

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

Kirsi Sandberg & Juho Härme

Contents

1 Johdanto	1
2 Tekstikappale ja sen topiikki	2
3 Tutkimusaineisto ja -metodi	3
4 Tulokset	6
4.1 Laadullinen analyysi	6
4.2 Määrällinen analyysi	8
5 Päätelmät	11

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 siis edustavan samaa tekstilajia (tekstilajin määritelmästä ks. esim. Honkanen 2012). Vaihtoraportti, 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. Monipuolisuus mainitaan usein myös arviointikriteeristöissä.

Lähestyessämme 2010-lukua Suomen oppilaitoksissa ja korkeakouluissa opiskelee kasvava joukko kirjoittajia (L1 ja L2), jolla ei ole opintojensa alkuvaiheessa opinnoissa menestymiseen tarvittavia tekstitaitoja. Kirjoitusohjeita ja -oppaita sekä suomenkielisiä mallitekstejä, ns. kielellisiä resursseja, löytyy verkosta nykyisin jo runsaasti, ja akateemisen kirjoittamisen (erikoiskielten) tarkastelun yhteydessä usein todetaan, että ”tekstilajitietoisuuden ja tekstilajituntemuksen lisääntyessä yksilön ilmaisupertuaari ja kielellinen valinnanvara laajenevat kattamaan myös yhteisön puhetavat (esim. Vuorijärvi 2013 s. 15; Bruce 2008 5-6, Solin 2006, 82 samoin Luukka 2002b, 14, 27)”. Sitä, millä tavalla ilmaisupertuaari konkreettisesti laajenee ja miten oppija pääsee tekstilajin kautta kiinni kielen rakenteiden varioiviin käyttötapoihin, ei Suomen kontekstissa ole kuitenkaan vielä juuri tutkittu. Funktionaalisista lähtökohdista tarkasteltuna yllä esitetty näkemys kaipaaisi myös empiiristä tutkimusta, koska rakenteiden hallinnan on osoitettu olevan kytköksissä tekstin laatuun (viite).

Tekstin kirjoittamisesta tekstikappaleiden tasolla saatavilla on enemmän aineistoa ja ohjeita siitä, miten se pitäisi tehdä, kuin deskriptiivistä tietoa todellisesta kielenkäytöstä (Komppa 2012: 121) (Chafe 1994: 296). Nykyisessä funktionaalista kielikäsitystä ja tutkivaa oppimista korostavassa pedagogisessa kehyksessä pääosin preskriptiivinen kuvaus tekstikonaisuuden ja -kappaleen rakentumisesta ei kuitenkaan 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).

Viime aikaisessa fennistisessä tutkimuksessa on tarkastelu varsinkin sitä, miten tekstikokonaisuus rakentuu funktionaalisten tai retoristen jaksojen pohjalta. On myös tarkennettu siihen, millä tavalla, eli millaisin kielen keinoin, kirjoittajat toteuttavat tekstin keskeisiä funktioita tai siirtoja (esim. Juvonen & kump. myös Hilkka). Näissä tarkasteluissa ei kuitenkaan puututa itse tekstikappaleen kirjoittamiseen eikä siihen, miten teksti temaattisesti rakentuu, vaikka tekstin temaattinen sisältö voi yhtälailla voi ohjata kirjoittajaa kirjoitusprosessin aikana. On nimittäin huomattavan paljon tekstilajeja (ks. Komppa & Vuorijärvi), joissa tekstin rakennetta voi kuvata topiikkipohjaiseksi. Näissä teksteissä 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.

Olemassa olevan tutkimustiedon rinnalle haluammekin tarkastella, miten kirjoittaja voi päästä alkuun kappaleen kirjoittamisessa, jos häntä ohjaa tieto tekstikappaleen topiikista, ts. siitä mitä kappaleen pitäisi pitää sisällään. Tutkimuksessa tarkasteltu 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 ja mitä rakenteita käyttäen kirjoittaja asettaa tekstikappaleen sisältötopiikkiin ja näin ollen miten kirjoittaja 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 roolista tekstikokonaisuudessa annetaan toistuvasti ohjeita oppikirjoissa ja kirjoittamisoppaissa. Paltridge (2002) on verrannut tutkielmaoppaita ja todellisten tutkielmien rakennetta ja toteaa, että yhtä mallia esitetään ja todellisuudessa variaatiota on huimasti. Voinee olettaa, että myös tekstikappaleiden tarkastelu tarjoaa kiinnostavaa aineistoa ohjeiden ja toteutusten välisestä todellisuudesta.

Tekstilingvistiksi tekstin kokonaisrakenteen analyysissä usein erotetaan tekstin funktionaaliset jaksot ja visuaaliset elementit, tekstikappale lukeutuu jälkimmäisiin. Näin niitä lähestytään lähinnä ortografisina merkkeinä, jotka eivät itsessään ole merkitseviä tekstin rakentumisen kannalta. Kirjoittajan kannalta kappalejako on kuitenkin nähdäksemme hyvin konkreettinen käytännön muovaama työkalu ja funktionaalinen kielenyksikkö. 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).

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). Tekstinosan ydin voi sijaita yhtä lailla ortografisen kappaleen alussa kuin muualla (Komppa,) ja kuten Juvonen sanoo, eri tekstityyppejä yhdistellen voi päästä samaan lopputulokseen, jolloin emme voi myöskään rakentaa mitään tiettyyn tekstilajiin soveltuvaa kaavaa. (Tuo pitää muotoilla tarkemmin, löperösti sanottu). Tekstikappaleen ensimmäinen

lause tai virke ei yksinään paljastakaan mitään siitä, millaisessa asemassa kappale hierakaisesti on tekstin kokonaisrakenteen kannalta. Kappaleiden välinen sidosteisuus, koheesio ja/tai koherenssi on kuitenkin saatava aikaan jollain tavalla, ja olemassa olevan tutkimuksen perustella voimme olettaa (kenties?) että tekstin sisältö saattaa olla jopa keskeisempi tekijä kuin tekstin kokonaisrakenne.

Tekstikappaleen ytimen tavoittelun sijaan kohdistammekin huomiomme siihen, miten kirjoittaja toteuttaa siirtymän aiheesta toiseen. Nimitämme siirtymän toteuttavia virkkeitä Juvosta (XX) mukaillen *topiikkia asettaviksi*. Tutkimuksemme kartoittaa eri tapoja asettaa topiikkia tarkastelemalla laajajhkosta aineistosta sitä, miten eri kirjoittavat ovat siirtyneet yhteen ja samaan topiikkiin.

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 esiteltyjä topiikkeja taas sellaiset tapaukset, jotka on jo tuotu aiemmin ja joihin nyt palataan. Lohkotopiikeilla viittamme tilanteeseen, jossa kirjoittaja käyttää hyväkseen johonkin edellä esiteltyyn topiikkiin kiinteästi liittyvää assosiaatioiden joukkoa viitatakseen entiteettiin, jota itseään ei vielä ole mainittu. Ajatus liittyy suoraan semiaktiivisuuden käsitteeseen Wallace Chafella (1987: 29), ja käytettävissä olemiseen Knud Lambrechtilla (1996: 99). Dikin (mts. 275) tästä antama esimerkki on lause *John gave a party last week, but the music was awful*, jossa *musiikki* on juhlien käsitteen avulla tekstiin tuotu lohkotopiikki. Toisaalta lohkotopiikissa voi nähdä yhteyden myös tekstin implisiittiseen teemaan (Shore 2008, ?).

Retorisen rakenteen teoria (ks. esim. Komppa ja jokin kv-lähde) rakentuu oletukselle, että tekstinosat ovat hierakaisessa suhteessa toisiinsa. Tästä huolimatta aineistopohjaisissa tutkimuksissa on paikallistettu huomattava joukko hierarkkisten tekstien rinnalla esiintyviä ns. moniytimisiä tekstejä, joiden osat ovat ennemmin rinnasteisia, keskenään tasavertaisia, eikä yhtä voida asettaa toisten ytimeksi (Komppa, ainakin s. 39). Vuorijärvi on päätenyt samanlaisiin löydöksiin siirtoanalyysin avulla. Vuorijärvi esittelee Paltridgeä (2002, 135) mukaillen topiikkipohjaisen rakenteen, jossa tekstin aihetta avataan ikään kuin osa kerrallaan. (Vuorijärvi. 75). Tämä topiikkipohjainen skeema kuvaa hyvin aineistonamme olevien matkaraporttien tekstilajia, ja käytämmekin topiikin käsitettä mukaillen mainittuja lähteitä. Toisin sanoen puhumme sisältötopiikista eli aiheesta tai aihepiiristä, josta tekstissä sanotaan jotakin.

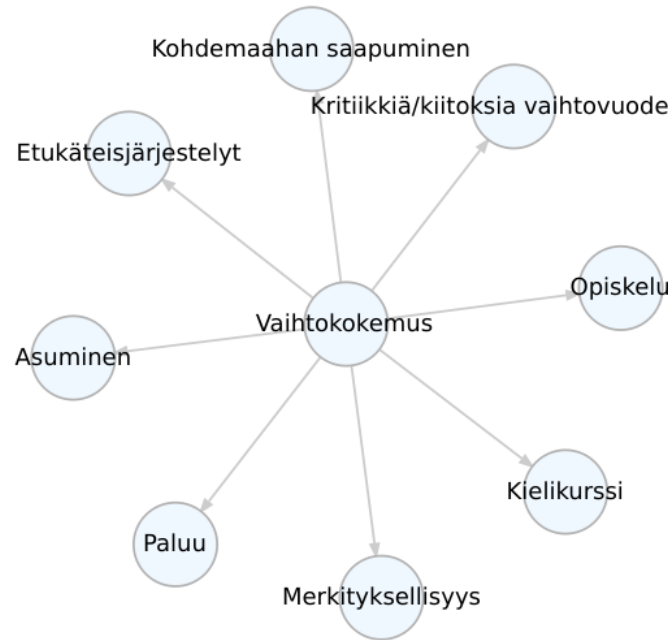
Johdanto, Topiikki 1, Topiikki 2, Topiikki 3 Päätelmä

Yliopiston antama kirjoitusohje kevyesti ohjaa tällaiseen tekstin rakentamiseen ja on näin mukana rakentamassa tutkimuksessa tarkasteltua tekstilajia. Jokainen tekstikappale aloittaa uuden tekstikokonaisuuden, joka ei niinkään asetu hierakiseen suhteeseen edeltävien tekstikappaleiden kanssa, vaan ennemminkin on joko sekventiaalisesti tai listamaisesti jäsentynyt tekstin diskiurssitopiikin alle. Ja tästä sitten voisimme nostaa yleiselle kirjoitustaiton tasoa kuvaavalle scenelle, jossa puhutaan kirjoittajan näkökulmista. Mutta tämä sitten, kun olen saanut sen osion valmiiksi.

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 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 (satunnaisotantaan 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:

¹Käyttäen kirjoittajien itsensä tekemää kappaleiden ortografista erottelua.

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 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 jäsennin on merkinnyt tutkimukseen valikoituneiden kappaleiden ensimmäisille *asua*-johdoksille, saadaan taulukon 2 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 2: 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).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 kappalejakoja – kirjoittajan omaa näkemystä

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

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...

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 puhtaan laadullisin perustein pohjaamalla tapausten keskinäiseen samankaltaisuuteen, tulkitaan osiossa 4.2. analyysin tulosta eli itse jaottelua määrällisten ominaisuuksien pohjalta.

4.1 Laadullinen analyysi

Taulukkoon 1 olemme koonneet edustavat esimerkit kustakin alaryhmästä. Niiden rinnalla on hyvin tiivis kuvaus topiikin asettamisen tavoista ja lisäksi muutama sana juuri kyseisen esimerkkivirkkeen kontekstista. Rajallisesta sivumäärästä johtuen syvennymme seuraavassa tarkemmin vain ryhmiin x, x, x, jotka osoittautuivat kirjoittamisen näkökulmasta erityisen kiinnostaviksi. Tarkemman käsittelyn tässä saavat esimerkit on taulukossa merkitty bbbb-merkillä.

Topiikkia asettava virke	Konteksti	Siirtymän toteutus
<i>Asuntoni sain vuokrattua Apartiksen(paikallinen TOAS) kautta.</i>	Käsittelee ennakkovalmisteluja yms. Lopuksi reflektoi osaamistaan kielikurssilla.	Alatopiikilla, vaikka ylätopiikkia ei ilmipantu.
<i>Vaihtoni aikana asuin Grandmontin asuntolassa.</i>	Kertoo, missä asui vaihtonsa aikana; kuvailee sisustuksen, kämppekaverit ja vuokran.	Esittelytopiikilla. Predikaattina on asua-verbi. Kertoo, missä kirjoittaja on asunut.
<i>Olin löytänyt asunnon ensimmäiseksi kuukaudeksi Airbnb:n kautta ennen Italiaan saapumista.</i>	Kertoo, mitä tapahtunut ennen kuvattavaa jaksoa. Kuvailee tarkemmin sijaintia, kulkuyhteyksiä yms.	Esittelytopiikilla, mutta aikarakenteltaan monitasoisempi kuin ryhmä 3.
<i>Asuminen, tai siis lähinnä asunnon hankkiminen, oli varmaankin vaihdon hankalin juttu.</i>	Arvioi edeltävässä kurssitarjonnan tasoa, paperiasioiden hoitumista yms.	Diskurssista johdetulla alatopiikilla, jota indikoi usein deverbaisubstantiivi. Luonnehtii kirjoittajan subjektiivista kokemusta.
<i>Asuntoloista valitsin Julius Raab Heimin vanhojen matkakertomusten perusteella.</i>	Kuvannut vaihtoon hakua ja kotimaan järjestelyjä. Maininnut jo paikallisista asuntolavaihtoehtoista.	Edeltävässä kappaleessa annetun tai hiukan kauempaa palautetun topiikin avulla. Kuuluvat toissijaisesti johonkin toiseen alaryhmään.
<i>Ulkomaisille opiskelijoille tarkoitettu asuntola on vain kolme vuotta vanha ja täten melko moderni.</i>	Edellinen kappale lennosta ja saapumisesta.	Muistuttavat 2. ryhmän tapauksia, kytkös implisiittisempi, usein väliotsikko. Luonnehtii asuntoa / asuntolaa konkreettisesti.

Topiikkia asettava virke	Konteksti	Siirtymän toteutus
<i>Itselläni kävi tuuri asunnon suhteen löytäessäni edullisen huoneen Chelseasta aivan Manhattanin keskeltä. Asuminen Tokiossa on aika kallista.</i>	Tekstin viimeinen kappale. Edellä arvioinut kokemustaan kokonaisuutena.	Alatopiikin avulla. Muistuttaa 1. ryhmää, mutta sisältönä verrataan omia kokemuksia toisten kokemuksiin.
<i>Suurlähetystön henkilökunta tarjosi resurssiensa puitteissa apuaan mm. maahantuloon, käytännön järjestelyihin sekä asunnon löytämiseen liittyvissä asioissa.</i>	Kappale melkein tekstin lopussa. Edellä kuvannut vapaa-ajan mahdollisuuksia. Edellä kuvannut hakuprosessia. Jatkaa vuokratasoon ja hintoihin.	4. ryhmän tapaan, mutta jakaen maailmantietoa, ei eksplikoiden omia kokemuksia. Esittelytopiikin avulla, mutta kirjoittaja ei ole agenttiivinen toimija. Fokuksessa joku muu keskeinen toimija.
<i>Lensin Wieniin sunnuntaina ja lento-kentältä suuntasin suoraan asunnolleni.</i>	Tekstin alkupuolella – edellä vain hakuprosessin kuvaus.	edellisiä implisiitisempi, kuitenkin uuden topiikin avulla. Kertova tekstityyppi ja aikamuoto yleensä imperfekti. Narraatio.
<i>Sitten pari sanaa asumisesta.</i>	Edellä yksityiskohtaisesti esim. seminaarien opintopisteistä.	metatekstin kautta. Uusi topiikki, joka tuodaan tekstiin eksplisiittisesti.
<i>Saapuessani Varsovaan minulla oli onneksi jo asunto tiedossa.</i>	Edellä kielenopiskelusta, jatkaa hintoihin.	uuden topiikin avulla, joka avausvirkkeen loppupuolella. Ajallinen jännite, katsotaan kuvastusta aikapisteestä taaksepäin.
<i>Lähes kaikki Bratislavan kauppa- ja korkeakoulun vaihto-opiskelijat asuivat Ekonom-nimisessä dormitoryssa.</i>	Kuvannut saapumista ja tutor-toimintaa.	7. ryhmän tapaan, mutta samaistuen, ei kontrastoiden muiden vaihto-opiskelijoiden toimintaan. Alaotsikko tyyppinen.

Taulukko 1: Esimerkit laadullisen analyysin tuloksena syntyneistä ryhmistä

Aineiston isoimman ryhmän muodostavat esimerkin 1 kaltaiset topiikkia asettavat virkkeet. Topiikki-indikaattorina on tässä *asuntoni*-sana. Se on siis tekstin ensimmäinen eksplisiittinen maininta topiikista; asumiseen tai asuntoon ei ole viitattu edeltävissä kappaleissa. Vaikka esimerkissä 1 ei täyty Dikin (1989: 275) kuvailema ehto siitä, että varsinainen ylätopiikki olisi erikseen esitelty lukijalle (vrt. myös Guijarro 2001: 110), *asuntoni* on tulkittu analyysissämme alatopiikiksi (ks. osio x.x edellä). Kirjoittaja nimittäin vaikuttaa operoineen tässä, kuten useimmissa muissakin ryhmän 1³ tapauksissa, yksinomaan tehtävänannon tarjoaman kysymyspatteriston ja saatavilla olevien mallitekstien puitteissa. Tästä seuraa, että asunto tai asuminen oletetaan lukijan käytettävissä olevaksi entiteetiksi (vrt. Chafe 1987: 29; Lambrecht 1996: 99), ja se on tuotu tekstiin ilman ennakkomainintoja tai esittelyjä. Voisi jopa sanoa, että esimerkin 1 kaltaiset tapaukset eivät niinkään aseta topiikkia, vaan ennemmin käyttävät jo asetetuksi oletettua topiikkia (vrt. implisiittinen teema Shore (2008)).

Seuraavan kiinnostavan ryhmän muodostavat esimerkin 2 kaltaiset tapaukset, joissa topiikki-indikaattorina on persoonamuotoinen *asua*-verbi. Virkkeiden pääasiallinen sisältö on yleensä informaatio siitä, missä kirjoittaja on vaihtonsa aikana asunut. Tapaukset ovat edellistä ryhmää selkeämmin johdantoja uuteen aiheeseen, ja Dikin jaottelun mukaisesti esimerkin 2 lopussa mainittu *asuntola* muodostaa (osiossa x.x. määritellyssä

³Viitaamme jatkossa ryhmiin tässä esitettyjen esimerkkien numeron perusteella.

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.

Kertominen ja kertova tekstityyppi ovatkin aineiston tekstilajille tyypillistä, mikä näkyy useammassakin alaryhmässä. Esimerkin 10 edustamassa ryhmässä asumisesta raportoidaan kertovan tekstityypin avulla, mikä käy annetusta esimerkistä esiin erityisen selvästi, koska kirjoittaja on käyttänyt ajan ilmausta. Aikamuoto ryhmän 10 virkkeissä on tyypillisesti imperfekti. 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.

Yhtälailla ryhmittelystämme on löydettävissä erilaisia tapoja aloittaa luonnehtimalla omia kokemuksiaan. Esimerkissä 4 topiikkia indikoi *asuminen*, mikä tähän ryhmään luokitelluille onkin varsin tyypillistä: jostakin varsinaisen diskurssitopiikin osa-alueelta tiivistetään oma kokemus kompleksiseen deverbali johdokseen, jota luonnehditaan predikatiivilauseella. Näissä tapauksissa kuvataan, miten *asunnon hakeminen oli helppoa* tai miten *kaikkein vaikeinta oli asunnon löytäminen*. Strategiana on itsestään käsin kertoa, miten asiat ovat sujuneet, omia onnistumisia ja epäonnistumisia arvioiden. Informaattiorakenteen 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.

[tähän seuraavaksi etäännytettyt kuvailut ja sitte lopetus linkkihommiin]

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. 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 välitotsikko.

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 Määrällinen analyysi

Tekstuaalisia funktioita ja kappaleen 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*).

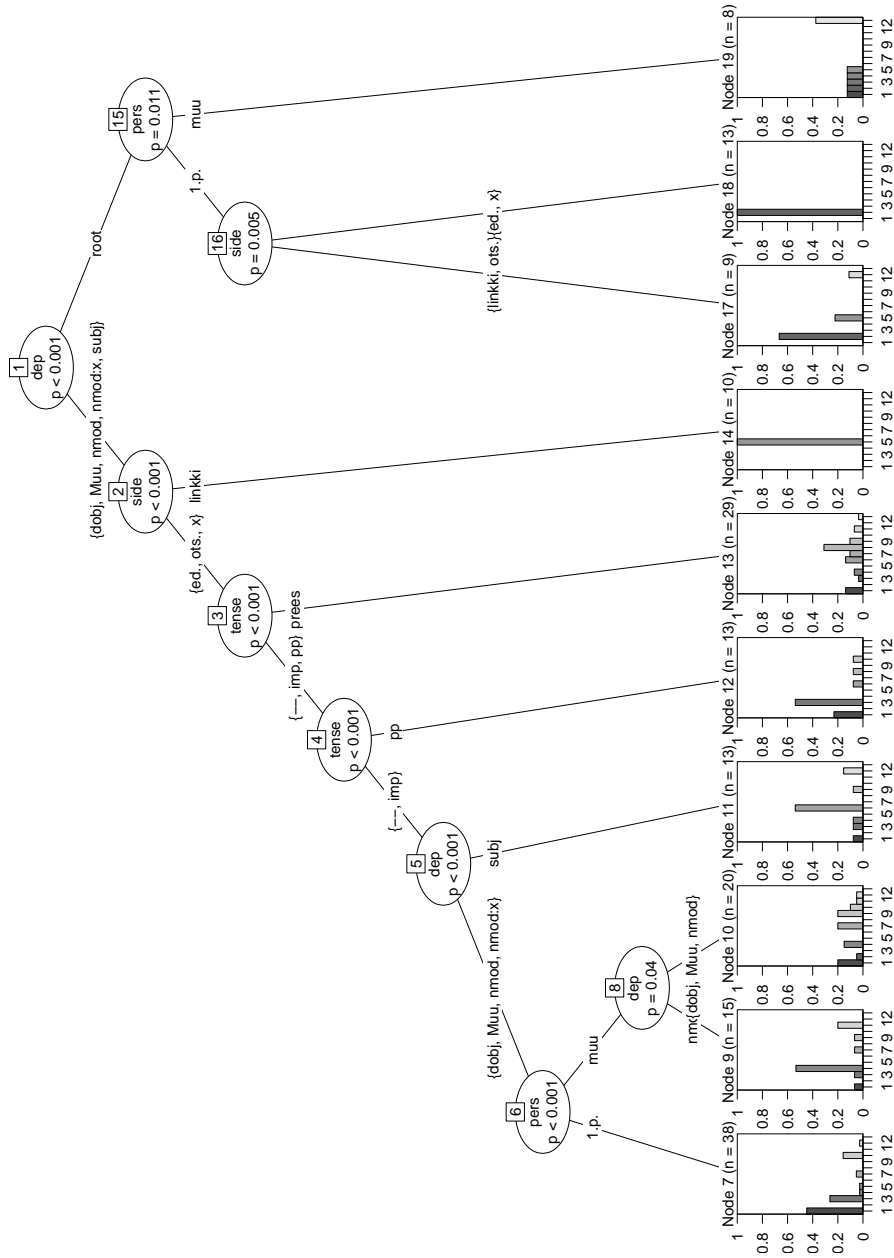
Ehdollisissa päätöksentekopuissa ajatuksena on, että (Levshina 2015, x) . Analyysissä määritellään yksi selitettävä muuttuja (vrt. Ketokivi 2015, x) ja (tavallisesti useita) selittäviä muuttujia, joiden avulla jonkin tapauksen päätymistä johonkin luokkaan pyritään ennustamaan. Nyt käsillä olevassa tutkimuksessa 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 2 luetellut dependenssiroolit, joita kuitenkin lavenettiin 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* (-).

Näiden muuttujien perusteella rakennettiin päätöksentekopuu R-ohjelmiston *party*-kirjastoon kuuluvaa *ctree*-funktia (Hothorn ym. 2006) käyttäen. Analyysin tulos on esitetty kuviossa 3.

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



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 noodin kohdalle on myös merkitty p-arvo (tästä tarkemmin).

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 kolmea⁵ 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

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.

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

⁵Päätöksentekopuun tuottaneessa algoritmossa indikaattorin sijainti osoittautui hieman yllättäenkin merkityksettömäksi.

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.

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

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 kirjoitelman 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.

Shore, Susanna 2008. Lauseiden tekstuaalisesta jäsennyksestä. *Virittäjä* 112:1, 24.

Vuorijärvi, Aino 2013. Tekstilaji ja yhteisö: Ammattikorkeakoulun opinnäytetyön diskussio tekstinä.