

EUROOPA ARVUTILINGVISTIKA KONVERENTSIL

30. märtsist kuni 3. aprillini toimus Kreeka pealinnas Ateenas järjekordne, kaheteistkümnes Rahvusvahelise Arvutilingvistika Assotsiatsiooni Euroopa haru (EACL) konverents (vt <http://www.eacl2009.gr/conference/>). Selliseid konverentse korraldatakse alates 1983. aastast, enamasti üle aasta.

Eestist olid kohal peale allakirjutanute veel Tartu Ülikooli doktorandid Siiri Pärkson ja Kristel Uiboaed.

Kahel esimesel päeval toimusid töö- ja õpipajad. Tööpajasid oli kokku kümme, neist ettekannete ja osavõtjate arvu poolest olid suurimad statistilise masintõlke ja loomuliku keele genereerimise alane. Osalejate hulgas oli populaarne ka intrigeeriva nimega tööpaja „Lingvistika ja arvutilingvistika koostoime: kas tõhus, halb või mõttetu?” Üheskoos otsiti vastuseid näiteks järgmistele küsimustele: 1) mida on andnud teineteisele lingvistika ja arvutilingvistika? 2) mida on õppida lingvistika ja arvutilingvistika senisest koostööst? 3) millist kasu saavad lingvistika ja arvutilingvistika teineteise paremast tundmisest? 4) missugused lingvistika ja arvutilingvistika seni puutumata alad on koostööks küpsed?

Jõuti järeldusele, et arvutilingvistika keskne eesmärk on endiselt iga loomuliku keele jaoks taaskasutatava forma-

liseeritud keeleteadmuse loomine ning et selle eesmärgini liikumine toimub sammhaaval. Hans Uszkoreit (Saarimaa Ülikool) märkis muuhulgas, et paljudes rakendustes kasutatakse juba mitmesuguseid grammatikaressursse, nt HPSG ERG (inglise keele ressursigrammatika, mis põhineb fraasistruktuurigrammatikal), LFG ParGram (mitme erineva keele paralleelne grammatika, mis põhineb leksikaal-funktsionaalsel grammatikal) ja inglise keele CCG (kombinatoorne kategoriaalne grammatika). Need rakendused on seni küll veel eksperimentaalsed, aga kui lingvistiline töötlus muutub efektiivsemaks, spetsiifilisemaks (suudab teha valikuid erinevate võimalike tõlgenduste hulgas), tõrkekindlamaks ja kiiremaks, siis oleme varsti ka esimeste tegelikus elus kasutatavate rakenduste tunnistajateks.

Õpipajasid toimus kokku neli: 1) „Kit-sendatud tingimuslikud mudelid loomuliku keele töötlusel”, 2) „Statistiline loomuliku keele töötlus infootsingus”, 3) „Hüvitusega masinõpe adaptiivsetes dialoogsüsteemides”, 4) „Kombinatoorne kategoriaalne grammatika loomuliku keele töötlusel”.

Põhikonverents algas kutsutud ettekandega, milles Ann Copestake (Cambridge'i Ülikool) põhjendas, miks loo-

muliku keele mõistmise modelleerimisel on asjakohane piirduda vähemalt esialgu pindmise semantilise analüüsiga.

Seejärel algasid ettekanded neljas paralleelsektsioonis. Nagu märkis konverentsi avamisel programmitöö esimees, Uppsala Ülikooli professor Joakim Nivre, saabus konverentsile kokku 360 ettekannet, millest vastu võeti 100 (28%). Kõige enam ettekandeid esitati (ja ka aktsepteeriti) kolmes valdkonnas: statistiline masintõlge, infootsing ja keeleressursid. See näitab ühtlasi, kuhu on praegu keskendunud uurimistöö. Ettekannetest oli pool suulisi ja pool stendiettekandeid.

Konverentsi teisel päeval leidsid teiste hulgas aset ka spetsiaalsed doktorantide sessioonid, kus sai kuulata kokku 12 ettekannet valmimisjärgus doktoritööde kohta. Igale tööle oli eelnevalt määratud kaks retsensenti. Doktorandile anti 20 min oma töö tutvustamiseks ja seejärel esinesid retsensendid, kes töid esile töö vored ja puudused, andes seega ettekandjale väärtuslikku tagasisidet.

Kolmas päev algas jälle kutsutud ettekandega: Franciska de Jong (Twente Ülikool) esines teemal „Loomuliku keele automaattõõtlus ja humanitaaria: vana sideme taaselustamine”. Ta rõhutas, et humanitaarteadused saavad aidata arvutilingvistikal püstitada ja

saavutada innovatiivseid uurimiseesmärke, nende hulgas on nii semantiline analüüs, multimodaalse info töõtlus, loomulikus keeles suhtlemine kui ka valminud süsteemide hindamine. Seejärel toimus demosessioon, kus esitleti 17 arvutiprogrammi mitmest arvutilingvistika valdkonnast: sisukokkuvõtete tegemine, mitmesõnaliste väljendite ekstraheerimine, loomulikus keeles suhtlemine, keelemängud jne. Allakirjutatud esitlesid koos Koblenz-Landau ülikooli arvutilingvistika professori Karin Harbuschiga programmi Elleipo, mis suudab nii saksa kui ka eesti keeles genereerida rindlausetest elliptilisi lauseid, kasutades mõlema keele puhul ühtesid ja samu reegleid korduvate lauseliikmete väljajätmiseks.

Kogu konverentsi jooksul said osavõtjad tutvuda näitusega, kus oma tegevust tutvustasid asutused ja firmad: Kreeka teadusuuringute keskus Demokritos, Ateena tehnoloogiakeskus, tehisintellekti ühing, aga ka Google.

Järgmine konverents sellest sarjast (mis on ühtlasi assotsiatsiooni ACL aastakonverents) toimub järgmisel aastal, 11.–16. juulini 2010 Uppsala Ülikoolis (vt <http://acl2010.org/>).

MARE KOIT,
HALDUR ÕIM