toim.) *Reflexive language: Reported speech and metapragmatics*. Cambridge: Cambridge University Press, 63–90.
- Hosiaisuus, Yrjö. 2003. *Tieteen termipankki. Kirjallisuuden sanakirja*. Helsinki: WSOY.
- Hyvärinen, M. 2012. Kertomuksen sosiaaliset lajit. Julkaisussa: V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiililä & M. Lounela (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 392–410.
- Ikola, O. 1961. *Das Referat in der finnischen Sprache. Syntaktisch-stilistische Untersuchungen*. Helsinki: Suomalainen tiedeakatemia.
- Juvonen, R. 2014. *Kirjoitelma ja tekijän ääni. Kehystämisen yhdyslauseet suomenkielisen ylioppilasaineen dialogisuuden hallinnassa*. Helsinki: Helsingin yliopisto.
- Kallio, J. 1978. Kolmannen persoonan dilemmat. *Sananjalka* 20, 52–74.
- Kalliokoski, J. 2005a. Referointi ja moniäänisyys kielenkäytön ilmiöinä. Julkaisussa: M. Haakana & J. Kalliokoski (toim.) *Referointi ja moniäänisyys*. Helsinki: Suomalaisen Kirjallisuuden Seura, 9–43.
- Kalliokoski, J. 2005b. Moniäänisyys ja koherenssi suomea toisena kielenä kirjoittavien teksteissä. Julkaisussa: M. Haakana & J. Kalliokoski (toim.) *Referointi ja moniäänisyys*. Helsinki: Suomalaisen Kirjallisuuden Seura, 224–257.
- Kallionpää, O. 2017. *Uuden kirjoittamisen opetus: Osallistavaa luovuutta verkossa*. Jyväskylä: Jyväskylän yliopisto.
- Kauppinen, A. 2012. Lapsenkielen tutkimus. Julkaisussa: V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiililä & M. Lounela (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 587–593.
- Komppa, J. 2012. *Retoriset rakenteet. Retorisen rakenteen teoria suomi toisena kielenä -ylioppilaskokeen kirjoitelman kokonaisrakenteen ja kappalejaon tarkastelussa*. Helsinki: Helsingin yliopisto.
- Kuiri, K. 1984. *Referointi Kainuun ja Pohjois-Karjalan murteissa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Köder, F. 2016. *Between direct and indirect speech. The acquisition of pronouns in reported speech*. Groningen: University of Groningen.
- Köder, F. & Maier, E. 2016. Children mix direct and indirect speech: Evidence from pronoun comprehension. *Journal of Child Language* 43, 843–866.
- Köder, F. & Maier, E. 2018. The advantage of story-telling: Children's interpretation of reported speech in narratives. *Journal of Child Language* 45, 541–557.
- Labov W. 1972. *Language in the inner city: Studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.

- Labov, W. 1982. Speech actions and reactions in personal narrative. Julkaisussa D. Tannen (toim.) *Analyzing discourse: Text and talk*. Washington, D. C.: Georgetown University Press, 219–247.
- Labov, W. & Waletzky, J. 1967. Narrative analysis: Oral versions of personal experiences. Julkaisussa: J. Helm (toim.) *Essays on the verbal and the visual arts*. Seattle: American Ethnological Society, 12–44.
- Leech, G. N. & Short, M. 1981. *Style in fiction. A linguistic introduction to English fictional prose*. London: Longman.
- Makkonen-Craig, H. 2011. Kirjoittajan kompetenssit ja äidinkielellä kirjoittaminen. Julkaisussa: A. Kauppinen, H. Lehti-Eklund, H. Makkonen-Craig & R. Juvonen (toim.) *Lukiolaisen äidinkieli. Suomen ja ruotsinkielisten lukioiden opiskelijoiden tekstimaisemat ja kirjoitustaitojen arviointi*. Helsinki: Suomalaisen Kirjallisuuden Seura, 62–91.
- Mar, R. A. 2004. The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia* 42, 1414–1434.
- Martin, J. R. & White, P. R. R. 2005. *The language of evaluation: Appraisal in English*. Hampshire: Palgrave Macmillan.
- Mayer, M. 1969. *Frog, where are you?* New York: Dial Press.
- Muhonen, M. 2019. *Kompleksiset lauserakenteet 6.-luokkalaisten Unelmien päivä -kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampere: Tampereen yliopisto.
- Mäkinen, L., Loukusa, S., Laukkanen, P., Leinonen, E. & Kunnari, S. 2014. Linguistic and pragmatic aspects of narration in Finnish typically developing children and children with specific language impairment. *Clinical Linguistics & Phonetics* 28: 6, 413–427.
- Niinimäki, A. 2015. *Moniääninen Moreeni. Referointi ja moniäänisyys Lauri Viidan Moreenissa*. Tampere: Tampereen yliopisto.
- Nippold, M. A. 2006. *Later language development: School-age children, adolescents, and young adults*. Texas: Pro-Ed.
- Nordqvist, Å. 2001. *Speech about speech. A developmental study on form and functions of direct and indirect speech*. Göteborg: Göteborg University.
- Pajunen, A. 2012. Kirjoittamistaitojen kehitys 8–12-vuotiailla. Alakoululaisten unelmakirjoitelmat. *Virtittäjä* 116: 4–32.
- Penttilä, A. 1957. *Suomen kielioppi*. Porvoo: WSOY.
- Rättyä, K. 2017. *Kielitiedon didaktiikkaa: Kielentäminen ja visualisointi sanaluokkien ja lauseenjäsenten opetusmenetelminä*. Helsinki: Helsingin yliopisto.
- Seppänen, L. 2015. *Vitosen kirjoittajat. Diskurssikompetenssit yhdeksäluokkalaisten heikoissa kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Helsinki: Helsingin yliopisto.
- Serratrice, L., Hesketh, A. & Ashworth, R. 2015. The use of reported speech in children's narratives. A priming study. *First Language* 35, 1: 68–87.
- Shore, S. 2005. Referoinnista projektioon ja metarepresentaatioon. Julkaisussa: M. Haakana & J. Kalliokoski (toim.) *Referointi ja moniäänisyys*. Helsinki: Suomalaisen Kirjallisuuden Seura, 44–82.
- Sintonen, M. 2011. *Suunnittelemisen vaikutus kertomuksen rakenteeseen viides- ja seitsemäsluokkalaisilla kirjoittajilla*. Suomen kielen pro gradu -tutkielma. Tampere: Tampereen yliopisto.

- Solin, A. 2006. Genre ja intertekstuaalisuus. Julkaisussa: A. Mäntynen, S. Shore & A. Solin (toim.) *Genre – tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura, 72–95.
- Stanzel, F. K. 1985 [1979]. *Theorie des Erzählens*. 3. durchgesehene Auflage. Göttingen: Vandenhoeck und Ruprecht in Göttingen.
- Strömqvist, S., Nordqvist, Å. & Wengelin, Å. 2004. Writing the frog story. Developmental and cross-modal perspectives. Julkaisussa: S. Strömqvist & L. Verhoeven (toim.) *Relating events in narrative. Volume 2. Typological and contextual perspectives*. Mahwah: Lawrence Erlbaum Ass., 359–394.
- Tiiri, M. 2021. *Lauserakenteiden kompleksisuus ammattikorkeakoululaisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Tolchinsky, L. 2004. The nature and scope of later language development. Julkaisussa: R. A. Berman (toim.) *Language development across childhood and adolescence*. Amsterdam: John Benjamins, 233–247.
- Vesala, S. 2012. *Kerronnan keinot alakoululaisten teksteissä*. Suomen kielen pro gradu -tutkielma. Tampere: Tampereen yliopisto.
- Viita, L. 1973 [1950]. *Moreeni*. Seitsemäs painos. Porvoo: WSOY.
- Vincent, D. & Perrin, L. 1999. On the narrative vs. non-narrative function of reported speech: A socio-pragmatic study. *Journal of Sociolinguistics* 3: 3, 291–313.
- Virtanen, M. T. & Juvonen, R. 2018. Kirjoitetun vuorovaikutuksen moninaisuus ja erityisyys. Alisteisen interrogatiivirakenteen sisältävät rakenteet blogivuorovaikutuksessa. Julkaisussa: S. Shore, T. Rahtu & M. T. Virtanen (toim.) *Kirjoitettu vuorovaikutus*. Helsinki: Suomalaisen Kirjallisuuden Seura, 179–213.
- Virtanen, M. T., Rahtu, T. & Shore, S. 2018. Moniäänisyys, vuorottelu ja kerronta. Julkaisussa: S. Shore, T. Rahtu & M. T. Virtanen (toim.) *Kirjoitettu vuorovaikutus*. Helsinki: Suomalaisen Kirjallisuuden Seura, 9–38.

## II

# Muodot ja rakenteet

Myöhemmän kielenkehityksen vaiheissa I osion luvut osoittivat, että sanaston diversiteetti kasvaa voimakkaasti ja semantiikan hallinta monipuolistuu. Kehitys näkyy niin sanojen määrän kasvuna, niiden konkreettis-jatkumon monipuolistumisena kuin erilaisten kompleksisten semanttisten ilmiöiden kielentämisen taitojen lisääntymisenä, on sitten kyse spatiaalisten suhteiden ilmaisuista, tekstin semanttisen koherenssin luomisesta tai moniäänisyyden ilmaisuista. Aikuistasoa ei tosin vielä peruskoulun aikana saavuteta.

Tämän osion luvuissa näkökulma on puhutun ja kirjoitetun kielen kehityksellisessä suhteessa ja kirjoitetun kielen rakenteessa. Lasten varhainen kielenkehitys on yksinomaan puhutun kielen haltuunottoa ja myöhemmän kielenkehityksen varhaisvaihe nojaa sekin vahvasti puhutun kielen malliin. Laasanen ja Pajunen keskittyvät pohtimaan puhutun ja kirjoitetun kielen keskinäistä suhdetta yleensä sekä muutoksen etene- mistä puhutusta kielestä kirjoitettuun kieleen, Honko tarkastelee muun muassa äänne–kirjain-vastaavuuden oppimista sen mukaan, miten se näkyy kirjoitetun kielen norminmukaisuuden kasvuna. Pajunen ja Vainio tarkastelevat ensin, miten puhutun kielen yksinkertainen rakenne kirjoitetussa kielessä iän mukaan monimuotoistuu eli kompleksis- tuu, ja seuraavaksi he pyrkivät selvittämään, mistä aineksista kielen



kompleksisuus oikeastaan syntyy. II osion joka luvussa kehitystä tarkastellaan aineistona kielen muotoyksiköt. Kun I osiossa analyysin lähtökohtana olivat merkitykset ja niillä ilmaistavien ilmiöiden analyysi, II osiossa on edetty poimimalla analysoitavaksi kielellisiä rakenteita. Ero heijastuu kielenhallinnan kehitysnopeuteen ja toisaalta tulosten kielitypologiseen vertailtavuuteen: muotojen tasolla kehitys on periaatteessa nopeampaa mutta vertailtavuus muihin kieliin heikompa. Ero näkyy myös ikäkausimediaanien esittämisessä: ilmiötasolla variaatio kasvaa oppimisen myötä, rakennetasolla voidaan helpommin esittää ikäkausittaisia yleistyksiä ainakin kertovasta tekstistä.

II osion ensimmäinen luku tarjoaa analyyseille taustan pohtimalla ja määrittelemällä puhutun ja kirjoitetun kielen suhdetta käsitteellisellä tasolla. Luvussa suhde nähdään siten resiprookkiseksi, että puhutulla ja kirjoitetulla kielellä on yhteinen ydin ja vaikutus voi kulkea molempiin suuntiin. Kyse ei siten ole täydestä erilaisuudesta mutta ei myöskään täydestä samanlaisuudesta. Suhdetta tarkastellaan yhtäältä teoreettisella tasolla ja toisaalta empiirisellä tasolla tutkimalla suomenkielisen aineiston avulla, miten sosiolingvistisin metodein todennettu puhutussa kielessä tapahtunut muutos näkyy kirjoitetussa kielessä. Jos nyt puhuttu ja kirjoitettu kieli edustavat samaa kielellistä systeemiä (tai ydintä) ja puhuttu kieli saadaan haltuun ensin, voisi olettaa, että kielenmuutoksen kontekstissa puhutussa kielessä jo toteutuneet tai meneillään olevat muutokset heijastuvat kirjoitettuun kieleen, vaikkakin viipeellä. Puhutun kielen vaikutus kirjoitettuun ei kuitenkaan ole näin suoraviivainen. Sama kaksisuuntainen vuorovaikutus ja toisaalta erilaistuminen näkyy kaikissa osion luvuissa. Tuloksista voi päätellä, että kirjoitettuun kielimuotoon kasvaminen ei käy itsestään eikä yksinomaan puhutun kielen perusteella. Tiivistäen on sanottava, että kirjoitettuun kielimuotoon kasvaminen edellyttää systemaattista opettelua ja opettamista.

# Meneillään oleva kielenmuutos puhutusta kielestä kirjoitettuun ja suomen possessiivisuffiksi

*Mikko Laasanen*

 <https://orcid.org/0000-0003-2654-7146>

*Anneli Pajunen*

 <https://orcid.org/0000-0001-8597-7491>

## Johdanto

Puhuttu ja kirjoitettu kieli ovat kielen eri modaliteetteja eli ilmenemismuotoja. Puhuttu kieli on akustinen ilmiö, jonka havaitseminen perustuu useimmiten kuuloaistiin. Kirjoitettu kieli on puolestaan fyysinen ilmiö, jonka havaitseminen perustuu useimmiten näköaistiin. Tästä ontologisesta erosta seuraa modaliteettien ominaisuuksiin liittyviä eroja, joista puolestaan voi seurata funktionaalisia eroja. Esimerkiksi kirjoitettu kieli on yleensä pysyvämpää kuin puhuttu kieli, joten kirjoitettua kieltä käytetään enemmän tiedon tallentamiseen. Eroista huolimatta puhuttu ja kirjoitettu kieli ovat kuitenkin saman ilmiön kaksi ilmentymää, minkä todistaa se, että samat kuvausmenetelmät sopivat kumpaankin (ks. Itkonen 2010, 117). Puhutun ja kirjoitetun kielen perimmäinen samuus on todennäköisesti helpommin havaittavissa kielen ylemmillä kuin alemmilla tasoilla. Esimerkiksi normaalipainotteinen suora sanajärjestys on suomessa puhutussa ja kirjoitetussa kielessä molemmissa tyyppiä *subjekti*

– *verbi – objekti*. Täten voi sanoa, että puhutun ja kirjoitetun kielen *peruskielioppi* on sama (ks. Hakulinen 2003, 5). Yksityiskohtaisemmassa tarkastelussa puhutun ja kirjoitetun kielen välillä voi havaita eroja, joiden perusteella voi puhua puhutun ja kirjoitetun kielen malleista. Lapsen ensimmäinen opittu modaliteetti on puhuttu kieli ja kirjoitettu kieli opitaan yleensä hieman myöhemmässä vaiheessa. Tästä seuraa se, että kielenoppiminen jatkuu vielä alkuvaiheen jälkeenkin (*myöhempi kielenkehitys*, ks. tarkemmin Pajusen ja Hongon esittelyä johdantoluvussa).

Puhutun kielen tutkimuksissa on havaittu, että suomen possessiivisuffiksin käyttö on vähentynyt (ks. esim. Kangasmaa-Minn 1986; Ikola 1992; Paunonen 1995; Korhonen 2006). Toisinaan possessiivisuffiksin käytön vähenemisen arvellaan olevan leviämässä myös kirjoitettuun kieleen (ks. esim. Ikola 1992; Paunonen 1995; Varjonen 1996; Niemi 2018). Tällöin oletuksena on siis se, että muutos alkaa puhutussa kielessä ja etenee jollain aikavälillä kirjoitettuun kieleen uuden puhutun kielen mallin vallatessa alaa aiemmalta kirjoitetulta mallilta. Kirjoitetun kielen osalta tilannetta ei ole kuitenkaan tutkittu, vaan havainnot perustuvat yksittäisiin esimerkkeihin. Koska puhekielen tutkimuksissa liitteettömyys on vahvistunut varsin nopeasti, voivat yksittäiset esimerkit antaa sen kuvan, että muutos on todellakin leviämässä kirjoitettuun kieleen. Kyse on kuitenkin subjektiivisista arvioista, ei tutkitusta tiedosta.

Sosiolingvistinen meneillään olevan kielenmuutoksen tutkimus on ensisijaisesti puhutun kielen tutkimusta, jossa nuorten ryhmä on usein avainasemassa muutoksen nopeuden ja suunnan tunnistamisen suhteen, sillä nuoret puhujat ovat usein muutoksen kärjessä (ks. esim. Labov 2001; Aitchison 2003, 739; Tagliamonte & D’Arcy 2009). Tästä huolimatta kehityksellisiä tekijöitä ei ole näissä tutkimuksissa juurikaan huomioitu (ks. Laasanen, Pajunen & Häikiö, arvioitavana). On täysin mahdollista, että nuorten ryhmään kuuluvat puhujat eivät (vielä) hallitse kirjoitetun kielen mukaista possessiivisuffiksin käyttöä, mikä saattaa vaikuttaa tutkimustuloksiin esimerkiksi hidastamalla muutoksen nopeutta. Toisaalta on myös mahdollista, että kirjoitetussa kielessä havaittu variaatio suffiksin käytön suhteen ei ole merkki muutoksesta vaan ainoastaan pidempiaikaisesta vaihtelusta ja kirjoittamisen muutoksista (ks. Weinreich, Labov & Herzog 1968, 188).

Tämän luvun tarkoituksena on selvittää, näkyykö kirjoitetussa kielessä merkkejä possessiivisuffiksin käytön vähenemisestä. Luku jakaantuu kahteen osaan. Ensimmäisessä osassa esitellään sosiolingvististä meneillään olevan kielenmuutoksen tutkimusta ja argumentoidaan puhutun ja kirjoitetun kielen perimmäisen samanlaisuuden puolesta. Tutkimuksen toisessa osassa esitellään ensin suomen possessiivisuffiksia ja sen tutkimusta puhutussa kielessä ja tämän jälkeen tarkastellaan possessiivisuffiksin käyttöä ja kehitystä kirjoitetussa kielessä aikavälillä 1940–2010 eriaikaisten korpusten avulla. Tämän tarkastelun vertailukohtana käytetään puhutussa kielessä tehtyjä havaintoja ja ennusteita suffiksin käytön nopeasta vähenemisestä. Tutkimuksen taustalla on oletus, että lapsen myöhempi kielenkehitys rakentuu aluksi puhutun kielen malliin yleiseen ja yksinkertaiseen perustuen mutta irtautuu ja laajenee siitä myöhemmin, kuten possessiivisuffiksien käytöllä pyrimme osoittamaan.

Kielimuotojen terminologiasta on kirjoitettu varsin paljon (ks. esimerkiksi Ikola 1972; 1986; Hiidenmaa 2005; Nuolijärvi & Sorjonen 2010, 15–18; Viinikka & Voutilainen 2013). Ensinnäkin tulee erottaa puhuttu ja kirjoitettu kieli omaksi kategoriakseen, kuten tekee myös Ikola (1972, 16). Siinä missä puhutun ja kirjoitetun kielen välinen ero on ontologinen modaaliteettiero, varsinaisten kielimuotojen, mm. kirjakieli, yleiskieli, puhekieli ja murre välinen ero ei ole. Tämä tarkoittaa sitä, että ainakin teoriassa jokainen näistä kielimuodoista voi ottaa puhutun tai kirjoitetun kielen määritteekseen eli voimme puhua esimerkiksi kirjoitetusta puhekielestä tai puhutusta kirjakielestä. Ongelmaksi saattaa kuitenkin muodostua käsite *kirjakieli* historiallisen painolastinsa vuoksi (so. alkujaan *kirjakieli* viittasi nimenomaan preskriptiiviseen kirjoitettuun kieleen). Esimerkiksi Ikola (1986, 5) pitää kirjakieltä ainoastaan yhtenä kirjoitetun kielen muotona. Preskription määrä on kuitenkin yksi tapa ymmärtää käsitteet: tietyssä mielessä käsitteet voidaan asettaa jatkumolle siten, että esimerkiksi kirjakieli tai kirjoitettu yleiskieli on jatkumon yhdessä päässä riippuen siitä, ymmärretäänkö kirjakieli yksinomaan kirjoitetun kielen muodoksi, yleiskielen synonyymiksi vai yleiskieltä vielä preskriptiivisemmäksi kielimuodoksi. Jatkumon toiseen päähän sijoittuvat murteet tai mahdollisesti erityyppiset aluepuhekielten muodot ja sosiaalisten ryhmien varieteetit. Tässä tutkimuksessa käytämme käsi-

tettä kirjakieli yhdessä kirjoitetun kielen kanssa siitä syystä, että monissa lähteissä lähes järjestään puhutaan kirjakielestä. Siten tuntuu turhalta korvata käsite kirjakieli esimerkiksi kirjoitetulla yleiskielellä, varsinkin kun määritelmistä ei vallitse yksimielisyyttä. Aivan vastaavasti käytämme käsitettä *murre* (pro *aluepuhekieli* tms.). Emme ota kantaa siihen, miten kielimuotojen käsitteet tulisi määritellä tai ymmärtää. Tärkeää on käsitteiden välinen suhde tai ero, joka on ymmärrettävissä, vaikka käsitteet itsessään eivät välttämättä olisi yksiselitteisiä.

## Meneillään olevan kielenmuutoksen tutkimus

Meneillään oleva kielenmuutos on monimutkainen tutkimuskohde, jossa yhdistyvät mm. kielentutkimuksen synkroninen ja diakroninen näkökulma sekä vakaa ja epävakaa kielellinen vaihtelu. Meneillään olevan kielenmuutoksen tutkimus on tyypillisimmillään ollut sosiolingvististä puhutun kielen tutkimusta. Historiallinen sosiolingvistiikka siirsi painopisteen takaisin historiaan ja alkoi tarkastella kirjoitettua kieltä. Meneillään olevan kielenmuutoksen tutkimus uudemmassa kirjoitetussa kielessä on kuitenkin jäänyt vähemmälle huomiolle, samoin kuin meneillään olevan muutoksen etenemisen tutkimus puhutusta kielestä kirjoitettuun kieleen. Tutkimuksemme paikkaa puutetta analysoimalla possessiivisuffiksin käyttöä ja kehitystä kirjoitetussa kielessä eri vuosikymmeninä sekä selvittämällä, näkykö kirjoitetussa kielessä merkkejä oletetusta possessiivisuffiksin käytön vähenemisestä. Possessiivisuffiksin käyttöä on tarkasteltu aiemmin yllättävän vähän kirjoitetun kielen kontekstissa (ks. Varjonen 1996). Lisäksi muutamassa opinnäytetutkimuksessa tarkastellaan suffiksien käyttöä S1- ja S2 -oppijoiden kielessä (ks. esimerkiksi Tiainen 2008; Turunen 2012), mutta koululaisten possessiivisuffiksien käytöstä ei ole tehty pitkäaikaistutkimusta, vaikka omistus- ja viitesuhteiden ilmaisukeinot ansaitsisivat tulla tutkituiksi.

Meneillään olevan kielenmuutoksen tutkimus kirjoitetussa kielessä leikkaa kahden metateoreettisen dimension lävitse, joista toinen koskee tutkimuksen ajallista etäisyyttä muutokseen ja toinen kielen ilmenemismuotoja. Näistä ensimmäistä on käsitelty laajemmin muualla (ks. Laasa-

nen 2016), joten pääpaino tässä yhteydessä on kielen ilmenemismuotojen tarkastelussa.

Historiallinen kielitiede tarkastelee kielen kehittymistä pitkällä aikavälillä vertaamalla synkronisia kielentiloja toisiinsa. Tarkasteltavana ovat yleensä tällöin jo päättyneet muutokset, ja koska tarkastelu toteutetaan ajallisesti kaukaa, on kuva muutoksesta usein ainakin kohtuullisen selkeä. Historiallinen kielitiede perustuu vertailuun ja rekonstruktioon, joita tukee erilaisten kirjoitettujen dokumenttien tutkimus (ks. Campbell 2013). Leonard Bloomfield ennakoii meneillään olevan kielenmuutoksen tutkimusta varsin tarkkaan (ks. Bloomfield 1951 [1933], 347; Laasanen 2016), mutta varsinainen meneillään olevan kielenmuutoksen tutkimus sai alkunsa vasta muutama vuosikymmen myöhemmin Labovin 1960-luvun sosiolingvististen tutkimusten myötä (ks. Labov 1963; 1966). Keskeiseksi tutkimusmetodiksi nousi näennäisaikametodi, jonka mukaan ”– 50-vuotiaiden kieli hetkellä  $x$  heijastelee 70-vuotiaiden kieltä vuonna  $x + 20$  ja 30-vuotiaiden kieli 70-vuotiaiden kieltä hetkellä  $x + 40$ ” (Kurki 2005, 33). Myöhemmin näennäisaikaa hyödyntävien näennäisaikatutkimusten rinnalle nousi vielä reaaliaikainen tutkimus, jossa muutoksen etenemistä voidaan seurata reaaliajassa esimerkiksi uusimalla aiemmin tehty näennäisaikatutkimus. Sosiolingvistinen meneillään olevan kielenmuutoksen tutkimus on ensisijaisesti synkronista kielentutkimusta, sillä eri ikäryhmiin perustuva vertailu tapahtuu yhden kielentilan sisällä. Ajallinen etäisyys muutokseen on huomattavasti pienempi kuin historiallisessa kielitieteessä yleensä, minkä takia kuva muutoksesta voi ajoittain olla epäselvä. Tästä seuraa se, että toisinaan voi olla hyvinkin vaikeaa erottaa muutosta kielellisestä variaatiosta.

Hiukan yksinkertaistaen voi todeta, että siinä missä historiallisen kielitieteen aineistona toimii (pääasiallisesti) kirjoitettu kieli erilaisten kielenmuistomerkkien muodossa, on tilanne päinvastainen sosiolingvistisen meneillään olevan kielenmuutoksen tutkimuksen kohdalla, jossa aineistona toimii puhuttu kieli erilaisten nauhoitteiden muodossa. Näitä kahta tutkimustyyppiä täydentää historiallinen sosiolingvistiikka, joka sai alkunsa Romainen (1982) sovellettua sosiolingvistiikan tutkimusmetodeja historiallisen syntaksin tutkimukseen. Historiallinen sosiolingvistiikka on muutoksen ja variaation tutkimusta, jossa aineistona

toimii jälleen kirjoitettu kieli, esimerkiksi kirjeet ja päiväkirjat (ks. Nevalainen & Raumolin-Brunberg 2012). Muutoksen tutkimuksen näkökulmasta ainoa mikä tästä kuvasta puuttuu, on meneillään olevan kielenmuutoksen tutkimus kirjoitetussa kielessä.

Näennäisaikaisissa tutkimuksissa muutoksen suunta ja nopeus lue-taan vertaamalla vanhempien puhujien puhetta nuorempien puhujien puheeseen, mikä asettaa nuoret puhujat kriittiseen asemaan muutoksen suunnan ja nopeuden tunnistamisen kannalta. Perinteisesti näennäis-aikametodin suurimmaksi ongelmaksi on nostettu ikäsidonaisuus (*age-grading*), joka tarkoittaa tietylle iälle tyypillistä mutta ohimenevää vaihetta (ks. Chambers 1995, 188). Yleensä kyse on nuorille puhujille tyypillisestä ilmiöstä, esimerkiksi suomessa *x*-grafeemin käyttö *ks*-yh-tymän paikalla 1980-luvulla (ks. Kurki 2013, 210). On huomattava, että ikäsidonaisuus ei viittaa suoraan kehityksellisiin tekijöihin, vaikka tä-mäkin olisi ainakin periaatteessa mahdollista. U-käyrässä sosiaalistu-minen erilaisten (ikä)ryhmien tyypillisiin kielenkäyttökoventioihin voisi selittää esimerkiksi sitä, että vältellään hyperkorrektiksi tulkittavia muo-toja, jollainen possessiivisuffiksikin voi olla ainakin tietyissä medioissa ja tekstilajeissa tai tuttavallisiksi oletetuissa vuorovaikutussuhteissa tai tilanteissa. Kehityksellisten tekijöiden huomiotta jättäminen saattaa kuitenkin vähentää näennäisaikaisten tutkimusten luotettavuutta, sillä riippuen tutkittavista kielenpiirteistä, nuorten puhujien kielellinen käyt-täytyminen saattaa heijastaa myös kehityksellisiä tekijöitä yhteisöllisen muutoksen (ks. Labov 1994, 84) lisäksi tai sijaan. Reaaliaikainen tutki-mus on näennäisaikaista tutkimusta huomattavasti luotettavampi muu-toksen tunnistamisessa, sillä siinä muutoksen tunnistaminen perustuu useampaan (yleensä kahteen, mutta ks. kuitenkin Lappalainen, Mustan-oja & O’Dell 2019; reaaliaikaisen tutkimuksen luotettavuudesta ks. myös Kuparinen, Mustanoja, Peltonen, Santaharju & Leino 2019) syn-kroniseen tilannekuvaan. Reaaliaikaisen tutkimuksen myötä on havait-tu, että yksilöiden kieli voi muuttua vielä aikuisiälläkin, jolloin puhutaan elinikäisestä muutoksesta (*lifespan change*, ks. Sankoff 2006; 2018). Kä-sitteenä elinikäinen muutos liittyy myös kielenoppimiseen ja -hallintaan myöhemmällä iällä (ks. esimerkiksi Ramscar, Hendrix, Love & Baayen 2013).

## Puhutun ja kirjoitetun kielen samuus ja erilaisuus

Ontologisesta näkökulmasta katsottuna puhuttu kieli on ensisijainen, kun taas kirjoitettu kieli on suhteellisen uusi keksintö (ks. Bloomfield 1951 [1933], 282; Hockett 1958, 4). Esimerkiksi suomen kielen kirjoitetun kielen katsotaan alkaneen kehittyä 1500-luvulla, jota ennen kieli oli ollut olemassa ainoastaan puhuttuina murteina.<sup>61</sup> Kirjoitettu kieli käsitteenä ei myöskään ota kantaa siihen, minkä tyyppisestä kielestä on kyse. Täten kirjoitettu kieli voi olla yhtä hyvin yleiskieltä kuin aluemurretta, mistä hyvänä esimerkkinä ovat Aku Ankka -sarjakuvien murreversiot. Kirjoitetun puhekielen voidaan myös katsoa hälventävän puhutun ja kirjoitetun kielen rajaa (ks. Viinikka & Voutilainen 2013, 18).

Perinteisen näkemyksen mukaan kirjoitettu kieli on puhuttua kieltä kuvaava merkkien järjestelmä. Tämän Aristoteleeseen asti palautuvan näkemyksen edustajiin kuuluvat niin Saussure (1962 [1916]), Bloomfield (1951 [1933]) kuin Hockettkin (1958). Näkemyksen edustajat korostavat puhutun kielen olevan lingvistiikan ensisijainen tutkimuskohde ja sallivat kirjoitetulle kielelle lähinnä metodologisen aseman kielen säilyttäjänä ja historiallisen lingvistiikan mahdollistajana. Näin ajattelee esimerkiksi Sturtevant (1917, 10). Vastakkaisen näkemyksen mukaan kirjoitettu kieli on kuitenkin huomattavasti itsenäisempi symbolisysteemi kuin perinteisen näkemyksen edustajat ajattelevat. Näin taas kirjoitti jo Penttilä (1932, 17–19), jonka mukaan varhaisin kirjoitus eli kuvakirjoitus ei vastannut puhutun kielen aineksia käytännössä lainkaan. Penttilä ei kuitenkaan pitänyt puhuttua ja kirjoitettua kieltä täysin itsenäisinä järjestelminä vaan näki niiden välillä varsin pitkällekin menevää vastaavuutta. Huomattavasti radikaalimman ajatuksen esittää Linell (1982, 44), jonka mukaan puhuttua ja kirjoitettua kieltä ei välttämättä hallitse yksi ja sama kielijärjestelmä (Saussuren *langue*) vaan kaksi osittain päällekkäistä järjestelmää (*la langue de la parole* ja *la langue de l'écriture*; ks. myös Goody 1977, 76–77; Kravchenko 2011). Sekä perinteinen että Linellin esittämä näkemys vaativat korjaamista. Kirjoitettu kieli ei redu-

61 Heikkilän (2017) tutkimuksen mukaan suomenkielisiä kirjoitettuja sanoja löytyy kuitenkin jo tätä varhaisemmista teksteistä.



soidu puhutun kielen kuvaukseksi, mutta se ei ole myöskään itsenäinen järjestelmä. Coulmas (1989) kirjoittaa, että luonnollinen kieli on yksityiskohtaisin ja täsmällisin symbolijärjestelmä, joka ihmisillä on käytössään. Olisi teoreettisesti mahdollista, että puhutun kielen rinnalle voitaisiin kehittää uusi, puhutusta kielestä täysin irrallinen ja itsenäinen kirjoitusjärjestelmä, joka kuitenkin vastaisi monimutkaisuudeltaan ja ilmaisuvoimaltaan puhuttua kieltä. Tämä olisi kuitenkin harvinaisen epäekonomista, sillä se käytännössä tarkoittaisi kielen rakentamista uudelleen. (Mts. 27.)

Puhutun ja kirjoitetun kielen takana oleva yhteinen ydin mahdollistaa sen, että kieliä, joita ainoastaan puhutaan, voidaan kuvata saman käsitteistön avulla kuin kieliä, joita myös kirjoitetaan (Itkonen 2004, 268). Tämän osoittaa todeksi P. inin kielioppi (*Aṣṭādhyāyī*, noin 400–500 eKr), jota Bloomfield (1970 [1929], 219) kutsui yhdeksi ihmisällyn suurimmista monumenteista. *Aṣṭādhyāyī* koostuu hieman alle 4000<sup>62</sup> säännöstä ja kuvaa Pāṇinin ajan puhuttua sanskritia (ks. Agrawala 1963 [1953], 353–354; Cardona 1976, 238–239; Scharfe 1977, 106). Kyse on synkronisesta kieliopista ja kielijärjestelmän (*langue*) kuvauksesta – siis juuri siinä mielessä kuin Saussure (1962 [1916]) myöhemmin kirjoitti. Tämä osoittaa sen, että *langue* ei ole Saussuren tai häntä seuraavien strukturalistien keksintö vaan kielenkuvauksen perimmäinen kohde sen alkuajoista lähtien (ks. Robins 1990 [1967], 154–155; Kiparsky 1994, 2917; Itkonen 2005, 4). Pāṇinin kielioppia ei kirjoitettu talteen noin 200 vuoteen vaan sen olemassaolo perustui puhtaasti oraaliseen perinteeseen (ks. Agrawala 1963 [1953], 294). Tästä huolimatta länsimainen kielitiede on Kiparskyn (1994, 2918) mukaan ottanut vaikutteita Pāṇinin kieliopista alusta lähtien ja tämä jatkuu vieläkin. Esimerkiksi 1800-luvun historiallisten kielitieteilijöiden käyttämä morfologinen analyysi perustuu Pāṇinin kielioppiin (mp.; ks. myös Bloomfield 1970 [1929], 221; Scharfe 1977, 115).

Pāṇinin kielioppi osoittaa vääräksi sellaiset näkemykset, joiden mukaan kielentutkimus ei olisi mahdollista ilman kirjoitusjärjestelmää (ks. Olson 1996 [1994], 68, 258, 277) tai että kielen (Saussuren *langue*) erot-

62 Tarkka määrä vaihtelee sen mukaan, mitkä säännöistä lasketaan yksittäisiksi säännöiksi.

taminen puheesta (Saussuren *parole*) johtuisi kirjoitetun kielen vaikutuksesta (ks. Harris 1980, 4–6). Vaikka kirjoitetun kielen tutkimus olisi-kin vaikuttanut puhutun kielen tutkimuksen metodeihin, on kirjoitetun kielen tutkimus ottanut vaikutteita puhutun kielen tutkimuksesta jo ennen kuin puhutun kielen tutkimus on ottanut vaikutteita kirjoitetun kielen tutkimuksesta. Vaikutuksen resiprookkisuus on osoitus siitä, että kirjoitetun ja puhutun kielen tutkimuksen metodologialla on sama perusta. Tämä taas johtuu tutkimuskohteiden perimmäisestä samuudesta. Pāṇinin kielioppi osoittaa vääriksi myös sellaiset näkemykset, joiden mukaan jatkumoluonteisen puheen analysoiminen diskreeteiksi entiteeteiksi olisi kirjoitetun kielen vaikutusta (ks. Harris 1980, 8–11; Dufva 2000, 173–175), sillä kieliopin lisänä on kolme muuta tekstiä, joista *akṣarasamāmnāya* (tai *śiva-sūtras*) luettelee ja luokittelee puhutun kielen foneemit niiden ominaisuuksien mukaan (Cardona 1976, 209–210; Kiparsky 1994, 2922–2923). Jatkumoluonteisen puheen analysoiminen diskreeteiksi yksiköiksi ei siis vaadi kirjoitettua kieltä tai ole sen seuraus vaan osoitus ihmisen taipumuksesta kategorisoida (ks. Boas 1964 [1911], 121) – tässä tulee hyvin esiin kielen tutkimuksen ja kielenoppimisen välinen yhteys.

Abercrombie (1967, 1–3) tekee eron kielen (*language*) ja kielen ilmene- mismuodon (*medium*) välille. Kielen ilmenemismuoto ei Abercrombien näkemyksen mukaan ole kieltä vaan ainoastaan väline kielelle. Kielitiede on tiede, joka tutkii ensisijaisesti kieltä, mutta tähän on laskettava mukaan myös kielen ilmenemismuoto itsessään. Kuten Romaine (1982, 18) toteaa, kirjoitetun kielen tutkimus tuottaa tietoa ei ainoastaan kielestä itsestään vaan myös kirjoitetusta kielestä ilmenemismuotona. Puhutun ja kirjoitetun kielen takana on yhteinen ydin, mutta ne ovat riittävän erilaisia ollakseen itsenäisiä tutkimuskohteita, minkä todistaa jo kirjoitetun ja puhutun kielen välisen suhteen suhteellisen pitkä tutkimushistoria. Esimerkiksi Akinnason (1982) mukaan kirjoitetun ja puhutun kielen välistä suhdetta on tutkittu ainakin 1920-luvulta lähtien, alussa ensisijaisesti vertailemalla eroja sanamäärissä. Syntaksin ja semantiikan tutkimus kehittyi kuitenkin suurin piirtein samanaikaisesti sanamäärien tarkastelun kanssa, vaikka ajallinen pääpaino tutkimuksella lieneekin vuosisadan jälkimmäisellä puoliskolla (mts. 99–104.). Viimeistään

1980-luvulla tutkimus alkoi laajentua diskurssin (tai tekstien) tutkimukseen (ks. esimerkiksi Tannen toim. 1982).

Kokonaisvaltainen kielentutkimus kattaa sekä puhutun että kirjoitetun kielen tutkimuksen. Toisinaan näkee vielä väitettävän, että kielitiede on keskittynyt tutkimaan vain joko kirjoitettua tai puhuttua kieltä ja että se tästä syystä on rajoittunutta. Linell (1982, 31) väittää, että nykyaikainen kielitiede on keskittynyt tutkimaan ainoastaan kirjoitettua kieltä, kun taas Coulmas (1989, iix) esittää aivan päinvastaista ja puhuu kirjoituksen tutkimuksen laiminlyömisestä. Tällaisiin julistuksiin tulisi todennäköisesti suhtautua vähintäänkin varauksella, koska ne on aina tehty tietystä perspektiivistä tietyt intressit mielessä. Mikäli kielitieteen historiaa tarkastellaan kokonaisuutena, jakaantuu puhutun ja kirjoitetun kielen tutkimus siten, että toinen esiintyy lähes aina ensisijaisena tutkimuskohteena aikakaudesta ja tutkimussuunnasta riippuen. Näin Romaine (1982, 20) saattoi täysin perustellusti kirjoittaa, että sosiolingvistisiä metodeja ei vielä 1980-luvulla ollut juurikaan sovellettu kirjoitetun kielen tutkimukseen. Kun näin sitten tapahtui, syntyi kokonainen tieteenala eli historiallinen sosiolingvistiikka (ks. Nevalainen & Raumolin-Brunberg 2012).

Tämän luvun näkökulmasta katsottuna puhutun ja kirjoitetun kielen vuorovaikutuksen perusteella voidaan olettaa, että puhutussa kielessä meneillään oleva muutos heijastuu viipeellä myös kirjoitettuun kieleen, sillä kirjoitettu kieli on yleensä konservatiivisempaa kuin puhuttu kieli. Samalla se kuitenkin antaa myös mahdollisuuden olettaa, että muutos ei toteudu sellaisenaan.

## Suomen possessiivisuffiksi

Omistusliite eli possessiivisuffiksi on suomen kielelle tyypillinen ilmiö (Karlsson 1998, 164–165; Miestamo 2006, 49; WAL 58A), jonka juuret ulottuvat todennäköisesti aina kantauraliin saakka (Mark 1925; Itkonen 1955; Ikola 1992, 176). Suomen kirjakielen isänä pidetty Mikael Agricola rakensi suomen kirjakielen Turun ja Viipurin seudun murteiden varaan, mutta hänen seuraajansa siirsivät painopisteen hämäläismurtei-

siin (Tuomi 1986, 22). Ehkä juuri tästä syystä omistusmuotojen käyttö hämäläismurteissa on lähimpänä kirjakielen kantaa (Ikola 1992, 193).<sup>63</sup> Vastaavasti kauimpana kirjakielen kannasta lienevät itämurteet. Yhdeksi syyksi possessiivisuffiksin käytön vähenemiselle onkin esitetty juuri itämurteiden vaikutuksen vahvistuminen (Ikola 1992, 193; Yli-Vakkuri 1993, 85). Toisaalta etenkin Helsingin puhekielen osalta yhdeksi tekijäksi on nostettu myös ruotsin kielen analyttinen vaikutus (Karlsson 1975, 62; Paunonen 1995, 526).

### POSSESSIIVISUFFIKSI KIRJAKIELESSÄ JA MURTEISSA

Kirjacielessä omistussuhdetta ilmaistaan possessiivisuffiksisilla omistusmuodoilla, joihin saattaa liittyä persoonapronominin genetiivimuoto (Setälä 1966 [1880], 87–88; Ikola 2001, 74–77). Lindénin (1959, 294–295) mukaan niissä tapauksissa, joissa suffiksi viittaa lauseen subjektiin tai subjektiin verrattavaan lauseenjäseneseen (= refleksiivinen käyttö), jää persoonapronomini pois (esimerkiksi *rikoin mukini*). Mikäli suffiksi taas viittaa muuhun kuin lauseen subjektiin (= ei-refleksiivinen käyttö), voidaan 1. ja 2. persoonan pronomini jättää pois, ellei sillä ole painoa (esimerkiksi *sait kirjani*). Kolmannen persoonan pronomini on sitä vastoin aina liitettävä omistusmuodon oheen (esimerkiksi *Teppo rikkoi hänen* [ei siis Tepon] *mukinsa*). Lindénin toteaa, että kirjacielessä 3. persoonan possessiivisuffiksi ei ole merkitystehtävältään yhtä itsenäinen kuin 1. ja 2. persoonan possessiivisuffiksi, sillä se kykenee ilmaisemaan omistajan ainoastaan viitattaessaan lauseen subjektiin, kun taas 1. ja 2. persoonan possessiivisuffiksi kykenee tähän kaikissa tapauksissa. Lindénin mukaan tätä piirrettä ei niinkään tavata murteissa, vaan hän laskee sen kuuluvaksi ensijaisesti kirjakieleen. Paunonen (1995, 505) toteaa Helsingin puhekielen omistusmuotoja käsittelevässä artikkelissaan, että koska 1. ja 2. persoonan persoonapronomininit eivät ole välttämättömiä merkityksen kannalta, ovat kieliopit ja kielioppaat suosittelleet jo pitkään pelkän suffiksin käyttöä. Murteissa tilanne näyttäisi kuitenkin toisenlaiselta, sillä myös 1. ja 2. persoonan omistusmuoto saa usein vahvikkeekseen perso-

63 Kiurun (1996) tutkimuksen mukaan possessiivisuffiksin käyttö on ollut kaikkein yleisintä ylä- ja alasatakuntalaismurteissa sekä Porin seudun murteissa.

napronominin niissä tapauksissa, missä omistusliite ei viittaa lauseen subjektiin (mp.; myös Lindén 1959, 299; Nuolijärvi 1986, 160). Lindén (1959, 299) päättelee, että tämmöisissä tapauksissa saattaa olla kyse siitä, ettei pelkkää possessiivisuffiksia koeta riittävän ilmaisuvoimaiseksi. Näiden tapausten lisäksi on vielä mainittava se seikka, että possessiivisuffiksia ei käytetä silloin, kun omistusmuoto ilmaisee perhettä, yhteisöä tai asuinpaikkaa vaan tällöin riittää pelkkä persoonapronominin genetiivimuoto (esim. *meidän äiti*, Setälä 1966 [1880], 88; Hakanen 2000; Ikola 2001, 75). Tämä sääntö näyttää olevan sama sekä kirjakiellessä että murteissa.

Murteet eivät tietenkään ole yhtenäisiä. Yllä on todettu, että todennäköisesti kauimpana kirjakielen kannasta ovat itämurteet. Esimerkiksi Latvalan (1899, 40–41) mukaan Pohjois-Savon murteessa suffiksin poisjättö on huomattavasti yleisempää kuin kirjakiellessä (esimerkiksi *Onko tämä minun harava?*). Ikolan (1992, 187) katsauksessa suffiksin poisjättö on ylipäätään moninkertaisesti yleisempää itämurteissa kuin länsimurteissa. Toisaalta Järvisen (2011) tutkimuksen mukaan liitteettömyys ei ole tuntematon ilmiö länsimurteissakaan ainakaan hämäläismurteiden osalta. Markin (1925, 241) mukaan ilmiö tunnetaan mm. Satakunnan, Etelä-Hämeen, Pohjois-Savon ja Kaakkois-Suomen alueilla puhutuista murteista.

#### POSSESSIIVISUFFIKSIN KEHITYS PUHUTUSSA KIELESSÄ

Possessiivisuffiksin kehitystä on tutkittu etenkin sosiolingvistisessä kontekstissa (mm. Turunen 1982; Nuolijärvi 1986; Paunonen 1995; Mitrinen 2005; Järvinen 2011). Yleinen mielipide tuntuu olevan, että possessiivisuffiksi elää puh kielessä rajua muutuskautta ja on väistymässä (esimerkiksi Kangasmaa-Minn 1986; Ikola 1992; Paunonen 1995). Esimerkiksi Paunonen (1995, 514) arvioi – kuten jo todettiin –, että possessiivisuffiksi katoaisi Helsingin puh kielestä 2040-luvun tienoilla. Toisaalta vastakkaisiakin näkemyksiä on esitetty. Yli-Vakkuri (1993, 83–84) ei usko, että omistusliitteiden produktiivinen käyttö tulisi suomen kielestä katoamaan, vaan näkee niiden vaihtelevassa käytössä erityyppisiä tekstistrategioita. Hurtta (1999, 88–91) puolestaan kritisoi Paunosen (1995) omistusmuotojärjestelmää koskevaa artikkelia siitä,

että aineistosta on rajattu pois juuri ne tapaukset, joissa possessiivisuffiksin asema on vahva, eli lauseenvastikkeet, nominaalirakenteet, adverbis ja toisen persoonan omistusmuodot. Yhtä kaikki, possessiivisuffiksin kirjakielen mukaisessa hallinnassa on havaittu puutteita myös ylioppilailailla ja lukioikäisillä (Leino 1992; Tiainen 2008). Enteilevästi Kangasmaa-Minn (1986) toteaaakin, että possessiivisuffiksi edustaa uhanalaista kielellistä ainesta, ja Ikola (1992, 176) kirjoittaa yleiskielen normijärjestelmän rappeutumisesta. Ennuste possessiivisuffiksin katoamisesta myös kirjoitetusta kielestä perustuu muun muassa näihin tutkimuksiin sen asemasta puhekielessä.

Helsingin puhekielessä possessiivisuffiksia ovat tarkastelleet ainakin Paunonen (1995 [1982]; 2005) ja Nuolijärvi (1986). Näistä Paunosen (1995 [1982]) tutkimus on näennäisaikainen tutkimus Helsingin puhekielestä ja sen kehittymisestä. Omistusmuotojen osalta tutkimuksessa esitetään, että suffiksiton *mun kirja* -tyyppi on valtaamassa alaa varsinkin nuorilla puhujilla ja leviämässä ei-refleksiivisestä asemasta refleksiiviseen asemaan. Paunonen (mts. 215) kuitenkin toteaa, ettei kyse ole uudesta ilmiöstä, sillä myös vanhemmilla puhujilla esiintyy sekä *minun kirja*- että *mun kirja* -tyyppejä, ja että liitteettömyys on täten kuulunut jo vuosisadan vaihteessa hahmottuvaan puhekieleen. *Mun kirja* -tyypin yleistymisen saa tukea myös Nuolijärven (1986) tutkimuksesta, jossa tarkastellaan Helsinkiin muuttaneiden eteläpohjalaisten ja pohjoissavolaisten kieltä heidän asuttuaan Helsingissä 10–20 vuotta. Tutkimuksen perusteella *mun kirja* -tyyppi on tunkeutunut muuttajien kieleen ja myös refleksiiviseen asemaan, jossa molemmissa murteissa tavataan ainoastaan *kirjani*-tyyppejä (mts. 166). Nuolijärvi (1986, 161–162, 166–168) kuitenkin huomauttaa, että itse tyyppi ei ole muuttajien murteelle vieras, sillä pohjalaismurteissa *mun kirja*- tyyppiä tavataan ei-refleksiivisessä asemassa, ja savolaismurteissa ei-refleksiivisessä asemassa käytössä on *minun kirja* -tyyppi. Paunosen (1995) reaaliaikainen tutkimus osoittaa, että aikaisemmat havainnot *mun kirja* -tyypin vahvenemisesta pitävät paikkansa. Esimerkiksi 1. persoonassa *mun kirja* -tyyppi on vahvistunut molemmissa asemissa noin 20 prosenttiyksikköä kaikki ikäluokat mukaan laskettuna (mts. 508). Nuorilla *mun kirja* -tyyppi on jo hallitseva molemmissa asemissa (mts. 511).

Tampereen puhekielessä possessiivisuffiksia ovat tarkastelleet ainakin Turunen (1982), Mitrunen (2005) ja Järvinen (2011). Lähempään tarkasteluun otamme tässä Järvisen reaaliaikaisen tutkimuksen Vaskiveden murteesta (2011). Tutkimuksen perusteella Helsingin puhekielestä tehdyt havainnot saavat ainakin jossain määrin tukea, sillä 1990-luvun vaskiveteläisten koululaisten ryhmässä liitteetön *mun kirja* -variantti on edennyt refleksiiviseen asemaan 1. persoonassa.<sup>64</sup> Toisaalta kokonaisuudessaan koululaisilla on vähemmän liitteettömiä muotoja kuin hiukan vanhemmilla nuorilla, joten Järvinen päättelee liitteettömien muotojen lisääntymisen taittuneen. Tätä tukee myös se havainto, että keski-ikäisillä liitteellisten muotojen määrä on lisääntynyt (mts. 94–96, 105–106, 113). Järvinen tarkastelee tutkimuksessaan myös muita omistusliitteellisiä muotoja kuten verbien nominaalimuotoja (esimerkiksi *juostessani*) sekä omistusliitteellisiä adverbjeja (esimerkiksi *tahallani*). Näissä tapauksissa omistusliitettä ei siis voi jättää pois eikä vaihtelua täten havaita. Järvisen (2011, 107) mukaan tällaisia muotoja oli vanhassa Vaskiveden aineistossa vanhoilla ja nuorilla 18 prosenttia kaikista omistusmuodoista ja keski-ikäisillä peräti 45 prosenttia. Uudessa Vaskiveden aineistossa nuorilla vaihteluttomia muotoja on 39 prosenttia, keski-ikäisillä 35 prosenttia ja koululaisilla noin 18 prosenttia kaikista omistusmuodoista. Mistään vähäisistä määristä ei siis ole kysymys, vaikkakin ikäryhmittäinen vaihtelu näyttäytyykin melko suurena.

Näistä tutkimuksista on syytä esittää joitakin huomioita. Hurta (1999, 88–91) on kritisoinut Paunosen (1995) tutkimusta siitä, että aineisto on rajattu siten, että jo rajauksen perusteella on odotuksenmukaista, että possessiivisuffiksit näyttäisivät olevan katoamassa. Hurta (mts. 91) viittaa Makkonen-Craigin (1996) tutkimukseen, joka ei tukisi sitä ajatusta, että possessiivisuffiksi olisi katoamassa kokonaan. Yllä on nähty, kuinka Järvisen tutkimuksessa vaihteluttomien omistusliitteellisten muotojen määrä on varsin suuri. On selvä, että hypoteettinen possessiivisuffiksin katoaminen suomen kielestä (ks. Karlsson 1975) voi tapahtua vasta sitten, kun suffiksi katoaa kaikista asemista – myös vaihteluttomista asemista. Tähän ei riitä liitteettömien muotojen leviäminen

64 Järvinen (2011, 105) tosin huomauttaa, että refleksiivisiä muotoja on kaikkiaan melko vähän.

refleksiiviseen asemaan i. persoonassa.<sup>65</sup> On kuitenkin huomattava, että Paunosen (1995) tutkimus käsittelee ainoastaan Helsingin puhekieltä, ei suomen puhekieltä kokonaisuudessaan. Tässä mielessä on vähintäänkin hämäävää, että Paunonen silti puhuu koko puhe-suomen kattavasta muutoksesta – mistä Hurttu Paunosen tutkimusta osuvasti kritisoi (1999, 88). Toisaalta on muistettava, että vaihtelun tutkiminen on täysin mahdotonta, ellei ole vaihtelua, mitä tutkia. On kuitenkin syytä pitää erillään aluepuhekielen tutkimus ja koko suomen kieltä ja sen erilaisia käyttäjiä kattavat ennusteet (muutoksen ennustamisesta tarkemmin, ks. Laasanen 2016).

Sekä Paunosen (1995 [1982]; 1995) että Nuolijärven (1986) tutkimuksissa Helsingin puhekielen tietynlaiseksi kuriositeetiksi esitetään *mun kirja* -tyypin esiintyminen refleksiivisessä asemassa. *Mun* ei kuitenkaan ole erillinen possessiivipronomini vaan pronominin *minun* hyvin yleinen pikapuhemuoto, jolle on monia vastineita myös murteissa, aivan samoin kuin *mua* on partitiivimuotoisen *minua*-sanana pikapuhemuoto. Järvinen (2011) puhuikin omassa tutkimuksessaan yleisesti *minun kirja* -tyypistä mutta mainitsee erikseen (mts. 73) koululaisten käyttävän yksinomaan pronominin *mun*-varianttia. Mikäli tyypit *minun kirja* ja *mun kirja* asetetaan rinnakkain, kyse ei siis ole siitä, että *minun kirja* -tyypin rinnalle on tullut uusi tyyppi, vaan siitä, että possessiivipronomini on redusoitunut muotoon *mun*. Tällä on loppujen lopuksi vain vähän tekemistä possessiivisuffiksin kehityksen kanssa.

Yllä on todettu, että possessiivisuffiksin kirjakielen mukaisessa hallinnassa on havaittu puutteita mm. ylioppilaskirjoituksissa (esimerkiksi Leino 1992). Yli-Vakkuri (1993, 78) toteaa, että virheitä tulee sitä enemmän, mitä vähemmän kielenkäyttäjä on joutunut tekemisiin kirjakielen kanssa. Paunonen (1995, 527) arvelee, että omistusliitteiden kirjakielen mukaista oikeaoppista käyttöä saattaa olla jopa vaikeata opettaa oppilaille, joiden omaan kieleen omistusliitteet eivät yksinkertaisesti kuulu. Oppilaille opetetaan kuitenkin aika paljon sellaisia asioita, jotka eivät

65 Vrt. Yli-Vakkuri (1992, 84–85): ”Jos omistusmuotojen vaihtelulla on jonkinasteinen vakiintunut tekstuaalis-pragmaattinen funktio, jos se toisin sanoen palvelee tekstistrategiaa, minusta on vaikea uskoa, että omistusliitteiden produktiivinen käyttö tulisi katoamaan suomen kielestä –”.



(vielä) kuulu heidän elämäänsä; ei ole mitään syytä olettaa, että possessiivisuffiksi olisi tässä joukossa poikkeus (ks. myös Hurttu 1995, 89–91).

Kyse ei ole kuitenkaan ainoastaan omistusjärjestelmän hallinnasta, vaan yleisemmin kielen sääntöjen tai normien natiivihallinnasta, jota ainakin suomen kielen osalta on tähän mennessä tutkittu auttamattoman vähän (ks. esimerkiksi Pajunen, Itkonen & Vainio 2015). Kuten Pajunen ja muut (mts. 160) toteavat, yleinen mielipide lienee se, että kaikki hallitsevat äidinkieltensä riittävästi. Tämä oletus kumpuaa kuin huomaamatta näkemyksestä, jossa jokaiselle kielenkäyttäjälle oletetaan yhtenäinen, homogeeninen kieliooppi tai kompetenssi, jota tutkii psykologian alaan kuuluva kieliooppioteoria. Hypoteettinen, kaikkia puhujia yhdistävä homogeeninen kompetenssi on tässä näkemyksessä teoreettinen välttämättömyys, joka mahdollistaa psykologisen ulottuvuuden saavuttamisen. Ongelmana on se, ettei tälle oletukselle ole perusteita. Tämä käy ilmi, kun kielen ajatellaan olevan yhdentyyppistä tietoa. Puhuja esimerkiksi tietää, että suomen kirjakielissä lauseen subjektiin viittaavan suffiksin yhteyteen ei liitetä possessiivipronominia, tai ei tiedä.<sup>66</sup> Mikäli puhuja tietää tämän, hän ei tee virhettä. Mikäli puhuja ei tiedä tai muista tätä, hän tekee virheen (mikäli ei arvaa oikein). Mikäli toinen puhuja tietää tämän, mutta toinen puhuja ei, eroavat nämä kaksi puhujaa toisistaan kirjakielen hallinnan suhteen. Toinen puhuja kuuluu siihen joukkoon, joka tekee ylioppilaskirjoituksissa virheen ja joka saa aikaan huolen yleiskielen normijärjestelmän rappeutumisesta, toinen puhuja taas ei.

Kirjakielen säännöt ovat kielen sääntöjä aivan samoin kuin murteidenkin säännöt. Se, että murteet eivät tunne pronominiin liikkakäytön välttämisperiaatetta (ks. Paunonen 1995, 505), ei tee niistä jollakin olennaisella tavalla erilaisia kuin kirjakieli. Molemmat sisältävät sääntöjä, mutta sääntöjen sisältö saattaa olla hiukan erilainen. Tässä onkin asian ydin: kaikki puhe virheistä ja puutteista ja normijärjestelmien rapautumisesta kertoo siitä, että kieli todellakin on yhdentyyppistä tietoa, jonka hallinnan suhteen puhujat varioivat. Variaation ulkoistaminen perfor-

66 Jackendoff (2002, 28) väittää täysin ristiriitaisesti, että puhujilla ei ole mitään tällaista tietoa, vaikka hänen oma esityksensä perustuu kokonaisuudessaan tällaisen tiedon olemassaololle ja käytölle (ks. Itkonen 2008, 16).

manssiin on teoreettisista lähtökohdista tehty *ad hoc* -ratkaisu, jonka avulla tutkimuskohteesta saadaan halutunlainen (vrt. vaihteleva kompetenssi, Ellis 2012).

Vaikka possessiivisuffiksin käytössä onkin havaittu selvää muutosta eri tutkimuksissa, näyttäisi näkemys ”dramaattisesta” ja ”rajusta” muutosvaiheesta sittenkin perustuvan yksittäisiin esimerkkeihin. ”Minä tahdon minun pillerit!”, kirjoittaa Kangasmaa-Minn (1986, 7); Paunonen (1995, 501) puolestaan viittaa ”– erään kaupanalan keskusjärjestön *meidän merkki* -iskulause[eseen]”. On kuitenkin kysyttävä, kuinka dramaattisesta muutoksesta voi olla kyse, jos liitteettömyys on kuulunut jo vuosisadan vaihteessa hahmottuvaan Helsingin puhekieleen (Paunonen 1995 [1982], 215), jos liitteettömyydestä on merkkejä jo vanhoissa murreaineistoissa (Järvinen 2011) tai jos jo vuonna 1966 Y. Lumio on ollut huolissaan omistusliitteiden poisjätöstä (Paunonen 1995, 501). Yksittäiset esimerkit voivat olla tietoinen tai tiedostamaton keino korostaa jonkin asian merkitystä, mutta loppujen lopuksi ne ovat ainoastaan yksittäisiä esimerkkejä.

Kangasmaa-Minn (1986, 7) toteaa kirjoituksessaan, että possessiivisuffiksi on taipuvainen katoamaan, koska se sijaitsee sanassa viimeisenä ja on toisinaan redundantti, kuten ilmauksessa *minun kalani*. Redundanttisen, sanassa viimeisenä esiintyvän aineksen poisjätto on kielen ekonomisuuden kannalta täysin rationaalista. Redundantit esiintymät ovat juuri niitä vaihtelevia tapauksia, joissa possessiivisuffiksin käytön katsotaan vähentyneen. Puhtaasti ekonomisuuden näkökulmasta on siis odotuksenmukaista, että suffiksin käyttö vähenee. Toisaalta kokeellinen tutkimus osoittaa, että possessiivijärjestelmän kaltaiset kongruenssisysteemit<sup>67</sup> eivät edusta täysin redundanttia tietoa vaan niillä on selvä funktio: ne nopeuttavat kielen prosessointia sitomalla yhteenkuuluvat ainekset yhteen (ks. Vainio, Hyönä & Pajunen 2008). Kuten yllä on jo todettu, suffiksi ei voi kuitenkaan kadota suomen kielestä, ennen kuin se katoaa myös ei-redundanteista asemista.

67 Kongruenssisysteemejä ovat mm. subjekti- ja objektikonjugaatio sekä omistus- ja pääsana-attribuuttikongruenssi.

## POSSESSIIVISUFFIKSIN KÄYTÖN KEHITYS MODALITEETISTA TOISEEN

Havainnot possessiivisuffiksin käytön vähenemisestä perustuvat ensisijaisesti puhuttuun kieleen. Kirjoitetusta kielestä on tarjolla pääsääntöisesti yksittäisiä esimerkkejä (ks. esimerkiksi Ikola 1992; Paunonen 1995), kuten jo todettiin. Yleinen mielipide lieneekin se, että kyse on puhutussa kielessä alkaneesta muutoksesta, joka hiljalleen levittäytyy kirjoitettuun kieleen (vrt. Tiainen 2008). Tämä on kuitenkin vain yksi mahdollinen vaihtoehto yhteensä neljästä teoreettisesti mahdollisesta vaihtoehdosta.

Ensimmäisen vaihtoehdon mukaan muutos on siis alkanut puhutussa kielessä, josta se on hiljalleen levinnyt kirjoitettuun kieleen. Liitteettömyys kirjoitetussa kielessä voi liittyä puhekielen tavoitteluun, huolimattomuuteen tai puutteelliseen sääntöjen hallintaan (ks. Varjonen 1996). Koska kirjoitettu kieli muuttuu yleensä hitaammin kuin puhuttu kieli, voisi ajatella, että kirjoitetun kielen tutkimus paljastaisi jonkin verran puhutun kielen tilannetta varhaisemmassa vaiheessa olevan muutoksen. Toisin sanoen myös kirjoitetussa kielessä tulisi havaita liitteettömyyden vahvistumista mutta ei välttämättä niin paljon kuin puhutussa kielessä.

Toisen vaihtoehdon mukaan muutos on alkanut kirjoitetussa kielessä, josta se on siirtynyt puhuttuun kieleen varsin nopeasti tai siten, että muutos on puhutussa kielessä äkillisesti kiihtynyt. Se, ettei liitteettömyyttä ole havaittu kirjoitetussa kielessä johtuisi tässä vaihtoehdossa siitä, ettei aihetta ole tutkittu riittävän tarkasti.

Kolmannen vaihtoehdon mukaan muutos on alkanut jokseenkin yhtäaikaaisesti sekä puhutussa että kirjoitetussa kielessä. Tässä vaihtoehdossa molemmissa kielen ilmenemistavoissa tulisi näkyä merkkejä liitteettömyyden vahvistumisesta, mutta tutkimuksen painopisteen ollessa puhutussa kielessä tilanne näyttää siltä, kuin muutos olisi alkanut puhutussa kielessä. Tämän vaihtoehdon erottaminen ensimmäisestä vaihtoehdosta lienee hyvin hankalaa, mutta voisi ajatella, että tässä tapauksessa kirjoitetussa kielessä esiintyisi jonkin verran ensimmäistä tapausta enemmän merkkejä liitteettömyydestä.

Ensimmäinen vaihtoehto lienee todennäköisin jo siitä syystä, että kyse on puhekielelle varsin tyypillisestä ilmiöstä ja tässäkin kirjoituk-

nessä on viitattu *mun*-muotoon persoonapronominin *minun* pikapuheumuotona. Kaksi muuta ovat kuitenkin teoreettisesti mahdollisia. Se, ettei niitä yleensä edes tuoda esiin, kertoo siitä, miten vahvasti ensisijaiseksi puhuttu kieli mielletään – aivan kuin huomaamatta. Mukaan on kuitenkin otettava myös neljäs vaihtoehto, jonka mukaan muutos ei ulottuisikaan kirjoitettuun kieleen, ja siinä ei odotuksenvastaisesti näkyisikään merkkejä possessiivisuffiksin käytön vähenemisestä verrattaessa eriaikaisia aineistoja. Kirjoitetussa kielessä havaittavat yksittäiset merkit suffiksin poisjätöstä olisivat tällöin osoitus (pitkäkestoisesta) variaatiosta.

## Possessiivisuffiksi kirjoitetussa kielessä

Possessiivisuffiksin käytön kehitys kirjoitetussa kielessä on käytännössä tutkimaton aihe. Kirjoitetun kielen osalta on esitetty vain yksittäisiä esimerkkejä ja huomioita, joiden mukaan puhutussa kielessä etenevä muutos on todennäköisesti leviämässä myös kirjoitettuun kieleen. Esimerkiksi Ikolan (1992, 181) ennusteen mukaan possessiivisuffiksien käyttö vähentyy sanomalehtikielessä. Koska tutkimuksemme on mukana 2000-luvun sanomalehtikorpuksia ja kirjoitelmia, voidaan Ikolan asettama ennuste testata. Mikäli esitetyt havainnot ja tehdyt ennusteet possessiivisuffiksin käytön vähenemisestä kirjoitetussa kielessä pitävät paikkansa, tulisi kirjoitetussa kielessä näkyä tästä selviä merkkejä. Mikäli tällaisia merkkejä ei löydy, voidaan ennusteet laskea epäonnistuneiksi, missä tapauksessa on pohdittava epäonnistumisen syitä. Tutkimuksemme taustana on siis yhtäältä puhutussa kielessä todettu muutos suffiksien käytön vähentymisen suuntaan ja toisaalta huomiot ja ennusteet suffiksien käytön vähenemisestä kirjoitetussa kielessä.

Koska tutkimuksemme käyttää massakorpuksia, eroavat tarkasteltavat yksityiskohdat ja kielenpiirteet tietyiltä osin aiemmista tutkimuksista. Esimerkiksi suffiksien käytön tarkastelu refleksiivisessä ja ei-refleksiivisessä asemassa, joka on tärkeässä asemassa useissa puhutun kielen tutkimuksissa, ei ole mahdollinen aineiston koon vuoksi. Toisaalta tutkimuksessa tarkastellaan yksityiskohtaisemmin mm. suffiksien käytön kehitystä sanaluokittain ja verbimuodoittain.

Aloitamme esittelemällä aineistot ja tutkimusmenetelmät, minkä jälkeen luomme yleiskatsauksen possessiivisuffiksien esiintymiin ja vaihteluun eri vuosikymmeninä. Tämän jälkeen tarkastelemme possessiivisuffiksien käyttöä ja kehitystä sanaluokittain; verbejä analysoimme myös niiden funktioiden kautta. Lisäksi tarkastelemme suffiksien kehitystä sellaisten niin sanottujen indeksisanojen yhteydessä, jotka ilmaisevat luontaisesti tai juridisesti omistettavia käsitteitä. Viimeiseksi katsomme vielä erikseen kolmannen persoonan tunnuksen vaihtelua ja muutosta. Lopuksi kokoamme tulokset ja pohdimme ennusteiden paikkansapitävyyttä.

### AINEISTO JA SEN KOODAUS

Tutkimuksen aineistona on sähköisiä, eri vuosikymmeniä edustavia korpuksia, joista on poimittu kaikki possessiivisuffiksilliset muodot seuraavan jaottelun mukaan: 1) yksikön 1. persoonan suffiksi, 2) monikon 1. persoonan suffiksi, 3) yksikön 2. persoonan suffiksi, 4) monikon 2. persoonan suffiksi sekä 5) kolmannen persoonan suffikit tunnuksen tyyppin mukaan jaotellen (*nsa*, *nsä*, vokaalin pidentymä+n). Kolmannen persoonan alaluokittelun syynä on pyrkimys tutkia erikseen, a) ovatko variantit (*Vn/nsV*) vapaassa vaihtelussa vai eivät kaikissa niissä tapauksissa, joissa molemmat periaatteessa ovat mahdollisia (kuten *taloaan* tai *taloansa*), ja toisaalta, b) onko varianteissa produktiivisuuseroja. Ajallisesti aineistoa on viimeiseltä 70 vuodelta ja se edustaa sekä sanomalehtikieltä että noin 18–20-vuotiaiden nuorten kieltä (ylioppilasainekirjoitelmia). Lisäksi 2000-luvun kieltä on sekä alakoulun oppikirjoista, peruskoululaisten ja nuorten aikuisten Unelmakirjoitelmista sekä fiktiosta. Pyrimme siten arvioimaan possessiivisuffikseja sekä ajallisesti että eri tekstilajeista ja -tyypeistä. Aineisto esitellään yksityiskohtaisesti tämän teoksen johdannossa.

Sanomalehtiaineistona on: a) Lauseopin arkiston (LA) materiaali, joka edustaa 1970-luvun sanomalehtikieltä mutta lisäksi kirjoitettua yleiskieltä Oulun korpuksen jakauman mukaan (ks. Saukkonen 1982). Otoksena on 20000 tekstisanaa (sanomalehdet, aikakauslehdet sekä ei-fiktiivinen kirjallisuus). b) Helsingin Sanomien vuosien 2001–2002 aineisto eli HS2000-korpus (Pajunen & Virtanen 2002a). Korpuksessa on noin 23,4 miljoonaa tekstisanaa. Lisäksi on c) miljoonan tekstisanan

otos Savon Sanomien vuoden 2001 aineistosta eli SS2000-korpuksesta (n. 8 miljoonaa tekstisanaa, Pajunen & Virtanen 2002b). Kirjoitelmakorpus (Pajunen 2004) on koostettu vuosina 1948–2003 ilmestyneistä valikoitujen ylioppilasaineiden kokoomateoksista ja sen sanamäärä on yhteensä miljoona tekstisanaa. Kirjoitelmakorpukselta on poimittu kolme eriaikaista otosta: 1948–1958 (144 752 tekstisanaa), 1968–1978 (272 816) ja 1988–1998 (215 850). Otoksiin viitataan taulukoissa useimmiten lyhenteillä 1950-luku, 1970-luku ja 1990-luku. Paikoin on käytössä koko miljoonan sanan kirjoitelma-aineisto. Kirjoitelmakorpus on koostettu ajallisin kriteerein ja se on sisällöllisesti samantapaista, vastaavasti 1970-luvun ja 2000-luvun lehtikieli ovat vertailtavissa. Lisäksi on vajaan puolen miljoonan tekstisanan aineisto, josta puolet koostuu 2000-luvun ensimmäisellä vuosikymmenellä käytössä olleista alakoulun eri oppiaineiden oppikirjoista (ks. Jaakola 2004)<sup>68</sup> ja puolet fiktiivisestä kirjallisuudesta (yht. 417 919 tekstisanaa). Joitain viittauksia tehdään Lauseopin arkiston murreaineistoihin, joista otoksena on 200 000 tekstisanaa (ks. Pajunen & Palomäki 1984). Kaikki aineiston osiot Lauseopin arkiston materiaalia lukuun ottamatta on koodattu samalla jäsentimellä (funktionaaliseen dependenssikielioppiin perustuva FDG-jäsentin, ks. Voutilainen 2002), ja aineiston joka osiosta on poimittu kaikki possessiivisuffiksilliset esiintymät samalla ContextMill-ohjelmalla (ks. Virtanen & Pajunen 2001). LA:n aineisto on koodattu manuaalisesti mutta se on moneen kertaan ja monella tapaa tarkistettuna luotettava. Kehityksellinen aineisto 2010-luvulta – eli peruskoululaisten ja nuorten aikuisten kirjoitelmista koostuva Unelma-aineisto – on mukana vertailussa vain suppeasti, koska sitä ei ole koodattu FDG-jäsentimellä muun aineiston tapaan. Unelma-aineistossa ei ylipäätään ole morfologista koodausta ja lisäksi sanamäärä ei ole kaikkien ikäryhmien osalta riittävä (aineisto esitellään tarkemmin tämän teoksen johdantoluvussa).

Automaattiset poimintaohjelmat sekä ali- että yligeneroivat jossain määrin. Puuttuvia muotoja ei pysty löytämään mutta kaikki yligeneroituneet on pyritty poistamaan tai niiden koodaus korjaamaan. Ohjelma

68 Oppikirja-aineiston on koontanut Mari Jaakola (2004) pro gradu -tutkielmaansa ja se on liitetty Kirjoitelmakorpuksen FDG-analysoituna. Kiitämme Mari Honkoa aineiston käyttöluvasta.

esimerkiksi poimii jonkin verran illatiivimuotoisia sanoja possessiivi-suffiksillisina. Lisäksi on ollut tarpeen oikaista paikoin sanaluokitusta ja esimerkiksi selvästi kiteytyneet substantiivi- ja pronomini muodot on korjattu adverbeiksi (*tavallaan, mielellään, kohdallaan; sinänsä, yksinään*). Jonkin verran löytyy satunnaisia muita sanaluokkavirheitä. Esimerkiksi muoto *elämääni* on korjattu verbimuodosta substantiivimuodoksi, koska konteksti paljastaa, että kyseessä ei ole tyyppi *elämääni elämää* vaan *elämääni kuuluu*. Aineistoon on voinut jäädä virheitä, mutta ne eivät todennäköisesti vähennä tilastollista luotettavuutta.

#### YLEISKATSAUS POSSESSIIVISUFFIKSIEN ESIINTYMIIN JA VAIHTELUUN ERI VUOSIKYMMENINÄ

Possessiivisuffiksien sisäinen jakauma eri persooniin osoittaa, että sanomalehtikieli on samantapaista vuosikymmenestä riippumatta: neutraalissa ja asiakeskeisessä tekstissä 3. persoonan suffiksi on täysin valitseva (ks. taulukko 32). Nykykielen lehtinäytteistä *Savon Sanomissa* 1. persoonaa käytetään selvästi vähemmän kuin *Helsingin Sanomissa*. Tämä voi heijastaa savon murteen oletuksen mukaan vähäisempää suffiksin käyttöä (vrt. Latvala 1899) tai sitten lehden rennompaa tyyliä. Sen sijaan kirjoitelma-aineisto ja LA:n koko otos eroavat selvästi lehtikielestä, koska 3. persoonan muotoja on 15–20 prosenttiyksikköä vähemmän kuin lehtikielessä. Ero on 1. persoonan hyväksi. Kun LA:n aineistosta erotetaan ei-fiktiivinen tietokirjallisuus, jossa on mm. muistelmia, lehtikielen aineistojen ero häviää. Muistelmateksteissä on ymmärrettävästi paljon *elämäni, kotimme* jne. ilmauksia. Oppikirjoissa ja fiktiossa tilanne on lehtikielen kaltainen mutta oppikirjoissa 2. persoonan jakauma eroaa runsaiden tehtävänantojen takia (*vihkosi* jne.).

Suffiksien jakauma näyttää 1970-luvun sanomalehtityyppisessä kirjoitetussa kielessä samantapaiselta kuin 2000-luvun sanomalehtimassakorpuksissa, eikä ole havaittavissa siirtymää 3. persoonan suffikseihin, mikä voisi viitata esimerkiksi suffiksin jonkin asteiseen kiteytymistendenssiin. Sen sijaan aineisto osoittaa, että 1. ja 3. persoonan suffiksien välillä on tekstilajieroa (vrt. taulukot 32 ja 33). Tulkintaa vahvistaa myös se, että LA:n pääkirjoituksissa on puolet vähemmän 1. persoonan suffikseja kuin uutis- ja reportaasiteksteissä. Toisin sanoen hyvin muodolli-

nen ja abstrakti tekstilaji vieroo i. persoonan suffiksija. Tulos on odotettava, koska pääkirjoituksia ei muutenkaan kirjoiteta minä-näkökulmasta (vaikka näkökulma sinänsä voi olla subjektiivinen). Kun sitten lehtikieltä verrataan muuhun aineistoon (kirjoitelmat, oppikirjat ja fiktio), sama korostuu. Etenkin kirjoitelmissa omakohtaisuus (*minun mielestäni*) ja muut valinnat (*kotimaamme luonto*) nostavat i. persoonan käyttöä ja laskevat kolmannen.

Taulukko 32. Possessiivisuffiksien jakauma otoksissa persoonan mukaan.

Korpus / Possessiivisuffiksi %	I. p. ps <i>kirjani, kirjamme</i>	2. p. ps <i>kirjasi, kirjanne</i>	3. p. ps <i>kirjansa</i>
LA 1970	37,1	1,1	61,7
LA lehdet 1970	16,4	1,7	81,9
Helsingin Sanomat	15,9	1,4	82,7
Savon Sanomat	12,6	1,4	86
Kirjoitelmat 1940–2000	30	2,6	67,3
Oppikirjat ja fiktio	16,8	14,6	68,5
Keskiarvo %	21,5	3,8	74,6

Taulukko 33. Possessiivisuffiksien jakauma otoksissa persoonan ja luvun mukaan.

Korpus / Possessiivisuffiksi %	Sg. I. p <i>kirjani</i>	Sg. 2. p <i>kirjasi</i>	Pl. I. p <i>kirjamme</i>	Pl. 2. p <i>kirjanne</i>	3. p <i>kirjansa</i>
LA 1970	19,6	0,1	17,5	1	61,7
LA lehdet 1970	2,5	0,3	13,8	1,4	81,9
Helsingin Sanomat	10,7	1,1	5,2	0,3	82,7
Savon Sanomat	7,2	1,1	5,4	0,3	86
Kirjoitelmat 1940–2000	16,2	2,1	13,8	0,5	67,3
Oppikirjat ja fiktio	14,2	12,8	2,6	1,8	68,5
Keskiarvo %	11,7	2,9	9,7	0,9	74,6

Aineistot eivät ole täysin verrannollisia mutta osoittavat sen, että possessiivisuffiksien käyttöä ja käytön muutoksia ei voi analysoida pelkästään yhdestä näytteestä, vaikka ajalliset erotkin olisivat näkyvissä, koska käyttö vaihtelee tekstilajeittain ja oletettavasti myös tekstityypeittäin. Aineisto osoittaa suffiksien käytön vaihtelevan myös tyylillisesti. Vanhimmissa teksteissä vilisee ilmaisuja kuten *kotimaamme* ja *taloutemme*, jotka 2000-luvulla ovat vähemmän käytettyjä (ks. taulukko 33). Aineisto nostaa esiin myös tendenssin teitittelystä sinutteluun: monikon 2. persoonan



muodot menettävät asemaansa yksikön 2. persoonan muodoille. Määrällisesti ero on pieni, mutta niin on 2. persoonan yleisyyskin vähäinen.

Sanaluokka tai sana, joka taipuu maksimaalisesti niissä taivutuskategorioissa, jotka sille ovat ominaisia, on yleensä paitsi prototyypinen myös produktiivinen (ks. Pajunen 2010). Esimerkiksi substantiivi voi suomessa taipua sijan, luvun, possessiivisuffiksin sekä sävy- ja kysymysliitepartikkelin mukaan, esimerkiksi *kirjoissasikohan*. Nominisanaluokat voi suomessa erottaa paitsi taipumisen määrän mukaan myös taivutuskategorioiden mukaan. Esimerkiksi adjektiivit taipuvat enemmän kieliopillisissa kuin lokaalisissa sijoissa ja adjektiivit tai pronominit eivät yleensä salli possessiivitaivutusta. Kun possessiivitaivutuksen eri persoonien muodot nyt arvioidaan sen mukaan, monessako eri sanojen taivutustyypeissä ne keskimäärin esiintyvät, pystytään arvioimaan, onko tietyn suffiksin käyttö enemmän produktiivista (mikä tahansa vaihtoehto, esimerkiksi sija tai luku) kuin kiteytyntä (vaihtoehdot kasautuvat harvempiin). *Helsingin Sanomien* aineistosta löytyy esimerkiksi lekseemistä *aivot* kaikkiaan yhteensä 17 eri suffiksin taivutusmuotoa (Sg1P: *aivoni, aivoissani, aivojeni*; Pl1P: *aivoissamme, aivojamme, aivojemme, aivomme*; Sg2P: *aivosi, aivoissasi, aivoistasi*; Pl2P: –; 3P: *aivoihinsa, aivojensa, aivonsa, aivoillaan, aivoissaan, aivoistaan, aivojaan*), mikä osoittaa, että sanaa käytetään monenlaisissa konteksteissa. Taivutusmuotoanalyysi on tehty HS2000-korpukselle (ks. taulukko 34) ja kirjoitelma-aineistolle.

Taulukko 34. Possessiivisuffiksilliset taivutusmuototyytit ja -esiintymät HS2000-korpussa.

Taivutusmuototyytit	Tyyppejä	Esiintymiä	Suhdeluku
Ensimmäinen	10 382	32 207	3,1
sg	5 720	21 742	3,8
pl	4 662	10 465	2,2
Toinen	1 467	2 791	1,9
sg	1 000	2 159	2,2
pl	467	632	1,4
Kolmas	35 130	167 902	4,8
an	14 767	63 137	4,3
än	4 740	27 151	5,7
nsa	11 739	54 807	4,7
nsä	3 884	22 807	5,9
Yhteensä	46 979	202 900	

Koko HS2000-aineistossa on suffiksillisia eri sanojen tyypejä perusmuodossa tai jossain taivutetussa muodossa 46 979 kpl; suurin osa on kolmannen persoonan taivutusmuotoja (ks. taulukko 34). Kun näiden esiintymien kokonaismäärä (202 900) jaetaan taivutusmuototyyppien määrällä, saadaan tulokseksi, että keskimäärin possessiivisuffiksillisella sanalla on 4,3 esiintymää. Kun sama suhdeluku lasketaan possessiivisuffiksin alatyypeistä, nähdään, että kolmannen persoonan suffiksit keskittyvät enemmän tiettyihin samoihin muotoihin – eli suhdeluku on korkeampi – kuin toisen ja ensimmäisen persoonan suffiksit. Suurin vaihtelevuus on monikollisissa suffiksissa ja suurin vaihtelemattomuus kolmannen persoonan *nsA*-tunnuksella merkityissä esiintymissä ja ääkkösmuodoissa (*-än* ja *-nsä*). Tämä luonnollisesti viittaa 3. persoonan muotojen kiteytymistendenssiin (funktio ei olisi enää läpinäkyvästi muodosta pääteltävissä).

Laskelmaa ei kannata tehdä aineiston kaikista osioista, koska ne ovat niin paljon pienempiä, että etenkin 1. ja 2. persoonan muotoja ei ole laskelmiin riittävästi. Persoonien välinen suhde on kuitenkin samanlainen sekä lehtikielessä että esimerkiksi kirjoitelmissa. Kirjoitelmissa 3. persoonan suhdeluku nousee ajan myötä mutta ei täysin lineaarisesti. Taivutusmuototyyppien ja esiintymien suhdeluku on eräänlainen diversiteetilaskelma, toisin sanottuna saatuihin arvioihin vaikuttaa voimakkaasti tekstimassan määrä (ks. esimerkiksi McCarthy & Jarvis 2007). Siksi ei ole itsestään selvä, että 3. persoonan tunnusten kiteytymistendenssi toimisi yleisemmin myös muutosennusteena, mutta se osoittaa silti, että tietyillä tunnuksilla merkittyjen sanojen toistuvuus on korkeampi kuin toisilla tunnuksilla merkittyjen.

Kehityksellisessä aineistossa eli Koulu- ja Unelmakorpuksessa possessiivisuffiksien käyttö keskittyy yksikön 1. persoonaan, mikä on kirjoitelmien teeman (”Unelmiäni päivä”) perusteella odotuksenmukaista. Silti kaikkien persoonien muodot ovat käytössä, nuorimmilla on jopa monikon 2. persoonan suffiksit hallinnassaan. Kehityksellisesti possessiivisuffiksien määrä lisääntyy. Alakoulun luokkien 1.–5. kirjoitelmissa suffiksi on 2,6 prosentissa sanaesiintymistä, kuudesluokkalaisilla ja yläkoululaisilla noin 4 prosentissa mutta nuorilla aikuisilla jo 5,8 prosenttia

sanaesiintymistä on suffiksillisia.<sup>69</sup> Vuoden 2006 kuudesluokkalaisilla suffikseja on hieman vähemmän kuin vuoden 2014 kuudesluokkalaisilla; tämä ero selittynee alueellisilla eroilla. Varhemmassa aineistossa kirjoittajat tulevat pohjalais- ja hämäläismurteiden alueelta, myöhemmässä puolet kirjoittajista on lounaismurteiden alueelta, ja erityisesti heidän teksteissään on paljon suffiksillisiä muotoja. Aineisto kuitenkin osoittaa, että 2010-luvunkin lapset ja nuoret käyttävät ja osaavat käyttää possessiivitaivutusta.

#### POSSESSIIVISUFFIKSIT JA SANALUOKKAKAUMAT

Possessiivisuffiksin kanssa esiintyvät sanat ovat aineistossa substantiiveja, verbejä, adverbeja ja lisäksi on muutamia luokkaan muu kerättyjä harvinaisempia tapauksia. Alustava sanaluokitus on FDG-jäsentimen tekemä ja se perustuu kontekstuaaliseen luokitukseen (ks. Voutilainen 2002). Luokituksessa esimerkiksi muoto *haastateltavia* on tulkittu substantiiviksi eikä adjektiiviksi tai verbiksi, koska sana kontekstissaan toimii subjektin tai objektin funktiossa. Tähän luokitukseen on tehty joitain korjauksia aineistoa analysoitaessa. Esimerkiksi kiteytyneet substantiivien tai adjektiivien muodot on tulkittu adverbeiksi, kuten jo mainittiin edellä. Siten HS2000-korpuksessa melko yleinen olotilaa ilmaiseva tyyppi *-immillAAn* on sanaluokaltaan adverbi (*hauskimmillaan, aktiivisimmillaan, äärimmillään*) eikä adjektiivi, vastaavasti tyyppi *ahneuksissaan* on adverbi eikä substantiivi. Erottelukriteerinä on useimmiten, että sana ei taivu kyseisen olotilanilmauksen merkityksessä maksimaalisesti eli kaikissa nominitaivutuksen muodoissa. *Mieli*-sanana monet sijamuodot ovat useimmiten adverbeja (*mielestään, mielelläni*). Spatiaalista tai topologista merkitystä ilmaisevat sarjat *alleni, altani* jne. on niin ikään tulkittu adverbeiksi.

Pronominiksi on luokiteltu lähinnä indefiniittisiä *kaikkinensa, toisiaan* ja *itseensä* -tyyppisiä muotoja mutta satunnaiset *minäni, minänsä* tyypit on tulkittu substantiiveiksi, koska niitä on käytetty *minuus*-tyyppisten lekseemien sijaan tai tapaan. Lukusanat aineistossa esiintyvät pääasiassa nominikannan yhteydessä (*45-sarjassa*) ja lähinnä itsenäisinä

69 Poiminnat Unelmakirjoittelusta on tehty sanaesiintymän muodon perusteella ja tarkistettu manuaalisesti.

lukusanoina esiintyvät ovat ordinaalilukusanoja (*neljäntenään*). Pronominit ja lukusanat ovat suffiksillisina marginaalisia ja vastaavasti adjektiivieista possessiivisuffiksin kanssa esiintyy vain osa (esimerkiksi *entisensä, kaltaisensa*).

Substantiivien ja verbien rajanveto on koskenut lähinnä partisiippeja. Kaksi kolmannelta aineiston sanamuodoista ja niiden esiintymistä on substantiiveja, runsaat kymmenen prosenttia adverbeja ja kymmenisen prosenttia verbejä (ks. taulukko 35). Ryhmään Muu kuuluvat pronominit, adjektiivit ja lukusanat. Verbimuodot ovat erilaisten rakenteiden, kuten lauseenvastikkeiden predikaattiosana. Niihin lukeutuvat seuraavat muodot: *tehdäkseni, kulkiessani* sekä partisiipit luvun ja sijamuodon mukaan eriteltynä ja lisäksi erilaiset *ma*-partisiipit. Substantiivin luokitus on annettu sanalle, jota ei ole ollut mitään syytä luokitella muihin sanaluokkiin.

Sanaesiintymät, joissa on possessiivisuffiksi, jakautuvat aineiston eri korpuksissa sanaluokkiin samantapaisesti (ks. taulukko 35). Substantiiveissa on ajallisessa vertailussa 1950-, 1970- ja 1990-lukujen kirjoitelmissa pientä laskua mutta 2000-luvun aineistoissa lähinnä vaihtelua. Adverbeissa on pientä nousevaa trendiä. Possessiivisuffiksilliset verbit liittyvät kompleksiseen kielenkäyttöön ja aineiston perusteella näiden määrässä ei ole (ajallista) muutosta vaan vaihtelua. Oppikirja-aineisto vältelee kompleksisia verbi-ilmauksia. Muissa suffiksimuodoissa näkyy erityisesti kirjoitelma-aineiston omakohtainen kirjoitustyyli, joka nostaa mm. *itse* ja *oma* -sanojen frekvenssejä (ks. luokka Muu). Vertailun perusteella voi todeta, että vaikka possessiivisuffiksien käytössä voi nähdä muutosta, kyse on ennemmin variaatiosta.

Taulukko 35. Possessiivisuffiksisen sanan esiintymien sanaluokkajakauma korpusaineistoissa.

Aineisto / Sanaluokka %	Adverbi	Substantiivi	Verbi	Muu
1950-luku	11,2	72	11,4	5,3
1970-luku	11,3	71,1	10,0	7,6
1990-luku	11,5	67,9	11,6	8,9
Oppikirjat	13,9	70,5	9	6,4
Fiktio	15,5	66,6	12,3	5,6
Savon Sanomat	16,8	66,2	11,6	5,3
Helsingin Sanomat	17	66,6	11,5	4,8

Lauseopin arkiston 1970-luvun yleiskielen teksteissä possessiivisuffiksit ovat olleet verbimuodoissa yleisempiä (16,3 % esiintymistä) kuin nyky-lehtikielessä, adverbeissa määrät ovat nykyisentapaisia (13,0 %) mutta substantiiveissa vähäisempiä (64,1 %). Ainoa puhutun kielen korpus, josta vastaavia laskelmia voi tehdä, on Lauseopin arkiston murrealueittain koottu korpus, jonka koehenkilöt ovat keskimäärin 1880-luvulla syntyneitä ja useimmiten 1960-luvulla haastateltuja. He siis edustavat aineiston vanhinta kerrostumaa, joskin aineisto on puhuttua kieltä. Suffiksiesiintymät eivät jakaudu sanaluokkiin murrealueittain samalla tavoin: substantiiveja taivutetaan suffiksillisina eniten hämäläismurteissa (67,6 % esiintymistä) ja vähiten savolaismurteissa (48,2 %) aiemmin esitettyjen oletuksien mukaan. Sen sijaan possessiivisuffiksillisia adverbeja ja ei-finiittisiä muotoja on huomattavasti enemmän savolaismurteissa (31,6 % ja 11,8 %) kuin hämäläismurteissa (16,1 % ja 4,9 %). Kokonaissanamäärään suhteutettuna (40 000 murrealueittain) possessiivisuffikseja on eniten juuri hämäläismurteissa (1,05 % sanaesiintymistä) mutta vähiten pohjalais- (0,55 %) ei savolaismurteissa (0,68 %).

Kuvaa saa tarkennettua analysoimalla kirjoitetun aineistoa sanaluokittain ja suffiksittain suhteessa kokonaissanamääriin (ks. taulukot 36–37).<sup>70</sup> Substantiiveissa monikon 1. persoonan käyttö vähenee nuorilla aikuisilla selvästi (eli *meidän maamme* -tyyppinen käyttö vähenee), ja toisen persoonan suffiksi on ylipäätään harvinainen missä tahansa korpuksessa (ks. taulukko 36). Oppikirjateksteissä yksikön 2. persoonan käyttö on paljon yleisempää kuin muissa osa-aineistoissa ja selittyy runsaalla oppilaan puhuttelulla muun muassa tehtävänannoissa (*kirjoita vastaus vihkoosi*). Oppikirja-aineisto poikkeaa muusta aineistosta myös sekä substantiiveihin että verbeihin liitettävän 3. persoonan suffiksin osalta ja tämäkin selittyy tekstilajieroista käsin. Kirjoitelmien substantiiveissa kolmannen persoonan suffiksi on yleinen vaikka laskeekin hieman ajan myötä, mutta kaikkein yleisin se on 2000-luvun fiktiossa. Tämä on alkuoletuksen vastaista, koska fiktio yleensä on lähempänä puhuttua kiel-

<sup>70</sup> Taulukoiden 36–37 laskelmia ei ole tehty HS2000-korpuksesta, koska sen kokonaissanamääriä ei voi laskea samalla tarkkuudella kuin muiden korpusten, eli korpusta vastaavasta tekstiaineistosta.

tä kuin sanomalehtikieli. Sanomalehtikielessä substantiivien 3. persoonan possessiivisuffiksin käyttö näyttää kirjoitelmiin verrattuna selvästi vähäisemmältä. Muiden persoonien osalta tilanne on vaihtelevampi ja taustalla on taas kerran sekä tekstilaji- että tyylieroja.

Taulukko 36. Substantiivien possessiivisuffiksisten esiintymien määrä kokonaissanamäärästä.

Aineisto / Possessiivisuffiksi %	-mme <i>kirjamme</i>	-nne <i>kirjanne</i>	-ni <i>kirjani</i>	-si <i>kirjasi</i>	-nsV/-Vn <i>kirjansa</i>
1950-luku	0,48	0,01	0,39	0,06	1,23
1970-luku	0,37	0,01	0,12	0,05	1,20
1990-luku	0,19	0,02	0,54	0,04	1,17
Oppikirjat	0,03	0,02	0,15	0,32	0,46
Fiktio	0,08	0,05	0,32	0,08	1,68
Lehtikieli	0,09	0,00	0,09	0,02	1,04

Verbien ei-finiittisissä muodoissa possessiivisuffiksit ovat selvästi harvinaisempia kuin substantiiveissa (ks. taulukko 37), vaikka yleisjakauma näyttää samantapaiselta. Erityisesti voi kiinnittää huomiota fiktion: sekä substantiivi- että verbiesiintymissä 3. persoonan suffiksi on yleisin juuri 2000-luvun fiktiossa. Monikon 1. persoonan käyttö vähenee ajallisesti ja yksikön 1. persoonassa on vaihtelua. Oppikirjojen tehtävänannot nostavat myös tiettyjä verbimuotoja (*luettuasi*). Kolmannen persoonan verbimuodoissa suffiksin käyttö vaihtelee.

Taulukko 37. Verbien possessiivisuffiksisten esiintymien määrä kokonaissanamäärästä.

Aineisto / Possessiivisuffiksi %	-mme <i>luettuamme</i>	-nne <i>luettuanne</i>	-ni <i>luettuani</i>	-si <i>luettuasi</i>	-nsV/-Vn <i>luettuaan</i>
1950-luku	0,05	0,00	0,07	0,00	0,26
1970-luku	0,02	0,00	0,02	0,01	0,22
1990-luku	0,03	0,00	0,09	0,00	0,24
Oppikirjat	0,00	0,00	0,01	0,04	0,09
Fiktio	0,00	0,00	0,06	0,01	0,35
Lehtikieli	0,01	0,00	0,02	0,00	0,19

Adverbeissa käytössä on lähinnä yksikön 1. ja 3. persoonan muotoja (ks. taulukko 38). Aineistokohtaiset erot näyttävät hyvin samanlaisilta kuin substantiivien ja verbienkin kohdalla. Ainoa poikkeus on sanomalehti-

aineisto (*Savon Sanomat*), jossa 3. persoonan possessiivisuffiksillisia ad-  
verbeja on käytössä suhteessa enemmän kuin substantiivien ja verbien  
kohdalla. Johtuuko tämä sitten vanhasta murre-edustuksesta, tekstin  
luonteesta vai mistä, jää selvittämättä. 1970-luvun kirjoitelmien tilanne  
ei kuitenkaan paljoo poikkeaa sanomalehtikielestä.

Taulukko 38. Adverbien possessiivisuffiksisten esiintymien määrä kokonaissanamäärästä.

Aineisto / Possessiivisuffiksi %	-mme mukaamme	-nne mukaanne	-ni mukaani	-si mukaasi	-nsV/-Vn mukaansa
1950-luku	0,02	0,00	0,08	0,01	0,32
1970-luku	0,01	0,00	0,02	0,00	0,29
1990-luku	0,02	0,00	0,09	0,01	0,32
Oppikirjat	0,00	0,00	0,04	0,04	0,18
Fiktio	0,01	0,01	0,09	0,01	0,40
Lehtikieli	0,00	0,00	0,02	0,00	0,31

Possessiivisuffiksien käytön mahdollinen muutos voi tulla esiin myös  
sanaluokkien sisäisenä. Substantiivien osalta pitäisi tehdä laadullista  
analyysiä, ja se ei ole aineiston koon vuoksi mahdollista. Verbiluokkaa  
voi sen sijaan analysoida sekä verbimuotojen että niiden funktioiden  
perusteella melko helposti (ks. taulukot 39 ja 40). Tarkastellut raken-  
teet ovat agenttirakenne (*tekemästäni työstä*), jotta-rakenne (*saadakseni  
aikaan*), samanaikaista (= SA) tai eriaikaista (= EA) referointia ilmaiseva  
rakenne (*tiesin kuulevani tai kuulleeni*) ja vastaavasti temporaalirakenne  
(*nähdessäni tai nähtyäni*). Mukana ovat luonnollisesti vain ne rakentei-  
den muodot, joissa possessiivisuffiksi on mahdollinen eli samasubjek-  
tilliset muodot; samasubjektiviisuus oikeastaan tarkoittaa refleksiivistä  
käyttöä. Varhemmin yleiskielessä ja kansanrunoudessa on esiintynyt  
myös erisubjektisiä suffiksillisiä muotoja (*sano sanoneeni, luulin tunte-  
vanne asian*), mutta näistä ei aineistossa ole jälkeäkään (ks. esimerkik-  
si Hakulinen 1979, 566). Tarkastelussa on lisäksi *tekevinään*-rakenne,  
ja kategoriaan Muu on luettu muutama muukin harvinaisempi tyyppi  
(kuten *tekemäisillään*). Käsittelemme erikseen taivutusmuotojen esiinty-  
mien sisäistä jakaumaa (ks. taulukko 39) ja osuutta kokonaissanamää-  
ristä (ks. taulukko 40).

Taulukko 39. Possessiivisuffiksillisten verbiesiintymien aineistokohtaiset jakaumat funktion mukaan.

Aineisto / Verbimuoto %	<i>kuulemani</i>	<i>kuullakseni</i>	<i>kuulevani</i> SA	<i>kuulleeni</i> EA	<i>kuullessani</i> SA	<i>kuultuani</i> EA	Muu
1950-luku	11,7	18,3	14,3	4,1	37,0	10,3	4,3
1970-luku	16,3	18,8	15,5	3,4	30,7	12,0	3,2
1990-luku	14,7	16,9	18,6	4,1	33,9	9,9	1,9
Oppikirjat	31,6	11,1	12,0	2,8	31,6	7,6	3,4
Fiktio	10,5	15,8	22,07	8,9	26,3	12,91	3,5
Lehtikieli (SS)	15,1	11,4	23,8	11,8	23,4	12,4	2,1

Verbiesiintymien jakauma kirjoitelmissa ja fiktiossa (ks. taulukko 39) poikkeaa jonkin verran lehtikielestä. Selvin ero on referentiaalista ja temporaalista samanaikaista tekemistä ilmaisevissa rakenteissa: lehtikielessä referentiaalinen on yleisin, muissa tekstilajeissa temporaalinen. Ero on luonnollinen, koska lehtitekstissä on keskeistä tarve referoida haastateltuja, mutta kertovissa tekstilajeissa suhteutetaan tapahtumia myös ajallisesti. *Jotta*-rakenne on vähäkäyttöisin oppikirjoissa ja lehtikielessä. Agenttipartisiippeja käytetään erityisesti oppikirjateksteissä. Kirjoitelma-aineistosta ei tule esiin selvää ajallista muutosta. Referentiaalinen samanaikaista tekemistä ilmaiseva rakenne lisääntyy ajan myötä, vastaava temporaalinen on lievästi laskusuuntainen, *jotta*-rakenne niin ikään vähenee mutta agenttipartisiipit lisääntyvät. Kun suffiksillisten verbiesiintymien määrä suhteutetaan otoksen sanamäärän (ks. taulukko 40), tilanne on edelleen varioiva. 1950-luvun aineistossa esiintymiä on enemmän kuin 1970-luvun aineistossa mutta 1990-luvun aineisto muistuttaa taas 1950-luvun osuuksia. Vähiten suffiksillisiä verbiesiintymiä on suhteessa sanamäärään lehtikielessä ja etenkin oppikirjoissa.

Taulukko 40. Possessiivisuffiksillisten verbiesiintymien määrä kokonaissanamäärästä.

Aineisto / Ps-verbimuoto %	<i>kuulemani</i>	<i>kuullakseni</i>	<i>kuulevani</i>	<i>kuulleeni</i>	<i>kuullessani</i>	<i>kuultuani</i>
1950-luku	0,05	0,08	0,06	0,02	0,16	0,04
1970-luku	0,05	0,05	0,05	0,01	0,09	0,03
1990-luku	0,06	0,07	0,08	0,02	0,15	0,04
Oppikirjat	0,05	0,02	0,02	0,00	0,05	0,01
Fiktio	0,04	0,06	0,09	0,04	0,11	0,05
Lehtikieli (SS)	0,03	0,02	0,05	0,03	0,05	0,03



Yleistykseenä voi todeta, että mitä henkilökohtaisempi näkökulma kirjoittajalla on tapahtumiin, sitä korkeammalle nousevat samasubjektiset rakenteet. Ero heijastuu suoraan oppikirjateksteihin, joissa lauseenvastikkeita on paitsi vähemmän, myös niiden jakauma poikkeaa muista aineistoista. Tapahtumien saman- tai eriaikaisuus ei sen sijaan niinkään liity näkökulman henkilökohtaisuuteen.

#### EROAVA JA EROAMATON OMISTUS

Possessiivisuffiksien käytön mahdollista muutosta voi tarkastella myös sellaisten sanojen avulla, joiden yhteydessä omistaminen on luonnollista (*inalienable*; eroamaton omistussuhde) tai perustuu juridiseen oikeuteen (*alienable*; eroava omistussuhde). Toinen mahdollisuus olisi ottaa tarkasteluun adverbija ja ottilanilmauksia, joissa possessiivisuffiksit ovat kietyntymässä vaihtelemattomaan muotoon ainakin joissain puhutun kielien varianteissa (*tein tahallani* > *tahallaan*; *olin raivoissani* > *raivoissaan*). Keskitymme tässä edelliseen ryhmään, koska omistussuhdetta ilmaisevia sanoja on koko aineistossa edustavammin kuin kyseisiä adverbeja. Lisäksi monissa lähteissä todetaan, että adverbeissa kongruoimattomia tapauksia on ollut vanhastaan.<sup>71</sup>

Vertaamme eroamatonta omistusta ilmaisevia, ruumiinosia tai sukulaisuhteita nimeäviä sanoja (*käsi, veli*) sanoihin, joiden yhteydessä ilmaistava omistussuhde on eroava (kuten *kirja, talo*), joko juridinen omistussuhde tai hallintasuhde (Heine 1997). Nimitämme näitä sanoja indeksisanoiksi, ja niiden avulla voi esittää hypoteesin, että eroamattoman (luonnollisen) omistuksen yhteydessä possessiivisuffiksi on odotuksenmukaisempi ja yleisempi kuin eroavan (juridisen). Lisäksi jälkimmäisistä on useammin syytä kirjoittaa myös muusta kuin omistamisen tai yhteenkuuluvuuden näkökulmasta. Kontekstissa voi tietysti aina olla kysymys omistussuhteen ilmaisemisesta, jota merkitään tai ollaan merkitsemättä possessiivisuffiksilla, tai sitten muusta käytöstä. Esimerkiksi sanalla *äiti* voidaan viitata omaan äitiin (*äitini*) tai voidaan puhua yleensä äideistä (*äidit ja lapset*).

71 Mark (1925, 242) antaa esimerkkejä vanhasta kirjasuomesta, murteista ja yleiskielestä (esimerkiksi Satakunta: *olin vain muuten vaitonaan*; Agr. *mine se yxinens olen*).

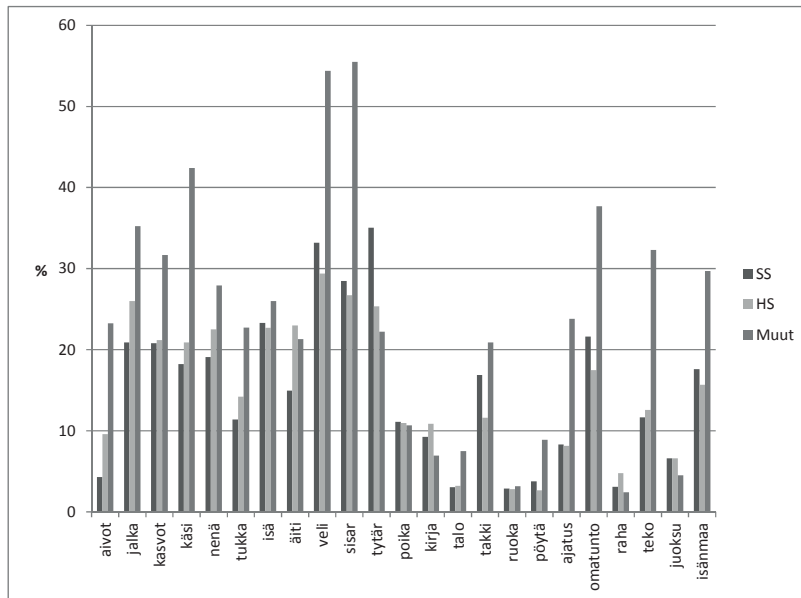
Olemme valinneet indeksisanoiksi etenkin sellaisia sanoja, joita aineiston eri osissa on tilastollisesti riittävästi. Indeksisanat on jaettu omistussuhteen luonteen mukaan neljään ryhmään eli ruumiinosia nimeäviin sanoihin (*aivot, jalka, kasvot, käsi, nenä, tukka*), sukulaissanoihin (*isä, poika, sisar, tytär, veli, äiti*), elottomia olioita nimeäviin sanoihin (*kirja, pöytä, ruoka, takki, talo*) sekä abstrakteihin sanoihin, jotka edustavat sekä eroavaa (*raha, teko, juoksu, isänmaa*) että eroamatonta (*ajatus, omatunto*) omistussuhdetta. Valituilla indeksisanoilla on koko aineistossa yhteensä 118 164 esiintymää, joista possessiivisuffiksillisia on 13 653 kpl (eli 11,5 %). Kun lehtikielessä (HS2000- ja SS2000-korpuksset) substantiivien suffiksiprosentti on keskimäärin 2–3 prosenttia, indeksisanojen valintaa voi pitää onnistuneena. Jaottelemme aineiston ensin lehtikieleen (HS ja SS) ja muihin teksteihin (kirjoitelmat, oppikirjat ja fiktio) ja käsittelemme sitten kirjoitelma-aineistoa myös ajallisesta näkökulmasta. Kirjoitelma-aineisto on indeksisanojen analyysissä mukana kokonaan (kaikki miljoona sanaesiintymää), jotta esiintymiä löytyisi riittävästi. Lopuksi kommentoimme lyhyesti kehityksellistä aineistoa.

Yleisero omistussuhteen laadun mukaan on kummassakin ryhmässä oletuksen mukainen (taulukko 41): possessiivisuffiksi merkitään eroamattoman omistussuhteen yhteydessä merkittävästi useammin kuin eroavan. 2000-luvun lehtikielessä ei ole eri aineistojen välillä suurta eroa. Sekä *Savon Sanomissa* että *Helsingin Sanomissa* suffiksi on suunnilleen yhtä yleinen toisaalta eroamattoman omistussuhteen (18–21 %) kuin toisaalta eroavan omistussuhteen ilmaisuissa (5–7 % esiintymistä). *Helsingin Sanomien* hienoinen yleisempiys johtunee toimituksen tietoisesta oikeakielisyyden vaalinnasta.<sup>72</sup> Fiktiossa, oppikirjateksteissä ja kirjoitelmissa (luokka Muut) tilanne poikkeaa lehtikielestä merkittävästi: sekä ruumiinosasanoissa (ero yli 15 prosenttiyksikköä) että abstraktisanoissa (ero yli 10 prosenttiyksikköä) possessiivisuffikseja on merkittävästi enemmän. Eroamattoman omistuksen osalta erot ovat pienet.

72 Huomaus perustuu 2000-luvun alussa HS:n toimituspäällikön antamaan tietoon, jonka mukaan toimituksen aamupalaverissa käsiteltiin päivittäin myös oikeakielisyyttä.

Taulukko 41. Indeksisanojen possessiivimuotojen määrä esiintymäluokittain.

Aineisto / Ps-indeksisanat %	Ruumiinosa <i>kätteni</i>	Sukulaisuus <i>isäni</i>	Eloton <i>kirjani</i>	Abstrakti <i>mieleni</i>
Savon Sanomat	18,2	18,8	5,7	6,8
Helsingin Sanomat	21,6	19,8	6,3	7,0
Muut	35,6	21,4	7,5	17,9



Kaavio 11. Indeksisanan possessiivimuotojen osuus sanan esiintymistä korpuksittain.

Jos käyttöä katsotaan indeksisanoittain (kaavio 11), nähdään, että erot eivät synny aivan tasaisesti, vaan joidenkin sanojen käyttökonteksti suosii possessiivisuffiksia yli omistussuhteen luonteen tai semanttisen luokan vaikutuksen. Esimerkiksi sanasta *takki* omistussuhde ilmaistaan useammin kuin muista elottoman luokan indeksisanoista. Sanoista *veli*, *sisar* ja *tytär* omistussuhde ilmaistaan useammin kuin sanoista *isä* ja *äiti*, *poika*-sanasta puhumattakaan. Kuten jo edellä nähtiin, muissa teksteissä omistussuhteen ilmaiseminen on yleisempää kuin lehtitekstissä mutta indeksisanoittainen tarkastelu kuitenkin osoittaa, että tämä ei päde kai-

kista yksittäisistä sanoista. Erot ovat hyvin pienet sanoissa *kirja, ruoka, juoksu, raha* ja *poika*, sanoissa *ruoka* ja *raha* omistussuhde ilmaistaan kaikissa osakorpuksissa harvoin. Abstraktisanoissa lehtikielen ja muiden tekstien suurimmat erot syntyvät sanoista *ajatus, omatunto, teko* ja *isänmaa*, mistä voi päätellä, että muissa teksteissä tyylilaji on lehtikieltä henkilökohtaisempi ja yksilöivämpi. Kaikissa ruumiinosasanoissa omistussuhde ilmaistaan muissa teksteissä yleisesti, ja ero lehtikieleen on hyvin selvä. Sukulaissanoista *veli* esiintyy muissa teksteissä useammin suffiksillisena kuin ilman, sanan *sisar* kokonaisfrekvenssi on oikeastaan liian matala johtopäätöksiin.

Mahdollista ajallista muutosta tarkastelemme luokasta Muu, jota edellä käsiteltiin kokonaisuutena. Nyt erottelemme ne aiempaan tapaan nuorten aikuisten kirjoitelmiin sekä oppikirjateksteihin ja fiktion. Kirjoitelmat on jaettu kolmista siten, että vanhimmassa osiossa on vuosien 1940–1960 kirjoitelmia, toisessa vuosien 1970–1980 ja uusien aineisto koostuu vuosien 1990–2003 kirjoitelmista. Indeksisanojen frekvenssi vaihtelee näissä osiossa 2 000–3 000 välillä. Fiktio- ja oppikirjateksteissä frekvenssi on 1 500–2 000.

Taulukko 42. Indeksisanojen possessiivimuotojen määrä eri aineistoissa.

Aineisto / Ps-indeksisanat %	Ruumiinosa <i>kätteni</i>	Sukulaisuus <i>isäni</i>	Eloton <i>kirjani</i>	Abstrakti <i>mieleni</i>
Kirjoitelmat 1940–1960	38,0	24,1	8,7	21,6
Kirjoitelmat 1970–1980	32,7	24,4	8,8	21,9
Kirjoitelmat 1990–2003	29,9	26,1	3,8	21,1
Oppikirjat	29,3	8,9	5,5	1,2
Fiktio	46,3	28,4	11,5	14,7

Kirjoitelmissa indeksisanojen suffiksimmäät eivät vaihtelee odotuksenmukaisesti, toisin sanoen suffiksien käyttö ei laske suoraviivaisesti 1940-luvulta 2000-luvulle, kuten puhutun kielen ennusteet antaisivat odottaa (ks. taulukko 42): yleisyys pysyy vuosikymmenestä toiseen aika samanlaisena sukulaissanoissa ja abstraktisanoissa, sukulaissanoissa on hienokseltaan jopa nousua ajan myötä. Eroavan omistussuhteen yhteydessä suffiksin käyttö on vähentynyt uusimmissa kirjoitelmissa, mutta esiintymiä ei ole riittävästi luotettavien päätelmien tekemiseksi. Ruu-

miinosanimien yhteydessä käyttö on laskenut tasaisesti. Kun kirjoitelmia verrataan fiktion, käy ilmi, että suffikseja käytetään juuri fiktiassa eniten kaikissa muissa ryhmissä paitsi abstraktisanoissa. Myös oppikirjatekstit toistavat jo aiemmin esiin tulleita eroja: suffikseja käytetään vähiten oppikirjateksteissä 2. persoonaa lukuun ottamatta, oli tarkastelunäkökulma mikä tahansa.

Jos possessiivisuffikseja arvioidaan kehityksellisestä aineistosta eli lasten ja nuorten aikuisten 2000-luvun Unelmakirjoitelmista (ks. Pajunen, Puranen & Yli-Paavola 2010; Pajunen 2016), tilanne näyttäytyy entisenlaisena sikäli, että suffiksien käyttö ei ole ennusteiden mukaisesti loppunut. Käyttö on edelleen runsaampaa eroamattoman kuin eroavan omistussuhteen yhteydessä (ks. taulukko 43) ja se lisääntyy kehityksellisesti. Unelma-aineistossa indeksisanojen esiintymiä on yhteensä 3 239 ja näistä possessiivisuffiksillisia on noin 20,2 prosenttia eli enemmän kuin kirjoitelmissa. Luvut ovat kuitenkin hieman harhauttavia, koska indeksisanoja on koko aineistossa suhteellisen vähän ajalliseen aineistoon verrattuna. Suffiksilliset sanat jakautuvat Unelma-aineistossa taulukon 41 indeksisanoihin hieman eri tavalla ja eroavuutta on myös indeksisanojen luokittaisissa frekvensseissä. Abstraktisanojen esiintymämäärä ei riitä arvioitavaksi: esimerkiksi *isänmaa* tai *omatunto* eivät liity lainkaan 2000-luvun unelmakontekstiin. Myös ruumiinosasanojen frekvenssit ovat vähän liian matalat, jotta tulos olisi aikuisia lukuun ottamatta luotettava. Pääasiassa sanat *käsi* ja *jalka* taipuvat suffiksillisina. Sen sijaan sukulaissanastoa on paljon (esiintymiä yht. 1 600) ja eniten niitä on nuorimmilla alakouluikäisillä. Tämä johtuu siitä, että lapsilla unelmat keskittyvät perhepiiriin, ja unelmia toteutetaan yhdessä isän, äidin ja sisarusten kanssa. Aikuisilla sukulaissanasto hajaantuu ja mainintoja on paljon vähemmän.

Taulukko 43. Indeksisanojen possessiivimuotojen määrä kehityksellisessä aineistossa eli Unelmakirjoitelmissa.

Unelma-aineisto / Ps-indeksisanat %	Ruumiinosa <i>kätäni</i> %	Sukulaisuus <i>isäni</i> %	Eloton <i>kirjani</i> %
Alakoulu 1.–5.	18,8	11,4	2,7
Alakoulu 6.	52,6	29,7	8,4
Aikuiset	58,1	61,1	10,3

Unelma-aineistoa analysoitaessa täytyy ottaa huomioon, että se on tuotettu jossain määrin muodollisessa kontekstissa joko oppitunnilla tai jossain oppilaitoksessa. Kielenhallinnan kannalta muodollisuus ei haittaa, vaan aineisto osoittaa paremmin, että nykylapset, nuoret ja nuoret aikuiset vielä hallitsevat possessiivisuffiksien käytön käyttöön pyrkiesään.<sup>73</sup> Suffiksin tuottaminen voi olla tavoitteena silloinkin, kun sanan taivutus ei onnistu (ks. esimerkki 1). Suffiksimuotojen käytön runsaus Unelmakirjoitelmissa liittyy referentiaalisten määräisten ilmauksien käyttöön ja tämän käytön kehittymiseen, kuten Kouvon ja muiden analyysistä tässä teoksessa käy ilmi. Kun nuorimmilla niin sanottu tuttu tieto esitellään ensimmäisen kautta (*äiti*), vanhemmilla se tehdään lisääntyvästi uudelleen maininnan kautta ja tarkenteellisena (*äitini*). Aineistosta näkyy kuitenkin hyvin, että suffikseja ei käytetä mekaanisesti. Jos esimerkiksi kirjoittajan äiti on jo esitelty lukijalle, häneen ei jatkuvasti viitata possessiivisuffiksilla. Sana *äiti* on perhepiirissä myös äidin nimi, mikä näkyy muun muassa taivutuksen – ja suffiksin – puuttumisesta (esimerkiksi *äitin*). Kirjoittaja, jonka isä tekstissä paljastuu isäpuoleksi, viittaa jatkossa oikeaan isäänsä aina possessiivisuffiksillisena. Samanlaista informaatiota välitetään lukijalle myös, kun puhutaan vaikkapa yleensä ruoasta ja toisaalta omasta lempiruoasta (*lempiruokani*), omasta työpöydästä (*työpöytäni*) jne. Kun esineeseen on omistussuhde tai sen tarkoitteeseen viittaussuhde ja suhdetta halutaan korostaa, sana taipuu suffiksillisenä.

Kirjoittamiskonteksti ja kirjoittamistehtävä vaikuttavat possessiivisuffiksien käyttöön ja käyttötapaan paljon. Myös kehityksellinen aineisto viittaa suffiksien käytön kontekstuaaliseen variaatioon.

1) *Mukana oli kaverini Miina ja isovelini Pete.* (3F15)<sup>74</sup>

73 Vastaava yleiskieleen pyrkiminen näkyy puhekielisen aineksen suhteellisena vähytenä Unelma-aineistossa (ks. Salonen 2012). Ne nuorimmat alakouluikäiset, jotka jo saavat tuotettua kertovaa tekstiä, käyttävät mm. subjekti–predikaattikongruenssia runsaasti, myös monikon 1. persoonassa. Murrosikä muuttaa tilannetta joksikin aikaa.

74 Tunniste luetaan seuraavasti: 3 = luokka-aste, F = tyttö, 15 = kirjoitelman tunniste korpuksessa.

Indeksisana voi kontekstissa liittyä omistussuhteen ilmaisuun myös ilman suffiksia. Oppikirjateksteissä ruumiinosia tarkastellaan useimmin elollisen ja inhimillisen rakenteeseen kuuluvina ja silloin kielenös on enemmän osa–kokonaisuussuhteisena (*ihmisellä on neljä raajaa, ihmisen aivot*) kuin possessiivisuffiksillisena omistussuhteena (*hänen aivonsa*). Muissa aineiston osissa omistussuhde ilmaistaan hyvin usein erisnimen tms. genetiivillä (*X:n käsi*). Aineistossa on myös sanontoja (*kengät jalassa, tallustella kahdella jalalla*) ja idiomeja (*jääädä jalkoihin*), joita käytetään ilman suffiksia. Omistussuhteen ilmaisuun riittää useimmiten pelkkä possessiivisuffiksi, mutta viittaussuhteesta riippuen kyseen tulee myös persoonanpronominin ja suffiksilla merkityn pääsanan yhdistelmä (*minun kirjani*). Tämä yhdistelmä on puhutun kielen varassa tehtyjen muutosennusteiden perusteella erityisen mielenkiintoinen, ja ennuste on, että suffiksi jää pois.

Kirjoitelma-aineistosta löytyy kaikkiaan 321 esimerkkiä persoonapronominin (*minun/meidän, sinun/teidän, hänen/heidän*) ja indeksisanan yhdistelmästä, ja näistä viidestä odotuksenmukainen suffiksi puuttuu puhutun kielen ennusteen mukaan (ks. 2a–e). Kolme ensimmäistä esimerkkiä on tyyppiä *meidän x* (2a–c) eli edustaa käyttötapa, joka tunnetaan jo vanhastaan suffiksittomana. Vanhimmissa aineiston osissa tosin on käytössä *meidän x:mme*-tyyppi. Esimerkiksi *maa*-sanana esiintymistä on suffiksillisia vanhimmissa kirjoitelmissa 44,6 prosenttia ja uusimmissa enää vain 13,6 prosenttia (vrt. taulukot 32 ja 33). Esimerkki 2d) vaikuttaa heikosti muodostetulta ja voi olla, että *oma*-sana on hankaloittanut rakenteen tuottoa. Lähinnä esimerkki 2e) edustaa puhutun kielen *mun tukka* -tyyppiä. Oireellisesti esimerkki on uusimmasta kirjoitelma-aineistosta. Yhden esiintymän todistusvoima on tosin satunnaisen luokkaa.

- 2a) ”Kyllä **meidän pojasta** lääkäri tulee, kunhan vain yrittää.”  
(Kirjoitelmat 1976, 647220)
- 2b) *Epäonnistumisen itkulle ei meidän talossa ollut tilaa* – –.  
(Kirjoitelmat 1996, III4870)
- 2c) *Kyllä meidän äitikin osaa ajaa, Tony uhosi.* (Fiktio)

2d) – – *nollia kuin hänen omissakaan rahoissa.*

(Kirjoitelmat 1971, 468457)

2e) *Loppua kohden hänen ajatukset ovat muuttuneet: – –.*

(Kirjoitelmat 1993, 1040062)

#### MUUTOKSIA TUNNUKSEN VALINNASSA

Tarkastelemme lopuksi vielä muutoksia possessiivisuffiksin tunnuksen tyypissä oletuksena, että tunnuksen lyheneminen voi helpottaa suffiksi-muodon kiteytymistä. Siten muoto *mielellänsä* tuntuisi henkilöviitteisemmältä kuin muoto *mielellään* ja jälkimmäinen on helpompi yhdistää 1. persoonaan (*tekisin sen mielellään*) kuin edellinen (*tekisin sen ?mielellänsä*). Tiedetään myös, että possessiivisuffiksien tunnusvariaatio on ollut vanhastaan hyvin suurta ja niiden kehityshistoria on pitkä.

Uralilaiskielissä erotetaan yleensä sekä absoluuttinen että possessiivinen deklinaatio, jotka eroavat paitsi sijojen määrässä myös taivutustunnuksissa. Possessiivisen taivutuksen tunnuksot ovat olleet yhteydessä persoonataivutukseen (Itkonen 1955), lisäksi possessiivisessä taivutuksessa oletetaan vanhastaan erotetun sekä omistajan että omistetun useus eri tunnussarjoilla (ks. Mark 1925; Itkonen 1955). Suomessa possessiivisen taivutuksen yksinkertaistuminen analogian ja kontaminaation tietä on yhdistänyt tunnuksia ja vähentänyt annettuja merkityksiä. Esimerkiksi iitin murteen lisäksi vanhassa kirjasuomessa näkyvä possessiivitaivutuksen vahva aste astevaihtelusanoissa liittyi yksikölliseen omistajaan ja heikko aste monikolliseen (Mark 1925, 243–244; Rapola 1965, 103–109). Enää astevaihtelua ei ole. Eri persoonissa tunnuksen valinta on varioinut aiemmin paljon etenkin puhutussa kielessä. Nykykielessä variaatio on pitkälti tasoittunut mutta näkyy edelleen sekä puhutun että kirjoitetun kielen kolmannen persoonan suffiksista, jonka tunnus on joko *nsA* tai vokaalin pidennys ja *n* (*Vn*). Markin (1925, 205) mukaan murteissa on ollut neljä eri 3. persoonan suffiksin tunnusta: a) *\*-ns3k* (eli murteissa mm. *-s ~ -ns ~ -sA ~ -sAk ~ -sA ~ -se*), b) *\*-ns3n* (mm. *-sen ~ -se ~ -sAn ~ -sAAn ~ -nsAn*), c) *\*-h3n* ja d) *\*-h3k*. Näistä b ja d ovat levikiltään kapea-alaisia mutta a ja c ovat esiintyneet (ja esiintyvät) (lähes) koko kielialueella; lähinnä (eräissä) lounaisissa murteissa käytös on vain a-tunnus eli *ns > -s* (ks. Mark 1925, 206).



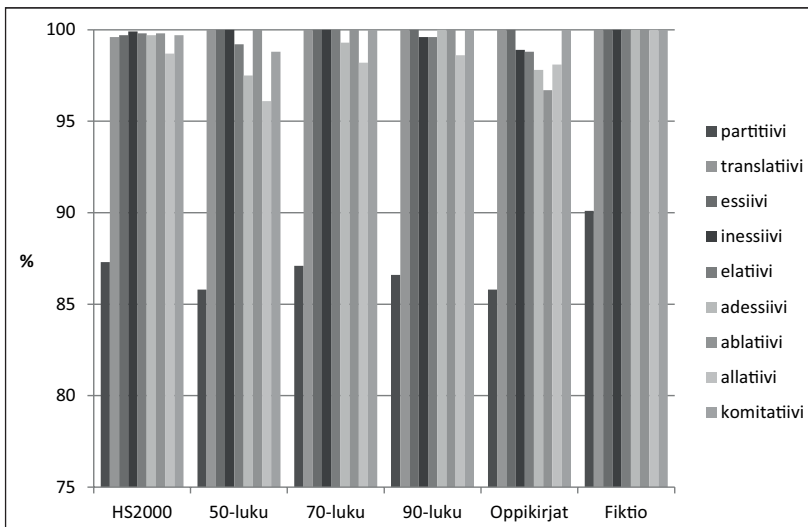
Tunnuksilla ei enää ole merkityseroa. Ne olivat komplementaarisesa distribuutiossa sekä 1800-luvulla että myöhemmin siten, että *nsA* esiintyi nominatiivissa, genetiivissä, akkusatiivissa ja illatiivissa (sekä aiemmin myös allatiivissa) eli lähinnä *n*-päätteisissä muodoissa, muissa sijoissa molemmat ovat lyhyen vokaalin jälkeen mahdollisia. Tunnuksien alkuperä on edelleen jossain määrin tulkinnallinen. Markin (1925, 31–53) mukaan Ahlquist esitti 1863, että *hän*-pronominiin palautuva *Vn* on alkuperäisempi ja *nsA* palautuu siihen. Budenz sen sijaan argumentoi 1869 ns. *n*-koaffiksiteorian puolesta, jonka mukaan *nsA* on monikollinen muoto (monta omistajaa) ja *hAn* (< \**sAn*) yksiköllinen (yksi omistaja). Sittemmin 1890-luvulla sekä Setälä että Paasonen puolisivat koaffiksiteoriaa, ja se tuntuu edelleen olevan muodossa tai toisessa hyväksytty. Szinneyi esitti 1900-luvun alussa, että kaikissa suomalais-ugrilaisissa kielissä olisi kaksi suffiksisarjaa omistajan määrän mukaan ja juuri *n*-elementti merkitsisi useutta. Itkonen (1955) oletti saamelaiskielten perusteella, että *n*-aines palautuu vanhaan duaalin tunnuksen, toisaalta Rapola (mp.) puolsi *k*-tunnusta monikon merkkinä.

Ei ole tiedossa, milloin sarjat olisivat suomessa langenneet yhteen, mutta niiden variaatio on edelleen olemassa joskin tasoittunut sanavartalon ja sijan mukaan. Lauri Hakulinen (1979, 112) katsoo, että vanha kirjasuomi nojasi yksin läntiseen *nsA*-tunnukseen – tosin huomioimatta, että tunnukset olivat myös länsimurteissa rinnakkaisia – ja että 1800-luvun alusta rinnalle olisi noussut muista murteista omaksuttu *Vn* niissä sijoissa, jotka päättyivät lyhyeen vokaaliin. Tämä tunnus olisi 1800-luvun puolivälin jälkeen yleistynyt. Häkkisen (1994, 221–224) mukaan vanhassa kirjasuomessa tunnus oli \**nsA(k)* ja itämurteissa oli käytössä sen variantti *sA*. Vhaelin 1700-luvun kieliopissa *Vn* liitetään savolaiseen *hVn*-tunnukseen – joka Markin mukaan kuitenkin oli laaja-alainen – ja Beckerin kieliopissa vuodelta 1824 tyyppi mainitaan rinnan *nsA*-tunnuksen kanssa, jota kuitenkin vielä pidetään ensisijaisena. Collanin kieliopissa 1840-luvulla *Vn* jo todetaan tavalliseksi sekä puhutussa että kirjoitetussa kielessä (ks. Häkkinen mp.). Nykynäkemyksen mukaan variaatiotilanne olisi ollut jo kauan ohi ja tunnukset olisivat komplementarisessa distribuutiossa sananlopun äännesuhteiden mukaan.

Lehtikielessä tunnuksien esiintymien jakauma on sekä substantiivies-

sa että verbeissä jonkin verran *nsA*-tunnuksen hyväksi (noin 55:45). Molempia tunnuksia käytetään produktiivisesti, useiden eri lekseemien yhteydessä ja monissa eri funktioissa. Sekä substantiivien että verbien osalta tunnukset ovat osaksi komplementaarisessa distribuutiossa, osaksi vapaassa vaihtelussa. Substantiivien osalta *nsA* on pakollinen nominatiivi-, genetiivi- ja illatiivisijoissa sekä partitiivisijassa *aa*-kontekstissa, joka muodollisesti usein lankeaa yhteen vastaavan illatiivimuodon kanssa (*hoitaa tehtävää(nsä)*, *uppoutua tehtävää(n/sä)*) ja jossa *Vn* ei siten ylipäätään ole mahdollinen. Muissa partitiiviesiintymissä tunnukset ovat vapaassa vaihtelussa (*lentokoneitansa ~ lentokoneitaan*). Kaikkien muiden sijojen osalta esimerkiksi Hakulinen, Vilkuna, Korhonen, Koivisto, Heino & Alho (2004, 125) katsovat, että *Vn* on ainoa vaihtoehto. Verbien osalta referentiaalisen lauseenvastikkeen ainoa mahdollinen tunnus on *nsA*, muissa verbimuodoissa variaatio on mahdollista, joskin epätodennäköistä. Agenttipartisiipin sijamuodot ja niin ikään possessiivisuffiksin tunnukset vaihtelevat kontekstin mukaan.

Partitiivisijassa tunnus vaihtelee odotuksenmukaisesti kaikissa osakorpuksissa (ks. kaavio 12). Se on *nsA* 10–15 prosentin todennäköisyydellä, mutta näistä vain muutama prosentti on aitoja variaatiotapauksia. Esimerkiksi HS2000-korpuksessa seuraavat esimerkit ovat vapaassa vaihtelussa: *ammattitaitoansa ~ ammattitaitoaan*, *ansioitansa*, *asioitansa*, *etujansa*, *hintojansa*, *kaveriansa*, *kohtaloansa*, *kotejansa*, *lastansa*, *nuortansa*, *peukaloansa*, *teitänsä*, *tietänsä* jne. Muissa sijamuodoissa eli essiivissä sekä sisäisissä ja ulkoisissa paikallissijoissa illatiivia lukuun ottamatta vapaa vaihtelu on edelleen mahdollinen mutta partitiivia vielä marginaalisempaa (translatiivi: *osoitteeksensa*; essiivi: *kuolemanansa*; inessiivi: *luolassansa*; elatiivi: *urastansa*, *työtaiidoistansa*, *juuristansa*; adessiivi: *tahollansa*, *pihoillansa*; ablatiivi: *maaltansa*; allatiivi: *lainoillensa*, *olallensa*; komitatiivi: *mielipiteinensä*, *pannuinensa*, *tuotteinensa*). Ajallisesti arvioiden ainakin nuorten kirjoitelmissa variaatio on vähentynyt: 1950-luvun tekstien paikallissijoissa on muutaman prosenttiyksikön verran variaatiota, joka 1990-luvulle tultaessa on lähes hävinnyt. Yllättäen vaihtelua on lähes 1950-luvun tapaan alakoulun oppikirjoissa mutta ei lainkaan 2000-luvun fiktiossa. Verbimuodoissa vapaa vaihtelu on satunnaista (*eläessänsä*; *puhdistaaksensa*).



Kaavio 12. 3. persoonan *Vn*-tunnuksen osuus vaihtelusta sijoittain.

Nykyvaihe näyttää siltä, että tunnusten täydennysjakauma on selkiytynyt, mutta etenkin partitiivissa ja paikallissijoissa on jonkin verran tunnusten vapaata vaihtelua. Eniten yleiskuvasta poikkeaa 1950-luvun aineisto, mutta koska lähes yhtä paljon vaihtelua on 2000-luvun oppikirja-aineistossa, ei voi todeta, että vapaa vaihtelu olisi kokonaan hävinnyt. Partitiivi on ainoa sija, jonka yhteydessä on sanamuodoittainen täydennysjakauma. Sananmuotoesiintymien ja sanaesiintymien perusteella laskien *nsA*-tunnus on yleisempi kuin *Vn*-tunnus, vaikka *Vn* esiintyy useammassa sijassa. Verbimuodoissa tunnukset ovat komplementaarissa distribuutiossa ja vapaata vaihtelua ei käytännössä enää ole (agenttipartisiippia lukuunottamatta). Täydennysjakauma ylläpitää kahta tunnusta, mutta verbien osalta voi olla, että tunnuksen funktio ennen pitkää käy epäselväksi. Tämä on sikäli epätoivottavaa, että tunnus erottelee sama- ja erisubjektisen referentiaalisen rakenteen (*kuuli sanovansa / (heidän) sanovan*).

Lopuksi voi todeta, että possessiivisuffiksien tunnusten variaatiota koskeva muutos on yksi harvoista suomen kielessä tapahtuneista muutoksista, jonka kesto voidaan arvioida vuosissa: variaation väheneminen ja komplementaarisen distribuution synty on vienyt yli 200 vuotta.

## YHTEENVETO JA JOHTOPÄÄTÖKSET

Olemme tarkastelleet possessiivisuffiksien käyttöä sekä kehitystä eri aikoina ja eri tekstilajeissa. Tutkimuksemme osoittaa, että possessiivisuffiksin käyttö on tarkasteltavina vuosikymmeninä kirjoitetussa kielessä samankaltaista eikä suuria muutoksia ole nähtävissä. Sen sijaan eri tekstilajien välillä voidaan havaita selkeitä eroja.

Lehtikielessä kolmannen persoonan suffiksi on vallitseva vuosikymmenestä toiseen, kun taas nuorten kirjoitelmissa, oppikirjoissa ja fiktiossa sen osuus on 15–20 prosenttiyksikköä pienempi. Oppikirjoissa korostuu yksikön toisen persoonan tunnus erilaisten tehtävänantojen takia, kirjoitelmissa puolestaan korostuu omakohtaisuus ensimmäisen persoonan muodossa. Sanaluokkajakauma paljastaa, että noin kaksi kolmasosaa kaikista suffiksillisista esiintymistä on substantiiveja, runsaat 10 prosenttia adverbija ja noin 10 prosenttia verbejä. Sanaluokkien sisällä on havaittavissa pientä laskua substantiiveissa ja vastaavasti pientä nousua adverbissa. Suffiksilliset verbit puolestaan liittyvät kompleksiseen kielenkäyttöön ja tekstilajien välillä voi havaita vaihtelua siten, että esimerkiksi oppikirjat välttävät kompleksisia verbi-ilmauksia. Sanaluokittainen tarkastelu osoittaa, että substantiiveissa monikon ensimmäisen persoonan käyttö vähenee selvästi. Yksikön toinen persoona on kauttaaltaan harvinainen lukuun ottamatta oppikirjoja. Yksikön kolmas persoona on yleisin 2000-luvun fiktiossa. Sekä verbeissä että adverbissa yleisjakauma näyttää samantapaiselta ja yksikön kolmannen persoonan suffiksi on jälleen yleisin fiktiossa. Verbiluokkien analyysi funktion mukaan osoittaa, että suffiksia esiintyy eniten finiittiverbin kanssa samanaikaista tekemistä ilmaisevissa rakenteissa. Myös tekstilajierot ovat jälleen esillä: referentiaalirakenne on yleisin sanomalehtikielessä ja temporaalirakenne muissa. Selvää ajallista muutosta ei ole havaittavissa. Kehityksellisesti suffiksit koululaisilla lisääntyvät.

Indeksisanojen analyysi osoittaa, että odotuksenmukaisesti suffiksi on yleisempi eroamattoman omistussuhteen yhteydessä kuin eroavan. Selkeää ajallista muutosta ei ole taaskaan havaittavissa vaan suffiksin yleisyys on vuosikymmenestä toiseen melko samanlainen, sukulaissanoissa voidaan havaita pientä nousua. Sen sijaan ruumiinosaa nimeävien sanojen yhteydessä käyttö on laskenut tasaisesti. Erittäin mielen-

kiintoisia puhutun kielen tutkimusten perusteella tehtyjen ennusteiden kannalta ovat tapaukset, joissa omistussuhteen ilmaisuun riittää pelkkä persoonapronomini, sillä ennusteiden mukaan suffiksi jäisi tällaisissa tapauksissa pois. Täysin ennusteiden vastaisesti 321:n esimerkin joukosta löytyy kuitenkin vain 5 tapausta, joissa suffiksi jää pois, ja vain yksi näistä on niin sanottu aito tapaus. Tämä luonnollisesti tarkoittaa, että melko muodollisessa kontekstissa kirjoittavat ovat vielä pystyneet käyttämään possessiivisuffikseja puhutun kielen konventioista poiketen. Epämuodollisemmissa konteksteissa he hyvinkin toimisivat toisin. Viimeiseksi analysoimme vielä yksikön kolmannen persoonan tunnuksen vaihtelua. Tuloksena on, että tunnusten täydennysjakauma on selkiytynyt. Vapaata vaihtelua esiintyy vain partitiivissa ja paikallisijoissa mutta vähän. Sanamuotoesiintymien perusteella *nsA*-tunnus on yleisempi, vaikka *Vn*-tunnus esiintyykin useammassa sijassa.

Puhutun kielen tutkimusten sekä kirjoitetussa kielessä tehtyjen huomioiden ja esimerkkien pohjalta tehtyjen ennusteiden perusteella possessiivisuffiksien käytössä tulisi näkyä selkeitä vähenemisen merkkejä. Tutkimuksemme ei tue näitä havaintoja. Täten esimerkiksi Ikolan (1992) ennuste suffiksien vähenemisestä lehtikielessä kumoutuu. Eritäin mielenkiintoinen tulos on myös se, että indeksisanoissa suffiksin käyttö on täysin ennusteiden vastaista, koska suffiksittomuuden leviämisen yhdeksi syyksi on esitetty sitä, että se vanhastaan puuttuu jo sellaisissa tapauksissa, joissa omistusmuoto ilmaisee perhettä, yhteisöä tai asuinpaikkaa. On siis mahdollista, että puhutussa kielessä havaittu muutostendenssi ei ole vielä ehtinyt levitä (muodolliseen) kirjoitettuun kieleen tai että se on hidastunut. Koska havaintoja suffiksittomuuden lisääntymisestä on jo 1960-luvulta, olisi muutoksen nopeus tässä tapauksessa huomattavasti hitaampi kuin mitä on esitetty puhutun kielen osalta. Toinen vaihtoehto on se, että tehdyt ennusteet perustuvat yksittäisiin esimerkkeihin ja niistä liian pitkälle vedettyihin johtopäätöksiin. On täten mahdollista, että muutos ei tule etenemään kirjoitettuun kieleen ainakaan sellaisenaan.

Tutkimuksemme osoittaa myös sen, että suffiksin käyttöä ja käytön kehitystä ei voi analysoida pelkästään yhden tekstilajin perusteella, sillä suffiksin käyttö vaihtelee sekä teksti- että tyylilajeittain. Suffiksin käyttöön

vaikuttaa myös sen funktio joko omistuksen tai referentiaalisen viittauksen ilmaisimena. Aineiston perusteella voi myös yleistää, että suffiksit ovat yleisempiä kertovassa kuin erittelevässä, arvioivassa tekstityypissä. Tulos antaa aihetta olettaa, että samankaltaisia eroja voi havaita muisakin kielenpiirteissä. Ajallisen muutoksen tutkimuksissa on siis huomioitava niin tekstilajien erot kuin tyylilliset erot, ainakin kirjoitetun kielen kohdalla. Tämän havainnon on aiemmin tehnyt mm. Biber (1998) kirjoitetun ja puhutun kielen välistä suhdetta käsittelevässä tutkimuksessaan. Mielenkiintoinen kysymys on, voiko tuloksesta vetää puhuttua kieltä koskevia johtopäätöksiä. Puhutussa kielessä ei voi erottaa tekstilajeja samalla tavalla kuin kirjoitetussa kielessä, mutta kuten Hakulinen (2003) toteaa, puhuttu kielikään ei ole yhtenäinen kielimuoto.

Yhdeksi possessiivisuffiksin käytön vähenemisen syyksi on esitetty sen redundanttisuus tietyissä asemissa, mihin tässäkin tutkimuksessa on viitattu. Toisaalta samassa yhteydessä on yleiskommenttina todettu, että possessiivijärjestelmän kaltaiset kongruenssisysteemit eivät edusta redundanttia tietoa, koska ne nopeuttavat prosessointia. Tätä tulisi ehkä pohtia laajemminkin: ovatko redundantteiksi ajatellut kielenpiirteet todella redundantteja? Ainakin muutoksen tutkimuksen kannalta tämä lienee olennainen kysymys. Esimerkiksi Lindén (1959) tuo esiin selvyyden, suhteiden täsmentämisen ja vahvistamisen sekä ilmaisuvoiman selittävinä tekijöinä pohtiessaan esimerkkien keinoin tapauksia, joissa suffiksi epäodotuksenmukaisesti joko jätetään pois tai liitetään mukaan. Tällaisissa tapauksissa olisi ehkä syytä pohtia myös prosessoinnin vaikutusta.

## Lopuksi

Tässä tutkimuksessa on tarkasteltu meneillään olevaa kielenmuutosta kirjoitetussa kielessä yhtäältä selvittämällä kirjoitetun ja puhutun kielen välistä suhdetta ja toisaalta analysoimalla possessiivisuffiksin käyttöä ja kehitystä eri vuosikymmeninä korpusten avulla. Kirjoitetun ja puhutun kielen välisen suhteen teoreettisessa tarkastelussa on korostettu kirjoitetun ja puhutun kielen samanlaisuutta ja argumentoitu sellaisia

näkemyksiä vastaan, jotka ylikorostavat niiden erilaisuutta. Possessiivisuffiksin korpusanalyysi puolestaan tuo esiin kirjoitetun ja puhutun kielen erilaisuuden, joka tosin kietoutuu niiden samanlaisuuden ympärille. Tämä voidaan ilmaista selkeämmin esimerkiksi toteamalla, että sekä kirjoitetussa että puhutussa kielessä käytetään possessiivisuffikseja (= samanlaisuus) mutta että niitä saatetaan käyttää hieman eri tavalla (= erilaisuus).

Tutkimuksemme tulokset eivät tue ennusteita, joiden mukaan possessiivisuffiksi olisi katoamassa suomen kielestä. Ikolan (1992) ennuste suffiksin käytön vähenemisestä ei tutkimuksemme mukaan osu oikeaan. Sen sijaan Paunosen (1995) ennusteeseen suffiksin katoamisesta Helsingin puhekielestä tutkimuksellamme ei voi ottaa suoraan kantaa. On kuitenkin vaikea kuvitella, että jos suffiksia käytetään jatkossa edes suurin piirtein entiseen tapaan esimerkiksi lehtikielessä, se tulisi kokonaan katoamaan puhekielestä, varsinkin kun otetaan huomioon tapaukset, joissa suffiksin asema on vahvempi.

Cameron (1997 [1990]) kritisoi sosiolingvistien tapaa selittää kielenpiirteiden ja kielenulkoisten tekijöiden välisiä korrelaatioita näkemyksellä, jonka mukaan kieli heijastaa yhteiskuntaa. Vaikka Cameronin varsin aiheellinen kritiikki kohdistuu ensi sijassa selittämiseen, voidaan myös itse näkemystä siitä, että kieli heijastaa yhteiskuntaa, kritioida. Voidaan esimerkiksi esittää, että kieli ei todellisuudessa heijasta yhteiskuntaa vaan on pikemminkin osa sitä, minkä seurauksena kieltä ei välttämättä tulisi tai voida erottaa yhteiskunnan rakenteista. On totta, että kieli ei ainoastaan heijasta yhteiskuntaa vaan myös vaikuttaa siihen takaisin, mutta heijastamisen kieltäminen on liioittelua. Otetaan esimerkiksi ilmaisut *kotimaamme* ja *taloutemme*, jotka oman tutkimuksemme mukaan olivat hyvin yleisiä 1950- ja 1970-luvun teksteissä. Tällaisten ilmausten käytön väheneminen johtuu yhteiskunnallisista (eli kielenulkoisista) tekijöistä, mutta nämä tekijät heijastuvat kieleen siten, että possessiivisuffiksien käyttö vähenee monikon ensimmäisessä persoonassa. Kielenulkoiset tekijät vaikuttavat puhujien kielelliseen käyttäytymiseen, minkä seurauksena kielessä tapahtuu muutoksia. Kieli ei ole täten erillään yhteiskunnasta mutta ei myöskään erottamaton osa sitä. Mikäli näin ei olisi, olisi autonominen kielitiede mahdotonta (mitä se ei ole).

Tutkimuksessamme tulee myös esiin kysymys kielenmuutoksen nopeudesta. Yllä on esimerkiksi todettu, että possessiivisuffiksien tunnus-ten osalta komplementaarisen distribuution synty on vienyt yli 200 vuotta. Suffiksien käytön vähenemistä puhutussa kielessä on kuvattu nopeasti tapahtuvaksi muutokseksi, jonka leviämisestä kirjoitettuun kieleen on merkkejä jo 1960-luvulta. Tästä on kuitenkin jo 60 vuotta, ja koska tulostemme perusteella suffiksien käytössä kirjoitetussa kielessä ei näy suurempia vähenemisen merkkejä, kovin nopeasta muutoksesta ei voi olla kysymys. On myös mahdollista, että puhutussa kielessä havaittu muutos ei koskaan tule edes etenemään kirjoitettuun kieleen vaan kyse on pysyvistä vaihtelusta, jossa suffiksin poisjätö kirjoitetussa kielessä on joko tarkoituksellista tai perustuu huolimattomuuteen ja kielenhallinnan puutteeseen.

Muutoksen nopeus on ylipäättään varsin monimutkainen kysymys. Sociolingvistisissä tutkimuksissa on todettu, että näennäisaikainen tutkimus saattaa aliarvioida muutoksen nopeutta (ks. esimerkiksi Labov 1994, 112; Sankoff 2006; Sankoff & Blondeau 2007). Toisaalta Wagner & Sankoff (2011) ovat havainneet, että näennäisaikainen tutkimus saattaa myös yliarvioida muutoksen nopeuden. Selittäväksi tekijäksi nostetaan yleensä yksilöiden käyttäytyminen: jos yhteisössä on meneillään muutos ja jos yksilöt muuttavat kieltään meneillään olevan muutoksen mukaisesti, on muutos nopeampaa kuin mitä näennäisaika antaisi olettaa. Vastaavasti jos yksilöt muuttavat kieltään yhteisössä meneillään olevan muutoksen vastaisesti, on muutos hitaampaa kuin mitä näennäisaika antaisi olettaa (ks. Paunonen 2005, 16; Laasanen 2016). Luultavasti tulkitoihin vaikuttaa myös se, missä vaiheessa oletettu muutos on (ks. Laasanen 2016). Lisäksi on huomattava, että näennäis- ja myös reaaliaikaisten tutkimusten ajallinen läheisyys muutokseen saattaa hämärtää kuvaa muutoksesta siinä mielessä, että varianttien edustuksissa havaitut muutokset saattavat saada muutoksen näyttämään hitaammalta tai nopeammalta kuin mitä se todellisuudessa on.

Meneillään oleva kielenmuutos puhutusta kielestä kirjoitettuun on monimutkainen ilmiö, sillä siinä sekoittuvat keskenään muutos, kehitykselliset tekijät ja modaliteettien väliset erot. Nämä tekijät korostuvat tarkasteltaessa nuoria puhujia, joita on jo pitkään pidetty muutoksen



airuina (ks. esimerkiksi Labov 2001; Aitchison 2003, 739; Tagliamonte & D'Arcy 2009). Nuoret puhujat ovat usein muutoksen kärjessä, mutta samanaikaisesti nuorten puhujien kielenkehitys saattaa olla vielä kesken ja he eivät välttämättä vielä hallitse kirjoitetun kielen mallin mukaista kielenkäyttöä yhtä hyvin kuin vanhemmat. Lisäksi aikuistuuksaan he usein muuttavat kielenkäyttöään kohti standardia (ks. Chambers 1995, 177–185) tai kuten Anttila (1989 [1972], 194) kirjoittaa, perinne voittoa. Muutoksen tunnistaminen tästä sekamelskasta saattaa olla varsin haasteellista.

## Lähteet

- Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Agrawala, V. S. 1963 [1953]. *India as known to Pāṇini: As study of the cultural material in the Aṣṭadhyāyī*. Varanasi: Prithvi Kumar.
- Aitchison, J. 2003. Psycholinguistic perspectives on language change. Julkaisussa: B. D. Joseph & R. D. Janda (toim.) *The handbook of historical linguistics*. Malden: Blackwell, 736–743.
- Akinnaso, F. N. 1982. On the differences between spoken and written language. *Language and speech* 25: 2, 97–125.
- Anttila, R. 1989 [1972]. *Historical and comparative linguistics*. Toinen, korjattu painos. Amsterdam: John Benjamins.
- Biber, D. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Bloomfield, L. 1951 [1933]. *Language*. New York: Henry Hold & Company.
- Bloomfield, L. 1970 [1929]. Review of Liebich. Julkaisussa: C. F. Hockett (toim.) *A Leonard Bloomfield Anthology*. Bloomington: Indiana University Press, 219–226.
- Boas, F. 1964 [1911]. On grammatical categories. Julkaisussa: D. Hymes (toim.) *Language in culture and society*. New York: Harper & Row Publishers, 121–123.
- Cameron, D. 1997 [1990]. Demythologizing sociolinguistics. Julkaisussa: N. Coupland & A. Jaworski (toim.) *Sociolinguistics: A reader and coursebook*. Houndmills: Macmillan Press Ltd., 55–67.
- Campbell, L. 2013. *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- Cardona, G. 1976: *Pāṇini: A survey of research*. The Hague: Mouton.
- Chambers, J. K. 1995. *Sociolinguistic theory: Linguistic variation and its social significance*. Oxford: Blackwell.
- Coulmas, F. 1989. *The writing systems of the world*. Oxford: Blackwell.
- D'Arcy, A. & Tagliamonte, S. A. 2015. Not always variable: Probing the vernacular grammar. *Language Variation and Change* 27, 255–285.

- Dufva, H. 1998. From "psycholinguistics" to a dialogical psychology of language: Aspects of the inner discourses. Julkaisussa: M. Lähteenmäki & H. Dufva (toim.) *Dialogues on Bakhtin: Interdisciplinary readings*. Jyväskylä: Jyväskylän yliopisto, 87–105.
- Dufva, H. 2000. Kirjoitettu kieli, kognitio ja emergenssi. Julkaisussa: U. Määttä, T. Nieminen & P. Pälli (toim.) *Emergenssin kielelliset kasvot*. Tampere: Tampereen yliopisto, 155–182.
- Ellis, N. 2012. Variable competence. *Linguistic Approaches to Bilingualism* 2: 3, 264–268.
- Goody, J. 1977. *The domestication of the savage mind*. Cambridge: Cambridge University Press.
- Hakulinen, A. 2003. Ovatko puhuttu ja kirjoitettu kieli erkaantuneet toisistaan? *Kielikello* 1, 4–7.
- Hakulinen, A., Villkuna, A., Korhonen, R., Koivisto, V., Heinonen, T. R. & Alho, I. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Hakulinen, L. 1979. *Suomen kielen rakenne ja kehitys*. Neljäs, korjattu ja lisäty painos. Helsinki: Otava.
- Hakanen, Aimo. 2000. Onks teijä heitin? Monikon persoonapronomineiksi luokiteltujen muotojen käytöstä suomen murteissa ja kirjakielessä. *Sananjalka* 42: 7–64.
- Harris, R. 1980. *The language-makers*. Ithaca: Cornell University Press.
- Heikkilä, M. 2017. *Keskiajan suomen kielen dokumentoitu sanasto ensiesiintymävuosiin*. Tampere: Mediapinta.
- Heine, B. 1997. *Possession: Cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press.
- Hiiidenmaa, P. 2005. Näkökulmia yleiskieleen. *Kielikello* 4, 5–11.
- Hockett, C. F. 1958. *A course in modern linguistics*. New York: The Macmillan Company.
- Hurttu, H. 1999. Variaationtutkimuksen myytit ja stereotyyptit. Julkaisussa: U. Määttä, P. Pälli & M. K. Suojanen (toim.) *Kirjoituksia sosiolingvistiikasta*. Tampere: Tampereen yliopisto, 53–101.
- Häkkinen, K. 1994. *Agricolasta nykykieleen: Suomen kirjakielen historia*. Porvoo: WSOY.
- Ikola, O. 1972. Kirjakieli ja puhekieli, yleiskieli ja murre. *Sananjalka* 14, 16–25.
- Ikola, O. 1986. Yleispuhekielemme synty. *Kielikello* 2, 5–7.
- Ikola, O. 1992. Omistusliitteiden harvinaistuminen. Julkaisussa: V. Yli-Vakkuri, M. Länsimäki & A. Nyman (toim.) *Yhteiskunta muuttuu, kieli muuttuu: Nykysuomen seuran 10-vuotisjuhlakirja*. Porvoo: WSOY, 176–196.
- Ikola, O. 2001. *Nykysuomen opas*. Turku: Turun yliopisto.
- Itkonen, E. E. 1955. Onko itämerensuomessa jälkiä duaalista? *Virittäjä* 59: 2, 161–175.
- Itkonen, E. 2004. Kielitiede tuuliajolla. *Kanava* 4–5, 266–268.
- Itkonen, E. 2005. *Ten non-European languages: An aid to the typologist*. Turku: Turun yliopisto.
- Itkonen, E. 2008. Concerning the role of consciousness in linguistics. *Journal of Consciousness Studies* 15: 6, 15–33.
- Itkonen, E. 2010. Maailman kielten erilaisuus ja samuus: Osa III. Turku: Turun yliopisto.
- Jaakola, M. 2004. Peruskoulun 2. ja 3. luokan oppikirjojen sanasto: Dynamiikasta luettavuuteen. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Jackendoff, R. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

- Järvinen, R. 2011. *Hämäläismurteiden omistusjärjestelmän muutos näennäis- ja reaaliajassa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Kangasmaa-Minn, E. 1986. Possessiivisuffiksin apologia. *Sananjalka* 28, 7–16.
- Karlsson, F. 1975. Suomen kielen tulevaisuus. *Sananjalka* 17, 51–66.
- Karlsson, F. 1998. *Yleinen kielitiede*. Helsinki: Yliopistopaino.
- Kiparsky, P. 1994. *Pañinian linguistics*. Julkaisussa: R. E. Asher & J. M. Y. Simpson (toim.) *The encyclopedia of language and linguistics*. Oxford: Pergamon Press, 2918–2923.
- Kiuru, Silva. 1996. Mitä murrenäytteet kertovat omistusliitteistä? *Virittäjä* 100: 82–94.
- Korhonen, Riitta. 2006. Ovatko omistusliitteet kadonneet? Kuukauden kielijuttu 5/2006. Kotimaisten kielten tutkimuskeskus. [https://www.kotus.fi/nyt/kolumnit\\_artikkelit\\_ja\\_esitelmat/kuukauden\\_kielijuttu\\_%282003\\_2011%29/ovatko\\_omistusliitteet\\_kadonneet](https://www.kotus.fi/nyt/kolumnit_artikkelit_ja_esitelmat/kuukauden_kielijuttu_%282003_2011%29/ovatko_omistusliitteet_kadonneet). Viitattu 1.2.2021.
- Kravchenko, A. 2011. How Huberto Maturana's biology of cognition can revive the language sciences. *Constructivist Foundations* 6, 352–362.
- Kuparinen, O., Mustanoja, L., Peltonen, J., Santaharju, J. & Leino, U. 2019. Muutosmallit Helsingin puhekielessä. *Sananjalka* 61, 30–56.
- Kurki, T. 2005. *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen. Kielenmuutosten seuramisesta ja niiden tarkastelussa käytettävistä menetelmistä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kurki, T. 2013. Näennäisaikametodi ja muutosmallit. *Sananjalka* 55, 202–212.
- Laasanen, M. 2016. Näennäisaikametodin taustaoletusten teoreettista tarkastelua. *Virittäjä* 120: 1, 57–83.
- Laasanen, M., Pajunen, A. & Häikiö, T. Arvioitavana. Kielen natiivihallinnan variaatio ja sosiolingvistinen meneillään olevan kielenmuutoksen tutkimus.
- Labov, W. 1963. The social motivation of sound change. *Word* 19, 273–309.
- Labov, W. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. 1994. *Principles of linguistic change, volume 1: Internal factors*. Oxford: Blackwell.
- Labov, W. 2001. *Principles of linguistic change, volume 2: External factors*. Oxford: Blackwell.
- Lappalainen, H., Mustanoja, L., O'Dell. 2019. Miten ja milloin yksilön kieli muuttuu? *Helsinginkiläisidiolektien muutos ja muutoksen tutkimuksen menetelmät. Virittäjä* 123: 4, 550–581.
- Latvala, S. 1899. *Lauseopillisia muistiinpanoja Pohjois-Savon murteesta*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Leino, P. 1992. Ylioppilaisaineiden kielivirheet: Mistä ne kertovat? Julkaisussa A. Julin, K. Parko, P. Tiuraniemi & S. Viinamäki (toim.) *Ylioppilaisaineita 1992*. Helsinki: Suomalaisen Kirjallisuuden Seura, 170–188.
- Lindén, Eva. 1959. Possessiivisuffiksien viittaussuhteista. Julkaisussa: P. Virtaranta, T. Itkonen & P. Pulkkinen (toim.) *Verba docent: Juhlakirja Lauri Hakulisen 60-vuotispäiväksi 6.10.1959*. Helsinki: Suomalaisen Kirjallisuuden Seura, 294–307.
- Linell, P. 1982. *The written language bias in linguistics*. Linköping: The University of Linköping.
- Makkonen-Craig, H. 1996. Yleispuhekielisydet lehtikielessä. Suomen kielen pro gradu -tutkielma. Helsingin yliopisto.

- Mark, J. 1925. *Die Possessivsuffixe in den Uralischen Sprachen: 1. Hälfte*. Helsinki: Suomalais-Ugrilainen Seura.
- McCarthy, M. & Scott, J. 2007. Vocd: A theoretical and empirical evaluation. *Language Testing* 24: 4, 459–488.
- Miestamo, M. 2006. Suomi maailman kielten joukossa eli mikä suomen rakenteessa onkaan erityistä? Julkaisussa: M. Harmanen & M. Siironen (toim.) *Kielioppi koulussa*. Helsinki: Äidinkielen opettajain liitto, 43–54.
- Mitrunen, H.-M. 2005. *Possessiivikongruenssi ja muut omistusliiteilmaukset Tampereen puhekielessä 1970- ja 1990-luvulla*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Nevalainen, T. & Raumolin-Brunberg, H. 2012. Historical sociolinguistics: Origins, motivations, and paradigms. Julkaisussa: J. M. Hernández-Campoy & J. C. Conde-Silvestre (toim.) *The handbook of historical sociolinguistics*. Malden: John Wiley & Sons, 22–40.
- Niemi, Heidi. 2018. Viikon 15 Tiedekysymys: Katoaako possessiivisuffiksi suomen kielestä? *Oulun ylioppilaslehti*. <https://www.oulu.fi/ylioppilaslehti/fi/viikon-15-tiedekysymys-katoaako-possessiivisuffiksi-suomen-kielesta/>. Viitattu 1.2.2021.
- Nuolijärvi, P. 1986. ”Ota minut sinun uniin” (Juice Leskinen): Nykysuomalaisen omistusmuotojärjestelmästä. Julkaisussa: P. Leino & J. Kalliokoski (toim.) *Kieli 1*. Helsinki: Helsingin yliopisto, 157–183.
- Nuolijärvi, P. & Sorjonen, M.-L. 2010. *Miten kuvata muutosta? Puhutun kielen tutkimuksen lähtökohтия murteenseuruuhankeen pohjalta*. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Olson, D. 1996 [1994]. *The world on paper: The conceptual and cognitive implications of writing and reading*. Cambridge: Cambridge University Press.
- Pajunen, A. 2004. Kirjoitelmakorpus. Morfosyntaktisesti analysoitu tietokanta, noin miljoona sanamuotoa.
- Pajunen, A. 2010. Sanojen synteettisyysasteesta suomen kielessä. *Virtittäjä* 114: 4, 481–501.
- Pajunen, A. 2016. Unelmakorpus. Peruskoululaisten ja nuorten aikuisten kirjoitelmia aiheesta Unelmieni päivä. Excel-tietokanta.
- Pajunen, A. & Palomäki, U. 1984. *Tilastotietoja suomen kielen rakenteesta 1*. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Pajunen, A. & Virtanen, P. 2002a. *Helsingin Sanomat* 2000–2001 [= HS2000-korpus]. Morfosyntaktisesti analysoitu tietokanta, 24 miljoonaa sanamuotoa.
- Pajunen, A. & Virtanen, P. 2002b. *Savon Sanomat* 2001 [= SS2000-korpus]. Morfosyntaktisesti analysoitu tietokanta, 8 miljoonaa sanamuotoa.
- Pajunen, A., Puranen, M. & Yli-Paavola, A. 2010. Koulukorpus. Alakoululaisten kirjoitelmia aiheesta Unelmieni päivä. Excel-tietokanta.
- Pajunen, A., Itkonen, E. & Vainio, S. 2015. Sanamerkituksen hallinta nuorilla aikuisilla. *Virtittäjä* 119: 2, 160–187.
- Paunonen, H. 1995 [1982]. *Suomen kieli Helsingissä: huomioita Helsingin puhekielen historiallisesta taustasta ja nykyvariaatiosta*. Helsinki: Helsingin yliopisto.
- Paunonen, H. 1995. Puhesuomen muuttuva omistusmuotojärjestelmä. *Virtittäjä* 99: 4, 162–200.
- Paunonen, H. 2005. Totta vai toiveajattelua: Näennäisaikainen muutos todellisen muutoksen kuvastajana. Julkaisussa: J. Vaattovaara, T. Suutari, H. Lappalainen & R. Grünthal

- (toim.) *Muuttuva muoto: Kirjoituksia Tapani Lehtisen 60-vuotispäivän kunniaksi*. Helsinki: Helsingin yliopisto, 13–53.
- Penttilä, A. 1932. Grafeema- ja foneemasuomen suhteesta. *Virittäjä* 36, 16–26.
- Ramscar, M., Hendrix, P., Love, B. & Baayen H. 2013. Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon* 8: 3, 450–481.
- Rapola, M. 1965. *Suomen kirjakielen historia pääpiirteittäin 1: Vanhan kirjasuomen kirjoitus- ja äänneosan kehitys*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Robins, R. H. 1990 [1967]. *A short history of linguistics*. Kolmas painos. Lontoo: Longman.
- Romaine, S. 1982. *Socio-historical linguistics: Its status and methodology*. Cambridge: Cambridge University Press.
- Salonen, I. 2012. *Puhekielen piirteiden esiintyminen alakoululaisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Sankoff, G. 2006. Age: Apparent time and real time. Julkaisussa: K. Brown (toim.) *The encyclopedia of language and linguistics. Volume 1*. Oxford: Elsevier, 110–116.
- Sankoff, G. 2018. Language change across the lifespan. *Annual Review of Linguistics* 4, 297–316.
- Sankoff, G. & Blondeau, H. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83: 3, 560–588.
- Saukkonen, P. 1982. *Oulun korpus. 1960-luvun suomen yleiskielen tutkimusmateriaali*. Oulu: Oulun yliopisto.
- de Saussure, F. 1962 [1916]. *Cours de linguistique générale*. Pariisi: Payot.
- Scharfe, Ht. 1977. *Grammatical literature. A history of Indian Literature, volume 5, fasc 2*. Wiesbaden: Otto Harrassowitz.
- Setälä, E. N. 1966 [1880]. *Suomen kielen lauseoppi*. Viidestoista painos. Helsinki: Otava.
- Sturtevant, E. H. 1917. *Linguistic change: An introduction to the historical study of language*. Chicago: The University of Chicago Press.
- Tagliamonte, S. & D’Arcy, A. 2009. Peaks beyond phonology: Adolescence, incrementation, and language change. *Language* 85: 1, 58–108.
- Tannen, D. (toim.) 1982. *Spoken and written language: Exploring orality and literacy*. Norwood: Ablex.
- Tiainen, S. 2008. *Sitten pitkän ajan kuluttua, kun sinusta on tullut jo pappa: Syntaksin ja sanaston vaihtelu opiskelijoiden kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Jyväskylän yliopisto.
- Tuomi, T. 1986. Murre ja yleiskieli. *Kielikello* 2, 22–23.
- Turunen, P. 1982. Monikon 3. persoonan kirjakielinen ja puhekielinen malli. Julkaisussa: K. Jonninen-Niilekselä (toim.) *Tampereen puhekieli tutkimuskohteena*. Tampere: Tampereen yliopisto, 161–189.
- Turunen, U. 2012. Suomi toisena kielenä -oppijoiden possessiivisuffiksin käyttö. Suomen kielen pro gradu -tutkielma. Jyväskylän yliopisto.
- Vainio, S., Hyönä, J. & Pajunen, A. 2008. Processing modifier-head agreement in reading: Evidence for a delayed effect of agreement. *Memory & Cognition* 36: 329–340.
- Viinikka, J. & Voutilainen, E. 2013. Ääniä ilmassa, merkkejä paperilla: Puhutun ja kirjoitetun kielen suhteesta. *Kielikello* 3, 18–21.

- Voutilainen, A. 2002. Conexor FDG [= functional dependency grammar] for Finnish: Linguistic description. Conexor: Natural knowledge. Moniste.
- Wagner, S. E. & Sankoff, G. 2011. Age grading in the Montréal French inflected future. *Language Variation and Change* 23, 275–313.
- WALS=The World Atlas of Language Structures Online. <https://wals.info>. Viitattu 20.2.2018.
- Varjonen, A.-M. 1996. "Meidänkin kannalta paras ratkaisu": Sanomalehtikielen omistusliitteistä. *Kielikello* 1996: 3.
- Virtanen, P. & Pajunen, A. 2001. ContextMill: The easy to use concordancer. Turku: Aurit Software.
- Weinreich, U., Labov, W. & Herzog, M. I. 1968. Empirical foundations for a theory of language change. Julkaisussa: W. P. Lehmann & Y. Malkiel (toim.) *Directions for historical linguistics*. Austin: University of Texas Press, 95–195.
- Yli-Vakkuri, V. 1993. Suomen kielen omistusliitteen tulevaisuus: Prognoosi vuosien 1991 ja 1992 ylioppilaskirjoitusten äidinkielen kokeen pohjalta. *Sananjalka* 34, 77–87.

# Kirjoittamisen tarkkuuden kehittyminen ja yhteys kirjoittajan kokonaiskielitaitoon alakouluiässä

*Mari Honko*

 <https://orcid.org/0000-0002-3240-622X>

## Johdanto

Kirjoittamisen taitodiskurssia painottava opetuksen, tutkimuksen ja arvioinnin perinne on Suomessa vahva. Tämä näkyy esimerkiksi menneiden vuosikymmenten opetus suunnitelmien perusteissa, oppimateriaaleissa ja arviointitottumuksissa sekä kielitieteen, erityispedagogiikan ja kasvatustieteen ja osin myös psykologian aloilla jo tehdyn tutkimuksen valinnoissa kuten aineistoissa, käytetyissä metodeissa ja teoreettisen viitekehysten hahmottelussa (ks. myös Kauppinen, Pentikäinen, Hankala, Kulju, Harjunen & Routarinne 2015). Taitodiskurssia painottaneessa opetuksessa korostuvat yksilön taidot. Kirjoittamista tarkastellaan ensisijaisesti kielellisten taitojen soveltamisena kirjoittamiseen, ja päähuomio kohdistuu kirjalliseen tuotokseen, kuten oikeinkirjoitukseen, tekstien yhtenäisyyteen, tarkkuuteen ja virheettömyyteen. Tätäkin diskurssia kiistatta tarvitaan, mutta sen asema ansaitsee uudelleen arviointia. (Ks. kirjoittamisen diskursseista Ivanič 2004.)

Taitodiskurssin puutteena voidaan pitää erityisesti siihen usein vähintään implisiittisesti liitettyä oletusta ikään kuin vaihtelemattomista, tilasteesta toiseen siirrettävissä olevista taidoista. Ongelmaksi taitodiskurssi muodostuu, mikäli sitä painotetaan opetuksessa ja arvioinnissa muiden näkökulmien kustannuksella ja mikäli unohdetaan yksilöiden erilaiset valmiudet, kirjoittamisen monet merkitykset ja päämäärät sekä kirjoittamistilanteiden ja -kontekstien vaihtelevuus. Kirjoittamisen opetuksen, arvioinnin ja tutkimuksen valinnat kytkeytyvät laajempaan kirjoittamiskäsitykseen, joka on muovannut niin kierrätettyjen käytänteiden, kirjoittamista koskevien uskomusten ja puhumisen tapojen kuin opetusta ohjaavien määräysasiakirjojen, arviointimallien ja oppikirjojenkin sisältöjä. (Ks. esimerkiksi Erra & Svinhuvud 2017). Käsite kirjoittamisesta ja kirjoitustaidosta on erityisesti tuoreimman opetussuunnitelmaudistuksen (OPH 2014) myötä laajentunut ja tutkimus moninaistunut jo tätä ennemmin. Yksilön taitoon ja erityisesti kirjoittamisen tarkkuuteen kiinnittyvä kirjoittamiskäsitys painottuu kuitenkin yhä vahvasti opetuksen ja arvioinnin käytänteissä sekä näistä käytävissä keskusteluissa, minkä olen havainnut myös työssäni kielen opettajana ja opettajankouluttajana. Näkökulma onkin tärkeä, sillä tekstittyneessä yhteiskunnassa toimimiseen tarvitaan vahvaa kirjoitustaitoa ja erilaisiin arviointitarpeisiin monenlaisia, myös yksilökeskeisiä, kirjoitustaidon arviointitapoja. Samalla tarvitaan ymmärrystä kirjoittamisen moninaisuudesta sekä kirjoitustaidon dynamisesta ja lukuisiin vaikuttaviin tekijöihin kietoutuvasta luonteesta (ks. esimerkiksi Ivanič 2004; Kauppinen ja muut 2015).

Vaikka taitodiskurssilla on Suomessa ollut kaikessa kieltenopetuksessa vahva asema, puuttuu tutkimukseen perustuvaa tietoa – johon opetusta ja arviointia voitaisiin suhteuttaa – kirjoittamistaidon kehitymisestä erityisesti alkuopetusvaiheen jälkeen. Alakouluikäisten oppilaiden kirjoittamisen tarkkuutta on Suomessa tutkimuksen keinoin aikaisemmin tarkasteltu lähinnä suomea äidinkielenä puhuvien lasten näkökulmasta. Tyypillisesti on tutkittu kontrolloitua kirjoittamista, kuten toisto- tai sanelukirjoittamisessa sanatasolla ilmenevää teknisen kirjoittamisen taitoa, ja kehyksenä on usein ollut lukemisen ja kirjoittamisen erityisvaikeuksiin liittyvä tutkimus (ks. esimerkiksi Kulju & Mäkinen 2009; Aro, Eklund, Leppänen & Poikkeus 2011; Kauppinen ja muut



2015). Myös alakouluikäisten tuottavaa kirjoittamista on jo aikaisemmin tutkittu erilaisin painoituksin (esimerkiksi Saarela 1997; Mäki 2002; Mäkihonko 2006; Pajunen 2012). Suomi toisena kielenä -oppijoiden kehittyvää kirjoitustaitoa on samassa ikävaiheessa kuitenkin tarkasteltu vasta muutamissa tutkimuksissa (ks. Halonen 2009; Honko 2013; Sikiö, Siekkinen & Holopainen 2016).

Tässä luvussa tarkastelen alakouluikäisten lasten tuottavaa kirjoittamista, erityisesti sen tarkkuuden kehittymistä yhdessä kirjoitustaidon kiivaassa kehitysvaiheessa vuosiluokilta 2–3 vuosiluokille 5–6. Lisäksi tarkastelen kirjoittamisen tarkkuuden suhdetta tuottavan kirjoittamisen muihin osataitoihin. Tutkimusryhmässä on sekä suomea äidinkielenä että toisena kielenä puhuvia koululaisia eri puolilta Suomea. S2-ryhmässä vertaan kirjoittamisen tarkkuutta ja sen kehittymistä myös kirjoittajan suomen kielen sekä oman äidinkielen taidon tasoon.

Tutkimukseni kytkeytyy osaksi kielitaidon arviointitutkimusta, tarkemmin osaksi kirjoitustaidon arviointia (Huhta & Hildén 2016). Kontrolloituun kirjoittamiseen nähden tuottava kirjoittaminen antaa kattavamman kuvan oppijan toiminnallisesta kirjoitustaidosta, esimerkiksi kyvystä ilmaista itseään kirjallisesti. Koulunkäynnin alkuvaiheessa kokonaisen tekstin tuottaminen ja keskittyminen yhtä aikaa sekä tekstin sisältöön että kielelliseen muotoon on useimmille kirjoittajille toisaalta vielä vaikeaa (Mäki 2002). Kirjoitustaidon mittarina tuottavan kirjoittamisen tehtävät ovatkin kirjoittajia varsin hyvin erottelevia, minkä vuoksi ne sopivat kontrolloitua kirjoittamista paremmin etenkin alkuopetusvaihetta myöhemmän vaiheen ja tyypillisesti kehittyvän kirjoitustaidon arviointiin (ks. myös Aro ja muut 2011). Kielitaidon vertailumittarina käytän kirjoittajien sanastollisia taitoja mittaavia sanastotehtäviä (Honko 2013) sekä kirjoittajilta ja heidän opettajiltaan kerättyä kyselyaineistoa.

Etenkin S2-oppilaiden oman äidinkielen taidot huomioivaa kielitaidon arviointitutkimusta on Suomessa toistaiseksi tehty hyvin niukasti. Kansallisen koulutuksen arviointikeskuksen selvityksen mukaan perusopetuksen päättövaiheen oppilaat ovat oman arvionsa mukaan yleensä saavuttaneet vähintään hyvät vuorovaikutustaidot sekä suomen kielellä että omalla äidinkielellään (Kuukka & Metsämuuronen 2016). Kielten käyttöalassa, taidoissa ja merkityksissä on kartoituksen mukaan yksilöl-

listä ja tilanteistakin vaihtelua, mutta kirjallisen viestinnän taidot ovat tarkastellussa ikäluokassa keskimäärin vahvemmat suomen kielessä kuin omassa äidinkielessä (mts. 73–77). Tähän vaikuttanee ennen kaikkea vähäisemmät mahdollisuudet saada opetusta, ohjausta ja harjoitusta omalla äidinkielellä lukemiseen ja kirjoittamiseen. Kartoituksessa ei kuitenkaan ole tarkasteltu monikielisiä resursseja yksilöllisellä tasolla eli esimerkiksi tietyn kirjoittajan kirjoitustaitoa omalla kielellään suhteessa suomen kielen kirjoitustaitoon.

Kriittisen pedagogiikan mukaan kirjoittamisella voidaan viitata taitoon mutta myös luovuuteen tai prosessiin ja sitä voidaan tarkastella lukuisista eri näkökulmista, kuten tekstilajien tuottamisena, sosiaalisten käytänteiden noudattamisena tai sosiopoliittisena toimintana (Ivaničín 2004 pohjalta Erra & Svinhufvud 2017). On siis muistettava, että kokonaisvaltaisessa kirjoittamisen tutkimuksessa ja opetuksessa kirjoittamisen tarkkuus on vain yksi osa kirjoitustaitoa ja että taitodiskurssi puolestaan on vain yksi näkökulma kirjoittamiseen (Ivanič 2004, 225). Vaikka tämä luku kohdistuu kirjoittamisen tarkkuuteen, tarkoitus ei ole vahvistaa jo ennestään vahvan taitodiskurssin asemaa kirjoittamisen opetuksessa. Sen sijaan tavoitteena on auttaa hahmottamaan kirjoittamisen tarkkuuden tyypillistä kehittymistä ja siihen vaikuttavia syitä osana kirjoittamisen kokonaistaitoa, erityisesti toisella kielellä kirjoitettaessa.

Kun tarkkuuden kehitystä havainnoidaan diakronisesti ja samalla otetaan huomioon kielenoppimisprosessin eteneminen muilla kielitaidon eri osa-alueilla, on mahdollista saada tietoa sekä tyypillisestä kielenkehityksestä että kielenkehityksen tukemisen tarpeista (ks. esimerkiksi Ellis & Barkhuizen 2005; Martin 2010; Martin, Mustonen, Reiman & Seilonen 2010; Slabakova 2016). Kirjoittamistaidon tyypilliseen kehityskulkuun ja haasteisiin tutustuminen auttaa esimerkiksi ymmärtämään, millaiset seikat ovat kirjoittamisen kannalta ajankohtaisia tietyssä ikävaiheessa. Tämän tutkimuksen tieto voikin auttaa paitsi tunnistamaan kirjoittamisen tarkkuuteen liittyvän tuen tarpeita peruskoulun alaluokilla myös kohdentamaan tukea oikein ja vähentämään tarpeetonta huolta tilanteissa, joissa kehkeytyvään kielitaitoon liittyy tilapäisesti myös epä-tarkkuus.

## TUTKIMUSKYSYMYKSET JA VIRHEEN KÄSITE

Tässä tutkimuksessa kirjoittamisen tarkkuutta tarkastellaan aluksi lingvistisen analyysin avulla. Sen jälkeen tuloksia peilataan kirjoittajakohtaisesti taustatietoihin sekä kirjoittamisen muihin osataitoihin. Tutkimuskysymyksiä on neljä:

- 1) Miten kirjoittamisen tarkkuus kehittyi seurantajaksolla alakoulun aikana?
- 2) Millaiset virheet ovat yleisimpiä ja mitkä tekijät niitä selittävät?
- 3) Onko kirjoittamisen tarkkuudessa eroja eri kirjoittajaryhmien välillä?
- 4) Onko kirjoittamisen tarkkuus yhteydessä muihin kirjoittamisen osataitoihin ja S2-ryhmässä tarkemmin kirjoittajan suomen kielen tai oman äidinkielen taitoon kielitaidon eri osa-alueilla?

Esittelen seuraavaksi tutkimuksen keskeiset käsitteet, perehdytän tarkkuuteen kehityksellisenä mittarina ja kokoan lyhyesti aiempaa tutkimusta. Aineisto ja tutkimusmetodi esitellään johdannon lopuksi. Tulokset esitetään analyysiluvuissa ja kootaan pohtien päätäntöluvussa.

Kun tutkitaan kielenkäytön tarkkuutta, otetaan samalla väistämättä kantaa siihen, että kieli ei aina täytä tarkan vaatimuksia. Kielellisiä epätarkkuuksia on nimetty ja määritelty lukemattomin eri tavoin, mutta eniten käytetty, joskin usein kovin ohuesti määritelty, lienee *virheen* käsite. Hieman yksinkertaistaen virheellä tarkoitetaan eroa odotuksenmukaisen ja toteutuneen muodon välillä eli esimerkiksi konventionaalisesta poikkeavaa kieliopillista muotoa, sanajärjestystä tai sanavalintaa. Englannin kieleen keskittyvässä toisen kielen oppimisen tutkimuksen perinteessä keskeistä on määrittelyssä ollut toistuvien ja satunnaisten poikkeamien erottaminen toisistaan (Corder 1967). Kiinnostavaa on, että suomen kielen taitoon kohdistuvissa tutkimuksissa on ollut varsin yleistä luokitella ja nimetä virheitä määrittelemättä kuitenkin itse virheen käsitettä. Ensimmäisessä suomi toisena kielenä -alan väitöskirjassa Martin (1995) kirjoittaa, että virheellä (*error*) on viitattu nimenomaan toisen kielen oppijan puheessa tai kirjoituksessa ilmenevään taidon puutteeseen, joka kontrastoituu äidinkielen puhujan tarkkaan tai korkeintaan lipsahduksia (*slips*) sisältävään kielenkäyttöön. Tämä näkemys ei nähdäkseni kuitenkaan ylety edellä mainittuun äidinkielen ke-

hityksen varhaisvaiheen tutkimukseen. Esimerkiksi Mäkihonko (2006) sekä Havukainen ja Salmi (2011) käyttävät virheen käsitettä synonyymina kirjoitusvirheelle tutkimuksissaan, jotka käsittelevät tuottavan kirjoitustaidon varhaisvaiheen kehittymistä.

On huomattava, että kielellisesti tarkkakaan muoto ei välttämättä ole sosiaalisesti hyväksyttävä, sillä kaikki virheettömät muodot eivät ole vaikiintuneet kaikkiin kielimuotoihin ja kielenkäyttötilanteisiin; virheellinen muoto puolestaan voi olla tilanteisesti aivan toimiva. Neutraalissa puheessa sanommekin esimerkiksi *juuri tänään*, *puolukoiden kanssa* ja *järjetön mies* mutta luultavasti emme *aivan tänään*, *poulukkojen kansa* tai *hullu ukko*, vaikka kaikki muodot olisivatkin käyttöyhteydessään ymmärrettäviä.

Virheet kiinnostavat sekä maallikoita että kielentutkijoita, koska ne poikkeavat tutusta tai tavanomaisesta ja ovat usein helposti havaittavissa. Kielivirheet mutta myös niistä puhuminen nostattaa helposti voimakkaita tunteita, sillä virheiden tarkasteluun liittyy aina myös arvottamista ja ideologista painolastia. Virheiden ilmeisin merkitys arjessa perustuu siihen, että ne voivat aiheuttaa vuorovaikutuksessa kitkaa kuten tarkennuspyyntöjä, oikean muodon hakemista, epäröintejä ja keskustelun katkeilua. Vahingollisimpia virheitä eivät kuitenkaan yleensä ole huomion kohteeksi nousevat virheet vaan ne, jotka sivuutetaan ja joiden seurauksena vuorovaikutukseen osallistuneet voivat tahoillaan jäädä väärän ymmärryksen varaan.

Koska vuorovaikutus perustuu yksittäisten sanojen ja sananmuotojen sijaan huomattavasti laajempien kielellisten ja ei-kielellisten kokonaisuuksien tulkintaan, kielivirheet eivät aina ole vaaraksi merkityksen välittymiselle. Tarkkuus on silti tulkinnan sujuvuuden kannalta eduksi ja tietyissä konteksteissa sitä suorastaan vaaditaan. Esimerkiksi lääkäriltä voidaan valta-asemansa ja vastuullisen työnkuvansa vuoksi edellyttää tarkempaa kielenkäyttöä kuin monessa muussa ammatissa toimivalta. Ohjaavien ja velvoittavien tekstien itsenäinen laatiminen ja tulkinta taas edellyttävät lääkäriltäkin suurempaa tarkkuutta kuin vaikkapa epämuodollisempi vuorovaikutus tai työparin tuella toimiminen (Tervola 2019).

Kielenkäytön tarkkuus on hyväksi silloin, kun on tarve ilmaista itseä, tulla ymmärretyksi ja ymmärtää muita täsmällisesti. Lisäksi sillä on merkitystä sosiaalisen aseman ja hyväksytyyden näkökulmasta: virheitä tekevä voidaan tulkita esimerkiksi huolimattomaksi ja välinpitämättömäksi, kieli- ja kulttuuritaidoltaan muita heikommaksi, joskus myös vähemmän luotettavaksi tai älykkääksi (ks. esimerkiksi Jessmer & Anderson 2001; Kreiner, Schnakenberg, Green, Costello, & McClin 2002; Vignovic & Thompson 2010). Jos virheitä tarkastellaan suomen kielen oppimisen kannalta, tarkkuuden kehittymistä voidaankin pitää yhtenä tavoittelemisen arvoisena ja edistymisestä kertovana kielitaidon piirteenä (ks. myös Brunni, Jantunen & Skantsi 2019), johon sisältyvät niin oikean rekisterin eli tilanteeseen sopivan kielimuodon valinta kuin kielellisten ilmausten rekisterinmukainen käyttö.

#### OPPIJANKIELEN TARKKUUS TUTKIMUKSEN KOHTEENA

Oppijankielentutkimuksessa virheanalyttinen lähestymistapa saavutti nopeasti vankan aseman 1960- ja 1970-lukujen taitteessa Stephen Corderin artikkelin *The significance of learner's errors* (1967) innoittamana ja on sittemmin saanut osakseen paljon sekä ansaittua että ansaitsematonta kritiikkiä. Matka virheanalyttisen tutkimuksen alkua ajoista nykyhetkeen on kuitenkin ollut pitkä: aluksi virheitä tarkasteltiin lähinnä ”täydellisen kielisysteemin” hallinnan puutteina ja niiden syitä etsittiin 1960-luvun kontrastiivisen analyysin hengessä melko suorasukaisesti esimerkiksi puhujan äidinkielestä (Lado 1957; ks. myös esimerkiksi Gass & Selinker 2008, 89–90).<sup>75</sup> Sittemmin on kiinnostuttu enemmän oppijankielen kehityksellisestä tarkastelusta.<sup>76</sup> Esimerkiksi DIALANG-arviointijärjestelmään liittyvässä tutkimuksessa tarkkuuden

75 L2-tutkimuskentällä ei silti liene koskaan vallinnut täyttä yhteisymmärrystä siitä, mille kielitaidon osa-alueille äidinkielen vaikutukset voivat virheinä ilmentyä ja milloin on kyse kielten välisten suhteiden aiheuttamista virheistä (*interlingual errors*), milloin kohdekielen systeemin sekoittamisesta (*intra-lingual errors*) – ja kuinka mahdollinen tai tarpeellinen tämän eron tekeminen kulloisessakin tilanteessa on.

76 Varhaisen virheanalyttisen tutkimuksen päämääriä on kritisoitu osin aiheettomastikin, sillä jo Corder (1967, 167) itse korosti virheiden tarkastelussa kielen oppijan tukemisen näkökulmaa: Kielen oppijan tekemät virheet ovat merkityksellisiä sekä kielen opettajalle, tutkijalle että oppijalle itselleen. Ne kertovat oppimisen vaiheista ja siitä, miten kieltä opitaan eli millaisia strategioita oppija esimerkiksi käyttää kieleen tutustuessaan.

(kielioppi, sanasto, oikeinkirjoitus) on havaittu korreloivan voimakkaasti aikuisten kielenoppijoiden sanavaraston koon kanssa (Alderson 2005, 55, 89). Samansuuntaisia tuloksia on saatu muillakin arviointimetoodeilla sekä aikuisten englanninoppijoiden kielestä (ks. esimerkiksi Meara & Milton 2003; Milton 2009, 171) että kouluikäisten S2-oppijoiden kielestä (Honko 2013).

Dynaamisuu den myötä tarkasteltava ilmiö on muuttunut: kontekstistaan irrotettujen virheiden sijaan on ryhdytty tutkimaan muun muassa kielenkäytön epäidiomaattisuutta, virheiden esiintymisympäristöjä sekä esiintymistiheyttä, kielellisen tarkkuuden suhdetta tekstin tulkintaan ja tavoitemuodon neuvottelemista vuorovaikutuksessa muiden kielenkäyttäjien kanssa. Korpustutkimuksen keinot mahdollistavat nykyään paljon sellaista, mihin ei 1960- ja 1970-luvuilla ollut pääsyä, ja 2000-luvulla virheiden analyysi yhdistetään yhä useammin laajempaan kielitaidon profilointiin sekä ohjatun tai spontaanin kielenkäytön vaikuttavuuden arviointiin.

Myös oppimis- ja kielikäsitteet ovat muuttuneet. Modernin virheanalyttisen tutkimuksen lähtöoletuksena on, että kaikki kielenkäyttäjät äidinkielen ja kielenkehityksen vaiheeseen katsomatta tekevät virheitä ja että tarkkuus ei myöskään kehity suoraviivaisesti – vaan kielellisten valmiuksien uudistumiseen liittyy kokeiluja ja haparointia (ks. suomen oppijoiden kirjoittamistaidon kehittämisestä esimerkiksi Seilonen 2013; Mustonen 2015; Brunni ja muut 2019). Virheiden syynä voi olla huolimattomuus sekä se, ettei edes syntyperäisten aikuisten kielitaito automaattisesti noudata mitään standardiolemusta edes kieliopin osalta (ks. myös Chipere 2009, 190–191). Toiminnalliseen kielitaitokäsitykseen kuuluu ajatus siitä, että virheet itsessään myös edistävät oppimista, sillä ne mahdollistavat epämukavuusalueella toimimisen ja siten uuden oppimisen ja voivat saada oppijan kiinnittämään huomiota kehkeytymässä oleviin ilmaisuihin. Toisaalta aiemman tutkimuksen pohjalta tiedetään, että virheet muuttuvat oppimisen edetessä: alkuvaiheessa tehdään enemmän rakenteellisia ja sanastollisia virheitä, myöhemmin poikkeavuudet liittyvät lähinnä ilmiöiden yleisyyteen ja tyypillisiin esiintymisympäristöihin (Martin 2015).

## TARKKUUDEN KEHITTYMINEN OSANA KARTTUVAA KIELITAITOA

Kirjoitustaidon ja kirjoittamalla viestimisen opettelua on verrattu uuden kielen opettelemiseen (esimerkiksi Halonen 2009, 330). Kirjoittamista opetteleva lapsi voi toisaalta hyödyntää kirjoittamisessa puhutun kielen pohjalta oppimaansa tietoa esimerkiksi suomen kielen sanastosta ja rakenteista, tekstilajeista sekä vuorovaikutuskulttuurista. Mitä monipuolisempi lapsen kielellinen ympäristö on varhaislapsuudessa ollut, sitä vahvempi perusta hänellä on kouluiässäkin toimia kielellä joustavasti erilaisissa tilanteissa, kartuttaa ilmaisuvarantoaan ja oppia uusia taitoja, kuten kirjoittamista. Esimerkiksi äänne–kirjain-vastaavuuden oppimisessa ja äännekestojen merkitsemisessä auttaa, jos jo osaa suomen kielen äänteet sekä niiden lyhyen ja pitkän keston merkitystä erottavan tehtävän. Kertomustekstin kirjoittamisessa puolestaan on hyötyä siitä, jos on tutustunut kertomuksiin kuuntelemalla ja puhumalla. Kakkoskielisellä puhujalla sekä suomen kielen systeeminen tieto että kielen käyttöön liittyvä kulttuurinen tieto on tyyppillisesti kapeampaa kuin äidinkielisellä suomenpuhujalla. Syynä tähän ei ole kakkoskielisyys tai monikielisyys itsessään vaan vähäisempi kokemus.

Kirjoitustaidon oppiminen edellyttää puhutun kielen ja kirjoitusjärjestelmän yhteyden ymmärtämistä. Suomen kielen oppijalle tämä tarkoittaa muun muassa latinalaisaakkoston ja vastaavien äänneiden yhteyden sekä kirjoitussuunnan hahmottamista ja oikeinkirjoitusnormeihin tutustumista sekä kykyä pukea omat ilmaisutarpeet kirjalliseen muotoon. Taitodiskurssin näkökulmasta kirjoitustaitoon kuuluukin Mäen (2002, 17) mukaan kolme keskeistä osa-aluetta: kirjoituksen koostaminen kirjoitusmerkkien eli grafeemien avulla, oikeinkirjoitus sekä sisältöjen tuottaminen eli varsinainen kirjoittaminen. Käsin kirjoittaessaan kirjoittamistaitoa harjoitteleva lapsi siis ikään kuin piirtää näkyviin ajatuksiaan, samalla kun hän vielä harjoittelee kirjainten ja sanahahmojen muodostamista, tekstin asettelua sekä tuotettavan tekstilajin erityispiirteitä.

Luku- ja kirjoitustaidon tyyppillisen kehityksen tunnistaminen on avain lukemisen ja kirjoittamisen erityisvaikeuksien tunnistamiseen (Matilainen 1993, 5). Kakkoskielisyys tuo kirjoittamiseen oman erityisvärinsä, johon vaikuttaa paitsi oppijan kielellinen tausta myös muun

muassa kulttuuriset ja sosiaaliset tekijät. Myös suomea äidinkielenään puhuva on koulun alkuvuosina vasta ottamassa kirjoittamista haltuun, ja kirjoituksen tuottaminen vie huomiota suunnitelmalliselta sisältöjen tuottamiselta (ks. esimerkiksi Matilainen 1993). Kirjoittamisen alkuvaiheessa on tyypillistä, etteivät sekä tarkka oikeinkirjoitus että eheän ajatuskokonaisuuden luominen onnistu yhtä aikaa ongelmitta (Mäki 2002, 40). Vaikka esimerkiksi kirjoittamisen sujuvuus kasvaa nopeasti, virheiden määrä ei alkuopetuksen aikana juuri vähene (Mäkihonko 2006). Toisluokkalaaisella voi vielä olla vaikeuksia esimerkiksi äännekestojen kirjoittamisessa ja aloitetun virkkeen jatkamisessa eheänä loppuun saakka ilman, että kyseessä olisi kielellinen erityisvaikeus (ks. myös Kulju & Mäkinen 2009). Tekstistä voi myös puuttua kirjaimia, ja sanavälien merkitsemisessä voi olla horjuntaa. Havukaisen ja Salmen (2011) aineistossa lähes puolet (45 %) toisluokkalaisten tuottavan kirjoittamisen epätarkkuuksista liittyi oikeinkirjoituksen konventioihin, 40 prosenttia äännekirjain-vastaavuuteen (mukaan lukien pitkien äännekestojen merkintä, lisämerkit, puuttuvat kirjaimet, rauniosanat, puuttuvat tavut, rotaatiot ja reversaalit) ja vajaa viidennes (17 %) lausetason ilmiöihin. Vähitellen, Matilaisen (1993, 70–72) mukaan viimeistään neljänneltä luokalta alkaen, kirjoittamisen suunnitelmallisuus lisääntyy ja oppilaalle alkaa muodostua syvempi käsitys sekä käsiteltävistä asiasisällöistä että tekstistä kokonaisuutena.

Tiedetään, että erityisesti geminaattojen kirjoittaminen on äidinkiellekin suomenpuhujalle alkuvaiheessa vaikeaa (Kulju & Mäkinen 2009). Tämä voi liittyä auditiiviseen hahmottamiseen. Pitkä vokaali toteutuu puheessa ympäristönsä lyhyeen vokaaliin nähden pitkänä kestona, mutta geminaatassa pitkää kestoa havainnoidaan esimerkiksi kluusiilien tapauksessa lähinnä sulkeumavaiheen eli hiljaisen jakson kautta. Suomi toisena kielenä -oppijan aikaisemmin oppimisessa kielissä kestoerolla ei välttämättä ole lainkaan merkitystä erottavaa tehtävää tai kestoero voi olla distinktiivinen vain joissakin äänneissä. On siis huomattava, että esimerkiksi juuri äännekestojen kirjoittamisen vaikeudet eivät liity ainoastaan kirjoitetun kielen normien hallinnan puutteisiin vaan myös kielenhallintaan. Vasta, kun oppija oppii kuulemaan esimerkiksi suomen kielen merkitysten rakentamisen kannalta keskeiset kestoerot,



hän voi oppia myös tuottamaan näitä eroja puheessa ja kirjoituksessa tarkasti.

Piper (2001, 23) kritisoi lapsuusiän toisen kielen oppimista käsittelevässä teoksessaan opettajia virhepoliiseina toimimisesta eli siitä, että virheet nähdään koulussa epäonnistumisen merkkinä. Nykyisen oppimiskäsityksen mukaan virheet ovat kuitenkin oppimisen edellytys ja osoitus käynnissä olevasta oppimisprosessista eli ikään kuin osittaisesta menestyksestä uuden asian käyttöön siirryttäessä (esimerkiksi Martin 2009, 324). Kun aktiivinen kielenoppija testaa osaamistaan ja kielestä tekemiään päätelmiä, virheilta ei voi eikä tarvitse välttyä. Virheet voivat olla myös osoitus jo jollain tavalla omaksutun asian uudelleen prosessoinnista, sillä ”takapakki ja uudelleen analysoiminen” ovat oppijankielelle tyypillisiä piirteitä, kuten Martin (2003, 83) asian ilmaisee.

Kielitaidon eri osa-alueisiin pureutuvan tutkimuksen perusteella rakennetun CAF-mallin mukaan kielitaito kehittyi koko ajan rinnakkain kolmella osa-alueella: samalla, kun opitaan tuottamaan uusia, kompleksisempia muotoja, jo haltuun otettujen muotojen tuottaminen käy sujuvammaksi ja tarkemmaksi (ks. esimerkiksi Housen & Kuiken 2009; Housen, Kuiken & Vedder 2012). Keskittyminen yhteen osa-alueeseen tyypillisesti kuitenkin heikentää muita ainakin väliaikaisesti, ja esimerkiksi kirjoitusnopeuden kasvattamisesta seuraa helposti sekä käytetyn sanaston ja rakenteiden yksinkertaistuminen että pulmia ilmaisun tarkkuudessa. Kouluiässä kirjoittaminen kehittyy monialaisesti, mikä omassa tutkimusaineistossani näkyy muun muassa tuottamisen runsautumisena sekä harvinaisten ja rakenteeltaan kompleksisten sanojen käytön tihentymisenä vuosiluokalta toiselle siirryttäessä samaan aikaan, kun ilmaisu tarkentuu ja tekstien kokonaisrakenne kehittyi (Honko 2013). Kielen sujuvuus, tarkkuus ja kompleksisuus eivät kuitenkaan etene samaan tahtiin tai suoraviivaisesti, ja osa kielelliseen ilmaisuun yhtä aikaa vaikuttavista tekijöistä voi olla vaikutuksiltaan vastakkaisia.<sup>77</sup> Yksi esimerkki on sanapituuden ja tarkkuuden suhde. Yleinen kielen

<sup>77</sup> Suomi toisena kielenä -oppijoiden kirjoittamistaitojen kehittymistä taitotasoitain on tutkittu mm. Cefling- ja Topling-hankkeissa Jyväskylän yliopistossa, ks. esimerkiksi Seilonen 2013, Mustonen 2015 ja Lesonen 2020, ensikielisten peruskouluikäisten kirjoitustaitoja puolestaan Myöhemmän kielenkehityksen hankkeessa (ks. Pajunen 2012).

kehittymisen tendenssi on, että oppimisen myötä ilmaisupituus kasvaa ja tarkkuus lisääntyy. Tämän tutkimuksen kertomusaineistossa tarkimpia ovat keskimäärin kuitenkin tekstit, joissa suhteellinen sanapituus on pieni.<sup>78</sup> Ristiriitaa selittäviä tekijöitä on monia: 1) Ensinnäkin lyhyet sanat ovat tyypillisesti kielen yleisimpiä ja tutuimpia sanoja, joiden kirjoitusasu hallitaan muita sanoja paremmin. 2) Tekstit, joissa keskimääräinen sanapituus on pieni, yleensä myös rakentuvat suhteellisen pienen sanamäärän varaan, jolloin hallittavien sanojen ja sananmuotojen määrä pysyy kohtuullisena. 3) Sanastoltaan monimuotoisissa teksteissä taas tarvitaan yleisten lisäksi harvinaisia sanoja. Harvinaisten sanojen käytön myötä sanapituus väistämättä kasvaa, mutta myös virheiden todennäköisyys lisääntyy.<sup>79</sup> (Ks. tarkemmin Malvern ja muut 2004, 173.) 4) Oppimisen alkuvaiheessa kirjoittamisen lyhytsanaisuus yhdistyy usein suffiksaation eli esimerkiksi sanojen taivuttamisen vähäisyyteen. Oppimisen edetessä kielen yleisimmät sanat ja tutut sanamuodot osataan jo hyvin, mutta etenkin harvinaiset sanat sekä sanojen yhdistyminen ja kontekstissaan oikean muodon valinta voivat aiheuttaa haasteita pitkään.

## Tarkkuuden mittaaminen

Tarkkuuden mittaaminen edellyttää mittayksikköä, jonka kautta osamista ja sen kehitystä tarkastellaan, ja valintaa yksikön esiintymien arvioinnissa. Määrittelen tässä luvussa oman lähestymistapani ja esittelen samalla tutkimuksen aineiston.

### MIKÄ ON VIRHE?

Se, mitä kielessä pidetään virheellisenä, on aina sopimuksenvaraista ja ajassa muuttuvaa. Sekä maallikot että kielen ammattilaiset voivat arvioi-

78 Lisäksi aikuismaisesta kirjoittamisesta poiketen tekstien rakenteellinen kompleksisuus yhdistyy alaluokilla keskimääräistä pienempään sanapituuteen, mitä mahdollisesti selittää tekstin tuottamisen kuormittavuus (Honko 2013, 339).

79 Pitkät sanat tarjoavat suuremman grafeemimäärän vuoksi suhteessa enemmän mahdollisuuksia virheisiin kuin lyhyet. Toisaalta sanan pituus ei automaattisesti tee siitä vaikeaa, mihin vaikuttaa suomen kielessä etenkin pitkien sanojen monimorfeemisuus.

da kieltä eri tavoin: toinen pitää virheellisenä (*inaccurate*) ilmausta, joka toisen mukaan on aivan kelvollinen (*acceptable*) tai korkeintaan kontekstissaan sopimaton (*inacceptable*). Virheet, jotka viestin yhdessä vastaanottajassa saavat aikaan negatiivisia tunnereaktioita, eivät vaikuta samalla tavalla kaikkiin saman viestin tulkitsijoihin. (Ks. esimerkiksi Pääkkönen 1985; James 1998; Boland & Queen 2015, 2016; Al-Mahrooqi & Al-Aghbari 2016). Äidinkieleltään muunkielinen esimerkiksi saattaa arvioida oppijankieltä sekä tarkkuuden että pragmaattisen sopivuuden suhteen jopa korkeasti koulutautunutta äidinkielistä puhujaa kriittisemmin (Al-Mahrooqi & Al-Aghbari 2016, 7).

Kirjoitetun suomen kielen tarkkuuden tarkastelun tekee erityisen vaikeaksi virheiden ongelmallinen luokittelu. Puutteet oikeinkirjoituksessa, morfologian hallinnassa ja sanojen sekä sanontojen merkitysten osaamisessa limittyvät toisinaan niin, että virheiden analysointi tarkkarajaisesti voi olla mahdotonta (ks. myös Tarnanen 2002, 127–128). Vaikka toisen ja vieraan kielen tutkimuksessa systemaattiset virheet (*errors*) sekä satunnaiset lipsut (*mistakes*) on usein pyritty erottamaan toisistaan,<sup>80</sup> ei tällainen eronteko kuitenkaan ole mahdollista analysoitaessa lyhyitä tekstinäytteitä, esimerkiksi tämän tutkimuksen aineistoa. En myöskään erottele kontekstiriippuvaisia ja kontekstista riippumattomia virheitä – eli esimerkiksi rektiosijojen väärin valittuja sijamuotoja (*me emme pidän*) väärin muodostetuista (*me emme pidän sinästä*).

James (1998, 64–78) erittelee kielellisten ilmausten tarkkuutta paitsi ilmausten laadun myös niiden hallinnan asteittaisen kehittymisen kautta. Toisessa ääripäässä ilmausta ei osata vielä lainkaan, toisessa ääripäässä sitä käytetään aina oikein. Tyypillisen ja epätyypillisen kielenkäytön erottavan linjan sen enempiä kuin säännönmukaisuuden rajapyykkien asettelu ei kuitenkaan ole yksiselitteistä. Tyylipoikkeamia, esimerkiksi puhekielen muotojen soveltamista kirjoitukseen, en ole omassa tutkimuksessani katsonut perustelluksi tulkita virheiksi lainkaan, sillä kirjoittajat vasta opettelevat kielen eri rekisterien konventioita. Lisäksi kertomusgenre jo itsessään mahdollistaa puhekielen käytön esimerkiksi

<sup>80</sup> Perusteena on pidetty mm. sitä, että lipsahdukset eivät noudata systematiikkaa, joka voitaisiin järjellä tavalla kuvata (ks. Corder 1967, 166–167).

henkilöhahmojen dialogeissa. Näyttää tosin siltä, että alakouluikäisten kertomustekstien dialogit ovat enemmän hypernormatiivisia ja noudattavat liian orjallisesti kirjoitetun kielen konventioita kuin ottavat vapauksia puhekielisyyden suuntaan, mikä muistuttaa Tannerin (2012) havainnotta oppikirjatekstien dialogeista.

Kirjoittamisen tarkkuutta mitattaessa on päädytty kolmenlaisiin menettelytapoihin (ks. esimerkiksi Ellis & Barkhuizen 2005). Holistinen arvio tekstin virheettömydestä on taloudellinen tapa saada tarkkuudesta yleiskäsitys. Se ei kuitenkaan anna lainkaan konkreettista tietoa kielellisen tuen tarpeesta eikä kerro sitä, kuinka laaja-alaisia pulmia kielen tarkkuudessa on. Jotkin virhetyypit, esimerkiksi sanojen sekoittuminen, saattavat myös painottua holistisessa arvioinnissa muita enemmän. Tämä ei kuitenkaan näy arviointituloksesta, joka on yleensä karkean mitta-asteikon mukainen numeerinen tai sanallinen arvio tarkkuuden tasosta. Arvioinnissa voidaan myös keskittyä tietyn piirteen tai tiettyjen piirteiden tarkkuuteen, jolloin saadaan tietoa juuri haluttujen kielenpiirteiden hallinnasta. Toisaalta valitut piirteet voivat edustaa hyvin suppeaa osaa oppilaan kieliopin tai sanaston hallinnasta puhumattakaan kokonaiskielitaidosta. Kolmas vaihtoehto tarkkuuden mittaamiseen on laskea virhetiheys eksplikoidulle virhejoukolle esittämällä näiden suhteen virheettömien sanojen tai lauseiden tai virkkeenkaltaisten T-yksiköiden prosenttiosuus. Tällöin ei yleensä oteta kantaa virheiden kokonaismäärään eikä virheiden laatua välttämättä eritellä tulosten esittelyn yhteydessä.

Yhdistämällä eri arviointitapoja, kuten tässä tutkimuksessa teen, on mahdollista verrata tarkkuutta erilaajuisissa aineistoissa ja pureutua tarkemmin muun muassa siihen, mihin muihin kielen osatekijöihin tarkkuus on yksilötasolla yhteydessä. Eri virhetyyppien jakauma voi toimia lähtökohtana toistuvien virhetyyppien tarkemmalle laadulliselle analyysille. Vastaavasti yksittäisten kielenpiirteiden tarkkuutta seuraamalla voidaan saada tietoa paitsi sen tyypillisestä vakiintumisajankohdasta oppijankieleen myös mahdollisista välivaiheista ja myöhemmistä uudelleenjärjestelyistä sekä tuen tarpeista silloin, kun tarkentumista ei tapahdu odotetulla tavalla.

Suomessa oppijankielen tarkkuutta on alimmilta vuosiluokilta lähtien perinteisesti havainnoitu sanelukirjoittamisen yhteydessä joko laskemalla oppijan tuottama virhemäärä tai virheettömien tekstisanojen määrä. Tilalle Kulju ja Mäkinen (2009, 511) ovat ehdottaneet sanan fonologisen kokonaisrakenteen hallinnan arvioimista. Diagnostiikan välineenä on käytetty myös erilaisia sanelukirjoituksen sovelluksia sekä muita strukturoituja testejä, kuten virheiden paikantamiseen ja epäsanoihin liittyviä tehtäviä. Näiden suositeltavuus on sidoksissa paitsi itse testien ominaislaatuun ja käyttötarkoitukseen myös koulukäynnin vaiheeseen ja testatavan ryhmän erityispiirteisiin (ks. esimerkiksi Santala, Saaristo-Rinta, Metsola & Klippi 2013). Tämän tutkimuksen verrokkikorpuksen perustuvassa tutkimuksessa Lehtinen (2013) erittelee S1-kirjoittajien tekstien tarkkuutta toisella, neljännellä ja kuudennella vuosiluokalla kirjoitetuissa Unelmakirjoitelmissa. Lehtinen ei kuitenkaan käsittele tekstien virheettiä suhteessa tekstien muihin ominaisuuksiin tai kirjoittamisen tarkkuutta yksilötason kielitaidon osa-alueena. Alakouluikäisten S2-oppiloiden teksteistä kattavaa tarkkuuden analyysiä ei tietääkseni ole tehty. Tässä luvussa havainnoidaan kirjoittamisen tarkkuutta monikielisessä lapsijoukossa sekä määrällisen että laadullisen analyysin avulla.

## AINEISTO JA ANALYYSIMETODI

Tutkimusaineistoni koostuu yhteensä noin 600 kertomustekstistä,<sup>81</sup> joista osa on suomea äidinkielenä puhuvien koululaisten ja osa edistyneiden,<sup>82</sup> lukuisia eri kieli- ja kulttuuriryhmiä edustavien suomi toisena kielenä -oppiloiden (S2) kirjoittamia (ks. myös Honko 2013). Kaikki kirjoittajat ovat yleisopetuksen oppilaita. Kertomuskeruun yhteydessä oppijoille painotettiin, että tekstin tuottamisessa sisältö on tärkein, mutta kehoitettiin silti pyrkimään kirjoitetun kielen sääntöjen noudattamiseen ja tarkistamaan teksti kirjoittamisen loppuksi. Tekstit kirjoitettiin käsin ja syötettiin myöhemmin Excel-tietokannaksi.

81 Tekstit on kirjoitettu yhtenäisellä teemalla *Unelmien päivä*.

82 Kaikki tutkimukseen osallistuneet suomi toisena kielenä -oppilaat ovat Suomessa syntyneitä tai ennen esikouluikää Suomeen muuttaneiden maahanmuuttajavanhempien lapsia, jotka ovat tutkimuksen alkuvaiheessa asuneet Suomessa vähintään neljä vuotta ja sen loppuvaiheessa vähintään seitsemän vuotta (Honko 2013).

Tarkkuuden analysoinnissa keskityn kuuteen kirjoittamisen ongelma-kohtaan (ks. myös taulukko 44):

- 1) pulmat äänteiden merkitsemisessä
- 2) sanojen yhteen ja erilleen kirjoittamisen virheet
- 3) taivutusvirheet (myös puuttuva taivutus ja virheellinen sanavartalo)
- 4) sanajärjestyksen virheet
- 5) virheellinen sanavalinta, taivuttamaton tai puuttuva sana
- 6) muu lause- tai virkerakenteen ongelma

Tarkkuuden analysoinnin taustaksi olen seurannut sekä suomea ensimmäisenä (S<sub>1</sub>) että toisena kielenä (S<sub>2</sub>) puhuvien lasten kielitaidon muutoksia myös muilla kirjoittamisen osa-alueilla, kuten kirjoittamisen sujuvuuspiirteissä ja käytetyssä sanastossa. Lisäksi olen kerännyt opettajilta tietoa lasten kokonaiskielitaidosta ja kielellisistä erityisvaikeuksista (ks. Honko 2013). Laadullisen virheanalyysin jälkeen vertaan kirjoittamisen tarkkuutta muihin kirjoittamisen osataitoihin tilastollisten analyysien avulla. S<sub>2</sub>-ryhmässä vertaan lasten yksilöllisiä kirjoittamistaitoja myös heidän kokonaiskielitaitoonsa käytettävissä olevan kyselytiedon avulla. Kyselytieto koostuu erillisestä suomen kielen sanaston hallintaan painotuvasta kielitestistä, S<sub>2</sub>-arvosanasta sekä lomakevastauksista, joissa suomen kielen ja oman äidinkielen opettajat arvioivat lapsen kielitaitoa kielitaidon eri osa-alueilla sekä mahdollisia kielellisiä erityisvaikeuksia. Olen käyttänyt määrällisissä analyyseissa lähinnä ei-parametrisia tilastollisia analyysejä, sillä ne soveltuvat aineistoni laatuun parametrisia testejä paremmin. Lisäksi olen tutkinut aineistoa sirontakuvioiden avulla mahdollisten epälineaaristen yhteyksien paljastamiseksi.

On syytä huomata, että vaikka kirjoittamisen tarkkuutta tarkastellaan käyttöpohjaisesti, virheiden syntymisen täsmällisiin syihin ei tässä tutkimuksessa päästä käsiksi. Esimerkiksi äännevirheet voivat olla seurausta foneettisen tietoisuuden puutteista (väärin kuultu sanahahmo), sanahahmon väärästä muistiedustuksesta (oikein kuultu sana muistetaan väärin) tai kirjoitusjärjestelmän hallintaan liittyvän dekodeauksen ongelmista (oikein muistettua sanaa ei vielä osata kirjoittaa oikein). Myös virheiden luokittelussa on toisinaan monitulkintaisuutta. Epävarmoissa tapauksissa olen katsonut ilmauksen mieluummin mahtuvan suomen

kielen variaation piiriin kuin poikkeavan siitä. Kun kirjoitettu muoto esimerkiksi reaalistuu partitiivia vaativassa lauseasemassa sanakirjamuodon kaltaisena taivuttamattomana sanana (*jalkapallo*), ei voida varmuudella sanoa, että kirjoittaja olisi pyrkinyt taivutusmorfologian käyttöön. Tapaukset on silti koodattu ainoastaan äännekestovirheenä, koska vokaaliyhtymän assimiloituminen (kuten *jalkapalloa* → *jalkapalloo*) on puhekielille tyypillinen piirre ja koska puhekielisyys muutoinkin voimakkaasti näkyy alakoululaisten kirjoittamisessa muun muassa juuri assimilaatioilmiöinä sekä loppukonsonanttien katona. Kielenhallinnan ja kirjoittamisen puutteet voivat toisin sanoen kirjoituksessa langeta yhteen, jolloin niitä ei voida luotettavasti erottaa yksittäisistä teksteistä.

Olen analysoinut tekstit yhdistetystä tekstikorpuksesta, jossa eri viiteryhmiä oppijoiden tekstit esiintyvät sekoitettuna. Analyysiä on ohjannut käyttöpohjaisuus, eli olen systemaattisesti ottanut huomioon myös poikkeavien muotojen esiintymiskontekstin. Tällainen lähestymistapa on oppijankielen tarkastelussa välttämätön, sillä vaikkapa urheiluselostusta jäljittelevä kielenkäyttö huudahduksissa *siisääään!* 'sisään' ja *maali!* 'maali' ei tietenkään ole osoitus kielen hallinnan puutteesta eivätkä ylipitkät vokaalit perusteltavissa kirjoituksen virheiksi.

Koska on selvää, että pelkkä virheiden lukumäärä ei eripituisia tekstejä tarkasteltaessa kerro kirjoittamisen tarkkuudesta koko totuutta, olen laskenut virheiden tekstitiheyden suhteessa kunkin tekstin kokonaisosan määrään ja osin myös suhteessa kyseisen kielenpiirteen kokonaisuuteen aineistossa.<sup>83</sup> Teknisesti koodaus on toteutettu niin, että olen voinut tarkastella aineistosta sekä virheellisten tekstisanojen karkeaa määrää (ottamatta kantaa virhetyyppiin tai eri virheiden mahdolliseen liittymiseen saneessa) että virheiden tarkempaa laatua ja kasautumista yksittäisiin tekstisanoihin. Virheiden tekstitiheyksiä olen sitten analysoinut sekä kirjoittaja- että kirjoittajaryhmäkohtaisesti.

Usein toistuvista virheistä välimerkkivirheitä ei tutkimuksessa ole koodattu virheiksi, eikä välimerkkien norminmukaista käyttöä voi vielä alaluokilla pitää odotuksenmukaisenaan.<sup>84</sup> Yksilölliset erot välimerk-

83 Ks. virheanalyysistä oppijankielen tutkimuksen yhteydessä esimerkiksi Gass & Selinker 1994, 66–74.

84 Koska sidossanojen käyttö on alaluokilla vielä melko vähäistä ja yksipuolista, voitaisiin useat virkerajamerkinnot korjata myös sidossanojen avulla.

kien käytössä ovat valtavan suuria, sillä vaikka jo tois- ja kolmasluokalaisissa on kirjoittajia, joille välimerkkien käyttö on sekä onnistunut virheettömästi että ollut monipuolista, eivät kaikki kirjoittajat ole käyttäneet välimerkkejä lainkaan tai niiden käyttö on yhä ollut horjuvaa ja tietämys esimerkiksi välimerkin sijainnista puutteellinen (ks. esimerkki 1).

1) *Olipakeran toive päivä. Aamulla minulle tuotiin sänkyyn. Lahjan kun avasin sen sieälä oli. uusi pallo sen jälkeen me mentiin. uimaan serenaan sielä oli. Vesiliukumäkiä ja kun me. Tultiin sieältä me lähdettiin Lasketelemaan minä otin mukaan, Oman lumi laudan sieälä oli. Hiihto hissejä siten me lähdetiin. pois ja mentiin linanmäelle, me mentiin vuoristorataan, ja kahvikupeihin ja peisotiin. ja karuselliin siten. sen jälkeen me lähdettiin. Vuokraamaan lefaa ja pidettiin.lefa ilta ja mentiin vasta. Kahelta yöllä nukumaan. minä vielä valvoin vähän aikaa ja nukahdin. Loppu. (3. lk. PS1)*

Ylemmillä vuosiluokilla päättövälimerkkien käyttäminen sujuu jo melko hyvin, mutta sivulausepilkkutuksen ja erikoismerkkien käytössä on yhä hyvin suuria yksilöiden välisiä eroja – kuten on välimerkit mahdollistavien rakenteiden käytössäkin. Runsaat välimerkkivirheet voivat siis olla merkki paitsi tarkkuuden ongelmista myös edistyneestä lause- ja virkerakenteiden käytöstä.

## Tarkkuus alakoululaisten kertomusteksteissä

Alakouluikäisten kirjoittaminen on kehittyvää ja tuottaa monenlaista tarkasteltavaa. Sanastossa ja rakenteissa on puhekielisyyttä ja ilmaukset voivat olla monitulkintaisia ja hankalasti ymmärrettäviä. Keskityn tässä yhteydessä tarkemmin sanatason ongelmiin, lähinnä normipoikkeamiin, ja tyypittelen niitä tiheyden, kirjoittajaryhmien ja toistuvuuden kannalta. Muita, esimerkiksi tyyliin liittyviä ongelmia enemmän esittelen taustaksi kuin analysoin virheiksi. Kertomusrakenne ei ole suurella osalla nuorimmista alakoululaisista vielä hallinnassa, samoin lause- ja virke-



rakenne voi ikätasolla tuottaa ongelmia, joten tarkastelen rakenneseikkoja tarkemmin vain vanhempien alakoululaisten kirjoitelmista.

#### PUHEKIELI OSANA KERTOMUSTEKSTEJÄ

Kertomustekstien poikkeamat kirjoitetusta yleiskielestä voidaan jakaa karkeasti kahteen pääluokkaan: normi- ja tyylipoikkeamiin. Normipoikkeamilla tarkoitetaan muotoja, jotka eivät kuulu pysyvänä osana mihinkään kohdekielen tunnettuun varianttiin vaan esimerkiksi syntyvät tilanteisina lipsahduksina tai ovat osoitus keskeneräisestä oppimisprosessista (esimerkit 2 ja 4). Tyylipoikkeamat puolestaan syntyvät kielen varianttien sekoittumisesta. Kirjoitetussa kielessä normipoikkeamia edustavat selkeimmin oikeinkirjoitusvirheet sekä virheet sanojen sijaituvuudessa ja vartalomuutoksissa, tyylipoikkeamia esimerkiksi varieteettien sekoittumisesta johtuvat leksikaaliset erikoisuudet, kuten leimallisesti puhekieliset ja lapsenkieleen kuuluvat sanat tai sananmuodot.<sup>85</sup> Etenkin nuorten koululaisten tekstejä arvioitaessa on usein olennaista erottaa puhekielisyydet (ks. esimerkit 2–3) muista poikkeamista, sillä vaikka lapset jo alaluokilla oppivat, että kielellä on useita ilmenemismuotoja, he eivät välttämättä pysty pitämään variantteja johdonmukaisesti erillään. Sen sijaan esimerkiksi sanan grafeemirakenteen virheet johtuvat yleensä sanahahmon vieraudesta tai foneemisen rakenteen dekodauksen ongelmista.

2) *Me juhlitaan* [vrt. \*me juhlivat] *ihanassa ravintolassa koko perhe.*

3) *Monet ihmetteli, että miksk mulla on kepit* 'kainalosauvat'.

vrt. virhemuodot \*ihmettelli, \*etta, \*mulle on, \*tikkut

Alakoululaisten teksteissä puhekielisyyys värittää kielen rakennetta ja näkyy myös sanavalinnoissa, vaikka kieli saattaa muutoin olla lähes virheetöntä. Puhekielisyyksien esiintyminen kertomusteksteissä on runsaimmillaan aivan kirjoitustaipaleen alussa mutta vähenee selvästi jo kolmannelle vuosiluokalle tullessa. Salosen (2012) tutkimus tosin osoittaa, että kehitys ei ole lineaarista vaan tarkentuvan, kirjoitetun yleis-

85 Puhekielen muotoja en tässä yhteydessä ole sisällyttänyt virheanalyysiin.

kielen piirteitä tunnollisesti noudattavan vaiheen jälkeen puhekielisyys uudelleen lisääntyy ja myös muuttaa muotoaan. Pojilla puhekielisyudet ovat alakoululaisten aineistossa hiukan yleisempiä kuin tytöillä ja S2-kirjoittajien teksteissä selvästi yleisempiä kuin S1-kirjoittajien teksteissä. Tässä suhteessa myös toisen ja vieraan kielen opiskelijoiden tekstit poikkeavat toisistaan, sillä vieraskielisessä ympäristössä kieltä opitaan tiiviimmin kirjoitetun kielenmuodon välittämänä ja oppijan arkiseksi pulmaksi voi muodostua ennemmin puhekielen varianttien puute kuin niiden ylikäyttö. Kielen eri rekisterien haltuun ottamisen myötä puhekielisyys muuttuu vähitellen yhä selvemmin tietoisesti käytetyksi työvälineeksi, jolla muun muassa elävöitetään dialogeja, puhutellaan lukijaa ja joskus jopa protestoidaan kirjoittamisen konventioita vastaan (ks. myös Halonen 2009 ja Niinimäki tässä teoksessa). Silti kielen varianttien vaihtelu on alakouluikäisten teksteissä usein vielä hallitsematonta.

#### KERTOMUSTEKSTIEN YMMÄRRETTÄVYYS

Kertova teksti on todettu kouluikäisille suhteellisen helpoksi tekstilajiksi; vähäisilläkin taidoilla onnistuminen on mahdollista, vaikka kirjoittaminen iän ja karttuvan kokemuksen myötä kehittyä ja esimerkiksi arvioiva, analysoiva ote yleistyy (ks. esimerkiksi Pajunen 2012). Tekstien päätehtävä eli kommunikatiivisuus täyttyy unelmapäiväteksteissäkin jo alaluokilla hyvin, sillä koko aineistossa ainoastaan muutaman yksittäisen tekstikohdan tulkinta jää epävarmaksi (esimerkiksi *tardo* 'automerkki').<sup>86</sup> Tekstien tulkitseminen edellyttää silti lukijalta maailmantietoa ja toisinaan myös kielellistä soveltamiskykyä. Tätä havainnollistaa seuraava kuudesluokkalaisten tytön kertomuskatkelma (ks. esimerkki 4), jossa limittyvät puhekielen redusoituneet muodot sekä taivutusmuotojen valinnan ja sanahahmon tarkkuuden pulmat:

4) *Ja unelmoin että kukaa meistä perheestä tai joku muu\_ ei tarvii olla jokin asma tai syöpä tai kaikki nämä jutut.* (6. lk. TS2, albania, 12 vuotta Suomessa)

86 Koko 60 000 saneen aineistossa epäselviä lekseemejä on vain kymmenkunta.

Toisinaan tulkinnanvaraisuuksissa on kyse laaja-alaisesta rakenteen tai tekstin jäsentelyn epäselvyydestä, mutta joskus myös yksittäisen leksikaalisen valinnan tulkinta aiheuttaa lukijalle päänvaivaa (esimerkiksi *Huoisin leikkiä 'haluaisin'/'voisin'*). Koheesion ja virketasoa laajemman kokonaisrakenteen pulmat on jätetty tarkkuusanalyysin ulkopuolelle. Tämäntyyppiset rakenteen pulmat vaikuttavat kuitenkin välillisesti kertomusrakenteen holistiseen arvioon, jota olen käyttänyt virhetihedyn vertailumuuttujana.

Joissakin virheellisen muodon sisältävissä tekstikohdissa tavoitemuodon tulkinta jää epävarmaksi. Esimerkissä 5 rakennevirheen taustalla saattaa olla taivutusmorfologian puutteellinen hallinta mutta myös tilanteeseen epäsovikivat leksikaaliset valinnat:

5) *Ja kato Big ben, juuri avatessa ikkunan verhot ja kivat naapurit.*  
(6. lk. PS2, vietnam)

Esimerkin 5 tulkinnaksi sopisi 'ja näin Big Benin heti, kun avasin ikkunanverhot' (Big Ben uutena puheenaiheena), mutta myös tulkinnat 'katsoin Big Beniä' tai 'ja katso' tai 'ja kuvittele' olisivat mahdollisia.<sup>87</sup> Esimerkissä 6 puolestaan subjektin tai subjektina toimivan nominilausekkeen merkitsemättä jättäminen on koheesion kannalta ongelmallista: Mikä on se, joka jatkui?

6) *Olin voittanut lotossa illalla urheiluauton ja veneen. Ja ei kuin vain jatkunut kun minä sain viestin, jossa luki ettei minun tarvitse enää koskaan käydä koulussa.* (6. lk. PS1)

Lauseiden ketjuttaminen on jo alaluokilla tyylillisesti ominaista muutamille taitavillekin kirjoittajille. Rakenteellisen ongelman ketjuttaminen muodostaa esimerkiksi silloin, kun lukijan odottama päälause jää virkkeestä puuttumaan ja semanttinen sisältö karkaa (ks. esimerkki 7):

<sup>87</sup> Merkityslaaientumista lapsenkielestä ks. esimerkiksi Tomasello & Akhtar 2000, 184.

7) *Kun meillä oli niin paljon rahaa ettemme voineet enää kantaa sitä mukana koska emme halunneet, että sitä valtavaa rahamäärää jonka voitimme, emme halunneet, että niitä varastetaan.* (6. lk. TS1)

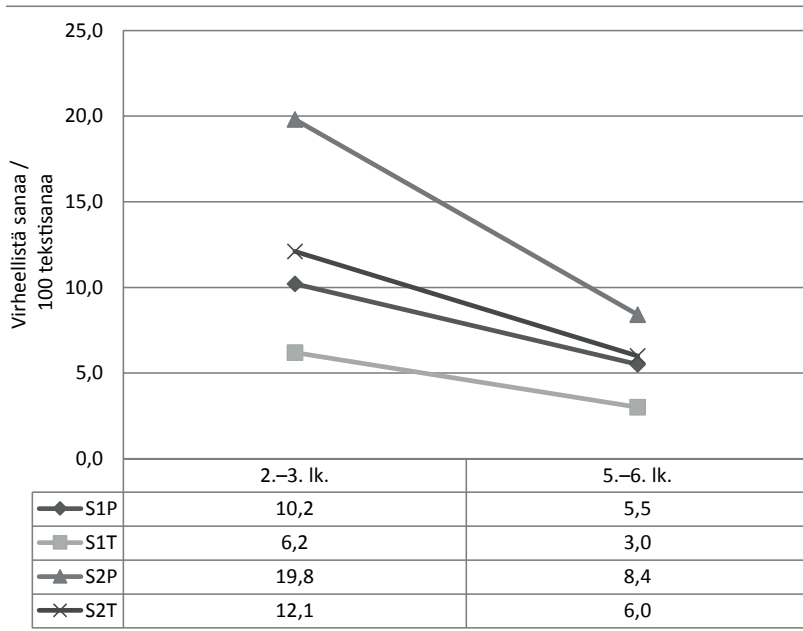
Alaluokilla monet tekstit ovat kuitenkin vielä sana- tai lausekelistoja, eikä lausetason rakenteellisille pulmille ole ollut luontevaa osoittaa määrää tai aina edes hahmoa (poikkeuksena jo mainitut erikseen koodatut virhetyypit). Siksi rakennevirheet on tässä tutkimusaineistossa koodattu ainoastaan viides- ja kuudesluokkalaisten kertomuksiin. Koska virkerakenteen puutteet yhdistyvät alaluokilla voimakkaasti muihin kielenhallinnan ja kirjoittamisen ongelmiin, ne heijastuvat joka tapauksessa tarkkuusanalyysin tuloksiin ja kertomusrakenteen holistiseen arvioon.

**VIRHETIHEYS SEKÄ SEN MUUTOKSET RYHMÄ- JA YKSILÖTASOLLA**  
Tarkkuuden yleisindikaattorina käytän virhetiheyttä, joka on laskettu virheiden määrän ja tekstisanojen kokonaissanemäärän suhteesta siten, että yksittäiseen saneeseen liittyvät eri virheet on huomioitu vain kerran.<sup>88</sup> Vaikka osa virheistä todennäköisesti on seurausta kielitaidon kompleksistumisesta (esimerkiksi rektiovirheet uusien verbien yhteydessä) ja kirjoitusjyvyyden kasvusta (esimerkiksi lipsut oikeinkirjoituksessa), ei silti ole tarkoituksenmukaista tai oletettavaa, että epätarkkuudet kielessä lisääntyisivät säännönmukaisesti tai hallitsisivat ilmaisua oppijoiden siirtyessä vuosiluokalta toiselle. Onkin mielenkiintoista seurata, miten oppijankielen tarkkuus muuttuu kolmen vuoden seuranta-jakson aikana, jolloin sekä sujuvuuspiirteissä (tuottamisrunsaus, lekseemi- ja sanepituus) että sanaston kompleksisuudessa (sanarakenne, harvinaisten sanojen käyttö ja käytetyn sanaston vaihtelevuus) tiedetään tapahtuvan kehitystä (ks. myös Honko 2013).

Suurin virhetiheys, keskimäärin noin 20 virheen sisältävää tekstisanaa sataa juoksevaa tekstisanaa kohden, on alaluokilla S2-poikien teks-

88 Tekstien tarkkuutta on L2-kontekstissa yleisesti tutkittu laskemalla virheettömien lauseiden, virkkeiden tai T-yksiköiden suhteellisia määriä (ks. esimerkiksi Unsworth 2008, 320). Nuorimpien kirjoittajien tekstiaineiston tarkkuuden tarkasteluun lause- ja virkeajattelu soveltuu huonosti ennen kaikkea siksi, että lause- ja virkerakenteiden laajuudessa ja kompleksisuudessa on suuria yksilöiden välisiä eroja.

teissä, pienin (noin kuusi sataa tekstisanaa kohden) puolestaan S1-tyttöjen teksteissä (kaavio 13). S1-poikien ja S2-tyttöjen ryhmien välillä ei lähtötasossa ole merkittävää eroa, vaikka eri virhetyypit painottuvatkin ryhmissä eri tavoin. Sen sijaan kieliryhmien sisällä tyttöjen ja poikien kielellinen tarkkuus poikkeaa toisistaan voimakkaasti, sillä poikien teksteissä virheellisiä sananmuotoja tulee tekstissä vastaan lähes puolet useammin kuin tyttöjen teksteissä.



Kaavio 13. Virhetiheyden muutos alakoulun alku- ja loppuvaiheessa kirjoittajaryhmittäin.<sup>89</sup>

Sanemäärään suhteutettu keskimääräinen virhetiheys laskee koko koululaisaineistossa jo toiselta kolmannelle vuosiluokalle lähes kolmanneksen

<sup>89</sup> Virhetiheys sisältää äänneiden merkitsemiseen ja äännevaihteluihin liittyvät virheet, yhteen ja erilleen kirjoittamisen virheet, taivutusvirheet (myös puuttuva taivutus), selkeät sanajärjestysvirheet ja sellaiset selkeät lausetason rakenteelliset virheet, joita ei voisi korjata välimerkein. Virhetiheys ei sisällä puhekielisyyksiä tai alkukirjain- ja välimerkkivirheitä.

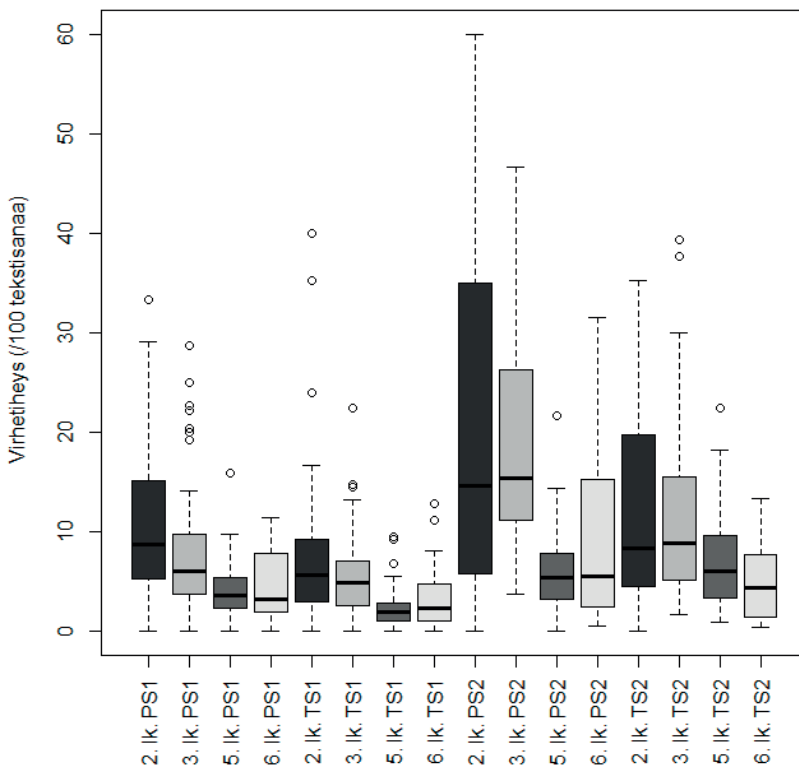
eli kahdestatoista noin kahdeksaan virheelliseen sanaan sataa juoksevaa tekstisanaa kohti. Ylemmillä vuosiluokilla virhetiheys on keskimäärin enää alle puolet alkuperäisestä eli noin viisi virheellistä tekstisanaa sataa juoksevaa sanaa kohti. Käytetyllä mittarilla arvioituna kirjoittaminen on viidennellä luokalla siis jo hyvinkin tarkkaa. Viidennen ja kuudennen vuosiluokan välillä virhetiheys ei kuitenkaan enää laske, eikä virhetyyppien keskinäisissä suhteissa tapahdu merkittävää muutosta.

Virhetiheyden muutos on selkeä kaikissa kirjoittajaryhmissä. Eniten kirjoittaminen on kolmen vuoden aikana tarkentunut S2-poikien ryhmässä, jossa lähtötaso oli heikoin. Kun vielä alaluokilla keskimäärin viidennes kaikista tekstisanoista sisälsi tässä ryhmässä yhden tai useamman virheen, on virhetiheys ylemmillä luokilla yli puolet pienempi. Absoluuttinen virheiden määrä on vähentynyt vielä enemmän, kun yksittäisiin saneisiin ei enää sisälly yhtä usein monia virheitä. Kaiken kaikkiaan virheellisten sanojen tekstitiheys on silti S2-oppijoilla vielä alakoulun loppuvaiheessa S1-oppijoita suurempi (Wilcoxonin testi  $W = 3718.5$ ,  $p < 0,001$ ) ja pojilla suurempi kuin tytöillä ( $W = 8161$ ,  $p < 0,01$ ).

Tuloksen tulkinnassa olennaista on muistaa, että lähes kaikki kirjoittajat ovat tutkimuksen ensimmäisessä vaiheessa vielä aloittelevia kirjoittajia, jolloin on luonnollista ja oletettavaa, että kirjoittamisen tarkkuudessa on paljon horjuntaa. Vaikka heikkojen kirjoittajien on perinteisesti havaittu poikkeavan vahvoista juuri suuremman virhemäärän perusteella, virheiden tiedetään myös kasautuvan tietyn tyyppisiin tekstiympäristöihin, esimerkiksi ylipitkiin tavuihin ja sanavartaloihin pääteaineksen sijaan (Kulju & Mäkinen 2009, 510–512, 522). Sekä aineis-toni laadullinen että määrällinen tarkastelu tukevat näitä havaintoja: edistyneiden alakouluikäisten teksteissä tavarakenteen karkeat virheet (kuten puuttuva, ylimääräinen tai tunnistamaton tavu) ovat vapaassa tuottamisessa kaikissa sana-asemissa ensinnäkin hyvin harvinaisia.<sup>90</sup> Lisäksi sana-asemalla ja sanan kokonaisrakenteella on merkitystä etenkin äänneiden kestoon liittyvien virheiden syntyymiseen: painottomassa asemassa epätarkkuudet ovat yleisempiä. Hajontakuvio (kaavio 14) puo-

90 Tutkimuksesta on rajattu pois dysfasia-diagnosin saaneet oppijat ja kokoaikaisessa erityisopetuksessa opiskelevat oppijat molemmista kieliryhmistä.

lestaan kertoo, että vaikka kirjoittajaryhmien väliset erot tarkkuudessa ovat suuret, jokaisessa kirjoittajaryhmässä on jo alaluokilla myös erittäin tarkkoja kirjoittajia.



Kaavio 14. Virhettiheyden kehitys eri oppijaryhmissä (julkaistu aiemmin Honko 2013, 153).

Jo varhaisessa kielenkehityksessä 2,5-vuotiailla S1-oppijoilla ilmenneet sanarakenteen prosessoinnin vaikeudet voivat näkyä kouluiässä lukutaidon tasossa (ks. tarkemmin Kulju & Mäkinen 2009, 512). Kirjoittamiseenkaan alkutaipaleella vaikeudet eivät ole satunnaisia, sillä aiemman tutkimuksen perusteella ne yhdistyvät muun muassa juuri hitaaseen ja takeltelevaan lukemiseen. Omassa tutkimuksessani kirjoittajaryhmien välinen järjestys kirjoittamisen tarkkuudessa säilyi samana alakoulun

edetessä. Näin on lähes kaikkien tarkasteltujen piirteiden hallinnassa, vaikka kussakin kirjoittajaryhmässä tapahtuu samaan aikaan huimaa kehitystä. Merkille pantavaa on, että myös yksilötasolla kirjoittamisen tarkkuus on 60 koululaisen seurantaryhmässä kohtalaisen stabiili ominaisuus, vaikka vertailtavana on kultakin kirjoittajalta vain yksi tekstipari. Spearmanin korrelaatioanalyysin tulos ( $r_s = 0,50$ ,  $p < 0,001$ ) kertoo, että virheettömmiin alaluokilla kirjoittaneet oppilaat ovat todennäköisimmin kirjoittaneet ikäluokkaan verrattuna tarkasti myös kolme vuotta myöhemmin.<sup>91</sup> Havainnot vastaavat lukemisen vaikeudesta saatua tutkimustietoa: vaikka yksittäisen kielenpiirteiden hallinnassa saattaa tapahtua nopeitakin muutoksia, tekstitaitojen kokonaisvaltainen kehittyminen on hidas prosessi, jossa roolit suhteellisen taitavien ja heikkojen kielenkäyttäjien välillä ovat melko pysyviä (ks. esimerkiksi Eloranta, Närhi & Ahonen 2009; Undheim 2009; Stothers & Klein 2010).

On muistettava, että kirjoittamisen tarkkuus ei aina ole merkki erinomaisesta kirjoittamisesta vaan se saattaa toisen ja vieraan kielen oppimisen kontekstissa yhdistyä tekstin leksikaaliseen ja rakenteelliseen yksinkertaisuuteen (ks. esimerkiksi Skehan 1998). Mukavuusalueelta kurkistelua – kuten monipuolisten sanavalintojen yhteyttä kirjoittamisen tarkkuuteen – voidaan tarkastella suuntaa-antavasti yksittäisten tekstien eri piirteitä vertaamalla. Tähän palataan seuraavassa analyysiluvussa, kun tekstien yksilöllistä virhetiheyttä verrataan muun muassa sanaston monimuotoisuuteen, harvinaisten sanojen osuuteen ja keskimääräiseen sanepituuteen.

## VIRHEETTÖMÄT TEKSTIT JA YKSILÖLLISET EROT

Kokonaan virheettömiä kirjoitelmia on 5.–6. luokan aineistossa vajaat 5 prosenttia kaikista kirjoitelmista ( $n_{PS1} = 4$ ,  $n_{TS1} = 9$ ) tai noin 6 prosenttia, mikäli alkukirjainvirheet jätetään huomioimatta. Ylemmillä vuosiluokilla kaikki virheettömät tekstit löytyvät S1-kirjoittajien aineistosta. Vaikka kirjoittamisen tarkkuus on kehityksellinen piirre, siihen liittyy

91 Esimerkiksi leksikaalisen diversiteetin eli sanaston monimuotoisuuden suoritustason pysyvyys on samassa aineistossa ja samoilla kirjoittajilla selvästi matalampi, vaikka sekin on tilastollisesti merkitsevä ( $r_s = 0,33$ ,  $p < 0,01$ ).



erottamattomasti sujuvuuden ja usein myös kompleksisuuden kärsiminen (Housen & Kuiken 2009, 465). Alimmilla vuosiluokilla virheetömät tekstit ovatkin molemmissa kieliryhmissä sekä sana- että lauserakenteiltaan yksinkertaisia ja tyypillisesti hyvin lyhyitä, monet vain yhden kokonaisen virkkeen mittaisia (ks. esimerkit 8–9).

8) *Minä unelmoin että minä osaisin mennä kauppaan.* (2. lk. PS2, albania)

9) *Minä nukuin, nukuin ja nukuin* (2. lk. PS1).

Silloin, kun tekstissä esiintyy virheitä, niitä esiintyy yleensä useissa sanoissa. S1-pojilla virheen sisältävien sanojen tyyppiarvot ovat yläluokilla 1 ja 3 (maksimi 18), S2-pojilla 1, 2 ja 8 (maksimi 34), S1- ja S2-tyttöillä 4 (maksimi arvot 21 ja 36). Tyyppiarvot ja maksimi arvot eivät erilaajuisissa aineistoissa ole suoraan vertailukelpoisia, mutta ne kuvaavat silti yksilöiden välisen vaihtelun suuruutta. Hajonta eri tekstien tarkkuudessa on erittäin suurta, sillä kun koko 5.–6. luokkien aineistossa virheen sisältävien tekstisanojen esiintymistiheys on noin 5,4 sataa tekstisanaa kohden, on keskihajonta lähes 5,0. S2-oppijoille monivirheisyys on kuitenkin tyypillisempää.<sup>92</sup>

Poikien kertomuksista vain kolme on kokonaan kirjoitettu virheettömällä yleiskielellä, sillä muissa puhekielen rakenteet ja sanasto joko liittävät kirjoitetun kielen muotoihin tai puhekielisyys hallitsee muotoja kauttaaltaan (esimerkki 10). Puhekielen konventioiden käyttö teksteissä on kuitenkin jo alaluokilla toisinaan tarkasti säädeltyä esimerkiksi niin, että se kattaa vain henkilöhahmojen puheenvuorot tai säilyy tyylillisesti eheänä minäkertojan äänessä läpi tekstin.

10) *Mä haluisin et mun elämäis ois ees yks koulupäivä jollo mua ei kiusattais. Ja mä toivoisin et saman päivän olis sisäliikkaa. Ja mä toivoisin et saman päivän ois ruokan perunoit ja nakkikastiket.* (3. lk. PS1, Uusimaa)

<sup>92</sup> Kun vähintään viisi virheen sisältävää tekstisanaa esiintyy 5. ja 6. vuosiluokalla keskimäärin puolella kirjoittajista, S1-ryhmässä osuus on 40 prosenttia ja S2-ryhmässä 64 prosenttia.

Kirjoitetun kielen konventioiden haltuun ottamisen haastavuudesta kertoo toisaalta se, kuinka myös hyvin taitavasti rakennetuissa teksteissä esiintyy epäsystemaattisesti käytettyjä puhekielen muotoja. Puhekielisyudet ovat lisäksi huomattavasti yleisempiä kertomusten loppu- kuin alkupuoliskoilla, mikä voi kertoa kirjoittajan väsymisestä tekstin loppua kohden.

S1-tytöillä virheettömät kertomukset ovat jo alaluokilla selvästi pidempiä kuin S1-pojilla, ja myös niistä löytyy puhekielelle tyypillisiä piirteitä. Poikien tekstien tapaan puhekielisyys esiintyy enemmän kertomusten loppu- kuin alkupuolella. S1-tyttöjen alaryhmässä on silti sekä alimmilla vuosiluokilla että alakoulun päättövaiheessa suhteessa eniten sellaisia vahvoja kirjoittajia, joiden teksteissä sujuvuus, tarkkuus ja kompleksisuus yhdistyvät: monet tekstit ovat sekä pitkiä, sanastoltaan ja rakenteiltaan monipuolisia että virheettömiä, ja kielen eri rekisterien käyttö on hallittua.

Sekä poikien suurempi virhetiheys että virheettömien tekstien pituus ja laatuero vahvistavat aiempien tutkimusten käsitystä poikien suuremmista vaikeuksista äidinkielellään kirjoittamisessa. Kielenhallinta on verrokkitestinä toimineessa sanastotestissä pojilla samaa tasoa kuin tytöillä tai jopa korkeampi (Honko 2013), mutta poikien kirjoittaminen on oikeakielisuuden ja rekisterivalintojen suhteen epätarkempaa ja rakenteeltaan eheät tekstit ovat harvinaisempia. Tekstitaitotutkimuksessa rajaukset kuten valittu genren tai tekstin käsittelemisen kanava voivat vaikuttaa tuloksiin. Silti sekä kansallisten arviointien että PISA-kartoitusten kautta on nähtävissä, että sukupuolten väliset erot luku- ja kirjoitustaidossa eivät tasoitu pelkästään tarkastelukontekstia modifioimalla (ks. esimerkiksi Lappalainen 2011; Arffman & Nissinen 2015).

#### YLEISIMMÄT VIRHETYYPIT

Virheiden luokittelu (ks. taulukot 44–45) osoittaa, että eniten tarkkuuden pulmia alakoululaisten kirjoituksessa aiheuttavat sanojen yhteen ja erilleen kirjoittaminen, kaksoiskonsonantit sekä isot ja pienet alkukirjaimet.<sup>93</sup>

93 Taulukoihin 44–45 on listattu vain kummankin keruuvaiheen 22 yleisintä virhetyppiä. Aineistoon koodatuista virhetyypeistä osa on jäänyt harvalukuisuuden vuoksi taulukon ulkopuolelle. Tällaisia piirteitä

Käsin kirjoitettaessa tyypillisiä ovat myös kirjainten piirtämiseen liittyvät haasteet. Lähes kaiken tyyppiset kirjoittamisen epätarkkuudet käyvät alakoulun loppuvaiheeseen tultaessa silti harvinaisemmiksi. Myös virheiden kokonaismäärä on pienempi, vaikka ylempien luokkien tekstiaineisto on sanamääräisesti noin kolmanneksen alempia vuosiluokkia suurempi. Muutaman yleisimmän virhetyypin (ylimääräinen ja puuttuva sanaraja, puuttuva geminaatta) osuus kaikista virheistä on kummassakin keruuvaiheessa suuri, mutta ylempillä luokilla virheiden jakauma on tässä suhteessa hiukan tasoittunut yhdysanavirheitä lukuun ottamatta. Taivutukseen ja sanavalintoihin liittyvät virheet ovat sikäli kiinnostavia, että vaikka nekin harvinaistuvat, niistä tulee muihin virheisiin nähden tyypillisempiä. Taivutuksen virheiden lisääntyminen voi liittyä siihen, että haltuun otetaan monimutkaisempia ilmaisekinoja samalla, kun kertomuksen sisällöllinen ja rakenteellinen tasokkuus kasvaa. Havainto vastaa dynaamisten systeemien viitekehyksessä tehtyä suomen kielen oppimisen kulkua esittelevää tutkimusta (Lesonen 2020).

Taulukko 44. Yleisimmät virhetyypit kertomusteksteissä 2.–3. luokalla.<sup>94</sup>

	Virhetyyppi	f	%	Kumulatiivinen tiheys / 100 sanaa
1	Ylimääräinen sanaraja	523	19,9	2,0
2	Puuttuva geminaatta	373	14,2	3,4
3	Puuttuva sanaraja	346	13,1	4,8
4	Alkukirjainvirhe <sup>95</sup>	236	9,0	5,7
5	Piirtovirhe	171	6,5	6,3
6	Puuttuva pitkä vokaali	116	4,4	6,8
7	Väärä taivutus	102	3,9	7,2
8	Puuttuva vokaali	77	2,9	7,5

ovat äänneiden erotteluun liittyvä vokaalin ja konsonantin sekoittuminen (esimerkiksi i/j) sekä sanan kirjoitusasuun liittyvät reversaali, rotaatio, reduplikaatio ja sanan osittainen toisto (kuten altaaseeneen 'altaaseen', voitisisin 'voittaisin').

94 Puhekielisyksistä virhetilastoissa ovat mukana ainoastaan virheellisenä lekseemivariantana tulkittavissa olevat tapaukset (*kun/kuin, vaan/vain, niin/tosi, joka/mikä/kuka*). "Muu lauserakenteen ongelma"-kohta kattaa sanatasoa laajemmat virhetyypit (esim. lausekkeen toisto tai puuttuva lauseke). Omina kohtinaan on koodattu yleisimmät lausetason ongelmat eli a) väärä sanajärjestys, b) puuttuva/ylimääräinen (yksittäinen) sana sekä c) virheellinen lauseenvastike. Kohta puuttuva/virheellinen taivutus sisältää myös kongruenssivirheet ja väärin valitut rektiosijat.

95 Alkukirjainvirheet eivät ole mukana virhetiheyttä koskeissa analyyseissa.

9	Puuttuva konsonantti	77	2,9	7,8
10	Väärä sana	77	2,9	8,1
11	Ylimääräinen geminaatta	73	2,8	8,3
12	Puuttuva sana (muu)	71	2,7	8,6
13	Puuttuva taivutus	62	2,4	8,9
14	Ylimääräinen pitkä vokaali	42	1,6	9,0
15	Väärä vokaali	41	1,6	9,2
16	Väärä konsonantti	37	1,4	9,3
17	Rauniosana	37	1,4	9,5
18	Irrallinen sananosa	68	2,6	9,7
19	Ylimääräinen vokaali	29	1,1	9,8
20	Ylimääräinen konsonantti	27	1,0	9,9
21	Soinnillisuus	25	0,9	10,0
22	Astevaihtelu	24	0,9	10,1
	Yhteensä	2634	100	

Taulukko 45. Yleisimmät virhetyypit kertomusteksteissä 5.–6. luokalla.

	Virhetyppi	Vrt. 2.–3. luokka	f	%	Kumulatiivinen tiheys /100 sanaa
1	Ylimääräinen sanaraja	1	323	19,0	0,9
2	Väärä taivutus	7	183	10,8	1,4
3	Puuttuva sanaraja	3	161	9,5	1,9
4	Alkukirjainvirhe	4	120	7,1	2,2
5	Puuttuva geminaatta	2	103	6,1	2,5
6	Väärä sana	10	93	5,5	2,8
7	Puuttuva pitkä vokaali	6	92	5,4	3,1
8	Piirtovirhe	5	82	4,8	3,3
9	Puuttuva sana (muu)	12	81	4,8	3,5
10	Ylimääräinen geminaatta	11	59	3,5	3,7
11	Puuttuva taivutus	13	56	3,3	3,8
12	Puuttuva konsonantti	9	46	2,7	4,0
13	Puuttuva vokaali	8	46	2,7	4,1
14	Ylimääräinen konsonantti	20	40	2,4	4,2
15	Muu lauserakenteen ongelma		40	2,4	4,3
16	Sanajärjestys		33	1,9	4,4
17	Astevaihtelu	22	33	1,9	4,5
18	Väärä konsonantti	16	31	1,8	4,6
19	Ylimääräinen pitkä vokaali	14	25	1,5	4,7
20	Väärä vokaali	15	22	1,3	4,7
21	Irrallinen sananosa	18	16	0,9	4,8
22	Rauniosana	17	13	0,8	4,8
	Yhteensä		1698	100	

## FREKVENTIT VIRHETYYPIT KIRJOITTAJARYHMITÄIN

Poikien kirjoittaminen on molemmissa kieliryhmissä ja useiden eri kielipiirteiden osalta keskimäärin epätarkempaa kuin tyttöjen: virheitä tehdään useammin (ks. kaavio 13). Tämä vastaa aikaisempien tutkimusten havaintoja kirjoitustaidosta eri ikävaiheissa (ks. esimerkiksi Mäki 2002; Pajunen 2012). Myös lukivaikeusriskiryhmässä oikeinkirjoitustaidoissa on havaittu alkuopetusvuosina eroja sukupuolten välillä tyttöjen eduksi, ja varhain lukemaan oppineilla pojilla kiinnostus kirjaimiin oli jo esiopetusvaiheessa tyttöjä vähäisempää ja tehtävää välttävä strategia tyttöjä yleisempi (Lerikkanen ja muut 2010). Koska poikien kirjoittaminen tarkastellussa aineistossa on myös sanamääräisesti niukempaa ja kertomusrakenteen saavuttaminen harvinaisempaa, on perusteltua kuvata poikien kirjoittamista tässäkin tutkimuksessa keskimäärin selvästi tyttöjä heikommaksi. Mikäli tarkastellaan yksinomaan kirjoittamisen virheitä, S1-pojilla vaikuttaa kuitenkin olevan ongelmia lähinnä saneen foneemisen esityksen koodaamisessa kirjoitukseksi,<sup>96</sup> kun S2-ryhmässä erot poikien ja tyttöjen välillä ovat vielä alakoulun alkuvaiheessa laaja-alaisempia. Tarkastelluissa S2-poikien teksteissä esimerkiksi sanojen taivuttamisen hallinta näyttää kehittyvän keskimäärin tyttöjä myöhemmin.

Maahanmuuttajataustaisia oppilaita opettaneet opettajat ovat listanneet tyypillisiksi S2-kirjoittamisen virhepiirteiksi pitkien vokaalien ja geminaattojen kestot, rektiovirheet, verbien ja adjektiiviattribuuttien kongruenssivirheet sekä korvakuulolta kirjoitettujen vasta opittujen sanojen virheellisen kirjoittamisen (Lyly 2008, 153).<sup>97</sup> Kirjoitetun kielen sijataivutuksessa lukioikäisillä S2-oppijoilla eniten ongelmia tuottaviksi piirteiksi on havaittu objektin sijalinnat ja paikallissijat (Elo 1998, 36–38). Mannisen ja Martinin (2000) mukaan venäjänkielisten ylioppilaskokelaiden tyyppivirheitä ovat kvantiteettiin, viittaussuhteisiin, sanavalintoihin, vokaalisointuun, kongruenssiin, objektiin, verbimuotoihin sekä sijataivutukseen liittyvät ongelmat. Vaikka S2-kirjoittajilla tiedetään olevan kirjoittamisen haasteita myös virketasolla muun muassa lausera-

96 Väite voidaan perustella myös sadutusnauhoitusten eli toisen ja kolmannen vuosiluokan oppilailta kerätyn suullisen kertomusaineiston avulla (ks. myös Honko 2017a ja 2017b).

97 Mahdollisia oman äidinkielen vaikutuksia kirjoittamisen puutteisiin opettajat eivät kuitenkaan välttämättä tarkemmin pysty erittelemään, mikä ei ole ihme, sillä tehtävä on lingvistillekin useimmiten vaikea.

kenteissa ja virkerajojen hahmottamisessa, suomen opettajat eivät pidä näitä leimallisina piirteinä juuri maahanmuuttajataustaisille oppilaille (Lyly 2008, 153). Kaikille oppilaille yhteisinä kirjoittamisen pulmapiirteinä opettajat ovat Lylyn kyselyssä maininneet välimerkkivirheet, virheet yhteen ja erilleen kirjoittamisessa sekä puhekielisyydet ja ongelmat viittaussuhteissa (*joka* ja *mikä*). (Lyly 2008, 153–154). Yleisistä virheistä on koululaiskirjoittamisen yhteydessä tapana mainita erikseen myös omistusliitteen merkitsemättä jättäminen (ks. esimerkiksi Elo 1998, 36; Turunen 2012). Tässä tutkimuksessa olen katsonut possessiivisuffiksin merkitsemättä jättämisen puhekielen vaikutukseksi ja varsinaisen virheanalyysin ulkopuolelle.

Kirjoittajaryhmittäinen tarkastelu kertoo, että sekä alakoulun alkuaikana loppuvaiheessa keskeiset virhetyypit ja niiden väliset voimasuhteet ovat ehkä hieman yllättäenkin pitkälti samanlaiset molemmissa kieliryhmissä. Äänneiden ja äännekestojen merkinnöissä, sanavalinnoissa ja taivutuksessa sekä erityisesti sanarajojen ja alkukirjainten merkinnöissä on pulmia paitsi S2-ryhmässä myös syntyperäisten kielenkäyttäjien ryhmässä. Unelmapäivätekstien virheitä ei toisin sanoen voida ainakaan ilman tarkentavaa laadullista analyysiä luontevasti luokitella siten, että jotkin virheistä olisivat vain S1- tai S2-kirjoittajien virheitä (ks. myös Hamp-Lyons 2003, 173), vaan kaikkia yleisimpiä virhetyyppejä esiintyy molemmissa kieliryhmissä. Näin on siitä huolimatta, että kyseiset virheet ovat S1-kirjoittajilla huomattavasti harvinaisempia ja liittyvät useammin puhekielen vaikutukseen eivätkä varsinaisesti kielenhallinnan puutteisiin eli esimerkiksi suomen foneemiston hallintaan. Kaikki S2-ryhmän kirjoittajat ovat asuneet Suomessa jo useita vuosia ja noin puolet heistä syntymästään saakka. Tämä tasoittaa mutta ei poista kieliryhmien välistä eroa.

Yleisimpien virhetyyppien kirjoittajaryhmäkohtaisiin taulukoihin (ks. taulukot 46 ja 47) on koottu koko aineiston tyypillisimmät virheet: toisen ja kolmannen luokan aineistossa 22 yleisimmän virhetyypin listalle mahtuvat kaikki koko korpuksessa yli 20 kertaa esiintyneet virheet, viidennen ja kuudennen luokan aineistossa kaikki yli 10 kertaa esiintyneet virheet. Luvuissa ei ole mukana puhekielisyyksistä johtuvia poikkeamia yleiskielen kirjoitusasuun eikä alkukirjainvirheitä, joiden määrä molem-

missa kieliryhmissä ja molemmissa keruuvaiheissa on hyvin suuri sekä virkkeenalkuisten että erisnimistä puuttuvien isojen kirjainten vuoksi. Juuri virkkeenalkuisten alkukirjainvirheiden koodaaminen olisi kuitenkin keinotekoista välimerkkivirheiden suuren määrän vuoksi.

Taulukko 46. Yleisimmät virhetyypit 2.–3. vuosiluokalla kirjoittajaryhmittäin.<sup>98</sup>

	Virhetyypin osuus %	S1	PS1	TS1	S2	PS2	TS2	Yht.
1	Puuttuva yhdyssana	27,4	24,2	30,3	10,4	9,4	11,8	19,6
2	Puuttuva sanaraja	16,9	20,7	13,5	9,2	9,5	8,6	13,4
3	Puuttuva geminaatta	11,2	10,7	11,7	21,9	23,4	20,0	16,1
4	Piirtovirhe	7,2	6,5	7,7	7,6	7,2	8,2	7,4
5	Irrallinen sananosa	4,1	4,5	3,8	1,5	2,0	0,9	2,9
6	Puuttuva vokaali	4,0	3,5	4,4	2,5	2,6	2,4	3,3
7	Puuttuva konsonantti	3,1	3,0	3,2	3,6	4,4	2,4	3,3
8	Väärä taivutus	3,1	2,8	3,3	5,9	5,1	7,1	4,4
9	Ylimääräinen geminaatta	3,1	3,7	2,6	3,2	3,1	3,3	3,2
10	Puuttuva pitkä vokaali	2,9	3,3	2,6	7,5	7,9	6,9	5,0
11	Väärä konsonantti	2,4	2,2	2,6	0,7	0,5	0,9	1,6
12	Väärä sana	2,2	1,7	2,7	4,6	3,8	5,8	3,3
13	Puuttuva sana	1,9	1,5	2,3	4,4	3,3	6,0	3,1
14	Soinnillisuus	1,7	1,5	2,0	0,3	0,2	0,4	1,1
15	Puuttuva taivutus	1,6	1,5	1,7	4,0	3,6	4,4	2,7
16	Rauniosana	1,4	2,3	0,6	1,8	2,8	0,4	1,6
17	Ylimääräinen vokaali	1,4	1,5	1,4	1,0	1,2	0,9	1,3
18	Väärä vokaali	1,4	1,7	1,1	2,3	2,0	2,7	1,8
19	Ylimääräinen konsonantti	1,2	0,5	1,8	1,1	1,2	1,1	1,2
20	Ylimääräinen pitkä vokaali	0,8	1,5	0,2	3,0	3,8	2,0	1,8
21	Astevaihtelun virhe	0,5	0,5	0,5	1,7	2,0	1,3	1,0
22	Väärä sanavartalo	0,4	0,7	0,2	1,7	1,2	2,4	1,0
	% yhteensä	100	100	100	100	100	100	100
	kpl yhteensä	1 258	599	659	1 059	608	451	2 317

98 Ei sisällä puhekielisyyskiä eikä alkukirjainvirheitä. Taulukossa on esitetty kunkin virhetyypin esiintymien osuus kaikista taulukkoon sisällytyistä virhetyypeistä. Tarkemmat selitteet löytyvät käsittelyluvuista.

Taulukko 47. Yleisimmät virhetyypit 5.–6. vuosiluokalla kirjoittajaryhmittäin.<sup>99</sup>

	Virhetyypin osuus %	S1	PS1	TS1	S2	PS2	TS2	Yht.
1	Puuttuva yhdyssana	32,4	29,5	35,1	11,2	13,6	9,4	19,2
2	Puuttuva sanaraja	16,4	21,5	11,3	6,6	7,5	5,9	10,3
3	Puuttuva sanaraja (muu)	6,1	4,4	7,4	4,6	6,1	3,6	5,2
4	Piirtovirhe	5,2	8,8	2,7	5,2	5,8	4,8	5,2
5	Puuttuva geminaatta	4,6	4,0	5,1	7,8	10,0	6,2	6,6
6	Ylimääräinen geminaatta	4,4	4,4	4,5	3,4	2,2	4,3	3,8
7	Väärä taivutus	3,9	3,2	4,5	16,4	14,8	17,6	11,7
8	Ylimääräinen konsonantti	3,5	2,0	4,8	1,9	0,7	2,8	2,6
9	Puuttuva konsonantti	3,2	4,4	2,4	2,8	2,9	2,7	2,9
10	Väärä sana	3,2	1,6	4,5	7,6	5,1	9,4	5,9
11	Puuttuva vokaali	2,9	4,0	2,1	3,0	4,4	2,0	2,9
12	Sanajärjestys	1,9	1,6	2,1	2,3	1,2	3,0	2,1
13	Väärä konsonantti	2,0	2,0	2,1	1,9	1,7	2,1	2,0
14	Puuttuva pitkä vokaali	1,7	0,8	2,4	8,4	6,3	9,9	5,9
15	Rakenne	1,7	1,6	1,8	3,1	2,9	3,2	2,6
16	Puuttuva taivutus	1,5	1,2	1,8	4,8	5,1	4,6	3,6
17	Pitkä vokaali	1,5	1,2	1,8	1,3	1,7	1,1	1,4
18	Ylimääräinen pitkä vokaali	1,4	1,2	1,5	1,7	1,0	2,3	1,6
19	Astevaihtelun virhe	0,5	0,4	0,6	3,1	2,9	3,2	2,1
20	Irrallinen sananosa	1,0	1,6	0,6	1,0	1,2	0,9	1,0
21	Sanarakenne	0,2	0,0	0,3	1,2	1,7	0,9	0,8
22	Ylimääräinen vokaali	0,8	0,8	0,9	0,6	1,2	0,2	0,7
	% yhteensä	100	100	100	100	100	100	100
	kpl yhteensä	592	251	336	975	412	563	1 567

## TOISTUVAT VIRHETYYPIT

Tässä alaluvussa esittelen tarkemmin aineistossa useimmin toistuvia virhetyyppejä: a) sanojen yhteen ja erikseen kirjoittamisen pulmia, b) ongelmia äännekestoissa, c) grafeemien valinnassa sekä d) grafeemien piirtämisessä ja siihen liittyen etu- ja takavokaalien erottamisessa, e) väärässä ja puuttuvassa taivutuksessa sekä f) sanajärjestyksessä ja g) sanavalinnoissa. Samalla esittelen kirjoittajaryhmien välisiä eroja kirjoittamisen tarkkuudessa ja sen kehittämisessä.

99 Ei sisällä puhekielisyyksiä eikä alkukirjainvirheitä. Taulukossa on esitetty kunkin virhetyypin esiintymien osuus kaikista taulukkoon sisällyvistä virhetypeistä.



#### SANARAJOJEN MERKITSEMINEN

Yhdyssanojen virheellinen merkintä eli osien kirjaaminen erillisinä on S1-oppijoilla tyypillisin virhe läpi alakoulun (*lajin vaihto koje, uima allas*), ja se on myös S2-ryhmässä kummassakin keruuvaiheessa toiseksi yleisin virhetyyppi. Jo alaluokilla yhdyssanoista on silti onnistuneesti kirjoitettu yhteen noin 60 prosenttia jokaisessa kirjoittajaryhmässä, ylemmillä vuosiluokilla poikien ryhmissä lähes 80 prosenttia ja tyttöjen ryhmissä yli 80 prosenttia. Tämä vastaa aikaisempia havaintoja yhdysanatajun kehittymisestä erityisesti kolmannen vuosiluokan jälkeen (Stenberg 2010).

Yhdyssanavirheet ovat lähes poikkeuksetta yhdyssubstantiivivirheitä, mikä johtuu muiden sanaluokkien yhdyssanojen vähäisestä esiintymismäärästä. Niihin yhdistyy molemmissa kieliryhmissä sekä tytöillä että pojilla muutakin kirjoittamisen epätarkkuutta – alemmilla luokilla noin 20 prosentissa tapauksista (*huipu auton, jatski kioski*), ylemmillä noin 10 prosentissa. Luvut peilaavat sekä pitkien sanojen kirjoittamisen vaikeutta, sillä virheitä on yhdyssanoissa enemmän kuin tekstisanoissa keskimäärin, että pitkienkin sanojen kirjoittamisen hallinnan kehittymistä alakoulun aikana.

Ylimääräisiä sanarajoja esiintyy alakoululaisten aineistossa myös yhdistämättömissä sanoissa, mutta nämä virheet eivät ole sattumanvaraisia. Ilmiö on yleisin koulunkäynnin alkuvaiheessa, jolloin niin liitepartikkeleita, johtimia kuin tavujakin kirjoitetaan välillä omiksi yksiköikseen. Molemmissa keruuvaiheissa useamman esiintymän tapauksia ovat additiivinen *-kin* (*ennen kin, kavereita kin*) sekä *-lAinen*-loppuiset adjektiivijohdokset (*kaiken laisia, saman laisia*). Toisella ja kolmannelle vuosiluokalle muutamit oppilaat myös tavuttavat vielä, tosin lähinnä paikallisesti, sanoja sanavälien avulla mahdollisesti oikeinkirjoituksen tai sanan taipumisen hahmottamisen helpottamiseksi (*per hee seen kuuluu, mun lem pi ruo ka on pa ler mon pasta, nukku ma an*). Vain hyvin pieni osa sanarajavirheistä viittaa silkkaan huolimattomuuteen, piittaamattomaan kirjoitustyyliin tai laaja-alaiseen sanarajojen hahmottamisen pulmaan (*siitäalkoisei kkailu*).

Sanojen ketjuttaminen on kertomusaineistossa tyypillisempää pojille kuin tytöille, jotka näyttävät oppivan sanarajojen tarkan merkitsemisen

keskimäärin varhemmin. Vaikka ketjuttaminen vähenee ylemmillä luokilla, puuttuu sanarajoja molemmissa kieliryhmissä enemmän poikien kuin tyttöjen teksteistä. Ketjutusvirheiden syntymiseen ja tulkintaan saattaa vaikuttaa tekstien kirjoittaminen käsin. Niissäkin on silti havaittavissa säännönmukaisuutta, kuten pikapuhemuototyyppisiä yhdistelmiä (*jolvoi*), määritteen ja edussanan yhdistelmiä (*kokopaikan, pahaolo, parituntia*) sekä taipumattomien pikkusanojen, esimerkiksi *kanssa*-postposition ja konjunktioiden sekä kieltoverbin yhdistämistä muihin sanoihin (*iskänkaa, juurikun, eituo*). Ketjuttamisvirheiden kasautuminen merkinnee, että sanojen ominaislaadulla ja toisaalta sanojen painottamisella on kirjoittamistavasta riippumattomia systemaattisia vaikutuksia siihen, mitkä sanat mielletään yhteen kuuluviksi.

Sanarajojen kirjoittamisen vaikeudet voivat olla satunnaisia lipsahduksia tai kertoa normien tuntemuksen puutteista tai piittaamattomuudesta normeja kohtaan. Myös sanan ominaispiirteet vaikuttavat sen kirjoittamisen vaikeuteen, sillä lyhyet ja nominatiivimuotoisen määritteen omaavat sanat kirjoitetaan yleensä muita tarkemmin (ks. myös Stenberg 2010). Sanojen yhteen ja erilleen kirjoittamisen taito kehittyä, kun kirjoitetun yleiskielen normit tulevat tutuiksi. Yhdyssanojen ja sanaliittojen kirjoittamisen tarkkuudella on kuitenkin tekstin pintarakennetta syvemmälle yltävä tehtävä: Sanarajojen tehtävä on auttaa kirjoituksen kuvaaman käsitteistön ja näistä syntyvien merkityskokonaisuuksien hahmottamista. Jos muodon ja merkityksen yhteys horjuu, lukija joutuu ponnistelemaan merkitysten tulkinnassa. Sanojen yhteen ja erilleen kirjoittaminen ei ole ”pelkkä oikeinkirjoitusasia”, vaan kyse on myös tavasta hahmottaa sanarakenteita ja jäsentää maailmaa kielen keinoin (Eronen, Maamies & Räikkälä 1996). Yhdyssanojen kirjoittaminen kertoo, miten hahmotamme maailmaa, ja tyypillisessä tilanteessa yhdyssana ilmaisee yhden käsitteen.

#### ÄÄNNEKESTOT

S2-ryhmässä eniten epätarkkuutta on alaluokilla geminaattojen kirjoittamisessa. S2-pojilla geminaatta on jäänyt yksinäiskonsonantiksi yli puolet useammin kuin yhdyssana kirjoittamatta yhteen. Ero on merkittävä, sillä S1-pojilla suhde on lähes yhtenevästä yhdyssanatiheydestä huoli-

matta päinvastainen. Myös S<sub>2</sub>-tytöillä geminaattavirheet ovat virheistä yleisimpiä, vaikka niiden tekstitiheys on pienempi kuin pojilla.

Suhteellisesti eniten vaikeuksia on tuottanut geminaatta *kk* (*jääkieko, karkia, marokoon*). Yleisiä ovat myös virheet merkinnöissä *ll, mm, nn, pp, ss* ja *tt* (*sielä, myöhemmin, Särkäniemeen, apelsiini, kansa, kilti*), harvinaisempia muutoinkin harvinaisemmissa yhdistelmissä *ff* ja *rr* (*lefaa, harastamaan*). Tuloksia voivat selittää oppijan äidinkielen ja mahdollisesti hallitsemien muiden kielten suomesta poikkeava äänteiden yhdistyminen: kaksoiskonsonantteja ei välttämättä ole tai niiden joukko on rajatumpi. Geminaattatyypeillä voi olla myös akustisia eroja, jotka vaikuttavat tuloksiin. Yksi vaikeutta selittävä tekijä saattaa lisäksi löytyä suomen kielen astevaihtelun yliyleistämistä.

Alaluokilla kesto-*virheitä* on toisinaan useita samassa sanassa (*tykän, pikusiskole, joukuessa*). Tähän voivat vaikuttaa niin pitkien sanojen kompleksisuus (prosessoinnin kuormittavuus), harvinaisuus (ei-tuttuus) kuin silkka sanapituuskin (virheitä mahtuu sanaan enemmän). Toisinaan kirjoittaja on hahmottanut, että sanaan kuuluu jokin pitkäkestoinen äänne, mutta pitkän keston sijainti on paikallistunut väärin (*etää, leenän, mennisime, semoisselle*). Tällaiset kompensaatiot näyttävät kasautuvan samoille kirjoittajille ja olevan ominaisia alimmilla vuosiluokilla käsillä olevalle kehitysvaiheelle. Myös oikein merkittyyn pitkään vokaaliin yhdistyy usein geminaattavirhe, ja vastaavasti vokaalin kesto-*virhe* reaalistuu usein sananmuodossa, jossa on geminaatta (*kunnella, liikkelle*). Ilmiö kuvastaa kesto-*jen* merkintöjen tarkentumisen asteittaista luonnetta: silloinkin, kun kirjoittaja tiedostaa kahdella kirjainmerkillä kirjoittamisen konvention idean ja erottaa ainakin osan puhutun kielen kesto-*eroista*, hän ei välttämättä erota pitkiä kesto-*ja* kaikista äänneistä tai kaikista sana-*asemista* tai merkitse kesto-*ja* säännönmukaisesti.

Sekä sukupuoli että kielitausta vaikuttavat geminaattavirheiden suhteelliseen määrään, vaikka yksilölliset erot ovat hyvin suuria. Alaluokilla geminaatto-*jen* onnistumisprosentti on S<sub>1</sub>-kirjoittajien teksteissä sekä tyttö-*jen* että poikien ryhmissä keskimäärin yli 95 kaikissa muissa yhtymissä paitsi *mm:ssä*, joka on pojilla jäänyt yksinäiskonsonantiksi noin joka kymmenennessä tapauksessa. S<sub>2</sub>-ryhmän pojilla geminaattavirheet ovat yleisimpiä: eniten virheitä ovat aiheuttaneet *kk* (onnistumisprosentti

noin 60 %, S<sub>2</sub>-tytöillä vastaavasti 85 %), *pp* (70 % / > 95 %), *ss* (70 %, > 95 %), *tt* (83 %, 93 %), *mm* (87 %, 71 %; mukana ei ole monikon 1. persoonaa korvaavia passiivimuotoja) ja *ll* (90 %, 93 %).

Vaikka geminaattavirheet ovat suhteellisen yleisiä, alaluokillakin vain viiden oppijan tekstissä niitä esiintyy runsaasti, yli 10 kappaletta. Kyseisistä oppijoista yksi on toisluokkalainen S<sub>1</sub>-tyttö, yksi kolmasluokkalainen S<sub>1</sub>-poika, kaksi arabiaa äidinkielenään puhuvaa toisluokkalaista poikaa ja yksi toisluokkalainen vironkielinen tyttö. Havainnot ovat linjassa aiemman tutkimustiedon kanssa: Geminaattojen tiedetään yleisesti tuottavan pulmia aloitteleville kirjoittajille (Kulju & Mäkinen 2009, 510–516, 524). Toisaalta geminaattavirheet saattavat olla merkki lukivaikkeudesta tai S<sub>2</sub>-ryhmässä suomen kielen kestoasteiden omaksumisen keskeneräisyydestä. Oli merkintöjen epätarkkuuden taustalla mitä syitä tahansa, niiden vaikutukset ovat alakoulun loppuun mennessä jo hyvin vähäiset: 5.–6. luokalla S<sub>2</sub>-poikienkin ryhmässä kaikkien geminaattojen merkintä on onnistunut yli 90-prosenttisesti. Vaikeimpia ovat alakoulun loppuvaiheessa tässä kirjoittajaryhmässä olleet *nn* (oikein 92 %) sekä *kk* (oikein 94 %).

Pitkien vokaalien merkinnät ovat kertomusteksteissä jo koulunkäynnin alkuvaiheessa huomattavasti tarkempia kuin pitkien konsonanttien merkinnät, ja jokaisen vokaalin merkinnät tarkentuvat edelleen alakoulun aikana. Vokaali itsessään vaikuttaa suuresti virheiden määrään: *yy:n* ja *öö:n* merkinnöissä ei esiinny virheitä, koska tavoitemuotoinkin nämä ovat harvinaisia. Määrällisesti eniten virheitä on molemmissa keruuvaiheissa aiheuttanut *aa:n* kirjoittaminen. Alaluokilla *ee:n* suhteellinen virheprosentti on kuitenkin suurin (noin 3,4 %). Myös sana-asemalla on merkitystä, sillä pitkät vokaalit redusoituvat kirjoituksessa lähinnä painottomista asemista sananlopuista ja keskeltä, eivät sanojen alusta. Alakoulun päättövaiheessa enää muutama tapauksista on sanassa (pää) painollisessa asemassa.

Pitkiin vokaaleihin liittyvät virheet ovat Suomessa pitkään asuneilla S<sub>2</sub>-oppijoilla huomattavasti yleisempiä kuin S<sub>1</sub>-ryhmässä. Vielä toisella ja kolmannella vuosiluokalla pitkän vokaalin redusoituminen lyhyeksi on lisäksi selvästi yleisempää S<sub>2</sub>-pojilla kuin S<sub>2</sub>-tytöillä ja hieman yleisempää pojilla myös S<sub>1</sub>-ryhmässä. Alakoulun loppuvaiheessa tyttöjen ja poikien välinen ero on kuitenkin kadonnut molemmissa kieliryhmissä.

Pitkien vokaalien ja konsonanttien kirjoittamisen haastavuus etenkin koulunkäynnin alussa on todettu myös lukuisissa sanelukirjoittamiseen perustuvissa aikaisemmissa tutkimuksissa. Näissä tutkimuksissa on havaittu, että sanan harvinaisuus ja taivutustyyppin tunnusmerkisyys sekä tavun pituus kasvattavat virheen todennäköisyyttä kestojen merkinnässä (ks. Kulju & Mäkinen 2009, 510–511). Sanelukirjoittamisessa geminaattojen ja konsonanttiyhtymien kirjoittaminen on aiemminkin todettu äidinkielisille aloitteleville kirjoittajille pitkien vokaalien ja diftongien kirjoittamista haastavammaksi (mts. 520).

Virhetilastoissa huomiota kiinnittää se, että ylimääräisten geminaattojen (esimerkiksi *halussi*, *päästä*, *omaann*) tekstitiheys on S1-kirjoittajien teksteissä yhtä suuri kuin puuttuvien pitkien vokaalien, vaikka S2-ryhmässä pitkien vokaalien merkintöjen puutteet ovat huomattavasti yleisempiä. Molemmissa kieliryhmissä ylimääräisistä geminaatioista suuri osa liittyy alakoulun päättövaiheessa jo joko astevaihtelun puuttumiseen (*pikkuserkkun*), ylyleistämiseen (*kauppungille*) tai liitepartikkelin sisältävien sananmuotojen kirjoitusasuun rajageminaation yhteydessä (*kenellekkään*). Ylimääräiset pitkät vokaalit (*uneessa*) puolestaan ovat molemmissa kieliryhmissä harvinaisia, ja suhteessa eniten tällaisia virheitä ovat tehneet S2-pojat alakoulun alaluokilla. Erityisryhmän muodostavat vierassanat, joiden toisen tavun vokaali tai toisen tavun alkukonsonantti on kirjoitettu pitkänä (*Austraaliaan*, *robootti*; *vamppyyri*, *internetin*).

Kertomusaineiston pohjalta voidaan arvioida, että myös osalla Suomessa syntyneistä S2-oppijoista on alakouluvaiheessa jonkin verran vaikeuksia paitsi pitkien äännekestojen merkinnöissä myös niiden havaitsemisessa. Akustisesti selvimmät tapaukset (painolliset sananalkuiset asemat) ovat alakoulun päättövaiheessa kaikissa kirjoittajaryhmissä jo hyvin hallussa. Horjunta sekä pitkistä äännekestoista lyhyisiin että lyhyistä pitkiin saattaa myös tarkoittaa, että oppijat, jotka tiedostavat äännekestojen merkitystä erottavan tehtävän ja merkitsemisen tärkeyden, päätyvät toisinaan myös ylikorjaamaan kirjoitustaan: he hahmottavat kestojen merkitsemisen systeemitasolla mutta eivät vielä kuule tai osaa merkitä kestoja tarkasti.

#### GRAFEEMIN VALINTA

Puuttuva konsonantti (muu kuin geminaattavirhe) on koko aineistossa yleisempi konsonanttiyhitymissä kuin vokaalien välissä. Puolivokaalin *j* puuttuminen tekijänimissä selittää jälkimmäisistä tapauksista noin puolet (*palvelia, tutkiana*). Konsonanttiyhitymissä etenkin akustisesti heikko *h* on jäänyt toistuvasti merkitsemättä (esimerkiksi *pereeni* 'perheeni', *voikaisi* 'voihkaisi'), samoin *η*-äänteen *n* (esimerkiksi *kunigas, sägystä*). Yleisiä nämä virheet eivät tutkimuksessa aineistossa silti ole. Ylemmillä vuosiluokilla selkein yksittäinen konsonanttien poisjääntiä aiheuttava tekijä näyttääkin olevan konsonantin akustisen havaittavuuden heikkous.

Virheellisesti valittujen konsonanttien joukko on hyvin kirjava. Läpinäkyvimpiä kirjoittamisprosessin kannalta ovat vierasperäisten kirjainten merkintöjen pulmat (*asvalttia* 'asfalttia', *twstä* 'tv:stä', *Nikaracuassa* 'Nicaraguassa') sekä *n/m*-merkintöjen epätarkkuus (*esin* 'esim.', *kauenpana* 'kauempana'), joita esiintyy kaikilla luokka-asteilla. Lähes yhtä yleisiä ovat alakoululaisten teksteissä satunnaisen oloiset kirjainten sekoittumiset (*joulupuppi* 'joulupukki', *koka* 'joka', *kuninhas* 'kuningas', *savoin* 'sanoin').

Syyt vokaalien sekoittumiseen pirstaloituvat samaan tapaan kuin virheellisten konsonanttien tapauksessa. Kirjaimen muodon epätarkkuus selittää muutaman tapauksista (*lappuun* 'loppuun'), ja osa *S2*-kirjoittajien vääristä vokaaleista näyttäytyy tekstissä todennäköisesti kielitaustasta johtuviksi. Vokaalit sekoittuvat silti vain harvakseltaan molemmissa kieliryhmissä.

#### PIIRTOVIRHEET JA VOKAALISOINNUN HALLINTA

Molemmissa kieliryhmissä yli 80 prosenttia kaikista piirtovirheistä on puuttuvia *ä:n* pisteitä, muita esimerkiksi puuttuva *t:n* viiva. Alimmilla vuosiluokilla *ä:n* ja *ö:n* pisteiden merkitsemättä jättäminen lienee melko yleisesti oikeinkirjoituksen lipsahdus, vaikka puutteellinen suomen kielen vokaaliharmonian hallinta saattaa ilmentyä tekstissä yhtenevällä tavalla, eikä vokaaliharmonian hallinnassa ilmene edistyneillä *S2*-oppijoilla tai äidinkielisillä kirjoittajilla juuri edes satunnaisia puutteita.

Molemmissa aineistoissa sananalkuiset ja sanansisäiset *ä:t* ovat redusoituneet *a:ksi* loppuasemaista *ä:tä* useammin, mikä on loogista

kirjoitussuuntaa ja kirjoitusjärjestystä ajatellen: monille kirjoittajille on tyypillistä palata merkitemään pisteet sanahahmon valmistuttua. Molemmissa kieliryhmissä *a/ä*-vaihtelua esiintyy sekä sanavartaloissa että sijapäätteissä ja enimmäkseen aivan tutuissa mutta myös harvinaisemmissa sanoissa. Molempien kieliryhmien teksteissä esiintyy tapauksia, joissa virheeseen yhdistyy muita äännetason virheitä. S2-ryhmässä useamman erityyppisen virheen kasautuminen samaan tekstisanaan on tosin kaiken kaikkiaan yleisempää. Ylemmillä vuosiluokilla piirtovirheet ovat samantyyppisiä kuin alaluokilla, mutta niiden tekstitiheys on molemmissa kieliryhmissä selvästi vähentynyt. S2-teksteissä niiden määrä on yhä tekstimassaan suhteutettuna noin kaksinkertainen verrokkiaineistoon verrattuna.

#### TAIVUTTAMISEN ONGELMAT

Kontekstissaan virheellisiä taivutusmuotoja esiintyy kaikkien kirjoittajaryhmien teksteissä. Taivutuksen puuttuminen (nomini perusmuodossa tai verbi 1. infinitiivissä) on jo alaluokilla kuitenkin harvinaisempi virhe kuin muuhun virheelliseen taivutusmuotoon päätyminen. Yleisimpiä taivutusvirheet ovat S2-ryhmässä, jossa siinäkin taivutuksen virhetiheys jää noin yhteen sanaan sataa tekstisanaa kohden. S1-kirjoitelmissa taivutusvirheiden esiintymistiheys laskee alakoulun aikana edelleen jo lähtökohtaisesti hyvin matalasta tasosta (tekstitiheys 0,2 → 0,1). S2-oppijoilla taivutusvirheiden tekstitiheys hiukan kasvaa (0,9 → 1,0), mutta kun tarvittavan taivutusaineuksen puuttuminen kokonaan on aiempaa harvinaisempaa, kokonaisuudessaan taivuttaminen on tarkentunut (1,4 → 1,3). Näin on siitä huolimatta, että teksteissä käytetty sanasto on monipuolistunut, lekseemi-, sane- ja tekstipituudet kasvaneet ja kompleksisten rakenteiden käyttö yleistynyt.

Nominien taivutusvirheistä eniten on objektivirheitä, joita esiintyy S2-kirjoittajilla paikallissijan valintavirheitä enemmän mutta S1-kirjoittajilla vain muutamia. Objektin sijavalinta ja muodostus tiedetäänkin yhdeksi suomen kielen oppijoita eniten haastavaksi rakennepiirteeksi (ks. esimerkiksi Martin 1995; Spoelman 2013). Nominivirheistä subjektin vääriä sijavalintoja sekä vääriä paikallissijamuotoja esiintyy niin ikään harvakseltaan molempien kirjoittajaryhmien teksteissä. Sekä nominien

että verbien taivutuksessa esiintyy myös jonkin verran yksikön ja monikon valintaan liittyviä luvun virheitä, vaikka niitäkin korpuksen kokoon nähden vähän. Lukuun ja subjektin sijavalintaan liittyviä virheitä S2-oppijoiden teksteissä ei esiinny juuri verrokkiryhmää useammin, paikallissijan valinnassa sen sijaan on vaikeuksia hiukan enemmän. Kongruenssivirheitä (esimerkit 11–12) esiintyy kaikkien kirjoittajaryhmien teksteissä niin alimmilla luokilla kuin alakoulun loppuvaiheessaakin. Kongruenssivirheet eivät kuitenkaan ole systemaattisia yhdelläkään kirjoittajalla, ja niiden kokonaismäärä on vähäinen.

11) *onhan meilläkin tänä viikollakin kivaa* (2. lk. PS1)

12) *Ei olisi koulua puoli vuoteen* (5. lk. TS2, vietnam, SS)

Verbintaivutuksessa pulmat kietoutuvat lähes aina kompleksisten rakenteiden, kuten kieltomuotojen ja verbiketjujen, yhteyteen. Satunnaisia puutteita on silti molemmissa kieliryhmissä myös predikaattiverbin persoonamuotojen valinnassa ja satunnaisesti moduksen ja tempuksen valinnassa.

Sanojen taivuttaminen ei ole kertaluonteisesti opittava ilmiö. Yksi ja sama kielenoppija saattaa ymmärtää jonkin taivutuspuutteen funktion, vaikka ei itse vielä käytä päätettä lainkaan, tai hän saattaa käyttää sitä puutteellisessa muodossa tai vain tutun konstruktion osana. Senkään jälkeen, kun tietty taivutusmuoto on ilmaantunut oppijan kirjoitukseen, käyttö ei siis välttämättä ole kattavaa. Äidinkielenään arabiaa puhuvan mutta Suomessa syntyneen toisluokkalaisen tytön kertomustekstin esimerkit 13–15 havainnollistavat ilmiötä: monisanaisissa nominilausekkeissa edussana on toistuvasti päätteetön. Toisaalta kongruoiva määrite taipuu näissäkin tapauksissa usein oikein ja myös edussanasta esiintyy virheettömästi taivutettuja objektimuotoja.

13) *Eka on se että haluaisi oman kannettava tietokone.* – – *toivoisi vielä oman leikki huone.*

14) *Menin kauppaan ja ostin vaatteita ja ostin kenkä ja sain jäätelöä.*

15) *Neljäs on se että saisi ison trampoliinin. Ja kultainen talo.*



Rektiovirheitä esiintyy niin ala- kuin yläluokilla ja molemmissa kieliryhmissä tytöillä ja pojilla, vaikkakin jälleen tiheämmin S2-teksteissä. Lähes kaikki rektiovirheet liittyvät verbitäydennyksien taivutusmuodon valintaan, vaikka muitakin esimerkkejä löytyy (*Talo oli tehty kultalla, se unelmoi kissaa; kolmekymmentä euron maksava, ennen lähdömmä, kaikki ikäiset* 'kaiken ikäiset').

Kaikki S2-oppijoiden teksteissä useita kertoja esiintyvät taivutusvirhetypit ovat sellaisia, joiden oppiminen aiemman tutkimuksen perusteella tiedetään haasteelliseksi (esimerkiksi Martin 1995; Ivaska 2009; Niiranen 2010; Reiman 2011; Spoelman 2013; Mustonen 2015). Subjektin sijavalintavirheissä huomiota kiinnittää partitiivisubjekti. Kieliryhmille yhteinen piirre on, että partitiivista esiintyy muutaman kerran sekä yliettä alikäyttöä. Mitenkään yleisiä subjektin sijavalintaan liittyvät pulmat eivät tässä edistyneiden kielinoppijoiden ryhmässä kuitenkaan ole. Paikallissijoista ovat tuottaneet vaikeuksia lähinnä olo- ja tulosijat (inessiivi, adessiivi, illatiivi, allatiivi), mutta kaiken kaikkiaan tämäntyyppiset valintavirheet ovat jo alaluokilla harvinaisia molemmissa kieliryhmissä, ja hyvin harvoin paikallissija on kokonaan merkitsemättä (ks. esimerkit 16–18). Lause- ja virkerakenteiden kompleksistuessa paikallissijojen valintaan liittyvien virheiden määrä on säilynyt ylemmillä vuosiluokilla ennallaan. Taivutusvirheet eivät toisin sanoen katoa ylemmillä luokilla, mutta yhä useammin ne ilmenevät aiempaa kompleksisempien lauserakenteiden yhteydessä, mikä vastaa myös Lesosen (2020) tuloksia.

16) *Oli kaunis ilma harjoitella golfia yksin kauniissa kentässä.*

(5. lk. TS1)

17) *Ja niinpä IsoBritannia siihen aion lähteä* (6. lk. PS2, vietnam, SS)

18) *kun Emilianäi jäätelöäsil tuli pääsä unelmin.* (2. lk. TS2, arabia, 5 vuotta Suomessa)

#### SANAJÄRJESTYS JA SANASTON TARKKUUS

Sanajärjestykseen liittyvät epätarkkuudet lisääntyvät ylemmillä luokilla. Se on hyvä esimerkki uuden asian haltuun ottamisen vaiheesta, jossa tarkkuuden kehittyminen seuraa kompleksistumiskehitystä. Sanajärjestyksen epätarkkuudet yhdistyvät lähes poikkeuksetta kasvavaan virke-

pituuteen useampilauseisissa virkkeissä, joissa lisäksi on saatettu käyttää kompleksisia rakenteita kuten lauseenvastikkeita ja verbiketjuja (ks. esimerkit 19 ja 20).

19) *Noustuani päätän mennä rinneravintolaan syömään, joka on Ursuksen yläpuolella.* (6.lk. S1P)

20) *Päätimme vielä yrittää liikunnanopettajalle tehdä metkun ja se onnistui.* (6.lk. S1T [A])

Ongelmavirkeissä on usein käytetty haastavaa kieltomuotoa ja sen yhteydessä indefiniittipronominia sekä erityistapauksissa vielä objektia. S2-kirjoittajille pulmia ovat verrokkiryhmää enemmän tuottaneet myös lausetta täydentävien ajanilmausten sijainti ja sisäinen rakenne. Koko aineiston tasolla sanajärjestys hallitaan kuitenkin molemmissa kieliryhmissä erittäin hyvin (vrt. muun muassa Elo 1998, 36).

Leksikaalisiin valintoihin liittyviä virheitä on totuttu pitämään oppijankielessä yleisinä ja ongelmallisina virheinä, koska niiden on ajateltu johtavan helposti väärin tulkintoihin (Read 2000, 200–201). Etenkin leksikaaliset virheet yhdessä rajoittuneen kokonaissanaston kanssa voivat myös ohjata lukijan tulkintaa voimakkaan negatiiviseksi tekstin mahdollisista muista ansioista huolimatta (Meara, Rodgers & Jakobs 2000, 346). Agustín Llachin (2011, 165, 172) tutkimuksessa kouluikäisten englanninoppijoiden kirjoitelmien laatu (holistinen arvio) korreloi voimakkaasti juuri leksikaalisten virheiden esiintymisen kanssa.<sup>100</sup> Alakoululaisten kirjoitelmissa leksikaaliset virheet eivät lähellekään aina silti aiheuta väärinymmärtämisen vaaraa, ja vain harvoin valittu lekseemi on tarkastelluissa kertomuksissa täysin sopimaton. Selkeästi virheellisiä sanavalintoja kertomusteksteissä esiintyy hyvin vähän, sekä alemmilla että ylempillä vuosiluokilla alle sata tapausta. Syntyperäisillä kirjoittajilla virheellisiksi merkittyjä sanavalintoja on molemmissa keruuvaiheissa saman verran kuin S2-kirjoittajilla. Tämä havainto poikkeaa osin toisen kielen oppimisen aiemmista tutkimuksista, joissa leksikaalisten vir-

<sup>100</sup> Yhteys ei välttämättä säily samankaltaisena oppimisen eri vaiheissa. Agustín Llachin (2011, 178) tutkimuksessa korrelaatio kääntyy osittain toisensuuntaiseksi ylempillä vuosiluokilla, mitä voi selittää leksikaalisen riskinoton positiivinen vaikutus tekstin sisällölliseen antiin.

heiden on havaittu olevan virheistä tyypillisimpiä (ks. Gass & Selinker 1994, 270–272, tuoreemmin esimerkiksi Pietilä 2012). Eroon vaikuttavat suomen kielen ominaispiirteet (taivutusjärjestelmän vaativuus), mutta myös poikkeavuudet tutkimustehtävissä ja tutkittujen oppijoiden iässä ja kielitaitotasossa sekä mahdollisesti itse virheiden luokittelutapa (Neff ja muut 2007, 208).

Kertomustekstit osoittavat, että vaikka oppijankieli ei leksikaalisten merkitysten suhteen aina ole kovin täsmällistä, sanavalintojen tarkkuus on pitkään Suomessa asuneiden S2-kirjoittajien teksteissä kuitenkin kehittyneempi kuin oikeinkirjoituksen tarkkuus. Tekstiyhteydessään suoranaisesti epäsopivia sanoja käytetään harvoin.

Rauniosanalla tarkoitetaan sellaista sananmuotoa, jonka oletettu tavoitemuoto on kaukana toteutuneesta sananmuodosta eli jonka kirjoitusasu poikkeaa tavoitemuodosta esimerkiksi tietyn grafeemimäärän verran (Matilainen 1985) tai jonka virhetyyppiä on virheiden lukuisuuden tai limitteisyyden vuoksi mahdotonta määrittellä tarkasti (Kulju ja Mäkinen 2009, 315). Omassa tutkimuksessani olen käyttänyt rauniosanan määrittelyyn kolmen grafeemin sääntöä: mikäli sanavartalossa on vähintään kolmen grafeemin poikkeama tavoitemuotoon tai virheet sanamuodon kirjoitusasussa ovat sellaisia, että lekseemi muutoin vaatii ympäröivän tekstin avulla päättelyä, se on määritetty rauniosanaksi. Vaikka käytetty rauniosanamääritelmä on tiukka, rauniosanat osoittautuvat kaikkien kirjoittajaryhmien teksteissä hyvin harvinaisiksi. Alemmilla vuosiluokilla rauniosanojen esiintyvyys on tekstimäärään suhteutettuna noin kolme kertaa suurempi kuin ylemmillä. Alimmilla luokilla rauniosanoja on enemmän poikien kuin tyttöjen teksteissä, mutta sukupuolen välinen ero tasoittuu alakoulun aikana: S1-kirjoittajilla rauniosanoja ei enää viidennellä ja kuudennella luokalla esiinny ja S2-ryhmässäkin pojilla huomattavasti aiempaa vähemmän.

Puuttuviksi sanoiksi on katsottu tapaukset, joissa lauserakenteen ongelmat voitaisiin korjata lisäämällä tekstiin puuttuva sana, yleensä subjekti, predikaatti tai objekti. S1-kirjoittajilla tällaiset tapaukset ovat alimmilla vuosiluokilla huomattavasti harvinaisempia kuin S2-kirjoittajilla, ja lauseesta puuttuu pääjäsenäni harvoin. Virkkeiden pidentyessä ja virkerakenteiden kompleksistuessa erityisesti verbiketjujen virheet

lisääntyvät (ks. esimerkit 21–23). S<sub>2</sub>-pojilla puuttuvien sanojen teksti-tiheys on alakoulun päättövaiheessa muita suurempi, mihin vaikuttaa sekä virheiden esiintyminen useammalla oppijalla että puuttuvan sanan tyyppivirhe muutamalla kirjoittajalla.

21) *Se on minun lempi pelaaja ja hän Ø minun mielestäni maailman paras.* (5. lk. PS<sub>2</sub>, albania, SS)

22) *Olen unelmoinut Ø delfiinin kouluttajaksi.* (5. lk. TS<sub>1</sub>)

23) *Minun unelmien päivä on, että juuri perjantaina meillä oppilailla on vapaapäivä ja Ø nukkua paljon* (6. lk. PS<sub>2</sub>, albania, SS)

## Tarkkuuden yhteys kirjoittamisen muihin osataitoihin ja kirjoittajan kokonaiskielitaitoon

Tarkastelen lopuksi kirjoittamisen tarkkuuden yhteyttä sanastohallintaan, suomen kielen taitoon sekä oman äidinkielen taitoon. S<sub>1</sub>-oppijat ovat mukana vertailussa soveltuvin osin.

### YHTEYS KIRJOITTAMISEN OSATAITOIHIN

Tarkkuuden vertailumuuttujina toimii kolme kertomustekstien kielellistä kehittyneisyyttä kuvaavaa muuttujaa: a) sanaston monimuotoisuus<sup>101</sup>, b) harvinaisten sanojen osuus<sup>102</sup> sekä c) tavoitemuodon mukaan korjattu sanepituus, jotka kaikki tiedetään kielellisen kehityksen mukaan karttuviksi ominaisuuksiksi (ks. tarkemmin esimerkiksi Malvern & Richards 2004; Crossley ja muut 2010). Tiiveimmin näistä tekstin ominaisuuksista kirjoittamisen tarkkuus on tutkitussa koululaisaineistossa yhteydessä sanaston monimuotoisuuteen: Mitä monipuolisempaa sanastoa kirjoitelmassa on käytetty, sitä pienempi virhetiheys siinä on todennäköisesti ollut. Toisen vuosiluokan jälkeen yhteys on kaikilla vuosiluokilla tilastollisesti merkitsevä tai erittäin merkitsevä ( $p < 0,01$  tai  $p < 0,001$ ). Yhteys selittyy etenkin S<sub>1</sub>- ja S<sub>2</sub>-kirjoitelmien voimakkailta keskimää-

101 D-indeksiin perustuva *sums of probabilities* -diversiteettiarvo.

102 Yleisyystasoa 1000 harvinaisemmat sanat, vertailukorpuksena CSC:n lekseemilista.

räisillä eroilla molempien piirteiden suhteen: sanaston monimuotoisuus on S1-kirjoitelmissa selvästi korkeampi ja virhetiheys pienempi (mediaanien erot erittäin merkitsevät). Alaryhmätasolla (PS1, TS1, PS2, TS2) yhteys on kuitenkin kohtalaisen voimakas ainoastaan viidennen vuosiluokan S1-tyttöjen ja kuudennen vuosiluokan S2-tyttöjen ryhmissä ( $r_s \leq -0,350$ ,  $p < 0,05$ ,  $n_{TS1} = 46$ ,  $n_{TS2} = 27$ ). Tulosta voi tulkita niin, että sekä kirjoittamisen tarkentuminen että sanaston käytön monimuotoistuminen ovat kehityksellisiä piirteitä, jotka tuovat osaamistason erot esille sitä selvemmin, mitä suuremmat ryhmän sisäiset taitoerot ovat ja mitä paremmin kirjoittajat ovat kirjoittaneet omalla taitotasollaan. Esimerkiksi heikko kirjoitusmotivaatio tai tarkkaavuus tilanteessa kuitenkin estää kirjoittajaa osoittamasta parhaita taitojaan; tosin motivaatiotekijät näyttävät vaikuttavan voimakkaammin tekstin tarkkuusominaisuuksiin kuin sanaston monimuotoisuuteen.

Saneen tavoitepituuden ja virhetiheyden välillä ei ole alimmilla vuosiluokilla lineaarista yhteyttä missään yksittäisessä kirjoittajaryhmässä, mutta koko kertomuskorpuksessa tarkkuus korreloi löyhästi sekä sanapituuden että runsaampien taivutuselementtien käytön kanssa toisella ja viidennellä vuosiluokalla ( $r_s \approx 0,4$ ,  $p < 0,05$ ). Kompleksisuus ja tarkkuus näyttävät silti kehittyvän eri tahdissa: toisella vuosiluokalla keskimääräisen sanepituuden ja virhetiheyden välinen yhteys on positiivinen, kolmannella vuosiluokalla enää hyvin heikosti positiivinen ja kääntyy viidennellä negatiiviseksi. Havaintoa on tulkittava siten, että koulunkäynnin alkuvaiheessa pisimpiä sananmuotoja tavoittelevissa kirjoitelmissa esiintyy keskimäärin myös eniten virheitä, ja vasta ylemmillä vuosiluokilla mekaanisten kirjoitustaitojen alkaessa vakiintua sanepituus lakkaa laajamittaisesti kasvattamasta virheiden todennäköisyyttä. Erityisesti koulun aloitusvaiheessa suomen kielen pitkät sanat kuormittavat sana- ja tavurakenteiden prosessointia, ja samoilla lapsilla on usein vaikeuksia sekä näiden tarkassa kirjoittamisessa että sujuvassa lukemisessa (Mäkinen 2007). Tulos on linjassa myös Malvernin kumppaneineen esittämän havainnon kanssa: oppimisen alkuvaiheessa oikeakielisyysvirheiden määrä on pienin teksteissä, joissa keskimääräinen sanapituus on pienin (Malvern ja muut 2004). Tämän tutkimuksen aineistossa korkea sanepituus alkaa (vasta) viidennellä vuosiluokalla yhdistyä alhaiseen

virheteriheyteen, ja silloinkin alaryhmistä ainoastaan S1-oppijoiden ryhmässä ( $r_s = -0,322$ ,  $p < 0,01$ ,  $n = 77$ ).

Kiintoisaksi osoittautuu myös harvinaisten sanojen tekstitiheyden ja kirjoittamisen tarkkuuden välinen yhteys. Koko aineiston tasolla mittarit asettavat tekstit toisella vuosiluokalla samansuuntaiseen paremmuusjärjestykseen: mitä enemmän harvinaisia sanoja tekstissä on käytetty, sitä suurempi tekstin virheteriheys todennäköisesti on ollut. Aikaisemmin samansuuntaisia havaintoja suomen sanojen yleisyyden ja niiden kirjoittamisen tarkkuuden suhteesta on tehnyt Pekonen (2007). Korrelaatio on kuitenkin hyvin matala eli harvinaisten sanojen käytön selitysvoima virheiden määrään on alhainen, vaikka suurehkoissa ryhmässä tulos on tilastollisesti merkitsevä ( $r_s = 0,212$ ,  $n = 198$ ). Yhteys muuttujien väliltä katoaa jo kolmannella vuosiluokalla ja korrelaatio pysyy lähellä nollaa alakoulun loppuun saakka. Ainoastaan S1-poikien ryhmässä harvinaisten sanojen runsas käyttö lisää virheiden todennäköisyyttä vielä ylemmillä vuosiluokilla.<sup>103</sup>

#### YHTEYS KIRJOITTAJAN SUOMEN KIELEN TAITOON

On mahdollista, että virheettömästi kirjoittavat koululaiset ovat kirjoittajina pedantteja mutta eivät erityisen taitavia kokonaiskielitaidoltaan. Aiempien tutkimusten perusteella kielitaidon tekijät kuitenkin kumuloi-tuvat niin, että kielitaidoltaan vahvat oppijat ovat tyypillisesti vahvoja hyvin monilla eri osa-alueilla (esimerkiksi Alderson 2005). Kielellisesti lahjakkaat eivät tosin aina motivoitu irrallisista kirjoittamistehtävistä, mikä aiemman tutkimuksen valossa pätee erityisesti poikiin. Tutkin omassa aineistossani alakoululaisten kielitaitoprofilin tasaisuutta asian valaisemiseksi kahden erityyppisen vertailumuuttujan avulla. Vertasin kirjoittamisen tarkkuutta ensinnäkin erillisellä testillä mitattuun suomen kielen sanaston hallintaan sekä toisekseen kyselytiedon<sup>104</sup> avulla arvioituun suomen kielen yleistason: lukemis-, puhumis- ja kirjoitus-taitoon, S2-arvosanaan sekä erityisvaikeuksiin.

103 Epälineaarinen yhteys muissa, suuremmissa kirjoittajaryhmissä on sinänsä looginen, sillä harvinaisten sanojen käyttöön ja niiden käytön tarkkuuteen vaikuttaa lukuisia ristikkäisiä tekijöitä.

104 S2-opettajan/luokanopettajan arvio S2-oppilaiden kielitaidosta.

Kirjoittajakohtainen kertomustekstin virheterveys korreloi jokaisella vuosiluokalla kohtalaisen voimakkaasti kirjoittajan leksikaalisten taitojen kanssa, mikäli tarkastellaan koko kirjoittajaryhmää ( $r_s \approx -0,5$ ,  $p < 0,001$ ; ks. taulukko 48). Yhteys on kaikilla vuosiluokilla negatiivinen eli tarkimmin kirjoittavat sanastollisilta taidoiltaan vahvimmat lapset.<sup>105</sup> Tulos on linjassa toisen kielen oppimisesta aiemmin havaitun kanssa (ks. esimerkiksi Llach 2011, 190, 192). Korrelaatio on kuitenkin kaukana maksimiarvostaan, mihin vaikuttaa riippuvuuden ehdollisuus. Vaikka vahva leksikon hallinta – sekä mahdollisesti tarkkaavuus ja tehtäviin suuntautuminen testitilanteessa – yhdistyy virheettömään tai melko virheettömään kirjoittamiseen, kaikki tarkat kirjoittajat eivät ole sanastollisilta taidoiltaan vahvoja.

Pienemmissä tyttöjen ja poikien alaryhmissä yhteys on aina samansuuntainen. Sanastotehtävistä parhaiten suoriutuneet lapset ovat todennäköisimmin myös tarkkoja kirjoittajia. Koska sanastotehtävät ovat edellyttäneet hyvin niukasti kirjoittamista, kirjoitustaito itsessään ei riitä yhteyden selittäjäksi, vaan syytä on haettava kielenhallinnan eroista. Erot tehtävöorientoituneisuudessa saattavat tosin osittain selittää suoriutumista molemmissa tehtävissä ja samalla myös alaryhmien välisiä korrelaatioeroja: etenkin sanastollisesti taitavat S1-pojat eivät näytä motivoituneen ylemmillä vuosiluokilla kirjoittamaan parhaan osaamisensa mukaisesti, ja juuri tässä ryhmässä taitojen välinen yhteyskin on heikoin.

Taulukko 48. Virheterveyden ja sanastotehtäväsuoriutumisen välinen yhteys ( $r_s$  = Spearmanin korrelaatio).

$r_s$	Toinen	Kolmas	Viides	Kuudes
Kaikki	-0,450***	-0,450***	-0,568**	-0,565***
S1	-0,400***	-0,197*	-0,286*	-0,462**
S2	-0,486***	-0,417**	-0,430***	-0,582***
Pojat (S1)	-0,261*	-0,144	-0,155	-0,165
Tytöt (S1)	-0,454***	-0,193	-0,418**	-0,686***
Pojat (S2)	-0,579**	-0,548	-0,269	-0,719***
Tytöt (S2)	-0,280	-0,181**	-0,471**	-0,297

105 On syytä huomauttaa, että sanastollisten taitojen arvioinnissa fonologisella tai morfologisella tarkkuudella ei ole ollut merkitystä.

Sekä S<sub>1</sub>- että S<sub>2</sub>-kirjoittajien opettajille lähetetyssä taustatietokyselyssä on kartoitettu oppijoiden kielellisiä erityisvaikeuksia. Näistä kyselymuuttujista muodostetun summamuuttujan oppijakohtaisen arvon vertaaminen saman oppijan tekstin virhetiheyyteen antaa lisätietoa eri oppijaryhmien kielellisten erityisvaikeuksien ja kirjoittamisen tarkkuuden välisestä suhteesta. Niissä kirjoittajaryhmissä, joissa sanastollisten taitojen ja kielen virheettömyyden yhteys osoittautui selvimmäksi, kielellisten erityisvaikeuksien ja tarkkuuden välillä ei ilmene yhteyttä. Tulosta selittää etenkin arvioitujen erityisvaikeuksien vähäinen määrä aineistossa.

S<sub>2</sub>-oppijoiden ryhmässä on ollut mahdollista tutkia myös kirjoittamisen tarkkuuden ja suomen kielen edellisen todistusarvosanan välistä yhteyttä. Niin S<sub>2</sub>-tytöillä kuin pojillakin edellinen suomen kielen todistusarvosana on yhteydessä kirjoittamisen tarkkuuteen (S<sub>2</sub>:  $r_s = -0,37$ ,  $p < 0,001$ ,  $n = 91$ ), mutta arvosanaan luonnollisesti vaikuttaa moni muukin asia, ja yhteys on heikompi kuin tarkkuuden tilastollisesti erittäin merkitsevä yhteys sanastollisiin taitoihin. Erityisen mielenkiintoinen on opettajan arvioiman kirjoitustaidon yhteys kirjoittamisen tarkkuuteen: tytöillä lineaarinen yhteys on kohtalaisen voimakas ja tilastollisesti merkitsevä ( $r_s = -0,438$ ,  $p < 0,001$ ,  $n = 63$ ) mutta poikien aineistossa sisäinen hajonta on suurempi eikä riippuvuutta ilmene ( $r_s = -0,136$   $p > 0,05$ ,  $n = 42$ ). Tulos tukenee havaintoa siitä, että pojat eivät ole osanneet tai halunneet osoittaa parasta osaamistaan unelmakirjoitelmaa kirjoittaessaan tai ainakaan tehneet sitä yhtä systemaattisesti kuin tytöt. Kirjoittamistaidon kartoituksissa olisikin jatkossa tärkeää kiinnittää erityistä huomiota kirjoittamistehtävään ja sen merkityksellisyyteen erilaisten kirjoittajien näkökulmasta. Toisaalta voi olla, että koska tyttöjen kirjoittaminen on keskimäärin poikien kirjoittamista kehittyneempää, juuri tyttöjen kirjoitustaidon arvioinnissa opettajat kiinnittävät enemmän huomiota myös tarkkuuspiirteisiin.

Opettajien arvioimista osataidoista suomen kielen puhetaito tai eurooppalaista viitekehystä mukaileva yleisempi taitotasoarvio eivät tarkastellussa aineistossa korreloi kirjoittamisen tarkkuuden kanssa. Pääsyynä ensimmäiseen lienee vertailuaineiston luonne: monet kirjoittamisen virheistä liittyvät nimenomaan kirjoitetun kielen normien hallintaan.



Lähes kaikki pikkulapsesta asti Suomessa asuneet koululaiset ovat lisäksi jo hyvin sujuvia puhujia, minkä vuoksi puhetaidon karkea arvio ei myöskään erottele oppilaiden kielenhallintaa S2-ryhmän sisällä (ks. mittarista tarkemmin Honko 2013). Eurooppalaista viitekehystä mukaillevan taitotasoarvion ja tarkkuuden heikkoon yhteyteen voidaan nähdä kaksi hyvin erityyppistä syytä, joista ensimmäinen liittyy toisen kielen oppijan kielenoppimisprosessin vaiheisiin: on mahdollista, että virheet lisääntyvät aaltomaisesti eri kielitaitotasolla, jolloin lineaarista tai edes monotonista yhteyttä ei korrelaatioanalyysillä voida havaita (ks. myös esimerkiksi Brunni ja muut 2019). Toinen mahdollinen selitys liittyy itse mittarin validiteettiin: alakouluikäisen suomenoppijan kielitaidon arvioinnissa viitekehysten mukaiset taitotasot eivät ole olleet kovin helppoja soveltaa.

Kielitaidoltaan vahva ei välttämättä hallitse kirjoitetun kielen normeja tai muusta syystä kirjoita tarkasti. Esimerkiksi kuudesluokkalaisten Suomessa syntyneen ja ensikieleltään vietnaminkielisen pojan kertomuksessa virheettömät fraasit ja idiomaattiset rakenteet (*pienestä asti, jonakin päivänä, mitä vielä, ne vasta ovatkin herkkuja*) sekä sofistikoituneet sanat (erityisesti harvinaisten sanojen käyttö) ja rakenteet (kuten *elämä ilman autoja*), konditionaalin ja perfektin käyttö, erilaiset sivulausetyypit ja lukuisat onnistuneet objektitavutukset vartalomuunnoksineen osoittavat oppijan paitsi käyttävän suomea monipuolisesti myös taitavan kompleksisia rakenteita. Toisaalta kirjoittamisessa on vielä puutteita muun muassa välimerkityksessä, alkukirjaimissa sekä yhteen ja erilleen kirjoittamisessa. Vaikka kielen rekisteritkin välillä sekoittuvat (*naarmuks*), teksti osoittaa, että kirjoitetun kielen konventioiden oppiminen on käynnissä. Tämä näkyy esimerkiksi verbien infiniittimuotojen standardinmukaisena viimeistelynä (*haaveillut, lähteä* vrt. *haaveillu, lähtee*) sekä joidenkin lähinnä kirjoitetussa kielessä esiintyvien konstruktioiden käyttönä. Juuri näiden konstruktioiden hiomattomuus (*eihän voi olla paha mennä, mitä mahtavempia*) toisaalta kuvastaa oppimisprosessin dynaamisuutta. Kun oppija käyttää ilmaisuväriä monipuolisesti, rohkeasti ja luovasti ja samalla vasta harjoittelee kirjoitetun kielen normeja, on luonnollista, että ilmaisuun sisältyy myös epäidiomaattisuutta ja epätarkkuutta.

## YHTEYS KIRJOITTAJAN OMAN ÄIDINKIELEN TAITOON

Koulunkäynnin alussa lasten luku- ja kirjoitustaidossa on suuria eroja, mutta alkuopetus tasaa eroja nopeasti, kuten tämänkin tutkimuksen aineisto osoittaa. Vaikka suomenkielisessä ympäristössä vietetty aika tukee kielitaidon kehittymistä erityisesti kouluikässä, myös koko peruskoulunsa Suomessa opiskelleilla S2-oppilailla voi olla puutteita suomen kielen taidoissa (myös Kuukka & Metsämuuronen 2016, 85). Kirjoitustaidon osalta yksilölliset erot ovat kuitenkin suuremmat oman äidinkielen taidossa. Koulun alkuvaiheessa vain harvat toisen kielen oppijat ovat ehtineet oppia kirjoittamaan ensikielellään. Noin 17 prosenttia monikielisistä oppilaista arvioi oman äidinkielen kirjoitustaitonsa vielä peruskoulun päättövaiheessa joko huonoksi tai olemattomaksi (Kuukka & Metsämuuronen 2016, 73). On huomattava, että myös yksilöllisissä mahdollisuuksissa toiminnallisen kielitaidon kehittämiseen on suuria eroja.

Tähän tutkimukseen osallistuneista vain noin joka kymmenes koululainen piti ensikielen kirjoitustaitoaan toisen tai kolmannen vuosiluokan kevätlukukaudella suomen kielen kirjoitustaitoaan parempana, ja kolme vuotta myöhemmin näin ajattelevien osuus oli edelleen hieman pienentynyt (Honko 2013). Siksi on ollut erityisen kiinnostavaa tarkastella sitä, miten varhaislapsuudesta asti Suomessa asuneiden S2-oppijoiden suomen kielen kirjoitustaito kehittyi suhteessa suomea ainoana kotikielenä puhuneisiin ikätovereihin ja toisaalta suhteessa S2-oppijoiden oman äidinkielen taitoon.

Noin puolet tutkimuksessa mukana olleista S2-oppilaista osallistui aineiston keruuhetkellä oman äidinkielen opetukseen. Näiden oppilaiden oman äidinkielen opettajista 21 vastasi kyselyyn, joka oli samansisältöinen kuin suomen kielen opettajille lähetetty kysely. Vähäiseen vastausmäärään vaikuttaa se, että kaikilla oppilailla ei ole mahdollisuutta tai halua osallistua järjestettyyn oman äidinkielen opetukseen sekä se, ettei opetusta etenkin vähän puhutuissa kielissä ja pienemmällä paikkakunnalla ollut tarjolla kaikille. Kysely ei myöskään tavoittanut kaikkia opettajia, eivätkä kaikki tavoitetut opettajat vastanneet kyselyyn. Vastausmäärä on tilastollisten analyysien kannalta hyvin pieni. Nostan tässä yhteydessä kuitenkin esille joitakin kiinnostavia havaintoja, jotka toimivat avauksi-

na keskusteluun ja mahdollisesti tukena laajempiin pitkittäistutkimuksiin tulevaisuudessa.

Opettajien arvioima oman äidinkielen taito viidennen ja kuudennen luokan keväällä ei ainakaan tässä pienessä tutkimusryhmässä vaikuttanut millään käytetyistä mittareista (tarkemmin Honko 2013) olevan yhteydessä saman ikävaiheen suomen kielen kirjoittamisen tarkkuuteen. Hieman yllättäen oman äidinkielen taidon ja aikaisemman eli toisen ja kolmannen luokan kirjoittamisen tarkkuuden väliltä paljastui kuitenkin melko voimakas ja tilastollisesti merkitsevä korrelatiivinen yhteys. Tämä yhteys ilmeni erikseen sekä opettajan arvioiman oman äidinkielen puhetaidon ( $r_s = -0,51$ ,  $p < 0,05$ ,  $n = 21$ ), kirjoitustaidon ( $r_s = -0,63$ ,  $p < 0,01$ ,  $n = 21$ ) että sanastollisten taitojen ( $r_s = -0,62$ ,  $p < 0,01$ ,  $n = 21$ ), edellisen oman äidinkielen todistusarvosanan ( $r_s = -0,56$ ,  $p < 0,01$ ,  $n = 20$ ) ja suomen kielen kirjoittamisen tarkkuuden välillä. Yhteys oli aina samansuuntainen: varhaisten kouluvuosien suomen kielen kirjoittamisen tarkkuus yhdistyi myöhemmin vahvaksi arvioituun oman äidinkielen taitoon.

Positiivinen yhteys saattaa tukea ajatusta toiminnallisen monikielisyiden eduista eli siitä, että yksilön vahva kielitaito yhdessä kielessä voi tukea muidenkin kielten taitojen kehittymistä. Kyse voi olla myös yleisemmin kielellisestä älykkyydestä tai taipumuksista, jotka syntyvät paljolti ympäristötekijöiden vaikutuksesta mutta osin myös geneettisesti (ks. esimerkiksi Kaprio 2006). Yhteyden katoaminen myöhemmässä vaiheessa voi merkitä ainakin kahta asiaa. Ensinnäkin kirjoittamisen tarkentuessa yläluokilla tarkkuusmittarin erottelevuus luonnostaan vähenee. Toiseksi varhaisen vaiheen vahva suomen kielen taito sekä ehkä myös käytettyjä arviointitapoja suosivat yksilölliset ominaisuudet, kuten tarkkaavuus, ovat voineet olla eduksi oman äidinkielen taidon kehittymiselle. Lisäksi on mahdollista, että opettajan kyselyyn vastaamisen yhteydessä soveltama arviointitapa on suosinut tarkkuutta tai sellaista tehtäväorientoituneisuutta, jota tämän tutkimuksen kieliaineiston keruussakin on vaadittu ja joka on erotellut oppilaita erityisesti alakoulun alkuvuosina.

## Lopuksi

Kirjoittaminen tarkentuu alakoulun aikana selvästi kaikissa kirjoittajaryhmissä. Voimakkainta kehitys on S2-pojilla, joilla kirjoittamisen epätarkkuus alakoulun alkuvaiheessa oli suurinta. Yleisimpiä ovat sekä alakoulun alku- että loppuvaiheessa sanojen yhteen ja erilleen kirjoittamisen haasteet. Etenkin alimmilla vuosiluokilla yleisiä ovat myös puuttuvat geminaatat ja pitkät vokaalit sekä kirjainten piirtämiseen liittyvät epätarkkuudet. Alakoulun loppuvaiheessa sanojen taivuttamiseen liittyvät virheet ovat jo hieman näitä yleisempiä. Puhekieli lomittuu kirjoitettuun yleiskieleen kaikissa kirjoittajaryhmissä – mutta osin hyvinkin hallitusti ja kerrontaa tukien.

Koko kertomusaineistossa virhetiheydellä ilmenee toisen vuosiluokan jälkeen negatiivinen yhteys kielenhallintaan, jonka mittarina tässä tutkimuksessa toimi sanaston käytön monimuotoisuus: virhetiheys on suhteessa suurin niillä kirjoittajilla, jotka käyttävät yksinkertaisinta sanastoa. Tätä selittää ennen kaikkea kieliryhmien välinen kielitaitoero: sanasto monimuotoistuu ja kirjoittaminen tarkentuu alakoulun aikana mutta S2-ryhmässä keskimäärin S1-ryhmää hitaammin. Tarkkuus ja sanarakenteen kompleksisuus puolestaan näyttävät kehittyvän eri tahdisia; pitkien sananmuotojen tavoittelu lisää virheiden todennäköisyyttä vain hieman ja vain alimmilla vuosiluokilla. Lineaarisen yhteyden puuttumista selittää muun muassa suomen kielen pitkien sanojen systemaattinen rakenne: monimorfeemisuus ei automaattisesti tee sanasta vaikeaa. Myös harvinaisten sanojen käyttö lisää virheiden todennäköisyyttä tutkimuksessa aineistossa vain aivan alakoulun alkuvaiheessa.

Tarkimmin kirjoittavat sanastollisilta taidoiltaan vahvat lapset. Tämän yhteyden voimakkuuteen voi kielenhallinnan lisäksi kuitenkin vaikuttaa myös yleinen tarkkaavuus ja tehtäväorientaatio. Opettajan arvioimien kielellisten erityisvaikeuksien määrään ei ilmene yhteyttä kummassakaan kieliryhmässä tai kummankaan kielen osalta (suomi, monikielisten oppilaiden oma äidinkieli), mitä osaksi selittää arvioitujen erityisvaikeuksien vähäinen määrä. Muista opettajan arviointiin nojautuvista suomen kielen taitomittareista ainoastaan edellinen todistusarvosana ja holistinen arvio kirjoitustaidosta ovat yhteydessä kirjoittamisen tarkkuu-

teen tutkimusaineistossa, näistäkin jälkimmäinen ainoastaan tyttöjen ryhmässä. On mahdollista, että tytöillä kirjoittamisen tarkkuus vastaa paremmin yleistä kirjoitustaidon tasoa ja että poikien ryhmässä tekstitaidoista irrotettavat tekijät, kuten tarkkaavuuden tai motivaation puute kirjoittamistilanteessa, heikentävät tätä yhteyttä. Kenties yllättävin, melko voimakas ja lisätutkimukseenkin rohkaiseva korrelatiivinen yhteys S2-aineistossa ilmeni koulunkäynnin alkuvaiheen suomen kielen kirjoitustaidon (virhetiheys) sekä myöhemmän oman äidinkielen taidon (puhuminen, kirjoittaminen, sanasto) välillä. Oman äidinkielen taitojen arviointi perustui useiden eri opettajien tekemiin ja useiden eri kielten taitoja koskeviin holistisiin arviointeihin.

Yleisen virhetiheyden laskeminen ja kaikentyyppisten oikeakielisyysvirheiden väheneminen tarkoittavat, että alakouluiässä kirjoittamisen tarkkuus kasvaa laaja-alaisesti siitä huolimatta, että kirjoittaminen yhtä aikaa sujuvoituu ja kompleksistuu. Vaikka pinnistely yhdellä kielitaidon osa-alueella saattaa hetkellisesti laskea suorituskyykyä toisella,<sup>106</sup> pidemmällä aikavälillä kehitys on kaikkiaan nousujohteista. Monet virheet selittyvät alaluokilla sillä, että kirjoitetun kielen konventioiden tuntemus on vielä puutteellista ja kirjoittamisen rutiinit puuttuvat. Esimerkiksi geminaattojen kirjoittamisessa äidinkielisilläkin on horjuntaa aluksi mutta myöhemmin vain satunnaisesti, kun taas yhdyssanojen kirjoittamisessa norminmukaisten konventioiden oppiminen ja käyttö on kompleksimpi ilmiö, joka haastaa myös aikuisia kirjoittajia.

Aineisto osoittaa myös selvästi, että poikien kirjoitustaito kehittyy keskimäärin tyttöjä hitaammin kaikilla tarkkuuden osa-alueilla ja että pitkäkään maassaoloaika ei ole aina auttanut kaikkia Suomessa koulupolkunsa aloittaneita S2-oppilaita saavuttamaan monipuolista ja tarkkaa kielellistä ilmaisu- ja esitystaitoa. Virheiden olemus ja esiintymisympäristötkin muuttuvat alakoulun aikana, mutta muutokset tapahtuvat eri kirjoittajilla ja kirjoittajaryhmittäin eri rytmisessä. Tarkkuustarkastelun tulokset vastaavat nykykäsitystä kielenomaksumisesta: eri kielenoppimisen tasoilla on

106 Esimerkiksi kirjoitustehtävän sisältöön suunnattu ennakkoharjoittelu saattaa viedä huomiota kielen tarkkuudesta ja yllättäen jopa kasvattaa virheiden määrää sen sijaan että vähentäisi niitä, ks. myös Skehan 1998, 69, 70–71.

omat haasteensa ja omat tyypilliset virhekeskittymänsä (ks. esimerkiksi Ellis & Barkhuizen 2005, 78). Voimakkainta kirjoittamisen tarkentuminen on juuri pojilla sekä toisen kielen oppijoiden ryhmissä, joissa lähtötaso koulunkäynnin alkuvaiheessa on matalin. Näin on siitä huolimatta, että juuri pojilla vaikuttaa olleen vaikeuksia osoittaa parasta osaamistaan.

Kielenoppijoille on luontevaa käyttää teksteissään ilmauksia, jotka he jo hallitsevat hyvin (ks. esimerkiksi Folse 2004, 130). Vaikuttaa siltä, että tämä pätee kirjoitetussa suomen kielessä erityisesti sanojen merkityksiin, sillä sanojen käyttö selvästi virheellisessä merkityksessä on tutkitussa kirjoitetun kielen aineistossa harvinaista. Kielen yleisimmät sanat kuitenkin kattavat hyvin suuren osan teksteistä, mikä väistämättä heikentää tekstien ilmaisuvoimaa. Tulosta voi pitää positiivisenakin, sillä sanaston käytön virheet ja epätäsmällisyydet on katsottu keskeisiksi, usein vuorovaikutuksen kannalta merkittävimmiksi kielitaidon puutteiksi (Ellis 1994, 65; Agustín Llach 2011, 91).<sup>107</sup> Myös kielenkäyttäjät itse merkitsevät puheessaan juuri sanastollisia ongelmia epäröintien ja puheen uudelleenmuotoilujen kautta (Halonen & Kokkonen 2008, 132) sekä mieltävät leksikaaliset virheet vakavimmiksi (Gass & Selinker 1994, 270).<sup>108</sup>

Oppimisen kannalta pelkästään tutun aineksen käyttäminen ei kuitenkaan ole etu. Virheitä voidaan nimittäin aina tarkastella paitsi osaamisen puutteina myös kielellisten resurssien karttumisesta kertovana ilmiönä (esimerkiksi Seilonen 2013; Mustonen 2015; Lesonen 2020). Esimerkiksi verbien finiittimuotojen virheitä ei synny, ellei oppija ylipäätään käytä infiniittisiä muotoja vaativia rakenteita. Tarkkuusorientoitunut tai ylenpalttisesti korjaavaa palautetta saava oppija voi oppia välttelemään virheitä ja epävarmoiksi kokemiaan ilmaisutapoja. Lopputuloksena voi syntyä virheettömiä tekstejä, mutta samalla tuottamisen sujuvuus ja

107 Huomattakoon, että suomen kielessä myös virheelliset rakenteet toki saattavat haitata viestin välitymistä esimerkiksi englannin kieltä useammin jo runsaamman taiputusmorphologian vuoksi. Lisäksi englanninkielisessä kirjallisuudessa leksikaalisista virheistä puhuttaessa sisällytetään virhetyyppiin usein myös mm. sanan oikeinkirjoituksen virheet ja toisinaan myös taiputuksen pulmat (ks. Agustín Llach 2011).

108 Leksikaalisia virheitä ja niihin suhtautumista on tosin tutkittu lähinnä aikuisten kielenoppijoiden kautta, ja kouluikäisen kielestä tietoa on huomattavasti vähemmän (ks. kuitenkin Agustín Llach 2011).

monimuotoisuus vähenevät eikä oppija pääse hyödyntämään oppimisen koko potentiaaliaan (ks. esimerkiksi Martin ja muut 2010).

Monipuolisen funktionaalisen kielitaidon saavuttamisen tavoitteen sisältyy Suomessa kirjallisten konventioiden ja muodollisen kielen hallinta, sillä sitä edellyttävät niin jatko-opinnot, työelämä kuin toimivan arjen viestintäkin (ks. myös Halonen & Kokkonen 2008, 123–124, 137–139; Kalliokoski 2006; 2008). Esimerkiksi suomi toisena kielenä -ylioppilaskokeesta tiedetään, ettei sisällöllisesti ansiokkaalla ja sekä rakenteellisesti että sanastollisesti monipuolisella kirjoittamisella yletä korkeisiin pistemääriin, mikäli kielivirheitä on tekstissä runsaasti (Elo 1998, 36). Toisaalta norminvastaisten muotojen käyttäminen voi toisinaan olla tietoista ja tarkoituksenmukaista (ks. esimerkiksi Halonen 2009; Lehtonen 2014).

Kielelliselle tarkkuudelle voidaan funktionaalisen kielikäsitteksen mukaisesti asettaa ainakin yksi laaja oppimistavoite: rakenteita ja sanastoa on opittava ymmärtämään ja tuottamaan niin hyvin, että merkityksen kannalta olennaiset erot tulevat selviksi (Martin 2003, 84; Halonen & Kokkonen 2008, 131). Koska kieltä ei kuitenkaan voi opetella tällaisten merkityksiä erottavien esimerkkien tai sääntöjen listoina, vaatimukseen sisältyy loppujen lopuksi valtaisa määrä opittavia asioita: sanoja, sanontoja, rakenteita, prosodiikkaa, sävyjä, tyylejä, käytänteitä, vaihtelua ja vaihtoehtoisuutta – mutta myös sääntöjä, vaihteluttomuutta ja vaihtoehdottomuutta. Vasta, kun yksilö hallitsee kieltä todella hyvin, hänen on mahdollista tietää, miten yksityiskohdat sitoutuvat kokonaisuuksiksi, kuinka paljon kokonaisuuksissa on jouston varaa ja mikä lopulta milloinkin on merkityksen kannalta olennaista. Tutkimuksen keinoin on esimerkiksi leksikaaliseen välttelyyn vaikea saada otetta.<sup>109</sup> Tarkkuus voi olla sidoksissa kielellisten taitojen ja tilannetekijöiden lisäksi myös yksilöllisiin kielenkäyttötapoihin. Erityisesti vieraalla kielellä kirjoitettaessa kohdekielen äänne-, muoto- ja lauserakenteen sekä kirjoitusjärjestelmän etäisyys jo osatuista kielistä mutta myös monet paljon laajemmat tekijät kuten kirjoittamisen konventiot ja vuorovaikutuskulttuuri vaikuttavat

109 Apuna voidaan käyttää esimerkiksi ääneenajattelumenetelmää tai tietoa ilmausten esiintymistodennäköisyyksistä. Rakennetuissa testiympäristöissä suomen kielen oppijoiden turvallisuushakuisuudesta on näyttöä (ks. esimerkiksi Nissilä 2011, 150).

toisaalta siihen, kuinka paljon opittavaa on. Mitä erilaisempaan järjestelmään ja kulttuuriin siirrytään, sitä enemmän kohdekielen mukaisten muotojen omaksuminen vaatii aikaa ja harjoittelua.

Kirjoitustaito ei kehity lineaarisesti tai toteudu samassakaan tekstissä aina samalla tavalla, ja etenkin rajatusta aineistosta tehdyt havainnot antavat kirjoittamisen tarkkuudestakin toisinaan ristiriitaisen kuvan. Tämän tutkimuksen tuloksista ei voi tehdä kauaskantoisia päätelmiä yksilön taidoista – tavoite on ollut ryhmätasolla –, sillä kultakin kirjoittajalta on tutkittu molemmissa vaiheissa vain yksi teksti. Lekin ja muiden (2008) mukaan juuri emotionaalinen teksti, kuten kirjoittaminen unelmien päivästä, kuitenkin vaatii tarkkojen ilmausten miettimistä ja vetää kirjoittajan huomion tyypillisesti tekstin paikallisiin seikkoihin globaalien sijaan. Siksi aineistoa voi pitää tutkimustehtävän kannalta perusteltuna.

Mäkisen (2007) tutkimustulokset rohkaisevat hyödyntämään virheanalyysien tuloksia erityisesti oppilaan puutteellisten kirjoittamisstrategioiden tunnistamiseen. Virheanalyysiin kytketyn sanarakenneanalyysin avulla on Kuljun ja Mäkisen (2008, 524–246) mukaan mahdollista saada tietoa erityisesti siitä, minkälaisia ongelmia heikoilla kirjoittajilla on sanarakenteen hahmottamisessa. Sanatasolla alakoululaisilla on vaikeuksia erityisesti suomen kielen pitkien äännekestojen kirjoittamisessa, mutta on muistettava, että oppimisen kannalta keskittyminen pelkästään näiden kirjoittamisen harjoitteluun ei ole tehokasta. Tarvitaan myös kielellisen tietoisuuden vahvistamista, esimerkiksi sanarakenteen tajun harjoittamista eri tavoin myös puhutun kielen kautta, sekä lukemisen ja kirjoittamisen taitojen tarkastelemista ja tukemista vuorovaikutteisesti.

On mahdollista, että tässä tutkimuksessa havaitun kaltaisia taitoeroja eri kirjoittajaryhmien välillä voitaisiin vähentää tunnistamalla kielellisen prosessoinnin ja opiskelutaitojen puutteita yli kielirajojen ja tukemalla tehokkaita lukemisen, kirjoittamisen ja opiskelun strategioita yhtä aikaa oppilaan kaikissa kielissä. Tiedetään, että sanarakenteen hahmottaminen vaikuttaa niin lukemisen kuin kirjoittamisen oppimiseen (ks. esimerkiksi Kulju & Mäkinen 2009). Tiedossani ei kuitenkaan ole tutkimuksia, joissa olisi tarkasteltu aloittelevien monikielisten lukijoiden ja kirjoittajien taitoja tai kielenhallintaa ja niiden tukemista yhtä aikaa suo-



men kielessä, omassa äidinkielessä ja mahdollisesti jo opiskeltavissa muissa kielissä. Tarvitaan lisätietoa myös siitä, suosiiko kirjoittamisen opetus ja arviointi tietäntyyppisiä opiskelu- tai vuorovaikutustapoja niin, että oppimisprofiilien tasoittamisen sijaan opetus voi jopa ruokkia kirjoittamisen taitoerojen syntymistä. Parhaimmillaan opetuksessa pystytään aidosti tukemaan erilaisia ja erilaisissa oppimisen vaiheissa olevia kirjoittajia, esimerkiksi tunnistamaan taitoerojen lisäksi yksilölle merkityksellisiä kirjoittamisen tapoja ja tavoitteita sekä muokkaamaan opetusta ja arviointia näiden havaintojen pohjalta.

Kirjoitetun kielen normien tuntemus karttuu kouluaikana nopeasti. Toiminnallisen kirjoitustaidon näkökulmasta sääntöpohjainen opetus ei kuitenkaan ole kirjoittamistaidon kehittymisen kannalta tehokkainta. Shoren ja Rapatin (2014) pohjalta voidaan esittää, että kirjoittamisen tarkkuutta lisää ja kirjoitustaitoa vahvistaa pintatason virheiden korjaamista tehokkaammin keskittyminen merkitysten selventämiseen. Tarkkuuden kannalta keskeistä on ymmärtää sääntöjen merkitys ja oppia ohjatusti harjoittelemalla, miten sääntöjä on tarkoitus soveltaa erilaisissa teksteissä ja vuorovaikutustilanteissa. Kirjoittamisen opettamisessa ja ohjauksessa on tärkeää tukea kirjoittamista tasapuolisesti sekä kognitiivisena, kulttuurisena että sosiaalisena pääomana ja toimintana (Luukka 2009). Samalla tulisi huomioida kirjoitustaidon kehityksen jatkumot ja kirjoitustaidon dynaaminen luonne eli esimerkiksi kirjoittamisen kytkökset kirjoittamismotivaatioon ja minäpystyvyyden kokemuksiin (Hankala, Harjunen, Kauppinen, Kulju, Pentikäinen & Routarinne 2015). Tähän ohjaavat myös voimassa olevat Perusopetuksen opetussuunnitelman perusteet (OPH 2014), joissa tuottamistaidot kytetään osaksi monilukutaitoa.

Kirjoitustaidon opettajilta ja arvioijilta vaaditaan herkkyyttä havainnoida kirjoittamisen yksilöllisiä tarpeita ja taitoja eli esimerkiksi sitä, että tietyt kirjoittamisen sisällöt ja taidot eivät ole yhtä merkityksellisiä, ajankohtaisia tai tarpeellisia kaikille kirjoittajille. Kaikissa tilanteissa ei myöskään tarvita samanlaista kirjoittamista. Kommunikatiivisen kirjoitustaidon näkökulmasta kaikki kirjoittamisen tekstilajit ja funktiot eivät esimerkiksi edellytä yhtä suurta tarkkuutta, eikä kaikkia tekstejä tuoteta yksin tai tulkita pelkästään kirjallisen tuotoksen perusteella.

## LÄHTEET

- Agustín Llach, M. P. 2011. *Lexical errors and accuracy in foreign language writing*. Bristol: Multilingual Matters.
- Alderson, C. 2005. *Diagnosing foreign language proficiency. The interface between learning and assessment*. London: Continuum.
- Al-Mahrooqi, R. & Al-Aghbari, K. 2016. Refusal strategies among Omani EFL students. *SAGE Open* October-December 2016, 1–10. <http://journals.sagepub.com/doi/pdf/10.1177/2158244016672907>. Viitattu 23.5.2018.
- Afffman, I. & Nissinen, K. 2015. Lukutaidon kehitys PISA-tutkimuksissa. Julkaisussa: J. Välijärvi & P. Kupari (toim.) *Millä eväillä osaaminen uuteen nousuun? PISA 2012 tutkimustuloksia*. Opetus- ja kulttuuriministeriön julkaisuja 2015: 6, 28–49.
- Aro, T., Eklund, K., Leppänen, P. & Poikkeus, A.-M. 2011. Lukivaikeusriskin arviointi ja lukivaikeuden tunnistaminen suomen kielessä. *Psykologia* 46, 92–98.
- Boland, J. E. & Queen, R. 2015. I think your going to like me: Exploring the role of errors in email messages on assessments of potential housemates. *Linguistics Vanguard* 1: 1, 283–293.
- Boland, J. E. & Queen, R. 2016. If you're house is still available, send me an email: Personality influences reactions to written errors in email messages. *Plos One*, 11: 3, e0149885. <https://doi.org/10.1371/journal.pone.0149885>.
- Brunni, S., Jantunen, J. & Skantsi, V. 2019. Korpusavusteinen virheanalyysi tarkkuuden kehityksestä EVK:n taitotasolla A2-B2. *Puhe ja kieli* 39: 3, 275–304.
- Chipere, N. 2009. Individual differences in processing complex grammatical structures. Julkaisussa: G. Sampson, D. Gil & P. Trudgill (toim.) *Language complexity as an evolving variable*. Oxford: Oxford University Press, 178–191.
- Corder, S. P. 1967. The significance of learner's errors. *International Review of Applied Linguistics* 5, 161–170.
- Crossley, S. A., Salsbury, T., McNamara, D. S. & Jarvis, S. 2010. Predicting lexical proficiency in language learner texts using computational indices. *Language Testing* 28: 4, 561–580.
- Ellis, R. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. & Barkhuizen, G. 2005. *Analysing learner language*. Oxford: Oxford University Press.
- Elo, E. 1998. Kvantiteetti vai kvaliteetti – vai molemmat? Hyvän kirjoittamisen syntaksia. Julkaisussa: N. Rekola & H. Korpela (toim.) *Toisella kielellä – näkökulmia lasten ja nuorten suomi toisena kielenä -opetukseen*. Helsinki: Opetushallitus, 35–40.
- Eloranta, A.-K., Närhi, V. & Ahonen, T. 2009. Oppimisvaikeuksien pysyvyys – seurantatutkimus yhdeksästä lukivaikeustaustaisesta nuoresta naisesta. *NMI-bulletin* 19: 1, 4–19.
- Eronen, R., Maamies, S. & Räikkälä, A. 1996. Yhdyssanat. *Kielikello* 1996: 4.
- Erra, S. & Svinhufvud K. 2017. Kirjoittamisen diskurssit lukion äidinkielen ja kirjallisuuden opetussuunnitelman perusteissa ja äidinkielen ylioppilaskokeen määräyksissä. *Virittäjä* 121: 3, 316–354.
- Folse, K. S. 2004. *Vocabulary myths. Applying second language research to classroom teaching*. Ann Arbor: The University of Michigan Press.

- Gass, S. M. & Selinker L. 1994. *Second language acquisition. An introductory course*. Hillsdale: Lawrence Erlbaum Ass.
- Gass, S. M. & Selinker, L. 2008. *Second language acquisition. An introductory course*. 3., uudistettu painos. Hoboken: Taylor & Francis.
- Halonen, M. 2009. Puhutun kielen variantit resurssina monikielisten koululaisten kirjoittelussa. *Virittäjä* 113: 3, 329–355.
- Halonen, M. & Kokkonen, M. 2008. ”Loppu-ännät puuttuu!” ”Toivotaan sujuvaa puhetta!” – Lapsi- ja aikuisoppijoiden kielenoppimis- ja arviointikontekstien erot. Julkaisussa: J. Tanner & M. Kokkonen (toim.) *Kakkoskieli 6. Suomenopetus, kielitaito ja tutkimus*. Helsinki: Helsingin yliopisto, 109–140.
- Hamp-Lyons, L. 2003. Writing teachers as assessors of writing. Julkaisussa: B. Kroll (toim.) *Exploring the dynamics of second language writing*. Cambridge: Cambridge University Press, 162–189.
- Hankala, M., Harjunen, E., Kauppinen, M., Kulju, P., Pentikäinen, J., & Routarinne, S. 2015. Tutkimuksesta tukea kirjoittamisen opettamiseen. Julkaisussa: E. Harjunen (toim.) *Tekstit puntarissa: Ajatuksia äidinkielen ja kirjallisuuden oppimistuloksista perusopetuksen päättöarvioinnissa 2014 ja 2010*. Helsinki: Kansallinen koulutuksen arviointikeskus, 73–84.
- Havukainen, M. & Salmi, L. 2011. *Oppilaiden tuottavan kirjoittamisen taso toisella luokalla*. Kasvatustieteen pro gradu -tutkielma. Itä-Suomen yliopisto.
- Honko, M. 2013. *Alakouluikäisten leksikaalinen tieto ja taito: toisen sukupolven suomi ja S1-verrokkit*. Tampere: Tampere University Press.
- Honko, M. 2017a. Sadutettu sanasto: puhutun kielen leksikaalinen diversiteetti arviointikohteena. Julkaisussa: M. Kuronen, P. Lintunen & T. Nieminen (toim.) *AFinLA-e*. Soveltavan kielitieteen tutkimuksia 10, 163–192.
- Honko, M. 2017b. Kieli- ja kielitaitokäsitykset tutkivan opettajan kenttäpäiväkirjamerkinnoissä. *Puhe ja kieli* 37: 4, 215–238.
- Housen, A. & Kuiken, F. 2009. Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics* 30: 4, 461–473.
- Housen, A., Kuiken, F. & Vedder I. (toim.) 2012. *Performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins.
- Huhta, A. & Hildén, R. 2016. Kielitutkinnot ja muu laajamittainen kielitaidon arviointi Suomessa. Julkaisussa: A. Huhta & R. Hildén (toim.) *AFinLA-e*. Soveltavan kielitieteen tutkimuksia 9, 3–26.
- Ivanič, R. 2004. Discourses of writing and learning to write. *Language and Education* 18: 3, 220–245.
- Ivaska, I. 2009. *Eksistentiaalilauseen ilmiäsuut edistyneiden suomenoppijoiden kirjoituksessa*. Suomen kielen pro gradu -tutkielma. Turun yliopisto.
- James, C. 1998. *Errors in language learning and use: Exploring error analysis*. Essex: Longman.
- Jessmer, S. L., & Anderson, D. 2001. The effect of politeness and grammar on user perceptions of electronic mail. *North American Journal of Psychology* 3: 2, 331–346.
- Kalliokoski, J. 2006. Tekstilajin taju ja toisella kielellä kirjoittaminen. Julkaisussa: A. Mäntynen, S. Shore & A. Solin (toim.) *Genre – tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura, 240–265.
- Kalliokoski, J. 2008. Suomi toisena kielenä ja koulussa kirjoittaminen. Julkaisussa:

- S. Routarinne & T. Uusi-Hallila (toim.) *Nuoret kielikuvassa. Kouluikäisten kieli 2000-luvulla*. Helsinki: Suomalaisen Kirjallisuuden Seura, 349–370.
- Kaprio, J. 2006. Lahjakkuuden monet ulottuvuudet. *Duodecim* 122: 23, 2871–2877.
- Kauppinen, M., Pentikäinen, J., Hankala, M., Kulju, P., Harjunen, E. & Routarinne, S. 2015. Systemaattinen katsaus perusopetusikäisten kirjoittamisen opetusta ja osaamista koskevaan tutkimukseen. *Kasvatus* 46: 2, 160–175.
- Kreiner, D. S., Schnakenberg, S. D., Green, A. G., Costello, M. J., & McClain, A. F. 2002. Effects of spelling errors on the perception of writers. *Journal of General Psychology* 129: 1, 5–17.
- Kulju, P. & Mäkinen, M. 2009. Sanelukirjoituksen arviointi – tyyppivirheistä sanarakenteiden fonologiseen analyysiin. *Virittäjä* 113: 4, 508–532.
- Kuukka, K. & Metsämuuronen J. 2016. *Perusopetuksen päättövaiheen suomi toisena kielenä (S2) -oppimäärän oppimistulosten arviointi 2015*. Helsinki: Karvi.
- Lado, R. 1957. *Linguistics across cultures*. Ann Arbor: The University of Michigan.
- Lappalainen, H. 2011. *Sen edestään löytää. Äidinkielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2010*. Helsinki: Opetushallitus. [http://koulutustoimikunnat.fi/download/132347\\_Sen\\_edestaan\\_loytaa.pdf](http://koulutustoimikunnat.fi/download/132347_Sen_edestaan_loytaa.pdf). Viitattu 31.5.2020.
- Lehtinen, R. 2013. *Kielen norminmukaisuus alakoululaisten kirjoitelmissä*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Lehtonen, H. 2015. *Tyylitellen: nuorten kielelliset resurssit ja kielen sosiaalinen indeksisyys monitieteisessä Helsingissä*. Helsinki: Helsingin yliopisto.
- Lerkkanen, M.-K., Poikkeus, A.-M., Ahonen, T., Siekkinen, M., Niemi, P. & Nurmi, J.-E. 2010. Luku- ja kirjoitustaidon kehitys sekä motivaatio esi- ja alkuopetusvuosina. *Kasvatus* 41: 2, 116–128.
- Lesonen S. 2020. *Valuing variability: Dynamic usage-based principles in the L2 development of four Finnish language learners*. Jyväskylä: Jyväskylän yliopisto.
- Lyly, T. 2008. Opettajien kokemuksia tekstien arvioinnista peruskoulussa. Julkaisussa: *Kakkoskieli* 6. Suomenopetus, kielitaito ja tutkimus. Helsinki: Helsingin yliopisto, 141–162.
- Malvern D., Richards, B. Chipere, N. & Durán, P. 2004. *Lexical diversity and language development: Quantification and assessment*. Hampshire: Palgrave Macmillan.
- Manninen, T. & Martin, S. 2000. Vieraskielisten ongelmia ylioppilasaineessa. Julkaisussa: S. Martin & H. Sulkala (toim.) *Tutkielmia oppijankielestä*, 73–90.
- Martin, M. 1995. *The map and the rope: Finnish nominal inflection as a learning target*. Jyväskylä: Jyväskylän yliopisto.
- Martin, M. 2003. Kieli on kuin lammikko – johdatusta toisen kielen oppimiseen. Julkaisussa: L. Nissilä, H. Vaarala & M. Martin (toim.) *Suolla suomea. Perustietoa maahanmuuttajien suomen kielen opettajille*. Helsinki: Äidinkielen opettajain liitto, 75–90.
- Martin, M. 2009. Yhteistä ja erityistä – johdantoa teemanumeron kirjoituksiin. *Virittäjä* 113: 3, 321–328.
- Martin, M. 2010. Sujuvuus, tarkkuus ja kompleksisuus kielitaidon kehityksessä. Julkaisussa: S. Stolt, M. Lehtihalmes, S. Tarvainen & K. Launonen (toim.) *Suomalainen monikielisyys ja sen haasteet*. Helsinki: Puheen ja kielen tutkimuksen yhdistys.
- Martin, M. 2015. VIRHE! Professori Maisa Martinin jäähyväisluento. Jyväskylän yliopisto 25.9.2015.

- Martin, M., Mustonen, S., Reiman, N. & Seilonen, M. 2010. *On becoming an independent user*. Julkaisussa: I. Bartning, M. Martin & I. Vedder (toim.) *Communicative proficiency and linguistic development, intersections between SLA and language testing research*. EUROSLA Monograph Series 1. European Second Language Association, 57–80.
- Matilainen, K. 1985. Lukemaanopettamisen menetelmien yhteydet oikeinkirjoitustaidon oppimiseen ensimmäisen lukuvuoden aikana. Kasvatustieteiden tiedekunnan tutkimuksia 5. Joensuu: Joensuun yliopisto.
- Matilainen, K. 1993. *Peruskouluilaisten kirjoittamisen hahmottamisstrategioiden kehittyminen ja kuvaus oppimisvaikeuksisten oppilaiden kirjoittamisen ja luetun ymmärtämisen strategioista*. Joensuu: Joensuun yliopisto.
- Meara, P. & Milton, J. 2003. *X\_Lex, The Swansea Levels Test*. Newbury: Express.
- Meara, P., Rodgers, C. & Jacobs, G. 2000. Vocabulary and neural networks in the computational assessment of texts written by second-language learners. *System* 28: 3, 345–354.
- Milton, J. 2009. *Measuring language vocabulary acquisition*. Bristol: Multilingual Matters.
- Mustonen, S. 2015. *Käytössä kehittyvä kieli. Paikat ja tilat suomi toisena kielenä -oppijoiden teksteissä*. Jyväskylä: Jyväskylän yliopisto.
- Mäki, H. 2002. *Elements of spelling and composition. Studies on predicting and supporting writing skills in primary grades*. Turku: Turun yliopisto.
- Mäkinen, M. 2007. Fonologinen analyysi toisluokkalaisten luku- ja kirjoitusprofileista. *Puhe ja kieli* 27: 4, 149–170.
- Mäkihönkö M. 2006. *Luetun ymmärtämisen ja tuottavan kirjoittamisen kehittyminen alkuopetuksen aikana*. Joensuu: Joensuun yliopisto.
- Neff, J., Ballesteros, F., Dafouz, E., Martinez, F., Rica, J.-P., Diez, M. & Prieto, R. 2007. A contrastive functional analysis of errors in Spanish EFL university writers' argumentative texts: A corpus-based study. Julkaisussa: E. Fitzpatrick (toim.) *Corpus linguistics beyond the word. Corpus research from phrase to discourse*. Amsterdam: Rodopi, 203–225.
- Niiranen, L. 2010. Tapaustutkimus kolmen suomenoppijan kompleksisista verbikonstruktiosta. Julkaisussa: A. Kaivapalu, P. Muikku-Werner & M.-M. Sepper (toim.) *Lähivörtlusi. Lähivertailuja* 20, 155–190.
- Nissilä L. 2011. *Viron kielen vaikutus suomen kielen verbien ja niiden rektioiden oppimiseen*. Oulu: Oulun yliopisto.
- OPH 2014. Perusopetuksen opetussuunnitelman perusteet. Helsinki: Opetushallitus.
- Pajunen, A. 2012. Kirjoittamistaitojen kehitys 8–12-vuotiailla. Alakouluilaisten nelmakirjoittelmat. *Virittäjä* 116: 1, 4–32.
- Pekonen, V. 2007. *Suomen kvantiteetti ja kirjoitusprosessi: Mallinnusta ja empiriaa*. Jyväskylä: Jyväskylän yliopisto.
- Pietilä, P. 2012. Lexical vs. syntactic competence in L2 English. *Journal of Language Teaching and Research* 3: 2, 314–320.
- Piper, T. 2001. *And then there were two: Children and second-language learning*. Toronto: Pippin Publishing Corporation.
- Pääkkönen, I. 1985. Kielivirheiden asema ja merkitys ylioppilasaineiden arvostelussa. Ylioppilaskirjoituksen aineiden arvostelijan puheenvuoro. *Kielikello* 1, 19–21.
- Read, J. 2000. *Assessing vocabulary*. Cambridge: Cambridge University Press.

- Reiman, N. 2011. Transitiivikonstruktio ikkunana syntaksin kehitykseen: Infiniittiset rakenteet ja passiivi taidon indikaattoreina S2-oppijoiden teksteissä. Julkaisussa: E. Lehtinen, M. Aaltonen, M. Koskela, E. Nevasaari & M. Skog-Södersved (toim.) *AFinLA-e Soveltavan kielitieteen tutkimuksia* 3, 142–157.
- Saarela, L. 1997. *Peruskoululaisten kirjoitelmien kehittyminen sanastotutkimuksen valossa*. Oulu: Oulun yliopisto.
- Salonen, I. 2012. *Puhekielen piirteiden esiintyminen alakoululaisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Santala, R., Saaristo-Rinta, S., Metsola, M. & Klippi, A. 2013. Suomenkielisten 13–16-vuotiaiden nuorten kielellisten taitojen arviointi. *Puhe ja kieli* 33: 1, 15–41.
- Seilonen, M. 2013. *Epäsuora henkilöön viittaaminen oppijansuomessa*. Jyväskylä: Jyväskylän yliopisto.
- Shore, S. & Rapatti, K. (toim.) 2014. *Tekstilajitaidot. Lukemisen ja kirjoittamisen opetus koulussa*. Helsinki: Äidinkielen opettajain liitto.
- Sikiö, S., Siekinen, M. & Holopainen, L. 2016. Literacy development among language minority background and dyslexic children in Finnish orthography context. *Reading Psychology* 37: 5, 706–727.
- Skehan, P. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Slabakova, R. 2016. *Second language acquisition*. Oxford: Oxford University Press.
- Spoelman, M. 2013. Prior linguistic knowledge matters: The use of the partitive case in Finnish learner language. Oulu: Oulun yliopisto.
- Stenberg, P.-R. 2010. *Yhdysänatajain kehittyminen ja lingvistinen tiedonrakentelu peruskoulusta yliopistoon*. Suomen kielen pro gradu -tutkielma. Itä-Suomen yliopisto.
- Stothers, M. & Klein, P. 2010. Perceptual organization, phonological awareness, and reading comprehension in adults with and without learning disabilities. *Annals of Dyslexia* 60, 209–237.
- Tarnanen, M. 2002. *Arvioija valokeilassa. Suomi toisena kielenä -kirjoittamisen arviointia*. Jyväskylä: Jyväskylän yliopisto.
- Tanner, J. 2012. *Rakenne, tilanne ja kohteliaisuus. Pyyntö S2-oppikirjoissa ja autenttisissa keskusteluissa*. Helsinki: Helsingin yliopisto.
- Tarnanen, M. 2003. Kielitaidon arviointi: valintoja ja neuvottelua. Julkaisussa: L. Nissilä, H. Vaarala & M. Martin (toim.) *Suolla suomea. Perustietoa maahanmuuttajien suomen kielen opettajille*. Helsinki: Äidinkielen opettajain liitto, 143–160.
- Tervola, M. 2019. *Maahanmuuttajalääkärien suomen kielen taito ja kielitaitotarpeet: Erityisalan kielitaidon näkökulma lääkärin työhön*. Tampere: Tampereen yliopisto.
- Tomasello, M. & Akhtar, N. 2000. Five questions for any theory of word learning. Julkaisussa: R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (toim.) *Becoming a word learner: A debate on lexical acquisition*. Counterpoints: cognition, memory, and language. Oxford: Oxford University Press, 179–186.
- Turunen, U. 2012. *Suomi toisena kielenä -oppijoiden possessiivisuffiksien käyttö*. Suomen kielen pro gradu -tutkielma. Jyväskylän yliopisto.
- Undheim, A. 2009. A thirteen-year follow-up study of young Norwegian adults with dyslexia in childhood: Reading development and educational levels. *Dyslexia* 15: 4, 291–303.

- Unsworth, S. 2008. Comparing child L2 development with adult L2 development: How to measure L2 proficiency. Julkaisussa: B. Haznedar & E. Gavruseva (toim.) *Current trends in child second language acquisition: A generative perspective*. Amsterdam: John Benjamins, 301–333.
- Vignovic, J. & Thompson, L. F. 2010. Computer-mediated cross-cultural collaboration: Attributing communication errors to the person versus the situation. *Journal of Applied Psychology* 95: 2, 265–276.

# Kielen rakenteen hallinta peruskoululaisilla ja nuorilla aikuisilla

*Anneli Pajunen*

© <https://orcid.org/0000-0001-8597-7491>

*Seppo Vainio*

## Johdanto

Kielenkehitys alkaa puhutusta kielestä ja myöhemmän kielenkehityksen ja kirjoittamisen alkuvaiheissa juuri puhuttu kieli on lapsilla mallina. Neljävuotiailla suullinen kerronta on hyppelehtivää, mutta jo viisivuotiaat osaavat jäsentää kerrontaansa episodeihin ja seitsenvuotiaat voivat jo olla taitavia suullisen tarinan kertojia (ks. Nieminen 2007; Suvanto 2012). Lapset kirjoittavat aluksi lyhyin ja yksinkertaisin lausein, jotka muistuttavat puhutun rakenteita ja kerrontaa. Osa 7-vuotiaista osaa myös kirjoittaa kertomusrakenteen mukaisesti, mutta vielä yhdeksänvuotiailla kirjoittamisen mallina voi olla puhutun tapa tuottaa kerrontaa (ks. Strömqvist, Nordqvist & Wengelin 2004). Suurimmalla osalla kirjoittamistaidot kehittyvät hitaasti ja niiden kehittyessä sana-, lause- ja kertomustason rakenteet kompleksistuvat. Tutkimme tässä yhteydessä juuri tätä rakenteellista kompleksistumiskehitystä vertailemalla eri-ikäisten peruskoululaisten kirjoitelmia keskenään ja edelleen nuorten aikuisten kirjoitelmiin. Tavoitteena on selvittää sekä ikäluokkatason että



kehityksellinen keskimääräinen osaaminen, jotta voidaan luoda ikään kuin kehityksellisiä mittareita, hahmottaa tavanomaisen osaamisen vaihteluvälejä. Tutkimusaineisto koostuu pääasiassa kertovaa tekstiä edustavista kirjoitelmista, jotka voivat olla rakenteeltaan paitsi hyvin yksinkertaisia ja yksitasoisia, myös monitasoisia, moniulotteisia ja moniäänisiä. Oletuksen mukaan keskimääräinen kehitys näkyy juuri rakenteen kompleksistumisena (vrt. Jagaiah, Olinghouse & Kearns 2020).<sup>110</sup> Toinen oletus on, että lapset ja nuoret käyttävät kirjoitelmissaan niitä keinoja ja resursseja, jotka he hallitsevat, joten kirjoitelmien analyysin avulla saadaan tietoa kielenkehityksen vaiheesta.

Kouluikäisten myöhemmän kielen kehitystä arvioidaan yleensä sana-, lauseke-, lause- ja tekstitason muuttujien avulla sekä lingvistisessä että psykolingvistisessä tutkimuksessa (ks. esimerkiksi Wolf Nelson & Van Meter 2007; Hall-Mills & Apel 2015; Price & Jackson 2015; Jagaiah ja muut 2020). Sanatason muuttujilla analysoidaan joko sanarakennetta tai sanastoa, lauseke- ja lausetasolla keskitytään sanaa ja lausetta laajempiin yksiköihin, muun muassa niiden syntaktiseen koostumukseen. Tekstitasolla voidaan tutkia esimerkiksi kokonaissanamääriä ja käyttää erilaisia holistisia onnistuneisuusarvioita. Tekstitalle kuuluvat myös tekstin rakenneosien analyysi, esimerkiksi kertomusrakenteen toteutuminen, aloitusten ja lopetusten analyysi, moniäänisyyden ja evaluoinnin analyysi (ks. esimerkiksi Labov & Waletzky 1967; Bamberg 1997; Suvanto 2012). Kehitystä voidaan arvioida myös sisällön ja koherenssin perusteella (ks. esimerkiksi Habermas & de Silveira 2008). Sisällönanalyysi luokittelee temaattista sisältöä esimerkiksi suhteessa kuvattavana oleviin tapahtumiin. Tämä toimii etenkin silloin, kun on jokin oletus näistä tapahtumista ja niiden järjestyksestä kuten esimerkiksi elämäkertakertomuksissa tai kuvaalisitoinneissa (vrt. Tuurin *Trajectoire*-aineisto tässä teoksessa). Temporaalisen ja lokaalisen koherenssin analyysi kirjaa kehystämistä aikaan ja paikkaan, kausaalisen koherenssin analyysi syy-seuraussuhteiden esittämistä. Referentiaalista koherenssia tutkimalla selvitetään, miten tarkoitteet tuodaan tekstiin ja miten niihin viitataan (ks. Givón 2020; tässä teoksessa Kouvo ja muut). Globaalikoherenssia

110 On syytä muistuttaa, että yksilötason kokonaiskehitystä ei voi arvioida yhden näytteen perusteella.

voidaan tutkia monesta näkökulmasta käyttämällä esimerkiksi T-yksikön rakenteen tai tekstijaksojen analyysiä. Tekstitypologiassa voidaan tarkastella esimerkiksi, miten kertova, kuvaileva tai argumentoiva teksti eroavat toisistaan. Eniten tiedetään kertovasta tekstistä, jonka analyysiin on valmiita välineitäkin (ks. Gillam & Pearson 2004; Schneider, Vis Dubé & Hayward 2005, Justice, Bowles, Kaderavek, Ukrainetz, Eisenberg & Gillam 2006; Graesser, McNamara & Kulikowich 2011). Tutkimustapa voi olla synkroninen, jolloin muutosta mitataan ikäluokkakohtaisia analyysijä vertaamalla. Muutosta voidaan tutkia myös pitkittäistutkimuksella, jolloin seurataan samoja koehenkilöitä pidemmän aikaa (ks. esimerkiksi Honko 2013). Voidaan myös tutkia suunnittelu- ja kirjoittamisprosessia (ks. esimerkiksi Strömqvist, Nordqvist & Wengelin 2004; Murtorinne 2005). Peruskoululaisten myöhemmän kielenkehityksen hankkeessa on ollut käytössä monia tutkimustapoja. Tässä yhteydessä keskitymme mikro- ja makrotason rakenneyksiköiden avulla rakenteelliseen kompleksistumiseen.

Suomessa peruskoululaisten kieltä on tutkittu kehityksellisestä näkökulmasta yllättävän vähän. Alakoululaisten kirjoitelmia on tutkittu kehityksellisesti jonkin verran (ks. esimerkiksi Matilainen 1989; Mäkihonko 2006; Pajunen 2012; Honko 2013)<sup>111</sup> mutta vanhemmista koululaisista tehdyt tutkimukset ovat kysymyksen asettelultaan toisenlaisia eikä niissä arvioida rakennetta kehityksellisestä näkökulmasta eikä suhteuteta osaaamista kielenkehitykseen (vrt. Murtorinne 2005; Ranta 2007; Juvonen, Kauppinen, Makkonen-Craig & Lehti-Eklund 2011; Harjunen, Marjanen & Karlsson 2019; toisin Leskinen, Savijärvi & Särkkä 1974 oppikoululaisista). Oppimistutkimusta on tehty sekä ala- että yläkoululaisista, mutta kielenhallinnan tutkimus perustuu näissä holistisiin opettaja-arviointeihin tai oikeinkirjoitukseen, ei kielen yksityiskohtaiseen analyysiin (ks. esimerkiksi Lappalainen 2003; 2008; Huisman 2006). Aikuistasoa saavutettavana tavoitetasona ei suomalaistutkimuksissa ole ollut mukana; aikuisten kielenhallinnan tutkimus on Suomesta ylipäättään puuttunut

111 Kulju, Kauppinen, Hankala, Harjunen, Pentikäinen & Routarinne (2017) toteavat, että kirjoittamista on eniten tutkittu esikoulusta toiseen luokkaan mutta tutkimuksia on hyvin vähän vanhemmista peruskoululaisista. Aikuiskirjoittamisen tutkimus puuttuu myös heidän mukaansa kokonaan.

(ks. kuitenkin Petälä 2018). Aikuisten kielenkäyttöä tai kirjoittamista koskeva kansainvälinenkin tutkimus on vähäistä, verrataan heitä sitten nuorempiin tai tutkitaan omana ryhmänään (ks. esimerkiksi Berman & Verhoeven 2002; Nippold, Hesketh, Duthie & Mansfield 2005; Nippold, Mansfield & Billow 2007; Nippold, Cramond & Hayward-Mayhew 2014; Price & Jackson 2015); aikuistutkimus liittyy lisäksi usein muistien muutosten ja ikääntymisen vaikutukseen (ks. esimerkiksi Bromley 1991; Kemper, Marquis & Thompson 2001). Suomessa on ehkä eniten tutkittu kertomuksia mutta lähinnä poikkeavan kielenkehityksen näkökulmasta (esimerkiksi Suvanto 2012; Mäkinen, Loukusa, Laukkanen, Leinonen, Kunnari 2014; Mäkinen 2014; Heikkola 2017) mutta muiden tekstityyppien tuottamista vähemmän (ks. esimerkiksi Örnmark 2011). Myös useita tekstilajeja vertailevat tutkimukset puuttuvat. Psykometrisiin eli mitattaviin muuttujiin perustuvia tutkimuksia ei ole useita, vaikka tällainen testaus on vakio muualla länsimaissa (vrt. Jagaiah ja muut 2020). Mitattavien muuttujien avulla on helppo seurata paitsi oppimista myös arvojen hajontaa, variaatiota ja mahdollista muutosta. Suomessa puhutaan nykyisin paljon kouluikäisten kielestä ja oppimistulosten muutoksesta, mutta esimerkiksi laaja seurantatutkimus puuttuu kokonaan ja samoin tällaiseen tutkittuun tietoon perustuvat toimenpiteet. Luku- ja kirjoitustaidon kartoitustutkimuksia kyllä on tehty pitkään, mutta ne eivät mittaa kielenhallintaa.

Kansainvälinen lasten ja nuorten kielenhallinnan ja kirjoittamisen tutkimus on laajaa, menetelmiltään vakiintunutta ja perustuu oleellisesti mittaamiseen. Lasten, nuorten ja nuorten aikuisten myöhempi kielenkehitys -hankkeen motivaatio on ollut kattaa tutkimusaukkoa, joka koskee suomenkielisten peruskoululaisten kielenhallinnan koko kehityskaarta. Tässä luvussa tavoitteena on osoittaa muutamia kielenkehityksen kanalta keskeisiä – sanaan, lauseeseen ja tekstiin perustuvia – muuttujia (tarkemmin seuraava jakso), joiden perusteella voi arvioida paitsi keskimääräistä kehitystä myös sukupuolittain eriytyvää ja yksilöllistä ikäkausittaista hallinnan variaatiota. Pääosin kyse on eri tason yksiköiden rakenteen monimuotoutumisesta ja toisaalta osaamisen hajonnan vähenemisestä iän mukaan. Aineistona on lähinnä kertovaa tekstiä edustavia kirjoitelmia, joita koehenkilöistä nuorimmatkin pystyvät ainakin peri-

aatteessa jo tuottamaan (vrt. Burns, de Villiers, Pearson & Champion 2012). Kertomuksien analyysissä voi tukeutua mainittuihin narratiivisiin normeihin. Analyysi perustuu alakoululaisilta kerättyyn Koulukorpukseen ja Unelmakorpukseen, joiden kirjoittajat ovat peruskoululaisia tai kuudes- ja yhdeksäsluokkalaisia sekä nuoria aikuisia (ks. aineiston ja koehenkilöiden esittelyä tämän teoksen johdannosta). Unelmakorpuksen koehenkilöt ovat osallistuneet myös morfologiseen ja semanttiseen testaukseen (ks. Vainio, Pajunen & Häikiö 2019).

Argumentaation kulku on seuraava: Ensimmäiseksi määritellään keskeinen käsitteistö ja esitellään aineisto ja koehenkilöt. Seuraavissa kolmessa luvussa esitellään tulokset, joiden avulla kuvataan paitsi kirjoitelmien kehityksellisiä ja rakenteellisia ominaisuuksia myös pohditaan eroja kuvaavien muuttujien ominaisuuksia. Ensimmäisessä käsitellään sanapituus, sanarakenne ja kokonaissanamäärä, joilla yleensä operationalisoidaan ikäluokkia erotteleva, iän tuoma kypsyys (*maturity*, ks. Wolf Nelson & Van Meter 2007; Olinghouse & Wilson 2013). Seuraavassa luvussa arvioidaan sanaluokkakoodaukseen perustuvia indeksejä, jotka laskevat muun muassa leksikaalisten ja kieliopillisten aineiden osuuksia. Viimeinen analyysiluku keskittyy syntaktiseen kompleksisuuteen. Lopuksi kootaan keskeiset tulokset.

## Muuttujat ja aineisto

Kehityksellinen näkökulma rakenteeseen edellyttää kielen yksiköiden kompleksistumisen tutkimusta. Kompleksisuus on asteittain kehittyvää ja sitä voidaan tutkia sekä horisontaalisella että vertikaalisella rakennetasolla. Edellinen liittyy esimerkiksi lauserakenteen monimuotoutumiseen ja jälkimmäinen tarkastelee muutoksia vaikkapa lausekkeen hierarkiassa. Määrittelemme seuraavaksi, mitä kompleksisuudella tarkoitetaan tässä tutkimuksessa ja millä aineistoin ja metodein tutkimus etenee.

**KEHITYSTÄ MITTAAVAT MUUTTUJAT KOMPLEKSISUUSASTEIKOLLA**  
Keskeinen kielenkehityksen trendi on yksinkertaisesta kompleksisempaan, on sitten kysymys sanastosta tai rakenteesta, puhumisesta tai kir-

joittamisesta. Yksinkertaisuus vaihtelee luonnollisesti arvioitavan kielen yksikön mukaan, ja se on helpoin määrittellä negatiivisesti. Esimerkiksi Givónin (2009, 3–7; 2020, 56–58) mukaan kompleksisuus on organisoidun systeemin ominaisuus, yksinkertaisesta puuttuu järjestyneisyys. Tässä yhteydessä tukeudumme toisaalta kielipiilliseen merkintään, jonka avulla verkoston järjestäytyneiden jäsenten suhteita voidaan ilmaista (ks. Itkonen, Pajunen & Suomi 2009; Pajunen 2010) ja toisaalta yleisyyteen. Kielen yksiköt, joita tutkimme, sijoittuvat joko horisontaaliselle tai vertikaaliselle tasolle, so. toisaalta sanat, lausekkeet ja lauseet ja toisaalta näiden tekstiin peräkkäin sijoittuvien yksiköiden sisäinen hierarkkinen rakenne. Johtamaton sana ja vain välttämättömistä jäsenistään koostuva lauseke ovat yksinkertaisia yksikköjä: niistä ei voi enää ottaa mitään pois kokonaisuuden muuttumatta. Tämä pätee sekä paradigmaattisesta että syntagmattisesta lausekkeesta. Johdetut sanat ja yhdyssanat, määritteitä sisältävät NP:t ja useammasta osasta ja/tai sidonnaisesta morfeemista koostuvat finiittimuodot ovat kompleksisia. Yleisimmät sanat ovat useimmiten rakenteeltaan yksinkertaisia ja leksikaalisen hierarkian yleistason sanoja, merkitykseltään usein hyvin konkreettisia. Kontekstissa ne ovat suhteellisen vähän taivutettuja (eli toimivat esimerkiksi subjektina tai objektina). Rakenteellinen kompleksisuusasteikko on toisin sanoen laaja: sekä sana-, lauseke- ja lausetason yksikköjä että näitä laajempia rakenteellisia kokonaisuuksia voidaan tarkastella asteikolla yksinkertainen–kompleksinen. Kompleksisten yksikköjen kehityksellinen merkitsevyys on osoitettu usean kielen analyysistä (ks. esimerkiksi Bromley 1991; Berman & Verhoeven 2002; Berman 2004; 2007; Nippold 2006; Nieminen 2007; Jagaiah ja muut 2020). Myös toisen tai vieraan kielen oppimisessa kompleksistuminen on varteenotettava tekijä mutta analyysi edellyttää aina lähde- ja kohdekielen rakenteiden vertailua, jos vertailulla ennustetaan esimerkiksi L2-kielen oppimista (ks. esimerkiksi Kaivapalu 2005). Aina ei näet ole ilmeistä, mitkä yksiköt vertautuvat toisiinsa eri kielissä (vrt. Vainio, Pajunen & Hyönä 2014; 2016). On vielä huomattava, että vaikka rakenteellinen kompleksistuminen on kehityksellistä, sen yhteys kirjoittamisen tasoon on huonosti tunnettu ja tutkimustulokset vaihtelevat (vrt. Beers & Nagy 2009, 186; Jagaiah ja muut 2020). Tasoa yleensä mittaavat holistiset arviot eivät

yksinään hevin voikaan toimia kehityksellisinä ja määriteltävinä mittareina. Eri arvioijien näkemuserot tekstin laadusta voivat myös olla suuret (Boewer, Béguin, Sanders & van den Bergh 2015).

Rajaamme kompleksistumisen tutkimisen tässä luvussa toisaalta sanatason, toisaalta lausetason yksiköihin. Rakenteellista kompleksistumista on Koulu- ja Unelmakorpuksen aineistoilla – siis samalla tavoin koodatusta aineistosta – tarkasteltu myös monessa opinnäytteessä. Esimerkiksi Mäkelä (2012) käsittelee NP:n kompleksistumiskehitystä eli hierarkkistumista, Nahkola (2012) yhdyslauseiden kehitystä, Oinonen (2008), Elomäki (2011), Salmenoja (2013) ja Rainoja (2014) yleisimmin rakenteellista kehitystä. S<sub>2</sub>-näkökulma on Hongolla (2013) ja Viljasella (2012). Haapalan (2008) pro gradu -tutkielmassa osoitettiin yleisen kielikokeen aineistolla verbilausekkeen predikaattiosan odotuksen mukainen kompleksistuminen taitotasoasteikoittain; vastaava kehitys on näkyvissä alakoululaisten kirjoitelmista (Keskinen 2012; Pajunen 2012). Tekstin rakenneosia ovat tutkineet muun muassa Sintonen (2011) ja Vesala (2012).

Tekstisanojen pituus ja niiden vähittäinen pidentyminen mittaa suomen kaltaisessa synteettisessä kielessä sekä sanojen taivutuksen että johdosten ja yhdyssanojen määrää. Suomen johtamattomien sanojen merkkimäärä – sanapituus – vaihtelee tyypillisimmin 3–7, vaikka yksittäisiä kahden tai yli seitsemän merkin mittaisia johtamattomia sanoja löytyy. Keskimääräisen sanapituuden kasvu toimii suomen kielessä yksinkertaisena kehityksellisenä mittarina (ks. Pajunen 2012; Honko 2013), koska lyhyet sanat ilmaisevat tyypillisesti perustason merkityksiä (esimerkiksi *juosta* ja *kaatua*) ja pitkät sanat spesifisempiä ja täsmällisempiä merkityksiä (esimerkiksi *juoksennella* ja *kompastua*) ja koska edellisten väheneminen ja jälkimmäisten lisääntyminen käy iän mukaan. Sanapituus korreloi tyypillisesti monen sisällöllisen muuttujan kanssa, kuten käytettyjen sanojen harvinaisuus, diversiteetti, leksikaalinen tiheys ja leksikaalinen spesifisyys (ks. esimerkiksi Malvern, Richards, Chipere & Durán 2008, 157; ks. myös Pajusen & Hongon analyysiä Unelmasanastosta tässä teoksessa). Kuvausta voi tarkentaa sanarakenteen analyysin avulla jakamalla aineiston sanat johtamattomiin, johdettuihin, yhdyssanoihin sekä esimerkiksi symboleihin ja sanaliittoihin. Sanatason

muuttujiin kuuluvat myös sanaluokat, joihin tyypillisesti perustuvat erilaiset tekstin ominaisuuksia mittaavat indeksit.

Sanaluokkaindeksejä on useanlaisia, ja niiden avulla voidaan antaa esimerkiksi tekstilajikohtaisia tekstin tyyppiä, laskea tekstin informatiivisuutta ja tutkia jonkin tarkkaan valitun ilmiön esiintyvyyttä (ks. esimerkiksi Read 2000; Berman, Nayditz & Ravid 2011; Price & Jackson 2015, 283). Niillä arvioidaan muun muassa leksikaalista tiheyttä eli leksikaalisten sanojen osuutta kokonaissanamäärästä (ks. O’Loughlin 1995). Niillä voidaan arvioida myös substantiivimäärän suhdetta verbimäärään (*noun-to-verb ratio*, ks. esimerkiksi Biber, Conrad & Reppen 1998) ja edelleen substantiivitiheyttä, joka vaihtelee tekstityypin mukaan. Toisaalta indeksit mittaavat informaatiotiivyyttä, leksikaalista rikkautta. Toisaalta indekseillä voi arvioida tekstin rakenteen luonnetta, esimerkiksi korkea finiittiverbipitoisuus viittaa rakenteellisesti yksinkertaiseen ja peruslauseilla etenevään tekstiin.<sup>112</sup>

Sanaluokkaindeksit edellyttävät luonnollisesti aineiston yksityiskohtaista ja yhtenäistä sanaluokitusta. Koko Unelma-aineisto on sanaluokiteltu ensin perussanaluokkiin (substantiivit, verbit, adjektiivit, adverbit, numeraalit, pronominit, konjunktiot ja interjektiot) ja edelleen alaluokkiin (muun muassa appellatiivit ja propriit, finiittiset ja ei-finiittiset verbimuodot). Lisäksi sanoja on analysoitu funktionaalisesti ja semanttisesti. Erityisen tarkkoja on oltu rajailmiöiden yhtenäisestä analyysistä. Esimerkiksi substantiiviparadigmasta irtautuneet adverbit on koodattu systemaattisesti samalla tavoin ja koodaus on tarkistettu arvioijareliabiliteettia laskemalla. Unelma-aineistosta on tehty sanaluokka-analyysia runsaasti myös opinnäytetöissä (ks. esimerkiksi Kuisma 2011 adjektiivien käytöstä; Keskinen 2012 modaaliverbeistä; Koivunen 2012 adverbeista; Nahkola 2012 konjunktioista; Palonen 2013 pronomineista). Tavoitteena on toisaalta osoittaa näiden muuttujien avulla tavanomaista kehitystä ja ikäluokkakohtaista variaatiota. Toisaalta pyrimme osoittamaan kertovan tekstin ikäluokkakohtaisia ja yleisiä tyyppiä, joissa hajonta vähenee kehityksellisesti.

112 Peruslauseella tarkoitetaan lausetta, joka kiistattomasti sisältää vain yhden predikaation (ks. Itkonen 2010, 203). Kyse voi olla intransitiivisesta tai transitiivisesta verbilauseesta tai *olla*-verbin kanssa järjestystä kopulalauseesta.

Käsitlemme tässä yhteydessä rakenteellista kompleksisuutta sanatason lisäksi myös lauserakennetta tutkimalla. Lauserakennetta analysoidaan niin sanottujen T-yksiköiden avulla (ks. Schneider ja muut 2005). T-yksikkö koostuu päälauseesta ja siihen liittyvistä finiittisistä ja ei-finiittisistä aineksista, jotka on selvästi merkitty alisteisiksi käyttämällä alisteisia elementtejä (komplementtilauseet, relatiivilauseet sekä konjunktiolla *koska*, *kun*, *vaikka* jne. alkavat adverbiaalilauseet), infinitiivejä tai lauseenvastikkeita (temporaali-, partisiippi- ja modaalirakenteet). Myös lausemuotoiset suorat ja epäsuorat lainaukset on laskettu alisteisiksi. Sen sijaan *minen*-johdoksia ja lausesubjektiin tai -objektiin teemasuhteessa olevia partisiippeja (esimerkiksi *lapsi istui kyynelehtivänä*) ei ole huomioitu alisteisina rakenteina, koska niiden yhteydessä on runsaasti tulkinnallisia tapauksia; tosin kumpikin ryhmä kasvaa kehityksellisesti. *minen*-johdoksia ja ei-finiittisiä aineksia arvioidaan kompleksisuuden tekijöitä analysoitaessa (ks. tämän teoksen loppuluku). T-yksikön laskentatapa pohjautuu semanttiseen tasoon eli propositioihin<sup>113</sup>, jotka siis rakenteen tasolla on koodattu joko päälauseeksi tai alisteiseksi ainekseksi. Kun jaottelu keskeiseen ja taustoittavaan ainekseen on kirjoittajan, tulos heijastaa hänen näkemystään ja tapaansa tiivistää sanottavaa. T-yksiköistä lasketaan koehenkilökohtaisesti niin keskimääräinen sanamäärä (= *mean length of terminal units* eli MLTU) kuin kompleksisuusindeksi (CI eli alisteisten yksikköjen suhde T-yksiköiden kokonaismäärään). MLTU ja CI ovat keskeisimpiä kehityksellisistä mittareista ja kaikkein yleisimmin käytettyjä (Jagaiah ja muut 2020); edellinen ilmaisee horisontaalisen tason kielenyksikköjen määrän ja jälkimmäinen vertikaalisen tason hierarkisuutta. Näiden mittareiden avulla pyrimme selvittämään kirjoitelmien syntaktisen tiivyyden (*syntactic density*), jonka tiedetään nousevan kehityksellisesti (ks. Nippold ja muut 2005; 2007; Berman 2007; Beers & Nagy 2009; 2011; suomesta ks. Petälä 2018; Muhonen 2019; Tiiri 2021).

Keskitymme tässä luvussa toisaalta rakenteellisiin muuttujiin, kuten sanapituus, sanarakenne ja erityyppinen rakenteellinen kompleksisuus, toisaalta sanaluokkien avulla laskettaviin indekseihin, joilla voidaan tar-

113 T-yksikköjen määrittely perustuu siis semanttiseen tasoon ja siksi analyysissä on irrelevanttia se, miten yksittäisissä kieliopeissa propositioiden reaalistumia luokitellaan.



kentaa lausetason muuttujien antamaa kuvaa tekstin luonteesta. Sanastomuuttujia ja kompleksisuustekijöitä käsitellään tässä teoksessa myös erikseen. Sukupuoli on muuttujana mukana tässäkin luvussa. Muuttujia käsitellään keskiarvolukujen, jakaumien ja niiden hajonnan vähenemisen kannalta sekä yleisyysluokitettuina. Todettakoon jo tässä vaiheessa, että ikäryhmittäinen hajonta tyyppillisesti kapenee iän myötä mutta ei häviä kokonaan. Oletus luonnollisesti on, että koulutuksen pitäisi vähentää hajontaa oleellisesti jo peruskouluvaiheessa (vrt. esimerkiksi Hadley & Ripoli 2012; Truckenmiller & Petscher 2020).

### AINEISTO JA TUTKIMUSMENETELMÄT

Aineistona on sekä Koulu- että Unelmakorpuksen kirjoitelmat (ks. tämän teoksen johdantoa), jotka on kirjoitettu samasta aiheesta eli unelmienvaiheesta samalla tehtävänannolla: kirjoita unelmiesi päivästä, jonka olet kokenut, haluaisit kokea tai voisit kokea maailmassa, missä mikä tahansa olisi mahdollista (ks. Pajunen 2012). Koulukorpuksen kirjoittajat ovat alakoulun luokilta 1–6. Tästä aineistosta ei ole laskettu kaikkia vertailulukuja, koska osa nuorimpien teksteistä on laskelmiin liian lyhyitä, jopa yksisanaisia. Unelmakorpuksen kirjoittajat ovat kuudesluokkalaisia ja yläkoululaisia sekä nuoria aikuisia. Koulukorpuksen kuudennen luokan aineisto on kirjoitettu kahdeksan vuotta aikaisemmin kuin Unelmakorpuksen kuudennen aineisto, ja tämä aikaero mahdollistaa jo kuudensien vertailua. Peruskoululaisten aineistoja vertaillaan nuorilta aikuisilta kerättyihin aineistoihin, jotka koostuvat toisaalta ammattikorkeakoululaisten sukupuolen mukaan vertaistetusta aineistosta ja toisaalta suomen kielen pääsykoeaineistosta, jossa naisten osuus kirjoittajista on runsaat 80 prosenttia. Kirjoittajia on yhteensä lähes 1900 ja heitä on joka ikäluokasta vähintään noin sata.

Peruskoululaisten aineistoista lasketaan yleisiä tunnuslukuja vuosiluokittain tai ryhmittäin mutta osasta aineistoa on tehty tarkempia analyysyjä, joihin pyrimme saamaan sukupuolittain satakunta koehenkilöä joka ryhmästä, mikä olisi edustava alaryhmäkoko myös jakaumien kannalta (ks. Nippold ja muut 2005, 1059); se ei täysin kaikissa ryhmissä onnistu. Osasta aineistoa on myös erilaisia kielellisiä tai muita vertailulukuja. Kuudes- ja yhdeksäsluokkalaiset sekä ammattikorkeakoululaiset

ovat osallistuneet elektronisiin sanaston hallintaa mittaaviin testeihin ja suomen kielen pääsykokeeseen osallistuneista on tietoa paitsi kieliopin ja kielenkäytön pääsykoeosioista myös ylioppilastutkinnon arvosanoista sekä sanastosta (ks. tarkemmin tämän teoksen johdanto). Aineistosta on koostettu myös koehenkilökohtainen taulukko, johon on koottu kaikki eri muuttujista lasketut arvot ja kaikki yhdestä koehenkilöstä saatavilla ollut tieto. Aineistoa analysoidaan laadullisesti, määrällisesti ja tilastollisesti. Tavoitteena on tässä yhteydessä antaa aineistosta yleiskuva, ja painopiste on kielellisen rakenteen kehityksellisessä muutoksessa. Tavoitteena on myös ryhmänsisäisen hajonnan esiin nostaminen. Tässä yhteydessä emme käsittele sanastoa, kirjoittamistavan muutoksia, kertomusskeemaa tai sisällöllisiä tekijöitä vaan keskitymme juuri laskettaviin rakennekeskeisiin muuttujiin. Laadullisempi ote on monissa muissa aineistoon perustuvissa tutkimuksissa. Vertaamme joitain tuloksia myös sanomalehtikielen vastaaviin. Ne on laskettu Savon Sanomien korpuksen (Pajunen & Virtanen 2002) miljoonan tekstisanan otoksesta ja niiden on tarkoitus havainnollistaa tekstityyppi- ja lajieroa.

## Kypsyys

Kypsyys eli kirjoitetun kielen konventioiden karttuva hallinta on helpoin operationalisoida pintafrekvenssien avulla. Pintafrekvenssillä lasketaan melko helposti tavoitettavia, koko tekstimassaa koskevia yksikköjä, joiden määrä tyyppillisesti kasvaa iän mukana tiettyyn oletustasoon. Matalat arvot liittyvät yleensä kehittyvään kieleen, korkeat arvot kuvaavat kompleksisempaa kieltä (ks. esimerkiksi Olinghouse & Wilson 2013). Pintafrekvenssien paljastama muutos on osin oireellinen, koska niihin vaikuttaa voimakkaasti kirjoitelman onnistuneisuus, kirjoittajan motivaatio ja toisaalta kokonaissanamäärä sekä luonnollisesti aihe ja tekstityyppi tai -laji (Pajunen 2012; Kyösti 2018). Tekstin pituus vaikuttaa erityisesti leksikaaliin indekseihin kuten diversiteettiarvoihin (vrt. esimerkiksi McCarthy & Jarvis 2007; Jarvis 2013). Omassa aineistossamme hyvin lyhyet kirjoitelmat vinouttavat pintafrekvenssejä ja tästä syystä mediaani tai aineiston uudelleen analysointi yleisyysluokittain antaa ikä-

luokasta paremman kuvan kuin esimerkiksi keskiarvo. Aineiston tilastollisesta analyysistä kaikkien lyhimmät kirjoitelmat (eli alle 50 sanaa) jätetään poikkeavina arvoina pois, jotta yleiskuva ei vääristyisi. Pintafrekvensseistä painotamme sanapituuden kasvua sekä keskimääräisen iän mukaan että sukupuolittain tarkasteltuina. Toiseksi tarkastelemme sanarakennemuutoksia ja kiinnitämme huomiota erityisesti johdosten käytön lisääntymiseen. Nämä kaksi ensimmäistä muuttujaa mittaavat leksikaalista kompleksistumista (Strömqvist, Johansson, Kriz, Ragnarsdóttir, Aisenman & Ravid 2002). Lisäksi tarkastelemme kokonaissanamääriä. Sanamäärä on lähinnä nuorimmilla kirjoittajilla kehityksellinen yksikkö, vanhemmilla kirjoittajilla on jo yksilöllistä taipumusta tiiviiseen tai laveaan kerrontaan. Pintafrekvenssien antamaa kuvaa pyrimme tarkentamaan muilla muuttujilla.

## SANAPITUUS

Tekstisanojen sanapituuden vaihtelu liittyy käytetyn sanaston laatuun ja taivutuksen määrään: Mitä enemmän käytetään kielen kaikkein yleisimpiä sanoja, sitä lyhyempiä sanat keskimäärin ovat. Mitä pidemmät sanat ovat kirjoittajilla käytössä, sitä kompleksisempia ne rakenteeltaan suomen tapaisessa kielessä ovat. Kompleksiset sanat ovat sekä johdoksia että yhdyssanoja ja lisäksi enemmän taivutettuja kuin yksinkertaisemat sanat. Näitä tekijöitä ei voi erotella pelkän sanapituus-muuttujan avulla. Oletus luonnollisesti on, että nuorimmat kirjoittajat käyttävät kaikkein lyhyimpiä sanoja. Sanapituudet on laskettu varovasti yleiskielistetyistä sanoista, jolloin esimerkiksi kirjoitusvirheet, puheenomainen loppuheitto tai yhdyssanavirheet eivät vaikuta tuloksiin (ks. taulukko 49). Aineiston kokonaissanamäärä on yli 230 000.

Taulukko 49. Tekstisanapituudet (= kirjainmerkkimäärä) Unelmakirjoitelmissa.

Ikäluokka / Tunnusluvut	Keskiarvo	Mediaani	Keskihajonta
Alakoulu 1.–5.	6,0	6	2,9
Alakoulu 6.	6,28	6	2,5
Yläkoulu 7.–8.	6,48	6	3,3
Yläkoulu 9.	6,59	6	3,2
Aikuiset, pääsykoe	6,99	7	3,3
Aikuiset, AMK	7,02	7	3,4

Tekstisanapituudet kasvavat kirjoitelmissa tasaisesti ikäluokasta ja luokkasteesta toiseen. Nuorimpien ja vanhimpien kirjoittajien välillä ero keskiarvoissa on yksi merkki. Se ei vaikuta suurelta mutta on merkityksellinen, koska taiputus usein lisää sanapituutta juuri yhdellä merkkillä (vrt. esimerkiksi genetiivi *-n*; partitiivi *-a*) ja koska johtamisen yhteydessä johtimen pituus vaihtelee yleensä 1–3 merkkiin. Toisin sanoen aikuiskirjoittajilla sanoissa on joko keskimäärin enemmän taiputusta tai sanarakenne on kompleksisempi. Sanapituuden mediaani on kaikilla peruskoululaisilla kuusi merkkiä ja aikuiskirjoittajilla seitsemän. Aikuiskirjoittajilla mediaani on sama kuin Savon Sanomien korpuksessa mutta sanomalehtikielessä sekä keskiarvo (7,8)<sup>114</sup> että keskihajonta (4) ovat korkeampia. Keskeinen syy eroihin lienee toinen tekstityyppi tai -laji. Unelmakirjoitelmissa myös sanapituuden vaihtelu (keskihajonta) lisääntyy iän mukaan mutta ei yhtä suoraan kuin sanapituudet. Leksseemipituudet ovat ikäryhmittäin tyyppillisesti yhtä merkkiä lyhyemmät (5,19–5,89).

Sukupuolen mukaan tarkasteltuna alakoulun luokkien 1.–5. pojat käyttävät keskimäärin aavistuksen pidempiä tekstisanoja kuin tytöt. Tyttöillä vaihteluväli on iän mukaan 5,9–6,9 ja pojilla 6,1–7,1. Ero kääntyy kuudennella luokalla tyttöjen hyväksi ja selittyy siten, että kaikkein nuorimmat pojat kirjoittavat paljon tyttöjä vähemmän ja usein substantiiveja luettelemalla. Toisin sanoen poikien aineistossa kieliopilliset – eli lyhyet – sanat ovat vähissä (ks. Pajunen 2012). Toisaalta ero kääntyy taas aikuisiässä miespuolisten hyväksi. Tämä liittyy miesten kompleksisempaan kirjoittamistapaan, joka tulee monen muuttujan kautta esille. Mediaanit ovat koko aineistossa ja sukupuolen mukaan erotellen samat, mutta poikien sanapituuksissa on enemmän hajontaa kuin tyttöjen.

Sanapituuden keskiarvo on laskettu myös koehenkilökohtaisesti jakamalla kuudennen, yhdeksännen ja aikuisten kirjoittamien sanojen merkkimäärä kirjoitelman sanamäärällä ja luokittelemalla aineisto saadun luvun mukaan keskimäärin alle viisimerkkisiin ja puoli merkkiä

114 Lauseopin arkiston 1970-luvun sanomalehtikielen sanapituuden keskiarvo oli 7,4–7,9 (ks. Pajunen & Palomäki 1984, 70–71), joten sanomalehtikielessä ei ole tapahtunut sanottavaa muutosta 30 vuoteen.

kerrallaan aina yli kahdeksanmerkkisiin (ks. taulukko 50). Näin saadaan hajonta näkyviin. Ryhmien sisäiset erot ovat suuremmat kuin ryhmien väliset mutta valtaosin hajonta keskittyy peruskoululaisilla välille 5,5–6,5 ja aikuisilla 6,5–7,5. Nämä ovat ikään kuin iänmukaisia tyypillisiä arvoja kertovassa tekstissä. Ääriarvot, jotka sijoittuvat alle 5,5 tai yli 7,5 ovat selvästi poikkeavia; niiden prosenttiosuus on tässä muuttujassa odotuksenmukainen (toisin sanoen 2–3 %).

Taulukko 50. Sanapituuksien keskiarvon mukainen jakauma.

Ikäluokka / Sanapituusluokat %	Alle 5 merkkiä	5	5,5	6	6,5	7	7,5	8 tai enemmän
Kuudes	0,6	2,9	18,8	39,6	23	9,6	4,1	1,3
Yhdeksäs	0,0	0,6	11,9	28,7	36,2	16,2	5,6	0,8
Aikuiset			1,2	13	40,3	31,1	12,5	1,8

Sanapituuksien vaihtelu liittyy tekstintuottamisen erilaisiin strategioihin ja toisaalta kehityseroihin. Tämä näkyy muun muassa sanapituuskeskiarvojen ja leksikaalisen indeksin korrelaatiosta: mitä pidemmät sanat kirjoitelmassa, sitä enemmän leksikaalisia sanoja. Lisäksi tuottamisongelmat ja motivaatio vaikuttavat myös sanapituuksiin. Esimerkiksi Kyöstin (2018) tutkimuksen kuudesluokkalaisilla sanapituuskeskiarvot vastasivat suunnilleen tämän tutkimuksen lukuja, mutta toisen kirjoituskerran motivaation nostattaminen nosti tyttöjen sanapituuksia ja laski poikien. Ääriarvoja saaneet tekstit ovat muilta ominaisuuksiltaan vaihtelevampia kuin lähellä ikäkeskiarvoa olevat. Esimerkiksi hyvin lyhyt teksti voi saada sekä matalan että korkean sanapituuskeskiarvon, joten pelkkä koehenkilön sanapituuskeskiarvo ei vielä kerro kaikkea. Lyhytsanaiset tekstit voivat olla puhekielisiä ja kertovia, pitkäsanaissa leksikaalinen tiheys eli leksikaalisten sanaluokkien yhteismäärä voi olla korkea, mutta teksti sinällään ei välttämättä ole kummassakaan tyypissä tasokas.

Esimerkkeihin 1–3 on valittu näyte kunkin ikäluokan lyhytsanaisestä tekstistä. Näytteet ovat kaikki naispuolisten kirjoittamia ja edustavat vertailtavuuden vuoksi hieman samantapaista tyyliä (mutta eivät onnistumisastetta): iän tuoma muutos näkyy sanapituuksissa selvästi. Kuudennen ja yhdeksännen luokan näytteet vaikuttavat puhekielisiltä ja suunnittelemattomilta, aikuisnäyte on kuvailevaa kerrontaa. Usein

puheenomaisessa, rakenteellisesti latteassa kerronnassa on naiskirjoittajilla lyhyitä sanoja. Mieskirjoittajilla ääripituudet johtuvat usein kirjoittamisongelmista: vähän tekstiä, substantiiviluetteloa ja niin edelleen. Kun sitten sanojen keskipituus nousee, se näkyy monissa tekstin ominaisuuksissa, muun muassa sekä leksikaalisessa että syntaktisessa tiheydessä. Vaikutelma siis on se, että ikäluokittain keskisanapituuksissa olevat tekstit ovat ominaisuuksiltaan neutraaleimpia ja ehkä ikätason osaamista parhaiten kuvaavia. Tätä oletusta voi tarkastella muiden muuttujien analyysin avulla.

Kuudes luokka (sanapituuden keskiarvo 5):

1) *Olipa kerran Tyttö, joka oli aina haaveillut omaa hevosta. Mutta hän ei saanut sitä kun he olivat hyvin hyvin köyhiä. Niillä ei ees melkeen riittäneet ruokaakaan rahaa.* (6F50)<sup>115</sup>

Yhdeksäs luokka (ka 5,6):

2) *Pyysin äitini hakemaan minut harkoista. Sitten hän tuli ja sanoi että lähdemme Suoniemelle katsomaan koiraa. Lähdimme ajamaan kohti Suoniemeä. Sitten saavuimme sinne missä koirat on.* (9F23)

Aikuiset (ka 6,2):

3) *Tunnen herääväni. Tunnen, kuinka aurinko on jo lämmittänyt osan yön kasteesta sen nousten nyt sumuverhoina peittämään vielä unessa olevia kasveja. Avaan silmäni ja olo on virkeä ja terve.* (AikF89)

## SANARAKENNE

Sanarakenteen perusteella sanasto voidaan jakaa yksi- ja monimorfeemiisiin ja jaolla voidaan arvioida ilmaisujen granulariteettia eli spesifisyysastetta. Oletus on, että yksimorfeemisten sanojen tarkkuus on matalampi kuin monimorfeemisten (Strömqvist ja muut 2002). Merkitykseltään

115 Esimerkinäytteissä ensimmäinen numero/osio viittaa luokka-asteeseen, ikäluokkaan tai ryhmään, toinen (F/M) kertoo sukupuolen ja kolmas ilmaisee kirjoitelman järjestysnumeron luokan mukaan numeroituna. Aikuisryhmään viitataan lyhenteellä Aik ja alaryhminä AMK viittaa ammattikorkeakouluisten aineistoon, Pääsy taas suomen kielen pääsykokeeseen osallistuneilta kerättyyn aineistoon.

rajatumia eli tarkempia voidaan käyttää harvemmissä konteksteissa kuin merkitykseltään rajaamattomampia, mutta rajaamattomampi merkitys jättää enemmän tulkinnanvaraa kuin rajattu. Sanan johtamattomuus, lyhyys ja yleisyys ovat erityisesti koululaissanastoa kuvaavia ominaisuuksia. Esimerkiksi lehtikorpuksen ja koululaissanaston yleisissä verbeissä on johtamattomia verbejä kuten *ajaa*, *maata* ja *syödä*. Harvinaisempien verbien joukossa on johdoksia kuten *ajella* ja *mussutella*, aikuissanastossa myös *lekotella*. Verbi *syödä* ilmaisee ravinnon nauttimisen, *mussutella* lisäksi syömistävän ja -keston. Oletus on myös, että monimorfeemisuus – leksikaalinen kompleksisuus – nousee kehityksellisesti. Erityisesti johdettujen sanojen katsotaan olevan tärkeitä lasten ja nuorten myöhemmässä kielenkehityksessä (ks. esimerkiksi Carlisle 2000; Kirby, Deacon, Bowers, Izenberg, Wade-Woolley & Parrila 2012; Truckenmiller & Petscher 2020, 61–62). Kielissä, joissa sananjohto on monimutkaista (vrt. Lieber & Štekauer 2014), johdosten oppiminen on erityisen hidasta ja siten kehityksellisesti sitäkin merkitsevämpää. Suomessa ja esimerkiksi hepreassa sananjohtoa opetellaan vielä nuoruusiällä ja aikuisenakin (ks. Ravid 1998; Vainio, Pajunen & Häikiö 2019; Pajunen & Itkonen 2019). Englannin tapaisessa kielessä, jossa sanoja johdetaan suomeen verrattuna vähän, sananjohto hallitaan toisella kymmenellä (Nippold & Sun 2008). Suomen verbiderivaatiolla tuotetaan paitsi spesifejä merkityksiä myös muunnellaan lauserakennetta. Nämä muunnokset edellyttävät kykyä nähdä tapahtumia milloin tekijän milloin vaikkapa kohteen, kokijan tai lähteen näkökulmasta ja vaikuttavat tekstin koherenssin luomiseen.

Unelmakirjoitelmien aineisto on koodattu sanarakenteen mukaan luokkiin johtamattomat, johdetut ja kiteytyneet sanat sekä yhdyssanat ja muut. Johtamattomien ja johdettujen sanojen erottelu perustuu läpinäkyvyyden periaatteelle ja pyrkii noudattamaan valistuneen maallikon tietämystä sanarakenteesta. Analyysi ei siis ole kielihistoriallista mutta ei myöskään perustu oletuksiin lasten johdostuntemuksesta. Useimmiten yksi- tai kaksitavuinen sana on johtamaton, mutta myös kolmitavuiset supistumavartaloiset verbit (*hakata*, *loukata*) on luettu johtamattomiin, samoin *-ea*-loppuiset adjektiivit. Johdos tarkoittaa siis pääasiassa kaksitavuista pidempää sanaa, jossa johdin on aika tavallisen tuntuinen,

helposti segmentoitavissa ja ainakin periaatteessa vielä produktiivinen. Yhdyssana on kahdesta tai useammasta johtamattomasta tai johdetusta sananosasta koostuva sana. Yhdyssanat on laskettu kirjoitusasultaan yleiskielistettyinä, joten määriin ei vaikuta oikeinkirjoituksen hallitsemattomuus (vrt. Hongon analyysi tässä teoksessa). Kiteytyneisiin sanoihin kuuluu sijataivutuksen pudokkaita, joilla nykykielen kannalta ei yleensä ole selvää kantasanaa. Luokassa on runsaasti pronominaalien paikallissijamuotoja (*siellä, täällä*) ja yleensä sanoja, jotka taivutettuina ovat adverbjeja, eivät kantasanansa mukaisesti esimerkiksi substantiiveja. Luokkaan muuten kuuluu sanaliittoja (kuten kirjojen, elokuvien ja pelien nimiä), symboleja (esimerkiksi \$, @) ja vierassanoja (esimerkiksi englanninkieliset sanat). Sanarakenteen analyysin avulla voidaan tarkentaa sanapituuksien vertailua. Sanarakenteen analyysi näyttää rakenteellisesti yksinkertaisten ja monimutkaisten sanojen määrän tarkasti mutta ei anna tietoa taivutuksen määrästä. Tarkastelemme tässä yhteydessä vain tekstisanojen sanarakennetta.

Johtamattomien tekstisanojen määrä on vallitseva koko aineistossa. Tämä on odotuksenmukaista, koska johtamattomat sanat ovat kielen yleisimpiä sanoja (ks. taulukko 51); lekseemitasolla aineiston tilanne on toinen. Johtamattomien tekstisanojen määrää selittää myös kirjoitelmien kertova luonne; abstraktiot, joita ilmaisevat sanat ovat tyypillisesti johdoksia, ovat vähissä. Johtamattomien tekstisanojen määrä kuitenkin laskee lähes lineaarisesti kirjoittajien luokka-asteen ja iän noustessa sukupuolesta riippumatta. Nuorimpien alakoululaisten ja aikuisten välinen ero on noin 15 prosenttiyksikköä, kuudesluokkalaisten ja aikuisten vajaa kymmenen prosenttiyksikköä ja yhdeksäsluokkalaisten ja aikuistenkin välillä on vielä selvä ero. Kun johtamattomien tekstisanojen määrä laskee, johdettujen tekstisanojen määrä vastaavasti nousee. Kehitys on nyt vielä selvemmin lineaarista. Johdettujen sanojen lisääntyminen johtuu morfologisen tietoisuuden kehittymisestä ja ilmaisujen spesifisyysasteen (granulariteetin) ja abstraktisuuden kasvusta. Sukupuoliero on jo odotuksen mukaan pieni, koska derivaatiomorfologian testauksessa samoilla koehenkilöillä ero ei ollut mainittava (ks. Vainio ja muut 2019). Toisaalta Kusnetsoffin (2017) mukaan sukupuolierot ovat vain derivaation tunnistamistehtävissä pienet, tuottamistehtävissä suuret.



Taulukko 51. Sanarakenteen jakauma ikäluokittain.

Ikäluokka / Sanarakennetyypit %	Johtamaton	Johdos	Yhdyssana	Kiteytynyt	Muut
1.–5. F	76,9	10,4	5,7	6,4	0,6
1.–5. M	75,2	10,2	7,1	5,9	1,5
Kuudes F	72	12,7	6,1	7,6	1,6
Kuudes M	71,9	11,1	7,3	7,4	2,3
Yhdeksäs F	70	15,1	6,4	7,5	0,9
Yhdeksäs M	68,1	14,4	8,2	7,7	1,7
AMK F	65,8	18,9	7,6	7,4	0,3
AMK M	64,3	19,2	8,6	7,5	0,4
Pääsy F	64,1	20,8	7,7	7	0,4
Pääsy M	62,9	21,4	8,4	6,7	0,6

Yhdyssanojen määrä on ensimmäisellä luokalla ja toisella aikuisryhmällä sama ja vaihtelu on eri ryhmissä pientä, miehillä määrät ovat korkeammat. Laadullisesti yhdyssanoissa tapahtuu muutosta. Nuorilla kirjoittajilla on paljon tavallisia yhdyssanoja (*aamupala, syntymäpäivä*) ja toisaalta satunnaismuodosteita (*karkkikylpy, jäätelömaa*), mutta vanhemmilla kirjoittajilla yhdyssanat ovat merkitykseltään spesifejä (*hautavajoama, sukulaiskiintiö*). Toisin sanoen ero on selvä lekseemitasolla. Sanomalehtikielessä yhdyssanojen osuus on 12,5 prosenttia eli selvästi korkeampi kuin nuorilla aikuisilla. Kiteytyneiden määrässä on hienoista nousua iän mukaan, mikä todennäköisesti viittaa tapahtumien tarkempaan ajalliseen ja paikalliseen kehystämiseen, koska juuri kiteytyneissä sanoissa on paljon ajan ja paikan adverbeja.

Yleisyysjakaumien tarkastelun keskitämme johtamattomien ja johdosten suhteeseen kuudennella ja yhdeksännellä luokalla sekä aikuisilla. Jakaumat on muodostettu luokittelemalla kokonaissanamäärä suhteessa johtamattomiin ja johdettuihin sanoihin. Johtamattomien luokat vaihtelevat alle 50 prosentista yli 90 prosenttiin kokonaissanamäärästä ja johdettujen alle viidestä prosentista yli 30 prosenttiin. Johtamattomien sanojen osuus peruskoululaisilla on yleisimmin 70–80 prosenttia tekstisanoista ja asettuu melko tasaisesti välille 60–80 prosenttia (ks. taulukko 52). Kuudennen ja yhdeksännen välinen ero on keskeisissä yleisyysluokissa noin 10 prosenttiyksikköä. Aikuisilla vaihtelu sijoittuu yleisimmin välille 60–70 prosenttia mutta yleisyysluokassa 50–60 on

jo melkoinen määrä aikuisista. Peruskoululaisten ja aikuisten välinen ero on keskeisissä yleisyysluokissa noin 30 prosenttiyksikköä ja se on siten huomattavasti suurempi kuin kuudennen ja yhdeksännen välinen ero. Näin suuret erot kirjoitelmissa käytetyn sanaston rakenteellisessa koostumuksessa vaikuttavat luonnollisesti suuresti ilmaisujen geneerisyyteen ja spesifisyyteen, konkreettisuuteen ja abstraktiuteen. Ääriarvoja (yli 80 % sanoista johtamattomia) on kuudesluokkalaisilla lähes 10 prosenttia eli väestötason kielellisen älykkyyden normaalijakaumaan nähden liian paljon.

Taulukko 52. Johtamattomien sanojen osuus kokonaissanamäärästä eri kirjoittajilla.

Ikäluokka / Johtamattomien yleisyysluokat %	Alle 50	50–59	60–69	70–79	80–89	90 %
Kuudes	0,9	3,2	34,2	52,4	8,3	0,9
Yhdeksäs	1,9	7,5	44,4	43,1	3,1	0
Aikuiset	0,9	24,9	60,8	13,4	0	0

Kuudesluokkalaisilla johdokset sijoittuvat enimmäkseen yleisyysluokkiin 5–15 prosenttia kokonaissanamäärästä, yhdeksäsluokkalaisilla luokkiin 10–20 prosenttia. Aikuisilla yleisyysluokka välillä 15–25 prosenttia on vallitseva: lähes puolella kirjoittajista tekstisanoista yli 20 prosenttia on johdettuja sanoja. Tämä aikuisilla yleisin luokka on peruskoululaisilla pieni. Nuorilla aikuisilla johdosten määrä tasaantuu ja vaihteluväli laskee välille 10–30 eli pysyttelee noin 20 prosenttiyksikössä. Tämä voi olla osoitus siitä, että aikuisten ryhmässä ei ole yhtä laajaa kirjoittajakirjoa kuin peruskoululaisten ryhmässä. Toisaalta kyse on myös derivatiivomorfologian hallinnan kehittymisestä. Ikäryhmittäin johdosjakauma on melko tasainen, mutta kuudesluokkalaisilla matalia arvoja on oletusjakaumaan nähden liikaa.

Esimerkkeihin 4–5 on valittu sekä kuudennelta että yhdeksänneltä näytteet, joissa on toisaalta mahdollisimman vähän johdoksia ja toisaalta mahdollisimman paljon. Johdosten suhteellinen osuus vaikuttaa melko luotettavalta kielenhallinnan kehitykselliseltä mittarilta.

Kuudes (johdoksia sanoista 2,3 %):

4a) *JA voisin muuttaa johonki sinne missä pelaan. Se olisi hienoa voisin ostaa ison talon. Tai sitten voisin pelata SM sarjaa sumessa.* (6M122)

Kuudes (johdoksia sanoista 21,5 %):

4b) *Unelmieni päivän viettäisin jossain lämpimässä maassa, jossain rikkaassa paratiisissa. Haluaisin tietysti lentää sinne. Haluaisin liittää ja lentää hurjaa vauhtia.* (6F10)

Yhdeksäs (johdoksia sanoista 4,1 %):

5a) *Unelmapäiväni olisi varmaan se kun saisin koiran. Olen aina toivonut koiraa mutten vielä ole saanut sitä. Haluaisin myös lähteä ystäväporukalla vaikka laivalle tai jonnekin kivaan lomapaikkaan.* (9F16)

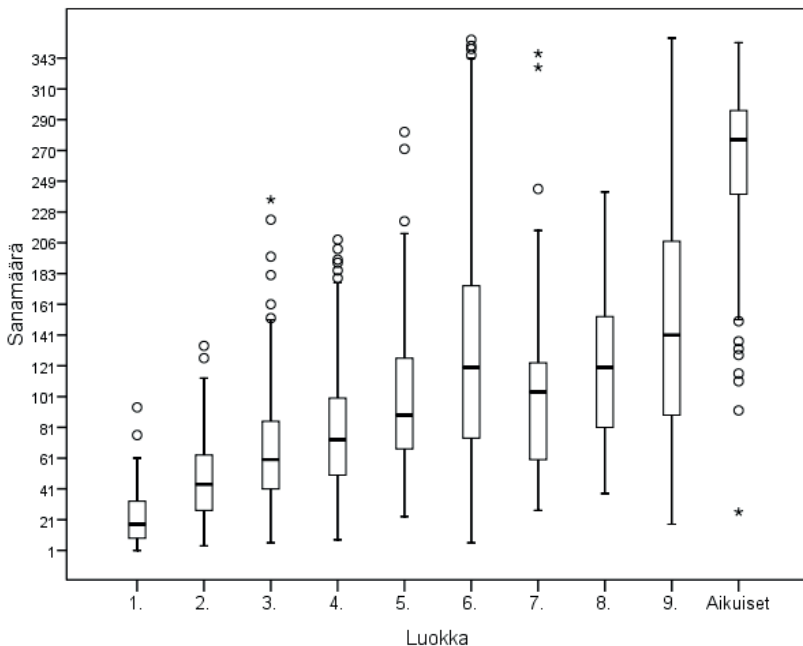
Yhdeksäs (johdoksia sanoista 24,84 %):

5b) *Ratsastuksen ja pipareitten jakojen jälkeen menisimme takaisin tallille ja koristelisimme tallin jouluisaksi. Jokaisen hevosen karsinan oveen tulisi iso pipari, jonka hevoset saisivat myöhemmin syödä.* (9F8)

## SANAMÄÄRÄN YKSILÖLLINEN VAIHTELU

Kirjoitelman sanamäärä on tekstitason mittari, jolla on helppo poimia erilleen ne kirjoittajat, jotka ovat selvästi alle ikäluokan ja sukupuolen yleisen produktiivisuuden (ks. Price & Jackson 2015, 285). Jos sanamäärä on matalissa ääriarvoissa, kyse voi olla kielihäiriön tuomasta ongelmasta. Mittaamisessa ylilyhyet tai ikäluokalle sanamääriltään muuten epätyypilliset kirjoitelmat antavat ongelmallisia arvoja. Esimerkiksi käyvät jo esiin tulleet keskisanapituudet, jotka nuorimmilla pojilla osoittautuvat epäodotuksenmukaisen korkeiksi. Syynä on kertovan tekstin korvannut substantiivivaltainen luettelo, josta kieliopilliset eli lyhyet sanat puuttuvat. On myös mahdollista, että lyhyissä kirjoitelmissa tulisi esiin vaikkapa kertovan tekstin kaikki rakenneosat. Tilastollisista monimuuttuja-analyyseista nämä ääriarvoja edustavat kirjoitelmat onkin poistettu, kuten jo edellä todettiin.

Iän vaikutus sanamääriin on todettu merkitykselliseksi useissa kielis-



Kaavio 15. Unelmakirjoitelmien sanamäärä ja sen hajonta kirjoittajaryhmittäin.

sä, mutta määrä sinänsä vaihtelee kielittäin (ks. Berman & Verhoeven 2002, 23). Syynä vaihteluun ovat ennen kaikkea kielitypologiset erot mutta varmaan myös tyylieroja on. Samasta temasta kirjoitettujen Unelmakirjoitelmien sanamäärä kasvaa lähes lineaarisesti koko peruskoulun ajan (ks. kaavio 15; ks. myös Pajunen 2012; Honko 2013).

Kuudennenluokan aineisto kaaviossa 15 poikkeaa silmin nähden muista. Syy on sen koostumus. Aineisto on kerätty vuosina 2006 ja 2014: jälkimmäinen ryhmä on laatinut pidempiä tekstejä, mikä nostaa paitsi mediaania myös hajontaa. Korkeampi kokonaissanamäärä ei kuitenkaan ole osoitus paremmasta tasosta vaan – tässä tapauksessa – yksinkertaisemmasta ja monisanaisemmasta rakenteesta (ks. Muhonen 2019). Vuoden 2014 kirjoittajia pyrittiin motivoimaan elokuvalippuja arpomalla ja motivaatio ehkä nosti sanamäärää mutta ei rakenteen kompleksisuutta. Kompleksisuus tarkoittaa myös tiiviyyttä, joten vuoden 2014 kirjoittajilla rakenne on paitsi yksinkertaisempi myös hajanaisempi. Rakenteellisesti

tarkasteltuna vuoden 2006 kuudesluokkalaiset muistuttavat enemmän vuoden 2014 yhdeksäsluokkalaisia kuin vuoden 2014 kuudesluokkalaisia. Yläkoulun seitsemännen ja kahdeksannen luokan kirjoittajia on muita vähemmän, joten nämä luvut eivät ole täysin edustavia.

Peruskoululaisten kirjoitelmien sanamäärä vaihtelee luonnollisesti tehtävyyppin ja tehtävänannon mukaan, motivaation lisäksi suunnittelu voi nostaa sanamääriä. Lisäksi tekstityyppi voi vaikuttaa sanamäärään. Usein katsotaan, että erittelevä teksti on lyhyempi kuin kertova (Koutsoftas & Gray 2012; Hall-Mills & Apel 2015). Aikuisten kirjoitelmissa sanamääräksi on ohjeistettu noin 300 sanaa. Toteutunut sanamäärä jää tämän alle, mutta hyvin monet kirjoittajat ovat pyrkineet kirjoittamaan täsmällisesti annetun ohjeen mukaan. Aikuisten kirjoitelmat on myös selvästi suunniteltu määrämittäisiksi. Vastaavasti peruskoululaisten kirjoitelmia jatketaan tarpeen vaatiessa kokonaisuudesta irrallisin lisätapahtumin. Vuoden 2014 kuudesluokkalaiset ovat turvautuneet juuri tällaisiin sattumalisyksiin. Suunnittelun rooli on siis erilainen koululaisilla ja aikuisilla (vrt. Sintonen 2011 peruskoululaisista).

Sanamäärän arvioimiseksi kirjoitelmat on uudelleen luokiteltu sanamääräryhmiin siten, että joka ryhmä kasvaa aina 50 sanalla (ks. taulukko 53). Luokitus osoittaa, että kuudesluokkalaisista tytöistä kolme neljästä kirjoittaa yli 50 mutta alle 200 sanan kirjoitelmia, pojista runsas 60 prosenttia. Neljännes kuudesluokkalaisista pojista kirjoittaa vielä hyvin lyhyesti. Yhdeksäsluokkalaisilla tytöillä sanamäärä vaihtelee useimmiten sadasta yli kahteen sataan mutta poikien sanamäärät ovat samoissa kuin kuudesluokkalaisten poikien; silti hyvin lyhyiden tekstien kirjoittajat ovat vähentyneet. Aikuiskirjoittajilla sukupuoli ei ole enää merkitsevä. Johtopäätös kuitenkin on, että kirjoittamistaitojen tai -halujen kehitys ei näytä koskevan kaikkia kirjoittajia on sukupuoli kumpi tahansa. On tietysti myös niin, että kaikkien osaamista ei eri syistä saada esiin. Pa-laamme sanamääriin T-yksikköjen yhteydessä.

Sukupuolen suuri rooli on suomen aineistossa hämmästyttävä. Esimerkiksi Bermanin ja Verhoevenin (2002, 23–24) tutkimuksessa ei löydetty sukupuolen mainittavaa vaikutusta sanamääriin. Tutkimuksessa verrattiin kahta eri tekstityyppiä (kertova ja erittelevä) ja modaliteettia (puhuttu ja kirjoitettu) edustavien tekstien avulla seitsemässä eri kielessä

Taulukko 53. Unelmakirjoitelmien sanamäärät kuudennella, yhdeksännellä ja nuorilla aikuisilla.

Ikäluokka / Sanamääräluokat %	1–50 sanaa	51–100	101–150	151–200	201–250	251–300	Yli 300
Kuudes F	0,6	18,8	31,3	23,1	13,1	6,3	6,9
Kuudes M	24,2	32	24,2	11,1	4,6	2,6	1,3
Yhdeksäs F	4,5	13,6	23,9	20,5	19,3	7,9	10,2
Yhdeksäs M	15,3	30,5	23,6	12,5	6,9	8,3	2,8
Aikuiset F	0,4	0,4	1,7	10,4	16,2	51,7	19,2
Aikuiset M	0	0	1	9,3	17,5	49,5	22,7

(englanti, hollanti, ranska, espanja, ruotsi, islanti ja heprea) neljän eri ikäluokan kirjoittamista. Ikäluokat olivat vastaavantapaiset kuin omassa aineistossamme. Ensimmäinen ryhmä olivat 9–10-vuotiaat, toinen 12–13-vuotiaat, 16–17-vuotiaat ja nuoret korkeakouluopiskelijat (20–30-vuotiaat). Saksalaisia yhdeksäsluokkalaista koskevassa DESI-tutkimuksessa tytöt olivat poikia parempia kaikissa mitatuissa osa-alueissa mutta suurimpia erot olivat tuottavissa osioissa kuten kirjoittamisessa (DESI-Konsortium 2008, 202–203); sanamääriä ei tutkimuksessa kommentoida. Myös Yhdysvaltain kansallisessa arvioinnissa (8. ja 12.-luokkalaisten) kirjoittamistaideoissa on sukupuolieroja tyttöjen hyväksi (ks. NAEP 2011; Writing 2011). Erot eivät kuitenkaan ole kovin suuret, sillä esimerkiksi 8.-luokkalaisten pistekeskisarvo on tytöillä 160 (maksimi 300) ja pojilla 140. Arviointikriteereistä ja näytteistä voi päätellä, että kokonaissanamäärä on parhaissa teksteissä korkeampi kuin heikommassa, mutta sanamäärä ei ole suoraan arviointikriteeri.<sup>116</sup> Wolf Nelsonin ja Van Meterin (2007, 301) tutkimuksessa sen sijaan sanamäärissä, T-yksiköiden ja konjunktioiden määrissä on oletuksen mukainen merkitsevä päävaikutus sukupuolen mukaan mutta eroa ei ole lausetyypeissä eikä oikeinkirjoituksessa.

<sup>116</sup> Arviointi perustuu kolmeen mittariin, jotka ovat ideoiden kehittäminen, ideoiden organisointi ja kieleen liittyvät resurssit ja konventiot. Jokaista arvioidaan kuusiportaisella asteikolla. Kirjoittajista puolet kuuluu ns. perustasolle (pistekeskisarvo 155), jota luonnehditaan seuraavasti: " – used some details but details did not always convey an experiment, while organization was somewhat loose, word choice clear rather than precise, and sentence structure relatively unvaried." Pistekeskisarvojen perusteella sekä tytöt että pojat kuuluvat keskimäärin tälle perustasolle. Noin 20 prosenttia kirjoittajista on perustason alapuolella, 30 prosenttia ylittää perustason.

PISA-tutkimuksen yhteydessä on tullut esiin, että Suomessa tulokset sukupuolittuvat voimakkaammin kuin muissa PISA-maissa (ks. PISA 2012; 2015; 2018). Opettajakommenttien mukaan suunta on alaspäin. Kyse saattaa olla myös länsimaisesta trendistä vaikkakin Suomessa erityisen voimakkaasta.<sup>117</sup> Toisaalta on tutkimustuloksia, jotka osoittavat, että vaikka tyttöjen kirjoitelmissa on enemmän sanoja kuin poikien, kirjoitelmien laatu ei silti eroa (ks. Babayiğit 2015, 34). Kokonaissanamäärä yksinään on ongelmallinen mittari mutta osoittaa kyllä ikäluokan keskimääräistä kehitystä etenkin nuoremmilla kirjoittajilla.

## Sanaluokkaindeksit

Sanaluokkaindeksejä on useita, ja niiden avulla voidaan laskea tekstityyppi- tai lajikohtaisia tekstin tyyppi-arvoja, laskea tekstin informatiivisuutta ja tutkia jonkin tarkkaan valitun ilmiön esiintyvyyttä (ks. esimerkiksi Berman, Nayditz & Ravid 2011; Price & Jackson 2015, 283; Hall-Mills & Apel 2015). Käsittelemme tässä yhteydessä substantiivi-verbi-suhdetta (= NTVR eli *noun-to-verb ratio*) ja leksikaalista tiheyttä (*lexical density*), joilla kehitystä voidaan arvioida ja jotka helposti osoittavat myös tekstien tyyppieroja. NTVR-arvo ilmaisee kompleksisuutta asteikolla yksinkertainen/puhekielinen vs. abstrakti, toisin sanoen korkea substantiiviosuus viittaa tekstin korkeaan abstraktiotasoon. Puhekielisessä tekstissä on paitsi lyhyet lauseet suhteessa paljon finiittiverbejä, abstraktissa tekstissä on suhteessa paljon substantiiveja. Tämä yleistys pätee erityisesti yhdistettynä korkeaan leksikaalisen tiheyteen sekä lauseiden korkeisiin sanamääriin. Lähinnä omaelämäkerrallisissa kertovissa teksteissä substantiivien ja verbien suhde on todennäköisesti lähellä yhden suhdetta yhteen.

Leksikaalinen tiheys ilmaisee leksikaalisten sanaluokkaesiintymien osuutta kokonaissanamäärästä. Leksikaalisia sanaluokkia ovat substantiivit, verbit ja adjektiivit sekä adverbeista ne, joiden kantasana ei ole prono-

117 Esimerkiksi Scandinavian Association of Cognitive Linguistics -konferenssissä vuonna 2017 keskusteltiin poikien heikoista kirjoittamistuloksista länsimaisena trendinä.

minaalinen. Leksikaalisten sanaluokkien vastakohta on kieliopilliset sanaluokat, kuten konjunktiot ja pronominit, joiden merkitys on tyypillisesti abstrakti ja joiden tehtävät ovat lähinnä syntaktisia. Leksikaalinen tiheys kuvaa tekstin informaatiopitoisuutta, sitä, kuinka paljon tarkoitteita ja tapahtumia tekstiin tuodaan ja kuinka paljon niitä luonnehditaan ja kehystetään aikaan ja paikkaan. Teemme myös sanaluokkakohtaisempia laskelmia kompleksisuuden erittelemiseksi tämän teoksen loppuluvussa.

Tekstiin liittyvän sanaluokkaindeksin arvo ei voi olla mikä tahansa: Jos NTVR-suhde (N/N+V) on hyvin korkea tai hyvin matala, kyse on substantiivi- tai verbiluettelosta ei kertovasta tai kuvailevasta tekstistä. Jos leksikaalinen tiheys (N+V+A+Adv/kokonaissanamäärä) on hyvin korkea, tekstistä puuttuu sidoksisuus ja se lähinnä luettelee tarkoitteita. Siten indeksien arvojen suuren vaihteluvälin voi tulkita osoittavan, että ikäluokan kaikki kirjoittajat eivät vielä kirjoita onnistunutta tekstiä. Indeksit ovat myös mitä ilmeisemmin tekstityyppi- ja lajisidonnaisia. Berman ja muut (2011) osoittavat, että erittelevän tekstin leksikaalinen tiheys on korkeampi kuin kertovan. Biberin ja muiden (1998) mukaan NTVR-suhde on englanninkielisessä tieteellisessä tekstissä 2,2:1 (eli substantiiveja on kaksinkertainen määrä verbeihin verrattuna), mutta fiktiossa ja puheessa lähempänä yhden suhdetta yhteen. Indeksien arvot ovat kielikohtaisia – tämä pätee erityisesti leksikaaliseen indeksiin – ja toisaalta niihin vaikuttaa myös valittu laskentaperiaate.

Unelmakirjoitelmat on sanaluokitettu tietokantaan pääsanaluokkiin ja alaluokkiin. Indeksit on laskettu pääsanaluokista muutamien poikkeuksin. Kehityksellisesti erottelevien pronominikantaisten adverbien suuri joukko ei ole laskelmissa mukana toisin kuin pronominaalidjektiivien pieni joukko. Propreja ei ole erotettu appellatiiveista. Kiteytymät eivät ole mukana NTVR-arvossa mutta kylläkin leksikaalisessa indeksissä. Verbeihin on luettu kaikki verbiparadigmassa taipuvat paitsi leksikaalistuneet partisiiipit. Toisin sanoen substantiivien ja verbien suhde on pyritty laskemaan juuri näissä funktioissa käytetyistä sanoista ja leksikaaliseen tiheyteen vaikuttaa kaikki ns. leksikaalinen aines. Indeksien arvoihin vaikuttaa kuitenkin myös se, erotellaanko yhden ja saman sanan leksikaalinen ja kieliopillinen käyttö vai ei (vrt. O’Loughlin 1995; Biber ja muut 1998); näin ei ole tehty omassa aineistossamme.

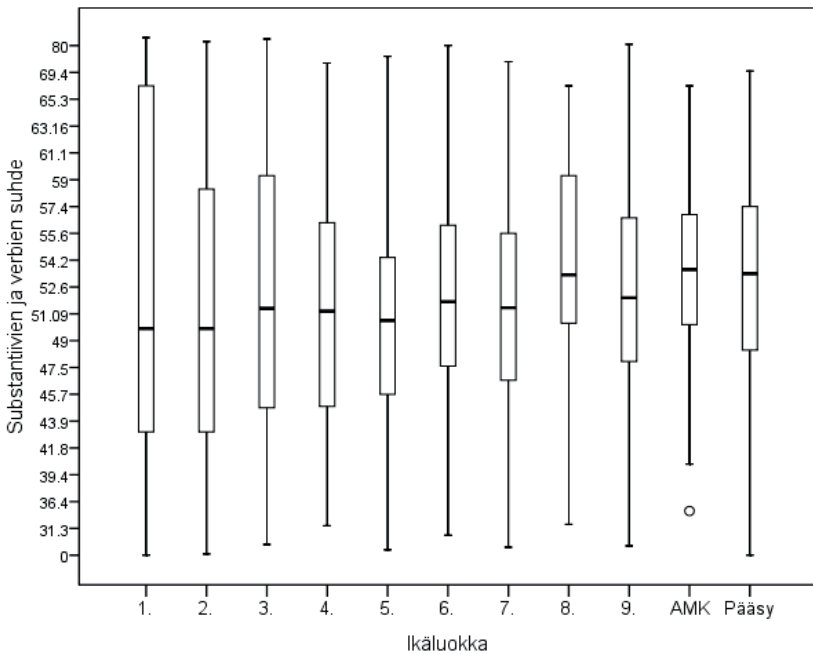


Pyrimme tässä yhteydessä osoittamaan jonkinlaisen luontevan vaihteluvälin kirjoitelma-aineiston avulla kyseisille indekseille kehityksellisen näkökulman lisäksi. Vertailulukuja ei ole valmiina isoista korpuksista mutta Savon Sanomien korpuksen miljoonan sanan otoksesta laskimme samoilla kriteereillä vastaavat arvot. Indeksit on laskettu 1717 kirjoitelmasta, joissa on yhteensä 218 849 tekstisanaa. Kirjoittajista 939 on naispuolisia, 778 miespuolisia.

#### SUBSTANTIIVIEN JA VERBIEN KESKINÄINEN SUHDE

Substantiivien ja verbien keskinäinen suhde eli NTVR näyttää Unelma-kirjoitelmissa oletuksen mukaan olevan – keskimäärin – melko lähellä yhden suhdetta yhteen erityisesti alakoululaisilla (ks. kaavio 16). Tämä tarkoittaa esimerkiksi sitä, että alimmilla luokilla kirjoitelmat ovat rakenteltaan yksinkertaisia, lähellä puhutun kielen mallia, lauseet ovat melko lyhyitä, subjektina toimii pronomini (ks. tarkemmin Pajunen 2012) ja kompleksiset rakenteet ovat vähissä. Substantiivien osuus alkaa nousta kolmannelta luokalta yläkoulun alkuun. Pieni notkahdus seitsemännen ja kahdeksannen luokan aineistoissa liittyy melko pieneen kirjoittajamäärään. Kuudennen ja yhdeksännen välillä ei ole mainittavaa eroa. Nuorilla aikuisilla NTVR on 1,1:1 (eli mediaani on 55). Kolme neljäsosaa kirjoitelmista ylittää yhden suhteen yhteen, keskeiset NTVR-mediaanit sijoittuvat lapsilla ja nuorilla välille 50–53, aikuiskirjoitelmissa välille 50–57, toisin sanoen substantiiveja on jonkin verran enemmän kuin verbejä. Tulos – substantiivien suhteellinen vähyys – on yllättävä, jos ajattelee, että sekä verbin argumentit että monet lokalisoivat ja kehystävät ainekset ilmaistaan suomessa substantiiveilla. Yksittäisessä lauseessa olettaisi aina olevan useita substantiiveja. Sanomalehtikielessä NTVR onkin huomattavasti korkeampi eli 65,9, ero viittaa selvästi tekstityyppi- ja lajieroon.

Tulosta selittävät monet kertovan kielen piirteet: minä-kerronta laskee substantiivien määrää ja vastaavasti erityisesti päivä-skeemassa kerronta etenee siirtymiä ilmaisevin yhden argumentin intransitiiviverbein (ks. Vesala 2012). Matalaa NTVR-arvoa selittää muun muassa runsas konjugoivien verbien ja pronominiin käyttö (vrt. esimerkki 6), vastaavasti korkeata NTVR-arvoa tuntuu selittävän luettelomaisuus ja ehkä ongelmat lauseiden muodostamisessa (vrt. 7).



Kaavio 16. Substantiivien ja verbien keskinäinen suhde (= NTVR) Unelmakirjoitelmissa (n = 1774).

NTVR on 35,3:

6) *Aamulla kun heräsin äiti oli kattanut aamupalaksi minun lempi murojani, joita en yleensä saa aamupalalla syödä. Kun olin syönyt, äiti kysyi: Menisimmekö serkkujen luokse? Minä tietenkin innokkaana sanoin: mennään!* (5F60)

NTVR on 64,3:

7) *Unelmieni päivä olisi jos voittaisin lotossa. Saisin paljon rahaa ja ostaisin kaikkea. Esim. Auton, kartanon, uima-altaan kartanoon, palkkaisin hovimestarin, kokin ja siivoijan. Ostaisin vielä huvijahdin ja karkkitechtaan.* (5M138)

NTVR-arvojen vaihteluväli on koko aineistossa suuri mutta mediaani asettuu välille 50–53,6, keskihajonta on 5,8. Kaikkein nuorimmilla kirjoittajilla hajonta on suurinta, ensimmäisellä luokalla mediaani on 50 mutta keskihajonta jopa 48,9. Keskeisten havaintojen osalta vaihteluväli kapenee alakoulun puolivälistä alkaen nopeasti ja osoittanee koherentin tekstin tuottamistaitojen syntyä ja kehitystä. Myös mediaanin nouseva suunta osoittaa kehityksellistä muutosta.

Kun NTVR-arvoa tarkastellaan uudelleenluokittelemalla havainnot ryhmiin alle 40–yli 65, normiarvot ovat 1–1,1:1 kaikilla muilla paitsi aikuisilla miehillä, joilla arvot ovat hieman korkeampia (ks. taulukko 54). Keskeiset arvot sijoittuvat välille 46–60 ja muut arvot ovat enemmänkin poikkeavia. Lisäksi substantiivien suhteellinen osuus kasvaa iän myötä. Mieskirjoittajilla vaihtelu on suurempi kuin naiskirjoittajilla mutta nousee iän mukaan nuoremmilla tavallisimmasta mediaaniluokasta 51–55 luokkaan 56–60. Kuudennen tytöt poikkeavat yleiskuvasta yhtä selvästi kuin aikuiset miehet: mediaanit vaihtelevat luokassa 46–55, mikä viittaa melko yksinkertaiseen tyyliin. Kuudennen ja yhdeksännen pojilla on paljon ääriarvoja eli osoituksia tekstinmuodostamisongelmista.

Taulukko 54. NTVR-suhde koehenkilöittäin Unelmakirjoitelmissa.

Ikäluokka / NTVR-luokat %	Alle 40	40–45	46–50	51–55	56–60	61–65	Yli 65
Kuudes F	2,5	7,5	30,0	32,5	16,9	7,5	3,1
Kuudes M	8,5	13,1	13,1	31,4	16,3	9,1	8,5
Yhdeksäs F	3,4	10,2	19,3	39,8	14,8	11,3	1,1
Yhdeksäs M	6,9	6,9	20,8	29,2	18,0	13,9	4,2
Aikuiset F	1,7	5,4	20,8	34,1	26,7	10,4	0,8
Aikuiset M	2,1	6,2	19,6	23,7	35,0	11,3	2,0

Kertovan tekstin ominaisuus näyttää olevan melko matala NTVR Biberin ja muiden (1998) ennusteen mukaan, lisäksi se nousee ikäluokittain hitaasti. Kun kertovassa tekstissä on substantiiveja hieman enemmän kuin verbejä ja kun kirjoittajien väliset ikäluokkakohtaiset erot ovat pienehköt, kerronta on melko tasapainoista. NTVR-arvoa olisi hyvä voida verrata myös muihin tekstilajeihin, mutta suomen kielestä vertailuaines Savon Sanomien aineistoa lukuun ottamatta puuttuu. NTVR-arvo olisi hyvä laskea myös sanojen funktion mukaan analysoiden.

## LESIKAALINEN TIHEYS

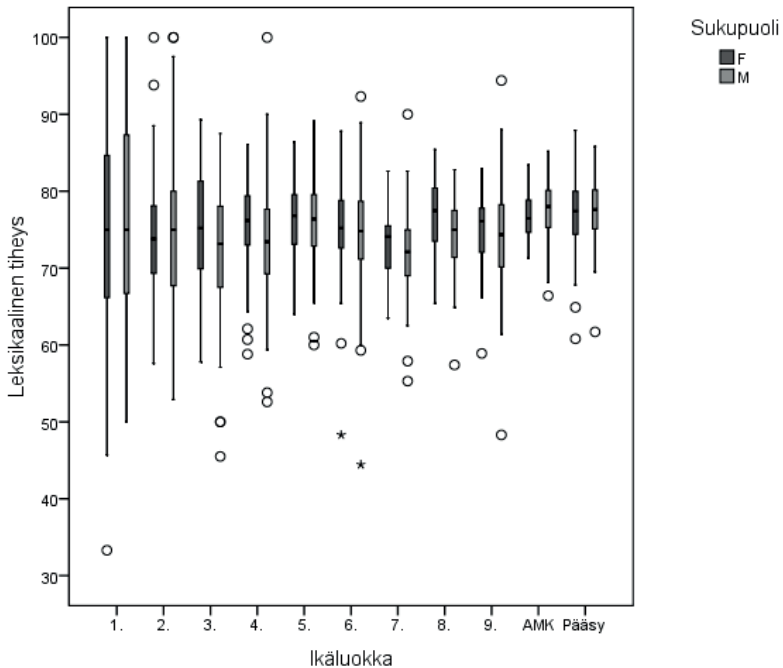
Leksikaalinen tiheys kuuluu leksikaalisen rikkauden mittareihin diversiteetin ja yleisyystasoaanalyysin lisäksi (ks. esimerkiksi Read 2000; Johansson 2008; Jarvis 2013; Honko ja muut 2019; Pajunen & Honko tässä teoksessa). Leksikaalinen tiheys mittaa niin sanottujen sisältösanojen eli kielenulkoisiin tarkoitteisiin ja tapahtumiin viittaavien sanojen määrää tekstissä. Se osoittaa, miten informaatiopitoista teksti on, mutta luonnollisesti tiheys jälleen vaihtelee eri tekstilajeissa ja kirjoittamistyyliessä. Lisäksi informatiivisuuteen vaikuttaa diversiteetin (eli eri leksiemien määrän) lisäksi leksikaalinen toisteisuus. Kaikista näistä indekseistä voi tehdä päätelmiä tekstin sisällön luonteesta, sen välittämän tiedon määrästä. Yleisesti tiedetään, että sisältö monipuolistuu kehityksellisesti (vrt. Heinonen 2018), mutta on syytä muistaa, että korkea tiheys ei sinänsä tarkoita hyvää tekstiä (vrt. Jarvis 2013; Honko ja muut 2019). Samoja indeksejä käytetään myös luettavuuden arviointiin.

Tiheys voidaan laskea vähintään kolmella eri tarkkuudella (Johansson 2008). Ensimmäinen ja tässä noudatettu on, että sisältösanojen esiintymät lasketaan funktioita huomioimatta, esimerkiksi *voida*-verbiä ei alaluokitella leksikaaliseen ja kieliopilliseen. Toinen vaihtoehto olisi analysoida referentiaalisesti käytetyt leksikaaliset ainekset erikseen ja kolmas olisi jättää muun muassa *olla*-verbi kokonaan huomiotta. Periaatteessa leksikaaliset sanaluokat eli substantiivit, adjektiivit ja verbit sekä myös adverbit ovat juuri sisältösanoja, konjunktiot, artikkelit ja niin edelleen kieliopillisia sanaluokkia. Sisältösanat voidaan kielissä määritellä monilla kriteereillä, ja vaikka etenkin substantiivit ja verbit ovat universaaleja, leksikaalista tiheyttä on vaikea laskea eri kielissä aivan samoin periaattein. Lisäksi tutkijoilla on omia valintojaan. Esimerkiksi O’Loughlin (1995, 228) ei lue leksikaalisiin sanoihin modaaliverbejä, *be*- ja *have*-verbejä eikä pronominaaleja.

Suomen kielessä modaaliverbien luokka ei ole täysin vakiintunut mutta voi kysyä, eikö modaaliverbitkin pitäisi erotella aineistosta, jos kerran modaaliverbit erotetaan. Modaaliverbien määrä verbiesiintymistä on aineiston aikuiskirjoittajilla hieman peruskoululaisia suurempi mutta niiden esiintymät kaikista verbiesiintymistä ovat tiheyslukujen kannalta pienet. Keskinen (2012) analysoi Unelmakirjoitelmien modaa-

listen ainesten käyttöä alakoululaisilla ja osoittaa muun muassa, että paitsi määrien hienoista kasvua muutosta tapahtuu myös funktioissa siten, että nuorimmat käyttävät modaaliverbejä useammin deonttises- sa (eli mitä saa tehdä) ja vanhemmat dynaamisessa funktiossa (mitä kykenee tekemään). Givónin (2009, 123) mukaan tämä on lasten kie- len kehityksessä yleinen trendi. *Olla*-verbi kopulana on eri asemassa kuin *olla*-verbi liittotempusten tai kiellon yhteydessä. Alakoululaisten aineistossa liittotempusten käyttö on vähäistä (15 % finiittisistä) mutta aikuisilla vajaat 30 prosenttia *olla*-esiintymistä kuuluu liittotempuksiin; toisaalta *olla*-verbi on muuten alakoululaisilla yleisempi kuin aikuisilla. Onkin todennäköistä, että leksikaalisten sanojen kieliopillisen käytön huomioiminen ei suuresti vaikuttaisi nykyjakaumaan. Substantiiveista ei erotella propreja, vaikka se voisi olla järkevää, koska ne eivät ole sisältösanoja vaan nimiä. Proprien osuus leksikaalisista sanoista laskee alakoululaisten viidestä prosentista nuorten aikuisten 1,7 prosenttiin. Adverbeista erotetaan pronominaaliadverbit, joiden määrä laskee ke- hityksellisesti (ks. Koivunen 2012), mutta pronominaaleja ei eroteta adjektiiveista, koska niiden määrä on olematon. Joka tapauksessa leksi- kaalinen tiheys vaihtelee arvioitavan kielen typologisten ominaisuuksien perusteella; suomessa tiheys on välttämättä aika korkea ja huomattavasti korkeampi kuin lähes analyttisessä englannissa. Tarkempi laskutapa saattaisi hieman laskea nuorimpien leksikaalista tiheyttä ja nostaa van- hempien. Lisäksi olisi hyödyllistä erotella leksikaalisten sanaluokkien edustajat konkreettisuusasteikon mukaisesti (vrt. Ravid 2006; Berman & Nir 2010); tähän jaotteluun tukeudutaan toisessa yhteydessä (ks. Paju- nen ja Honko, tämä teos).

Unelmakirjoitelmat ovat samasta aiheesta ja ohjaavat omaelämäker- tomiseen – kuten on jo todettu. Pääosin ne edustavat kertovaa tekstiä, jossa on kirjoittajakohtaisesti ja iän mukaan lisääntyvästi kuvailua ja erittelyä. Tekstien keskinäisessä vertailussa leksikaalinen tiheys on in- formatiivinen indeksi ja todennäköisesti kuvaa yleisemminkin synteet- tisen suomen kielen kertovan tekstin ominaisuuksia. Tiheys sijoittuu peruskoululaisilla keskimäärin noin 75 prosenttiin eli yleensä noin 75 prosenttia kertovien kirjoitelmien tekstisanoista on leksikaalisia (ks. kaavio 17). Aikuisilla mediaani on 77 prosenttia ja keskeiset havainnot



Kaavio 17. Leksikaalinen tiheys ja sukupuoli.

asettavat välille 75–80 prosenttia (vrt. esimerkit 8 ja 9); erot selittynevät todennäköisimmin kirjoittajien eri kehitysvaiheista ja kertovan ja toisaalta kertovan ja erittelevän tekstin eroista. Keskeisten havaintojen osalta peruskoulun alaluokilla on enemmän hajontaa kuin ylempillä. Vaikka tekstin leksikaalinen tiheys olisi korkea (kuten esimerkissä 9), NTVR-arvo voi olla matala (esimerkissä 9 51,6). Arvioimme mittareiden ristikkäisyyttä seuraavassa luvussa.

Leksikaalinen tiheys on 73,7 (kolmannen luokan mediaani):

8) *Unelmieni päivä oli päivä joka oli kesälomalla. Unelmieni päivä oli kun isi ja äiti veivät minut ja pikku siskoni vaassalandiaan. Olimme vaasa Landiassa kaksi yötä. Meidän huone oli neljännessä kerroksessa. Aina illalla ja aamulla menimme uimaan.* (3F37)

Leksikaalinen tiheys on 79,2 (nuorten aikuisten mediaanin yläpuolella):

9) *Oli kaunis syyskuun päivä New Yorkissa. Ilma oli selkeä, ja tulevan syksyn saattoi aistia kipakasta tuulesta. Kävelin Brooklynin katuja take away -kahvi kädessäni höyryten. Olin tullut New Yorkiin muutama viikko aiemmin.* (PääsyF76)

Leksikaalisen tiheyden mediaani on peruskoulussa tytöillä hieman korkeammalla kuin pojilla, aivan nuorimpia lukuun ottamatta. Molemmilla tiheyden vaihteluväli etenkin keskeisten havaintojen osalta iän mukaan pienenee. Ääriarvoja on kaikissa ryhmissä. Aikuis kirjoittajilla molemmissa kirjoittajaryhmissä tiheys on miehillä korkeampi kuin naisilla.

Leksikaalinen tiheys keskittyy yleisyysluokkiin 70–80 kaikissa ikäryhmissä mutta nousee iän mukaan (taulukko 55). Yhdeksäsluokkalaisten tyttöjen jakauma jo lähenee aikuisten vastaavaa. Peruskoulussa sekä kuudennen että yhdeksännen pojilla jakauma näyttää aika samanlaiselta ja kummassakin ryhmässä on paljon ääriarvoja, mikä jälleen viittaa ongelmiin tekstinmuodostuksessa. Vain kolmannes peruskoulun pojista sijoittuu luokkaan 75–80 prosenttia. Luokassa alle 65 prosenttia arvot ovat selvästi poikkeavia kaikilla kirjoittajilla. Ylimmässä luokassa on melko paljon kaikkien ikäryhmien kirjoittajia mutta kuudennen pojilla korkea leksikaalinen tiheys viittaa luetteloon, aikuisilla erittelyyn. Sanomalehtikielessä leksikaalinen tiheys on miljoonan sanan otoksesta laskettuna hieman korkeampi kuin kertovassa tekstissä eli 81,5. Kun lehti kielen indeksi vertailun vuoksi lasketaan 100 000 tekstisanan otoksista, käy ilmi, että luku pysyttelee aika vakiona.

Taulukko 55. Leksikaalinen tiheys koehenkilöittäin Unelmakirjoitelmissa.

Ikäluokka / Leksikaalisen tiheyden luokat %	Alle 65	65–69	70–74	75–79	80 tai yli
Kuudes F	1,2	10	35,6	35,6	17,5
Kuudes M	3,9	16,4	30,3	29,6	19,7
Yhdeksäs F	1,1	10,2	28,4	48,8	11,4
Yhdeksäs M	6,9	16,7	33,3	26,3	16,7
Aikuiset F	0,8	3,7	22,5	48,3	24,6
Aikuiset M	1,0	4,1	16,5	49,5	28,8

## SANALUOKKAINDEKSIT JA HOLISTINEN ARVIO

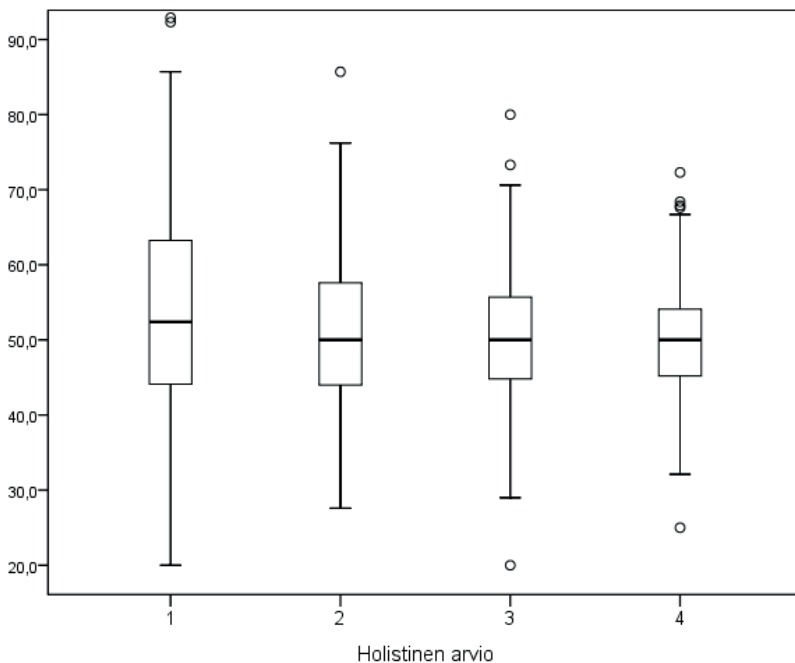
Käsitys sekä NTVR-arvon että leksikaalisen tiheyden vakiintumisesta tekstin onnistuneisuuden myötä vahvistuu, kun alakoululaisten aineis-  
tosta tehty tasoluokitus (ks. taulukko 56) ristiintaulukoidaan sanaluokka-  
indeksien kanssa (kaaviot 18–21). Tasoja on nuorimmilla (luokat 1.–5.)  
neljä (ks. Pajunen 2012): luettelo tai lista (kaavioissa numero 1), epäloo-  
ginen tai muuten epäonnistunut, kesken tai muuta sellaista (numero  
2), juonirunko eli tapahtumia luetteleva skripti (numero 3) ja kertomus-  
skeemaa noudattava (numero 4). Bermanin (2007, 348) ja Hickmannin  
(2003) mukaan noin kymmenvuotiaat lapset ovat jo sisäistäneet kerto-  
musrakenteen mallin eli skeeman ja hallitsevat myös viittaussuhteet  
kertomuksen osallistujiin sekä aikaan ja paikkaan. Todettakoon lyhyesti,  
että jopa kuudesluokkalaiset 2000-luvun suomalaislapset jäävät kauaksi  
näistä taidoista. Kuudensista luokista on tehty myös hieman tarkempi  
holistinen luokitus, joka erottelee tasoluokan 2 toisaalta epäloogiseksi  
(2a), toisaalta otsikon vierestä kirjoitetuksi (2b). Lisäksi luokan 3 skriptit  
on jaettu kahtia hyvin pelkistettyihin (3a) ja hieman kertovampiin (3b).  
Skriptien joukko kasvaa selvimmin kehityksellisesti, mutta kertomus-  
skeemaa noudattavia kirjoitelmia on odotusarvoon nähden alakoulun  
ylimpienkin tytöillä melko vähän pojista puhumattakaan. Vuoden 2006  
kuudennen luokan tytöt ovat kirjoittaneet enemmän kertomuksia kuin  
vuoden 2014 tytöt, joille skripti on tyypillisin. Osa tutkijoista pitää kerto-  
musrakenteen tapaisten holististen mittarien käyttöä arveluttavana, kos-  
ka arviointi on vaikeaa, skeemat varioivat kulttuurisesti ja niin edelleen.  
Etusijalle asettuisi koherenssia luovien keinojen käytön vertailu (Burns  
ja muut 2012, 132). Toisaalta kuitenkin koherenssi ja kertomusrakenne  
suhteutuvat toisiinsa (vrt. Kouvo ja muut, tämä teos).

Kertomusskeemaa noudattavissa kirjoitelmissa sekä NTVR:n että  
leksikaalisen tiheyden keskeiset arvot (25–75 %) vaihtelevat mediaanin  
ympäriällä vähän ja vastaavat aikuisarvoja (ks. kaaviot 18–21). Muissa ho-  
listisen arvioinnin tyypeissä vaihtelua on paljon enemmän, erityisesti  
poikien teksteissä on ääriarvoja. Tarkastelu osoittaa, että kertovalla teks-  
tillä todella on tyypilliset leksikaalisten indeksien arvot. Jopa hyvissä  
skripteissä arvot vaihtelevat enemmän kuin kertomusskeemaa noudat-  
tavissa.

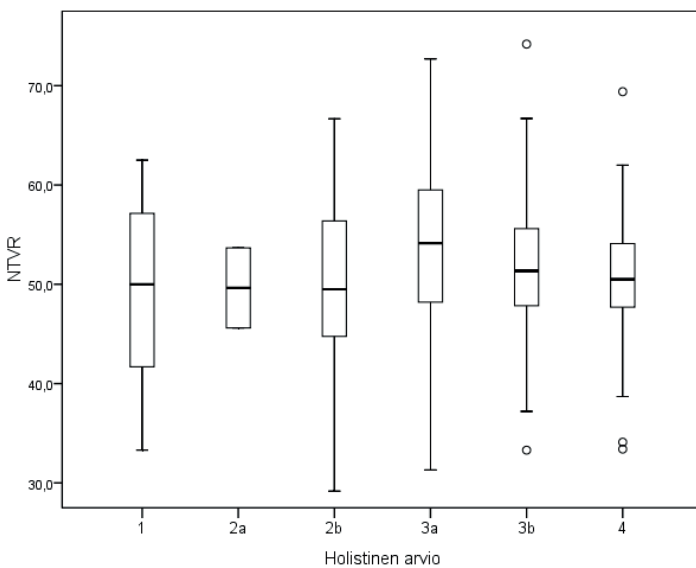


Taulukko 56. Alakoululaisten Unelmakirjoitelmien holistinen arvio.

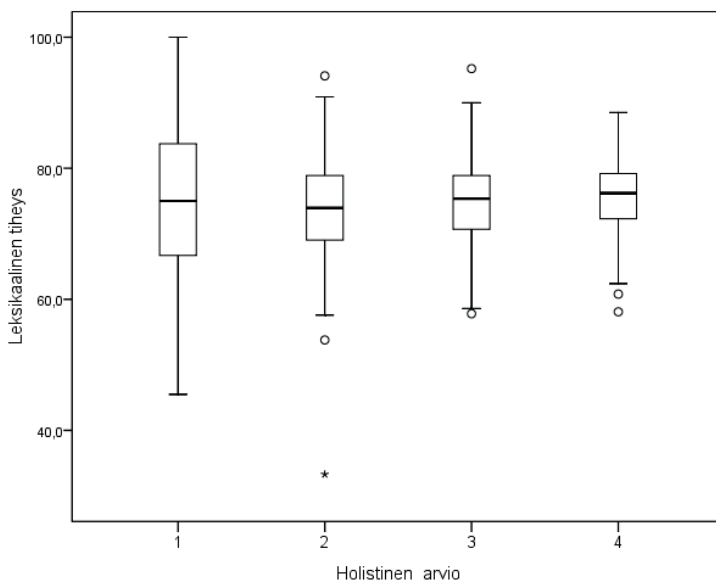
Ikäluokka / Tasoluokka %	Luettelo/ Lista	Epälooginen/ Kesken	Skripti	Kertomus
Ensimmäinen F	38,6	28	17,5	15,8
Ensimmäinen M	62,5	17,8	19,6	0,0
Toinen F	18,7	16,5	38,5	26,4
Toinen M	50,0	19,0	25,0	5,9
Kolmas F	11,6	15,9	43,5	29,0
Kolmas M	28,2	10,6	43,5	17,6
Neljäs F	16,8	2,0	44,2	37,0
Neljäs M	31,3	19,3	42,0	7,2
Viides F	2,7	2,7	44,6	50,0
Viides M	16,4	22,8	31,6	29,1
Kuudes F	0	5,6	59,4	35,0
Kuudes M	5,9	11,2	71,8	11,2



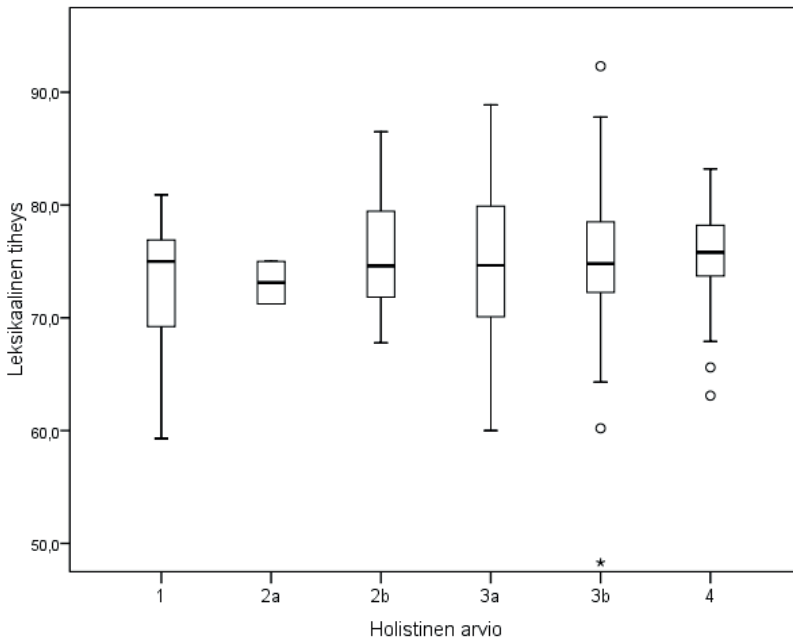
Kaavio 18. Alakoululaisten 1.–5. kirjoitelmien NTVR holistisin tasoluokittain.



Kaavio 19. Kuudesluokkalaisten kirjoitelmien NTVR ja tarkempi holistinen arvio.



Kaavio 20. Alakoulun luokkien 1.–5. kirjoitelmien leksikaalinen tiheys tasoluokittain.



Kaavio 21. Kuudesluokkalaisten kirjoitelmien leksikaalinen tiheys tasoluokittain.

## T-yksiköt ja lauserakenteen kompleksisuus

T-yksikköön eli terminaaliyksikköön tiivistetään itsenäisten propositioiden merkitysisältö suhteessa pääpropositioon. Se kertoo siten koehenkilön näkemyksen tapahtumien keskinäisistä ajallisista tai esimerkiksi kausaalisista suhteista. Jotta analysoija ei pyri tulkitsemaan kirjoittajan päämääriä, T-yksikön koostamisperiaatteet ovat tiukat. T-yksikkö rakentuu aina kirjoittajan päälauseeksi merkitsemään lauseeseen ja ottaa jäsenikseen sille muodollisesti alisteiseen suhteeseen merkityt lauseet. T-yksikkö vastaa jossain määrin virkkeen käsitettä mutta voi olla lausetta laajempi ja toisaalta virkettä suppeampi yksikkö. Ero virkkeisiin näkyy etenkin rinnasteisten päälauseiden kohdalla: Lauseet *Pekka istui ja Matti seisoi* kuuluvat samaan virkkeeseen mutta eri T-yksiköihin (ks. Schneider

ja muut 2005). Sen sijaan lauseet *Pekka istui Matin seistessä* tai lauseet *Pekka istui ja ajatteli kohta nukahtavansa* kuuluvat paitsi samaan virkkeyseen myös samaan T-yksikköön.<sup>118</sup>

T-yksikköjen avulla saadaan kaksi kehityksellistä mittaria. Ensinnä sen avulla lasketaan kehityksellisessä tutkimuksessa keskeisin vertailuyksikkö eli T-yksikön keskipituus sanoina (= MLTU) ja toiseksi T-yksilöiden avulla mitataan lauserakenteen kompleksisuutta niin sanotulla kompleksisuusindeksillä (= CI). Tämä indeksi ilmaisee, montako proposition palautuvaa lausetta tai ei-finiittistä rakennetta on keskimäärin tiivistetty samaan T-yksikköön eli kuinka tiivistä ja informaatiokulun kannalta painotettua tekstiä kirjoittaja tuottaa. T-yksikön keskipituus ja kompleksisuusindeksi ilmaisevat siis molemmat kompleksisuutta. Matala arvo viittaa kummassakin yksinkertaiseen ja vastaavasti korkea kompleksiseen rakenteeseen, mutta keskipituus ilmaisee enemmän kielenyksikköjen määrää, kompleksisuusindeksi niiden hierarkkista suhdetta.

T-yksiköt määritellään siis juuri sen mukaan, miten alisteisuus on merkitty tekstiin ilmimorfeemeilla. Esimerkiksi rinnasteiset relatiivilauseet otetaan huomioon saman T-yksikön jäsenenä vain silloin, kun molemmat on merkitty relatiivisanalla alisteisiksi (*mies joka x ja joka y*). Unelmakirjoitelmissa on suhteellisen vähän tällaisia rakenteita, koska mahdollisista rinnasteisista sivulauseista jälkimmäisen muodollinen alistuksen merkki usein puuttuu, vaikka semanttisesti yksikköjen suhde olisi ymmärrettävä alisteiseksi. Esimerkki 10 on tässä suhteessa edustava: ajatuksen löytyminen edellyttää, että loppuosa luetaan muodossa: *tai jotka itse kiusaavat*.

10) (T-1) *Haluan puhua lapsille ja nuorille, joita koulukiusaataan* (T-2)  
*tai he itse kiusaavat.* (9F16)

On tapana sanoa, että T-yksiköillä mitataan syntaktista tiivyyttä (*syntactic/clausal density*, ks. Nippold ja muut 2014), kirjoittajan kykyä tiivistää

<sup>118</sup> Hankkeessa on erikseen tutkittu myös virkkeen koostumusta ja lauseiden sanamääriä. Molemmissa muutokset ovat kehityksellisiä. Ks. alakoululaisista Elomäki 2011; yläkoululaisista Oinonen 2008.

sanomaansa ja suhteuttaa vähemmän tärkeä tärkeämpään. T-yksiköillä voidaan myös arvioida kirjoittajan tapaa luoda tekstiinsä koherenssia.

Kirjallisuudessa T-yksikköjen määrittely vaihtelee jonkin verran (vrt. esimerkiksi Scott & Windsor 2000; Berman & Verhoeven 2002; Nippold ja muut 2005; 2007; Price & Jackson, Schneider ja muut 2005). Toisinaan otetaan huomioon alisteisina rakenteina vain sivulauseet eli niin sanotut nominaaliset (argumenttiasemassa olevat komplementtilauseet), relatiiviset ja adverbiaaliset sivulauseet (ks. esimerkiksi Nippold ja muut 2005). Toisinaan otetaan huomioon sivulauseiden lisäksi kaikki propositiota kielentävä ja alisteiseksi tulkittava infiniittinen aines (Schneider ja muut 2005). Kolmas vaihtelu liittyy alisteisen rakenteen position huomioimiseen. Esimerkiksi sivulauseet voidaan jakaa vasemmalle tai oikealle haaroviin; vasemmalle haarovuus on oletuksen mukaan kompleksisempaa kuin oikealle, ainakin se on prosessoinnin ja muistin kannalta vaativampaa. T-yksiköt ovat yleensä käytössä kirjoitettua kieltä analysoitaessa ja pienten lasten puhetta analysoidaan niin sanottujen C-yksikköjen avulla, joilla voi ottaa epätäydelliset rakenteet paremmin huomioon (ks. Nieminen 2007). Noudatamme tässä yhteydessä Schneiderin ja muiden (2005) analyysiä, joka on alun perin tehty englannin kieleen mutta joka on helposti sovellettavissa myös suomeen. Huomioimme T-yksikköön kaikki alisteiset sivulauseet, lauseenvastikkeet, rektion omaisesti pääverbiin liittyvät infinitiivit, suorat ja epäsuorat lausemuotoiset lainat ja niin edelleen. Arvioimme kirjoitelmia T-yksikköjen kokonaismäärän, niiden sanamäärän sekä niistä lasketun kompleksisuusindeksin kautta. T-yksikköjen jäsenten asemaa suhteessa päälauseeseen arvioidaan hankkeessa erikseen (ks. Muhonen 2019; Tiira 2021).

## T-YKSIKKÖJEN MÄÄRÄ KIRJOITELMISSA JA NIIDEN KESKIPITUUS SANOINA

T-yksikköistä voidaan laskea paitsi niiden kokonaismäärä myös niissä olevien sanojen keskimääräinen määrä. T-yksiköt ilmaisevat osaltaan, miten laajoja kokonaisuuksia kirjoittaja pitää mielessään ja yhdistää samaan syntaktiseen kokonaisuuteen ja kuinka tiukkaan hän ne osaa tai haluaa sitoa toisiinsa. Tekstin eri yksikköjen sidoksisuuden kasvu lisää samalla tekstin koherenssia, ja kehityksellisesti koherenssin lisääntymi-

nen on erityisen mielenkiintoista (ks. esimerkiksi Berman & Verhoeven 2002; Givón 2020). Koherenssia lisää muun muassa se, että useita tapahtumia suhteutetaan toisiinsa. Esimerkissä 11a kyse voi olla samanlaisista tai mahdollisesti eriaikaisistakin tapahtumista, tapahtumajärjestystä ilmaisevan ikonisuuden kannalta kyse voisi olla peräkkäisistä tapahtumista. Esimerkissä 11b kyse on ainoastaan peräkkäisistä. Siten 11b on koherentimpi kuin 11a ja antaa lukijalle vähemmän tulkinnanvaraa. Ei olekaan yllättävää, että syntaktisesti kiinteämmin kontekstiinsa sidottu rakenne on lukijan kannalta myös helpompi (eli nopeampi) prosessoida kuin vastaava irrallisempi rakenne (ks. Hyönä & Vainio 2001). Nuorilla kirjoittajilla tapahtumien aikasuhteen ilmaisu ei tosin välttämättä aina onnistu. Tekstin koherenssia lisää myös samasta henkilöstä pidempään kertominen, mikä näkyy esimerkiksi samaan viittaavien referentiaalisten ketjujen pitenemisenä (ks. tässä teoksessa Kouvon ja muiden analyysiä; Givón 1983).

11a) *Söin ja tein töitä.*

11b) *Syötyäni tein töitä.*

T-yksiköt on laskettu alakoulun ja yläkoulun viimeisiltä luokilta eli kuudennelta ja yhdeksänneltä sekä nuorten aikuisten ryhmästä. Alakoulun muut luokat on sivuutettu, koska nuorimpien aineistossa on niin paljon lauseenmuodostusongelmia. Kuudensien aineistossa on 314, yhdeksänsien aineistossa 160 ja nuorten aikuisten aineistossa 340 kirjoitelmaa. T-yksikköjä on aineistossa kaikkiaan 22 683 ja sanoja niissä yhteensä 155 017. Kuudesluokkalaisilla on kirjoitelmissaan keskimäärin 20 T-yksikköä, tytöillä enemmän kuin pojilla, ja yhdeksäsluokkalaisilla 22. Nuorilla aikuisilla T-yksikköjen määrä vaihtelee 30–40. Siten paitsi sanamäärät myös T-yksikköjen määrä nousee kehityksellisesti. Suurin osa peruskoulun pojista kirjoittaa 10–20 T-yksikön kirjoitelmaa mutta huomattava osa vain alle kymmenen (ks. taulukko 57). Peruskoulun tytöistä kolmannes kirjoittaa 10–20 ja toinen kolmannes 20–30 T-yksikköä kirjoitelmaansa, joten ero vastaavan ikäisiin poikiin on selvä. Aikuisilla miehillä yleisin luokka on 30–40, naisilla T-yksikköjen määrä hajaantuu enemmän kuin miehillä.

Taulukko 57. T-yksikköjen määrä Unelmakirjoitelmissa.

Ikäluokka / T-yksikköluokat %	Alle 10	10–19	20–29	30–39	40–49	50 tai yli
Kuudes F	5	29,4	28,7	20,6	10,0	6,3
Kuudes M	31,6	36,8	17,8	8,6	2,6	2,6
Yhdeksäs F	6,8	26,1	32,9	14,8	12,5	6,8
Yhdeksäs M	19,4	34,7	27,8	9,7	4,2	4,2
Aikuiset F	0,4	3,7	23,3	36,6	29,2	6,6
Aikuiset M	0	6,2	17,5	43,3	21,6	11,3

Sukupuolen vaikutus kirjoitelmien kokonaissanamääriin, kuten jo edellä tuli ilmi, ja vielä T-yksikköjenkin määriin on suuri mutta T-yksikköjen sanamäärä vaihtelee sukupuolen mukaan vähemmän. Keskipituus (= MLTU) nousee hitaasti iän myötä: kuudesluokkalaisilla on keskimäärin 6,3 sanaa yhdessä yksikössä, yhdeksäsluokkalaisilla 6,6 ja nuorilla aikuisilla 7,6. Nuorimmilla on jopa 3–4 sanan T-yksikköjä, aikuisilla nämä ovat vähissä (ks. taulukko 58). Viiden sanan T-yksiköt vähenevät iän mukaan kehityksellisesti ja 6–8 sanan lisääntyvät. Peruskoululaisilla yleisin keskipituus on viisi sanaa, nuorilla aikuisilla seitsemän sanaa. Kirjoitelmat, joissa T-yksiköt ovat hyvin lyhyitä (keskimäärin kolme tai neljä sanaa), ovat tyypillisiä kuudesluokkalaisille ja yhdeksännen pojille. Hyvin pitkiä (9–10 sanaa) on etenkin aikuisilla miehillä: lähes kolmanneksessa miesten kirjoitelmista keskisanapituus on yhdeksän tai enemmän. T-yksikön sanamäärää voi selittää toisaalta rakenteellisella kompleksisuudella (paljon alisteisia rakenteita), toisaalta kuvailevuudella, runsaalla ajallisaikallisella kehystämällä ja niin edelleen. MLTU ja kompleksisuusindeksi korreloivat vahvasti, kuten Pajunen ja Vainio loppuluvussa osoittavat.

Taulukko 58. T-yksiköiden keskipituus (MLTU) sanoina Unelmakirjoitelmissa.

Ikäluokka / MLTU-luokat %	3 sanaa	4	5	6	7	8	9	10 tai enemmän
Kuudes F	1,2	14,4	33,8	27,5	12,5	5,6	3,1	1,9
Kuudes M	1,3	17,7	30,3	21,7	15,1	5,9	2,6	5,3
Yhdeksäs F	1,1	7,9	29,5	23,9	17,0	12,5	6,8	1,1
Yhdeksäs M	0	19,4	26,4	20,8	13,9	12,5	2,8	4,2
Aikuiset F	0	1,7	10,4	25,8	27,9	16,6	8,3	9,1
Aikuiset M	0	2,1	10,3	18,5	24,7	16,5	14,4	13,4

Unelmakirjoitelmien tuloksia voi verrata lehtitekstiin, vaikka koodaus ei olekaan täysin samanlainen. Savon Sanomien korpuksessa sanat on jaoteltu virkkeisiin, ei T-yksiköihin, jolloin muun muassa rinnasteiset lauseet (*Pekka söi ja Matti nukkui*) kuuluvat samaan yksikköön. Eroilla saattaa olla virkkeiden sanamääriin nostava vaikutus. Näin laskettuna Savon Sanomien miljoonan sanan otoksessa on yli 80 000 virkettä ja virkkeen keskimääräinen sanamäärä on 11,8 (mediaani on 11 ja keskiahajonta 6,8). Ero aikuiskirjoittajiin on noin neljä sanaa. Ero käy selvimmän ilmi kumulatiivisista prosentteista. Naiskirjoittajilla sanamäärä 3–9 kattaa T-yksiköistä 91 prosenttia iän mukaan laskien, mutta Savon Sanomissa vastaava sanamäärä kattaa virkkeistä vain 41,6 prosenttia. Ero on niin suuri, että se on merkityksellinen, vaikka laskukriteerit vaikuttaisivatkin lukuihin. Ero tarkoittaa, että lehtitekstin kompleksisuus on sekä lauseke- että lausetasolla huomattavasti korkeampi kuin Unelmakirjoittajilla. Ero liittyy paitsi tekstin ja sisältöjen luonteeseen, myös kirjoittajan kokeneisuuteen.

#### KOMPLEKSISUUSINDEKSI

Kompleksisuusindeksi mittaa kielenkäyttäjän kykyä tiivistää sanomaansa yhteen rakenteelliseen yksikköön ja samalla kykyä painottaa tapahtumien keskinäistä tärkeyttä ja arvioida niitä. Kompleksisuusindeksille (= CI) saadaan arvo laskemalla kirjoitelman T-yksikköjen ja alisteisten rakenteiden määrä yhteen ja jakamalla saatu luku T-yksikköjen määrällä (ks. Schneider ja muut 2005). Jos indeksi on lähellä yhtä, kirjoitelman kompleksisuus on hyvin matala ja teksti etenee päälausein. Jos indeksi lähenee kahta, keskimäärin joka T-yksikössä on yksi alisteinen rakenne ja kompleksisuus on verrattain korkea. Indeksillä vaihtelee Unelma-aineistossa tyypillisimmin 1,5–1,7, kuudennen mediaani on 1,53, yhdeksännen 1,58 ja aikuisilla 1,65–1,7. Matalin arvo on yksi ja tällaisessa tekstissä ei siis ole alisteisia lauseita mutta ei myöskään infinitiivejä. Korkein arvo on yli kolme. Kompleksisuus vaihtelee sekä kirjoittajien iän että sukupuolen mukaan, mikä näkyy selvästi, kun Unelmakirjoitelmat jaotellaan kompleksisuusluokkiin (ks. taulukko 59). Hyvin korkeat arvot ovat poikkeuksellisia ainakin kertovassa tekstissä.



Taulukko 59. Kompleksisuusindeksin arvot luokittain Unelmakirjoitelmissa.

Ikäluokka / CI-luokat %	1–1,19	1,2–1,39	1,4–1,59	1,6–1,74	1,75–1,99	2 tai yli
Kuudes F	3,7	26,2	24,4	19,4	15,0	11,2
Kuudes M	8,5	19,1	25,0	17,1	14,5	15,8
Yhdeksäs F	3,4	14,8	34,1	15,9	15,9	15,9
Yhdeksäs M	4,2	15,3	34,7	15,3	19,4	11,1
Aikuiset F	0	11,6	30,0	23,7	18,3	16,2
Aikuiset M	1,0	6,2	18,5	28,8	25,8	19,6

Peruskoululaisilla on kirjoitelmissaan hyvin matalia kompleksisuusarvoja jonkin verran, eniten kuudennen luokan pojilla. Kuudennen luokan tytöistä neljännes käyttää melko puhutun kielen mukaista rakennetta. Aineistossa kuudennen luokan tyttöjen sekä aikuismiesten kirjoitelmia lukuun ottamatta yleisin on rakennetyyppi, jossa joka toinen T-yksikkö koostuu päälauseesta ja sille alisteisesta rakenteesta, joka toinen pelkää päälauseesta. Kompleksisemmissä rakenteissa alkaa aikuisten kirjoitelmien osuus nousta. Erityisesti aikuisten miesten kirjoitelmille on ominaista korkea kompleksisuus. Sanomalehtikielestä ei ole varmoja vertailutietoja, mutta Savon Sanomien miljoonan sanan otoksesta laskelemia voi yrittää tehdä. Silloin kyllä joutuu turvautumaan virkemääriin, jättämään rinnasteiset lauseet luvuista ja jossain määrin arvioimaan myös lauseenvastikkeiden määriä. Sitaatteja ei ole mahdollista huomioida. Näillä varauksilla kompleksisuusindeksi nousee lehtikielessä 1,8:aan. Voi olettaa, että todennäköisimmin lehtikielen joka virkkeessä on jokin alisteinen elementti.

Kompleksisuuseroille on mitä ilmeisimmin monia syitä. Näyttää ensinnäkin siltä, että miesten Unelmakirjoitelmissa pyritään erittelevämpään kerrontaan, monissa aineiston kirjoitelmissa naiset enemmän kertovat ja kuvailevat kuin erittelevät. Erittelevässä, argumentoivassa tekstissä indeksi on korkeampi (ks. Berman & Verhoeven 2002, 35; Jagaiah ja muut 2020, 2595–2596). Toiseksi miesten kirjoitelmille on ominaista ilmaista vähemmän tärkeät asiat taustalla (eli käyttää lauseenvastikkeita), naisten kirjoitelmissa käytetään enemmän rakenteeltaan yksinkertaisia T-yksikköjä. Kolmanneksi miesten kirjoitelmissa käytetään useammin eräänlaista tapahtumajärjestyksen mukaista rytmittämistä: *menen syö-*

*mään – syötyäni menen ulkoilemaan – ulkoiltuani lähden nukkumaan.* Naisten kirjoitelmissa jätetään tapahtumajärjestys useammin pääteltäväksi; tällainen teksti ei silti ole ilman koherenssia, koska koherenssi voi rakentua tapahtumien luonnolliseen tai loogiseen aikajärjestykseen. Lukijalle etenkin kausaalisuhteiden merkitsemättömyys on silti haaste. Tosiasiassa tietysti sekä miesten että naisten kirjoitelmissa käytetään molempia kyseisiä kirjoittamisstrategioita mutta eri yleisyydellä. Naiskirjoittajien kirjoitelmat ovat T-yksikköinä laskettuna hieman pidempiä kuin miesten, mutta koska kompleksisuuden ja T-yksikköjen määrä korreloi negatiivisesti, miesten tekstit ovat keskimäärin syntaktisesti tiiviimpiä kuin naisten.

Esimerkit 12–15 havainnollistavat mittarein kuvattua tyyppien eroa erittelevästä kerronnasta (12) kuvailevaan (13) ja toisaalta tapahtumien suhteita tarkasti ilmaisevaan (14) vs. ilmaisemattomaan (15) kerrontaan. Esimerkki 15 edustaa yksinkertaista rakennetta, jossa lukijan tehtäväksi jää päätellä tapahtumien keskinäiset suhteet. Esimerkit ovat yhdeksänneltä luokalta. Ohimennen todettakoon, että esimerkki 12 on erittelevyydessään poikkeus Unelmakirjoitelmissa. Aikuisaineiston esimerkit 16–17 osoittavat, että kirjoitelman luonteeseen vaikuttaa risteiviä tekijöitä, joten yksittäinen arvo ei aina ole riittävän informatiivinen. Asian selventämiseksi onkin syytä erikseen tutkia myös kompleksisuustekijöitä tarkemmin.

Erittelevä teksti, MLTU = 7,88, CI = 1,72:

12) *Päivä sekoitetaan usein vuorokauteen. Päivä ei kuitenkaan ole ajanjakso joka käsittää 24 tuntia, vaan matka aamusta iltaan, aurinnoususta auringonlaskuun. Mikä taasen on unelmaa? No, oma mielipiteeni asiaan: unelma – haave, toive, miltein unen omainen. Unelmiäni päivä, mitä tämä konteksti voisi käsittää? Päivä josta uneksin, vai päivä jona unelmoidaan. (9M50)*

Kertova ja kuvaileva teksti, MLTU = 6,17, CI = 1,34:

13) *Aamulla herään valoon. Ulkoa kuuluu viserrystä, mutta talossa on hiljaista. Äiti ei ole vielä herännyt, hyvä niin, tämä päivä olisi minun. Astun ulos ja lempeä tuuli toivottaa minut tervetulleeksi ulkoilmaan. Naapuritkaan eivät ole vielä heränneet, – –. (9F65)*

Rakenteeltaan kompleksinen teksti, MLTU = 5,83, CI = 1,87:

14) *Kun heräsin kuulin laineiden liplattavan ja aaltojen kuohahtamisen niiden osuessa rantakivikkoon. Avasin silmäni ja näin lokiin kaartelevan yläpuolellani. Nousin ylös ja huomasin olevani hiekkarannalla, joka loppui toisella puolella kivikkoon. Päästyäni ylös lähdin kävelemään rantaa pitkin löytääkseni syötävää nälkäni.* (9M147)

Rakenteeltaan yksinkertainen teksti, MLTU = 4,67, CI = 1,14:

15) *On aamu hotelli Presidentissä. Auringon säteet pyrkivät huoneeseeni sälekaihtimien väleistä. Olen juuri herännyt suureen päivään. Huonekaverini ovat jo ylhäällä. Tänään se tapahtuu, mistä olen aina haaveillut. Se on salibandyn MM-finaali. En voi uskoa tätä todeksi, mutta totta se on. Kotiyleisön edessä täydellä Hartwall areenalla, ja Ruotsi vastassa. Huippua!* (9F103)

Korkea NTVR (63,3), korkea leksikaalinen tiheys (81), korkea CI (1,73) ja korkea MLTU (8,1):

16) *Saapuessani treenikämpän pihaan minut havahduttaa ajatuksistani tuttu, hieman katkeran kuuloinen ääni: "Jaahas, jaksaisikohan arvokitarataiteilija roudata tänään edes omat soittokampeensa auttoon? Me ollaan täällä jo kaikki muu roudattu." "No jos nyt sitten tämän kerran", vastaan laiskasti hymyillen ja retuutan kaiutinkaappini suurieleisesti ähisten auton takakonttiin.* (PääsyM46)

Korkeahko NTVR (58,4), korkea leksikaalinen tiheys (85,4), matala CI-arvo (1,1) ja MLTU (4,2):

17) *Varpuset visertävät. Auringon luomat säteet tuovat hohkaavaa lämpöä iholle. Raotan varovaisesti silmiäni tähän kauniiseen aamuun. Omenapuiden tuoksu takapihalla monenkirjavien pudonneiden lehtien ympärillä huumaa mielen. Otan varovaisen hörpyn Earl Grey -teetä ja avaatan sanomalehden ensimmäisen sivun. "Ihanaa", tuumin. Päivä koittaa. Käyn suihkussa ja höylään partani.* (PääsyM166)

## Lopuksi

Keskeinen kehityksellinen trendi kirjoittamisessa on yksinkertaisesta kompleksiin. Trendi ilmaisee samalla kehitystä kielenhallinnassa. Rakenteen tasolla kyse on siitä, että sanat, lausekkeet ja lauseet pitenevät ja niiden monimuotoisuus lisääntyy, tekstin tasolla sanamäärä kasvaa. Kyseinen trendi heijastuu monien kielenyksiköiden kautta, osin risteävästi tyylillisen kehityksen kanssa. Tyyliä voi tutkia vain holistisesti arvioimalla, mutta siihen ei ole yksiselitteisiä, kaikkien hyväksymiä kriteerejä. Toinen ongelma on, että kompleksisuus kasvaa kehityksellisesti, mutta kompleksisin rakenne ei välttämättä ole paras. Ei ole tutkittua tietoa siitä, mikä on kehityksellisen kompleksisuuden ja vaikkapa tyyllitelevän pelkistykseen suhde. Kun miehet ja naiset kirjoittavat kompleksisuudeltaan erilaista tekstiä, kyse voi olla monesta asiasta, esimerkiksi erilaisesta suhtautumisesta ohjeistukseen; ylioppilaskirjoituksiin valmentautuviahan usein kehoitetaan välttämään kompleksisia rakenteita esimerkiksi epäonnistumisvaaran vuoksi.<sup>119</sup> Tyydyimme tässä yhteydessä kokoamaan analyysin keskeiset tulokset, ja jatkamme kompleksisuuden ongelman arvioimista tämän teoksen loppuluvussa.

Sanapituudet kasvavat kehityksellisesti ja tarjoavat nopean keinon arvioida ikätaso ja tekstityyppi huomioiden kirjoitelman kompleksisuutta. Esimerkiksi tämän luvun keskimääräinen sanapituus on 7,6, mikä ylittää kirjoitelmien aikuiskirjoittajan keskiarvon mutta vastaa suunnilleen sanomalehtikeskiarvoa. Sanapituus viittaa argumentoivaan, erittelevään ja melko kompleksiin tekstiin. Alakouluikäisillä kirjoittajilla keskiarvo on 6, ja se nousee hitaasti peruskoulun loppuun mennessä 6,5:een, aikuiskirjoittajilla keskiarvo on 7. Jos samassa kertovassa tms. tekstissä kirjoittaja jää selvästi alle ikäluokkansa keskiarvon, tilannetta pitää arvioida tarkemmin. Myös sanarakenteet kompleksistuvat kehityksellisesti, ja tämä näkyy erityisen hyvin johdosten määrän kasvusta. Nuorimmilla kirjoittajilla johdosten esiintymämäärä kaikista sanoista on 10 prosenttia, yhdeksäsluokkalaisilla 15 prosenttia ja nuorilla aikuisilla 20 prosenttia. Kun sanamäärä aineistossa on yli 200 000, erot ovat tilastollisesti

119 Tieto on peräisin suomen kielen opiskelijoilta.

hyvin merkitseviä. Johdokset tuovat tekstiin spesifejä merkityksiä. Yhdyssanojen määrässä ei vastaavaa kehitystä tapahdu, vaikka yhdyssanojen laatu muuttuu kehityksellisesti.

Leksikaaliset indeksit ovat hyvin tekstityypiriippuvaisia. Kertovassa tekstissä NTVR eli substantiivi–verbi-indeksi on lapsilla 1:1 ja nousee nuorilla aikuisilla 1,1:1:een. Kertovassa tekstissä indeksi on siis odotuksen mukaisesti melko matala. Leksikaalinen tiheys on synteettisessä suomen kielessä korkea, koska itsenäiset sanat eivät ilmaise kieliopillisia merkityksiä vaan taivutus. Kertovassa tekstissä tiheys on kaiken ikäisillä 75–80 prosenttia kokonaissanamäärästä. Erittelevän tekstin leksikaalinen tiheys on jonkin verran korkeampi. Lapsilla keskiarvo on 75 prosenttia, aikuisilla 77 prosenttia eli ero on pieni. Taas on kuitenkin todettava, että tarina on lasten ja aikuisten osalta erilainen. Lapsilla leksikaalinen tiheys on korkea, mutta kuten Pajunen ja Honko tämän teoksen toisessa luvussa osoittavat, se ei ole yhteydessä esimerkiksi leksikaaliseen diversiteettiin, yhdeksäsluokkalaisilla tiheys ja diversiteetti alkavat nousta yhdessä ja aikuisilla yhteys on hyvin selvä. Toisin sanoen aikuisilla on sitä enemmän eri lekseemejä, mitä korkeampi leksikaalisten sanaluokkien osuus on tekstimassasta. Siten aikuisteksteissä korkea leksikaalinen tiheys osoittaa tekstin olevan informaatiopitoista ja semanttisesti rikasta, mutta lapsilla on teksteissään toisteisuutta.

Syntaktinen tiheys niin ikään kasvaa kehityksellisesti. Sitä mittasimme T-yksikköjen keskipituuden (MLTU) ja lauserakenteen kompleksisuusindeksin avulla. MLTU:n mediaani on kuudesluokkalaisilla 6 sanaa, yhdeksäsluokkalaisilla 6,2 ja aikuisilla 7,4. Kuudesluokkalaisten T-yksiköissä on keskimäärin 1,5 lausetta, yhdeksäsluokkalaisilla 1,58 ja aikuisilla 1,67. Lauserakenteen kompleksistuminen tarkoittaa informaation tiiviimpää esittämistä ja sitä nostavat esimerkiksi sivulauseet ja infiniittiset rakenteet. Sivulauseista nominaaliset viittaavat evaluoinnin lisääntymiseen, relatiiviset muun muassa tarkempaan diskurssientiteettien kuvaukseen sekä adverbiaaliset ajan, ehtojen ja syiden esittämiseen.

Kaikkia annettuja keskiarvolukuja voidaan käyttää ikäkausitason arvioinnissa, toiset luvut vain ovat helpompia tavoittaa kuin toiset. Esimerkiksi leksikaalisen indeksin laskeminen edellyttää tarkkaa sanaluokka-analyysiä, mutta sanapituuskeskiarvot on helppo laskea tekstin-

käsittelyohjelmalla. Kokemus kuitenkin opettaa arvioimaan. Esimerkiksi johtamattomien ja johdettujen sanojen suhdetta on melko helppo tarkkailla. Vastaavasti voi opetella arvioimaan keskiarvojen taustatekijöitä eri ikäryhmissä. Esimerkiksi kompleksisuusindeksiin vaikuttavien infiniittisten muotojen puuttuminen on lapsilla kehityksellinen ongelma, aikuisilla taas yksinkertaisen rakenteen suosimisen merkki. Hyvin yksinkertainen tai hyvin kompleksinen tyyli edustavat molemmat syntaktisesti yksipuolista tyyliä (vrt. Beers & Nagy 2009, 187).

Myös sanamäärä kasvaa kehityksellisesti ja tasaisesti. Sanamäärää mittaamme sekä kirjoitelmien kokonaissanamääränä että T-yksikköjen määränä. Sanamäärä on yleisesti käytetty kehitysmittari mutta monella tavalla ongelmallinen: Se riippuu tehtävästä, tehtävänannosta, tekstin luonteesta ja ennen kaikkea kirjoittajan motivaatiosta ja pystyvyysuskosta. Ongelmallisinta on kuitenkin, että tekstin kompleksisuus ja tiiviys laskee sekä kokonaissanamääriä että T-yksiköiden määriä. Toisin sanoen lattea, kehityksellisesti ikään kuin heikompi rakenne tuottaa enemmän sanoja kuin kehityksellisesti parempi, kompleksinen rakenne (vrt. Mühöner 2019).<sup>120</sup> Sanamäärämittari ei myöskään huomioi tekstin koherenttiutta. Suurin ongelma on, että sanamäärä mittarina suosii tyttöjä, joiden tuottaminen sinänsä on helppoa; naisten prosessointinopeus on keskimäärin miehiä nopeampaa älykkyydeltään vertaistetussa aineistossa ja se näkyy sekä lukemisessa että kirjoittamisessa (ks. Caramata & Woodcock 2006; Keith, Reynolds, Patel ja Ridley 2008; Beal, Holdnack, Saklofske & Weiss 2016).

Kirjoittamisen sukupuolierot ovatkin selvät tyttöjen hyväksi, jos arvio perustuu yksin sanamääriin. Tytöt myös saavuttavat kertomusrakenteen konventiot poikia aikaisemmin, kehitys toisin sanoen kulkee eriaikaisesti. NTVR ja leksikaalinen tiheys sekä johdosten määrät eivät eroa sanotavasti sukupuolen mukaan. MLTU ja kompleksisuusindeksi ovat miespuolisilla kirjoittajilla hieman korkeammat kuin naispuolisilla, ja monet mieskirjoittajat suosivat kompleksista tyyliä. On tutkimustuloksia, joiden mukaan MLTU ja kompleksisuusindeksi ovat korkeampia juuri

120 On mielenkiintoista, että Nieminen (2007) sai jo 2–3-vuotiaiden lasten syntaktisen rakenteen arvioinnissa esiin pituuden ja kompleksisuuden käänteisen suhteen.

niillä opiskelijoilla, joiden koulumenestys on parempi (vrt. Truckenmiller & Petscher 2020), mutta toisaalta syntaktisen kompleksisuuden ja laadun välistä yhteyttä pidetään inkonsistenttina (Beers & Nagy 2009, 189–190). On myös mahdollista, että suhde ei ole inkonsistentti vaan riippuvainen iästä ja tekstin tyypistä (Crowhurst 1980) tai että se on vain viitteellinen (Crowhurst 1980; 1983).

Toinen kehityksellinen trendi, joka näkyy koehenkilökohtaisessa tarkastelussa jokaisen tarkastellun rakenneyksikön osalta, on ikäluokka-kohtaisen hajonnan kaventuminen. Nuoremmilla mittarien arvojen hajonta on aina suurempi kuin vanhemmilla, vaikka monen mittarin osalta on luonnollista, että hajontaa on hieman. Mittarin ääriarvot viittaavat ongelmiin, ikäluokan keskiarvolukujen ympärillä pyörivät arvot taas viittaavat ennemmin tekstinrakentamisen keinojen eroihin. Koulutuksen tehtävä on vähentää ikäluokittain näkyviin tulevaa suurta hajontaa aivan samalla tavalla kuin sen pitäisi saada aikaan ikäkausittain tasaisesti nouseva muutos tekstitaidoissa. Esimerkiksi monen mittarin arvot eivät eroa kuudesluokkalaisten ja yhdeksäsluokkalaisten poikien välillä, vaikka tyttöjen kehitys on tasaisesti etenevä. Tämän ”suoran viivan” täytyy olla koulun – opetussuunnitelmien, kouluviranomaisten ja viime kädessä yhteiskunnan – ongelma, ei pelkästään itse poikien. Erot eivät näet selity kielellisen älykkyyden eroilla.

## LÄHTEET

- Babayigit, S. 2015. The dimensions of written expression: Language group and gender differences. *Learning and Instruction* 35, 33–41.
- Bamberg, M. (toim.) 1997. *Narrative development: Six Approaches*. Mahwah: Lawrence Erlbaum Ass.
- Beal, L., Holdnack, J., Saklofske, D. & Weiss, L. 2016. Practical considerations in WISC-V interpretation and intervention. Julkaisussa: L. Weiss, D. Saklofske, J. Holdnack & J. Prifitera (toim.) *WISC V. Clinical use and interpretation*. Cambridge: Academic Press, 63–93.
- Beers, S. F. & Nagy, W. E. 2009. Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing* 22, 185–200.
- Beers, S. F. & Nagy, W. E. 2011. Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing* 24, 183–202.

- Berman, R. A. 2004. Between emergence and mastery: The long developmental route of language acquisition. Julkaisussa: R. A. Berman (toim.) *Language development across childhood and adolescence*. Amsterdam: John Benjamins, 9–34.
- Berman, R. A. 2007. Developing linguistic knowledge and language use across adolescence. Julkaisussa: E. Hoff & M. Shatz (toim.) *Blackwell handbook of language development*. Malden: Blackwell Publishing, 347–367.
- Berman, R. A. & Verhoeven, L. 2002. Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language and Literacy* 5: 1, 2–42.
- Berman, R. A. & Nir, B. 2010. The lexicon in writing–speech-differentiation. *Written Language & Literacy* 13: 2, 183–205.
- Berman, R. A., Nayditz, R. & Ravid, D. 2011. Linguistic diagnostics of written texts in two school-age population. *Written Language and Literacy* 14: 2, 161–187.
- Biber, D., Conrad, S. & Reppen, R. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Bouwer, R., Béguin, A., Sanders, T. & van den Bergh, H. 2015. Effects of genre on the generalizability of writing scores. *Language Testing* 32: 1, 83–100.
- Bromley, D. B. 1991. Aspects of written language production over adult life. *Psychology and Aging* 6: 2, 296–308.
- Burns, F. A., de Villiers, P. A., Pearson, B. Z. & Champion, T. B. 2012. Dialect-neutral indices of narrative cohesion and evaluation. *Language, Speech, and Hearing Services in Schools* 43, 132–152.
- Camarata, S. & Woodcock, R. 2006. Sex differences in processing speed: Developmental effects in males and females. *Intelligence* 34, 231–252.
- Carlisle, J. F. 2000. Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing* 12, 169–190.
- Carlisle, J. F. 2003. Morphology matters in learning to read. *Reading and Writing* 12, 169–190.
- Crowhurst, M. 1980. Syntactic complexity in narration and argument at three grade levels. *Canadian Journal of Education* 5, 6–13.
- Crowhurst, M. 1983. Syntactic complexity and writing quality: A review. *Canadian Journal of Education* 8: 1, 1–16.
- DESI-Konsortium. 2008. *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Weinheim und Basel: Beltz Verlag.
- Elomäki, K. 2011. *Kirjoitetun kielen rakenteellisten kompleksistuminen kolmas- ja neljäsluokkalaisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Gillam, R. & Pearson, M. 2004. *Test for narrative language*. Austin: Pro-Ed.
- Givón, T. (toim.) 1983. *Topic continuity in discourse: Quantified cross-language studies*. Amsterdam: John Benjamins.
- Givón, T. 2009. *The genesis of syntactic complexity*. Amsterdam: John Benjamins.
- Givón, T. 2020. *Coherence*. Amsterdam: John Benjamins.
- Graesser, A. C., McNamara, D. S. & Kulikowich, J. M. 2011. Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher* 40: 5, 223–234.
- Haapala, T. 2008. *Finiittiverbeistä verbiketjuihin: verbityimen kompleksistuminen S2-oppijoiden kielessä*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.



- Habermas, T. & de Silveira, C. 2008. The development of global coherence in life narratives across adolescence: Temporal, causal and thematic aspects. *Developmental Psychology* 44: 3, 707–721.
- Hadley, P. A. & Ripoli, M. 2012. A lifespan perspective on individual differences in grammatical abilities. *Linguistic Approaches to Bilingualism* 2: 3, 269–272.
- Hall-Mills, S. & Apel, K. 2015. Linguistic feature development across grades and genre in elementary writing. *Language, Speech, and Hearing Services in Schools* 46, 242–255.
- Harjunen, E., Marjanen, J. & Karlsson, J. 2019. Äidinkielen pieni pitkittäisarviointi 2014–2017. Perusopetuksen päätöstä lukion päättöön. Julkaisut 4: 2019. Kansallisen koulutuksen arviointikeskus Karvi.
- Heikkola, L. M. 2017. *Kognitiivinen uupumus ja kieli. Semispontaanien narratiivien määrällinen ja laadullinen analyysi*. Åbo: Åbo Akademi University.
- Heinonen, S. 2018. *Narratiivisuuden sisällöllinen kehittyminen aiheanalyysin kautta tarkasteltuna: kvantitatiivinen tutkimus aiheiden esiintymisen vaihtelusta iän ja sukupuolen mukaan "Unelmieni päivä" -kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Hickmann, M. 2003. *Children's discourse: Person, time, and space across languages*. Cambridge: Cambridge University Press.
- Honko, M. 2013. *Alakouluikäisen leksikaalinen tieto ja taito. Toisen sukupolven suomi ja S1-verrokki*. Tampere: Tampereen yliopisto.
- Honko, M., Jarvis, S. & Vainio, S. 2019. Lukijat sanaston monimuotoisuutta määrittä-mässä: leksikaalisen diversiteetin tarkastelua määrällisen ja laadullisen tutkimuksen rajapinnassa. *Virtittäjä* 123: 1, 44–78.
- Huisman, T. 2006. *Luen, kirjoitan ja ratkaisen. Peruskoulun kolmasluokkalaisten oppimistulokset äidinkieliessä ja matematiikassa*. Oppimistulosten arviointi 7/2006. Helsinki: Opetushal-litus.
- Hyönä, J. & Vainio, S. 2001. Reading morphologically complex clause structures in Finnish. *European Journal of Cognitive Psychology* 13, 451–474.
- Itkonen, E. 2010. *Maailman kielten erilaisuus ja samuus. Osa III*. Turku: Turun yliopisto.
- Itkonen, E., Pajunen, A. & Suomi, K. 2009. Maailman kielten luokituksen ongelmia. *Virtittäjä* 113: 2, 285–296.
- Jagaiah, T., Olinghouse, N. G. & Kearns, D. M. 2020. Syntactic complexity measures: vari-ation by genre, grade-level, students' writing abilities, and writing quality. *Reading and Writing* 33, 2577–2638.
- Jarvis, S. 2013. Capturing the diversity in lexical diversity. *Language Learning* 63: Suppl/1, 87–103.
- Johansson, V. 2008. Lexical diversity and lexical density in speech and writing: A develop-mental perspective. *Department of linguistics and phonetics working papers* 53, 61–79. Lund: Lund University.
- Justice, L. M., Bowles, R., Kaderavek, J. K., Ukrainetz, T., Eisenberg, S. & Gillam, R. 2006. The index of narrative micro-structure (INMIS): A clinical tool for analyzing school-aged children's narrative performance. *American Journal of Speech-Language Pathology* 15, 1–15.
- Juvonen, R., Kauppinen, A., Makkonen-Craig, H. & Lehti-Eklund, H. 2011. Mitä lukiolais-ten kirjoittamisesta on tutkittu. Julkaisussa: A. Kauppinen, H. Lehti-Eklund, H. Makko-

- nen-Craig & R. Juvonen (toim.) *Lukiolaisten äidinkieli. Suomen ja ruotsinkielisten lukioiden opiskelijoiden tekstimaisemat ja kirjoitustaitojen arviointi*. Helsinki: Suomalaisen Kirjallisuuden Seura, 27–61.
- Kaivapalu, A. 2005. *Lähdekieli kielenoppimisen apuna*. Jyväskylä: Jyväskylän yliopisto.
- Keith, T. Z., Reynolds, M. R., Patel, P. G., Ridley, K. P. 2008. Sex differences in latent cognitive abilities ages 6 to 59. Evidence from the Woodcock-Johnson III tests of cognitive abilities. *Intelligence* 36, 502–525.
- Kemper, S., Marquis, J. & Thompson, M. 2001. Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. *Psychology and Aging* 16: 4, 600–614.
- Keskinen, J. 2012. ”Jos joskus pitäisi keksiä joku mielikuvitusmaailma, niin aivan varmasti jokaisella ihmisellä olisi ainakin omat unelmat: siitähän jokainen unelmoi.” *Modaalisuus alakouluikäisten lasten unelmakirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Kirby, J. R., Deacon, H. S., Bowers, P., Izenberg, L., Wade-Woolley, L. & Parrila, R. 2012. Children’s morphological awareness and reading ability. *Reading and Writing* 25: 2, 389–410.
- Koivunen, J. 2012. *Adverbien käyttö alakoululaisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Koutsoftas, A. & Gray, S. 2012. Comparison of narrative and expository writing in students with and without language-learning disabilities. *Language, Speech, and Hearing Services in Schools* 43, 395–409.
- Kuisma, J. 2011. *Ominaisuuden ilmausten käytön kehittyminen alakouluikäisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Kulju, P., Kauppinen, M., Hankala, M., Harjunen, E., Pentikäinen, J. & Routarinne, S. 2017. Reviewing Finnish studies on writing in basic education. Towards a pedagogy for diversity. Julkaisussa: N. Pyyry, L. Tainio, K. Juuti, R. Vasquez & M. Paananen (toim.) *Changing subjects, changing pedagogies: Diversities in school and education*. Helsinki: Helsingin yliopisto, 110–127.
- Kusnetsoff, T. 2017. *Suomenkielisten nuorten morfologinen tietoisuus ja sen yhteys luetun ymmärtämiseen*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Kyösti, J. 2018. *Motivoinnin vaikutuksia kuudesluokkalaisten kirjoittamiin kertomuksiin*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Labow, W. & Waletzky, J. 1967. Narrative analysis: Oral versions of personal experiences. Julkaisussa: J. Helm (toim.) *Essays on the verbal and the visual arts*. Seattle: American Ethnological Society, 12–44.
- Lappalainen H.-P. 2003. *Osaat lukea – miten osaat kirjoittaa? Perusopetuksen 6. vuosiluokan suorittaneiden äidinkielen ja kirjallisuuden oppimistulosten arviointi 2002*. Helsinki: Opetushallitus.
- Lappalainen, H.-P. 2008: *On annettu hyviä numeroita. Perusopetuksen 6. vuosiluokan suorittaneiden äidinkielen ja kirjallisuuden oppimistulosten arviointi 2007*. Helsinki: Opetushallitus.
- Leskinen, H., Savijärvi, I. & Särkkä, T. 1974. *Koululaisten kirjallisen ilmaisutaidon kehittymisen*. Jyväskylä: Jyväskylän yliopisto.

- Lieber, R. & Štekauer, P. (toim.) 2014. *The Oxford handbook of derivational morphology*. Oxford: Oxford University Press.
- Malvern, D. D., Richards, B. J., Chipere, N. & Durán, P. 2008 [2004]. *Lexical diversity and language development. Quantification and assessment*. New York: Palgrave Macmillan.
- Matilainen, K. 1989. *Kirjoitustaidon kehittyminen neljän ensimmäisen kouluvuoden aikana*. Joensuu: Joensuun yliopisto.
- McCarthy, P. M. & Jarvis, S. 2007. *Vocd: A theoretical and empirical evaluation*. *Language Testing* 24: 4, 459–488.
- Muhonen, M. 2019. *Kompleksiset lauserakenteet 6.-luokkalaisten Unelmieni päivä -kirjoittelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Murtorinne, A. 2005. *Tuskan hauskaa! Tavoitteena tiedostava kirjoittaminen. Kirjoitusprosessi peruskoulun yhdeksännellä luokalla*. Jyväskylä: Jyväskylän yliopisto.
- Mäkelä, K.-L. 2012. *Substantiivilausekkeiden kompleksistuminen alakouluikäisten Unelmieni päivä -kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Mäkihönko, M. 2006. *Luetun ymmärtämisen ja tuottavan kirjoittamisen kehittyminen alkuopetuksen aikana*. Joensuu: Joensuun yliopisto.
- Mäkinen, L. 2014. *Narrative language in typically developing children, children with specific language impairment and children with autism spectrum disorder*. Oulu: University of Oulu.
- Mäkinen, L., Loukusa, S., Laukkanen, P., Leinonen, E. & Kunnari, S. 2014. Linguistic and pragmatic aspects of narration in Finnish typically developing children and children with specific language impairment. *Clinical Linguistics & Phonetics* 28: 6, 413–427.
- National Assessment of Educational Progress (= NAEP). 2011. Writing assessment. [http://nationsreportcard.gov/writing\\_2011](http://nationsreportcard.gov/writing_2011). Viitattu 13.9.2019.
- Nahkola, T. 2012. *Yhdylauseiden kehitys kouluikäisen kielessä: tutkimus 1.–6.-luokkalaisten kirjoittelusta*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Nieminen, L. 2007. *A complex case. A morphosyntactic approach to complexity in early child language*. Jyväskylä: University of Jyväskylä.
- Nippold, M. A. 2006 [3. painos]. *Later language development: School-age children, adolescents and young adults*. Austin: Pro-Ed.
- Nippold, M. A., Hesketh, L. J., Duthie, J. K. & Mansfield, T. C. 2005. Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research* 48, 1048–1064.
- Nippold, M. A., Mansfield, T. & Billow, J. L. 2007. Peer conflict explanations in children, adolescents, and adults: Examining the development of complex syntax. *American Journal of Speech-Language Pathology* 16, 179–188.
- Nippold, M. & Sun, L. 2008. Knowledge of morphologically complex words: A developmental study of older children and young adolescents. *Language, Speech, and Hearing Services in Schools*, 39, 365–373.
- Nippold, M. A., Cramond, P. M. & Hayward-Mayhew, C. 2014. Spoken language production in adults: Examining age-related differences in syntactic complexity. *Clinical Linguistics & Phonetics* 28: 3, 195–207.
- Oinonen, S. 2008. *Rakenteellinen kompleksistuminen peruskoulun seitsemäs-, kahdeksas- ja yhdeksäsluokkalaisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.

- Olinghouse, N. G. & Wilson, J. 2013. The relationship between vocabulary and writing quality in three genres. *Reading and Writing* 26, 45–65.
- O’Loughlin, K. 1995. Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language Testing* 12, 217–237.
- Pajunen, A. 2010. Sanojen synteettisyysasteesta suomen kielessä. *Virtittäjä* 114: 4, 481–501.
- Pajunen, A. 2012. Kirjoittamistaitojen kehitys 8–12-vuotiailla: alakoululaisten unelmakirjoittelmat. *Virtittäjä* 116: 1, 3–34.
- Pajunen, A. & Palomäki, U. 1984. *Tilastotietoja suomen kielen rakenteesta* 1. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Pajunen, A. & Virtanen, P. 2002. SS2001-korpus, Savon Sanomat 2001. FDG-parserilla analysoitu sanomalehtikorpus, Context-Mill-muotoinen tietokanta.
- Pajunen, A., Itkonen, E., Vainio, S. 2015. Sanamerkitkityksen hallinta nuorilla aikuisilla. *Virtittäjä* 119: 2, 160–187.
- Pajunen, A., Itkonen, E. & Vainio, S. 2016. Nuorten aikuisten kyky määritellä sanoja. *Virtittäjä* 120: 4, 477–515.
- Pajunen, A. & Itkonen, E. 2019. Intuition and beyond: A hierarchy of descriptive methods. Julkaisussa A. Mäkilähde, V. Leppänen, E. Itkonen (toim.) *Norms and normativity in language and in linguistics*. Amsterdam: John Benjamins, 213–234.
- Palonen, O. 2013. *Pronominit 8–12-vuotiaiden kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- PISA 2012 ensituloksia*. Opetus- ja kulttuuriministeriön julkaisuja 2013: 20. <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/75271/okm20.pdf?sequence=1&isAllowed=y>. Viitattu 15.11.2014.
- PISA 15 ensituloksia*. Huipulla pudotuksesta huolimatta. Opetus- ja kulttuuriministeriön julkaisuja 2016: 41. <http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/79052/okm41.pdf>. Viitattu 25.4.2017.
- PISA 2018*. Suomi lukutaidossa parhaiden joukossa. Opetus- ja kulttuuriministeriön tiedote 3.12.2019. <https://minedu.fi/-/pisa-2018-suomi-lukutaidossa-parhaiden-joukossa>. Viitattu 24.6.2020.
- Petälä, I. 2018. *Kertomuksen rakenne, evaluaatio ja kompleksisuus aikuisten käännekohta-kertomuksissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Price, J. R. & Jackson, S. C. 2015. Procedures for obtaining and analyzing writing samples of school-age children and adolescents. *Language, Speech, and Hearing Services for Schools* 46, 277–293.
- Rainoja, T. 2014. *Lausetajun kehitys ja lausepituuksien kasvu 2.–6.-luokkalaisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Ranta, T. 2007. *Kirjoittamisprosessi teksteinä: tekstilingvistinen näkökulma abiturienttien tekstintuottamismenettelyihin*. Joensuu: Joensuun yliopisto.
- Ravid, D. 2006. Semantic development in textual contexts during school years: Noun scale analyses. *Journal of Child Language* 33, 791–821.
- Ravid, D. & Avidor, A. 1998. Acquisition of derived nominals in Hebrew: Developmental and linguistic principles. *Journal of Child Language* 25, 229–266.
- Read, J. 2000. *Assessing vocabulary*. Cambridge: Cambridge University Press.

- Salmenoja, E. 2013. *Rakenteellinen kompleksistuminen yläkouluikäisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Scott, C. M. & Windsor, J. 2000. General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech Language and Hearing Research* 43: 2, 324–339.
- Schneider, P., Vis Dubé, R., & Hayward, D. 2005. The Edmonton narrative norms instrument. <http://www.rehabresearch.ualberta.ca/enni>. Viitattu 15.9.2018.
- Sintonen, M. 2011. *Suunnittelemisen vaikutus kertomuksen rakenteeseen viides- ja seitsemäsluokkalaissa kirjoittajilla*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Strömqvist, S., Johansson, V., Kriz, S., Ragnarsdóttir, H., Aisenman, R. & Ravid, D. 2002. Toward a cross-linguistic comparison of lexical quanta in speech and writing. *Written Language and Literacy* 5, 49–69.
- Strömqvist, S., Nordqvist, Å. & Wengelin, Å. 2004. Writing the frog story: Developmental and cross-modal perspectives. Julkaisussa: S. Strömqvist & L. Verhoeven (toim.) *Relating events in narrative. Volume 2: Typological and contextual perspectives*. Mahwah: Lawrence Erlbaum Ass., 359–394.
- Suvanto, A. 2012. *Lapsi tarinaa rakentamassa. Kielihäiriöisten lasten kerrontataidot ja niiden kuntouttaminen*. Oulu: Oulun yliopisto.
- Tiirö, M. 2021. Lauserakenteiden kompleksisuus ammattikorkeakouluisten kirjoitelmissa. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Truckenmiller, A. J. & Petscher Y. 2020. The role of academic language in written composition in elementary and middle school. *Reading and Writing* 33, 45–66.
- Vainio, S., Pajunen, A. & Hyönä, J. 2014. L1 and L2 word recognition in Finnish: Examining L1 effects on L2 processing of morphological complexity and morphophonological transparency. *Studies in Second Language Acquisition* 2014: 1, 1–30.
- Vainio, S., Pajunen, A. & Hyönä, J. 2016. Processing modifier-head agreement in reading L1 and L2 Finnish: An eye-tracking study. *Second Language Research* 32: 1, 3–24.
- Vainio, S., Pajunen, A. & Häikiö, T. 2019. Acquisition of Finnish derivational morphology: School-age children and young adults. *First Language* 39: 2, 139–157.
- Vesala, S. 2012. *Kerronnan keinot alakouluikäisten teksteissä*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Viljanen, M. 2012. *8–9-vuotiaiden S1- ja S2-oppilaiden syntaktisten valmiuksien kehittyminen*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Writing 2011. = The Nation's Report Card. Writing 2011. National assessment of educational progress at grades 8 and 12. U.S. Department of Education.
- Wolf Nelson, N. & Van Meter, A. M. 2007. Measuring written language ability in narrative samples. *Reading and Writing Quarterly* 23: 3, 287–309.
- Örnmark, M. 2011. Tyylitaju ja genretietoisuus 9. luokalla. Julkaisussa: E. Harjunen, R. Juvonen, J. Kuusela, B. Silén, M. Sääkslahti & M. Örnmark (toim.) *Miten peruskoulu-laiset kirjoittavat? Näkökulmia ja kysymyksiä. Perusopetuksen 9. luokan äidinkielen ja kirjallisuuden oppimistulosten seuranta-arvioinnin aineistoa 2010*. Raportit ja selvitykset 2011: 2, 40–51. <https://www.oph.fi/julkaisut>. Viitattu 22.4.2020.

# Syntaktista rakennetta tiivistävät ja kompleksistavat tekijät peruskoululaisilla ja nuorilla aikuisilla

*Anneli Pajunen*

ORCID <https://orcid.org/0000-0001-8597-7491>

*Seppo Vainio*

## Johdanto

Syntaktisen rakenteen kehitystä tarkastellaan tyypillisesti sen tiivistymisen näkökulmasta. Tiivistyminen tarkoittaa, että asioiden väliset suhteet esitetään siten, että painotetaan enemmän relevantteja kuin irrelevantteja tekijöitä, tapahtumia ei aina esitetä tapahtumajärjestyksen vaan esimerkiksi tärkeysjärjestyksen mukaan, pohditaan asioiden välisiä syy-seuraussuhteita, arvioidaan tapahtumia ja niin edelleen (ks. esimerkiksi Beers & Nagy 2011). Tiivistäminen heijastuu syntaktiseen rakenteeseen ja näkyy muun muassa rakenteellisen hierarkian ja sanamäärien kasvusta. Kielellisesti se ilmaistaan käyttämällä nominalisaatiota, monijäsenisiä lausekkeita, sivulauseita, lauseenvastikkeita ja infiniittisiä rakenteita, toisin sanoen rakenteita, jotka ovat ominaisia kirjoitetulle kielelle ja joita rakenteeltaan yksinkertaisemmassa puhutussa kielessä käytetään vähemmän. Nämä rakenteelliset keinot opitaan hallitsemaan kehityksellisesti, ja niiden monipuolinen hallintaan saaminen kestää useimmilla aikuisilla.

suuteen. Tiivistyminen ja sen mukanaan tuoma rakenteen kerroksisuus eli kompleksisuus on huomioitava paitsi kehityksellisessä tutkimuksessa myös, kun tutkitaan kielen oppimista (ks. Lu 2009), lukemista ja kirjoittamista (ks. esimerkiksi Leikin & Assayag Bouskila 2004; Nippold, Hesketh, Duthie & Mansfield 2005; Nippold, Mansfield & Billow 2007), kielen ymmärtämisen- ja tuottamisprosesseja (ks. Demberg & Keller 2008; Scontras, Badecker, Shank, Lim & Fedorenko 2014) tai kun laaditaan arviointikriteerejä (vrt. Graesser, McNamara & Kulikowich 2011). Kompleksisuus huomioidaan tyypillisesti myös tekstien vaikeutta arvioitaessa.

Kompleksisuutta käsitellään kirjallisuudessa monessa yhteydessä ja määrittelystä ei ole yksimielisyyttä (ks. esimerkiksi Givón 1991; Dahl 2004; Itkonen 2009). Kompleksisuudesta voidaan puhua myös monen eri tason ilmiönä, niin sana-, lauseke-, lause- kuin tekstitason ilmiönä, ja esimerkiksi termejä syntaktinen, kognitiivinen, lingvistinen ja rakenteellinen kompleksisuus voidaan käyttää lähes synonyymeinä (ks. Szmrecsanyi 2004). Kompleksisuutta ei siten ole yksiselitteistä operationalisoida. Givón (2009, 3) määrittelee kompleksisuuden organisoituneiden entiteettien ominaisuudeksi, jolloin kompleksisuus nähdään asteikollisena, lähtökohtana esimerkiksi juuri peruslausetaso. Rakenteen eli objektiivisen kompleksisuuden kannalta kriteerinä voi olla paitsi lauserakenteiden tiivistyminen myös niiden ei-lineaarisuus. Kompleksinen (lause)rakenne on se, joka ei koostu pelkästään peruslauseen eli intransitiivi-, transitiivi- tai kopulalauseen predikaatista ja sen argumenteista, ja ei-lineaarisia ovat esimerkiksi sisäkkäiset rakenteet. Toimijaan suhteutuvaa kompleksisuutta (*agent related*) voidaan arvioida oppimisen tai ymmärtämisen tuottamasta vaivasta.

On tärkeä ymmärtää, että objektiivinen ja toimijaan suhteutuva kompleksisuus eivät ole suorassa suhteessa toisiinsa.<sup>121</sup> Siksi sellaiset ennakkoletukset kuin ”vähempi on helpompi” tai ”selkeästi ajateltu on selkeästi ilmaistu” ovat ongelmallisia. Oletetaan esimerkiksi, että rakenne, jossa on vähän kielipiillistä merkintää ja lyhyet lauseet ja sanat, on yksinkertainen (= ei-kompleksinen), mutta samalla oletetaan, että tämä sama

121 Esimerkiksi Lapp, Moss, Grant & Johnson (2015) käsittelevät kompleksisuutta monipuolisesti koulukontekstissa sekä opettamisen että opiskeluvaihtoehtojen kannalta.

rakenne on helppo tuottaa tai ymmärtää. Prosessointitutkimus on osoittanut, että on oikeasti vaativaa, jos tekstistä puuttuu koherenssin ilmimerkintä ja jos lukija joutuu itse tekstin koherenssin päättelemään. Itkosen (2010, 9) mukaan parataktisten lauseiden S<sub>1</sub> ja S<sub>2</sub> suhde voi olla peräkkäisyys, seuraamus, samanaikaisuus tai vastakohta, lisäksi ne voivat olla ajallisessa, finaalisessa tai alisteisessa suhteessa toisiinsa. Nämä kaikki suhteet lukija joutuu päättelemään, jos niitä ei merkitä näkyviin. Esimerkiksi McNamaran, Louwersen ja Graesserin (2002, 11) mukaan lausepituus – siis lauseen sanamäärä – on lauseiden tai rakenteiden kompleksisuusvertailussa ongelmallinen käsite. ”Vähempi on helpompi” -periaatehan johtaa siihen, että lyhyt lause tai virke olisi helpompi ymmärtää kuin pitkä. Esimerkki 1a sisältää kaksi lyhyttä virkettä, esimerkki 1b ilmaisee saman asian yhdellä pitkällä virkkeellä, ja jos siis sanamäärä sellaisenaan olisi kompleksisuuskriteeri, 1a olisi helpompi ymmärtää kuin 1b:

- 1a) *One part of the cloud develops a downdraft. Rain begins to fall.*
- 1b) *One part of the cloud develops a downdraft, which causes rain to fall.*

Esimerkissä 1a ei kuitenkaan anneta kielellistä vihjettä lauseiden välisestä kausaalisuhteesta ja lukijan pitää luoda tämä suhde pelkän lauseiden läheisyyden perusteella. Esimerkissä 1b sana *causes* toimii kohesiivisena lauseiden suhdetta ilmaisevana keinona (ks. McNamara ja muut 2002, 11). Tämän voi tulkita muun muassa siten, että lausetason kompleksisuus vaikuttaa eri tavoin kuin tekstitasoon. Toisaalta Hyönä ja Vainio (2001) osoittavat suomen osalta, että vaikka saman asiasisällön ilmaisevissa lauseissa molemmissa olisi niiden välisen suhteen ilmaisin (vrt. 2a ja 2b), lukijan kannalta toinen – ja tässä tapauksessa rakenteellisiin kriteereihin kompleksisemmaksi oletettu – voi olla helpompi (= luetaan nopeammin).<sup>122</sup> Eroa selittävät rakenteiden frekvenssierot: esimerkin 2b

122 Tutkimus on toteutettu silmänliikkeitä mittaamalla. Metodilla päästään kiinni lukijan tiedostamattomaan prosessointiin eli alemman tason visuaaliseen prosessointiin, vaiheeseen ennen tietoista, semanttista prosessointia (ks. esimerkiksi Vainio 2006).



konstruktio on suomen kielessä yleisempi kuin esimerkin 2a *jotta*-lause. Temporaalirakenteen ja *kun*-lauseen suhde on päinvastainen, koska nyt *kun*-lause on paitsi yksinkertaisempi (= vähemmän tiivis) myös yleisempi.

2a) *Jotta uhri toipu/isi, hänen on saatava pikaisesti lisää verta.*

2b) *Toipuakseen uhrin on saatava pikaisesti lisää verta.*

Oletus ”selkeästi ajateltu on selkeästi – yksinkertaisesti – ilmaistu” si-  
vuuttaa tekstin abstraktiustason ja asiayhteyden, tuottamisen ja ymmär-  
tämisen sekä vielä lukijan tietotason. Se sitoo asiantuntijan ajattelun  
maallikon ajatteluun väittämällä, että selkeää on vain se, mitä maallikko  
– asiantuntijan oletuksen mukaan – ajattelee ja ymmärtää. Esimerkiksi  
Preiss, Castillo, Flotts ja San Martin (2013) osoittavat, että kirjoittajilla on  
taipumus kirjoittaa vaikeista asioista kompleksisemmin kuin helpoista.  
Nippold toistaa monessa yhteydessä, että sisällön käsitteellinen vaikeus  
näkyvät rakenteen kompleksisuutena (ks. esimerkiksi Nippold 2009).  
Lapp, Moss, Grant ja Johnson (2015) erottavat sekä kvantitatiivisen, mi-  
tattavissa olevan että toimijaan suhteutuvan kompleksisuuden mutta  
lisäksi vielä kvalitatiivisen eli sisältöön – käsitteelliseen tasoon – liittyvän  
kompleksisuuden.

Tutkimme tässä luvussa kehityksellisestä näkökulmasta syntaktisen  
rakenteen tiivistymistä arvioimalla pääasiassa kuudes- ja yhdeksäsluok-  
kalaisten Unelmakirjoitelmia vertailukohtana nuoret aikuiset. Sana-  
luokkakohtaista analyysiä teemme koko peruskoululaisten aineistosta  
eli Koulu- ja Unelmakorpuksista (aineistosta tarkemmin tämän teoksen  
johdannossa). Lisäksi on vertailulukuja sanomalehtikielestä. Rajaamme  
kompleksisuuden rakenteelliseksi (= objektiiviseksi) ja arvioimme toi-  
saalta formatiiveja – kieliopillisesti ilmaistuja merkityksiä – ja erottelem-  
me toisaalta argumenttirakenteen ja sen välttämättömät jäsenet muista.  
Toimijakeskeinen kompleksisuus on osin pääteltävissä formatiivien  
määrän kehityksellisestä noususta ja argumenttirakenteen rikastumi-  
sesta, mutta sitä emme tutki. Unelmakirjoitelmat eivät ole sisällöllisesti  
kovin kompleksisia ja suuri osa on kertovia. Heinosen (2018) mukaan  
kirjoitelmissa käsitellään keskimäärin kolmea aihtta, aiheiden vaihtelu-  
väli on 1–8. Aihteiden määrä laskee iän mukaan, ja kun sanamäärä sa-

malla kasvaa, aiheita käsitellään aiempaa perusteellisemmin. On todennäköistä, että tämä näkyy kompleksisuustekijöistä kuten sanaluokkaosuuksista. Lisäksi kirjoittajien kehitysaste vaihtelee aineistossa hyvin paljon. Tavoite on tuoda kaikki tämä vaihtelu esiin ja samalla osoittaa keskimääräinen ikäluokkatasolla hallittu kompleksisuus ja sen osatekijät. Yksilötasolla on mahdollista, että tyylillinen yksinkertaistaminen edellyttää kompleksisuuden eri asteiden hallintaa ja että taito tuottaa hyvää tekstiä edellyttää koko kielellisen repertuaarin hallintaa. Tätä kehitystä ei tosin voi todistaa synkronisella aineistolla.

Määrittelemme ensin keskeiset käsitteet, sitten analysoimme nominaalisia sanaluokkatekijöitä, jotka voisivat vaikuttaa syntaktisen rakenteen tiivyyteen ja sanamääriin. Seuraavaksi vertailemme kompleksisuutta nostavia verbaalisia tekijöitä ikäluokittain ja eri kirjoittajilla. Nominaalisten ja verbaalisten sanaluokkatekijöiden analyysi on sekä laadullista että yleisyyksiä kuvailevaa – ei varsinaista tilastollista analyysiä –, koska aineistona olevissa kirjoitelmissa on tilastoanalyysiä hankaloittavia risteäviä tekijöitä, esimerkiksi erilaisia tyylillisiä valintoja. Kuvaileva ote on välttämätön myös ryhmien sisäisen suuren vaihtelun takia. Yksi tavoite onkin tutkia, mitä tästä vaihtelusta voi kompleksisuuden osalta sanoa. Lopuksi pohditaan kompleksisuutta kehitykselliseltä kannalta ja kootaan tulokset.

## Syntaktisen rakenteen kehityksellinen kompleksisuus

Syntaktisen rakenteen kompleksisuutta voi arvioida muun muassa määrittelemällä yksinkertaisuus negaation kautta: yksinkertainen on se, mitä ei voi selvin kriteerein osoittaa kompleksiseksi.<sup>123</sup> Esimerkiksi kieliopillisen merkinnän määrällä (ks. Itkonen 2009, 190–208; Itkonen, Pajunen & Suomi 2009) voidaan erottaa kompleksisemmat yksiköt yksinkertaisemmista: edellisissä merkintää on, jälkimmäisissä ei

123 Nykyisin kompleksisuutta arvioidaan myös teknisesti tiivistämällä eli ns. kompressiomenetelmällä (ks. Sadeniemi, Kettunen, Lindh-Knuutila & Honkela 2008; Ehret & Szmracsanyi 2016). Tässä menetelmässä ei nähdä laadullisia eroja.

ole. Syntaksissa vain pääsanasta koostuva (endosentrinen) lauseke on yksinkertainen yksikkö. Vastaavasti yhden predikaation suorittava peruslause – intransitiivi- tai transitiivilause –, jossa on vain pakolliset argumentit ja yksiosainen finiittinen verbi, on yksinkertainen yksikkö (peruslausetyypeistä ks. Itkonen 2008). Yksinkertainen yksinäislause eli päälause noudattaa hallitsevan verbin argumenttirakennetta välttämättömyyden jäsenineen (ks. Pajunen 2001), kompleksinen rakenne on tästä argumenttirakenteeseen kuulumattomin jäsenen rikastettu. Rinnastus ja ketjuttaminen on tekstitasolla yksinkertaisempaa kuin alistus: alistus on aina merkitty jollain formatiivilla.

Yksinkertaiset yksiköt ovat yleensä yleisiä, mutta yleisyys on ongelmallinen kriteeri ensinnäkin, koska siitä ei ole etenkään suomen kielen osalta tarpeeksi tietoa, ja toiseksi, koska yleisyystekijät vaihtelevat muun muassa tekstilajeittain. Myös tuttuus on yksi yleisyystekijä. Äidinkielen kielenpuhujan kokemusta rakenteellisesta yleisyydestä, siis subjektiivista yleisyyttä tai tuttuutta, on tutkittu olemattomasti. Vainio, Pajunen, Hyönä ja Bertham (tulossa) vertailevat kokeellisesti morfologisesti yksinkertaisia ja kompleksisia transitiivi- ja intransitiiviverbejä sisältävien lauseiden prosessointia aikuisilla ja osoittavat, että lukijan kannalta (kohteen olotilan muutosta kielentävä) transitiivilause on perustavampi – nopeampi prosessoida – kuin vastaava intransitiivilause; ero tosin ei ole suuri. Lasten kielenoppimisessa yksinkertaiset, leksikaalisesti rajautuvat rakenteet ovat ensisijaisia; suuri osa ensiverbeistä on transitiivisia (Tomasello 2003; Diessel 2004). Toisaalta intransitiiviverbejä on pidetty niin sanottuna semanttisena primitiivinä (keskustelusta ks. esimerkiksi Wierzbicka 1980).

Syntaktisia kompleksisuuskriteereitä ovat siten ainakin lausekkeiden, lauseiden tai virkkeiden sanamäärä sekä merkinnän määrää nostava sisäinen hierarkkisuus, ei-finiittiset ainekset ja ei-lineaarinen lauserakenne. Esimerkiksi Szmrecsanyin (2004) mukaan syntaktisen kompleksisuuden mittareina lauseen sanamäärä ja toisaalta NP:n hierarkkinen rakenne eli mittarit, joiden työläys analysoijan kannalta on aivan erilainen, tuottavat kohtuullisen samanlaisia tuloksia; tulos on saatu verrattain pienellä aineistolla. Beers ja Nagy (2009; 2011) toteavat, että syntaktinen kompleksisuus eriytyy tekstin tyyppin mukaan ja että sitä täytyy

mitatakin eri tavoin. Kertovien tekstien kompleksisuutta pitäisi mitata virkkeiden sanamääristä, erittelevien tekstien taas lauseiden sanamääristä. Näin eriteltyinä syntaktinen kompleksisuus ennustaisi heidän mukaansa myös kirjoituksen kehityksellistä tasoa. Beaman (1984, 45) toteaa, että tekstin syntaktinen kompleksisuus suhteutuu alisteisten rakenteiden määrään, tyyppiin ja hierarkkiseen syvyyteen ja vastaavasti syntaktisesti yksinkertaisessa tekstissä on enemmän rinnasteisia rakenteita. Givón (1991, 347) katsoo, että alisteiset rakenteet viittaavat merkitympiin ja siten kognitiivisesti raskaampiin kategorioihin. Scontras ja muut (2014) toteavat, että ei-lokaaliset viittaussuhteet ovat vaativampia tuottaa ja prosessoida kuin lokaaliset. Lisäksi ainakin englannissa ei-lokaalisessa kontekstissa subjektiin viittaava relatiivipronomini on helpompi prosessoida – ja ilmeisesti helpompi tuottaa – kuin objektiin viittaava.

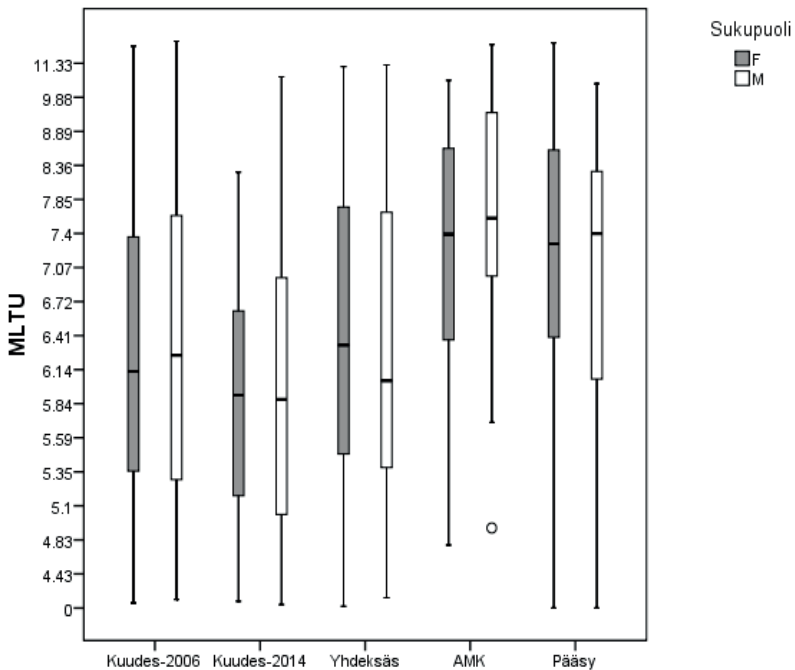
Oman analyysimme lähtökohtana on edellisen luvun antama tieto siitä, että peruskoululaisten kirjoitelmissa sekä virkkeisiin vertautuvien T-yksiköiden sanamäärä (= MLTU eli *mean length of terminal units*) että lauserakenteiden tiiviys (= CI eli kompleksisuusindeksi) kasvavat. Tavoitteemme on arvioida kompleksisuuden osatekijöitä ja sen vaihtelua tarkemmin: mistä aineksista kompleksisuus kulloinkin koostuu ja mitä niistä voi päätellä. Keskitymme kuudes- ja yhdeksäsluokkaisten kirjoitelmiin mutta kehitystrendejä osoittaaksemme otamme erityisesti sanaluokkatasolla huomioon myös muita peruskoululaisia. Vertailukohtana ovat ylioppilastutkinnon suorittaneet nuoret aikuiset. Ei kuitenkaan ole syytä ajatella, että kehitykselliset trendit sellaisenaan osoittavat esimerkiksi, mikä aikuisen lukijan kannalta on helppoa, mikä vaikeata (vrt. Taylor, Greenberg, Laures-Gore & Wise 2012). Emme myöskään tässä yhteydessä puutu varsinaisesti tekstien rakenteeseen, mikä edellyttäisi muun muassa globaalikoherenssin tutkimista.

Keskitymme virkettä muistuttaviin T-yksiköihin ja selvitämme niiden koostumusta ikäkausittain ja kehityksellisesti. T-yksikkö koostuu oletuksen mukaan suhteessa kontekstiin itsenäisestä yksinäislauseesta ja siihen liittyvistä finiittisistä ja ei-finiittisistä aineksista, jotka on merkitty alisteisiksi käyttäen alistuskonjunktioita tai relatiivi- ja interrogatiivipronomineja sekä ei-finiittisiä aineksia (T-yksiköistä ks. Schneider, Vis Dubé & Hayward 2005). Myös lausemuotoiset suorat ja epäsuorat lainat

lasketaan alisteisiksi. Substantivoituja verbejä, kuten *minen*-johdoksia, ei huomioida, vaikka niiden määrä kasvaa kehityksellisesti, koska lukumäärä sinänsä on pieni. T-yksikkö on aina kirjoittajan hierarkkiseksi kokonaisuudeksi merkitsemä yksikkö, ja se ei voi koostua löysästi ketjutuvista lauseista. Siten T-yksikkö on virkettä kehityksellisempi yksikkö, koska sen tuottajan pitää suunnitella yksikkö tiettyä merkitystä ilmaisevaksi kokonaisuudeksi. T-yksiköistä on laskettu koehenkilökohtaisesti niin keskipituus sanoina (MLTU) kuin kompleksisuusindeksi (CI eli alisteisten yksikköjen suhde T-yksilöiden kokonaisuuteen).

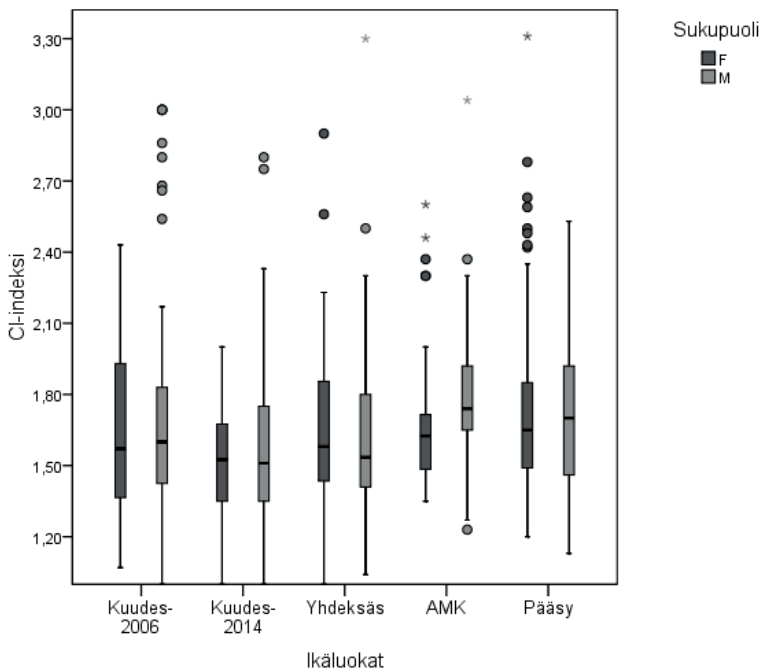
Syntaktisen rakenteen kehityksellinen kompleksistuminen ilmenee T-yksiköiden keskipituuden kasvusta. Unelmakorpuksen aineistossa MLTU:n mediaani on peruskoulun kuudesluokkalaisilla 6 ja yhdeksäsluokkalaisilla 6,2. Nuorilla aikuisilla mediaani on 7,4. Sukupuolen mukaiset erot ovat kaikissa ryhmissä aika pienet. T-yksikön keskipituus on hyvin yleisesti kehitykselliseksi arvioitu mittari (ks. esimerkiksi Nippold 2006; Nippold ja muut 2005; 2007; Wolf Nelson & Van Meter 2007; Hall-Mills & Apel 2015; Mäkinen, Loukusa, Nieminen, Leinonen & Kunnari 2014; Jagaiah, Olinghouse & Kearns 2020), jonka sanamääriä nostaa sekä lauserakenteen rikastuminen että sen tiivistyminen. Eri kieliä koskevat tulokset eivät ole suoraan vertailukelpoisia, koska esimerkiksi kieliopillinen merkitys ilmaistaan yhdessä kielessä affiksilla, toisessa sanalla. Vastaavasti lauserakenteen tiivistä mittaava kompleksisuusindeksi nousee kehityksellisesti. Sen mediaani on kuudennella 1,55, yhdeksännellä 1,58 ja nuorilla aikuisilla 1,67. Molemmissa indekseissä arvojen koehenkilökohtainen vaihtelu osoittaa (ks. kaaviot 22 ja 23), että tuloksia ei pidä tulkita liian yksioikoisesti. Lisäksi vaihtelua on myös nuorten aikuisten tavoitetasoa edustavilla kirjoittajilla. Ääriarvoja on erityisesti kuudennen pojilla ja toisaalta naisten ryhmissä. Tilastolliselta kannalta muutokset molemmissa mittareissa ovat merkitseviä.

Erittelemme tässä luvussa, mitkä eri tekijät kompleksisuusmittareita nostavat ja miten ne vaikuttavat. Aineistossa on vastakkaisia kirjoittamisstrategioita, jotka vaikeuttavat tilastollista analyysiä. Tarkastelemme seuraavaksi, miten paljon sanaluokkakohtaisten frekvenssitietojen avulla voi arvioida sitä, mikä nostaa T-yksiköiden sanamääriä tai kompleksisuusindeksin arvoja. Sanaluokista tarkasteltavina ovat adjektiivit (kuvauk-



Kaavio 22. T-yksiköiden keskipituus sanoina (MLTU) ja hajonta Unelmakorpuksessa.

sen rikastaminen), substantiivit ja pronominit (viittaussuhteet) sekä verbit, koska oletettavasti juuri näiden sanaluokkien määrät vaikuttavat joko T-yksiköiden keskipituuteen tai kompleksisuusindeksiin ja molempiinkin. Adjektiiveilla ilmaistaan pääasiassa peruslauseen ytimen ulkopuolisia aineksia; substantiivien, pronominiin ja verbien keskinäiset suhteet liittyvät merkityksen spesifioimiseen ja ilmaisun tiivistymiseen. Lisäksi tutkimme alisteisten rakenteiden yleisyyttä paitsi niiden määristä myös funktioittain eriteltyinä. Arvioimme tuloksia sekä ikäryhmitäin, sukupuolen mukaan että kirjoittajakohteisesti ja suhteutamme niitä mahdollisuuksien mukaan myös T-yksiköiden keskipituuksiin ja kompleksisuusindeksiin sekä tarvittaessa muihinkin indekseihin. Vertaamme peruskoululaisia nuoriin aikuisiin. Sanaluokkatietoja on kaikista Unelmakirjoitelmista, mutta T-yksiköt ovat muuttujana vain



Kaavio 23. Kompleksisuusindeksin mediaani ja hajonta Unelmakorpuksessa.

Unelmakorpuksessa, jossa kuudesluokkalaisten ja nuorten aikuisten kirjoitelmia kumpiakin on yli kolme sataa, muissa peruskouluryhmissä on noin 150 kirjoitelmaa.

## Nominaalisten sanaluokkien vaikutus sanamäärään ja rakenteen kompleksisuuteen

Yksittäiset sanaluokat voivat vaikuttaa toisaalta T-yksikön keskipituuteen, toisaalta kompleksisuusindeksiin (vrt. Graesser, McNamara, Louwerse & Cai 2004). Tarkastelemme ensin nominaalisia sanaluokkia ja pyrimme arvioimaan niiden vaikutusta mainittuihin indekseihin ja osaksi tekstin rakenteeseen. Tarkastelemme ohessa myös sanaluokkien laadullisia ominaisuuksia (vrt. Price & Jackson 2015, 283). Alkualetus on, että

nominaaliset sanaluokat – kuten adjektiivit ja pronominit – nostavat keskipituuksia mutta eivät kompleksisuusindeksiä. Tyypillisesti juuri ei-argumenttiasemissa käytettävien sanaluokkien edustajat kuten adjektiivit ja adverbit näyttävät lisääntyvän kehityksellisesti, vaikka luonnollisesti niiden leksikaalinen diversiteetti ja funktioiden erilaistuminen osoittavat kehityksen yksityiskohtaisemmin (ks. tarkemmin Kuisma 2011; Pajunen 2012). Substantiivien määrästä ja substantiivien ja pronominiin keskinäisestä suhteesta voi arvioida tekstin yleisempää luonnetta esimerkiksi asteikolla konkreettinen–abstrakti. Substantiivien vaikutus kompleksisuuteen voi kuitenkin olla kahtalainen muun muassa, koska ne toimivat paitsi finiittiverbien myös ei-finiittisten verbien argumentteina. Partisiippeja ja infinitiivejä käsittelemme lähinnä verbiluokan alakategorioina. Adverbien funktio teksteissä on moninainen ja erityisesti temporaaliset ja lokaaliset adverbit lisääntyvät kehityksellisesti. Niiden funktiona on kehystää kerrontaa ajallisesti ja paikallisesti. Ne siis voisivat nostaa sanamäärää, mutta sivuutamme ne tässä yhteydessä.

#### ADJEKTIIVITIHEYS

Adjektiivit spesifioivat tarkoitteita rajaamalla niiden alaa ja samalla lisäävät tekstin kuvailevuutta. Adjektiivien semanttiset alaluokat eivät näiden perusfunktioiden osalta sanottavasti poikkea toisistaan mutta kehityksellisesti niissä on eroa. Muutosta tapahtuu erityisesti relationaalisista adjektiiveista absoluuttisiin, esimerkiksi arvottavista (*kiva, hyvä*) eksakteihin (*kolmiosainen, puinen*) (ks. Kuisma 2011). Unelmakirjoitelmista oletus on, että joillakin kirjoittajilla juuri adjektiivit nostavat T-yksiköjen keskipituutta. Tähän viittaa muun muassa se, että käyttö voi olla lähes kliseistä. Esimerkiksi nuoret aikuiset luonnehtivat ilmaa, paikkakuntaa, päivää, salaattia, taloa ja vettä raikkaaksi (vrt. *raikas juoma, luonnonkaunis ja raikas Kaustinen*). Erityisesti attribuuttiasemaiset nostavat keskipituuksia; predikoivat adjektiivit ovat kopulalauseen pakollisia jäseniä eivätkä tuo ylimääräistä lauseeseen. Silti myös kopulalauseet lisääntyvät kertovassa tekstissä kehityksellisesti (ks. Sintonen 2011). Alakouluikaisilla yleisimmät viisi adjektiivia (*kiva, oma, hyvä, uusi ja iso*) kattavat 24,3 prosenttia adjektiiviesiintymistä ja yhden esiintymän adjektiivit eli HL-adjektiivit (kuten *kapea, karmiva, karvainen, kastanjanruskea*) 10 pro-



senttia. Nuorilla aikuisilla yleisimmät viisi adjektiivia (*hyvä, pieni, oma, uusi ja suuri*) kattavat 16 prosenttia adjektiiviesiintymistä ja HL-adjektiivit (*kahvintuoksuinen, kaikenikäinen, kaksitoistasylinterinen, kalpea*) 16,5 prosenttia. Sanomalehtitekstissä (SS2002, Pajunen & Virtanen 2002) samat suhteet ovat 14 prosenttia ja 5,2 prosenttia ja yleisimmät adjektiivit ovat *pieni*-sanaa lukuun ottamatta samoja kuin aikuisilla. Adjektiivien diversiteetti nousee kehityksellisesti, mutta korpus sinänsä on liian pieni, jotta luvut vielä vakiintuisivat. Adjektiivimääriä arvioitaessa mukana ovat kaikki leksikaaliset adjektiivit mutta myös pronominaaliadjektiivit (esimerkiksi *sellainen*), joiden määrä kuitenkin on hyvin pieni.

Adjektiiviesiintymien määrä eli adjektiivitiheys on laskettu kirjoitelmi- en kokonaissanamäärästä koehenkilöittäin ja ikäryhmittäin sekä sukupuolittain (ks. taulukko 60). Alakoulun tytöillä on adjektiiveja selvästi enemmän ja monipuolisemmin käytössä kuin pojilla; yläkoulussa ero on vielä suurempi. Alakoulun ensimmäisellä ja toisella luokalla on vielä paljon kirjoitelmia, joissa ei ole lainkaan adjektiiveja. Näiden osuus laskee nopeasti ja alakoulussa adjektiivien määrä nousee eniten ryhmässä 1–5 prosenttia sanamäärästä. Tosin pojilla on pitkään kokonaan adjektiivittomia kirjoitelmia. Kuudesluokkalaisilla tytöillä tyypillinen määrä on 6–10 prosenttia kokonaissanamäärästä ja pojilla 1–5 prosenttia. Suhde on yhdeksäsluokkalaisilla samantapainen mutta hieman korkeampi. Peruskoululaisista reippaasti alle puolella adjektiiviesiintymiä on 6–10 prosenttia sanoista, nuorista aikuisista kaksi kolmasosaa yltää tähän määrään. Viidenneksellä naispuolisista aikuisista adjektiiveja on jopa 15–20 prosenttia sanaesiintymistä. Pääsykoeaineistossa adjektiiveja on käytetty jonkin verran enemmän kuin ammattikorkeakouluaineistossa. Savon Sanomissa adjektiivitiheys esiintymistä on 8 prosenttia, joten korkeimmat Unelma-adjektiiviprosentit ovat leimallisen korkeita. Vastaavasti korkeat adjektiivien käyttämättömyysprosentit paria ensi kouluvuotta lukuun ottamatta ovat kehityksellisesti merkitseviä.

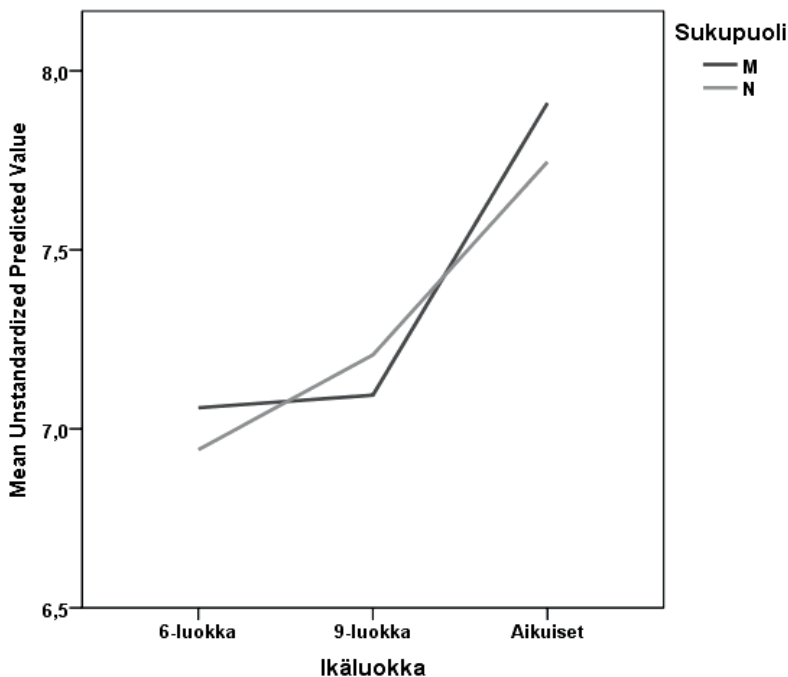
Arvioimme sanaluokkakohtaista esiintymistiheyttä myös laskemalla kunkin sanaluokan tiheyden korrelaatiota pääosin T-yksikön keskipituu- teen ja kompleksisuusindeksin arvoihin. Tulokset esitetään ikäluokittain ja sukupuolittain eriteltynä kaaviomuodossa (ks. kaaviot 24 ja 25 adjektiiveista). Tarkastelu keskittyy kuudensiiin, yhdeksänsiiin ja nuoriin ai-

Taulukko 60. Adjektiivitiheys Unelmakirjoitelmissa.

Ikäluokka / Kirjoitelma %	0	1–5	6–10	11–15	16–20	21–25 %
1.–2. F	26,3	40,9	18,9	8,7	2,9	2,2
1.–2. M	38,6	35,7	14,3	5,7	3,6	2,1
3.–4. F	6,1	50,3	31,3	10,4	1,8	0
3.–4. M	16,7	44,0	28,6	7,7	3,0	0
Viides F	2,7	65,3	24,0	6,6	1,3	0
Viides M	13,9	55,7	24,0	5,1	1,2	0
Kuudes F	2,5	44,1	45,9	5,7	1,9	0
Kuudes M	4,6	56,3	33,1	5,3	0	0,7
Yhdeksäs F	2,3	30,2	60,5	7,0	0	0
Yhdeksäs M	2,8	50,0	44,4	2,8	0	0
Aikuiset F	0	10,0	67,1	22,5	0,4	0
Aikuiset M	0	12,4	74,6	11,3	1,7	0

kuisiin, joilla adjektiivitiheys on jo vakiintuneempi kuin nuorimmilla alakoululaisilla. Nämä kahden kielellisen ja kahden demografisen muutujan kaaviot näyttävät ikä- ja sukupuolierot aika suoraan mutta ikäluokakokohtaiset vaikutukset tulevat tarkemmin esiin vasta, kun suhteutetaan tiheysmuuttujien korrelaation kautta syntyvät arvot tiheysmuuttujien mediaaneihin. Toisin sanoen vertailu ei nyt perustu yksin kuvaan vaan y-akselin ja sanaluokkamediaanin arvoihin eli siihen, onko kahden kielellisen muuttujan korrelaatiosta laskettu tiheysarvo yli vai ali sen mediaanin. Adjektiivimediaanit ovat kuudennella 6,1 (tytöt) ja 4,9 (pojat); yhdeksännellä 7,5 ja 5,6 sekä aikuisilla 9,3 ja 8,8.

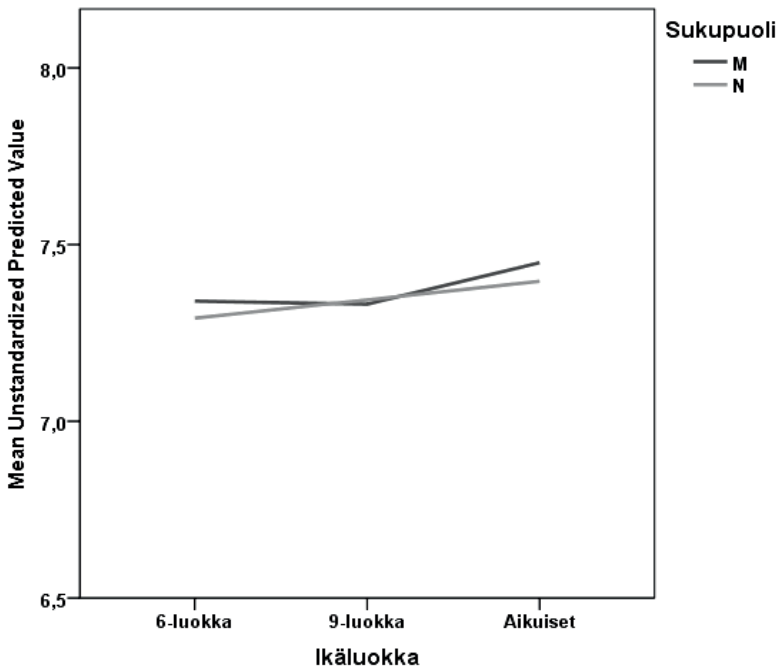
Adjektiivit nostavat kuudensilla T-yksikön sanamääriä, pojilla hieman enemmän kuin tytöillä (ks. kaavio 24). Tästä voisi päätellä, että adjektiivitiheyden nousu erityisesti pojilla on kehityksellinen, koska juuri alakoulussa on paljon adjektiiveja käyttämättömiä poikia. Muissa ikäryhmissä vaikutus on käänteinen yhdeksännen poikia lukuun ottamatta. Tämä johtunee siitä, että kehittyneimmillä kirjoittajilla on adjektiivien käytössä niin suurta vaihtelua, että vaikutus ei tule ryhmätasolla näkyviin. Nuorimmilla kirjoittajilla adjektiivien korkea määrä liittyy myös lauserakenteen kompleksisuuteen (ks. kaavio 25). Tämän voi katsoa viittaavan siihen, että kuudesluokkalaisten keskivertotekstiä rikastetaan adjektiiveilla ja kompleksisilla lauserakenteilla, lähinnä 1. infinitiivillä ja sivu-



Kaavio 24. Adjektiivitiheys ja T-yksikköjen sanamäärä. Mediaani: kuudes tytöt 6,1 ja pojat 4,9, yhdeksäs 7,5 ja 5,6 ja aikuiset 9,3 ja 8,8.

lauseilla. Nuorilla aikuisilla adjektiivitiheys ei enää nosta kompleksisuutta. Yhdeksän tytöt muistuttavat adjektiivitiheydessä nuoria aikuisia. Kaikkiaan adjektiivien käytöllä on indekseihin eniten vaikutusta nuorimmilla kirjoittajilla.

Adjektiivien ja substantiivien suhteellinen osuus nousee ikäluokittain tasaisesti ja samalla substantiivia rajaavien adjektiivien – modifioinnin – määrä nousee. Multikollineaarisuus estää näiden muuttujien korrelaation laskemisen. Finiittiverbien määrä on nuoremmilla kirjoittajilla korkeampi kuin vanhemmilla, mutta peruskoulun kuudennella ja etenkin pojilla adjektiivitiheys ja finiittiverbien määrä kulkevat käsi kädessä. Aikuisilla ja yhdeksän tytöillä korrelaatio on negatiivinen: adjektiiveja on keskimääräistä vähemmän, jos finiittiverbien määrä on korkea. Sukupuolittain ero on suurin yhdeksäsluokkalaisilla.



Kaavio 25. Adjektiivitiheys ja kompleksisuusindeksi. Mediaani: kuudes tytöt 6,1 ja pojat 4,9, yhdeksäs 7,5 ja 5,6 ja aikuiset 9,3 ja 8,8.

Esimerkki 3 edustaa aikuisnäytettä, jossa on paljon adjektiiveja, esimerkki 4 on päinvastainen. Vaikutelma on molemmissa tylillinen. Esimerkissä 3 sävy on rento. Vain yksi adjektiiveista predikoi. Esimerkissä 4 lauseet ovat lyhyitä ja sävy toteava. Esimerkissä 5 adjektiivien määrä on melko korkea mutta monet ovat predikoivassa funktiossa ja eivät kasvata keskipituutta samalla tavalla kuin näytteessä 3. Tosin myös predikoivat lauseet luovat tekstiin kuvailevuutta.

Adjektiiveja 16,7 % sanoista, MLTU = 14, kompleksisuusindeksi = 2,1:  
3) *Turhanpäiväisen naturalin jälkeen olisi aika siirtyä lounaalle. Lounaaksi valikoituisi grillattua härkää ja ehkä tilkkanen hieman nahkaista ja kielellä viipyilevää punaviinia. Täyttävän ja maittavan lounaan jälkeen olisi aika siirtyä isommalla porukalla saunomaan, – –.* (AMKM33)

Adjektiiveja 4,7 % sanoista, MLTU = 4,9, kompleksisuusindeksi = 1,2:

4) *Minä herään myöhään aamupäivällä. Myöhäisellä aamupäivällä tarkoitan ennen puoltapäivää, mutta jälkeen kymmenen. Linnut eivät laula. Syön aamiaisen. En mitään outoa tai kummallista; muroja tai puuroa sekä lasillisen maitoa. Lähden käymään kaupassa.* (AMKM58)

Adjektiiveja 10,8 % sanoista, MLTU = 6, kompleksisuusindeksi = 1,4:

5) *Unelmieni päivääni voisi kuvitella arkiseksi, minulle se kuitenkin on täynnä niitä asioita mistä nautin ja mistä saan voimaa työssä käyntiin ja opiskeluuni. Tällainen yhteinen vapaapäivä on harvinainen meidän taloudessa, joten koen sen sitäkin arvokkaammaksi.* (AMKF92)

Adjektiivien määrä nousee kehityksellisesti. Sukupuolten välinen ero pysyy samansuuntaisena alakoulusta lähtien siten, että naiskirjoittajilla on aina enemmän adjektiiveja kuin mieskirjoittajilla. Vaikutus MLTU:hun on ryhmätasolla vähäinen nuorimpia lukuun ottamatta mutta yksilötasolla vaihteleva. Nuorilla adjektiivit liittyvät kompleksiseen lauserakenteeseen mutta vanhemmilla eivät.

## SUBSTANTIIVITIHEYS

Substantiiveista valtaosa on kaikilla kirjoittajilla appellatiiveja. Alakoululaisilla kymmenen yleisintä substantiivilekseemiä kattaa kaikista 17 prosenttia ja 30 yleisintä jo kolmanneksen. Aikuisilla kymmenen yleisintä kattaa substantiiviesiintymistä saman verran mutta 30 yleisintä vain neljänneksen. Diversiteetti siis kasvaa selvästi iän myötä (ks. myös Laine-Leinonen 2013); aikuisilla suurin osa substantiiveista on HL-lekseemejä, lapsilla joka kymmenes substantiivi. Alakoululaisilla substantiiveista kymmenisen prosenttia on propreja mutta nuorilla aikuisilla vain nelisen prosenttia. Kirjoitelman aihe vaikuttaa suuresti yleisyyteen, siten kaikilla substantiivit *päivä* ja *unelma* ovat yleisimpien joukossa. Alakoulukäisillä yleisiä substantiiveja ovat lisäksi *äiti*, *kaveri*, *koti*, *koira* ja *hevonen*, aikuisilla *aika*, *elämä*, *hetki* ja *ystävä*. Aikuisilla siis abstraktit käsitteet nousevat konkreettisten rinnalle. Ravidin (2006) heprean kielestä

tehdystä tutkimuksesta kävi ilmi, että abstraktien substantiivien osuus kasvaa vähitellen ja muutos on selvintä nuoruusiässä; narratiiveissa abstraktistuminen on hitaampaa kuin erittelevissä teksteissä. Substantiiveista *minen*-johdokset ovat selvästi abstrakteja ja myös sikäli nuorille kirjoittajille hankalia, että niiden käyttö edellyttää kykyä tiivistää ja fokusoida peruslauseen sisältämää informaatiota. *Minen*-substantiiveja on kaikenikäisten kirjoitelmissa, mutta ne ovat kaikissa kirjoittajaryhmissä pääasiallisesti HL-lekseemejä. Alakoulun luokilla 1.–5. *minen*-johdoksia on 0,07 prosenttia kokonaissanamäärästä, kuudennella 0,2, yläkoulun luokilla 0,3 ja nuorilla aikuisilla 0,6, toisin sanoen myös abstraktien *minen*-substantiivien määrä nousee kehityksellisesti. Bermanin, Nayditzin ja Ravidin (2011, 174) mukaan abstraktien substantiivien osuus lisääntyy lasten ja nuorten kertomuksissa, jos kirjoittajan sosio-ekonominen tausta on korkea. Suomen aineisto ei ole sosioekonomisesti tasakoosteinen.

Substantiiveiksi on Unelma-aineistossa koodattu appellatiivit ja propriit, jotka eivät ole muodoltaan kiteytyneitä, toisin sanoen taipuvat periaatteessa sijaituvuuden eri muodoissa ja toimivat lauseessa kaikissa argumenttiasemissa. Jo NTVR-indeksi (eli substantiivien ja verbien suhde) ja leksikaalinen tiheys osoittivat edellisessä luvussa substantiivien määrän kehityksellisen nousun. Substantiiveja voidaan tarkastella myös suhteessa kokonaissanamäärään eli substantiivitiheytenä (*nominal density*). Ravidin (2006, 792) mukaan substantiivitiheys kasvaa erityisesti nuoruusvuosina ja korreloi leksikaalisen diversiteetin kanssa; teksti on myös syntaktisesti sitä kompleksisempaa mitä enemmän siinä on substantiiveja ja adjektiiveja. Suomen aineistossa adjektiivimäärä ei vaikuta aikuisilla lausetason kompleksisuuteen vaan lisää sanoja lauseketasolla ja T-yksikössä. Substantiiviosuudet on laskettu vajaasta 2000 kirjoitelmasta, mutta Ravidin tutkimuksessa on kaikkiaan vain 80 kirjoitelmaa, joten voi olla, että erot tasaantuvat suuremmassa aineistossa.

Substantiivitiheys vaihtelee kirjoitelmien kokonaissanamäärästä 12 prosentista 46 prosenttiin, mutta mediaaneista vaihtelu ei sanottavasti näy: alakoulussa mediaani on 30 prosenttia ja nuorilla aikuisilla 31,5 prosenttia. Muutos näkyy kuitenkin yksilövariaation kautta. Kuudesluokkalaisilla substantiiveja on kirjoitelmissa lähes yhtä usein 20–30 prosenttia kokonaissanamäärästä kuin 30–40 prosenttia, ja iän myötä jälkimmäinen

ryhmä kasvaa nopeasti (ks. taulukko 61). Nuorimmilla alakoulun kirjoittajilla sekä hyvin pienet että hyvin suuret substantiivimäärät liittyvät ongelmateksteihin. Näitä poikkeavia arvoja on alimmilla luokilla 20–40 prosentissa kirjoitelmista. Aikuisilla miehillä substantiivien osuus selvästi suurempi kuin naisilla. Ero näkyy jo yhdeksäsluokkalaisilla aivan Ravidin tulosten mukaisesti. Savon Sanomissa substantiivitiheys on 45 prosenttia kokonaissanamäärästä. – Todettakoon ohimennen, että korkea substantiivimäärä korreloi positiivisesti nuorten aikuisten aineistossa pääsykoepisteisiin, vastaavasti korkea pronomini- ja finiitti-verbien osuus korreloi negatiivisesti.<sup>124</sup> Toisin sanoen korkeilla pisteillä pääsykokeen läpäissyt on kirjoittanut Unelmakirjoitelman, jossa substantiivimäärä on korkea mutta ei-kompleksiseen tekstiin viittaavien finiittiverbien määrä matala.

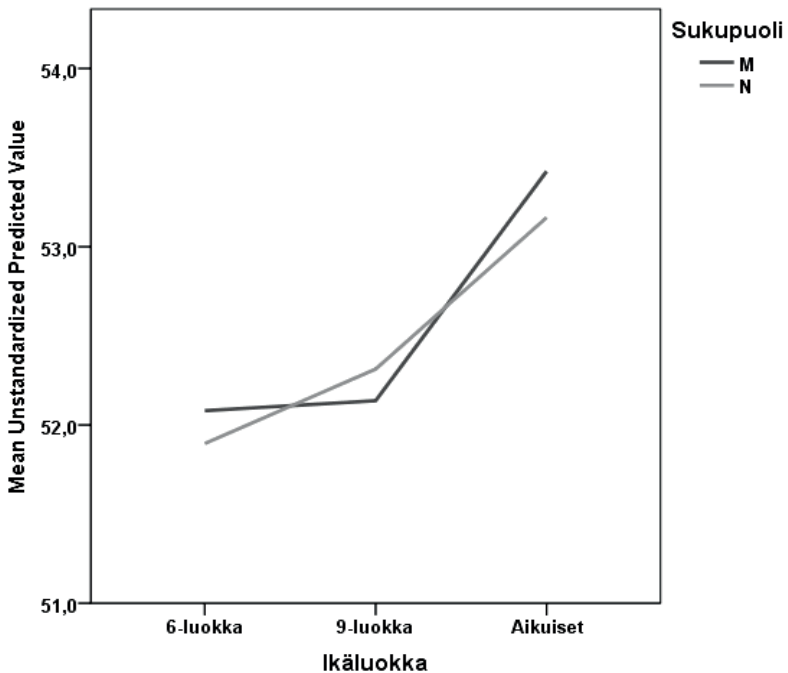
Taulukko 61. Substantiivitiheys Unelmakirjoitelmissa.

Ikäluokka / Kirjoitelma %	1–10	11–20	21–30	31–40	Yli 40 %
1.–2. F	1,4	8,7	37,2	34,3	18,2
1.–2. M	1,4	8,6	32,8	27,1	30
3.–4. F	0	6,7	49,1	36,2	8,0
3.–4. M	0	10,7	39,9	35,7	13,7
Viides F	0	4	41,3	49,3	5,3
Viides M	0	3,8	46,8	38,0	11,4
Kuudes F	0	1,8	46,9	45,0	6,2
Kuudes M	0	5,9	41,2	43,1	9,8
Yhdeksäs F	0	2,2	45,4	45,4	6,8
Yhdeksäs M	0	4,2	33,3	54,2	8,3
Aikuiset F	0	2,3	37,2	60,4	0
Aikuiset M	0	1,7	26,3	68,4	3,5

Substantiivien ja T-yksikön sanamäärän suhdetta ei pysty suoraan laskemaan, joten tarkastelemme substantiivitiheyttä substantiivi–verbisuhteen (= NTVR) kautta. Laskentatapa on sama kuin edellä adjektiivien yhteydessä. NTVR-suhde kuvaa tekstin substantiivipitoisuutta suhteessa verbien määrään, ja se on kertovassa tekstissä oletuksen mukaan hie-

124 Osa nuorista aikuisista kirjoitti Unelmakirjoitelman yliopiston pääsykokeen yhteydessä. Heistä on monenlaisia vertailutietoja.

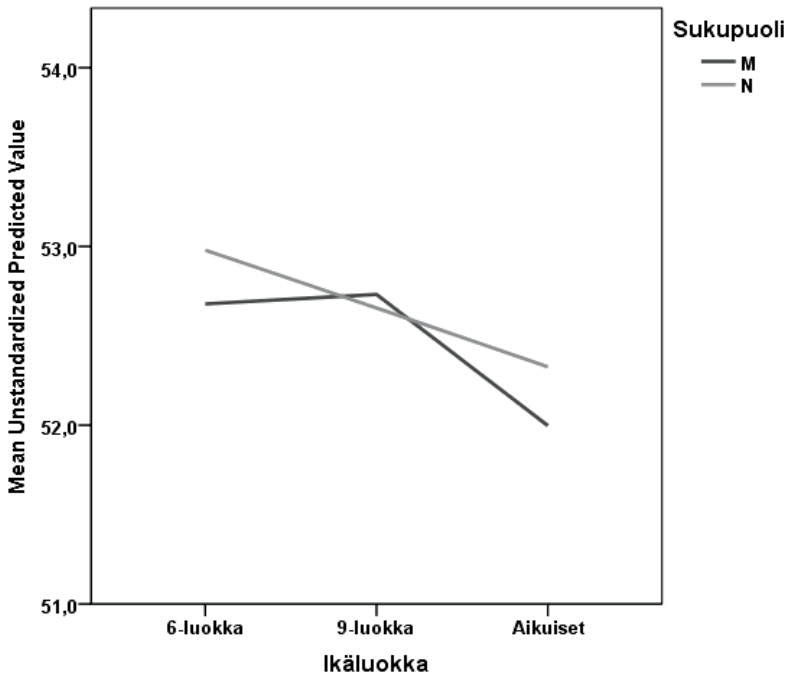
man yli yhden suhdetta yhteen. NTVR ja T-yksikön sanamäärien suhde on negatiivinen mieskirjoittajilla – eli korkea substantiivimäärä laskee T-yksikön sanamääriä –, mutta naiskirjoittajilla vaikutus on positiivinen vaikka vähäinen (ks. kaavio 26). NTVR-mediaani on kuudennella tytöillä 51,2 ja pojilla 52,6, yhdeksännellä 51,5 ja 53 sekä aikuisilla 53,4 ja 54,5.



Kaavio 26. NTVR-suhde ja T-yksikköjen sanamäärä. NTVR-mediaani: kuudes 51,2 ja 52,6, yhdeksäs 51,5 ja 53 sekä aikuiset 53,4 ja 54,5.

Peruskoululaisilla NTVR on nouseva kompleksisessa tekstissä yhdeksännnen poikia lukuun ottamatta (ks. kaavio 27). Aikuisilla NTVR-suhde laskee kompleksisessa tekstissä. Ero voi johtua komplementtilauseiden edustumisesta joko sivulauseena tai lauseenvastikkeena. – NTVR-mediaani on kuudensilla 51,2 ja 52,6, yhdeksänsillä 51,5 ja 53 sekä aikuisilla 53,4 ja 54,5.





Kaavio 27. NTVR ja kompleksisuusindeksi. NTVR-mediaani: kuudes 51,2 ja 52,6, yhdeksäs 51,5 ja 53 sekä aikuiset 53,4 ja 54,5.

Esimerkit 6–9 kuvaavat taulukon 61 jaottelun mukaan vaihtelevaa substantiivimäärää vähästä keskimääräiseen ja paljoon. Kaikki esimerkit ovat kuudesluokkalaisilta. Jos substantiivien määrä on hyvin pieni (ks. esimerkki 6), teksti on selvästi heikkoa. Näytteissä 7–9 pelkkä substantiivien määrä ei riitä jakamaan niitä tason mukaan, mutta kun substantiivien suhteellinen osuus nousee, teksti joko irtautuu minä-kerronnasta (8) tai tarkoitteita tuodaan tekstiin enemmän ja samalla kerronta tulee täsmällisemmäksi (9).

Vähän substantiiveja (10–20 % sanamäärästä): MLTU = 3,4, kompleksisuusindeksi = 1, leksikaalinen tiheys = 59,3, substantiivi-verbi-indeksi = 33,3:

6) *Mä lennän ja törmään puuhun ja sit oon koomassa ja sit mä he-  
rään ja voitan lotossa ja ostan jotakin turhaa.* (6M4)

Substantiiveja (20–30 % sanamäärästä): MLTU = 6,6, kompleksisuusindeksi = 1,6, leksikaalinen tiheys = 73,8, substantiivi-verbi-indeksi = 49,5:

7) *Minun mielestäni unelmieni päivä olisi tällainen: vesihiihtäisin, saisin ratsastaa hepalla ja olla serkullani yötä. Serkkuni nimi on Anna. Hänellä on kauniit ruskeat hiukset ja siniset silmät. Hänen perheensä asuu omakotitalossa, jossa on valtava takapiha. Leikimme erittäin usein ja erityisesti lomilla.* (6F14)

Keskimääräisesti substantiiveja (30–40 % sananmäärästä): MLTU = 6,1, kompleksisuusindeksi = 1,6, leksikaalinen tiheys = 67,5, substantiivi-verbi-indeksi = 53,9:

8) *Unelmien päivä olisi päivä Helsingissä – Astun junasta pois Helsingin yhdellä rautatieasemalla. Näen paljon ihmisiä, mutta yksi on yli muiden. Hän seisoo kolmen muun miehen kanssa asemalla. Jotkut ihmiset pyytävät niiltä nimmareita.* (6F40)

Hyvin paljon substantiiveja (yli 40 %): MLTU = 4,7, kompleksisuusindeksi = 1,2, leksikaalinen tiheys = 87,8, substantiivi-verbi-indeksi = 62,1:

9) *Unelmieni päivän voisin viettää Hawajilla kolme viikkoa. Makaisin auringon alla päivät pitkät! Ihailisin eksoottisia maisemia. Se olisi minulle kuin unelmien täyttymys, paratiisi. Seuraavana päivänä voisin shoppailla päivän ilmaiseksi.* (6F5)

#### PRONOMINITIHEYS

Unelmakirjoitelmien pronomini-jakauma persoona-, demonstratiivi- ja indefiniittipronomineihin sekä lausetta aloittaviin relatiivisiin ja interrogatiivisiin pronomineihin muuttuu iän mukaan; laskelmissa ei ole mukana pronominaaleja. Persoona- ja demonstratiivipronominit ovat hallitsevia ensimmäisen ja toisen luokan kirjoitelmissa, *minä*-pronomini menettää asemiaan kolmannella ja neljännellä, viidennellä ja kuuden-

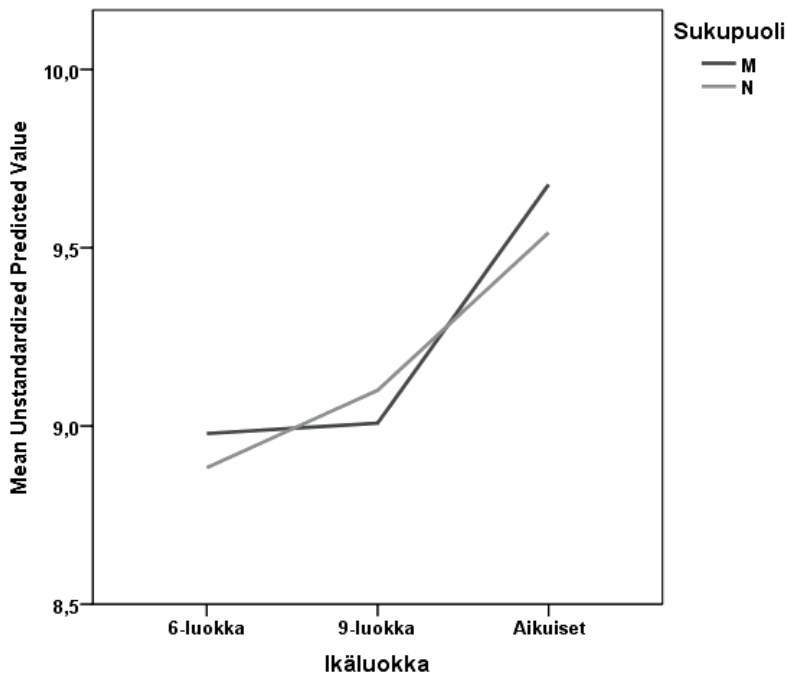
nella *hän*- ja *se*-pronominini ovat yleisimmät. Myöhemmin nousee erityisesti indefiniittipronominien luokka yleisimmäksi. Näiden pronominien vaikutusalana on pääsanalan alan tietynlainen kvalifiointi, joten ne muistuttavat tässä suhteessa adjektiiveja. Relatiivi- ja interrogatiivipronominien osuus kasvaa ensimmäisen luokan noin 8 prosentista tasaisesti aikuisten runsaaseen 15 prosentin osuuteen koko pronominiiluokasta. Lehtiteksteissä lausetta aloittavien pronominien osuus on vielä runsaampi eli 19 prosenttia pronominiiluokasta. Sekä indefiniittisten yksilöivien ja kvantifioivien pronominien että sivulauseita aloittavien pronominien määrän kasvu on selvästi kehityksellistä, mutta muutokset persoonapronomineissa liittyvät kertovaan tekstiin ja kertojan roolien monipuolistumiseen. Persoonapronomininit eivät oletettavasti vaikuta tekstin kompleksisuuteen.

Peruskoulun pojilla on kirjoitelmissaan vähän vähemmän pronomineja kuin tytöillä mutta aikuisryhmässä tilanne on päinvastainen (taulukko 62). Tavanomaisin määrä asteikolla vähästä paljooon on ensimmäistä ja toista luokkaa lukuun ottamatta 6–10 prosenttia sanamäärästä. Pienin osuus kasvaa iän myötä, joskin mieskirjoittajilla hieman eri tavoin kuin naiskirjoittajilla. Naiskirjoittajilla kehityslinja on melko tasainen. Kirjoitelmia on pienimmässä pronominiyhmässä vähemmän kirjoittajien iän noustessa, ja toiseksi suurin määrä vastaavasti nousee. Aikuiset mieskirjoittajat – melko hyvät mieskirjoittajat – eivät sanottavasti poikkea tästä linjasta, joskin naiskirjoittajien tekstejä on enemmän luokassa 6–10 prosenttia sanamäärästä ja mieskirjoittajien luokassa 11–15 prosenttia sanamäärästä. Sen sijaan peruskoulun pojat poikkeavat kuvasta: heillä on sekä pieniä että suuria pronominiääriä enemmän kuin muilla kirjoittajilla. Mediaani pysyttelee kaikilla muilla paitsi yhdeksännnen pojilla melko samoissa lukemissa. Pronominitiheys on sanomalehtikielessä 5,3 prosenttia. Ero Unelmakirjoitelmiin liittyy pääosin tekstin luonteeseen, mutta heikoilla kirjoittajilla pronominitiheys voi nousta myös löysien viitesuhteiden tai pronominiketjujen vuoksi.

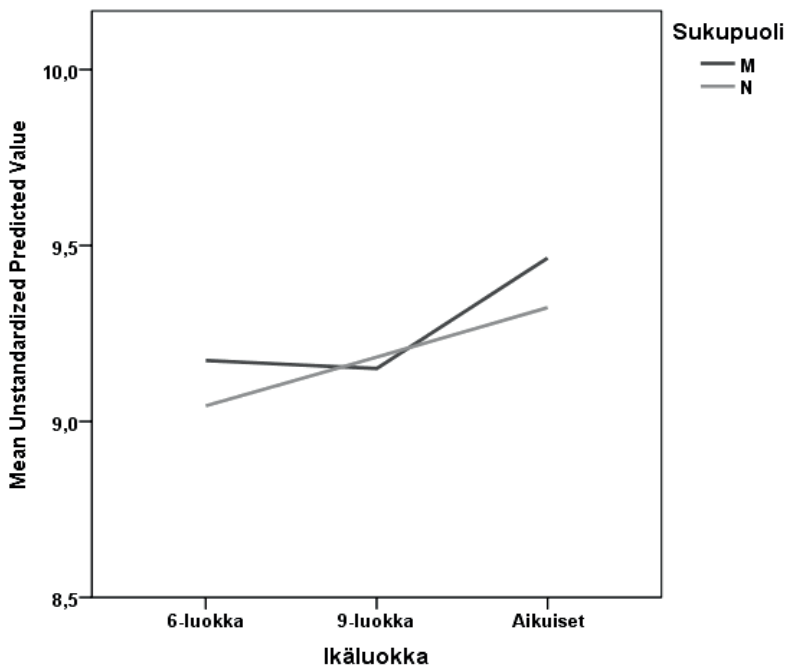
Pronomininit nostavat MLTU:ta kaikissa muissa ryhmissä paitsi kuudennen luokan tytöillä (ks. kaavio 28). Aikuisilla nousu on hyvin selvä mutta miehillä suurempi kuin naisilla. Taustalla on todennäköisimmin indefiniittisten, relatiivisten ja interrogatiivisten pronominien vaikutus.

Taulukko 62. Pronominitiheys Unelmakirjoitelmissa.

Ikäluokka Kirjoitelma %	0-5	6-10	11-15	16-20	Yli 20 %
1.-2. F	25,5	26,3	21,9	13,8	12,4
1.-2. M	38,6	22,1	18,6	7,1	13,6
3.-4. F	16,5	32,5	35,0	7,9	7,9
3.-4. M	18,4	28,0	28,6	18,4	6,5
Viides F	9,3	37,3	44,0	6,6	2,6
Viides M	12,6	49,5	22,8	15,2	0
Kuudes F	11,2	48,7	33,7	5,6	0,6
Kuudes M	14,5	48,0	24,3	11,8	1,3
Yhdeksäs F	7,9	55,7	29,5	3,4	3,4
Yhdeksäs M	22,2	47,2	19,4	9,7	1,4
Aikuiset F	6,1	57,2	30,9	5,6	0
Aikuiset M	7,2	52,6	36,1	4,1	0



Kaavio 28. Pronominitiheys ja T-yksikköjen sanamäärä. Pronominimediaanit: kuudes 8,8 ja 8,8, yhdeksäs 8,7 ja 7,8 sekä aikuiset 8,9 ja 8,8.



Kaavio 29. Pronominitiheys ja kompleksisuusindeksi. Pronominimediaanit: kuudes 8,8 ja 8,8, yhdeksäs 8,7 ja 7,8 sekä aikuiset 8,9 ja 8,8.

Jos tekstissä on paljon pronomineja, myös sen kompleksisuus on noususuuntainen (ks. kaavio 29), mutta tytöillä vähemmän kuin pojilla. – Pronominimediaanit ovat kuudensilla 8,8 ja 8,8, yhdeksänsillä 8,7 ja 7,8 sekä aikuisilla 8,9 ja 8,8.

Esimerkkitekstit 10–11 ovat yhdeksänneltä luokalta ja osoittavat, miten erilaiset tekijät vaikuttavat muuttujan arvoihin. Kirjoitelmassa 10 pronominiäärää laskee verbiin koodattu *minä*-kerronta, kirjoitelmassa 11 sitä nostaa muun muassa kvanttorien runsaus. Tekstin laatimistapa vaikuttaa siis ratkaisevasti.

Vähän pronomineja (2,2 % sanamäärästä), leksikaalinen tiheys = 76,4, NTVR-indeksi = 56,4, MLTU = 5,9, kompleksisuusindeksi = 1,5:

10) *Unelmapäivä alkaisi sillä että heräisin pirteänä ja hyvin nukku-  
neena. Aurinko paistaisi ikkunasta ja palvelia toisi aamupalan sän-  
kyyn ja ilmoittaisi että olen voittanut lotossa jättipotin. Syömisen jäl-  
keen laittaisin mukavat vaatteet päälleni ja menisin aamu-uinnille ulos  
uima altaaseen. (9F15)*

Suhteessa paljon pronomineja (16,4 % sanamäärästä), leksikaali-  
nen tiheys = 65,7, NTVR-indeksi = 54,1, MLTU = 9,6, kompleksisi-  
suusindeksi = 1,8:

11) *Unelmieni päivä olisi sellainen että ei olisi koulussa sääntöjä koko  
päivänä ja että kaupassa saisi hakea kaikkea ilmaseks. Sitten Saksa  
pyyhittäis maailmankartalta ja että kaikki Aasialaiset teloitettaisiin ei  
sillä että mulla ois jotain aasialaisia vastaan mutta se on ärsyttävä  
maan-osa. (9M83)*

## Verbaalisten sanaluokkien vaikutus rakenteen sanamääriin ja kompleksisuuteen

Verbeissä samojen lekseemien toisteisuus on luonnollisesti paljon suu-  
rempi kuin substantiiveissa. Kaikkein yleisimpiä verbejä käytetään en-  
nemmin syntaktisessa tai kielipollisessa funktiossa kuin leksikaalisessa,  
mikä nostaa sekä sanamääriä että kompleksisuutta. Yleisimmät ovat  
sekä lapsilla että aikuisilla pitkälti samoja verbejä: *olla, saada, haluta,  
tulla*. Alakoulun luokilla 1.–5. yleisimmät seitsemän verbilekseemiä kat-  
tavat verbiesiintymistä lähes puolet, nuorilla aikuisilla sama verbimäärä  
kattaa vain runsaan kolmanneksen esiintymistä. Samat lekseemit hallit-  
sevat sanomalehtikielen yleisyystilastoa ja kattavat myös kolmanneksen  
verbiesiintymistä. Alakoulussa HL-verbilekseemien osuus esiintymistä  
on 2,6 prosenttia (esimerkiksi *vääntyä, ylittää, yllättyä, ystävyystyö, älämöl-  
listä*), nuorilla aikuisilla 4,3 prosenttia (esimerkiksi *ähistä, äksyillä, äkätä,  
ärsyttää, äyskäröidä*). Miljoonan sanan lehtikielessä HL-verbilekseemien  
osuus esiintymistä on enää vajaa prosentti.

Verbien osuus kokonaissanamäärästä on kaikissa ikäryhmissä saman-  
tapainen: mediaani on hieman alle 30 prosenttia. Nuorimmilla verbien

tavallisin yleisyysluokka on 20–29 prosenttia sanamäärästä, viidennen pojilla ryhmä 30–39 prosenttia on yleisin, mutta kuudesluokkalaisista lähtien yleisyysluokka 20–29 prosenttia on vallitseva (ks. taulukko 63). Vaihtelu on myös sukupuolesta riippumatta melko vähäistä alakoulun jälkeen mutta kaikkiaan kirjoittajien määrä on luokassa 20–29 prosenttia nouseva. Ilmeisesti iän myötä muiden aineiden osuus sanamassasta kasvaa verbien kustannuksella ja oletettavasti nostaa paitsi sanamääriä myös kompleksisuutta. Sanomalehtikielessä verbiesiintymiä on 21,9 prosenttia kaikista sanaesiintymistä. Verbejä käytetään monissa eri funktioissa ja on todennäköistä, että verbien syntaktinen ja leksikaalinen variaatio ikään kuin häviää sanaluokan sisään. Peruskoulun pojilla on kaikilla poikkeavia arvoja ja prosenttiosuus on normaalijakaumaan verraten hieman liian korkea.

Taulukko 63. Verbitiheys Unelmakirjoitelmissa.

Ikäluokka / Kirjoitelma %	0–9	10–19	20–29	30–39	40 % tai yli
1.–2. F	3,6	6,5	48,5	32,6	8,7
1.–2. M	6,5	4,3	39,6	32,4	17,3
3.–4. F	0	3,6	58,9	35,6	1,8
3.–4. M	0,6	8,9	56,4	32,1	1,8
Viides F	0	0	56,0	42,7	1,3
Viides M	0	5,1	39,2	54,4	1,3
Kuudes F	0	1,2	63,1	33,7	1,9
Kuudes M	0	6,1	57,4	35,1	1,4
Yhdeksäs F	0	1,1	65,9	32,9	0
Yhdeksäs M	1,4	1,4	65,3	29,2	2,8
Aikuiset F	0	0,4	75,6	23,9	0
Aikuiset M	0	1,1	74,7	24,2	0

Kun arvioidaan T-yksikön rakenteen kompleksisuuteen vaikuttavia tekijöitä, kannattaa analyysi pohjata erityisesti T-yksikön verbien asemaan ja funktioihin. Yksinkertaisin rakenne koostuu finiittiverbillisestä yksinäislauseesta mutta kompleksinen rakenne lisäksi sekä finiittiverbillisistä alisteisista lauseista että rakenteista, joissa on ei-finiittistä ainesta. Hallitsevan lauseen ja alisteisen aineksen suhde on asymmetrinen. Esimerkiksi Langackerin (2014) mukaan hallitsevan ja alisteisen aineksen suhde voidaan määritellä neljän kriteerin perusteella, jotka ovat täydellisyys

(*completeness*), sisältyvyys (*containment*), prominenssi (*prominence*) ja saavutettavuus (*accessability*). Alisteinen rakenne on siis epäitsenäinen, sisältyy laajempaan kokonaisuuteen, ei ilmaise pääväitettä eikä johda toiseen elementtiin tai pohjasta sen esiintymistä. Langackerin määritelmä katsoo suhdetta useammalta kielen tasolta (ks. myös esimerkiksi Herlin, Visapää & Kalliokoski 2014). Tässä yhteydessä T-yksiköiden analyysi pohjaa syntaktis-semanttiseen analyysiin, Schneiderin ja muiden (2015) ohjeistukseen mutta myös siihen, miten kirjoittaja on tekstinsä rakentanut. Unelmakorpuksen kirjoitelmien T-yksiköistä kuudessa kymmenestä on alisteista ainesta, joko ei-finiittisiä muotoja tai alisteisia lauseita (ks. taulukko 64). Määrä kasvaa kehityksellisesti. Kuudensilla on myös selvä ajallinen ero siten, että vuoden 2006 kuudesluokkalaisilla alisteisia aineksia on T-yksiköissä huomattavasti enemmän kuin vuoden 2014 kuudesluokkalaisilla. Ero on tytöillä lähes kymmenen prosenttiyksikköä, pojilla runsaat viisi.

Taulukko 64. Alisteisen ja ei-finiittisen aineksen osuus T-yksikköjen kokonaismäärästä.

Ikäluokka / Alisteinen aines T-yksiköissä	F %	M %
2006-Kuudes	60,0	61,3
2014-Kuudes	51,5	55,7
Yhdeksäs	61,3	62,2
Aikuiset	67,3	72,7

Alisteisen aineksen perusteella laskettu kompleksisuusindeksi ei erottele ei-finiittisiä ja finiittisiä alitusrakenteita toisistaan, vaikka näissä voi olla kirjoittamisstrategian, ikäluokan tai yksilön kannalta suuria eroja. Voi myös olettaa, että esimerkiksi runsas propositionaalisten asenteiden ilmaisu (eli modaaliverbien käyttö) ja toisaalta sivulauseiden käyttö vaikuttavat kompleksisuuteen eri tavoin. Vastaavasti voi olla, että näissä on kehityksellistä eroa. Tarkastelemme T-yksiköiden sisäistä jakautumista tarkemmin jakamalla verbisanaluokan finiittiverbeihin, verbin rektio-infinitiiveihin (eli *a-* ja *ma-*infinitiivit) ja muihin ei-finiittisiin aineksiin. Paitsi finiittistä ja ei-finiittistä käyttöä, tarkastelemme myös alisteisten sivulauseiden käyttöä kirjoitelmissa. Muuttujat ovat entiseen tapaan ikäluokka, sukupuoli ja yksilövariaatio. Oletuksena on, että rakenteellinen kompleksisuus lisää myös T-yksikön sanamäärää.



## FINIITTISET JA EI-FINIITTISET VERBIT

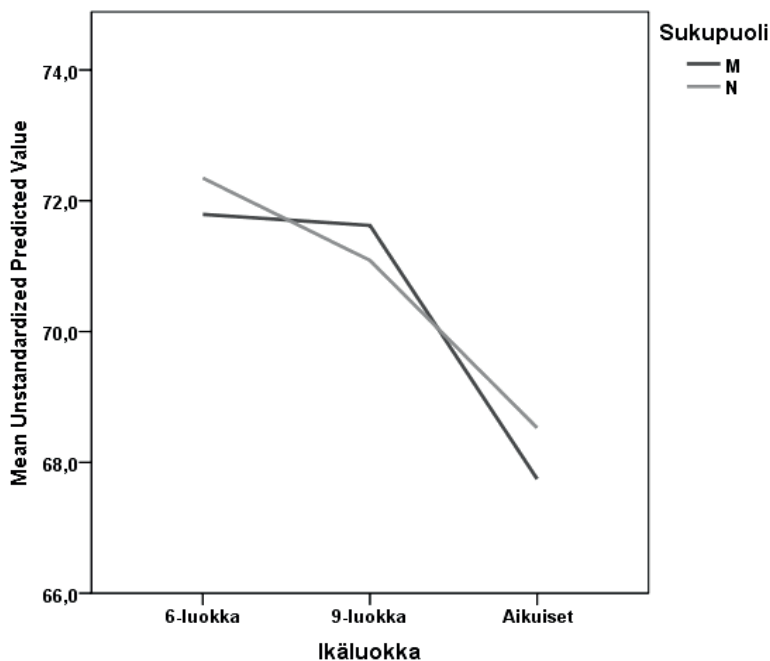
Yleisesti ottaen mitä enemmän kirjoittaja käyttää finiittisiä muotoja, sen yksinkertaisempaa – ja oletettavasti puhutun kielen rakenteen mukaisempaa – tekstiä hän tuottaa. Mitä enemmän hän käyttää ei-finiittisiä muotoja, sen kompleksisempaa ja kehittyneempää – edelleen oletuksen mukaan – hänen kielenkäyttönsä on. Verbisanaluuksissa on alaluokiteltu finiittifunktioiden käyttöihin verbimuotoihin (*olen, en [ole], ollaan*) sekä ei-finiittisiin muotoihin, jotka on edelleen jaettu verbin rektioinfinitiivien sekä muihin (*e*-infinitiivi, partitiivit sekä verbin kieltomuoto). Verbin hallitsemista ei-finiittisistä monet ovat modaaliverbien yhteydessä käytettyjä leksikaalisia verbejä. Muut ei-finiittiset koostuvat pääasiassa verbin liittomuodoista mutta myös eri lauseenvastikkeista. Tämä viimeinen muut-ryhmä on kehityksellisesti mielenkiintoisin, koska sekä liittotempukset, kieltomuodot että lauseenvastikkeet lisääntyvät peruskoululaisten kirjoitelmissa kehityksellisesti (vrt. Pajunen 2012). Tosin voi olla, että kiellon osalta vähäisyys liittyy unelmakontekstiin.

Finiittiverbien osuus koko verbimäärästä laskee nuorimpien keskimäärin 80 prosentista vanhimpien runsaaseen 60 prosenttiin eli suunnilleen 20 prosenttiyksikköä. Ero on merkitsevä suuri. Lehtikielessä finiittiosuus on 62,1 prosenttia. Peruskoulussa pojilla on finiittiverbejä hieman enemmän kaikista verbiesiintymistä kuin tytöillä ja nuorilla aikuisilla tilanne on päinvastainen, vaikka ero on pieni. Yksilöllinen vaihtelu on finiittiverbienkin osuudessa suuri (taulukko 65). Peruskoululaisista kuudesluokkalaisilla ja yhdeksännen pojilla on lähes finiittiverbien varassa eteneviä kirjoitelmia vielä aika paljon (ja yli odotuksen), nuorilla aikuisilla sen sijaan ei lainkaan. Ryhmä 60–69 prosenttia verbeistä finiittisiä kasvaa selvimmin iän mukaan ja kattaa aikuisilla yli puolet kirjoitelmista. Tässäkin muuttujassa miehillä tilanne kääntyy aikuisiällä päinvastaiseksi, sillä nyt naisilla on vallitsevasti finiittiverbillisten kirjoitelmien osuus korkeampi kuin miehillä.

Finiittiverbien määrä vaikuttaa selvästi T-yksikön keskipituuteen ja kompleksisuuteen. Ero on selvin lapsilla ja nuorilla aikuisilla mutta päinvastainen (ks. kaaviot 30 ja 31). Peruskoululaisilla korkea finiittiverbien määrä laskee T-yksikön sanamäärää ja kompleksisuusindeksiä: kertomus voi edetä yksinkertaisin peruslausein. Aikuiskirjoittajilla sekä sa-

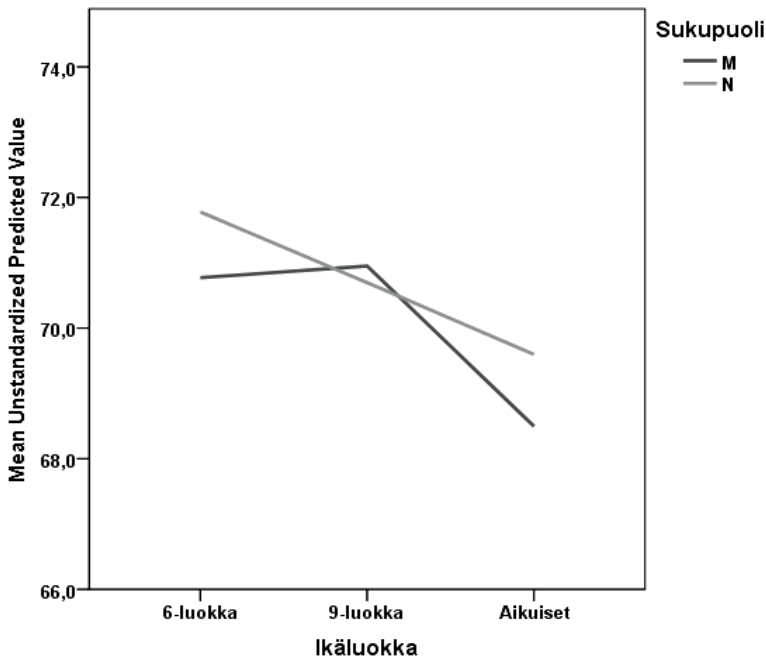
Taulukko 65. Finiittisten verbien tiheys Unelmakorpuksessa.

Ikäluokka / Kirjoitelma %	Yli 90	80–90	70–79	60–69	50–59	20–49 %
Kuudes F	5,6	26,9	43,1	20,0	4,4	0
Kuudes M	8,5	31,4	28,7	21,6	5,9	3,9
Yhdeksäs F	1,1	12,5	38,6	32,9	14,8	0
Yhdeksäs M	6,9	16,6	43,1	22,2	11,1	0
Aikuiset F	0	0,4	23,4	51,0	23,4	1,7
Aikuiset M	0	1,0	12,4	54,6	31,9	0



Kaavio 30. Verbeistä finiittisiä ja T-yksikköjen sanamäärä. Finiittiverbien mediaani-  
osuus: kuudes 75,3 ja 77,8, yhdeksäs 70,6 ja 73,9 sekä aikuiset 65,1 ja 63,2.

namäärä että rakenteen kompleksisuus nousevat finiittiverbien määrän  
noudesta. Sama linja näkyy jo yhdeksäsluokkalaisilla mutta heikkona.  
Erot syntynevät sivulauseista. Kuudennella on paljon kirjoitelmaa, joissa  
sivulauseita on hyvin vähän, jolloin finiittiverbi esiintyy yksinkertaises-  
sa T-yksikössä. Vanhemmilla kirjoittajilla sivulauseet kattavat jo puo-



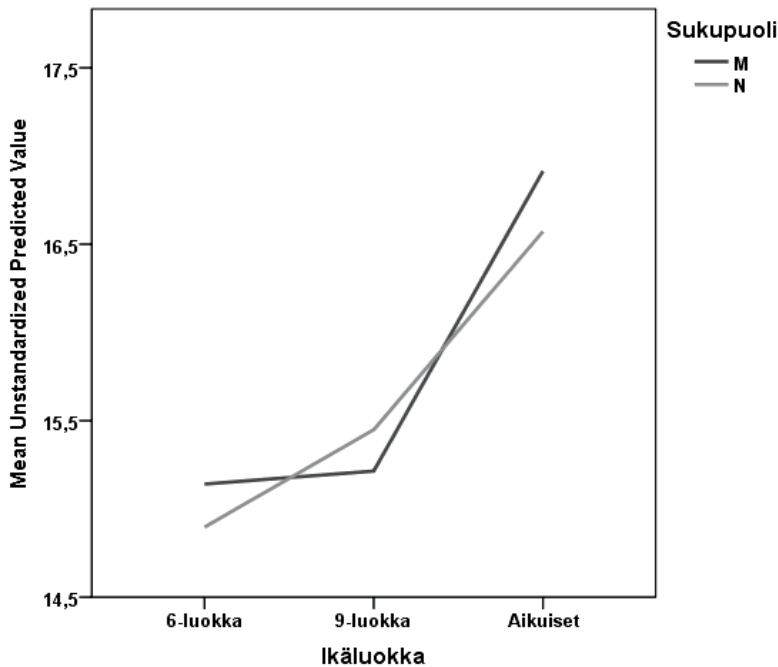
Kaavio 31. Verbeistä finiittisiä ja kompleksisuusindeksi. Finiittiverbien mediaaniosuus: kuudes 75,3 ja 77,8, yhdeksäs 70,6 ja 73,9 sekä aikuiset 65,1 ja 63,2.

let kaikista alisteisista aineksista, joten T-yksikkö on monilauseinen. Kehityksellisesti ero finiittiverbien osuudessa on aineiston suurimpia. Sukupuolen mukaan tilanne ei eroa kovin paljoa paitsi, että mieskirjoittajilla on ääriarvoja enemmän. – Finiittiverbien mediaaniosuus kaikista verbimuodoista on kuudensilla 75,3 ja 77,8, yhdeksänsillä 70,6 ja 73,9 sekä aikuisilla 65,1 ja 63,2.

Verbin hallitsemien rektioinfinitiivien osuus nousee Unelmakorpuksen kirjoitelmissa kehityksellisesti. Keskeisin tiheysluokka on 11–20 prosenttia verbiesiintymistä rektioinfinitiivejä (ks. taulukko 66). Tätä pienempi ja suurempi luokka pienenee iän mukaan. Sukupuolen mukaan muutos menee taas kerran hieman eri tavoin. Kuudennen työllä rektioinfinitiivien määrä on luokassa 0>10 prosenttia muihin verrattuna suuri, ja suuremmissa luokissa miesten luvut ovat usein korkeammat kuin naisten. Peruskoulun pojilta rektioinfinitiivit voivat puuttua koko-

Taulukko 66. Rektioinfinitiivien tiheys Unelmakorpuksessa.

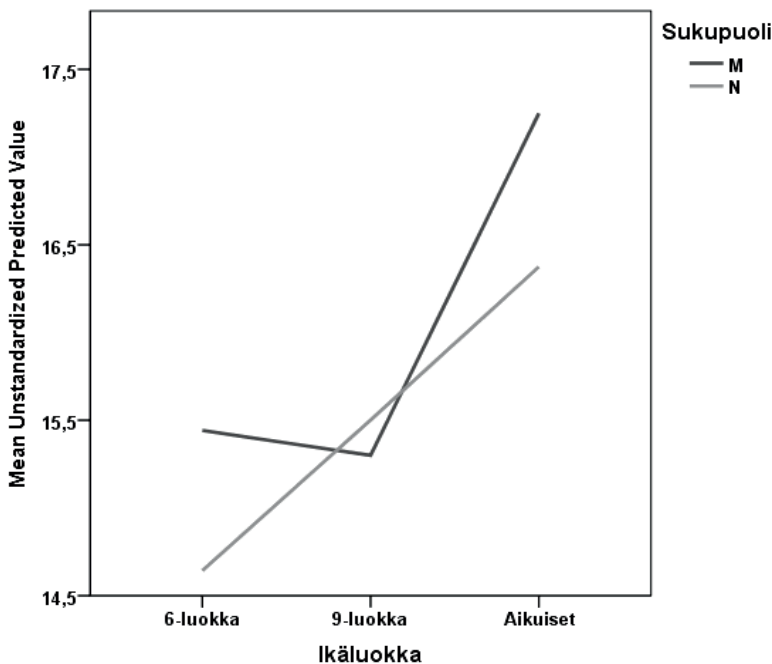
Ikäluokka / Kirjoitelma %	0	0>10	11–20	21–30	31–40	Yli 40 %
Kuudes F	1,2	33,7	46,9	15,6	1,9	0,6
Kuudes M	9,8	17,6	41,2	17,0	8,5	5,9
Yhdeksäs F	1,1	20,4	45,4	26,1	6,8	0
Yhdeksäs M	2,8	13,9	50,0	26,4	5,5	1,4
Aikuiset F	0	18,5	65,5	14,3	1,7	0
Aikuiset M	0	15,3	58,2	24,5	2,0	0



Kaavio 32. Rektioinfinitiivit ja T-yksikköjen sanamäärä. Rektioinfinitiivien mediaani: kuudes 13,7 ja 14,3, yhdeksäs 14,3 ja 15 sekä aikuiset 14,5 ja 15,9.

naan, kuudennen pojilla epäodotuksenmukaisen usein (lähes 10 % odotuksenmukaisen 2–3 % sijaan). Lehtikielessä rektioinfinitiivien osuus verbiesiintymistä on 14,3 prosenttia.

T-yksikön keskipituus nousee rektioinfinitiivien määrän kasvaessa ja eritoten aikuisryhmässä (ks. kaavio 32), mutta kasvua on vähän yh-



Kaavio 33. Rektioinfiitiivit ja kompleksisuusindeksi. Rektioinfiitiivien mediaani: kuudes 13,7 ja 14,3, yhdeksäs 14,3 ja 15 sekä aikuiset 14,5 ja 15,9.

deksännen pojilla. Vastaavasti rakenteen kompleksisuus kasvaa rektioinfiitiivien myötä kaikilla ikäryhmillä (ks. kaavio 33) mutta merkityksellisimmin aikuisryhmissä. Tilanne on samantapainen nais- ja mieskirjoittajilla, mutta nostava vaikutus on melko pieni kuudennen luokan tytöillä ja yhdeksännnen pojilla. Kuudennen tytöillä on paljon yksinkertaisia rakenteita, ja osa yhdeksännnen pojista on ollut alle ikäluokan tason moneksi. – Rektioinfiitiivien mediaani on kuudensilla 13,7 ja 14,3, yhdeksänsillä 14,3 ja 15 sekä aikuisilla 14,5 ja 15,9.

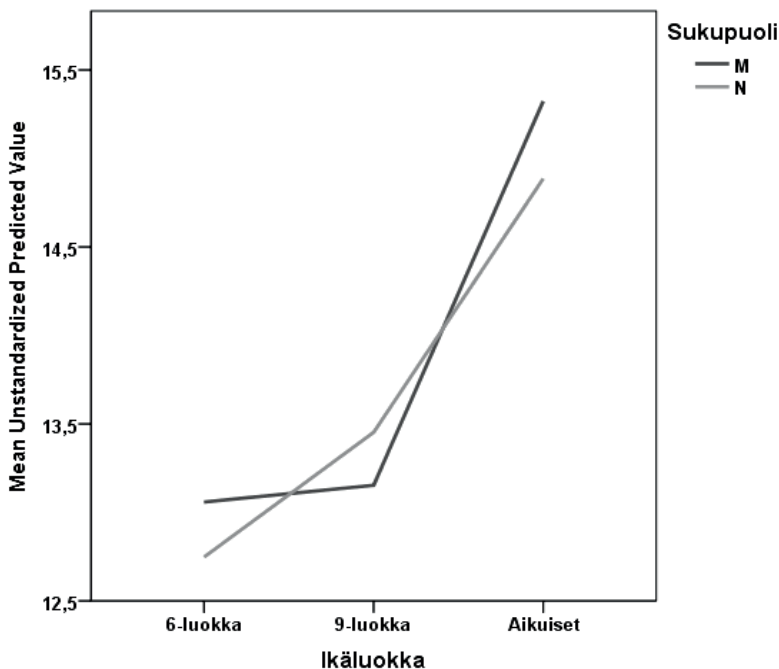
Finiittiverbien ja rektioinfiitiivien lisäksi verbeihin kuuluu joukko muita verbimuotoja: *e*-infiitiivi (*juosten, tankatessa*), verbin kieltomuoto (*[en] ole*) ja partisiipit. Kehityksellinen ero tulee näkyviin myös, kun katsotaan näiden muiden jäljelle jääneiden osuutta kaikista verbiesiintymistä (ks. taulukko 67). Nuorimmilla muita verbifunktioita on eniten

Taulukko 67. Muiden ei-finiittisten verbimuotojen tiheys Unelmakirjoitelmissa.

Ikäluokka / Kirjoitelma %	0	1–5	6–10	11–15	16–20	Yli 20 %
Kuudes F	8,1	15,0	32,5	25,6	12,5	6,2
Kuudes M	26,8	18,9	26,1	15,0	7,8	5,2
Yhdeksäs F	5,7	7,9	22,7	20,4	20,4	22,7
Yhdeksäs M	19,4	9,7	25,0	23,6	12,5	9,7
Aikuiset F	0	0	6,3	13,4	34,4	45,8
Aikuiset M	0	2,0	4,1	17,3	26,5	50,0

yleisyysluokassa 6–10 prosenttia verbeistä; vanhimmilla taas yleisyysluokassa yli 20 prosenttia. Sukupuolen mukaan on edelleen näkyvissä huojuunaa. Kuudennella on poikia, joiden kielenkäyttötaidot ovat selvästi ikäluokasta jäljessä. Yhdeksännellä on vastaava ryhmä, josta ei voi tietää, mikä ongelmia aiheuttaa, todennäköisesti sekä puutteet taidoissa että asenne ja/tai ulkoinen motivaatio. On syytä taas muistuttaa, että nämä samat kuudennen ja yhdeksännen pojat pärjäivät lähes tyttöjen veroisesti tai paremmin morfologisessa ja semanttisessa testauksessa, joten kyse ei ole yksin kielenhallinnasta (vrt. Truckenmiller & Petscher 2020). Lehtikielessä muiden verbiesiintymien osuus on 23,5 prosenttia eli tässäkin nuoret aikuiset ovat lähellä lehtikielen ominaisuuksia.

Muut verbimuodot nostavat T-yksikön keskipituutta voimakkaasti kuudesluokkalaisilla sekä yhdeksännen pojilla, muilla vaikutus on vähäisempi ja aikuisilla selvästi laskeva (ks. kaavio 34). Vastaavasti ne nostavat ja toisaalta laskevat kompleksisuutta samoissa ryhmissä (ks. kaavio 35), mikä erityisesti aikuisten osalta tuntuu ensin hämmästyttävältä. Laskeva vaikutus liittyy partisiippien monifunktioisuuteen. Muista verbimuodoista partisiippien osuus on suurin ja aikuisten partisiipeista kolme neljäsosaa on muissa funktioissa kuin lauseenvastikkeiden predikaattina. Vain lauseenvastikkeiden odottaisi nostavan kompleksisuutta. – Muiden verbimuotojen mediaanit ovat kuudensilla 9,5 ja 5,5, yhdeksänsillä 13,7 ja 9,2 sekä aikuisilla 19,4 ja 19,7.

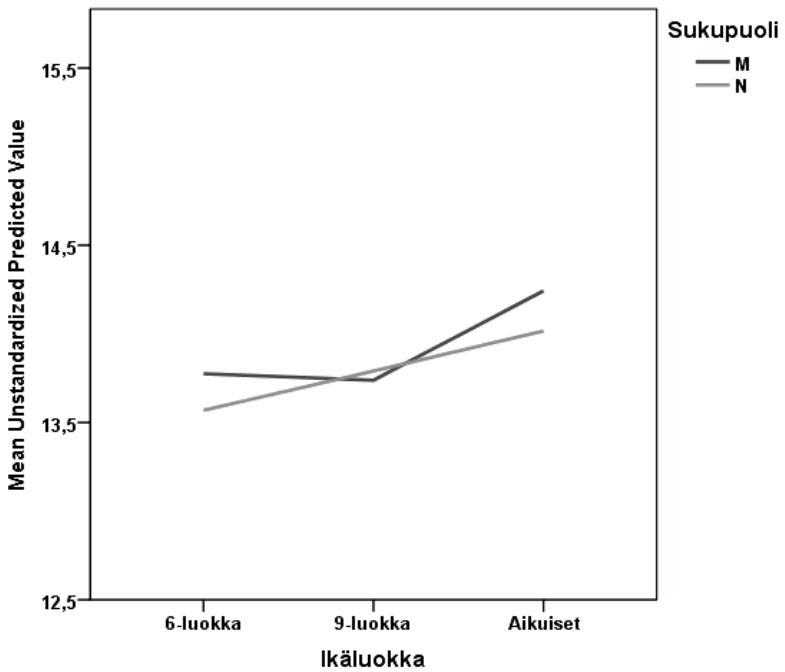


Kaavio 34. Muut ei-finiittiset ja T-yksikköjen sanamäärä. Muiden verbimuotojen mediaanit: kuudes 9,5 ja 5,5, yhdeksäs 13,7 ja 9,2 sekä aikuiset 19,4 ja 19,7.

Yksilötasolla lauseenvastikkeet nostavat nuorten aikuisten kirjoitelmissa kompleksisuutta, vaikka se ei keskiluvuista vielä tullut ilmi. Esimerkin 12 kirjoitelmassa on kaikkiaan 14 muuta ei-finiittistä muotoa. Toisaalta muiden ei-finiittisten määrä voi olla korkea ilman, että kompleksisuus on korkealla, jos käytössä on paljon tempusta ilmaisevia partisiippeja (ks. esimerkki 13).

Kompleksisuusindeksi = 1,74, muiden ei-finiittisten osuus verbeistä = 37,1 prosenttia:

12) *Korkealla sillalla näkymä siniselle Välimerelle tulee näkyviin vielä viimeisen kerran ja sitten katoaa tien kaartuessa sisämaahan. – – Py-sähdyks pienellä huoltoasemalla tuskin housuissaan pysyvä nuorten miesten tankatessa autoja ja toinen idyllisen pikkukylän kahvilassa*



Kaavio 35. Muut ei-finiittiset ja kompleksisuusindeksi. Muiden verbimuotojen mediaanit: kuudes 9,5 ja 5,5, yhdeksäs 13,7 ja 9,2 sekä aikuiset 19,4 ja 19,7.

*cappuccinon ja voisarven kera. Auringon alettua jo laskeutua avautuu edessä näkymä Comon vehreille kukkuloille ja kauas horisonttiin jatkuvaan --. (AMKM43)*

Kompleksisuusindeksi = 1,29, muiden ei-finiittisten osuus verbeistä = 29,1 prosenttia:

13) *Olen koko elämäni haaveillut omasta täydellisestä hääpuvustani, maailman sädehtivimmästä timanttisormuksesta ja ikimuistoisesta hääjuhlasta. --Vaikka avioliitto ei ole minulle elämän keskipiste, olen kuitenkin täysin ihastunut häihin juhvana, suomalaiset häät on yksinkertaisesti täydellinen juhla kaikkine perinteineen. (PääsyF255)*



Lopuksi voi todeta, että finiittiverbien ja ei-finiittisten ainesten suhde on käänteinen. Finiittiverbien määrä laskee kehityksellisesti 30 prosenttiyksikköä ja vastaavasti ei-finiittisten nousee. Ei-finiittiset ainekset eivät kuitenkaan sijoitu aina samoihin kirjoitelmiin vaan jakaantuvat alatyypin mukaan. Rektioinfinitiivit ovat yleisiä tulevaisuuteen viittaavissa kirjoitelmissa, mutta partiippien määrä nousee menneestä kertovissa. Ei-finiittiset ainekset nostavat selvästi rakenteen (ja tietysti myös verbimuodon) kompleksisuutta mutta erot ovat usein selvimmät yksilötasolla.

Ei-finiittisten ainesten nousua selittävät pitkälti verbiluokan laadulliset erot. Kehityksellisesti lisääntyvät erityisesti lauseargumentin kanssa järjestyvät eli mentaaliverbit ja puheaktiverbit (esimerkiksi *tuntea*; *nähdä*, *kuulla*; *ajatella*; *sanoa*; ks. Pajunen 2001) tekoja, tapahtumia ja tiloja suoraan kielentävien kustannuksella. Kaikki lauseargumentin sallivat verbit ovat syntaksiltaan ja semantiikaltaan kompleksisia. Ne kielentävät propositionaalisia asenteita, niillä referoidaan toista ääntä, ilmaistaan abstrakteja tunteita ja niin edelleen. Tekstitasolla lauseargumentti voi realistua komplementtilauseena, eri- tai samasubjektisena infinitiivinä tai partiippiirakenteena (esimerkiksi *luulin että*; *ajattelin mennä*, *käskin heidän lähteä*; *luulin olevani*). Näiden verbien osuus kasvaa kehityksellisesti kuudennelta yhdeksännelle ja edelleen nuoriin aikuisiin (ks. myös Muhonen 2019; Tiiri 2021). Kaikista verbiesiintymistä mentaali- ja puheaktiverbien esiintymien osuus on kuudennella 16 prosenttia ja yhdeksännellä 18 prosenttia mutta nuorilla aikuisilla jo yli 22 prosenttia, eli lisääntyminen samantapaisissa kirjoitelmissa on tasaista (ks. taulukko 68). Myös ryhmän sisäiset suhteet muuttuvat: kuudesluokkalaisilla ryhmän verbeistä 40 prosenttia on puheaktiverbejä, yhdeksäsluokkalaisilla noin kolmannes ja nuorilla aikuisilla enää vajaa neljännes (ks. taulukko 69). Erityisesti evaluoivien mentaaliverbien osuus kasvaa kehityksellisesti, mutta referoivien puheaktiverbien osuus laskee. Tämä tarkoittaa Niinimäen luvussa esitettyä analyysiä mukaillen, että hieman vanhemmat kirjoittajat pohtivat asioita, esittävät omia mielipiteitään ja ilmaisevat tunteitaan, mutta nuoremmat kirjoittajat referoivat puheakteja. Mentaaliverbeistä joka ryhmässä käytetään eniten kognitioverbejä, mutta ainoastaan emotioverbien määrä kasvaa lineaarisesti. Lehtikielessä

mentaaliverbien osuus on hieman suurempi kuin kirjoitelmissa mutta lähellä nuorten aikuisten osuutta. Sen sijaan puheaktiverbien osuus on suurempi ja tämä johtuu luonnollisesti lehtikielen runsaasta referoinnista.

Taulukko 68. Mentaali- ja puheaktiverbien osuus verbiesiintymistä ja kokonaissanamäärästä.

Ikäluokka / Mentaali- ja puheaktiverbi	Verbiesiintymistä	Kokonaissanamäärästä
	%	%
Kuudes	16	4,5
Yhdeksäs	18	5
Aikuiset	22,3	6,1
Lehtikieli	23,7	5

Taulukko 69. Esiintymien jako mentaali- ja puheaktiverbeihin.

Ikäluokka / Verbiluokka	Mentaaliverbiesiintymät	Puheaktivesiintymät
	%	%
Kuudes	59,1	40,9
Yhdeksäs	71,8	28,1
Aikuiset	76,8	23,1
Lehtikieli	64,1	35,8

## FINIITTIVERBILLISET SIVULAUSEET

Myös T-yksikön alisteiset, finiittiverbin sisältävät lauseet nostavat syntaktisen rakenteen kompleksisuutta. Niiden määrä nousee kehityksellisesti oletettavasti, koska rinnastus on ensisijaista tai helpompaa kuin alistus (ks. Itkonen 2009, 189). Nahkola (2012, 42) analysoi alakoululaisten Koulukorpuksen kirjoitelmia (luokat 1.–6.) ja osoittaa, että rinnastuskonjunktoiden osuus lauseita yhdistävistä elementeistä laskee ensimmäiseltä kuudennelle 15 prosenttiyksikköä, adverbiaalisten elementtien osuus nousee kymmenisen prosenttiyksikköä ja relatiivisten osuus kahdentuu vajaasta viidestä kymmeneen prosenttiin. Adverbiaalisista eniten lisääntyvät kontrastiivista, kausatiivista ja finaalista funktiota ilmaisevat; temporaalikonjunktioita käytetään adverbiaalisista silti eniten (Nahkola 2012, 44). Vielä aikuisillakin temporaalet ja kausaaliset adverbiaalilauseet vallitsevat (Tiira 2021, 59). Ja vaikka rinnastuksen rooli heikkenee alakoulun aikana, rinnastus on toki kaikkein yleisin tapa yh-

distää lauseita (noin 50–60 %). Tietysti myös kyky rinnastaa lauseita on kehityksellistä, minkä oivaltaa nopeasti, kun tarkastelee nuorimpien kirjoittajien atomistisia lauseluetteloita. Tarkastelemme tässä yhteydessä peruskoulun kuudes- ja yhdeksäsluokkalaisten alisteisten lauseiden määriä ja funktioita ja vertaamme niitä nuorten aikuisten vastaaviin. Joitain vertailuja esitämme myös vastaavista lauseenvastikkeista, mutta ne perustuvat rajatumpaan aineistoon.<sup>125</sup> Lisäksi esitämme joitain vertailulukuja lehtikielen lauserakenteista, mutta ne perustuvat jossain määrin arvioihin.

Lauseita yhdistävät elementit koostuvat kirjoitelmissa konjunktioista ja relatiivi- ja interrogatiivipronomineista. Konjunktiot on edelleen luokiteltu yksinkertaisen kielioppiluokituksen mukaan alisteisiin ja rinnasteisiin. Kaikki alistavalla elementillä alkavat lauseet nostavat lauserakenteen kompleksisuutta. Rinnastavat on tässä jätetty käsittelemättä, koska ne merkitään vaihtelevasti T-yksiköihin.<sup>126</sup> Tarkastelemme erikseen alisteisten sivulauseiden jakaumaa komplementtilauseisiin (todennäköisimmät argumenttiasemaiset sivulauseet, *sanoi että*), relatiivilauseisiin sekä temporaalisiin, kausaalisiin, ehtoa ilmaiseviin ja muihin niin sanottuihin adverbialisiin alustuslauseisiin. Monista kielistä on esitetty, että juuri relatiivilauseiden määrä lisääntyisi kehityksellisesti (ks. Berman & Verhoeven 2002, 34–35), mutta suomen osalta alakoulussa adverbialilauseiden määrä nousi vielä enemmän kuin relatiivilauseiden (ks. Nahkola 2012, 42). Alisteisten sivulauseiden funktioita analysoitaessa on syytä muistaa, että alustuskonjunktio ei aina kerro, missä funktiossa lause T-yksikössä on (vrt. Herlin 1998; Visapää 2012; Cristofaro 2014).

125 Kuudesluokkalaisten osalta analyysissä ovat mukana vain ne kuudesluokkalaisten kirjoitelmat, jotka ovat kertomusmaisia eli noin 60 % kaikista kirjoitelmista (Muhonen 2019). Aikuisten osalta mukana on vain ammattikorkeakouluilaisten aineisto eli noin kolmasosa aikuisaineistossa (Tiiro 2021). Tosin Tiiron aineisto jakautuu tasan sukupuolen mukaan. Rajatummasta aineistosta lasketut luvut eivät täysin vastaa kaikkia koko aineistosta laskettuja. Esimerkiksi lauseenvastikkeiden määrä on Muhosella korkeampi kuin koko kuudennen aineistossa ja kaikilla aikuisilla hieman matalampi kuin Tiiron aikuisilla.

126 Jos subjekti ilmaistaan verbin persoonataivutuksella, rinnasteiset lauseet kuuluvat eri T-yksiköihin (esimerkiksi *menen ja teen*), toisin kuin esimerkiksi englannissa (vrt. Schneider ja muut 2005). Kolmannessa persoonassa vastaava ei ole mahdollista ja rinnasteiset lauseet kuuluvat nyt samaan T-yksikköön (*hän menee ja tekee*).

Aineiston kuudesluokkalaisten T-yksiköitä analysoinut Muhonen (2019) toteaa, että konjunktioiden käytössä on vielä vakiintumattomuutta.

Alisteisten sivulauseiden tiheys on laskettu kirjoitelman kokonaissanamäärästä koehenkilöittäin (ks. taulukko 70). Valtaosalla peruskoululaisista on kirjoitelmassaan 1–5 alisteista sivulauseita. Kuudennella on vielä iso ero tyttöjen ja poikien välillä, mutta se tasaantuu jonkin verran yhdeksänteen mennessä. Aikuis kirjoittajista noin puolella on 1–5 ja puolella 6–10 sivulauseita; sukupuoliero sivulauseetiheydessä on pieni. Sivulauseiden määrä toisin sanoen kasvaa selvän kehityksellisesti. Myös Kauppinen (1976), Vikainen (1982), Lundin (1983), Närhi (1990) ja Virolainen (2003) ovat todentaneet tämän kehityksen varhaisemmista koululaiskirjoitelmista. Samalla tutkimukset myös osoittavat, että sivulauseiden osuus kaikista lauseista on kuitenkin jatkuvasti laskenut. Muhonen (2019) toteaa selvän eron vuosien 2006 ja 2014 kuudesluokkalaisten (kertomusmaisissa) Unelmakirjoitelmissa sekä määrissä että hierarkkisuuksessa.

Lauseenvastikkeiden väheneminen on ollut erityisen suuri. Virolaisen 2000-luvun kirjoitelmissa lauseenvastikkeita on yhdeksäsluokkalaisilla 4,5 prosenttia kaikista lauseista mutta Närhin 1980-luvun aineistossa jopa 15,5 prosenttia. Tuloksia ei voi suoraan verrata tämän tutkimuksen aineistoon, mutta jos otetaan huomioon T-yksikköjen keskimääräinen kompleksisuus, voidaan Tiiron (2021) ja Muhosen (2019) laskelmien perusteella arvioida, että kuudesluokkalaisilla on lauseenvastike 0,9 prosentissa kertovan tekstin lauseista (T-yksikköjä 4 327 x CI 1,55) ja aikuisilla 7,8 prosentissa (T-yksikköjä 3 868 x CI 1,67). Unelmakirjoitelmissa tosin kaikenikäisillä kirjoittajilla on ongelmia niin sivulauseiden rinnastamisessa kuin lauseenvastikkeiden muodostamisessa ja käytössä. Toisin sanoen vaikka sivulauseiden ja lauseenvastikkeiden määrä kasvaa koehenkilöillä kehityksellisesti, kyky muodostaa niitä on heikennyt. – Sanomalehtikielessä näyttäisi joka kolmannessa virkkeessä olevan yksi alisteinen lause.

Taulukko 70. Alisteisten lauseiden tiheys Unelmakirjoitelmissa ikäluokittain.

Ikäluokka / Kirjoitelma %	o kpl	1–5	6–10	11–15	Yli 15
Kuudes F	0	70	26,8	1,9	1,3
Kuudes M	8,5	54,6	24,3	7,2	5,3
Yhdeksäs F	3,4	79,5	15,9	1,1	0
Yhdeksäs M	8,3	69,4	18	2,8	1,4
Aikuiset F	0	50,4	47,1	2,5	0
Aikuiset M	1	53,1	43,9	2	0

Sivulauseiden (n = 5 787) alaluokittelu osoittaa, että juuri relatiiviset sivulauseet lisääntyvät kehityksellisesti myös suomen kielessä, toisin sanoen jo kuudennelta alkava kehitys jatkuu aikuiskirjoitelmiin (ks. taulukko 71). Ero kuudes- ja yhdeksäsluokkalaisten välillä on lähes kahdeksan prosenttiyksikköä, kuudesluokkalaisten ja aikuisten välillä yli 13 prosenttiyksikköä ja yhdeksäsluokkalaisten ja ammattikorkeakouluilaisten välillä kuusi prosenttiyksikköä, yhdeksäsluokkalaisten ja pääsykokeeseen osallistuneiden välillä ero on vain kolme prosenttiyksikköä. Pääsykoeaineiston kirjoittajat ovat pääosin naisia, ammattikorkeakouluilaisten aineistossa on miehiä ja naisia yhtä paljon. Komplementtilauseissa muutos ei ole lineaarinen, mutta sen sijaan adverbiaalilauseiden määrä laskee lineaarisesti.

Taulukko 71. Alisteisten lauseiden jakauma funktioittain ja ikäluokittain Unelmakorpukessa.

Ikäluokka / Sivulauseetyppi %	Komplementtilause	Adverbiaalilause	Relatiivilause
Kuudes F	22	59,8	18,2
Kuudes M	22,1	55	22,8
Yhdeksäs F	24,4	43	32,6
Yhdeksäs M	24	43,4	32,6
Aikuiset F	21	42,8	36,2
Aikuiset M	17,9	40,6	41,4

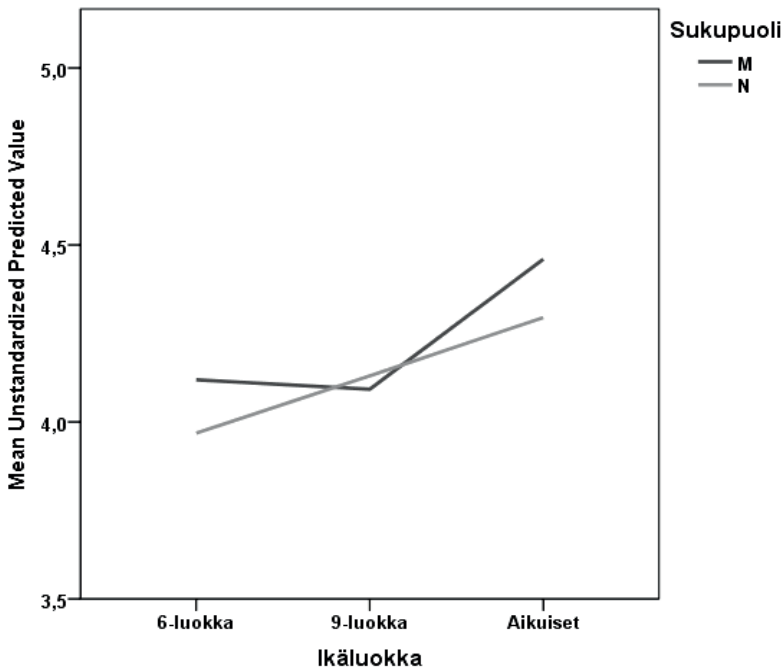
Komplementtilauseiden määrissä ei näy suuria kehityksellisiä eroja eri ikäluokkien välillä toisin kuin niiden alatyypeissä. Nuorimmilla kirjoittajilla *että*-lause on vallitseva komplementtilauseen tyyppi (Nahkola 2012, 36; Muhonen 2019, 36), mutta vanhemmilla kirjoittajilla epäsuora

kysymyslause ja samanaikaista tekemistä ilmaiseva lauseenvastike lisääntyvät. Tämä on oletetun kompleksistumiskehityksen mukaista (ks. Noonan 2007, 114–115). Lisäksi kuten Niinimäki osoittaa tässä teoksessa, suora esitys laskee selvän kehityksellisesti kuudennelta aikuisryhmiin ja *että*-lauseella ilmaistu epäsuora esitys kasvaa. Samalla komplementtia hallitsevan verbin tyyppi muuttuu puheaktista evaluointia ilmaiseviin kognitioverbeihin ja tämäkin muutos on kehityksellinen.

Myös relatiivilauseiden funktioissa ikä tuo mukanaan eroja. Tiiron (2021) mukaan kuvailevia (rajaavia ja ei-rajaavia) relatiivilauseita on aikuisilla eniten (65 % kaikista), lähinnä narratiivisia, tarinaa eteenpäin vieviä on vajaa kolmannes, ja loput ilmaisevat subjektiivista mielipidettä. Erityisesti narratiiviset relatiivilauseet näyttävät lisääntyvän aikuiskirjoittajilla kuudesluokkalaisiin verrattuna (Muhonen 2019), vaikka aikuiskirjoittajilla on kaikkia relatiivilauseen tyyppejä enemmän kuin kuudesluokkalaisilla. Ero vaihtelee 5–10 prosenttiyksikköä. Relatiivilauseissa myös ei-lineaariset rakenteet eli kiilalauseet lisääntyvät; niitä on aikuis teksteissä noin viisi prosenttia kaikista relatiivilauseista (Tiiron 2021). Muissa sivulauseityypeissä ei-lineaariset rakenteet ovat poikkeuksellisia.

Adverbiaalilauseista enemmistö eli neljä viidestä on aikuisilla joko temporaalisessa tai kausaalisisessa funktiossa ja ne ovat yhtä yleisiä (Tiiron 2021, 60). Kuudennella adverbiaalilauseista yli 60 % on temporaalisia (Muhonen 2019, 30–32), mutta kausaaliset kasvavat selvän kehityksellisesti. Temporaalisten lauseiden ja etenkin lauseenvastikkeiden funktio on kerronnallinen ja koherenssia luova silloin, kun ne sijaitsevat päälauseen edellä; 30 prosenttia temporaalisista sivulauseista ja 70 prosenttia vastaavista lauseenvastikkeista on etuasemaisia. Taka-asemaisten funktio on kuvailla tai täydentää. Tiiron mukaan temporaalilauseen asema vaikuttaa myös sen aikasuhteeseen: etuasemassa oleva ilmaisee useimmiten peräkkäistä aikasuhdetta (esimerkiksi *syötyäni menin nukkumaan*), taka-asemassa oleva ennemmin päällekkäistä (esimerkiksi *luin syödessäni*). Kehityksellisesti peräkkäisiä suhteita ilmaistaan ennen päällekkäisiä (Aksu-Koç & von Stutteheim 1994, 392; Tiiron 2021, 67).

Sivulauseiden funktioiden muutokset eivät vaikuta T-yksiköiden sanamääriin tai kompleksisuusindeksiin toisin kuin niiden määrän kasvu, mutta lauseenvastikkeet voivat pudottaa T-yksikön sanamäärää. Sanoma-



Kaavio 36. Alisteiset sanamäärästä ja kompleksisuusindeksi. Alistavien mediaani: kuudes 3,7 ja 3,2, yhdeksäs 3,5 ja 3,3 sekä aikuiset 3,5 ja 3,8.

lehtikielen miljoonan sanan otoksessa relatiivi- ja adverbiaalilauseiden osuudet ovat lähellä nuorten aikuisten lukuja (35 % ja 37 %) mutta komplementtilauseita on 10 prosenttiyksikköä enemmän lehtikielessä kuin nuorilla aikuisilla. Syynä lienee jälleen sanomalehtikielen referoinnin tarve, joka jo todettiin edellä puheaktiverbien yhteydessä.

Kuudesluokkalaisten ja aikuisten aineistossa on selvät sukupuolen mukaiset erot sivulauseiden funktioissa: naiskirjoittajat käyttävät enemmän adverbiaalilauseita, mieskirjoittajat relatiivilauseita. Aikuisaineistossa on yli viiden prosenttiyksikön ero juuri relatiivilauseiden määrissä, ja jo tämä voisi selittää aikuisten mieskirjoittajien korkeampaa kompleksisuusindeksiä. Yhdeksäsluokkalaisten aineistossa ei ole sukupuolen mukaan selittyviä eroja, vaikka hyvin monissa muissa kompleksisuusmuuttujissa juuri yhdeksänneen pojat ovat eronneet yleiskuvasta. Lau-

seenvastikkeet kuuluvat selvästi mieskirjoittajien kieleen (Tiiro 2021). Kaikilla kirjoittajilla sekä MLTU että kompleksisuusindeksi (ks. kaavio 36) nousevat alistavien sivulauseiden lisääntymisen myötä, nousu on suurinta aikuiskirjoittajilla ja pienintä kuudennen tyttöillä.<sup>127</sup> – Alistavien sanojen mediaaniosuus kaikista sanoista on kuudensilla 3,7 ja 3,2, yhdeksänsillä 3,5 ja 3,3 sekä aikuisilla 3,5 ja 3,8.

Peruskoululaisten ja aikuisten välillä on siis eroa sivulauseiden määrässä ja funktioissa. Lisäksi aineistoa selailemalla näkee nopeasti, että etenkin kuudesluokkalaisilla on vielä paljon epävarmuutta paitsi konjunktioiden valinnassa myös lause- ja virkerakenteiden luomisessa (samoin Muhonen 2019). Kuudennen esimerkkitekstit 14–15 ovat melko pitkiä, ja ne ovat suunnilleen ikätasoisten tyttöjen kirjoittamia. Esimerkissä 14 virkerakenne on toisteinen, mikä näkyy muun muassa relatiivilauseista. Esimerkki 15 edustaa aikajärjestyksessä etenevää yksinkertaista rakennetta.

Kompleksisuusindeksi = 1,67, alisteisia 15, rinnasteisia 2, sanoja 279:  
14) *Mario taas kiinnitti huomionsa aivan muuhun. Hän oli suoraan pulahtanut veteen ja uinut pienelle saarelle, joka kökötti aivan järven keskellä. Minä hyppäsin veteen hyppytornin korkeimmasta kohdasta. Kun olin hypännyt veteen niin lähdin uimaan kohti saarta missä Mario odotteli minua. Lopulta parin minan kuluttua pääsin saarelle, jossa Mario odotteli minua. Mario, joka oli peläyttämässä minua plumpautti veteen, koska oli kiivennyt puuhun ja oksa oli katkennut. Minä aloin nauramaan suu kuivana hänelle, Mario oli nolona, koska rannalla oli poikia, jotka katsoivat meitä iloisina. (6F25).*

Kompleksisuusindeksi = 1,09, alisteisia 2, rinnasteisia 1, sanoja 241:  
15) *Aamupalan jälkeen menemme talliin ja puunaamme hevoset ja lähemme maastolenkille. Maastolenkin varrella on pieni lampi ja pysähdyimme uittamaan hevosia siihen. Olimme uimassa n. yhden tunnin ja sen jälkeen jatkoimme matkaa. Maastolenkin jälkeen veimme hevo-*

127 Kaavioon 36 ei ole erikseen huomioitu lauseenvastikkeita. Ne nostaisivat erityisesti mieskirjoittajien CI-indeksiä.



*set talliin. Tallissa me siivosimme karsinat ja teimme kaikki tallityöt, pistimme hevoset laitumelle ja menimme itekin syömään. Ruoaksi oli pizzeria ja minulle tietysti gluteinitonta koska olen keliakaikikko. Loppu päivällä ratsastimme kentällä ja kaikki meni hyvin. Aika loppui hie-  
man kesken, mutta siinä oli suurin osa siitä päivästä. (6F28)*

## Kompleksisuus kehityksellisesti: mikä nostaa tai laskee kompleksisuutta

Tarkastelemme lopuksi kompleksisuuden kehityksellistä nousua verbaalilla keskeisten sanaluokkaesiintymien, verbaalisten kompleksisuustekijöiden sekä eräiden muiden indeksien suhteellisen osuuden ja arvojen vaikutusta T-yksikön keskipituuden (MLTU) ja kompleksisuusindeksin (CI) mediaanin mukaan jaotellen.

Sanaluokista on otettu huomioon adjektiivit, substantiivit ja pronominit sekä verbit ja verbeistä erikseen infinitiivit ja partisiipit. Esiintymäosuudet on jaettu sen mukaan, miten ne sijoittuvat MLTU:n mediaanin ala- tai yläpuolelle. Kuudennella mediaani on 6, yhdeksännellä 6,2 ja nuorilla aikuisilla 7,4 (ks. taulukko 72). Tulos on, että infinitiivejä, adjektiiveja ja pronomineja on enemmän yli mediaanin tapauksissa, joten voi olettaa, että näiden sanaluokkien jäsenet voivat nostaa T-yksikköjen keskimääräistä sanamäärää. Verbien suhteellinen osuus sen sijaan laskee yli mediaanin tapauksissa joka ikäluokassa. Partisiippien määrä nousee aikuisryhmässä mutta vaikutus MLTU:n sanamäärään on pieni. Substantiiveja on enemmän silloin, kun MLTU on alle ikäluokan mediaanin. Tämä on hieman epäodotuksenmukaista, koska sekä leksikaalinen tiheys että NTVR nousevat kehityksellisesti; substantiivit vaikuttavat näissä molemmissa eniten. Ikäryhmittäin erot eri sanaluokkien jäsenten esiintymämäärissä eivät ole niin suuret, että ne voisivat huomattavasti vaikuttaa T-yksiköiden keskipituuteen verbejä ehkä lukuun ottamatta. Verbien määrä laskee kehityksellisesti, ja vastaavasti substantiivien, adjektiivien ja pronominiinien määrä nousee. Infinitiivien määrä laskee kehityksellisesti, mikä tuntuu epäodotuksenmukaiselta. Lasku kuitenkin selittyy, kun ottaa huomioon, että nuorimmilla modaali-verbit (esimer-

kiksi *haluan mennä*) nostavat kompleksisuutta enemmän kuin aikuisilla. Muutoksia kaikkien käsiteltyjen sanaluokkien esiintymissä voi pitää kehityksellisinä, mutta erityisesti substantiivien ja verbien osuus on altis muuttumaan MLTU:n keskipituuden mediaanin mukaan: määrä on molemmilla suurempi, jos MLTU on alle mediaanin, ja silloin kyse on lähinnä yksinkertaisilla peruslauseilla etenevästä tekstistä.

Taulukko 72. Sanaluokkaesiintymien vaikutus T-yksikön sanamääriin (alle/yli MLTU-mediaanin).

MLTU-mediaani / Sanaluokka %	Adjek- tiivit	Substan- tiivit	Prono- minnit	Verbit	Infini- tiivit	Parti- siipit
Kuudes MLTU alle 6	5,7	30,6	8,1	29,6	5,1	1,9
Yhdeksäs MLTU alle 6,23	5,6	31,7	7,4	29,2	4,3	1,9
Aikuiset MLTU alle 7,41	8,8	31,9	8,6	28,7	4,8	3,2
Kuudes MLTU yli 6	5,4	29,7	9,6	27,6	6,5	2,2
Yhdeksäs MLTU yli 6,23	7,3	29,9	9,3	27,2	5,4	1,9
Aikuiset MLTU yli 7,41	9,5	31,6	9,5	25,9	5,1	3,4

Varsinaisiksi kompleksisuustekijöiksi on laskettu verbaaliset ainekset, joita tarkastellaan alaluokittamalla ne muodon ja funktion mukaisesti (ks. taulukko 73). Lisäksi kompleksisuustekijöihin lukeutuvat alisteiset elementit, joita arvioimme suhteessa kaikkiin lauseita yhdistäviin elementteihin. Beersin ja Nagyn (2009) mukaan kompleksisuustekijöiden vaikutus T-yksikön keskipituuteen ei olisi kertovassa tekstissä odotuksenmukainen. Unelmakirjoittelusta näkyy selvästi, että finiittiverbien määrä laskee ikäluokkatason mediaanin noustessa ja toisaalta rektio-infiniittivejä ja muita ei-finiittisiä verbejä on joka ikäluokassa enemmän kirjoitelmissa, joiden MLTU on yli ikäluokkamediaanin. Vastaavasti lausetta alistavien elementtien määrä kaikista yhdistävistä on suurempi kirjoitelmissa, joiden MLTU on yli mediaanin. Kuudesluokkalaisilla alisteiset lauseet ovat sekä ali että yli mediaanin tapauksissa keskeisempi kompleksisuustekijä kuin muilla ryhmillä, sillä vaikka relatiivilauseet lisääntyvät kehityksellisesti, adverbialilauseet vähenevät.

Taulukko 73. Kompleksisuustekijöiden vaikutus MLTU:hun (alle/yli mediaaniin).

MLTU-mediaani / Kompleksisuustekijä %	Finiitti- verbit	Rektio- infinitiivit	Muut ei-finiittiset	Alistavat yhdistävistä
Kuudes MLTU alle 6	80	12,5	6,8	33,3
Yhdeksäs MLTU alle 6,23	75	12,5	10,5	25
Aikuiset MLTU alle 7,41	67,0	13,6	18,7	31,3
Kuudes MLTU yli 6	73,3	15,6	9,1	42,3
Yhdeksäs MLTU yli 6,23	69,2	16,6	12,8	35,9
Aikuiset MLTU yli 7,4	61,7	16,8	21,2	41,5

Kolmanneksi on arvioitu johdoksia ja johtamattomia sanoja sekä eri sanaindeksejä T-yksikön ikäluokkamediaaniin (taulukko 74). Johdosten määrän tiedetään nousevan kehityksellisesti ja vastaava nousu koskee myös leksikaalista indeksiä, NTVR:ta sekä pronomini–substantiivi-indeksiä. Unelmakorpuksessa kuudesluokkalaisten sanarakenne ei muutu odotuksen mukaan suhteessa MLTU:n ikäluokkamediaaniin, mutta johdosten määrä kyllä kasvaa jonkin verran MLTU:n noustessa yhdeksäsluokkalaisilla ja aikuisilla. Toisin sanoen vaikka johdokset lisääntyvät kehityksellisesti, ikäryhmittäin paljon johdoksia käyttävät eivät kirjoita kovin paljon pidempiä T-yksiköitä. Vastaavasti NTVR eriytyy vähäisesti, ja vaikka suhde on korkeamman MLTU:n teksteissä korkeampi, ero on silti pieni. Leksikaalinen tiheys laskee pidemmissä T-yksiköissä. Tämä selittynee pronominiin lisääntymisellä, koska ne vähentävät leksikaalisten sanojen käytön tarvetta. Pronominiin suhteellinen osuus niiden ja substantiivien yhteisosuudesta kasvaa T-yksikön pidentyessä. Ero on sanaindekseistä selvin.

Taulukko 74. Sanaindeksit ja MLTU (alle/yli mediaaniin).

MLTU-mediaani / Sanaindeksit %	Johdokset	Johta- mattomat	NTVR	Leksikaalinen tiheys	Pronomini- substantiivi- indeksi
Kuudes MLTU alle 6	11,7	71	51,7	76,1	21,1
Yhdeksäs MLTU alle 6,23	13,3	69,4	51,2	76,8	19,8
Aikuiset MLTU alle 7,41	19,7	65,6	52,7	78	21,1
Kuudes MLTU yli 6	11	71,7	51,9	74,2	25,4
Yhdeksäs MLTU yli 6,23	15,8	69,5	53	74,2	23,7
Aikuiset MLTU yli 7,41	20,8	62,5	54,6	76,6	23,4

Kompleksisuusindeksiä (= CI) arvioidaan samalla tavoin kuin MLTU:ta eli jakamalla aineisto indeksin ikäluokkamediaanin avulla kahtia alle mediaanin ja yli mediaanin ryhmiin. Kuudennella mediaani on 1,55, yhdeksännellä 1,58 ja aikuisilla 1,67. Adjektiivien käyttö laskee kompleksisuuden myötä mutta erot ovat jälleen pienet (ks. taulukko 75). Substantiiveja on enemmän alle mediaanin tapauksissa eli siis rakenteellisesti yksinkertaisemmassa tekstissä mutta kehityksellisesti määrä nousee kummassakin tapauksessa. Substantiivien rooli on toistamiseen vastoin ennako-oletusta, sillä Ravidin (2006, 792) mukaan teksti on syntaktisesti sitä kompleksisempää mitä enemmän siinä on substantiiveja ja adjektiiveja. Verbien määrä on matalampi alle mediaanin tapauksissa, vaikka kehityksellisesti määrä laskee kummassakin ryhmässä. Myös infinitiivien käyttö on vähäisempää alle mediaanin tapauksissa, ja yli mediaanin tapauksissa näkyy kehityksellinen nousu, samoin on tilanne partisiippien osalta, mutta niiden kokonaisfrekvenssit ovat kaikissa ryhmissä pienet.

Taulukko 75. Sanaluokkaesiintymien vaikutus kompleksisuusindeksiin (alle/yli mediaanin).

CI-mediaani / Sanaluokka %	Adjektiivit	Substantiivit	Pronomininit	Verbit	Infinitiivit	Partisiipit
Kuudes CI alle 1,55	5,8	32,8	8,5	27,9	4,6	1,9
Yhdeksäs CI alle 1,58	6,5	31,7	7,5	27,8	3,7	1,7
Aikuiset CI alle 1,67	9,3	32,9	8,6	26,9	4,3	3,2
Kuudes CI yli 1,55	5,4	28,7	9,5	29,6	7,3	2,4
Yhdeksäs CI yli 1,58	6,7	30,0	9,3	28,4	6,1	2,11
Aikuiset CI yli 1,67	9,0	31,1	9,5	27,8	5,8	3,4

Finiittiverbit, rektioinfinitiivit ja muut ei-finiittiset on laskettu verbien kokonaissanamäärästä (ks. taulukko 76); alistavien konjunktioiden osuus on laskettu kaikista lauseista ja lausekkeista yhdistävistä elementeistä. Näillä ns. kompleksisuustekijöillä on välttämättä vaikutusta kompleksisuusindeksiin, koska kyse on piirteistä, joilla kompleksisuus on lause- tasolla määritelty. Mielenkiintoisempi tietysti onkin kysymys, paljonko tekijät eri ikäryhmissä vaikuttavat ja onko jokin tekijä näkyvämpi jossain tietystä ikäluokassa. Finiittiverbien määrä laskee selvän kehityksellisesti, sillä eniten finiittiverbejä käytetään kuudensien kirjoitelmissa ja vähiten aikuiskirjoitelmissa. Lasku kompleksisuuden mukaan kulkee vastaavas-

ti. Rektioinfinitiivien määrä nousee huomattavasti mentäessä yli mediaanin tapauksiin, ja vaikka määrissä on kehityksellinen kasvu, se ei ole kovin suuri. Muiden ei-finiittisten määrä nousee sekä kehityksellisesti että kompleksisuuden myötä, tosin aikuisilla ero on pieni. Alistavat elementit lisääntyvät yli mediaanin tapauksessa ja nousu on 10 prosenttiyksikön luokkaa kaikissa ikäryhmissä.

Taulukko 76. Kompleksisuustekijöiden vaikutus kompleksisuusindeksiin (alle/yli mediaanin).

CI-mediaani / Kompleksisuustekijä %	Finiittiverbit	Rektioinfinitiivit	Muut ei-finiittiset	Alistavat yhdistävistä
Kuudes CI alle 1,55	81,2	11,3	6,3	33,3
Yhdeksäs CI alle 1,58	76,5	11,6	10,3	25,0
Aikuiset CI alle 1,67	67,4	12,5	19,6	30,8
Kuudes CI yli 1,55	71,9	17,8	9,1	43,3
Yhdeksäs CI yli 1,58	68,2	18,4	13,0	35,7
Aikuiset CI yli 1,67	61,0	17,9	20,0	42,2

Sanarakenteen kompleksisuus ei näytä vaikuttavan lausetason kompleksisuuteen kovin selvästi (taulukko 77). Johdoksia on enemmän alle mediaanin tapauksissa, mutta toisaalta niiden määrä kasvaa kehityksellisesti molemmassa tapauksissa. Johtamattomien sanojen määrä laskee kehityksellisesti mutta lausetason kompleksisuus hienokseltaan nostaa johtamattomien määrää. Tämä mitä todennäköisimmin johtuu pronomien, modaaliverbien ja yhdistävien elementtien lisääntymisestä, koska ne ovat pääosin johtamattomia ja lyhyitä sanoja. Vastaava kehitys näkyy pronomien ja substantiivien keskinäisestä suhteesta. Pronomien käyttö laskee hieman kehityksellisesti mutta nousee kompleksisuuden myötä. Leksikaalisissa indekseissä kompleksisuuserot ovat pieniä mutta pääasiassa NTVR-suhde ja leksikaalinen tiheys laskevat kompleksisuuden myötä. Toisin sanoen kompleksisessa tekstissä on vähemmän substantiiveja kuin ei-kompleksisessa. Tämä voi heijastaa kompleksisen tekstin ei-finiittisten ainesten tiheyttä ja toisaalta ei-kompleksisen tekstin tyyppiominaisuuksia. Yleensä esimerkiksi erittelevän tekstin kompleksisuus on matalampi ja leksikaalinen tiheys korkeampi kuin kertovan, mutta Unelma-aineistossa erittelevämpiä tekstejä on liian vähän, jotta oletusta voisi näyttää toteen.

Taulukko 77. Sanaindeksit ja kompleksisuusindeksi (alle/yli mediaanin).

CI-mediaani / Sanaindeksit %	Johdokset	Johtamat- tomat	NTVR	Leksikaali- nen tiheys	Pronomini-sub- stantiivi-indeksi
Kuudes CI alle 1,55	11,9	70,6	53,8	75,8	21,2
Yhdeksäs CI alle 1,58	13,9	68,9	53,3	76,5	19,7
Aikuiset CI alle 1,67	20,0	64,4	54,8	78,0	20,6
Kuudes CI yli 1,55	10,8	72,5	50,0	74,7	25,3
Yhdeksäs CI yli 1,58	15,1	70,3	51,5	74,6	23,2
Aikuiset CI yli 1,67	20,2	64,1	52,7	76,8	23,9

Lopuksi voi vielä tarkastella MLTU:n ja kompleksisuusindeksin keskinäistä suhdetta niiden mediaanien avulla (ks. taulukot 78–79). Lauserakenteen kompleksisuus nousee selvästi T-yksikön sanamäärän kasvaessa ja vastaavasti sanamäärä nousee kompleksisuuden kasvaessa. Erot ovat hyvin selviä. On itsestään selvä, että kompleksisuus nostaa T-yksikön sanamääriä, mutta ei ole yhtä odotuksenmukaista, että korkempi MLTU liittyy korkeampaan kompleksisuuteen. Todennäköisesti kyse on siitä, että kirjoitelmat kuitenkin pääasiassa edustavat yhtä ja samaa tekstityyppiä, kertomuksia (tai niihin pyrkiviä tekstejä). Tulos osoittaa, että Beersin ja Nagyn hypoteesi siitä, että kertovan tekstin mittarina MLTU ei olisi pätevä, ei pidä paikkaansa suomen aineistossa.

Taulukko 78. Indeksien suhde a:b.

MLTU	Kompleksisuusindeksi
Kuudes MLTU alle 6	1,4
Yhdeksäs MLTU alle 6,23	1,42
Aikuiset MLTU alle 7,41	1,56
Kuudes MLTU yli 6	1,75
Yhdeksäs MLTU yli 6,23	1,8
Aikuiset MLTU yli 7,41	1,83

Taulukko 79. Indeksien suhde b:a.

Kompleksisuusindeksi	MLTU
Kuudes CI alle 1,55	5,29
Yhdeksäs CI alle 1,58	5,5
Aikuiset CI alle 1,67	6,7
Kuudes CI yli 1,55	6,79
Yhdeksäs CI yli 1,58	7,43
Aikuiset CI yli 1,67	8,5

## Lopuksi

Puhutun kielen rakenne ei yleensä ole kovin hierarkkinen eli kerroksellinen eivätkä lauseet keskimäärin ole kovin pitkiä. On näkemyksiä, joiden mukaan puhutussa kielessä yhden argumentin ilmaukset olisivat vallitsevia (vrt. Bybee & Hopper 2001). Kirjoitetussa kielessä rakenne on yleensä tiiviimpi ja informaatiota voidaan pakata yksittäisiin yksiköihin kerroksellisemmin. Kyse on ensi sijassa modaliteettierosta, joka syntyy kuulo- ja näköaistin eroista ja on yhteydessä lyhytkestoisen muistin kapasiteettiin: kielen auditiivinen ja visuaalinen prosessointi ovat hyvin erilaisia toimintoja ja rasittavat muistia eri tavoin. Myöhemmän kielenkehityksen aikana pitäisi oppia ymmärtämään näitä modaliteettieroja ja toimimaan kummankin ehdoin kielen rakenteen tasolla ja lisäksi sopivin kielen resurssein. Tässä luvussa olemme tarkastelleet Unelmakirjoitelmien avulla, miten kielen rakennetta tiivistetään oletuksenmukaisena lähtökohtana varhaisen kielenkehityksen aikana opittu puhutun kielen rakenne ja miten nuorempien ja vanhempien kirjoittajien tiivistämisen keinot eroavat toisistaan. Tiivistäminen toimii yksinkertaisesta kompleksiin sekä horisontaalisella eli rakenteita rikastavalla että vertikaalisella eli hierarkkisuuutta nostavalla tasolla. Rajoitimme tässä yhteydessä rakenteelliseen, objektiiviseen kompleksisuuteen ja sivuutimme toimijaan suhteutetun kompleksisuuden, joka edellyttäisikin kokeellista tutkimusotetta. Rakenteellinen kompleksisuus määriteltiin eri yksiköihin pakatun tiedon moninaistumisena siten, että yksinkertaiseen muotoon aina lisätään sitä tarkentavaa ainesta tai siihen alisteiseksi – tavalla tai toisella – merkittäviä uusia yksiköjä. Siten johtamaton sana, endosentrisen lauseke ja yksinkertainen vain välttämättömistä argumenteista koostuva tilaa, tapahtumaa tai tekoa ilmaiseva lause ovat yksinkertaisia yksiköjä ja niiden kompleksistaminen voi tapahtua lisäämällä niihin uusia sanoja tai formatiiveja eli kieliopillisesti ilmaistuja merkityksiä. Samat rakenteelliset keinot ovat kompleksistamisen välineitä tekstityypistä riippumatta, mutta suhteet vaihtelevat.

Tarkastelumme keskittyi jonkinasteisesti kertovaan tekstiin. Tavoitteena oli luoda kuva ikäkausittain tyypillisestä rakenteen kompleksistamisesta ja toisaalta löytää kehityksellisesti muuttuvia keinoja tiivistää ja

rikastaa kertovaa tekstiä. Onko tuloksena oleva teksti helppoa, selkeätä tai ymmärrettävää, ei ole määriteltävissä rakenteesta käsin asteikolla yksinkertainen–kompleksinen eikä tutkijan näkemys yksin riitä sitä kertomaan. Lyhyesti voi tulosten perusteella todeta, että iän myötä kompleksinen kirjoitetun kertovan tekstin rakenne alkaa muistuttaa sanoma-lehtikielen rakennetta.

Luvussa tarkasteltiin kompleksisuutta hyödyntäen niin sanotun T-yksikön käsitettä ja arvioitiin paitsi T-yksikön sanamäärien myös sen propositioon palautuvien alisteisten rakenteiden kasvua. Sanamäärätarkastelu keskittyi sanaluokkakokohtaisiin analyyseihin, joiden avulla pyrittiin selvittämään T-yksikön sanamäärien koostumusta, keskiarvoa (MLTU) ja kasvua. Hierarkiatasolla yksinkertainen T-yksikkö ilmaisee yhden merkityksen (väitteen), kompleksinen useampia siten, että nämä rakennetta kompleksistavat merkitykset ilmaistaan suhteessa pääväitteeseen alisteisina käyttäen keinoina joko sivulauseita tai ei-finiittisiä rakenteita. T-yksiköiden hierarkkista rakennetta tutkittiin niin sanottua kompleksisuusindeksiä (CI) hyödyntäen laskemalla, kuinka monia ja minkälaisia alisteisia propositioon palautuvia rakenteita niissä keskimäärin on. T-yksikön sanamäärän kasvu ja siihen sisältyvien alisteisten rakenteiden kasvava määrä toisaalta kertovat kompleksistumisesta. Kielenkäyttäjät itse on määritellyt T-yksikön rajat pisteitä käyttämällä ja ainoastaan päälauseita rinnastavat virkkeet on analyysissä erotettu kahdeksi yksiköksi pisteistä piittaamatta, koska ne eivät ole keskenään alisteisia. T-yksikkö siis kertoo, miten kielenkäyttäjät on halunnut rikastaa ja nivoa tapahtumia yhteen ja selittää niiden ajallisia, kausaalisia ja spatiaalisia suhteita. T-yksikkö on siten kehityksellinen yksikkö, ja analyysi paljastaa osaltaan tekstin globaalikoherenssia ja sen vähittäistä kehittymistä iän mukaan.

Sanaluokkakokohtaisessa analyysissä tutkittiin adjektiivien, substantiivien, pronominien ja verbien suhteellisia osuuksia, osuuksien muutoksia eri ikäryhmissä ja vaikutusta sanamääriin. Ikäryhminä olivat osaksi mukana kaikki alakoululaiset tai vain kuudennet, yläkoululaisista yhdeksäsluokkalaiset sekä nuoret aikuiset. Kaikkien tarkasteltujen sanaluokkien osuuksissa tapahtuu kehityksellisiä muutoksia: adjektiivien määrä nousee kehityksellisesti, ja kertovassa tekstissä tavoitemäärä on 6–10 prosenttia sanaesiintymistä. Naiskirjoittajilla adjektiivimäärä on kaikissa



ikäluokissa korkeampi kuin mieskirjoittajilla. Substantiivien määrä niin ikään kasvaa, kunnes niitä on suunnilleen 30–40 prosenttia sanaesiintymistä. Pronominien rooli on kahtalainen ja persoonapronominit ja muut pronominit vaikuttavat kompleksistumiseen eri tavoin. Persoonapronominit lisääntyvät kyllä nuorimmilla kirjoittajilla kehityksellisesti samalla kun kertovan tekstin määrä kasvaa, mutta ne eivät pidennä T-yksikköä eivätkä nosta kompleksisuutta. Muut pronominit määrittävät usein substantiiveja ja lisäävät silloin lausekkeiden sanamääriä. Pronominien tavoitemäärä kertovassa tekstissä on 5–10 prosenttia sanamäärästä. Verbien osuus kokonaissanamäärästä on kaikissa ikäryhmissä samantapainen: mediaani on hieman alle 30 prosenttia, ja kuudesluokkalaisista lähtien yleisyyden luokka 20–30 prosenttia on vallitseva.

T-yksikön keskipituuteen (MLTU) sanoina eri sanaluokat vaikuttavat eri tavoin. Lisäksi on ikäkausi- ja sukupuolivaihtelua. Adjektiivit nostavat nuorimmilla MLTU:ta mutta eivät enää niinkään muilla. Myös substantiivien vaikutus on sanamääriä nostava mutta enemmän nais- kuin mieskirjoittajilla. Pronominien vaikutus on positiivinen. Verbimäärät pysyvät samoissa, mutta jos verbiluokka jaetaan finiittisiin ja ei-finiittisiin, löydetään aineiston suurimmat kehitykselliset kompleksistumiserot: finiittiverbien määrä laskee iän myötä yli 20 prosenttiyksikköä ja ei-finiittisten määrä kasvaa. Kuudesluokkalaiset kirjoittavat aika vahvasti finiittiverbien varassa pää- ja sivulauseita käyttäen ja rikastavat tekstiään adjektiivein. Myös naiskirjoittajilla finiittiverbien määrä on yleensä korkeampi kuin mieskirjoittajilla. Finiittiverbien korkea määrä yleensä laskee T-yksikön sanamääriä, koska teksti etenee yhden proposition T-yksiköllä eli yksinäislauseilla. Tällaisen tekstin globaalikoherenssi täytyy lukijan yleensä pitkälti päätellä. Rektioinfinitiivit ja muut ei-finiittiset ainekset lisääntyvät kehityksellisesti, niiden tyyppimäärä on tiheysluokassa 10–20 prosenttia verbiesiintymistä. Muita ei-finiittisiä on peruskoululaisilla 5–10 prosenttia verbiluokasta, mutta aikuisilla määrä on jo yli 20 prosenttia. Verbiluokan sisällä siirtymä finiittisistä ei-finiittisiin nostaa T-yksikön sanamääriä ja on hyvin selvä kehityksellinen piirre. Ei-finiittiset ainekset tosin koostuvat nuorilla eri aineksista kuin vanhemmilla. Erityisesti kuudesluokkalaisilla on paljon modaaliverbeihin

liittyviä infinitiivimuotoja (esimerkiksi *haluan/saan mennä*), mutta iän myötä ei-finiittiset ainekset monipuolistuvat.

Jos T-yksikön sanamäärät kasvavat peruskoululaisilla pääosin kehityksellisesti, ne nuorimmat kirjoittajat, joilta puuttuvat määrittävät sanat kuten adjektiivit tai ylipäättään sanat, jotka eivät ole argumentiasemassa, eivät ole vielä saavuttaneet ikäluokkansa keskiaverto-osaamista. Tilanne on nuorilla aikuisilla moninaisempi, koska heillä T-yksikön sanamäärä-vaihtelu liittyy usein tyyliä vaihteluihin. Osa aikuisista kirjoittaa kompleksisesti, osa hyvin yksinkertaisesti; edellisillä on korkea sanamäärä ja kompleksisuus, jälkimmäisillä T-yksikössä on vähän sanoja ja samoin lauseita. Kehityksellisesti kompleksisempi tyyli on parempi mutta laadullisesti tyyliä ei voi arvottaa. Peruskoulun kuudensilla on vielä jonkin verran poikkeavia arvoja, peruskoulun tytöillä tulee esiin finiittiverbejä ja adjektiiveja suosiva kirjoitustyyli. Peruskoulun kuudennen ja yhdeksännen poijista odotuksenmukaista suurempi osuus edustaa niin sanottua suoraa viivaa tai nollalinjaa monen muuttujan suhteen: kirjoitelmissa ei ole mitään tekstiä rikastavaa tai kompleksistavaa ainesta ja muuttujien arvot eivät sijoitu nousevalle linjalle. Jopa substantiivi-verbisuhde voi olla epätavallinen, mikä antaa viitteitä siitä, että tavoite tai tulos ei ole kertovaa tekstiä. Jopa kymmenisen prosenttiakin näistä poijista kirjoittaa ongelmatekstiä, vaikka odotusarvo olisi korkeintaan 2–3 prosenttia.

Nominaaliset sanaluokat eivät yleensä vaikuta suoraan T-yksikön hierarkkiseen kompleksisuuteen mutta välillisesti esimerkiksi korkea substantiivimäärä voi olla seurausta useista alisteisista rakenteista. Vastavasti korkea pronominiäärä voi liittyä runsaaseen ei-finiittisten tai alisteisten lauseiden määrään. Adjektiivit eivät nosta kompleksisuutta mutta liittyvät nuorilla tytöillä kompleksiseen tekstiin. Ei-finiittiset ainekset ja sivulauseet nostavat kompleksisuusindeksiä keskeisesti sanamäärien lisäksi. Rektioinfinitiivien korkea määrä liittyy usein tulevaisuushakuiseen tekstiin, kirjoitetaan siitä, mitä halutaan tulevaisuudessa saada, minne mennä ja mitä tehdä, mutta muilla ei-finiittisillä rakenteilla voidaan käsitellä myös menneisyyttä ja tuoda esiin ajallisia kerrostumia. On selvä, että jälkimmäinen on kehityksellisesti merkitsevempi kuin edellinen.

Alisteiset sivulauseet nostavat sekä sanamääriä että kompleksisuusindeksiä kaikilla kirjoittajilla. Alisteiset lauseet jaettiin komplementti-, adverbiaali- ja relatiivilauseisiin. Komplementtilauseet ovat puheakti- ja mentaaliverbien lauseargumentteja, ne eivät vanhemmilla peruskoululaisilla enää lisäännny kehityksellisesti, mutta siirtymä referoivasta puheaktista propositionaalisia asenteita, tunteita ja ajatuksia kielentäviin mentaaliverbeihin on kehityksellinen. Lausemuotoisten adverbiaalilauseiden määrä laskee kehityksellisesti. Kertovassa tekstissä temporaaliset ja kausaaliset sivulauseet ovat keskeiset adverbiaalilauseiden ryhmät. Sekä komplementtilauseet että temporaalilauseet voidaan korvata lauseenvastikkeilla ja aikuiskirjoitelmissa temporaalirakenteita on kolmen kertaa enemmän kuin referentiaalisia rakenteita; niitä on erityisesti miesten kirjoitelmissa. Temporaalinen lauseenvastike kasvaa kehityksellisesti etuasemaisena, narratiivisena tai päälauseen kanssa myös takasemassa päällekkäistä tapahtumaa kielentävänä. Monista kielistä esitetty relatiivilauseiden kehityksellinen kasvu on totta myös suomen kielen kohdalla. Siirtymä sivulauseen kuvailevasta funktiosta narratiiviseen on kehityksellinen ja vastaavasti sekä ei-lineaaristen ja etuasemaisten relatiivilauseiden lisääntyminen.

Myöhemmän kielenkehityksen tutkimushankkeessa yksi lähtökohtaoletus oli, että tilastolliset eli allomorfiset sukupuolierot jakautuvat Gaussin käyrän mukaisesti siten, että enemmistö ihmisistä sijoittuu jakauman keskialueelle, jossa on paljon päällekkäisyyttä. Äärialueille sijoittuu 2–3 prosenttia ihmisistä. Sukupuolieroja on arvioitu hyvin monella tapaa hankkeessa ja tulosten yleistys on, että jakauma on monen ominaisuuden suhteen odotuksenmukainen. Erityisesti sanastossa ja sekä semanttisen että morfologisen tiedon osalta erot ovat suhteellisen pienet tosin siten, että naisilla hallinta kulkee nopeampaa tahtia kuin miehillä, ja siten, että miesten tuloksia sijoittuu ääriarvoille – asteikon molempiin päihin – enemmän kuin naisten. Nämä tulokset ovat eräänlaisia monissa yhteyksissä esiin tulleita yleisiä totuuksia. Tekstitasolla substantiivien ja verbien suhteellinen osuus ja T-yksiköiden keskipituus eivät ole sukupuolittain erottelevia tekijöitä mutta jossain määrin erilais-tavia ovat adjektiivien osuus, joka tytöillä ja naisilla on aina korkeampi kuin pojilla ja miehillä, ja toisaalta ei-finiittisten ainesten osuus verbi-

luokasta, joka pojilla ja mieskirjoittajilla on keskimäärin korkeampi kuin tytöillä ja naisilla. T-yksikköjen rakenteellinen kompleksisuus on niin ikään miehillä selvästi korkeampi kuin naisilla ja korkeampi kompleksisuus tulee esiin jo kuudennen luokan pojilla, vaikkakaan heillä ei aina ole kompleksisuuden ilmaisen keinoja hallinnassaan.

Unelmakirjoitelmat edustavat kehityksellisesti kertovaa tekstiä. Kuudesluokkalaisista kirjoittajista 60 prosenttia ei vielä täysin hallitse kertomusskeemaa – jonka hallintaa odotetaan jo 10-vuotiailta – ja heidän kirjoittamansa tekstit ovat paitsi lyhyitä, myös melko yksinkertaisia. Aiheita voi olla käsittelyssä useitakin, mutta yksittäisestä aiheesta kirjoitetaan vähän, mikä pirstaloi myös rakennetta. Yhdeksäsluokkalaiset kirjoittavat pääosin kuudesluokkalaisia pidempiä ja hieman kompleksisempia kertomuksia. Tyttöjen tekstit alkavat ominaisuuksiltaan muistuttaa nuorten aikuisten tekstejä, pojat tuottavat paljon myös ongelmatekstejä. Nuorten aikuisten kirjoitelmien rakenteelliset ominaisuudet muistuttavat monessa suhteessa sanomalehtikielen ominaisuuksia. Erot näkyvät parhaiten kirjoitelmien adjektiivien suuresta määrästä, korkeammasta verbitiheydestä ja referointiin liittyvien keinojen vähemmästä määrästä.

Unelmakirjoitelmien ja lehtitekstin erot syntyvät osin tekstityyppi-eroista. Unelmakirjoitelmat ovat pääosin kertomuksia – aikuisilla melko hyviä – ja eritteleviä tekstejä on melko vähän. Lehtiteksti koostuu monista tekstilajeista, vaikka kertovan tekstin osuus on suuri. Erityisesti uutiset on kirjoitettu kertomustyyliin, samoin monet raportit ja haastattelut. Lähinnä esimerkiksi arvioivat tekstit ja pääkirjoitustekstit edustavat erittelevää ja arvioivaa tekstityyppiä. Siten on ymmärrettävää, että nuorten aikuisten kirjoitelmilla ja sanomalehtiteksteillä on ehkä odotettua vähäisemmät rakenteelliset ja tilastolliset erot. Kun molempien näytteet on kerätty 2000-luvun alusta, ne edustanevat samaa kielellistä kompleksisuuden muotia. Rakenteellinen kompleksisuus näet paitsi kasvaa kehityksellisesti, myös vaihtelee ajallisesti. Siten 1970-luvun kompleksinen muoti poikkeaa 2010-luvun yksinkertaisemmasta muodista. Mutta jos rakenne on paitsi yksinkertainen myös hajanainen, kyse ei ole enää muodista vaan kehityksen ongelmista. Kuudesluokkalaiset kirjoittivat vuonna 2006 unelmapäivästään tiiviimmin kuin vuoden 2014 kuudesluokkalaiset, ero näkyy nimenomaan globaalikoherenssin

hajanaistumisena. Tämä on yksi merkki trendin laskusuuntaisuudesta suurten matalien ääriarvojen lisäksi.

Lopuksi voi kysyä, onko koululaisten kirjoitelmista ilmenevä rakenteellinen kielenhallinta odotuksenmukainen. Vastaus on kyllä, jos kiinnitetään huomio ikäryhmittäiseen keskimääräiseen osaamiseen. Vastaus on ei, jos kiinnitetään huomio ääriarvojen odotuksenmukaista suurempaan määrään. Vastaus on vähintään epäröivä, jos kiinnitetään huomio kompleksisuustekijöiden yksipuolistuvaan ja haparoivaan käyttöön. Ja lopuksi, vaikka heikkoa suoriutumista voi selittää esimerkiksi (ulkoisen) motivaation puutteella, kirjoittajaksi kehittyä vain kirjoittamalla. Siksi huolimatta kielenhallinnan tasosta, heikko kirjoittaminen voi olla opiskelun ja urapyrkimysten este vielä aikuisenakin.

## LÄHTEET

- Aksu-Koç, A. & Stutterheim C. 1994. Temporal relations in narrative: Simultaneity. Julkaisussa: R. A. Berman & D. Slobin (toim.) *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale: Lawrence Erlbaum Ass., 839–878.
- Beaman, K. 1984. Coordination and subordination revised: Syntactic complexity in spoken and written narrative discourse. Julkaisussa: D. Tannen & R. Freedle (toim.) *Coherence in spoken and written discourse*, Norwood: Ablex Publ., 45–80.
- Beers, S. F. & Nagy, W. E. 2009. Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing* 22, 188–200.
- Beers, S. F. & Nagy, W. E. 2011. Writing development in four genres from grade three to seven: Syntactic complexity and genre differentiation. *Reading and Writing* 24, 183–202.
- Berman, R. & Verhoeven, L. 2002. Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language and Literacy* 5: 1, 2–42.
- Berman, R., Nayditz, R. & Ravid, D. 2011. Linguistic diagnostics of written texts in two school-age population. *Written Language and Literacy* 14: 2, 161–187.
- Bybee, J. & Hopper, P. 2001. Introduction to frequency and the emergence of linguistic structure. Julkaisussa: J. Bybee & P. Hopper (toim.) *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 1–24.
- Cristofaro, S. 2014. Is there really a syntactic category of subordination. Julkaisussa: L. Visapä, J. Kalliokoski & H. Sorva (toim.) *Contexts of subordination. Cognitive, typological and discourse perspectives*. Amsterdam: John Benjamins, 73–91.
- Dahl, Ö. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
- Demberg, V. & Keller, F. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109, 193–210.

- Diessel, H. 2004. *The acquisition of complex sentences*. Cambridge: Cambridge University Press.
- Ehret, K. & Szmrecsanyi, B. 2016. An information-theoretic approach to assess linguistic complexity. Julkaisussa: R. Baechler & G. Seiler (toim.) *Complexity, isolation and variation*. Berlin: De Gruyter, 71–94.
- Givón, T. 1991. Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language* 15: 2, 335–370.
- Givón, T. 2009. *The genesis of syntactic complexity*. Amsterdam: John Benjamins.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. 2011. Coh-matrix: Providing multilevel analyses of text characteristics. *Educational Researcher* 40: 5, 223–234.
- Graesser, A. C., McNamara, D. S., Louwse, M. M. & Cai, Z. G. 2004. Coh-matrix: Analysis of text cohesion and language. *Behavior Research Methods, Instruments, & Computers* 36: 2, 193–202.
- Hall-Mills, S. & Apel, K. 2015. Linguistic feature development across grades and genre in elementary writing. *Language, Speech, and Hearing Services in Schools* 46, 242–255.
- Heinonen, S. 2018. *Narratiivisuuden sisällöllinen kehittyminen aiheanalyysin kautta tarkasteltuna: kvantitatiivinen tutkimus aiheiden esiintymisen vaihtelusta iän ja sukupuolen mukaan "Unelmiäni päivä" -kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Herlin, I. 1998. *Suomen kun*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Herlin, I., Visapää, L., Kalliokoski, J. 2014. Introduction. Julkaisussa: L. Visapää, J. Kalliokoski & H. Sorva (toim.) *Contexts of subordination. Cognitive, typological and discourse perspectives*. Amsterdam: John Benjamins, 1–17.
- Hyönä, J. & Vainio, S. 2001. Reading morphologically complex clause structures in Finnish. *European Journal of Cognitive Psychology* 13, 451–474.
- Itkonen, E. 2008. *Maailman kielten erilaisuus ja samuus. Osa I*. Turku: Turun yliopisto.
- Itkonen, E. 2009. *Maailman kielten erilaisuus ja samuus. Osa II*. Turku: Turun yliopisto.
- Itkonen, E. 2010. *Maailman kielten erilaisuus ja samuus. Osa III*. Turku: Turun yliopisto.
- Itkonen, E., Pajunen, A. & Suomi, K. 2009. Maailman kielten luokituksen ongelmia. *Virtittäjä* 113: 2, 285–296.
- Jagariah, T., Olinghouse, N. G. & Kearns, D. M. 2020. Syntactic complexity measures: variation by genre, grade-level, students' writing abilities, and writing quality. *Reading and Writing* 33, 2577–2638.
- Kauppinen, S. 1976. *Kielenkäyttötaito ja koululaisten kirjoitelmista mitattu virkerakenteen kehitys*. Helsinki: Helsingin yliopisto.
- Kuisma, J. 2011. *Ominaisuuden ilmausten käytön kehittyminen alakouluikäisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Laine-Leinonen, J. 2013. *Koulukorpuksen leksikko: 1.–6.-luokkalaisten aktiivinen sanavarasto*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Langacker, R. 2014. Subordination in a dynamic account of grammar. Julkaisussa: L. Visapää, J. Kalliokoski & H. Sorva (toim.) *Contexts of subordination. Cognitive, typological and discourse perspectives*. Amsterdam: John Benjamins, 17–72.
- Lapp, D., Moss, B., Grant, M. & Johnson, K. 2015. *A Close look at close reading*. Alexandria: Association for Supervision & Curriculum Development.

- Leikin, M. & Assayag Bouskila, O. 2004. Expression of syntactic complexity in sentence comprehension: A comparison between dyslexic and regular readers. *Reading and Writing* 17: 7–8, 801–822.
- Lu, X. 2009. Automatic measurement of syntactic complexity in child language acquisition. *International Journal of Corpus Linguistics* 14: 1, 3–28.
- Lundin, T. 1983. *Aikuisopiskelijoiden ja peruskoulun yläastetta käyvien oppilaiden kirjoitelmien virkerakenteesta*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- McNamara, D. S., Louwerse, M. M. & Graesser, A. C. 2002. Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. University of Memphis. [www.academica.edu/2878021/Coh-Metrix](http://www.academica.edu/2878021/Coh-Metrix). Viitattu 31.8.2016.
- Muhonen, M. 2019. *Kompleksiset lauserakenteet 6.-luokkalaisten Unelmieni päivä -kirjoitelmista*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Mäkinen, L., Loukusa, S., Nieminen, L., Leinonen, E. & Kunnari, S. 2014. The development of narrative productivity, syntactic complexity, referential cohesion and event content in four- to eight-year-old Finnish children. *First Language* 34: 1, 24–42.
- Naikola, T. 2012. *Yhdyslauseiden kehitys kouluikäisen kielessä: Tutkimus 1.–6.-luokkalaisten kirjoitelmista*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Nippold, M. A. 2006 [3. painos]. *Later language development: School-age children, adolescents and young adults*. Austin: Pro-Ed.
- Nippold, M. A. 2009. School-aged children talk about chess: Does knowledge drive syntactic complexity. *Journal of Speech, Language, and Hearing Research* 52, 856–871.
- Nippold, M. A., Hesketh, L. J., Duthie, J. K. & Mansfield, T. C. 2005. Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research* 48, 1048–1064.
- Nippold, M. A., Mansfield, T. & Billow, J. L. 2007. Peer conflict explanations in children, adolescents, and adults: Examining the development of complex syntax. *American Journal of Speech-Language Pathology* 16, 179–188.
- Noonan, Michael. 2007 [1985]. Complementation. Julkaisussa: T. Shopen (toim.) *Language typology and syntactic description. Volume III: Complex constructions*. Cambridge: Cambridge University Press, 52–150.
- Närhi, L. 1990. *Lausemaisuuksien asteet 9.-luokkalaisten kirjoitelmissa. Tyttöjen ja poikien tekstien vertailua*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Pajunen, A. 2001. *Argumenttirakenne. Asiantilojen luokitus ja verbien käyttäytyminen suomen kielessä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Pajunen, A. 2012. Kirjoittamistaitojen kehitys 8–12-vuotiailla: alakoululaisten unelma-kirjoitelmien. *Virittäjä* 116, 3–34.
- Pajunen, A. & Virtanen, P. 2002. Savon Sanomat 2001 [SS2000-korpus]. Morfosyntaktisesti analysoitu tietokanta, 8 miljoonaa sananmuotoa.
- Pajunen, A., Vainio, S., Hyönä, J. & Bertram, R. Tulossa. The role of derivative morphemes and lexical frequency in accessing transitivity information in reading: An eye tracking study in Finnish. Julkaisussa: A. Mäkilähde, T. Alho, V. Leppänen & A. Pajunen (toim.) *Unity and diversity of linguistics*. Routledge.

- Preiss, D. D., Castillo, J. C., Flotts, P. & San Martin, E. 2013. Assessment of argumentative writing and critical thinking in higher education: Educational correlates and gender differences. *Learning and Individual Differences* 28, 193–203.
- Price, J. R. & Jackson, S. C. 2015. Procedures for obtaining and analyzing writing samples of school-age children and adolescents. *Language, Speech and Hearing Services for Schools* 46, 277–293.
- Ravid, D. 2006. Semantic development in textual contexts during the school years: Noun scale analyses. *Journal of Child Language* 33, 791–821.
- Sadeniemi, M., Kettunen, K., Lindh-Knuutila, T. & Honkela, T. 2008. Complexity of European Union languages: A comparative approach. *Journal of Quantitative Linguistics* 15: 2, 185–211.
- Schneider, P., Vis Dubé, R., & Hayward, D. 2005. The Edmonton Narrative Norms Instrument. <http://www.rehabresearch.ualberta.ca/enni>. Viitattu 15.9.2018.
- Scontras, G., Badecker, W., Shank, L., Lim, E. & Fedorenko E. 2014. Syntactic complexity effects in sentence production. *Cognitive Science* 39: 3, 1–25.
- Sintonen, M. 2011. *Suunnittelemisen vaikutus kertomuksen rakenteeseen viides- ja seitsemäsluokkalaisilla kirjoittajilla*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Szmrecsanyi, B. M. 2004. On operationalizing syntactic complexity. Julkaisussa: G. Purnelle, C. Faron & A. Dister (toim.) *Le poids des mots. Proceedings of the 7th international conference of textual data statistical analysis*. Louvain-la-Neuve, March 10–12, 2004, Volume 2. Louvain-la-Neuve, Presses universitaires de Louvain, 1032–1039.
- Tiiri, M. 2021. *Lauserakenteiden kompleksisuus ammattikorkeakoululaisten kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Tomasello, M. 2003. *Constructing a language. A usage-based theory of language acquisition*. Cambridge: Harvard University Press.
- Truckenmiller, A. J. & Petscher Y. 2020. The role of academic language in written composition in elementary and middle school. *Reading and Writing* 33, 45–66.
- Vainio, S. 2006. Lausetason kokeellisesta tutkimuksesta: Tuloksia, haasteita ja tulevaisuuden näkymiä. *Puhe ja kieli* 26: 1, 3–21.
- Vesala, S. 2012. *Kerronnan keinot alakoululaisten teksteissä*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Vikainen, I. 1982. Lausetajun kehityksestä. Julkaisussa: M. Larmola (toim.) *Kouluikäisen kieli*. Helsinki: Suomalaisen Kirjallisuuden Seura, 45–56.
- Virolainen, S. 2003. *Alisteisten rakenteiden käyttö peruskoulun seitsemännen ja yhdeksännen luokan oppilaiden kirjoitelmissa*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto.
- Wierzbicka, A. 1980. *Lingua mentalis. The semantics of natural language*. Sydney: Academic Press.
- Wolf Nelson, N. & Van Meter, A. M. 2007. Measuring written language ability in narrative samples. *Reading and Writing Quarterly* 23: 3, 287–309.



## Kirjoittajat

**Honko, Mari**, FT, soveltavan kielentutkimuksen dosentti, tutkijatohtori, Jyväskylän yliopisto.

🔗 <https://orcid.org/0000-0002-3240-622X>

**Kouvo, Anna-Kaisa**, FM, äidinkielen ja kirjallisuuden lehtori.

**Laasanen, Mikko**, FM, yleinen kielitiede, FT suomen kieli, Tampereen yliopisto.

🔗 <https://orcid.org/0000-0003-2654-7146>

**Niinimäki, Anneli**, FT, äidinkielen lehtori.

🔗 <https://orcid.org/0000-0003-0050-10>

**Pajunen, Anneli**, FT, yleisen kielitieteen dosentti, suomen kielen professori, Tampereen yliopisto.

🔗 <https://orcid.org/0000-0001-8597-7491>

**Tuuri, Emilia**, FM, väitöskirjatutkija, Tampereen yliopisto.

🔗 <https://orcid.org/0000-0002-8298-6875>

**Vainio, Seppo**, FT, yleinen kielitiede ja psykologia, Turun yliopisto.

## Abstract

*Later language development in Finnish. School-age children and young adults*

The topic of the book is the incremental growth of linguistic knowledge from lexical to structural-cum-textual during the so-called later language development. Language mastery does not presuppose any acquaintance with prescriptive grammar but, instead, concerns the core of language which the so-called consensus principle applies to: the most frequent words and structures are mastered with certainty by everybody, but uncertainty increases as less frequent and more variable phenomena are taken into consideration. It is the goal of the study to make explicit the knowledge that is common to school children of different age groups, and to show how it develops both in its core and in its fringe areas. The mastery of less common aspects exhibits considerable statistical variation. The research embodies methodological pluralism insofar as it has been carried out by means both of the corpus method and the experimental method. Here experimental subsumes writing tasks, paper-and-pencil tests, and behavior under experimental conditions. The amount of participants native in Finnish varies from 300–2000. The book has a bipartite structure: mastery of meanings (Part I), and mastery of forms (Part II).

## Hakemisto

### A

abstraktiotaso 102, 390, 424  
abstraktistuminen 69, 91, 102, 437  
  abstraktistumiskehitys, ks. kehitys  
adjektiivi 69, 77, 85–88, 96–99, 100, 104,  
  202, 272, 274, 275, 374, 382, 390, 395, 396,  
  428, 429, 431–438, 442, 464–467, 471–475  
adverbi 33, 34, 43, 69, 77, 85–88, 96, 98, 99,  
  104, 111, 129, 175, 202, 208, 261, 262, 270,  
  274–278, 280, 291, 374, 383, 384, 390, 391,  
  395, 396, 417, 431  
adverbiaali 124  
aikajärjestys 409, 463  
aikajäsennys 169, 175  
aikamuoto 168, 232  
aikarakenne 34  
aikasuhte 405, 461  
aisti 90, 249, 470  
akseli 134, 135, 149  
aktiiviteetti (*activity*) 88, 109, 113, 116  
ala (*region*) 110, 121  
alustus, ks. myös lause 403, 426, 457  
analogia 14, 102, 287  
ankkurointi 160  
appellatiivi 374, 391, 436, 437  
argumentti 78, 392, 422, 426, 431, 470  
  lauseargumentti 456, 474  
artikkeli 74, 159, 160, 170, 189, 190, 194, 196,  
  395  
asteikko 20, 21, 30, 40, 48, 89, 93, 94, 159,  
  160, 186, 193, 241, 372, 389, 390, 431, 442,  
  471, 471  
annettuusasteikko (*givenness scale*) 159,  
  162, 192, 193  
elollisuusasteikko 89  
koherenssiasteikko 160, 170  
kompleksisuusasteikko 41, 371, 372, 396  
konkreettisuusasteikko 89–94, 100, 396  
saavutettavuusasteikko 193  
taitotasoasteikko 373  
astevaihtelu 287, 331, 334, 335, 338, 340

### D

*de dicto* 201  
*de re* 201  
deiksis 127, 136, 200, 212, 241  
deiktinen elementti 60, 232, 234, 241  
deiktinen keskus 232  
derivaatio 13, 14, 70, 95, 382, 383  
dimensio 97, 98, 252  
diskursisreferentti 40, 156–160, 164, 191  
distribuutio  
  komplementaarinen distribuutio 288,  
  289, 290, 295  
  semanttinen distribuutio 114, 131, 134,  
  143, 148  
diversiteetti  
  leksikaalinen diversiteetti 15, 43, 51, 64,  
  69, 80–84, 98, 100, 101, 123, 135, 202,  
  242, 247, 273, 327, 362, 373, 395, 412, 416,  
  431, 432, 436  
diversiteettitunnusluku, ks. myös indeksi  
  68, 71, 80, 347, 377

### E

elisitaatio 116, 117, 123, 127, 150, 151, 206,  
  239  
  kuvaelisitaatio 205–207, 229, 239  
elisitaatiokehys 109, 110, 111, 116–118, 129,  
  136, 140, 144, 148  
ennuste 18, 41, 65, 251, 261, 263, 267, 268,  
  283, 284, 286, 292, 294, 394  
  muutosennuste 273, 286  
ensikieli 8, 9, 18, 22, 352, 353  
esitys  
  eläytymisesitys 203, 236  
  epäsuora esitys 45, 200, 201, 203, 204,  
  205, 206, 207, 208, 211, 212, 213, 224, 225,  
  226, 229, 230, 231, 232, 234, 235, 236, 237,  
  239, 241, 461  
  sekaesitys 231, 232, 233  
  suora esitys 45, 201, 202, 204, 205, 2006,  
  2007, 213, 224, 225, 226, 228, 229, 230,

231, 232, 239, 240, 241, 461  
vapaa epäsuora esitys 45, 203, 209, 211,  
212, 224, 231, 233, 234, 235, 236, 238, 241  
Eurooppalainen viitekehys 351, 352

## F

FDG-jäsenin 28, 269, 274, 419  
fiktio 27, 46, 223, 268, 270, 271, 275, 276,  
277, 278, 279, 281, 283, 284, 286, 289, 291  
formatiivi 424, 426, 470  
frekvenssi  
perhefrekvenssi 65, 66, 67  
pintafrekvenssi 377, 378  
taivutusmuotofrekvenssi 72  
taivutusperhefrekvenssi 66, 68  
frekvenssittekijä 65, 68, 69, 71, 72, 80, 100

## G

geminaatta, ks. myös konsonantti 311, 330,  
331, 334, 335, 339, 340, 341  
grafeemi 254, 310, 313, 320, 335, 341, 346  
grammi 111, 115, 120–135, 137–140, 142, 143,  
146–149, 151  
väylägrammi 122, 123, 142, 148  
granulariteetti 40, 93, 95, 99, 100, 102, 114,  
115, 122, 135, 137, 150, 381, 383

## H

hallinta 9–11, 13, 14, 16–22, 25, 26, 28, 29,  
35, 40, 41, 42, 45, 46, 48, 49, 54, 55, 59,  
60, 61, 64, 65, 67, 71, 91, 95, 100, 102,  
105, 109, 115, 158, 167, 178, 179, 181, 192,  
193, 200, 216, 225, 238, 241, 244, 247, 254,  
261, 263, 264, 266, 273, 299, 308, 311,  
314–319, 322, 327, 332, 333, 336, 341, 349,  
350, 351, 358, 385, 425, 474, 475  
kielenhallinta 9–11, 13, 14, 20, 23, 24, 26,  
30, 40–42, 47–49, 54, 63, 68, 86, 108,  
165–167, 242, 248, 285, 295, 311, 318, 323,  
329, 333, 350, 352, 355, 359, 369, 370, 385,  
411, 453, 476  
natiivihallinta 264  
sanastonhallinta 29, 61, 64, 65, 85, 106,  
347, 349, 377

semanttis-pragmaattinen hallinta 167, 192  
hierarkia 62, 86, 88, 94, 95, 161, 162, 180,  
372, 421, 471  
saavutettavuushierarkia (*accessibility  
hierarchy*) 163, 164, 172  
hierarkiataso  
geneerinen taso 86, 127, 136, 176, 177, 181,  
188, 185  
perustaso 86, 94, 95, 373  
spesifitaso 94  
HL-lekseemi (*hapax legomenon*) 88, 92, 431,  
432, 436, 437, 445  
*Holistic spatial semantics* 109

## I

ikäkausi 11, 12, 26, 48, 200, 203, 221, 248,  
370, 412, 472  
ikätaiso 15, 20, 23, 62, 71, 320, 381, 411, 463  
aikuistaso 9, 109, 167, 192, 247, 369  
ikäluokkataso 23, 24, 62, 108, 109, 425,  
465  
väestötaso 12, 20, 21, 23, 27, 49, 385  
yksilötaso 12, 21, 23, 43, 83, 315, 316, 323,  
327, 368, 425, 436, 454, 456  
ilmenemismuoto 199, 200, 211, 249, 252,  
253, 257, 320  
imitatiivi 202, 207  
indeksi  
D-indeksi 347  
kompleksisuusindeksi 375, 403, 404,  
406–408, 412, 413, 427–432, 435, 436,  
440, 441, 444, 445, 447, 448, 450, 452,  
454, 455, 461–464, 467–469, 471, 473, 474  
leksikaalinen indeksi 377, 380, 391, 412,  
466, 468  
pronomini–substantiivi-indeksi 466, 469  
sanaluokkaindeksi 24, 41, 371, 374, 375,  
394, 390–392, 395, 396, 398, 399, 466,  
469  
substantiivi–verbi-indeksi 84, 412, 441  
indeksisana 46, 47, 268, 280, 281, 282, 283,  
284, 286, 291, 292  
infinitiivi 342, 375, 404, 407, 431, 433, 447,  
452, 456, 464, 465, 467, 473

- rektionifinitiivi 447, 448, 450–452, 456, 465–468
- informaatio, ks. myös tieto 84, 157, 164, 297, 285, 374, 412, 437, 470
- spatiaalinen informaatio 130, 134, 138, 139
- uusi informaatio 158, 159, 163, 164, 170
- väyläinformaatio 124, 150
- intertekstuaalisuus 45, 199, 200, 204, 207, 208, 211, 212, 213, 214, 216, 239, 240, 246
- ironia 199, 200
- J
- jakso 16, 17, 157, 162, 182, 226, 369
- jaksotus 163
- johdin 13, 68, 76, 336, 379, 382
- johdos 13, 14, 64, 66–68, 70, 72, 86, 97, 336, 373, 375, 378, 382–386, 411–413, 428, 437, 466, 468, 469
- johtaminen, ks. myös derivaatio 13, 14, 67, 85, 97, 379
- johtoilmaus, ks. myös lause 211, 225, 228, 235, 238
- juoni 100, 168, 174, 184, 185, 187, 202, 203, 206, 226, 240, 241
- juonirunko, ks. myös skripti 156, 174, 175, 187, 399
- K
- kantasana 66–68, 383, 390
- kategoria 63, 88, 105, 110, 121, 128, 168, 232, 236, 251, 272, 278, 427, 431
- kattavuus 65, 71, 73, 74, 76
- kausaalisuhde 409, 423, 471
- kehitys 60, 71, 75, 82, 86, 91, 101, 149, 158, 159, 165, 166, 173, 179, 189, 191, 199, 200, 204, 207, 208, 216, 217, 229, 238, 242, 247, 248, 251, 252, 260, 266, 267, 268, 291–293, 296, 305, 313, 320, 323, 326, 327, 367, 368, 370, 371, 373, 374, 383, 388, 390, 394, 411–414, 421, 425, 460, 468, 477
- abstraktistumiskehitys 43
- kielenkehitys 9–18, 23, 42, 46, 59, 111, 208, 217, 247, 250, 251, 296, 305, 309, 312, 326, 367, 369, 370, 371, 382, 470, 474
- kompleksistumiskehitys 344, 367, 373, 461
- kehityskaari 10, 239, 370
- kehitystrendi 48, 101, 167, 427
- kehystys 150, 207
- kehystäminen 19, 25, 119, 368, 384, 406
- kerronta 25, 29, 33, 34, 40, 45, 46, 49, 86, 99, 100, 101, 135, 155, 166, 169, 174, 175, 188, 199, 200, 202, 203, 205, 209, 217, 219, 224–229, 231–233, 235–238, 355, 367, 378, 380, 381, 392, 394, 408, 409, 431, 444
- minä-kerronta 209, 392, 440, 444
- kerrontahetki 211, 230, 232, 241
- kertoja 45, 157, 168, 174, 175, 184, 199, 200, 201, 208, 209, 211, 214, 215, 217, 219, 221, 222, 227, 229, 230–237, 241, 367, 442
- minä-kertoja 45, 209, 219, 231, 234–237, 328
- kertomus 17, 34, 45, 61, 99, 100, 104, 117, 155, 156, 158, 166–170, 172–176, 182–188, 194, 197, 199, 201, 206, 208, 209, 210, 214, 216–222, 226, 228–231, 233, 235, 240, 242, 310, 323, 328–330, 345, 352, 370, 371, 399, 400, 417, 419, 420, 437, 448, 469, 475
- kertomusrakenne 16, 17, 33, 40, 44, 45, 99, 100, 155, 156, 158, 160, 162, 166, 169, 178, 181, 188, 192, 225, 238, 242, 319, 322, 323, 332, 367, 368, 399, 413
- kertomusrakenteen osa
- aloitusosa 200, 205, 209, 212, 218–220, 368
- lopetusosa 200, 205, 212, 218, 220, 240, 368
- orientaatio-osa 162, 218, 219, 226, 227
- kesto, ks. myös virhe 310, 311, 325, 332, 338, 339, 340
- äännekesto 310, 311, 333, 335, 337, 340, 359
- kielen sääntö 14, 204, 226, 256, 260, 264, 266, 316, 358, 360
- kielenmuutos 248, 250–254, 293, 295
- kielenoppiminen 80, 250, 254, 257, 305, 352, 356, 426
- kielentutkimus 252, 253, 256–258, 308
- kielentäminen 10, 11, 13, 15, 17–19, 21, 24,

- 25, 28, 29, 40–43, 45, 46, 51, 59, 60, 63, 99, 103, 108–112, 115, 120, 132, 133, 135, 137, 139, 150, 151, 157, 202–205, 238, 239, 242, 245, 247  
kielentämissstrategia, ks. strategia  
kieli  
analyttinen kieli 73, 259, 396  
kirjoitettu kieli 8, 9, 11, 15, 16, 18, 19, 25, 28, 45–49, 59, 63, 157, 158, 167, 190, 194, 201–205, 208, 216, 226, 238, 239, 241, 242, 247–258, 266–268, 270, 287, 288, 291–296, 311, 315, 316, 320, 328, 329, 332, 337, 351, 352, 355–357, 360, 377, 404, 415, 421, 470  
puhuttu kieli 8, 9, 11, 12, 15, 16, 18, 19, 28, 41, 45–48, 59, 65, 72, 84, 188, 189, 191, 201–203, 205, 207, 226, 229, 230, 232, 238, 241, 247–258, 260, 266, 267, 276, 280, 283, 285–288, 292–295, 310, 338, 353, 359, 362, 367, 388, 392, 408, 421, 448, 470  
kielihäiriö 36, 47, 386  
kielioppi 9, 16, 85, 111, 191, 224, 236, 242, 250, 256, 257, 259, 264, 269, 288, 309, 315, 375, 377  
kielitaito 15, 24, 47, 61, 304–306, 308–310, 312, 315–317, 323, 346, 347, 349, 352–358  
kielitieto 9, 10, 23, 30, 38, 264, 310  
intuitiivinen tieto 8  
sanatieto 11–15, 26, 29, 30, 43, 59, 60, 61, 63, 66  
tiedostamaton tieto 8  
tietoinen tieto 8  
yhteinen tieto 9, 157, 191  
kielitypologia 44, 128, 387  
kielto, ks. myös verbi 396, 448  
kieltomuoto 343, 345, 448, 452  
kiintopiste 44, 110, 111, 113, 114, 117, 119, 120–126, 128–134, 136–146, 148, 149  
kirjoitustaito 10, 14, 23, 63, 83, 101, 194, 197, 216, 238, 242, 303–305, 307, 310, 312, 329, 332, 348–351, 353–356, 359, 360  
kiteytymä 77, 101, 270, 273, 287, 391  
koheesio 164, 166, 322  
koherenssi 17, 25, 40, 44, 49, 59, 156–160, 164, 180, 181, 188, 191–193, 205, 227, 238, 247, 368, 382, 399, 404, 405, 409, 423, 461  
globaalikoherenssi 41, 368, 427, 471, 472, 475  
kausaalinen koherenssi 368  
kielellinen koherenssi 45, 238  
koherenssiasteikko, ks. asteikko  
referentiaalinen koherenssi 19, 25, 60, 156, 158, 160, 164, 165, 167, 180, 188, 191, 368  
spatiaalinen koherenssi 25, 368  
tekstikoherenssi 159, 192, 193  
temaattinen koherenssi 25  
temporaalinen koherenssi 25  
kohteliaisuusoletus 162  
kokemuksellisuus 208, 209, 226, 239  
kompetenssi 213, 217, 264, 265  
kompleksistuminen, ks. myös kehitys 16, 19, 41, 323, 368, 369, 371–373, 378, 412, 415, 428, 471, 472  
kompleksisuus 35, 49, 55, 108, 117, 143, 246, 248, 312, 323, 328, 329, 338, 348, 355, 363, 371, 372, 375, 390, 391, 403, 407–409, 411, 413, 422–427, 431, 433, 434, 442, 444–449, 453, 454, 456, 458, 459, 464, 465, 467–473, 475  
kompleksisuusindeksi, ks. indeksi  
leksikaalinen kompleksisuus 382  
objektiivinen kompleksisuus 422, 424, 470  
rakenteellinen kompleksisuus 35, 166, 313, 375, 387, 402, 403, 406, 442, 424, 425, 430, 446, 452, 457, 470, 475  
syntaktinen kompleksisuus 19, 371, 414, 426, 427  
toimijaan suhteutuva kompleksisuus 422, 424, 470  
kompleksisuustekijä 376, 409, 425, 464–468, 476  
konditionaali 214–216, 235, 240, 352  
kongruenssi 265, 285, 299, 332  
konjunktio 69, 74, 85, 86, 87, 230, 232, 337, 374, 375, 389, 391, 395, 427, 457, 458, 459, 463, 467

- konsonantti, ks. myös geminaatta 318, 329, 330, 331, 334, 335, 337, 338, 339, 340, 341
- konstruktio 102, 109, 244, 343, 352, 364, 365, 424
- kontaminaatio 287
- koodaus 26, 33, 268, 269, 318, 371, 374, 407
- koululainen 31, 38, 39, 43, 45, 64, 72, 75–80, 82, 88, 92, 93, 95, 98, 99, 101, 102, 208, 215, 217, 221, 231, 236, 252, 262, 263, 291, 304, 316, 320, 327, 349, 352, 353, 369, 388, 476, 477
- alakoululainen 30, 35, 41, 64, 80, 94, 96, 123, 159, 160, 187, 188, 197, 200, 202, 213, 224, 241, 246, 318, 319, 320, 321, 329, 336, 341, 345, 349, 359, 363, 369, 371, 373, 383, 392, 396, 399, 400, 403, 417, 433, 436, 457, 471
- ammattikorkeakoululainen 31, 32, 38, 71, 80, 83, 226, 376, 381, 420, 432, 458, 460
- peruskoululainen 10, 11, 14, 15, 21, 23, 26, 29, 30, 31, 34–37, 40, 43, 48, 49, 71, 76, 79, 82, 101, 202, 207, 208, 233, 268, 269, 367, 369, 370, 371, 376, 379, 380, 384, 385, 388, 395, 396, 408, 424, 427, 429, 432, 439, 448, 459, 463, 472–474
- yläkoululainen 30, 35, 41, 64, 135, 211, 242, 273, 369, 376, 403, 471
- kuvakulma 118, 135, 138, 141, 143, 144
- kuvaus 62, 110, 138, 139, 142, 173–175, 177, 182, 183, 185–188, 221, 235–237
- kuviteltavuus 89, 90
- kvanttori 444
- kypsyys 17, 157, 371
- kysely 26, 27, 35, 39, 304, 317, 333, 351, 353, 354
- käsiteanalyysi 26, 68
- käyttöpohjainen kielenoppimisteoria 80, 317, 318
- L
- lause
- adverbiaalilause 375, 457, 458, 460–462, 465, 474
  - alustuslause 458
  - että*-lause 230, 231, 237, 460, 461
  - intransiiviilause 422, 426
  - johtolause 202, 204, 207, 212, 227, 228, 230, 242
  - jotta*-lause 424
  - jälkilause 235
  - kehyslause 200, 226, 233, 240
  - ketjulause 15
  - kiilalause 461
  - komplementtilause 375, 404, 439, 456, 458, 460, 462, 474
  - kopulalause 16, 374, 422, 431
  - kun*-lause 242, 424
  - kysymyslause 232, 461
  - liikelause 118, 119
  - nominaalilause 221
  - peruslause 16, 84, 374, 422, 426, 429, 437, 448, 465
  - päälause 322, 375, 402, 404, 407, 408, 426, 461, 471, 474
  - relatiiviilause 74, 375, 403, 458, 460, 461, 462, 463, 465, 474
  - sivulause 403, 404, 412, 421, 439, 442, 447, 449, 457, 458, 459, 460, 461, 462, 463, 471, 472, 473, 474
  - transiiviilause 426
  - verbilause 374
  - yhdyslause 233
  - yksinäislause 426, 427, 446, 472
- lauseenjäsen 259
- lauseenmuodostus 78, 405
- lauseenvastike 49, 231, 261, 275, 280, 289, 330, 345, 375, 404, 408, 421, 439, 448, 453, 454, 458, 459, 461, 463, 474
- lauseke 131, 132, 134, 158, 172, 173, 176, 178, 190, 221, 322, 323, 330, 368, 371, 372, 407, 411, 418, 421, 422, 426, 437, 467, 470, 472
- nominilauseke 322, 343
- substansiiviilauseke 172, 192, 193, 418
- verbilauseke 373
- lauseluettelo 34, 458
- lausetyyppi
- peruslausetyyppi 426
  - sivulausetyyppi 352, 460, 461

- lehtikieli 46, 79, 269, 270, 271, 273, 276–279, 281, 283, 288, 291, 292, 294, 377, 445, 448, 451, 453, 456, 457, 458, 462  
sanomalehtikieli 72, 101, 103, 267, 268, 270, 277, 278, 291, 377, 379, 384, 392, 398, 408, 424, 442, 445, 446, 459, 462, 471, 475  
lekseemi 28, 32, 64–67, 69, 70–80, 86, 88, 91, 92, 94, 95–98, 101, 123, 134, 143, 145, 148, 227, 272, 274, 289, 321, 323, 342, 345–347, 379, 383, 384, 395, 412, 436, 445  
leksikalisointi 115, 117  
leksikko 59, 62, 80, 165, 208, 238, 350  
liike 29, 34, 44, 96, 108–114, 116–119, 121–124, 127, 128, 131–133, 138, 142, 144, 146–149  
liikerata 44, 112, 114, 117, 122–125, 128, 133, 135, 138–140, 142, 146–148, 150  
liiketilanne 29, 43, 44, 108–110, 112, 113, 115, 117, 118, 123–125, 133, 135–137, 140, 142, 143, 147–149  
liitepartikkeli 66, 272, 336, 340  
liitteettömyys 250, 260, 261, 265, 336  
lista 169, 182, 323, 399, 400  
lokaatio 110, 113, 114, 122  
  translokaatio 110, 113, 119  
luettelo 24, 34, 83, 173, 174, 175, 176, 182, 183, 185, 186, 187, 188, 381, 386, 391, 398, 399, 400, 458  
lukeminen 11, 18, 36, 62, 217, 238, 303, 305, 310, 326, 327, 348, 349, 359, 365, 418, 422  
luokitus, semanttinen 65, 68, 85, 88, 92, 96  
luokka, semanttinen 15, 69, 89, 90–99, 102, 110, 282, 431
- M**  
maininta  
  ensimaininta 120, 172, 173, 178–180, 191, 193, 194, 285  
  uudelleenmaininta 40, 163, 172, 173, 177, 178, 180, 187, 190, 193, 194, 285  
mallintaminen 14  
merkintä 85, 157, 164, 170, 192, 204, 207, 241, 311, 318, 333, 336, 338, 339, 340, 341, 372, 422, 423, 425, 426
- mielen teoria (*theory of mind*) 157, 165, 192, 200  
minä  
  kertova minä 208, 209, 211, 218, 236, 239  
  kokeva minä 208, 209, 211, 219, 236, 237  
MLTU (*mean length of terminal units*), ks.  
  myös pituus 375, 403, 406, 409, 410, 412, 413, 427–429, 435, 436, 440, 441, 444, 445, 463–467, 469, 471, 472  
modaliteetti 249, 250, 266, 388  
modaliteettiero 8, 16, 46, 203, 251, 295, 470  
modus 216, 232, 240, 343  
monimuuuttujamenetelmä 26, 386  
moninäkökulmaisuuus 45, 199  
monimorfeemisuus 313, 355, 381, 382  
morfologia 9, 19, 22, 28, 42, 314  
  derivaatiomorfologia 19, 20, 383, 385  
  taivutusmorfologia 318, 322, 357  
motivaatio 18, 32, 37, 48, 49, 88, 215, 223, 348, 356, 360, 363, 370, 377, 380, 387, 388, 413, 453, 476  
  sisäinen motivaatio, ks. myös pystyvyys-  
  usko 32  
  ulkoinen motivaatio 18, 23, 32, 36, 37, 87  
murre 28, 68, 129, 240, 251, 252, 255, 258–264, 270, 274, 276, 278, 280, 287, 288  
muunnos 241, 352, 382  
muutos  
  deiktinen muutos 204, 206, 207, 241  
  elinikäinen muutos (*lifespan change*) 12, 254  
  kehityksellinen muutos 88, 92, 96, 377, 394, 471  
  meneillään oleva muutos 46, 231, 248, 250–254, 258, 293, 295  
  paikanmuutos 108, 110, 113  
  tilanmuutos 113, 116  
myöhempi kielenkehitys 8–16, 18, 42, 46, 59, 111, 188, 208, 217, 247, 250, 251, 312, 367–370, 382, 470, 474
- N**  
nimeäminen 29, 63, 102, 236  
nominiaalimuoto 262



- nominalisaatio 421
- normi 9, 225, 231, 233, 234, 264, 310, 311,  
320, 337, 351, 352, 360, 371
- normijärjestelmä 261, 264
- norminmukaisuus 29, 35, 47, 53, 247, 318,  
356, 363
- normipoikkeama 41, 319, 320
- NTVR (*noun to verb ratio*), ks. myös indeksi  
84, 374, 390–394, 397, 399, 400, 401, 410,  
412, 413, 437–440, 444, 445, 464, 466,  
468, 469
- näennäisaika 295
- näennäisaikamethodi 27, 253, 254
- näennäisaikatutkimus 253, 254, 261, 295
- O
- objekti
- olio ja esine -objekti 47, 59, 88–90, 92,  
115, 120, 132 137
- kieliopillinen objekti 113, 124–126, 128,  
139, 144–148, 163, 274, 332, 342, 343, 345,  
346, 352, 372, 375, 427
- oikeinkirjoitus 20, 232, 302, 309, 310, 311,  
314, 320, 323, 332, 336, 337, 341, 346, 357,  
364, 369, 383, 389
- omaelämäkerta 34, 99, 203, 226, 368, 396
- omistusliite, ks. myös possessiivisuffiksi  
258, 260, 262, 263, 265, 333
- omistussuhde 259, 280–286, 291, 292
- onnistuneisuusarvio 368
- osa–kokonaisuussuhde 89, 286
- osallistuja 18, 45, 113, 160, 169, 175, 199,  
203, 399
- P
- paikka 16, 88, 96, 98, 99, 108, 113, 114, 123,  
129, 135, 136, 138, 140, 141, 142, 144, 145,  
149, 155, 160, 168, 172, 174, 214, 368, 384,  
391, 399
- paradigma 77, 372, 374, 391
- partisiippi 70, 72, 275, 375, 391, 431, 448,  
452–456, 464, 465, 467
- agenttipartisiippi 279, 289, 290
- peittoluku 101
- perhe
- derivaatioperhe 67, 68
- perhefrekvenssi, ks. frekvenssi
- sanaperhe 68
- substantiiviperhe 66
- taivutusperhe 66, 67, 72, 78–80
- verbiperhe 66
- perspektiivi 60, 200, 230, 241, 258
- pituus 71, 83, 157, 313, 329, 340, 345, 348,  
373, 377, 379, 381, 413
- keskipituus, ks. myös MLTU 403, 404,  
406, 412, 428, 430–432, 435, 448, 451,  
453, 464, 465, 472, 474
- lausepituus 419, 423
- sanapituus 68, 312, 313, 338, 348, 371, 373,  
375, 378–381, 383, 386, 406, 411, 412
- tekstipituus 81, 83, 342
- polaarisuus 89
- possessiivisuffiksi, ks. myös omistusliite  
19, 25, 28, 41, 46, 47, 173, 250–252, 254,  
258–268, 270–282, 284–287, 289–295, 333
- produktiivisuus 14, 68, 268, 386
- profiili 18, 101, 349, 360, 364
- prominenttijärjestys 163
- pronominaali 97, 98, 383, 391, 395, 396,  
432, 441
- pronomini 45, 85, 86, 87, 129, 156, 160,  
164, 165, 166, 170, 172, 173, 180, 189, 191,  
192, 193, 194, 204, 259, 260, 264, 272,  
274, 275, 288, 374, 391, 392, 429, 431, 438,  
441, 442, 444, 445, 464, 465, 466, 467,  
468, 471, 472
- demonstratiivipronomini 171, 173, 189,  
194, 441
- indefiniittipronomini 166, 171, 274, 345,  
441, 442
- interrogatiivipronomini 427, 441, 442, 458
- persoonapronomini 157, 172, 232, 259,  
260, 267, 286, 292, 297, 442, 472
- pronominitiheys, ks. tiheys
- relatiivipronomini 74, 427
- universaalipronomini 173, 178, 179
- propositio 375, 402–404, 471, 472
- propositionaalinen asenne 96, 447, 456, 474

- propri 70, 374, 391, 396, 436, 437  
prosessointi 43, 65, 164, 165, 167, 265, 293,  
312, 338, 348, 359, 404, 423, 426, 470  
prosessointinopeus 19, 413  
puhe 9, 18, 46, 156, 158, 170, 200, 202, 203,  
205, 207–209, 211, 215, 217, 222, 224, 225,  
228–233, 235, 236, 242, 254, 257, 306, 307,  
311, 312, 357, 391, 404  
puheakti 211, 224, 228, 236, 456, 461  
puhehetki 9, 230, 232, 233  
puhekieli 150, 189, 250–252, 255, 259, 260–  
263, 265, 266, 294, 314, 318, 320, 321, 328,  
329, 333, 355  
puhekielisuus 215, 315, 318, 319, 320, 321,  
324, 328, 329, 330, 333, 334, 335  
puhetilanne 18, 200, 230–232, 234, 241  
puheyhteisö 164  
puhuttelu 212, 213, 218, 224, 240  
pystyvyyssusko (*self-efficacy*), ks. myös moti-  
vaatio 18, 23, 36, 223, 413  
minäpystyvyys 360
- R
- rajaamattomuus 112, 125, 131, 134  
rajanylitys (*boundary crossing*) 111, 115, 116,  
120, 127, 128, 129, 137, 141, 142, 144, 145  
rajanylitysilmäus 116, 127  
rajanylityskonteksti 116  
rajanylitysrajoite 116  
rajanylitystilanne 149  
rajattuus (*boundedness*) 43, 44, 109–114, 117,  
124, 125, 134, 138, 143, 147  
rakenne  
argumenttirakenne 97, 424, 426  
agenttirakenne 278  
esittelyrakenne 156, 170  
*jotta*-rakenne 278, 279  
kokonaisrakenne 17, 68, 193, 218, 244,  
312, 316, 322, 325  
lauserakenne 15, 16, 41, 48, 170, 200, 245,  
246, 328, 330, 331, 344, 346, 358, 371, 375,  
382, 402, 403, 412, 422, 426–428, 433,  
436, 458, 469  
modaalirakenne 375  
partisiippirakenne 375, 456  
referentiaalirakenne 291  
sanarakenne 15, 32, 33, 41, 48, 53, 68, 77,  
78, 86, 87, 323, 326, 335, 337, 355, 359, 363,  
368, 371, 373, 375, 379, 381–384, 411, 466,  
468  
tavurakenne 325, 348  
temporaalirakenne 49, 278, 291, 375, 424,  
474  
virkerakenne 317, 319, 323, 344, 346, 463  
rauniosana 70, 311, 331, 334, 346  
reaaliaikametodi 24  
reaaliaikatutkimus 253, 254, 261, 262, 295,  
298  
referointi 45, 200, 201–210, 212, 224–226,  
229–234, 236, 241, 278, 457, 462, 475  
reitti 29, 120, 124, 125, 143, 153  
kulkureitti 91, 111, 136, 137, 139, 140, 141,  
144, 145, 149  
rekisteri 10, 16, 62, 72, 127, 150, 210, 308,  
314, 321, 329, 352  
rektio 364, 404  
resurssi 23–25, 43, 44, 59, 60, 108–111, 120,  
148–150, 168, 203, 305, 357, 362, 363, 368,  
389, 470  
rikkaus 65, 374, 395  
rinnastus 230, 426, 457  
roolileikki 205
- S
- salienssi 145, 149, 160, 192  
samanlaisuusoletus 20  
sana  
abstraktisana 85, 92, 95, 102, 281, 283,  
284  
asiasana 95, 102  
johtamaton sana 67, 68, 77, 86, 87, 95, 97,  
100, 372, 373, 382–385, 413, 466, 468–470  
kieliopillinen sana 68, 73, 84, 85, 111, 379,  
386, 391, 395  
leksikaalinen sana 71, 96, 374, 380, 390,  
391, 395, 396, 412, 466  
objektisana 92, 95, 102  
ruumiinosasana 281, 283, 284, 291

- sisältösana 68, 84, 395, 396  
sukulaissana 281, 283, 284, 291  
vierassana 340, 383  
yhdyssana 13, 66–68, 77, 86, 87, 95, 97,  
101, 334–337, 356, 372, 373, 378, 382–384,  
412  
sanahahmo 15, 310, 317, 320, 321, 342  
sanajärjestys 160, 170, 171, 204, 249, 306,  
317, 330, 331, 335, 344, 345  
sanaliitto 70, 337, 373, 383  
sanaluokka 13, 28, 32, 33, 41, 68, 69, 71, 75,  
84–88, 95, 245, 267, 268, 270, 272, 274,  
275, 276, 278, 291, 336, 371, 374, 375, 390,  
412, 427–433, 446, 464, 465, 467, 471, 472  
kieliopillinen sanaluokka 85, 395  
leksikaalinen sanaluokka 96, 380, 390,  
391, 395, 396  
nominiaalinen sanaluokka 272, 425, 430,  
431, 473  
pääsanaluokka 85, 94, 391  
sanaluokkaindeksi, ks. indeksi  
verbaalinen sanaluokka 425, 445, 447, 448  
sanamerkitys 40, 55, 59, 63, 64, 102  
sananomukaisuuskriteeri 201  
sanaraja 80, 330, 331, 333–337  
sanasto 8, 9, 13–16, 19, 21, 22, 30, 42,  
43, 52, 54, 59, 61–66, 68–77, 79–87, 90,  
93–95, 97, 99, 100–102, 242, 247, 309,  
310, 312, 313, 315, 317, 319, 323, 327–329,  
342, 344, 347, 348, 355–358, 368, 371, 373,  
377, 378, 381, 385, 474  
abstraktisanasto 62, 91, 102  
aikuissanasto 101, 382  
koululaissanasto 79, 101, 382  
sanastonhallinta, ks. hallinta  
spatiaalinen sanasto 106, 151  
substantiivisanasto 94, 95, 102  
verbisano 102  
sanatieto, ks. tieto  
sanavalinta 62, 241, 306, 317, 320, 327, 330,  
332, 333, 335, 345, 346  
sanavarasto 10, 13, 104, 309  
sanavartalo 13, 76, 288, 317, 325, 334, 342,  
346  
sanelukirjoitus 303, 316, 340  
sanue 66, 67, 68  
satelliitti 112  
satelliittikehyskieli 112, 116, 149, 150  
sidoksisuus 230, 391, 404  
sija 66, 123, 124, 128–130, 148, 272, 287–  
290, 292, 344  
paikallissija 129, 150, 152, 289, 290, 292,  
332, 342–344, 383  
rektiosija 314, 330  
sijakieli 112, 149  
sijamuoto 114, 274, 275, 289, 314, 342, 383  
sijataivutus 320, 332, 383, 437  
sijavalinta 170, 342, 343, 344  
sitaatti 213, 224, 408  
skeema 17, 33, 34, 242, 399  
kertomusskeema 17, 34, 96, 100, 212, 238,  
242, 377, 392, 399, 475  
skripti 96, 169, 170, 174, 399, 400  
päiväskripti 90, 100  
sosiolingvistiikka 252, 253, 258  
spatiaalinen siirtymä 112–115, 128, 130–132,  
135, 144, 146  
spesifisyys 18, 40, 45, 373, 385  
spesifisyysaste (*granularity*) 381, 383  
stimulus 24, 43, 44, 111, 116, 126, 148, 150  
strategia 60, 72, 100, 116, 168, 169, 263,  
308, 332, 359, 364, 380  
hahmottamisstrategia 169, 364  
kerronstrategia 72, 231  
kielentämisstrategia 24, 25, 43, 142, 359  
kirjoittamisstrategia 359, 409, 428, 447  
subjekti 163, 166, 170, 180, 181, 249, 259,  
260, 264, 274, 285, 322, 342–344, 346,  
372, 375, 392, 427, 458  
substantiivi 33, 66, 69, 75, 77, 78, 84–94,  
102, 106, 111, 129, 150, 164, 171, 172, 176,  
189, 190, 272, 274–278, 281, 288, 289,  
291, 374, 379, 383, 390–396, 429, 431, 434,  
436–438, 440, 441, 445, 464–468, 471,  
472, 474  
substantiivilauseke, ks. lauseke  
substantiiviperhe, ks. perhe  
substantiivisanasto, ks. sanasto

- substantiivitiheys, ks. tiheys  
substantiivi–verbi-indeksi, ks. indeksi  
suffiksaatio 313  
sujuvuus 307, 311, 312, 317, 323, 328, 329, 357, 363  
sukupuoli 11, 12, 19–23, 39, 43, 49, 77, 81, 82, 101, 103, 176, 177, 186, 187, 189, 196, 210, 212, 329, 332, 333, 338, 346, 370, 376, 378, 379, 381, 383, 386, 388, 389, 397, 406, 407, 413, 416, 428, 429, 432, 434, 436, 446, 447, 450, 453, 458, 462, 472  
sukupuoliero 11, 19, 20, 22, 26, 43, 46, 49, 105, 186, 383, 389, 413, 433, 459, 474  
suomi toisena kielenä 29, 31, 35, 38, 154, 244, 252, 304, 306, 309, 311, 312, 316, 317, 321, 323–325, 328, 332, 333, 336–349, 351–353, 355, 356, 373  
suoritus (*accomplishment*) 113, 146  
suunnittelu 16, 31, 32, 100, 102, 162, 369, 388  
suunta (*direction*) 43, 44, 109, 110–114, 119, 123, 125, 127–151, 153  
vertikaalinen suunta 127, 133, 146, 147
- T
- taitodiskurssi 302, 303, 305, 310  
taitotaso 64, 215, 312, 346, 348, 351, 352, 361, 373  
taivutusmuoto 66–68, 78, 101, 272, 273, 278, 321, 342–344  
tajunnanvirta 234, 236, 241  
tapahtuma 16, 24, 28, 29, 33, 34, 43, 44, 49, 59, 69, 85, 86, 88, 89, 93, 96, 99, 100, 105, 109, 134, 155–157, 159, 160, 164, 167, 168, 169, 174, 175, 182, 184, 185, 187, 188, 203, 204, 209, 211, 218, 223, 225, 234, 241, 279, 280, 368, 382, 384, 391, 395, 399, 402, 405, 407, 409, 421, 456, 470, 471, 474  
tapahtumajärjestys 25, 33, 34, 156, 191, 405, 408, 409, 421  
tarina 9, 10, 16, 17, 25, 28, 32, 33, 157, 165, 175, 178, 201, 207, 209, 218, 219, 226, 231, 232, 367, 412, 461  
sammakkotarina 207, 229, 230, 239  
tarkenne 171, 179, 172, 173, 178, 180, 189, 190, 191, 193, 285  
pronominitarkenne 172, 173, 190, 191  
se-tarkenne 158, 163, 172, 188, 189, 190, 191, 194  
tarkkaavuus 348, 350, 354, 355, 356  
tarkkuus 47, 59, 109, 302–310, 312–319, 321–329, 335, 337, 344, 346–349, 351, 352, 354–356, 358–360, 381  
epätarkkuus 305, 306, 311, 323, 325, 330, 336, 337, 339, 341, 344, 352, 355  
tarkoite 43, 44, 86, 93, 94, 102, 156, 158, 161, 164, 169, 170–173, 177, 181, 188–191, 193, 194, 221, 285, 368, 391, 395, 431, 440  
tuttu tarkoite 170, 173, 176, 178, 179, 190, 191, 194  
uusi tarkoite 44, 188, 191  
tavoitemuoto 34, 203, 309, 322, 339, 346, 347  
tavu 311, 325, 336, 340  
teko 90, 92, 96, 97, 282, 283, 456, 470  
teksti  
argumentoiva teksti 24, 34, 369, 408, 411  
arvioiva teksti 35, 100, 293, 475  
erittelevä teksti 33, 157, 160, 203, 293, 388, 391, 397, 408, 409, 412, 427, 437, 468, 475  
kertova teksti 16, 17, 33–35, 41, 44, 45, 49, 72, 84, 99, 100, 155–160, 167–170, 173, 180, 200, 202, 209, 214, 225, 248, 279, 321, 368–370, 374, 380, 386, 390, 391, 394, 396–399, 407, 409, 411, 412, 427, 431, 438, 442, 459, 465, 469, 470–475  
kuvaileva teksti 369, 380, 391, 409  
tekstilaji 17, 47, 72, 167, 168, 181, 193, 200, 209, 224, 254, 268, 271, 276, 277, 279, 291, 292, 293, 305, 310, 321, 360, 370, 394, 395, 426, 475  
tekstimerkitys 59, 60  
tekstittömyys 159  
tekstityyppi 16, 17, 25, 33–35, 74, 97, 99, 167–169, 174, 183, 186, 187, 193, 194, 202, 203, 206, 237, 271, 293, 370, 374, 377, 379, 388, 390–392, 411, 469, 470, 475

- tempus 200, 204, 232, 240, 241, 343, 396, 448, 454
- testi 11–13, 19, 20, 27–30, 35–39, 49, 66, 68, 76, 77, 85, 89, 91, 101, 106, 117, 164, 165, 316, 317, 329, 349, 377
- testihäly 165
- tieto
- kulttuurinen tieto 157, 163, 310
  - maailmantieto 114, 130, 159, 321
  - sanatieto 9, 11–15, 26, 29, 30, 43, 59, 60, 61, 63, 66
  - semanttinen tieto 20–22, 26, 42, 101, 474
  - tiedostamaton tieto 8
  - tietoinen tieto 8
  - tuttu tieto 44, 160, 163, 285
  - uusi tieto 160, 162, 163, 166
  - yhteinen tieto 9, 191
- tietoisuus
- morfologinen tietoisuus 53, 104, 383, 417
  - foneettinen tietoisuus 317
  - metakielinen tietoisuus 220
- tiheys 16, 177, 186, 189, 319, 330, 331, 337, 373, 381, 395, 396, 398, 412, 432, 433, 445, 451, 453, 460, 464, 466, 468, 469
- adjektiivitiheys 431–435
  - esiintymistiheys 178, 194, 309, 328, 342, 432
  - leksikaalinen tiheys 84, 390, 391, 395–399, 401, 402, 410, 412, 413, 437, 440, 441, 468
  - pronominitiheys 441–444
  - sivulauseitiheys 459
  - substansiivitiheys 24, 374, 436–438
  - tekstitiheys 318, 325, 338, 340, 342, 347, 349
  - verbitiheys 446, 449
  - virhetiheys 315, 316, 322–327, 329, 330, 342, 347–350, 355, 356
- tiiviyys 19, 84, 374, 375, 387, 403, 413, 425, 427, 428
- tila 43, 85, 88, 92, 93, 96, 133, 456, 470
- toiminnan esittäminen (*re-enactment*) 201
- toisteisuus 76, 78, 86, 87, 95, 169, 395, 412, 445
- topiikki 157, 159, 166
- topologinen relaatio 121, 274
- tuottaminen 17, 45, 49, 62, 63, 157–159, 164–167, 178, 193, 206, 211, 217, 239, 240, 285, 304, 305, 310–313, 316, 325, 357, 370, 413, 424
- tuttuus 13, 29, 66, 86, 159, 160, 161, 173, 189, 191, 192, 338, 426
- oletettu tuttuus 158, 161–164, 192
  - pääteltävä tuttuus 161
- T-yksikkö (*terminal unit*) 156, 226, 315, 323, 369, 375, 388, 389, 402–409, 412, 413, 427–434, 437–439, 443, 446–451, 453, 454, 457–459, 461, 464–466, 469, 471–475
- täydennysjakauma 290, 292
- U
- U-käyrä 239, 254
- V
- variaatio 23–26, 42–44, 46, 53, 60, 65, 79, 108–111, 115, 116, 120, 136, 148, 149, 165, 166, 207, 227, 248, 250, 253, 264, 267, 275, 285, 287–290, 318, 370, 374, 437, 446, 447
- vastaanottaja 155–157, 159–163, 166, 167, 169, 171, 172, 191, 192, 200, 203, 218, 233, 314
- verbi 66, 67, 69, 70, 74–76, 78–80, 84–87, 95–97, 99, 112, 116, 124, 126–128, 133, 135–138, 143, 146, 148, 150, 168, 174, 176, 189, 202, 207, 209, 227, 228, 233, 250, 262, 268, 274–278, 289–291, 323, 332, 342, 343, 352, 357, 364, 374, 382, 390–396, 426, 428, 429, 431, 437, 438, 444–450, 452–456, 458, 461, 464, 465, 467, 471, 472, 474
- deiktinen verbi 127, 136
  - emootioverbi 96, 456
  - finiittiverbi 291, 390, 426, 431, 434, 438, 447–450, 452, 456, 457, 465–468, 472, 473
  - intransiiviverbi 392, 426
  - kieltoverbi 337
  - kognitioverbi 96, 97, 456, 461
  - kommunikaatioverbi 96
  - liikeverbi 12, 13, 34, 76, 77, 96, 112, 113, 127

- mentaaliverbi 96, 456, 457, 474  
modaaliverbi 374, 395, 396, 447, 448, 464, 468, 472  
puheaktiverbi 202, 207, 208, 227, 229, 230, 231, 456, 457, 462  
suuntaverbi 127, 136, 150  
tapahtumaverbi 43, 96, 99  
tapaverbi 115, 116, 126, 135, 138, 150  
tekoverti 96, 97  
tilaverbi 96, 97  
väyläverbi 115, 124, 126–128, 142, 150  
verbikehyskieli 112, 115, 116, 150  
verbiketju 343, 345, 346  
verbiluokka 34, 97, 126, 150, 278, 291, 431, 456, 457, 472  
verbimuoto 78, 267, 270, 275–279, 289, 290, 332, 374, 448, 450, 452–456  
viitekehys (*frame of reference*), spatiaalinen 109, 110, 113, 124, 132, 133, 302  
viittaus, kielellinen 40, 45, 159, 165, 166, 180, 188, 211–214, 216, 236, 238, 293  
viittausketju 16, 157, 189  
viittausuhde 40, 156, 158, 160, 165, 167, 171, 191, 192, 206, 207, 238, 241, 285, 286, 298, 332, 333, 399, 427, 429  
virhe  
alkukirjainvirhe 324, 327, 330, 331, 333–335, geminaattavirhe 338, 339, 341  
kongruenssivirhe 330, 332, 343  
piirtovirhe 330, 331, 334, 335, 341, 342  
sanajärjestysvirhe 324  
sanarajavirhe 336  
yhdys-sanavirhe 330, 336, 378  
äännekestopvirhe 318, 338  
virheettömyys 302, 315, 351  
virre 15, 156, 159, 175, 176, 210, 215, 219, 311, 322, 323, 328, 345, 346, 402, 403, 407, 408, 423, 426, 427, 429, 459, 471  
virkeraja 225, 318, 333  
virkerakenne, ks. rakenne  
vokaali 268, 287, 288, 311, 318, 330–332, 334, 335, 338–341, 355  
vokaaliharmonia 341  
vokaalisointu 332, 341  
vuorovaikutus 45, 46, 62, 199, 213, 217, 218, 231, 233, 240, 242, 246, 248, 258, 307, 309, 357  
välimerkki 202, 207, 225, 228, 229, 241, 318, 319, 324  
väylä (*path*) 43, 44, 108, 109, 110–116, 119–122, 124–129, 131–138, 140–142, 144–148, 150, 151  
väyläinformaatio 124, 150
- Y  
ydin 11, 78, 248, 256, 257, 264, 415, 429  
yhtenäisyys 44, 169, 191, 302  
yksimielisyysperiaate (*consensus principle*) 9  
yksinkertaistuminen 287, 312  
yksinkertaisuus 100, 327, 372, 425  
yleisyytasto 15, 65, 66, 72, 75–79, 94, 96, 99, 101, 102, 347  
ymmärtäminen 14, 40, 49, 53, 76, 104, 157, 158, 159, 163, 164, 165, 203, 206, 239, 310, 422, 424
- Ä  
älykkyys, kielellinen 12, 20, 21, 23, 27, 354, 385, 414  
ääni 209, 229, 233, 234, 235, 236, 244  
henkilön ääni 209, 233  
kertojan ääni 232, 233  
minän ääni 234  
vieras ääni 234, 236  
äänisyys  
kaksiäänisyys 234  
moniäänisyys 15, 19, 35, 40, 45, 59, 60, 199–201, 203–209, 211–214, 216–218, 222–224, 226, 232–239, 242, 277, 368  
ääne–kirjain-vastaavuus 47, 247, 310, 311

