

Sanaston erot yksikielisten suomea ja kaksikielisten suomea ja ruotsia
puhuvien lasten välillä tarinankerrontatehtävässä

Heidi Saarijärvi-Lautala
Pro gradu -tutkielma
Ohjaaja: Raymond Bertram
Turun yliopisto
Yhteiskuntatieteellinen tiedekunta
Psykologian ja logopedian laitos
Logopedia
Kesäkuu 2022

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

TURUN YLIOPISTO

Yhteiskuntatieteellinen tiedekunta, psykologian ja logopedian laitos

SAARIJÄRVI-LAUTALA, HEIDI: Sanaston erot yksikielisten suomea ja kaksikielisten suomea ja ruotsia puhuvien lasten välillä tarinankerrontatehtävässä.

Pro gradu -tutkielma, 45 s., 8 liites.

Logopedia

Kesäkuu 2022

Kaksikielisyys on yleistä, ja arvioiden mukaan jopa kolmannes maailman väestöstä puhuu päivittäin kahta tai useampaa kieltä. Suomessa on kaksi kansalliskieltä, ja suomalaisista noin 5 prosenttia puhuu ruotsia äidinkielenään. Kokonaisuudessaan kaksikielisyys on kasvava ilmiö – äidinkielenään muuta kuin suomea, ruotsia tai saamea puhuvien lasten määrä on viimeisten kymmenen vuoden aikana kasvanut huomattavasti, ja kaksikielisten lasten määrä myös puheterapiavastaanotoilla on kasvusuunnassa. Sanaston eroista yksi- ja kaksikielisten välillä on kuitenkin melko vähän tutkimusta, varsinkin suomen ja ruotsin osalta. Yksi- ja kaksikielisten käyttämän sanaston eroista saatavaa tietoa voitaisiinkin hyödyntää esimerkiksi puheterapeuttista arviointia tai kuntoutusta suunniteltaessa. Lisäksi tietoa voitaisiin käyttää puheen sujuvuuden tutkimuksessa, sillä puhuja saattaa käyttää itselleen helpompia sanoja sujumattomuuksia aiheuttavien sanojen sijasta, ja siksi sujumattomuuksien määrän lisäksi olisi hyvä huomioida käytetyn sanaston yleisyys ja monipuolisuus.

Tämän pro gradu -tutkimuksen tarkoituksena oli selvittää, onko yksikielisten suomea puhuvien ja samanaikaisesti kaksikielisten suomea ja ruotsia puhuvien suomalaisten lasten sanaston monipuolisuudessa sekä käytettyjen sanojen yleisyydessä eroja, ja jos on, millaisia. Tämä tutkimus on osa Speech disfluencies in bilingual speakers (Sibils) -tutkimusprojektia, jossa tarkastellaan yksi- ja kaksikielisten lasten puheen sujumattomuuksia. Tutkimus toteutettiin vertailemalla 24 yksikielistä ja 20 kaksikielistä 6–11-vuotiasta lasta. Tutkittavia pyydettiin kertomaan tarina kuvakirjan kuvien pohjalta. Tarinoissa käytetyn sanaston monipuolisuutta tarkasteltiin Guiraudin indeksin sekä edistyneen Guiraudin indeksin avulla, ja sanojen yleisyyttä tutkittiin tarkastelemalla tarinoissa esiintyneiden sanojen esiintymistiheyksiä korpuksessa.

Tutkimuksen tulosten perusteella yksikieliset lapset käyttivät merkitsevästi enemmän harvinaisia sanoja kuin kaksikieliset, ja kaksikieliset käyttivät merkitsevästi enemmän yleisiä sanoja kuin yksikieliset. Yksi- ja kaksikielisten lasten käyttämien sanastojen monipuolisuudessa ei ollut tilastollisesti merkitsevää eroa. Suomen kielen hyvä taitotaso kaksikielisten suomea ja ruotsia puhuvien suomalaisten keskuudessa saattaa vaikuttaa siihen, että sanaston monipuolisuudessa ei ollut eroja, mutta erot suomenkielisen syötteen määrässä saattavat silti vaikuttaa sanaston yleisyyden osalta löytyneihin eroihin.

asiasanat: yksikielisyys, kaksikielisyys, sanavarasto

Sisällysluettelo

1 Johdanto	5
1.1 Kaksikielisyys	7
1.1.1 Kaksikielisyyden määritelmä	7
1.1.2 Kaksikielisyyden piirteitä	8
1.1.3 Ruotsia ja suomea äidinkielenään puhuvat suomalaiset	10
1.2 Sanavaraston koko ja siihen vaikuttavat tekijät.....	12
1.2.1 Sanavaraston kehitys yksi- ja kaksikielisillä.....	12
1.3 Kerrontataidot ja kertomukset	14
1.3.1 Kerrontataitojen kehitys	15
1.3.2 Kerrontatehtävät arviointimenetelmänä.....	15
1.4 Sanaston arviointi.....	17
1.4.1 Sanaston monipuolisuus	17
1.4.2 Sanaston yleisyys.....	19
1.5 Tutkimuskysymykset	20
2 Menetelmät	21
2.1 Koehenkilöt	21
2.2 Kerrontatehtävä	23
2.3 Aineistonkeruu	24
2.4 Kysely vanhemmille.....	25
2.5 Sanaston arviointi.....	26
2.5.1 Sanaston monipuolisuuden arviointi	27
2.5.2 Sanaston yleisyyden arviointi.....	27
2.6 Tutkimuksen eettisyys.....	27
3 Tulokset	28
3.1 Sanojen monipuolisuus	28
3.2 Sanojen yleisyys.....	29
4 Pohdinta	31
4.1 Sanaston eroavaisuudet.....	31
4.1.1 Sanaston monipuolisuus	31
4.1.2 Sanaston yleisyys.....	32
4.2 Luotettavuus.....	34

4.3 Lopuksi.....	36
Lähteet	38
Liitteet.....	46
Liite 1. Vanhemman suostumuslomake	46
Liite 2. Lapsen suostumuslomake	46
Liite 3. Kysely vanhemmille	46

1 Johdanto

Tämän tutkimuksen tarkoituksena on selvittää yksikielisten suomea puhuvien ja kaksikielisten suomea ja ruotsia puhuvien lasten sanaston eroja suomenkielisessä tarinankerrontatehtävässä. Tarinoista tarkastellaan käytetyn sanaston monipuolisuutta ja sanojen yleisyyttä. Tutkimus on osa *Speech disfluencies in bilingual speakers (Sibils)* -tutkimusprojektia, jossa tarkastellaan puheen sujumattomuuksia yksi- ja kaksikielisillä lapsilla. Kaksikielisten lasten aineisto on kerätty jo aiemmin projektissa (Brandt-Hägglom, 2021; Valo, 2021), joten tässä tutkimuksessa aineistoa kerätään yksikielisiltä suomenkielisiltä lapsilta.

Suomen kansalliskielet ovat suomi ja ruotsi, ja suomalaisista noin 5 prosenttia puhuu äidinkielenään ruotsia (Perustuslaki 1999/731, 7 §; Suomen virallinen tilasto, 2021b). Romainen (2017) mukaan ruotsia äidinkielenään puhuvilla suomalaisilla saattaa olla kielellisistä vähemmistöistä jopa paras laillinen asema maailmassa. Ruotsia äidinkielenään puhuvia suomalaisia asuu Suomessa eniten Ahvenanmaalla, Pohjanmaalla sekä Uudellamaalla (Kuntaliitto, 2017). Lisäksi myös muiden kuin suomea, ruotsia tai saamea puhuvien määrä Suomessa on globalisoitumisen myötä kasvanut. 1990-luvun alussa vieraskielisiä asui Suomessa vielä alle 50 000, mutta vuonna 2020 vieraskielisiä oli jo yli 432 000, eli lähes 8 prosenttia väestöstä (Suomen virallinen tilasto, 2021a). Ihminen voi ilmoittaa väestötietojärjestelmään äidinkielekseen vain yhden kielen, joten tilastoinnit eivät kuitenkaan anna täydellistä kuvaa kaksi- tai monikielisten ihmisten määrästä Suomessa (Tammenmaa, 2020). Folktingetin eli Suomenruotsalaisten kansankäräjien (Tammenmaa, 2020) mukaan on arvioitu, että kolme neljäsosaa suomea ja ruotsia puhuvista kaksikielisistä lapsista rekisteröidään ruotsinkielisiksi.

Kaksikielisten määrä väestöstä on siis kasvanut viime vuosina, joten myös kaksikielisten lasten määrä puheterapiavastaanotoilla on todennäköisesti kasvusuunnassa. Suurin osa suomenkielisestä kielellisten taitojen tutkimuksesta ja menetelmistä keskittyy kuitenkin yksikielisiin (Lauonen, 2007). Lisää suomenkielistä tutkimustietoa kaksikielisyyden vaikutuksista lasten kielellisiin taitoihin tarvitaankin puheterapeuttisen arvioinnin ja kuntoutuksen tueksi, sillä kaksikielisten arvioiminen on usein haastavampaa kuin yksikielisten (Hoff, 2009). Kaksikielisten lasten kohdalla voi esimerkiksi herätä kysymys, liittyvätkö kielelliset hankaluudet esimerkiksi kehitykselliseen kielihäiriöön vai eikö lapsi vain ole saanut tarpeeksi altistusta kielelle (Boerma & Blom, 2017). Kielellisten vaikeuksien yli- ja alidiagnosointi ovatkin yleisempiä ongelmia monikielisten lasten kohdalla verrattuna yksikielisiin lapsiin (Laing & Kamhi, 2003).

Kaksikielisen lapsen kielellisiä taitoja olisi hyvä arvioida hänen molemmilla kielillään (Pearson ym., 1993). Puheterapeutin olisi hyvä joko osata itse kyseisiä kieliä tai käyttää apuna tulkkia tai esimerkiksi lapsen vanhempaa, mutta tulkin tai vanhemman välityksellä lapsen kielellisten taitojen arvioiminen voi silti olla hankalaa (Launonen, 2007). Suuri osa puheterapeuteista tuntee epävarmuutta monikielisten lasten kielellisten taitojen arvioinnissa (Kaiser & Rautakoski, 2013). On myös hankala määritellä, miten hyvin puheterapeutin pitäisi osata kieltä, että hän voisi arvioida ja kuntouttaa asiakasta kyseisellä kielellä (Launonen, 2007). Sanaston eroista yksi- ja kaksikielisten välillä on tehty useita tutkimuksia, mutta monet tutkimuksista käsittelevät eroja englannin- ja espanjankielisten lasten sanastojen välillä (Kunnari ym., 2016). Äidinkielenään suomea sekä suomea ja ruotsia puhuvien lasten sanastoja taas on tutkittu vähemmän. Saatavia tietoja sanaston eroista yksi- ja kaksikielisillä voitaisiin siis mahdollisesti hyödyntää kliinisessä työssä.

Kaksi- ja monikielisyys voidaan määritellä hyvin eri tavoilla. Selvyyden vuoksi tässä tutkimuksessa käytetään pääosin käsitettä kaksikielisyys, jonka voidaan ajatella kattavan myös kolmea tai useampaa kieltä käyttävät ihmiset (Wei, 2000). Tässä tutkimuksessa vertaillaan yksi- ja kaksikielisiä 6–11-vuotiaita suomalaislapsia, joista yksikieliset lapset puhuvat äidinkielenään vain suomea, ja kaksikielisillä lapsilla tarkoitetaan samanaikaisesti kaksikielisiä suomea ja ruotsia äidinkielenään puhuvia lapsia. Tutkimuksessa lapsia pyydettiin kertomaan kuvakirjan kuvien perusteella tarina, joka litteroitiin ja analysoitiin. Tarinoista tarkasteltiin käytetyn sanaston monipuolisuutta sekä sanojen yleisyyttä. Tässä tutkimuksessa ei vertailla keskenään eri kieliä, vaan tarkastellaan nimenomaan yksi- ja kaksikielisyyden vaikutusta lasten sanastoon suomenkielisissä tarinoissa. Suomessa ensimmäisen vieraan kielen tai toisen kotimaisen kielen opetus alkaa yleensä peruskoulun kolmannella vuosiluokalla (Opetushallitus, 2014), ja lisäksi lapset saattavat oppia vierasta kieltä esimerkiksi televisio-ohjelmista tai peleistä. Tämän mahdollisen varhaisen kielitaidon ei ajatella vaikuttavan oleellisesti tässä tutkimuksessa, mutta lapsen osaamia kieliä kysytään silti lapsen huoltajalta.

Tämä tutkimus on osa *Speech disfluencies in bilingual speakers (Sibils)* -tutkimusprojektia, jossa tarkastellaan pääasiassa yksi- ja kaksikielisten lasten puheen sujumattomuuksia. Samasta aineistosta tehdyn toisen pro gradu -tutkielman (Riihioja, 2021) mukaan puheen sujumattomuuksien määrässä tai laadussa ei ollut tilastollisesti merkitsevää eroa yksi- ja kaksikielisten lasten välillä. Tulos eroaa aiemmista tutkimuksista, joiden mukaan kaksikielisillä

havaitaan usein yksikielisiä enemmän puheen sujumattomuuksia, etenkin heidän heikommalla kielellään (Eggers ym., 2019). Riihiojan (2021) tutkimuksen kaksikielisillä on varsin hyvä suomen kielen taito, mikä saattaa vaikuttaa siihen, että yksi- ja kaksikielisten lasten puheen sujumattomuuksien väliltä ei löytynyt eroa. Kuitenkin on mahdollista, että kaksikielinen välttää itselleen vaikeita ja sujumattomuuksia aiheuttavia sanoja tai tavuja, ja käyttää niiden sijasta itselleen helpompia sanoja (Lowe ym., 2017). Sujumattomuuksien määrä ei siis välttämättä kerro koko totuutta yksi- ja kaksikielisten lasten puheen sujuvuuden eroista tai samankaltaisuuksista, joten siksi tässä tutkimuksessa tarkastellut erot käytetyn sanaston yleisyydessä ja monipuolisuudessa voivat antaa lisää tietoa myös puheen sujumattomuuksiin liittyvien tutkimusten pohjalle.

1.1 Kaksikielisyys

Kaksi- tai monikielisyys on yleistä – on arvioitu, että maailman väestöstä jopa kolmannes käyttää kahta tai useampaa kieltä päivittäin (Wei, 2000). Jos kaksikielisyyttä ajatellaan sen laajimmassa merkityksessä, ja mukaan lasketaan kaikki esimerkiksi koulussa vierasta kieltä oppineet, niin puhtaasti yksikielisiä ihmisiä on Wein (2000) mukaan oikeastaan hyvin vähän. Suuri osa kielitieteellisestä tutkimuksesta kuitenkin koskee yksikielisyyttä (Romaine, 2017).

1.1.1 Kaksikielisuuden määritelmä

Kaksikielisyys voidaankin määritellä useilla eri tavoilla. Sen voidaan ajatella tarkoittavan kahden tai useamman kielen samantasoista osaamista, mutta kaksikielisyydellä voidaan myös tarkoittaa sitä, että aikuinen on oppinut puhumaan muuta kieltä kuin äidinkieltään niin, että pärjää sen avulla (Bialystok, 2001; Wei, 2000). On vaikea määritellä, missä kohtaa kahta kieltä osaava ihminen muuttuu yksikielisestä kaksikieliseksi, minkä takia kaksikielisuuden käsite laajenikin 1900-luvun aikana huomattavasti (Mackey, 2000). Kielitaito ei välttämättä myöskään ole samantasoista kielen eri osa-alueilla, mikä vaikeuttaa kaksikielisuuden määrittelyä (Wei, 2000; Mackey, 2000). Mackeyn (2000) mukaan kaksikielisyttä määriteltäessä tulisikin ottaa huomioon se, miten paljon ja millaisissa tilanteissa ihminen käyttää osaamiaan kieliä, miten hän vaihtelee kielten välillä sekä miten kielet vaikuttavat toisiinsa.

Yksi yleinen tapa määritellä kaksikielisyys on funktionaalinen kaksikielisyys, jonka mukaan ihminen on kaksikielinen, kun hän pystyy toimimaan arkipäiväisissä tilanteissa kahdella tai useammalla eri kielellä (Wei, 2000). Kaksikielisyys voidaan määritellä myös esimerkiksi

omaksumisajankohdan tai kielten osaamisen kautta. Samanaikaisella eli simultaanilla kaksikielisyydellä tarkoitetaan tilannetta, jossa lapsi oppii kahta kieltä syntymästään tai lähes syntymästään saakka, kun taas peräkkäisesti kaksikieliset oppivat toisen kielen ensimmäisen kielen oppimisen jälkeen (Hoff, 2009). Tasapainoisessa kaksikielisyydessä ihminen osaa molempia kieliä lähes yhtä hyvin, mutta ihminen voi myös osata toista kieltä paremmin, jolloin kieli on vahvempi eli dominantimpi (Pearson, 2008). Kielten dominanssi voi myös muuttua, jos ihminen esimerkiksi muuttaa toisen kieliselle alueelle ja alkaa käyttää kyseistä kieltä enemmän. Tässä tutkimuksessa kaksikielisillä lapsilla tarkoitetaan samanaikaisesti kaksikielisiä suomea ja ruotsia äidinkielenään puhuvia lapsia. Suurin osa tutkimuksen kaksikielisistä lapsista on tasapainoisesti kaksikielisiä. Mukana on kuitenkin myös lapsia, joilla suomi tai ruotsi on dominantimpi, mutta hekin osaavat molempia kieliä hyvin.

Sekä yksilö että yhteiskunta voivat olla monikielisiä (Romaine, 2017), ja kielellinen ympäristö vaikuttaakin kielitaitoon oleellisesti. Esimerkiksi lapsen molemmat vanhemmat voivat puhua eri kieltä, mutta sen lisäksi heillä voi olla eri yhteinen kieli, tai perheen kieli voi olla eri kuin ympäristön kieli (Hoff, 2009). Kaikista kaksikielisessä ympäristössä kasvaneista lapsista ei kasva tasapainoisesti kaksikielisiä aikuisia, mihin vaikuttavat Pearsonin (2007) mukaan esimerkiksi kielellisen syötteen määrä sekä ympäristön asenteet kieltä kohtaan. Ympäristön kieli on Dallerin ym. (2003) mukaan usein yksilön vahvempi kieli, sillä koulutus on yleensä ympäristön kielellä ja lisäksi sitä kuulee useammassa yhteyksissä. Ihminen voikin käyttää kieliä eri tilanteissa, ja esimerkiksi lapsi, joka käyttää kotona ja koulussa eri kieliä, osaa kielistä todennäköisesti hyvin erilaisia sanoja, eli myös sanasto voi olla kontekstikohtaista (Bialystok, 2001; Hoff, 2009; Wei, 2000). Kahden kielen ylläpitäminen ei välttämättä aina ole yksinkertaista, esimerkiksi tilanteessa, jossa ainoastaan lapsen toinen vanhempi puhuu tiettyä kieltä (Wei, 2000). Kieli liittyy usein myös identiteettiin, ja Wein (2000) mukaan osa maahanmuuttajista haluaa unohtaa kotimaansa ja äidinkieltensä, kun taas osa haluaa ylläpitää kieltään ja kulttuuriaan.

1.1.2 Kaksikielisyyden piirteitä

Wein (2000) mukaan etenkin 1900-luvun ensimmäisellä puoliskolla vallalla oli ajatus, jonka mukaan kaksikielisyys vaikutti negatiivisesti kielelliseen ja kognitiiviseen kehitykseen. Ajateltiin muun muassa, että kaksikielisillä on pienempi sanavarasto ja heidän on vaikeampi käyttää kieliä kieliopillisesti oikein kuin yksikielisillä (Wei, 2000). Wein (2000) mukaan nykyään tällaiset ajatukset on kumottu, ja on esimerkiksi huomattu, että kaksikieliset saattavat olla jopa

parempia kommunikoimaan kuin yksikieliset. Wei (2000) viittaa tekstissään Bakerin ja Prys Jonesin (1998) määrittelemiin kaksikielisyyden hyötyihin, joissa painottuvat mahdollisuus kommunikoida molemmilla kielillä, mahdollisesti parempi sensitiivisyys kommunikoijana, laajemmat työmahdollisuudet sekä mahdollisuus tutustua syvemmin useampiin kulttuureihin. Kaksikielisten kulloinkin käyttämä kieli riippuu monista tekijöistä, esimerkiksi keskustelukumppanista, puheenaiheesta ja paikasta. Samoja kieliä osaavat kaksikieliset ihmiset saattavat keskustellessaan vaihdella käyttämiään kieliä, mitä kutsutaan koodinvaihdoksi (code-switching) (Wei, 2000). Koodinvaihto ei yleensä aiheuta kieliopillisesti vääriä virkkeitä, vaan puhuja täydentää virkkeitä toisen kielen sanoilla kieliopillisesti oikein (Hoff, 2009; Wei, 2000). Kielten siirtovaikutuksella (language transfer) taas tarkoitetaan tilannetta, jossa ihmisen osaamien kielten piirteet vaikuttavat toisiinsa (Gass & Selinker, 1992).

Hoffin (2009) mukaan samanaikaisesti kaksikielisillä lapsilla kielellinen kehitys saattaa olla hie- man hitaampaa kuin yksikielisillä lapsilla, sillä he kuulevat yhtä kieltä vähemmän kuin yksikieliset, joiden koko syöte on tietyllä kielellä. Ero kuitenkin saattaa kadota 10 ikävuoteen mennessä. Samanaikaisen kaksikielisyyden lisäksi toisen kielen voi oppia myös myöhemmin elämässä, mutta natiivitasoista kielitaitoa on iän kasvaessa vaikeampi saada (Berken ym., 2016; Hoff, 2009). Saavutettavissa oleva taso kuitenkin vaihtelee eri kielen osa-alueilla (Berken ym., 2016). Berkenin ym. (2016) katsauksen mukaan tutkimusten välillä on kuitenkin erimielisyyttä siitä, onko niin sanottua kielen oppimisen kriittistä kautta olemassa, vai onko kielen oppimisen optimaalinen aika ennemminkin lineaarisesti iän myötä heikkenevä. Aivot muokkautuvat toisen kielen oppimisen myötä vielä aikuisenakin, mutta aivojen plastisiteetti on määrältään ja laadultaan erilaista kuin lapsena (Berken ym., 2016). Yksi- ja kaksikielisten aivoista onkin löydetty eroja, mutta erojen syyt ja niiden mahdolliset vaikutukset vaativat vielä lisää tutkimusta (Lehtonen, 2021). Berkenin ym. (2016) mukaan tutkijoilla vallitsee kuitenkin konsensus siitä, että erot yksi- ja kaksikielisten aivoissa näkyvät enemmänkin aivoalueiden aktivaation eroina kuin kielellisten alueiden sijainnin eroina. Toki kielen oppimiseen vaikuttavat iän lisäksi monet muutkin tekijät, kuten henkilön motivaatio, persoonallisuus ja ympäristöstä saatava kielellinen syöte (Cook, 2003). Cookin (2003) mukaan aikuiset voivatkin oppia toista kieltä jopa lapsia paremmin, sillä heillä on paremmat opiskelutaidot ja aiempaa tietoisuutta kielten piirteistä.

1.1.3 Ruotsia ja suomea äidinkielenään puhuvat suomalaiset

Tässä tutkimuksessa suomea äidinkielenään puhuvia suomalaislapsia verrataan suomea ja ruotsia äidinkielenään puhuviin kaksikielisiin suomalaislapsiin. Suomen kansalliskielet ovat suomi ja ruotsi, ja lain perusteella muun muassa viranomaispalveluita on oikeus saada sekä suomeksi että ruotsiksi (Perustuslaki 1999/731, 7 §). Ruotsinkielisten osuus väestöstä on laskenut 1950-luvun 8,6 prosentista 2020-luvun 5,2 prosenttiin (Suomen virallinen tilasto, 2021b). Suomen kaikki 16 ruotsinkielistä kuntaa sijaitsevat Ahvenanmaalla, kun taas kaksikielisiä kuntia on Suomessa 33, ja niissä 18:ssa suomi on enemmistökielenä (Kuntaliitto, 2017). Ruotsia äidinkielenään puhuvia asuu kuitenkin myös monissa yksikielissä suomenkielissä kaupungeissa (Suomen virallinen tilasto, 2021b).

Tällä hetkellä väestötietojärjestelmään voi merkitä vain yhden kielen, ja Folktinget on ilmaissut huolta siitä, että ruotsinkielisten palveluiden saatavuus voisi heikentyä entisestään, jos kaksikielillä olisi mahdollisuus merkitä molemmat kielet äidinkielikseen (Council of Europe, 2019). Euroopan neuvoston neuvoa-antava komitea (Council of Europe, 2019) onkin huolissaan ruotsin kielen aseman heikentymisestä Suomessa, sillä komitean mukaan asenteet ruotsinkielistä vähemmistöä ovat muuttuneet negatiivisemmiksi, eivätkä ruotsinkieliset suomalaiset aina saa palveluita perustuslain mukaisesti omalla kielellään. Voidaankin olettaa kaksikielisten suomalaisten kohtaavan suomea puhuvia usein ja joutuvan käyttämään suomen kieltä, mikä saattaa näkyä kaksikielisten suomalaisten hyvänä suomen kielen taitotasona.

Suomen kieleen on lainautunut runsaasti sanoja ruotsin kielestä, mutta sanoja on usein ääntämyksellisten erojen takia sopeutettu suomen kieleen sopivammiksi (Häkkinen, 1997). Kielessä on kuitenkin myös eroja, ja Seymourin ym. (2003) mukaan suomea pidetään kielenä ortografialtaan hyvin läpinäkyvänä, kun taas ruotsi on ortografialtaan monimutkaisempi. Suomen ja ruotsin kielten välisten erojen lisäksi Suomessa puhuttava ruotsi eroaa Ruotsissa puhuttavasta ruotsista muun muassa intonaatioltaan ja sanastoltaan, ja erot näkyvät erityisesti puhekielessä (Reuter, 2012).

Yksikielisiä suomea tai ruotsia ja kaksikielisiä suomea ja ruotsia äidinkielenään puhuvia lapsia on 2000-luvulla alettu vertailla toisiinsa enemmän. Tutkimuksissa on tarkasteltu esimerkiksi kielellisissä tehtävissä suoriutumista, kehityksellisen kielihäiriön riskiä (mm. Westman ym., 2008), lukitaitoja ennustavia tekijöitä (mm. Vataja ym., 2021) sekä kertomusten

makrostruktuuria (mm. Kunnari ym., 2016). Yksi- ja kaksikielisten lasten sanaston eroavaisuuksia etenkin kerrontatehtävissä on kuitenkin tarkasteltu vain vähän.

Westman ym. (2008) tutkivat yksikielisiä ruotsia puhuvia sekä kaksikielisiä suomea ja ruotsia puhuvia suomalaislapsia. Kaksikielisistä lapsista suurin osa oli ruotsia dominantisti puhuvia, ja tutkimuksessa ei ollut mukana suomea dominantisti puhuvia lapsia. Tutkimuksessa huomattiin, että kaksikielisyys ei ollut yhteydessä kehityksellisen kielihäiriön riskiin. Kaksikieliset lapset suoriutuivat kuitenkin yksikielisiä lapsia heikommin kahdella osuudella kahdeksasta, eli kehonosien nimeämisessä ja lausetoistossa ruotsiksi. Westmanin ym. (2008) mukaan tämä saattaa liittyä siihen, että kaksikielisten pitää omaksua enemmän sanoja kuin yksikielisten.

Kunnari ym. (2016) tarkastelivat tutkimuksessaan yksikielisiä suomea sekä tasapainoisesti kaksikielisiä suomea ja ruotsia äidinkielenään puhuvien keski-ikäisten 5;8-vuotiaiden lasten kertomusten makrostruktuuria. Struktuurin kompleksisuudessa ei nähty eroa yksi- ja kaksikielisten lasten välillä eikä myöskään kaksikielisten lasten eri kielissä. Kuitenkin kaksikieliset lapset suoriutuivat struktuurin osalta heikommin suomenkielisessä kerronta- kuin uudelleenkerrontatehtävässä. Uudelleenkerrontatehtävässä lapset käyttivät kielestä riippumatta monimutkaisempaa struktuuria kuin tarinankeksintätehtävässä. Kunnari ym. (2016) huomauttavat, että tarinankerrontatehtävä saattaa paljastaa kaksikielisten lasten kielten välillä eroja, jotka uudelleenkerronnassa jäisivät löytymättä, sillä kerrontatehtävä voi olla lapselle haastavampi.

Vatajan ym. (2021) ruotsiksi toteutetussa tutkimuksessa yksikielisillä ruotsia sekä samanaikaisesti kaksikielisillä suomea ja ruotsia puhuvilla 1-luokkalaisilla lapsilla huomattiin, että nopean sarjallisen nimeämisen tehtävät sekä kirjaintietoisuus ennustivat parhaiten lukitaitojen kehittymistä sekä yksi- että kaksikielisillä. Yksikieliset lapset suoriutuivat kaksikielisiä lapsia merkittävästi paremmin ruotsinkielisessä nopean sarjallisen nimeämisen tehtävässä. Yksi- ja kaksikielisten lasten väliltä ei kuitenkaan löydetty merkittävää eroa lukitaidoissa.

Näissä tutkimuksissa yksi- ja kaksikielisten suomalaislasten välillä löytyi vain pieniä eroja. Suomea ja ruotsia äidinkielenään puhuvat kaksikieliset osaavat yleisesti ottaen molempia kieliä hyvin ja tasapainoisesti. Aiempien tutkimusten perusteella voidaan olettaa, että kaksikielisten lasten kokonaissanavarastot ovat samankokoisia tai jopa suurempia kuin yksikielisten lasten sanavarastot, mutta kun kaksikielisten lasten eri kielten sanavarastoja tarkastellaan erikseen,

voidaan niiden koossa löytää eroja. Seuraavissa kappaleissa tarkastellaan sanavaraston kokoa ja kehitystä tarkemmin.

1.2 Sanavaraston koko ja siihen vaikuttavat tekijät

Lapsi sanoo yleensä ensimmäiset sanansa noin vuoden ikäisenä. Lasten ensimmäiset sanat ovat konkreettisia ja käytännönläheisiä, ja sanasto kehittyy kohti harvinaisempia ja abstraktimpia sanoja (Loukusa, 2007). Aikuisen ihmisen sanavarasto on kielestä riippuen kymmenien, ellei jopa satojen tuhansien sanojen kokoinen. Keuleersin ym. (2015) mukaan etenkin ikä, koulutus ja monikielisyys ovat tärkeitä sanavaraston kokoon vaikuttavia tekijöitä. Mitä vanhempi, kouluttautuneempi tai enemmän kieliä osaava, sitä suurempi ensimmäisen kielen sanavarasto ihmisellä Keuleersin ym. (2015) tutkimuksen mukaan oli. Brysbaertin ym. (2016) mukaan sanavaraston kokoon vaikuttaa myös esimerkiksi se, miten paljon ihminen lukee tai katsoo televisiota.

Sanavarastoa voidaan yksinkertaisimmillaan ajatella yksittäisten sanojen ja niiden merkitysten varastoksi, mutta sanan tai sanavaraston määrittely ei ole yksiselitteistä (Read, 2000). Esimerkiksi Hoff (2009) määrittelee sanan symboliksi, jolla on oma merkityksensä, ja jota voidaan käyttää viittaamaan tiettyihin asioihin. Sanan osaaminen on kuitenkin jatkumo, ja sanat liittyvät aina toisiinsa (Nagy & Scott, 2000). Brysbaertin ym. (2016) mukaan sanamäärä eri listauksissa voi vaihdella suuresti sen perusteella, otetaanko mukaan esimerkiksi sanan taivutusmuodot, yhdyssanat, fraasit, saman sanan eri merkitykset tai murre sanat. Sanavarastoa ja nimeämistaitoja voidaan arvioida muun muassa erilaisilla sanalistoihin perustuvilla testeillä. Ihmisen hallitsemaa sanavarastoa voidaan arvioida hyvin monesta eri näkökulmasta, ja arvioinnissa voidaan keskittyä niin ymmärtävään kuin tuottavaankin sanavarastoon (Read, 2000). Sanavarastoa saatetaan tutkimusten lisäksi arvioida esimerkiksi lasten tai afasiasta kärsivien aikuisten kielellisiä taitoja arvioitaessa.

1.2.1 Sanavaraston kehitys yksi- ja kaksikielisillä

Sanavaraston kehittymisen nopeus ja sanavaraston koko ovat hyvin yksilöllisiä niin yksikielisillä kuin kaksikielisilläkin lapsilla, mikä asettaa oman haasteensa lasten vertailuun. Hoffin (2009) mukaan sanavaraston kehittymisen erojen taustalla olevat syyt voivat liittyä lapsen ympäristöön tai lapsen ominaisuuksiin. Ympäristötekijöitä ovat esimerkiksi lapsen saaman kielellisen syötteen määrä ja laatu (Hoff, 2009) sekä perheen sosioekonominen asema (Fernald ym., 2013). Lapsen ominaisuuksiin liittyviä tekijöitä taas ovat esimerkiksi ikä ja sukupuoli (Simonsen

ym., 2014) sekä persoonallisuus, muiden kielellisten taitojen kehittymisen nopeus ja fonologisen muistin kapasiteetti (Hoff, 2009).

Joidenkin tutkimusten mukaan kaksikielisillä lapsilla voi olla pienemmät sanavarastot yksittäisillä osaamillaan kielillä kuin yksikielisillä verrokeilla, mutta kun kaksikielisiltä lapsilta huomioidaan heidän sanavarastonsa molemmilla kielillä, sanavaraston koko on sama tai jopa isompi kuin yksikielisillä lapsilla (Hoff ym., 2012; Pearson ym., 1993). Tätä on perusteltu sillä, että kaksikielisten kielellinen syöte jakautuu kahdelle eri kielelle, jolloin he altistuvat vähemmän yksittäiselle kielelle (Hoff ym., 2012; Keuleers ym., 2015). Keuleersin ym. (2015) mukaan lähi-kielillä omaksuttu sanasto kuitenkin kasvattaa sanavaraston kokoa myös ensimmäisellä kielellä, ja tämä kasvu korvaa vähentyneen altistumisen vaikutuksen. Lisäksi mitä paremmin ihminen osaa jotakin kieltä, sitä harvemmin hän kohtaa itselleen tuntemattomia sanoja, jolloin toisen kielen aiheuttama vähentynyt altistuminen ensimmäiselle kielelle ei vaikuta niin paljoa (Keuleers ym., 2015).

Pearson ym. (1993) tutkivat 0;8–2;6-vuotiaita englantia sekä englantia ja espanjaa puhuvia lapsia, ja huomasivat, että yksikielisten ja kaksikielisten lasten välillä ei ollut tilastollisesti merkitsevää eroa siinä, miten nopeasti he omaksuivat varhaista sanastoa, kun kaksikielisiltä huomioitiin molemmat kielet. Samaan tulokseen päätyivät myös Hoff ym. (2012), jotka tutkivat 1;10–2;6-vuotiaita lapsia. Yhtä kieltä tarkasteltaessa yksikieliset lapset olivat merkittävästi edistyneempiä sanaston ja kieliopin osalta, mutta kun kaksikielisiltä lapsilta otettiin huomioon molemmat kielet, eroa ei nähty. Kaksikieliset lapset siis saattavat oppia yksittäisiä kieliä hie-man yksikielisiä lapsia hitaammin. Pearson ym. (1993) huomauttaakin, että kaksikielisten lasten kielellistä kehitystä arvioitaessa pitäisi aina huomioida lapsen kielellinen suoriutuminen molemmilla kielillä.

De Houwer ym. (2014) sen sijaan eivät löytäneet eroa 1;1–1;8-vuotiaiden yksikielisten hollantia ja kaksikielisten hollantia ja ranskaa puhuvien lasten tuottavan sanavaraston koossa, ja he suosittelivatkin, että erojen löytyessä pitäisi tarkistaa muut vaikuttavat tekijät kuin kaksikielisyys. Kaksikieliset 1;8-vuotiaat lapset kuitenkin ymmärsivät eri sanoja yhteensä merkitsevästi enemmän kuin yksikieliset lapset, ja tutkijoiden mukaan tähän vaikuttaa kaksikielisten lasten kuuleman kielellisen syötteen suurempi variaatio (De Houwer ym., 2014). Myös Thordardottirin (2011) tutkimuksen mukaan samanaikaisesti kaksikielisillä englantia ja ranskaa puhuvilla lapsilla kielelle altistumisen määrä oli vahvasti yhteydessä sekä tuottavaan että

vastaanottavaan sanavarastoon. Molemmille kielille saman verran altistuneet lapset suorituivat yksikielisten veroisesti vastaanottavan sanaston osalta, mutta tuottavan sanaston kohdalla tarvittiin hieman enemmän kyseiselle kielelle altistumista. Thordardottirin (2011) mukaan tähän saattavat vaikuttaa muun muassa kielten tasapainoinen asema Montrealissa sekä se, että kielet ovat lähellä toisiaan.

Monet tutkimukset ovat keskittyneet vertailemaan eroja yksikielisten ja kaksikielisten omaksumien sanojen määrissä. Kutsuki (2021) kuitenkin tutki sitä, omaksuvatko 2;0–3;3-vuotiaat japania ja englantia äidinkielenään puhuvat kaksikieliset lapset samantyyppisiä sanoja kuin pelkkää japania äidinkielenään puhuvat yksikieliset lapset. Tutkimuksen mukaan molempien ryhmien lapsilla omaksutut sanat olivat jakautuneet samoihin sanaluokkiin, mutta kaksikieliset lapset tuottivat vähemmän vuorovaikutukseen liittyviä sanoja. Tähän saattoi vaikuttaa se, että kaksikielisillä lapsilla oli vähemmän altistumista toiselle kielistä, ja altistumistilanteet eivät olleet niin monipuolisia, jos toista kieltä kuuli esimerkiksi vain kotona. Tutkimuksen otoskoko kuitenkin oli pieni (N=12) ja kaksikielisillä lapsilla kielille altistuminen ei ollut tasaista. Kaksikieliset lapset olivat myös keskimäärin 8 kuukautta yksikielisiä vanhempia, mikä Kutsukin (2021) mukaan viittaa siihen, että kaksikieliset lapset tarvitsevat yksikielisiä lapsia enemmän aikaa sanaston omaksumiseen. Tutkimuksen tulokset ovatkin lähinnä suuntaa antavia.

Tutkimusten mukaan kaksikielisten lasten kokonaissanavaraston koko ei siis yleensä eroa yksikielisten lasten sanavaraston koosta, mutta kun kaksikielisten lasten eri kielten sanavarastoja tarkastellaan erikseen, yksittäisen kielen sanavaraston koko saattaa olla pienempi kuin yksikielisellä lapsella. Tämä saattaa näkyä myös tämän tutkimuksen kaksikielisillä lapsilla. Hypoteesina tässä tutkimuksessa onkin, että suomea ja ruotsia puhuvat kaksikieliset lapset käyttäisivät hieman yksinkertaisempaa sanastoa kuin yksikieliset suomea puhuvat lapset.

1.3 Kerrontataidot ja kertomukset

Tässä tutkimuksessa yksi- ja kaksikielisiä lapsia pyydettiin kertomaan kertomus kuvakirjan perusteella. Kertomuksella tarkoitetaan jonkin tapahtumasarjan kuvausta suullisesti tai kirjallisesti (Tieteen termipankki, 2021). Kertomuksia voi olla hyvin monenlaisia, ja ne voivat olla sekä spontaaneja että houkuteltuja (Suvanto & Mäkinen, 2011). Suvanto ja Mäkinen (2011) jakavat tekstissään kertomukset henkilökohtaisiin kertomuksiin, skripteihin sekä fiktiivisiin kertomuksiin. Henkilökohtaisilla kertomuksilla tarkoitetaan omakohtaisten tapahtumien kuvausta, ja omista kokemuksista kertominen on ihmisille yleensä luonnollista. Skripteillä tarkoitetaan

selkeitä ja informatiivisia kertomuksia siitä, mitä tietyssä tilanteessa yleensä tapahtuu. Fiktiovisillä kertomuksilla taas tarkoitetaan yleensä viihdyttäviksi tarkoitettuja tarinoita. (Suvanto & Mäkinen, 2011.) Kertomuksissa olennaista on niiden yhtenäisyys, mikä tekee tarinasta tilanteeseen nähden mielekkään, ja sen takia tarinoissa onkin yleensä aina tiettyjä kokonaisuuden kannalta tärkeitä elementtejä (Schneider ym., 2006; Mäkinen & Kunnari, 2009).

1.3.1 Kerrontataitojen kehitys

Kertomukset ovat osa lapsuutta, ja yleensä lapset kuulevat erilaisia tarinoita sekä koulussa että vapaa-ajallaan (Botting, 2002; Schneider ym., 2006). Kertomusten kertominen sen sijaan vaatii monenlaisten kielellisten ja kognitiivisten taitojen käyttöä (Mäkinen & Kunnari, 2009; Schneider ym., 2006). Lapsella tulee olla tietoa tarinan struktuurista ja kontekstista sekä kokemuksia kertomuksen aiheista, ja lisäksi hänen pitää osata käyttää tietoja hyväkseen kertoessaan tarinaa (Hudson & Shapiro, 1991). Kerrontataidot kehittyvätkin varhaisesta lapsuudesta varhaisaikuisuuteen asti (Mäkinen & Kunnari, 2009). Kertomustyyppien vaativuudessa on myös eroja, ja yhtenäisen ja järkevän tarinan kertominen esimerkiksi kuvasarjan perusteella vaatii ihmiseltä kognitiivisia taitoja henkilökohtaisten kertomusten kertomista enemmän (Suvanto & Mäkinen, 2011).

Justice ym. (2006) huomasivat tutkimuksessaan, että kertomusten mikrostrukturi kehittyi etenkin 5–10-vuotiailla lapsilla, ja myös Schneiderin ym. (2006) mukaan taidot kehittyivät huomattavasti ennen kouluikää. Hudson ja Shapiro (1991) taas huomasivat tutkimuksessaan, että alakouluikäisten keksimissä tarinoissa oli jo melko paljon rakennetta, mutta iän myötä lasten tarinat kuitenkin monipuolistuivat ja pitenivät, ja niissä oli nähtävissä enemmän strukturia. Botting (2002) kerää artikkelissaan yhteen useiden tutkimusten tuloksia, joiden mukaan nuorempien lasten tarinat ovat yleensä lyhyempiä, niissä on vähemmän eri sanoja, niissä on helpompi syntaksi ja yksinkertaisempi rakenne. Tämän takia onkin tärkeää, että kahta ryhmää toisiinsa verrattaessa ryhmät ovat mahdollisimman samankaltaisia iältään.

1.3.2 Kerrontatehtävät arviointimenetelmänä

Boerman ym. (2016) mukaan kerrontataitoja tutkitaan usein joko tarinan uudelleenkertontatehtävän tai tarinanmuodostustehtävän avulla. Uudelleenkertontatehtävässä koehenkilö kuulee tarinan ja hänen tulee kertoa se uudestaan, kun taas tarinanmuodostustehtävässä koehenkilön tulee keksiä tarina itse esimerkiksi kuvasarjan avulla (Boerma ym., 2016). Kerrotusta tarinasta voidaan tarkastella niin käytettyä makro- kuin mikrostrukturiakin.

Makrostruktuurilla tarkoitetaan tekstin pääajatusta, kun taas mikrostruktuurilla tarkoitetaan esimerkiksi lauseiden merkityksiä ja virkkeiden välisiä merkityssuhteita (Mäkinen & Kunnari, 2009). Mäkisen ja Kunnarin (2009) mukaan kerrontatehtäviä käytetään kliinisessä puheterapiatyössä etenkin kuvailevan tiedon lähteenä, sillä suomenkielisiä ja suomalaiseen kulttuuriin sopivia kerronnan arviointimenetelmiä ei juuri ole. Puheterapeuttien kokemuksen ja ammattitaidon merkitys ovatkin suuressa roolissa (Smolander ym., 2016). Kerrontatehtävillä pystytään arvioimaan esimerkiksi lauseiden yhdistelyä ja laajempaa kokonaisuuksien hallintaa, jota yksittäisten lausettehtävien perusteella ei pystytä arvioimaan (Mäkinen & Kunnari, 2009).

Boerman ym. (2016) mukaan kerrontatehtäviä voidaan käyttää myös kaksikielisten lasten kielellisten taitojen arvioinnissa. Normitettut testit saattavat olla ongelmallisia kielellisesti tai kulttuurillisesti moninaisista taustoista tulevien kohdalla, sillä testit on usein normitettu tietynkielillä ja -maalaisilla lapsilla, ja ne saattavat olettaa, että lapset ovat altistuneet tietynlaisille sanastoille ja kokemuksille elämänsä aikana (Laing & Kamhi, 2003). Ongelma kaksikielisten kielellisiä taitoja arvioitaessa on myös se, että kielen taitotaso ei välttämättä ole samanlainen kaikissa konteksteissa, ja siksi kaksikielisiä testattaessa tulisikin huomioida se, millaisissa tilanteissa ihminen osaamiaan kieliä käyttää (Wei, 2000). Lisäksi hankaluutta voi aiheuttaa testajan tai testattavan käyttämä kieli, ja se, millaista kieltä testattavalta lapselta oletetaan (Laing & Kamhi, 2003). Ei-standardoitu tarinankerrontatehtävä saattaa siis olla kaksikielisiä lapsia arvioitaessa parempi tutkimusmenetelmä kuin standardoitu testi.

Tutkimuksissa käytetään usein fiktiivisiä kertomuksia (Suvanto & Mäkinen, 2011). Bottingin (2002) mukaan narratiivit ovat yleensä strukturoidumpia kuin spontaanipuhe, joten niitä on helpompi käsitellä. Lisäksi spontaanikeskustelussa keskustelukumppanilla on merkittävä rooli, ja sitä on vaikea kontrolloida (Botting, 2002). Samasta kuvasarjasta kerrottujen tarinoiden vertailu toisiinsa on myös yleensä helpompaa kuin esimerkiksi henkilökohtaisten tarinoiden (Suvanto & Mäkinen, 2011). Lasten kielellisestä kehityksestä on myös olemassa paljon tutkimustietoa, joten narratiivien avulla ihmisryhmiä voidaan myös vertailla toisiinsa (Botting, 2002). Suvannon ja Mäkisen (2011) mukaan tutkimuskäytössä fiktiivisten kertomusten etu on myös se, että lapsen voi olla hankala alkaa kertoa esimerkiksi henkilökohtaista kertomusta vieraalle aikuiselle, mutta kuvien perusteella tarinan kertominen voi olla helpompaa. Kaikille lapsille tarinoiden kertominen ei kuitenkaan ole tuttua, mikä vaikuttaa myös heidän kykyynsä keksiä tarina tutkimustilanteessa (Suvanto & Mäkinen, 2011). Tässä tutkimuksessa lapsia pyydettiin

muodostamaan tarina tietyn kuvasarjan perusteella, jotta yksi- ja kaksikielisiltä lapsilta saatiin kerättyä yhtenäisiä ja vertailtavia tarinoita.

1.4 Sanaston arviointi

Tämän tutkimuksen tavoitteena on selvittää, millaisia samankaltaisuuksia tai eroja yksikielisen suomea puhuvien sekä kaksikielisten suomea ja ruotsia puhuvien lasten sanavalintojen välillä on heidän kertoessaan tarinaa kuvasarjan perusteella. Käytetyn sanaston arviointiin on käytetty kahta tekijää, sanaston monipuolisuutta ja sanojen yleisyyttä. Sanaston monipuolisuudella voidaan tarkoittaa sitä, miten paljon tekstissä on erilaisia sanoja eli leksikaalista variaatiota (Johansson, 2008). Sanaston yleisyydellä taas tarkoitetaan sitä, miten yleisiä tai harvinaisia sanoja tekstissä on käytetty (Read, 2000). Sanaston monipuolisuus ja yleisyys ovat tiiviissä yhteydessä toisiinsa, ja Johnsonin ym. (2016) mukaan sekä harvinaisten sanojen että monipuolisten sanojen käyttö on yhteydessä parempaan suoriutumiseen kirjoitustehtävässä toisella kielellä. Read (2000) määrittelee monipuolisuuden ja yleisyyden leksikaalisen rikkouden (lexical richness) osatekijöiksi. Muita tekijöitä Readin (2000) mukaan ovat leksikaalinen tiiviyys, eli sisältösanojen osuus kokonaissanamäärästä, sekä virheet sanaston käytössä. Seuraavissa kappaleissa näitä esitellään lisää.

1.4.1 Sanaston monipuolisuus

Sanaston monipuolisuudella voidaan tarkoittaa muun muassa sitä, miten paljon erilaisia sanoja ihminen osaa, tai miten paljon eri sanoja tekstissä esiintyy (Johansson, 2008). Ihmisen osaamien sanojen määrän lisäksi oleellista on se, miten hän käyttää sanastoaan esimerkiksi tarinoita kertoessaan. Voidaan siis väittää, että parempi sanaston hallinta johtaa monipuolimpaan sanaston käyttöön. Sanaston monipuolisuutta voidaan arvioida useilla eri menetelmillä. Arviointiin on käytetty muun muassa itsearviointia, mutta koska sanan osaaminen on jatkumo (Nagy & Scott, 2000), niin ihmiset myös arvioivat osaamansa sanat eri tavoilla (Daller ym., 2003). Tämän takia itsearviointi ei olekaan kovin luotettava tapa arvioida sanaston monipuolisuutta (Daller ym., 2003). Monipuolisuuden mittaaminen ei ole yksinkertaista, minkä takia on kehitetty erilaisia matemaattisia malleja sanaston monipuolisuuden arviointiin. Silti kaikissa menetelmissä on yleensä omat huonot puolensa (Van Hout & Vermeer, 2007). Sanaston monipuolisuuden arvioimisen menetelmiä on kritisoitu esimerkiksi siitä, etteivät ne ota huomioon käytetyn sanaston yleisyyttä tai vaikeutta (Vermeer, 2000), minkä takia tässä tutkimuksessa tarkastellaan erikseen sanaston yleisyyttä.

Monien menetelmien, kuten Type-token ration (TTR) ja siihen perustuvien The moving-average type-token ration (MATTR) (Covington & McFall, 2010) ja Guiraudin indeksin, perusajatuksena on verrata tekstissä esiintyvien eri sanojen (types) määrää tekstin kokonaissanamäärään (tokens). TTR:ssä erilaisten sanojen määrä jaetaan tekstin kokonaissanamäärällä, jolloin saadaan TTR-arvo. Käytännössä, mitä suurempi TTR-arvo on, sitä enemmän aineistossa on leksikaalista variaatiota (Nation & Meara, 2010). Johanssonin (2008) mukaan puhutuissa aineistoissa TTR-arvo on yleensä kirjoitettua tekstiä matalampi, sillä puheessa on enemmän toistoa kuin kirjoitetussa tekstissä.

TTR on paljon käytetty menetelmä, mutta sitä on myös kritisoitu paljon (Daller ym., 2003), ja esimerkiksi Vermeer (2000) pitää TTR:ää huonona sanaston monipuolisuuden mittarina. Ongelma TTR:ssä ja monissa muissa sanaston monipuolisuuden arviointiin käytetyissä menetelmissä on se, että analysoitavan tekstin pituus vaikuttaa saatuun arvoon. Pidemmässä teksteissä TTR-arvo on usein lyhyitä tekstejä pienempi, sillä mitä pidempään teksti jatkuu, sitä todennäköisempää on, että sana on jo esiintynyt aiemmin kyseisessä tekstissä (Daller ym., 2003). Covington ja McFallin (2010) mukaan tätä ongelmaa on yritetty ratkaista muun muassa standardoimalla tekstin pituus ja muokkaamalla menetelmää ottamaan huomioon tekstin pituus. Tässä tutkimuksessa tekstin pituuden standardointi ei ollut vaihtoehto, sillä lapset saivat kertoa tarinan haluamallaan tavalla. Onkin kehitetty erilaisia tekstin pituuden huomioivia menetelmiä, esimerkiksi MATTR, jossa TTR lasketaan tietyn liikkuvan ikkunan sisällä olevista sanoista (Covington & McFall, 2010), sekä Guiraudin indeksi, jossa TTR lasketaan käyttämällä nimittäjänä kokonaissanamäärän neliöjuurta (Daller ym., 2003). Neliöjuuren avulla pidempi teksti saa korkeamman arvon kuin lyhyt teksti TTR:n ollessa teksteissä sama (Daller ym., 2003).

Sanaston monipuolisuutta voidaan siis arvioida monella tavalla, mutta tässä tutkimuksessa päädyttiin käyttämään Guiraudin indeksia. Guiraudin indeksissä tekstin saama arvo lasketaan jakamalla tekstissä esiintyvien eri sanojen määrä kokonaissanamäärän neliöjuurella. Myös Guiraudin indeksia on kritisoitu siitä, että muiden TTR:ään perustuvien metodien ohella sekkään ei ratkaisisi tekstin pituuden tuomaa ongelmaa kokonaan (Hess ym., 1989). Vermeerin (2000) ja Dallerin ym. (2003) mukaan Guiraudin indeksi kuitenkin ottaa tekstin pituuden huomioon paremmin kuin perinteinen TTR, ja on siksi parempi menetelmä sanaston monipuolisuuden arvioinnissa.

Ihmisillä, jotka käyttävät enemmän harvinaisia sanoja, on todennäköisesti parempi kyseisen kielen taitotaso ja suurempi sanavarasto (Daller ym., 2003). Tämän takia Dallerin ym. (2003) tutkimuksen mukaan niin sanotut edistyneet (advanced) menetelmät toimivat sanaston monipuolisuuden arvioinnissa paremmin kuin perinteiset tavat arvioida sanaston monipuolisuutta. Edistyneissä versioissa TTR:stä ja Guiraudista arviointiin otetaan mukaan jokaiselta koehenkilöltä ainoastaan sanat, jotka eivät ole esimerkiksi kyseisen kielen 1000 tai 2000 yleisimmän sanan joukossa (Daller ym., 2003). Tässä tutkimuksessa sanaston monipuolisuutta arvioitiin Guiraudin indeksin lisäksi myös edistyneellä Guiraudin indeksillä.

1.4.2 Sanaston yleisyys

Sanaston monipuolisuuden arviointi on siis haastava tehtävä. Vermeer (2000) suosittaleekin tekstissään tarkastelemaan monipuolisuuden lisäksi käytettyjen sanojen yleisyyttä. Sanan yleisyyttä voidaan arvioida eri tavoilla, mutta perinteisesti sitä on arvioitu tarkastelemalla tarinoissa käytetyn sanaston esiintymistä jostakin laajemmasta aineistosta tehdyssä tekstipanikissa eli korpuksessa (Gernsbacher, 1984). Korpuksia voidaan rakentaa eri lähteissä, esimerkiksi sanomalehdissä tai keskustelupalstoilla, esiintyvistä teksteistä. Korpuksissa sanojen yleisyys mitataan yleensä lemموjen mukaan, sillä muuten esimerkiksi sanojen taivutusmuodot sekoittaisivat yleisyyden laskemista (Nation & Meara, 2010). Nationin ja Mearan (2010) mukaan puhutun kielen arvioinnissa käytettävän korpuksen tulisi sisältää aineistoa myös puhutusta kielestä, jotta korpus olisi mahdollisimman hyvä sanaston yleisyyden arviointiin, mutta korpuksia on usein koottu nimenomaan kirjoitetuista teksteistä.

Gernsbacher (1984) vetää tekstissään yhteen aiempia tutkimuksia, joiden perusteella korpuksissa esiintyvät yleiset ja harvinaiset sanat eroavat toisistaan monilla tavoilla. Yleiset sanat ovat myös foneemisilta ominaisuuksiltaan harvinaisia sanoja yleisempiä, eli niissä esiintyy tumpia äänneyhdistelmiä, ja lisäksi korpuksissa yleiset sanat ovat usein konkreettisempia ja enemmän itsenäisiä merkityksiä sisältäviä kuin harvinaiset sanat (Gernsbacher, 1984). Tanaka-Ishii ja Teradan (2011) mukaan esimerkiksi englannin kielessä yleisimpiä sanoja ovat kuitenkin funktiosanat, joilla ei ole omaa merkitystä, kun taas tutuimpia ovat sanat, joilla on itsenäinen merkitys, eli esimerkiksi substantiiveja ja verbejä. Yleisesti ottaen kuitenkin yleiset sanat ovat tutumpia kuin harvinaiset sanat (Vermeer, 2000). Tuttuus ja yleisyys ovatkin käsitteinä lähellä toisiaan. Sanan täytyy olla tuttu voidakseen olla yleinen, mutta kaikki tutut sanat eivät silti välttämättä ole tekstiaineistoissa yleisiä (Tanaka-Ishii & Terada, 2011). Tanaka-Ishii ja

Teradan (2011) tutkimuksen mukaan tuttuus ja yleisyys korreloivat keskenään, mutta korrelaatio on suurempaa puhutun kielen kuin kirjoitetun kielen kanssa. Tämä sopii myös siihen, että ihmiset käyttävät puheessaan yleisempiä sanoja kuin kirjoittaessaan (Berman & Nir, 2010; Hayes, 1988). On myös huomattu, että ihminen prosessoi yleisiä sanoja nopeammin ja tehokkaammin kuin harvinaisia sanoja, mutta prosessointinopeus kuitenkin vaihtelee yksilöllisesti (Brysbaert ym., 2018; Gernsbacher, 1984).

Sanaston yleisyyttä voidaan arvioida myös vertaamalla aineistossa esiintyviä sanoja listoihin kyseisessä kielessä esiintyvistä yleisistä sanoista (Tanaka-Ishii & Terada, 2011). Yleisyyslistoja voidaan käyttää hyväksi esimerkiksi valittaessa sanastoa kielen opetukseen (Keller, 1991). Sanojen yleisyys vaihtelee eri kielten ja kulttuurien välillä (Keller, 1991), eikä samoja sanalistoja siten voi käyttää erikielissä tutkimuksissa. Suomen kielen yleisimpien sanojen lista on vuodelta 1979 (Saukkonen, 1979), joten tässä tutkimuksessa sanaston yleisyyttä päädyttiin arvioidaan tarkastelemalla tarinoissa käytetyn sanaston esiintymistä Finnish Internet Parsebank-korpuksessa (2021; Luotolahti ym., 2015), joka sisältää yhteensä 3 400 000 058 sanaa, joista 2 813 810 000 on eri sanoja. Tämän jälkeen sanat jaettiin analysoinnin helpottamiseksi neljään yleisyysryhmään.

1.5 Tutkimuskysymykset

Tämän tutkimuksen tavoitteena on selvittää, onko yksikielisten suomea puhuvien ja kaksikielisten suomea ja ruotsia puhuvien suomalaislasten sanaston monipuolisuudessa sekä käytettyjen sanojen yleisyydessä eroja, ja jos on, niin millaisia. Monipuolisuutta arvioidaan sekä koko aineiston osalta että jättämällä pois hyvin yleiset sanat, eli sanat, joiden yleisyys Finnish Internet Parsebank-korpuksessa (2021) on yli 1000 per miljoona sanaa. Yleisyyden osalta taas arvioidaan, jakautuvatko tarinoissa käytetyt sanat yleisiin ja harvinaisiin sanoihin samalla vai eri tavalla yksi- ja kaksikielisillä lapsilla. Yleisillä sanoilla tarkoitetaan tässä sanoja, joiden yleisyys korpuksessa on yli 100 per miljoona sanaa, ja harvinaisilla sanoilla taas sanoja, joiden yleisyys on alle 10 per miljoona sanaa.

Aiempien tutkimusten perusteella yksi- ja kaksikielisten lasten sanavarastot eivät yleensä eroa toisistaan koon osalta, kun kaksikielisiltä otetaan huomioon molemmat kielet, mutta kun kaksikielisten lasten eri kielten sanavarastoja tarkastellaan erikseen, sanavaraston koko yhdellä kielellä saattaa olla pienempi kuin yksikielisillä lapsilla. Tässä tutkimuksessa kaksikielisiltä lapsilta tarkastellaan heidän käyttämänsä sanastoa vain suomeksi, joten voidaan olettaa, että

yksikielisten lasten käyttämässä sanastossa olisi monipuolisempia ja harvinaisempia sanoja. Toisaalta Suomessa kaksikielisten suomen kielen taidot ovat yleensä hyvällä tasolla, joten voi olla, että eroja ei siksi löydy.

Tutkimuskysymykset ovat:

1. Onko käytetyn sanaston monipuolisuudessa eroja yksi- ja kaksikielisten lasten välillä suomenkielisessä kerrontatehtävässä?
2. Onko käytettyjen sanojen yleisyydessä eroja yksi- ja kaksikielisten lasten välillä suomenkielisessä kerrontatehtävässä?

2 Menetelmät

Tutkimus toteutettiin osana *Speech disfluencies in bilingual speakers (Sibils)* -tutkimusprojektia, jossa tarkastellaan puheen sujumattomuuksia yksi- ja kaksikielisillä puhujilla. Aiemmin projektissa on vuonna 2020 kerätty samanaikaisesti kaksikielisten suomea ja ruotsia äidinkielenään puhuvien lasten aineisto (N=20) (Brandt-Häggblom, 2021; Valo, 2021). Yksikielisten suomea äidinkielenään puhuvien lasten aineiston keräsivät Julia Riihioja ja Heidi Saarijärvi-Lautala kevään 2021 aikana. Aineisto kerättiin aiempaa aineistoa vastaavasti 1–3-luokkalaista tyypillisesti kehittyneiltä lapsilta.

Aineistoa varten koehenkilöitä pyydettiin kertomaan kuvakirjan perusteella tarina ja lisäksi lasten vanhempia pyydettiin täyttämään taustatietokysely. Saatu aineisto litteroitiin ja analysoitiin. Sanaston arviointiin on olemassa monia menetelmiä ja lähestymistapoja, ja tässä tutkimuksessa aineistosta arvioitiin käytetyn sanaston yleisyyttä ja monipuolisuutta. Yleisyyttä arvioitiin tarkastelemalla aineiston sanaston yleisyyttä Finnish Internet Parsebank-korpuksessa (2021). Sanaston monipuolisuutta taas arvioitiin käyttämällä tavallista ja edistynyttä Guiraudin indeksiä. Guiraudin indeksi laskettiin jakamalla tarinoissa esiintyvien eri sanojen määrä tarinan kokonaissanamäärän neliöjuurella. Edistynyttä Guiraudin indeksiä varten taas aineistosta poistettiin yleisimmät sanat, minkä jälkeen indeksi laskettiin uudestaan. Seuraavissa kappaleissa menetelmiä esitellään tarkemmin.

2.1 Koehenkilöt

Yksikielisten lasten aineisto kerättiin 24 yksikieliseltä äidinkielenään suomea puhuvalta lapselta helmi-maaliskuun 2021 aikana. Lapset olivat iältään 7;0–11;1-vuotiaita.

Poissulkukriteerejä olivat kaksikielisyys sekä riski epätyypilliselle kielenkehitykselle PaBiQ-kyseilyn (COST Action IS0804, 2011; Tuller, 2015) perusteella. Lisäksi aineistona tutkimuksessa oli aiemmin Sibils-projektissa vuonna 2020 kerätty kaksikielisten suomea ja ruotsia äidinkielenään puhuvien lasten aineisto (Brandt-Hägglom, 2021; Valo, 2021), josta tähän tutkimukseen otettiin mukaan kaksikielisten lasten suomenkielinen osuus. Kaksikielisiä lapsia aineistossa oli yhteensä 20, ja iältään he olivat 6;7–11;6-vuotiaita. Yhteensä tutkimuksessa oli siis 44 koehenkilöä, jotka olivat keskimäärin 8;7-vuotiaita (taulukko 1). Yksikielistä lapsista tyttöjä oli 15 (62,5 %) ja poikia 9, kun taas kaksikielisten ryhmässä tyttöjä oli 14 (70,0 %) ja poikia 6. Kaksikielistä lapsista vanhempien mukaan 12 lasta (60 %) puhui molempia kieliä yhtä paljon ja olivat tasapainoisesti kaksikielisiä, ja neljällä lapsella suomi (20 %) ja neljällä ruotsi (20 %) oli lapsen dominantti kieli. Kaksikieliset lapset olivat kotoisin Manner-Suomesta, eli yksikään lapsi ei asunut kokonaan ruotsinkielisessä kunnassa.

Aineiston analysointi aloitettiin tarkastelemalla yksikielisten ja kaksikielisten ryhmien samankaltaisuutta iän ja sukupuolijakauman osalta. Ikä oli molemmissa ryhmissä normaalisti jakautunut ($p > .05$), joten ryhmien välisiä eroja tarkasteltiin riippumattomien otosten T-testillä. Testin perusteella ryhmät eivät iältään eronneet toisistaan tilastollisesti merkitsevästi: $t(42) = -.655$, $p = .516$. Khiin neliö -testin perusteella ryhmät eivät myöskään sukupuolen osalta eronneet toisistaan tilastollisesti merkitsevästi: $\chi^2(1) = .273$; $p = .601$.

Taulukko 1. Yksikielisten ja kaksikielisten koehenkilöiden perustiedot

	N	Sukupuoli		Ikä (kk)
		Tytöt	Pojat	
Yksikieliset	24	15	9	84–133 (ka 101.5, kh 11.9)
Kaksikieliset	20	14	6	79–138 (ka 104.1, kh 14.8)
Yhteensä	44	29	15	79–138 (ka 102.7, kh 13.2)

2.2 Kerrontatehtävä

Varsinaisessa aineistonkeruussa käytettiin Mercer Mayerin kuvasatukirjan saduista joko ”Sammakko menee ravintolaan” (Frog goes to dinner; Mayer, 1974) tai ”Sammakon oma seikkailu” (Frog on his own; Mayer, 1973) -satua. Sadut ovat 32 sivua pitkiä ja niissä ei ole tekstiä. Heilmannin ym. (2016) mukaan ”Sammakon oma seikkailu” -tarinassa on 27 kohtausta, 8 episodua ja 6 hahmoa, kun taas ”Sammakko menee ravintolaan” -tarinassa on 23 kohtausta, 11 episodua ja 11 hahmoa. Heilmannin ym. (2016) mukaan tarinoiden välillä oli vain hieman eroa siinä, miten paljon eri sanoja sekä miten paljon ja pitkiä ilmaisuja lapset käyttivät kertoessaan tarinoita. Kuitenkin sekä englantia että espanjaa äidinkielenään puhuvat lapset käyttivät keskimäärin 4–6 eri sanaa enemmän kertoessaan ”Sammakko menee ravintolaan” -tarinaa kuin ”Sammakon oma seikkailu” -tarinaa (Heilmann ym., 2016).

Kertomuksen kertominen valikoitui aineistonkeruumenetelmäksi aiemmin projektissa kerätyn aineiston perusteella. Lisäksi kuten edellä esiteltiin, kerrontatehtävät voivat olla kaksikielisille lapsille standardoituja testejä oikeudenmukaisempi tutkimusmenetelmä (Laing & Kamhi, 2003), ja kuvasarjoista keksittyjä kertomuksia on spontaanipuhetta helpompaa vertailla (Suvanto & Mäkinen, 2011). Kaksikieliset lapset kertoivat toisen sadun suomeksi ja toisen ruotsiksi. Se, kumman sadun yksikielinen lapsi sai, määräytyi siten, että yksikielisten lasten aineisto sovitettiin iän ja sukupuolen mukaan mahdollisimman samankaltaiseksi kaksikielisten lasten aineiston suomenkielisen osuuden kanssa.

Yksikielisten ryhmässä tarinat jakautuivat tasan, eli molemmat tarinat kerrottiin 12 kertaa (taulukko 2). Kaksikielisillä ”Sammakon oma seikkailu” -tarina kerrottiin 9 kertaa ja ”Sammakko menee ravintolaan” -tarina 11 kertaa. Yksikielisten lasten tarinat olivat 129–899 sanaa pitkiä (ka 378,3 sanaa, kh 149.05), kun taas kaksikielisten lasten kertomat tarinat olivat 96–494 sanaa pitkiä (ka 319,5 sanaa, kh 97.03). Sekä yksikielisillä että kaksikielisillä ”Sammakko menee ravintolaan” -tarinat olivat keskimääräiseltä sanamäärältään ”Sammakon oma seikkailu” -tarinoita pidempiä, kuten myös Heilmannin ym. (2016) tutkimuksessa.

Taulukko 2. Perustiedot kerrontatehtävästä

	N	Tarina		Sanamäärä	Sanamäärä per tarina	
		Seikkailu	Ravintola		Seikkailu	Ravintola
Yksikieliset	24	12	12	129–899 (ka 378.3, kh 149.1)	129–474 (ka 335.0, kh 105.2)	254–899 (ka 421.7, kh 176.9)
Kaksikieliset	20	9	11	96–494 (ka 319.5, kh 97.0)	181–380 (ka 287.0, kh 63.5)	96–494 (ka 346.1, kh 113.7)

2.3 Aineistonkeruu

Aineistoa kerättiin 1–3-luokkalaisilta yksikielisiltä suomea äidinkielenään puhuvilta lapsilta. Koehenkilöiden hankinta aloitettiin ottamalla sähköpostitse yhteyttä varsinaissuomalaisiin ja satakuntalaisiin alakouluihin ja pyytämällä rehtoreita välittämään tutkimuksen infokirje esimerkiksi Wilma-järjestelmän kautta oppilaiden vanhemmille. Kirjeessä vanhempia pyydettiin ottamaan yhteyttä tutkijoihin sopivan tutkimusajan valitsemiseksi. Vanhemmille toimitettiin sähköpostitse linkki Doodle-ajanvarauskalenteriin sekä vanhemman ja lapsen suostumuslomake tutkimukseen osallistumiselle (liite 1, liite 2). Suostumus tutkimukseen osallistumista varten annettiin vastaamalla sähköpostiin. Vanhemmille lähetettiin myös ohjeet etäyhteyden käyttöä varten. Lisäksi vanhempia pyydettiin täyttämään sähköinen Webropol-kysely, jossa kysyttiin lapsen taustatietoja sekä lapsen kielen kehitykseen liittyviä tekijöitä (liite 3).

Aineistonkeruu tapahtui Zoom-ohjelman avulla etäyhteydellä. Aineistoa oli keräämässä kaksi tutkijaa, jotka keräsivät kahden ensimmäisen lapsen aineistot yhdessä, jotta voitiin varmistaa yhtenäiset käytännöt tutkimustilanteissa. Tutkimustilanteen alussa lapsen kanssa juteltiin lapselle tutuista aiheista, esimerkiksi harrastuksista tai koulusta. Tarkoituksena oli, että lapsi rentoutuisi varsinaista tehtävää varten ja lisäksi saatiin kerättyä lapsen spontaanipuhetta mahdollista jatkotarkastelua varten. Yksikielisten lasten osalta aineistonkeruu kesti noin 15–20 minuuttia, josta noin 5 minuuttia käytettiin jutteluun.

Tutkimus toteutettiin etäyhteyden avulla, joten tutkimustilanteessa kirja jaettiin lapsen näytölle Zoom-ohjelman näytönjakamistoiminnon avulla. Lapselle näytettiin ensin koko satu

lapsen omassa tahdissa niin, että lapsi sai kertoa, kun halusi sivua käännettävän. Tämän jälkeen kuvakirja katsottiin uudestaan läpi ja lasta pyydettiin samalla kertomaan tarina tutkijalle omin sanoin. Myös tässä vaiheessa lapsi sai itse valita etenemistahdin. Tarvittaessa lasta kannustettiin jatkamaan kerrontaa neutraaleilla ilmauksilla, kuten ”mitä sitten tapahtui?”. Tarinankerronnan jälkeen lapselta kysyttiin vielä hänen mielipidettään kuvakirjasta ja mahdollisia esiin nousseita kysymyksiä. Osallistumisen jälkeen lapselle toimitettiin sähköpostilla lahjaksi koodi, jolla hän sai valita GoGift-palvelusta haluamansa 5 euron arvoisen lahjakortin.

Tutkimustapaamiset nauhoitettiin Zoom-ohjelman avulla. Aineistonkeruun jälkeen nauhoitteet litteroitiin ja samalla niistä poistettiin tunnistetiedot, kuten nimet. Aineisto jaettiin tutkijoiden kesken sattumanvaraisesti puoliksi litterointia varten luotettavuuden lisäämistä varten, ja lisäksi epäselvät kohdat litteroitiin yhdessä. Aineisto litteroitiin lasten puheen osalta eksaktisti, eli litteraattiin sisällytettiin esimerkiksi puheessa ilmenevät täytesanat, toistot ja kesken jääneet ilmaisut. Tutkijan ja mahdollisen vanhemman osalta aineisto litteroitiin sanatarkasti ja tarvittaessa referoiden, jotta aineistoa on helpompi tulkita. Lisäksi litteraateista jätettiin pois tarinaan tutustumisen sekä kerronnan aikana esiintyneet sivun vaihtamiseen liittyvät ilmaisut. Nauhoitteet ja litteraatit säilytettiin Turun yliopiston tietoturvalisessa Seafire-pilvipalvelussa ja ne poistettiin tutkimuksen valmistuttua.

2.4 Kysely vanhemmille

Ennen Zoom-tapaamista vanhemmille toimitettiin sähköinen Webropol-kysely, jossa kysyttiin lapsen perustietoja sekä selvitettiin kielelliseen kehitykseen liittyviä taustatekijöitä, lähisuvussa esiintyviä kielellisiä haasteita ja lapsen kielellistä ympäristöä (liite 3). Kyselylomakkeilla voidaan saada luotettavaa ja hyödyllistä tietoa kaksikielisen lapsen kielitaidosta sekä kielitaitoon vaikuttavista tekijöistä lapsen molempien kielten osalta (Tuller, 2015). Osa tässä tutkimuksessa käytetystä kyselystä perustuu muokattuun versioon Questionnaire for Parents of Bilingual Children (PaBiQ) -kyselystä (COST Action IS0804, 2011; Tuller, 2015). Logopedian opiskelijat Brandt-Hägglom ja Valo käänsivät PaBiQ-kyselyn suomeksi kaksikielisten lasten tutkimista varten. Riihioja ja Saarijärvi-Lautala muokkasivat Brandt-Hägglomin ja Valon käännöksen sekä alkuperäisen kyselyn perusteella yksikielisten lasten vanhemmille lähetetyn kyselylomakkeen.

Kyselyä muokattiin sopimaan paremmin yksikielisten lasten arviointiin, ja siksi alkuperäisestä kyselystä otettiin pois kysymykset, joissa käsiteltiin lapsen ja perheen puhumien kielten

käytön jakautumista. Lisäksi joitakin kyselyn kysymyksiä muutettiin koskemaan vain yhtä kieltä. Esimerkiksi tässä tutkimuksessa käytetyn kyselyn kysymyksissä 13–18 vanhempaa pyydetään vertaamaan lastaan muihin samanikäisiin lapsiin, kun taas alkuperäisessä PaBiQ-kyselyssä vertailu pyydetään tekemään kaikkien lapsen osaamien kielten osalta erikseen. Lisäksi kyselyyn lisättiin kysymykset 21 ja 22, joissa kysytään lapsen puheterapiakontakteista, sekä kysymys 24, jossa kysytään lapsen sairauksista tai vammoista. Nämä kysymykset lisättiin, jotta saataisiin vielä enemmän tietoa yksikielisten lasten taustoista ja mahdollisista tuloksiin vaikuttavista tekijöistä. Perhetietoihin myös lisättiin mahdollisuus kertoa äidin ja isän lisäksi muista perheeseen mahdollisesti kuuluvista aikuisista, esimerkiksi vanhempien uusista puolisoista. Vastaajista yksi oli kertonut kolmannesta aikuisesta. Lisäksi kyselyn kieliasua muokattiin hieman, sillä alkuperäinen PaBiQ-kysely oli tarkoitettu tehtäväksi haastatteluna (Tuller, 2015). Kyselyvastaukset saatiin kaikkien yksikielisten lasten ja 18 kaksikielisen lapsen vanhemmalta.

Alkuperäisen PaBiQ-kyselyn sekä tässä tutkimuksessa käytetyn kyselyn avulla voitiin myös laskea riskipisteitä, eli miten suuri riski lapsella on ei-tyypilliselle kielenkehitykselle. Arvo laskettiin sen perusteella, milloin lapsi on sanonut ensimmäisen sanansa ja yhdistänyt sanoja lauseeksi, onko vanhemmilla ollut huolta lapsen kielellisestä kehityksestä, sekä onko lapsen lähi-perheessä erilaisia kielellisiä vaikeuksia. PaBiQ:n perusteella yli 19 pistettä tarkoittaa, että kielellinen kehitys on todennäköisesti tyypillistä, kun taas alle 15 pistettä voidaan nähdä merkkinä mahdollisesta ei-tyypillisestä kielenkehityksestä. Lapsista kukaan ei saanut alle 15 pistettä, ja vain kaksi yksikielistä lasta sai 15–19 pistettä, toinen 16 ja toinen 17 pistettä. Kyseisten lasten aineistot eivät merkittävästi poikenneet muusta aineistosta, joten heidät sisällytettiin tutkimukseen. Riskipisteiden osalta ryhmät eivät olleet normaalisti jakautuneet ($p < .05$), joten ryhmien välistä eroa tutkittiin epäparametrisella Mann-Whitney U-testillä. Testin mukaan ryhmät eivät myöskään riskipisteiden osalta eronneet toisistaan tilastollisesti merkitsevästi: $U(42) = 273.500$, $Z = 1.547$, $p = .122$.

2.5 Sanaston arviointi

Tässä tutkimuksessa aineistosta arvioitiin yksi- ja kaksikielisten koehenkilöiden käyttämän sanaston monipuolisuutta ja yleisyyttä. Aineiston analysointiin käytettiin IBM SPSS Statistics 28 -ohjelmaa sekä Microsoft Excel -ohjelmaa. Toistot tai epäröinnit voivat aiheuttaa haasteita sanaston arvioinnissa, sillä ne vaikuttavat esimerkiksi kokonaissanamäärään (Nation & Meara, 2010). Tässä tutkimuksessa päädyttiin olemaan laskematta mukaan kesken jääneitä sanoja

sekä epäröintejä, sillä ne vaikuttaisivat olennaisesti sanojen määrään aineistossa. Kaikki aineistojen sanat muutettiin analysointia varten perusmuotoon, minkä jälkeen laskettiin, miten monta kertaa kukin sana esiintyi kyseisen henkilön aineistossa.

2.5.1 Sanaston monipuolisuuden arviointi

Tässä tutkimuksessa sanaston monipuolisuuden arviointiin käytettiin Guiraudin indeksiä, joka on muunneltu versio TTR-menetelmästä. Guiraudin indeksissä tekstin saama arvo lasketaan jakamalla tekstissä esiintyvien eri sanojen määrä kaikkien sanojen määrän neliöjuurella. Tämän lisäksi sanaston monipuolisuutta arvioitiin myös edistyneellä Guiraudin indeksillä. Edistyneitä Guiraudin indeksiä varten koehenkilöiden käyttämistä sanoista poistettiin sanat, joiden esiintymistiheys Finnish Internet Parsebank-korpuksesta (2021) oli yli 1000 per miljoona sanaa. Tämän jälkeen Guiraudin indeksi laskettiin uudestaan. Yksi- ja kaksikielisten lasten aineistojen monipuolisuutta verrattiin toisiinsa sekä tavallisen että edistyneen Guiraudin indeksin avulla.

2.5.2 Sanaston yleisyyden arviointi

Tässä tutkimuksessa sanaston yleisyyttä arvioitiin hakemalla kaikille koehenkilöiden käyttämille sanoille niiden yleisyys käyttämällä Finnish Internet Parsebank-korpusta (2021), joka sisältää yhteensä 3 400 000 058 sanaa. Jokaiselle sanalle myös laskettiin sanan yleisyys per miljoona sanaa kyseisen korpuksen perusteella. Sanojen esiintymistiheys vaihteli nollan ja useiden miljoonien välillä, joten sanat jaettiin analysoinnin selkeyttämiseksi neljään ryhmään sen perusteella, mikä on yksittäisen sanan esiintymistiheys miljoonassa sanassa. Ryhmiin jakaminen myös helpotti aineiston analysointia kaksisuuntaisen riippumattomien otosten t-testin avulla. Ryhmät olivat <10 ; ≥ 10 ja <20 ; ≥ 20 ja <100 sekä ≥ 100 per miljoona sanaa. Tämän jälkeen laskettiin, kuinka monta kappaletta ja prosenttia sanoista kuului kuhunkin ryhmään. Lopuksi yksi- ja kaksikielisten koehenkilöiden aineistoja verrattiin toisiinsa käytetyn sanaston yleisyyden osalta.

2.6 Tutkimuksen eettisyys

Tutkimuksen eettisyyden ja perheiden vapaaehtoisuuden varmistamiseksi lapsilta ja heidän vanhemmiltaan pyydettiin sähköpostitse kirjallinen suostumus tutkimukseen osallistumiselle (liite 1, liite 2). Suostumuslomakkeessa myös kerrottiin, että osallistuminen on vapaaehtoista ja heillä on halutessaan oikeus keskeyttää osallistumisensa missä vaiheessa tahansa. Lisäksi perheille kerrottiin tutkimuksen etenemisestä ja heidän tietojensa sekä kerätyn aineiston

käsittelystä ja säilytyksestä. Tutkimuksen teossa käytettiin Turun yliopiston tietoturvallista versiota Zoom-ohjelmasta aineistonkeruuseen sekä Turun yliopiston suojattua Seafile-pilvipalvelua aineiston säilyttämiseen. Litterointivaiheessa kaikki yksittäisen ihmisen tunnistamisen mahdollistavat tiedot poistettiin aineistosta ja tiedostot koodattiin lasten henkilökohtaisilla tunnistenumeroilla. Lisäksi tutkimuksen teon päätyttyä nauhoitteet ja litteraatit poistettiin. Koehenkilöillä oli kuitenkin oikeus pyytää tietojensa poistoa missä vaiheessa tahansa, myös ennen tutkimuksen valmistumista.

Turun yliopiston ihmistieteiden eettisen toimikunnan ihmistieteellisten tutkimusten jaosto hyväksyi tutkimusprojektin tutkimussuunnitelman 17.12.2019 päivätyssä lausunnossaan. Lisäksi projektille haettiin tutkimuslupa Turun kaupungilta ja tutkimukselle tehtiin Henkilötietojen suojaa koskeva vaikutustenarviointi (DPIA) tieteellisessä tutkimuksessa.

3 Tulokset

Tutkimuksen tavoitteena oli selvittää, onko yksikielisten suomea puhuvien ja kaksikielisten suomea ja ruotsia puhuvien suomalaislasten tarinankerrontatehtävässä käyttämissä sanoissa eroja. Tarkoituksena oli tarkastella eroja käytetyn sanaston monipuolisuudessa ja yleisyydessä.

3.1 Sanojen monipuolisuus

Sanaston monipuolisuuden arvioinnissa käytettiin Guiraudin indeksiä sekä edistynyttä Guiraudin indeksiä. Guiraudin indeksi laskettiin jakamalla eri sanojen määrä tarinan kokonaissanamäärän neliöjuurella. Edistynyttä Guiraudin indeksiä varten taas aineistosta poistettiin kaikki sanat, joiden yleisyys Finnish Internet Parsebank-korpuksessa (2021) on yli 1000 per miljoona. Suurempi Guiraudin indeksin arvo tarkoittaa, että aineistossa on enemmän leksikaalista variaatiota.

Yksikielisten lasten tarinat olivat keskimäärin 378,33 sanaa pitkiä, kun taas kaksikielisten lasten kertomat tarinat olivat keskimäärin 319,50 sanaa pitkiä, joten yksikielisten lasten tarinat olivat keskimäärin kaksikielisten tarinoita pidempiä. Yksikielisillä lapsilla Guiraudin indeksi oli keskimäärin 5.70 ja kaksikielisillä 5.38. Yksikielisillä on lukumäärällisesti Guiraudin indeksin perusteella siis keskimääräisesti kaksikielisiä enemmän leksikaalista variaatiota tekstissään. Ryhmät eivät kokonaissanamäärien perusteella olleet normaalisti jakautuneita ($p < .05$), joten ryhmien välisiä eroja tarkasteltiin Mann-Whitney U-testillä, jonka perusteella ryhmien välillä ei

kuitenkaan ollut tilastollisesti merkitsevää eroa ($U(44) = 186.000$, $Z = -1.273$, $p = .203$, $\eta^2 = .038$). Sanaston monipuolisuudessa ei siis Guiraudin indeksin perusteella ole tilastollisesti merkitsevää eroa yksikielisten ja kaksikielisten lasten ryhmien välillä. Tulokset, keskiarvot ja keskihajonnat löytyvät taulukosta 3.

Yksikielisillä edistynyt Guiraudin indeksi oli keskimäärin 6.33 ja kaksikielisillä 5.88. Yksikielisillä lapsilla on lukumäärällisesti siis myös edistyneen Guiraudin indeksin perusteella keskimäärin kaksikielisiä enemmän leksikaalista variaatiota tekstissään. Ero yksikielisten ja kaksikielisten välillä oli edistyneen Guiraudin perusteella suurempi kuin tavallisella Guiraudilla laskettuna, mutta Mann-Whitney U-testin perusteella ryhmien välinen ero ei ollut tilastollisesti merkitsevää ($U(44) = 172.000$, $Z = -1.603$, $p = .109$, $\eta^2 = .060$), vaikka efektikoko olikin kohtalainen. Sanaston monipuolisuudessa yksi- ja kaksikielisten lasten ryhmien välillä ei siis ole tilastollisesti merkitsevää eroa Guiraudin indeksin tai edistyneen Guiraudin indeksin perusteella.

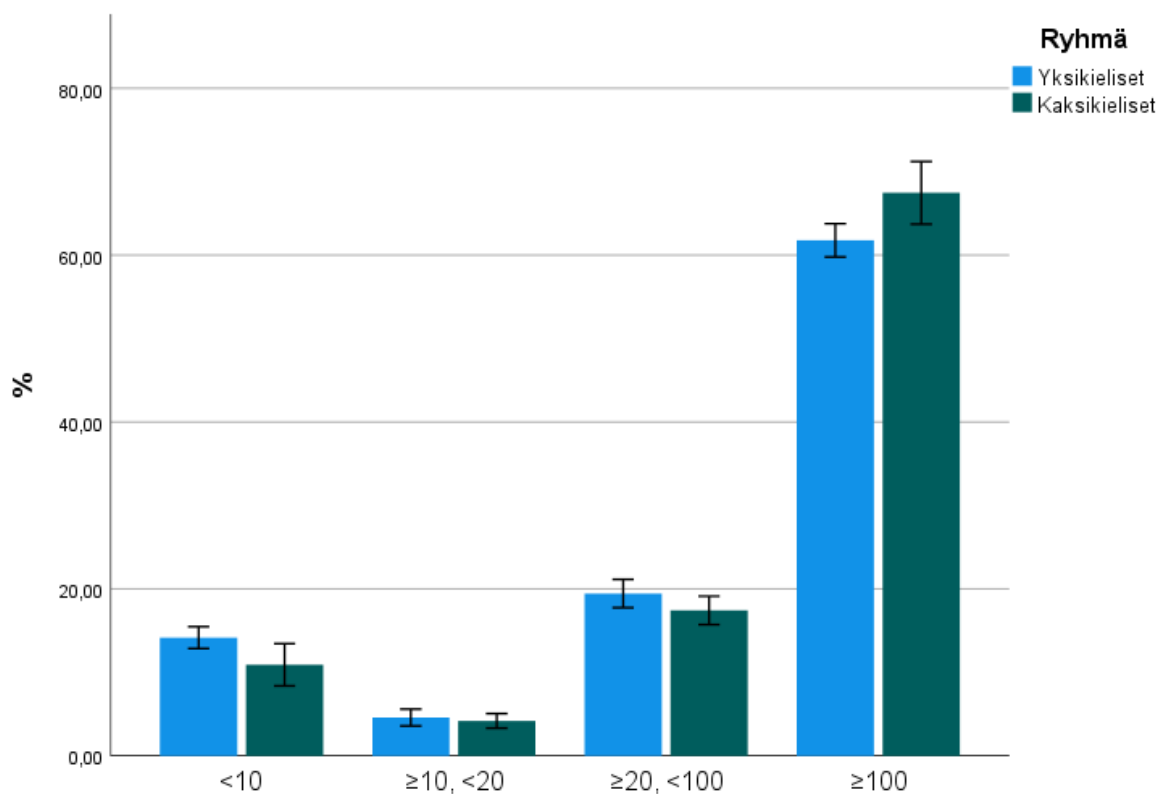
Taulukko 3. Yksi- ja kaksikielisten koehenkilöiden sanamäärät, Guiraudin indeksit ja edistyneet Guiraudin indeksit sekä keskiarvot (ka) ja keskihajonnat (kh)

	Sanoja yhteensä	Eri sanoja	Guiraudin indeksi	Edistynyt Guiraudin indeksi
Yksikieliset	129–899 (ka 378.3, kh 149.1)	49–175 (ka 110.0, kh 30.7)	4.21–7.96 (ka 5.70, kh .99)	4.28–8.31 (ka 6.33, kh .94)
Kaksikieliset	96–494 (ka 319.5, kh 97.0)	42–150 (ka 96.1, kh 26.5)	4.29–8.43 (ka 5.38, kh .98)	3.97–8.58 (ka 5.88, kh 1.06)

3.2 Sanojen yleisyys

Sanojen yleisyyttä arvioitiin tarkastelemalla aineiston sanojen esiintymismääriä Finnish Internet Parsebank-korpuksessa (2021). Tämän jälkeen jokaisen tutkittavan sanat jaettiin neljään ryhmään (<10; ≥10 ja <20; ≥20 ja <100 sekä ≥100 per miljoona sanaa) sanojen esiintymistiheyden perusteella. Sanojen yleisyyden osalta tarkasteltiin, miten lasten käyttämät sanat jakautuivat eri yleisyysryhmiin.

Ryhmien välisiä eroja sanojen jakautumisessa eri yleisyysryhmiin tarkasteltiin käyttämällä kaksisuuntaista riippumattomien otosten t-testiä. Ryhmäkokojen pienuuden takia varmistettiin, että muuttujat olivat normaalisti jakautuneita ($p > .05$). T-testin perusteella yksikielisten ja kaksikielisten lasten ryhmien välillä oli tilastollisesti merkitsevä ero <10 -ryhmässä ($t(42) = 2.384, p = .024, d = .74$) ja ≥ 100 -ryhmässä ($t(42) = -2.795, p = .009, d = -.86$), ja efektikoko oli molemmissa suuri. Yksikielisten käyttämistä sanoista 14,17 prosenttia kuului harvinaisimpien sanojen ryhmään, eli niitä esiintyi korpuksessa alle 10 per miljoona sanaa, kun taas kaksikielillä kyseiseen ryhmään kuului 10,92 prosenttia sanoista. Kaksikielisten käyttämistä sanoista 67,47 prosenttia kuului yleisiin sanoihin, eli sanoihin, joita korpuksessa esiintyi yli 100 per miljoona sanaa, kun taas yksikielillä vastaava lukema oli 61,79 prosenttia. Sen sijaan tilastollisesti merkitsevää eroa ei löytynyt ≥ 10 ja <20 -ryhmässä ($t(42) = .615, p = .542, d = .19$) eikä ≥ 20 ja <100 -ryhmässä ($t(42) = 1.740, p = .089, d = .54$). T-testin tulosten perusteella siis yksikieliset lapset käyttivät tilastollisesti merkitsevästi enemmän harvinaisia sanoja (<10 per miljoona), kun taas kaksikieliset lapset käyttivät tilastollisesti merkitsevästi enemmän yleisiä sanoja (≥ 100 per miljoona). Sanojen jakautuminen yleisyysryhmiin löytyy myös kuviosta 1.



Kuvio 1. Yksi- ja kaksikielisten käyttämien sanojen prosentuaalinen jakautuminen yleisyysryhmiin (per miljoona sanaa) sekä 95 % luottamusvälit.

Mikäli tarkastellaan käytettyjen sanojen jakautumista vain alle ja yli 100 kertaa miljoonassa esiintyviin sanoihin, havaitaan, että yksikieliset lapset käyttivät tilastollisesti merkitsevästi enemmän harvinaisempia (<100 per miljoona) sanoja ($t(42) = 2.793$, $p = .009$), kun taas kaksikieliset käyttivät tilastollisesti merkitsevästi enemmän yleisempiä (≥ 100 per miljoona) sanoja ($t(42) = -2.795$, $p = .009$).

4 Pohdinta

Tutkimuksen tarkoituksena oli siis vertailla yksi- ja kaksikielisten suomalaislasten käyttämän sanaston monipuolisuutta ja yleisyyttä tarinankerrontatehtävässä. Tämän tutkimuksen tulosten perusteella yksi- ja kaksikielisten lasten käyttämän sanaston monipuolisuudessa ei ollut tilastollisesti merkitsevää eroa. Sen sijaan ryhmien väliltä löytyi ero käytetyn sanaston yleisyydessä: yksikieliset lapset käyttivät tilastollisesti merkitsevästi enemmän harvinaisia sanoja kuin kaksikieliset, ja kaksikieliset lapset käyttivät tilastollisesti merkitsevästi enemmän yleisiä sanoja kuin yksikieliset lapset.

4.1 Sanaston eroavaisuudet

4.1.1 Sanaston monipuolisuus

Sanaston monipuolisuuden eroja tarkasteltiin käyttämällä Guiraudin indeksiä, joka lasketaan jakamalla tarinoissa esiintyneiden eri sanojen määrä tarinan kokonaissanamäärän neliöjuurella. Neliöjuurella jakaminen vähentää tarinan pituuden vaikutusta, mikä on monien sanaston monipuolisuuden arviointimenetelmien kohdalla ongelmana, sillä tekstin jatkuessa on todennäköisempää, että sana on jo esiintynyt tekstissä aiemmin (Daller ym., 2003). Tämä oli tarpeen, sillä yksikielisillä lapsilla tarinat olivat keskimäärin noin 60 sanaa pidemmät kuin kaksikielisillä. Ryhmien välisiä mahdollisia eroja tarkasteltiin myös edistyneellä Guiraudin indeksillä, joka Dallerin ym. (2003) mukaan soveltuu sanaston monipuolisuuden arviointiin tavallisia menetelmiä paremmin. Edistyneessä Guiraudin indeksissä aineistosta poistettiin sanat, joiden yleisyys oli yli 1000 per miljoona sanaa.

Suurempi Guiraudin indeksin arvo tarkoittaa sitä, että tekstissä on enemmän leksikaalista variaatiota (Nation & Meara, 2010). Yksikielisillä lapsilla oli sekä Guiraudin indeksin että edistyneen Guiraudin indeksin perusteella keskimäärin hieman kaksikielisiä enemmän leksikaalista variaatiota tarinoissaan, mutta ero ei kummallakaan menetelmällä laskettuna ollut tilastollisesti merkitsevä. Edistyneessä Guiraudin indeksissä efektikoko oli kuitenkin kohtalainen.

Yleisten sanojen poistaminen aineistosta ei siis vaikuttanut tuloksiin merkittävästi. Kunnari ym. (2016) eivät löytäneet eroa myöskään tarinoiden rakenteen monimutkaisuudessa suomalaisten yksi- ja kaksikielisten lasten tai kaksikielisten lasten eri kielten välillä. De Houwerin ym. (2014) mukaan erojen löytyessä yksi- ja kaksikielisten välillä tulisikin ennen johtopäätöksiä tarkistaa, ettei kaksikielisyyden sijaan eroja aiheuta jokin muu tekijä, kuten esimerkiksi lapsen ikä tai perheen sosioekonominen asema.

Muun muassa Hoffin ym. (2012) ja Pearsonin ym. (1993) mukaan yksi- ja kaksikielisten lasten sanavarastojen koossa ei ole eroa, jos kaksikielisiltä huomioidaan heidän molemmat kielensä, mutta yksittäisen kielen sanavarasto saattaa olla pienempi kuin yksikielisen lapsen sanavarasto. Tämän tutkimuksen kaksikieliset lapset olivat samanaikaisesti kaksikielisiä, ja suurin osa lapsista oli tasapainoisesti kaksikielisiä, jolloin voidaan olettaa heidän kielitaitonsa olevan molemmilla kielillä samalla tai lähes samalla tasolla. Ruotsin kielen asema Suomessa on heikentynyt (Council of Europe, 2019), ja ruotsinkielisten määrä vähentynyt (Suomen virallinen tilasto, 2021b), joten voidaan olettaa, että kaksikieliset joutuvat käyttämään paljon myös suomea. Dallerin ym. (2003) mukaan ympäristön kieli on usein myös yksilön vahvempi kieli, sillä sitä kuulee useammassa yhteyksissä. Tämän tutkimuksen kaksikielisten ryhmän lapset asuivat Manner-Suomessa ja heidän osaltaan tarkasteltiin sanastoa nimenomaan suomen kielen osalta, joten tulokset saattavat liittyä suomen hyvään asemaan ja ihmisten hyvään suomen kielen taitoon myös kaksikielisissä kunnissa.

Myös arviointitapa ja tarinoiden pituus saattavat vaikuttaa tuloksiin. Guiraudin indeksin perusteella yksikielisillä lapsilla oli lukumäärällisesti hieman enemmän leksikaalista variaatiota tekstissään. Lapset kuitenkin kertoivat tarinan samoista kahdesta kuvakirjasta, joten lapsen mahdollisuus vaikuttaa tarinan kulkuun oli pieni, ja voi olla, että spontaanissa kerronnassa eroja olisi saattanut löytyä. Lisäksi, vaikka lapset kertoivat tarinat samoista kuvakirjoista, lyhimmän ja pisimmän tarinan ero oli yli 800 sanaa. Vaikka Guiraudin indeksissä neliöjuuren ottaminen kokonaissanamäärästä vähentääkin tekstin pituuden merkitystä, esimerkiksi Hess ym. (1989) kritisoi, ettei menetelmä poista tekstin pituuden vaikutusta kokonaan.

4.1.2 Sanaston yleisyys

Sanaston yleisyyttä on usein tutkittu tarkastelemalla sanaston esiintymistä erilaisissa korpukissa tai kyseisen kielen yleisimpien sanojen listoissa (Gernsbacher, 1984; Tanaka-Ishii & Terada, 2011). Suomen kielelle ei ole muodostettu luotettavia ja ajantasaisia yleisimpien sanojen

listoja, joten tässä tutkimuksessa sanastoa verrattiin korpukseen. Sanojen frekvenssit korpuksessa vaihtelivat nollan ja useiden miljoonien välillä, joten analysoinnin helpottamiseksi laskettiin sanan yleisyys miljoonassa sanassa ja sen jälkeen sanat jaettiin neljään yleisyysryhmään. Yleisyyden osalta arvioitiin siis, jakautuivatko yksi- ja kaksikielisten käyttämät sanat eri yleisyysryhmiin samalla vai eri tavalla.

Tulosten perusteella yksikieliset lapset käyttivät kaksikielisiä lapsia enemmän harvinaisia sanoja, kun taas kaksikieliset lapset käyttivät yksikielisiä lapsia enemmän yleisiä sanoja. Johnsonin ym. (2016) tutkimuksen mukaan tieto yleisten sanojen tuottamisesta oli yhteydessä parempaan kirjoitussuoritukseen toisella kielellä. Kuitenkin varsinainen yleisten sanojen käyttö oli negatiivisessa yhteydessä toisen kielen kirjoitussuoritukseen, eli suurempi määrä harvinaisia sanoja tekstissä oli yhteydessä parempaan suoriutumiseen kirjoitustehtävässä. Dallerin ym. (2003) mukaan ihmiset, joilla on suurempi sanavarasto ja parempi kielen taitotaso, käyttävät enemmän harvinaisia sanoja kuin ihmiset, joilla kielen taitotaso on heikompi tai sanavaraston koko pienempi. Tämän tutkimuksen perusteella yksi- ja kaksikielisten lasten välillä ei ole merkittävää eroa heidän käyttämänsä sanaston monipuolisuuden suhteen, mutta saattaa kuitenkin olla, että kaksikielisten taitotaso tai sanavaraston koko eivät silti ole aivan samalla tasolla yksikielisten lasten kanssa. Kertomuksen lyhyys ja strukturoitu rakenne voivat vaikuttaa siihen, ettei eroja monipuolisuuden osalta löytynyt. Sama asia voidaan kuitenkin ilmaista eri tarkkuudella, esimerkiksi ”se söi sen” tai ”mies söi leivän”, joten voi olla, että kaksikieliset lapset käyttivät yksinkertaisempaa ja samalla yleisempää sanastoa.

Thordardottirin (2011) tutkimuksessa molemmille kielille saman verran altistuneet suoriutuivat yksikielisten veroisesti etenkin vastaanottavan sanaston osalta, mutta tuottavan sanaston osalta samaan suoriutumiseen tarvittiin kaksikielisiltä hieman enemmän altistumista kyseiselle kielelle. Esimerkiksi kielten tasapainoinen asema ja kielten läheisyys saattavat Thordardottirin (2011) mukaan vaikuttaa tuloksiin. Hoffin ym. (2012) ja Keuleersin ym. (2015) mukaan kaksikielisen lapsen kuuleman kielellisen syötteen jakautuminen kahdelle kielelle saattaa vaikuttaa siihen, että kaksikielisillä lapsilla voi olla yksikielisiä lapsia pienempi sanavarasto yksittäisellä kielellä. Todennäköisesti sanavarastosta tällöin puuttuvat nimenomaan harvinaisimmat sanat, sillä yleiset sanat ovat pääosin harvinaisia sanoja tutumpia (Vermeer, 2000). Ruotsinkielisten osuus väestöstä on nykyään vain reilut 5 prosenttia, ja ruotsia äidinkielenään puhuvia suomalaisia asuu laajasti ympäri Suomea (Kuntaliitto, 2017; Suomen virallinen tilasto,

2021b). Suomessa kaksikielisten suomen kielen taidot ovat siis oletettavasti hyvällä tasolla ja suomella on hyvä asema myös kaksikielisissä kunnissa, mutta kaksikieliset lapset eivät välttämättä silti saa yhtä paljon kielellistä syötettä suomeksi kuin yksikieliset lapset, sillä syöte jakautuu suomen ja ruotsin kielten välille.

Monet menetelmät eivät toimi kaksikielisten lasten arvioinnissa, sillä ne on usein normitettu yksikielisille ja tietynmaalaisille lapsille, ja jotka ovat kokeneet samanlaisia asioita elämässään (Laing & Kamhi, 2003). Tässä tutkimuksessa tarkasteltiin kuitenkin Suomessa asuvia yksi- ja kaksikielisiä lapsia, joten kulttuuristen erojen voidaan ajatella olevan pieniä.

4.2 Luotettavuus

Tutkimuksen yhdeksi haasteeksi nousi koehenkilöiden löytäminen. Otokokojen pienuus onkin yleinen ongelma yksi- ja kaksikielisiä lapsia vertailevissa tutkimuksissa (esim. Kutsuki, 2021), ja myös tässä tutkimuksessa suuremmilla otoskoilla olisi saatu paremmin yleistettäviä tuloksia. Koulujen haluttomuuteen osallistua vaikutti osaltaan varmasti se, että kontaktien perusteella koulut olivat jo valmiiksi kuormittuneita koronapandemian ja siitä seuranneen etäopetuksen takia. Lisäksi hankaluutta lisäsi se, että tutkittavia etsittiin koulujen kautta, joten viestin saaneen rehtorin lisäksi myös vanhempien piti innostua osallistumaan lastensa kanssa.

Tutkimuksen koehenkilöiden hakuun käytetty tapa saattoikin vaikuttaa siten, että tutkimukseen mahdollisesti valikoitui keskimäärin kielellisiltä taidoiltaan parempia lapsia. Noin kolmasosa tutkimuksen yksikielisistä koehenkilöistä kerättiin tutkijoiden lähipiiristä laittamalla viesti tutkimuksesta tutkijoiden omille Facebook-sivuille, joten osa lapsista oli jommallekummalle tutkijalle tuttuja. Lapsia ei kuitenkaan testannut lapselle tuttu tutkija, sillä se olisi saattanut vaikuttaa lapsen käytökseen ja hänen kertomaansa tarinaan. Kaksikielisten osalta suomea ja ruotsia puhuvien kaksikielisten käyttäminen koehenkilöinä oli tutkimuksen toteutuksen ja koehenkilöiden löytämisen kannalta hyödyllistä, sillä tarpeeksi suuren otoksen löytäminen muista kaksikielisten ryhmistä olisi voinut olla vielä haastavampaa.

Aineistonkeruu tapahtui etäyhteydellä, joten se vaati kodilta laitteen ja enemmän aikaa verrattuna siihen, jos aineistonkeruu olisi voitu suorittaa kouluissa. Infokirjeessä vanhempia ohjeistettiin, että lapsi osallistuisi tutkimuksen tekoon tietokoneella, mutta kaksi lasta osallistui kännykän avulla. Kännykän ruudulla tarina näkyy pienempänä, mikä saattaa hankaloittaa yksityiskohtien huomaamista. Muuten ainakaan yksikielisten kohdalla ei esiintynyt

aineistonkeruuta merkittävästi haitanneita teknisiä ongelmia. Aineiston kerääminen etäyhteydellä sai aikaan sen, että olosuhteet osallistumisen aikana olivat lapsilla erilaiset, sillä osa oli yksin rauhallisessa paikassa ja osalla keskittymiseen saattoi vaikuttaa esimerkiksi sisarusten läsnäolo. Osa lapsista oli tilanteessa ilman vanhempaa, ja osa halusi vanhemman olevan läsnä. Toisaalta etäyhteyden avulla lapsi sai olla omassa kotiympäristössään ja tutkija oli ”kauempana” lapsesta, mikä joidenkin lasten kohdalla saattoi vähentää jännitystä osallistumista kohtaan. Pääosin lapset kertoivat tarinan mielellään ja toimivat hyvässä yhteistyössä tutkijoiden kanssa.

Tämän tutkimuksen luotettavuutta parantaa se, että kaikki aineistonkeruun vaiheet, lähetetyt viestit, materiaalit sekä analyysit on kirjattu ylös. Lisäksi aineistoa oli keräämässä ja litteroimassa kaksi tutkijaa, mikä antoi tutkijoille mahdollisuuden esimerkiksi keskustella epäselvistä asioista yhdessä. Kaksikielisten ja yksikielisten lasten aineistoa keräämässä olivat eri tutkijat, mikä saattaa näkyä erilaisina tapoina kerätä tai käsitellä aineistoa. Toisaalta aineiston analysointi aloitettiin molempien ryhmien osalta litteroiduista, mutta muuten käsittelemättömistä aineistoista, joten kaikki analyysit on toteutettu molemmille ryhmille samalla tavalla.

Luotettavuutta parantaa myös se, että yksikielisten lasten ryhmä sovitettiin iältään ja sukupuoleltaan mahdollisimman samankaltaiseksi kuin ensin kerätty kaksikielisten lasten ryhmä. Yksi- ja kaksikielisten ryhmät eivät eronneet toisistaan tilastollisesti merkitsevästi riskipisteiden, iän tai sukupuolen osalta. Myös se, kumman sadun lapsi sai, määräytyi aiemman aineiston perusteella, kuitenkin niin, että puolet lapsista sai toisen ja puolet toisen tarinan. Kahden sadun käyttäminen mahdollisti yksikielisten lasten kohdalla myös sisarusten osallistumisen samasta paikasta ilman, että toisen lapsen kertoma tarina vaikutti merkittävästi toisen lapsen tarinaan.

Monet kaksikieliset ryhmät voivat olla ryhminä hyvin heterogeenisiä, mutta väliin tulevia muuttujia saattaa olla vähemmän suomea puhuvien yksikielisten sekä suomea ja ruotsia puhuvien kaksikielisten välillä, mikä voi parantaa tutkimuksen luotettavuutta. Soveri ym. (2011) esimerkiksi eivät löytäneet yksikielisten suomea puhuvien ja kaksikielisten suomea ja ruotsia puhuvien aikuisten väliltä eroa sosioekonomisen aseman osalta, ja myös tässä tutkimuksessa suomea sekä suomea ja ruotsia puhuvien perheiden voidaan ajatella olevan keskenään melko samanlaisia esimerkiksi sosioekonomiselta asemaltaan. Tässä tutkimuksessa löydettiin eroja käytetyn sanaston yleisyydessä yksikielisten ja suomen kielen taidoiltaan taitavien

kaksikielisten väliltä ryhmillä, jotka tulevat hyvin samanlaisista taustoista. Voidaankin olettaa, että erot sanavalinnoissa olisivat suuremmat, jos ryhmät eivät olisi niin samanlaisista taustoista, kuten voisi olla esimerkiksi maahanmuuttajien kohdalla. Tämä tulisi huomioida, kun vertaillaan yksikielisten ja kaksikielisten suoriutumista esimerkiksi puheen sujuvuuden tutkimuksissa.

Vaikka tulokset ovatkin pienten otoskokojen takia suuntaa antavia, antavat ne pohjaa jatkotutkimuksille. Tarvetta yksi- ja kaksikielisten lasten sanaston käyttöön liittyville tutkimuksille on, sillä kaksi- ja monikielisten lasten määrä on kasvanut globalisoitumisen myötä. Jatkossa suomea ja ruotsia äidinkielenään puhuvien lasten lisäksi tutkimusta olisikin hyvä laajentaa koskemaan muita kielipareja, jolloin voitaisiin tarkastella kielten vaikutusta tuloksiin. Suomea ja ruotsia puhuvien kaksikielisten kohdalla olisi hyvä tarkastella myös lasten ruotsin kielen taitoja, ja ottaa mukaan myös yksikielisten ruotsia puhuvien lasten ryhmä. Tutkimusta voisi laajentaa myös tarkastelemalla esimerkiksi pienempiä lapsia tai aikuisia.

Tämä tutkimus on osa tutkimusprojektia, jossa tarkastellaan puheen sujumattomuuksia yksi- ja kaksikielisillä lapsilla. Samasta aineistosta tehdyssä Riihiojan (2021) tutkimuksessa ei löytynyt eroja yksi- ja kaksikielisten lasten puheen sujumattomuuksien määrässä tai laadussa, mikä eroaa aiemmista tutkimuksista, joissa on havaittu, että kaksikielisillä on usein enemmän puheen sujumattomuuksia kuin yksikielisillä. Puheen sujumattomuuksia ja etenkin niiden eroja tai samankaltaisuuksia yksi- ja kaksikielisillä olisikin hyvä tarkastella myös niin, että huomioidaan sujumattomuuksien lisäksi käytetyn sanaston yleisyys ja monipuolisuus, sillä ihminen saattaa välttää sujumattomuuksia aiheuttavia sanoja ja korvata ne helpommilla sanoilla (Lowe ym., 2017).

4.3 Lopuksi

Kaksi- ja monikielisten lasten määrä Suomessa ja puheterapiavastaanotoilla on viime vuosina kasvanut. Lisäksi suomalaisista noin 5 prosenttia puhuu äidinkielenään ruotsia. Kuitenkaan tutkimusta suomea ja ruotsia äidinkielenään puhuvien kaksikielisten ja suomea äidinkielenään puhuvien yksikielisten lasten sanaston eroista ei juuri ollut ennen tätä tutkimusta. Tutkimustuloksia voitaisiin hyödyntää esimerkiksi puheterapia-arvion tai kuntoutuksen suunnittelun tukena, mutta myös muun muassa puheen sujumattomuuksia tutkittaessa, sillä pelkän sujumattomuuksien määrän lisäksi olisi tärkeää huomioida myös käytetyn sanaston yleisyys ja monipuolisuus.

Tässä tutkimuksessa tarkasteltiin yksi- ja kaksikielisten lasten käyttämän sanaston monipuolisuuden ja yleisyyden eroja suomenkielisessä tarinankerrontatehtävässä. Yksikielisten suomea ja kaksikielisten suomea ja ruotsia äidinkielenään puhuvien lasten väliltä ei löytynyt tilastollisesti merkitsevää eroa heidän käyttämänsä sanaston monipuolisuuden osalta, mikä voi johtua esimerkiksi suomen kielen hyvästä taitotasosta myös kaksikielisissä perheissä. Toisaalta kuitenkin huomattiin, että yksikieliset käyttivät harvinaisimpia sanoja kaksikielisiä enemmän, kun taas kaksikieliset lapset käyttivät yleisimpiä sanoja kaksikielisiä enemmän. Tähän saattaa liittyä kaksikielisten lasten kuuleman kielellisen syötteen jakautuminen yhden kielen sijasta kahdelle kielelle. Ryhmien väliltä löytyi eroa käytettyjen sanojen yleisyydessä, vaikka ryhmät tulevat samankaltaisista taustoista, joten eri kaksikielisten ryhmällä tutkittaessa erot saattaisivat olla suurempia. Tulokset ovat kuitenkin suuntaa antavia, joten lisää tutkimusta aiheesta tarvitaan.

Lähteet

- Berken, J. A., Gracco, V. L. & Klein, D. (2016). Early bilingualism, language attainment, and brain development. *Neuropsychologia*, 98, 220–227. <https://doi.org/10.1016/j.neuropsychologia.2016.08.031>
- Berman, R. & Nir, B. (2010). The lexicon in writing–speech-differentiation: Developmental perspectives. *Written Language & Literacy*, 13(2), 183–205. <https://doi.org/10.1075/wll.13.2.01ber>
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511605963>
- Boerma, T. & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible?. *Journal of Communication Disorders*, 66, 65–76. <https://doi.org/10.1016/j.jcomdis.2017.04.001>
- Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F. & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *International Journal of Language & Communication Disorders*, 51(6), 626–638. <https://doi.org/10.1111/1460-6984.12234>
- Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, 18(1), 1–21. <https://doi.org/10.1191/0265659002ct224oa>
- Brandt-Häggblom, A. (2021). Kartläggning av talflytsstörningar hos svensk- och finsktalande barn i Finland. Pro gradu -tutkielma. Åbo Akademi.
- Brybaert, M., Mander, P. & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, 27(1), 45–50. <https://doi.org/10.1177/0963721417727521>
- Brybaert, M., Stevens, M., Mander, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, 7, 1116. <https://doi.org/10.3389/fpsyg.2016.01116>

- Cook, V. (2003). Linguistics and second language acquisition: One person with two languages. Teoksessa Aronoff, M. & Rees-Miller, J. (toim.). *The Handbook of Linguistics*. Blackwell, 488–511. <https://doi.org/10.1002/9780470756409>
- COST action IS0804. (2011). Questionnaire for Parents of Bilingual Children (PaBiQ). Viitattu 22.1.2022. <https://www.bi-sli.org/pabiq>
- Council of Europe. (2019). Fifth opinion on Finland. Advisory Committee on the Framework Convention for the Protection of National Minorities (ACFC). Viitattu 2.4.2022. <https://rm.coe.int/5th-op-finland-en/16809839e4>
- Covington, M. A. & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100. <https://doi.org/10.1080/09296171003643098>
- Daller, H., Van Hout, R. & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24(2), 197–222. <https://doi.org/10.1093/applin/24.2.197>
- De Houwer, A., Bornstein, M. H. & Putnick, D. L. (2014). A Bilingual-Monolingual Comparison of Young Children's Vocabulary Size: Evidence from Comprehension and Production. *Applied Psycholinguistics*, 35(6), 1189–1211. <https://doi.org/10.1017/S0142716412000744>
- Eggers, K., Van Eerdenbrugh, S. & Byrd, C. T. (2019). Speech disfluencies in bilingual Yiddish-Dutch speaking children. *Clinical Linguistics & Phonetics*, 34(6), 576–592. <https://doi.org/10.1080/02699206.2019.1678670>
- Fernald, A., Marchman, V. A. & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2), 234–248. <https://doi.org/10.1111/desc.12019>
- Finnish Internet Parsebank. (2021). Internet_Parsebank_4B-korpus. http://epsilon-it.utu.fi/nse/run.cgi/first_form

- Gass, S. M. & Selinker, L. (1992). Introduction. Teoksessa Gass, S. M. & Selinker, L. (toim.). *Language transfer in language learning: Revised edition*. John Benjamins Publishing Company, 1–17. <https://doi.org/10.1075/lald.5>
- Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, 113, 256–281. <https://doi.org/10.1037//0096-3445.113.2.256>
- Hayes, D. P. (1988). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language*, 27(5), 572–585. [https://doi.org/10.1016/0749-596X\(88\)90027-7](https://doi.org/10.1016/0749-596X(88)90027-7)
- Heilmann, J. J., Rojas, R., Iglesias, A. & Miller, J. F. (2016). Clinical impact of wordless picture storybooks on bilingual narrative language production: A comparison of the ‘Frog’ stories. *International Journal of Language & Communication Disorders*, 51(3), 339–345. <https://doi.org/10.1111/1460-6984.12201>
- Hess, C. W., Haug, H. T. & Landry, R. G. (1989). The reliability of type-token ratios for the oral language of school age children. *Journal of Speech, Language, and Hearing Research*, 32(3), 536–540. <https://doi.org/10.1044/jshr.3203.536>
- Hoff, E. (2009). *Language development* (4. painos). Wadsworth Cengage Learning.
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M. & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1–27. <https://doi.org/10.1017/S0305000910000759>
- Hudson, J. A. & Shapiro, L. R. (1991). From knowing to telling: The development of children's scripts, stories, and personal narratives. Teoksessa McCabe, A. & Peterson, C. (toim.). *Developing Narrative Structure*. Lawrence Erlbaum Associates, Inc, 89–136.
- Häkkinen, K. (1997). Kuinka ruotsin kieli on vaikuttanut suomeen?. *Sananjalka*, 39(1), 31–54.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund University, Department of Linguistics and Phonetics, Working papers*, 53, 61–79.

- Johnson, M. D., Acevedo, A., & Mercado, L. (2016). Vocabulary knowledge and vocabulary use in second language writing. *TESOL Journal*, 7(3), 700–715. <https://doi.org/10.1002/tesj.238>
- Justice, L., Bowles, R., Kaderavek, J., Ukrainetz, T., Eisenberg, S. & Gillam, R. (2006). The Index of Narrative Microstructure: A Clinical Tool for Analyzing School-Age Children’s Narrative Performances. *American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association*, 15, 177–191. [https://doi.org/10.1044/1058-0360\(2006/017\)](https://doi.org/10.1044/1058-0360(2006/017))
- Kaiser, E. & Rautakoski, P. (2013). Monikielisten lasten puheterapeuttinen tutkiminen Suomessa. *Puheterapeutti*, 1, 23–26.
- Keller, H. H. (1991). Word frequency and pedagogical value: contrasting textbook vocabulary lists with word frequency counts. *Slavic and East European Journal*, 35(2), 228–244. <https://doi.org/10.2307/308316>
- Keuleers, E., Stevens, M., Mandera, 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), 1665–1692. <https://doi.org/10.1080/17470218.2015.1022560>
- Kunnari, S., Välimaa, T. & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Applied Psycholinguistics*, 37(1), 123–144. <https://doi.org/10.1017/S0142716415000442>
- Kuntaliitto. (2017). Ruotsin- ja kaksikieliset kunnat: taustatietoa 2008–2017. Viitattu 27.1.2022. https://www.kuntaliitto.fi/sites/default/files/media/file/2017-02-ruotsin-ja-kaksikieliset-kunnat_0.pdf
- Kutsuki A. (2021). Do Bilinguals Acquire Similar Words to Monolinguals? An Examination of Word Acquisition and the Similarity Effect in Japanese-English Bilinguals' Vocabularies. *European Journal of Investigation in Health, Psychology and Education*, 11(1), 168–182. <https://doi.org/10.3390/ejihpe11010014>

- Laing, S. P. & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech & Hearing Services in Schools*, 34(1), 44–55. [https://doi.org/10.1044/0161-1461\(2003/005\)](https://doi.org/10.1044/0161-1461(2003/005))
- Launonen, K. (2007). Monikielisyyden haasteet kielihäiriöiden tunnistamisessa ja kuntoutuksessa. *Virittäjä*, 111(2), 240–243.
- Lehtonen, M. (2021). Introduction to Bilingualism: The Multilingual Brain. Luento Turun yliopistossa 18.3.2021.
- Loukusa, S. (2007). 3–9-vuotiaiden lasten suoriutuminen Bostonin Nimentätestistä. *Puhe ja kieli*, 27, 141–149.
- Lowe, R., Helgadottir, F., Menzies, R., Heard, R., O'Brian, S., Packman, A. & Onslow, M. (2017). Safety behaviors and stuttering. *Journal of Speech, Language, and Hearing Research*, 60(5), 1–8. https://doi.org/10.1044/2016_JSLHR-S-16-0055
- Luotolahti, J., Kanerva, J., Laippala, V., Pysälo, S. & Ginter, F. (2015). Towards universal web parsebanks. Teoksessa Nivre, J. & Hajičová, E. (toim.). *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 211–220.
- Mackey, W. F. (2000). The description of bilingualism. Teoksessa Wei, L. (toim.). *The bilingualism Reader*, 3, 25. Routledge, 22–50. <https://doi.org/10.4324/9780203461341>
- Mayer, M. (1973). *Frog on his own*. Dial press.
- Mayer, M. (1974). *Frog goes to restaurant*. Dial press.
- Mäkinen, L. & Kunnari, S. (2009). Lasten kerrontataitojen arvioiminen. *Puhe ja kieli*, 29, 103–120.
- Nagy, W. E. & Scott, J. A. (2000). Vocabulary Processes. Teoksessa Kamil, M. L., Mosenthal, P. B., Pearson, P. D. & Barr, R. (toim.). *Handbook of Reading Research*, 3, 269–284. <https://doi.org/10.4324/9781410605023>
- Nation, P. & Meara, P. (2010). Vocabulary. Teoksessa Schmitt, N. (toim.). *An Introduction to Applied Linguistics*. Routledge, 34–52. <https://doi.org/10.4324/9780203783733>

- Opetushallitus. (2014). Perusopetuksen opetussuunnitelman perusteet 2014. Viitattu 6.5.2021. <https://eperusteet.opintopolku.fi/#/fi/perusopetus/419550/tiedot>
- Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, 28(3), 399–410. <https://doi.org/10.1017/S014271640707021X>
- Pearson, B. Z. (2008). *Raising a Bilingual Child*. Living Language.
- Pearson, B. Z., Fernández, S. C. & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43(1), 93–120. <https://doi.org/10.1111/j.1467-1770.1993.tb00174.x>
- Perustuslaki 11.6.1999/731, 7 §. Viitattu 27.1.2022. <https://www.finlex.fi/fi/laki/ajantasa/1999/19990731#L2P17>
- Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732942>
- Reuter, M. (2012). Swedish as a pluricentric language. Teoksessa Clyne, M. (toim.). *Pluricentric Languages: Differing Norms in Different Nations*. De Gruyter Mouton, 101–116. <https://doi.org/10.1515/9783110888140.101>
- Riihioja, J. (2021). Puheen sujumattomuudet suomen kielessä kaksi- ja yksikielisillä lapsilla. Pro gradu -tutkielma. Turun yliopisto.
- Romaine, S. (2017). Multilingualism. Teoksessa Aronoff, M. & Rees-Miller, J. (toim.). *The Handbook of Linguistics*, 460. John Wiley & Sons, Inc., 541–556. <https://doi.org/10.1002/9781119072256.ch26>
- Saukkonen. (1979). *Suomen kielen taajuussanasto = A frequency dictionary of Finnish*. WSOY.
- Schneider, P., Hayward, D. & Vis Dubé, R. (2006). Storytelling from pictures using The Edmon- ton Narrative Norms Instrument. *Journal of Speech-Language Pathology and Audiology*, 30, 224–238.
- Seymour, P. H., Aro, M., Erskine, J. M. & Collaboration with COST Action A8 Network. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174. <https://doi.org/10.1348/000712603321661859>

- Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S. & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*, 34(1), 3–23. <https://doi.org/10.1177/0142723713510997>
- Smolander, S., Kunnari, S. & Laasonen, M. (2016). Näkökulmia kielellisten taitojen arviointiin ja kielellisen vaikeuden tunnistamiseen monikielisellä lapsella. *Puhe ja kieli*, 36(1), 57–75.
- Soveri, A., Laine, M., Hämäläinen, H. & Hugdahl, k. (2011). Bilingual advantage in attentional control: Evidence from the forced-attention dichotic listening paradigm. *Bilingualism*, 14(3), 371–378. <https://doi.org/10.1017/S1366728910000118>
- Suomen virallinen tilasto (SVT). (2021a). Vieraskieliset. Tilastokeskus. Viitattu 22.3.2022. <https://www.stat.fi/tup/maahanmuutto/maahanmuuttajat-vaestossa/vieraskieliset.html>
- Suomen virallinen tilasto (SVT). (2021b). Väestörakenne. Tilastokeskus. Viitattu 26.3.2022. <http://www.stat.fi/til/vaerak/index.html>
- Suvanto, A. & Mäkinen, L. (2011). Lasten kerrontataitojen kehitys. Teoksessa Loukusa, S. & Paavola, L. (toim.). *Lapset kieltä käyttämässä: Pragmaattisten taitojen kehitys ja sen häiriöt*. PS-kustannus, 63–82.
- Tammenmaa, C. (toim.). (2020). Oikeusministeriö. Usean kielen merkitseminen väestötietojärjestelmään -selvitys. Oikeusministeriön julkaisuja, selvityksiä ja ohjeita 2020:8. Viitattu 2.4.2022. https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/162056/OM_2020_8.pdf
- Tanaka-Ishii, K. & Terada, H. (2011). Word familiarity and frequency. *Studia Linguistica*, 65, 96–116. <https://doi.org/10.1111/j.1467-9582.2010.01176.x>
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4), 426–445. <https://doi.org/10.1177/1367006911403202>

- Tieteen termipankki 15.12.2021: Kirjallisuudentutkimus: kertomus. Viitattu 15.12.2021.
<https://tieteentermipankki.fi/wiki/Kirjallisuudentutkimus:kertomus>
- Tuller, L. (2015). Clinical Use of Parental Questionnaires in Multilingual Contexts. Teoksessa Armon-Lotem, S., de Jong, J. & Meir, N. (toim.). *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, 13. Multilingual Matters, 301–330.
<https://doi.org/10.21832/9781783093137>
- Valo, M. (2021). Puheen tyypillinen sujumattomuus suomea ja ruotsia puhuvilla kaksikielisillä lapsilla. Pro gradu -tutkielma. Turun yliopisto.
- Van Hout, R. & Vermeer, A. (2007). Comparing measures of lexical richness. Teoksessa Daller, H., Milton, J. & Treffers-Daller, J. (toim.). *Modelling and Assessing Vocabulary Knowledge*. Cambridge University Press, 93–115.
<https://doi.org/10.1017/CBO9780511667268.008>
- Vataja, P., Lerkkanen, M. K., Aro, M., Westerholm, J., Risberg, A. K. & Salmi, P. (2021). The predictors of literacy skills among monolingual and bilingual Finnish–Swedish children during first grade. *Scandinavian Journal of Educational Research*, 1–17.
<https://doi.org/10.1080/00313831.2021.1942191>
- Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17(1), 65–83. <https://doi.org/10.1177/026553220001700103>
- Wei, L. (2000). Dimensions of bilingualism. Teoksessa Wei, L. (toim.). *The Bilingualism Reader*, 3, 25. Routledge, 2–21. <https://doi.org/10.4324/9780203461341>
- Westman, M., Korkman, M., Mickos, A. & Byring, R. (2008). Language profiles of monolingual and bilingual Finnish preschool children at risk for language impairment. *International Journal of Language & Communication Disorders*, 43(6), 699–711.
<https://doi.org/10.1080/13682820701839200>

Liitteet

Liite 1. Vanhemman suostumuslomake

Lasten kielellisiä taitoja tutkiva tutkimus

Vanhemman suostumuslomake

Tällä lomakkeella pyydämme sinua antamaan lapsellesi luvan osallistua Zoomin välityksellä toteutettavaan tutkimukseen, jossa tutkitaan lasten kerrontataitoja. Tutkimustilanne alkaa pienellä keskustelulla lapsen kanssa liittyen tavallisiin päivittäisiin aiheisiin, kuten harrastuksiin. Sen jälkeen lapsi kertoo tarinan kuvakirjaan perustuen. Tutkimus on osa laajempaa tutkimusprojektia, jossa vertaillaan kaksikielisten ja yksikielisten lasten kielellisiä taitoja. Tutkimukseen osallistuminen kestää noin 30 minuuttia.

Kerrotut tarinat kirjoitetaan myöhemmin puhtaaksi tarkempaa analysointia varten. Jotta tarinat voidaan kirjoittaa puhtaaksi, tutkimustilanne äänitetään ja videoidaan. Äänityksissä ei mainita lapsesi nimeä, vaan tiedot kirjataan tunnistenumeron avulla. Valmiista tutkielmistamme ei siis voi tunnistaa yksittäistä lasta. Videotallenteet poistetaan analyysien jälkeen. Äänitallenteet ja puhtaaksi kirjoitetut kertomukset säilytetään Turun yliopiston ylläpitämässä pilvipalvelussa, ja materiaali on käytettävissä vain tämän projektin tutkijoille.

Voitte halutessanne kysyä kysymyksiä tutkimukseen liittyen. Tutkimukseen osallistuminen on vapaaehtoista ja tutkimus voidaan keskeyttää missä vaiheessa tahansa ilman seurauksia. Sinulla ja lapsellasi on myös oikeus pyytää tietojen poistamista tutkimustilanteen jälkeen.

Tämän lisäksi pyydämme sinua vastaamaan pieneen kyselyyn, jossa selvitetään perustietoja lapsesta ja hänen kielellisestä kehityksestään. Kysely lähetetään sinulle sähköisesti, ja sen voi täyttää tietokoneella, tabletilla tai puhelimella. Osallistuminen ja tietojen käsittely noudattavat samoja ehtoja kuin edellä mainittiin. Kyselyn tietoja käytetään lasten vertailuun ryhmätasolla.

Annan lapselleni luvan osallistua tutkimukseen.

Suostun itse osallistumaan tähän tutkimukseen vastaamalla kyselyyn.

Liite 2. Lapsen suostumuslomake

Lapsen suostumuslomake

Voitko osallistua Zoomin kautta tehtävään tutkimukseen, jossa tehtävänäsi on katsoa kuvakirjaa ja kertoa mitä sinulle tulee kirjan kuvista mieleen? Lisäksi kysymme sinulta muutamia kysymyksiä esimerkiksi harrastuksistasi. Voit kysyä mieleen tulevia kysymyksiä milloin tahansa, ja voimme tarvittaessa keskeyttää tutkimuksen, jos et halua jatkaa. Tutkimus kestää noin 30 minuuttia. Voitko siis osallistua tähän tutkimukseen?

Olen tietoinen oikeuksistani ja annan suostumukseni osallistua tutkimukseen.

Liite 3. Kysely vanhemmille

Kysely vanhemmille



Pakolliset kentät on merkitty (*) ja ne pitää täyttää lomakkeen lähettämiseksi.

Hei!

Kiitos, että te ja lapsenne olette lupautuneet osallistumaan tutkimukseemme!

Olemme tosiaan logopedian opiskelijoita, ja teemme pro gradu –tutkielmiamme Turun yliopiston ja Åbo Akademin tutkimuksessa. Tutkimuksessa verrataan 7–9–vuotiaiden yksikielisten suomea puhuvien lasten puheen sujuvuutta ja kielellisiä taitoja kaksikielisten suomea ja ruotsia puhuvien lasten taitoihin. Tutkielmien tarkoituksena on selvittää, miten kaksikielisten ja yksikielisten lasten puheen sujuvuus ja muut kielelliset taidot eroavat toisistaan.

Tällä kyselylomakkeella on tarkoitus kerätä perustietoja tutkimukseen osallistuvasta lapsesta sekä hänen kielellisistä taidoistaan. Kyselyn täyttäminen kestää noin 5–10 minuuttia. Tutkimukseen osallistuminen on vapaaehtoista, ja osallistumisen voi keskeyttää missä vaiheessa tahansa.

Lapset saavat pienen lahjan osallistuttuaan tutkimukseen ja vanhemmilla on mahdollisuus tutustua tutkielmiin, kun ne valmistuvat.

Kiitos, että mahdollistatte tutkimuksemme!

Ystävällisin terveisin,
Julia Riihioja ja Heidi Saarijärvi

Jos teillä on kysyttävää tutkimukseen liittyen, niin meidät tavoittaa sähköpostiosoitteista julia.m.riihioja@utu.fi ja heidi.s.saarijarvi@utu.fi.

Perustietoja

1. Lomakkeen täyttäjän nimi ja suhde lapseen *

*

2. Lapsen nimi *

*

3. Lapsen tunnistenumero *

4. Lapsen syntymäaika (pp/kk/vvvv) *

*

5. Lapsen syntymämaa *

Suomi

Jokin muu, mikä?

6. Jos lapsi ei ole syntynyt Suomessa, milloin olette tulleet Suomeen?

Vuonna

7. Lapsen lähiperhe

Kerro tekstikenttiin pyydetyt tiedot lapsen lähiperheeseen kuuluvista aikuisista.

Aikuinen 1:
suhde lapseen,
ylin suoritettu
koulutusaste,
äidinkieli

Aikuinen 2 (jos
perheeseen
kuuluu toinen
aikuinen): suh-
de lapseen, ylin
suoritettu kou-
lutusaste, äidin-
kieli

Aikuinen 3 (jos perheeseen kuuluu kolmas aikuinen): suhde lapseen, ylin suoritettu koulutusaste, äidinkieli

Aikuinen 4 (jos perheeseen kuuluu neljäs aikuinen): suhde lapseen, ylin suoritettu koulutusaste, äidinkieli

8. Onko tutkimukseen osallistuvalla lapsella sisaruksia? *

Ei ole

Kyllä on, minkä ikäisiä?

Tämän sivun kysymykset liittyvät lapsesi kielen kehitykseen.

9. Mitä kieliä lapsesi puhuu (enemmän kuin yksittäisiä sanoja)? *

Suomi

Ruotsi

Englanti

Muu, mikä?

10. Mitä kieltä lapsesi puhuu parhaiten? *

Suomea

Ruotsia

Englantia

Muuta, mitä?

11. Minkä ikäisenä lapsesi sanoi ensimmäiset sanansa? *

Lapsen ikä (vuodet ja kuukaudet)

*

12. Minkä ikäisenä lapsesi yhdisti ensimmäistä kertaa sanoja lyhyiksi lauseiksi? (esim. "anna pallo") *

Lapsen ikä (vuodet ja kuukaudet)

*

13. Kuinka hyvin lapsesi pystyy mielestäsi ilmaisemaan itseään verrattuna muihin saman ikäisiin lapsiin?

- 0: Ei kovin hyvin tai ei yhtä hyvin kuin ikätoverit
- 1: Vähän heikommin kuin ikätoverit
- 2: Yhtä hyvin kuin ikätoverit
- 3: Todella hyvin tai paremmin kuin ikätoverit

14. Kuinka laaja lapsesi sanavarasto mielestäsi on verrattuna muihin saman ikäisiin lapsiin?

- 0: Selvästi suppeampi kuin ikätovereilla
- 1: Vähän suppeampi kuin ikätovereilla
- 2: Yhtä laaja kuin ikätovereilla
- 3: Todella laaja tai laajempi kuin ikätovereilla

15. Kuinka hyvin lapsesi mielestäsi osaa muodostaa oikeaoppisia lauseita verrattuna muihin saman ikäisiin lapsiin?

- 0: Ei kovin hyvin tai ei yhtä hyvin kuin ikätoverit

- 1: Hieman heikommin kuin ikätoverit
- 2: Yhtä hyvin kuin ikätoverit
- 3: Todella hyvin tai paremmin kuin ikätoverit

16. Oletko huomannut lapsesi puheessa enemmän sujumattomuutta kuin muilla saman ikäisillä lapsilla?

- 0: Selvästi enemmän kuin ikätovereilla
- 1: Vähän enemmän kuin ikätovereilla
- 2: Saman verran kuin ikätovereilla
- 3: Vähemmän kuin ikätovereilla

17. Oletko tyytyväinen lapsesi kykyyn ymmärtää puhetta?

- 0: En lainkaan tyytyväinen
- 1: En kovin tyytyväinen
- 2: Melko tyytyväinen
- 3: Erittäin tyytyväinen

18. Ovatko lapsesi lukemisen ja kirjoittamisen taidot samalla tasolla kuin muilla saman ikäisillä?

- 0: Selvästi alemmalla tasolla kuin ikätoverit
- 1: Lähes samalla tasolla kuin ikätovereilla
- 2: Samalla tasolla kuin ikätovereilla
- 3: Korkeammalla tasolla kuin ikätovereilla

19. Olitko huolissasi lapsesi kielen kehityksestä ennen 3–4 vuoden ikää? *

Esimerkiksi puhumaan oppimisen hitauden, puheen ymmärtämisen haasteiden, puheen epäselvyyden tai änkytyksen takia

En ollut

Kyllä olin, miksi?

Joku muu oli. Kuka ja miksi?

20. Oletko tällä hetkellä huolissasi lapsen kielen kehityksestä? *

Voit kertoa halutessasi lisää.

En ole

Kyllä olen, miksi?

21. Onko lapsesi ollut koskaan puheterapiajaksolla tai puheterapeutin arviossa? *

Ei ole

Kyllä on

22. Miksi ja milloin lapsesi on ollut puheterapiajaksolla tai puheterapeutin arviossa? *

*

23. Onko lapsellasi ollut toistuvia korvatulehduksia tai onko hänellä kuulovaikeuksia? *

Ei ole

On. Kerro lisää:

24. Onko lapsellasi todettu jokin häiriö, sairaus tai vamma? *

Esimerkiksi oppimisvaikeus, ADHD tai autismikirjon häiriö.

Ei ole

Kyllä on, kerro lisää:

Seuraavissa kolmessa kysymyksessä merkitse, mikäli kyseisiä vaikeuksia esiintyy jollain lapsen perheenjäsenistä.

25. Vaikeuksia lukemisessa ja/tai kirjoittamisessa *

- Isällä
- Äidillä
- Sisaruksella
- Ei kenelläkään yllä olevista

26. Ymmärtämisen vaikeuksia, kun muut puhuvat kyseisen henkilön äidinkieltä *

- Isällä
- Äidillä
- Sisaruksella
- Ei kenelläkään yllä olevista

27. Vaikeuksia ilmaista itseään suullisesti (ääntäminen, lauseiden muodostus, sananlöytäminen jne.) äidinkielellään *

- Isällä
- Äidillä
- Sisaruksella
- Ei kenelläkään yllä olevista

28. Muuta kerrottavaa tai kommentoitavaa?

Suuri kiitos osallistumisestanne!

Jos teille tulee myöhemmin kysyttävää tutkimukseen liittyen, niin meidät tavoittaa sähköposti-osoitteista

julia.m.riihioja@utu.fi ja heidi.s.saarijarvi@utu.fi