

## UURIMUSE OBJEKTIKS KOLLOKATSIOONID

**Liisi Piitsa. Sagedamate inimest tähistavate sõnade kollokatsioonid eesti keeles. (Dissertationes linguisticae Universitatis Tartuensis 23.) Tartu: Tartu Ülikooli Kirjastus, 2015. 162 lk. 5 lisa.**

Liisi Piitsa doktoritöö „Sagedamate inimest tähistavate sõnade kollokatsioonid eesti keeles” selgitab, kuidas kollokatsioonisuhetes avalduvad inimese sugu ja vanus. Uurimuse eesmärgiks on analüüsida sõna distributiivset käitumist ja semantilisi puutepunkte ehk seda,

millised on inimest tähistavate sõnade suhted teiste sõnadega ja milline on paradigmaatiliste suhete väljendumine süntagmaatilisel tasandil. Töö lähtekoht on sõnastatud juba 1950. aastate ameerika strukturalistide arusaamas, et sõna distributsioon peegeldab sõna tähendust.<sup>1</sup> Sellest on mitmed teisedki

---

<sup>1</sup> Z. S. Harris, Distributional structure. – Word. Journal of the Linguistic Circle of New York 1954, kd 10, nr 2–3, lk 146–162; Z. S. Harris, Co-occurrence and transformation in linguistic structure. – Language 1957, kd 33, nr 3, lk 283–340.

keeleuurijad aastate jooksul kirjutanud, ja Piits viitab eriti varaste saksa ja šveitsi lingvistide ning ameerika ja briti lingvistide uurimustele.

Piitsa väitekiri kuulub nende kollokatsioonide ja semantika uurimuste hulka, mida on viimase aastakümne jooksul avaldatud eesti keele uurimise valdkonnas. Piits ise on kirjutanud magistritöö „Sõnade *mees*, *naine* ja *inimene* esinemisümbrusest ja stiilivärvingust” (2004) ja pärast seda avaldanud mõned artiklid eesti sõnade *mees* ja *naine* kollokatsioonidest.<sup>2</sup> Teistel eesti lingvistidel on olnud erinevaid lähtekohti: Kadri Jaanits<sup>3</sup> ja Heinike Heinsoo<sup>4</sup> on uurinud kollokatsioone keeleõppe vaatenurgast, Heiki-Jaan Kaalep ja Kadri Muischnek arvutilingvistika seisukohast<sup>5</sup>, Kristel Uibo<sup>6</sup> huvi kollokatsioonide vastu on

seotud murdejoonte kaardistamisega ja Jelena Kallase<sup>7</sup> uurimusel on leksikograafiline aspekt.

Väitekiri koosneb neljast peatükist, kokkuvõttest ja ingliskeelsest resümeeist ning viiest lisast CD-l, mille tabelites on näha kollokatsioonirühmade kvantitatiivsed andmed valitud sõnade järgi. Töö struktuur on üldiselt tasakaalukas ja läbimõeldud. Esimene peatükk (lk 9–20) on lühikene sissejuhatus, kus autor käsitleb mõistete *kollokatsioon* ja *kollokatsioonikobar* eripära oma töö seisukohast. Sissejuhatus on minu arvates lühem kui väitekirjades enamasti ning esitleb varasemat uurimistööd ja meetodeid üsna napilt. Lugeja saab ikkagi teada, et Piits tunneb oma eriala ja selle varasemaid uurimissuundi, sest ta kirjeldab mitmekülgset uurimuse teaduslikku tausta ja varasemate uurijate töid teistes peatükkides analüüsi käigus. Piits tunneb hästi neid uurimusi, mis on tähtsad semantika ja kollokatsioonide analüüsi jaoks. Eriti meeldis mulle see, et ta kirjeldab ja kommenteerib seda, kuidas arusaamine kollokatsioonist on arenenud, ja tsiteerib ning refereerib ka vanemate uurijate töid. Ja mitte ainult ei refereeri, vaid käsitleb varasemas kirjanduses esitatud tulemusi ja järeldusi. Niimoodi tuleb välja, kuidas uurimus rajaneb eelmistel tulemustel ja arutelul, ja see, et meie „ued” avastused on olnud juba varem uurijate laual. Selline realistlikkus on hea tööle, mis ikkagi kasutab uusi meetodeid ja mahukamaid materjale, kui enne oli võimalik.

Teises peatükis „Inimest tähistavate sõnade tähendusest” (lk 21–44) vaatleb autor, kuidas eesti keele kesksete sõnaraamatute leksikograafid on analüüsinud ja klassifitseerinud sõnade *ema*, *inimene*, *isa*, *laps*, *mees*, *naine*, *poeg*,

<sup>7</sup> J. Kallas, Eesti keele sisusõnade süntagmaatilised suhted korpus- ja õppeleksikograafias. (Humanitaarteaduste dissertatsioonid 32.) Tallinn: Tallinna Ülikooli Kirjastus, 2013.

<sup>2</sup> L. Piits, Sõnade *mees* ja *naine* kollokatsioonide võrdlemise võimalusi eesti, saksa ja inglise keele korpustes. – Keel ja Arvuti. (Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 6.) Toim Mare Koit, Renate Pajusalu, Haldur Õim. Tartu, 2006, lk 105–114; L. Piits, Kas eesti *mees* on mees või inimene? – Ariadne Lõng: nais- ja meesuuringute ajakiri 2005/2006, nr 1/2, lk 3–10; L. Piits, Distributional Hypothesis: Words for 'human being' and their Estonian collocates. – *Trames: Journal of the Humanities and Social Sciences* 2013, kd 17, nr 2, lk 141–158.

<sup>3</sup> K. Jaanits, Leksikaalsetest kollokatsioonidest soome ja eesti keeles. Magistritöö Tartu Ülikooli läänemeresoome keelte õppetoolis. Tartu, 2004.

<sup>4</sup> H. Heinsoo, Adjektiivide tajumine ja õpetamine. – Lähivõrdlusi. Lähivertailuja 2010, nr 19, lk 120–135.

<sup>5</sup> H-J. Kaalep, K. Muischnek, Eesti keele püsiühendid arvutilingvistikas: miks ja kuidas. – Eesti Rakenduslingvistika Ühingu aastaraamat 2009, nr 5, lk 157–172.

<sup>6</sup> K. Uibo, Statistilised meetodid murdekorpuse ühendverbide tuvastamisel. – Eesti Rakenduslingvistika Ühingu aastaraamat 2010, nr 6, lk 307–326; K. Uibo, Kollokatsioonilised meetodid ja konstruktsioonilise varieerumise tuvastamine. – Eesti ja soome-ugri keeleteaduse ajakiri. *Journal of Estonian and Finno-Ugric Linguistics* 2013, kd 4, nr 1, lk 185–204.

*poiss, tüdruk* ja *tütar* tähenduste esitust sõnaraamatutes ja tesaurus, ning teeb valiku peamistest põhitähendustest, võttes need kollokatsiooniseoste korpusepõhisel analüüsimisel võrdlusaluseks. Ülevaade antakse „Eesti keele seletava sõnaraamatu”, „Eesti keele põhisõnavara sõnastiku” ja Eesti Wordneti põhjal, mis on kindlasti olulised allikad selle töö jaoks. Kui on küsimus semantikast, on sõnaraamatutes esitatud tähenduste klassifitseerimine oluline lähtekoht, kuigi peab alati küsima, kuidas lähendada leksikograafiat ja korpuslingvistikat. See küsimus jääb Piitsa töös tagaplaanile.

Väitekirja 3. ja 4. peatükis kirjeldab Piits selgelt, milliste meetodite ja korpustega on tulemused saadud. Analüüsis kasutatud materjal on kogutud Tartu Ülikooli eesti keele koondkorpuse ajakirjandustekstide korpusest, milles oli viis miljonit tekstisõna aastatest 1990–2001. Kui uuritakse erinevate kollokatsioonide kvantitatiivset esinemist, mis on Piitsa töö eesmärgiks, on vaja kasutada suuri korpusi. Kõigest tuleb välja, et Piits tunneb oma materjali ja meetodeid hästi ning hindab ka nende piiranguid.

Kolmandas peatükis „Sagedamate inimest tähistavate sõnade kollokatsioonikobarate võrdlus” (lk 45–69) kirjeldab Piits, kuidas inimest tähistavate sõnade tähendus kajastub sõna üldises distributiivses käitumises, ja otsib seost süntagmaatilise ja paradigmaatilise tasandi vahel. Peatükis analüüsitakse, kui suur osa valitud otsisõnade vasakule ja paremale hargnevatest kollokaatidest kattusid, võrreldes sõnade 30 sagedamat kollokaati. Osa tulemustest ei ole üllatav, näiteks see, et adjektiivid moodustasid vasakule hargnevatest kollokaatidest 17 %, paremale jäi vaid 2 %, või samuti see, et verbid moodustasid vasakpoolsetest kollokaatidest 24 % ja parempoolsetest 43 %, on eesti lugejale selge, sest ta oskab intuitiivselt oma

emakeele süntaksit. Lugeja võib küsida, mida uut see analüüs välja toob. Kõige tähtsamad ei olegi täpsed arvud, vaid see, kuidas sellise meetodiga ja laiema korpusega saab kindlustada ka oodatavaid nähtusi. Sõnade süntagmaatilisest kontekstist võib teha järeldusi antonüümia kohta ja sõnade sarnasus avaldub nende kontekstide kokkulangevuses. Sellised tulemused on Piits saanud selle väiksema korpuse põhjal, aga ta pigem tahab näidata leksikogrammatiliste mallide abil, milline süntaktiline seos võib otsisõna ja kollokaadi vahel esineda. Selleks on olnud vaja kontrollida kolme hüpoteesi, mida autor teeb neljandas peatükis.

Neljas peatükk „Sagedamad inimest tähistavad sõnad kahes leksikogrammatilises konstruktsioonis” (lk 70–131) on minu arvates uurimuse tähtsaim ja huvitavaim osa, kus uuritakse, kuidas kajastuvad kollokatsiooniseostes mainitud sõnadega tähistatud isikute vanus ja sugu. Esitatakse kolm hüpoteesi: esiteks – inimest tähistavate sõnade kollokatsioonides avalduvad sugudevahelised erinevused, teiseks – nende sõnade kollokatsioonid peegeldavad võimusuhteid ühiskonnas, ja kolmandaks – kollokatsioonides avaldub otsisõnaga tähistatud tüüpilise isiku vanus. Analüüsi objektiks on adjektiivsed laiendid ja verbikonstruktsioonid. Analüüsis kasutatakse Sketch Engine'i tarkvara<sup>8</sup> ja Jelena Kallase loodud arvutigrammatikat.

Neljas peatükk sisaldab väga palju detaile ja tekst on vahel natuke monotoonne, kuigi on vältimatu kirjeldada detaile, et otsisõnade erinevused välja tuleksid. Ikkagi oleks olnud hea, kui ka

<sup>8</sup> Sketch Engine'i eesti mooduli sisendiks on u 203 miljoni sõna suurune Tartu Ülikooli eesti keele koondkorpus. Suurima osa moodustavad ajakirjandustekstid, aga korpused on ka ilukirjandustekste, seadustekste, jututubade ja foorumite tekste ning teadusartikleid. Tekstid on pärit ajavahemikust 1990–2008.

neid erinevusi, nagu adjektiivsete laien-  
dite liigitumist semantilistesse tüüpi-  
desse (VANUS, FÜÜSIS, PSÜÜHIKA, HINNANG,  
SOTSIAAL, PEREKOND) ja kollokaatverbide  
jagunemist gruppidesse (nt aktiivseid  
füüsilisi tegevusi tähistavad verbid, ag-  
ressiivseid tegevusi tähistavad verbid,  
emotsionaalset käitumist tähistavad  
verbid jne), oleks kirjeldatud mingi-  
suguste konstellatsioonidena, mitte loe-  
teluna teksti sees. Nii oleks kohe näha  
iga sõnapaari kollokatsioonide erine-  
vused ja neid oleks parem võrrelda kui  
teksti sees.

Piitsa analüüsid on hoolikalt tehtud.  
Autor kirjeldab valitud materjale, prog-  
ramme ja meetodeid korralikult. Ta os-  
kab siduda näiliselt erinevad peatükid  
nii, et järgmine põhineb eelmise pea-  
tüki tulemustel ja nähtuste analüüsid  
tugevdavad tervikut ja üldisi järeldusi.  
Mitmed tabelid ja joonised selgitavad  
enamasti hästi ja selgesti tulemusi ja on  
muidugi vältimatud. Piitsa tekst edeneb  
sujuvalt, juhul kui muukeelne lugeja os-  
kab seda piisavalt hinnata.

Lühikeses kokkuvõttes (lk 132–136)  
on nii kesksed tulemused kui ka järeldu-  
sed esitatud selgesti ja lugeja saab hea  
üldpildi. Arutelu tulemuste ja tulevas-  
te uurimuste vajaduse üle jääb tähele-  
panu alt kõrvale. Sellist arutelu oleks  
olnud vaja kas või selle pärast, et Eestis  
on kollokatsioone uuritud erinevatest  
lähtekohtadest ja eesti keele struktuur  
erineb paljudest teistest keeltest. Kuna  
Liisi Piitsa uurimus on rohkem kvanti-  
tatiivne kui kvalitatiivne, siis minu ar-  
vates seetõttu autor eriti ei arutle, mida  
tulemused ühiskonnast paljastavad.  
Keeleuurija on nendes kohtades üldse

väga ettevaatlik, aga kaalutlevaid järeldusi on võimalik teha. Muidugi on valik autori käes; deskriptiivne uurimus on ju võimalik ja küllaldane. Aga ootasin, et autor võtaks ajakirjanduskeele valikute suhtes natuke rohkem seisukohti, kui küsimus on inimese vanuse ja soo distri-  
butsiooni kirjeldamises ajalehtedes, mis on alati meie silmade ees ja mis avaldavad mõju inimeste tunnete, arvamuste ja mõtteviisidele. Teine küsimus on, mida peaks uurima selle töö tulemuste põhjal ehk mis oleks järgmine samm. Kas interdistsiplinaarne kvalitatiivne uurimus sotsioloogidega või veel sügavamale korpuslingvistika meetodisse minemine? Selliseid mõtteid Piitsa välja ei too, kuigi on tavaline, et sellel tasemel tuleb selle üle arutleda. Neist tahaks ka lugeja rohkem teada.

On alati hea, kui teaduslikus töös on kasutatud oma keelt oma keele nähtusi analüüsides. Kuna osa Piitsa tulemustest oleks huvipakkuv ka ulatuslikumale teadlaskonnale, oleks väga soovitatav, et autor avaldaks oma tulemused ka teistes keeltes. Kuigi ingliskeelne kokkuvõte „Collocations of the most frequent Estonian words for 'human being'” (lk 149–155) annab hästi teada, millised on Piitsa tulemused, oleks hea kirjeldada neid ka laiemalt, eriti selle pärast, et soome-ugri keelte sõnamuutmine ja süntaks on indoeuroopa keeltest erinev, nagu autor ise tihti rõhutab. See ettepanek on loomulikult autori tulevase tööprogrammi jaoks.

PIRKKO NUOLIJÄRVI