

# Sata ja yksi tapaa ilmaista topiikkia

*Kirsi Sandberg & Juho Härme*

## Contents

<b>1 Johdanto</b>	<b>1</b>
<b>2 Tutkimusaineisto ja -metodi</b>	<b>1</b>
<b>3 Yleiskuva</b>	<b>2</b>
<b>4 Zoomaus</b>	<b>3</b>
<b>5 Päätelmiä</b>	<b>3</b>
<b>Lähteet</b>	<b>3</b>

## 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. Osa heistä on suorittanut opintonsa S2-oppimäärän mukaisesti, mutta opiskeluun liittyvien tekstilajien ja ns. akateemisten genrejen hallinta tuottaa hankaluuksia myös suomalaisen peruskoulun suomi äidinkielenä -oppimäärän suorittaneille (ks. esim. Kuusela 2010). 2010-luvulla kielten ja kirjoittamisen opetuksessa korostuu tekstilähtöisyys ja funktionaalinen tapa tarkastella kieltä (ks. esim. opsit & muita lähteitä). Tässä kehyksessä perinteinen, Suomenkin kontekstissa pääosin preskriptiivinen kuvaus (Komppa 2012: 121) 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 miten eri tavoilla tämä on tehty.

Tutkimusperustaisen kirjoittamisenopetuksen tueksi tarvitaan käyttöpohjaista, kuvailevaa tutkimusta. Yksi keskeinen kriteeri mm. äidinkielen ylioppilaskokeen arvioinnissa on kappaleiden sujuva liittyminen toisiinsa (Ylioppilastutkintolautakunta 2017), mutta saatavilla on enemmän tietoa siitä, miten se pitäisi tehdä, kuin todellisesta kielenkäytöstä (Komppa sanoo tämän sivulla 121 ja viittaa (Chafe 1994: 296))

Tarve syntaksin ja tekstin rajapinnan analyysille (lähdeviite, eu-kartoitus) -> sovellusmahdollisuudet

## 2 Tutkimusaineisto ja -metodi

Tutkimuksen ensimmäisessä vaiheessa aineistoon valikoidut tekstit analysoitiin kappaleittain<sup>1</sup> siten, että jokaiselle kappaleelle merkittiin topiikki eli se, mistä kappale kertoo (vrt. Lambrechtin määritelmä). Yhteensä neljäkymmentä tekstiä<sup>2</sup> kattaneen kokeiluaineiston perusteella päätettiin keskittyä *asumisesta* kertoviin kappaleisiin siitä syystä, että asumista käsiteltiin lähes jokaisessa analysoidussa näytetekstissä ja että asuminen aiheena havaittiin tavallisesti melko selvärajaiseksi. Koska tekstejä analysoi kaksi eri henkilöä (kumpikin 20 tekstiä), näyteaineiston avulla testattiin myös, kuinka samanlaisia tulkinnat kappaleiden aiheista olivat. (tähän joku maininta niistä tuloksista).

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

<sup>2</sup>Tarkista määrä

Kun kaikki tekstit kattava varsinainen kappalekohtainen analyysi saatiin päätökseen, lopulliseksi aineistoksi muodostui 415 asumisesta kertovan kappaleen sisältävää tekstiä. Tekstit sinänsä ovat keskimäärin 945 sanaa pitkiä ja sisältävät kaiken kaikkiaan keskimäärin 13 kappaletta, joista vähintään yhden topiikiksi on edellä kuvatussa tutkimusvaiheessa analysoitu asuminen.

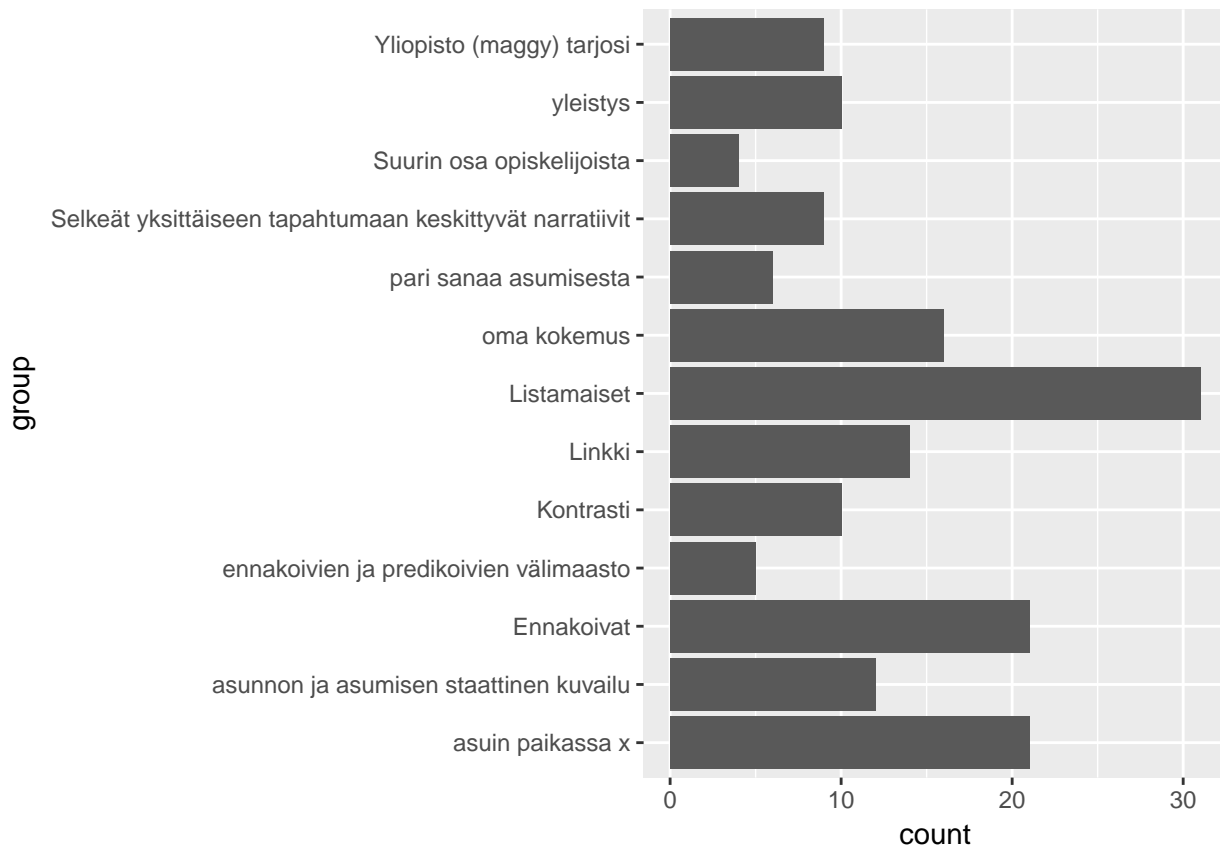
Asumisesta kertovien kappaleiden syntaktisesta rakenteesta saatiin yleiskuva annotoimalla kappaleet koneellisesti dependenssijäsenteimellä (ks. Haverinen ym. 2014). Koneellisesti tuotettu annotointi toimi lähtökohtana, kun kappaleita ryhdyttiin luokittelemaan eri ryhmiin sen mukaan, miten niissä indikoitiin asumistopiikkia. Luokittelu tapahtui seuraavien vaiheiden mukaisesti:

1. Määriteltiin, että kaikista potentiaalisesti asumista indikoivista sanoista otetaan mukaan ainoastaan varsinaiset asua-verbin johdokset, koska tämä oli selvimmin rajattavissa oleva kategoria. Tarkasteltaessa topiikki-indikaattoreita tarkasteltiin siis sitä, minkälaisissa rakenteissa kunkin kappaleen ensimmäinen asua-johdos (kuten verbi *asua* tai substantiivit *asuminen*, *asunto*, *asuntoasia*, *asuntomarkkinat*) esiintyi.
2. Jokaisesta kategoriasta valittiin tarkasteltavaksi puolet (tai vähintään x) Tätä kautta mukaan valikoitui eniten niitä, joita... yms. Tämä joukko analysoitiin siten, että mukaan otettiin vain tapaukset, joissa tulkintamme mukaan toteutettiin alatopiikin funktiota.
3. Kun jokaisesta syntaktisesta kategoriasta oli saatu kategorian suhteellisen koon mukainen (tai suurin mahdollinen) määrä alatopiikkitapauksia, nämä analysoitiin tarkemmin ja ryhmiteltiin eri tapauksiksi.

Analyysin varsinaiset tulokset kunkin syntaktisen kategorian osalta on kuvattu tarkemmin seuraavassa.

### 3 Yleiskuva

Tänne yleisiä havaintoja, esim.



## 4 Zoomaus

Tänne tarkempi zoomaus, etenkin linkki-ryhmästä ja lsitamaiset-ryhmästä

## 5 Päätelmiä

Ensin: edellä tehdyn analyysin perusteella saadut ryhmät kootusti

Sitten:

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

## Lähteet

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

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

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

Kuusela, Jorma 2010. Kun kirjoittaminen ei suju. Teoksessa: Elina Harjunen, Riitta Juvonen, Jorma Kuusela, Beatrice Silén, M Sääkslahti, & M Örnmark (eds.) *Miten peruskoululaiset kirjoittavat?* Vsk. 9. Opetushallitus. Saatavissa: [http://oph.fi/download/132345\\_Miten\\_peruskoululaiset\\_kirjoittavat.pdf](http://oph.fi/download/132345_Miten_peruskoululaiset_kirjoittavat.pdf) [viitattu 19.12.2017].