



Lukijat sanaston monimuotoisuutta määrittämässä

Leksikaalisen diversiteetin tarkastelua määrällisen ja
laadullisen tutkimuksen rajapinnassa

MARI HONKO, SCOTT JARVIS JA SEPPO VAINIO

1 Johdanto

Sanasto on jatkuvasti kasvavana ja uudelleenmuotoutuvana kategoriana kiinnostava kohde kielellisen diversiteetin tutkijoille. Yleensä tekstin leksikaalista diversiteettiä eli sanastollista monimuotoisuutta on pidetty myönteisenä ominaisuutena paitsi siksi, että se mahdollistaa tarkkojen merkityssisältöjen ilmaisemisen, myös siksi, että sen on havaittu olevan yhteydessä tekstin laadukkuuteen sekä sen tuottajan – puhujan tai kirjoittajan – ominaisuuksiin laajemminkin (ks. seuraavaa lukua ja tarkemmin esim. Malvern, Richards, Chipere & Durán 2004). Hosmanin (2002: 374) mukaan leksikaalinen diversiteetti myös tekee tekstistä vastaanottajan kannalta kiinnostavamman.

Leksikaalinen diversiteetti on yksi keskeisistä toimivan vuorovaikutuksen osatekijöistä. Jotta tarkkojen merkitysten ilmaiseminen onnistuu, kielenkäyttäjien on hallittava sanastoa laajasti ja heillä on oltava myös tietoa sanojen käytöstä. Suppea valikoima puolestaan tarkoittaa kielenkäytön toisteisuutta: samojen sanojen käyttämistä yhä uudelleen. Harkitsematon toisto voi olla toimivan vuorovaikutuksen este, sillä se hämärtää tekstin merkityssuhteita (Witte & Faigley 1981: 197–199, 202). Harkitulla toistolla on sen sijaan vuorovaikutuksessa tarpeellisia funktioita, koska sen avulla voidaan muun muassa lisätä tekstin sidosteisuutta ja tehostaa haluttujen merkitysten välittymistä (Halliday & Hasan 1976: 288–290; Halliday 2004: 571–573).

Tutkimuksessa leksikaalista diversiteettiä on tietyn kielen sisällä mahdollista lähestyä kahdesta tarkastelukulmasta: rajaamalla tarkastelu itse tekstiin tiiviisti liittyviin kriteereihin (kuten tekstilajiin, tekstin ikään tai kohderyhmään) tai suuntaamalla huomio eri yksilöiden tapaan käyttää kieltä. Tässä artikkelissa keskitymme yksilönäkökulmaan ja siinä erityisesti leksikaalisen diversiteetin mittaamiseen ja arviointiin. Empiirinen aineistomme on kaksiosainen ja koostuu ensinnäkin S1- ja S2-oppimäärää opiskelevien koululaisten kirjoittamista suomenkielisistä kertomusteksteistä sekä toiseksi lukijoiden niille antamista leksikaalisen diversiteetin arvioista (ks. alalukuja 4.1–4.2). Olemme

tutkineet näiden subjektiivisten arvioiden yhdenmukaisuutta sekä yhteyttä kertomustekstien sanaston piirteisiin tilastollisten analyysien avulla (ks. tarkempia tutkimuskysymyksiä luvussa 2). Tällä tavalla selvittämme, kuinka yhdenmukaisia inhimillisten arvioijien käsitykset leksikaalisesta diversiteetistä ovat ja mihin sanaston piirteisiin käsitykset näyttävät nojautuvan.

Leksikaalinen diversiteetti on ominaisuus, joka tietyn varauksin karttuu kehittyvän kielitaidon myötä (ks. alalukua 3.2), minkä vuoksi sen mittaamista on aiemmin hyödynnetty erityisesti kielitaidon arviointiin liittyvässä tutkimuksessa. Tutkimusta on motivoinut yhtäältä tarve saada lisää tietoa kielenoppimisen etenemisestä ja toisaalta tarve kehittää entistä parempia arviointimenetelmiä.¹ Myös omalla tutkimuksellamme on aineistonsa sekä osin tutkimustehtävänsä ja metodiikkansakin puolesta kytköksiä oppijankielen tutkimukseen ja kielitaidon arvioinnin tutkimukseen. Vaikka käytämme tutkimuksessa oppijankielen aineistoa, emme tässä artikkelissa kuitenkaan tarkastele leksikaalista diversiteettiä kehityksellisenä piirteenä. Tutkimustamme ei myöskään voi pitää tyypillisenä arviointitutkimuksena, sillä se ei kohdistu kielenosaamisen tasoon, kielenoppimisen etenemiseen tai arvioinnin laadukkuuteen, jotka ovat kielitaidon arvioinnin tutkimuksen tavallisia kohteita.² Tutkimuksemme voi kuitenkin tuoda uutta tietoa ja tarjota uusia avauksia esimerkiksi suomi toisena kielenä -arviointitutkimukseen, mistä osoituksena voidaan pitää aiempaa lupaavaa tutkimusta leksikaalisen diversiteetin mallintamisesta englanti toisena kielenä -oppijoiden parissa (ks. Crossley, Salsbury, McNamara & Jarvis 2011a, 2011b). Siksi esittelemme taustaksi aiempaa tutkimusta myös kielenoppimisen ja arvioinnin kontekstista.³

Metodisesti lähestymistapamme on jopa lähempänä ekologian kuin lingvistiikan tutkimusperinnettä, sillä vaikka molemmat tieteenalat näkevät diversiteetin tärkeänä ilmiönä ja niin ekologinen ote kuin ekologiset metaforatkin ovat tulleet kielen tutkimukseenkin viimeistään 2000-luvun alussa (ks. esim. van Lier 2004), ekologian piirissä diversiteetin tilastollisesta mallintamisesta on huomattavasti enemmän kokemusta (Jarvis 2013a: 99).⁴ Inhimillisten arvioijien käyttäminen leksikaalisen diversiteetin tutkimuksessa on kansainvälisestikin vielä varsin tuore ilmiö. Sitä ei aiemmin ole sovellettu suomen kielen tutkimukseen, ja sille on löydettävissä ainoastaan muutama englannin kielen kontekstissa tehty verrokkitutkimus.⁵ Mielestämme se kuitenkin tar-

1. Leksikaalisen diversiteetin mittaamisella on kuitenkin ollut tutkimuksessa muitakin sovelluskohteita, kuten tietyn kielenpuhujan (kirjailijan, poliitikon, rikoksesta epäillyn) yksilöllisen tyylin selvittäminen.

2. Arviointitutkimusta erityisesti kielitaidon arvioinnin näkökulmasta erittelevät esimerkiksi Huhta ja Hildén (2015).

3. Kiitämme artikkelin nimettömiä arvioijia tutkimuksellista viitekehystä koskeneesta palautteesta ja lukuisista muista yksilöidyistä korjausehdotuksista, jotka ovat auttaneet meitä käsikirjoituksen jallostamisessa nykyiseen muotoonsa. Kiitämme lisäksi professori Anneli Pajusta tutkimuksen aineiston keruun tukemisesta.

4. Ks. tarkemmin esim. Cardinale, Duffy, Gonzalez, Hooper, Perrings, Venail, Narwani, Mace, Tilman, Wardle, Kinzig, Daily, Loreau, Grace, Larigauderie, Srivastava & Naeem 2012; Chao & Jost 2012.

5. Tiedossamme ei ole tutkimuksia, joissa inhimillistä arviointia olisi systemaattisesti käytetty nimenomaan suomenkielisten tekstien sanaston holistisessa arvioinnissa tai rajatun tiettyjen sanastollisten piirteiden tarkastelussa, vaikka sanastoa onkin voitu pitää yhtenä arviointiin vaikuttavana tekijänä. Se kuitenkin tiedetään, että niin koulutettujen kuin kouluttamattomienkin arvioijien näkemykset tekstien laadukkuudesta voivat poiketa toisistaan huomattavasti. Arviointituloksiin vaikuttavat arvioitavaan aineistoon ja arvioijiin itseensä liittyvien tekijöiden lisäksi muun muassa arviointitehtävä ja arviointikriteerit

joaa teoreettisesti perustellun keinon päästä lähemmäksi validia leksikaalisen diversiteetin kuvausta ja empiiristä tutkimusta (ks. myös lukua 3).

Tutkimuksemme lähtökohta on samantapainen kuin Crossleyn, Salsburyn, McNamaran ja Jarvisin (2011a, 2011b), Crossleyn, Salsburyn ja McNamaran (2012) sekä Jarvisin (2013b) tutkimuksissa, joissa on tarkasteltu englanninkielisiä tekstejä ja niiden arviointia. Tutkimuksissa havaittiin, että inhimillisten arvioijien käsitykset leksikaalisesta taidokkuudesta (*lexical proficiency*⁶) voidaan ennakoida jopa 60 prosentin tarkkuudella vain neljää tekstipiirrettä tarkastelemalla (Crossley ym. 2011b). Paras eli arvioita tarkimmin ennustava malli saatiin puhekielen aineistosta, joka on perusluonteeltaan omaa kirjoitetun kielen tekstiaineistoamme moniulotteisempi ja jonka holistisiin arvioihin voi olla vaikutusta muun muassa puheen prosodisilla piirteillä. Valtaosa myös muista kuin inhimillisiä arvioijia hyödyntäneistä leksikaalista diversiteettiä koskevista tutkimuksista käsittelee englannin kieltä, eivätkä tulokset ole suoraan sovitettavissa morfologiselta rakenteeltaan ja sananmuodostusperiaatteiltaan hyvin erilaiseen suomen kieleen (ks. myös Johansson 2008: 65). Tämän artikkelin yksi keskeinen tavoite onkin avata leksikaalisen diversiteetin tutkimuksen metodiikkaa nimenomaan suomen kielen kontekstissa.

Artikkeli rakentuu seuraavasti: Tutkimuskysymysten esittelyn (luku 2) jälkeen kokoamme luvussa 3 aiemman tutkimuksen antia eli leksikaalisen diversiteetin määrittelytapoja, tutkimushavaintoja sekä kriittisiäkin arvioita käytetyistä analyysimenetelmistä. Luvussa 4 kuvaamme tutkimuksen aineiston ja analyysimetodit. Tulokset esitämme luvussa 5 ja niiden keskeiset sisällöt implikaatioineen kokoamme päätösluvussa 6, jossa myös pohdimme jatkotutkimusmahdollisuuksia.

2 Tutkimuksen tavoitteet, tutkimuskysymykset ja tutkimuksen kulku

Tutkimuksemme empiirisessä osassa olemme tarkastelleet suomenkielisten kokonaistekstien leksikaalista diversiteettiä arviointimenetelmällä, jota toistaiseksi on kielitieteellisessä diversiteettitutkimuksessa hyödynnetty hyvin vähän.⁷ Menetelmä perustuu tekstin ja tulkinnan vertailuun: aikuisten lukijoiden käsityksiä koululaisten kirjoittamien kertomustekstien diversiteetistä verrataan tekstien sanastosta havaittaviin piirteisiin, joihin oletamme käsitysten nojautuvan.

Tutkimuksemme tuo uutta tietoa kahdella tavalla: Ensinnäkin käytämme leksikaalisen diversiteetin arvioinnissa uudenlaisia menetelmiä, joiden toimivuudesta tarvitaan tutkimukseen perustuvaa lisänäyttöä. Toiseksi uutta on se, että arvioinnin kohteena on nimenomaan suomenkielinen aineisto. Tutkimuskysymyksiä on kaksi:

tai niiden puute (ks. esim. Crossley & McNamara 2011; Guo, Crossley & McNamara 2013). Harjusen (2015) tutkimus, jossa tarkastellaan pohtivan tekstin arvioinnin ongelmia, osoittaa, että ainakin silloin, kun arviointitehtävä on väljä, eri arvioijien huomio kohdistuu osin hyvin erilaisiin ja eri tason ilmiöihin; tekstien ohella arvioija saattaa arvioida esimerkiksi tekstin kirjoittajan kypsyyttä tai sitoutuneisuutta kirjoitustehtävään.

6. Tutkimuksessa ei siis tarkalleen ottaen yritetty ennakoida sanaston monimuotoisuuden (*lexical diversity*) arvioita.

7. Toinen vaihtoehto olisi tarkastella kokonaisteksteistä poimittuja tekstinäytteitä, mihin kuitenkin liittyy useita epävarmuustekijöitä (ks. alalukua 3.3).

- 1) Kuinka yhdenmukaisia lukijoiden käsitykset ovat suomenkielisen tekstiaineiston leksikaalisesta diversiteetistä? Kysymykseen vastataan vertaamalla toisiinsa eri arvioijien antamia tekstikohtaisia pisteytyksiä tilastollisen analyysin avulla.
- 2) Mitkä tekstin sanastolliset piirteet vaikuttavat parhaiten selittävän lukijoiden käsityksiä sanaston monimuotoisuudesta? Piirteitä etsitään teoriaohjautuvasti rakentamalla regressioanalyysin avulla usean selittävän muuttujan regressiomalli. Mallin pohjana käytetään tekijöitä, joiden aikaisemman tutkimuksen perusteella tiedetään potentiaalisesti vaikuttavan leksikaaliseen diversiteettiin.

Tutkimuksen ensimmäisessä vaiheessa tarkastelimme lukijoiden (N = 23) käsityksiä tekstien sanastosta holististen arviointien avulla. Toisessa vaiheessa erittelimme tekstien kielellistä rakennetta niiden sanastoa kuvaavien muuttujien avulla, numeeritimme muuttujat ja kokosimme muuttujakohtaiset tiedot tietokannaksi. Kolmannessa vaiheessa tutkimme tilastollisten analyysien avulla, mihin sanaston piirteisiin tekstin lukijoiden käsitykset sanastollisesta monimuotoisuudesta tutkitussa aineistossa ensisijaisesti perustuvat, ja pyrimme rakentamaan käsityksiä selittävän mallin. Täydellisessä mallissa muuttujat selittäisivät holistiset arviot kokonaan eli eri arvioijien käsitykset (joihin malli perustuu) olisivat keskenään täysin yhdenmukaiset – jokaisessa kertomuksessa erikseen. Näin suuri yhdenmukaisuus on tosin lähinnä teoreettinen ajatus eikä saa tukea aikaisemmasta tutkimuksesta (ks. esim. Jarvis 2017).

3 Leksikaalinen diversiteetti tutkimuksen kohteena

Leksikaalisen diversiteetin määrittely ja käyttö tutkimuksessa on ollut kirjavaa. Tässä luvussa esittelemme leksikaalista diversiteettiä sekä aikaisemman tutkimuksen kohteena että oman tutkimuksemme rajausten näkökulmasta.

3.1 Mitä diversiteetti on?

Leksikaalisen diversiteetin käsite on kielitieteellisessä kirjallisuudessa ollut käytössä jo 1930-luvun lopulta (*diversity of vocabulary*, Carroll 1938), mutta sen määritelmä on yhä vakiintumaton ja tutkimuksen tarkastelunäkökulmasta riippuvainen (ks. Jarvis 2013b: 15). Osin vakiintumattomuus johtuu siitä, että leksikaalista diversiteettiä on aiemmassa tutkimuksessa lähestytty kahdesta toisistaan poikkeavasta näkökulmasta eli teksti- ja yksilönäkökulmasta käsin. Tekstinäkökulmaa painotettaessa sanaston monimuotoisuutta on tarkasteltu joko yleisesti kielen tarjoamien ilmaisuvarojen tai konkreettisen tekstin sisältämän sanaston näkökulmasta. Silloin synonyymisina ilmauksina leksikaaliselle diversiteetille on käytetty esimerkiksi sanaston *variaatiota*, *vaihtelua* ja *joustavuutta* sekä laiveammin *eroja* tekstien sanastossa.

Yksilönäkökulmaa painotettaessa puhutun tai kirjoitetun tekstin sanastoa on käytetty lähinnä välineenä, jonka tavoitteena on saada ote tekstin tuottaneen yksilön kieli-

taidosta. Tällöin sanaston monimuotoisuuden on ajateltu tarjoavan kielitaidon kuvaamiseen keinon, joka mahdollistaa myös yksilöiden välisen vertailun ja yksilön kielitaidossa tapahtuvien muutosten havainnoinnin. Kielitaitoa ajatellen sanaston monimuotoisuuden on katsottu antavan tietoa muun muassa yksilön sanavaraston laajuudesta ja kompleksisuudesta, kielellisestä kypsyydestä, verbaalisesta luovuudesta sekä kyvystä käyttää leksikaalisia resursseja tehokkaasti. (Ks. esim. Malvern ym. 2004; Jarvis 2013a.) Sekä teksti- että yksilönäkökulmaa painottavissa tutkimuksissa analyysimetodinä on lähes poikkeuksetta ollut aineistopohjainen tekstianalyysi⁸, jossa huomiota on kiinnitetty lähinnä aineistossa käytetyn sanaston määrään ja toisteisuuteen.

Määrää ja toisteisuutta on diversiteettitutkimuksissa tyypillisesti mitattu tekstin kokonaissanemäärän (määrä) ja erilaisten sanojen määrän (runsaus) avulla. Diversiteetin kannalta merkitystä on katsottu olevan myös sanojen erityisyydellä tarkastellussa aineistossa sekä jakauman tasaisuudella eli tekstin sanojen toisteisuudella suhteessa toisiinsa. Sironnalla puolestaan tarkoitetaan diversiteettitutkimuksessa tietyn sanan eri esiintymien etäisyyttä tai kasautumista tarkasteltavassa tekstissä. Koko sanaston vaihtelevuutta, joka kuvaa sekä eri sanojen määrää että sijoittumista, on esitetty kuudenneksi keskeiseksi leksikaalisen diversiteetin osatekijäksi. (Ks. tarkemmin alalukua 4.4; Jarvis 2013b: 25.)

Tässä artikkelissa termiä *leksikaalinen diversiteetti* käytetään synonyymisesti *sanaston monimuotoisuuden* kanssa (ks. myös Malin 2012). Sen sijaan leksikaaliseen diversiteettiin toistuvasti rinnastetun ja suomenkielisessäkin tutkimuskirjallisuudessa esiintyvän *leksikaalisen rikkauden* katsomme Jarvisin (2013b: 21, 38) lailla omaksi erilliseksi käsitteekseen, tekstin eri sanojen määräksi, joka tosin saattaa olla leksikaalisen diversiteetin yksi osatekijä (ks. alalukua 5.3). Rajaus on perusteltu, sillä alun perin leksikaalisella rikkauksella on kielitieteellisessä tutkimuksessa tarkoitettu juuri sanojen määrää yksilön mentaalileksikossa (Yule 1944) – ja samalla vain yhtä diversiteetin oletetuista osatekijöistä (ks. myös alalukua 4.4).

3.2 Yksilönäkökulma leksikaaliseen diversiteettiin

Sanaston moninaisuudella on itseisarvoa, joka ei ensisijaisesti liity kieltä käyttäviin yksilöihin vaan perustuu kielen vuorovaikutuksellisiin tehtäviin. Yhteistä etenkin yksilönäkökulmaa painottaville sanaston monimuotoisuuden tutkimuksille kuitenkin on, että runsaaseen diversiteettiin katsotaan yhdistyvän positiivisia arvoja, jotka kuvaavat myös kielenkäyttäjän ominaisuuksia – eivät pelkästään kyseistä kieltä, tekstiä tai vuorovaikutustilannetta. Tällaisia runsaaseen leksikaaliseen diversiteettiin yhdistettyjä piirteitä ovat muun muassa käsitykset puhujan korkeasta statuksesta (ks. Hosman 2002: 374–375) sekä kommunikatiivisesta kyvykkyydestä (ks. esim. Burroughs 1991). Näyttää myös siltä, että kuulijan myötämielisyyden puhujaa kohtaan on tietyn ehdoin yhteydessä puheen leksikaaliseen diversiteettiin (Bradac & Wisegarver 1984). Selkeimpiä tulokset ovat olleet var-

8. Korpuspohjaisia analyysimenetelmiä on esitelty mm. Tognini-Bonelli (2001: 65). Kaikkia aineistoja ei voida pitää varsinaisina korpuksina, jolloin korpuspohjaisuuden sijaan olisi täsmällisempää puhua aineistopohjaisuudesta (ks. Jantunen 2009: 102).

sinaisissa kielitaitotutkimuksissa, joissa on tutkittu sanaston käytön eroja aikuisten ja lasten, eri-ikäisten lasten, tyypillisesti ja epätyypillisesti kehittyvien lasten, ensikielisten ja toisen kielen puhujien sekä oppimisen eri vaiheissa olevien toisen kielen puhujien välillä.⁹

Tutkimuksissa, joissa leksikaalista diversiteettiä on tarkasteltu kehityksellisenä piirteenä, sen on – tosin kirjavasti määriteltynä – ajateltu antavan tietoa erityisesti sanavaraston koosta: mitä suurempi yksilön mentaalinen leksikko tietyllä hetkellä on, sitä suurempi kielellinen potentiaali hänellä on käytettävissään ja sitä todennäköisemmin hän myös käyttää sanastollisia resurssiaan tehokkaasti (Malvern ym. 2004: 5; Jarvis 2013b: 17; Wang 2014: 84). Kielitaito ei kehity tyhjiössä, ja jo pikkulapsivaiheen interventiotutkimukset ovat osoittaneet, että hoivapuheen sanaston variointi (ei ainoastaan sen määrä) selittää sanastollisten taitojen kehittymistä. Hoivapuheen laatuun taas vaikuttaa vahvasti esimerkiksi perheen sosioekonominen asema, mutta sosioekonomista asemaa määräävämpi tekijä lapsen kielenkehityksen kannalta on silti kielellinen ympäristö, joka voi olla yksilötasolla rikas tai köyhä riippumatta sosioekonomisesta asemasta (Hoff 2003; Black, Peppé & Gibbon 2008; Rowe 2012).

Vaikka havaitut yhteydet eivät ole lineaarisia, ne ilmenevät riittävän suuressa ja heterogeenisessä aineistossa lähes aina samansuuntaisina ja käyttöperusteista kielenomaksumisen teoriaa tukien: suurempi kielenkäyttökokemus ja monipuolisempi kieliympäristö yhdistyvät monimuotoisemman sanaston käyttöön niin puheessa kuin kirjoituksessakin, ja uusien sanojen oppiminen edellyttää niiden toistumista erilaisissa muodoissa, ympäristöissä ja käyttötilanteissa.¹⁰ Poikkeuksen muodostavat tutkimukset, joissa on tarkasteltu kielellistä attritiota, sillä sanaston monimuotoisuuden on havaittu vähenevän sekä muunkielisessä ympäristössä tapahtuvan ensikielen hii-pumisen että afasian yhteydessä.¹¹

Sanaston vaihtelevuus voi olla haitaksi, jos kohderyhmää, viestintätilannetta ja viestinnän tavoitteita ei oteta huomioon. Sanaston säästeliäs käyttö varoinnin sijaan on tarkoituksenmukaista vaikkapa selkokielisissä teksteissä sekä tuotettaessa eritteleviä tekstejä, joissa on pyrittävä sisällölliseen yksiselitteisyyteen esimerkiksi johdonmukaisten terminologisten valintojen kautta. Leksikaalinen diversiteetti onkin havaittu keskimäärin runsaammaksi kertovissa teksteissä tai tekstiosuuksissa kuin muissa tutkituissa tekstityypeissä.¹² Hallitsemattoman diversiteetin ongelmat korostuvat puhutussa vuorovaikutuksessa, jossa tilanneherkkyys on onnistuneen kommunikaation edellytys ja jossa esimerkiksi tekstin vähäisempi prosessointiaika ja tarve kompen-

9. Ks. Grobe 1981; Watkins, Kelly, Harbers & Hollis 1995; Scott & Windsor 2000: 332; Jarvis 2002: 74–76; Dewaele & Pavlenko 2003: 132–133; Durán, Malvern, Richards & Chipere 2004: 229–234; Johansson 2008: 71, 75; Unsworth 2008: 317–318, 322; Gui 2010; Yu 2010; Berman, Nayditz & Ravid 2011; Crossley, Salsbury & McNamara 2012; Higgins, Xi, Zechner & Williamson 2011; Housen, Schoonjans, Janssens, Welcomme, Schoonheere & Pierrard 2011; Kormos 2011.

10. Vermeer 2001; Tomasello 2003: 5, 15, 46–48, 80–81; Berman 2007: 349; Müller Gathercole & Hoff 2007: 114–118.

11. Bastiaanse & Jonkers 1998; Korpijaakko-Huuhka 2003: 170–173; MacWhinney, Fromm, Holland, Forbes & Wright 2010; Crepaldi, Ingignoli, Verga, Contardi, Semenza & Luzzatti 2011; Schmid & Jarvis 2014.

12. Vertailtavina tiedottava, suostutteleva (ohjaava), vertaileva (erittelevä) ja argumentoiva (kantaa ottava) tekstityyppi (ks. esim. Johansson 2008; Olinghouse & Wilson 2013; Sadeghi & Dilmaghani 2013). Yksilönäkökulmaa painottava tutkimus ei voi sivuuttaa myöskään tekstinäkökulmaa, koska jo aineiston laatu (vs. muut tekstityypit) voi diversiteettitutkimuksessa vaikuttaa tutkimustuloksiin.

soida kuulemista vaikeuttavan taustahälyn vaikutuksia edellyttävät yleensä kirjoitetun kieltä enemmän redundanssia (Bremer, Broeder, Roberts, Simonot & Vasseur 1993: 159–161). Myös kielellisen ilmaisun suunnitteluun suunnatuilla voimavaroilla näyttää olevan merkitystä sanaston varioinnille, mikä on havaittu tutkimalla valehtelun vaikutusta tekstin rakentumiseen (Dulaney 1982; Carpenter 1990).

Suomen kielen alalla leksikaalista diversiteettiä on tutkittu hyvin niukasti. Tuoreet tutkimusjulkaisut on tehty kielenoppimisen kontekstissa, ja niissä on keskitytty kirjoitetun leksikaalisen diversiteetin muutoksiin osana oppijankielen kehitystä. Kouluikäisten kielitaitoon pureutuvan tutkimushankkeen (Later language development, Tampereen yliopisto, joht. prof. Anneli Pajunen) seurantatutkimuksen (Honko 2013) mukaan alakouluvaiheen kielenkehitys näkyy kirjoitetussa kielessä lähes suoraviivaisesti sanaston lisääntyvänä variointina sekä yksilö- että ryhmätasolla ja niin ensikielisten kuin toisen kielen oppijoiden ryhmässä. Pitkästä maassaoloajasta ja usean vuoden koulunkäynnistä huolimatta sanaston käyttö on toisen kielen oppijoiden ryhmässä kuitenkin merkittävästi toisteisempaa. Tämän voi ajatella kertovan leksikaalisten taitoerojen pysyvyydestä yksilötasolla sekä leksikaalisen diversiteetin potentiaalista suomen kielen oppimisen tutkimuksessa ja arvioinnissa. Cefling-hankkeessa (Jyväskylän yliopisto, joht. prof. Maisa Martin) puolestaan on tutkittu lukuisten kielenpiirteiden esiintyvyyttä suomi toisena kielenä -oppijoiden teksteissä, jotka edustavat eri taitotasoa¹³. Cefling-aineistosta pro gradu -tutkielmansa tehnyt Malin (2012) havaitsi, että yläkoululaisten ja aikuisten kirjoittamissa teksteissä sanasto monipuolistuu taitotasolta ylempälle siirtyäessä, mutta kasvu ei ole lineaarista eikä kaikissa tekstilajeissa yhtä voimakasta. Tarkastellussa aineistossa olennainen muutos oli lisäksi havaittavissa vasta ylempien taitotasojen B ja C välillä, mikä vastaa Wangin (2014) tuloksia englannin kielen oppijoiden kirjoittamisesta: kielenomaksumisen alkuvaiheessa kielitaidon kehittymiseen ei välttämättä vielä saada otetta leksikaalista diversiteettiä tarkastelemalla.

Saarela (1997), Pajunen (2012) ja Honko (2013) sekä Laine-Leinonen (2013) ovat – osin toisistaan poikkeavia menetelmiä käyttäen – havainneet sanastollisen monimuotoisuuden lisääntyvän peruskouluikäisten lasten teksteissä koulunkäynnin edessä lapsen kielitaustasta riippumatta. Tutkimusten mukaan tekstit samaan aikaan yleensä myös pitenevät selvästi, niissä käytetään aiempaa useammanlaisia sekä kompleksisempia rakenteita (esimerkiksi yhdyslauseita ja monimorfeemisia sanoja) ja ne toteuttavat paremmin tekstilajinsa ominaispiirteitä. Erityisesti alakouluikäisissä sekä tekstitaitojen karttumisen että kognitiivinen kehitys on kiivasta. Riittävän suuressa aineistossa muutokset näkyvät kasvavina arvoina useimpia kehityksellisiä mittareita käytettäessä, ja vasta alakoulun jälkeen ilmaisun tiivistyminen alkaa toimia tätä kasvutendenssiä vastaan (ks. Pajunen 2012: 7).

Suomalaisessa koulujärjestelmässä erot tyttöjen ja poikien välisissä tekstitaidoissa ovat viime vuosina herättäneet toistuvasti huomiota ja huolta (Kupari, Sulkunen, Vetteranta & Nissinen 2012; Pajunen 2012; Jakku-Sihvonen 2013; Arffman & Nissinen 2015). Myös kirjoitelmien sanastosta tehdyt havainnot ovat samansuuntaisia: vaikka sanas-

13. Taitotasot määriteltiin kielten oppimisen, opettamisen ja arvioinnin yhteisen eurooppalaisen viitekehyksen mukaan.

ton käyttö monipuolistuu kertovassa kirjoittamisessa kielitaustasta riippumatta ja erikseen sekä tyttöjen että poikien ryhmissä, monipuolisinta sanaston käyttö on tutkimusten mukaan tyypillisesti tyttöjen kirjoittamissa teksteissä (Saarela 1997; Honko 2013; Laine-Leinonen 2013).

Suomen kielen leksikaalisesta diversiteetistä parinkymmenen viime vuoden aikana tehdyt tutkimukset ovat kohdistuneet lähes yksinomaan kirjoitettuun oppijankieleen ja erityisesti kertoviin teksteihin. Metodisesti tutkimukset ovat noudatelleet kansainvälisen tutkimuksen ratkaisuja: Malin (2012) käytti tutkimuksessaan TTR-sanatoisteisuuslukua (*type/token-ratio*), MTLD-rikkauslukua ja Shannonin indeksiä, joka on ekologiainkin piiristä tuttu entropiaan perustuva tunnusluku.¹⁴ Saarela (1997) hyödynsi tutkimuksessaan Malinin tapaan Shannonin indeksiä, Laine-Leinonen (2013) puolestaan sekä Shannonin indeksiä että erilaisia rikkauslukuja (ks. mts. 82). Pajunen (2012) tarkasteli muutoksia koululaisten käyttämien verbilekseemien jakaumassa, ja Honko (2013) käytti tutkimuksessaan perinteisten sanatoisteisuuslukupien lisäksi muun muassa sanojen esiintymistodennäköisyyteen perustuvaa SOP-diversiteetti-indeksiä. Metodiikka kaipaisi kuitenkin päivittämistä. Syitä päivittämisen tarpeeseen sekä omaa ehdotustamme diversiteettitutkimuksen metodikeskusteluihin ja -kokeiluihin käsittelemme yksityiskohtaisemmin seuraavissa alaluvuissa.

3.3 Leksikaalisen diversiteetin tutkimuksen metodiikkaa

Kielitieteellisen tutkimuksen piiriin muotoutui jo 1900-luvun alkupuolella vahva perinne: sanaston monimuotoisuutta arvioitiin joko suoraan tai epäsuorasti mittaamalla tietyn tekstinäytteen sana–sane-suhdetta. Hyvin suosituksi arviointimenetelmäksi nousi nopeasti etenkin TTR-arvo (*type/token ratio*), joka tarkoittaa tekstin erilaisten sanojen (lekseemien/lemmujen) määrän jakamista tekstisanojen (saneiden) kokonaismäärällä. TTR:n ja sen sovellusten suosio on kestänyt vuosikymmeniä. Miller (1981) on nähnyt TTR:llä jopa diagnostista potentiaalia esittäessään, että riittävä sanatoisteisuus ($TTR < 0,50$) voisi auttaa kielihäiriöiden tunnistamisessa (ks. Coene, Schauwers, Gillis, Rooryck & Govaerts 2010: 8). Jo varhaisista sana–sane-suhdetta leksikaalisen diversiteetin (Carroll 1938) sekä sanaston ”joustavuuden” tai ”vaihtelevuuden” (Johnson 1939, 1944) mittarina¹⁵ esitelleistä julkaisuista käy kuitenkin ilmi yksi TTR:n keskeinen ominaisuus, nimittäin sen riippuvuus otoskoosta. Lyhyessä tekstissä sanaston toistuminen on epätodennäköisempää kuin pitkässä, ja siksi TTR saa lyhyissä teksteissä suhteessa suurempia arvoja. Ratkaisuksi erikokoisten aineistojen vertailuun on jo varhain esitetty TTR:n laskemista tasamittaisista otoksista (ks. esim. Johnson 1944; Baker 2010: 21). Tätä MSTTR:ksi (*mean segmental type/token ratio*) nimitettyä sovellusta on pidetty yhtenä harvoista teknisesti onnistuneista yrityksistä tasoittaa aineistojen kokoeroja (Jarvis 2013a: 91; 2013b: 17). Tutkimustulosten vertailtavuuden ongelma on kuitenkin säilynyt myös

14. Palaamme näihin termeihin hieman myöhemmin, ks. TTR:stä alalukua 3.3 ja MTLD:stä alalukua 4.4.

15. Sanalla *mittari* viittamme tässä tekstissä erilaisiin arviointityökaluihin, joilla leksikaalista diversiteettiä on pyritty mittaamaan. Esimerkiksi TTR ja rakentamamme tilastollinen malli ovat tällaisia mittareita.

MSTTR:ää käyttäneissä tutkimuksissa lähinnä siksi, että eri tutkimuksissa on päädytty erilaajuisiin otoksiin – esimerkiksi 200:aan tai 2 000 sanaan (ks. myös Jarvis 2002: 58). Toinen olennainen puute menetelmässä on se, että tasamittaisia otoksia käytettäessä osa aineistosta jää lähes aina tarkastelun ulkopuolelle.

Vähitellen TTR-pohjaisista mittareista on pyritty hiomaan yhä tarkempia eli ennen kaikkea yhä paremmin otoskoon vaihtelut huomioon ottavia arviointityökaluja. Silti näidenkin mittarien (mm. Yulen [1944] indeksi, Shannonin [1948] indeksi, Guiraud'n [1954] indeksi, Herdanin [1964]) indeksi ja Uberin [1978] indeksi) on yksi toisensa jälkeen todettu kärsivän esikuviansa puutteista (ks. tarkemmin esim. Jarvis 2002). Sofistikoituneimpien indeksien, kuten Yulen (1944) ja Simpsonin (1949) indeksien sekä vocd D -diversiteettimittarin, avulla on samalla pureuduttu sanaston esiintymistodennäköisyyteen tutkituissa teksteissä (ks. esim. McCarthy & Jarvis 2007; Jarvis 2013a, 2013b). Jarvisin (2013a: 94) mukaan ainoat otoskoosta todellisuudessa riippumattomat indeksit ovat silti yhä jo aiemmin mainittu Johnsonin MSTTR sekä McCarthyn (2005) MTLD, jotka itse asiassa ovat toimintaperiaatteiltaan toistensa peilikuvia (ks. myös McCarthy & Jarvis 2010). MSTTR perustuu TTR-keskiarvon laskemiseen vakioittaisista tekstikatkelmista, MTLD puolestaan TTR:n vakiointiin. Siksi myös menetelmien perusongelma on sama: koska menetelmät eivät käsittele tekstejä yhtenä kokonaisuutena, niitä ei myöskään voida pitää täysin validina tapana tutkia tai arvioida kokonaistekstejä. (Ks. tarkemmin Jarvis 2013a: 94.)

Tekstien pirstaloiminen ei tosin ole diversiteettitutkimuksen suurin ongelma. Kun vaivihkaa tutkimuksessa on ajautettu eräänlaiseen metodilaatikkoon eli tutkimaan yhä uudestaan diversiteettimittarien soveltuvuutta erilaajuisien aineistojen analysointiin ja niillä saatavien tulosten korrelointia kielitaitomuuttujiin. Samalla on lähes kokonaan unohtunut, millaiseen teoreettiseen taustaan – jos mihinkään – kehitellyt mittarit nojaavat tai mikä niiden tarttumapinta leksikaaliseen diversiteettiin itseensä on (Jarvis 2013a: 94–95; 2013b: 13, 17–18). Ongelmaksi on siis muotoutunut tutkimuksen sisäisen validiteetin puute; eihän ole merkityksellistä, kuinka hyvin tietty diversiteettimittari soveltuu erilaisten ja erimittaisten aineistojen vertailuun, jos mittari ei mittaa haluttua asiaa. Kun tutkitaan leksikaalista diversiteettiä, ensimmäisen esitettävän kysymyksen tulisikin nähdäksemme olla, missä (ei mistä) tietyn tekstin sanaston monimuotoisuus rakentuu. Onko se puhtaasti tekstin ominaisuus, määrittykö se vasta tekstin tulkitsijan mielessä vai onko kyse näiden vaihtoehtojen yhdistelmästä?¹⁶

Jos diversiteetti olisi vain tekstin ominaisuus eikä sillä olisi vaikutusta kielen vuoro-vaikutuksellisten funktioiden toteutumiseen, tekstien tulkintaan tai niiden tuottajasta muodostuviin käsityksiin, sen tarkastelu menettäisi olennaisesti kiinnostavuuttaan. Vain kieli – rajatummin kielen piirteiden kuten sanaston vaihtelu –, joka välittää merkityksiä, on sosiaalisessa mielessä olemassa. Jos tietyn tekstin sanaston vaihtelun lisääminen esimerkiksi nimettyjen käsitteiden paradigmaattisia suhteita hyödyntämällä (*rakennus, kivitalo, raunio, koti* vs. *talo*) ei vaikuttaisi tekstin välittämiin merkityksiin, vaihtelu olisi olemassa tekstissä mutta ei tulkinnoissa. Esimerkiksi sanaston toisteisuuden etuja ja haittoja

16. Samoja kysymyksiä ovat Zipfin (1935) pohjalta jo aiemmin pohtineet Jarvis (2017) sekä Castañeda-Jiménez ja Jarvis (2014), vaikkakaan eivät täsmälleen samasta näkökulmasta tai samoin argumentein.

arvioitaessa nimenomaan tulkinta on kuitenkin ratkaisevassa asemassa, ja siksi tekstin leksikaalisen diversiteetin määrittelyssä sen tulkinnan huomioon ottaminen on perusteltua, ellei jopa välttämätöntä. Tulkinta taas on vahvasti kontekstisidonnaista eli muun muassa tekstilajista, lukemisen tarkoituksesta ja tulkitsijan ominaisuuksista riippuvaista. Jos kontekstitekijät jätetään huomiotta, saatetaan diversiteettitarkasteluissa saada matemaattisesti päteviä mutta kielenkäyttötilanteiden kannalta järjettömiä tuloksia.

Yksi tapa edistää leksikaalisen diversiteetin määrittelyä on korostaa inhimillisen tulkinnan merkitystä. Leksikaalisen diversiteetin mittareita kehittämään pyrkivissä tutkimuksissa näin on toistaiseksi kuitenkin tehty vain harvoin. Tulkinnot eivät toki voi olla täysin subjektiivisia eikä minkään tekstin sanaston monimuotoisuus toisaalta siis pelkästään sen vastaanottajan mielessä syntyvä ominaisuus. Jotta kielen merkityksiä välittävä tehtävä voisi toteutua myös eri yksilöiden välillä, pelkästään yksilölliset ja tilannesidonnaiset tulkinnot eivät siis tarjoa ratkaisua.

Huomionarvoista on, että vaikka sekä tekstiin että tulkintaan kapeasti rajautuvalla tutkimuksella on selvät puutteensa, leksikaalista diversiteettiä on tutkimuksissa tähän asti lähestytty melkein aina pelkästään jommastakummasta näkökulmasta. Itse uskomme, että ratkaisu piilee näkökulmien yhdistämisessä, toisin sanoen tekstin lingvististen piirteiden ja subjektiivisen tulkinnan yhdistämisessä.

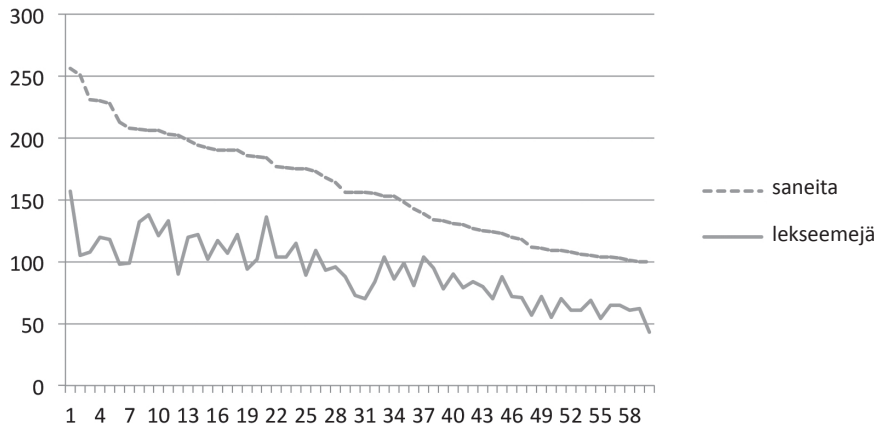
4 Aineistot ja analyysimenetelmät

Leksikaalisen diversiteetin arvioinnin lähtökohtana on aina teksti tai laajempi tekstiaineisto. Itse arviointi voi olla teksti- tai kokijälähtöistä tai näitä tapoja yhdistävää. Seuraavissa alaluvuissa esitellään tarkemmin tutkimuksen aineistot sekä niiden analysoinnissa käytetyt menetelmät.

4.1 Arvioinnin kohteena oleva tekstiaineisto

Aineisto, johon leksikaalisen diversiteetin arviot kohdistuvat, koostuu 60 lyhyestä kertomustekstistä. Tekstit ovat viides- ja kuudesluokkalaisten eli 10–12-vuotiaiden koululaisten yhtenevällä tehtävänannolla kirjoittamia temaattisia kertomuksia aiheesta ”unelmien päivä” ja osa suurempaa tekstikorpusta (Honko 2013). Alun perin tekstit kirjoitettiin käsin mutta syötettiin myöhemmin tietokannaksi ja lemmattiin, minkä jälkeen niiden sanastollista rakennetta analysoitiin lukuisten eri muuttujien avulla sekä sane- että lekseemitasolla. Tekstien pituus vaihtelee 100 saneesta 256 saneeseen; niiden keskipituus on noin 159 sanetta (keskihajonta n. 42 sanetta). Koska aivan lyhyet tekstit olisivat soveltuneet sisällöllisen suppeuden ja koherenssin puutteen vuoksi huonosti vertailukelpoiseen holistiseen arviointiin, ensimmäinen määräävä kriteeri tekstiaineiston otannassa oli juuri tekstin riittävä pituus (vähintään 100 sanetta). Lyhyissä teksteissä sanasto ei myöskään ehdi toistua. Lyhyiden tekstien arviointi olisi toisin sanoen helposti johtanut tutkimustehtävän kannalta epäluotettavaan tulkintoihin, sillä diversiteetin arviointi perustuu aiemman tutkimuksen perusteella ainakin osittain juuri sanaston toisteisuuteen (ks. Jarvis 2013b: 22–23).

Tekstin pituuden tiedetään olevan suoraan sidoksissa lukuisiin sanastollisiin piirteisiin, kuten erilaisten sanojen määrään, kerran esiintyvien sanojen määrään sekä yksilöllisten sanojen määrään. Myös omassa tekstiaineistossamme lekseemien¹⁷ määrä mukalee odotuksenmukaisella tavalla tekstien kokonaispituutta eli sanemäärää (kuvio 1).



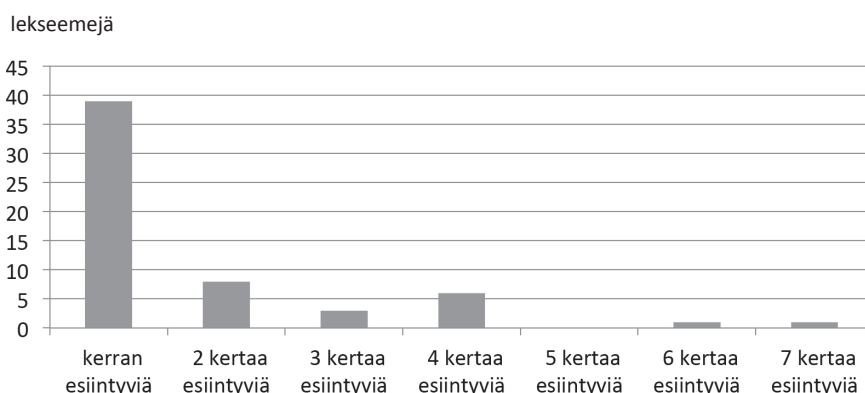
Kuvio 1.

Tekstiaineiston pituusjakauma. Kuviossa on esitetty kunkin kertomustekstin kokonaispituus saneina (katkoviiva, vaihteluväli 100–256) sekä siinä esiintyvien eri sanojen määrä lekseemeinä (yhtenäinen viiva, vaihteluväli 43–157). Y-akseli osoittaa saneiden ja lekseemien lukumäärän tekstikohtaisesti, ja tekstit on asetettu kuviossa pituusjärjestykseen kokonais-sanemäärän perusteella. Havaintoarvoja yhdistävä viiva on kuvion lukemista helpottava visuaalinen valinta.

Lyhyehköissä teksteissä pituutta kartutetaan siis pääosin uusilla sanoilla sen sijaan, että toistettaisiin jo käytettyjä sanoja. Tästä kertoo myös aineiston hapaks legomenon -sanojen eli tekstissä ainoastaan kerran esiintyvien sanojen suuri tekstikohtainen määrä, keskimäärin noin 65 sanaa tekstiä kohden. Suuri on niin ikään hapaks legomenon -sanojen tekstikohtainen osuus kaikista lekseemeistä (ka. n. 69 %) ja saneista (ka. n. 41 %). Keskimäärin kaksi viidestä kertomustekstien sanoista on toisin sanoen sellaisia, ettei niitä ensiesiintymisen jälkeen käytetä tekstissä enää uudelleen. Kertomustekstien keskimääräinen sanepituus (6,0 grafeemia) ja lekseemipituus (5,2 grafeemia) sen sijaan ovat pienemmät kuin suomenkielisissä teksteissä keskimäärin, mikä selittyy ennen kaikkea tekstisanaston morfologisen rakenteen suhteellisella yksinkertaisuudella (ks. tarkemmin Honko 2013: 169–171).

17. Lekseemi on sanan abstraktio, johon kuuluvat sanan ortografiset, foneettiset, kieliopilliset ja semanttiset variaatiot, kuten eri taivutusmuodot ja virheellisesti kirjoitetut muodot. Lekseemiä edustaa (esim. pituusanalyysseissa) sen hakumuoto: nomineilla yksikön nominatiivi ja verbeillä A-infinitiivin lyhyt muoto. Johdokset ja yhdyssanat (eli sananmuodostuskeinoilla muodostetut sanat) katsotaan leksikaalistuneiksi riippumatta niiden vakiintuneisuudesta kielenkäyttöön. (Ks. tarkemmin lekseemin määrittelystä tämän aineiston käsittelyssä Honko 2013: 52–59.)

Zipfin lain mukaan suhteellisen pieni sanajoukko kattaa kielestä riippumatta valtaosan tekstimassasta: luonnollisten kielten n :nneksi yleisimmän sanan tekstitiheys on kääntäen verrannollinen lukuun n .¹⁸ Zipfin lain toimintaperiaate käy ilmi myös lyhyen esimerkkitekstimme sanastoa tarkastelemalla. Noin kaksi kolmasosaa (67 %) lekseemeistä esiintyy tekstissä ainoastaan kerran ja vajaa seitsemännes kaksi kertaa, kun kolmesti tai useamman kerran esiintyviä sanoja on vain muutama (kuvio 2).



Kuvio 2.
Lekseemien frekvenssijakauma esimerkkitextomuksessa.

Tekstipituuden lisäksi taustamuuttujista kontrolloitiin kirjoittajan ikä (vuosiluokka), sukupuoli ja ensikieli (suomi/muu) siten, että näiden kolmen muuttujan suhteen kaikkia kombinaatioita oli aineistossa tasamäärä. Näin toimittiin, sillä aineistosta haluttiin mahdollisimman edustava ja juuri kyseisten muuttujien voitiin aiemman tutkimuksen perusteella olettaa vaikuttavan merkittävästi tekstien sanaston monimuotoisuuteen. Ennen leksikaalisen diversiteetin holistista arviointia tekstien kieliasu huollettiin ja tekstisanojen rakenne muokattiin todennäköisten tavoitemuotojen mukaiseksi.¹⁹ Tämä tehtiin, jotta arvioijien huomio ei kiinnittyisi arviointitehtävän kannalta epäolennaisiin tarkkuuden puutteisiin, jotka vielä alakouluikäisten kirjoittamisessa ovat melko yleisiä sekä kakkoskielillä että ensikielillä kirjoittajilla.

18. Zipfin lain toi suuren yleisön tietoisuuteen jo 1930-luvulla amerikkalainen tutkija George Kingsley Zipf. Keskustelu sen toteutumisesta eri kielissä ja toteutumisen reunaehdoista kuitenkin jatkuu yhä (ks. esim. Xiao 2008, jossa käsitellään Zipfin lain soveltuvuutta kiinan kielen tarkasteluun).

19. Kaikki korjaukset tehtiin juoksevan tekstin rinnalle Excel-tietokannassa. Pääosin korjauksia tehtiin virheelliseen taivutusmorfologiaan sekä virheellisiin äännekestoisiin ja puuttuviin tai väärin kirjaimiin (ks. korjausperiaatteista myös Honko 2013: 144–145). Virheellisiä tai epäkonventionaalisia sanavalintoja (joiden määrä lopulta osoittautui vähäiseksi) ei sen sijaan muutettu, jotta tekstien leksikko säilyisi alkuperäisenä. Muutoksista on keskusteltu tutkimusryhmässä, ja lopulliset korjaukset on tehnyt yksi henkilö.

4.2 Lukijoiden antamat diversiteettiarviot

Tekstiaineiston sanaston monimuotoisuudesta pyydettiin holistinen arvio 20 yliopistopiskelijalta, joilla kaikilla oli suomen kieli joko opintojen pää- tai sivuaineena. Lisäksi aineistoon sisällytettiin kolme pilottivaiheen vastausta, koska kyselylomaketta tai arvioitavia tekstejä ei ollut tarpeen muokata pilotin jälkeen. Näistä korkeakoulututkinnon suorittaneista informanteista yksi oli suorittanut myös suomen kielen yliopisto-opintoja, mutta kahdella muulla ei ollut lingvististä taustaa. Arvioijat, 4 miestä ja 19 naista, olivat iältään 19–50-vuotiaita (keski-ikä 29 vuotta). Vastaaajien kokonaismäärä aineistossa oli siis lopulta 23, ja kukin arvioija suoritti tehtävänsä loppuun eli arvioi kaikki 60 kertomustekstiä.

Arviot kerättiin sähköisen lomakkeen avulla. Vastaaajia ohjeistettiin lukemaan jokainen kertomusteksti nopeasti kokonaisuudessaan ja arvioimaan kunkin kertomuksen sanaston diversiteettiä välittömästi lukemisen jälkeen. Arviot pyydettiin antamaan asteikolla 0–10 siten, että 0 on matalin ja 10 korkein mahdollinen arvo. Huomiota pyydettiin kiinnittämään ainoastaan tekstin *sanaston diversiteettiin*, ei esimerkiksi kirjoittajan kielitaidon tasoon tai kirjoittamisen yleiseen laadukkuuteen. Diversiteetin käsitettä havainnollistettiin sekä läheisten käsitteiden (*sanaston diversiteetti/moninaisuus/variaatio*) avulla että kuvailemalla hyvin vähäisen ja hyvin runsaan leksikaalisen diversiteetin tekstejä (*sanasto erittäin toisteista – sanasto moninaista*).²⁰ Lisäksi arviointia tukemaan annettiin valmiiksi arvioitu esimerkkikertomus, jonka tarkoitus oli paitsi havainnollistaa arviointitehtävää myös auttaa arviointiasteikon kohdistamisessa. Esimerkkikertomus (diversiteettiarvo = 5) valittiin siten, että se oli käytetyn verrokkimittarin (*sums of probabilities*) mukaan kaikkein lähimpänä laajemman tekstikorpuksen (N = 224) aineiston sanastollisen diversiteetin ääriarvojen (25,7 ja 37,7) keskiarvoa. *Sums of probabilities* on matemaattinen tunnusluku, joka perustuu tekstin kunkin yksittäisen sanan esiintymistodennäköisyyteen etukäteen asetetun parametrin mukaisissa tekstikatkelmissa (McCarthy & Jarvis 2007). Muut tekstit pyydettiin arvioimaan suhteessa esimerkkikertomuksen sanaston diversiteettiin.

- (1) Esimerkkikertomus: sanaston diversiteetti = 5 (asteikolla 0–10)

Haluaisin, että isona voitaisin lotossa, että saisin ison talon ja hienon auton ja että saisin oman firman. Kun ajaisin hienolla autolla, niin kaikki tytöt juoksisivat perässä. Minun isä omistaa jo yhden firman. Olisinpa samanlainen kuin isäni isona. Haluaisin myös hyväksi jalkapallonpelaajaksi tai hyväksi urheilijaksi, jotta saisin hienot lihakset, jotta kaikki katsoisivat minua. Haluai-

20. Ohjeiden sanavalintoja mietittiin huolellisesti. Esimerkiksi sanaa *rikas* ei käytetty monimerkityksisyyden ja osin harhaanjohtavien konnotaatioiden vuoksi: toisinaan *rikas* (*rich*) viittaa lingvistisessä kirjallisuudessa hyvin laajasti erilaisiin sanastollisiin ominaisuuksiin, joihin voi sisältyä esimerkiksi tarkkuus, harvinaisuus ja sanaluokka (ks. esim. Read 1997: 314; 2000: 200–201), toisinaan etenkin varhaisemmassa tutkimuksessa taas puhtaasti sanojen lukumäärään tai sanatoisteisuuteen. *Vaihteleva* voisi olla suomenkielisenä terminä ohjeissa hyvin toimiva, mutta se saattaisi myös ohjata lukijan kiinnittämään huomiota paikalliseen vaihteluun enemmän kuin koko tekstin sanastoon.

sin actionnäyttelijäksi. Jos minulle jäisi vielä rahaa, niin antaisin ne hyvän-
tekeväisyyteen, jotta ihmiset eivät kuolisi nälkään tai köyhyyteen. Ja haluaisin,
että koulua ei olisi aina tai että koulut alkaisivat aina klo 10. Haluaisin
päästä lukion ja yliopiston läpi, jotta minun vanhempani olisivat onnellisia.
Haluaisin olla hyvä koulussa, että saisin diplomin.

Tekstit avattiin arvioitavaksi vasta, kun arvioijat olivat tutustuneet tehtävänantoon ja vahvistaneet ymmärtävänsä sen. Arvioijille kerrottiin, että vaikka arviointitehtävässä tulee olla johdonmukainen, arvioita ei pidä miettiä liikaa, sillä tutkimuksen kannalta tärkein on lukijan intuitiivinen käsitys sanaston diversiteetistä. Arviointinäkömän alussa lukijoita muistutettiin vielä siitä, että hyvin vähäisen diversiteetin tekstissä sanaston käyttö on erittäin toisteista, runsaan diversiteetin tekstissä puolestaan on käytetty moninaista sanastoa.²¹

4.3 Ohjelmat ja laskentaympäristöt

Tilastolliset analyysit toteutettiin pääosin SPSS-tilasto-ohjelmalla (versio 21.0). Hajontakuvioiden sekä osin normaalisuustesteissä ja korrelaatioanalyysissä käytettiin myös R-ohjelmistoa (R Core Team 2014). Kuvioiden piirtämisessä ja tunnuslukujen laskemisessa käytettiin lisäksi Excel-taulukkolaskentaohjelmistoa (versio 2010). Korrelaatioanalyysillä sekä hajontakuvioiden avulla tutkittiin muuttujien keskinäisiä riippuvuussuhteita. Regressioanalyysin tarkoituksena oli selvittää, kuinka hyvin leksikaalisen diversiteetin teoreettisesti johdetut osatekijät ennustavat lukijoiden tekemiä leksikaalisen diversiteetin arvioita. Lisäksi tarkasteltiin, mikä osatekijöiden joukko vastaisi parhaiten lukijoiden arvioita. Oletuksena oli, että parhaan regressiomallin²² osatekijät kuvastavat tekijöitä, joita lukijat käyttävät arvioidessaan tekstin leksikaalista diversiteettiä.

4.4 Muuttujat ja niiden esikarsinta

Jotta ihmisten arvioijien – lukijoiden tai kuulijoiden – diversiteettikäsitteitä voidaan verrata arvioitujen tekstien kielellisiin piirteisiin, on ensin valikoitava mallissa testattavat piirteet teoreettiseen perustaan nojautuen. Käytimme regressiomallissa muuttujina objektiivisia sanastomittareita nähdäksemme, selittävätkö mittarit ihmisten siitä tekemiä arvioita. Tässä tutkimuksessa muuttujat valittiin Jarvisin teoreettisen mallin muuttujien perusteella (2013b: 22–25, ks. myös 2013a: 101–102). Kuusi huomiotavaa muuttujaa ovat samat kuin alaluvussa 3.1 esitellyt diversiteetin osatekijät, ja ne operationalisoitiin tilastollisia analyysejä varten seuraavalla tavalla:

21. Arviointiohje oli pääpiirteiltään sama kuin Jarvisin (2013b) englanninkielisessä aineistossa käyttämä, mutta suomenkielisessä versiossa lukijalle annettiin diversiteetistä hieman enemmän tietoa.

22. Paras mahdollinen malli on sellainen, jossa on korkein R-neliö (*R-squared*). Osatekijöiden tulee olla tilastollisesti merkitseviä, eikä niiden välinen keskinäinen korrelaatio saa olla liian korkea (*multicollinearity*). Hypoteesimme kannalta parhaassa mallissa kaikki teoreettisesti perustellut osatekijät olisivat mukana, mutta toisesta näkökulmasta (esimerkiksi tehokkuuden kannalta) paras malli saavutetaan pienimmällä osatekijöiden määrällä niin sanotun Ockhamin partaveitsi -periaatteen mukaisesti.

Tarkasteltavien yksiköiden

- 1) **määrä** (*volume*): saneiden määrä eli tekstipituus (rajattu vähintään 100 sanan pituisiin teksteihin), mahdolliset arvot 100–X.
- 2) **runsaus** (*abundance*): eri sanojen määrä, mahdolliset arvot 1–X.
- 3) **erityisyys** (*disparity*): sanojen merkitys, funktio ja yksilöllisyys, mahdolliset arvot 0–X (riippuen siitä, kuinka monta sanaa tekstissä on). Jos tekstissä olisi 100 sanaa ja jokainen sana olisi erityinen, tekstin erityisyysarvo olisi 100. Erityisten sanojen (kuten *todella*, *kunnes*, *kohti*) tarkemmat valintakriteerit ovat muotoutuneet aiemman tutkimuksen pohjalta seuraavasti:
Sana on erityinen, jos se
 1. merkityksensä ja käyttöfunktionsa puolesta joko
 - sijoittaa asioita tai tekoja aikaan tai paikkaan
 - selventää asioiden, tapahtumien tai tekstin osien välisiä suhteita TAI
 - määrittää (tai määrällistää) tekemistä ja laatua (johtamattomat adverbit, funktiosanat)
 2. esiintyy aineistossa vähintään kolmella eri kirjoittajalla
 3. esiintyy korkeintaan kolmanneksessa (33 %) korkeimman diversiteetti-luokan teksteistä TAI korreloi diversiteettiarvioiden kanssa vähintään välttämättä ($r \geq 0,15$). Erityiset sanat ovat usein, joskaan eivät aina, kielessä suhteellisen harvinaisia (ks. suomen sanojen yleisyydestä esim. Kielipankki 2004).
- 4) **jakauman tasaisuus** (*evenness*): sanojen frekvenssijakauma, mahdolliset arvot 0,08–1,00.
- 5) **vaihtelevuus** (*variability*): MTLD, mahdolliset arvot 10–X.
MTLD on sanojen monipuolisuusluku, joka on osoittautunut luotettavaksi ja otoskoosta riippumattomaksi keinoksi mitata sanaston monimuotoisuutta. Mitä suurempi MTLD-indeksin arvo on, sitä tiheämpi tekstin sanasto on. Menetelmä perustuu sanojen peräkkäiseen järjestykseen ja jokaisen yksittäisen saneen omaan TTR-arvoon (McCarthy & Jarvis 2010: 384). MTLD:n laskennassa olennaista on, kuinka monta peräkkäistä sanaa tekstissä keskimäärin kohdataan, ennen kuin lekseemien ja saneiden välinen suhde laskee tietyn rajan alle.²³ (Ks. tarkemmin esim. mt.; suomeksi Malin 2012.) Laskenta poikkeaa sana–sane-suhteen perinteisistä mittaustavoin, sillä pelkkä TTR-arvojen keskiarvon laskeminen ei toisi esille sanojen levittymistä tekstin eri osiin.
- 6) **sironta** (*dispersion*): sanojen sijoittuminen tekstiin (sironta 20,5), mahdolliset arvot 0–99.
Optimaalinen sironta syntyy, kun sanan esiintymät sijoittuvat tekstiin mahdollisimman kauas toisistaan – eli säännöllisin välein sen sijaan, että ne kerääntyisivät yhteen. Sironnassa laskettiin, kuinka monta sanakimppeä on

23. TTR-arvot lasketaan tekstin ensimmäisestä saneesta eteenpäin, kunnes TTR ensimmäisen kerran putoaa alle asetetun raja-arvon (0,72). Laskennasta tallennetaan muistiin, kuinka monta peräkkäistä sanaa tarvitaan, ennen kuin TTR alittaa tämän rajan. Seuraavaksi TTR lasketaan tekstin toisesta saneesta eteenpäin, kunnes TTR jälleen putoaa alle raja-arvon (0,72). Sama toistetaan, kunnes päästään tekstin viimeiseen saneeseen. Tämän jälkeen koko laskenta uusitaan takaperin eli tekstin viimeisestä saneesta kohti ensimmäistä sanetta. Lopulta lasketaan keskiarvo kaikista mittauksista eli tekstijaksojen pituuksista.

aina 20 saneen sisällä. Sanakimpuksi määritellään tilanteet, joissa kaksi samaa lekseemiä edustavaa sanetta esiintyy 20 saneen mittaisen tekstikatkelman sisällä. Numero 5 indeksin nimessä tarkoittaa, että aineiston viittä yleisintä sanaa ei oteta huomioon, koska näiden sanojen (esim. *olla*) taajuudella ei tunnu olevan vaikutusta diversiteettiarvoihin.

Aluksi regressiomallin laskemisessa olivat mukana kaikki kuusi teoreettisesti relevanttia muuttujaa. Laskennan edetessä muuttujia otettiin mallista pois yksitellen ja niitä myös lisättiin takaisin. Näin tehtiin, kunnes päädyttiin parhaaseen malliin, jossa kaikki mukaan otetut muuttujat olivat tilastollisesti merkitseviä eivätkä tuottaneet multikollineaarisuusongelmia (toleranssi > 0,2, VIF [*Variance Inflation Factor*] < 5; ks. esim. Gerbing 2014: 230).

4.5 Korrelaatioanalyysit

Korrelaatioanalyysissa käytettiin kolmea eri analyysia: Spearmanin järjestyskorrelaatioanalyysia (rS) sekä soveltuvin osin Pearsonin korrelaatioanalyysia (rP) ja Cronbachin alfaa. Mahdollisten epälineaaristen yhteyksien löytämiseksi aineistoa tarkasteltiin lisäksi hajontakuvioiden avulla muuttujapareittain. Spearmanin ja Pearsonin korrelaatioanalyysia käytettiin kahden muuttujan saamien arvojen keskinäisen riippuvuuden tutkimiseen. Spearmanin järjestyskorrelaatioanalyysia käytettiin silloin, kun parametristen testien edellytykset eivät täyttyneet; sitä sovellettiin verrattaessa tekstien diversiteettiarvoja muuttujiin, joiden arvot eivät tarkastellussa aineistossa noudattaneet normaalijakaumaa. Molemmat analyysit voivat tuottaa arvoja väliltä 0–1 siten, että suuri arvo viittaa voimakkaaseen riippuvuuteen. Normaalisuuden testaamisessa käytettiin Shapiro–Wilk-testiä ja normalisuutta havainnoitiin lisäksi histogrammien sekä standardoitujen jäännösten sirontakuvion avulla. Lukijoiden diversiteettiarviot (koko aineisto, $W = 0,984$; $p = 0,619$), kerran esiintyvien hapaks legomenon -sanojen määrä ($W = 0,976$; $p = 0,296$) ja tekstitiheys ($W = 0,988$; $p = 0,820$), rakenteeltaan yksinkertaisten sanojen tekstitiheys (monolekseemit, $W = 0,992$; $p = 0,958$) sekä tekstien keskimääräinen lekseemipituus ($W = 0,974$; $p = 0,245$) ovat riittävän normaalisia, jotta parametristen testien käyttäminen on mahdollista. Tekstipituuksien, yksilöllisten sanojen määrän ja tekstitiheyden, harvinaisten sanojen tekstitiheyden ja keskimääräisen sanepituuden jakaumat sen sijaan eivät ole normaaliset ($p < 0,05$). Cronbachin alfaa (Cronbach 1951, ks. myös Henson 2001) käytettiin tutkittaessa eri lukijoiden arvioiden yhdenmukaisuutta. Se perustuu muuttujien välisiin korrelaatioihin ja lukumäärään. Myös Cronbachin alfa saa arvoja väliltä 0–1 siten, että suuri arvo kertoo mittarin, tässä tapauksessa diversiteettiarvioiden, yhdenmukaisuudesta.

4.6 Regressioanalyysi

Tutkimuksessa hyödynnettiin myös regressioanalyysia. Usean muuttujan lineaarisella sekamallilla määritettiin, kuinka hyvin kuuden edellä mainitun muuttujan yhdistelmä selittää lukijoiden diversiteettiarvioita. Regressioanalyysin riippuva eli selitet-

tävä muuttuja oli diversiteettiarvio kustakin käytetystä tekstistä, ja riippumattomia eli selittäviä muuttujia olivat määrä, runsaus, erityisyys, jakauman tasaisuus, vaihtelevuus ja sironta.

Keskeinen tilastollinen testi regressioanalyysissa on R-neliö (Larson-Hall 2010: 191), jonka arvot vaihtelevat välillä 0–1,0. Arvo 0 tarkoittaa, että malli ei pysty lainkaan ennustamaan riippuvan muuttujan vaihtelua, ja vastaavasti 1,0 tarkoittaa, että malli ennustaa riippuvan muuttujan vaihtelun kokonaan. Käytännössä malli ei juuri koskaan ole kummankaan ääripään mukainen. Arvoja, jotka ovat suurempia kuin 0,40, pidetään mallin kannalta merkitsevinä. Esimerkiksi Crossleyn, Salsburyn, McNamaran ja Jarvisin (2011b) mallissa R-neliö oli pilottiaineistossa 0,62 ja tutkimusaineistossa 0,60, Jarvisin (2013b) mallissa puolestaan 0,49 (inhimilliset arviot leksikaalisesta diversiteetistä). Jos R-neliö on yli 0,50, malli ennustaa yli puolet riippuvan muuttujan vaihtelusta. Mallia voi kuitenkin parantaa muuntamalla mallin muuttujien mittaustapaa, lisäämällä uusia muuttujia tai tekemällä muutoksia mallin muuttujajoukkoon. Koska mallin R-neliö kasvaa lisääessä muuttujia, SPSS antaa muokatun R-neliön, joka korjaa tämän vinoutuman. Muokattu R-neliö on kerroin, jota käytettiin tämän tutkimuksen regressiomallin arvioinnissa.

Tyypillisesti regressioanalyysissa tarkastellaan mallin muuttujien välistä multikollinearisuutta sen vuoksi, että suuri multikollinearisuus estää kunkin yksittäisen muuttujan vaikutuksen arvioinnin. Jos muuttujat korreloivat voimakkaasti, voi olla vaikeaa määritellä yksittäisen muuttujan vaikutusta malliin ja valita mallin parasta muuttujajoukkoa. Näistä syistä tämän tutkimuksen analyysissa olivat mukana kaikkien riippumattomien muuttujien parien korrelaatiot (ks. verkkoliitettä ja lukuja 5–6). Huomattakoon, että keskinäinen korrelaatio ei kuitenkaan vaikuta kokonaisuutena mallin sopivuuteen tai mallin kykyyn ennustaa riippumattoman muuttujan vaihtelua (Baguley 2012: 450).

5 Tulokset

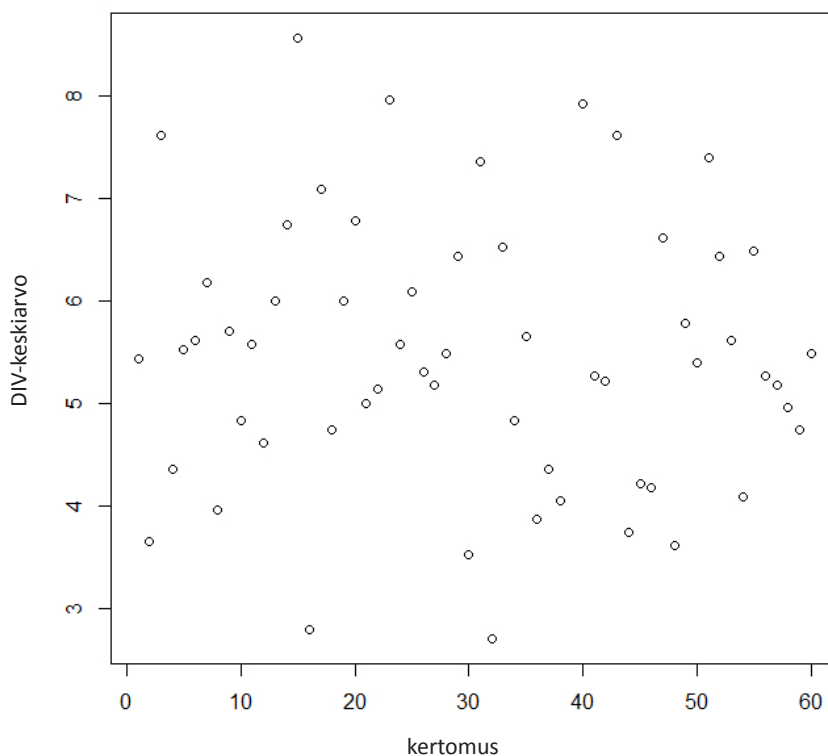
Tässä luvussa esittelemme tilastollisten analyysien tulokset. Aluksi tarkastelemme diversiteettiarvioiden yleistasoja, sen jälkeen arvioiden keskinäistä yhdenmukaisuutta ja lopuksi arvioita selittäviä sanastollisia tekijöitä.

5.1 Leksikaalinen diversiteetti: arvioiden yleistaso ja esimerkit

Diversiteettiarvioiden yleistaso asettuu melko lähelle arviointiasteikon keskilinjaa: eniten on annettu keskimääräisiä arvoja (5 ja 6). Arvioijat kuitenkin käyttivät arviointiasteikkoa koko laajuudessaan, eli yksittäiset arvioijat antoivat yksittäisille kertomuksille myös hyvin matalia (1) ja hyvin korkeita (10) arvoja. Verkkoliitteen taulukoon 1 on koottu aineiston keskeiset tunnusluvut. Lisäksi liitteen kuviossa 1 esitetään kertomuskohtaisten diversiteettikeskiarvojen jakauma histogrammin avulla.

Olennaista on, että arvioidun tekstin järjestysnumerolla ei hajontakuvion (kuvio 3) tai järjestyskorrelaatioanalyysin perusteella ole yhteyttä arviointilinjaan ($r_S = -0,08$;

$p = 0,569$) eli yksittäisen arvioijan linja ei ole systemaattisesti muuttunut tehtävän edessä esimerkiksi harjaantumis- tai väsymiseffektin seurauksena. Tämä on merkityksellinen ja metodin kannalta lupaava havainto erityisesti siksi, että arvioitujen tekstien määrä on suhteellisen suuri (vrt. Crossley, Salsbury & McNamara 2012; Jarvis 2013b).



Kuvio 3.

Diversiteetti-arvion yhteys arvioitavan kertomuksen arviointijärjestykseen. X-akseli kertoo kertomuksen järjestysnumeron arvioinnissa ja y-akseli arvioiden kertomuskohtaisen keskiarvon.

Pienin kertomuskohtainen keskiarvo (2,70) on kertomuksella, joka arvioitiin aineiston puolivälissä (järjestysnumero 32, esim. 2). Suurin keskiarvo (8,57) puolestaan on arviointijärjestyksessä 15. kertomuksella (esim. 3). Arvioiden välinen keskihajonta on sekä vähäisen että runsaan leksikaalisen diversiteetin teksteissä samaa, melko mallista tasoa (kh 1,0–2,0). On syytä muistaa, että näidenkin esimerkkitekstien sanastoa voitaisiin tutkia lukuisilla vaihtoehtoisilla tavoilla. Kokijälähtöisyys eli lukijan käsitysten tarkasteleminen on kuitenkin käyttämämme leksikaalisen diversiteetin määritelmän mukaan perustelluin tapa, sillä vain inhimillisen tulkinnan kautta merkityksellisytyvä kieli on sosiaalisessa mielessä olemassa.

(2) Esimerkkikertomus: sanastoltaan köyhimmäksi arvioitu teksti

Ajattelin huvipuistoa, matkoja ja kesää, mikä oli kivoin. Ensin halusin huvipuistoon, sitten matkoille ja lopuksi halusin, että tulisi kesä. Mitä kesällä voi tehdä, ajattelin. Kesällä voisi mennä rannoille, huvipuistoon, matkalle ja olla kavereiden kanssa ulkona leikkimässä. Se olisi kivaa. Ajattelin, että koska tulisi kesä. Siihen on vielä paljon aikaa. Ja sitten parannuin ja menin kouluun. Aika meni nopeasti koulussa. Sitten tuli kesä ja me menimme rannoille, huvipuistoon ja matkoille ja olin kavereittenkin kanssa ulkona.

(3) Esimerkkikertomus: sanastoltaan monimuotoisimmaksi arvioitu teksti

Raikas tuuli täyttää huoneen. Nousen vuoteesta, pukeudun ja menen parvekkeelle. Alapuolellani aukeaa suuri, kullankeltainen auringonkukkaniitty ja horisontissa siintää meri. Laskeudun alas kartanon portaita ja menen ulos. Laskeudun rantaan pitkin polkua, johon on upotettu kivilaattoja. Polku kulkee kukkuloiden, joilla kasvaa omena- ja kirsikkapuita, välistä ja auringonkukkaniittyjen lävitse. Olen rannassa, joka puolella kirkkaan vihreää ruohoa ja suuria koivuja. Kiipeän yhteen niistä. Tuuli keinuttaa oksaa, jolla istun. Tunnen oloni mukavaksi ja sitten nukahdan. Herään tunnin kuluttua joutsenten lauluun.

Sanastoltaan monimuotoisimmaksi arvioidun tekstin (esim. 3) informaatiotiheys on koko aineiston suurin, mutta sen pituus on vain keskimääräistä tasoa. Koko arvioidussa aineistossa tekstipituuden yhteys leksikaaliseen diversiteettiin on melko heikko ($rS = 0,48$), vaikka sen yhteys samojen tekstien holistiseen, tekstilajipiirteitä painottavaan arvioon on voimakas (Honko 2013: 296–298). Havainnot vahvistavat ajatusta siitä, että lukijat todella pystyvät pyydetessä keskittymään tekstien sanastoon esimerkiksi tuottamisen runsauden tai tekstin kokonaisrakenteen sijaan. Mikäli arviointitehtävän suorittaminen ei onnistuisi, myöskään sanastollisiin piirteisiin kiinnittyvän, vain muutamasta piirteestä koostuvan mallin ei pitäisi selittää arvioiden vaihtelua.

Sen sijaan kerran esiintyvien sanojen määrä ($rP = 0,84$; $p < 0,001$) ja tekstitiheys ($rP = 0,66$; $p < 0,001$), yksilöllisten sanojen määrä ($rS = 0,71$; $p < 0,001$) ja tekstitiheys ($rS = 0,54$; $p < 0,001$) sekä harvinaisten sanojen suhteellinen osuus eri yleisyystasoja käytettäessä ovat keskimääräistä suuremmat paitsi esimerkkiteksteissä 3 myös yleisesti koko aineiston runsaan leksikaalisen diversiteetin teksteissä (fb1000 eli tuhat yleisintä sanaa, fb2000 ja fb9996, $rS = 0,42$ – $0,61$; $p < 0,001$). Tämä tarkoittaa, että sanastoltaan monimuotoiseksi arvioiduissa teksteissä esiintyy muita tekstejä tiheämmin sanoja, jotka eivät toistu siinä tai muiden kirjoittajien teksteissä ja jotka eivät kuulu suomen kielen yleisimpien sanojen joukkoon (vertailukorpuksena Suomen sanomalehtikielen taajuussanasto, CSC 2004). Myös sanapituuksissa on eroa: saneet ja lekseemit ovat muihin teksteihin nähden keskimäärin pidempiä ($rS = 0,51$ ja $rP = 0,57$; $p < 0,001$). Lisäksi morfologiselta rakenteeltaan kompleksisten sananmuotojen ($rP = 0,43$; $p < 0,001$) tekstitiheys on sanastoltaan monimuotoisissa teksteissä keskimääräistä suurempi. On

kuitenkin hyvä muistaa, että mainittujen piirteiden välillä vallitsee suomen kielen syntetisyyden vuoksi monisyinen yhteyksien verkosto. Esimerkiksi harvinaiset sanat ovat usein sekä pitkiä ja morfologiselta rakenteeltaan kompleksisia että merkitykseltään spesifisiä avointen sanaluokkien sanoja. Lukijoiden herkkyyks sanojen yleisyydelle ei myöskään ole yksiselitteinen, sillä vaikka tekstin suuri hyvin yleisten sanojen osuus (fb1000 sekä fb2000) ja toisaalta vertailukorpuksesta kokonaan puuttuvien sanojen suuri osuus heijastuvat diversiteettiarvioiden tasoon, näiden väliin asettuva yleisyystaso (fb3000) ei erottele lukijoiden käsityksiä leksikaalisesta diversiteetistä lainkaan. Lisäksi muuttujien väliset yhteydet eivät välttämättä ole pelkästään lineaarisia silloinkaan, kun yhteyksiä korrelaatioanalyysin avulla ilmenee. Aineistossamme hajontakuvioiden tarkastelu ei kuitenkaan tuo esille epälineaarisia yhteyksiä diversiteettiarvioiden ja muiden muuttujien välillä.

5.2 Arvioiden keskinäinen yhdenmukaisuus

Diversiteettiarvioiden ($N = 23$) yhdenmukaisuutta tutkittiin Cronbachin alfa-kertoimen avulla. Cronbachin alfan suurin mahdollinen arvo on yksi (1,0). Tilastollisessa tutkimuksessa tarkasteltavan aineiston riittävän yhdenmukaisuuden kannalta kohtuullisen suotuisana on totuttu pitämään arvoja 0,7–0,9, erinomaisena ja tutkimuksen reliabiliteetin kannalta tavoittelemisen arvoisena puolestaan arvoja, jotka ovat tuota suurempia (Heikkilä 1998: 187; Kline 1999; DeVellis 2012: 109–110; Jarvis 2013a: 102; ks. kuitenkin myös analyysin rajoituksista esim. Zinbarg, Revelle, Yovel & Li 2005). Omassa aineistossamme alfan arvo 0,959 ylittää selvästi tavoitteeksi asetetun reliabiliteettitason. Tulos kertoo, että eri arvioijien tekemät arviot olivat yleisesti ottaen hyvin samansuuntaisia eli käsitykset arvioitujen tekstien leksikaalisesta diversiteetistä olivat varsin yhdenmukaisia.

Yksittäisten tekstien arvioissa oli kuitenkin myös hajontaa. Siksi kunkin arvioijan vastausten yhdenmukaisuutta yleiseen linjaan, tässä tapauksessa muiden arvioijien kertomuskohtaisten arvioiden keskiarvoihin, arvioitiin myös erikseen käyttämällä Pearsonin korrelaatiokerrointa. Pearsonin korrelaatiokerroin sai arvoja väliltä 0,54–0,85, ja kaikki korrelaatiot olivat tilastollisesti erittäin merkitseviä ($p < 0,001$). Tulos kertoo, että vaikka joidenkin arvioijien linja poikkesi toisista muita voimakkaammin, kaikkien yksittäisten arvioijien arviot korreloivat kuitenkin vähintään tyydyttävällä tasolla yleisen näkemyksen kanssa.

5.3 Tilastollinen malli lukijoiden diversiteettikäsitysten selittäjänä

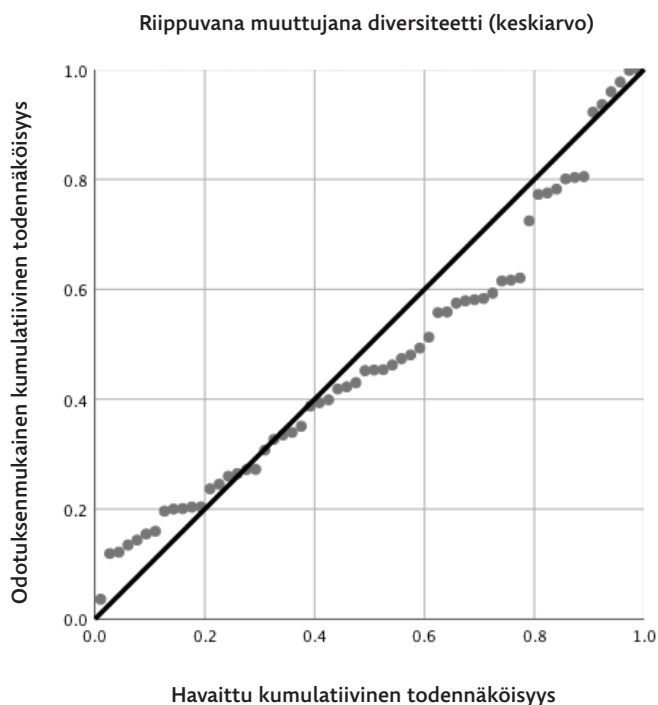
Kuten edellä mainittiin, käytimme SPSS:n usean muuttujan lineaarista regressioanalyysia sen testaamiseen, selittääkö kuuden laskennallisen muuttujan malli leksikaalisen diversiteetin inhimillisiä arvioita. Inhimilliset arviot muodostivat analyysin riippuvan muuttujan, ja kaikki kuusi mallin ennustavaa muuttujaa olivat riippumattomia muuttujia. Mallin selitysvoimaksi saatiin $R^2 = 0,737$ (korjattu R :n neliö), mikä ilmenee taulukosta 1.

Taulukko 1.

Kuuden muuttujan mallin koonti.

R	R ²	Korjattu R ²	Estimaatin keskivirhe
.874	.763	.737	.680

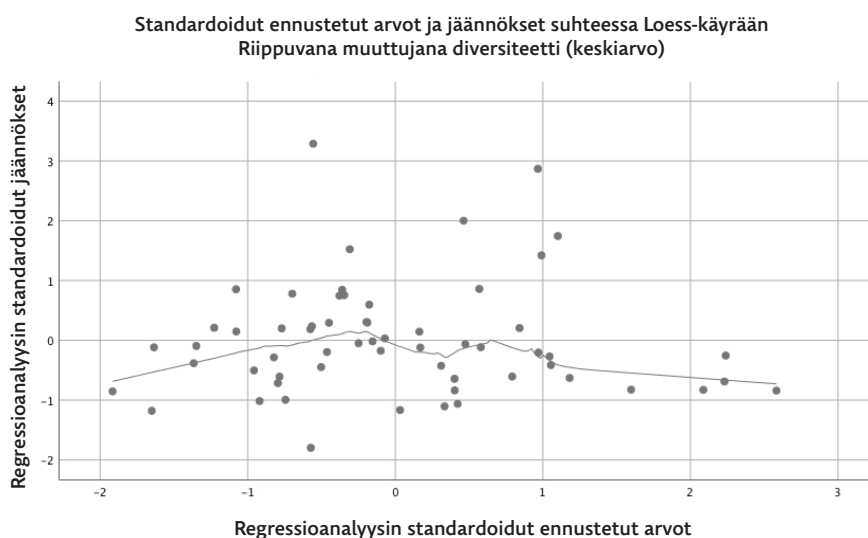
Regressioanalyysin edellyttämä mallin normalisuusoletus tarkistettiin yleisesti käytetyllä menetelmällä, jossa standardoidut jäännökset ja normaalijakauman odotusmukaiset arvot asetetaan yhteen ja esitetään sirontakuvion avulla (ks. kuviota 4). Tavallisesti regressiomalleissa oletetaan, että jäännökset (ei raakadata) vastaavat normaalijakaumaa (Baguley 2012: 325). Normaalijakaumasta poikkeavat arvot näkyvät kuviossa jäännösten käyrämäisenä muotona ja erityisesti siten, että joko toinen tai molemmat kuvaajan arvosarjan ääripäistä (*tails*) käyristyvät pois päin normaalijakauman suorasta linjasta (mts. 325–326). Kuvioista 4 nähdään, että jotkut standardoidut jäännökset poikkeavat hieman normaalijakaumasta, mutta kokonaisuutena trendi on – molemmat ääripäät mukaan lukien – että standardoidut jäännökset vastaavat jotakuinkin odotusmukaisia arvoja. Näin ollen aineisto täyttää normalisuusoletuksen vaatimukset.



Kuvio 4.

Sirontakuvi standardoiduista jäännöksistä suhteessa normaalijakaumaan.

Regressioanalyysin tärkeimpiä perusoletuksia on, että ennustavien eli riippumattomien muuttujien ja riippuvan muuttujan välinen suhde on lineaarinen. Tämä voidaan tarkistaa asettamalla yhteen standardoidut ennustetut arvot ja standardoidut jäännökset (UCLA 2017) ja sovittamalla sirontakuviioon Loess-käyrä (ks. kuviota 5). Lineaarisuusoletuksen kannalta on tärkeää, että jäännökset ovat jakautuneet molemmille puolille arvoa $y = 0$. Kuviosta 5 nähdään, että Loess-käyrä on lähellä nollaa, kun $-1 < x < 1$ eli alueella, missä valtaosa jäännöksistä on. Loess-käyrä kallistuu kohti arvoa $y = -1$, kun $x < -1$ ja $x > 1$. Näillä alueilla havaintopisteitä on kuitenkin niukasti ja Loess-käyrä pysyy hyväksyttävällä tasolla. Kokonaisuutena kuvaaja täyttää lineaarisuusoletuksen.



Kuvio 5.

Loess-käyrä sijoitettuna standardoitujen ennustettujen arvojen ja jäännösten sirontakuviioon.

Taulukoissa 2–3 on esitetty riippumattomia muuttujia koskevat tilastotiedot. Taulukon 2 p-arvot näyttävät osoittavan, että muuttujista vain *runsaus* vaikuttaa merkittävästi malliin, jossa kaikki selittävät muuttujat ovat mukana. Keskeistä on kuitenkin huomata, että muuttujien multikollineaarisuuden eli keskinäisen riippuvuuden mittari VIF (*variance inflation factor*) osoittaa muutamia yli 10:n suuruisia arvoja, mikä viittaa liialliseen multikollineaarisuuteen (UCLA 2017; ks. taulukkoa 3). Tämän perusteella taulukon 2 kertoimet (korrelaatiokertoimia lukuun ottamatta) ovat todennäköisesti epätarkkoja (Baguley 2012: 449–451; ks. myös Baguley 2013), ja on syytä laskea useampia vaihtoehtoisia malleja.

Taulukko 2.
Mallin kertoimet (koko malli).

Muuttujat	Kertoimet		Standardoidut	t	p	Korrelaatiot		
	Standardoimattomat	Keski- virhe				Beta	Kontrollimaton	Osittainen
(Vakio)	9.811	8.878		1.105	.274			
Määrä	-.014	.009	-.464	-1.525	.133	.478	-.205	-.102
Runsaus	.047	.018	.875	2.582	.013	.778	.334	.172
Vaihtelevuus	.004	.005	.093	.705	.484	.703	.096	.047
Sironta	-.041	.023	-.177	-1.768	.083	-.668	-.236	-.118
Erityisyys	.102	.055	.201	1.844	.071	.742	.245	.123
Tasaisuus	-6.817	9.331	-.101	-.731	.468	.295	-.100	-.049

Taulukko 3.
Multikollineaarisuuden tilastotiedot (koko malli).

Muuttujat	Toleranssi	VIF
Määrä	.048	20.705
Runsaus	.039	25.757
Vaihtelevuus	.255	3.928
Sironta	.443	2.257
Erityisyys	.375	2.665
Tasaisuus	.233	4.292

Koska taulukon 3 VIF-arvot viittaavat liialliseen multikollineaarisuuteen muuttujien *määrä* ja *runsaus* välillä, regressioanalyysi tehtiin kerran myös ilman muuttujaa *määrä* ja kerran ilman muuttujaa *runsaus*. Ensin mainitussa analyysissä mallin korjattu R^2 oli 0,730, eikä multikollineaarisuus muodostunut ongelmaksi (VIF enimmillään 3,575). Tässä mallissa p-arvot muuttujille *runsaus*, *sironta* ja *erityisyys* olivat kaikki tilastollisesti merkitseviä ($p < 0,05$). Jälkimmäisessä analyysissä multikollineaarisuus ei ollut ongelma (VIF enimmillään 3,286), mutta mallin selitysaste oli hieman matalampi ($R^2 = 0,709$). Tässä mallissa kaikki jäljellä olleet muuttujat paitsi *tasaisuus* olivat merkitseviä tasolla $p < 0,05$.

Koska *tasaisuus* ei vaikuttanut malleihin tilastollisesti merkitsevällä tasolla, laskettiin vielä kolme uutta mallia: yksi, jossa olivat mukana kaikki muut muuttujat paitsi *tasaisuus*, toinen, josta puuttuivat *tasaisuus* ja *määrä*, sekä kolmas, josta puuttuivat *tasaisuus* ja *runsaus*. Ensimmäisen mallin korjattu R^2 oli korkein (0,739), mutta muuttujat *määrä* ja *runsaus* osoittivat liiallista multikollineaarisuutta (*määrän* VIF = 11,271

ja runsauden VIF = 19,720) samoin kuin alkuperäisessä mallissa. Toisen mallin korjattu R^2 oli lähes yhtä korkea, 0,734. Mallissa ei ollut multikollineaarisuuongelmia (VIF enimmillään 2,617), ja kaikki neljä muuttujaa (*runsaus*, *vaihtelevuus*, *sirona* ja *erityisyys*) vaikuttivat malliin (merkitsevyytaso $p < 0,05$). Kolmannessakaan mallissa ei ollut multikollineaarisuuongelmia (VIF enimmillään 2,260), mutta selitysaste laski hieman ($R^2 = 0,713$).

Kuusi edellä kuvattua regressiomallia on esitetty kootusti taulukossa 4 alenevan selitysvoiman (R^2) mukaisesti.

Taulukko 4.
Kuuden regressiomallin vertailu.

	Malli	Korjattu R^2	Multikollineaarisuus
1	kaikki muuttujat paitsi tasaisuus	.739	korkea
2	koko kuuden muuttujan malli	.737	korkea
3	kaikki muuttujat paitsi määrä ja tasaisuus	.734	matala
4	kaikki muuttujat paitsi määrä	.730	matala
5	kaikki muuttujat paitsi runsaus ja tasaisuus	.713	matala
6	kaikki muuttujat paitsi runsaus	.709	matala

Aiemman kirjallisuuden pohjalta rakennettu kuuden selittävän muuttujan malli (malli 2) kärsii esitettyjen tilastotietojen mukaan multikollineaarisuudesta. Multikollineaarisuus ei Baguleyn (2012: 450; ks. myös UCLA 2017) mukaan vaikuta mallin selitysvoimaan (R^2) mutta on ongelmallista muuttujien mahdollisten päällekkäisvaikutusten kannalta. Siksi luotettavimpina voidaan pitää malleja 3–4: kummassakaan ei ilmene multikollineaarisuutta, mutta selitysvoima on lähes mallin 2 veroinen. Näistä malleista puuttuu osa alkuperäisen mallin muuttujista (mallista 3 *määrä* ja *tasaisuus*, mallista 4 *määrä*). Vaikka tämänkaltaista hypoteettisen mallin karsintaa ei aina pidetä suositeltavana (Baguley 2012: 457), katsomme menettelymme perustelluksi: eri muuttujien suhteita leksikaalisen diversiteetin inhimillisiin arvioihin on vaikea selvittää malleissa, joissa ilmenee multikollineaarisuutta. Kaikkein paras on esitettyjen ehtojen perusteella malli 3, joka koostuu tekstisanaston runsaudesta, vaihtelevuudesta, sironasta ja erityisyydestä.

6 Keskustelua ja pohdintaa

Tässä artikkelissa olemme tarkastelleet leksikaalisen diversiteetin muotoutumista kertomustekstien inhimillisten arvioiden sekä näihin perustuvan tilastollisen mallinnuksen avulla ja pyrkinet löytämään vastauksen kahteen kysymykseen: 1) kuinka yhdenmukaisia inhimillisten arvioijien käsitykset leksikaalisesta diversiteetistä ovat ja

2) mihin sanastollisiin piirteisiin arviointi näyttää nojautuvan? Kokoamme seuraavaksi yhteen tutkimuksen keskeiset tulokset ja pohdimme niitä sekä niihin mahdollisesti vaikuttaneita tekijöitä aikaisemman tutkimuksen valossa. Lopuksi palaamme leksikaalisen diversiteetin määrittelyyn erittelemällä sen suhdetta sanastollisen tiedon subjektiivisiin ja objektiivisiin piirteisiin ja esitämme muutamia ajatuksia mahdollisista jatkosovelluksista.

Tutkimukseen osallistuneiden vapaaehtoisten arvioijien näkemykset arvioitujen tekstien leksikaalisesta diversiteetistä osoittautuivat pitkälti yhteneviksi. Vaikka arvioijat käyttivät annettua kymmenportaista arviointiasteikkoa koko laajuudessaan ja vaikka joidenkin arvioijien näkemykset poikkesivat yleisestä arviointilinjasta enemmän kuin toisten, jokaisen arvioijan näkemys vastasi yleistä linjaa vähintään tyydyttävästi. Suurehkosta arvioitavien tekstien määrästä huolimatta arviointilinja ei muuttunut tehtävän suorittamisen aikana ja kaikki arvioijat suorittivat tehtävänsä loppuun saakka.

On lähes varmaa, että leksikaalisen diversiteetin holistinen arviointi perustuu aina lukuisiin tekijöihin, joista kaikki – kuten arvioijan vireystaso ja kielellinen kokeneisuus – eivät ole suoraan riippuvaisia arvioitavasta tekstistä (ks. myös Jarvis 2017). Siksi voidaan pitää yllättävänä, että suurehkon lukijajoukon näkemykset arvioitujen tekstien leksikaalisesta diversiteetistä ovat lähellä toisiaan ja että vain neljän melko helposti laskehtavissa olevan piirteen eli sanaston runsauden, sironnan, erityisyyden ja vaihtelevuuden avulla voitiin ennustaa oikein 73 % tekstiaineistomme leksikaalisen diversiteetin inhimillisistä arvioista. Näistä piirteistä koostuvan parhaan mallimme ennustavuus on hyvä paitsi absoluuttisesti – malli selittää lähes kolme neljäsosaa lukija-arvioiden vaihtelusta – myös aiempien tutkimusten (Crossley ym. 2011a; Jarvis 2013b) tuloksiin verrattuna. Tilastollisen analyysin perusteella lukija-arvioijan käsitys kertomustekstien leksikaalisesta diversiteetistä rakentuu ennen kaikkea neljän piirteen varaan: tekstin eri sanojen (lekseemien) määrään, saman sanan esiintymien sijoittumiseen tekstissä suhteellisen kauas toisistaan (yhteen kerääntymisen sijaan), sanojen erityisyyteen (merkitys, funktio, yksilöllisyys) sekä uusien sanojen tasaiseen ilmaantumiseen tekstissä. Huomioiduista piirteistä tekstin pituus (sanemäärä) tai sanojen frekvenssijakauman tasaisuus eivät näytä tuovan merkittävää lisää mallin selitysvoimaan, eivätkä ne tässä mielessä vaikuta olevan yhtä kiinteä osa leksikaalista diversiteettiä; tosin ne mahdollisesti kuvaavat osin samoja asioita kuin muut neljä piirrettä.

Osa tutkimustulosten eroista saattaa selittyä tutkimusten metodisilla eroavaisuuksilla. Puhekielen arviointiin perustuva lupaava tutkimus (Crossley ym. 2011b) on motivoinut omaa tutkimustamme, mutta sitä ei voi arviointiin vaikuttavien tekijöiden erojen vuoksi pitää tulostensa puolesta suoraan vertailukelpoisena tutkimuksellemme, jonka aineisto on kirjoitettua kieltä. Arvioijajoukkomme (N = 23) puolestaan on valikoitu ja huomattavasti laajempi kuin kummassakaan aiemmassa kirjoitettuun englannin kieleen kohdistuneessa verrokkitutkimuksessa (Crossley ym. 2011a; Jarvis 2013b). Ero on merkityksellinen, sillä suurempi arvioijajoukko voi tasoittaa ääriarvioiden vaikutusta ja lisätä siten omalta osaltaan tutkimuksen pätevyttä.²⁴ Käyttämämme arviointiasteikko on tar-

24. Jarvisin (2013b) tutkimuksessa tekstit (N = 50) arvioitiin ristiin siten, että kunkin yksittäisen arvioijan (N = 11) arvioima tekstimäärä (N = 10–20) oli selvästi pienempi kuin omassa tutkimuksessamme

kempi kuin Crossleyn ja kumppaneiden tutkimuksissa (10-portainen vs. 5-portainen) mutta sama kuin Jarvisilla.

Tutkimusten välillä on eroja myös arvioitavien tekstien käsittelytavassa sekä tilastolliseen analyysiin valituissa muuttujissa (ks. Crossley ym. 2011a: 567–570; Jarvis 2013b: 31–34). Oman tekstiaineistomme kielellinen tarkkuus yhdenmukaistettiin ennen diversiteettiarvioiden tekemistä, mikä oletettavasti on vähentänyt arvioijien huomion kiinnittymistä esimerkiksi kakkoskielisyydestä tai lukivaikeudesta johtuviin puutteisiin teksteissä. Molemmista aiemmissa tutkimuksissa kaikki tekstit olivat ESL-oppijoiden kirjoittamia, mutta Crossleyn tutkimuksessa tekstien kieliasua ei mahdollisesti²⁵ muokattu ennen arviointia. Tällä ratkaisulla saattaa olla olennainen vaikutus tuloksiin, kun tiedetään, että kielen omaksumisessa eri osa-alueet – kuten juuri leksikaalinen monimuotoisuus ja tarkkuus – kehittyvät samaan aikaan (Bates & Goodman 1997) ja että juuri tarkkuuden puutteet usein kiinnittävät tekstin tulkitsijan huomion muiden piirteiden kustannuksella (Lightbown & Spada 1990). Tällöin diversiteettitulkintoja sellittäväksi tekijäksi saattaa osoittautua kehityksellisten piirteiden korrelointi itsessään. Jarvisin aineistona oli oman tutkimuksemme tapaan joukko koululaisten yhtenevällä tehtävänannolla kirjoittamia tekstejä. Kirjoittajat olivat kuitenkin kaikki englantia vieraana kielenä -oppijoita. On syytä huomata, että tekstin tuottajan kielitaidolla voi olla tuloksiin vaikutusta, vaikka arvioitavien tekstien kieliasu olisikin arviointeja varten yhdenmukaistettu.

Pelkästään oman tutkimuksemme ja edellä esiteltyjen muutamien verrokkitutkimusten varassa ei voida esittää täsmällisiä suosituksia sopivasta tekstien ja arvioijien määrästä ja laadusta tai arviointiasteikosta silloin, kun leksikaalisen diversiteetin arvioinnissa käytetään inhimillisiä arvioijia. Tutkimuksemme kuitenkin osoittaa, että lingvistisesti orientoitunut ja motivoitunut arvioija jaksaa riittävän hyvin keskittyä jopa 60 lyhyen tekstin arvioimiseen yhdellä kertaa annettujen ohjeiden mukaisesti. Käyttämämme 10-portainen asteikko ei vaikuta olevan valikoidulle arvioijajoukolle liian hienosyinen, vaan tutkimuksemme lukijat ovat käyttäneet sitä laaja-alaisesti ja yhdenmukaisella tavalla. Noin 20 hengen arvioijajoukko on aiempiin tutkimuksiin verrattuna huomattavasti suurempi ja tutkimuksemme tarpeisiin ilmeisen riittävä. Kaikki havainnot lisäävät käyttämiemme rajausten uskottavuutta ja samalla perustelevat niiden soveltamista mahdollisissa jatkotutkimuksissa. Vaikka kunkin tutkimukseen osallistuneen lukijan arviointilinjaa voidaan esitettyjen tilastollisten analyysien perusteella pitää luotettavana, arviot eivät kuitenkaan ole täysin yhdenmukaisia. Siksi arviointimenetelmän (lukijat tekstien diversiteettiä arvioimassa) soveltaminen esimerkiksi kielitaidon arviointiin edellyttäisi lisätutkimusta, kuten sen selvittämistä, miten arvioijien kouluttaminen, modaaliteetti tai tekstityyppi vaikuttavat arviointiin.

Kun arvioijia pyydetään arvioimaan kielennäytteitä tietyn, ennalta määrätyn piirteen kautta, vaarana on, että he perustavat arvionsa johonkin toiseen tekstejä leimaa-

(N = 60) ja jokaisen tekstin arvioi ainoastaan 2–3 arvioijaa, mikä voi heikentää tulosten luotettavuutta. Crossleyn, Salsburyn, McNamaran ja Jarvisin (2011a) tutkimuksessa tekstimäärä oli suurempi mutta arvioijia vain kolme.

25. Tekstien kieliasua kyllä yhdenmukaistettiin ennen niiden automaattista analyysia Coh-Metrix-ohjelmalla, mutta ei ole varmaa, kumpi versio teksteistä arvioitiin.

vaan piirteeseen tai tekstistä syntyvään yleisvaikutelmaan (ks. myös Harjunen 2015). Tässä tutkimuksessa arvioinnin laatuun pyrittiin vaikuttamaan viidellä tavalla: 1) Jo ennen kertomustekstien lähettämistä arvioitavaksi niiden kieliasu yhdenmukaistettiin. 2) Arvioijaksi pyydettiin suomen kielen asiantuntijoita, joilla oletettavasti on sekä kiinnostusta että taitoa suomenkielisten tekstien analysointiin. 3) Arvioitava piirre, leksikaalinen diversiteetti, selitettiin ja havainnollistettiin arviointitehtävän yhteydessä.²⁶ 4) Arvioijille tarjottiin mahdollisuus saada lyhyt palaute arvioinnistaan, mikä saattoi parantaa sitoutuneisuutta tehtävän suorittamiseen. Käytännössä palaute tarkoitti mahdollisuutta saada tietoa omasta arviointilinjasta suhteessa arvioinnin yleislinjaan. 5) Arvioijia muistutettiin arvioinnin kohteesta eli siitä, että tarkoitus ei ole arvioida kirjoittajan kielitaidon tasoa tai tekstin yleistä laadukkuutta. Tällainen ennakoimatonkaan arviointikäyttäytyminen ei silti välttämättä olisi ongelma – onhan tarkoitus avoimesti tarkastella myös metodin toimivuutta.

Koska arvioitu tekstiaineisto koostui usean eri kirjoittajan samalla tehtävänannolla kirjoittamista teksteistä, annetut diversiteetti-arviot kohdistuvat välillisesti myös kirjoittajien kielitaitoon. Tekstiaineiston koostamisessa tämä otettiin huomioon valitsemalla aineistoon tasamäärä tyttöjen ja poikien sekä S1- ja S2-oppimäärää ja eri vuosiluokilla opiskelevien koululaisten tekstejä. On selvää, että hyvin heterogeenisessä kirjoittajaryhmässä tuotettujen tekstien leksikaalisen diversiteetin vaihtelu on suurempaa kuin tasaisessa kirjoittajaryhmässä. Tässä mielessä omaa kirjoittajaryhmäämme ei kuitenkaan voi pitää erityisen heterogeenisenä, sillä kirjoittajien ikäjakauma oli verrattain tasainen eikä mukana ollut alkeistason kielenoppijoita tai erityisopetuksen piiriin kuuluvia oppilaita.

Artikkelin alussa pohdimme leksikaalisen diversiteetin yhteyttä sanaston toisteisuuteen ja toimme esille, kuinka toisteisuus on yhteydessä vähäiseen diversiteettiin. Kuten jo Jarvis (2013a: 101) on todennut, leksikaalista diversiteettiä ei tutkimuksemme perusteella silti ole syytä pitää vastakohtana leksikaaliselle toisteisuudelle, sillä toisteisuus on tekstin objektiivisesti mitattava ominaisuus ja mitä ilmeisimmin vain yksi leksikaalisen diversiteetin osatekijä. Sen sijaan diversiteetti on mielestämme ensisijaisesti kokemusperäisen leksikaalisen redundanssin vastapari, jonka olemus tavoitetaan parhaiten juuri tutkimalla kokemusperäisiä havaintoja (taulukko 5; ks. myös Jarvis 2013b: 19–21).

Taulukko 5.

Leksikaalisten konstruktioiden taksonomia (Jarvis 2013b: 21).

	objektiivinen	subjektiivinen
uutuus	vaihtelevuus (<i>variability</i>)	diversiteetti (<i>diversity</i>)
toistuminen	toisteisuus (<i>repetition</i>)	redundanssi (<i>redundancy</i>)

26. Emme toisaalta halunneet harjoituttaa arvioijia tai antaa heille liian paljon ohjeistusta, koska halusimme nähdä, minkälaisia intuitioita heillä itsellään on leksikaalisesta diversiteetistä.

Aina inhimillisten arvioijien käyttäminen ei ole mahdollista tai taloudellista, ja silloin inhimillistä arviointia jäljittelevä, valikoituihin kielenpiirteisiin perustuva automaattinen arviointi voi tarjota toimivan vaihtoehdon. Oikein koostettuna ja käytettynä tällainen työkalu voi tutkimuksemme perusteella olla luotettava ainakin tietyn tyyppisiä tekstiaineistoja arvioitaessa: käyttämässämme kertomustekstiaineistossa vain neljään kielenpiirteeseen perustuva automaattinen arviointi tavoitti inhimillisen arvioijan kokemuksen kertomustekstien leksikaalisesta diversiteetistä varsin hyvin, ja sovelluksia on mahdollista kehittää edelleen myös muiden kielen ilmiöiden ympärille palvelemaan esimerkiksi oppijankielen tutkimuksen ja kielitaidon arvioinnin tarpeita (ks. myös Crossley, Clevinger & Kim 2017). Jarvis (2013a: 102) muistuttaa kuitenkin, että ennen tilastollista mallintamista jokaista piirrettä on voitava mitata luotettavasti ja vielä niin, että vältetään muuttujien välinen multikollinearisuus eli voimakas keskinäinen korrelaatio.

Aineistomme lopullisessa neljän muuttujan mallissa ei ollut multikollinearisuuden suhteen ongelmia, vaikka niitä oli lähtökohtana olleessa kuuden muuttujan mallissa. Mallin uskottavuutta lisää myös se, että kaikki muuttujat ovat teoreettisesti perusteltuja; multikollinearisuuskkin on kaikkein ongelmallisinta silloin, kun regressioanalyysissä päädytään muuttujajoukkoon pelkästään tilastollisista eikä teoreettisista syistä. Mikäli mallia haluttaisiin jatkossa soveltaa esimerkiksi kielitaidon taso- tai itsearviointiin, muuttujien arvojen laskemisen automaattisuutta täytyisi kuitenkin lisätä. Käyttämässämme mallissa tämä tarkoittaa ennen kaikkea tekstiaineiston automaattista lemmausta ja sen tarkkuuden varmistamista sekä erityisten sanojen poiminta-periaatteen kehittäminen tai uudelleenmäärittelyä.

Suomessa sanastontutkimukseen liittyy vielä paljon tutkimattomia alueita. Jatkossa mallin testaamista olisikin hyvä laajentaa eri tekstilajien teksteihin ja testata sitä kirjoituksen lisäksi puhuttuun kieleen. Mallista voidaan myös pyrkiä tekemään entistä vahvempi kokeilemalla esimerkiksi parempaa tasaisuusmittaria, uutta erityisyysmittaria, laajempaa aineistoa tai lisää muuttujia. Oppijankielen tutkimuksen yhteydessä olisi kiinnostavaa tutkia myös sitä, toimiiko malli kielitaidon kehityksen arviointityökaluna – ja jos toimii, millä ehdoilla. Leksikaalisen diversiteetin lukija-arviot näyttävät olevan löyhästi yhteydessä arvioidun tekstin sanarakenteen kompleksisuuteen. Tulos ei ole yllättävä, sillä vähänkin pidempien tekstien leksikaalista diversiteettiä kasvattamaan tarvitaan yleensä myös kielen harvinaisia sanoja, jotka usein ovat sananmuodostuskeinoilla muodostettuja ja siksi monimorfeemisia. Jatkossa olisi kuitenkin mahdollista perehtyä myös siihen, vaikuttaako tekstisanojen morfologinen rakenne, esimerkiksi kieliopillisten morfeemien runsas tai monipuolinen käyttö, diversiteettiarvioihin. Arvioijajoukon valikoituneisuus ja koko, aineiston laatu ja määrä sekä tehtävän ohjeistukseen ja palautteenantoprosessiin liittyvät tekijät voivat olennaisesti vaikuttaa arvioinnin tarkkuuteen ja välillisesti myös arvioiden perusteella rakennettuun leksikaalisen diversiteetin malliin. Tämä tulee muistaa menetelmän mahdollisissa jatkosovelluksissa.

Lähteet

- ARFFMAN, INGA – NISSINEN, KARI 2015: Lukutaidon kehitys Pisa-tutkimuksissa. – Jouni Välijärvi & Pekka Kupari (toim.), *Millä eväillä uuteen nousuun? PISA 2012 tutkimustuloksia* s. 28–49. <http://urn.fi/URN:ISBN:978-952-263-334-7>.
- BAGULEY, THOMAS 2012: *Serious stats. A guide to advanced statistics for the behavioral sciences*. New York: Palgrave Macmillan.
- 2013: *Multicollinearity and collinearity (in multiple regression) – a tutorial*. <http://psychologicalstatistics.blogspot.com/2013/11/multicollinearity-and-collinearity-in.html> (18.1.2018).
- BAKER, PAUL 2010: *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- BASTIAANSE, ROELIEN – JONKERS, ROEL 1998: Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia. – *Aphasiology* 12 s. 951–969.
- BATES, ELIZABETH – GOODMAN, JUDITH C. 1997: On the inseparability of grammar and lexicon. Evidence from acquisition, aphasia and real-time processing. – *Language and Cognitive Process* 12 s. 507–584.
- BERMAN, RUTH A. 2007: Developing language knowledge and language use across adolescence. – Erika Hoff & Marilyn Shatz (toim.), *Handbook of language development* s. 346–367. London: Blackwell Publishing.
- BERMAN, RUTH A. – NAYDITZ, RONIT – RAVID, DORIT 2011: Linguistic diagnostics of written texts in two school-age populations. – *Written Language and Literacy* 14 s. 161–187.
- BLACK, ESTHER – PEPPÉ, SUE – GIBBON, FIONA 2008: The relationship between socioeconomic status and lexical development. – *Clinical Linguistics & Phonetics* 22 s. 259–265.
- BRADAC, JAMES J. – WISEGARVER, RANDALL 1984: Ascribed status, lexical diversity and accent. Determinants of perceived status solidarity, and control of speech style. – *Journal of Language and Social Psychology* 3 s. 239–255.
- BREMER, KATHARINA – BROEDER, PETER – ROBERTS, CELIA – SIMONOT, MARGARET – VASSEUR, MARIE-THERESE 1993: Ways of achieving understanding. – Clive Perdue (toim.), *Adult language acquisition. Cross-linguistic perspectives. Vol II. Field methods* s. 153–195. Cambridge: Cambridge University Press.
- BURROUGHS, ELIZABETH I. 1991: Lexical diversity in listeners' judgments of children. – *Perception and Motor Skills* 73 s. 19–22.
- CARDINALE, BRADLEY – DUFFY, EMMET – GONZALEZ, ANDREW – HOOPER, DAVID – PERRINGS, CHARLES – VENAIL, PATRICK – NARWANI, ANITA – MACE, GEORDINA – TILMAN, DAVID – WARDLE, DAVID – KINZIG, ANN – DAILY, GRETCHEN – LOREAU, MICHEL – GRACE, JAMES – LARIGAUDERIE, ANNE – SRIVASTAVA, DIANE – NAEEM, SHAHID 2012: Biodiversity loss and its impact on humanity. – *Nature* 486 (7401) s. 59–67.
- CARPENTER, RONALD H. 1990: The statistical profile of language behavior with Machiavellian intent or while experiencing caution and avoiding self-incrimination. – Robert W. Rieber & William A. Stewart (toim.), *The language scientist as expert in the legal setting. Issues in forensic linguistics* s. 5–17. Annals of the New York Academy of Sciences 606. New York: The New York Academy of Sciences.
- CARROLL, JOHN B. 1938: Diversity of vocabulary and the harmonic series law of word-frequency distribution. – *The Psychological Record* 2 s. 379–386.
- CASTAÑEDA-JIMÉNEZ, GABRIELA – JARVIS, SCOTT 2014: Exploring lexical diversity in second language Spanish. – Kimberly L. Geeslin (toim.), *The handbook of Spanish second language acquisition*. Chichester, West Sussex: John Wiley & Sons.

- CHAO, ANNE – JOST, LOU 2012: Diversity measures. – Alan Hastings and Louis Gross (toim.), *Encyclopedia of theoretical ecology* s. 203–207. Berkeley: University of California Press.
- COENE, MARTINE – SCHAUWERS, KAREN – GILLIS, STEVEN – ROORYCK, JOHAN – GOVAERTS, PAUL J. 2010: Genetic predisposition and sensory experience in language development. Evidence from cochlear-implanted children. – *Language and Cognitive Processes* s. 1–19.
- CREPALDI, DAVIDE – INGINOLI, CIARA – VERGA, RUGGERO – CONTARDI, ANTONELLA – SEMENZA, CARLO – LUZZATTI, CLAUDIO 2011: On nouns, verbs, lexemes, and lemmas. Evidence from the spontaneous speech of seven aphasic patients. – *Aphasiology* 25 s. 71–92.
- CRONBACH, LEE J. 1951: Coefficient alpha and the internal structure of tests. *Psychometrika* 16 s. 297–334.
- CROSSLEY, SCOTT, A. – CLEVINGER, AMANDA – KIM, YOUJIN 2017: The role of lexical properties and cohesive devices in text integration and their effect on human ratings of speaking proficiency. – *Language Assessment Quarterly* 11 s. 250–270.
- CROSSLEY, SCOTT, A. – MCNAMARA, DANIELLE S. 2011: Understanding expert ratings of essay quality. Coh-Matrix analyses of first and second language writing. – *International Journal of Continuing Engineering Education and Life-Long Learning* 21 s. 170–191.
- CROSSLEY, SCOTT, A. – SALSURY, TOM – MCNAMARA, DANIELLE, S. 2012: Predicting the proficiency level of language learners using lexical indices. – *Language Testing* 29 s. 243–263.
- CROSSLEY, SCOTT A. – SALSURY, TOM – MCNAMARA, DANIELLE S. – JARVIS, SCOTT 2011a: Predicting lexical proficiency in language learner texts using computational indices. – *Language Testing* 28 s. 561–580.
- 2011b: What is lexical proficiency? Some answers from computational models of speech data. – *TESOL Quarterly* 45 s. 182–193.
- CSC – Tieteen tietotekniikan keskus 2004: Suomen sanomalehtikielen taajuussanasto [tekstikorpus]. Kielipankki. <http://urn.fi/urn:nbn:fi:lb-201405272>.
- DEVELLIS, ROBERT F. 2012: *Scale development. Theory and applications*. Los Angeles: Sage.
- DEWAELE, JEAN-MARC – PAVLENKO, ANETA 2003: Productivity and lexical diversity in native and non-native speech. A study of cross-cultural effects. – Vivian J. Cook (toim.), *L2 effects on the L1* s. 120–141. Clevedon: Multilingual Matters.
- DULANEY, EARL F. JR. 1982: Changes in language behavior as a function of veracity. – *Human Communication Research* 9 s. 75–82.
- DURÁN, PILAR – MALVERN, DAVID – RICHARDS, BRIAN – CHIPERE, NGONI 2004: Developmental trends in lexical diversity. – *Applied Linguistics* 25 s. 220–242.
- GERBING, DAVID V. 2014: *R Data analysis without programming*. New York, NY: Routledge.
- GROBE, CARY 1981: Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. – *Research in the Teaching of English* 15 s. 75–85.
- GUI, LIN 2010: A contrastive study of lexical proficiency between L1 and L2 compositions via computerized assessment. – *Foreign Language Teaching and Research* 42 s. 445–450.
- GUO, LIANG – CROSSLEY, SCOTT A. – MCNAMARA, DANIELLE S. 2013: Predicting human judgements of essay quality in both integrated and independent second language writing samples. A comparison study. – *Assessing Writing* 18 s. 218–238.
- HALLIDAY, MICHAEL A. K. 2004: *An introduction to functional grammar*. Toimittanut Christian M. I. M. Matthiessen. Kolmas painos. London: Edward Arnold.
- HALLIDAY, MICHAEL A. K. – HASAN, RUQAYA 1976: *Cohesion in English*. English Language Series 9. London: Longman.

- HARJUNEN, ELINA 2015: Inhimillinen lukija – tekstin arvioinnin ongelma? – Elina Harjunen (toim.), *Tekstit puntarissa. Ajatuksia äidinkielen ja kirjallisuuden oppimistuloksista perusopetuksen päättöarvioinnissa 2014 ja 2010* s. 97–114. Helsinki: Kansallinen koulutuksen arviointikeskus. <https://karvi.fi/publication/tekstitpuntarissa/> (1.4.2018).
- HEIKKILÄ, TARJA 1998: *Tilastollinen tutkimus*. Helsinki: Edita.
- HENSON, ROBIN K. 2001: Understanding internal consistency reliability estimates. A conceptual primer on coefficient alpha. – *Measurement and Evaluation in Counseling and Development* 34 s. 177–189.
- HIGGINS, DERRICK – XI, XIAOMING – ZECHNER, KLAUS – WILLIAMSON, DAVID 2011: A three-stage approach to the automated scoring of spontaneous spoken responses. – *Computer Speech and Language* 25 s. 282–306.
- HOFF, ERIKA 2003: The specificity of environmental influence. Socioeconomic status affects early vocabulary development via maternal speech. – *Child Development* 74 s. 1368–1378.
- HONKO, MARI 2013: *Alakouluikäisten leksikaalinen tieto ja taito. Toisen sukupolven suomi ja S1-verrokki*. Acta Universitatis Tamperensis 1865. Tampere: Tampere University Press. <http://urn.fi/URN:ISBN:978-951-44-9251-8>.
- HOSMAN, LAWRENCE A. 2002: Language and persuasion. – James P. Dillard & Michael Pfau (toim.), *The persuasion handbook. Developments in theory and practice* s. 371–390. Thousand Oaks, CA: Sage.
- HOUSEN, ALEX – SCHOONJANS, ELS – JANSSENS, SONJA – WELCOMME, AURELIE – SCHOONHEERE, ELLEN – PIERRARD, MICHEL 2011: Conceptualizing and measuring the impact of contextual factors in instructed SLA – the role of language prominence. – *IRAL—International Review of Applied Linguistics in Language Teaching* 49 s. 83–112.
- HUHTA, ARI – HILDÉN, RAILI 2015: Kielitutkinnot ja muu laajamittainen kielitaidon arviointi Suomessa. – Ari Huhta & Raili Hilden (toim.), *Kielitaidon arviointitutkimus 2000-luvun Suomessa*. AFinLA-e. Soveltavan kielitieteen tutkimuksia 9. Jyväskylä: AFinLA. <http://urn.fi/URN:NBN:fi:jyu-201703291797>.
- JAKKU-SIHVONEN, RITVA 2013: *Sukupuolenmukaista vaihtelua koululaisten oppimistuloksissa ja asenteissa*. Koulutuksen seurantaraportit 2013:5. Helsinki: Opetushallitus. https://www.oph.fi/julkaisut/2013/sukupuolenmukaista_vaihtelua_koululaisten_oppimistuloksissa_ja_asenteissa (1.4.2018).
- JANTUNEN, JARMO HARRI 2009: Ei pelkästään mielikuvituksen puutteen vuoksi – Kieliaineistojen systemaattinen käyttö tutkimuksessa. – *Virittäjä* 113 s. 101–113.
- JARVIS, SCOTT 2002: Short texts, best fitting curves and new measures of lexical diversity. – *Language Testing* 19 s. 57–84.
- 2013a: Capturing the diversity in lexical diversity. – *Language Learning* 63 s. 87–106.
- 2013b: Defining and measuring lexical diversity. – Scott Jarvis & Michael Daller (toim.), *Vocabulary knowledge. Human ratings and automated measures* s. 13–44. Amsterdam: John Benjamins Publishing.
- 2017: Grounding lexical diversity in human judgments. – *Language Testing* 34 s. 537–553.
- JOHANSSON, VICTORIA 2008: Lexical diversity and lexical density in speech and writing. A developmental perspective. – *Working Papers* 53 s. 61–79. Lund University, Dept. of Linguistics and Phonetics.
- JOHNSON, WENDELL 1939: *Language and speech hygiene. An application of general semantics*. Ann Arbor, MI: Edwards Brothers.
- 1944: Studies in language behavior. I. A program of research. – *Psychological Monographs* 56 s. 1–15.

- KLINE, PAUL 1999: *The handbook of psychological testing*. London: Routledge.
- KORMOS, JUDIT 2011: Task complexity and linguistic and discourse features of narrative writing performance. – *Journal of Second Language Writing* 20 s. 148–161.
- KORPIJAAKKO-HUUHKA, ANNA-MAIJA 2003: *Kyllä se lintupelotintaulujuttu nyt siinä on käsittelyssä. Afaattisten puhujien kielellisiä valintoja sarjakuvatehtävässä*. Helsinki: Helsingin yliopiston fonetiikan laitoksen julkaisuja 46. <http://urn.fi/URN:ISBN:952-10-1469-5>.
- KUPARI, PEKKA – SULKUNEN, SARI – VETTENRANTA, JOUNI – NISSINEN, KARI 2012: *Enemmän iloa oppimiseen. Neljännen luokan oppilaiden lukutaito sekä matematiikan ja luonnontiedon osaaminen. Kansainväliset PIRLS- ja TIMSS-tutkimukset Suomessa*. Jyväskylä: Koulutuksen tutkimuslaitos. <https://jyx.jyu.fi/bitstream/handle/123456789/40574/1/978-951-39-5011-8.pdf> (1.4.2018).
- LAINÉ-LEINONEN, JAANA 2013: *Koulukorpuksen leksikko. 1.–6.-luokkalaisten aktiivinen sanavarasto*. Suomen kielen pro gradu -tutkielma. Tampereen yliopisto. <http://urn.fi/urn:nbn:fi:uta-1-23229>.
- LARSON-HALL, JENIFER 2010: *A guide to doing statistics in second language research using SPSS*. New York: Routledge.
- LIER, LEO VAN 2004: *The ecology and semiotics of language learning – a sociocultural perspective*. Boston: Kluwer Academic.
- LIGHTBOWN, PATSY M. – SPADA, NINA 1990: Focus-on-form and corrective feedback in communicative language teaching. Effects on second language learning. – *Studies in Second Language Acquisition* 12 s. 429–448.
- MACWHINNEY, BRIAN – FROMM, DAVIDA – HOLLAND, AUDREY – FORBES, MARGARET – WRIGHT, HEATHER 2010: Automated analysis of the Cinderella story. – *Aphasiology* 24 s. 856–868.
- MALIN, ESSI 2012: *Suomi toisena kielenä -oppijoiden sanaston kehittyminen taitotasolta toiselle siirryttäessä*. Suomen kielen pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:jyu-201211223060>.
- MALVERN, DAVID – RICHARDS, BRIAN – CHIPERE, NGONI – DURÁN, PILAR 2004: *Lexical diversity and language development. Quantification and assessment*. Houndmills, Hampshire: Palgrave Macmillan.
- MCCARTHY, PHILIP M. 2005: *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)*. Memphis, TE: University of Memphis.
- MCCARTHY, PHILIP M. – JARVIS, SCOTT 2007: Vocd. A theoretical and empirical evaluation. – *Language Testing* 24 s. 459–488.
- 2010: MTL, vocd-D, and HD-D. A validation study of sophisticated approaches to lexical diversity assessment. – *Behavior Research Methods* 42 s. 381–392.
- MILLER, JON F. 1981: *Assessing language production in children. Experimental procedures*. Baltimore, MD: University Park Press.
- MÜLLER GATHERCOLE, VIRGINIA C. – HOFF, ERIKA 2007: Input and the acquisition of language. Three questions. – Erika Hoff & Marilyn Shatz (toim.), *Handbook of language development* s. 107–127. London: Blackwell.
- OLINGHOUSE, NATALIE G. – WILSON, JOSHUA 2013: The relationship between vocabulary and writing quality in three genres. – *Reading and Writing. An Interdisciplinary Journal* 26 s. 45–65.
- PAJUNEN, ANNELI 2012: Kirjoittamistaitojen kehitys 8–12-vuotiailla. Alakoululaisten unelma-kirjoitelmat. – *Virittäjä* 116 s. 4–32.
- R Core Team 2014: R. A language and environment for statistical computing. R Foundation for

- Statistical Computing. Vienna, Austria. <http://www.R-project.org/> (8.8.2014).
- READ, JOHN 1997: Vocabulary and testing. – Norbert Schmitt & Michael McCarthy (toim.), *Vocabulary – description, acquisition and pedagogy* s. 303–320. Cambridge: Cambridge University Press.
- 2000: *Assessing vocabulary*. Cambridge: Cambridge University Press.
- ROWE, MEREDITH L. 2012: A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. – *Child Development* 83 s. 1762–1774.
- SAARELA, LEENA 1997: *Peruskoululaisten kirjoitelmien kehittyminen sanastotutkimuksen valossa*. Acta Universitatis Ouluensis. B Humaniora 25. Oulu: Oulun yliopisto.
- SADEGHI, KARIM SCOTT – DILMAGHANI, SHOLEH KARVANI 2013: The relationship between lexical diversity and genre in Iranian EFL learners' writings. – *Journal of Language Teaching and Research* 4 s. 328–334.
- SCHMID, MONICA – JARVIS, SCOTT 2014: Lexical access and lexical diversity in first language attrition. – *Bilingualism. Language and Cognition* 17 s. 729–748.
- SCOTT, CHERYL M. – WINDSOR, JENNIFER 2000: General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. – *Journal of Speech, Language and Hearing Research* 43 s. 324–339.
- TOGNINI-BONELLI, ELENA 2001: *Corpus linguistics at work*. Studies in corpus linguistics 6. Amsterdam: Benjamins.
- TOMASELLO, MICHAEL 2003: *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- UCLA 2017. Institute for digital research and education. Introduction to regression with SPSS lesson 2. SPSS regression diagnostics. <https://stats.idre.ucla.edu/sas/modules/sas-learning-moduleintroduction-to-the-features-of-sas/> (18.1.2018).
- UNSWORTH, SHARON 2008: Comparing child L2 development with adult L2 development. How to measure L2 proficiency. – Belma Haznedar & Elena Gavruseva (toim.), *Current trends in child second language acquisition. A generative perspective* s. 301–333. Amsterdam: John Benjamins.
- VERMEER, ANNE 2001: Breadth and depth of vocabulary in relation to acquisition and frequency of input. – *Applied Psycholinguistics* 22 s. 217–234.
- WANG, XUAN 2014: The relationship between lexical diversity and EFL writing proficiency. – *University of Sydney Papers in TESOL* 9 s. 65–88.
- WATKINS, RUTH – KELLY, DONNA – HARBERS, HEIDI – HOLLIS, WENDY 1995: Measuring children's lexical diversity. Differentiating typical and impaired language learners. – *Journal of Speech and Hearing Research* 38 s. 1349–1355.
- WITTE, STEPHEN P. – FAIGLEY, LESTER 1981: Coherence, cohesion, and writing quality. – *College composition and communication* 32 s. 189–204.
- XIAO, HANG 2008: On the applicability of Zipf's law in Chinese word frequency distribution. – *Journal of Chinese Language and Computing* 18 s. 33–46.
- YU, GUOXING 2010: Lexical diversity in writing and speaking task performances. – *Applied Linguistics* 31 s. 236–259.
- YULE, G. UDNY 1944: *The statistical study of literary vocabulary*. Cambridge: Cambridge University Press.
- ZINBARG, RICHARD – REVELLE, WILLIAM – YOVEL, IFTAH – LI WEN 2005: Cronbach's α , Revelle's β , and McDonald's ω H. Their relations with each other and two alternative conceptualizations of reliability. – *Psychometrika* 70 s. 123–133.
- ZIPF, GEORGE KINGSLEY 1935: *The psycho-biology of language*. Boston: Houghton-Mifflin.

SUMMARY

Readers' perceptions of lexical diversity: Examining lexical diversity at the interface between quantitative and qualitative research

This article examines the construct of lexical diversity while focusing on research methodology and the potential for lexical diversity to be used as an index of language proficiency. The study gives attention to questions of inter-rater reliability, the effects of texts' lexical characteristics on raters' lexical diversity ratings, and which set of features best accounts for raters' perceptions of lexical diversity.

The present study focuses on lexical diversity in Finnish, and it does this by comparing the lexical characteristics of texts with how they are perceived. The data consists of narrative texts written by school children, as well as lexical diversity ratings assigned to the same texts by adult raters. Each text ($n = 60$) was rated by 23 raters, whose ratings were tested statistically for inter-rater reliability. A regression model was then used to investigate which lexical features the raters relied on while assessing the texts' levels of lexical diversity.

The results show that the raters' lexical diversity ratings were highly consistent with one another (Cronbach's $\alpha = 0.959$). Despite the large number of texts they were asked to rate, all raters rated each of the 60 texts, and their intra-rater consistency remained high from the beginning to the end of the rating task. The results have important implications for the construct definition of lexical diversity: of all the lexical features examined in the present study, four alone suffice to account for nearly three quarters (roughly 73%) of the variance in the ratings. The results suggest that raters' perceptions of lexical diversity are strongly associated with 1) the number of different words in a text (abundance), 2) the intervals between occurrences of the same word (dispersion), 3) the semantic quality of individual words in the text (specialness) and 4) the overall degree of repetitiveness in the text (variety).

Lukijat sanaston monimuotoisuutta määrittämässä: Leksikaalisen diversiteetin tarkastelua määrällisen ja laadullisen tutkimuksen rajapinnassa

Artikkelissa tarkastellaan leksikaalisen diversiteetin eli tekstin sanastollisen monimuotoisuuden rakentumista. Tavoitteena on esitellä leksikaalisen diversiteetin tutkimuksen metodiikkaa ja osoittaa sen potentiaali kielitaidon arvioinnin välineenä. Tutkimuksessa selvitetään, kuinka yksilöllisiä arvioijien käsitykset tietyn tekstin sanastollisesta monimuotoisuudesta ovat, missä määrin arvioijien käsityksiä voidaan selittää tekstien sanastollisilla piirteillä ja mitkä näistä piirteistä ovat käsitysten selittäjinä tärkeimpiä.

Artikkelissa leksikaalista diversiteettiä tutkitaan tekstejä ja niiden tulkintaa vertailemalla. Aineisto koostuu koululaisten kirjoittamista kertomusteksteistä ja aikuisten lukijoiden teksteille antamista sanastollisen monimuotoisuuden arvioista. Kunkin tekstin (yht. 60) arvioi 23 arvioijaa, joiden vastausten yhdenmukaisuutta tarkasteltiin tilastollisesti. Tämän jälkeen selvitettiin tätä tutkimusta varten rakennetun tilastollisen mallin avulla, millaisiin sanastollisiin piirteisiin inhimilliset arvioijat kiinnittävät huomiota tekstien sanastollista monimuotoisuutta arvioidessaan.

Tulokset osoittavat, että arvioijien käsitykset yksittäisten tekstien sanastollisesta monimuotoisuudesta ovat hyvin yhdenmukaisia (Cronbachin alfa = 0,959). Arviointien luotettavuutta lisää se, että arvioitavien tekstien suuresta määrästä huolimatta kaikki arvioijat suorittivat tehtävän loppuun saakka eikä arviointilinja olennaisesti muuttunut tehtävän aikana. Leksikaalisen diversiteetin määrittelyn kannalta on olennaista, että tekstien yksittäisistä sanastollisista muuttujista neljä riittää selittämään lähes kolme neljäsosaa (n. 73 %) arvioiden vaihtelusta. Tutkitun aineiston perusteella lukijoiden tulkinta tekstin sanastollisesta monimuotoisuudesta tukeutuu vahvasti 1) tekstin eri sanojen määrään (*runsaus*), 2) tietyn sanan esiintymien välisiin etäisyyksiin tekstissä (*sironta*), 3) sanojen laatuun (*erityisyys*) sekä 4) tekstin sanastolliseen tiheyteen (*vaihtelevuus*), joka perustuu uusien sanojen tasaiseen ilmaantumiseen tekstissä.

Kirjoittajien yhteystiedot (addresses):

Mari Honko: etunimi.h.sukunimi@jyu.fi

Scott Jarvis: etunimi.sukunimi@utah.edu

Seppo Vainio: etunimi.sukunimi@utu.fi

Mari Honko on suomen kielen opettaja Jyväskylän yliopiston avoimessa yliopistossa. Hän työskenteli tätä tutkimusta tehdessään suomen kielen yliopistonlehtorina Jyväskylän yliopistossa.

Scott Jarvis on kielitieteen professori Utahin yliopistossa Yhdysvalloissa.

Seppo Vainio toimii tutkijana Turun yliopistossa ja Novia-ammattikorkeakoulussa.