

EMAKEELE SELTSIS

Emakeele Seltsi sügiskoosolekud tutvustasid eesti keele elektroonilisi ressursse ning nende kasutamisevõimalusi. Esimene kõnekoosolek peeti 18. oktoobril 2007 Tartus.

Pille Esloni ettekande teemaks oli „Õppijakeele korpused ja keeleõpe”. Esineja andis ülevaate teiste maade õppijakeele korpustest ja nende kasutamisest ning tutvustas lähemalt Tallinna

Ülikooli vahekeele korpust. See on loodud uurimistöö ning eesti keele kui teise keele/võõrkeele õppe eesmärgil. Korpus on praegu vene emakeelega keeleõppijate kirjalike tekstide kogu, mida on kavas täiendada suulise kõne näidete ning samuti muulastest keeleõppijate keelenäidetega. Kõneleja rääkis lähemalt tekstide märgendamise ning sellega seotud vealiikidest.

Jaak Vilo ettekande pealkiri oli „Keeletehnoloogia viimine uurimistööst arvutikasutaja teenistusse”. Esineja kõneles keeletehnoloogia sihtprogrammist ning selle oodatavatest tulemustest. Keeletehnoloogia raames tuleb tegelda kõnetuvastuse, kõnesünteesi, masintõlke, inimese-masina dialoogisüsteemide, infootsingu lahendustega. Selleks on vaja mitmesuguseid alusuuringuid, tuleb koostada kirjaliku ja suulise keele korpuse jne, millest kõneleja ka ülevaate andis.

Kadri Muischnek rääkis teemal „Tänapäeva kirjaliku eesti keele korpused Tartu Ülikoolis”. Ta esitas ideaalina korpust kui polüfunktsionaalset elektrooniliselt talletatud tekstikogu, millesse kuuluvad tekstid on valitud eesmärgipäraselt, nii et nendest koosnev tervik annab tõepärase pildi kogu keelest, selle hetkeseisust või muutumisest. 1980. aastate korpuse koostamise põhimõtted tähendavad, et selline korpus (kirjalikud tekstid, proosa, täiskasvanute kirjutatud tekstid, lugemiseks kirjutatud tekstid jms) annab ülevaate eeskätt haritumate keeletekasutajate korrigeeritud keelest. Aastate 1890–1990 läbilõikekorpus on representatiivne sinna valitud tekstiklasside – ajalehetekstide ja ilukirjandustekstide – osas. Eesti keele koondkorpus sisaldab ajalehtede ja ajakirjade tekste, ilukirjandus-, seadus- ja teadustekste, Riigikogu stenogramme, jututubade tekste. Valmiv tasakaalus korpus sisaldab võrdses mahus ajalehe-, ilukirjandus- ja teaduskeelt ning võimaldab võr-

relda kirjaliku keele kolme tähtsamat tekstiklassi.

Margit Langemets andis ülevaate Eesti Keele Instituudi elektroonilisest keelevarast, selle mitmekülgsest ning rohketest olemasolevatest ja tulevastest kasutamisevõimalustest. Võimaliku esituskujuna nimetas ta e-teksti, digitaalseid helilinte, andmebaase, sõnastikusüsteeme jne. M. Langemets vaatles keelekihi elektroonilise keelevara kujunemist ja olukorda. Murdeid ja sugulaskeeli on helilindistatud, EKI veebileheküljel saab kasutada sõnastikke, leidub ka e-tekste. Keelenõu andmebaasi põhjal on sündinud avalik, laialt kasutatav keelenõuvakk, kättesaadavad on vajalikud keelehooldeallikad. Loodud on leksikograafi töökeskkonna tuum, sõnastike haldussüsteem EELEX.

Teine koosolek teemal „Eesti keele elektroonilised ressursid” peeti 29. novembril Tallinnas.

Asta Õim tutvustas eesti kõnekäändude ja fraseologismide andmebaasi, selle seisuga ja perspektiive. Internetipõhine andmebaas on loodud eri arhiivide materjalide alusel ning esindab põhiliselt vanema keele väljendeid. Andmebaas sisaldab 160 000 kõnekäänukirjet, millest väike osa on suure sagedusega (nt *nagu hane selga vesi*), suure hulga kohta on aga vaid üks üleskirjutus (nt *aeg luubis*). Päringuid saab esitada eri sõnade esinemise kohta, tüübi, sisu, aja jms järgi. See võimaldab näiteks jälgida ka väljendite tähenduse muutumist ajas.

Liina Lindström andis ülevaate eesti murrete korpusest. Eesti murrete korpus on sündinud Tartu Ülikooli ja Eesti Keele Instituudi koostöös ning selle eesmärgiks on teha murdematerjalid uurijale elektrooniliselt kättesaadavaks. Korpus koosneb murdelindistustest, foneetilises ja lihtsustatud transkriptsioonis murdetekstidest, morfoloogiliselt märgendatud tekstidest jne. Tekste on kõigest murretest. Korpusele

toetudes saab uurida sagedasi nähtusi, teksti kasutuskontekstis, võrrelda eri murdeid jne.

Andriela Rääbise ettekanne kandis pealkirja „Tartu Ülikooli eesti suulise keele korpus ja dialoogikorpus”. Suulise keele korpusesse on materjali kogutud 1997. aastast alates. Korpus on liigendatud kolmelt aluselt: argi- ja institutsionaalne suhtlus; dialoogid ja monoloogid; silmast silma, telefoni- ja meediasuhtlus. Korpusesse ei koguta kindlate situatsioonitüüpide materjali, enamik on audiolindistused, videosid on vähe. Tekstid on transkribeeritud, näidatud on mitmesuguseid suhtlushäälit-susi, kõne omadusi, pealerääkimisi jne. Eesti dialoogikorpus EDiC on loodud kahel eesmärgil: uurida inimestevahe-list suhtlust ning modelleerida inimese ja arvuti suhtlust.

Külli Habicht rääkis Tartu Ülikooli vana kirjakeele korpus(t)est. Vanast kirjakeelest on talletatud läbilõige pea-aegu 400-aastasest keele- ja kultuuri-loost, XVI sajandist XIX sajandi kesk-paigani. Vana kirjakeele eripära on teiste uurimisobjektidega võrreldes K. Habichti sõnutsi selle tõlgendamise keeru-kus ning ainese suur keeleline varieeruvus. Korpus pakub uurimisvõimalusi etümoloogidele, piiblitõlke ajaloo uurija-tele, grammatika- ja leksikauurijatele jne. Esiialgu on päringud võimalikud vaid märgendamata tekstist, märgenda-tud teksti päringusüsteem on loomisel.

ANNIKA HUSSAR