

# Sata ja yksi tapaa ilmaista topiikkia

*Kirsi Sandberg & Juho Härme*

## Contents

<b>1 Johdanto</b>	<b>1</b>
<b>2 Teoriaa (otsikko auki)</b>	<b>2</b>
<b>3 Tutkimusaineisto ja -metodi</b>	<b>4</b>
<b>4 Yleiskuva</b>	<b>5</b>
<b>5 Zoomaus</b>	<b>7</b>
<b>6 Päätelmiä</b>	<b>8</b>
<b>Lähteet</b>	<b>8</b>

## 1 Johdanto

Pari sanaa asumisesta - Eli miten yliopisto-opiskelijoiden kirjoittamissa vaihtoraporteissa asetetaan asuminen tekstikappaleen topiikiksi?!!

Opiskelijavaihdosta kirjoitettava vaihtoraportti on institutionaalinen teksti, jonka kirjoittamiseen yliopisto antaa lähinnä sisältöä koskevan, viitteellisen ohjeistuksen. Näillä verkossa julkaistavilla raporteilla on selkeä funktio: tukea vaihtojaksoa suunnittelevia opiskelijoita päätöksenteossa ja omaan tulevaan vaihtojaksoonsa valmistautumisessa. Vaikka kirjoitusohjeita ja suomenkielisiä mallitekstejäkin on tämän tekstin tapaan löydettävissä verkosta nykyisin runsaasti, opiskeluun ja työelämään liittyvien tekstilajien hallinta tuottaa hankaluuksia sekä peruskoulun(ja lukion) suomi äidinkielenä -oppimäärän suorittaneille että opintonsa S2-oppimäärän mukaisesti käyneille. Oppilaitoksissa ja korkeakouluissa opiskelee siis kasvava joukko kirjoittajia, jolla ei ole hallinnassaan opinnoissa menestymiseen tarvittavia tekstitaitoja.

Matkaraportti, kuten tekstilajit yleensäkin on dynaaminen ja voi toteutua kielellisesti äärettömän monella eri tavalla (onhan riittävän yleistä tietoa, ei kai kaipaa viiterimpsua, dynaamisuudesta puhuu esim. Vuorijärvi 2003, 13), ja tekstiä, jossa kieltä käytetään varioiden ja monipuolisesti mutta lajinsa konventioita riittävässä määrin noudatellen arvostetaan ja pidetään lukijan näkökulmasta luettavana ja mielenkiintoisena. Vuorijärven mukaan (2013, 15) uusissa institutionaalisissa tilanteissa ja tekstilajikonventioiden haltuunotossa ” tarvitaan kielenkäytön sosiokulttuuristen käytänteiden tuntemusta, tekstilajin tajua (viittaus Kalliokoskeen, lisää tähänkin jos jää tekstiin)” riippumatta siitä, käytetäänkö kieltä L1- vai L2-kontekstissa. Avoimeksi kuitenkin jää, mistä kirjoittaja saa repertuaariinsa sen rakenteiden, ilmaisujen ja lekseemien valikoiman, josta varioidaan tekstilajin mukaista tekstiä - ja millä perustellaa valinnat tehdään. Kysymys on luonnollisesti akuutimpi L2-kontekstissa.

Fennistiikassa on tutkittu kielen rakenteiden ja tekstin rajapintaa, mutta useimmiten (esim. Juvosen vk, sivunumerot tähän & siinä mainitut lähteet) kuvaamalla joillain kriteereillä valittujen rakenteiden erilaisia funktioita tekstissä. Sen sijaan koulukirjoittamista ei ole mainittavasti tutkittu (Komppa 2012: 9–10), ja tutkimusperustaisen kirjoittamisenopetuksen tueksi tarvitaan nähdäksemme lisää käyttöpohjaista, kuvailevaa tutkimusta. Lisää tähän lähteeksi myös katsaus lukio-ikäisten kirjoittamisen tutkimukseen & Juvosen vk. Tekstin kirjoittamisen osalta, tekstikappaleiden tasolla, saatavilla on siis toistaiseksi enemmän tietoa siitä, miten se pitäisi tehdä, kuin todellisesta kielenkäytöstä (Komppa 2012: 121) (Chafe 1994: 296). Lisää vielä tähän side EU-kartoitukseen ja tutkimustarpeeseen sovellusten kehittämistä varten.

Akateemisen kirjoittamisen oppimisen ja opettamisen yhteydessä tekstin ja kielen rakenteiden & leksikon rajapintaa on käsitteellistetty eri tavoin riippuen tutkimuksen viitekehystä. Diskurssikompetenssi (Bhatia, Bruce), kirjoittamisen kompetensseista laajemmin suomeksi Makkonen-Craig. Teoreettisissa tarkasteluissa myös usein todetaan, että ”tekstilajitietoisuuden ja tekstilajituntemuksen lisääntyessä yksilön ilmaisurepertuaari ja kielellinen valinnanvara laajenevat kattamaan myös yhteisön puhetavat (esim. Bruce 2008 5-6, Solin 2006, 82 samoin Luukka 2002b, 14, 27)”. Vuosikymmenen loppua lähetyttäessä kirjoittamisen tutkimus kuitenkin päätyy useasti ratkaisuun, että vasta rakenteiden erilaisten käyttötapojen konkreettinen hallinta oman tuottamisen tasolla vie myös tekstin kohti tekstilajin mukaista tavoitetta. Näin ajatellen ei olisi kyse dikotomiasta vaan kimpusta tms., jota voidaan eritellä ja tarkastella eri suunnista. (tajunnanvirtaa iltapäivästä)

Nykyisin kielten ja kirjoittamisen opetuksessa ja asetetuissa tavoitteissa korostuu tekstilähtöisyys ja funktionaalinen tapa tarkastella kieltä (ks esim. opsit & parit ÄOL vuosikirjat, Kirsi lisää viitteitä tähän). Tässä kehyksessä perinteinen, pääosin preskriptiivinen kuvaus tekstikokonaisuuden ja -kappaleen rakentumisesta ei 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, tekstitaitoja korostavan kouluopetuksen pohjalta ymmärrys tekstilajista, mutta repertuaarissa ei olekaan sopivaa rakennetta, jolla pystyisi itse tuottamaan tavoitemuotoisen tekstin ja esim. käymään dialogia lukijansa kanssa (ks. Juvonen 2012). Ja me siis haluamme lähestyä tätä niin päin, että kirjoittaja tietää mitä haluaisi tehdä, mutta ei sitä, millaisia rakenteita tai funktioita käyttäen se olisi mahdollista.

Sen rinnalle, että tarkastellaan, miten jokin retorinen kuvio reaalistuu rakenteiden avulla, haluamme tutkia, miten kirjoittaja voi päästä alkuun kappaleen kirjoittamisessa, jos hän tietää, mitä kappaleen pitäisi pitää sisällään. Eli toisin sanoen: Jos tiedän, ja niin kuin tämän aineiston kirjoittajat ovat tienneet, että minun on tässä raportissa kerrottava siitä, miten olin järjestänyt asumiseni opiskelijavaihdon aikana, millaisia erilaisia tapoja minulla on se tehdä. Tekstilaji on sinänsä selvä, apuna on satoja raportteja ja kohtalaisen tarkat kirjoitusohjeet. ”Tekstit eivät synny tyhjiössä, vaan pikemminkin ne liikkuvat laumoina ja ovat suhteessa toisiinsa monin tavoin” (???).

(Tässä tutkimuksessa siis tarkastellaan, millä tavalla uudessa tekstikappaleessa indikoidaan kappaleen rakentamaan alatopiikkia. Yksi tyypillinen viestinnällinen tarve, joka uudella kirjoittajalla voisi olla, on tekstin diskurssitopiikin käsitteleminen ja jaottelu alatopiikkeihin. Miten tämä tehdään, miten rakenteita varioidaan?)

Analyysin avulla haetaan vastausta seuraaviin tutkimuskysymyksiin: 1) Millaisia syntaktisia ja funktionaalisia tapoja yliopisto-opiskelijat käyttävät saman topiikin asettamiseen? 2) Miten tällaisia .. soveltaa kirjoittamisessa ja sen opettamisessa (-> vastauksena taksonomia)

Muutama sana johdannon loppuun aineistosta ja menetelmistä. Tähän perustelut aineistovalinnalle.

Akateemiset genret ovat kompleksisia (Bhatia 2012), ja siksi kirjoittamisprosessiin vaikuttaa aina lisäksi kirjoitettavan sisällön kompleksisuus. Tämä heijastuu väistämättä myös tekstin tuottamiseen (Bereiter and Scardamalia 1987; Camp 2012). Tämä tutkimuksen aineistovalinnassa haluttiin etsiä analysoitavaksi tekstejä, joiden rakentumisessa tehtävän kuormittavuus ei muodostuisi keskeiseksi tekijäksi. Vaihtoraporttien sisältö - kirjoittajan oma toiminta, kokemukset ja ajanjakson arviointi - on kirjoittajalle saavutettavissa vähäisin ponnistuksin ja tarjolla on seikkaperäiset ja havainnolliset ohjeet, tehtävä itsessään ei ole ollut kuormittava. Lisäksi kirjoittamiseen on tarjolla seikkaperäinen ohjeistus ja malliksi satoja aikasemmin kirjoitettuja, verkossa julkaistuja tekstejä.

Miksi ko. aineisto? -Iso, keskenään yhtenäinen tekstiaineisto -Kirjoitettu johonkin funktioon -Ei ammattikirjoittajia tai editoituja tekstejä, mutta tottuneita kirjoittajia -Tehtävä itsessään ei ole ollut kongnitiivisesti kuormittava ja malleja on ollut saatavilla runsaasti

## 2 Teoriaa (otsikko auki)

Korkeakouluopiskelijat kirjoittavat paljon, ja opintojen aikana kirjoitetuilla teksteillä (luentopäiväkirjat, esseet, opinnäytteet jne.) on selkeä funktio. Tekstien muotoa ohjaa joukko konventioita. Vuorijärvi tiivistää

tekstilajin käsitteestä ja sosiaalisesta luonteesta: “Tekstit eivät synny tyhjiössä, vaan pikemminkin ne liikkuvat laumoina ja ovat suhteessa toisiinsa monin tavoin” (???). Suomalaisissa oppilaitoksissa ja korkeakouluissa akateemiseen kirjoittamiseen ja tekstin tuottamiseen ei ole syvennytty samassa määrin kuin esim. USA:ssa, jossa akateemista kirjoittamista tyypillisesti opetetaan tehostetummin ja on myös tutkittu pitkään (???). Kirjoittamisen opetus- ja opiskelukulttuurit eroavat siis huomattavasti toisistaan. Kuitenkin jotain jo on: Ammattikorkeakoulun opinnäytetyön diskussio-osan siirtotyyppien analyysillä Vuorijärvi on saanut selville joukon kielellisiä resursseja, joista kielenkäyttäjän on mahdollista valita sopivimmat. Vuorijärvi toteaa, että näitä valintoja ohjaa lähinnä tekstilaji, opinnäytetyö on sekä vakiintunut että dynaaminen (s. 213-214). ”Tämä työ asettuu monessa suhteessa myös erikoiskielten tutkimuksen kontekstiin, joissa motivaationa on hankkia deskriptiivistä tietoa akateemisista tai professionaalisista tekstilajeista. Näissä tuloksia ja sovelluksia mallinnetaan erityisesti L2-käyttäjien näkökulmasta, mutta aineisto on syntyperäisten tuottamaa. (Vuorijärvi 96).”

Erityisesti englanninkielisessä kirjoittamisen opetuksen perinteessä on ollut tapana käyttää topic sentence -käsitettä kuvaamaan virkettä, jolla kirjoittaja ilmaisee kappaleen tärkeimmän ajatuksen. Suomeksi tämä kääntyy nykyisellään useimmiten ydinvirkkeeksi (tähän lähteitä). Ydinvirkkeiden todellisesta yleisyydestä ollaan tutkijoiden keskuudessa kuitenkin montaa mieltä. ks. Smith 2008). Tekstikappaleen ytimen tavoittelun sijaan tarkoituksenamme onkin tässä tutkimuksessa selvittää, minkälaisen vihjeiden avulla lukija johdatellaan kappaleen aiheeseen. Näitä vihjeitä nimitämme indikaattoreiksi. Tavoitteenamme oli siis kartoittaa rakenteet, joita varioimalla kirjoittajat indikoivat kappaleen topiikkaa. Yksi keskeinen kriteeri mm. äidinkielen ylioppilaskokeen arvioinnissa on kappaleiden sujuva liittyminen toisiinsa (Ylioppilastutkintolautakunta 2017). Lopputuloksena hahmottelemme asumiseen liittyvien topiikki-indikaattorien taksonomiaa

Tekstikappale tarkastelun kohteena (Olin ajatellut tätä teoriaosan alle, mutta eikös olekin yksi tulos jo se, millaisia tekstit ovat kokonaisrakenteeltaan. . .)

Tutkimusaineisto rajautuu ulkoisesti raportin kirjoittamiseen annetun ohjeistuksen ja raporttien julkaisualustan perusteella. Tekstin kokonaisrakenteen analyysissä on tyypillistä erottaa toisaalta funktionaaliset jaksot ja toisaalta visuaaliset elementit, tekstikappale lukeutuu jälkimmäiseen. Tässä tutkimuksessa tarkastelun kohteeksi otettiin tekstikappaleet, joka ei ehkä ole se tyypillinen tapa lähestyä tekstiä. Tehty valinta perustui tutkimuksen ensimmäisessä vaiheessa tehtyihin havaintoihin. (Miten pitkästi avaamme tätä?)

Vaihtoraportit ovat kokonaisrakenteeltaan pääosin topiikkipohjaisia - retorisen rakenteen teorian läpi tarkasteluna moniytimisiä. Näin ollen jokainen tekstikappale aloittaa uuden tekstikokonaisuuden, joka ei niinkään asetu hierakiseen suhteeseen edeltävien tekstikappaleiden kanssa, vaan ennemminkin on joko sekventiaalisesti tai listamaisesti jäsentynyt tekstin päätopiikin alle.

Vuorijärvi käyttää käsitettä topiikkipohjainen rakenne, jonka alle eri tavalla rakentuneet opinnäytetyötekstit asettuvat (s. 75). Tekstin aihetta siis avataan osa kerrallaan. Topiikilla tarkoitetaan tässä yhteydessä sisältötopiikkaa eli aihetta tai aihepiiriä, josta tekstissä sanotaan jotakin. Vuorijärvi esittelee (s. 75-76) Paltridgen 2002, 135 mukaisen skeeman Johdanto Topiikki 1 Topiikki 2 Topiikki 3 . . . . . Päätelmät

Tämän mallin rinnalla Vuorijärvi tarkastelee perinteistä, tutkimuslähtöistä mallia, jossa on osat: johdanto, menetelmä ja aineisto, tulokset ja pohdinta. Noin puolet Vuorijärven aineistosta asettuu tutkimuslähtöiseen malliin kategoriaan. (s. 81)

Retorisen rakenteen teoriassa (Komppa) eritellään tekstin osat ja kartoitetaan näiden osien tyypillisesti hierakinen suhde (ydin & satelliitit). Tämän tutkimuksen aineisto sisältää pääosin tekstejä, joiden osat ovat moniytimisessä suhteessa, jolloin osat ovat ennemmin rinnasteisia, keskenään tasavertaisia, eikä yhtä voida asettaa toisten ytimeksi. Komppa s. 39 antaa esimerkin sarjasta sekvenssisuhteita, jos haluamme hyödyntää.

JA sitten voisi jatkaa Dik & alatopiikit tähän omaksi kokonaisuudekseen, mutta tämän kirjoittaminen lienee viisasta, kun tulososa on hahmollaan.

- Raportti ja raportointi eivät esiinny monialaisten yliopistojen akateemisen kirjoittamisen ohjeistuksissa (ks. esim. Kielijelppi, Opiskelijoiden tekstilajeja) samalla tavalla kuin ammattikorkeakoulujen ohjeistuksissa, joissa sekä opinnäytetyön raportointi että harjoitustöistä kirjoitettavat raportit ovat keskeisiä tekstejä (ks. esim. Jyväskylän ammattikorkeakoulu). Yliopiston dokumentaatiossa tekstejä

nimetäänkin matkakertomuksiksi, mutta ohjeistuksessa annetut tukikysymykset viittaavat sisällöltään raportin kirjoittamiseen (ks. tarkemmin analyysiluvut). Myös tekstille eksplisiittisesti määritellyt tavoitteet, tukea vaihtoa harkitsevien päätöksentekoa ja auttaa valmistautumisessa, painottavat odotuksia raporttimaisen tekstin suuntaan.

Käyttämämme menetelmä muistuttaa prosessina Vuorijärven analyysiprosessin vaiheita: sivut (116-117).

Satunnaisesti valitun 30 tekstin otoksesta voi tehdä karkean luokittelun kolmeen erityyppiseen kokonaisrakenteeseen. Tyypittely on tehty kirjoittajan tekemän tekstikappalejaon perusteella.

Tyyppi 1 Kirjoittaja aloittaa tekstin vastaamalla ohjeistuksen ensimmäisen kohdan kysymyksiin. Jos lopussa on evaluoiva jakso, se jää irralliseksi.

Tyyppi 2 Kirjoittaja aloittaa johdannolla ja sen jälkeen siirtyy vastaamaan kysymyksiin. Tekstikappaleet ovat kuitenkin irrallisia toisistaan ja muistuttavat avoimen kyselylomakkeen vastauksia. Lopun arviointi jää hiukan pinnalliseksi.

Tyyppi 3 Kirjoittaja aloittaa johdannolla, ja raportoiva ote säilyy teksti läpi niin, että myös arvioiva jakso lopussa yhdistyy tekstiin luontevasti.

### 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”.

Tässä käsiteltävä lopullinen tutkimusaineisto kattaa kaikkiaan 415 opiskelijoiden laatimaa tekstiä, jotka tutkimuksen ensimmäisessä vaiheessa annotoitiin kappaleittain<sup>1</sup> siten, että jokaiselle kappaleelle merkittiin topiikki. Yhteensä kaksikymmentä tekstiä käsitteäen 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,

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

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

Jos tutkitaan tarkemmin niitä kategorioita, jotka syntaktinen jäsenin 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 määriteltyjen kriteerien pohjalta, vaan sillä perusteella, mitkä tapaukset muistuttivat toisiaan.

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

## 4 Yleiskuva

Edellä kuvatun laadullisen analyysin tuloksena tutkittavaksi valitut 168 tekstiä jaoteltiin 13 ryhmään. Kustakin ryhmästä on seuraavassa annettu yksi edustava esimerkkivirke. Esimerkkien tarkempiin konteksteihin pureudutaan alempana yksityiskohtaisemmin. Jo tässä vaiheessa on syytä huomauttaa, että osassa tapauksista asumisesta kertovaa kappaletta edelsi väliotsikko.

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

- (1) Asuntoni sain vuokrattua Apartiksen( paikallinen TOAS) kautta. (31 kpl)
- (2) Lensin Wieniin sunnuntaina ja lento-kentältä suuntasin suoraan asunnolleni. (9 kpl)
- (3) Vaihtoni aikana asuin Grandmontin asuntolassa. (21 kpl)
- (4) Sitten pari sanaa asumisesta. (6 kpl)
- (5) Asuntoloista valitsin Julius Raab Heimin vanhojen matkakertomusten perusteella. (14 kpl)
- (6) Asuminen Tokiossa on aika kallista. (10 kpl)
- (7) Olin löytänyt asunnon ensimmäiseksi kuukaudeksi Airbnb:n kautta ennen Italiaan saapumista. (21 kpl)
- (8) Saapuessani Varsovaan minulla oli onneksi jo asunto tiedossa. (5 kpl)
- (9) Asuminen, tai siis lähinnä asunnon hankkiminen, oli varmaankin vaihdon hankalin juttu. (16 kpl)
- (10) Suurlähetystön henkilökunta tarjosi resurssiensa puitteissa apuaan mm. maahantuloon, käytännön järjestelyihin sekä asunnon löytämiseen liittyvissä asioissa. (9 kpl)
- (11) Lähes kaikki Bratislavan kauppakorkeakoulun vaihto-opiskelijat asuivat Ekonom-nimisessä dormitoryssa. (4 kpl)
- (12) Itse hankin asunnon jo etukäteen, ja paikallinen kaverini kävi sitä katsomassa ja hakemassa avaimen jo etukäteen. (10 kpl)
- (13) Asuminen tapahtuu pienessä huoneessa yhdessä kämppiksen tai kahden kanssa. (12 kpl)

Esimerkissä 6 indikaattorina on koko lauseen ensimmäisenä sijaitseva *asuminen*-sana. Tämän ryhmän teksteille tyypillisesti esimerkki asettaa uuden topiikin esittämällä yleisen tason huomioita käsiteltävästä ilmiöstä. Toisin kuin esimerkin 9 tapauksessa, kirjoittaja ei siis lähde liikkeelle omista kokemuksistaan, vaan käsittelee tilannetta laajemmin, tavoitteenaan eräässä mielessä jakaa maailmantietoa lukijalleen. Esimerkin 6 edustama ryhmä kattaa kaikkiaan 10 tapausta (5,95 % koko aineistosta), joista 2:ssa (mukaan lukien esimerkki 6) asumisesta kertovaa kappaletta edeltää väliotsikko *Asuminen*.

Rakenteellisesti tarkasteltuna... y.3.p Tämän kyseisen esimerkin ympäristö

Esimerkissä 1 indikaattorina on *asuntoni*-sana. Tämän ryhmän tekstit muodostavat aineiston määrällisesti laajimman ryhmän ja leimallista indikaattorin sisältäville lauseille on, että indikaattori on useimmiten lauseen objekti ja hyvin usein myös lauseen ensimmäinen sana. Tämän ryhmän teksteissä indikaattori edustaa yleensä kontekstissaan ns. uutta tietoa, eli asumiseen, asuntoon, asuntolaan tms. ei ole viitattu edeltävässä tekstissä. Tätä esimerkkiä edeltävässä tekstissä kirjoittaja on käsitellyt ensin valimistelujaan kotoa käsin, sitten kielikurssia melko seikkaperäisesti, vaihto-opiskelijoille järjestettyä vapaa-ajan ohjelmaa ja lopuksi vielä reflektoinut omaa osaamistaan valitsemalla kielikurssilla. Pitää vielä lisätä, minkä verran tässä ryhmässä on väliotsikoita. Hiukan kesken siis on.

Esimerkissä 2 indikaattorina on *asuntolaan*-sana. Tämän ryhmän teksteissä asuminen raportoidaan kertovan tekstityyppin avulla, joka käy tästä esimerkistä esiin erityisen selvästi, koska kirjoittaja on käyttänyt ajan ilmausta. Aikamuoto näissä tapauksissa on tyypillisesti imperfekti. Kesken tämäkin.

Esimerkissä 5 indikaattorina on *asuntoloista*-sana. Ilman kontekstia tarkasteluna tapaus vaikuttaa samalta kuin listamainen esimerkki, mutta laajemman analyysin tuloksena selvisi, että kirjoittaja on jo aiempaan, kirjoittaessaan etukäteisjärjestelyistä ja kohdemaahan saapumisesta, rakentanut kuvaa paikallisista asuntolavaihtoehdoista ja asuntolassa elämisestä. Tämän ryhmän tekstit sijoittuvatkin toissijaisesti myös johonkin toiseen alaryhmään, mutta ovat erottuneet omasta alaryhmästään kontekstin tarkastelun vaiheessa. Eli kirjoittaja on edeltävässä tekstissään kirjoittanut sisään asumisen tavalla tai toisella ja jollakin lailla monimutkaisemmin kuin vain maininnut sen. Kesken tämänkin, siirsinkin jo kokeeksi kumminkin.

Esimerkissä 7 edellisessä kappaleessa jotain, muttei linkkiä

Esimerkissä 3 väliotsikko jotain

Esimerkissä 8 väliotsikko

Vertaa esimerkkiä 9 esimerkkiin 6

Jäljempänä tarkasteltavien x-tapausten jälkeen suurimmat ryhmät olivat

## 5 Zoomaus

Tämän analyysin tarkoituksena oli ottaa muutama tyypillinen eri asteisesti listamainen tapaus ja sen jälkeen asettaa se rinnakkain samaa ilmiötä toteuttavien muunlaisten ratkaisutapojen kanssa. Tässä on nyt kuitenkin yleistä tekstien rakenteen hahmottelua, ehkei mahdu/haluta lopulliseen, mutta ajatusten järjestelmiseksi ainakin tässä vaiheessa.

### 1. tapaus: puhdas listatyypinen (id732)

Ennen tätä kappaletta kirjoittaja on käsitellyt valmistelujaan kotoa käsin, sitten kielikurssia melko seikkaperäisesti, vaihto-opiskelijoille järjestettyä vapaa-ajan ohjelmaa ja lopuksi vielä reflektoi omaa osaamistaan valitsemallaan kielikurssilla. Näiden alatopiikkien (?) jälkeen kirjoittaja alkaa uuden tekstikappaleen seuraavasti:

Esimerkki x.x

Asuntoni sain vuokrattua Apartiksen (paikallinen TOAS) kautta. Vaihtareille on varattu kaksi erillistä taloa, toinen (Beauregard) sijaitsee aivan kaupungin keskustassa juna-aseman ja yliopiston päärakennuksen vieressä ja toinen (Bellevue) puolestaan noin 20 minuutin kävelymatkan päässä keskustasta. Asuntohakemus toimitettiin myös kaikkien muiden hakemuspaperien yhteydessä eli hakuprosessi oli siis senkin osalta tehty todella helpoksi. –

Shoren (2008, 42) sanoin tätä tapausta voisi kuvailla niin, että tekstin teemankulun hahmottaminen perustuu tietoihin, jotka "eivät ole tekstissä ilmipantuja vaan lähinnä kielenulkoista tietoa (viite). Teema on siis tekstin etenemisen kannalta (enemmän tai vähemmän uusi): se ei perustu selkeisiin merkityssuhteisiin vaan kehyksiin, skeemoihin. yms. pääteltävä teema siis." Tällä kirjoittajalla tuntuu olevan sama strategia koko ajan, koska jatkaa topiikeilla 'vaihtarit' ja 'asuntohakemus'. Lukijan kannalta tekstin koherennssi perustuu siis siihen, että opiskelijavaihdon kehys on lähtökohtaisesti tuttu tai lukija on lukenut joko raportin ohjeistuksen tai tutustunut muihin samalla alustalla julkaistuihin teksteihin.

Tekstin kokonaisrakenne tällä, kuten usealla muullakin aineiston kirjoittajalla, on Vuorijärven termiä käyttäen topiikkipohjainen. Raportissa käsitellään sisältö topiikki kerrallaan, eivätkä nämä sekvenssit ole hierarkkisessa suhteessa keskenään. (ks. myös Komppa). JA TULISIKO TÄHÄN LINKKI YLEISKUVAN TULOSSIIN, ELI ETTÄ USEIN LAUSEENALKUINEN OBJEKTI ON SE SYNTAKTINEN RAKENNE, JOTA HYÖDYNNETÄÄN?

Jos tekisimme näin, niin sitten tästä voisi jatkua suoraan jonkinlainen kuvio ja esimerkit, millaisia muita funktioita samalla rakenteella tehtään. (Tämä olisi ehkä jopa mielekäs kohta sille.)

Hiearkkisen rakenteen sijaan aineistosta löytyy kuitenkin kirjoittajia, jotka rakentavat tekstinsä temporaalisten sekvenssien avulla. Tämän strategian käyttö luo raporttiin narratiivin.

Esimerkki x.x. (id 399, linkitetty & narratiivinen)

"Hong Kongiin saavuttuani kaikki sujui täysin mutkattomasti ja haluan antaa CUHK:lle täyden kympin vaihtaritoiminnan hyvästä hoitamisesta. Saapuessani yliopistolle kaikki vaihtarit majoitettiin ensimmäisen viikon ajaksi samaan collegeen, jotta meidän olisi mahdollisimman helppo tutustua toisiimme ja jotta vaihtarien tapahtumien koordinointi olisi mahdollisimman helppoa. Ensimmäinen viikko oli niin kutsuttu Orientation week, joka oli täynnä kevyitä luentoja koskien mm. yliopiston tai asuntolan "pelisääntöjä" tai kuinka selvitä hengissä Hong Kongissa. Lisäksi kv-toimisto oli järjestänyt meille joinain iltoina ohjelmaa, kuten esimerkiksi Tervetuliaisillallisen. Ensimmäinen viikko on siis täydellinen tilaisuus tutustua uusiin vaihtareihin ja päästä nauttimaan vaihtokokemuksesta jo heti ensimmäisestä päivästä lähtien.

Orientation weekin lopussa muutettiin lopulliseen asuntolaan. Minä päädyin ennalta tekemieni valintojen pohjalta (no mandatory meal etc.) Shaw Collegen I-house 2:seen, johon liittyy paljon hyviä ja huonoja puolia. Ensinnäkin yliopistolla asuminen on ylivoimaisen käytännöllistä ja halpaa yksityisen asunnon vuokraamiseen verrattuna. – "

Tässä topiikkia indikoi 'asuntola', joka on lauseessa täydennyksenä toimiva paikan adverbiaali. Lause linkittyy edelliseen kappaleeseen (Shoren sanoin, mutta meille Dikin käsittein, kun teoriaosa on kirjoitettu) orientoivana sivuteemana toimivan ajan adverbiaalilla. Edellisessä kappaleessa on kuitenkin puhuttu

orientaatioviikosta ja majoittamisestakin, eli tämä kirjoittaja ei ensimmäisen esimerkin tavoin nojaa tekstin ulkopuoliseen tietoon, vaan itse kirjoittamaansa edeltävään tekstijaksoon.

Ja vielä kolmas esimerkki, joka asettuisi tulkintani mukaan jollakin jatkumolla näiden kahden välimaastoon.

Esimerkki x.x tekstin id 618

Kirjoittaja on edellisessä kappaleessa kuvannut saapumistaan kohdemaahan ja järjestelyjä siihen liittyen. Kappaleessa on yksi maininta asumisjärjestelyistä, mutta se tulee esille yleisen aikatauluttamisen yhteydessä, joten se ei sinänsä orientoi millään tavalla seuraavan kappaleen topiikkia.

“Soluasunnon itselleni hain Saksasta wg-gesuchtin nettisivujen kautta. Saksalaisen perheen vuokraamassa solussa asui lisäksi kolme muuta. Vaikkei kämppiksien kanssa yhteiselo aina ilman ongelmia onnistunutkaan, suosittelen ehdottomasti ainakin harkitsemaan opiskelija-asuntolan sijaan soluasunnossa asumista.–”

## 6 Päätelmiä

Ensin: edellä tehdyn analyysin perusteella saadut ryhmät kootusti

Sitten:

- Kuinka monta eri tapaa oikeasti löydettiin?
- mihin tällaista voisi soveltaa?
- mitä voidaan sanoa tekstilajista / kirjoittajista / yms.?
- Testaa “valvottua klusterianalyysiä” käyttäen tässä tehtyä luokittelua pohjana
  - Manuaalisen analyysin perusteella todella voidaan paitsi määrittää ryhmät, joihin indikointitavat voi jakaa, myös opettaa algoritmi, joilla muita tekstejä voisi ryhmitellä samoin perustein.

## Lähteet

Chafe, Wallace 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

Haverinen, Katri, Jenna Nyblom, Timo Viljanen, Veronika Laippala, Samuel Kohonen, Anna Missilä, Stina Ojala, Tapio Salakoski & Filip Ginter 2014. Building the essential resources for Finnish: the Turku Dependency Treebank. *Language Resources and Evaluation* 48:3, 1–39.

Komppa, Johanna 2012. Retorisen rakenteen teoria suomi toisena kielenä-ylioppilaskokeen kirjoitelman kokonaisrakenteen ja kappalejaon tarkastelussa.

Levshina, Natalia 2015. *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company.