

# ***suuri, iso vai kookas? tärkeä, keskeinen vai merkittävä?***

## **Miten ensikielisyys vaikuttaa lähisynonymisten adjektiivien valintaan kaunokirjallisuuskontekstissa**

**Niina Kekki**  
**Turun yliopisto**

**Ilmari Ivaska**  
**Turun yliopisto**

### **Abstrakti**

Tarkastelemme tässä artikkelissa monimuuttuja-sekamallin avulla, minkä adjektiivin lähisynonymisista pesueista *suuri/iso/kookas* ja *tärkeä/keskeinen/merkittävä* S2- ja S1-puhujat valitsevat kaunokirjallisessa tekstikontekstissa. Edistyneiden kielenoppijoiden lähisynonymien käytön ja opettamisen ongelmat on tunnistettu jo varhain. Lähisynonymivalinnat on havaittu toisen kielen oppimisen tutkimuksessa yhdeksi yleisimmistä leksikaalisen tason syistä sille, että vielä edistyneidenkin kielenoppijoiden kielenkäyttö voi vaikuttaa kontekstille epätyypilliseltä. Aiempi tutkimus on lisäksi osoittanut, että kaunokirjallisuuden lukeminen mahdollistaa sattumanvaraisen sanaston oppimisen niin ensikielillä kuin toisen/vieraan kielen puhujilla.

Käyttöpohjaisen näkemyksen mukaan kieli ja sen oppiminen pohjautuvat luonnollisen kielen käyttöön. Kielenkäytön ytimessä ovat konstruktiot eli muoto-merkitys-käyttö-kombinaatiot, jotka muodostavat toisiinsa kytkeytyvän verkoston. Vertaamme S2- ja S1-puhujien adjektiivivalintoja niiden tyypilliseen käyttöön konstruktionäkökulmasta eli selvittämme, mitkä muuttujat selittävät sanavalintaa konstruktiossa, jossa adjektiivi toimii pääsanansa määrittäneenä. Varsinainen aineisto on kerätty kyselytutkimuksella, mutta käytämme valintaa selittävinä taustamuuttujina laajaa kirjoa erilaisia taustamuuttujia. Näihin kuuluu valinnanalaisen konstruktion ominaisuuksia (mm. sijamuoto), kielenkäyttäjäkohtaisia muuttujia (mm. kielitausta ja lukutottumukset) sekä useita korpuksista johdettuja, tarkastelunalaisen konstruktion tyypillistä käyttöä kuvaavia muuttujia (mm. adjektiivien ja niiden pääsanojen yhteisesiintymien suhteelliset vahvuudet kielen eri rekistereissä).

Tutkimuksen tuloksena havaittiin, että suomeksi lukevat S2-puhujat valitsevat adjektiiveja alkuperäisen kaunokirjallisen kontekstin mukaisemmin kuin muilla kielillä lukevat. S1-puhujilla ei vastaavaa vaikutusta ole. Konstruktioiden osalta tutkimus vahvistaa aiempaa tietoa siitä, että L1- ja edistyneet L2-puhujat hahmottavat konstruktioita huomattavan samantapaisesti. Tarkasteltujen kollokaatioiden tekstilajipreferenssillä ei ole suurta merkitystä korkeakoulutetuilla vastaajilla, mahdollisesti koska kyselyaineisto mittaa enemmän sanaston reseptiivistä kuin produktiivista hallintaa. Tulokset osoittavat, että tilastollisena menetelmänä käytetty monimuuttujainen sekamalli sopii tutkimuksessa tarkasteltavan ilmiön selittämiseen. Malli toimii molemmissa pesueissa hyvin, mutta kiinnostavasti se selittää selvästi paremmin *suuri*-pesueen adjektiivivalintojen vaihtelua. Synonymiaa käsittelevissä korpustutkimuksissa on usein tarkastelussa vain yksi lähisynonymipari tai -pesue, joten tämän tutkimuksen pohjalta voi esittää, että synonymiaa käsittelevässä tutkimuksessa yhden lähisynonymipesueen tarkastelun pohjalta ei

voi kovin luotettavasti tehdä synonymiaa yleisesti koskevia tulkintoja. Lisäksi tutkimuksen tulokset kannustavat käyttämään kaunokirjallisuutta aikuisten kielenopetuksessa.

**Avainsanat:** kielenoppiminen, suomi, suomi toisena kielenä, synonymia, monimuuttujasekamalli, kaunokirjallisuus

## 1 Johdanto

Jos kirjoittaa internetin hakukoneeseen kysymyksen ”mitä eroa on sanoilla *iso* ja *suuri*”, saa hakutuloksina keskustelupalstaketjuja, joissa kielenkäyttäjät pohtivat näiden sanojen semantiikkaa. Vaikka ensikielinen puhuja osaa yleensä käyttää tällaisia lähisynonyymisia<sup>1</sup> sanoja intuitiivisesti, niiden välisen eron sanallistaminen voi olla vaikeaa. Kielenoppijalle puolestaan voi tuottaa päänsäivä valita ilmaisu lähisynonyymien pesueesta, jossa ilmausten välillä on vain hyvin hienovaraisia merkityseroja. Ilmiön hankaluudesta huolimatta lähisynonyymien tunteminen mielletään yhdeksi sujuvan kielenkäytön tunnuspiirteeksi: ne osoittavat tekstilajien ja kielirekisterien tuntemusta, ja niiden avulla kieleen saa vaihtelua ja oikeanlaista sävyä (Edmonds & Hirst 2002; Jarvis 2013; Danglli & Abazaj 2014).

Edistyneiden kielenoppijoiden synonyymisten ilmausten käytön ja opettamisen ongelmat on tunnistettu jo varhain (esim. Martin 1984), ja lähisynonyymivalinnat on havaittu yhdeksi yleisimmistä leksikaalisen tason syistä sille, että vielä edistyneidenkin kielenoppijoiden kielenkäyttö voi vaikuttaa kontekstille epätyypilliseltä (Hasselgren 1994: 245). Vaikka kielenoppijoiden vaikeudet lähisynonyymien omaksumisessa ovat hyvin tiedossa, valtaosa synonyymiatutkimuksesta on kuitenkin keskittynyt ensikielisten (englannin- tai muiden indoeurooppalaisten kielten puhujien) kielenkäyttöön (Liu & Zhong 2016: 242).

Epäodotuksenmukaisesti käytetyt lähisynonyymit saattavat vääristää merkityksiä tai tuottaa epätoivottuja sävyeroja (Danglli & Abazaj 2014: 632). Sanan osaamiseen ei riitä, että tietää sanan määritelmän tai käännöksen eri kielillä, vaan osaaminen vaatii myös tietoa siitä, millaisia ovat sanan eri muodot, minkä sanojen kanssa se tyypillisesti esiintyy ja missä konteksteissa sitä on sopivaa käyttää (Alderson ym. 2015: 101). Kaunokirjallisuuden lukemista pidetään yleisesti tehokkaana tapana oppia kieltä ja erityisesti uutta sanastoa (ks. esim. Krashen 1993; Kim 2004), oli kyse sitten ensikielestä tai myöhemmin opittavasta kielestä. Aiempi tutkimus on osoittanut, että kaunokirjallisuuden lukeminen mahdollistaa sattumanvaraisen sanaston oppimisen niin ensikielisillä kuin toisen/vieraan kielen puhujilla (esim. Reynolds 2015).

Tutkimme tässä artikkelissa määrällisin keinoin sitä, minkä lähisynonyymisen adjektiivin suomea ensikielensä (S1) ja edistyneellä tasolla<sup>2</sup> toisena kielenä (S2) puhuvat valitsevat kaunokirjallisessa tekstikontekstissa sekä millainen yhteys kaunokirjallisuuden lukemisella on synonyymisen adjektiivin valintaan aikuisilla S2- ja S1-puhujilla. Tarkastelun kohteena on kaksi lähisynonyymipesuetta: adjektiivi *suuri* ja sen lähisynonyymit *iso* ja *kookas* sekä adjektiivi *tärkeä* ja sen lähisynonyymit *keskeinen* ja *merkittävä*. Nämä ovat tutkimuksessa käytetyn kaunokirjallisuuskorpuksen yleisimmät synonyymiset adjektiivipesueet (ks. tarkemmin luku 3.1).

<sup>1</sup> Koska absoluuttista synonymiaa pidetään yleisesti kielenkäytölle epätodennäköisenä (esim. Cruse 1986; Lyons 1968), käytämme tutkimuksessa termiä *lähisynonyymi* kuvamaan sanoja, jotka ovat merkitykseltään ainakin osin niin samankaltaisia, että kielenkäyttäjät ymmärtävät niiden viittaavan samaan asiaan.

<sup>2</sup> Eurooppalaisen viitekehyksen taitotasolla B–C.

Esimerkiksi monelle suomenoppijalle tuttu Kielitoimiston sanakirja (jatkossa KS, 2022) ei erottele näiden pesueiden variantteja toisistaan tyylillisten tai affektiivisten piirteiden perusteella. *tärkeä* mainitaan sekä *keskeisen* että *merkittävän* synonyyminä. *keskeinen* eroaa merkitykseltään kahdesta muusta pesueen adjektiivista lokatiivisissa yhteyksissä (*keskeinen sijainti*), ja on siten polyseemisempi kuin kaksi muuta pesueen tarkastelunalaista varianttia. KS (2022) esittelee *suuri*-adjektiivin variantit saman ensisijaisen esimerkin avulla: ”*suuri* ja vahva mies; *iso* mies; *kookas* mies”. Sekä *suurella* että *isolla* on sanakirjan mukaan polyseemistä käyttöä vaikutukseltaan huomattavien tai merkittävien asioiden kuvaamisessa, mitä ei esiinny *kookkaan* määritelmässä. Suomenoppijoiden apuvälineeksi kehitetty korpuspohjainen ConLexis-sanakirja (2013) tarkentaa yleisanakirjasta löytyviä tietoja esittelemällä, mitkä ovat *suuren* ja *ison* tyypillisiä kollokaatteja. *suuren* kollokaatit se erittelee abstraktimmiksi ja *ison* konkreettisemmiksi. *kookas* on ConLexiksen mukaan huomattavasti harvinaisempi kuin *iso* ja *suuri*. Näistä tarkennuksista huolimatta suomenoppijoiden käyttämien sanakirjojen sana-artikkelit saattavat jättää mielikuvan, että osa tai kaikki lähisynonyymeistä ovat useimmissa tilanteissa keskenään vaihtoehtoisia.

Vertaamme S2- ja S1-puhujien adjektiivivalintoja niiden tyypilliseen käyttöön konstruktionäkökulmasta eli pyrimme selvittämään, mitkä muuttujat selittävät sanavalintaa konstruktiossa, jossa adjektiivi toimii pääsanansa määrittäneenä (esim. *tärkeä/keskeinen/merkittävä kokous*, *suuri/iso/kookas hankinta*). Varsinainen aineisto on kerätty kyselytutkimuksella, mutta käytämme valintaa mahdollisesti selittävinä taustamuuttujina laajaa kirjoa erilaisia taustamuuttujia, kuten valinnanalaisen konstruktion ja sen tyypillisen käytön ominaisuuksia sekä kielenkäyttäjäkohtaisia muuttujia. Näihin kuuluvat S1- ja S2-puhujien kaunokirjallisuuden lukeneisuutta tarkastelevat muuttujat. Tarkastelemme muuttujien vaikutusta monimuuttuja-sekamallin avulla, mikä mahdollistaa useiden keskenään erilaisten muuttujien vaikutuksen sekä niiden välisten suhteiden samanlaisen tarkastelun.

Vertaamme tutkimuksessa S1- ja S2-puhujien kielenkäyttöä. Emme kuitenkaan arvota kielimuotoja suhteessa toisiinsa emmekä näe, että olisi olemassa jonkinlaista ”täydellistä”, ensikielistä kielimuotoa, jota kohti kielenoppija voisi pyrkiä (ks. esim. Deshors ym. 2018). Aiemmassa synonymiatutkimuksessa on huomattu, että ensikieliset puhujat eivät aina valitse samaa synonyymia samaan kontekstiin (Liu 2013). Enemmistön tekemä valinta on siitä huolimatta yleensä paras indikaattori ilmaisun kontekstisidonnaisesta merkityksestä (Liu & Zhong 2016: 244). Sovellamme siis tutkimuksessamme kontrastiiviselle oppijankielen tutkimukselle tyypillistä kuvailevaa lähestymistapaa, jossa S1- ja S2-puhujien erojen tarkastelun tarkoitus on kuvailla kielimuotoja ja niissä esiintyvää tyypillistä vaihtelua, ei arvioida niitä suhteessa toisiinsa tai ottaa kantaa yksittäisten ilmausten tai valintojen norminmukaisuuteen (Granger 2015).

## 2 Tutkimuksen teoreettiset lähtökohdat

### 2.1 Kaunokirjallisuuden lukemisen vaikutus sanaston oppimiseen

Lähestymme kieltä ja sen oppimista tässä artikkelissa käyttöpohjaisen näkemyksen mukaan, eli näemme, että sekä kieli itsessään että sen oppiminen pohjautuvat kielen käyttöön (esim. Tomasello 2003). Käyttöpohjaisessa näkökulmassa sosialisatio on merkittävässä osassa kielenoppimista: osallistumalla vuorovaikutukseen kielen puhujien ja kieliympäristön kanssa oppija tulee osaksi kieliyhteisöä. Kielen oppiminen tarkoittaa siis pohjimmiltaan sitä, että oppija oppii tapoja tulla toimeen (kohdekielisen) maailman ja

sen merkitysten kanssa. (van Lier 2000: 246–247.) Kaunokirjallisuuden lukeminen voi monin tavoin toimia porttina kohdekielisen maailman ymmärtämiseen ja siitä osalliseksi tulemiseen (Kekki ym. 2023). Yleisesti uskotaan, että kaunokirjallisuuden ekstensiivinen eli omaehtoinen ja -tahtinen lukeminen kasvattaa niin ensikielisen kuin kielenoppijankin sanavarastoa tarjoamalla lukijalle mielekkään kontekstin suuremman sanamäärän kohtaamiseen, kuin ohjatussa opetustilanteessa olisi mahdollista (esim. Pietilä & Merikivi 2014; Reynolds 2015).

Sanojen omaksumista kaunokirjallisuutta lukemalla on tutkittu esimerkiksi mitaamalla, kuinka hyvin kielenoppijat oppivat kaunokirjallisessa tekstissä esiintyviä tai tekstiin lisättyjä tekaistuja (*nonce*) sanoja. Tulokset näistä tutkimuksista osoittavat, että sanavarasto voi kasvaa kaunokirjallisuutta lukemalla ainakin jonkin verran (esim. Waring & Nation 2004; Nation 2018; Reynolds & Ding 2022). Uusien sanojen omaksumisen todennäköisyyteen vaikuttaa muun muassa se, kuinka merkityksellisessä kontekstissa uusi sana tekstissä esiintyy (Webb 2008), kuinka usein ja tasapuolisesti sana esiintyy eri teksteissä sekä missä määrin uusi sana muistuttaa jotakin sanaa oppijan muissa osaamisissa kielissä (Reynolds & Ding 2022).

Ensikieliset aikuiset näyttävät oppivan uusia sanoja kaunokirjallisuutta lukemalla tehokkaammin kuin kielenoppijat (Reynolds 2015: 121). Waringin ja Takakin (2003) tutkimukseen osallistuneet englantia vieraana kielenä -oppijat mielsivät tekaistun sanan olevan usein synonyyminen jollekin heidän jo osaamalleen (opittavan kielen) sanalle. Tämä antaa viitteitä siitä, että kaunokirjallisuuden lukeminen voi kasvattaa lukijan synonyymisten sanojen tuntemusta. Tekaisujen sanojen oppimista tarkastelevissa tutkimusasetelmissa luettavat tekstit ovat kuitenkin usein lyhyitä ja niiden lukemista mitataan laboratoriomaisissa olosuhteissa (Reynolds 2022). Kysymykseksi jää, millaisia pitkäaikaisvaikutuksia kokonaisten romaanien omaehtoisella lukemisella on sanavarastoon.

Monikielisten aikuislukijoiden kohdalla mielenkiintoinen kysymys on, millaisia ovat monella kielellä lukemisen vaikutukset, esimerkiksi millä tavoin vieraalla kielellä lukeminen vaikuttaa ensikielen hallintaan. Auttaako vieraalla kielellä lukeminen vain kohdekielen oppimisessa vai voiko se myös tukea (tai haitata) ensikielen taitoa vielä aikuisiälläkin? Yhä useammat suomea ensikielensä puhuvat nuoret lukevat vain englanniksi (Bono 2023), mikä on herättänyt opettajissa huolta ensikielenä puhuttavan suomen rapautumisesta. Esimerkiksi lukioikäisten unkarilaisnuorten vieraan kielen (englanti, ranska ja/tai venäjä) oppimisen vaikutusta ensikielen kirjoitustaidon kehittymiseen tutkineet Istvan Kecskes ja Tünde Papp (2000: 19, 30) esittävät pitkittäistutkimuksensa pohjalta, että vieraan kielen osaaminen edistyneellä tasolla vaikuttaa ensikielen käytön tapaan. Vaikutus näkyy ensisijaisesti ensikielen hienosyisemmässä käytössä, kuten monipuolisempien lauserakenteiden ja selektiivisemmän sanaston hallinnassa. Vieraan kielen oppiminen vaikuttaa ensikieleen kuitenkin vain, jos oppimisprosessi on tarpeeksi intensiivinen ja oppijan oma motivaatio korkea.

Prosessit, jotka mahdollistavat uusien sanojen oppimisen kaunokirjallisuutta lukemalla, ovat edelleen pääasiassa tuntemattomia (Pietilä & Merikivi 2014: 29). Tutkimus on lisäksi keskittynyt nuorempiin ja usein ensikielisiin lukijoihin, joten erityisesti monikielisten aikuisten osalta tutkimusta on vain vähän. Tutkimuksissa, joissa sanaston oppimista testataan viiveellä lukutapahtuman jälkeen, on huomattu, että heti lukemisen jälkeen muistetut uudet sanat unohtuvat usein nopeasti (esim. Waring & Takaki 2003). Monet sanaston satunnaista omaksumista käsittelevät tutkimukset käsittelevät ja testaavat opittavia sanoja yksittäisinä yksiköinä. Reynolds (2015: 124) huomauttaakin, että sanastonhallintaa käsittelevissä tutkimuksissa pitäisi kiinnittää enemmän huomiota kohdesa-

noja ympäröiviin, toistuviin malleihin. Omassa asetelmassamme pyrimme hahmottamaan konstruktio pohjaisen lähestymistavan avulla, onko kaunokirjallisuuden lukeneisuudella pitkäaikaista vaikutusta lähisynonymian sanaston käyttöön, ja kuinka muulla kuin omalla ensikielellä lukeminen vaikuttaa tilanteeseen.

## 2.2 Konstruktio pohjainen näkökulma lähisynonymiaan ja sen oppimiseen

Tämän artikkelin tutkimusasetelma perustuu ajatukseen kielen konstruktio pohjaisuudesta eli siitä, että kielen yksiköissä muoto, merkitys ja käyttö ovat erottamattomia (Goldberg 2003). Erilajaiset rakenteet morfeemeista lausetyypeihin ja sitäkin laajempiin yhdistelmiin voidaan mieltää konstruktioiksi. Esimerkiksi adjektiivimäärittteen sisältävä substantiivilauseke on konstruktio, jonka käyttöön keskitymme tässä tutkimuksessa. Aiempi tutkimus antaa viitteitä siitä, että sekä ensikieliset puhujat että kielenoppijat hahmottavat tällaisia muoto-merkitys-käyttö-kombinaatioita eli konstruktioita huomattavan samantapaisesti (esim. Gries & Wulff 2005; Ivaska & Bernardini 2020). Ronald Langackerin juurtumisteorian (Langacker 1987: 59–60) mukaan mitä useammin konstruktioa käytetään, sitä automaattisemmaksi sen prosessointi muodostuu ja sitä hyväksyttävämpi se on puhujan intuitiolle. Vaikka pelkkä altistuminen kielelle ei riitä selittämään kielen oppimisen prosessia (Ellis 2008: 375), moniulotteinen tilastollinen analyysi kielenkäyttäjien synonyymien konstruktioiden käytöstä tuottaa kielenoppimisen ymmärtämisen kannalta tärkeää tietoa siitä, millaisia valintamahdollisuuksia muoto-merkitys-käyttö-yhdistelmien käytössä on.

Lähisynonymian tutkimus lähtee yleensä liikkeelle synonyymien ilmausten välisen erojen selvittämisestä. Oletus on, että jos kielessä on käytössä kaksi merkitykseltään samankaltaista muodoltaan poikkeavaa ilmausta, ne on otettu käyttöön erilaisten strategioiden seurauksena ja samankaltaisuudesta huolimatta eroavat jollain lailla merkityksensä ja käyttönsä osalta (De Jonge 1993: 253). Erottavat piirteet voivat olla kielensisäisiä tai -ulkoisia, ja erojen suuruus vaikuttaa siihen, että jotkin lähisynonyymiset ilmaukset ovat synonyymisempiä kuin toiset (Cruse 1986: 267).

Kielensisäisistä piirteistä lähisynonyymivalintaa voi ohjata esimerkiksi tekstilaji: Jarmo Jantusen (2001: 172) tarkastelemassa akateemisessa tekstikorpuksessa käytetään melko tasaisesti *tärkeä*-pesueen monia variantteja, kun taas kaunokirjallisuuskorpuksessa *tärkeä* on huomattavasti kaikkia muita variantteja yleisempi. Bob De Jongen (1993: 534) italian lähisynonyymisiä *parere* ja *sembrare* ('näyttää'; 'vaikuttaa') -verbien käyttöä tarkastelleen tutkimuksen mukaan verbit eivät ole mielivaltaisesti vaihtoehtoisia tai ainoastaan kielenkäyttäjän preferenssiin perustuvia, vaan niiden välillä on selkeitä funktionaalisia eroja. Tiheyttä ilmaisevien suomen kielen adjektiivien polysemiaa tarkastellut Anni Jääskeläinen (2023) havaitsi puolestaan, että *tiheän* ja *tiuhan* polyseemiset merkitykset ovat sidoksissa niiden käyttöön substantiivin määrite- tai adverbialikonstruktioissa. Kiinnostavasti siis ”jo yhden sanan eri polyseemisten merkitysten välillä voi olla suuria eroja esiintymiskonstruktioissa” (Jääskeläinen 2023: 399).

Korpuslingvistiikan menetelmät ovat mahdollistaneet viime vuosikymmeninä uudenlaisen tarkastelutavan lähisynonyymien tyypillisen käytön tunnistamiseksi. Erityisesti on tutkittu lähisynonyymien verbien käyttöä (esim. italiassa De Jonge 1993; venäjässä Divjak & Gries 2006; suomessa Arppe 2008; englannissa Deshors & Gries 2016) ja kokoa kuvaavia adjektiiveja (esim. Biber ym. 1998; Gries & Otani 2009). Näiden tutkimusten tulokset antavat tukea teorialle, jonka mukaan leksikaalisten yksikköjen merkitys muodostuu suurelta osin kontekstin perusteella (Firth 1957). Tämä on erityisen selvää, kun tutkitaan



suhteellisia adjektiiveja, jollaisia tässä artikkelissa tarkasteltavat *tärkeä*- ja *suuri*-pesueiden adjektiivit ovat. Suhteellisten adjektiivien tulkinta riippuu vahvasti siitä pääsanasta, johon ne liitetään (Hakulinen ym. 2004: § 605): esimerkiksi *pieni koira* ja *iso rotta* voivat olla kokoluokaltaan samanlaiset, mutta pääsanoja on totuttu kuvaamaan erilaisilla skaaloilla.

Korpuspohjainen suomen kielen synonymiatutkimus on edelleen vähäistä ja tarvitaan enemmän tietoa siitä, kuinka erilaiset korpuslingvistiset menetelmät soveltuvat suomen kaltaiseen agglutinoivaan kieleen. Lisäksi suurin osa lähisynonyymeja koskevasta tutkimuksesta keskittyy yhteen synonymipariin (esim. De Jonge 1993; Jantunen 2001; Kekki & Ivaska 2022) tai -pesueeseen (esim. Divjak & Gries 2006; Arppe 2008; Gries & Otani 2009). Ilmiö kaipaa lisävalotusta esimerkiksi sen suhteen, millä tavoin yhtä tietyssä konstruktiossa, kuten substantiivinmäärittäenä, esiintyvää synonymipesuetta koskevia tuloksia voidaan soveltaa muihin pesueisiin.

Toisen kielen tutkimuksessa on huomattu, että toisen kielen edistyneetkin käyttäjät voivat osata oppimaansa kieltä kieliopillisesti hyvin mutta tehdä käyttökontekstille epätyypillisiä valintoja (Ivaska 2015), kuten käyttää epätavanomaisia synonyymeja. Ensikieliset puhujat käyttävät tarkempia ilmaisuja, kun taas kielenoppijat turvautuvat usein esiintyviin sanoihin ja ilmaisuihin (Granger 2004: 132). Tämä voi tuottaa epätavanomaisia sanojen yhdistelmiä, kuten *keskeinen kokous*. Kielenoppijoiden ylikäyttämiin sanoihin on viitattu termillä leksikaaliset nallekarhut (Hasselgren 1994; Jantunen 2015): ne opitaan tunnistamaan neutraaleiksi jo kielenoppimisen alkutaipaleella ja tuntuvat siksi helpolta käyttää sillä seurauksella, että niiden käyttö laajenee myös konteksteihin, joissa ensikielinen puhuja valitsisi jonkin toisen ilmauksen (Ringbom 2007).

Siltä osin kun oppijankielen lähisynonyymien käyttöä on tutkittu, tutkimus on keskittynyt erityisesti leksikaaliseen kollokaationäkökulmaan eli siihen, minkä pääsanan yhteydessä mitäkin lähisynonyymia käytetään. Esimerkiksi Ching-Yin Lee ja Jyi-Shane Liu (2009: 252) esittävät kiinankielisten englanninoppijoiden lähisynonyymisia substantiiveja ja adverbeja koskevan tutkimuksensa pohjalta, että lähisynonyymien tyypillisten kollokaattien tunteminen on avainasemassa lähisynonyymien merkityseron ymmärtämisessä ja käytön oppimisessa. Konstruktiotajattelun mukaan sanasto ei kuitenkaan ole irrallaan kieliopista tai syntaksista, vaan sanojen merkityksiä pitää lähestyä myös ne huomioiden *kollostruktuurallisesti* (Stefanowitsch & Gries 2003: 210). Konstruktionäkökulman huomioivien tutkimusten tulokset viittaavat siihen, että vähemmän käytettyjen lähisynonyymivarianttien omaksumiseen voi vaikuttaa negatiivisesti konstruktion kieliopillinen monimutkaisuus: oppijat siis turvautuvat ”oletusarvoiseen” varianttiin silloin, kun kognitiivinen prosessointikuorma on jo valmiiksi suuri (Deshors & Gries 2014; Deshors 2016; Kekki & Ivaska 2022). Jotta voimme ymmärtää kielenkäyttäjien tekemiä valintoja, tarvitsemme tietoa paitsi siitä sanasta, jota adjektiivi määrittää, myös määritekonstruktion kieliopillisista ominaisuuksista ja laajemmasta tekstikontekstista, jossa se esiintyy.

### 2.3 Tutkimuskysymykset

Lähestymme tutkimuksen tavoitetta seuraavien tutkimuskysymysten avulla:

- 1) Millaisia adjektiivivalintoja S1- ja S2-puhujat tekevät alkuperäiseen kaunokirjallisen tekstin kirjoittajaan verrattuna?
- 2) Millainen yhteys kaunokirjallisuuden lukeneisuudella on siihen, millaisia adjektiivivalintoja S1- ja S2-puhujat tekevät?
- 3) Mitkä tekstilajikäyttöön ja adjektiivimääritekonstruktion liittyvät muutujat vaikuttavat valintaan?

Kaunokirjalliset tekstit ovat kielellä luotua sanataidetta, joka voi kirjailijasta riippuen olla lähellä aikakautensa arkista kielenkäyttöä – tai hyvin kaukana siitä. Kaunokirjallisuuden tekstilajipiirteiden vuoksi oletamme, että sekä S1- että S2-puhujat tekevät osin erilaisia valintoja kuin kaunokirjallisen kontekstin alkuperäinen kirjoittaja, ja että vapaa-ajallaan kaunokirjallisuutta lukevat tekevät keskenään samansuuntaisempia valintoja.

### 3 Aineisto ja metodi

#### 3.1 Aineistona korpuspohjainen synonymiakysely

Selvittääksemme kaunokirjallisuuden lukeneisuuden vaikutusta S1- ja S2-puhujien synonyymisten adjektiivien käyttöön, laadimme kaunokirjallisuuskorpusta hyödyntävän verkkokyselyn (ks. liite A; osin vastaavaa kokeellista menetelmää ovat hyödyntäneet esimerkiksi Antti Arppe ja Juhani Järvikivi (2007) sekä Dilin Liu ja Shouman Zhong (2016)). Kysely koostuu i) yleisistä taustatietokysymyksistä, ii) kaunokirjallisuuden lukemistottumuksia kartoittavista kysymyksistä, iii) monivalintakysymyksistä, jotka pakottavat valitsemaan yhden vaihtoehdon kolmen lähisynonyymisen adjektiivin pesueesta annettuun kaunokirjalliseen kontekstiin sekä iv) avoimista kysymyksistä, jotka selvittävät vastaajan syitä adjektiivien valinnalle. Tässä artikkelissa keskitymme määrällisen analyysimenetelmän avulla selvittämään, kuinka kaunokirjallisuuden lukemistottumukset liittyvät monivalintakysymyksissä tehtyyn adjektiivin valintaan. Monivalintakysymysten rakenne molempien lähisynonyymipesueiden osalta näkyy kuviossa 1.

|                                                                                                                                                                                                                |          |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| <p>”Luulenpa olevani”, Švejk vastasi, ”sillä isänikin oli Švejk ja äitini Švejkin rouva. En voi tuottaa heille niin _____ häpeää, että kieltäisin nimeni”. Hellä hymy häivähti tutkintatuomarin kasvoilla.</p> |          |
| <input type="checkbox"/>                                                                                                                                                                                       | isoa     |
| <input type="checkbox"/>                                                                                                                                                                                       | kookasta |
| <input checked="" type="checkbox"/>                                                                                                                                                                            | suurta   |

|                                                                                                                                                                                                                                                 |            |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <p>Lopullinen päämäärä on tietysti kapina. Mutta minä en ole niin _____ mies, että tietäisin miten se tehdään. Tiedän vain että se on tehtävä ja nuo pirut ajettava helvettiin viimeiseen nurkkaan, niin totta kuin Jumala meitä auttakoon.</p> |            |
| <input type="checkbox"/>                                                                                                                                                                                                                        | tärkeä     |
| <input checked="" type="checkbox"/>                                                                                                                                                                                                             | merkittävä |
| <input type="checkbox"/>                                                                                                                                                                                                                        | keskeinen  |

**Kuvio 1.** Kaksi esimerkkiä aineistona käytettävän kyselyn monivalintakysymysten rakenteesta. Esimerkeissä on merkittynä kaunokirjallisen tekstin alkuperäinen adjektiivi-valinta.

Tarkasteltavien lähisynonyymien osalta kysely on rakennettu korpusvetoisesti. Kyselyn pohjana olemme käyttäneet InterCorp Finnish v12 -korpuksen (Fárová & Vavřín 2019) suomenkielisiä kaunokirjallisuustekstejä. Valitsimme tarkasteltaviksi adjektiiveiksi *suuren* ja *tärkeän*, koska ne ovat korpuksen frekventeimmät adjektiivit, joilla on korpuksessa kolme vähintään 50 kertaa esiintyvää lähisynonyymia. *tärkeä/keskeinen*-lähisynonyymiparin käyttöä on tutkittu aiemmin S1-puhujien (Jantunen 2001; Vanhatalo 2003) sekä S2-puhujien predikatiivikonstruktion käytön (Kekki & Ivaska 2022) näkökul-

masta. Tässä artikkelissa keskitymme lähisynonyymien käyttöön substantiivin määritteinä, minkä lisäksi laajennamme synonyymiparista saatua tietoa ottamalla mukaan kolmannen synonyymivariantin ja vertailemalla tuloksia toiseen lähisynonyymipesueeseen.

Kyselyn pilotissa oli mukana neljä lähisynonyymia per synonyymipesue, mutta niiden määrä päätettiin karsia kolmeen, jotta vastausaika pysyisi kohtuullisena. Kyselyn monivalintakysymykset (ks. liite A) on laadittu tarkasteltavien lähisynonyymien kollokaattiesiintymien pohjalta: jokaiselle lähisynonyymille on valittu korpuksesta 10 esimerkkikatkelmaa, jossa adjektiivi määrittää sille tyypillistä pääsanaa eli substantiivikollokaattia. Kollokaatit etsittiin korpuksesta saatavien MI-arvojen (*mutual information score*) avulla, jotka mittaavat kahden leksikaalisen yhteisesiintymän todennäköisyyttä verrattuna sanojen esiintymiseen yksinään (Church & Hanks 1990). Mitä korkeampi MI-arvo, sitä todennäköisemmin pääsana esiintyy tutkitun adjektiivin kanssa kuin sattumanvaraisesti. Koska pesueiden vähiten frekventeillä adjektiiveilla (*kookas, keskeinen*) ei ollut riittävästi merkitseviä kollokaatteja, näiden adjektiivien tapauksessa esimerkkikatkelmiin on otettu mukaan myös konteksteja, joissa adjektiivi esiintyy minkä tahansa substantiivin määritteenä. Katkelmat on rajattu noin kolmen virkkeen mittaisiksi, ja niistä noin puolet on alun perin kirjoitettu suomeksi, noin puolet on suomeksi käännettyä kirjallisuutta. Osasta katkelmia on mahdollista tunnistaa alkuperäinen kaunokirjallinen teos esimerkiksi katkelmassa esiintyvien henkilöhahmojen nimien perusteella.

Vastausväsymyksen minimoimiseksi sekä adjektiivivaihtoehdot että kysymysten järjestys on satunnaistettu, kuitenkin niin, että kaikki vastaajat vastasivat ensin *suuri*-pesueen (30 katkelmaa) ja sitten *tärkeä*-pesueen (30 katkelmaa) monivalintakysymyksiin. Kyselyn keskimääräinen vastausaika oli 30 minuuttia. Vastaajia ohjattiin tekemään synonyymivalintansa seuraavanlaisella ohjeistuksella: ”Seuraavat esimerkit ovat kaunokirjallisista teksteistä. Valitse sana, joka mielestäsi sopii parhaiten kontekstiin. Johonkin kysymykseen voi olla mielestäsi enemmän kuin yksi oikea vastaus, mutta valitse se vaihtoehto, joka tuntuu sopivimmalta intuitiosi mukaan”.

### 3.2 Kyselyn osallistujat

Kyselyä<sup>3</sup> levitettiin suomalaisten korkeakoulujen postituslistojen kautta, sillä tavoitteena oli kerätä vastauksia mahdollisimman vertailukelpoiselta joukolta korkeakoulutettuja S1- ja S2-puhujia. Kyselyn vastaajina oli 234 ensikielistä suomenpuhujaa ja 28 suomenoppijaa. Molempien ryhmien kaikki vastaajat olivat korkeakoulutettuja. Vastaajamäärä on odotuksenmukainen ottaen huomioon, kuinka monta aikuisena suomea oppinutta, edistynyttä kielenkäyttäjää Suomen korkeakoulujen piirissä voi arvella olevan suhteessa suomea ensikielensä puhuviin. Vastaajien iän keskiarvo oli 31,4 vuotta (keskihajonta = 9,6), ja selvä enemmistö vastaajista (73,7 %) oli naisia. Naisten suuri osuus vastaajissa voi kertoa siitä, että heissä on ollut muita sukupuolia enemmän kyselyn aiheesta kiinnostuneita. Samoin kysely on saattanut kiinnostaa lukemista keskivertoa enemmän harrastavia ihmisiä. Molempien ryhmien vastaajista noin puolet vastaajista raportoi lukevansa kaunokirjallisuutta vapaa-ajallaan paljon: S1-puhujista 121 (51,7 %) ja S2-puhujista 14 (50 %) lukee vähintään muutamia kertoja viikossa. ”Ei koskaan” lukevia oli S1-vastaajista ainoastaan 1 (0,4 %) ja S2-vastaajista 2 (7,1 %).

<sup>3</sup> Kyselyn toteuttamisessa noudatettiin Tutkimuseettisen neuvottelukunnan (2019) ohjeistusta. Kyselyyn vastattiin anonymisti. Aineiston käsittelyn yhteydessä on varmistettu, ettei vastaajien tunnistaminen ole mahdollista myöskään epäsuorasti esimerkiksi taustatietoja yhdistämällä. Kyselyyn vastaaminen perustui vapaaehtoisuuteen eikä vastaamisesta saanut palkkiota. Vastaajilta kerättiin informoitu kirjallinen suostumus vastausten käyttöön tutkimusaineistona.



Suomenoppijoiden osalta kysely suunnattiin saateviestissä edistyneen tason kielenoppijoille. Koska kaunokirjallisuudelle tyypillinen rikas kieli ja synonymia kielellisenä ilmiönä ylipäättään eivät vielä kuulu alkeistason kielenoppijoiden osaamisprofiiliin, mahdolliset alkeistason vastaajat olisivat luultavasti joutuneet turvautumaan vahvasti arvaamiseen vastauksissaan (Liu & Zhong 2016: 243). Tieto S2-vastaajien kielitaidosta perustuu heidän omaan arvioonsa: S2-vastaajista 14 (50 %) raportoi olevansa Eurooppalaisen viitekehyksen B- ja 14 (50 %) C-tasolla. Kaikki S2-vastaajat olivat oppineet suomea vasta aikuisiällä. S2-vastaajat raportoivat 14 eri ensikieltä (albania, arabia, englantia, espanja, italia, kiina, kurdi, persia/farsi, puola, saksa, serbokroatia, turkki, unkari, venäjä). Koska yksittäisellä kielellä on aineistossa niin vähän puhujia eikä kieliä saa mielekkäästi ryhmiteltyä, ensikieliä ei voitu käyttää analyysin muuttujina. Mahdollisen ensikielen siirtovaikutuksen tarkastelun sijaan tutkimus keskittyy kuvaamaan edistyneen oppijansuomen tyypillisyyksiä saatavilla olevan, ensikielitaustaltaan heterogeenisen aineiston avulla.

### 3.3 Monimuuttuja-sekamalli

Käytämme tässä tutkimuksessa pääasiallisena tilastollisena menetelmänä monimuuttujaista logistista regressio-sekamallia. Tällaisten mallien avulla pyritään mallintamaan jonkin vastemuuttujan arvo tarkastelemalla samanaikaisesti useita selittäviä muuttujia. Sekamalleissa osa selittävästä muuttujista on niin sanottuja kiinteitä muuttujia eli joko numeerisia tai sellaisia kategorisia muuttujia, joiden kaikki mahdolliset arvot tiedetään etukäteen. Osa puolestaan on niin sanottuja satunnaismuuttujia eli sellaisia kategorisia muuttujia, joiden kaikkia mahdollisia arvoja ei tiedetä etukäteen ja jotka voivat eri aineisto-otannassa saada myös muita arvoja. Monimuuttujamallin etuna verrattuna muuttujien tarkasteluun yksittäin on se, että se mahdollistaa yhden muuttujan tarkastelun siten, että muiden muuttujien vaikutus on vakioitu.

Tilastollisten regressiomallien soveltuvuutta arvioidaan usein kolmen eri suureen avulla: niin sanottu  $R^2$ -arvo kuvaa sitä, missä määrin malli kykenee tavoittamaan aineistossa havaittavan vaihtelun. Suureen arvot vaihtelevat välillä 0–1, jossa suuremmat arvot indikoivat vaihtelun paremmin tavoittavaa mallia. Sekamalleissa tämä suure jakautuu edelleen marginaaliseen ja ehdolliseen osaan, joista ensimmäinen kuvaa kiinteiden muuttujien osuutta ja jälkimmäinen kiinteiden ja satunnaismuuttujien kokonaisuutta.  $R^2$ -suurelle ei yleensä aseteta minkäänlaisia raja-arvoja, sillä nämä vaihtelevat tutkimusasetelmien ja aineistojen välillä suuresti. Tämän lisäksi arviointiin käytetään usein niin sanottua C-arvoa, joka kuvaa sitä, missä määrin malli soveltuu vastemuuttujansa mallintamiseen. Suureen arvot vaihtelevat välillä 0,5–1, ja 0,8 ylittävien arvojen katsotaan yleisesti indikoivan, että malli soveltuu kulloiseenkin tehtävään (esim. Gries 2021: 327). Mallin toimivuutta voidaan arvioida niin ikään sen perusteella, miten hyvin se kykenee ennustamaan vastemuuttujan arvot – tässä tutkimuksessa siis sitä, ovatko kielenkäyttäjät valinneet alkuperäisessä kontekstissa käytetyn adjektiivin. Tässä tapauksessa mahdolliset arvot ovat siis välillä 0–1, joissa suuremmat arvot kuvaavat paremmin ennustavaa mallia. Tässä vertailun minimiarvona pidetään usein vastemuuttujan yleisempää arvoa eli sitä arvoa, johon päästään arvaamalla aina yleisin vaihtoehto. Kuvaamme seuraavaksi kunkin tässä tutkimuksessa käytetyn muuttujan, perustellen samalla niiden roolin tässä tutkimuksessa. Tämän jälkeen esittelemme käyttämämme sekamallin yhtenä kokonaisuutena.<sup>4</sup>

<sup>4</sup>Tilastollinen analyysi on toteutettu R-ohjelmointiympäristössä (R Core Team 2022). Kyselylomake (Liite A), tilastollisessa analyysissä käytetty aineisto ja R-koodi ovat saatavilla osoitteessa: <<https://osf.io/d2b9y/>>

Tässä tutkimuksessa vastemuuttuja on luonteeltaan binäärinen: se kuvaa sitä, onko vastaaja valinnut lähisynonymien adjektiivien joukosta alkuperäisen käyttökontekstin mukaisen adjektiivin (alkuperäinen kirjailijan tai teoksen suomentajan **valinta**). Tutkimuksen keskeisenä tavoitteena on selvittää ensikielisyyden ja kaunokirjallisuuden lukemisen vaikutusta ja niiden välistä yhteyttä suhteessa lähisynonymien adjektiivien valintaan. Muuttuja **profiili** kuvaa, käyttääkö vastaaja suomea ensikielensä vai toisena kielenä.

Kaunokirjallisuuden lukeneisuuden määrää selvitetään kahdella muuttujalla: **kuinka usein** ja **kuinka monta kirjaa vuodessa keskimäärin vastaajat lukevat**. Pietilän ja Merikiven (2014) tutkimuksessa englantia vapaa-ajallaan lukevat nuoret pärjäisivät merkitsevästi paremmin vieraana kielenä opittavan englannin sanaston laajuutta mitaavissa testeissä kuin nuoret, jotka eivät lukeneet koulun ulkopuolella. Nuorten vieraan kielen sanavarasto oli myös sitä laajempi, mitä useammin he lukivat. (Pietilä & Merikivi 2014: 33–34.) Kummankin muuttujan osalta vastaajat valitsivat kyselyssämme seitsemän sanallisesti kuvatun vaihtoehdon väliltä parhaiten omia tottumuksiaan kuvaavan (ks. tarkemmin taulukko 2). Olemme muuntaneet muuttujat numeeriseksi siten, että sen pienin arvo (1) vastaa kyselyn määrällisesti pienintä vaihtoehtoa (esim. *En koskaan lue vapaa-ajallani kaunokirjallisuutta*) ja suurin arvo (7) vastaa kyselyn määrällisesti suurinta vaihtoehtoa (esim. *Luen vapaa-ajallani joka päivä*). Lukeneisuuden määrän lisäksi mallissa käytetään selittävänä muuttujana sitä, **millä kielellä** vastaajat pääasiassa lukevat. Tämän muuttujan avulla selvitetään, onko lähisynonymien käytön kannalta väliä, lukeeko kaunokirjallisuutta nimenomaan sillä kielellä, jonka osaamista tutkitaan. Tämän kaksiarvoisen muuttujan arvoina ovat siis *suomi* ja *muut kielet*.

Useat tutkimukset nostavat esiin kollokaatioiden tärkeyden paitsi lähisynonymien sanojen (esim. Xiao & McEnery 2006), myös yleisesti toisen kielen kohdekielisen käytön (esim. Lu 2017) oppimisessa. Synonymipesueiden kunkin adjektiivin ja sen pääsanan välistä yhteyttä mallinnetaan tässä tutkimuksessa kolmella muuttujalla: i) käytetyn adjektiivin ja sen kanssa esiintyvän pääsanan eli **kollokaation assosiaation vahvuudella** kaunokirjallisuuskorpuksessa, ii) **kollokaation kokonaisfrekvenssillä** kaunokirjallisuuskorpuksessa ja iii) **kollokaation tekstilajipreferenssillä**, eli sillä, missä tekstilajissa käytetyn kollokaation välinen assosiaatio on voimakkain. Sanojen välisen yhteisesiintymisen on osoitettu kiertyvän vahvasti kielitaitoon ja kielen idioomaattisuuteen – ja vapaan valinnan ja idioomaattisen valinnan välisen jatkumon kartoittamista (engl. *open choice principle* vs. *idiom principle*) pidetään usein yhtenä modernin korpuskielitieteen lähtökohtana (ks. esim. Sinclair 1991: 109–110). Sanojen välisen assosiaation mittaamiseen on käytetty monia eri mittareita, joista monissa ongelmana on se, että ne yhdistävät kaksi toisistaan erillistä ilmiötä – assosiaation vahvuuden ja yhteisesiintymisen käyttöfrekvenssin – niin, että näiden vaikutusta ei voida erotella toisistaan (Gries 2022). Olemmekin tässä tutkimuksessa erottaneet nämä toisistaan niin, että assosiaation mittarina käytetään alkuperäisessä tekstissä käytetyn adjektiivin ja pääsanan suhdetta kuvaavaa logaritmistä vetosuhdetta (*log(arithmetic) odds ratio*; laskukaavasta, ks. Gries 2022: 7) ja frekvenssin mittarina yhteisesiintymien määrää kaunokirjallisuuskorpuksessa.

Lähisynonymien käytön erot ovat osin tekstilajisidonnaisia, sillä sekä lähisynonymien adjektiivien että niiden kanssa esiintyvien pääsanojen jakauma on erilainen eri tekstilajeissa (Biber ym. 1998; Biber & Conrad 2005). Tekstilajipreferenssin laskemiseksi haimme kyselyssä esiintyvät kollokaatiot neljästä kirjoitetun kielen korpuksesta, jotka näkyvät taulukossa 1. Tämän jälkeen laskimme vastaajan valitseman adjektiivin ja kunkin

pääsanän muodostaman kollokaation assosiaation voimakkuuden kussakin korpuksessa edellä kuvatulla tavalla. Muuttujan arvona on se tekstilaji, jossa assosiaatio oli kaikkein voimakkain eli jossa yhteisesiintyminen on kaikkein tiiveintä.

**Taulukko 1.** Tekstilajipreferenssin laskemiseen käytetyt korpuslähteet

|                          |                                                                                                     |
|--------------------------|-----------------------------------------------------------------------------------------------------|
| kaunokirjallisuustekstit | kaannossuomi-korp ( <i>akateemiset tekstit, tietotekstit, biografiat</i> jätetty haun ulkopuolelle) |
| sanomalehtitekstit       | lehdet90ff-v2 (valittuna <i>muut lehdet; tiedelehdet</i> jätetty haun ulkopuolelle)                 |
| akateemiset tekstit      | e-thesis-fi (valittuna <i>väitöskirjat</i> )                                                        |
| internetkeskustelut      | suomi24-2001-2020-korp                                                                              |

Koska tutkimuksen keskiössä on ensikielisyyden vaikutus lähisynonyymivalintaan, kielenkäyttäjäprofiili on sisällytetty malliin ns. vuorovaikutusmuuttujana kaikkiin muihin edellä kuvattuihin kiinteisiin muuttujiin. Näin mallin avulla voidaan tarkastella kunkin muuttujan osalta sitä, eroavatko S1- ja S2-kielenkäyttäjät siltä osin toisistaan.

Yksittäiset adjektiivivalinnat eivät ole kyselyssä riippumattomia toisistaan, sillä yhtäältä yhden vastaajan vastaukset voivat heijastaa idiosynkraattisia piirteitä ja toisaalta vastaukset tiettyyn kysymykseen ovat väistämättä sidoksissa ko. käyttökontekstiin myös muilta kuin edellä kuvattujen muuttujien osalta. Otamme nämä riippuvuussuhteet huomioon sisällyttämällä **vastaajan** ja **kysymyksen** yksilöllisen tunnusteen malliin satunnaismuuttujina. Kielen rakenteellisena selittävänä tekijänä käytämme **sijamuotoa**. Aiemman tutkimuksen (Jantunen 2001; Kekki & Ivaska 2022) perusteella sijamuoto vaikuttaa merkitsevästi siihen, käyttääkö puhuja akateemisessa tekstissä *tärkeää* vai *keskeistä*. Kyselyyn poimitut tekstikatkelmat on valittu adjektiivin ja pääsanän perusteella, ja aineistossa ei esiinny kummassakaan synonyymipesueessa kaikkia sijamuotoja. Tästä syystä sijamuoto on sisällytetty malliin satunnaismuuttujana, jolloin laskennallinen malli ei käsittele aineistossa havaittujen vaihtoehtojen joukkoa sulkeisena. Muuttujien tyypit ja tasot on koottu taulukkoon 2.

**Taulukko 2.** Analyysissä käytetyt muuttujat ja niiden tasot

| Tyyppi                            | Muuttuja                                                                         | Tasot                                                                                                                                                                                                                                                   |
|-----------------------------------|----------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Vastemuuttuja                     | vastaako vastaajan<br>VALINTA<br>alkuperäistä kontekstia                         | kyllä, ei                                                                                                                                                                                                                                               |
| Kielenkäyttäjä-<br>kohtaiset      | PROFIILI                                                                         | S1 (n=234), S2 (n=28)                                                                                                                                                                                                                                   |
|                                   | KUINKA MONTA<br>KIRJAA LUKEE<br>VUODESSA                                         | 0–1 (1), 2–5 (2), 6–10 (3), 11–15 (4),<br>6–20 (5), 21–30 (6), >30 (7)                                                                                                                                                                                  |
|                                   | KUINKA USEIN<br>LUKEE                                                            | en koskaan (1), harvemmin (2),<br>noin kerran kuukaudessa (3),<br>muutamia kertoja kuukaudessa (4),<br>noin kerran viikossa (5),<br>muutamia kertoja viikossa (6),<br>joka päivä (7)                                                                    |
|                                   | MILLÄ KIELELLÄ<br>LUKEE                                                          | suomeksi, muulla kielellä kuin suomeksi                                                                                                                                                                                                                 |
| Muista<br>korpuksista<br>johdetut | ALKUPERÄISEN<br>KOLLOKAATION<br>FREKVENSSEI<br>kaunokirjallisuus-<br>korpuksessa | <i>suuri</i> -pesue: 0...115<br><i>tärkeä</i> -pesue: 0...127                                                                                                                                                                                           |
|                                   | KOLLOKAATION<br>TEKSTILAJI-<br>PREFERENSSEI                                      | kaunokirjallisuus<br>internetkeskustelut<br>akateemiset tekstit<br>sanomalehtitekstit                                                                                                                                                                   |
| Konstruktio-<br>kohtaiset         | ALKUPERÄINEN<br>kirjailijan/suomentajan<br>VALINTA                               | <i>suuri, iso, kookas</i><br><i>tärkeä, keskeinen, merkittävä</i>                                                                                                                                                                                       |
|                                   | adjektiivin ja sen<br>pääsan<br>ASSOSIAATION<br>VAHVUUS                          | <i>suuri</i> -pesue: 0,935...9,704<br><i>tärkeä</i> -pesue: 1,133...8,392                                                                                                                                                                               |
| Satunnais-<br>muuttujat           | määritettävän<br>san<br>SIJAMUOTO                                                | <i>suuri</i> -pesue: ablatiivi, adessiivi, allatiivi, ge-<br>netiivi, illatiivi, inessiivi, instruktiivi, nomi-<br>natiivi, partitiivi<br><br><i>tärkeä</i> -pesue: adessiivi, elatiivi, genetiivi,<br>inessiivi, nominatiivi, partitiivi, translatiivi |
|                                   | KYSYMYS                                                                          | Q1–Q60                                                                                                                                                                                                                                                  |
|                                   | VASTAAJA                                                                         | N=262                                                                                                                                                                                                                                                   |

## 4 Tulokset

### 4.1 Mallien kokonaiskuva

**suuri-pesue.** Ensimmäisen mallin avulla pyrimme mallintamaan edellä (ks. luku 3.3) kuvattujen muuttujien avulla sitä, mitkä seikat korreloivat sen kanssa, valitseeko kielenkäyttäjä *suuri*-pesueesta sen adjektiivin, jota kulloinkin tarkasteltavassa kaunokirjallisesa teoksessa on vastaavassa kontekstissa käytetty. *suuri*-pesuetta koskeva malli toimii kaikkiaan erittäin hyvin: kiinteiden muuttujien selitysastetta kuvaava marginaalinen  $R^2$ -arvo on 0,50, kun myös satunnaismuuttujat sisältävä ehdollinen  $R^2$ -arvo on peräti 0,91. Mallin C-arvo on 0,957 ja sen kyky ennustaa vastemuuttujan oikea arvo on 0,90 (vertailuna käytettävä vastemuuttujan yleisemmän arvon todennäköisyys – se, että vastaaja on valinnut alkuperäistä kontekstia vastaavan adjektiivin – on 0,62). Kaikkiaan diagnostiset tulokset osoittavat mallin soveltuvan hyvin tarkastelunalaisen ilmiön kuvaamiseen, mikä osaltaan myös vahvistaa sen avulla saatavien tulosten luotettavuutta.

Taulukko 3 kuvaa mallin kunkin muuttujan ja vuorovaikutuksen osalta sitä, onko niiden vaikutus<sup>5</sup> kokonaisuuteen tilastollisesti merkitsevä. Mallin yhdeksästä kiinteästä muuttujasta neljän vaikutus on tilastollisesti merkitsevä. Lisäksi viiden muuttujan vuorovaikutus vastaajan ensikielisyiden kanssa on tilastollisesti merkitsevä, mikä tarkoittaa kyseisen muuttujan vaikutuksen olevan erilainen ensikielenään suomea käyttävien vastaajien ja suomea toisena kielenä käyttävien vastaajien kohdalla. *suuri*-pesueen kohdalla esimerkiksi vastaajan ensikielisyys ( $p = 0,021$ ) ja vastaajan vuodessa lukemien kirjojen määrä vuorovaikutuksessa ensikielisyiden kanssa ( $p = 0,026$ ) osoittautuivat tilastollisesti merkitseviksi muuttujiksi. Tarkastelemme osioissa 4.2–4.4 tarkemmin yksittäisten muuttujien osuutta kokonaisuudessa keskittyen erityisesti niihin muuttujiin, joiden vaikutus on tilastollisesti merkitsevä.

**tärkeä-pesue.** Toisessa mallissa käytimme vastaavia muuttujia mallintamaan sitä, valitseeko vastaaja *tärkeä*-pesueesta saman adjektiivin, jota on käytetty kaunokirjallisuudessa tarkastelunalaisissa kontekstissa. Tämä malli ei yllä aivan *suuri*-pesuetta kuvaavan mallin lukemiin, mutta sen voidaan silti katsoa kykenevään aineistossa havaitun vaihtelun kohtalaisen hyvin: marginaalinen  $R^2$ -arvo on 0,16 ja satunnaismuuttujat sisältävä ehdollinen  $R^2$ -arvo on 0,57. C-arvo on 0,819, mikä ylittää hyvän sovellettavuuden rajana pidetyn arvon 0,8. Se kykenee ennustamaan vastemuuttujan arvon todennäköisyydellä 0,79, kun vertailuna käytettävä vastemuuttujan yleisemmän arvon todennäköisyys on 0,62. Taulukko 3 kuvaa mallin kunkin muuttujan ja vuorovaikutuksen osalta sitä, onko niiden vaikutus kokonaisuuteen tilastollisesti merkitsevä. Mallissa neljän kiinteän muuttujan vaikutus on tilastollisesti merkitsevä, minkä lisäksi neljän muuttujan vuorovaikutus vastaajan ensikielisyiden kanssa on tilastollisesti merkitsevä. Myös tämän pesueen osalta tarkastelemme yksittäisiä muuttujia tarkemmin osioissa 4.2–4.4.

<sup>5</sup> Käytämme regressioanalyysien vakiintuneen tavan mukaan termiä *vaikutus* (engl. *effect*) kuvaamaan muuttujien välistä suhdetta. Termi ei ota kantaa muuttujien mahdollisiin kausaalisuhteisiin.



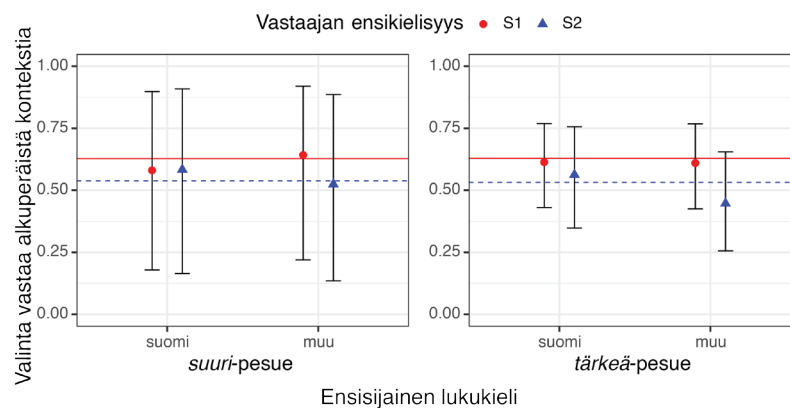
**Taulukko 3.** Uskottavuusosamäärä-testien (LRT) tulokset tutkittujen pesueiden osalta. Kynnysarvoa 0,05 alemmat arvot (merkitty asteriskilla \*) tulkitaan tilastollisesti merkitseviksi alkuperäisen kontekstin adjektiivivalinnan kannalta. Df = vapausaste<sup>6</sup>

| Muuttuja                                                  | Df         | suuri-pesue                                        |           | tärkeä-pesue                                       |           |
|-----------------------------------------------------------|------------|----------------------------------------------------|-----------|----------------------------------------------------|-----------|
|                                                           |            | LRT                                                | P-arvo    | LRT                                                | P-arvo    |
| Vastaajan ensikielisyys                                   | 1          | 5,337                                              | 0,021*    | 15,016                                             | 0,0001*   |
| Vastaajan ensisijainen lukukieli                          | 1          | 3,457                                              | 0,063     | 1,039                                              | 0,308     |
| Vuorovaikutus: Vastaajan ensikielisyys                    |            | 4,076                                              | 0,043*    | 4,649                                              | 0,031     |
| Vastaajan vuodessa lukemien kirjojen määrä                | numeerinen | 0,035                                              | 0,851     | 1,672                                              | 0,196     |
| Vuorovaikutus: Vastaajan ensikielisyys                    |            | 4,962                                              | 0,026*    | 0,908                                              | 0,341     |
| Vastaajan lukemisen frekvenssi                            | numeerinen | 0,148                                              | 0,700     | 0,097                                              | 0,756     |
| Vuorovaikutus: Vastaajan ensikielisyys                    |            | 5,404                                              | 0,020*    | 0,715                                              | 0,398     |
| Valitun adjektiivin ja sen pääsanan välinen assosiaatio   | numeerinen | 1093,542                                           | < 0,0001* | 415,055                                            | < 0,0001* |
| Vuorovaikutus: Vastaajan ensikielisyys                    |            | 0,110                                              | 0,740     | 5,154                                              | 0,023*    |
| Valitun adjektiivin ja sen pääsanan tekstilajipreferenssi | 2          | 405,977                                            | < 0,0001* | 99,627                                             | < 0,0001* |
| Vuorovaikutus: Vastaajan ensikielisyys                    |            | 3,864                                              | 0,145     | 8,150                                              | 0,017*    |
| Pääsanalan esiintymisfrekvenssi kaunokirjallisuudessa     | numeerinen | 78,539                                             | < 0,0001* | 0,522                                              | 0,470     |
| Vuorovaikutus: Vastaajan ensikielisyys                    |            | 30,090                                             | < 0,0001* | 1,063                                              | 0,302     |
| Alkuperäisessä kontekstissa käytetty adjektiivi           | 2          | 5,463                                              | 0,065     | 12,866                                             | 0,002*    |
| Vuorovaikutus: Vastaajan ensikielisyys                    |            | 55,164                                             | < 0,0001* | 3,242                                              | 0,198     |
| Sijamuoto (satunnaismuuttuja)                             |            | Estimaatin keski-hajonta<br>S1: 0,744<br>S2: 1,076 | 0,0001*   | Estimaatin keski-hajonta<br>S1: 0,389<br>S2: 0,555 | 0,915     |
| Korrelaatio ensikielisyyden ja sijamerkin välillä         |            | 0,10                                               | < 0,0001* | 0,94                                               | 0,812     |
| Kysymys (satunnaismuuttuja)                               |            | Estimaatin keski-hajonta:<br>3,637                 | < 0,0001* | Estimaatin keski-hajonta:<br>1,722                 | < 0,0001* |
| Vastaaja (satunnaismuuttuja)                              |            | Estimaatin keski-hajonta:<br>0,094                 | 0,809     | Estimaatin keski-hajonta:<br>0,216                 | 0,013*    |

<sup>6</sup> Vapausaste tarkoittaa niiden muuttujien määrää, jotka voivat vaihdella. Esimerkiksi vastaajan ensikielisyys -muuttujan Df = 1, koska vaihtoehtoja on kaksi (S1 tai S2) ja alkuperäisessä kontekstissa käytetyn adjektiivin Df = 2, koska vaihtoehtoja on 3 (*suuri/iso/kookas* tai *tärkeä/keskeinen/merkittävä*).

Tarkastelemme seuraavaksi yksittäisten muuttujien vaikutusta siihen, onko vastaaja valinnut saman adjektiivin kuin alkuperäisen tekstin kirjoittaja. Esittelemme rinnakkain tulokset *suuri-* ja *tärkeä-*pesueista. Käsitlemme ensin lukemistottumuksiin liittyvät muuttujat ja sen jälkeen käyttöä, merkitystä ja muotoa kuvaavat konstruktiokohtaiset muuttujat. Kuvio 2–9 kuvaavat tarkasteltujen muuttujien vaikutusta lähisynonymien adjektiivin valintaan. Selittävien muuttujien arvot ovat kuvioissa vaakakselilla ja vastemuuttujan (eli vastaajan adjektiivivalinnan suhteen alkuperäisen kontekstin adjektiiviin) arvot pystyakselilla. Pystyakselin arvot vaihtelevat välillä 0–1 niin, että suuremmat arvot tarkoittavat suurempaa todennäköisyyttä valita sama adjektiivi, jota on käytetty alkuperäisessä kaunokirjallisessa kontekstissa. Merkintä S1 tarkoittaa kuvioissa suomea ensikielenään puhuvia ja S2 suomea toisena kielenä puhuvia vastaajia.

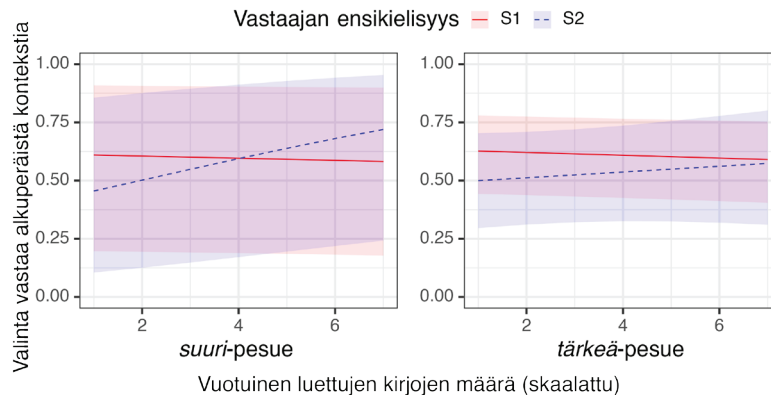
## 4.2 Lukemistottumukset



**Kuvio 2.** Vastaajan ensisijaisen lukukielen vaikutus adjektiivivalintaan<sup>7</sup>

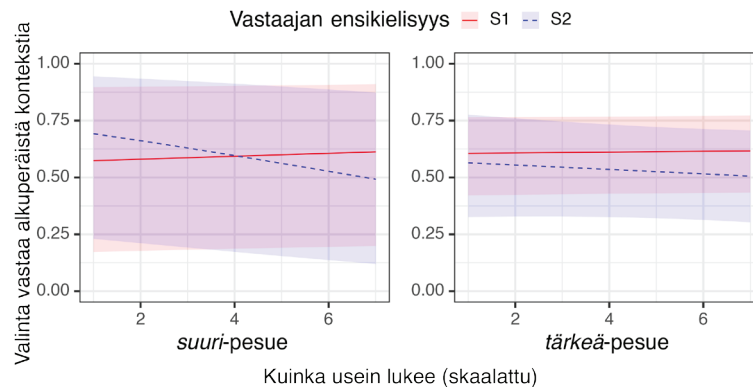
Vastaajan ensisijainen lukukieli näyttää kiinnostavan tendenssin monikielisen lukemisen hyödyllisyydestä kuviossa 2. Molempien pesueiden kohdalla ne S2-puhujat, jotka raportoivat lukevansa tällä hetkellä ensisijaisesti suomeksi, ovat valinneet adjektiiveja samansuuntaisemmin alkuperäisen kaunokirjallisen kontekstin kanssa. S1-puhujien osalta päinvastoin muulla kielellä kuin suomeksi lukeminen näyttää tuottavan kontekstinmukaisempia valintoja. Ero on *suuri-*pesueen osalta lähellä tilastollisesti merkitsevää. Tulos voi viitata siihen, että vieraan kielen osaaminen edistyneellä tasolla näkyy oman ensikielen hienosyisemmässä käytössä, kuten Kecskes ja Papp (2000) esittävät. Vieraalla kielellä lukeminen voisi siis parantaa herkkyyttä oman ensikielen käytössä. S1- ja S2-puhujien tendenssit ensisijaisen lukukielen vaikutuksesta kertovat hieman eri asiasta, koska mallissa käytetyllä muuttujalla ei saada S2-puhujien osalta selville, tapahtuuko muulla kielellä kuin suomeksi lukeminen vastaajien ensikielellä vai jollakin muulla heidän osaamallaan kielellä, kuten englanniksi. Erilaisten lukukieliasetelmien tarkasteluun tarvittaisiin tätä tutkimusta laajempi aineisto.

<sup>7</sup> Punainen pallo ja sininen kolmio kuvaavat kyseisen muuttujan estimaattia eli keskimääräistä vaikutusta monimuuttujamallissa. Pystysuuntainen virhepalkki kuvaa 95 % luottamusväliä eli sitä väliä, jolla arvojen oletetaan korkeintaan vaihtelevan eri aineistoa käytettäessä. Vaakasuuntaiset viivat kuvaavat alkuperäisen kontekstin mukaisten vastausten mallista riippumatonta keskiarvoa. Punainen yhtenäinen vaakasuuntainen viiva kuvaa S1-vastaajien keskiarvoa ja sininen katkoviiva S2-vastaajien keskiarvoa. Sama merkintätapa on käytössä myös kuvioissa 7, 8 ja 9.



**Kuvio 3.** Vastaajan vuodessa lukemien kirjojen määrän vaikutus adjektiivivalintaan

Myös vuodessa luettujen kirjojen määrä kuviossa 3 näyttää samansuuntaisen tendenssin molempien pesueiden osalta, vaikka *tärkeä*-pesueessa ero ei olekaan tilastollisesti merkitsevä. S2-puhujien osalta mitä enemmän vastaaja lukee kirjoja, sitä yhdenmukaisempia hänen adjektiivivalintansa ovat alkuperäisen kontekstin kanssa. Ensikielisten aikuisten on havaittu voivan oppia kaunokirjallisuutta lukemalla keskimäärin enemmän sanoja kuin kielenoppijoiden (Reynolds 2015: 121). S1-puhujilla kaunokirjallisuuden lukemisen määrä ei tässä tutkimuksessa kuitenkaan näytä vaikuttavan adjektiivivalintojen kontekstinmukaisuuteen. S1-puhujilla on aikuisiällä suomea oppineita pidempi altistumisaika suomenkieliselle kaunokirjallisuudelle. Täten heidän aikuisiällä lukemiensa kirjojen määrä ei välttämättä enää vaikuta tässä kyselyssä tarkasteltuun sanaston osaan, joka kuuluu sekä tarkasteltujen synonyymisten adjektiivien että niiden kanssa esiintyvien pääsanojen osalta melko arkiseen perussanastoon.

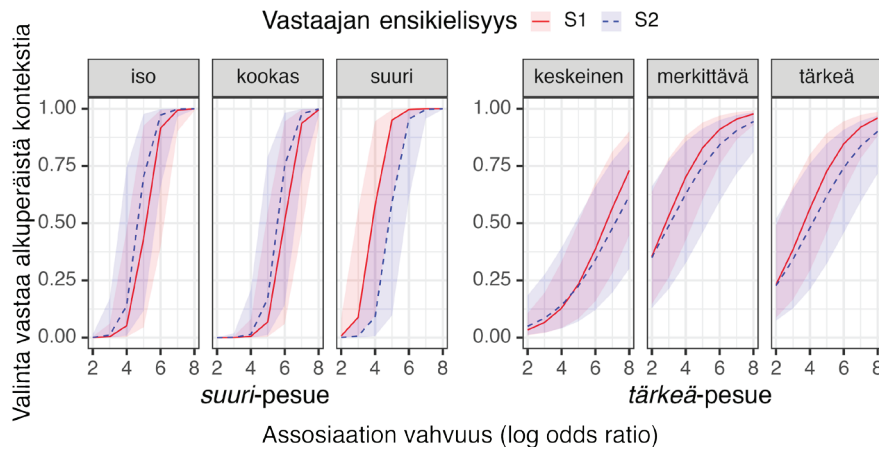


**Kuvio 4.** Vastaajan lukemiseen käyttämän ajan vaikutus adjektiivivalintaan

Kuvio 4:n kuvaama tendenssi on jossain määrin epäodotuksenmukainen erityisesti suhteessa edelliseen kuvioon 3. Siinä missä vuodessa luettujen kirjojen suurempi määrä nostaa S2-puhujilla adjektiivivalintojen yhdenmukaisuutta alkuperäisen kontekstin kanssa, yksittäisten lukutuokioiden suurempi määrä näyttää puolestaan laskevan sitä. Luettujen kirjojen ja lukutuokioiden kasvavat määrät ovat aiemmissa tutkimuksissa (esim. Pietilä & Merikivi 2014) molemmat korreloineet positiivisesti sanaston laajuuden kanssa nuorilla vieraan kielen oppijoilla. Vastaajat, jotka raportoivat kyselyssämme lukevansa kaunokirjallisuutta ”joka päivä” saavat luettua hyvin eri määrän kirjoja: monet heistä

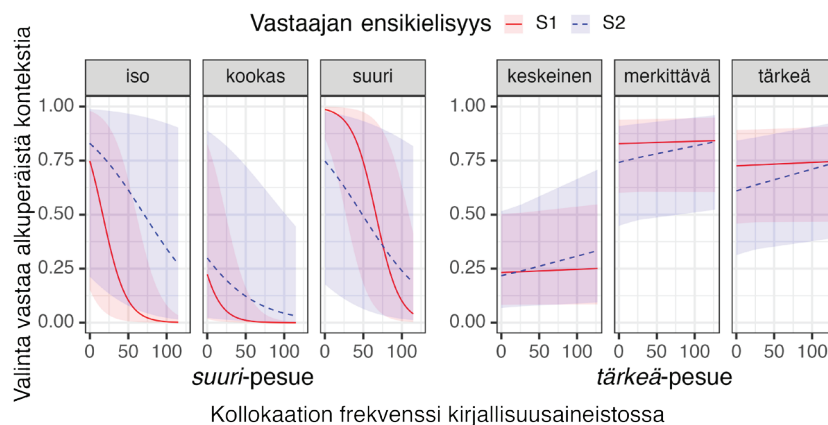
lukevat yli 30 kirjaa vuodessa, mutta jotkut vain 2–5. Synonyymisen sanaston oppimisen näkökulmasta tärkeää näyttäisi siis olevan nimenomaan se, kuinka paljon kaunokirjallista kielen mallia eli syötöstä saa, eli kuinka monta kokonaisteosta lukee.

### 4.3 Tyypillinen käyttökonteksti



**Kuvio 5.** Adjektiivin ja sen pääsanan eli kollokaation assosiaation vahvuus S1- ja S2-vastaajien adjektiivivalintojen erottimena

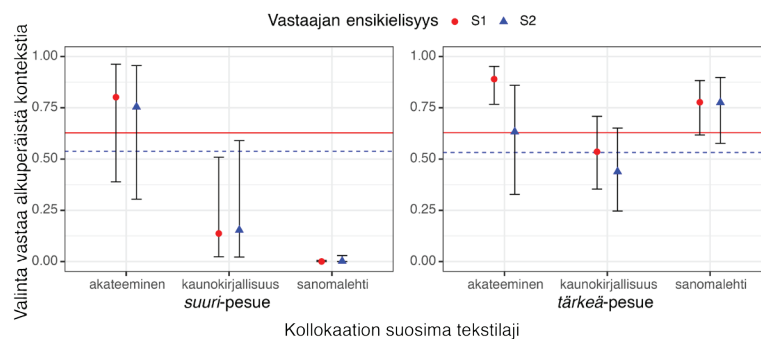
Kuviossa 5 näkyvä tendenssi vahvistaa aiemmasta tutkimuksesta (esim. Gries & Wulff 2005) saatua tietoa siitä, että ensikieliset puhujat ja edistyneet kielenoppijat hahmottavat konstruktoita hyvin samantapaisesti: mitä voimakkaammin alkuperäisessä kontekstissa käytetty adjektiivi ja sen kanssa esiintyvä pääsana ovat assosioituneet toisiinsa kaunokirjallisuuskontekstissa ylipäätään, sitä todennäköisempää on, että sekä S1- että S2-vastaajat valitsevat alkuperäisen kontekstin mukaisia adjektiiveja. Kuvioista näkyy niin ikään *suuri*-pesueen *suuri*-variantin kohdalla ja *tärkeä*-pesueen kaikkien varianttien osalta se, että S1-puhujat ovat tälle assosiaativihjeelle hieman edistyneitä S2-oppijoita herkempiä.



**Kuvio 6.** Kollokaation yleisyys S1- ja S2-vastaajien adjektiivivalintojen erottimena

*tärkeä*-pesueen kohdalla kollokaation absoluuttisen määrän yleisyys vaikuttaa odotuksenmukaisesti kuviossa 6: mitä yleisempi alkuperäisessä kontekstissa käytetty kollokaatio on kaunokirjallisuuskontekstissa ylipäätään, sitä samankaltaisempia valintoja sekä

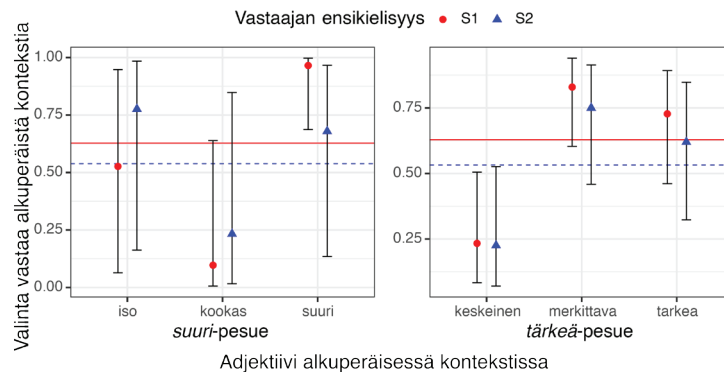
S1- että S2-vastaajat tekevät alkuperäisen kontekstin kanssa. *suuri*-pesueen kohdalla tilanne näyttää ensi alkuun päinvastaiselta, ja adjektiiv-substantiivikombinaation absoluuttinen frekvenssi näyttää aiheuttavan molemmilla vastaajaryhmillä erisuuntaisia vastauksia kuin alkuperäisessä kontekstissa. Ero selittyy kuitenkin sillä, että kyselyyn valikoituneet *suuri*-pesueen pääsanat ovat ylipäättään hyvin yleisiä ja monikäyttöisiä, eli kollokaatioilla ei ole suuresta frekvenssistään huolimatta kovinkaan vahvaa assosiaatio-suhdetta. Todella frekventtejä, yli sata kertaa esiintyviä kollokaatioita on molemmilla pesueilla vain yksi, joten kuvio on melko hypoteettinen sen suhteen, millainen vaikutus todellisella frekvenssillä oikeasti on valintaan. Tämä vahvistaa aiempaa käsitystä (esim. Gries & Deshors 2014: 111) siitä, että pelkkä esiintymien frekvenssi ei riitä tilastolliseen mallintamiseen vaan ilmiön ymmärtämiseksi tarvitaan esimerkiksi assosiaatiolaskuja (ks. kuvio 5 edellä).



**Kuvio 7.** Kollokaation tekstilajipreferenssi S1- ja S2-vastaajien adjektiivivalintojen erottimena

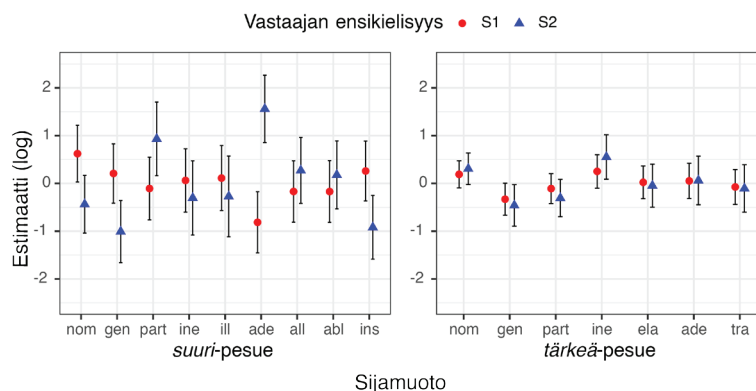
Kollokaation preferoimaa tekstilajia tarkasteltiin neljän kirjoitetun tekstilajin avulla: kaunokirjallisuus, akateemiset tekstit, sanomalehtitekstit ja internetkeskustelut. Koska yksikään kollokaatio ei preferoinut voimakkaimmin internetkeskusteluja, se ei ole näkyvissä kuviossa 7. Kuviossa näkyy kiinnostavasti, että tekstilajipreferenssi on ohjannut S1- ja S2-vastaajien valintoja hyvin samansuuntaisesti molempien adjektiivipesueiden osalta. Pesueiden väliset erot kuvaavat mahdollisesti jossain määrin tekstilajien välisiä eroja sen suhteen, mitä eri tekstilajeissa puhutaan: koosta puhuminen saattaa olla tyyppillistä nimenomaan akateemisille teksteille, kun taas merkityksestä puhutaan tasaisemmin kaikissa tekstilajeissa. Aiemmassa tutkimuksessa (esim. Ivaska 2015) edistyneiden suomenoppijoiden kirjoittamien akateemisten tekstien rakenne-eroja ensikielisten teksteihin on perusteltu nimenomaan eriasteisella akateemisen kirjoittamisen kielenkäytön tuntemuksella. Tämän tutkimuksen aineiston perusteella tekstilajilla ei kuitenkaan näytä olevan kovin suurta eroa korkeakoulutetuilla vastaajilla, mikä johtunee siitä, että aineisto kuvaa ennemmin synonyymisen sanaston reseptiivistä kuin produktiivista hallintaa.

#### 4.4 Alkuperäisen sanan ominaisuudet



**Kuvio 8.** Yleiskatsaus adjektiivivalintojen vastaavuudesta alkuperäiseen kontekstiin

Yleiskatsaus vastaajien adjektiivivalinnoista suhteessa alkuperäisen kaunokirjallisen tekstin kirjoittajan tekemään valintaan kuviossa 8 osoittaa selvän eron *suuri*- ja *tärkeä*-pesueiden välillä. *tärkeä*-pesueessa vähiten frekventti *keskeinen* noudattaa molemmilla vastaajaryhmillä samaa linjaa, mutta kiinnostavasti yhdenmukaisimmin alkuperäisen kontekstin kanssa on käytetty varianttia *merkittävä*, ei siis frekventteintä *tärkeää*. *suuri*-pesueen osalta muuttuja ei ollut tilastollisesti merkitsevä, mutta sen eroa toiseen pesueeseen voi silti pitää kiinnostavana. S1-vastaajat ovat käyttäneet pesueen frekventteintä varianttia eli *suurta* kontekstinmukaisimmin. S2-vastaajilla *iso* nousee useimmin kontekstinmukaisesti valituksi, mikä saattaisi viitata siihen, että *iso* mielletään pesueen monikäyttöisimmäksi variantiksi. Molemmissa vastaajaryhmissä epäfrekventteintä *kookasta* on valittu vähiten samankaltaisesti alkuperäisen kontekstin kanssa. Osasyynä *kookkaan* ja *keskeisen* vähemmän kontekstinmukaiseen valintaan lienee se aineistonkeruuseen liittyvä seikka, että pesueiden yleisemmistä adjektiiveista poiketen osa näiden sanojen yhteydessä esiintyvistä pääsanoista ei ollut yhtä vahvasti kyseiseen adjektiiviin assosioituneita. Selvää on, että eri varianttien kokonaisfrekvenssin vaikutus valintaan ei ole yksiselitteinen eivätkä S2-vastaajat eroa S1-vastaajista sen suhteen yhdenmukaisesti, vaan vaikutus näyttää vaihtelevan synonyymipesueittain.



**Kuvio 9.** Konstruktion sijamuoto S1- ja S2-vastaajien adjektiivivalintojen erottimena

Kuvio 9 visualisoi sijamuodon vaikutusta siihen, ovatko vastaajat valinneet alkuperäiskontekstia vastaavan adjektiivin. Koska sijamuoto on käytetyssä mallissa luonteeltaan satunnaismuuttuja, ei sen vaikutusta tule vertailla suoraan edellä kuvattujen



muuttujien kanssa. Muuttujan vaikutusta ja kuviota 9 voi tarkastella siitä näkökulmasta, miten kukin muuttujan arvo suhteellisesti vaikuttaa adjektiivin valintaan: nollan ylittävät arvot kuvaavat sijamuotoja, joissa esiintyvät adjektiivit valittiin alkuperäiskontekstin mukaisesti keskimääräistä yleisemmin ja nollan alittavat arvot puolestaan sijamuotoja, joissa esiintyvät adjektiivit poikkesivat keskimääräistä yleisemmin alkuperäiskontekstista. Huomionarvoista on erityisesti se, että *suuri*-pesueen kohdalla sijamuotojen välistä vaihtelua on paljon enemmän kuin *tärkeä*-pesueessa. Lisäksi on mielenkiintoista, että *suuri*-pesueessa S1- ja S2-vastaajat poikkeavat monissa kohdin toisistaan selvästi, kun taas *tärkeä*-pesueessa tällaista eroa ei ole. Tämä tukee ja tarkentaa taulukon 3 perusteella tehtyä päätelmää, jonka mukaan *suuri*-pesueessa S1- ja S2-vastaajien välillä on tilastollisesti merkitsevä ero mutta *tärkeä*-pesueessa ei. Vastaajaryhmien välinen ero on selvä yhtäältä *suuri*-pesueen nominatiivissa ja instruktiivissa esiintyneiden tapausten kohdalla, joissa S1-vastaajat ovat valinneet selvästi S2-vastaajia useammin alkuperäisen kontekstin mukaisen adjektiivin. Toisaalta partitiivissa ja erityisesti adessiivissa esiintyneiden adjektiivien kohdalla S2-vastaajat ovat valinneet selvästi S1-vastaajia useammin alkuperäisen kontekstin mukaisen adjektiivin.

## 5 Johtopäätökset

Olemme tässä artikkelissa käsitelleet kokeellisen kyselyaineiston pohjalta tehdyn tilastollisen regressiomallin avulla sitä, minkä lähisynonymymisen adjektiivin ensikieliset suomenpuhujat ja edistyneet suomenoppijat valitsevat kaunokirjalliseen teksti-kontekstiin substantiivin määriteasemassa. Tutkittavina synonymimipesueina olivat *suuri/iso/kookas* ja *tärkeä/merkittävä/keskeinen*, jotka ovat kaunokirjallisuuskorpuksen yleisimmät adjektiivipesueet. Tarkastelimme valintaa suhteessa alkuperäisen kirjoittajan tekemään valintaan, vastaajien kaunokirjallisuuden lukemistottumuksiin sekä adjektiivimääritekonstruktion ominaisuuksiin ja tekstilajikäyttöön.

Vastauksena ensimmäiseen tutkimuskysymykseen huomattiin sekä S1- että S2-vastaajien adjektiivivalintojen eroavan jossain määrin kaunokirjallisten tekstien alkuperäisten kirjoittajien valinnoista. Lisäksi kielitaustalla on tilastollisesti merkitsevä ero valintoihin. Useat lukemistottumuksia ja adjektiivien konstruktiokäyttäytymistä mittaavat muuttujat osoittautuivat tilastollisesti merkitseviksi. Valintaan pakottavan kyselytestin avulla voidaan tarkastella lähisynonymymisten adjektiivien reseptiivistä hallintaa, mikä tuottaa osin erinäköisiä tuloksia kuin aiemmat L2-puhujien synonymymisten ilmausten produktiiviseen käyttöön keskittyneet korpustutkimukset.

Toisena tutkimuskysymyksenä kysyimme, millainen yhteys kaunokirjallisuuden lukeneisuudella on S1- ja S2-puhujien adjektiivivalintoihin. Lukemistottumusten osalta voidaan todeta, että kaunokirjallisuuden lukeminen vaikuttaa samansuuntaisesti tarkasteltavaan ilmiöön eli adjektiivin valintaan määritekonstruktiossa. Edistyneillä S2-puhujilla valinnat ovat sitä kontekstinmukaisempia, mitä enemmän he lukevat kohdekielellä eli suomeksi. S1-puhujilla vastaavaa yhteyttä ei mallinnuksen perusteella voi osoittaa. Sen sijaan ensikielisillä aikuisilla vieraalla kielellä lukeminen saattaa tehdä herkemmäksi kaunokirjallisuuskontekstin mukaisille valinnoille.

Myös kolmannessa tutkimuskysymyksessä esittämämme tekstilajipreferenssiä ja kollokaation assosiaatiota kuvaavat muuttujat vaikuttivat samansuuntaisesti molemmissa pesueissa. Alkuperäisen sanan ominaisuudet, eli kaunokirjallisessa kontekstissa alun perin käytetty adjektiivi ja konstruktion sijamuoto, puolestaan vaikuttivat eri tavoin

tarkastelujen pesueiden kohdalla. Tämä on erityisen tärkeä havainto jatkotutkimuksen kannalta, kun mietitään, kuinka synonyymisten konstruktoiden rakennetta tarkastelevien tutkimusten tuloksia voidaan yleistää koskemaan synonymiaa laajemmin.

Tulokset osoittavat, että tilastollisena menetelmänä käytetty monimuuttujainen logistinen regressio-sekamalli sopii tutkimuksessa tarkasteltavan ilmiön selittämiseen. Malli toimii molemmissa pesueissa hyvin, mutta kiinnostavasti se selittää selvästi paremmin *suuri*-pesueen adjektiivivalintojen vaihtelua. Ero selittyy ennen kaikkea sillä, että aiemman tutkimuksen perusteella identifioituilla, lukutottumuksiin ja konstruktiokehtaisiin ilmiöihin kohdistuneilla kiinteillä muuttujilla on selvästi pienempi rooli *tärkeä*-pesueen valintoja mallinnettaessa. Kysymys- ja vastaajakohtaiset satunnaismuuttujat sen sijaan selittävät karkeasti ottaen yhtä suuren osan vaihtelusta kummassakin pesueessa, eli kyselyn rakenne tuskin selittää pesueiden välistä eroa. Synonymiaa käsittelevissä korpustutkimuksissa on usein tarkastelussa vain yksi synonyymipari tai -pesue. Tämän tutkimuksen pohjalta voi esittää, että synonyymiutta käsittelevässä tutkimuksessa tulisi huomioida yksittäisten lähisynonyymisten pesueiden sijaan useita eri synonyymipesueita, jotta sanakohtaisen vaihtelun sijaan olisi ylipäättään mahdollista päästä tarkastelemaan synonymiaa laajempaa ilmiötä.

Tutkimuksessa käytettyyn kyselyasetelmaan liittyy joitakin rajoituksia, jotka on hyvä ottaa huomioon tulosten tulkinnassa. Vastausväsymyksen ja arvaamiseen turvautumisen minimoimiseksi adjektiivivalintaan pakottavat monivalintakysymykset ja niiden vastausvaihtoehdot on satunnaistettu, mutta kyselyalustasta johtuvista teknisistä syistä pesueet esitettiin kuitenkin kaikille vastaajille samassa järjestyksessä: *suuri*-pesue ennen *tärkeä*-pesuetta. Vastausväsymys on siis saattanut olla suurempi jälkimmäisen pesueen kohdalla, mikä voi heijastua huonompina arvoina pesueen mallintamisen toimivuudessa. Kysymys- ja vastaajakohtaiset satunnaismuuttujat kattavat kuitenkin karkeasti ottaen yhtä paljon vaihtelusta pesueesta riippumatta, joten kyselyn rakenne tuskin on tässä kovinkaan merkittävässä osassa. S1- ja S2-vastaajaryhmät olivat hyvin erikokoisia (N = 234 vs. N = 28), joten S1-vastaajaryhmää voi pitää edustavampana kuin S2-ryhmää. Monimuuttujamalli ottaa ryhmien kokoeron kuitenkin lähtökohtaisesti huomioon, joten erikokoisten vastaajaryhmien ei pitäisi merkittävästi vaikuttaa tulosten luotettavuuteen. Yhteen vastausvaihtoehtoon pakottavassa kyselyssä ei lisäksi saada huomioitua synonyymisten ilmausten välimaastoa, siis tapauksia, joissa ensikielinen vastaaja voisi todennäköisesti käyttää jompaakumpaa tai mitä tahansa annetuista vaihtoehdoista (Gries & Deshors 2014; Deshors & Gries 2016; Kekki & Ivaska 2022). Välimaasto saattaa selittää niitä kyselyn kohtia, joiden valinnoissa on erityisen paljon vaihtelua vastaajien välillä.

Lähisynonyymisten ilmausten avulla kielenkäyttäjä voi esimerkiksi osoittaa tuntevansa tekstilajien välisiä eroja ja tuottaa hienosyisiä sävyeroja. Tutkimuksen tulokset kannustavat käyttämään kaunokirjallisuutta aikuisten kielenopetuksessa, jotta näiden sävyerojen tunnistaminen mahdollistuisi. Tutkimuksessa käytetty kysely (ks. liite A) voi lisäksi toimia opetusmateriaalina sellaisenaan ja auttaa opettajaa tarkastelemaan yhdessä suomenoppijoiden kanssa, mitkä tekijät vaikuttavat adjektiivin valintaan kaunokirjallisuuskontekstissa.

## Korpuslähteet

- InterCorp - Finnish, version 12: Fárová, Lenka & Vavřín, Martin. 2019. Institute of the Czech National Corpus, Faculty of Arts, Charles University.
- kaannossuomi-korp: Mauranen, Anna. 1999. Käännössuomen korpus, Korp [tekstikorpus]. Kielipankki.
- suomi24-2001-2020-korp: City Digital Group. 2021. Suomi24 virkkeet -korpus 2001-2020, Korp-versio [tekstikorpus]. Kielipankki.
- e-thesis-fi: Helsingin yliopisto. 1999. Helsingin yliopiston suomenkielisen E-thesiksen Korp-versio [tekstikorpus]. Kielipankki.
- lehdet90ff-v2: Helsingin yliopisto. 2017. 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä -korpus, versio 2 [tekstikorpus]. Kielipankki.

## Lähteet

- Alderson, Charles & Haapakangas, Eeva-Leena & Huhta, Ari & Nieminen, Lea & Ullakonoja, Riikka. 2015. *The diagnosis of reading in a second or foreign language*. New York: Routledge, Taylor & Francis Group.
- Arppe, Antti. 2008. *Univariate, bivariate and multivariate methods in corpus-based lexicography: A study of synonymy*. Helsingin yliopisto. (Väitöskirja).
- Arppe, Antti & Järvikivi, Juhani. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2). 131–159. <https://doi.org/10.1515/CLLT.2007.009>
- Biber, Douglas & Conrad, Susan. 2005. Register Variation: A Corpus Approach. Teoksessa Schiffrin, Deborah & Tannen, Deborah & Hamilton, Heidi (toim.), *The Handbook of Discourse Analysis*, 175–196. Wiley. <https://doi.org/10.1002/9780470753460.ch10>
- Biber, Douglas & Conrad, Susan & Reppen, Randi. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511804489>
- Bono, Susanna. 2023. 14-vuotias Alma Hurme lukee vain englanniksi – ”Suomeksi samat kirjat saattavat kuulostaa vähän hassuilta”. YLE. <https://yle.fi/a/74-20043198> (Viitattu 16.8.2023.)
- Church, Kenneth & Hanks, Patrick. 1990. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics* 16(1). 22–29.
- Conlexis. 2013. <http://wiki.osao.fi/conlexis/> (Viitattu 1.8.2023.)
- Cruse, David. 1986. *Lexical Semantics*. Cambridge University Press.
- Danglli, Leonard & Abazaj, Griselda. 2014. Lexical Cohesion, Word Choice and Synonymy in Academic Writing. *Mediterranean Journal of Social Sciences* 5(14). 628–632. <https://doi.org/10.5901/mjss.2014.v5n14p628>
- De Jonge, Bob. 1993. The existence of synonyms in a language: Two forms but one, or rather two, meanings? *Linguistics* 31(3). 521–538. <https://doi.org/10.1515/ling.1993.31.3.521>
- Deshors, Sandra. 2016. Inside phrasal verb constructions: A co-varying collexeme analysis of verb-particle combinations in EFL and their semantic associations. *International Journal of Learner Corpus Research* 2(1). 1–30. <https://doi.org/10.1075/ijlcr.2.1.01des>
- Deshors, Sandra, & Gries, Stefan. 2016. Profiling verb complementation constructions across New Englishes: A two-step random forests analysis of *ing* vs. *to* complements. *International Journal of Corpus Linguistics* 21(2). 192–218. <https://doi.org/10.1075/ijcl.21.2.03des>
- Deshors, Sandra & Götz, Sandra & Laporte, Samantha. 2018. Linguistic innovations in EFL and ESL: Rethinking the linguistic creativity of non-native English speakers. Teoksessa Deshors, Sandra & Götz, Sandra & Laporte, Samantha (toim.), *Benjamins Current Topics*, 1–20. John Benjamins Publishing Company. <https://doi.org/10.1075/bct.98.01des>
- Deshors, Sandra & Gries, Stefan. 2014. A case for the multifactorial assessment of learner language: The uses of *may* and *can* in French-English interlanguage. Teoksessa Glynn, Dylan & Robinson, Justyna (toim.), *Human Cognitive Processing*, 179–204. John Benjamins Publishing Company. <https://doi.org/10.1075/hcp.43.07des>
- Divjak, Dagmar & Gries, Stefan. 2006. Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1). 23–60. <https://doi.org/10.1515/CLLT.2006.002>
- Edmonds, Philip & Hirst, Graeme. 2002. Near-Synonymy and Lexical Choice. *Computational Linguistics* 28(2). 105–144. <https://doi.org/10.1162/089120102760173625>

- Ellis, Nick. 2008. Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. Teoksessa Robinson, Peter & Ellis, Nick (toim.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, 382–415. Routledge. <https://doi.org/10.4324/9780203938560-24>
- Firth, John. 1957. A Synopsis of Linguistic Theory, 1930–55. *Studies in Linguistic Analysis, Special Volume of the Philological Society*. 1–31.
- Goldberg, Adele. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences* 7(5). 219–224. [https://doi.org/10.1016/S1364-6613\(03\)00080-9](https://doi.org/10.1016/S1364-6613(03)00080-9)
- Granger, Sylviane. 2004. Computer Learner Corpus Research: Current Status and Future Prospects. Teoksessa Connor, U. & Upton, T. (toim.), *Applied Corpus Linguistics: A Multidimensional Perspective*, 123–145. <https://doi.org/10.1163/9789004333772>
- Granger, Sylviane. 2015. Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research* 1(1). 7–24. <https://doi.org/10.1075/ijlcr.1.1.01gra>
- Gries, Stefan & Otani, Naoki. 2009. Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal* 2009(34). 121–150.
- Gries, Stefan. 2021. *Statistics for Linguistics with R* (3rd edition). De Gruyter Mouton.
- Gries, Stefan. 2022. What do (some of) our association measures measure (most)? Association? *Journal of Second Language Studies* 5(1). 1–33. <https://doi.org/10.1075/jsls.21028.gri>
- Gries, Stefan & Deshors, Sandra. 2014. Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora* 9(1). 109–136. <https://doi.org/10.3366/cor.2014.0053>
- Gries, Stefan & Wulff, Stefanie. 2005. Do foreign language learners also have constructions? *Annual Review of Cognitive Linguistics* 3. 182–200. <https://doi.org/10.1075/arcl.3.10gri>
- Hakulinen, Auli & Vilkuna, Maria & Korhonen, Riitta & Koivisto, Vesa & Heinonen, Tarja-Riitta & Alho, Irja (toim.). 2004. *Ison suomen kieliopin verkkoversio*. Kotimaisten kielten tutkimuskeskus. <https://kaino.kotus.fi/visk/etusivu.php> (Viitattu 1.6.2023)
- Hasselgren, Angela. 1994. Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics* 4(2). 237–258. <https://doi.org/10.1111/j.1473-4192.1994.tb00065.x>
- Ivaska, Ilmari. 2015. *Edistyneen oppijansuomen konstruktioiirteitä korpusvetoisesti: Avainrakennanalyysi*. (Annales Universitatis Turkuensis C 409). Turun yliopisto. (Väitöskirja).
- Ivaska, Ilmari & Bernardini, Silvia. 2020. Constrained language use in Finnish: A corpus-driven approach. *Nordic Journal of Linguistics* 43(1). 33–57. <https://doi.org/10.1017/S0332586520000013>
- Jantunen, Jarmo. 2001. ”Tärkeä seikka” ja ”keskeinen kysymys”: Mitä korpuslingvistinen analyysi paljastaa lähisynonymyista? *Virittäjä* 105(2). 170–192.
- Jantunen, Jarmo. 2015. Oppimiskontekstin vaikutus oppijanpragmatiikkaan: Astemääritteet leksikaalisina nallekarhuina. *Lähivõrdlusi. Lähivertailuja* 25. 105–136. <https://doi.org/10.5128/LV25.05>
- Jarvis, Scott. 2013. Capturing the Diversity in Lexical Diversity: Lexical Diversity. *Language Learning* 63. 87–106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Jääskeläinen, Anni. 2023. Adjektiivien polysemiaa konstruktioissa. *Virittäjä* 127(3). 316–344. <https://doi.org/10.23982/vir.84893>
- Kecskes, Istvan & Papp, Tünde. 2000. *Foreign language and mother tongue*. Lawrence Erlbaum Associates.
- Kekki, Niina & Ivaska, Ilmari. 2022. The use of synonymous adjectives of Finnish as a second language learners: Applying the MuPDAR(F) approach. *International journal of learner corpus research* 8(1). 67–96. <https://doi.org/10.1075/ijlcr.21006.kek>
- Kekki, Niina & Jytälä, Riitta & Parente-Čapková, Viola. 2023. Kulttuurisen kielenoppimisen jäljillä: Aikuisten suomenoppijoiden kokemuksia lukupiirissä. *Sananjalka* 65(65). 196–214. <https://doi.org/10.30673/sja.115631>
- Kielitoimiston sanakirja. 2022. <https://www.kielitoimistonsanakirja.fi> (Viitattu 1.8.2023)
- Kim, Myonghee. 2004. Literature Discussions in Adult L2 Learning. *Language and Education* 18(2). 145–166. <https://doi.org/10.1080/09500780408666872>
- Krashen, Stephen. 1993. *The power of reading: Insights from the research*. Libraries Unlimited.
- Langacker, Ronald. 1987. *Foundations of cognitive grammar. Vol. 1: Theoretical prerequisites*. Stanford University Press.
- Lee, Ching-Yin & Liu, Jyi-Shane. 2009. Effects of Collocation Information on Learning Lexical Semantics for Near Synonym Distinction. *International Journal of Computational Linguistics & Chinese Language Processing* 14(2). 205–220. <https://aclanthology.org/O09-4004>
- van Lier, Leo. 2000. From input to affordance: Social-interactive learning from an ecological perspective. Teoksessa Lantolf, James (toim.), *Sociocultural Theory and Second Language Learning*, 245–260. Oxford University Press.



- Liu, Dilin. 2013. Salience and construal in the use of synonymy: A study of two sets of near-synonymous nouns. *Cognitive Linguistics* 24(1). 67–113. <https://doi.org/10.1515/cog-2013-0003>
- Liu, Dilin & Zhong, Shouman. 2016. L2 vs. L1 Use of Synonymy: An Empirical Study of Synonym Use/Acquisition. *Applied Linguistics* 37(2). 239–261. <https://doi.org/10.1093/applin/amu022>
- Lu, Yuanwen. 2017. *A corpus study of collocation in Chinese learner English*. Routledge.
- Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139165570>
- Martin, Marilyn. 1984. Advanced Vocabulary Teaching: The Problem of Synonyms. *The Modern Language Journal* 68(2). 130–137.
- Nation, Paul. 2018. Reading a whole book to learn vocabulary. *ITL - International Journal of Applied Linguistics* 169(1). 30–43. <https://doi.org/10.1075/itl.00005.nat>
- Pietilä, Päivi & Merikivi, Riikka. 2014. The Impact of Free-time Reading on Foreign Language Vocabulary Development. *Journal of Language Teaching and Research* 5(1). 28–36. <https://doi.org/10.4304/jltr.5.1.28-36>
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing* [Software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reynolds, Barry. 2015. A Mixed-Methods Approach to Investigating First- and Second-Language Incidental Vocabulary Acquisition Through the Reading of Fiction. *Reading Research Quarterly* 50(1). 111–127. <https://doi.org/10.1002/rrq.88>
- Reynolds, Barry. 2022. Situated incidental vocabulary acquisition: The effects of in-class and out-of-class novel reading. *Applied Linguistics Review* 13(5). 705–733. <https://doi.org/10.1515/applirev-2019-0059>
- Reynolds, Barry & Ding, Chen. 2022. Effects of word-related factors on first and second language English readers' incidental acquisition of vocabulary through reading an authentic novel. *English Teaching: Practice & Critique* 21(2). 171–191. <https://doi.org/10.1108/ETPC-05-2021-0049>
- Ringbom, Håkan. 2007. *Cross-linguistic similarity in foreign language learning*. Multilingual Matters.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford U.P.
- Stefanowitsch, Anatol & Gries, Stefan. 2003. Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Vanhatalo, Ulla. 2003. Kyselytestit vs. Korpuslingvistiikka lähisynonyymien semanttisten sisältöjen arvioinnissa: Mitä vielä keskeisestä ja tärkeästä? *Virittäjä* 107(3). 351–369.
- Waring, Rob & Nation, Paul. 2004. Second language learning and incidental vocabulary learning. *Angles on the English Speaking World* 4. 97–110.
- Waring, Rob & Takaki, Misako. 2003. At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language* 15(2). 130–163.
- Webb, Stuart. 2008. The effects of context on incidental vocabulary learning. *Reading in a Foreign Language* 20(2). 232–245.
- Xiao, Richard & McEnery, Tony. 2006. Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective. *Applied Linguistics* 27(1). 103–129. <https://doi.org/10.1093/applin/ami045>

## Liite A Tutkimuskysely

Kyselylomake (Liite A) on saatavilla osoitteessa: <https://osf.io/d2b9y/>

## Yhteystiedot:

Niina Kekki  
Turun yliopisto  
[niina.kekki@utu.fi](mailto:niina.kekki@utu.fi)

Ilmari Ivaska  
Turun yliopisto  
[ilmari.ivaska@utu.fi](mailto:ilmari.ivaska@utu.fi)