

# Pari sanaa asumisesta – miten yliopisto-opiskelijoiden kirjoittamissa vaihtoraporteissa asetetaan tekstikappaleen topiikki?

*Kirsi Sandberg & Juho Härme*

## Contents

<b>1 Johdanto</b>	<b>1</b>
<b>2 Tekstikappale ja sen topiikki</b>	<b>2</b>
<b>3 Tutkimusaineisto ja -metodi</b>	<b>3</b>
<b>4 Tulokset</b>	<b>6</b>
4.1 Laadullinen analyysi . . . . .	6
4.2 Tilastollinen tarkastelu . . . . .	10
<b>5 Päätelmät</b>	<b>12</b>

This paper explores the various strategies writers use to shift from one topic to another based on a corpus of student exchange reports. The main questions addressed in the study are 1) what linguistic means students use to set a new topic and 2) how these means could be defined functionally and syntactically. The data for the study consists of xxx reports from exchange period abroad. Both qualitative and statistical approaches were adopted in order to capture the complexities of the phenomenon. The paper describes 13 separate functional groups of topic-setting strategies. The statistical analysis conducted on these groups shows that majority of them can be drawn from three simple micro-level variables. This information can be used to develop teaching and writing practises especially in L2 context but to some degree in L1 also.

Avainsanat: topiikki, kirjoittaminen, korpusmentelmät

Keywords: topic, writing, corpus-based methods

## 1 Johdanto

Opiskelijavaihdosta kirjoitettava vaihtoraportti on institutionaalinen teksti, jonka kirjoittamiseen yliopisto antaa sisältöpainotteisen ohjeistuksen. Näillä verkossa julkaistavilla raporteilla on selkeä funktio: tukea vaihtojaksoa suunnittelevia opiskelijoita päätöksenteossa ja omaan tulevaan vaihtojaksoonsa valmistautumisessa. Niiden voi katsoa edustavan samaa tekstilajia (tekstilajin määritelmästä ks. esim. (Juvonen ym. 2012)) tähän vielä Suvin määritelmä mukaan. Vaihtoraportti, kuten tekstilajit yleensäkin on dynaaminen ja voi toteutua kielellisesti äärettömän monella eri tavalla, ja tekstiä, jossa kieltä käytetään varioiden ja monipuolisesti mutta lajinsa konventioita riittävässä määrin noudatellen arvostetaan ja pidetään lukijan näkökulmasta luettavana ja mielenkiintoisena.

Lähestyessämme 2020-lukua Suomen oppilaitoksissa ja korkeakouluissa opiskelee kasvava joukko kirjoittajia (L1 ja L2), jolla ei opintojensa alkuvaiheessa ole opinnoissa menestymiseen tarvittavia tekstitaitoja. Akateemisen kirjoittamisen (laajemmin erikoiskielten) tarkasteluun liittyy usein käsitys, että tekstilajitietoisuus ja -tuntemus edesauttaa ja laajentaa kielenkäyttäjän kielellistä repertuaaria (Vuorijärvi 2013: 15) ja siinä mainitut lähteet. Sitä, miten oppija pääsee tekstilajin kautta kiinni kielen rakenteiden varioiviin käyttötapoihin, ei Suomen kontekstissa ole nähdäksemme tutkittu. Funktionaalisista lähtökohdista tarkasteltuna esitetty näkemys

kaipaisi kuitenkin empiiristä tutkimusta, koska myös rakenteiden hallinnan on osoitettu olevan kytköksissä tekstin laatuun (ks. esim. (Jisa & Tolchinsky 2009))

Nykyisessä funktionaalista kielikäsitystä ja tutkivaa oppimista korostavassa pedagogisessa kehyksessä tarjolla olevat pääosin preskriptiiviset ohjeistukset tekstikappaleen rakentumisesta (ks. esim. xxx) eivät palvele tarkoitustaan. Ns. uusi kirjoittaja on aktiivinen tiedonhakija ja tekstien tutkija, joka ei enää kysy, miten minun pitäisi tämä tehdä vaan millä eri tavoilla tämä on tehty. Tällä kirjoittajalla voi olla nykyisen, tekstilajitietoisen kouluopetuksen pohjalta ymmärrys tekstilajista, mutta repertuaarissa ei olekaan sopivaa rakennetta, jolla pystyisi itse tuottamaan tavoitemuotoisen tekstin ja esim. käymään dialogia lukijansa kanssa (ks. (Juvonen 2014a)).

Viime aikaisessa kirjoittamista käsittelevässä fennistisessä tutkimuksessa on tarkasteltu sitä, miten tekstikokonaisuus rakentuu funktionaalisten tai retoristen jaksojen pohjalta. On myös tarkennettu siihen, millaisin kielisin keinoin kirjoittajat toteuttavat tekstin rakenteen kannalta keskeisiä funktioita tai siirtoja (ks. esim. (Juvonen 2014a)(Komppa 2012)(Paldanius 2017)(Vuorijärvi 2013)). Hierarkkisesti rakentuneiden tekstien rinnalla eää huomattavan paljon tekstilajeja (Hiippala 2013)(Komppa 2012)(Vuorijärvi 2013)), joissa tekstin rakennetta voi kuvata enemmän topiikkipohjaiseksi ja temaattisen sisällön pohjalta rakentuvaksi kuin hierarkkisesti. Rakenne muodostuu sekventiaalisesti tai lineaarisesti peräkkäin asetetuista sisältötopiikeista, jotka kytkeytyvät tavalla tai toisella diskurssitopiikkiin. Hierarkkisten tekstirakenteiden rinnalla tekstin rakentaminen ns. sisältö edellä vaikuttaisikin olevan yhtä lailla relevantti kirjoitusstrategia myös korkeakoulussa.

Saatavilla olevan tutkimustiedon rinnalle haluammekin selvittää, miten kirjoittaja voi päästä alkuun kappaleen kirjoittamisessa, jos häntä ohjaa tieto tekstikappaleen topiikista, ts. siitä mitä kappaleen pitäisi pitää sisällään. Meitä kiinnostaa, millä tavalla ja mitä rakenteita käyttäen kirjoittaja asettaa tekstikappaleen sisältötopiikkiin ja siirtyy tekstitopiikista toiseen. Analyysin avulla haetaan vastausta seuraaviin tutkimuskysymyksiin:

- 1) Millaisia tapoja yliopisto-opiskelijat käyttävät saman topiikin asettamiseen?
- 2) Miten näitä tapoja voidaan määritellä toisaalta funktionaalisesti, toisaalta syntaktisesti?

Tutkimus toteutetaan korpusvetoisesti yhdistelemällä laadullisia (osio 4.1) ja määrällisiä (osio 4.2) menetelmiä.

## 2 Tekstikappale ja sen topiikki

Tämä tutkimus on soveltava ja jo lähtökohdiltaan tähtää kirjoittamisen oppimisen ja opettamisen kehittämiseen sekä sovellusmahdollisuuksiin. Se asettuu erikoiskielten kontekstiin, mutta lähestyy ennemminkin tekstin, kielen rakenteiden ja leksikon jakamia rajapintoja kuin tekstin kokonaisuutta tai tekstilajin muodostumista sinänsä. Samasta lähtökohdasta on myös valittu keskeinen analyysiyksikkö - tekstikappale - joka on toistaiseksi harvakseltaan ollut tutkimuksen kohteena (ks. kuitenkin x ja y) mutta jonka kirjoittamiseen ohjeistetaan oppikirjoissa ja kirjoittamisoppaissa.

Kirjoittajan kannalta kappalejako on hyvin konkreettinen työkalu ja funktionaalinen kielenyksikkö, vaikka tutkimuksessa niitä lähestytään lähinnä ortografisina merkkeinä, jotka eivät itsessään ole merkitseviä tekstin rakentumisen kannalta. Kirjoittamisoppaissa ja ohjeistuksissa annetaan kuvauksia siitä, millä tavalla teksti rakentuu kappaleistaan, ja millaisia kappaleiden – sisällön ja funktion tasolla – olisi hyvä olla (viitteet tähän vielä). Lisäksi kappaleiden koherentti toisiinsa sitominen on tyypillisesti arviointikriteeri, esim. ylioppilaskirjoituksissa (viite).

Englanninkielisessä kirjoittamisen opetuksen perinteessä on ollut tapana käyttää topic sentence -käsitettä kuvaamaan virkettä, jolla kirjoittaja ilmaisee kappaleen tärkeimmän ajatuksen (ks. esim. Smith 2008). Suomeksi tämä kääntyy nykyisellään ydinvirkkeeksi, joka yhdessä tukivirkkeiden kanssa muodostaa kappaleen. Tekstikappaleen ytimen tavoittelun sijaan kohdistammekin huomiomme siihen, miten kirjoittaja siirtyy tekstissään aiheesta toiseen. Nimitämme siirtymän toteuttavia virkkeitä *topiikkia asettaviksi* vrt. (Juvonen 2014b).

Käyttämämme topiikin käsite pohjaa Simon Dikin funktionaalisessa kieliopissaan (1989) antamaan esitykseen, jota vastaavassa tekstikappaleiden rakenteen analyysissä on soveltanut mm. Jesus Guijarro (2001). Dikin

mukaan (mts. 266) kaikilla diskursseilla (joka Dikin määrittelemänä tarkoittaa mitä tahansa koherenttia tekstiä) on se ominaisuus, että ne kertovat jostakin. Tämä ominaisuus, jota nimitämme *aiheisuudeksi* (*aboutness*), koskee paitsi tekstiä kokonaisuutena, myös siitä eroteltuja alakokonaisuuksia, kuten kappaleita ja virkkeitä. Määrittelemmekin topiikin tämän artikkelin puitteissa siksi, mistä teksti, kappale tai virke kertoo (vrt. Lambrecht 1996: 118).

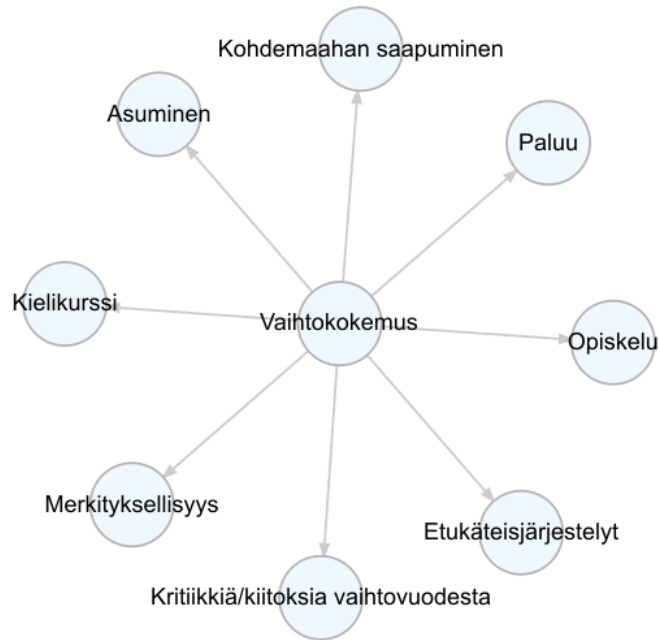
Tekstin sisällä topiikit muodostavat Dikin (mt. 271, vrt. myös Givón 1983) jaottelussa topiikkiketjuja, joiden katsomme tässä artikkelissa koostuvan *esittelytopiikeista* (Dikillä *new topic*), *esitellyistä topiikeista* (Dikillä *given topic*) ja *lohkotopiikeista* (Dikillä *subtopic*). Esittelytopiikkeja ovat tapaukset, joissa kirjoittaja tuo jonkin aiheen tekstiin ensimmäistä kertaa ja esiteltystä topiikkeja taas sellaiset tapaukset, jotka on jo tuotu aiemmin ja joihin nyt palataan. Lohkotopiikeilla viittamme tilanteeseen, jossa kirjoittaja käyttää hyväkseen johonkin edellä esitellyyn topiikkiin kiinteästi liittyvää assosiaatioiden joukkoa viitatakseen entiteettiin, jota itseään ei vielä ole mainittu.

Ajatus lohkotopiikista kytkeytyy Wallace Chafen semiaktiivisuuden käsitteeseen (1987: 29), ja käytettävissä olemiseen Knud Lambrechtilla (1996: 99). Esimerkki Dikiltä (mts. 275) on lause *John gave a party last week, but the music was awful*, jossa *musiikki* on juhlien käsitteen avulla tekstiin tuotu lohkotopiikki. Lohkotopiikkien muodostumisessa keskeisiä ovatkin sanojen väliset merkityssuhteet kuten hyponymia ja meronymia (vrt. Shore 2008: 33). Shoren esityksessä suomen tekstuaalisesta jäsennyksestä assosiaatioihin perustuviin tekstin osien välisiin suhteisiin suhtaudutaan varauksella; ilmiötä kuvataan implisiittisen teeman käsitteellä (Shore 2008, ?).

Huolimatta siitä, että tekstuaalisia suhteita, joita on haasteellista ankkuroida eksplisiittiisiin tekstin osiin, on kuvattu eri viitekehyksissä eri tavoin ja suomen kontekstissa ehkä myös vältetty kuvaamasta, niiden merkitys kirjoittajalle on keskeinen. Aineistopohjaisissa tutkimuksissa on nimittäin paikallistettu huomattava joukko hierarkkisesti rakentuvien tekstien rinnalla esiintyviä ns. moniylimisiä tekstejä, joiden osat ovat enemmän rinnasteisia, keskenään tasavertaisia, eikä yhtä voida asettaa toisten ytimeksi (Hiippala 2013)(Komppa 2012)(Vuorijärvi 2013). Vuorijärvi esittelee Paltridgeä [täydennä lähdeluetteloon] mukaillen topiikkipohjaisen rakenteen, jossa tekstin aihetta avataan ikään kuin osa kerrallaan (Vuorijärvi 2013). Tämä topiikkipohjainen skeema kuvaa hyvin aineistonamme olevien matkaraporttien tekstilajia, ja topiikin käsittemme rinnastuu monessa suhteessa mainittujen lähteiden käsitteistykseen. Toisin sanoen puhumme sisältötopiikista eli aiheesta tai aihepiiristä, josta tekstissä sanotaan jotakin.

### 3 Tutkimusaineisto ja -metodi

Nyt käsillä olevan tutkimuksen aineisto koostuu siis yliopisto-opiskelijoiden vaihtokokemuksestaan kirjoittamista teksteistä. Tarkka tehtävänanto tekstien laatimiseen oli annettu erillisissä ohjeissa, joissa opiskelijoita pyydetään kirjoittamaan ”matkakertomus, jossa kuvaillaan ja arvioidaan opiskelua ulkomaisessa yliopistossa tai harjoittelua ulkomailla”. Ohjeet sisältävät kahdeksankohtaisen listan mahdollisesti käsiteltävistä asioista (mm. *etukäteisjärjestelyt*, *asuminen* ja *merkityksellisyys*). Listaa kehoitetaan tulkitsemaan toteamalla, että ”seuraavat ohjeet eivät ole sitovia, mutta ne on hyvä lukea läpi ennen kirjoittamista”. Vaikka ohjeet eivät siis velvoita noudattamaan juuri tiettyä sisältörakennetta, ylivoimaisesti suurin osa teksteistä on rakennettu juuri ohjetekstin ehdottaman aiherajauksen perusteella. Tekstien rakenne on siis kuvion 1 kaltainen: niiden diskurssitopiikkina on vaihtokokemus ylipäättään, ja tätä ylemmän tason topiikkia käsitellään jakamalla se alatopiikkeihin, joista tavallisimmat on lueteltu kuviossa.



Kuvio 1: Matkakertomustekstien diskurssitopiikki ja alatopiikit

Tässä käsiteltävä lopullinen tutkimusaineisto kattaa kaikkiaan 492 opiskelijoiden laatimaa tekstiä, jotka tutkimuksen ensimmäisessä vaiheessa annotoitiin kappaleittain<sup>1</sup> siten, että jokaiselle kappaleelle merkittiin topiikki. Yhteensä kaksikymmentä tekstiä käsittäneen kokeiluaineiston perusteella päätettiin keskittyä *asumisesta* kertoviin kappaleisiin siitä syystä, että asumisesta kirjoitettiin lähes jokaisessa analysoidussa näytetekstissä ja koska asuminen aiheena havaittiin tavallisesti melko selvärajaiseksi. Koska tekstit analysoi kaksi eri henkilöä (artikkelin tekijät), näyteaineiston avulla testattiin myös, kuinka samanlaisia tulkinnat kappaleiden aiheista olivat. Asumiskappaleiden osalta arvioijien tulkinnat olivat samoja 45 tapauksessa 47:stä.

Kokeiluaineiston jälkeen koko tutkimusaineisto käytiin läpi siten, että teksteistä etsittiin vain kappaleita, joiden topiikki oli tulkittavissa asumiseksi. Kun tämä kaikki tekstit kattava varsinainen annotointiprosessi saatiin päätökseen, lopulliseksi aineistoksi muodostui 415 vähintään yhden asumisesta kertovan kappaleen sisältävää tekstiä. Tekstit ovat keskimäärin 945 sanaa pitkiä ja sisältävät kaiken kaikkiaan keskimäärin 13 kappaletta. Prosessin tuloksena saatu tutkimusaineisto on siis kohtalaisen kokoinen, ja sitä voidaan mahdollisesti jatkossa käyttää erilaisten kvantitatiivisten päätelmien pohjana. Nyt käsillä olevaa tutkimusta varten päädyttiin kuitenkin hyödyntämään aineistoa ennen muuta kvalitatiivisesti. Tätä varten aineistosta eroteltiin tarkempaa analyysia silmällä pitäen käsiteltäväksi painotettu satunnaisotanta.

Satunnaisotannan painottamisella pyrittiin varmistamaan, että analysoitava otos edustaa mahdollisimman kattavasti koko tutkimusaineistoa (satunnaisotantaaan liittyvistä periaatteista ks. esim. Levshina 2015: 8). Otoksen edustavuuden takaamiseksi kaikki 415 tekstiä kattava korpus rajattiin ja annotoitiin ensin automaattisesti niin pitkälle kuin mahdollista. Prosessi piti sisällään seuraavat vaiheet:

1. Määriteltiin, että analysoitavien kappaleiden topiikin asettavat virkkeet luokitellaan sen perusteella, mikä sana toimii *topiikki-indikaattorina* eli ensimmäisenä leksikaalisena vihjeenä kappaleen topiikista. Analysoitavien kappaleiden määrää rajattiin siten, että mukaan otettiin ainoastaan sellaiset tapaukset, joissa topiikki-indikaattorina oli joskin *asua*-verbin johdos (kuten *asunto*, *asuntoasia* tai *asuminen*) tai *asua*-verbi itse.
2. Huomio kohdistettiin vain niihin 338 tapaukseen, joissa *asua*-johdos sijaitsee ensimmäisen asumisesta kertovan kappaleen ensimmäisessä virkkeessä. Tällä varmistettiin se, että tarkastelun kohteeksi tulee

<sup>1</sup>Käyttäen kirjoittajien itsensä tekemää kappaleiden ortografista erottelua.

- mahdollisimman paljon juuri niitä virkkeitä, joissa kirjoittaja todella siirtyy vanhasta topiikista uuteen.
3. Mukaan valikoituneille kappaleille suoritettiin automaattinen syntaktinen ja morfologinen jäsennys sekä lemmatisointi hyödyntämällä Turun yliopistossa kehitettyä dependenssijäsennintä (Haverinen ym. 2014) ja siihen liittyviä komponentteja.

Jos tutkitaan tarkemmin niitä kategorioita, jotka syntaktinen jäsennin on merkinnyt tutkimukseen valikoituneiden kappaleiden ensimmäisille *asua*-johdoksille, saadaan taulukon 1 mukainen jakauma.

	Frekvenssi	Selitys	Esimerkki	N
dobj	94	objekti	<i>Hankin asunnon</i>	47
root	60	predikaattiverbi	<i>Asuin kerrostalossa</i>	30
nmod	50	luokittelematon adverbiaali	<i>Asunnon suhteen</i>	25
nmod:gobj	34	nominaalisen verbimuodon objekti	<i>Asunnon hankkiminen</i>	17
nsubj	32	subjekti	<i>Asuminen järjestyi</i>	16
nmod:poss	29	omistusrakenteen kohde	<i>Asuntoloiden vuokrataso</i>	14
nsubj:cop	21	kopulalauseen subjekti	<i>Asuntola on uusi</i>	10
muu	18	esim. infinitiivitäydennykset	<i>Halusin asumaan paikallisten kanssa</i>	9
Yht.	338	—	—	168

Taulukko 1: Asua-johdosten jakautuminen eri syntaktisiin kategorioihin.

Taulukon Frekvenssi-sarakkeesta nähdään paitsi se, että yhteensä edellä kuvattujen kolmen annotointi- ja rajausvaiheen jälkeen jäljelle jäi 338 tekstiä, myös se, että selvästi tavallisimpia topiikki-indikaattoreita olivat objektina esiintyvät *asua*-johdokset. Kategoriat, joiden edustajia oli yhteensä alle kymmenen, niputettiin yhdeksi *muu*-kategoriaksi. Varsinainen satunnaisotanta tehtiin taulukossa esitettyjen kategorioiden perusteella siten, että jokaisesta taulukossa luetellusta ryhmästä (mukaan lukien *muu*-ryhmä) valittiin satunnaisesti ryhmän kokonaisfrekvenssin puolikasta vastaava määrä tapauksia (ks. taulukon viimeinen sarake).<sup>2</sup> Nämä 168 tapauksesta otettiin varsinaisen laadullisen analyysin kohteeksi.

Laadullisen analyysin tavoitteena oli ryhmitellä aineistosta erotellut 168 tekstiä siten, että yhteen ryhmään päätyisivät aina ne tavat asettaa topiikki, jotka syntaktisilta tai funktionaalisilta ominaisuuksiltaan muistuttavat toisiaan. Jokaista topiikkia asettavaa virkettä tutkittiin erikseen laajemmassa tekstiyhteydessään, minkä jälkeen sitä verrattiin muihin virkkeisiin ja määriteltiin, oliko kyseessä jokin uusi, muista erottuva tapausjoukko vai ennemminkin jokin jo muodostettuun joukkoon kuuluva tapaus. Tapauksia ei siis ryhmitelty minkään ennalta määriteltujen kriteerien pohjalta, vaan sillä perusteella, mitkä tapaukset muistuttivat toisiaan.

Metodimme muistuttaa jonkin verran Vuorijärven (2013: 116) mainitsemaa korpuslingvistiksessä tekstianalyysissä käytettyä menetelmää (ks. Biber ym. 2007), jossa analyysiin otettavat tekstit käydään ensin yksitellen läpi, minkä jälkeen kukin teksti pyritään pilkkomaan tiettyä viestitätarkoitusta toteuttaviin, erikseen nimettäviin jaksoihin. On kuitenkin olennaista huomata se lähtökohtaero, ettei nyt käsillä olevassa tutkimuksessa niinkään etsitä eri tekstijaksojen viestintätarkoituksia, vaan liikutaan puhtaasti topiikkien tasolla kysymällä, miten jokin uusi aihe esitellään lukijalle. Toinen ero on siinä, että tarkasteltavien tekstijaksojen rajauksessa lähtökohtana pidetään kirjoittajan itsensä tekemää ortografista kappalejakoa – kirjoittajan omaa näkemystä siitä, milloin tekstissä siirrytään sellaiseen uuteen jaksoon, joka ansaitsee tulla erotetuksi edeltävästä tekstistä. Kolmanneksi tämän tutkimuksen metodin erottaa edellä mainituista se, ettei analyysin tuloksena syntyneitä ryhmiä pyritä nimeämään (vrt. Vuorijärvi 2013: 117), vaan ne määritellään ainoastaan suhteessa toisiin teksteihin.

...vielä kappale ryhmittelyn perusteista ja siitä, ettei semant. kriteerejä käytetty...

<sup>2</sup>Satunnaisotannan tekninen toteutus dokumentoitiin git-versionhallintaan ja on nähtävillä osoitteessa x (tarkistettu 8.1.2017)

## 4 Tulokset

Edellä kuvatun laadullisen analyysin tuloksena tutkimusaineiston tekstit jaoteltiin 13 ryhmään, joita osiossa 4.1. luonnehditaan yhden edustavan esimerkin avulla. Kuten metodin kuvauksen yhteydessä todettiin, tässä kuvatut eri topiikin asettamistapoja edustavat ryhmät on muodostettu systemaattisesti sijoittamalla toisaan muistuttavat tapaukset samaan kategoriaan. Ryhmien väliset rajat eivät luonnollisestikaan ole suoraviivaisia, ja jaottelun olisi voinut tehdä muillakin tavoilla. Eri ryhmät ovat lisäksi enemmän tai vähemmän läheisessä yhteydessä toisiinsa, ja näitä linkkejä on tässä esitetyssä kuvauksessa pyritty osoittamaan. Vaikka varsinainen ryhmäjako perustuu laadulliseen samankaltaisuuteen, tulkitaan osiossa 4.2. analyysin tulosta eli itse jaottelua tilastollisin menetelmin.

### 4.1 Laadullinen analyysi

Laadullisen analyysin tuloksena päädyttiin melko suureen määrään eri kategorioita. Jaottelun hienojakoisuudella haluttiin säilyttää intuitio tiettyjen topiikin asettamistapojen muita suuremmasta samankaltaisuudesta. Määrittelemämme 13 kategoriaa voitaisiin luonnollisesti koota myös laajemmiksi ryppäiksi. Yksi ryhmiä vertailemalla esiin nouseva yhdistävä piirre on kirjoittamisen tutkimuksessa (viite  $x, y, z$ ) käytetty jaottelu kirjoittajakeskeisiin, *XXXXkeskeisiin* ja *XXXXkeskeisiin*. Esittelemmekin erottelemamme topiikin asettamistavat seuraavassa jaoteltuna löyhästi näiden kolmen kategorian mukaisiin alalukuihin. Esittely tehdään käyttämällä jokaisesta ryhmästä mallina yhtä kappaleen avaavaa virkettä, jonka laajempi konteksti on nähtävillä liitteessä 1.

#### 4.1.1 Kirjoittajakeskeiset lähestymistavat

Kirjoittajakeskeisiksi on tutkimuskirjallisuudessa kutsuttu... Omassa aineistossamme tällaisiksi topiikin asettamistavoiksi voidaan luokitella karkeasti ottaen  $N$  ryhmää, mikä tarkoittaa, että kyseessä oli selvästi suurin tässä käsiteltävistä kolmesta kategoriasta. Kirjoittajakeskeisistä ryhmistä suurin – ja samalla koko tutkimuksen yleisin – kategoria ovat esimerkin 1 kaltaiset tapaukset, joihin jatkossa viitataan nimellä *listamaiset*.

(1) Asuntoni sain vuokrattua Apartiksen( paikallinen TOAS) kautta.

Informaatorakenteellisesti tarkasteltuna esimerkin 1 *asuntoni* voitaisiin lukea lohkotopiikiksi, joskaan esimerkissä ei täyty Dikin (1989: 275) kuvailema ehto siitä, että varsinainen ylätopiikki olisi erikseen esitelty lukijalle (vrt. myös Guijarro 2001: 110). Asumiseen tai asuntoon ei itse asiassa ole viitattu edeltävissä kappaleissa, vaan topiikkia asettava virke on samalla ensimmäinen maininta itse topiikista. Kirjoittaja operoi tässä ja useimmissa muissa listamaiset-ryhmän tapauksissa niin selkeästi tehtävänannossa ehdotetun rakenteen puitteissa, että hän olettaa lukijan automaattisesti jakavan käsityksen siitä, mitä aiheita tekstin diskurssitopiikkiin, vaihtokokemukseen, liittyvien asioiden joukko pitää sisällään. Voisi jopa sanoa, että esimerkin 1 kaltaiset tapaukset eivät niinkään aseta topiikkia, vaan ennemmin käyttävät jo asetetuksi oletettua topiikkia.

Esimerkin 2 kaltaisissa tapauksissa on kyse lohkotopiikin sijasta esittelytopiikista. Tavallisimmin esittelyn kohteena on kirjoittajan asunto, joka tuodaan tekstiin imperfektimuotoisen *asua*-verbin avulla, minkä takia ryhmälle annettiin nimi *asuin x:ssä*:

(2) Vaihtoni aikana asuin Grandmontin asuntolassa.

*Asuin paikassa x* -tyyppiset rakenteet ovat hyvin suoraviivaisia keinoja esitellä paikka  $x$ , josta tällaisella virkkeellä alkava kappale kertoo. Esimerkiksi tässä esitellyn virkkeen tapauksessa kirjoittaja jatkaa kappalettaan kuvailemalla asuntonsa sisustuksen, kämppäkaverit ja vuokran suuruuden.

Kolmannen kirjoittajakeskeisen ryhmän virkkeet ovat ajallisesti muita monitasoisempia. Nämä tapaukset on ryhmitelty yhteen, koska niissä kaikissa pohjustetaan asumisesta kertomista jollakin asumista edeltävään

aikaan liittyvällä tapahtumalla. Tällä perusteella nimitämme tämän ryhmän tapauksia *ennakoiviksi*. Ryhmää edustaa esimerkki 3:

(3) Olin löytänyt asunnon ensimmäiseksi kuukaudeksi Airbnb:n kautta ennen Italiaan saapumista.

Esimerkissä 3 asumistopiikin pohjustus tapahtuu luomalla pieni jännite: jos kerran ensimmäiseksi kuukaudeksi oli löydetty Airbnb-asunto, miten asumisasia lopullisesti ratkaistiin? Avausvirkkeen jälkeen kappaleessa selviää, että Airbnb-ratkaisusta muodostuikin lopulta asunto koko vaihdon ajaksi, ja kappaleen loppuosassa kirjoittaja siirtyy kuvailemaan tarkemmin asunnon yksityiskohtia. Informaatorakenteeltaan ryhmä 3 onkin hyvin samanlainen kuin edellinen ryhmä: koko kappaleen topiikkina oleva *asunto* tuodaan tekstiin esittelytopiikkina ensimmäisessä virkkeessä.

Esimerkin 4 edustamassa *narratiiviset*-ryhmässä asumisesta raportoidaan hieman ennakoivat-ryhmää muistuttavasti kertovan tekstityypin avulla:

(4) Lensin Wieniin sunnuntaina ja lento-kentältä suuntasin suoraan asunnolleni.

Näissäkin tapauksissa topiikki-indikaattori on informaatorakenteen kannalta esittelytopiikki, joka esimerkin narratiivisuuden vuoksi tulee asetetuksi muita tässä tarkasteltuja uuden topiikin sisältäviä ryhmiä implisiit-  
tisemmin.

Osa satunnaisotantaan päätyneistä lauseista muistutti toisaalta ennakoivia, toisaalta narratiivisia tapauksia liittymättä kuitenkaan erityisen tiiviisti kumpaankaan ryhmään. Näitä *ennakoivat/narratiiviset*-tapauksia edustaa muun muassa seuraava virke:

(5) Saapuessani Varsovaan minulla oli onneksi jo asunto tiedossa.

Myös esimerkissä 5 indikaattorisanan (*asunto*) rooli informaatorakenteen kannalta on esittelytopiikki. Esimerkin 4 tavoin myös esimerkissä 5 kuvataan ajallinen jännite, mutta jännitteen ja sen ratkaisemisen suunta on eri: Siinä missä esimerkki 3 esitti kappaleen alussa avoimen tilanteen ja myöhemmässä kappaleessa tilanteen ratkaisun, aloittaa esimerkki 5 ratkaisusta (saavuttaessa asunto oli jo löydetty) ja käyttää lopun kappaleen takaumamaisesti sen kuvailuun, miten ratkaisuun päästiin (kirjoittaja oli etsinyt asuntoa valmiiksi Facebookin kautta).

Esimerkki 6 kuvaa *oma kokemus*-ryhmäksi nimeämäämme tapausten joukkoa. Näissä tapauksissa kirjoittaja käsittelee asumista kuvaamalla sitä omakohtaisena prosessina, jonka onnistumista tai epäonnistumista voidaan arvioida.

(6) Asuminen, tai siis lähinnä asunnon hankkiminen, oli varmaankin vaihdon hankalin juttu.

Esimerkissä 6 topiikkia indikoi deverbaalisubstantiivi (tässä *asuminen*), mikä ryhmälle onkin tyypillistä: kenties suoraviivaisin tapa kuvailla omaa kokemusta jostakin varsinaisen diskurssitopiikin osa-alueesta on niputtaa kokemus nominalisoituun verbimuotoon ja luonnehtia tätä predikaatiivilauseella. Näissä tapauksissa kuvataan, miten *asunnon hakeminen oli helppoa* tai miten *kaikkein vaikeinta oli asunnon löytäminen*. Informaatorakenteen kannalta *oma kokemus*-ryhmän tapaukset ovat monesti listamaiset-ryhmän tapaan lohkotopiikkeja. Esimerkki 6 on kuitenkin esimerkkiä 1 selvemmin topiikkia asettava virke, sillä siinä asumisen rooli yhtenä vaihtokokemuksen osana on ilmaistu eksplisiittisesti.

Oma kokemus -ryhmää lähellä ovat esimerkin 7 kaltaiset tapaukset, joissa niissäkin usein reflektoidaan omien valintojen onnistumista. Esimerkin 7 edustamassa *kontrastia ilmaisevassa* ryhmässä omaa kokemusta kuitenkin verrataan eksplisiittisesti toisten kokemuksiin.

(7) Itselläni kävi tuuri asunnon suhteen löytäessäni edullisen huoneen Chelseasta aivan Manhattanin keskeltä.

Informaatorakenteeltaan ryhmän 7 tapaukset ovat nekin tyypillisesti lohkotopiikkeja, ja olettavat indikaattorisanan implikoiman topiikin jossain määrin esimerkin 1 kaltaisesti jo asetetuksi.

#### 4.1.2 Depersonalisoidut (??) lähestymistavat

Depersonalisoiduilla lähestymistavoilla viitataan... Kirjallisuudessa...

Laadullisen analyysin tuloksena laadituista ryhmistä kenties selkeimmin kirjoittajan omasta kokemuksesta irtautuvat esimerkin 8 kaltaiset tapaukset, joita nimitämme *yleistäviksi*.

(8) Asuminen Tokiossa on aika kallista.

Samoin kuin esimerkissä 6, myös esimerkissä 8 topiikki-indikaattorina on koko lauseen ensimmäisenä sijaitseva *asuminen*-sana. Erona näiden esimerkkien välillä on, että siinä missä esimerkki 6 on informaatorakenteen kannalta omasta vaihtokokemuksesta lohkaistu lohkotopiikki, ankkuroituu esimerkki 8 siihen fyysiseen paikkaan, jossa vaihto tapahtui. Ryhmän 8 kirjoittajat eivät siis niinkään lähde liikkeelle omista kokemuksistaan, vaan käsittelevät tilannetta laajemmin, tavoitteenaan eräässä mielessä jakaa maailmantietoa lukijalleen.

Yleistys kirjoittajan oman kokemuspäiriin ulkopuolelle voidaan kuitenkin tehdä myös toisella tavalla. Esimerkissä 9 tämä tapahtuu keskittymällä johonkin ulkoiseen agenttiin, jonka toimet vaikuttavat kirjoittajaan yleisesti ja asumisen järjestymiseen erityisesti.

(9) Suurlähetystön henkilökunta tarjosi resurssiensa puitteissa apuaan mm. maahantuloon, käytännön järjestelyihin sekä asunnon löytämiseen liittyvissä asioissa.

Esimerkissä 9 on informaatorakenteen kannalta katsottuna kyse esimerkkien 2, 3 ja 11 tapaan esittelytopiikista, mutta tässä esittely toteutetaan hyödyntämällä jotakin tekstin kannalta keskeistä ulkoista toimijaa (kuten suurlähetystön henkilökuntaa), minkä takia tämä tapausjoukko nimettiin *ulkoinen toimija* -ryhmäksi. Näitä lähellä ovat esimerkin 10 kaltaiset *suurin osa*-ryhmäksi nimetyt tapaukset, joissa asumistopiikki esitellään kertomalla jotakin siitä, miten kirjoittajan kuvaamassa tilanteessa yleensä toimitaan tai miten suurin osa toimii:

(10) Lähes kaikki Bratislavan kauppakorkeakoulun vaihto-opiskelijat asuivat Ekonom-nimisessä dormitoryssa.

Suurin osa -ryhmän virkkeet ovat lähellä kontrastiryhmää, johon luetuissa tapauksissa oleellista on kirjoittajan itsensä vertailu muihin. Esimerkin 10 edustamille tapauksissa kirjoittaja ei kuitenkaan yleensä kuvaa omaa toimintaansa mitenkään yleiselle toimintatavalle vastakkaisena, vaan useimmiten noudattaa myös itse kuvailemaansa enemmistön mallia. Virke 10 on todella vain keino johdattaa lukija kappaleen varsinaiseen aiheeseen, *Ekonom*-nimiseen asuntolaan (esittelytopiikki), jonka sijainnista, vuokratasosta ja muista ominaisuuksista kappale kertoo.

Depersonaloituuihin tapauksiin laskettiin myös seuraavan esimerkin kaltaiset *staattinen kuvailu* -ryhmäksi luokitellut tapaukset:

(11) Ulkomaisille opiskelijoille tarkoitettu asuntola on vain kolme vuotta vanha ja täten melko moderni.

Esimerkissä 11 kirjoittaja esittelee uuden topiikin lähtemällä liikkeelle konkreettisesti asuntonsa luonnehtimisesta (vrt. esimerkki 6, jossa luonnehditaan kirjoittajan kokemusta). Informaatorakenteeltaan nämä tapaukset ovat monesti lähellä ryhmää 2, sillä molemmissa kappaleen ensimmäinen virke esittelee lukijalle sen, minkä yksityiskohtaiseen kuvailuun koko loppukappale keskittyy. Esimerkissä 11 topiikki asetetaan kuitenkin jossain määrin implisiittisemmin, ja hyppäys edellisestä aiheesta (esimerkin 11 tapauksessa HIV-testin teettäminen) on tyypillisesti rajumpi. Tämä selittää sitä, että monet ryhmän 11 tapauksista (ei kuitenkaan tässä esitetty esimerkki) sisältävät asumiskappaletta edeltävän väliotsikon, kuten *Asuminen* tai *Asumisjärjestelyt*.

#### 4.1.3 Metatason (??) lähestymistavat

Metatason lähestymistavat olivat tutkimusaineistossa harvassa, vaikka oletus on, että.. Tähän on otettu kaksi ryhmää.



Ensimmäinen metatason lähestymistavoista nimettiin *intertekstuaaliset*-ryhmäksi sillä perusteella, että näissä tapauksissa topiikki esiteltiin käyttämällä hyväksi jotakin jo aiemmin mainittua viittauskohdetta kuten esimerkissä 12.

(12) Asuntoloista valitsin Julius Raab Heimin vanhojen matkakertomusten perusteella.

Esimerkissä 12 indikaattorina on *asuntoloista*-sana. Ilman kontekstia tarkasteltuna tämäkin tapaus vaikuttaa samalta kuin listamainen esimerkki, mutta laajemman analyysin tuloksena selvisi, että kirjoittaja on jo etukäteisjärjestelyistä ja kohdemaahan saapumisesta raportoidessaan rakentanut kuvaa paikallisista asuntolavaihtoehtoista ja asuntolassa elämisestä. Esimerkin 12 *asuntolat* on siis jo mainittu ja kyseessä on tässä kohtaa informaatorakenteen kannalta esittely- tai lohkotopiikin sijasta esitelty topiikki.

Oman, selvästi muista erottuvan ryhmänsä muodostavat ne esimerkin 13 edustamat tapaukset, joissa kirjoittaja eksplisiittisesti ilmaisee siirtyvänsä käsittelemään asumista.

(13) Sitten pari sanaa asumisesta.

Tähän *eksplisiittiset*-ryhmään kuuluvien esimerkin 13 kaltaisten johdantovirkkeiden voisi ajatella jossain määrin korvaavan väliotsikoita tekstissä, vaikka yksi tähän ryhmään luokiteltu virke (*Monella varmaan herää jo kysymys, entäs asuminen?*) tekeekin poikkeuksen siinä, että teksti sisältää sekä johdantovirkkeen että väliotsikon. Nämä esimerkit edustavat luonnollisesti esittelytopiikkeja ja ovat aiheeseen johdattamisen implisiittisyys-eksplisiittisyys-asteikon eksplisiittisessä ääripäässä (vrt. päinvastaisessa ääripäässä sijaitseva ryhmä 1).

#### 4.1.4 Kokoavasti

Edellä käsitelty 13 ryhmää on tiivistetty taulukkoon X:

Ryhmän nimi	Tapauksia yhteensä	Topiikin laji	Näkökulma
Listamaiset	31	lohkotopiikki	kirjoittajakeskeinen
Asuin x:ssä	21	esittelytopiikki	kirjoittajakeskeinen
Ennakoivat	21	esittelytopiikki	kirjoittajakeskeinen
oma kokemus	16	esittelytopiikki	kirjoittajakeskeinen
Intertekstuaaliset	14	esitelty topiikki	metataso
asunnon ja asumisen staattinen kuvailu	12	esittelytopiikki	depersonalisoitu
Kontrasti	10	lohkotopiikki	kirjoittajakeskeinen
Yleistys	10	esittelytopiikki	depersonalisoitu
Ulkoisen toimija	9	esittelytopiikki	depersonalisoitu
Narratiiviset	9	esittelytopiikki	kirjoittajakeskeinen
Eksplisiittiset	6	esittelytopiikki	metataso
Ennakoivat/narratiiviset	5	esittelytopiikki	kirjoittajakeskeinen
Suurin osa	4	esittelytopiikki	depersonalisoitu

Kuten edellä metodin kuvauksen yhteydessä todettiin, tässä kuvatut eri topiikin asettamistapoja edustavat ryhmät on muodostettu sijoittamalla aina toisiaan muistuttavat tapaukset samaan kategoriaan. Ryhmien väliset rajat eivät luonnollisestikaan ole suoraviivaisia, ja jaottelun olisi voinut tehdä muillakin tavoilla. Eri ryhmät ovat lisäksi enemmän tai vähemmän läheisessä yhteydessä toisiinsa, ja näitä linkkejä on tässä esitetyssä kuvauksessa pyritty osoittamaan. Seuraavaksi nyt esiteltujen ryhmien rakenteellisia tyyppipiirteitä tutkitaan tarkemmin määrällisestä näkökulmasta. Argumenttimme on, että vaikka ryhmät on muodostettu puhtaasti tulkitsemalla eri tekstien välisiä samankaltaisuuksia, on suurimmalle osalle ryhmistä melko yksinkertaista antaa myös muodolliset, lähinnä rakenteeseen perustuvat määritelmät, joiden soveltaminen voi joillekin kirjoittajille tai opettajille olla abstrakteja funktionaalisten ominaisuuksien kuvailuja helpompaa.

## 4.2 Tilastollinen tarkastelu

Edellä (taulukko ) esitimme tavan ryhmitellä asumisesta kertovat tekstikappaleet sen mukaan, miten kirjoittaja toteuttaa siirtymän uuteen aiheeseen. Seuraavassa sovellamme laadullisen analyysin tuloksena saatuun ryhmäjakoon tilastollisia menetelmiä, tarkemmin ottaen ehdollisia päätöksentekopuita (conditional inference trees).

Ehdolliset päätöksentekopuut ovat regressiotyyppinen luokittelumenetelmä, jossa yhden tai useamman selittävän muuttujan assosiaatiota selitettävään muuttujaan (termeistä ks. Ketokivi 2015: 131) tarkastellaan sarjana tutkimusaineistoa jaottelevia päätöksiä (Levshina 2015: 291). Meidän tutkimukseemme sovellettuna tämä tarkoittaa sen selvittämistä, mitkä aineiston 168 tapauksen ominaisuudet (selittävät muuttujat) parhaiten ennustavat sen, mihin kolmestatoista ryhmästä (selitettävä muuttuja) tapaus on luokiteltu. Käytössämme ovat taulukon 2 mukaiset selittävät muuttujat.

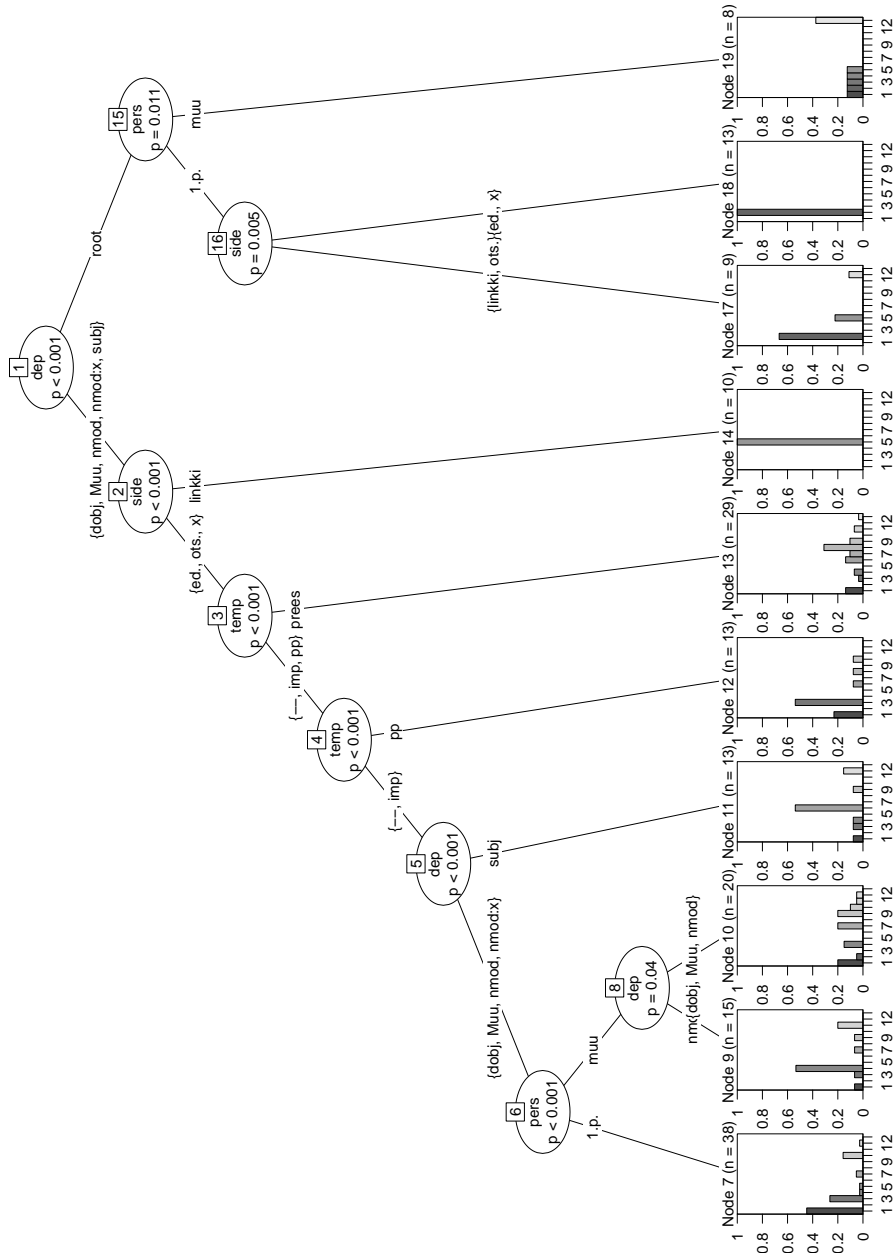
Muuttujan nimi ja lyhenne	Muuttujan arvot
Koneellisen jäsentimen topiikki-indikaattorille asettama syntaktinen rooli (dep)	Arvot esitetty taulukossa 1, kuitenkin lavennettuna siten, että arvot <i>nsubj</i> ja <i>nsubj:cop</i> yhdistettiin yhdeksi subjektin kategoriaksi ( <i>subj</i> ) ja arvot <i>nmod:poss</i> sekä <i>nmod:gobj</i> yhdeksi <i>nmod:x</i> -kategoriaksi, jolla viitattiin kaikkiin kompleksisiin nominaalirakenteisiin ( <i>asunnon hankinta</i> , <i>asumisen järjestäminen</i> ).
Topiikki-indikaattorin sijainti virkkeessä (sij)	Kaksi arvoa a) Tapaukset, joissa topiikki-indikaattori on koko virkkeen ensimmäisenä. b) Muut tapaukset.
lauseen predikaattiverbin persoona (pers)	Kaksi arvoa: a) Tapaukset, joissa predikaattiverbi on ensimmäisessä persoonassa ja b) Muut tapaukset.
Side muuhun tekstiin (side)	Neljä arvoa: a) Tapaukset, joissa topiikki asetetaan viittaamalla selkeästi johonkin edellä muodostettuun linkkiin ( <i>linkki</i> ) b) Tapaukset, joissa asumistopiikkia on pohjustettu edellisessä kappaleessa, muttei niin selvästi, että kyseessä olisi linkkitapaus ( <i>edel.</i> ) c) Tapaukset, joissa asumiskappaletta edeltää väliotsikko ( <i>ots</i> ) d) Tapaukset, joissa mitään edellä kuvatuista siteistä ei ole ( <i>x</i> ).
Topiikki-indikaattorin morfologinen rakenne (morf)	Kolme arvoa: a) finiittiverbit b) infiniivimuodot (mukaan lukien minen-johdokset) c) substantiivit.
Aikamuoto (tense)	Neljä arvoa: a) preesens, imperfekti, pluskvamperfekti/perferkti <sup>3</sup> , ellipsi (–).

Taulukko 2: Tilastollisen analyysin selittävät muuttujat

Tässä käytetty päätöksentekopuu rakennettiin R-ohjelmiston *party*-kirjastoon kuuluvaa *ctree*-funktia (Hothorn ym. 2006) käyttäen.<sup>4</sup> Analyysin tulos on esitetty kuviossa 3.

<sup>3</sup>Perfektitapauksia aineistossa oli vain yksi ja sekin konditionaalissa, joten molemmat suomen liittoaikamuodot yhdistettiin samaan kategoriaan.

<sup>4</sup>viittaus githubiin



Kuvio 3: Topiikkia asettavien ryhmien rakenne luokittelupuun avulla tarkasteltuna

Kuviossa 3 tutkimusaineisto on jaettu kymmeneen ryhmään (kuvion alimmalla rivillä esitetyt pylväskaaviot) edellä kuvattujen selittävien muuttujien perusteella. Kuvio esittää selittävien muuttujien vaikutuksen sarjana päätöksiä (kuvion ympyrämuotoiset noodit), jotka on numeroitu yhdestä 19:ään, esimerkiksi: *jos jonkin tapauksen dependenssirooli on predikaatti (noodi 1) ja pääverbin persoona ensimmäinen (noodi 15) ja jos lisäksi tapausta ei edellä väliotsikko, kyseessä on aina ryhmään 2 kuuluva tapaus (noodi 18)*. Jos taas dependenssirooli on predikaatti (noodi 1), joka ei kuitenkaan esiinny ensimmäisessä persoonassa, on kyseessä useimmiten ryhmän 13 tapaus, vaikka muutkin vaihtoehdot ovat mahdollisia. Jokaisen kuvion haarauman kohdalle on myös merkitty p-arvo, joiden avulla voidaan arvioida kunkin haarauman merkitsevyyttä (Levshina 2015: 292). Korkein p-arvo kuviossa 3 on noodilla 15, joka sekin on vain noin yhden prosentin luokkaa, mikä tarkoittaa että kuviossa esitetyt jaot ovat kaiken kaikkiaan merkitseviä (Baayen 2012: 69; p-arvoista ks. esim. Levshina 2015: 10).

Oleellista tässä esitetyssä päätöksentekopuussa on, että vaikka itse luokittelu ryhmiin tehtiin puhtaasti artikkelin kirjoittajien tulkinnan perusteella ja kustakin ryhmästä annettiin edellä funktionaalinen kuvaus, on ryhmiä mahdollista kuvailla melko pitkälle myös rakenteellisten ominaisuuksien valossa. Karkeasti ottaen voidaan todeta, että suurin osa laadullisen analyysin tuloksena laadituista kolmestatoista ryhmästä pystytään päättämään hyödyntämällä kolmea<sup>5</sup> yksinkertaista rakenteellisen mikrotason muuttujaa: verbin persoonaa, aikamuotoa ja indikaattorisanan dependenssiroolia. Taulukossa 3 on tiivistetty näiden muuttujien vaikutus topiikkia asettavan virkkeen päätymiseen johonkin laadullisessa analyysissä määritellyistä ryhmistä.

Verbin persoona	Verbin aikamuoto	Indikaattorisanan dependenssirooli	ryhmä	luokittelupuun loppunoodit
1.	<i>imperfekti</i>	<i>ei predikaatti</i>	1	7
1.	–	<i>predikaatti</i>	2	17 ja 18
–	<i>pl.perf/perf.</i>	–	3	12
<i>muu kuin 1.</i>	<i>imperfekti</i>	<i>osa kompleksista lauseketta</i>	4	9
–	<i>imperfekti</i>	<i>subjekti</i>	6	11
–	<i>preesens</i>	<i>ei predikaatti</i>	8	13
<i>muu kuin 1.</i>	–	<i>predikaatti</i>	13	19
<i>muu kuin 1.</i>	<i>imperfekti</i>	<i>ei predikaatti, ei subjekti, ei osana kompleksista lauseketta</i>	7 ja 9	10

Taulukko 3: Kvantitatiivisesti pääteltävissä olevat ryhmät

Taulukon mukaan ryhmä 1 voidaan rakenteiden tasolla määritellä siten, että siinä on nominaalinen topiikki-indikaattori, jonka pääverbi esiintyy ensimmäisessä persoonassa ja imperfektissä. Ryhmän 4 puolestaan erottaa muista se, että asumisaiheeseen viitataan siinä jollakin kompleksisella lausekkeella, kuten *asunnon hankinta* tai *asumisen järjestäminen*. Ryhmä 3 taas voidaan määritellä ennen kaikkea aikamuodon pohjalta: siihen kuuluvat tapaukset, joissa verbi on pluskvamperfektissä. Ainoastaan ryhmien 10, 11 ja 12 osalta ei tässä esitetyn päätöksentekopuun perusteella voida esittää selkeitä piirteitä, jotka johtaisivat tapauksen sijoittamiseen juuri näihin ryhmiin. Lisäksi on huomattava, että esimerkin sijoittaminen ryhmään 5 on perustunut täysin artikkelin kirjoittajien tulkintaan: se, että tulevaa topiikkia on eksplisiittisesti pohjustettu aiemmassa tekstissä, ei ole pääteltävissä siitä, minkälaisia rakenteita topiikkia asettavassa virkkeessä on käytetty.

Päätelmäluvun puolelle se, miä tästä kaikesta seuraa.

## 5 Päätelmät

Tarkastelimme tässä tutkimuksessa tekstikappaleen topiikin muodostumista korpusvetoisesti, kahdella menetelmällä. Tavoitteenamme oli selvittää, millaisia tapoja yliopisto-opiskelijat käyttävät matkaraportteissaan saman topiikin asettamiseen ja miten näitä tapoja voidaan määritellä toisaalta funktionaalisesti, toisaalta syntaktisesti.

Tutkimuksen aineistona oli yliopisto-opiskelijoiden kirjoittamista matkakertomuksista koostettu korpus, ja osa artikkelin löydöksistä liittyikin eittämättä myös tekstilajille ominaisiin ilmaisutapoihin, vaikka tutkimuksen keskiössä olivat virkkeet ja niiden väliset suhteet. Käyttämämme aineisto ja menetelmät eivät tarjoa vastauksia siihen, miksi kirjoittajat ovat päätyneet ehkä hieman luettelomaisestikin kirjoittamaan ohjeessa annetuista aiheista sen sijaan, että olisivat kirjoittaneet hierarkkisesti rakentuneen tekstin.

Laadullisen analyysin tuloksena erotimme 13 erilaista tapaa toteuttaa kappaleenalkuinen siirtymä topiikista toiseen. Pääpiirteissään näiden 13 ryhmän eroavaisuudet palautuvat 1) siihen, toteuttaako kappale kertovaa,

<sup>5</sup>Päätöksentekopuun tuottaneessa algoritmossa indikaattorin sijainti osoittautui hieman yllättäenkin merkityksellömäksi, samoin kuin morfologinen rakenne.

kuvailevaa vai erittelevää tekstityyppiä (vrt. Juvonen 2014b: 81–82) 2) siihen, millaisessa suhteessa aloitusvirkkeessä esitetty topiikki on aiempaan tekstiin ja 3) siihen, millaisesta näkökulmasta käsin kirjoittaja lähestyy tekstinsä aihetta. Tilastollisessa rakenteiden tason analyysissä suurin osa laadullisen analyysin tuloksena laadituista ryhmistä pystyttiin lopulta päättämään hyödyntämällä kolmea yksinkertaista muuttujaa: predikaatin persoonaa ja aikamuotoa sekä indikaattorisanan dependenssiroolia.

Esimerkiksi tekstilajeja tai argumentointia käsitellessään opettajan tai oppimateriaalin tekijän on usein turvauduttava melko abstrakteihin komponentteihin ja määritelmiin. Yhtenä tämän tutkimuksen tuloksena ovatkin funktionaalisten kuvausten lisäksi tuotetut rakenteiden tason määritelmät, jotka tarjoavat vaihtoehdoisen lähestymistavan kappaleiden rakentamisen opettamiseen ja omaksumiseen. Tämä lähestymistapa voi olla erityisen hedelmällinen suomea toisena tai vieraana kielenä kirjoittaville: konkreettiset kielenaineksiin viittaavat ohjeet kuten “rakenna lause yksikön kolmannessa persoonassa” tai “muodosta käsiteltävästä asiasta minen-substantiivi ja kuvaile sitä adjektiivilla” voivat monissa tilanteissa olla helpommin lähestyttäviä kuin abstraktit kuvaukset siitä, minkälaisia kappaleiden tulisi olla. Tässä tuotetut ryhmittelyt voisivatkin toimia lähtölaukauksena myös erilaisia todellisista teksteistä poimittuja mallirakenteita esittelevälle fraasipankille (vrt. Chitez ym. 2015). Juuri tällaiset materiaalit istuvat hyvin ajatukseen uudesta kirjoittajasta, joka ei niinkään kysy, miten asiat pitäisi tehdä, vaan etsii malleja siitä, millä eri tavoilla on jo tehty.

Baayen, R. H. 2012. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: Cambridge university press.

Biber, D., U. Connor & T. A. Upton 2007. *Discourse on the move: Using corpus analysis to describe discourse structure* Vsk. 28. John Benjamins Publishing.

Chafe, W. 1987. Cognitive constraints on information flow. Teoksessa R. S. Tomlin (toim.) *Coherence and Grounding in Discourse : Outcome of a Symposium, Eugene, Oregon, June 1984*. Typological Studies in Language 11. Amsterdam: John Benjamins, 21–25.

Chitez, M., C. Rapp & O. Kruse 2015. Corpus-supported academic writing: how can technology help? Teoksessa F. Helm L. Bradley & S. Thoučny. (toim.) *Critical CALL-Proceedings of the 2015 EUROCALL Conference, Padova, Italy*. Dublin, Ireland: Research-publishing.net.

Dik, S. C. 1989. *The Theory of Functional Grammar, Part I: The Structure of the Clause*. Functional grammar series 9. Dordrecht: Foris Publications.

Givón, T. 1983. *Topic continuity in discourse*. John Benjamins Publishing Company.

Guijarro, A. J. M. 2001. Topicality chains in two discourse genres. *Estudios Ingleses de la Universidad Complutense* (9), 103–128. <http://revistas.ucm.es/index.php/EIUC/article/view/EIUC0101110103A>.

Haverinen, K., J. Nyblom, T. Viljanen, V. Laippala, S. Kohonen, A. Missilä, S. Ojala, T. Salakoski & F. Ginter 2014. Building the essential resources for Finnish: the Turku Dependency Treebank. *Language Resources and Evaluation* 48 (3), 1–39.

Hiippala, T. 2013. *Modelling the structure of a multimodal artefact*. Helsinki: Helsingin yliopisto. URN: ISBN:978-952-10-9427-9.

Hothorn, T., K. Hornik & A. Zeileis 2006. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 15 (3), 651–674.

Jisa, H. & L. Tolchinsky 2009. Developing a personalized stance through linguistic means in typologically different languages. Written expository discourse. *Written Language & Literacy* 12 (1), 1–25.

Juvonen, R. 2014a. *Kirjoitelma ja tekijän ääni: kehystämisen yhdyslauseet suomenkielisen ylioppilasaineen dialogisuuden hallinnassa*. Helsinki: Helsingin yliopisto. URN:ISBN:978-951-51-0264-5.

Juvonen, R. 2014b. Näkökulma kirjoitelman dialogisuuteen: Kuinka- ja miten-yhdyslauseet ylioppilasaineessa. *Virittäjä* 118 (1), 72.

Juvonen, R., M. Virtanen & E. Voutilainen 2012. Fennistisen tekstilajitutkimuksen suuntia. *Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiilikä & Mikko Lounela (toim.), Genreanalyysi. Tekstilajitutkimuksen*

*käsikirja* 453–469.

Ketokivi, M. 2015. *Tilastollinen päättely ja tieteellinen argumentointi*. Helsinki: Gaudeamus.

Komppa, J. 2012. *Retorisen rakenteen teoria suomi toisena kielenä-ylioppilaskokeen kirjoitelman kokonaisrakenteen ja kappalejaon tarkastelussa*. Helsinki: Helsingin yliopisto. URN:ISBN:978-952-10-8164-4.

Lambrecht, K. 1996. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge studies in linguistics 71. Cambridge: Cambridge university press.

Levshina, N. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Paldanius, H. 2017. Historian esseevastauksen funktionaalinen rakenne. Teoksessa S. Latomaa, E. Luukka, & N. Lilja (toim.) *AFinLAn vuosikirja 2017*. Suomen soveltavan kielitieteen yhdistyksen julkaisuja 75. Jyväskylä: Suomen soveltavan kielitieteen yhdistys.

Shore, S. 2008. Lauseiden tekstuaalisesta jäsennyksestä. *Virittäjä* 112 (1), 24–65.

Vuorijärvi, A. 2013. *Tekstilaji ja yhteisö: Ammattikorkeakoulun opinnäytetyön diskussio tekstinä*. Helsinki: Helsingin yliopisto. URN:ISBN:978-952-10-8928-2.