

Sata ja yksi tapaa ilmaista topiikkia

Kirsi Sandberg & Juho Härme

Contents

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

1 Johdanto

Oppilaitoksissa ja korkeakouluissa opiskelee kasvava joukko kirjoittajia, jolla ei peruskoulun päättötodistuksesta huolimatta ole hallinnassaan opinnoissa menestymiseen tarvittavia tekstitaitoja ja riittävää repertuaaria kirjoitetulle ilmaisulle tyypillisiä rakenteita. Opiskeluun liittyvien tekstilajien ja ns. akateemisten genrejen hallinta tuottaa hankaluuksia sekä osalle suomalaisen peruskoulun suomi äidinkielenä -oppimäärän suorittaneita (???) että opintonsa S2-oppimäärän mukaisesti (lähdeviite s2-kartoitukseen) suorittaneita. Aino Vuorijärvi (???) kirjoittaaakin, että uusissa institutionaalisissa tilanteissa ja näihin liittyvien tekstilajikonventioiden haltuunotossa ”kielenkäytön sosiokulttuuristen käytänteiden tuntemusta, tekstilajin tajua” tarvitsevat yhtäläisesti syntyperäiset kielenkäyttäjät. Vuorijärvellä on tässä lähdeviitteitä, muista tarkastaa vielä muotoilu.

2010-luvulla kielten ja kirjoittamisen opetuksessa korostuu tekstilähtöisyys ja funktionaalinen tapa tarkastella kieltä (ks esim. opsit & parit ÄOL vuosikirjat, Kirsi lisää viitteitä tähän). Tässä kehyksessä perinteinen, Suomenkin kontekstissa pääosin preskriptiivinen kuvaus tekstikokonaisuuden ja -kappaleen rakentumisesta (Komppa 2012: 121) 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 hallussaan monilukutaitojen harjoittelun myötä ymmärrystä tekstilajista, mutta repertuaarissa ei ole sopivaa rakennetta, jolla pystyisi itse tuottamaan tavoitemuotoisen tekstin. On tosin kirjoitettu (ja tämän vielä tarkistan), että ”tekstilajitietoisuuden ja tekstilajituntemuksen lisääntyessä yksilön ilmaisurepertuaari ja kielellinen valinnanvara laajenevat kattamaan myös yhteisön puhettavat (esim. Bruce 2008 5-6, Solin 2006, 82 samoin Luukka 2002b, 14, 27)”. Mutta on olemassa myös tästä poikkeavia näkemyksiä.

Tutkimusperustaisen kirjoittamisenopetuksen tueksi tarvitaan siis nähdäksemme käyttöpohjaista, kuvailtavaa tutkimusta. Johanna Komppa tarkastelee suomi toisena kielenä -ylioppilaskirjoitusten kirjoitelmia tekstikokonaisuuksina, deskriptiivisestä näkökulmasta. Kansainväliseen kenttään verrattuna Suomessa ei koulukirjoittamista ole mainittavasti tutkittu (Komppa 2012: 9–10). Lisää tähän lähteeksi myös katsaus lukio-ikäisten kirjoittamisen tutkimukseen. Saatavilla on 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.

(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 alatopiikeihin. 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.

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 topiikkia. 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 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

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 enemmän 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.

Aineisto ja menetelmät

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

Miksi tämä aineisto? -Iso, keskenään yhtenäinen tekstiaineisto -Kirjoitettu johonkin funktioon -Ei ammatik kirjoittajia tai editoituja tekstejä, mutta tottuneita kirjoittajia -Tutkimuslupa hankittu

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 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:

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 jäsenennys sekä lemmatisointi hyödyntämällä Turun yliopistossa kehitettyä dependenssijäsenennintä (Haverinen ym. 2014) ja siihen liittyviä komponentteja.

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

	Frekvenssi	Selitys	Esimerkki	N
doobj	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

¹Käyttäen kirjoittajien itsensä tekemää kappaleiden ortografista erottelua.

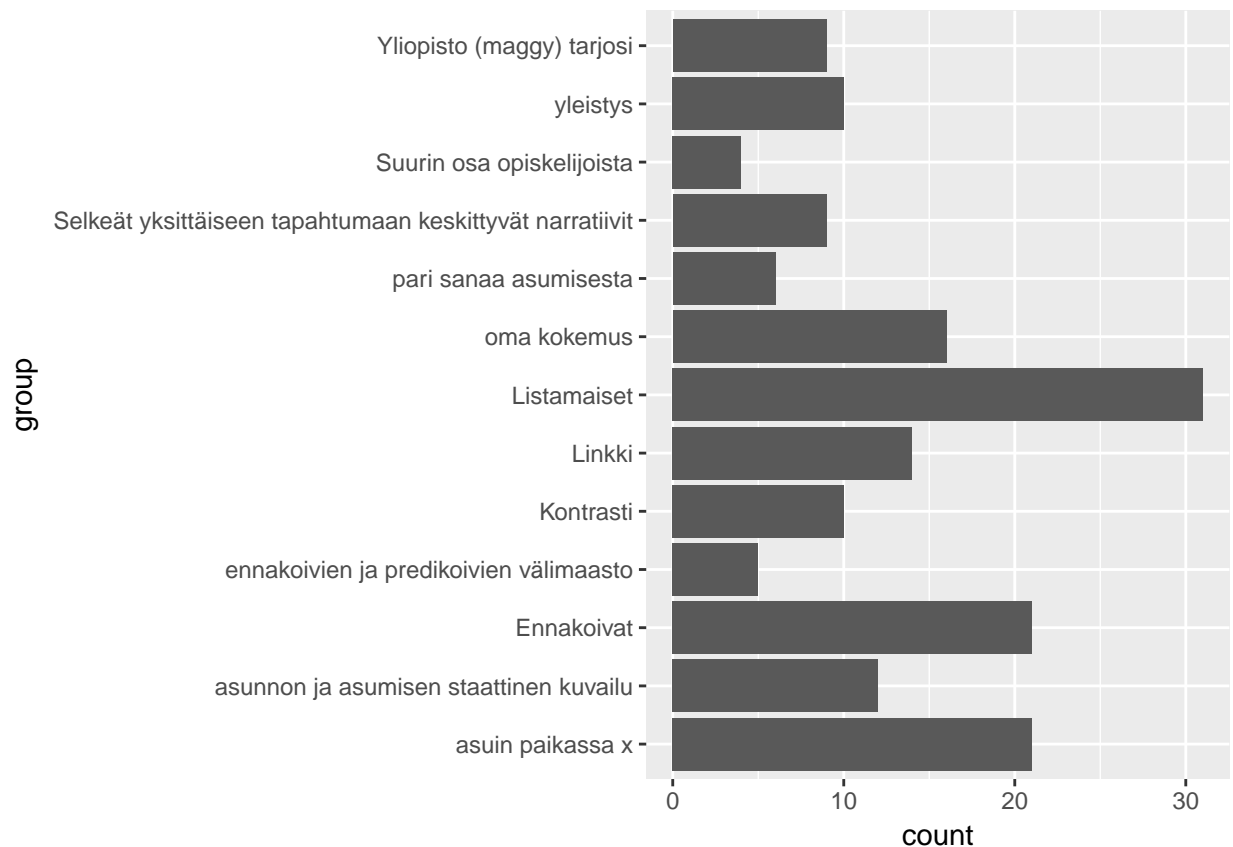
Taulukko 1: Asua-johdosten jakautuminen eri syntaktisiin kategorioihin.

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).² Nämä 168 tapauksesta otettiin varsinaisen laadullisen analyysin kohteeksi.

Aineistosta eroteltujen 168 tekstin laadullisen analyysin tavoitteeksi asetettiin tapausten ryhmitteleminen

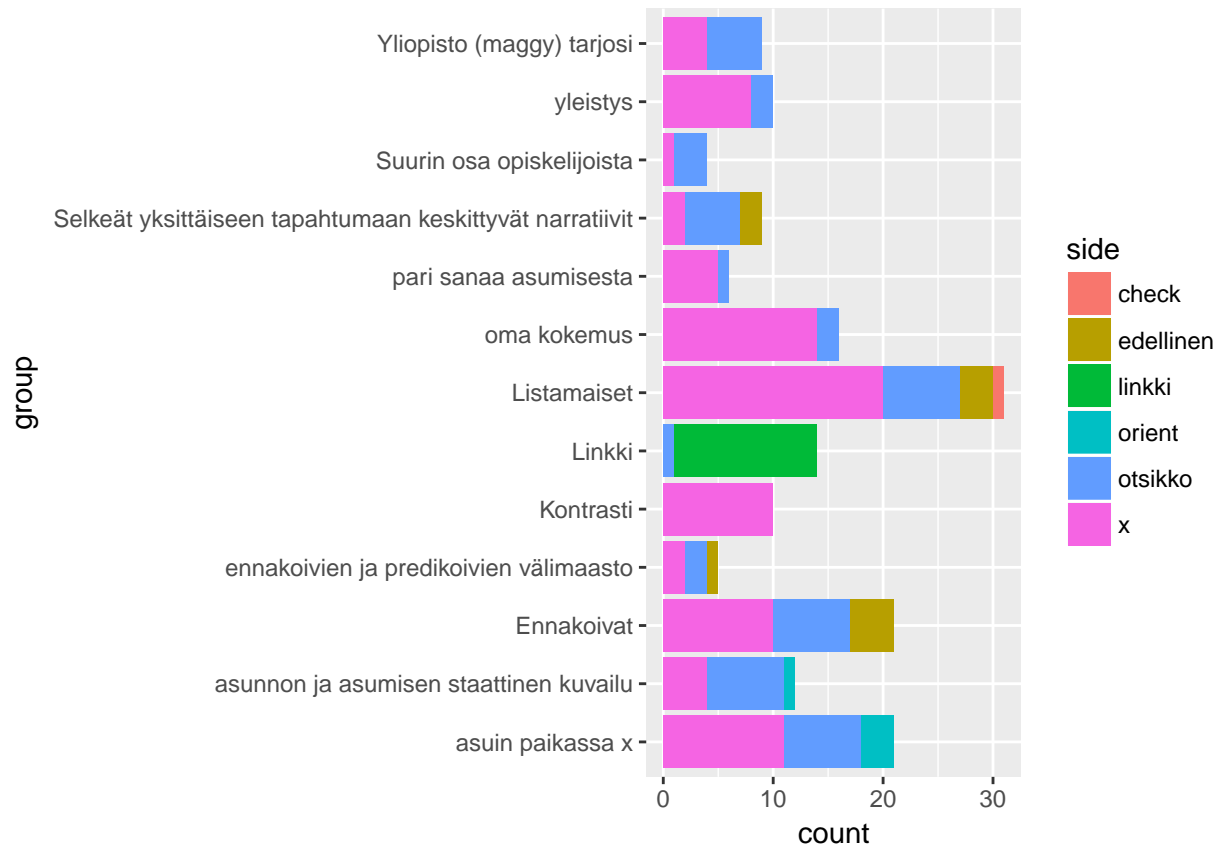
4 Yleiskuva

Tännne yleisiä havaintoja, esim.



Jos muuten rinnastelee vähän sitä, mitkä siteet linkittyvät mihinkin ryhmään:

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



5 Zoomaus

Tänne ajatus siitä, että otetaan yksi tai useampi esimerkkilausetta ja käydään läpi vaihtoehtoisia ilmaisutapoja

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 kirjoittelman kokonaisrakenteen ja kappalejaon tarkastelussa.

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