

ARVUTI- JA ÕPPELEKSIKOGRAAFIA ALANE VÄITEKIRI

Jelena Kallas. Eesti keele sisusõnade süntagmaatilised suhted korpus- ja õppeleksikograafias. Tallinna Ülikool, humanitaarteaduste dissertatsioonid 32. Tallinn 2013. 185 lk.

4. aprillil 2013 kaitses Jelena Kallas Tallinna Ülikooli humanitaarteaduste doktorinõukogus edukalt väitekirja „Eesti keele sisusõnade süntagmaatilised suhted korpus- ja õppeleksikograafias”. Töö juhendajateks olid Asta Õim ja Margit Langemets ning oponentideks Raili Pool ja Kadri Muischnek.

Doktoritöös on lõimitud kaks aktuaalset rakenduslingvistilist uurimissuunda: korpusleksikograafia ja võõrkeeleõppijatele mõeldud sõnastike koostamine.

Viimase kümnendi jooksul on keelekorpused, sh eesti keele korpused saavutanud piisava suuruse selleks, et toimida keelekirjelduste, sh leksikograafia adekvaatse alusmaterjalina. Samas on korpuste suurus muutnud vajalikuks nendest keelise info hankimise automatiseerimise. Kaitsitud dissertatsiooni teemavalik ongi sellest vaatepunktist aktuaalne ja vajalik, nimelt on autor leksikograafilise tarkvara Sketch Engine¹ formaalses raamistikus läbi analüüsinud eesti keele sisusõnade – substantiivide, adjektiivide, adverbide ja verbide – süntagmaatilised suhted ning töö rakendusliku tulemusena on valminud Sketch Engine'i eesti keele moodul.

Sketch Engine on leksikograafiline tarkvara, mis sõelub keelekorpusest välja päringusõna leksikograafiliselt relevantse info ja esitab selle sõnavisandi (ingl *word sketch*) kujul, mis peaks andma piisavalt infot ükskeelse sõnaraa-

matu sõnaartikli kirjutamiseks. Eesti keeles on seda tarkvara varem tutvustatud ka dissertandi osalusel kirjutatud artiklis.²

Dissertant oli seadnud endale eesmärgiks eesti keele substantiivide, adjektiivide, adverbide ja verbide süntagmaatiliste suhete korpus- ja õppeleksikograafilise käsitluse teoreetilise raamistiku loomise ning rakenduslikuks eesmärgiks süntagmaatiliste suhete korpuspõhise tuvastamise ja nende esituse õppeotstarbelises eesti keele sõnastikus.

Sellisena osutus töö teema väga laiaks: käsitleda kõigi sisusõnade süntagmaatilisi suhteid nii korpus- kui ka õppeleksikograafias. Ehk oleks tööle teoreetilisest aspektist kasuks tulnud materjali kitsendamine, näiteks piirdumine ainult ühe sõnaliigi süntagmaatiliste suhete analüüsiga, mis oleks võimaldanud enam süvitsiminekut. Kuid rakenduslikust küljest on tarkvara Sketch Engine eesti keele mooduli toimimiseks oluline, et reeglistik oleks loodud kõigi sisusõnade jaoks.

Töö põhiliseks teaduslikuks panuseks on teoreetilise raamistiku loomine eesti substantiivide, adjektiivide, adverbide ja verbide süntagmaatiliste suhete korpusleksikograafiliseks analüüsiks ning nende suhete esitamiseks õppeleksikograafias. Töö näitab ühtlasi, et leksikaalse üksuse poolt tingitud süntaktilisi mustreid (töös esitatud sõnavisandite kujul) saab tuvastada ka süntaktiliselt märgendamata korpusest, kasutades ainult morfoloogilist infot ja sõnajärge. See on ootuspärasem fikseerituma sõnajärgjega keelte puhul, kuid mõnevõrra üllatuslikult kehtib ka

¹ <http://www.sketchengine.co.uk> (7. XI 2013); A. Kilgarriff, P. Rychly, P. Smrz, D. Tugwell, The Sketch Engine. – EURALEX 2004 Proceedings, Lorient, France, lk 105–115.

² J. Kallas, M. Tuulik, M. Jürvis- te, Leksikograafilise tarkvara Sketch Engine eesti keele moodul. – ESUKA – JEFUL 2012, nr 3–2, lk 57–77.

suhteliselt vaba sõnajärjega eesti keele puhul. Esile tuleks tõsta väitekirja rakenduslikku väärtust – leksikograafidele töövahendi loomist.

Töö algab sissejuhatusega, mis annab ülevaate uurimuse taustast, olulisematest mõistetest, eesmärkidest, materjalist, meetodist ja töö ülesehitusest. Järgneb korpusleksikograafia teoreetilistele aspektidele pühendatud teine peatükk, kus käsitletakse põhjalikumalt ka eesti keele sõnavisandite grammatika koostamiseks lahendamist vajanud küsimusi.

Väitekirja mahukaim osa on kolmas peatükk, milles autor esitab nn sõnavisandite grammatika sisusõnade sõnaliikide kaupa, lähtudes olulisematest eesti keele grammatikakäsitlustest. Kogu peatükis on olulisel kohal kaasmoodustaja mõiste, mille all mõeldakse otsisõnaga süntagmaatilises suhtes olevat sõna, ning edaspidise käsitluse mõistmise seisukohalt on oluline see, et kirjeldatavas sõnavisandite grammatikas võib kaasmoodustajaks olla nii otsisõna laiend kui ka põhi. Kolmanda peatüki vahekokkuvõtetes esitatud eri sõnaliikide (substantiivi, verbide jt) sõnavisandite näited illustreerivad hästi väitekirjas esitatud reeglite abil sõnavisandite koostamise võimalust.

Verbide sõnavisandite juures pööratakse eraldi tähelepanu perifrastilistele verbidele, mille märgendamine ja ka sõnavisandite abil tuvastamine ei ole lihtne, kuna sisendkorpuse morfoloogiliste märgendite süsteemis afiksaaladverbe muudest adverbidest ei eristata; Jelena Kallase loodud reeglistikus on probleem lahendatud afiksaaladverbide leksikoni abil. Nõustuda tuleb siinkohal väitekirja autori seisukohaga, et perifrastilistele verbide tähendus ei pruugi keeleõppijale olla läbipaistev ja vähemalt õppesõnastikes tuleks need esitada eksplitsiitselt omaette üksusena.

Kolmanda peatüki viimane alaosa annab ülevaate sõnavisandite hinda-

misest. Hindamisel on kasutatud Adam Kilgarriffi välja töötatud metoodikat, mida on rakendatud ka nt jaapani, sloveeni, inglise ja taani keele sõnavisandite hindamisel. Hinnatud on programmi väljundi täpsust, mis näitab õigesti tuvastatud süntagmaatiliste suhete osakaalu kõigi tuvastatud suhete hulgast, kusjuures „õigesti tuvastatud” on defineeritud kui leksikograafiliselt relevantne, st leksikograafi hinnangul vajalik ja kasulik. Evalveerimise tulemustest selgub, et tarkvara Sketch Engine jaoks dissertandi koostatud eesti keele reeglite komplekti abil tuvastatud sõnavisandite täpsus on rangelt hinnates (kõigi seitsme hindaja hinnangud langetades kokku) 70 % ja hindajate enamuse arvamusel 78 %. See tulemus on parem kui Sketch Engine'i taani keele mooduli täpsus ja enam-vähem sama mis sloveeni ja inglise keele mooduli täpsus.

Väitekirja neljas peatükk esitab põhjaliku ülevaate süntagmaatiliste suhete esitusviisidest õppeleksikograafias üldiselt; viies peatükk on sisu poolest neljanda peatüki jätk, kirjeldades süntagmaatiliste suhete esitusviise eesti keele õppesõnastikes ja andmebaasides. Alustades ülevaatega eesti keele reksioonivalimikes ja -sõnastikes kasutatud süntagmaatiliste suhete esitusviisidest, jõuab autor väitekirja seisukohast väga olulise õppesõnastikuni – eesti keele põhisõnavara sõnastikuni. Just sedalaadi õppesõnastike koostamine ongi väitekirjas kirjeldatud sõnavisandite grammatika üks peamisi rakendusvaldkondi.

Kuna käsitletav põhisõnavara sõnastik on suunatud eelkõige A1-tasemelt A2- ja B1-tasemele areneda soovivatele eesti keele õppijatele, on igati asjakohane selles peatükis antud ülevaade eesti keele A2- ja B1-keeleoskustaseme sõnavara- ja grammatikapädevusest. Kuna eesti keele kohta puuduvad praegu uuringud, missuguseid leksikaalgram-

matilisi konstruktsioone eri keeleoskustasemel õppijad eesti keeles tegelikult kasutavad ja kuidas konstruktsioonid taseme kasvades muutuvad, saab lähtuda ainult üldistest keeleoskustaseme kirjeldustest ja olemasolevatest sõnavaraloenditest. Tasemekirjeldustele ja sõnavaraloenditele toetudes teeb autor relevantseid kokkuvõtteid sõnavara ning leksikogrammatiliste konstruktsioonide kohta, mis peaksid sisalduma A2- ja B1-tasemel õppijatele mõeldud õppesõnastikus. Väitekirjas kirjeldatakse nimetatud keeleoskustasemel õppijatele mõeldud põhisõnavara sõnastiku märksõnade loendi ning sõnaartiklite koostamise põhimõtteid arusaadavalt ja ammendavalt. Sõnastiku koostamispõhimõtted arvestavad keeleõppija vajadusi, väga oluline on, et sõnastikus arvestatakse sõnade koosinemise sagedust ning tuuakse välja teksti loomiseks vajalikud leksikogrammatilised konstruktsioonid.

Väitekirja kui terviku seisukohast oluline on viienda peatüki lõpuosas esitatud info selle kohta, kuidas väitekirjas kirjeldatud sõnavandideid on võimalik põhisõnavara sõnastiku koostamisel ära kasutada. Selles osas näitab autor kujukalt, kuidas saab uurimuse teoreetilisi tulemusi praktilises leksikograafia-töös rakendada ja millist kasu tõuseb sõnavandide grammatikast õppesõnastike koostamisel. Sõnavandide abil

kooruvad välja nii eesti keele sisusõnade reksioonistruktuurid kui ka tähendusjaotused. Autor näitab *saabuma*-verbi abil, kuidas korpuspõhiselt tuvastatud süntagmaatiliste suhete põhjal saab teha esialgseid järeldusi ka verbi argumentide semantiliste rollide kohta. Seejuures on autor korpusandmete kasutamise osas vajalikul määral kriitiline.

Sõnavandide grammatika on lisaks leksikograafiale kindlasti kasutatav ka eesti keele kui võõrkeele õppimiseks mõeldud õppematerjalide koostamisel. Sõnavandideid saaks tulevikus ära kasutada harjutusvara koostamisel, vähendades sel viisil õppematerjalide looja idiolekti mõju keeleülesannetele, nii nagu korpusete kasutamine vähendab leksikograafi keelekasutuse mõju sõnaraamatunäidetele.

Jelena Kallase doktoritööl on märkimisväärselt palju edasiarendamise võimalusi. Autor peab ise olulisimaks arengusuunaks töös esitatud põhimõtete järgi koostatud õppesõnastike sarja loomist. Sel viisil suureneks sõnastike roll keeleõppeprotsessis ning sõnastikud hakkaksid üle võtma keeleõpiku ning ka õpetaja funktsiooni. Jääme huviga ootama nii esimest põhisõnavara sõnastikku kui ka kõiki järgnevaid.

KADRI MUISCHNEK,
RAILI POOL