

Lukutaitoisten kehitysvammaisten aikuisten sanamerkitysten hallinta

Idastiina Valtasalmi
Tampereen yliopisto

Tiivistelmä

Artikkelissa tarkastellaan selkokielisiä tekstejä lukevien kehitysvammaisten aikuisten sanamerkitysten hallintaa. Tutkimuksen tavoitteena on selvittää, miten tarkasti selkolukijat tunnistavat sanamerkityksiä ja millaista osaamistason vaihtelua heillä on. Toisena tavoitteena on selvittää, onko lukutottumuksilla sekä sanojen yleisyydellä, tuttuudella ja konkreettisuustasolla vaikutusta heidän semanttiseen osaamiseensa. Tutkimusmenetelmänä on selektiivinen sana-assosiaatiotesti, johon osallistui 24 kehitysvammaista aikuista ja verrokkiryhmänä 24 korkeakouluopiskelijaa. Testissä osallistujille esitettiin sanaverkkoja, joissa on ärsykesanojen lisäksi kuusi muuta sanaa. Osallistujien tehtävänä oli valita kullekin ärsykesanalle kolme lähimerkityksisintä vastinetta. Osallistujien vastauksia arvioitiin suhteessa toisiinsa ja sanojen sanakirjamerkityksiin. Selkolukijoiden lukutottumuksia ja lukemisen painottumista selkokieleen ja yleiskieleen selvitettiin kyselyllä. Tuloksista ilmenee, että selkolukijoiden sanastonhallinnassa on suuria eroja yksilötasolla. Ryhmätasolla he tunnistivat todennäköisimmin vastinesanoja, joilla on mahdollisimman monta selkokielelle tyypillistä piirrettä. Konkreettisia sanoja osattiin paremmin kuin abstrakteja. Samalla hierarkiatasolla olevat vastinesanat tunnistettiin helpommin kuin geneerisen tason vastineet. Korkea objektiivinen frekvenssi saattoi liittyä parempaan sanojen tunnistamiseen, mutta selkolukijat tunsivat myös harvinaisia sanoja, jotka olivat arkikielestä tuttuja. Selkolukijoilla oli enemmän merkityksen erottelun vaikeuksia kuin verrokeilla, mutta yleisimmin vastinesanoista tunnistettiin ainakin osa. Tämän perusteella selkokieleessä voidaan käyttää synonyymejä, kun ne valitaan harkiten. Lukutottumusten vaikutusta selkolukijoiden semanttiseen osaamiseen on syytä tutkia vielä tarkemmin.

Avainsanat: selkokieli, kehitysvammaiset, sanastonhallinta, sana-assosiaatio

Abstract

This article examines the lexical knowledge of adults with intellectual disabilities who read texts in easy Finnish. The first aim of this article is to investigate how accurately easy language users recognize word meanings and what kind of variation there is in their lexical competence. The second aim is to investigate if reading habits and word features, such as frequency, familiarity and concreteness, have an effect on their word knowledge. The method used was a selective word association task. Participants included 24 adults with intellectual disabilities and 24 university students as a control group. During the study, test items were presented in the form of word webs that included a trigger word and six comparison words. The participants were asked to match each trigger word with three comparison words that were most closely related in meaning. The responses were assessed in relation to one another and dictionary meanings of the words. Additionally, a survey was conducted to obtain information about easy language users' reading habits and preferences regarding reading in easy Finnish and standard Finnish. The results show considerable variation in word knowledge between individual easy language users. As a group, however, they were more likely to choose correct comparison words that had several features of easy language. Concrete words were known better than abstract words. More specific words that were on the same hierarchical level were found more easily than words that shared general features with the target word. High objective frequency may have contributed to correct choices, but easy language users also knew low-frequency words if they were familiar from everyday language. Easy language users had more difficulties in distinguishing word meanings than the control group, but they usually chose at least some of the comparison words correctly. Based on this, synonyms can be used in easy language if they are chosen with care. The effect of reading habits on easy language users' word knowledge requires further studies.

Keywords: easy language, intellectual disabilities, lexical knowledge, word association

1 Johdanto

Selkokieli määritellään Suomessa seuraavasti: ”Selkokieli on suomen kielen muoto, joka on mukautettu sisällöltään, sanastoltaan ja rakenteeltaan yleiskieltä luettavammaksi ja ymmärrettävämmäksi. Se on suunnattu ihmisille, joilla on vaikeuksia lukea tai ymmärtää yleiskieltä.” (Leskelä 2019b: 93.) Selkokielen kohderyhmiin kuuluu Suomessa arviolta 650 000–750 000 ihmistä eli 11–14 % väestöstä (Juusola 2019). Leskelä & Lindholm (2012: 12–31) määrittelevät selkokielen kohderyhmiä selkokielen tarpeen taustalla olevan

syyn perusteella: selkokielen tarve voi johtua synnynnäisistä neurobiologisista syistä, kielitaidon heikentymisestä sairauden tai vamman vuoksi tai väliaikaisesta tuen tarpeesta kielen oppimisvaiheessa, eli selkokielen tarve voi olla joko tilapäistä tai pysyvää. Selkokielestä hyötyvien ihmisten kielellisen osaamisen taso vaihtelee, minkä vuoksi selkokieltä on alettu kehittää perustason lisäksi helpommalla ja vaativammalla vaikeustasolla (Leskelä 2019a). Helppo selkokieli on suunnattu lukijoille, joiden kielitaito on hyvin rajallinen, kun taas vaativa selkokieli sopii lukijoille, joille perustason selkokieli on liian helppoa mutta yleiskieli on liian vaikeaa. Toistaiseksi on kuitenkin määritelty vain perustason selkokielen kriteerit (Selkokeskus 2018), joita käytetään mahdollisimman laajalle lukijajoukolle sopivan yleistason selkokielen arviointiin.¹ Jotta selkokieltä voidaan kehittää edelleen, on tarpeellista tutkia myös selkokielen käyttäjiä. Näin saadaan tietää, millainen kieli on heille luettavaa ja ymmärrettävää.

Tässä artikkelissa tarkastellaan selkotekstejä lukevien kehitysvammaisten aikuisten (jatkossa *selkolukijoiden*) sanamerkitysten hallintaa. Tarkastelu oletettavasti avaa näkymiä heidän kielitaitoonsa laajemminkin, sillä sanastonhallinnan tiedetään korreloivan luetun ymmärtämisen ja muun kielellisen osaamisen kanssa (Milton 2009: 170–192; Schmitt 2010: 3–8). Tutkimuksen kohteena on sanatiedon syvyys eli tieto, jota selkolukijoilla on sanoista. Tutkimuksen ulkopuolelle jätetään sanatiedon leveys, joka puolestaan tarkoittaa sanavaraston laajuutta (ks. esim. Schmitt 2014). Siinä missä sanatiedon leveyttä voidaan kuvailla kvantitatiivisesti, syvyys on vaikeammin hahmotettavissa, sillä se voi tarkoittaa monenlaista tietoa sanojen muodosta, merkityksistä ja verkostoista (Nation 2001: 27; Read 2004; suomeksi ks. esim. Honko 2013). Vaikka leveys ja syvyys ovat omia ulottuvuuksiaan, ne eivät ole toisistaan irrallaan, sillä sanavaraston karttuessa uudet sanat asettuvat osaksi mentaalileksikossa olevaa sanojen verkostoa (Aitchison 1987; Read 2004: 221). Tästä näkökulmasta katsottuna sanatiedon syvyys tarkoittaa tietoa semanttisista verkostoista ja kykyä tunnistaa merkitykseltään yhteenkuuluvia sanoja.

Tutkimuksen tavoitteena on selvittää, miten hyvin selkolukijat tunnistavat sanojen lähimerkityksisiä vastineita. Tutkimusmenetelmänä on sanassosiaatiotesti, jossa osallistujat valitsevat ärsykesanalle samaan semanttiseen verkostoon kuuluvia sanoja (ks. esim. Read 1993; 2012). Osallistujina on selkolukijoita ja korkeakouluopiskelijoita, joiden vastauksia arvioidaan suhteessa toisiinsa ja sanojen sanakirjamerkityksiin. Tutkimuksen lähesty-

¹ Kriteerit sopivat parhaiten informatiivisten selkotekstien arviointiin.

mistapa on vastaanottajalähtöinen, eli tutkimuksen kohteena on osallistujien kielellinen osaaminen ja kognitiiviset kyvyt (vrt. Leskelä 2019a: 370–371). Vastaavanlaista selkokielen kohderyhmien kielitaidon tutkimusta on tehty kansainvälisestikin melko vähän, ja monet aiemmista tutkimuksista ovat keskittyneet lähinnä selkotekstien ymmärrettävyyden tarkasteluun (Sutherland & Isherwood 2016; Chinn & Homeyard 2017; Pappert & Bock 2019).

Tämä artikkeli avaa uusia näkökulmia suomalaiseen selkokielen tutkimukseen, sillä kehitysvammaisia osallistujia ei valittu kielelliseen testiin lääketieteellisten diagnoosien vaan lukutottumusten perusteella. Kehitysvammaiset henkilöt muodostavat kognitiivisilta kyvyiltään ja lukutaidoltaan vaihtelevan ryhmän (Vinni 1998; Ratz & Lenhard 2013; Di Blasi ym. 2018), ja on selvää, että lukijakunnan heterogeenisuus täytyy ottaa huomioon myös selkoikieltä tutkittaessa. Tässä tutkimuksessa kehitysvammaiset aikuiset osallistuivat sana-assosiaatiotestiin, jonka tarkoituksena on tuottaa uutta tietoa paitsi koehenkilöryhmän ja verrokkiryhmän välisistä sanastonhallinnan eroista myös koehenkilöryhmän sisäisistä eroista. Menetelmää on käytetty Suomessa aiemminkin, sillä Pajunen ym. (2015) ovat käyttäneet sana-assosiaatiotestiä nuorten aikuisten sanamerkitysten hallinnan testaamiseen ja Tervola ym. (2020) maahanmuuttajalääkäriä kielitaidon testaamiseen. Oma tutkimukseni on osoitus testiformaatin sopivuudesta selkokielen eri kohderyhmille, sillä kehitysvammaiset henkilöt ovat kohderyhmä siinä missä suomen kieltä opiskelevat maahanmuuttajatkin.

Tässä artikkelissa etsitään vastauksia seuraaviin kysymyksiin: Miten tarkasti kehitysvammaiset selkolukijat hallitsevat sanamerkityksiä, ja millaista osaamistason vaihtelua heillä on? Vaikuttavatko lukukokemuksen määrä sekä testisanojen yleisyys, tuttuus ja konkreettisuustaso heidän suoriutumiseensa sana-assosiaatiotestissä? Kysymyksiin vastataan tarkastelemalla osallistujien valitsemien vastinesanojen määrää yksilökohtaisesti ja sanakohtaisesti. Hypoteesina on, että selkolukijoiden vastauksissa on enemmän hajontaa kuin verrokkien ja että harvinaiset ja abstraktit sanat ovat selkolukijoille vaikeampia kuin yleiset ja konkreettiset. Artikkelin lopussa pohditaan, miten tutkimuksen tuloksia voidaan käyttää selkokielen kehittämiseen.

2 Osallistujat ja menetelmät

Tampereen alueen ihmistieteiden eettinen toimikunta teki tutkimukselle eettisen ennakkoarvion ja antoi myöntävän lausuntonsa kesäkuussa 2018. Tut-

kimus toteutettiin Etelä-Suomen alueella keväällä 2019 ja koehenkilöt tavoitettiin ryhmistä, joissa kehitysvammaisille aikuisille tarjotaan osallisuutta edistävää toimintaa. Verrokkiaineisto kerättiin korkeakouluopiskelijoilta saman vuoden syksyllä. Tutkimuksen osallistujat esitellään luvussa 2.1 ja sana-assosiaatiotestistä kerrotaan luvussa 2.2.

2.1 Osallistujat

Tutkimukseen osallistui 24 äidinkieleltään suomenkielistä kehitysvammaista aikuista, jotka lukivat selkotekstejä säännöllisesti tai satunnaisesti. He edustavat Leskelän & Lindholmin (2012) jaottelun mukaisesti selkokielen kohde-ryhmää, jonka selkokielen tarve on pysyvää ja johtuu neurobiologisista syistä. Osallistujien mediaani-ikä oli tutkimuksen tekohetkellä 37 vuotta, ja he jakautuivat ikäryhmiin seuraavasti: yhdeksän osallistujista oli 20–29-vuotiaita, viisi 30–39-vuotiaita, kuusi 40–49-vuotiaita ja neljä 50–55-vuotiaita. Osallistujille kerrottiin, että tutkimukseen osallistuminen on vapaaehtoista, ja heiltä pyydettiin suostumus tutkimukseen. Kaikki selkolukijat olivat aktiivisia aikuisia, jotka päättivät osallistumisestaan itse.

Verrokkiaineisto kerättiin 24 korkeakouluopiskelijalta, joiden mediaani-ikä oli tutkimuksen tekohetkellä 24,5 vuotta. He jakautuivat ikäryhmiin seuraavasti: kahdeksantoista oli 20–29-vuotiaita, viisi 30–39-vuotiaita ja yksi 40–49-vuotias. Korkeakouluopiskelijat valittiin vertailuryhmäksi edustamaan aikuistason kielenhallintaa (vrt. Pajunen ym. 2015). Korkeakouluopiskelijoiden käyttö vertailuryhmänä on perusteltua myös selkokielen käyttöön liittyvän epäsymmetrian vuoksi: selkokieltä tuottavat ihmiset eivät tavallisesti itse kuulu selkokielen kohde-ryhmiin, joten heidän käsityksensä sanojen vaikeustasosta voivat olla erilaisia kuin eri-ikäisten selkokielen käyttäjien. Kahden erilaisen kielenkäyttäjärühmän vertailusta saadaan tietoa epäsymmetriaan liittyvistä eroista.

Kehitysvammaisia selkolukijoita ei valittu tutkimukseen lääketieteellisten diagnoosien vaan lukutottumusten perusteella. He vastasivat tutkimuksen alussa taustakyselyyn, jolla selvitettiin lukuharrastuksen määrää ja lukemisen painottumista selkokieleeseen tai yleiskieleen. Kyselyn pistemäärä vaihteli välillä 6–24, ja mitä enemmän pisteitä vastaaja sai, sitä enemmän ja monipuolisemmin hän luki kirjoja, lehtiä ja muita tekstejä selko- ja yleiskielellä. Likert-tyyppiset kyselyt on todettu päteväksi tavaksi testata kehitysvammaisia nuoria ja aikuisia (Hartley & MacLean 2006). On kuitenkin havaittu, että vamman laatu voi vaikuttaa osallistujien kykyyn vastata kyselyyn, mikä saat-

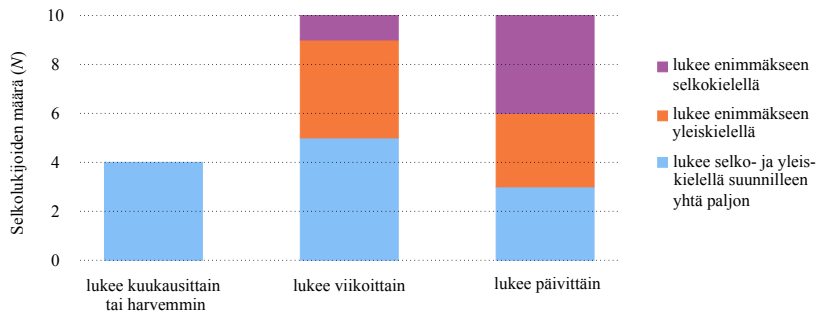
taa näkyä puuttuvina vastauksina ja taipumuksena valita myönteisimpiä vaihtoehtoja (Hartley & MacLean 2006: 815–818). Luotettavuutta ja vastausten määrää voidaan lisätä varmistamalla, että osallistujat ymmärtävät, miksi ja miten kyselyyn vastataan (Hartley & MacLean 2006: 818–823).

Tässä tutkimuksessa selkolukijoille kerrottiin, että kyselyllä selvitetään heidän lukutottumuksiaan ja että kaikenlaiset tottumukset ovat yhtä arvokkaita. Ennen kyselyä he tutustuivat selkokieliseen tutkimustiedotteeseen, ja myös kysely toteutettiin selkokielellä.² Osallistujat lukivat kyselyä itsenäisesti ja vastasivat kaikkiin kohtiin yksiselitteisesti ja täydellisesti. Kyselyssä käytettiin selkokieltä, koska on havaittu, että kehitysvammaiset henkilöt ymmärtävät selkokielisiä tekstejä paremmin kuin yleiskielisiä (Vinni 1998). Kehitysvammaisilla henkilöillä on usein lukemisen sujuvuuteen ja luetun ymmärtämiseen liittyviä vaikeuksia, mutta sanaston ja kielen rakenteiden hallinnassa sekä lukemisen taidoissa on yksilöllisiä eroja (Fajardo ym. 2014).

Kyselyn ajatuksena oli, että monipuolinen lukijaprofiili kertoisi lukukokeneisuuden myötä karttuneesta kielellisestä osaamisesta, kun taas yksipuolinen lukijaprofiili tarkoittaisi vähäisempää kokeneisuutta ja mahdollisesti heikompaa kielellistä osaamista. Kyselyvastauksista piirtyy selkolukijoiden lukijaprofiilien jatkumo, joka koostuu lukemisen toistuvuudesta ja lukemisessa painottuvasta kielimuodosta. Vähiten lukevat osallistujat lukivat kuukausittain tai harvemmin, enemmän lukevat osallistujat lukivat viikoittain, ja eniten lukevat osallistujat lukivat päivittäin. Lukijoiden kesken oli vaihtelua siinä, lukivatko he useammin selkokielellä vai yleiskielellä vai käyttivätkö he kumpaakin kielimuotoa suunnilleen yhtä paljon. Selkolukijoiden lukukokeneisuus ja kielimuotojen painottuminen esitetään kuviossa 1.

Kymmenen osallistujista luki päivittäin, kymmenen viikoittain ja neljä kuukausittain tai harvemmin. Viidellä lukeminen painottui selkokieleen, seitsemällä yleiskieleen ja kahdellatoista ei kumpaankaan, eli he lukivat selko- ja yleiskielisiä tekstejä suunnilleen yhtä usein. Kuukausittain tai harvemmin lukevat osallistujat lukivat kokonaisuudessaan hyvin vähän ja käyttivät selko- ja yleiskieltä suunnilleen yhtä paljon. Viidellä viikoittain lukevista osallistujista lukeminen ei painottunut erityisesti kumpaankaan kielimuotoon, ja neljällä lukeminen painottui yleiskieleen. Eniten selkokieltä lukevia löytyi päivittäin lukevien joukosta, mitä selittää heidän aktiivinen lukuharrastuksensa: he kertoivat lukevansa selkokirjoja useammin kuin muut. Seitsemän päivittäin luke-

² Selkokeskus tarkasti tutkimustiedotteen ja kyselyn kieli- ja ulkoasun ja myönsi niille selkotunnuksen. Selkotunnus myönnetään teksteille, jotka ovat selkokielen kriteerien mukaisia.



Kuvio 1. Selkukojien lukukokoneisuus ja kielimuotojen painottuminen lukemisessa

vista osallistujista käytti selkokieltä lähes pelkästään tai yleiskielen rinnalla, mutta mukana oli kolme osallistujaa, joilla lukeminen painottui yleiskieleen ja selkokielen tarve oli vähäistä. Osallistujien lukutottumukset siis vaihtelivat määrällisesti ja laadullisesti, mutta kaikkia yhdisti jonkinlainen selkokielen tarve.

2.2 Selektiivinen sana-assosiaatiotesti

John Read (ks. esim. 1993; 2012) kehitti 1990-luvulla selektiivisen sana-assosiaatiotestin (*word associates format, WAF*), jolla tutkitaan sanamerkitysten hallintaa. Testissä osallistujille esitetään ennalta valittuja sanoja, ja heidän tehtävänä on tunnistaa ja valita yhteenkuuluvia sanoja hämäyssanojen joukosta. Yhteenkuuluvien sanojen suhteet voivat olla paradigmaattisia (synonyymit, vieruskäsitteet sekä ylä- ja alakäsitteet), syntagmaattisia (kollokaatit) tai analyyttisiä (vastinesana edustaa osaa ärsykesanan merkityksestä). Kun osallistujat valitsevat sanoja, he vertailevat niitä toisiinsa, mikä edellyttää tietoa sanojen välisistä suhteista. Testitulokset siis kertovat kielenhallinnasta yleisemminkin kuin yksittäisten sanamerkitysten osalta. Testissä sanat esitetään ilman lauseyhteyttä, eli tutkimuksen kohteena on kontekstivapaa leksikaalinen tieto (vrt. Pajunen ym. 2015: 166–167). Jos osallistujat tunnistavat yksittäisiä sanoja, heidän sanatietonsa on merkityksen hallinnan tasolla eikä perustu tilannesidonnaisuuteen, jossa merkitykset tunnistetaan kontekstista (vrt. Schoonen & Verhallen 2008: 213). Koska testi voidaan kohdentaa tietyn-

laisiin sanojen välisiin suhteisiin, se valittiin myös tähän tutkimukseen, jossa tarkastellaan lähimerkityksisten sanojen hallintaa.

Ennen selektiivisen testin kehittämistä sanamerkitysten hallintaa tutkittiin vapaaseen assosiaatioon perustuvilla testeillä, joissa sanoja tuotetaan ärsyke-sanojen pohjalta (ks. esim. Meara 2009). Vapaita assosiaatioita tutkinut Meara (2009: 22–23) kuvailee sana-assosiaatioille tyypillisiä piirteitä kielellisen kehityksen ja kielenoppimisen eri vaiheissa seuraavasti: Aikuiset natiivipuhujat tuottavat yleensä paradigmaattisia assosiaatioita (esim. *koira* > *eläin*), kun taas lapset ovat taipuvaisia tuottamaan syntagmaattisia assosiaatioita (esim. *koira* > *haukkua*) tai äänneassosiaatioita, joissa vastaus heijastelee ärsykesanan muotoa (esim. *katala* > *matala*). Kielenoppijat puolestaan tuottavat vaihtelevampia assosiaatioita kuin natiivipuhujat, ja heidän vastauksensa saattavat olla epäodotuksenmukaisia esimerkiksi puutteellisesta kielitaidosta johtuvien väärinymmärrysten vuoksi. Vapaissa sana-assosiaatioissa on siis kehityksellisiä eroja, jotka kertovat osallistujien leksikon laadusta. Jos vastaavanlaisia eroja tulee näkyviin myös toteutustavaltaan rajatummassa selektiivisessä testissä, niitä voidaan pitää yhtä lailla informatiivisina.

Selektiivisestä sana-assosiaatiotestistä on tehty monenlaisia versioita kielitaidoltaan erilaisille osallistujaryhmille mm. sanaluokkaa, vastinesanojen määrää ja vastine- ja hämäyssanojen laatua vaihdellen (ks. esim. Qian 1999; Greidanus ym. 2005; Schoonen & Verhallen 2008; Horiba 2012). Testin luotettavuutta vähentää mahdollinen arvaaminen, mutta luotettavuutta on pyritty lisäämään kehittämällä pisteytystapoja, joissa virheelliset valinnat vähentävät pistemäärää tai valinnat pisteytetään sanaverkkokohtaisesti (pisteytystavoista ks. Schmitt ym. 2011). Read (2012: 312) on kuitenkin havainnut, että arvaamisessakin käytetään jonkinlaista sanatietoa, jonka perusteella valintoja tehdään. On myös havaittu, että osallistujat saattavat jättää testikohdan tyhjäksi sen sijaan, että arvaisivat tuntemattomia sanoja (Qian & Schedl 2004: 44; Read 2012: 311).

Sanaverkkokohtaista kaikki oikein -pisteytystä (engl. *all or nothing*) käytetään tässä tutkimuksessa yhtenä tarkastelutapana (ks. § 4.1 ja 4.2). Siinä vastaaja saa pisteen, kun hän valitsee ärsykesanan kaikki vastinesanat oikein ja jättää hämäyssanat valitsematta (vrt. Schoonen & Verhallen 2008). Schmitt ym. (2011: 119) ovat todenneet, että kaikki oikein -pisteytys sopii erityisen hyvin kuuden verrokkisanan sana-assosiaatiotesteihin, kun tavoitteena on selvittää vastaajien välisiä osaamiseroja. Pisteytystapa on kuitenkin ankara, ja arvaamalla on vaikea tehdä kaikki valinnat oikein (vrt. Schoonen & Verhallen 2008: 223–224). Ankaruutta voidaan pitää sekä etuna että haittana,

sillä vastaajat eivät voi näyttäytyä pätevämpinä kuin ovat, mutta samalla heidän osittainen sanatietonsa jää huomiotta (Schmitt ym. 2011: 122). Siksi tässä tutkimuksessa tarkastellaan myös osittain oikeita valintoja, jotta saadaan selville ryhmänsäisiä eroja.

Selkolukijoiden testisuorituksia arvioidaan suhteessa lukukokeneisyyteen, ja osallistujia arvioidaan myös suhteessa toisiinsa (ks. § 4.1 ja 4.2). Lisäksi selkolukijoiden ja verrokkien testisuorituksia vertaillaan ja arvioidaan suhteessa sanojen sanakirjamerkityksiin (ks. § 3.2 ja 4.5), sanojen yleisyyteen ja tuttuuteen (ks. § 4.3) sekä konkreettisuustasoon (ks. § 4.4). Selkolukijoiden ja verrokkien vertailua voidaan pitää tutkimuksen vahvuutena: vaikka osa sanoista valittaisiin arvaamalla ja testi yliarvioisi kummankin osallistujaryhmän osaamista, vertailusta saadaan tietoa sanamerkityksen hallinnan laadullisista eroista (ks. Bogaards 2000: 496).

3 Tutkimuksen toteutus

Tutkimusta varten tehtiin synonymiaan perustuva sana-assosiaatiotesti, jonka alustavassa aineistossa oli 312 sanaa. Alustava sana-aineisto kerättiin etsimällä *MOT Kielitoimiston sanakirjasta* ja *MOT Synonyymisanakirjasta* sanoja, joilla on paljon synonyymejä. Vaikka testi ei ollut selkokielen, jo alustavia valintoja tehdessä tukeuduttiin Selkomittarin (Selkokeskus 2018: 9) periaatteisiin eli suositettiin konkreettisia perussanoja, joiden arvioitiin olevan yleisiä ja mahdollisimman monille tuttuja. Vaikka selkokielessä neuvotaan käyttämään yleistä ja lukijoille tuttua sanastoa, ei ole täysin selvää, miten yleisyyttä ja tuttuutta pitäisi arvioida (Leskelä & Kulkki-Nieminen 2015: 48–49; Leskelä 2019b: 131). Kielentutkimuksessa sanaston yleisyyttä tarkastellaan usein frekvenssillä eli laskemalla sanojen esiintymiä korpuksesta, mutta siitä ei selviä, miten usein sanoja käytetään muissa yhteyksissä ja miten hyvin sanat tunnetaan (Schmitt 2010: 67–68). Selkokielen tutkimuksessa tuttuuden arvioinnin lähtökohdaksi on ehdotettu muun muassa sanaston prototyypillisyyttä ja keskeisyyttä selkokielen kohderyhmien elämäalueilla, mutta käsitykset sanojen tuttuudesta perustuvat ainakin jossakin määrin oletuksiin (Leskelä 2019a: 376; 2019b: 131–134).

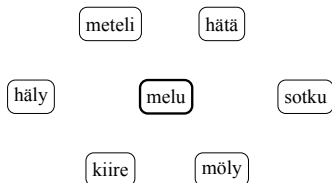
Tässä tutkimuksessa alustavan sana-aineiston tuttuutta selvitettiin tuttuustestauksella, johon osallistui 23 vammautonta nuorta aikuista.³ He arvioivat

³ Tuttuustestauksen osallistujat ja sana-assosiaatiotestien verrokot olivat kaksi eri ryhmää. Kiitän kaikkia tutkimukseen osallistuneita.

sanoja asteikolla: 1 = ”en tunne sanaa”, 2 = ”en ole sanasta ihan varma”, 3 = ”sana vaikuttaa tutulta”, 4 = ”sana on tuttu”, 5 = ”käytän sanaa itse” (vrt. Pajunen ym. 2015: 162). Ajatuksena oli, että osallistujien omaan arvioon perustuva tuttuustuntuma (ks. Balota ym. 2001: 639–640) ilmaisee arkikielen ja harvinaisen sanaston tunnettuutta paremmin kuin objektiivinen frekvenssi, sillä korpuksessa tällaisia sanoja ei välttämättä esiinny. Tuttuustestauksen perusteella sana-assosiaatiotestiin valittiin 20 ärsykesanaa ja 120 verrokkisanaa, jotka olivat tuttuustestauksen osallistujille tuttuja tai joita he käyttivät itse. Yksi verrokkisanoista esiintyi testissä kaksi kertaa, joten testissä oli yhteensä 139 erilaista substantiivia. Testiin valitut sanat olivat pääasiassa lyhyitä ja johtamattomia. Ärsykesanojen mediaanipituus oli 5,5 kirjainta ja verrokkisanojen 5 kirjainta.

Testisanojen yleisyyttä tarkasteltiin *Helsingin Sanomien* numeroista koostuvasta HS2000-korpuksesta (Pajunen & Virtanen 2002) poimittujen frekvenssien perusteella. Ärsykesanojen frekvenssi oli 5–480 esiintymää miljoonassa ja mediaani oli 26. Kahdestakymmenestä ärsykesanasta 7 oli hyvin yleisiä (> 30 esiintymää miljoonassa), 11 melko yleisiä tai yleisiä (10–30 esiintymää miljoonassa) ja 2 melko harvinaisia (< 10 esiintymää miljoonassa ja < 5 esiintymää miljoonassa). Verrokkisanojen frekvenssin vaihteluväli oli suurempi kuin ärsykesanojen, ja niiden yleisyys oli 0,3–2 306 esiintymää miljoonassa ja mediaani oli 10. Noin 30 % verrokkisanoista oli frekvenssiltään hyvin yleisiä, noin 20 % melko yleisiä tai yleisiä ja 50 % melko harvinaisia. Frekvenssi ei siis ollut määräävä kriteeri verrokkisanojen valinnassa, vaan testiin valittiin sanoja, jotka täyttivät tuttuuden kriteerin. Näin testiin päätyi sanoja, kuten *pipo*, *muki* ja *rapa*, jotka olivat korpuksessa harvinaisia mutta tuttuustestauksen perusteella tuttuja. Toisaalta testiin tuli myös muutama hyvin harvinainen sana, jotka eivät tuttuustestauksen perusteella olleet kovinkaan aktiivisessa käytössä. Näitä olivat esimerkiksi *koloitus* ja *viitsa*.

Semanttisesti ärsykesanat jakautuivat konkreettisiin, melko konkreettisiin, melko abstrakteihin ja abstrakteihin. Ajatuksena oli, että siirryttäessä konkreettisesta kohti abstraktia sanojen tarkoitteiden havaittavuus eri aisteilla vähenee (Brysaert ym. 2013). Lisäksi sanat jaettiin semanttisiin alaluokkiin käsitteellisesti, esim. esinesanat, kasvisanat ja ainesanat (Miller 1998). Konkreettisia sanoja oli testissä eniten, sillä selkokielessä pyritään suosimaan niitä (Leskelä 2019b: 132–133). Testin konkreettisimpia sanoja olivat laskettavat esinesanat *laatikko*, *hattu*, *sirpale*, *kuppi*, *korppi* ja *allas*, kasvisana *oksa* sekä inhimilliset roolinimet *ystävä* ja *johtaja*, eli mukana oli elollisiin ja elottomiin tarkoitteisiin viittaavia sanoja. Melko konkreettisia olivat kollektiivisana



Kuvio 2. Esimerkki sanaverkosta

joukko, ainesana *kura* sekä spatiaalisiin rakennelmiin viittaavat *mutka* ja *reitti*, joiden tarkoitteet voidaan havaita eri aisteilla mutta joita ei voida laskea tai käsitellä yhtä helposti kuin konkreettisten sanojen tarkoitteita. Melko abstrakteja sanoja olivat *haju*, *kipu* ja *melu*, jotka viittasivat tarkoitteisiin, jotka voidaan aistia mutta joita ei voida nähdä tai koskettaa. *Kilpailu*, *käsky*, *ongelma* ja *vilppi* viittasivat abstrakteihin tarkoitteisiin, jotka eivät ole tarkkarajaisia mutta joista voi syntyä esimerkiksi tilanteisiin liittyviä mielikuvia.

Testiin valitut sanat asetettiin sanaverkon muotoon (ks. kuviota 2), ja osallistujien tuli yhdistää vastinesanat keskellä olevaan ärsykesanaan (vrt. Schoonen & Verhallen 2008: 218–219).

Sanaverkoissa vastine- ja hämäyssanojen yhteismäärä oli kuusi, ja jokaisesta sanaverkosta tuli valita kolme lähimerkityksisintä sanaa.⁴ Testissä oli 20 ärsykesanaa, joista jokaiselle oli 3 vastinesanaa, eli oikeita valintoja oli koko testissä 60. Hämäyssanoja oli kahdenlaisia: osalla oli ärsykesanaan jokin mahdollinen merkitysyhteys ja osa oli merkitykseltään yhteensopimattomia. Oletuksena oli, että yhteensopimattomat hämäyssanat helpottavat testiä, kun taas jokin mahdollinen merkitysyhteys erottelee osaamista (Greidanus & Nienhuis 2001: 572–573).

3.1 Aineisto

Sana-assosiaatiotestiä esiteltiin osallistujille suullisesti ja varmistettiin, että he osaavat vastata siihen. Testi tehtiin kynä–paperi-suorituksena, ja ohje sanaverkon täyttämiseen oli seuraavanlainen: ”Piirrä viiva niiden sanojen välille, jotka tarkoittavat samaa tai melkein samanlaista asiaa.” Selkolukijat harjoittelivat vastaamista yhdellä sanaverkolla, jota ei käytetty tutkimuksen aineistona. Testissä sanaverkot esitettiin omilla sivuillaan, joten osallistujat

⁴ Kuviossa 2 vastinesanoja ovat *meteli*, *möly* ja *häly* ja hämäyssanoja *hätä*, *sotku* ja *kiire*.

näki vait yhen verkon kerrallaan. Testistä käytettiin kahta samansisältoistä versiota, joista toisessa sanaverkot olivat käänteisessä järjestyksessä, sillä pyrkimyksenä oli vähentää mahdollisia väsymyksen vaikutuksia testin loppupuolella. Testin kummassakin puoliskossa oli suunnilleen samantyyppisiä ärsykesanoja, jotta vaikeustaso pysyi samanlaisena läpi testin. Testin tekemiseen käytettävää aikaa ei rajoitettu.

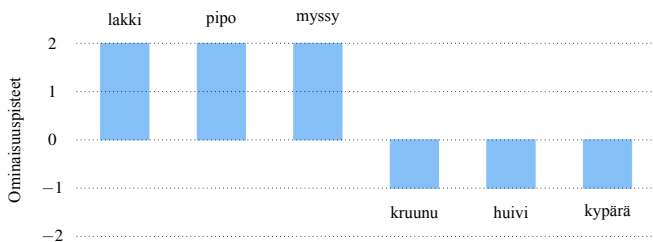
Jokainen selkolukija täytti testissä 20 sanaverkkoa, eli aineistossa oli yhteensä 480 vastausta. Näistä 464 oli tehtävänannon mukaisia. Aineistossa oli kahdeksan täytettyä sanaverkkoa, joissa oli valittu vähemmän kuin kolme sanaa, mutta puutteelliset vastaukset pidettiin analyysissä mukana, koska tutkimuksessa tarkasteltiin myös osittaista sanatietoa. On mahdollista, että sanat jäivät valitsematta, koska vastaaja ei osannut tehdä valintaa eikä myöskään halunnut arvata. Kahdeksassa täytetyssä sanaverkossa valinnat jäivät epäselviksi tai puuttuivat kokonaan,⁵ mutta luvussa 4.4 käydään läpi, mihin sanoihin nämä liittyivät. Verrokkihenkilöitä osallistui testiin 24, mutta yhden vastaukset poistettiin analyysivaiheessa.⁶ Analysoitavaksi jäi siis 460 (eli 23×20) täytettyä sanaverkkoa, joista viidessä oli valittu kaksi sanaa.

3.2 Sanojen ominaisuuspisteet ja läheisyysarviot

Sanat pisteytettiin mukailamalla Pajusen ym. (2015) käyttämää menetelmää, jota kuvaillaan tässä luvussa. Menetelmässä vastine- ja hämäyssanoille annetaan ominaisuuspisteet, joilla ilmaistaan niiden suhdetta ärsykesanaan. Pisteiden tarkoituksena on osoittaa jokaiselle ärsykesanalle kolme lähimerkityksistä vastinetta sanakirjamerkityksen perusteella. Tässä tutkimuksessa ärsykesanan kanssa samalla hierarkiatasolla oleville vastinesanoille annettiin korkeimmat pisteet, sillä testi perustui synonymiaan. Synonymia määriteltiin samalla tavalla kuin *MOT Synonymisanakirjassa*: samaa merkitsevien sanojen lisäksi synonyymeiksi laskettiin lähivastineet, jotka kuuluvat joiltakin merkityspiirteiltään yhteen ja jotka voidaan korvata toisillaan ainakin joissakin konteksteissa. Siksi testissä oli myös geneerisen tason vastineita, eli synonyymit eivät olleet täysin tarkkoja.

⁵ Mukana oli tyhjä sanaverkko, tulkinnanvaraisia merkintöjä sekä vastauksia, joissa oli valittu enemmän kuin kolme sanaa. Jälkimmäisessä syynä saattoi olla joko vahinko tai pyrkimys valita mahdollisimman monta sanaa, jotta osa valinnoista menisi oikein (vrt. Zhang & Koda 2017: 19). Tähän viittaavat mm. vastaukset, joissa puolet tai yli puolet valinnoista kohdistui hämäyssanoihin.

⁶ Puolet valinnoista poikkesi tehtävänannosta.



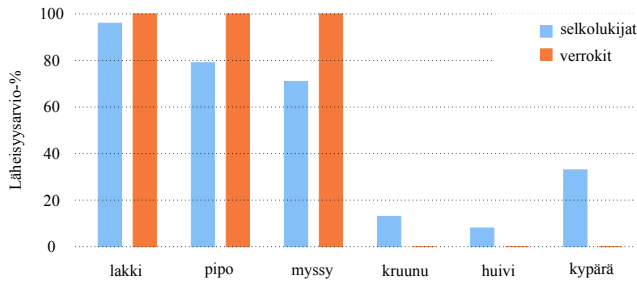
Kuvio 3. Ärsykesanan *hattu* verrokkisanojen ominaisuuspisteet

Pisteytystä varten ärsykesanojen merkitykset katsottiin *MOT Kielitoimiston sanakirjasta* ja tarvittaessa laajemmat määritelmät *Nykysuomen sanakirjasta*. Vastine- ja hämäyssanojen sanakirjamerkityksiä vertailtiin ärsykesanojen merkityksiin ja sanat pisteytettiin asteikolla $-2 \dots 2$ seuraavasti:

- -2 = Hämäyssanalla ja ärsykesanalla ei ole merkitysyhteyttä (esim. *oksa-tolppa*).
- -1 = Hämäyssana ja ärsykesana eroavat merkitykseltään, mutta niillä voi olla jokin yhteys (esim. *johtaja-sankari*). Sanat voivat olla myös lähi-merkityksisiä, jolloin niitä erottaa jokin tietty ominaisuus, joka vastaa-
jan täytyy tunnistaa (esim. *hattu-kruunu*, ks. kuviota 3).
- 1 = Vastinesana ja ärsykesana ovat geneerisessä suhteessa toisiinsa (esim. *allas-säiliö*).
- 2 = Vastinesana ja ärsykesana ovat lähes tai täysin samamerkityksiset (esim. *melu-meteli*).⁷

Selkolukijoiden ja verrokkiryhmän testisuoritusten perusteella vastine- ja hämäyssanoille tehtiin läheisyysarviokaaviot, joista selviää, kuinka moni vastaajista valitsi kunkin sanan (vrt. Pajunen ym. 2015: 171). Kaaviossa valinnat ilmaistaan prosentteina osallistujien määrästä eli selkolukijoilla 24:stä ja verrokkiryhmällä 23:sta. Epäselvät ja puuttuvat valinnat merkittiin nollassi,

⁷ Pajunen ym. (2015: 168–169) asteikossa on mukana myös 0, joka annettiin sanoille, jotka eivät käy vastineiksi yleisimmän merkityksensä perusteella mutta joilla on yhteinen sivumerkitys. Omassa testissäni tällaisia sanoja ei ollut.



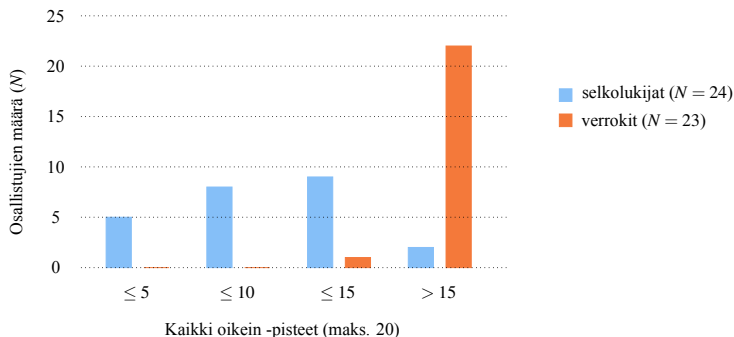
Kuvio 4. Ärsykesanan *hattu* vastinesanojen läheisyysarviot

minkä vuoksi läheisyysarviot jäivät joissakin sanoissa odotettua matalammiksi. Verrokkihenkilöiden tulokset osoittivat testin toimivaksi, sillä korkeimmat läheisyysarviot osuivat kaikissa sanaverkoissa sanoille, joilla oli korkeimmat ominaisuuspisteet (vrt. kuvioita 3 ja 4).

Esimerkkisanan *hattu* kaikki verrokkisanat sisältävät laajasti ajateltuna merkityksen 'päähine', mutta osallistujien tehtävänä oli valita vain kolme lähimerkityksistä sanaa. Heidän tuli tunnistaa, että *lakki*, *pipo* ja *myssy* ovat kuvullisia päähineitä, jossa on tai ei ole lieriä tai lippaa, kun taas hämäyssanoissa *kruunu*, *huivi* ja *kypärä* on lisänä merkityspiirteitä, jotka vievät niitä kauemmas ärsykesanan merkityksestä ('vallan tunnus', 'pään suojana pidettävä kangaskappale', 'kova suojapäähine'). Kuviosta 4 selviää, että kumpikin osallistujaryhmä valitsi yleisimmin vastinesanat *lakki*, *pipo* ja *myssy*, mutta ainoastaan verrokkit olivat valinnoista täysin samaa mieltä. 33 % selkoluki-joista piti hämäyssanaa *kypärä* sopivana vastineena, ja valituiksi tulivat myös *kruunu* ja *huivi*.

4 Tulokset

Luvussa 4.1 selvitetään ryhmienvälisiä eroja vertailemalla, kuinka moneen sanaverkkoon osallistujat valitsivat kaikki vastinesanat oikein. Testisuorituk-
sia tarkastellaan myös oikein valittujen vastinesanojen määrän perusteella, jotta saadaan tarkempi käsitys ryhmiensisisäisistä eroista. Luvussa 4.2 selvi-
tetään sanojen vaikeustasoa vertailemalla, kuinka monet valitsivat kuhunkin
ärsykesanaan kaikki vastinesanat oikein. Lisäksi tutkitaan osittaista sanatietoa
vastauksista, joissa osallistujat valitsivat kaksi vastinesanaa oikein. Luvuissa



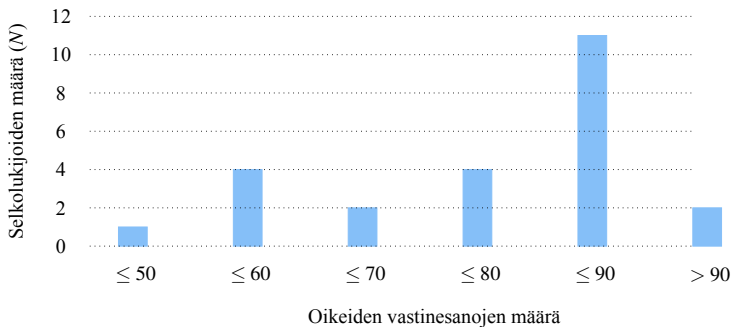
Kuvio 5. Osallistujien testisuoriutuminen kaikki oikein -pisteiden perusteella

4.3 ja 4.4 tarkastellaan sanojen yleisyyden, tuttuuden ja konkreettisuustason yhteyttä osallistujien tekemiin valintoihin. Luvussa 4.5 arvioidaan osallistujien kykyä valita vastinesanoja sanakirjamerkitysten mukaisesti.

4.1 Testisuoriutuminen ja osaamistason vaihtelu

Kaikki oikein -pisteytyksessä osallistujat saivat pisteitä sanaverkoittain, ja pisteiden maksimimäärä oli 20. Tuloksesta ilmenee, että verrokkiryhmän testisuoriutuminen oli tasaista ja lähes kaikki tekivät kaikki valinnat oikein yli 15 sanaverkkoon (ks. kuviota 5). Pisteet vaihtelivat välillä 15–20, ja mediaani oli 19, eli testi oli verrokeille helppo, joskaan ei itsestään selvä. Selkolukijoiden pisteet vaihtelivat välillä 1–16 ja mediaani oli 10, joten kokonaisten sanaverkkojen löytäminen oli heille vaikeampaa.

Parhaiten kokonaisia sanaverkkoja löytäneet selkolukijat ($N = 2$) saivat yli 15 pistettä, eli he pärjäsivät testissä siinä missä suurin osa verrokeistakin. Sanaverkkoja löysivät hyvin myös yli 10 pistettä saaneet selkolukijat ($N = 9$), ja heidän tuloksensa olivat verrokkiryhmän pistetason rajapinnassa, sillä yksi verrokkihenkilö asettui tähän samaan kategoriaan. Kahdeksan selkolukijaa täytti oikein yli viisi mutta enintään kymmenen sanaverkkoa, ja viisi täytti oikein enintään viisi sanaverkkoa. Selkolukijoiden testisuoriutumisessa oli siis selvästi enemmän vaihtelua kuin verrokeilla ja erot ääripäiden välillä olivat suuret.



Kuvio 6. Selkolukijoiden testisuoriutumisen oikeiden vastinesanojen määrän perusteella

Koska edellä esitetty ei kuvaile testisuoriutumista osittain oikein täytettyjen sanaverkkojen osalta, osaamistason eroja arvioidaan myös tarkastelemalla oikein valittujen vastinesanojen kokonaismäärää, joka oli 60 (20 ärsykesanaa \times 3 vastinesanaa). Kaikki verrokkit valitsivat yli 90 % vastinesanoista oikein, ja vaihteluväli 92–100 % ja mediaani oli 98. Selkolukijat valitsivat 45–93 % vastinesanoista oikein, ja mediaani oli 80. Yli puolet selkolukijoista löysi vastinesanoja hyvin tai erinomaisesti, kun taas hieman alle puolelle vastinesanojen löytäminen oli vaikeampaa (ks. kuviota 6).

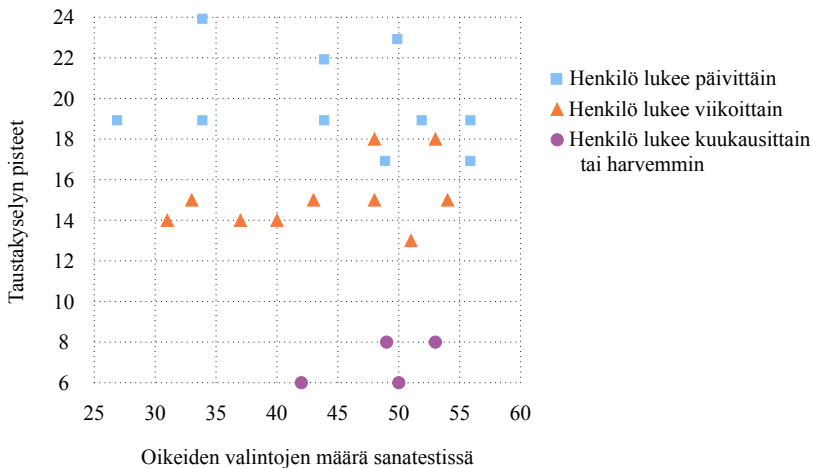
Parhaiten pärjänneet selkolukijat ($N = 2$) valitsivat yli 90 % vastinesanoista oikein, ja hyvin pärjänneet ($N = 11$) valitsivat yli 80 % sanoista oikein. Näiden vastaajien sanamerkitysten hallinta on hyvällä tasolla, kun huomioidaan, että osa sanoista ei ollut erityisen helppoja yleiskielen, saati sitten selkokielen kriteereillä arvioiden. Kuusi selkolukijaa löysi vastinesanoja kohtalaisesti eli valitsi 61–80 % sanoista oikein. Viidelle testi oli vaikea tai erittäin vaikea, ja he valitsivat enintään 60 % vastinesanoista oikein. Selkolukijoiden ryhmänsisäinen osaamistason vaihtelu on odotuksenmukaista, sillä he edustivat iän ja taustakyselyn perusteella varsin heterogeenistä lukijaryhmää. Vaihtelu antaakin aiheutta todeta, että testi oli vähiten pisteitä saaneille selkolukijoille liian vaikea eikä mitannut heillä sanatiedon syvyyttä vaan ennemminkin heidän tuntemiensa ärsykesanojen määrää. On kuitenkin mahdollista, että parhaiten pärjänneet selkolukijat eivät olisi erottuneet helpommalla testillä, joten testin vaikeustasoa voidaan siksi pitää onnistuneena (vrt. Greidanus ym. 2005: 225–226).

Sanatestissä osaaminen oli yhtenäisintä selkolukijoilla, jotka lukivat selko- ja yleiskielisiä tekstejä suunnilleen yhtä usein (ks. kuviota 1). 12 lukijasta 10 valitsi testissä yli 70 % vastinesanoista oikein. Selkokielistä tai yleiskielistä painottavien lukijoiden oikeat vastaukset jakautuivat vaihtelevammin eri vastausprosenttien välille, ja suoriutuminen oli muutenkin hajanaisempaa. Eniten heikkoja tuloksia (enintään 60 % vastinesanoista oikein) saivat osallistujat, jotka lukivat enimmäkseen yleiskielellä. Vaikuttaa siis, että osa selkolukijoista lukee yleiskielisiä tekstejä, vaikka ne olisivat tavalla tai toisella liian vaikeita, kun taas toisilla sanastonhallinta on yleiskielen lukemiseen riittävällä tasolla. Tätä ajatusta tukee myös taustakyselyssä ilmennyt yleiskielisten lehtien suuri suosittuus: 18 selkolukijaa ilmoitti lukevansa niitä viikoittain tai päivittäin, ja yleiskielisiä lehtiä luettiin useammin kuin selkokielisiä. Yleiskielisten lehtien suosiota selittänevät niiden aihepiireihin ja saatavuuteen liittyvät syyt, sillä lehtiä on saatavilla helposti ja monipuolisesti. Yleiskielisiä juoruja naistenlehtiä pidettiin kiinnostavina, mikä nosti lukemisintoa. Selkokielellä tällaisia lehtiä ei ole saatavilla.⁸ Yleiskielisiä kirjoja puolestaan saatettiin lukea siksi, että vastaajat eivät tieneet, mistä selkokirjoja voi hankkia.

Oikein valittujen vastinesanojen määrän vertailusta taustakyselyn tuloksiin selviää, että hyviin tuloksiin sanatestissä ylsivät niin vähän kuin paljonkin lukevat selkolukijat (ks. kuviota 7). Kysely ei kuitenkaan ollut tarpeeksi tarkka, jotta sen perusteella voitaisiin arvioida lukukokeneisuuden vaikutusta semanttiseen osaamiseen tarkemmin.

Sanatestin parhaimmat tulokset sai kaksi päivittäin lukevaa osallistujaa, joista toinen luki enimmäkseen yleiskielellä ja toinen selkokielellä. Heistä enimmäkseen yleiskielellä lukeva sai taustakyselystä vain 17 pistettä eli kokonaispistemäärä oli pienempi kuin joillakin viikoittain lukevilla. Todellisuudessa hän, samoin kuin toinenkin kyselystä 17 pistettä saanut osallistuja, luki yleiskielisiä tekstejä päivittäin, ja kokonaispistemäärä jäi matalaksi, koska selkokielellä lukeminen oli hyvin vähäistä. Selkokielen tarve oli siis lähinnä tilannesidonnaista ja liittyi tiettyihin tekstilajeihin, minkä vuoksi säännöllinen lukuharrastus toteutui yleiskielen varassa. Toinen sanatestistä huipputuloksen saanut osallistuja puolestaan luki selkotekstejä päivittäin ja yleiskielisiä viikoittain, eli parhaimmat tulokset saavutettiin melko erilaisilla lukijaprofiileilla.

⁸ Selkokielisiä lehtiä ilmestyy Suomessa kaksi: ajankohtaislehti *Selkosanommat* ja aikakauslehti *Leija*.



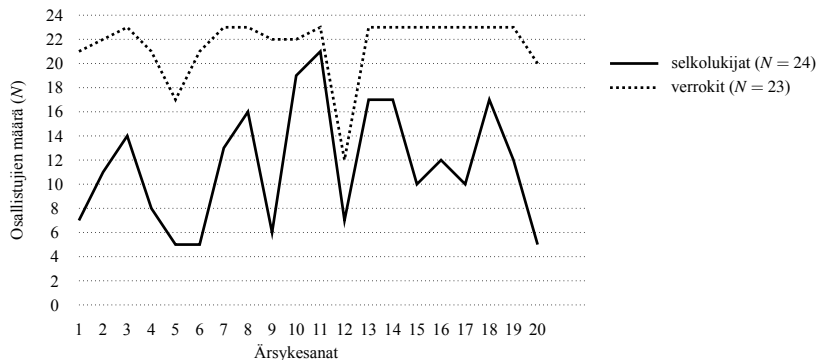
Kuvio 7. Selkolukijoiden lukukokoneisuus ja semanttinen osaaminen

4.2 Sanamerkitysten hallinta oikeiden vastausten perusteella

Seuraavaksi selvitetään, kuinka monet osallistujista valitsivat kuhunkin ärsykesanaan kaikki vastinesanat oikein. Tuloksesta ilmenee, että verrokkiryhmän sanakohtaiset valinnat olivat melko samanlaisia, mutta mukana oli myös muutama poikkeus, joihin he valitsivat kaikki oikeat vastinesanat huomattavasti harvemmin kuin muihin (ks. katkoviivaa kuviossa 8).

Vain 12 verrokkia teki kaikki valinnat oikein ärsykesanaan 12 (*korkki*) ja 17 ärsykesanaan 5 (*oksa*), joita käsitellään tarkemmin luvussa 4.4. Myös selkolukijoiden oikeiden vastausten määrät laskivat näissä sanoissa (ks. yhtenäistä viivaa kuviossa 8), joten tehtävä oli kummallekin osallistujaryhmälle vaikea. Kaikki verrokkit valitsivat yhteentoista ärsykesanaan kaikki vastinesanat oikein ja suorittivat *oksa*- ja *korkki*-sarjoja lukuun ottamatta muutkin sanaverkot varsin sujuvasti. Verrokkiryhmän sanakohtaisten kaikki oikein -suoritusten vaihteluväli oli *korkki*-sarjasta johtuen suuri eli 12–23.

Myös selkolukijoiden kaikki oikein -testisuoritusten vaihteluväli oli suuri eli 5–21, mutta heillä se ei johtunut yksittäisistä poikkeuksista, vaan vaihtelua oli ärsykesanoittain läpi testin (ks. yhtenäistä viivaa kuviossa 8). Huomattavaa on, että selkolukijoiden testisuoritukset vaihtelivat jyrkästi myös sanoissa,

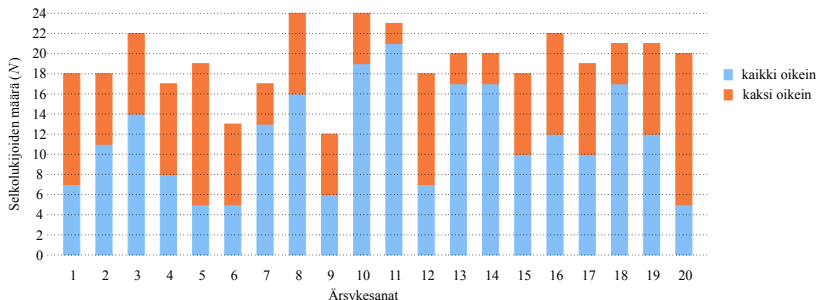


Kuvio 8. Kaikki oikein vastanneiden osallistujien määrä ärsykesanoittain

joissa verrokkien suoritukset pysyivät lähes tasaisina. Sanojen merkitykset olivat selkolukijoille kuitenkin tutumpia kuin tämän perusteella voidaan päätellä, sillä he saattoivat tuntea sanan verkostoa osittain ja valita kaksi vastinesanaa oikein (ks. kuviota 9).

Selektiivisen sana-assosiaatiotestin luotettavuutta tarkastelleet Schmitt ym. (2011) ovat havainneet, että kuuden verrokkisanan testissä kolmea oikein valittua vastinesanaa voidaan pitää osoituksena sanamerkityksen hallinnasta. Kahden oikean vastinesanan valitsemista he pitivät osittaisena sanamerkityksen hallintana, joskin testi saattaa yliarvioida osaamista, jos osallistujat ovat arvanneet. Jos testi yliarvioi selkolukijoiden osaamista, se yliarvioi yhtä lailla myös verrokkiryhmän osaamista. Testissä verrokkit valitsivat kaksi vastinesanaa oikein niille ärsykesanoille, joille he eivät valinneet kolmea vastinesanaa oikein (vrt. kuvioon 8). Selkolukijoilla oli siis osittaiseksi tulkittavaa sanatietoa enemmän kuin verrokeilla, mutta jotkut sanamerkitykset olivat heille kokonaan vieraita.

Kuviosta 9 selviää, että kaksi sanaverkkoa oli selkolukijoille vaikeampia kuin muut, sillä ärsykesanojen 6 (*vilppi*) ja 9 (*kipu*) oikeita vastinesanoja löydettiin vähemmän. Parhaiten selkolukijat löysivät ärsykesanojen 8 (*johtaja*), 10 (*kuppi*) ja 11 (*ystävä*) vastineita. Mahdollisia syitä oikeiden valintojen määrän vaihteluun tarkastellaan seuraavissa alaluvuissa.

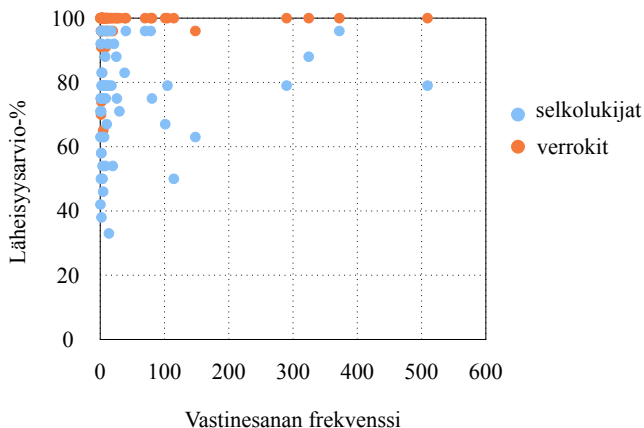


Kuvio 9. Vähintään kaksi vastinesanaa oikein valinneiden selkolukijoiden määrä ärsykesanoittain

4.3 Sanojen yleisyyden ja tuttuuden vaikutus osallistujien tekemiin valintoihin

Vastinesanojen frekvenssit ristiintaulukoitiin läheisyysarvioiden kanssa, jotta voitiin tarkastella sanojen yleisyyden yhteyttä osallistujien tekemiin valintoihin. Vastinesanojen frekvenssit olivat 0,3–509,2 esiintymää miljoonassa, ja hieman yli puolet sanoista oli melko harvinaisia (< 10 esiintymää miljoonassa). Harvinaisempia vastinesanoja oli verrattain paljon, mikä johtui konkreettisten esinesanojen suuresta määrästä: monet niistä olivat sanomalehtikorpuksessa harvinaisia mutta arkisista tilanteista tuttuja. Sanat kuuluivatkin lähes kaikkien tuttuustestaukseen osallistuneiden nuorten aikuisten produktiiviseen sanavarastoon, ja ”käytän sanaa itse” -vaihtoehdon valintojen vastaajien osuus oli eri sanojen kohdalla 67–100 % (mediaani oli 100). Sana-assosiaatiotestin tulosten perusteella sanat olivat tuttuja myös sanatestin verrokkiryhmälle, sillä he valitsivat vastinesanoja frekvensseistä riippumatta oikein, kun taas selkolukijoiden valinnoissa oli hajontaa (ks. kuviota 10).

Kuviosta 10 selviää, että selkolukijat valitsivat frekvenssiltään erilaisia sanoja vaihtelevasti. Tulos viittaa siihen, ettei yleisyys ollut tärkein valintoihin vaikuttava tekijä, vaan myös muilla sanojen vaikeustasoon vaikuttavilla ominaisuuksilla, kuten tuttuudella ja konkreettisuudella, oli kumulatiivinen vaikutus heidän osaamiseensa. Yli 75 % selkolukijoista valitsi hyvin yleisiä sanoja oikein, joten korkea frekvenssi (yli 200 esiintymää miljoonassa) saattoi tukea osaamista. Harvinaisempien vastinesanojen tarkastelu osoittaa, että selkolukijat joko onnistuivat tai eivät onnistuneet valitsemaan niitä, joten pelkkä



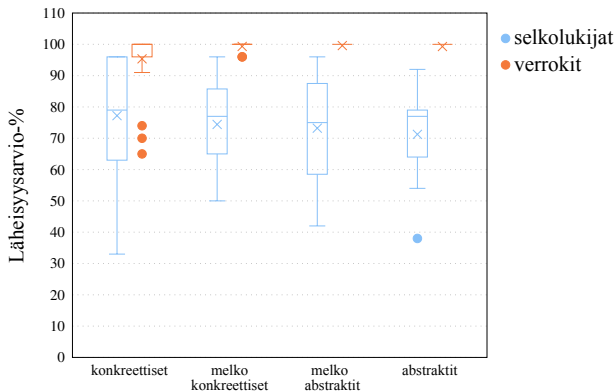
Kuvio 10. Vastinesanojen frekvenssin yhteys läheisyysarvioihin

matala frekvenssi itsessään ei tehnyt sanasta vaikeaa. Tällöin osaamista selittikin tuttuus, eli selkolukijat osasivat valita sanoja, joiden tuttuustuntuma oli todettu tuttuustestauksessa hyvin korkeaksi. ”Käytänkö sanaa itse” saattaakin olla hyödyllinen kysymys selkokieleen sopivan sanaston arvioinnissa. Yli 80 % selkolukijoista osasi melko harvinaisista sanoista varsinkin konkreettisia esinesanoja, mutta mukana oli myös muutama melko abstrakti sana. Vähiten osattujen sanojen joukkoon kertyi lähinnä abstrakteja ja melko abstrakteja sanoja.

4.4 Konkreettisuustason vaikutus osallistujien tekemiin valintoihin

Testissä oli mukana konkreettisia, melko konkreettisia, melko abstrakteja ja abstrakteja sanoja, joille osallistujien tuli löytää oikeita vastineita. Tuloksesta ilmenee, että selkolukijat valitsivat konkreettisimpien sanojen vastinesanoja varmemmin kuin abstraktien, kun taas verrokkit valitsivat pääasiassa oikeita vastinesanoja kaikilla konkreettisuusluokituksen tasoilla (ks. kuviota 11).

Verokkiryhmän läheisyysarvioiden mediaanit olivat kaikilla konkreettisuusluokituksen tasoilla 100 % ja vaihteluvälit olivat pienet, kun taas selkolukijoiden vastauksissa oli enemmän hajontaa ja mediaanit olivat 80 % tai sen alle. Hajontaa selittää testisuoriutumisen vaihtelu yksilötasolla, sillä jotkut



Kuvio 11. Vastinesanojen konkreettisuustason vaikutus osallistujien läheisyysarvioihin⁹

selkolukijoista löysivät vastinesanoja enemmän kuin toiset. Lisäksi selkolukijoiden aineistossa oli mukana kahdeksan nollaksi pisteytettyä sanaverkkoa, joissa valinnat jäivät epäselviksi. Näistä kaksi osui abstraktille sanalle *vilppi* ja yksi sanalle *käsky*. Kolme osui melko abstrakteille sanoille *kipu*, *haju* ja *melu*. Konkreettisissa ja melko konkreettisissa sanoissa nollaksi pisteytettyjä sanaverkkoja oli kummassakin yksi sanoilla *korkki* ja *reitti*, eli konkreettisten vastinesanojen valinta oli tässäkin mielessä sujuvampaa. Voidaankin sanoa, että konkreettisissa sanoissa oli selkolukijoille enemmän yhteistä tarttumapintaa ja suurempi joukko löysi vastinesanoja varmemmin oikein.

Konkreettisia sanoja olivat laskettavat esineet *laatikko*, *hattu*, *sirpale*, *kuppi*, *korkki* ja *allas*, inhimilliset roolinimet *ystävä* ja *johtaja* sekä kasvisana *oksa*. Mielenkiintoisimpia näistä olivat inhimilliset roolinimet, joita selkolukijat osasivat hyvin. *Ystävä* oli testin helpoin sana (ks. taulukon 1 kohtaa 1), mikä ei ole yllättävää, sillä sana on yleinen ja arkikielestä tuttu. Selkolukijat tunsivat hyvin myös *johtaja*-sanan vastineita (ks. taulukon 1 kohtaa 2), joten inhimillisten roolinimien tutuus selkolukijoille voisi olla mielenkiintoinen tutkimuksen kohde.

⁹ Kuviaita 11–12 luetaan seuraavasti: Ruudun yläreuna tarkoittaa yläneljänneistä ja alareuna alaneljänneistä. Ruudun sisällä oleva viiva tarkoittaa mediaania ja rasti keskiarvoa. Janojen päät kuvaavat pienintä ja suurinta arvoa, ja janan ulkopuolella olevat pisteet kuvaavat poikkeavia arvoja.

Taulukko 1. Esimerkkejä osallistujaryhmien valitsemista sanoista^a

	Vastaajat	Ärsyke- sana	Vastinesanat ja vastaajien määrä (%)	Hämäyssanat ja vastaajien määrä (%)
1	selkolukijat	ystävä	kaveri (96), toveri (92), kumppani (96)	vartija (0), tuomari (8), hallitsija (8)
	verrokit	ystävä	kaveri (100), toveri (100), kumppani (100)	vartija (0), tuomari (0), hallitsija (0)
2	selkolukijat	johtaja	päällikkö (96), esimies (75), pomo (96)	isäntä (8), sankari (4), opettaja (21)
	verrokit	johtaja	päällikkö (100), esimies (100), pomo (100)	isäntä (43), sankari (0), opettaja (0)
3	selkolukijat	oksa	havu (58), risu (79), vitsa (63)	latva (54), taimi (46), tolppa (0)
	verrokit	oksa	havu (74), risu (100), vitsa (100)	latva (17), taimi (9), tolppa (0)
4	selkolukijat	korkki	tappi (75), tulppa (79), tukko (46)	kansi (50), luukku (17), läppä (21)
	verrokit	korkki	tappi (70), tulppa (100), tukko (65)	kansi (43), luukku (0), läppä (17)
5	selkolukijat	mutka	käänne (100), kaarre (100), kurvi (100)	kuoppa (0), ryppy (0), lommo (0)
	verrokit	mutka	käänne (88), kaarre (79), kurvi (75)	kuoppa (33), ryppy (17), lommo (8)
6	selkolukijat	kura	muta (79), rapa (71), loka (63)	lumi (17), räntä (58), rae (13)
	verrokit	kura	muta (100), rapa (100), loka (100)	lumi (0), räntä (0), rae (0)
7	selkolukijat	kipu	tuska (63), särky (54), kolotus (42)	kuume (29), tauti (42), sairaus (58)
	verrokit	kipu	tuska (96), särky (100), kolotus (100)	kuume (0), tauti (0), sairaus (4)
8	selkolukijat	haju	lemu (92), käry (79), löyhy (75)	sumu (13), höyry (17), viima (13)
	verrokit	haju	lemu (100), käry (100), löyhy (100)	sumu (0), höyry (0), viima (0)
9	selkolukijat	vilppi	harhautus (38), huijaus (63), petos (54)	pila (50), vahinko (46), toive (21)
	verrokit	vilppi	harhautus (96), huijaus (100), petos (96)	pila (9), vahinko (0), toive (0)
10	selkolukijat	ongelma	vaikeus (79), pulma (92), hankaluus (79)	ohje (4), vinkki (21), opastus (25)
	verrokit	ongelma	vaikeus (100), pulma (100), hankaluus (100)	ohje (0), vinkki (0), opastus (0)

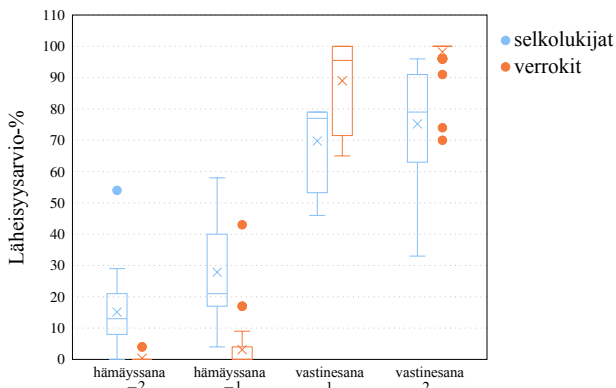
^a Luvut on pyöristetty lähimpään kokonaislukuun.

Konkreettisista sanoista vaikeimpia olivat *oksa* ja *korkki*. On mahdollista, että *oksa*-sarjan ongelmat liittyivät merkityksen tulkinnan vaikeuteen, ja sanaverkko olisi voinut olla vastine- ja hämäyssanojen osalta selkeämpikin (ks. taulukon 1 kohtaa 3). Osallistujat saattoivat tehdä valintoja partonomisen tiedon pohjalta, jolloin he valitsivat kokonaisuuteen kuuluvia osia. Toisaalta vastinesanat *vitsa*, *havu* ja *risu* olivat harvinaisia, mikä saattoi saada osallistujat suosimaan hieman yleisempiä hämäyssanoja *latva* ja *taimi*. Myös *korkki*-sarja osoittautui haastavaksi kummallekin osallistujaryhmälle, sillä vastinesana *tukko* oli vaikea löytää (ks. taulukon 1 kohtaa 4). Tässä sanaverkossa selkolukijat valitsivat oikeiden vastinesanojen ohessa kaikkia hämäyssanoja.

Melko konkreettisten sanojen luokassa olivat kollektiivisana *joukko*, ainesana *kura* sekä spatiaalisiin rakennelmin viittaavat *mutka* ja *reitti*. Sanaverkot olivat selkolukijoille vaikeustasoltaan suunnilleen samanlaisia, mutta parhaiten he tunsivat *mutka*-sanankokoa (ks. taulukon 1 kohtaa 5). Myös *kura*-sanankokoa he päätyivät osittain oikeisiin valintoihin, mutta valitsivat myös hämäyssanoja *räntä*, *lumi* ja *rae* (ks. taulukon 1 kohtaa 6). Selkolukijoiden vastauksissa olikin nähtävissä merkityksen erottelun vaikeuteen viittaavia piirteitä, kun sanojen merkitykset olivat lähellä toisiaan.

Melko abstraktien sanojen luokassa olivat asianimet *melu* ja *haju* sekä elollisista predikoitavaan tilaan viittaava *kipu*, joka oli selkolukijoille näistä kolmesta vaikein (ks. taulukon 1 kohtaa 7). *Kipu*-sanaverkossa selkolukijoiden kolmen yleisimmän valinnan joukkoon päätyikin hämäyssana, *sairaus*. Vastinesana *kolotus* puolestaan oli heille vieras, mikä on odotuksenmukaista, sillä sana on selkokielen vastainen eli harvinainen ja merkitykseltään läpinäkymätön. *Kipu*-sanaverkon vaikeustasoa saattoi nostaa hämäyssanojen semanttinen läheisyys ärsykesanaan, ja selkolukijoiden vastauksissa oli tässäkin viitteitä merkityksen erottelun vaikeudesta. Näin oli myös *haju*-sanaverkossa, jonka vastinesanoja he osasivat hyvin, mutta valintoja osui myös kaikille hämäyssanoille (ks. taulukon 1 kohtaa 8).

Abstrakteja sanoja olivat *kilpailu*, *käskey*, *ongelma* ja *vilppi*, joista *vilppi* oli selkolukijoille vaikein. Sanan vaikeus näkyi yhtenä epäselvänä ja yhtenä tyhjäksi jätettynä vastauspaperina sekä runsaina hämäyssanojen valintoina (ks. taulukon 1 kohtaa 9). *Ongelma* puolestaan oli selkolukijoille abstrakteista sanoista tutuin: sen vastinesanat tunnistettiin suurelta osin oikein, joskin valintoja meni myös kaikille hämäyssanoille. Voidaankin sanoa, että abstraktien sanojen luokassa selkolukijoiden ja verrokkien erot olivat suurimmillaan: verrokkitunnistivat lähes kaikki vastinesanat täysin oikein, kun taas selkolukijat tunnistivat niitä harvemmin. Vaikka selkolukijat valitsivat kaikissa sanaver-



Kuvio 12. Läheisyysarviot ja verrokkisanojen semanttinen yhteys ärsykesanaan

koissa vastinesanoja osittain oikein, valintoja osui myös kaikille hämäyssanoille, joillekin jopa runsaasti.

4.5 Osallistujien kyky valita lähimerkityksisiä vastineita

Selkolukijoiden ja verrokkiryhmän kyky valita vastinesanoja oli samansuuntainen, sillä kummankin ryhmän läheisyysarviot nousivat sanojen ominaisuuspisteiden myötä. Verrokkit kuitenkin valitsivat vastinesanoja varmemmin oikein. Kummallakin osallistujaryhmällä matalimmat läheisyysarviot osuivat ärsykesanan merkityksen kanssa yhteensopimattomiin hämäyssanoihin ja korkeimmat puolestaan lähimerkityksisille sanoille (ks. hämäyssanaa -2 ja vastinesanaa 2 kuviossa 12).

Verrokkit valitsivat lähimerkityksisen vastineen (vastinesana 2) lähes aina ja geneerisen vastineen (vastinesana 1) noin 95 % todennäköisyydellä, eli he tunsivat sanamerkitykset tarkasti. Selkolukijat valitsivat lähimerkityksisen vastineen keskimäärin 80 % todennäköisyydellä ja geneerisen hieman alle 80 % todennäköisyydellä, eli oikean vastinesanan valitseminen oli heille vaikeampaa, mutta hekin tekivät valintoja enimmäkseen oikein. Geneerisen tason vastineet olivat kummallekin osallistujaryhmälle hankalampia kuin samalla hierarkiatasolla olevat vastineet. Verrokkiryhmässä geneerisen tason vastineiden hankaluus näkyi muun muassa haluttomuutena valita näitä sanoja, ja he saattoivat valita sanaverkosta vain kaksi tarkempaa vastinetta ja jättää

kolmannen kokonaan valitsematta. Geneeristen vastineiden valitsematta jättämisen taustalla on oletettavasti sama syy, jonka Pajunen ym. (2015: 173) kohtasivat tutkimuksessaan: verrokkihenkilöt todennäköisesti tunsivat sanojen merkitykset, mutta eivät osanneet ajatella, että geneeristen ja spesifisten vastineiden välillä on merkitysyhteys, koska kouluopetuksessa korostetaan synonymiaa ja hierarkkisista rakenteista puhutaan vähemmän. Geneeriset vastineet olivat siis kummallekin osallistujaryhmälle vaikeampia, ja osallistujat suosivat tarkempia vastineita.

Osallistujaryhmien suurimmat erot liittyivät hämäyssanoihin, sillä verrokot eivät juurikaan valinneet merkitykseltään täysin yhteensopimattomia sanoja, kun taas selkolukijat valitsivat niitä useammin (ks. hämäyssanaa –2 kuviossa 12). On siis mahdollista, että selkolukijat eivät tunteneet sanojen merkityksiä tai he yhdistivät vastinesanoja ärsykesanoihin sattumanvaraisen assosiaatioiden perusteella. Assosiaatioksi tulkittavia valintoja oli esimerkiksi *kuppi*-sarjassa, jossa 13 % selkolukijoista valitsi hämäyssanan *pilli* (”pillillä voidaan juoda kupista”). Vastaavanlaisia virhevalintoja oli myös *allas*-sanaverkossa, jossa 54 % selkolukijoista arveli oikeaksi vaihtoehdoksi hämäyssanaa *putki* (”altaassa voi olla putki”). Näissä assosiaatiot eivät liittyneet ärsykesanan ydinmerkitykseen vaan ennemminkin kielenulkoiseen tietoon ja kokemuksiin (ks. Aitchison 1987: 44), joten vastausten ei voida ajatella mittaavan sanamerkityksen hallintaa. Toisaalta nähtävissä oli myös merkityksen erotteluun liittyviä vaikeuksia varsinkin hämäyssanoissa, joilla oli jokin merkitysyhteys ärsykesanaan: verrokkiryhmällä merkityksen erotteluun liittyviä vaikeuksia oli joissakin sanaverkoissa, mutta selkolukijoilla niitä oli enemmän (ks. hämäyssanaa –1 kuviossa 12).

5 Yhteenvedo ja pohdinta

Tässä artikkelissa on tarkasteltu kehitysvammaisten selkolukijoiden sanamerkitysten hallintaa. Tavoitteena on ollut selvittää, miten tarkasti he tunnistavat sanojen lähimerkityksisiä vastineita ja miten sanojen piirteet vaikuttavat heidän osaamiseensa. Seuraavaksi vedetään yhteen tutkimuksen tulokset ja pohditaan, millaisia näkökulmia ne avaavat selkolukijoiden kielitaitoon ja selkokieleen yleisesti.

Ensimmäinen tutkimuskysymys koski selkolukijoiden sanamerkitysten hallinnan tarkkuutta ja osaamistason vaihtelua. Tulokset osoittavat, että kehitysvammaisten selkolukijoiden sanastonhallinnassa on merkittäviä eroja, eikä

yksilön kielellisestä osaamisesta voidaan tehdä päätelmiä sen perusteella, että hän kuuluu tiettyyn selkokielen kohderyhmään. Ryhmätasolla sen sijaan voidaan tehdä yleistyksiä siitä, millaiset sanat ovat todennäköisimmin ymmärrettäviä mahdollisimman monille. Osaamistason vaihtelu herättääkin kysymyksen selkokielen kohderyhmien määrittelystä, sillä kohderyhmiä määritellään selkokielen tarpeen taustalla olevan syyn sekä tilapäisyyden ja pysyvyyden perusteella. Olisiko kohderyhmien määrittelyssä tarpeellista korostaa myös ulottuvuutta, joka luonnehtii yksilön tarvitseman kielellisen tuen määrää? Kysymys on aiheellinen, sillä tähän tutkimukseen osallistui ihmisiä, jotka kuuluvat samaan selkokielen kohderyhmään mutta joiden kielellisessä osaamisessa on suuria eroja. Leskelä (2019a) hahmottelee selkokieltä kolmella eri vaikeustasolla: helppo taso, perustaso ja vaativa taso. On perusteltua olettaa, että tähän tutkimukseen osallistuneista selkolukijoista jokainen löytäisi näiltä vaikeustasoilta itselleen sopivaa lukemista ja olisi määriteltävissä lukijaksi, joka useimmiten tai ajoittain tarvitsee helpon tason, perustason tai vaativan tason selkokieltä. Jotta tämä voisi toteutua, selkokielen vaikeustasoja tulisi määritellä tarkemmin ja tuottaa tekstejä eri vaikeustasoilla.¹⁰

Toinen tutkimuskysymys nosti tarkastelun kohteeksi selkolukijoiden testisuoriutumiseen vaikuttaneet kielenulkoiset ja kielelliset tekijät. Kielenulkoisista tekijöistä tarkasteltiin selkolukijoiden lukutottumuksia eli lukemisen toistuvuutta ja selko- ja yleiskielen käyttöä. Tulosten perusteella osallistujien lukutottumuksilla ei ollut yhteyttä semanttiseen osaamiseen, mutta koehenkilöitä oli testissä melko vähän, ja he olivat keskenään kovin erilaisia. Tulos saattaisi olla toisenlainen, jos aihetta tutkittaisiin suuremmalla ja tarkkarajaisemmalla otannalla, jossa yksittäisten osallistujien vaikutus kokonaisuuteen ei olisi niin voimakas. Laajemmassa tutkimuksessa voitaisiin myös tarkastella useampia semanttiseen osaamiseen mahdollisesti vaikuttavia kielenulkoisia tekijöitä. Tämän tutkimuksen hedelmällisimmäksi anniksi nouseekin selkolukijoiden lukutottumusten ja semanttisen osaamisen moninaisuuden kuvaaminen: osaamistason vaihtelu kertoo ryhmänsisäisistä eroista, joita on tarpeellista tutkia tulevaisuudessa myös muiden kohderyhmien osalta.

Selkolukijoiden testisuoriutumiseen vaikuttaneista kielellisistä tekijöistä keskeisimpiä olivat sanaston konkreettisuus, yleisyys ja tuttuus, joita tarkasteltiin synonymiaan perustuvalla testillä. Koska suurin osa selkolukijoista tunnisti sanaverkoista ainakin osan vastinesanoista oikein, voidaan pitää toden-

¹⁰ Eri vaikeustasoilla kirjoitettuja tekstejä on jo olemassa, mutta vaikeustasoja ei ole määritelty tarkasti eikä niiden ilmoittamiseen ole vakiintunutta käytäntöä.

nettuna, että perustason selkokielessä synonyymejä voidaan käyttää, kun ne valitaan harkiten (vrt. Leskelä 2019b: 137). Selkokielessä synonyymien käytön tulee olla johdonmukaista, luontevaa ja maltillista (Selkokeskus 2018: 10), ja tämän tutkimuksen tulokset tarjoavat suuntaviivoja näiden periaatteiden toteuttamiseen. Kielelliseltä osaamiseltaan eritasoiset selkolukijat tunnistivat todennäköisimmin vastinesanoja, joille kertyi mahdollisimman monta selkokielelle tyypillistä piirrettä. Konkreettisuudesta oli hyötyä, sillä konkreettisten sanojen vastineita tunnistettiin paremmin kuin abstraktien. Samalla hierarkiatasolla olevat vastinesanat tunnistettiin helpommin kuin geneeriset vastineet, mutta selkolukijoilla oli enemmän merkityksen erottelun vaikeuksia kuin verrokeilla, jolloin hämäyssanoja saatettiin erehtyä luulemaan vastinesanoiksi. Tulokset antavat viitteitä siitä, että läpinäkyvämmät vastinesanat olivat selkolukijoille hankalampia tunnistaa. Korkea frekvenssi saattoi liittyä sanojen parempaan tunnistamiseen, mutta selkolukijat tunsivat myös matalafrekvenssisia sanoja, jotka olivat arkikielestä tuttuja.

Koska selkolukijoiden kielellisen osaamisen erot ovat suuria, selkokielen sanaston tuttuuden arvioinnissa on tyydyttävä epävarmuuteen ja etsittävä sanoja, jotka ovat tuttuja mahdollisimman monille. Usein selkokirjoittajat tekevät sanastoon liittyviä valintoja oman tuntumansa perusteella ja saavat palautetta valintojen onnistuneisuudesta vasta jälkikäteen. Tässä tutkimuksessa sanoille tehtiin tuttuustestaus, josta selvisi, että selkolukijat osasivat sanoja, jotka kuuluivat tuttuustestaajien aktiiviseen sanavarastoon. Voisiko subjektiivinen frekvenssi siis olla yksi lähestymistapa sanojen tuttuuden arviointiin? Tällöin selkokirjoittajat voisivat pohtia subjektiivista frekvenssiä kahdella kysymyksellä, joista ensimmäinen olisi tuttuutta kartoittava ”kuinka usein käytän sanaa itse” ja toinen olisi yleisyyttä pohtiva ”kuinka usein kohtaan sanan ja missä tilanteissa” (vrt. Balota ym. 2001).¹¹ Vaikka selkokirjoittaja lähestyisikin näitä kysymyksiä oman tuntumansa perusteella, voidaan olettaa, että aktiivisessa käytössä olevat sanat, joita havaitaan ja tuotetaan usein monissa eri konteksteissa, ovat tuttuja mahdollisimman monille.

¹¹ Balota ym. (2001: 641) arvioivat subjektiivista frekvenssiä eli sanan käytön ja kohtaamisen toistuvuutta seitsenportaisella asteikolla: 1 = ”en koskaan”, 2 = ”kerran vuodessa”, 3 = ”kerran kuukaudessa”, 4 = ”kerran viikossa”, 5 = ”joka toinen päivä”, 6 = ”kerran päivässä” ja 7 = ”useita kertoja päivässä”.

Lähteet

- Aitchison, Jean. 1987. *Words in the mind: An introduction to the mental lexicon*. Oxford: Blackwell.
- Balota, David A. & Pilotti, Maura & Cortese, Michael J. 2001. Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition* 29. 639–647. DOI: 10.3758/BF03200465.
- Bogaards, Paul. 2000. Testing L2 vocabulary knowledge at a high level: The case of the Euralex French tests. *Applied Linguistics* 21. 490–516. DOI: 10.1093/applin/21.4.490.
- Brysbaert, Marc & Warriner, Amy B. & Kuperman, Victor. 2013. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46. 904–911. DOI: 10.3758/s13428-013-0403-5.
- Chinn, Deborah & Homeyard, Claire. 2017. Easy read and accessible information for people with intellectual disabilities: Is it worth it? A meta-narrative literature review. *Health Expectations* 20. 1189–1200. DOI: 10.1111/hex.12520.
- Di Blasi, Francesco & Buono, Serafino & Città, Santina & Costanzo, Angela & Zoccolotti, Pierluigi. 2018. Reading deficits in intellectual disability are still an open question: A narrative review. *Brain Sciences* 8. DOI: 10.3390/brainsci8080146.
- Fajardo, Inmaculada & Ávila, Vicenta & Ferrer, Antonio & Tavares, Gema & Gómez, Marcos & Hernández, Ana. 2014. Easy-to-read texts for students with intellectual disability: Linguistic factors affecting comprehension. *Journal of Applied Research in Intellectual Disabilities* 27. 212–225. DOI: 10.1111/jar.12065.
- Greidanus, Tine & Beks, Bianca & Wakely, Richard. 2005. Testing the development of French word knowledge by advanced Dutch- and English-speaking learners and native speakers. *The Modern Language Journal* 89. 221–233. DOI: 10.1111/j.1540-4781.2005.00276.x.
- Greidanus, Tine & Nienhuis, Lydius. 2001. Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *The Modern Language Journal* 85. 567–577. DOI: 10.1111/0026-7902.00126.
- Hartley, Sigan L. & MacLean, William E. 2006. A review of the reliability and validity of Likert-type scales for people with intellectual disability. *Journal of Intellectual Disability Research* 50. 813–827. DOI: 10.1111/j.1365-2788.2006.00844.x.
- Honko, Mari. 2013. *Alakouluikäisten leksikaalinen tieto ja taito: Toisen sukupolven suomi ja S1-verrokit*. Tampere: Tampere University Press.
- Horiba, Yukie. 2012. Word knowledge and its relation to text comprehension: A comparative study of Chinese- and Korean-speaking L2 learners and L1 speakers of Japanese. *The Modern Language Journal* 96. 108–121. DOI: 10.1111/j.1540-4781.2012.01280.x.

- Juusola, Markku. 2019. *Selkokielen tarvearvio 2019*. Helsinki: Selkokeskus, Kehitysvammaliitto.
- Leskelä, Leelaura. 2019a. Helppoa vai vaativampaa selkokieltä: Selkokielen mittaaminen ja vaikeustasot. *Puhe ja kieli* 39. 367–393. DOI: 10.23997/pk.75679.
- 2019b. *Selkokieli: Saavutettavan kielen opas*. Helsinki: Kehitysvammaliitto.
- Leskelä, Leelaura & Kulkki-Nieminen, Auli. 2015. *Selkikirjoittajan tekstilajit*. Helsinki: Kehitysvammaliitto.
- Leskelä, Leelaura & Lindholm, Camilla. 2012. Näkökulmia kielellisesti epäsymmetriseen vuorovaikutukseen. Teoksessa Leskelä, Leelaura & Lindholm, Camilla (toim.), *Haavoittuva keskustelu: Keskustelunalyttisia tutkimuksia kielellisesti epäsymmetrisestä vuorovaikutuksesta*, 12–31. Helsinki: Kehitysvammaliitto.
- Meara, Paul. 2009. *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: John Benjamins.
- Miller, George A. 1998. Nouns in WordNet. Teoksessa Fellbaum, Christiane (toim.), *WordNet: An electronic lexical database*, 23–46. Cambridge, MA: MIT Press.
- Milton, James. 2009. *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters. DOI: 10.21832/9781847692092.
- MOT Kielitoimiston sanakirja*. E. p. Helsinki: Kotimaisten kielten keskus & Kielikone.
(<https://mot-kielikone-fi.libproxy.tuni.fi/mot/uta/netmot.exe>). (Luettu 2020-02-19).
- MOT Synonyymisanakirja*. E. p. Helsinki: Kielikone. (Toim. Harri Jäppinen).
(<https://mot-kielikone-fi.libproxy.tuni.fi/mot/uta/netmot.exe>). (Luettu 2020-02-19).
- Nation, I. S. P. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nyky-suomen sanakirja*. 1951–1961. Helsinki: WSOY.
- Pajunen, Anneli & Itkonen, Esa & Vainio, Seppo. 2015. Sanamerkityksen hallinta nuorilla aikuisilla. *Virittäjä* 119(2). 160–188.
- Pajunen, Anneli & Virtanen, Patrik. 2002. Helsingin Sanomat 2000–2001 (= HS2000). (Morfosyntaktisesti analysoitu hakutietokanta, 31 miljoonaa sanamuotoa).
- Pappert, Sandra & Bock, Bettina M. 2019. Easy-to-read German put to the test: Do adults with intellectual disability or functional illiteracy benefit from compound segmentation? *Reading & Writing* 33. 1105–1131. DOI: 10.1007/s11145-019-09995-y.
- Qian, David D. 1999. Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review* 56. 282–307. DOI: 10.3138/cmlr.56.2.282.
- Qian, David D. & Schedl, Mary. 2004. Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing* 21. 28–52. DOI: 10.1191/0265532204lt273oa.

- Ratz, Christoph & Lenhard, Wolfgang. 2013. Reading skills among students with intellectual disabilities. *Research in Developmental Disabilities* 34. 1740–1748. DOI: 10.1016/j.ridd.2013.01.021.
- Read, John. 1993. The development of a new measure of L2 vocabulary knowledge. *Language Testing* 10. 355–401.
- 2004. Plumbing the depths: How should the construct of vocabulary knowledge be defined? Teoksessa Bogaards, Paul & Laufer, Batia (toim.), *Vocabulary in a second language: Selection, acquisition and testing*, 209–227. Amsterdam: John Benjamins.
- 2012. Piloting vocabulary tests. Teoksessa Fulcher, Glenn & Davidson, Fred (toim.), *The Routledge handbook of language testing*, 307–320. London: Routledge.
- Schmitt, Norbert. 2010. *Researching vocabulary: A vocabulary research manual*. London: Palgrave Macmillan.
- 2014. Size and depth of vocabulary knowledge: What the research shows. *Language Learning* 64. 913–951. DOI: 10.1111/lang.12077.
- Schmitt, Norbert & Ng, Janice W. C. & Garras, John. 2011. The word associates format: Validation evidence. *Language Testing* 28. 105–126. DOI: 10.1177/0265532210373605.
- Schoonen, Rob & Verhallen, Marianne. 2008. The assessment of deep word knowledge in young first and second language learners. *Language Testing* 25. 211–236. DOI: 10.1177/0265532207086782.
- Selkokeskus. 2018. *Selkomittari*. Helsinki: Kehitysvammaliitto. (https://selkokeskus.fi/wp-content/uploads/2018/10/SELKOMITTARI_2018_11.10.18.pdf). (Luettu 2020-09-24).
- Sutherland, Rebekah J. & Isherwood, Tom. 2016. The evidence for easy-read for people with intellectual disabilities: A systematic literature review. *Journal of Policy and Practice in Intellectual Disabilities* 13. 297–310. DOI: 10.1111/jppi.12201.
- Tervola, Maija & Vainio, Seppo & Häikiö, Tuomo. 2020. Potilassanaston hallinta terveydenhuoltoalalla: Maahanmuuttajataustaisten lääkäreiden ja asioimistulkkiopiskelijoiden sanamerkityksen tarkkuus. *Virtittäjä* 124(1). 4–32. DOI: 10.23982/vir.70136.
- Vinni, Irja. 1998. *Tekstistä selkoa: Kehitysvammaisten aikuisten selkokielisten ja yleiskielisten tekstien ymmärtäminen*. Helsinki: Kehitysvammaliitto.
- Zhang, Dongbo & Koda, Keiko. 2017. Assessing L2 vocabulary depth with word associates format tests: Issues, findings, and suggestions. *Asian-Pacific Journal of Second and Foreign Language Education* 2. 1–30. DOI: 10.1186/s40862-017-0024-0.

Yhteystiedot:

Idastiina Valtasalmi

Suomen kieli

PL 300

33014 Tampereen yliopisto

sähköposti: idastiina(dot)valtasalmi(at)tuni(dot)fi