

Kerttu Mattila

**KAHDEN SUOMEN KUVALEHDEN  
KIRJOITTAJAN VERTAILU**  
Kvantitatiivinen tyylintutkimus

Informaatioteknologian ja viestinnän tiedekunta

Kandidaatintutkielma

Heinäkuu 2020

# TIIVISTELMÄ

Kerttu Mattila: Kahden Suomen Kuvalehden kirjoittajan vertailu: kvantitatiivinen tyylintutkimus  
Kandidaatintutkielma  
Tampereen yliopisto  
Suomen kielen tutkinto-ohjelma  
Heinäkuu 2020

---

Tutkimuksen tarkoituksena on vertailla kahden kirjoittajan tyyliä. Tutkimuksessa halutaan selvittää, voiko kirjoittajan tyyliä tunnistaa ja kirjoittajia erottaa toisistaan kappaleiden, virkkeiden ja lauseiden pituuksien perusteella. Tutkimuksen taustana on tyylintutkimus sekä aiemmat tutkimukset suomen tekstivirkkeiden ja -lauseiden pituuksista.

Tutkimuksessa vertaillaan kahta Suomen Kuvalehden Näkökulma-palstan vakituista kirjoittajaa. Molemmilta kirjoittajilta on valittu neljä tekstiä kyseiseltä palstalta. Tutkimuksessa käytetään kvantitatiivista tutkimusmenetelmää. Aineiston teksteistä on laskettu kappale-, virke-, lause- ja sanamäärät, sekä kappaleiden pituudet virkkeinä ja virkkeiden ja lauseiden pituudet sanoina. Kappaleiden, virkkeiden, lauseiden sekä erikseen pää- ja sivulauseiden pituuksien jakaumista on tehty graafiset kaaviot ja kaikista jakaumista on laskettu kolme tunnuslukua (keskiarvo, mediaani ja moodi).

Tutkimuksen tulokset paljastavat, että näiden piirteiden valossa on mahdollista löytää kirjoittajien välille selkeitäkin eroavaisuuksia. Toinen kirjoittajista käyttää keskimäärin pidempiä kappaleita sekä lyhyempiä virkkeitä ja lyhyempiä lauseita kuin toinen kirjoittaja. Kaikkien piirteiden, kuten sivulauseiden, osalta eroa ei tämän aineiston valossa löytynyt. Tulosten vahvistamiseksi ja lisätiedon saamiseksi aihe vaatisi laajempaa ja syvempää jatkotutkimusta. Jatkotutkimuksessa on mahdollista hyödyntää tehokkaasti tietotekniikkaa ja tilastollisia menetelmiä laajemman aineiston analysoimiseksi.

Avainsanat: kvantitatiivinen tyylintutkimus, kappale, virke, lause

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

## Sisällys

1	JOHDANTO .....	3
1.1	Tavoitteet.....	3
1.2	Aineisto .....	4
1.3	Tutkimuskysymykset.....	5
2	TUTKIMUKSEN TAUSTA.....	6
2.1	Tyyliintutkimus .....	6
2.2	Tekstilaji.....	7
2.3	Tilastoa suomen tekstivirkkeiden ja -lauseiden pituuksista .....	7
3	MENETELMÄ JA KÄSITTEET .....	8
3.1	Tutkimusmenetelmä .....	8
3.2	Keskeiset käsitteet .....	9
4	ANALYYSI.....	10
4.1	Virkkeiden määrät kappaleissa.....	10
4.2	Sanojen määrät virkkeissä .....	12
4.3	Sanojen määrät lauseissa .....	13
5	LOPUKSI .....	15
6	LÄHTEET .....	18
7	TAULUKOITA .....	20

# 1 JOHDANTO

## 1.1 Tavoitteet

Tutkimuksessani tarkastelen tekstin rakentumista lauseiksi, virkkeiksi ja kappaleiksi. Aineistoni koostuu yhteensä kahdeksasta tekstistä Suomen Kuvalehden Näkökulma-palstalta. Tarkoitukseni on selvittää, voiko tekstin tai kirjoittajan tyyliä tunnistaa lause-, virke- ja kappalepituuksien avulla. Näkökulmana tutkimuksessani on kahden eri kirjoittajan tyylin vertailu saman tekstilajin teksteissä. Analysoimalla aineistoani pyrin selvittämään, muodostuuko saman kirjoittajan eri tekstien välille yhtäläisyyksiä, jotka samalla eroavat toisen kirjoittajan tekstien piirteistä.

Tutkimuksessani keskityn virkkeistä muodostuviin tekstikappaleisiin, virkkeiden pituuksiin sekä lausepituuteen. Tutkimuksessani on kvantitatiivinen ote. Tutkimuksessa ei ole tarkoitus muodostaa kokonaiskuvaa tekstien ja kirjoittajien tyylistä, vaan syventyä valitsemaani näkökulmaan.

Tutkimuksen taustalla on oma kiinnostukseni tyyliintutkimusta ja forensista lingvistiikkaa kohtaan. Valitsin tämän näkökulman, koska haluan selvittää, minkälainen vaikutus tekstin rakenteellisilla piirteillä on tyylin muodostumiselle. Tyyliintutkimusta on tehty paljon, mutta tutustuessani aiempaan tutkimukseen löysin enemmän kaunokirjallisuudesta kuin asiateksteistä tehtyä tyyliintutkimusta, ja näkökulma on painottunut enemmän esimerkiksi sanatasoon. Tutkimukseni voi tuoda tyyliintutkimukseen pienen lisän.

Tyyliintutkimus tuo tietoa siitä, minkälaisia tyyliä käytetään. Kirjoittajan ja tekstin tyyli voi vaikuttaa muun muassa tekstin tulkintaan, uskottavuuteen ja ymmärrettävyyteen. Tutkimukseni tuo lisätietoa siitä, ovatko rakenteelliset piirteet osana tekstin tyylin muodostumista tutkimukseni aineistossa. Tutkimuksen avulla voin tarkastella, eroavatko kahden eri ammattikirjoittajan yleispiirteittäin samantyyppiset tekstit toisistaan. Jos menetelmä tuottaa tulosta, voisi se tietotekniikkaa hyväksi käyttäen olla tehokas tapa tunnistaa kirjoittajien tyyliä suuristakin aineistoista.

Teoreettisesti ajateltuna on mahdollista erottaa toisistaan kaksi eri kirjoittajaa, vaikka tekstilaji ja konteksti ovatkin samat. Tämä johtuu siitä, että kaikilla kielenkäyttäjillä on oma erilainen tapansa käyttää kieltä. Tähän vaikuttaa eletty elämä ja hankittu kielellinen varanto. Periaatteessa kieltä voi käyttää miten vain kieliyhteisön sallimissa rajoissa, mutta käytännössä ihmiset useimmiten tekevät kirjoittaessaan valintoja, jotka ovat juuri heille tyyppillisiä. (Rentola 2019: 11.)

## 1.2 Aineisto

Keräsin tutkimusaineistoni Suomen Kuvalehden verkkosivujen näköislehdestä tammikuussa 2019. Suomen Kuvalehti on vuodesta 1917 lähtien ilmestynyt kuvallinen viikkolehti. Verkkopalvelu mukaan luettuna sillä on viikoittain yli 350 000 lukijaa. Lehti pyrkii tuottamaan korkealaatuista ja riippumatonta sisältöä, vaikka toisaalta se on historiansa aikana ollut tukemassa hallitsevaa yhteiskuntajärjestystä ja kutakin vallinnutta ajatussuuntaa. (Vares & Siltala 2016: 9–10, 437.)

Valitsin kaksi Näkökulma-palstan säännöllistä kirjoittajaa, joilta molemmilta otin tarkasteltavakseni kunkin kirjoittajan palstalla julkaistut neljä peräkkäistä tekstiä. Keräämäni lehtijutut on julkaistu vuosina 2018–2019. Näkökulma-palstan tekstit muistuttavat tyyli-ilailtaan uutiskomenttia tai asiantuntija-artikkelia.

Kommentti on teksti, joka on kirjoitettu uutisen oheen. Kirjoittajalla on aiheeseen henkilökohtainen ote, ja kirjoituksessa esitetään kannanotto tai useampia uutisen aihepiiriin liittyen. (Makkonen-Craig 2005: 210.) Näkökulma-palstan nimi viittaa kommenttitekstilajiin. Lisäksi aineistoni teksteissä näkyy kirjoittajan läsnäolo. Toisaalta kyseisessä lehdessä ei aina ole ollut selkeästi tiettyä uutista, johon näkökulma olisi kirjoitettu.

Muilta piirteiltään aineistoni tekstit muistuttavat asiantuntija-artikkelia. Asiantuntija-artikkeli pyrkii objektiivisuuteen, mutta toisaalta teksti voi olla persoonallista. Se sisältää lähtökohtaisesti vähemmän mielipiteitä ja enemmän tietoa kuin kolumni. (Suhola, Turunen & Varis 2005: 109–110.) Asiantuntija-artikkelin aihe on useimmiten ajankohtainen ja yhteiskunnallinen, teksti on selkeästi asiapitoinen mutta pohtiva. Tyypillisesti tekstin yhteydestä löytyy tieto kirjoittajan taustasta. (Makkonen-Craig 2005: 210.)

Näkökulma-palsta on jokaisessa Suomen Kuvalehden numerossa. Palstalla on useita vierailevia kirjoittajia, jotka eivät ole Suomen Kuvalehden toimittajia. Kirjoittajat tuovat oman asiantuntemuksensa kautta näkökulman johonkin ajankohtaiseen asiaan. Tutkimukseeni valitut tekstit ovat kahdelta palstan vakituiselta kirjoittajalta: Anna-Lena Laurénilta ja Hanna Ojaselta.

Näen aineiston soveltuvan tähän tutkimukseen, sillä kyseessä olevat tekstit ovat samaa tekstilajia ja niiden oletettu lukijakunta on sama. Tämä mahdollistaa kirjoittajien tyylien eroavaisuuden tutkimisen. Palstalla, jolta olen aineistoni kerännyt, on vakioformaatti, minkä vuoksi tekstien pituudet vaihtelevat vain vähän. Tekstien sanamäärä vaihtelee välillä 514–639 sanaan. Aineistona on leipäteksti, koska tutkimuksen kohteena on varsinaisen tekstin tyyli. Otsikoiden ja ingressien tutkimukseen tarvittaisiin laajempi aineisto.

Olen valinnut aineistokseni tekstit, jotka ovat ajallisesti lähellä toisiaan, jotta voin minimoida ajan vaikutuksen. Kielenkäytön konventiot muuttuvat ajan myötä. Vaikka lehtikieltä voidaan pitää myös hitaasti muuttuvana kielenkäytön alueena (Mäkinen 1992: 214), toisaalta lehtitekstien odotetaan reagoivan herkästi yhteiskunnallisiin tyylinmuutoksiin (Mäkinen 1992: 222). Jotta omat mielipiteeni tai ymmärrykseni jonkin tekstin sisällöstä eivät vaikuttaisi tiedostamatta aineiston valintaan, valitsin aineistooni tekstit ennen kuin olin täysin perehtynyt niihin.

Tutkimuksessani käytän teksteistä lyhenteitä, jotka ovat seuraavat:

L1: Anna-Lena Laurén 4.12.2018: Provokaatiosta toiseen

L2: Anna-Lena Laurén 9.11.2018: Siirtomaaksi ei pidä suostua

L3: Anna-Lena Laurén 12.10.2018: Venäläiset passiostoksilla

L4: Anna-Lena Laurén 7.9.2018: Historiaa kaivetaan uusiksi

O1: Hanna Ojanen 4.1.2019: Lasten vuosi alkaa nyt

O2: Hanna Ojanen 23.11.2018: Kansanturvallisuuden äärellä

O3: Hanna Ojanen 21.9.2018: Tanskalaiset takaovella

O4: Hanna Ojanen 26.10.2018: Uskonyhteisöt YK:n apuna

### 1.3 Tutkimuskysymykset

1. Voiko kahden eri kirjoittajan saman tekstilajin tekstejä erottaa toisistaan lause-, virke- ja kappalepituuksia analysoimalla?
2. Miten kirjoittajien tekstit eroavat toisistaan näiden piirteiden osalta?
3. Minkälaisia tyylilleen ominaisia valintoja lauseiden, virkkeiden ja kappaleiden pituuksissa kirjoittajat tekevät?
4. Onko saman kirjoittajan teksteistä löydettävissä sellaisia yhteneväisiä rakenteellisia piirteitä, joiden perusteella kirjoittajan voi tunnistaa?

Analysoin tutkimuksessani lauseiden, virkkeiden ja kappaleiden pituuksia ja pyrin selvittämään, eroavatko kirjoittajien tekstit toisistaan näiden osalta. Jos kirjoittajien väliltä löytyy eroa, selvitän, millä tavalla kirjoittajien tekstit eroavat toisistaan. Tarkoitukseni on saada selkoa siitä, onko tällaisen kvantitatiivisen tutkimuksen keinoin mahdollista havaita kirjoittajien tekemän tyylillisiä valintoja myös lauseiden, virkkeiden ja kappaleiden pituuksien osalta.

Jos pystyn havaitsemaan kirjoittajille ominaisia tyylillisiä valintoja, ovatko ne niin selkeästi toisistaan eroavia, että tällaisen tutkimuksen avulla olisi mahdollista tunnistaa tekstin kirjoittaja? Onnistuisiko kirjoittajan tunnistaminen esimerkiksi monien kirjoittajien joukosta vertailemalla tekstiä joukkoon tekstejä, joiden kirjoittajat tiedetään? Tai olisiko jopa mahdollista selvinneen kirjoittajaprofiilin kautta hakea suuresta tekstimassasta yhteensopivuuksia?

## 2 TUTKIMUKSEN TAUSTA

### 2.1 Tyylintutkimus

Tyylin perusteella eroteltuja tekstilajeja voidaan kutsua tyylilajeiksi. Tekstin kirjoittaja pyrkii osin tietoisesti noudattamaan eri tyylilajien normeja ja toisaalta noudattaa niitä osittain alitajuisesti. (Saukkonen 1984: 30, 84.) Tekstissä on useimmiten tarkoitus viestiä onnistuneesti ja ilmaista retorisesti tehokkaasti ja vaikuttavasti. Kirjallisen tekstin tyyli syntyy siitä, että kirjoittaja pyrkii kohti tätä. (Saukkonen 2001: 147.)

Jokaisella tekstillä on oma kontekstinsa (Saukkonen: 1984: 157), joka vaikuttaa tyylin muodostumiseen. Konteksti on tekstin ulkoinen todellisuus. Se voidaan käsittää tilanteen ja tekstin välisenä yhteytenä tai tilanteen ja kielenkäytön yhteytenä. (Ojala 1971: 6.) Tyyli konkretisoituu vasta valmiissa tekstituotoksessa (Ojala 1971: 7). Tyyli syntyy sellaisista kirjoittajan valinnoista, jotka tekstin rakenne ja konteksti huomioiden ovat mahdollisia, esimerkiksi sanavalinnoista (Ojala 1971: 9–10). Kielelliset odotukset tekstin kontekstista, tekstityypistä ja -lajista sekä kielen yleiset säännöt ohjaavat kirjoittajaa, jolloin kirjoittaessa ei voi valita sanoja, sanamuotoja ja lauserakenteita täysin vapaasti (Verdonk 2002: 6).

Teksteissä voi esiintyä tyylillistä vaihtelua, koska saman tai lähes saman asiasisällön voi ilmaista monin eri tavoin (Saukkonen 1984: 115). Tekstin tyyli muodostuu kokonaisuudesta. Tarkastellessa tyyliä tästä näkökulmasta voidaan sanoa, ettei yhden tekstin eri osilla ole omia erillisiä tyylejä. (Saukkonen 1984: 115.)

Koska suomi on synteettinen ja suffiksipitoinen kieli, se tarjoaa paljon valintavaihtoehtoja, joilla voidaan vaikuttaa tekstin tyyliin. Syntaktisista piirteistä tärkeimpiä, jotka mahdollistavat tyylivaihtelua, ovat sanajärjestys, lauseen ja lauseenosien liittymät sekä lauserakenteen kompleksisuus. (Ojala 1971: 30.) Tämänkaltaista tarkempaa analyysia olisi myös mahdollista tehdä aineistostani jatkotutkimuksena.

## 2.2 Tekstilaji

Tekstilajin määrittely ei ole yksinkertaista (Shore & Mäntynen 2006: 11). Joukkoa samankaltaisia tekstejä voi kutsua tekstilajiksi vain, jos kieliyhteisö tunnistaa tekstijoukon tekstilajiksi. Tekstin lähettäjän ja vastaanottajan on molempien tunnistettava, että kyseinen teksti kuuluu johonkin tekstilajinormiin. (Saukkonen 2001: 61.) Saman tekstilajin eri teksteillä on keskenään yhteisiä piirteitä ja ne muistuttavat osittain toisiaan, mutta eivät ole keskenään täysin samanlaisia. Toisaalta eri tekstilajit sekoittuvat toisiinsa, sisältävät toistensa piirteitä ja muodostuvat osittain samoista aineksista. (Shore & Mäntynen 2006: 11.)

Tekstilajin ja tyylin yhteys voidaan ymmärtää niin, että samaa tekstilajia edustavat tekstit ovat kyseisen tekstilajin tyyllisiä variantteja (Saukkonen 2001: 190). Toisaalta myös tekstin rakenne voi olla yhtenä tekijänä määrittämässä, mihin tekstilajiin teksti kuuluu tai mihin se ei voi kuulua (Mäntynen 2006: 42).

## 2.3 Tilastoa suomen tekstivirkkeiden ja -lauseiden pituuksista

Vesa Heikkinen, Outi Lehtinen ja Mikko Lounela (2001) analysoivat Euroopan Unionin LE-PAROLE-hankkeeseen kuuluvan korpuksen tekstiaineistoa, joka koostuu pääasiassa 90-luvun sanomalehti-, romaani- ja tietokirjateksteistä. Tekstinaineiston ajallinen ero suhteessa omaan aineistooni ja virhemahdollisuus huomioiden voin varauksella käyttää tutkimusta vertailupohjana omasta aineistostani saamilleni keskipituuksille.

Heikkisen, Lehtisen ja Lounelan (2001) tutkimuksen mukaan suomenkielisen tekstin keskimääräinen lausepituus on 6,8 sanaa ja virkepituus 11,1 sanaa. Erikseen sanomalehtiaineistosta saadut tulokset olivat keskimäärin 7,1 sanaa lauseessa ja 11,2 sanaa virkkeessä. He vertasivat saamiaan tuloksia 60-luvulta kerättyyn yleiskieliseen tekstikorpukseen ja arvioivat, että kirjoitetussa suomen kielessä virkkeet mahdollisesti lyhenevät. Lausepituus tutkimusten mukaan on pysynyt samansuuruisena, keskimäärin kuudesta seitsemään sanaa lauseessa.

Hakulinen, Karlsson ja Vilku (1980) ovat myös tutkineet kvantitatiivisin menetelmin neutraalin asiaproosan syntaksia. He ovat saaneet virkkeen keskiarvoiseksi pituudeksi 1,8–2 lausetta. Vaihtelu johtuu lauseen määrittelyn rajatapauksista, kuten elliptisistä ilmauksista. (Hakulinen ym. 1980: 136.) Virkkeen keskimääräiseksi pituudeksi he ovat saaneet 13,3 sanaa (Hakulinen ym. 1980: 138) ja lauseen pituudeksi keskimäärin 6,6 sanaa (Hakulinen ym. 1980: 137).



## 3 MENETELMÄ JA KÄSITTEET

### 3.1 Tutkimusmenetelmä

Tarkastelen tekstin rakentumista kappaleiksi, virkkeiksi ja lauseiksi. Vertailen kahta eri kirjoittajaa saadakseni selville, rakentavatko nämä tekstinjäsennyksen ja syntaksin piirteet tyyliä. Pyrin selvittämään, muodostuuko saman kirjoittajan eri tekstien välille yhtäläisyyksiä kappaleiden, virkkeiden ja lauseiden pituuksien osalta, ja näkyykö samalla eroa toisen kirjoittajan teksteihin nähden.

Ensin tarkastelen kumpaakin kirjoittajaa erikseen. Vertailen saman kirjoittajan neljää tekstiä toisiinsa, jotta voin sanoa, löytyykö niistä yhtäläisyyksiä. Toisin sanoen, näkyykö kirjoittajan idiolekti kaikissa neljässä tekstissä. Jos molemmille kirjoittajille on määritettävissä ominainen tyyli tämän aineiston perusteella, voin vertailla kirjoittajien tyylien eroavaisuuksia.

Tutkimuksessa käytän kvantitatiivista tutkimusmenetelmää. Aineiston teksteistä olen laskenut sana-, kappale-, virke- ja lausemäärät sekä kappaleiden, virkkeiden ja lauseiden pituudet. Kappaleiden pituudet on laskettu virkkeissä, virkkeiden ja lauseiden pituudet sanoissa. Lauseet on eroteltu pää- ja sivulauseiksi ja omaksi ryhmäkseen verbittömät ilmaukset. Olen havainnollistanut saadut tulokset graafisesti piirtämällä jakaumista kaaviot käyttämällä Microsoft Excel -taulukkolaskentaohjelmaa. Lisäksi olen laskenut jakaumista kolme yleistä tunnuslukua (keskiarvo, mediaani ja moodi). Jakaumien ja saatujen tunnuslukujen avulla olen verrannut saman kirjoittajan tekstejä keskenään ja eri kirjoittajia toisiinsa.

Kun tehdään jonkin piirteen perusteella vertailua kahden tekstin välillä, tuloksena voi olla, että kyseinen piirre esiintyy vain toisessa tekstissä, piirre esiintyy molemmissa, mutta on toisessa yleisempi, tai piirre esiintyy molemmissa yhtä paljon. Myös nollafrekvenssi, eli piirrettä ei esiinny lainkaan, on huomioitava. (Saukkonen 1984: 119.) Tekstit voivat siis osoittautua jonkin piirteen osalta samanlaisiksi sillä perusteella, että kyseistä piirrettä ei esiinny kummassakaan tekstissä lainkaan. Huomionarvoista on myös, että ero voi johtua satunnaisvaihtelusta. Tilastollisilla testeillä voidaan laskea, kuinka todennäköisesti otannoissa havaittava ero heijastaa koko aineistoa. (Esim. Holopainen & Pulkkinen 2015.)

Erojen tilastollista merkitsevyyttä olen tarkastellut laskemalla t-testin Microsoft Excel -ohjelmalla. Analyysin yhteydessä ilmoitettu p-arvo kertoo, millä todennäköisyydellä arvot ovat peräisin samanlaisesta jakaumasta. Jos todennäköisyys on pienempi kuin 0,05, eroa pidetään yleisesti tilastollisesti merkitsevänä (esim. Holopainen & Pulkkinen 2015: 183). Tilastollisesti merkitsevän

eron raja on kuitenkin tapauskohtainen. Pienemmästä aineistosta on usein vaikeampi todistaa tilastollisesti merkitsevää eroa t-testillä. Lisäksi käyttämäni t-testi olettaa likimain normaalijakautuneen aineiston (Holopainen & Pulkkinen 2015: 182). Joissain tapauksissa (esim. joidenkin tekstien virkkeet ja useiden tekstien kohdalla sivulauseet) tutkimukseni aineisto ei noudata normaalijakaumaa. Nämä tekijät saattavat heikentää tilastollisen testin luotettavuutta.

Tutkimuksessa on pyritty selvittämään, eroavatko eri kirjoittajan tekstit toisistaan valittujen muuttujien perusteella. Tuloksena voi kuitenkin olla myös se, ettei kirjoittajien väliltä löydy tällä menetelmällä eroavuuksia, tai ettei saman kirjoittajan tekstien väliltä löydy selkeitä yhteneväisyyksiä.

## 3.2 Keskeiset käsitteet

**Virke, lause** ja **lauseke** muodostavat hierarkkisen järjestelmän, jossa virke muodostuu lauseista ja lause lausekkeista (Löflund 2002: 1, 3). Lauseke koostuu pääsanasta ja mahdollisista pääsanana määritteistä, joita voi olla yksi tai useampi (Löflund 2002: 3). Lause puolestaan on kokonaisen ajatuksen muodostava sanojen yhdistelmä, joka sisältää finiittiverbin ja useimmiten yhden tai useamman nominaalisen jäsenen. Lauseessa sanojen välillä vallitsee neksus- tai määritesuhde. (Löflund 2002: 3–4.) Tutkimuksessani olen laskenut lauseiksi myös verbittömät ilmaukset, mutta olen jättänyt ne pois pää- ja sivulausejaottelusta. Virke on tekstin rakenteen ortografinen käsite (ISK 2004 § 864). Virkkeeksi määritellään usein yhden tai useamman kirjoitetun lauseen muodostama yksikkö, joka alkaa isolla alkukirjaimella ja loppuu päättövälimerkkiin (Löflund 2002: 4). Toisaalta se voi muodostua myös yksittäisestä sanasta, joka ei täytä lauseen määritelmää (ISK 2004 § 864).

**Kappale** on kirjoitetun tekstin rakenteen jäsenyskeino. Kappale voi pituudeltaan olla yhdestä sanasta useisiin kokonaisiin sivuihin, mutta kappalejaolta useimmiten odotetaan loogista tekstin jäsenystä asiakokonaisuuksien mukaan. Eri tekstilajeissa käytetään keksimäärin hyvin eri pituisia kappaleita. (Anhava 1987.)

**Tyyli** muodostuu kielellisistä valinnoista. Tässä tutkimuksessa tyyliä lähestytään dualistisen käsityksen kautta, eli erilaisina tapoina esittää sama asia. (Voutilainen 2012: 77–78.) **Tekstilaji** nähdään tässä joukkona tekstejä, jotka tuottavat ja joilta kielenkäyttäjät edellyttävät tiettyjä kulttuurisidonnaisia kielenkäyttötapoja (Voutilainen 2012: 86–87).

Lasketut **muuttujan arvot** (kuten yhdessä aineistoni tekstissä esiintyvien lauseiden pituudet sanamäärän mukaan laskettuna) muodostavat **jakauman**. Jakaumasta lasketaan **tunnuslukuja**, jotka antavat siitä tietoa kukin eri näkökulmasta. (Holopainen & Pulkkinen 2015: 78.)

Tunnuslukuista tutkimuksessani käytän **aritmeettista keskiarvoa (Ka)**, **mediaania (Md)** ja **moodia (Mo)**. Keskiarvo saadaan laskemalla havaintoarvot yhteen ja jakamalla saatu tulos havaintojen lukumäärällä (Holopainen & Pulkkinen 2015: 83). Arvojen mukaan järjestetyn havaintoaineiston keskimäinen havainto on mediaani, jos havaintoarvoja on pariton määrä. Tapauksessa, jossa arvoja on parillinen määrä, esitetään mediaanina toinen keskimmäisistä arvoista tai molemmat. (Holopainen & Pulkkinen 2015: 80.) Jos arvoja on parillinen määrä, tässä tutkimuksessa käytän mediaanina keskimmäisistä arvoista suurempaa. Keskiarvo ja mediaani antavat tietoa keskimääräisestä arvosta. Jos ne eivät ole samat, havaintoaineisto ei ole symmetrinen. (Holopainen & Pulkkinen 2015: 84–85.) Jos esimerkiksi yhden tekstin lausepituuksien keskiarvo on mediaania huomattavasti suurempi, yksi mahdollisuus on, että tekstissä on yksittäinen tai yksittäisiä pitkiä lauseita. Moodi on havaintoaineiston yleisin muuttujan arvo (Holopainen & Pulkkinen 2015: 79). Se antaa tässä tutkimuksessa tietoa esimerkiksi siitä, minkä mittaiset lauseet ovat kussakin tekstissä yleisimpiä.

## 4 ANALYYSI

### 4.1 Virkkeiden määrät kappaleissa

	L1	L2	L3	L4	O1	O2	O3	O4
<b>Sanoja</b>	639	514	528	578	579	569	584	554
<b>Kappaleita</b>	29	23	23	26	22	27	28	21
<b>Virkkeitä</b>	72	53	53	63	42	52	49	54
<b>Ka</b>	2,5	2,3	2,3	2,4	1,9	1,9	1,7	2,6
<b>Md</b>	2	2	2	2	2	2	2	3
<b>Mo</b>	1	2	1 ja 2	2	1	2	2	3

Taulukko 1. Virkkeiden määrä tekstien kappaleissa

Sanoja: kunkin tekstin sanamäärä

Kappaleita: kappaleiden lukumäärä kussakin tekstissä

Virkkeitä: virkkeiden lukumäärä kussakin tekstissä

Ka: kappaleen keskiarvopituus virkkeinä kussakin tekstissä

Md: kappaleen mediaanipituus virkkeinä

Mo: kuinka monta virkettä on yleisin kappaleen pituus

Tässä työssä käytettyjen aineistojen formaatti on lehtijuttu, jossa sanamäärä on rajattu, joten kirjoittajien eroja ei tässä suhteessa ole mielekäästä tutkia. Tekstit ovatkin keskenään melko samanmittaisia, vaikka Laurénin tekstien kesken onkin suurempaa hajontaa kuin Ojasen tekstien. Laurénin tekstien leipätekstiosuuksien sanamäärä vaihtelee välillä 514–639 ja Ojasen tekstien sanamäärä välillä 554–584.

Molemmilla kirjoittajilla tekstien kappalemäärät vaihtelevat vähän. Laurénin teksteissä on 23–29 kappaletta ja Ojasen teksteissä 21–28 kappaletta. Laurénilla on hieman enemmän kappaleita (keskimäärin 25,25 kappaletta tekstissä) kuin Ojasella (keskimäärin 24,5), mutta kappaleiden määrä teksteissä ei merkittävästi erota kirjoittajia toisistaan.

Sen sijaan kirjoittajien tekstien virkkeiden määrät eroavat toisistaan. Ero ei kuitenkaan ole tilastollisesti merkitsevä ( $p = 0,09$ ). Laurénin teksteissä on 53–72 virkettä, kun Ojasen teksteissä on 42–54 virkettä. Laurénin teksteissä on keskimäärin yli 10 virkettä enemmän kuin Ojasen teksteissä. Koska kirjoittajilla on teksteissään lähes yhtä paljon kappaleita, mutta Laurénin teksteissä virkkeitä on enemmän, on Laurénin tekstien kappaleissa keskimäärin enemmän virkkeitä kuin Ojasen tekstien kappaleissa.

Laurénin kappaleiden virkemäärien jakaumat (esitetty liitteissä, taulukko 4) muistuttavat toisiaan kaikissa neljässä tekstissä. Kappaleiden pituudet painottuvat hyvin lyhyisiin, yhden ja kahden (paitsi tekstissä L4 kahden ja kolmen) virkkeen kappaleisiin, mutta Ojasen teksteihin verrattuna Laurénin teksteissä on tasaisemmin myös useamman virkkeen kappaleita. Kappaleita, joissa on yhdestä kahteen virkettä, on kussakin tekstissä 14–17. Kappaleita, joissa on kolmesta neljään virkettä, on puolestaan 8–10. Viiden virkkeen kappaleita ei ole, tai niitä on korkeintaan kaksi tekstissä.

Ojasen teksteistä kolme (O1, O2, O3) ovat vielä Laurénin tekstejä selkeämmin painottuneet yhden ja kahden virkkeen kappaleisiin. Useamman virkkeen kappaleita on vain vähän. Yhden tai kahden virkkeen kappaleita on Ojasella kolmessa tekstissä tekstiä kohden 17–25, kun taas kolmen tai neljän virkkeen kappaleita on 3–5, viiden virkkeen kappaleita ei lainkaan. Yksi teksti (O4) poikkeaa kuitenkin selvästi muista kolmesta. Poikkeavassa tekstissä jakauma on symmetrisempi ja painottunut kolmen virkkeen kappaleisiin, ja lisäksi tekstissä on myös yksi viiden virkkeen kappale.

Kun otetaan molempien kirjoittajien kaikki tekstit huomioon, kappaleiden virkemäärissä löytyi eroja. Ero kirjoittajien välillä kappaleiden osalta on myös tilastollisesti merkitsevä ( $p = 0,01$ ). Kaikissa teksteissä kappaleiden pituudet ovat kuitenkin enintään 4 tai 5 virkkeen mittaisia, joten tekstien kappaleet pysyttelevät samoissa rajoissa. Kummallakaan kirjoittajalla ei siten ole todella montaa virkettä yhdessä kappaleessa.

## 4.2 Sanojen määrät virkkeissä

	L1	L2	L3	L4	O1	O2	O3	O4
<b>Virkkeitä</b>	72	53	53	63	42	52	49	54
<b>Sanoja</b>	639	514	528	578	579	569	584	554
<b>Ka</b>	8,9	9,7	10,0	9,2	13,8	10,9	11,9	10,3
<b>Md</b>	8	9	10	8	12	11	12	9
<b>Mo</b>	6 ja 9 ja 11	11	5	7	10	5	8	9
<b>Sanoja virkkeissä</b>	1–20	1–23	3–21	3–28	4–34	4–23	4–27	2–28

Taulukko 2. Virkkeiden pituudet

Virkkeitä: virkkeiden lukumäärä kussakin tekstissä

Sanoja: sanojen lukumäärä kussakin tekstissä

Ka: virkkeiden keskiarvopituus sanoina kussakin tekstissä

Md: virkkeiden mediaanipituus

Mo: yleisimmin esiintyvä virkkeen sanamäärä

Sanoja virkkeissä: sanojen määrän vaihteluväli virkkeissä

Laurénin teksteissä on enemmän virkkeitä kuin Ojasen teksteissä ja Laurénin tekstien virkkeet ovat keskimäärin Ojasen virkkeitä lyhyempiä. Keskimittaiset virkkeet ovat Ojasella keskimäärin hiukan pidempiä kuin Laurénilla. Ero on lähellä tilastollisesti merkitsevyyden rajaa sekä virkkeiden keskiarvopituuksissa ( $p = 0,05$ ), että mediaanipituuksissa ( $p = 0,04$ ).

Lyhimpiä ja pisimpiä virkkeitä lukuun ottamatta molempien kirjoittajien teksteissä on yhtä lailla eripituisia virkkeitä. Virkkeiden sanamääräjakaumat on esitetty graafisesti liitteissä (kaaviot 1–8). Kirjoittajia vertailtaessa Ojasella virkkeet ovat tasaisemmin jakautuneet eri mittaisiksi, kun taas Laurénin teksteissä lyhyemmät virkkeet ovat selvemmin painottuneet. Kun tarkastellaan virkkeitä, joissa on 5–11 sanaa, Laurénilla sen pituisia virkkeitä on keskimäärin noin 60 % kaikista tekstien virkkeistä. Ojasella puolestaan tämän pituisia virkkeitä on keskimäärin vain noin 46 % kaikista tekstien virkkeistä. Molemmilla kirjoittajilla on kuitenkin tässä suhteessa yksi hieman poikkeava teksti. Laurénilla yhdessä tekstissä (L4) virkkeet ovat muita tekstejä vielä vahvemmin painottuneet 5–11 sanan mittaisiin virkkeisiin (noin 68 % kaikista tekstin virkkeistä). Ojasella

poikkeava teksti (O4) on tällä mittarilla lähempänä Laurénin tekstejä kuin muut Ojasen tekstit. Siinä lyhyitä 5–11 sanan mittaisia virkkeitä on noin 54 %.

Laurén käyttää teksteissään myös yhden sanan virkkeitä, joita Ojasella ei ole lainkaan. Laurénin voidaankin nähdä käyttävän tyylikeinona yksisanaisia huudahduksia ja kommentointeja. Ojasella kolmessa tekstissä (O1, O2, O3) lyhyin virke on neljän sanan mittainen. Neljänneistä tekstistä (O4) löytyy kuitenkin kolme 3 sanan mittaista virkettä ja yksi 2 sanan pituinen virke. Laurénilla puolestaan 1–3 sanan virkkeitä on jokaisessa tekstissä, kaikissa neljässä teksteissä yhteensä 15 virkettä. Yhden sanan virkkeitä on Laurénilla kahdessa tekstissä (L1, L2) yhteensä kolme virkettä.

Ojasen pisimmät virkkeet ovat puolestaan keskimäärin selvästi pidempiä kuin Laurénilla, ja pitkiä virkkeitä on Ojasella myös enemmän. Ojasen jokaisessa tekstissä on yli 20 sanan virkkeitä. Näitä on neljässä tekstissä yhteensä 16. Laurénin teksteistä kolmessa (L2, L3, L4) on yli 20 sanan virkkeitä, yhteensä neljä. Yli 25 sanan virkkeitä Laurénilla on vain yksi (tekstissä L4), kun taas Ojasella näitä on kolmessa tekstissä (O1, O3, O4) yhteensä viisi.

Kun vertaillaan kirjoittajien tekstejä virkkeiden sanamäärien osalta, löytyy selkeitä eroja kirjoittajien välille. Ero on myös tilastollisesti merkitsevä ( $p = 0,00002$ ).

### 4.3 Sanojen määrät lauseissa

	L1	L2	L3	L4	O1	O2	O3	O4
<b>Lauseita yhteensä</b>	112	92	96	102	78	89	95	88
<b>Päälauseita</b>	75	58	54	71	49	57	59	60
<b>Sivulauseita</b>	32	30	42	29	27	26	31	24
<b>Verbittömiä</b>	5	4	0	2	2	6	5	4
<b>Sanoja päälauseissa</b>	1–15	2–11	1–12	2–15	2–34	1–15	2–16	1–28
<b>Päälauseiden ka- pituus</b>	6,1	5,4	4,9	5,7	7,5	6,7	6,4	6,4
<b>Päälauseiden md</b>	6	5	4	5	6	6	6	6
<b>Päälauseiden mo</b>	5	4	4	5	5	5	6	5
<b>Sanoja sivulauseissa</b>	2–10	2–14	2–16	3–11	2–27	3–19	3–11	3–11
<b>Sivulauseiden ka- pituus</b>	5,3	6,6	6,3	5,8	7,2	6,5	6,1	6,5
<b>Sivulauseiden md</b>	5	5	6	5	7	5	6	6
<b>Sivulauseiden mo</b>	5	5	4	5	7	4	6	9
<b>Sanoja verbittömissä lauseissa</b>	1–4	1–3		1	6–12	1–5	2–6	2–6

Taulukko 3. Lauseiden pituudet

Lauseita tarkastellessa ero kirjoittajien välillä ei ole yhtä selvä kuin virkkeiden kohdalla. Kuitenkin vastaavasti kuten virkkeet, myös Laurénin lauseet ovat keskimäärin lyhyempiä kuin Ojasella.

Laurénin pisin lause on 16 sanan mittainen, kun taas Ojasen teksteissä pisin on 34 sanan pituinen. Laurénilla ei ole yksittäisiäkään pitkiä lauseita, vaikka yksittäisiä pitkiä virkkeitä hänkin käyttää. Ojasella jotkin yksittäiset pitkät virkkeet koostuvat yhdestä lauseesta. Ojasella onkin yli 16 sanan mittaisia lauseita kolmessa tekstissä (O1, O2, O4) yhteensä kuusi. Lauseiden keskiarvopituudet eroavat, mutta ero ei ole suurempi kuin virkkeiden kohdalla. Ero päälauseiden keskiarvopituuksissa on tilastollisesti merkitsevä ( $p = 0,01$ ), kun taas kirjoittajien tekstien sivulauseiden keskiarvopituudet eivät eroa tilastollisesti merkitsevästi ( $p = 0,2$ ).

Lauseiden sanamääräjakaumat on esitetty graafisesti liitteissä (kaaviot 9–16). Kaavioista näkee, että Laurénin tekstit muistuttavat paljon toisiaan. Eniten on neljän tai viiden sanan mittaisia lauseita. Lyhyempiä ja pidempiä lauseita on melko tasaisesti vähenevä määrä, mitä kauemmas yleisimmästä arvosta mennään, kuitenkin niin, että yleisintä arvoa pidempiä lauseita on enemmän kuin lyhyempiä. Jokaisessa tekstissä on myös vähintään kaksi yhden sanan lausetta ja pisin lause on 14–16 sanan mittainen. Ojasen tekstit ovat keskenään erilaisempia. Yksi teksti (O3) muistuttaa Laurénin tekstejä lauseiden pituuksien osalta. Muissa Ojasen teksteissä lauseiden pituudet ovat jakautuneet tasaisemmin erimittaisiksi lauseiksi kuin Laurénin teksteissä. Osassa Ojasen teksteistä (O2, O4) löytyy yhden sanan lauseita, ja tekstien pisimmät lauseet vaihtelevat välillä 16–34.

Keskiarvoiltaan Ojasen tekstien pää- ja sivulauseet ovat suunnilleen yhtä pitkiä. Laurénilla vaihtelua on enemmän: kahdessa teksteistä päälauseet ovat lyhyempiä kuin sivulauseet, yhdessä tekstissä toisin päin ja yhdessä tekstissä ne ovat melko samanmittaisia. Sivulauseiden ja päälauseiden suhteellisissa määrissä kirjoittajat eivät eroa toisistaan. Laurénin teksteissä sivulauseita on keskimäärin noin 33 % kaikista lauseista, kun vastaavasti Ojasen teksteissä sivulauseita on keskimäärin noin 31 %.

Päälauseiden yleisin pituus on Laurénilla 4 tai 5 sanaa, Ojasella 5 tai 6 sanaa. Tämä ei tuo kirjoittajien välille juurikaan eroa. Laurénilla sivulauseissa yleisin lausepituus on sama 4 tai 5 sanaa, mutta Ojasen tekstien sivulauseissa vaihtelua on enemmän. Eniten on 4, 6, 7 tai 9 sanan mittaisia sivulauseita. Voi olla, että Laurénilla on enemmän syntaktisesti toisiaan muistuttavia sivulauseita, ja Ojasella taas enemmän varioivia sivulauseita, mutta sitä ei voida suoraan tämän perusteella sanoa.

Päälauseiden sanamääräjakaumat on esitetty graafisesti liitteissä (kaaviot 17–24). Kirjoittajien päälauseet eroavat toisistaan siten, että Laurénin tekstien päälauseiden sanamääräjakaumalla on selkeämpi huippu ja lauseet ovat vahvemmin kasautuneet lyhyemmiksi päälauseiksi kuin Ojasella. Laurénin pisimmät päälauseet ovat 15 sanan mittaisia. Ojasella on myös yksittäisiä pitkiä lauseita, mutta yli 15 sanan mittaisia päälauseita on kuitenkin yhteensä vain kuusi.

Molemmilla on yhden sanan päälauseita, mutta ei jokaisessa tekstissä, ja yhteensä niitä on vain viisi (Laurén) ja kaksi (Ojanen). Päälauseet kirjoittajilla eroavat tilastollisesti merkitsevästi ( $p = 0,0004$ ).

Kirjoittajien teksteissä sivulauseiden pituudet eivät eroa erityisen paljon toisistaan (esitetty liitteiden kaavioissa 25–32). Ojasella on yksittäisiä pidempiä sivulauseitakin kahdessa tekstissä (O1, O2), mutta muutamasta pitkästä sivulauseesta ei voi vetää johtopäätöksiä. Mahdollisesti isompi aineisto voisi tuoda asiaan selvyyttä. Laurénin kaikki sivulauseet ovat 2–16 sanan mittaisia, ja Ojasellakin melkein kaikki sivulauseet ovat 2–11 sanan pituisia (12 sanan pituisia ja pidempiä yhteensä neljä). Ero kirjoittajien välillä sivulauseissa ei ole myöskään tilastollisesti merkitsevä ( $p = 0,2$ ).

Kumpikin kirjoittaja käyttää verbittömiä ilmaisuja, jotka ovat huudahdustyypisiä, kuten ”Hyvä!” tai omiksi virkkeikseen erotettuja lausekkeita, kuten ”Toinen tiedotusministeriöstä, toinen ulkomaalaisvirastosta.” Niiden määrien perusteella ei voida selkeästi erotella kirjoittajia tämän aineiston valossa. Vaikuttaisi siltä, että Ojanen käyttää kyseistä tehokeinoa hieman enemmän kuin Laurén, ja Ojasen teksteissä on myös pidempiä verbittömiä ilmauksia kuin Laurénin teksteissä. Olen laskenut nämä mukaan kaikkiin lauseisiin, vaikka kaikki verbittömät ilmaukset eivät varsinaisesti täytä lauseen kriteerejä. Sivu- ja päälausevertailusta olen jättänyt verbittömät ilmaukset kokonaan pois.

## 5 LOPUKSI

Ensimmäinen tutkimuskysymykseni oli, voiko kahden eri kirjoittajan saman tekstilajin tekstejä erottaa toisistaan lause-, virke- ja kappalepituuksia analysoimalla? Valitsemani näkökulma ja tutkimusmenetelmä tuotti joidenkin piirteiden osalta kahden kirjoittajan välille selviä ja myös tilastollisesti merkitseviä eroja. Näin ei ollut kaikkien piirteiden kohdalla, mutta kokonaisuudessaan sanoisin, että ainakin tämän kyseisen aineiston kohdalla kaksi kirjoittajaa on mahdollista erottaa toisistaan analysoimalla lauseiden, virkkeiden ja kappaleiden pituuksia. On mahdollista ajatella, että näillä kahdella kirjoittajalla on omanlaisensa tyyli kirjoittaa, ja se näkyy myös heidän tavassaan jakaa tekstiä kappaleiksi, virkkeiksi ja lauseiksi.

Toinen ja kolmas kysymykseni olivat, miten kirjoittajien tekstit eroavat toisistaan valitsemieni piirteiden osalta ja minkälaisia tyylliseen ominaisia valintoja lauseiden, virkkeiden ja kappaleiden pituuksissa kirjoittajat tekevät. Keskeisimmät tulokset käyn läpi seuraavaksi.

Kappaleiden määrät kirjoittajien teksteissä ovat melko yhteneväiset. Sen sijaan jonkinlainen ero näkyy kirjoittajien välillä virkkeiden määrissä. Laurénilla virkkeitä on yhteensä



enemmän kuin Ojasella, ja siten myös Laurénin kappaleissa on keskimäärin enemmän virkkeitä kuin Ojasen kappaleissa.

Sekä Laurén että Ojanen kirjoittavat paljon lyhyitä kappaleita. Tekstien pisimmät viiden virkkeen kappaleet ovat vain yksittäisiä. Molemmilla kirjoittajilla neljästä tekstistä kolmessa on eniten yhden ja kahden virkkeen kappaleita. Toisaalta myös aiemman tutkimuksen mukaan oppikirjoista, sanomalehtitekstistä ja kaunokirjallisuudesta keskimäärin lyhyimmät kappaleet löytyvät juuri sanomalehtiteksteistä. Voidaan nähdä, että sanomalehtitekstissä kappaleiden pituuteen vaikuttaa asiakokonaisuuksien mukaan jäsentelyn lisäksi myös ymmärrettävyys, toimitustyön vaatimukset ja myös graafiset ja visuaaliset vaikutukset. (Anhava 1987.)

Laurénin ja Ojasen tekstien kappaleet eroavat toisistaan siten, että vaikka molemmilla on lyhyitä kappaleita, Ojasen teksteissä painottuvat yhden ja kahden virkkeen kappaleet vielä selkeämmin. Laurén käyttää tasaisemmin eri pituisia kappaleita.

Virkkeiden pituuksissa kirjoittajat eroavat siten, että Ojasen tekstien virkkeet ovat keskimäärin pidempiä kuin virkkeet Laurénin teksteissä. Ojasen teksteissä virkkeiden keskimääräinen sanamäärä on kolmessa tekstissä lähellä Heikkisen, Lehtisen ja Lounelan (2001) saamia tuloksia suomenkielisten tekstien keskimääräisistä virkkeen pituuksista. Yhdessä Ojasen teksteistä se puolestaan on lähellä vanhemmista tutkimuksista (Hakulinen ym. 1980) saatuja tuloksia. Laurénin teksteissä virkkeiden keskimääräinen sanamäärä jää lyhyemmäksi verrattuna aiempiin tutkimuksiin. Toisaalta Heikkinen ym. (2001) varovasti ennustavat virkkeiden lyhentymistä ajan myötä suomen kielessä, jolloin saamani tulokset eivät ole suuresti ristiriidassa aiemman tutkimuksen kanssa.

Molemmilla kirjoittajilla on sanamääriltään eri pituisia virkkeitä, eikä kummallakaan kirjoittajalla ole selkeää yhtä yleisintä virkepituutta, joka toistuisi useammassa tekstissä, eikä kummallakaan kirjoittajalla ole tekstiä, jossa virkkeet olisivat erityisen painottuneet sanamäärältään tietynmittaisiin. Kirjoittajien virkkeet eroavat siten, että Laurénilla on enemmän lyhyempiä (5–11 sanan) virkkeitä kuin Ojasella. Laurénilla on myös enemmän erittäin lyhyitä virkkeitä, ja Ojasella puolestaan on selvästi enemmän pitkiä (yli 20 sanan) virkkeitä.

Verrattaessa omia tuloksiani aiempaan tutkimukseen (Heikkinen ym. 2001; Hakulinen ym. 1980) näkyy lausepituuden kohdalla samansuuntainen tulos kuin virkepituuden kohdalla. Ojasen teksteissä keskimääräinen lauseiden sanamäärä on melko yhteneväinen suomen tekstilauseiden keskimääräisten pituuksien kanssa, kun taas Laurénin teksteissä lauseiden keskimääräiset sanamäärät jäävät alle yleisen keskiarvon.

Kirjoittajien lauseiden pituudet eivät eroa toisistaan yhtä selkeästi kuin virkkeiden pituudet. Molemmilla kirjoittajilla suurin osa lauseista on 3–9 sanan mittaisia. Kuitenkin erona on se,

että Laurenin lauseet ovat sanamäärien osalta hyvin yhteneväisiä keskenään, kun Ojasella tekstit eroavat toisistaan enemmän. Toinen näkyvä ero kirjoittajien välillä on se, että Ojanen käyttää teksteissään yksittäisiä pitkiä lauseita (yli 16 sanan mittaisia), joita Laurén ei käytä lainkaan.

Sivu- ja päälauseiden suhteellinen osuus on kirjoittajilla hyvin lähellä toisiaan. Päälauseet ovat painottuneet Laurénilla hieman lyhyemmiksi kuin Ojasella. Ojasella puolestaan on yksittäisiä pitkiä päälauseita, joita Laurénilla ei ole. Sivulauseiden osalta ei tämän tutkimuksen aineiston perusteella ollut löydettävissä selkeitä eroavaisuuksia. Verbittömiä ilmauksia kumpikin kirjoittaja käyttää teksteissään jonkin verran. Ojasen teksteissä näitä oli hieman enemmän ja ne olivat pidempiä.

Viimeinen tutkimuskysymykseni oli, onko saman kirjoittajan teksteistä löydettävissä sellaisia yhteneväisiä rakenteellisia piirteitä, joiden perusteella kirjoittajan voi tunnistaa. Tämän tutkimuksen puitteissa ei ole mahdollista vielä vastata, voiko tutkimuksessani tarkasteltujen tekstinjäsenyksellisten ja syntaktisten valintojen perusteella tunnistaa tekstin kirjoittaneen henkilön eroon muista kirjoittajista. Jotta kysymykseen saisi vastauksen, olisi aineistoon otettava mukaan useampia eri kirjoittajia. Tällaisen tutkimuksen valossa voisi selvittää, muodostuuko näiden piirteiden osalta jonkinlaisia kirjoittajaryhmiä, joilla on samankaltainen tyyli, vai olisiko jopa mahdollista erottaa kaikki kirjoittajat toisistaan. Kirjoittajan tunnistamista tekstiä tutkimalla hyödynnetään esimerkiksi rikostutkinnassa.

Tutkimuksessani myös ilmeni molemmilla kirjoittajilla joidenkin piirteiden puolesta poikkeavia tekstejä, jotka erosivat kirjoittajan muista teksteistä. Laajentamalla aineistoa useampiin teksteihin samalta kirjoittajalta voisi selvittää, oliko kyseessä tosiaan erikoinen poikkeus, vai onko kirjoittajilla taipumus kirjoittaa erityylisiä tekstejä enemmänkin. Tilastollisten testien tekeminen suuremmalla aineistolla puolestaan antaisi varmemman tuloksen.

Jatkotutkimuksella olisi aineiston laajentamisen lisäksi mahdollisuus myös syventyä aiheeseen. Tarkempi tilastollinen tarkastelu antaisi mahdollisesti lisätietoa. Lisäksi olisi mahdollista syventyä tarkastelemaan esimerkiksi kirjoittajien käyttämiä lausetyyppejä ja lausekkeiden rakentumista ja selvittää, tuovatko ne eroa kirjoittajien välille.

## 6 LÄHTEET

- ANHAVA, JAAKKO 1987: Kappalejaon tehtävä ja sen opettaminen. – *Kielikello* 3/1987.
- HAKULINEN, AULI – KARLSSON, FRED – VILKUNA, MARIA 1980: *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus*. Helsinki: Helsingin yliopisto.
- ISK = HAKULINEN, AULI – VILKUNA, MARIA – KORHONEN, RIITTA – KOIVISTO, VESA – HEINONEN, TARJA RIITTA – ALHO, IRJA 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- VOUTILAINEN, EERO 2012: Tyyli. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiirilä & Mikko Lounela (toim.), *Genreanalyysi: tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus.
- HEIKKINEN, VESA – LEHTINEN, OUTI – LOUNELA, MIKKO 2001: Kuvia kirjoitetusta suomesta. – *Kielikello* 3/2001.
- HOLOPAINEN, MARTTI – PULKKINEN, PEKKA 2015: *Tilastolliset menetelmät*. Helsinki: Sanoma Pro Oy.
- LÖFLUND, JUHANI 2002: *Suomen syntaksin oppikirja: suomen ja ruotsin lauserakenteiden vertailua*. Turku: Åbo Akademi.
- MAKKONEN-CRAIG, HENNA 2005: *Toimittajan läsnäolo sanomalehtitekstissä Näkökulmia suomen kielen dialogisiin passiivilauseisiin*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- MÄKINEN, KIRSTI 1992: Muuttuuko lehtikieli? Muuttaako lehtikieli? – Valma Yli-Vakkuri, Maija Länsimäki & Aarre Nyman (toim.), *Yhteiskunta muuttuu – kieli muuttuu: Nykysuomen Seuran 10-vuotisjuhlakirja*. Juva: WSOY.
- MÄNTYNYNEN, ANNE 2006: Näkökulmia tekstin ja tekstilajin rakenteeseen. – Anne Mäntynen, Susanna Shore & Anna Solin (toim.), *Genre - tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- OJALA, AATOS 1971: *Johdatus lingvistiseen tyylintutkimukseen*. Jyväskylä: Jyväskylän yliopisto.
- RENTOLA, ROOSA 2019: *Kielestä kiinni: kuinka kielentutkimusta käytetään rikosten selvittämisessä?* Sastamala: Warelia.
- SAUKKONEN, PAULI 2001: *Maaailman hahmottaminen teksteinä: tekstirakenteen ja tekstilajien teoriaa ja analyysia*. Helsinki: Yliopistopaino.
- SAUKKONEN, PAULI 1984: *Mistä tyyli syntyy?* Helsinki: WSOY.  
Jyväskylä: Oy Finn Lectura Ab.
- SHORE, SUSANNA – MÄNTYNYNEN, ANNE 2006: Johdanto. – Anne Mäntynen, Susanna Shore & Anna Solin (toim.), *Genre - tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- SUHOLA, AINO – TURUNEN, SEPPO – VARIS, MARKKU 2005: *Journalistisen kirjoittamisen perusteet*.

VARES, VESA – SILTALA, SAKARI 2016: *Sanan ja kuvan vuosisata*. Helsinki: Kustannusosakeyhtiö Otava.

VERDONK, PETER 2002: *Stylistics*. Oxford: Oxford University Press.

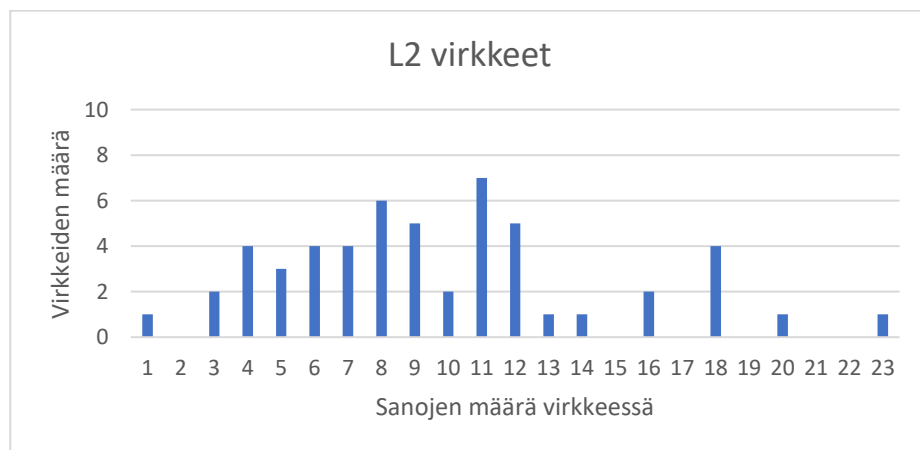
## 7 TAULUKOITA

Teksti	Kappaleiden pituudet virkkeiden määrän mukaan					yhteensä kappaleita	yhteensä virkeitä
	1	2	3	4	5		
L1	9	8	3	7	2	29	72
L2	6	8	5	4	0	23	53
L3	7	7	5	3	1	23	53
L4	5	10	7	3	1	26	63
<b>L yht.</b>	27	33	20	17	4	101	241
O1	9	8	3	2	0	22	42
O2	8	14	4	1	0	27	52
O3	12	13	1	2	0	28	49
O4	4	5	9	2	1	21	54
<b>O yht.</b>	33	40	17	7	1	98	197

Taulukko 4. Virkkeiden määrä kappaleissa



Kaavio 1. Tekstissä L1 esiintyvien virkkeiden sanamääräjakauma



Kaavio 2. Tekstissä L2 esiintyvien virkkeiden sanamääräjakauma



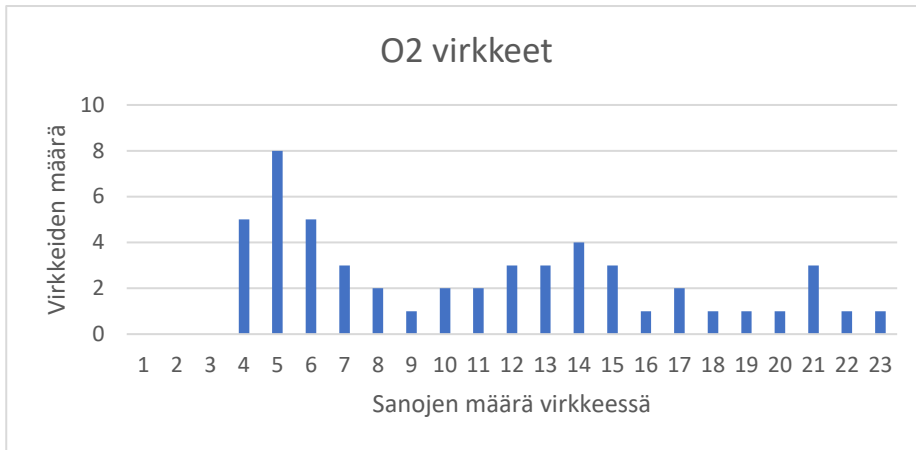
Kaavio 3. Tekstissä L3 esiintyvien virkkeiden sanamääräjakauma



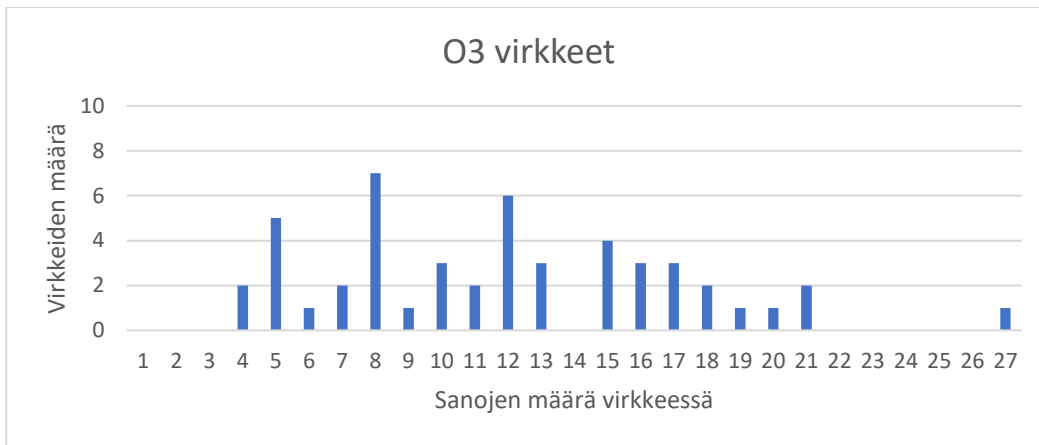
Kaavio 4. Tekstissä L4 esiintyvien virkkeiden sanamääräjakauma



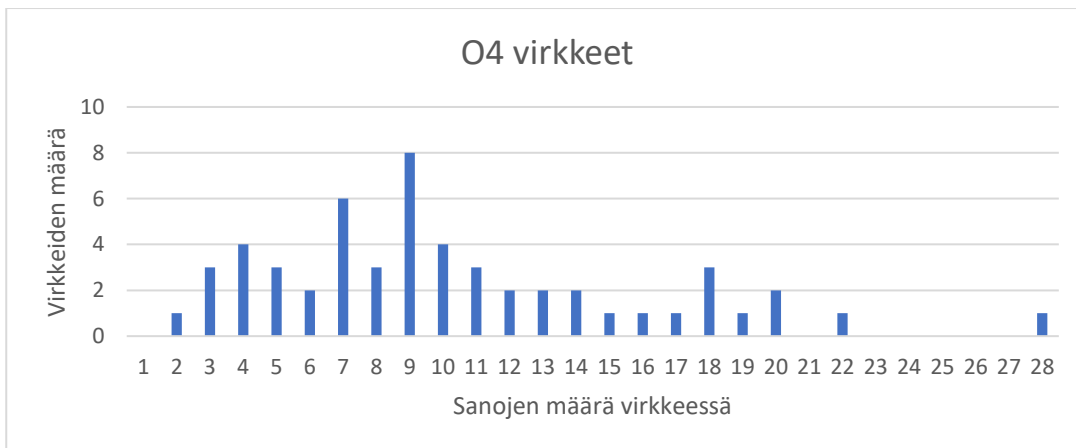
Kaavio 5. Tekstissä O1 esiintyvien virkkeiden sanamääräjakauma



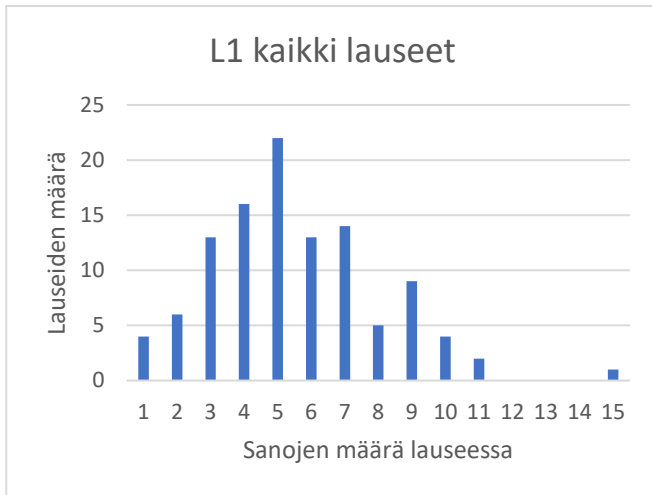
Kaavio 6. Tekstissä O2 esiintyvien virkkeiden sanamääräjakauma



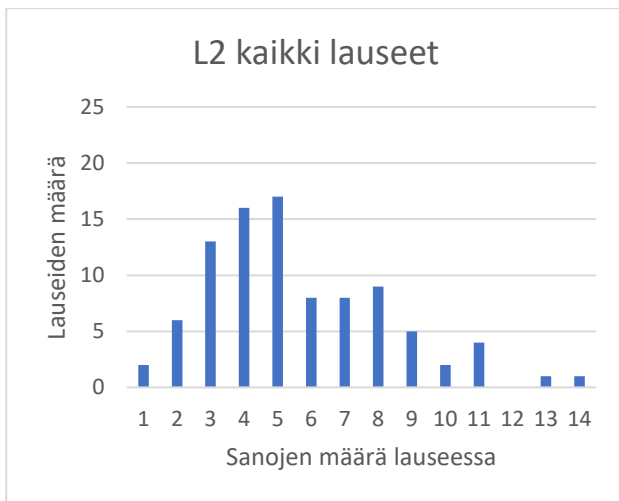
Kaavio 7. Tekstissä O3 esiintyvien virkkeiden sanamääräjakauma



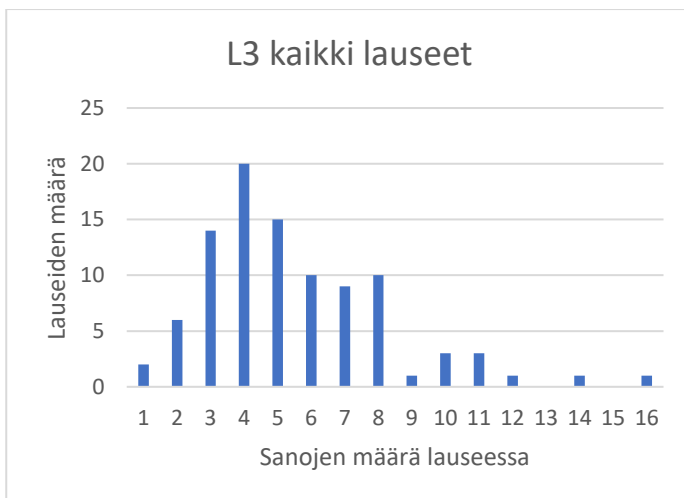
Kaavio 8. Tekstissä O4 esiintyvien virkkeiden sanamääräjakauma



Kaavio 9. Tekstissä L1 esiintyvien kaikkien lauseiden sanamääräjakauma

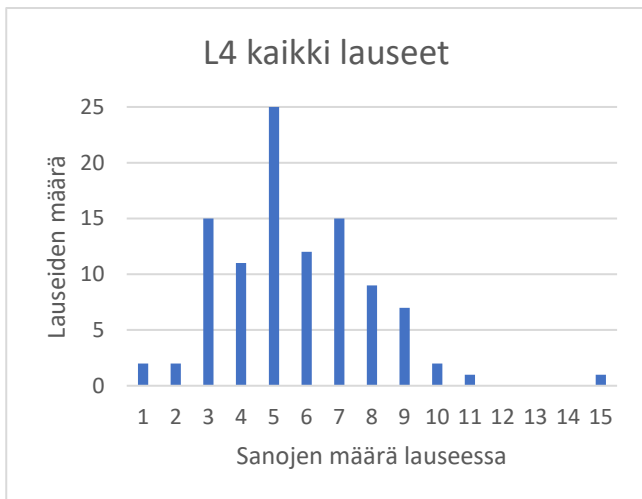


Kaavio 10. Tekstissä L2 esiintyvien kaikkien lauseiden sanamääräjakauma

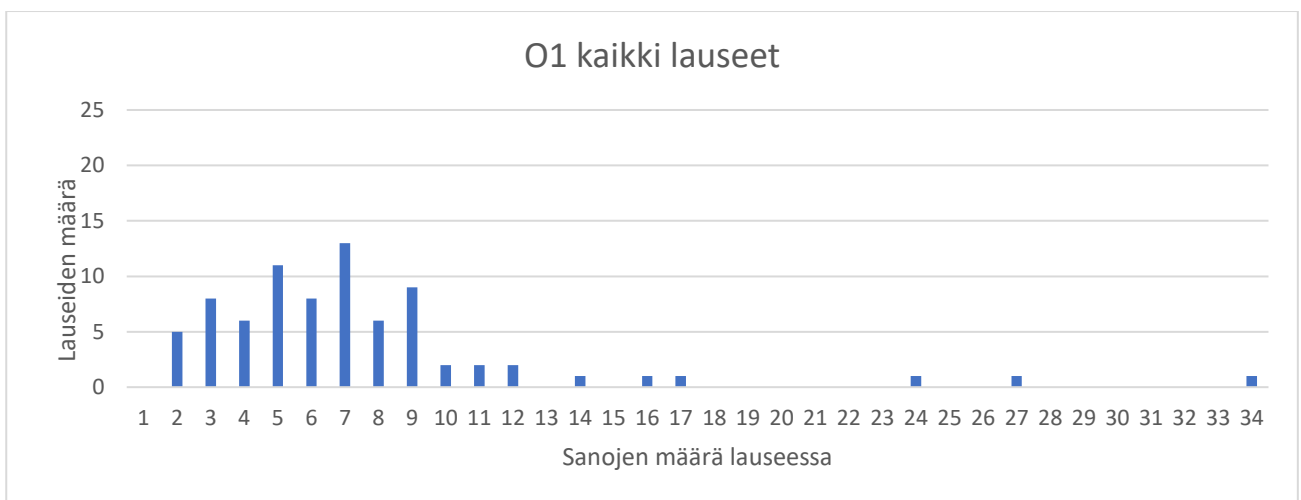


Kaavio 11. Tekstissä L3 esiintyvien kaikkien lauseiden sanamääräjakauma

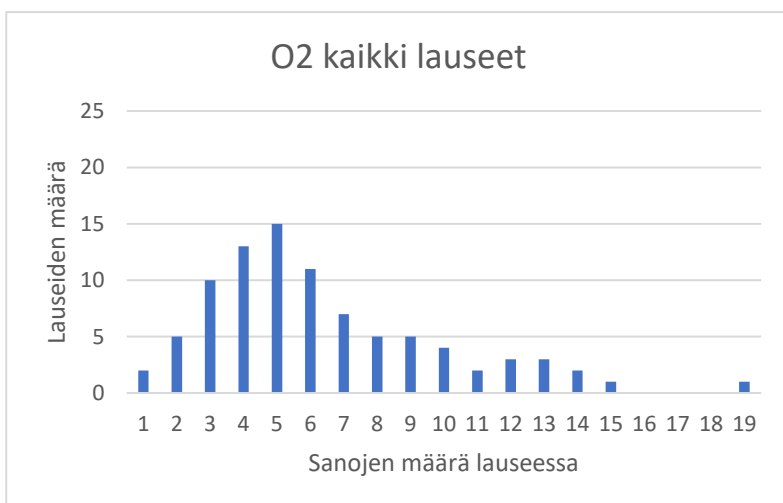




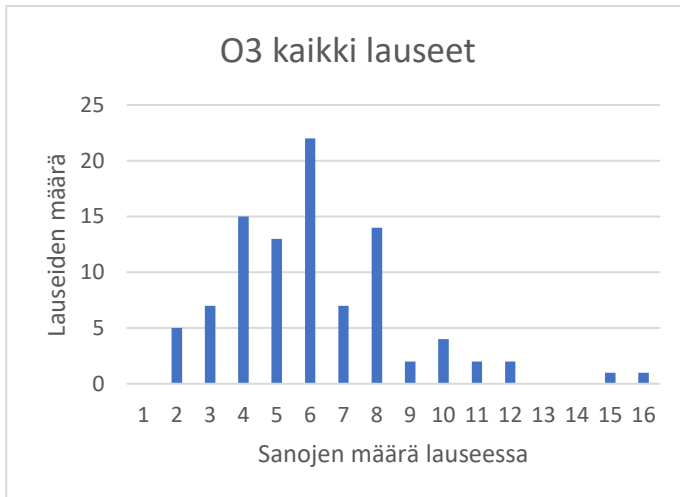
Kaavio 12. Tekstissä L4 esiintyvien kaikkien lauseiden sanamääräjakauma



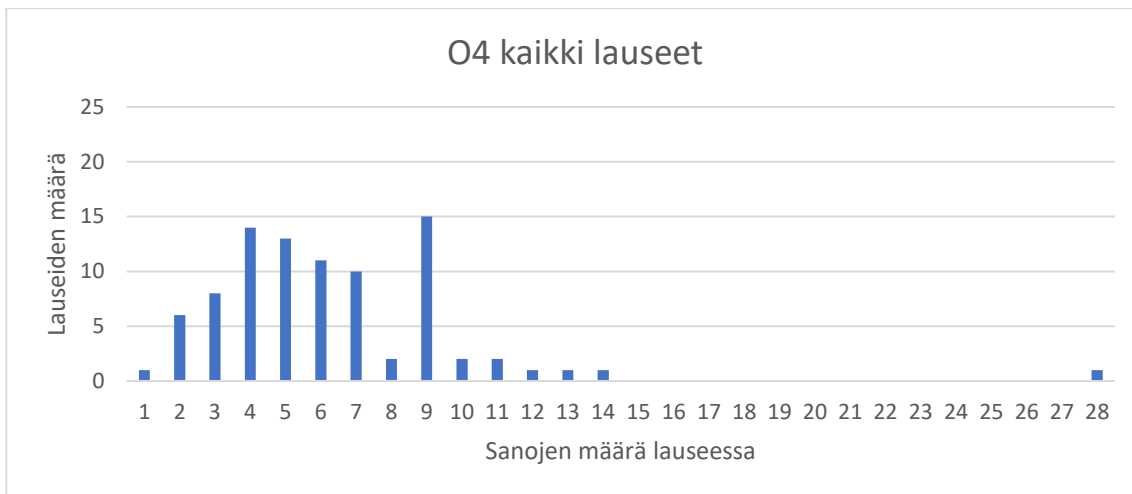
Kaavio 13. Tekstissä O1 esiintyvien kaikkien lauseiden sanamääräjakauma



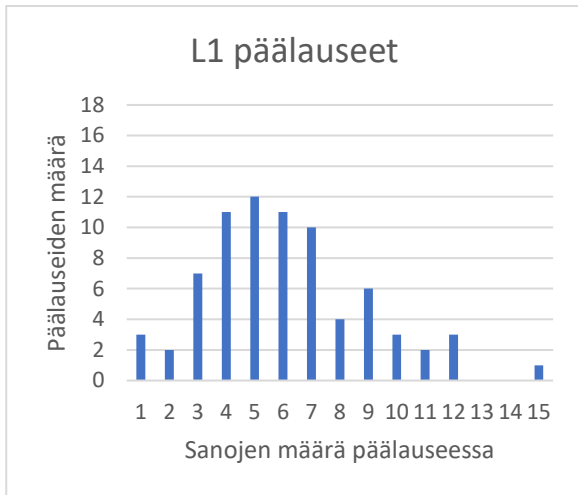
Kaavio 14. Tekstissä O2 esiintyvien kaikkien lauseiden sanamääräjakauma



Kaavio 15. Tekstissä O3 esiintyvien kaikkien lauseiden sanamääräjakauma



Kaavio 16. Tekstissä O4 esiintyvien kaikkien lauseiden sanamääräjakauma



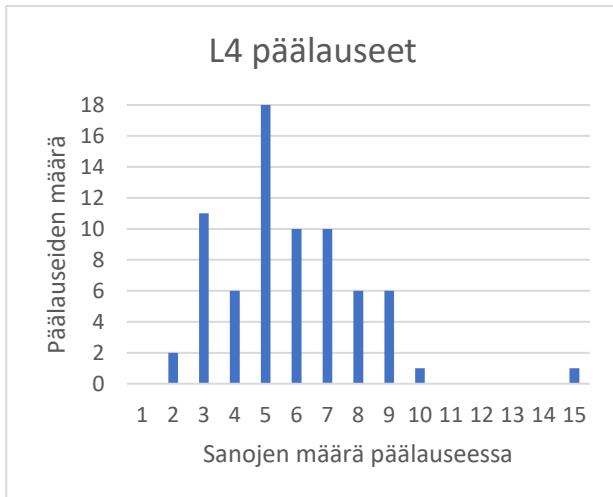
Kaavio 17. Tekstissä L1 esiintyvien päälauseiden sanamääräjakauma



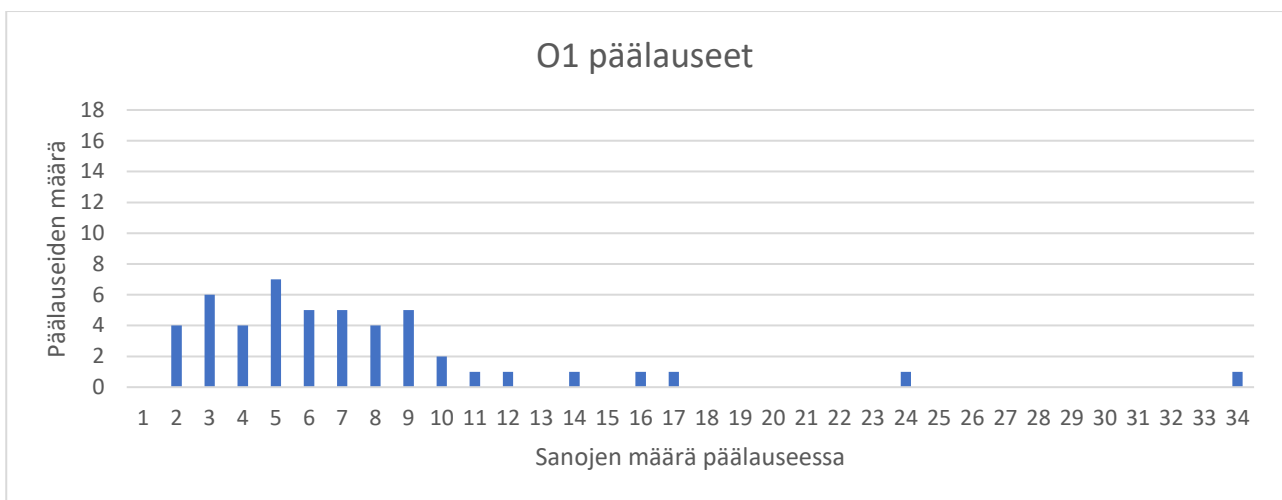
Kaavio 18. Tekstissä L2 esiintyvien päälauseiden sanamääräjakauma



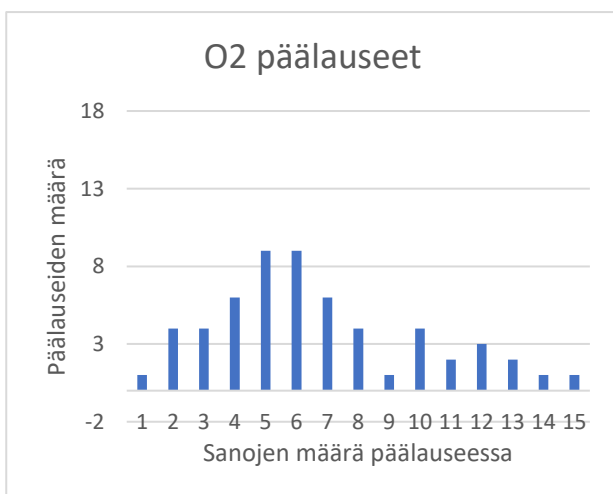
Kaavio 19. Tekstissä L3 esiintyvien päälauseiden sanamääräjakauma



Kaavio 20. Tekstissä L4 esiintyvien päälauseiden sanamääräjakauma



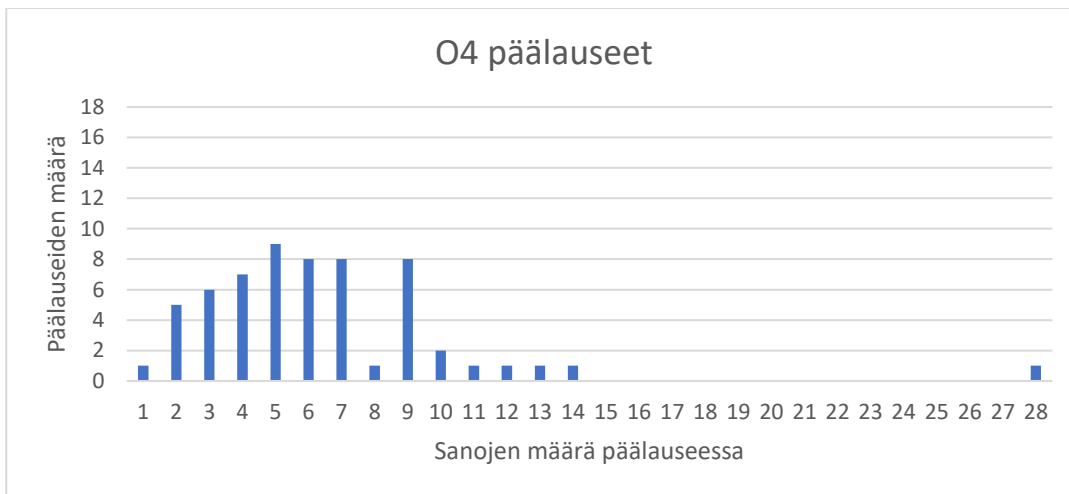
Kaavio 21. Tekstissä O1 esiintyvien päälauseiden sanamääräjakauma



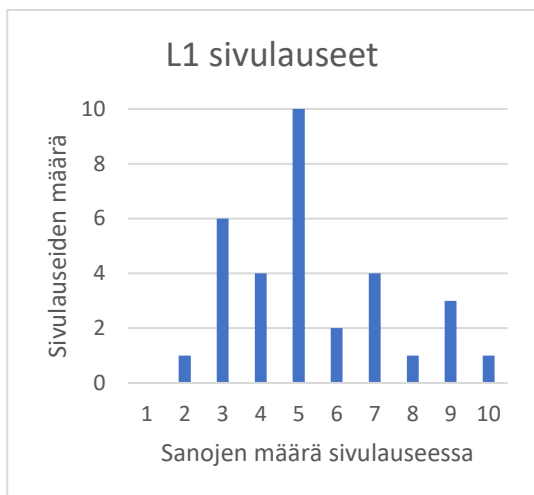
Kaavio 22. Tekstissä O2 esiintyvien päälauseiden sanamääräjakauma



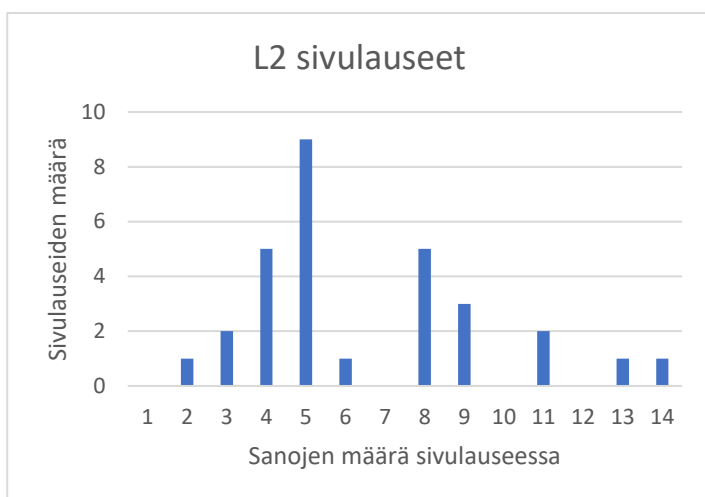
Kaavio 23. Tekstissä O3 esiintyvien päälauseiden sanamääräjakauma



Kaavio 24. Tekstissä O4 esiintyvien päälauseiden sanamääräjakauma



Kaavio 25. Tekstissä L1 esiintyvien sivulauseiden sanamääräjakauma



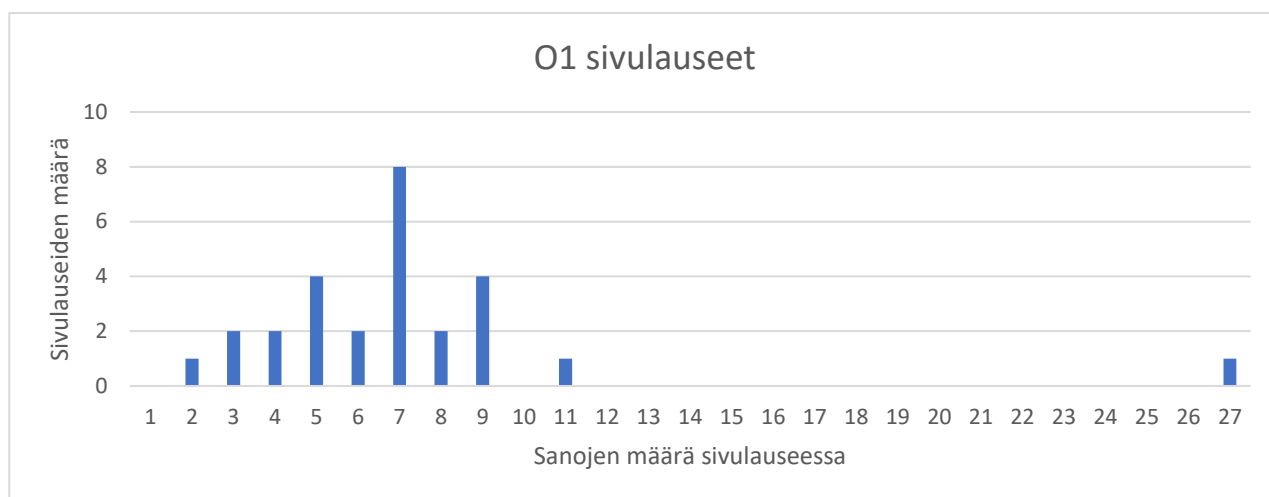
Kaavio 26. Tekstissä L2 esiintyvien sivulauseiden sanamääräjakauma



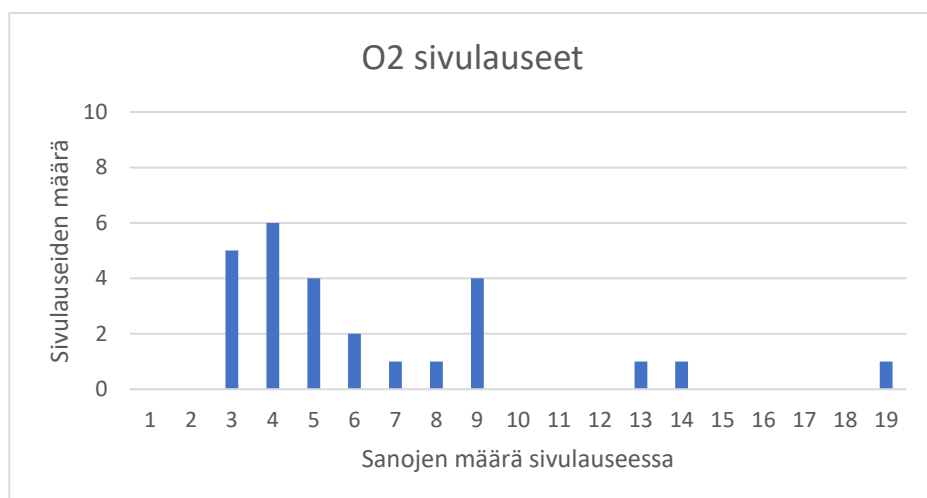
Kaavio 27. Tekstissä L3 esiintyvien sivulauseiden sanamääräjakauma



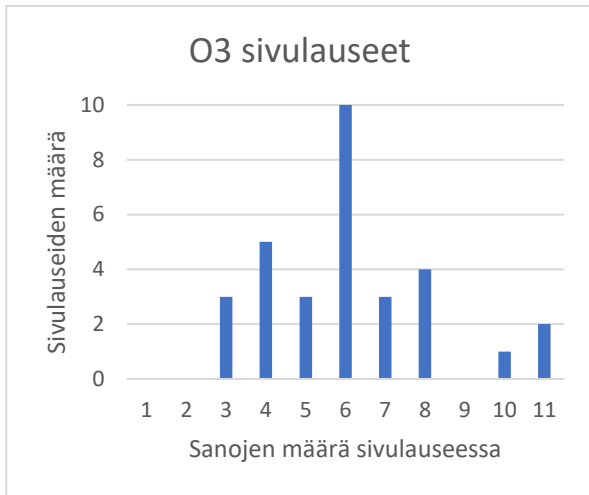
Kaavio 28. Tekstissä L4 esiintyvien sivulauseiden sanamääräjakauma



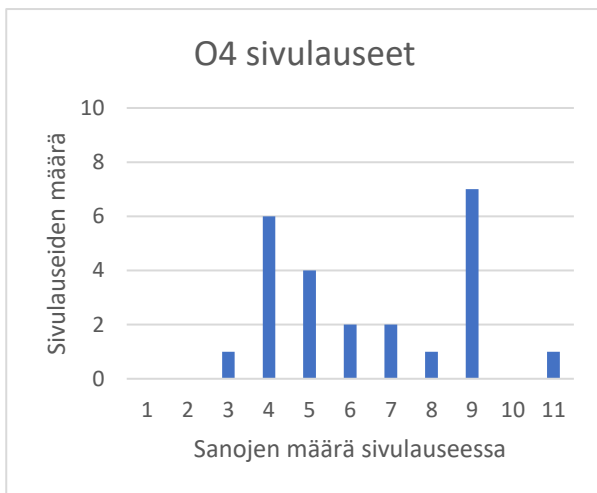
Kaavio 29. Tekstissä O1 esiintyvien sivulauseiden sanamääräjakauma



Kaavio 30. Tekstissä O2 esiintyvien sivulauseiden sanamääräjakauma



Kaavio 31. Tekstissä O3 esiintyvien sivulauseiden sanamääräjakauma



Kaavio 32. Tekstissä O4 esiintyvien sivulauseiden sanamääräjakauma