

## RAHVUSVAHELINE KEELERESSURSSIDE KONVERENTS

Mitu Tallinna ja Tartu keeletehnoloogi osales rahvusvahelisel konverentsil LREC-2008<sup>1</sup> (Language Resources and Evaluation Conference), mis toimus 28.–30. mail iidises Maroko linnas Marrakechis. Konverents leidis aset Maroko kuninga Mohammed VI patronaazi all.

Seda konverentside sarja korraldab ELRA<sup>2</sup> (European Language Resources Association), mis loodi 1995. aastal eesmärgiga koguda ja levitada keele-ressursse: keeleandmete kogusid ja keele automaattöötuse tarkvara. Esimese konverentsini jõuti 1998. aastal ja see toimus Granadas. Sestpeale on LREC-i korraldatud igal paarisaastal, käesolev oli sarja kuues ja toimus esmakordselt väljaspool Euroopat. Tegu on selle valdkonna ühe olulisema konverentsiga, mida näitab ka osalejate ülisuur arv, 1111.

Konverentsi programmikomitee, mida juhtis Pisa Arvutilingvistika Insti-

tuudi direktor Nicoletta Calzolari, pidi tegema valiku enam kui 900 saadetud artikli hulgast. Kokku peeti 645 ettekannet.

Konverentsil tutvustati innovatiivseid keele automaattöötuse meetodeid, samuti erinevate keelte jaoks loodud ressursse, uusi suundi, suuri projekte ja initsiatiive. Üldtunnustatud on keele-ressursside oluline roll keeletehnoloogias. Just keeleressursside teeneks peetakse seda, et keeletehnoloogia on saavutanud küpsuse ja rakendusteks vajaliku tõrkekindluse.

Keeletöötuses domineerivad andmepõhised meetodid. Teemade mitmekesisus võimaldas kuulajatel saada hea ülevaate valdkonna arengutest ja aitas kaasa uute ideede sünnile.

Põhiteemad olid järgmised.

– Semantika, teadmus, ontoloogiad, sisu analüüsimine erinevast vaatenurgast, alates esitamisest ja annoteerimisest ja lõpetades hankimisega, ruumilis-ajalisest infost viitamiseni ja diskursu-

<sup>1</sup> <http://www.lrec-conf.org/lrec2008/>

<sup>2</sup> <http://www.elra.info/>

sest kuni emotsioonide, tunnete ja arvamusteni.

– Kõrvuti semantikaga pälvisid tähelepanu ka süntaktiliselt annoteeritud keeleressursid.

– Peale keeleressurssidega seotud tööriistade ja süsteemide oli hästi esindatud info ekstraheerimine, nimega üksuste tuvastamine, dokumentide liigitamine, küsimustele vastamine ja sisukokkuvõtete koostamine.

– Kõne valdkonnas käsitleti raadiouudiste töötlust, kõneleja tuvastamist, kõnesünteesi, kõnetuvastust, dialoogi, erinevate keelte jaoks loodud kõnekorpuse, aga ka tunnete ja emotsioonidega seotud teemasid.

– Multimodaalsus ja multimeediумid videotöötlustest kuni viipekeeleni, suhtlus ja dialoog multimodaalses keskkonnas, emotsioonid.

– Mitu ettekannet oli pühendatud suurtele infrastruktuuridele, arhitektuuridele, suurtele projektidele, standardiseerimisele.

– Palju ettekandeid käsitles leksikone, korpuse ja nendega manipuleerimise erinevaid tööriistu.

– Süsteemide (nii kirjaliku kui ka suulise keele) evalveerimine ja ressurside valideerimine, evalveerimismetodid.

– Palju ettekandeid oli multimodaalsetest ressursidest ja süsteemidest, masintõlke rakendustest, samuti keelevariantidest.

– Tähtsal kohal olid terminoloogia probleemid, kõige enam oli esindatud biomeditsiini terminoloogia.

Suulise ettekandega esinesid koos Mark Fidel ja Heiki-Jaan Kaalep (paralleelkorpuste töötlemise eksperimentidest), stendiettekanded olid Einar Meistril koos Jaak Viloga (eesti keele tehnoloogia arendamisest) ning Mare Koidul koos Tiit Hennoste, Olga Gerasimenko, Riina Kasterpalu, Andriela

Räabise ja Krista Strandsoniga (suulise eesti keele korpustest kui intelligentsete kasutajaliideste loomise ressursist).

Alates 2004. aastast antakse välja ELRA asutaja ja esimese presidendi Antonio Zampolli (1937–2003) mälestuseks loodud preemiat, millega tunnustatakse silmapaistvat panust keeletehnoloogias. 2008. aastal pälvis preemia Sheffieldi Ülikooli professor Yorick Wilks, kes konverentsi avaistungil esines kokkuvõtliku ettekandega oma 45-aastasest keeletehnoloogiaalasest tegevusest.

Kaks kutsutud ettekannet käsitles veebi semantika lähtena (Ricardo Baeza-Yates, firma Yahoo) ning valetamise avastamist kõnes (Julia Hirschberg, Columbia Ülikool).

Konverentsile eelnesid õpi- ja töötoad 26.–27. mail ning järgnesid töötoad 31. mail ja 1. juunil. Ühe õpitubadest – „Mitteverbaalsed inforessursid konstruktiivse dialoogi juhtimiseks” – viis läbi Kristiina Jokinen, kes on TÜ intelligentsete kasutajaliideste külalisõppejõud.

Töötoas OntoLex 2008 esinesid koos Neeme Kahusk, Kadri Kerner ja Heili Orav (eesti keele ontoloogiast). Konverentsi eel toimus ka järjekordne Euroopa Liidu keeleressursside ja -tehnoloogia infrastruktuuri projekti CLARIN<sup>3</sup> koosolek, kus Eestit esindas TÜ arvutiteaduse instituudi juhataja Tiit Roosmaa.

LREC-i konverentsi ja töötubade toimetised ilmusid elektroonilise väljaandena CD-l ja need on kavas kättesaadavaks teha ka veebis.

MARE KOIT

<sup>3</sup> <http://www.clarin.eu/>