

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

Kirsi Sandberg & Juho Härme

Contents

1 Johdanto	1
2 Teoriaa (otsikko auki)	3
3 Tutkimusaineisto ja -metodi	4
4 Tulokset	7
4.1 Yleiskatsaus tavoista asettaa topiikki	7
4.2 Puu	10
4.3 Kirjoittajan näkökulma	12
5 Päätelmät	12
Lähteet	13

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. Honkanen 2012). 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, Nieminen, Honkanen etc.), 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 perustella 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. Tekstin kirjoittamisen osalta, tekstikappaleiden tasolla, saatavilla on siis toistaiseksi enemmän tietoa siitä, miten se pitäisi tehdä, kuin deskriptiivistä tietoa todellisesta kielenkäytöstä

(Komppa 2012: 121) (Chafe 1994: 296). Lisää vielä tähän side EU-kartoitukseen ja tutkimustarpeeseen sovellusten kehittämistä varten. Tutkimusperustaisen kirjoittamisenopetuksen tueksi tarvitaan nähdäksemme lisää käyttöpohjaista, kuvailevaa tutkimusta.

(hypätäänkö tässä hiukan sfääreihin??) 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 ilmaisuvarustusta ja kielellinen valinnanvara laajenevat kattamaan myös yhteisön puhuttavat (esim. Bruce 2008 5-6, Solin 2006, 82 samoin Luukka 2002b, 14, 27)”. Vuosikymmenen loppua lähestyttäessä kirjoittamisen tutkimus kuitenkin päättyy 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. (tajuunavirtaa 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” (2013: 69).

(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)

(muistissa tässä vain itsellä: nämä eivät ole ns. konventionaalisen raportin muotoisia, raportti ei monialaisen yliopiston repertuaarissa ole keskeinen, sen sijaan tekstit ovat keskenään saman muotoisia ja funktioltaan samanlaisia, joten niitä voisi nimetä omaksi tekstilajikseen, vaihtoraportiksi tai vaihtokertomukseksi.)

Usein puhutaan tekstin makro- ja mikrotasosta. Vaikka tarkastelemme tekstikappaleita ja rajaamme vieläkin pienempään eli avausvirkkeeseen, tutkimuskysymyksemme on ennemmin makro- kuin mikrotasolla. Nimittäin, teksti on jakautunut myös topiikkiin, ylätopiikki jakautuu alatiopikkeihin sisällönkin puolesta, ei ainoastaan funktioiden. Ja kirjoittaja voi yhtäläisesti rakentaa tekstiä niiden avulla. Ja tarvita konkreettista ohjeistusta siihen, miten tämä kielen tasolla tehdään.)

Korkeakoulun tekstilajit ja niiden hallinta (voisiko olla yläotsikko?)

Tekstikappaleet ja niiden väliset suhteet (alaotsikko)

Tämä tutkimus on alusta alkaen rakennettu tähdäten opiskelun ja opettamisen sovellusmahdollisuuksiin. Samasta lähtökohdasta on myös valittu keskeinen analyysiyksikkö, joka on siis tekstikappale. Tekstin kokonaisrakenteen analyysissä on tyypillistä erottaa toisaalta funktionaaliset jaksot ja toisaalta visuaaliset elementit, tekstikappale lukeutuu jälkimmäiseen. Valintamme poikkeaa useimmista lähestymistavoista, sillä tyypillisesti tekstikappaleita ei tarkastella itsenäisinä kokonaisuuksina tekstilingvistiksessä siirtoanalyysissä (pari lähde) eikä retorisen rakenteen analyysissä (pari lähde), jotka Suomen kontekstissa ovat keskeisimmät lähestymistavat opiskelutekstien analyysiin. Toisin sanoen tekstikappaleita pidetään lähinnä ortografisina merkkeinä, jotka eivät itsessään ole merkitseviä tekstin rakentumisen kannalta.

Kirjoittajan kannalta kappalejako on kuitenkin hyvin konkreettinen ja käytännön sanelema työkalu. Kirjoittamisoppaissa ja ohjeistuksissa annetaan kuvauksia siitä, millä tavalla teksti rakentuu kappaleistaan, ja millaisia kappaleiden – sisällön ja funktion tasolla – olisi hyvä olla. Lisäksi kappaleiden koherentti toisiinsa sitominen on tyypillisesti arviointikriteeri, esim. ylioppilaskirjoituksissa (viite). Kirjoittamisprosessin kuluessa kirjoittaja voi kirjoittaa sisältöä eri topiikeista ja vasta esim. editointivaiheessa rakentaa tekstiä hierarkkiseksi kokonaisuudeksi. ”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.)”

Vuorijärvi on siirtotyyppien analyysillä saanut selville joukon kielellisiä resursseja, joista kielenkäyttäjän on mahdollista valita sopivimmat ja hänen mukaansa näitä valintoja ohjaa lähinnä tekstilaji, tässä tapauksessa opinnäytetyö(s. 213-214). Vuorijärvi tarkastelee työssään ammattikorkeakoulun opinnäytetyöiden tekstuaalista rakentumista, erityisesti diskussio-osiota. Hän kuvaa perinteisen, tutkimuslähtöisen mallin, jossa on osat: johdanto, menetelmä ja aineisto, tulokset ja pohdinta. Noin puolet Vuorijärven aineistosta asettuu tutkimuslähtöiseen malliin kategoriaan. (s. 81). Lisäksi Vuorijärvi esittelee Paltridgeä mukaillen topiikkipohjaisen rakenteen (me siis jatkossa myös)(s. 75). Tekstin aihetta siis avataan osa kerrallaan. Topiikilla tarkoitetaan tässä yhteydessä sisältötopiikkia 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

(tällaiseen skeemaan yliopiston antama ohjekin kevyesti ohjaa, tosin johdantoa ja päätelmiä ei mainita, mutta ne ovat ikään kuin läsnä annetuissa topiikeissa)

Myös retorisen rakenteen teoriassa pyritään pääsemään käsiksi tekstin osien välisiin suhteisiin, ja teoria rakentuu oletukselle, että tekstin osat ovat hierarkkisessa suhteessa toisiinsa. Tässä tutkimuksessa ei kuitenkaan ole tutkittu tekstin osien välisiä suhteita, vaan liikutaan topiikkien tasolla. Kuten Komppakin kirjoittaa, tekstinosa ydin voi sijaita yhtäläillä ortografisen kappaleen alussa tai muualla ja kuten Juvonen sanoo, eri tekstityyppejä yhdistellen voi päästä samaan lopputulokseen, ei oletuksemme voi olla, että ortografisen kappaleen ensimmäinen lause tai virke yksinään paljastaisikaan mitään siitä, millaisessa asemassa kappale hierarkkisesti on tekstin kokonaisrakenteen kannalta. Yhtäläisesti kuin että mitä tekstissäni seuraavaksi teen, kirjoittajalla voi olla ajatus siitä, mistä asiasta, eli topiikista haluaisi seuraavaksi kirjoittaa. Hän siis

tarvitsee myös sisällöllisen linkin, joka on mukana rakentamassa tekstin kokonaiskoherenssia. Lähdemme siis oletuksesta, että topiikit eivät ole sidoksissa vaiheisiin, jaksoihin tai muihin tekstin rakenteen kannalta keskeisiin osiin.

Vaihtoraportit ovat kokonaisrakenteeltaan pääosin topiikkipohjaisia - retorisen rakenteen teorian läpi tarkasteluna moniymitisiä. 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. Tämän tutkimuksen aineisto sisältää pääosin tekstejä, joiden osat ovat moniymitisisessä suhteessa, jolloin osat ovat enemmän rinnasteisia, keskenään tasavertaisia, eikä yhtä voida asettaa toisten ytimeksi. Komppa s. 39 antaa esimerkin sarjasta sekvenssisuhteita, jos haluamme hyödyntää.

(Lopeta tämä kappale tähän) 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 (???).

Topiikin asettaminen (alaotsikko)

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 topiikkia. Yksi keskeinen kriteeri mm. äidinkielen ylioppilaskokeen arvioinnissa on kappaleiden sujuva liittyminen toisiinsa (Ylioppilastutkintolautakunta 2017). Lopputuloksena hahmottelemme asumiseen liittyvien topiikki-indikaattorien taksonomiaa

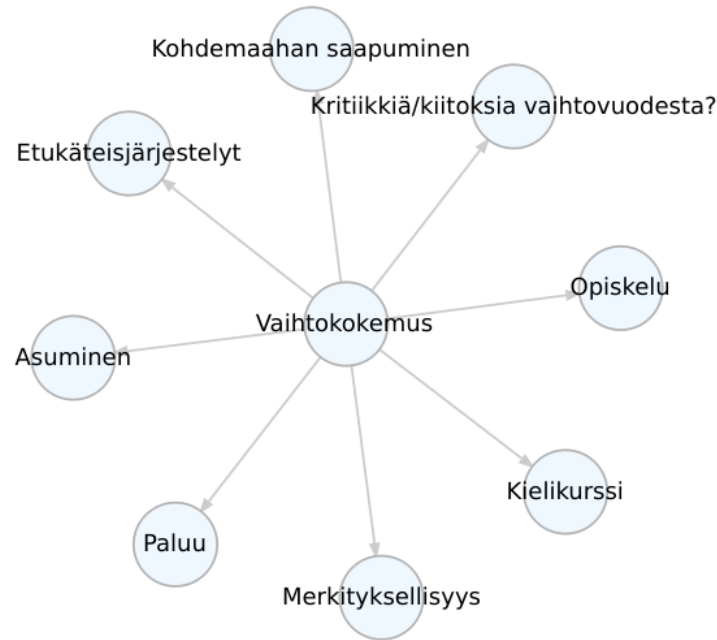
Samoin Paltridge (2002) on verranut tutkielmaoppaita ja todellisten tutkielmien rakennetta ja toteaa, että yhtä mallia esitetään ja todellisuudessa variaatiota on huomasti. Tässä meidän tapauksessamme jo yliopiston antama ohjeistus ohjaa kirjoittajaa Paltridgen kuvaamaan topiikkipohjaiseen jäsennykseen (jota myös Komppa ja Vuorijärvi ovat löytäneet) sen sijaan, että odotukseksi asetettaisiin konventiaalisien raporttien malli, johon ohjataan esim. ammattikorkeakoulujen kirjoitusoppaissa (ks. esim. viiten Jyväskylän yliopiston sivuille, viite olemassa jo.)

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

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

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 kahdeksanakohtaisen 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 aiheajauksen 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 kappaleittain¹ 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

¹Käyttäen 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 sarakke).² 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...

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

4 Tulokset

Edellä kuvatun laadullisen analyysin tuloksena tutkittavaksi valitut 168 tekstiä jaoteltiin 13 ryhmään. Osiossa 4.1. kutakin ryhmää luonnehditaan yleisellä tasolla tarkastelemalla lähemmin yhtä edustavaa esimerkkiä. Vaikka varsinainen ryhmäjako on tehty puhtaasti laadullisin perustein pohjaamalla tapausten keskinäiseen samankaltaisuuteen, tulkitaan osiossa 4.2. analyysin tulosta eli itse jaottelua määrällisten ominaisuuksien pohjalta. Osio 4.3 peilaa tässä tehtyä ryhmittelyä aiempiin tutkimuksiin ja siihen, miten tässä funktionaalisiin perustein tehty jaottelu osuu yksin sen kanssa, miten tekstejä on toisaalla jaoteltu kirjoittajan näkökulman perusteella.

4.1 Yleiskatsaus tavoista asettaa topiikki

Käymme tässä alaluvussa laadullisen analyysin tuloksena syntyneet ryhmät läpi niiden koon mukaisessa järjestyksessä suurimmasta pienimpään. Aineiston laajimman ryhmän, kaikkiaan 31 tapausta, muodostavat esimerkin 1 kaltaiset topiikkia asettavat virkkeet. Ryhmässä on jonkin verran tekstejä (7 kpl / 22.58 %), joissa varsinaista asumisesta kertovaa kappaletta tai kappaleita edeltää aiheeseen johdattava väliotsikko kuten *Asuminen* tai *Asuminen Istanbulissa* – näin ei kuitenkaan ole tässä esitetyn virkkeen tapauksessa.

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

Esimerkin 1 topiikki-indikaattorina on *asuntoni*-sana, joka tälle ryhmälle leimallisesti on lauseen alussa sijaitseva objekti. Asumiseen tai asuntoon ei ole viitattu edeltävissä kappaleissa, vaan topiikkia asettava virke on samalla ensimmäinen maininta itse topiikista: esimerkiksi 1 edeltävässä tekstissä kirjoittaja on käsitellyt ensin valmistelujaan kotoa käsin, sitten kielikurssia melko seikkaperäisesti, vaihto-opiskelijoille järjestettyä vapaa-ajan ohjelmaa ja lopuksi vielä reflektoinut omaa osaamistaan valitsemallaan kielikurssilla. Informaatorakenteellisesti tarkasteltuna esimerkin 1 *asuntoni* voidaan nähdä alatopiikkina (ks. osio x.x edellä), vaikka esimerkissä 1 ei täyty Dikin (1989: 275) kuvailema ehto siitä, että varsinainen ylätopiikki olisi erikseen esitelty lukijalle (vrt. myös Guijarro 2001: 110). Kirjoittaja kuitenkin operoi tässä ja useimmissa ryhmän 1³ 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. Asunto tai asuminen nähdään toisin sanoen lukijan käytettävissä olevana entiteettinä (vrt. Chafe 1987: 29; Lambrecht 1996: 99), mikä mahdollistaa sen tuomisen tekstiin ilman ennakkomainintoja tai esittelyjä. Voisi jopa sanoa, että esimerkin 1 kaltaiset tapaukset eivät niinkään aseta topiikkia, vaan ennemmin käyttävät jo asetetuksi oletettua topiikkia.

Toiseksi suurimman laadullisen analyysin tuloksena syntyneistä ryhmistä muodostavat esimerkin 2 kaltaiset tapaukset, joita tutkimusaineistossa on 21 kappaletta. Esimerkkiä 2 samoin kuin 6 muuta tämän ryhmän virkettä (yht. 33.33 % ryhmän tapauksista) edeltää asumiseen viittaava väliotsikko.

(2) Vaihtoni aikana asuin Grandmontin asuntolassa.

Kaikissa ryhmän 2 tapauksista topiikki-indikaattorina on persoonamuotoinen *asua*-verbi, ja virkkeiden pääasiallinen sisältö on yleensä informaatio siitä, missä kirjoittaja on vaihtonsa aikana asunut. Tämän ryhmän tapaukset ovat edellistä ryhmää selkeämmin nimenomaan johdantoja uuteen aiheeseen, ja Dikin jaottelun mukaisesti esimerkin 2 lopussa mainittu *asuntola* muodostaa (osiossa x.x. määritellyssä termin teknisessä merkityksessä) uuden topiikin (1989: 269). Vastaavat *asuin paikassa X* -tyyppiset rakenteet ovat siis keinoja esitellä *paikka X*, josta tällaisella virkkeellä alkava kappale kertoo. Esimerkin 2 tapauksessa kirjoittaja jatkaa kappalettaan kuvailemalla asuntonsa sisustuksen, kämppäkaverit ja vuokran suuruuden.

Esimerkin 3 kaltaisten tapausten muodostama ryhmä on täsmälleen samankokoinen kuin ryhmä 2 (21 tapausta) – sattumalta myös otsikoitujen kappaleiden määrä on identtinen (7 kappaletta).

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

³Viitaamme jatkossa ryhmiin tässä esitettyjen esimerkkien numeron perusteella.

Esimerkin 3 edustamat virkkeet ovat ajallisesti muita monitasoisempia. Lukija johdatetaan kappaleen topiikkiin kuvaamalla jotakin, mikä on tapahtunut ennen sitä ajanjaksoa, jolle varsinainen käsiteltävä topiikki sijoittuu. Esimerkin 3 tapauksessa luodaan 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. Kappaleen loppuosassa kirjoittaja kuvailee tarkemmin asunnon yksityiskohtia: sijaintia, kulkuyhteyksiä ja tyypillisyyttä kohdekulttuurissa. Informaatorakenteeltaan ryhmä 3 onkin hyvin samanlainen kuin edellinen ryhmä: koko kappaleen topiikkina oleva *asunto* esitellään uutena topiikkina ensimmäisessä virkkeessä.

Neljäntenä ryhmänä tarkastellaan seuraavanlaisia tapauksia, joissa topiikki asetetaan luonnehtimalla jotenkin kirjoittajan henkilökohtaista kokemusta aiheesta:

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

Esimerkissä 4 topiikkia indikoi *asuminen*, mikä tähän ryhmään luokitelluilla kaikkiaan 16 tapauksella onkin tyypillistä: kenties suoraviivaisin tapa kuvailla omaa kokemusta jostakin varsinaisen diskurssitopiikin osa-alueesta on niputtaa kokemus nominalisoituun verbimuotoon ja luonnehtia tätä predikatiivilauseella. Näissä tapauksissa kuvataan, miten *asunnon hakeminen oli helppoa* tai miten *kaikkein vaikeinta oli asunnon löytäminen*. Kirjoittajan näkökulma onkin usein koko tekstin ja diskurssitopiikinkin tasolla subjektiivinen – strategiana on kertoa käsiteltävästä asiasta sen kautta, miten asiat ovat sujuneet, omia onnistumisia ja epäonnistumisia arvioiden. Informaatorakenteen kannalta nämä tapaukset ovat monesti ryhmän 1 tapaan alatopiikkeja. Esimerkki 4 on kuitenkin esimerkkiä 1 selvemmin topiikkia asettava virke, sillä siinä asumisen rooli yhtenä vaihtokokemuksen osana on ilmaistu eksplisiittisesti. Vain kahdessa ryhmän 4 tapauksessa asumisesta kertovaa kappaletta edeltää väliotsikko.

Esimerkissä 5 indikaattorina on *asuntoloista*-sana. Ilman kontekstia tarkasteltuna tämäkin tapaus vaikuttaa samalta kuin listamainen esimerkki, mutta laajemman analyysin tuloksena selvisi, että kirjoittaja on jo aiempaan, kirjoittaessaan etukäteisjärjestelyistä ja kohdemaahan saapumisesta, rakentanut kuvaa paikallisista asuntolavaihtoehtoista ja asuntolassa elämisestä. Tämän ryhmän tekstit sijoittuvatkin toissijaisesti myös johonkin toiseen alaryhmään, mutta ovat erottuneet omasta alaryhmästään kontekstin tarkastelun vaiheessa.

- (5) Asuntoloista valitsin Julius Raab Heimin vanhojen matkakertomusten perusteella.

Esimerkin 5 kaltaisia aiempaan kontekstiin tietoisesti sidottuja tapauksia on aineistossa 14 kappaletta. Näissä tapauksissa kyse on virketasolla tarkasteltuna joko annetuista topiikeista – jos kontekstuaalinen linkki on muodostettu suhteellisen lähellä tässä käsiteltävää varsinaista asumiskappaletta – tai, jos edellinen maininta on selvästi kauempana, palautetuista topiikeista (Dik 1989: 270–271). X:ssä tämän ryhmän virkkeessä kappaletta edeltää väliotsikko.

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

Esimerkissä 6 kirjoittaja esittelee uuden topiikin lähtemällä liikkeelle konkreettisesti asuntonsa luonnehtimisesta (vrt. esimerkki 4, jossa luonnehditaan kirjoittajan kokemusta). Vastaavia asuntolaa tai asumista staattisesti kuvaavia topiikin asettavia virkeitä on kaikkiaan 12 kappaletta. Informaatorakenteeltaan nämä tapaukset ovat monesti lähellä ryhmää 2, sillä molemmissa kappaleen ensimmäinen virke esittelee lukijalle sen, minkä yksityiskohtaiseen kuvailuun koko loppukappale keskittyy. Esimerkissä 6 topiikki asetetaan kuitenkin jossain määrin implisiittisemmin, ja hyppäys edellisestä aiheesta (esimerkin 6 tapauksessa HIV-testin teettäminen) on tyypillisesti rajumpi. Tämä selittää sitä, että peräti 58.33 % ryhmän 6 tapauksista (ei kuitenkaan tässä esitetty esimerkki) sisältää asumiskappaletta edeltävän väliotsikon.

Esimerkki 7 edustaa kaikkiaan 10 tapausta käsittävää ryhmää, jonka ominaispiirteenä on kirjoittajan itsensä vertailu muihin vaihdossa olleisiin.

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

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

samoja piirteitä kuin esimerkissä 4: myös tässä reflektoidaan sitä, mikä omassa kokemuksessa oli helppoa ja mikä vaikeaa. Yhdessäkään ryhmän 7 tapauksessa ei ole asumiseen viittaavaa väliotsikkoa.

Samoin kuin esimerkissä 4, myös esimerkissä 8 topiikki-indikaattorina on koko lauseen ensimmäisenä sijaitseva *asuminen*-sana. Erona näiden esimerkkien välillä on, että siinä missä esimerkki 4 on informaattiorakenteen kannalta omasta vaihtokokemuksesta lohkaistu alatopiikki, ankkuroituu esimerkki 8 siihen fyysiseen paikkaan, jossa vaihto tapahtui:

(8) Asuminen Tokiossa on aika kallista.

Ryhmän 8 kirjoittajat eivät lähde liikkeelle omista kokemuksistaan, vaan käsittelevät tilannetta laajemmin, tavoitteenaan eräässä mielessä jakaa maailmantietoa lukijalleen. Esimerkin 8 edustama ryhmä kattaa kaikkiaan 10 tapausta, joista kahdessa (mukaan lukien esimerkki 8) asumisesta kertovaa kappaletta edeltää väliotsikko *Asuminen*.

Yleistys kirjoittajan oman kokemuspiirin ulkopuolelle voidaan kuitenkin tehdä myös toisella tavalla. Esimerkissä 9 ja kahdeksassa muussa tapauksessa 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 informaattiorakenteen kannalta katsottuna kyse esimerkkien 2, 3 ja 6 tapaan uudesta topiikista, mutta sen esittely toteutetaan hyödyntämällä jotakin tekstin kannalta keskeistä toimijaa, tässä tapauksessa suurlähetystön henkilökuntaa. Kuten ryhmän 7 kohdalla, myös ryhmässä 9 yli puolessa tapauksia (55.56 %) asumisesta kertovaa kappaletta edeltää väliotsikko.

Esimerkin 10 edustamassa ryhmässä asumisesta raportoidaan kertovan tekstityypin avulla, mikä käy tästä esimerkistä esiin erityisen selvästi, koska kirjoittaja on käyttänyt ajan ilmausta.

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

Aikamuoto ryhmän 10 virkkeissä on tyypillisesti imperfekti. Näissäkin tapauksissa topiikki-indikaattori on informaattiorakenteen kannalta uusi topiikki, joka esimerkin narratiivisuuden vuoksi tulee asetetuksi muita tässä tarkasteltuja uuden topiikin sisältäviä ryhmiä implisiittisemmin. Kaikkiaan ryhmään 10 kuuluu 9 tapausta, joista niistäkin suurin osa (5 kpl) sisältää väliotsikon.

Oman, selvästi muista erottuvan ryhmänsä muodostavat ne esimerkin 11 edustamat 6 tapausta, joissa kirjoittaja eksplisiittisesti ilmaisee siirtyvänsä käsittelemään asumista.

(11) Sitten pari sanaa asumisesta.

Sitten pari sanaa asumisesta -tyyppisten johdantovirkkeiden voisi ajatella jossain määrin korvaavan väliotsikoita tekstissä, vaikka yksi tähän ryhmään luokiteltu virke (*Monella varmaan herää jo kysymys, entäs asuminen?*) tekeekin poikkeuksen. Nämä 6 esimerkkiä edustavat Dikin jaottelussa uusia topiikkeja (1989: 268), ja ovat aiheeseen johdattamisen implisiittisyys-eksplisiittisyys-asteikon eksplisiittisessä ääripäässä (vrt. päinvastaisessa ääripäässä sijaitseva ryhmä 1).

Laadullisessa analyysissä luokiteltiin omaksi pieneksi ryhmäkseen myös esimerkin 12 kaltaiset 5 tapausta:

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

Nämä topiikkia asettavat tapaukset sijoittuvat ryhmien 2, 3 ja 10 välimaastoon. Kuten mainituissa ryhmissä, myös näissä tapauksissa kyseessä on varsinainen aiheeseen johdettava virke, jossa kappaleen topiikki esitellään uutena avausvirkkeen loppupuolella. Myös ryhmässä 12 kuvataan ajallinen jännite, mutta se esitetään usein eri järjestyksessä kuin ryhmässä 3. Siinä missä esimerkki 3 esitti kappaleen alussa avoimen tilanteen ja myöhemmässä kappaleessa tilanteen ratkaisun, aloittaa esimerkki 12 ratkaisusta (saavuttaessa asunto oli jo löydetty) ja käyttää lopun kappaleen takaumamaisesti sen kuvailuun, miten ratkaisuun päästiin (kirjoittaja oli etsinyt asuntoa valmiiksi Facebookin kautta).

Pienimmän mutta silti tarkastelemisen arvoisen ryhmän aineistossamme muodostivat esimerkki 13 ja kolme muuta vastaavaa tapausta:

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

Nämä topiikkia asettavat virkkeet ovat lähellä ryhmää 7, johon luetuissa tapauksissa oleellista on kirjoittajan itsensä vertailu muihin. Ryhmässä 13 kirjoittaja lähtee liikkeelle siitä, miten yleensä tehdään tai miten suurin osa tekee. Tämä voitaisiin esittää vastakohtana omalla kokemuksella, mutta ryhmän 13 tapauksille on tyypillistä, että myös kirjoittaja noudattaa kuvailtua enemmistön toimintatapaa. Virke 13 on todella vain keino johdattaa lukija kappaleen varsinaiseen aiheeseen, *Ekonom*-nimiseen asuntolaan (uusi topiikki), jonka sijainnista, vuokratasosta ja muista ominaisuuksista kappale kertoo. Yhtä lukuun ottamatta kaikki tämän ryhmän tapaukset sisältävät myös asumiseen viittaavan väliotsikon.

Kuten edellä metodin kuvauksen yhteydessä todettiin, tässä kuvatut eri topiikin asettamistapoja edustavat ryhmät on muodostettu sijoittamalla aina 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. 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 opettajalle olla abstrakteja funktionaalisten ominaisuuksien kuvailuja helpompaa.

4.2 Puu

Tekstuaalisia funktioita ja ja kappaletason elementtejä kuvattaessa on usein.. Esimerkiksi Vuorijärvellä kuvaus “blablaa”. Jos opettaja haluaa konkretisoida eri tekstistrategioita, hänen on tavallisesti turvauduttava vain abstrakteihin määritelmiin. Yhtenä tämän tutkimuksen tarkoituksena on kuitenkin tuottaa myös toisenlaista, rakenteiden tason tietoa... Tätä tietoa on kerätty seuraavassa soveltamalla laadullisen analyysin tuloksena saatuun ryhmäjakoon määrällisiä menetelmiä, tarkemmin ottaen ehdollisia päätöksentekopuita (*conditional inference trees*).

Ehdollisissa päätöksentekopuissa ajatuksena on, että (Levshina 2015, x) . Analyysissa määritellään yksi selitettävä muuttuja (vrt. Ketokivi 2015, x) ja (tavallisesti useita) selittäviä muuttujia, joiden avulla jonkin tapauksen päätymistä johonkin luokkaan pyritään ennustamaan. Nyt käsillä olevassa tutkimuksessa selitetävän muuttujan muodostavat laadullisen analyysin tuloksena luodut topiikin asetustapojen 13 ryhmää. Yksittäisen matkakertomuksen päätymistä johonkin näistä ryhmistä on suorittamassamme luokittelupuuanalyysissä tarkasteltu seuraavien selittävien muuttujien avulla.

1. *Topiikki-indikaattorin dependenssirooli*. Tämän muuttujan avulla määriteltiin alkuperäinen satunnaisotanta. Sen arvojen pohjana ovat taulukossa 1 luetellut dependenssiroolit, joita kuitenkin lavennettiin siten, että arvot *nsubj* ja *nsubj:cop* yhdistettiin yhdeksi subjektin kategoriaksi (*subj*) ja arvot *nmod:poss* sekä *nmod:gobj* yhdeksi *nmod:x*-kategoriaksi, jolla viitattiin kaikkiin kompleksisiin nominaalirakenteisiin (*asunnon hankinta*, *asumisen järjestäminen*).
2. *Topiikki-indikaattorin sijainti virkkeessä* jakaa aineiston kahteen luokkaan: tapauksiin, joissa topiikki-indikaattori on koko virkkeen ensimmäisenä ja muihin tapauksiin.
3. Myös *lauseen predikaattiverbin persoona* jaettiin kahteen kategoriaan: tapauksiin, joissa predikaattiverbi on ensimmäisessä persoonassa ja muihin tapauksiin.
4. *Side muuhun tekstiin* -muuttujalla on kolme kategoriaa: tapaukset, joissa asumistopiikkia on pohjustettu jo edellisessä kappaleessa muttei niin selvästi, että kyseessä olisi linkkitapaus (*edel.*); tapaukset, joissa topiikki asetetaan viittaamalla selkeästi johonkin edellä muodostettuun linkkiin (*linkki*); tapaukset, joissa asumiskappaletta edeltää väliotsikko (*ots*) sekä tapaukset, joissa mitään edellä kuvatuista siteistä ei ole (*x*).
5. *Topiikki-indikaattorin morfologinen rakenne* erottelee toisistaan finiittiverbit, infiniivimuodot (mukaan lukien minen-johdokset) ja substantiivit.
6. *Aikamuoto*-muuttuja pitää sisällään arvot *preesens*, *imperfekti*, *pluskvamperfekti/perferkti*⁴ ja *ellipsi* (-).

⁴Perfektitapauksia aineistossa oli vain yksi ja sekin konditionaalissa, joten molemmat suomen liittoaikamuodot yhdistettiin samaan kategoriaan.

Figure 1 displays a graphical model and its corresponding 15 node marginal distributions. The graphical model on the left shows a hierarchical structure of nodes (1-15) representing linguistic features and their dependencies. Node 1 is the root, with children Node 2 (side) and Node 15 (pers). Node 2 has children Node 3 (tense) and Node 16 (side). Node 3 has children Node 4 (tense) and Node 17 (linkki). Node 4 has children Node 5 (dep) and Node 18 (side). Node 5 has children Node 6 (pers) and Node 19 (muu). Node 6 has children Node 7 (pers) and Node 8 (dep). Node 7 has children Node 9 (pers) and Node 10 (dep). Node 9 has children Node 11 (pers) and Node 12 (dep). Node 11 has children Node 13 (pers) and Node 14 (dep). Node 13 has children Node 15 (pers) and Node 16 (side). Node 15 has children Node 17 (linkki) and Node 18 (side). Node 17 has children Node 19 (muu) and Node 20 (dep). The 15 marginal distributions on the right show the probability of each feature value for each node, with the x-axis representing the feature value and the y-axis representing the probability. The distributions are labeled Node 1 through Node 15.

Kuviossa 3 tutkimusaineisto on jaettu kymmeneen ryhmään (kuvion alimmalla rivillä esitetyt pylväskaaviot) edellä kuvattujen selittävien muuttujien perusteella. Kuvio esittää eri selittävien muuttujien vaikutuksen selitettävään muuttujaan sarjana päätöksiä (kuvan ympyränmuotoiset 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 noodin kohdalle on myös merkitty p-arvo (tästä tarkemmin).

Oleellista tässä esitetyssä päätöksentekopuussa on, että vaikka itse luokittelu ryhmiin tehtiin puhtaasti artikkelin kirjoittajien tulkin 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 yksinkertaista rakenteellisen mikrotason muuttujaa: verbin persoonaa, aikamuotoa ja indikaattorisanan dependenssiroolia. Taulukossa 2 on tiivistetty näiden muuttujien vaikutus topiikkia asettavan virkkeen päätymiseen:

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 2: Kvantitatiivisesti pääteltävissä olevat ryhmät

Taulukossa 2 on esitetty... Lisäksi...

Huomaa: sijainnilla ei tässä puussa merkitystä..

4.3 Kirjoittajan näkökulma

“Lingvististen piirteiden tanssi”

Tänne Kirsin ajatus kolmijaosta (writer oriented etc.)

5 Päätelmät

1. Kirjoittamisen opetuksen kannalta ajateltuna. Työkaluja nimenomaan sen listaamiseen, *miten on tehty*. HUOM>> oppilaalle itselleen arvioitavaksi, mikä sopii tekstilajiin ja omaan tyyliin ja kulloinkin käsiteltävään aiheeseen.

Tähän analyysia siitä, miksi tietyt rakenteet vallitsevia. Kertooko esim siitä, ettei tietoisesti rakenneta, vaan vedetään assosiatiivisesti?

– Mitä hyötyä sitten siitä, että voi jostain kuulla erilaisista tavoista aloittaa kappale?

2. Konkreettinen sovellus fraasipankissa tms.

Lähteet

- Biber, Douglas, Ulla Connor & Thomas A. Upton 2007. *Discourse on the move: Using corpus analysis to describe discourse structure* Vsk. 28. John Benjamins Publishing.
- Chafe, Wallace 1987. Cognitive constraints on information flow. Teoksessa: Russell S. Tomlin (ed.) *Coherence and Grounding in Discourse : Outcome of a Symposium, Eugene, Oregon, June 1984*. John Benjamins Publishing Company.
- Chafe, Wallace 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.
- Dik, Simon C. 1989. *The Theory of Functional Grammar, Part I: The Structure of the Clause, Vol. 9 of Functional Grammar Series*. Dordrecht: Foris Publications.
- Guijarro, A. Jesús Moya 2001. Topicality chains in two discourse genres. *Estudios Ingleses de la Universidad Complutense* 9: 103–128.
- 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.
- Hothorn, Torsten, Kurt Hornik & Achim Zeileis 2006. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 15:3, 651–674.
- Ketokivi, Mikko 2015. *Tilastollinen päättely ja tieteellinen argumentointi* 2. laaj. laitos. Helsinki: Gaudeamus.
- Komppa, Johanna 2012. Retorisen rakenteen teoria suomi toisena kielenä-ylioppilaskokeen kirjoittelman kokonaisrakenteen ja kappalejaon tarkastelussa.
- Lambrecht, Knud 1996. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge studies in linguistics vol. 71. Cambridge: Cambridge university press.
- Levshina, Natalia 2015. *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company.
- Vuorijärvi, Aino 2013. Tekstilaji ja yhteisö: Ammattikorkeakoulun opinnäytetyön diskussio tekstinä.