

## EESTI VAHEKEELE KORPUSE SÜGISSEMINAR

13. novembril 2008. aastal toimus Tallinna Ülikoolis juba kolmas eesti vahekeele korpuse töörühma korraldatud sügisseminar. Seekordne teema oli „Korpusuuringute metodoloogia ja märgendamise probleemid”. Nagu peakorraldaja Pille Eslon lootis, jagati nii Tartu ja Tallinna kui ka Oulu ja Turu uurijad kokku toonud seminaril asjatundlikke kogemusi ja nii mõnelgi teemal tekkis elav arutelu.

Seminari avas Reili Arguse (TLÜ) ettekanne, mis sisaldas ülevaadet üldkognitiivsetest, keeleomastest ja tehnilistest aspektidest, mida tuleb arvestada harva esinevate või suhteliselt hilja ja raskemini omandatavate keelenüansside uurimisel. Lähemalt käsitleti eesti keele aspektilisuse ja ajamorfoloogia omandamise uurimist kolme erineva katse abil.

Pille Eslon (TLÜ) rääkis, et korpusest tulenev keeleanalüüs pakub põhimõtteliselt uusi võimalusi kasutada kvalitatiivseid ja kvantitatiivseid analüüsimeetodeid ning integreerida saadud keelekirjeldusi ja teooriaid rakenduslikel eesmärkidel. Ettekandes tutvustati, millistele tulemustele on jõutud näiteks objektnoomeni käändekasutuse võrdlemisel eesti kirjakeele ja õppijakeele korpuste alusel jälgimaks eesti kirjakeeles toimuvat ning avastamaks selles varjatult kulgevaid muutusi.

Helena Metslang (TÜ, TLÜ) kirjeldas sõnajärje näitel katset analüüsida eesti vahekeele korpuse materjali TÜ ESTCG analüsaatori abil ning ühendada saadud tulemused vigade käsitsi-märgendusega.

Krista Liini (TÜ) grammatikakorrektori prototüüp tuvastab eestikeelsetes tekstides puuduvaid ja üleliigseid komasid. Vead leitakse kitsenduste grammatika reeglite abil, mis on koostatud ning mida on testitud interneti-

portaalidest kogutud väära komakasutusega kommentaaride korpuse põhjal.

Mark Fishel ja Heiki-Jaan Kaalep (TÜ) tutvustasid ideed luua spetsiaalne korpuste kombineerimise tarkvara, mis aitaks ühendada ühe ja sama korpuse eri versioone.

Ilmari Ivaska ja Kirsti Siitonen (Turu Ülikool) rääkisid Turu Ülikoolis loodavast korpusest, mis hõlmab edasi jõudnud soome keele õppijate keelenäiteid. Korpuse kumuleeruva märgendamise esimesel etapil keskendutakse morfoloogiale, seejärel alustatakse veatüüpide märgendamist.

Tiit Hennoste (TÜ) jagas suulise keele korpuse märgendamise kogemusi, rääkides eri kirjeldustasandeil sagedamini ette tulevatest probleemidest.

Annekatriin Kaivapalu (TLÜ) juhtis tähelepanu introspektiivsetele meetoditele, mis võivad aidata täiendada õppijakeele analüüsi sihtkeele produtseerimise seisukohast. Lähemalt oli juttu eesti vahekeele korpuse sõnajärjevigade psühholingvistilise reaalsuse väljaselgitamisest testi, nn valjusti mõtlemise ja intervjuude abil.

Ingrid Krall (TLÜ) polemiseeris esimese kursuse vene emakeelega üliõpilaste eestikeelsete kirjalike tekstide kogu põhjal eesti vahekeele korpuse ning Kadri Sõrmuse koostatud emakeeleõppija korpuse veaklassifikatsioonidega, rõhutades lingvistilise ja struktuuritaksonoomia vaieldamatut tähtsust.

Katre Õim (TLÜ, Eesti Kirjandusmuuseum) tutvustas moodust, kuidas käsitleda fraseologisme formaalsete tunnuste asemel mõistesüsteemide alusel. Näiteks eesti kõnekäändude ja fraseoloogismide andmebaasi tarbeks loodud mõistesüsteem aitab kokku viia semantiliselt seotud väljendeid, esile tuua väljendite tähenduse kujunemisjärke jm.

Jarmo Harri Jantunen (Oulu Ülikool) demonstreeris kahest või enamast

sõnast koosnevatele fraseoloogilistele üksustele keskendudes, kuidas kasutada selliste järjendite uurimisel korpuse sõnaloendeid. Ettekandes kirjeldati, mida uut on programmi WWW-Lemmie 2.0 abil teada saadud näiteks soome keele *kovin*-tüüpi määruste (*kovin paljon*, *kovin suuri*) kohta soome keeles.

Ettekannetevahelist keskustelu elavdasid Leo Võhandu tähelepanekud keelevälistelt erialadelt. Seminari materjalidega saab tutvuda aadressil <http://evkk.tlu.ee>.

KATRE ÕIM