

Sata ja yksi tapaa ilmaista topiikkia

Kirsi Sandberg & Juho Härme

1 Johdanto

2 Otsikko

TEstiä

- lista 1
- lista 2

2.1 Otsikko 2

llkljlkj

2.1.1 Otsikko 3

3 Tutkimusaineisto ja -metodi

Tutkimuksen ensimmäisessä vaiheessa aineistoon valikoidut tekstit analysoitiin kappaleittain¹ siten, että jokaiselle kappaleelle merkittiin topiikki eli se, mistä kappale kertoo (vrt. Lambrechtin määritelmä). Yhteensä neljäkymmentä tekstiä² 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).

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 dependenssijäsentimellä (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 alatopiikkitaupauksia, nämä analysoitiin tarkemmin ja ryhmiteltiin eri tapauksiksi.

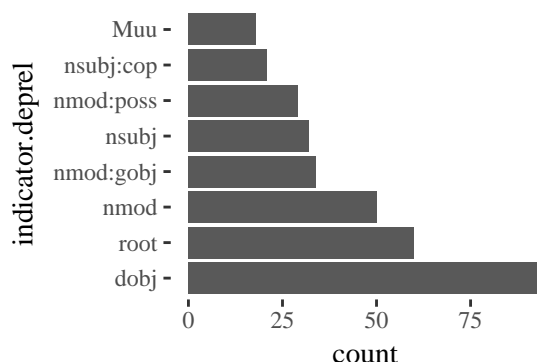
¹Käyttäen kirjoittajien itsensä tekemää kappaleiden ortografista erottelua.

²Tarkista määrä

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

4 Alatopiikkitapaukset syntaktisten kategorioiden mukaan ryhmiteltyinä

Asua-verbi ja sen johdokset esiintyvät tutkimusaineistossa tavallisimmin jo asumisesta kertovan kappaleen ensimmäisessä virkkeessä (81,45 % tapauksista). Juuri nämä ovat myös potentiaalisimpia kohteita alatopiikin esiintymisen kannalta, minkä takia edellisessä osiossa kuvattu satunnaisotanta päätettiin kohdistaa näihin tapauksiin. Lopputuloksena oli kaiken kaikkiaan 338 asumisesta kertovan kappaleen sisältävää tekstiä. Näissä itse asua-johdoksen syntaktinen kategoria eli jäsentimen sille asettama dependenssirooli vaihteli kuvion 1 mukaisesti:



Kuvio 1: Asua-johdosten jakautuminen eri syntaktisiin kategorioihin.

Kuvion 1 mukaan tavallisin asumista ilmaisevan topiikki-indikaattorin dependenssirooli on edellä kuvatulla tavalla rajatussa tutkimusaineistossa objekti (*dobj*). Jonkin verran muista kategorioista erottuvat myös predikaattiverbit (*root*) sekä tarkemmin määrittelemättömät adverbialit (*nmod*). Eri kategorioista tutkittaviksi valikoitujen tapausten määrä on esitetty taulukossa 1:

Taulukko 1: Syntaktiset kategoriat ja satunnaisotannan koko

Kategoria	Yht.	Otanta
dobj	94	47
root	60	30
nmod	50	25
nmod:gobj	34	17
nsubj	32	16
nmod:poss	29	14
nsubj:cop	21	10
Muu	18	9

4.1 Predikaattiverbit

Predikaattiverbitapaukset ovat harvoin..

4.2 Objektit

4.3 jne

5 Päätelmiä

- Kuinka monta eri tapaa oikeasti löydettiin?
- mihin tällaista voisi soveltaa?
- mitä voidaan sanoa tekstilajista / kirjoittajista / yms.?

Lähteet

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.