

KAITSTUD DOKTORITÖÖD

9. aprillil kaitses Hele-Andra Kuulmets Tartu Ülikoolis doktoritöö „Cross-Lingual Transfer Learning and Evaluation in Low-Resource Settings” („Keeltevaheline siirdeõpe ja selle hindamine väheste ressursidega oludes”). Juhendaja oli prof Mark Fišel (TÜ), oponendid prof Barbara Plank (Müncheni Ludwig Maximiliani ülikool, Saksamaa) ja filosoofiadoktor Jindřich Helcl (Oslo ülikool, Norra).

Meie igapäevaellu üha enam sekkuva tehisaru üks põhikomponente on keelemudel, tänu millele suudab tehisaru inimkeele peeneid nüansse mõista. Selleks et keelemudel inimkeelt hästi mõistaks, tuleb seda treenida suurte tekstikogumite peal. Enamik maailmas kõneldavaid keeli pole teksti kujul vajalikus mahu kättesaadavad. Seetõttu on tehisaru võimalused paljudes keeltes piiratud, süvendades ressursirikaste ja ressursivaeste keelte vahelist tehnoloogilist ebavõrdsust. Õnneks on ka väiksemate tekstikogumitega keeltele lahendus olemas. Nimelt on täheldatud, et keelemudelite treenimisel korruga paljude keelte

materjali peal paraneb nende oskus treeningandmestikus vähem esindatud keeltest aru saada. Seda nähtust nimetatakse keeltevaheliseks teadmussiirdeks: mudel õpib näiteks ingliskeelsetest tekstidest omandatud teadmisi kasutama eesti keeles vastamisel. Kui minna veel tehnilisemaks, siis juhtub mitmekeelsel treenimisel see, et eri keelte matemaatilised esitused mudeli sees muutuvad üksteisega sarnasemaks, mis põhjustabki keeltevahelist teadmiste siiret.

Doktoritöös uuritakse, kuidas tugevdada keelemudelites keeltevahelist teadmussiiret, et parandada keelemudelite eesti keele oskust. Töö jaguneb kaheks osaks: esimeses uuritakse meetodeid väikeste keelemudelite kasutamiseks konkreetse ülesande lahendamisel ja teises seda, kuidas õpetada peamiselt ingliskeelsetel tekstidel treenitud suurele keelemudelile eesti keelt. Töö peamine järeldus on, et mitmekeelne treenimine, isegi kui kasutada ainult sünteetilisi andmeid, võib märgatavalt parandada keelemudeli oskust eesti keeles mitmesuguseid ülesandeid lahendada. See tulemus osutab, et keelte-

vahelise teadmussiirde oskuslik ärakasutamine aitab väiksematele keeltele tehisaru maailmas parema esindatuse tagada. (<https://dspace.ut.ee/items/f7ddfa34-fdd6-43bf-bf7c-5cde3852e8a9>)

24. aprillil kaitses Riku Erkkilä Helsingi ülikoolis doktoritöö „The Semantics of Udmurt Spatial Cases“ („Udmurdi kohakäänete semantika“). Juhendajad olid prof Riho Grünthal (Helsingi ülikool), dots Svetlana Edygarova (Helsingi ülikool), emeriitprofessor Elena Skribnik (Helsingi ülikool), oponent prof Laura Janda (Tromsø ülikool).

Uurali keelte hulka kuuluva ja Venemaa Euroopa osa kirdeosas kõneldava udmurdi keele üks kesksemaid morfosüntaktilisi tunnuseid on kohakäänete suur hulk. Väitekiri esitab kognitiivse lingvistika raamistikus tehtud kvalitatiivse korpusanalüüsi udmurdi kohakäänete kohta. Tööl oli kolm eesmärki: 1) kirjeldada udmurdi keele kohakäänete semantikat ning iga käände semantilist struktuuri; 2) rakendada kognitiivse lingvistika meetodeid udmurdi keele uurimiseks, mida pole varem tehtud, et kontrollida kognitiivse keeleteaduse teoreetiliste väidete üldistusjõudu; 3) kontrollida, kas algselt inglise eessõnade semantika analüüsiks loodud metodoloogiat saab kasutada kohakäänete semantika analüüsimiseks ning seda vajadusel kohandada eessõnade

ja kohakäänete morfosüntaktilisi erinevusi arvestades.

Analüüsiks on valitud põhimõttelise polüseemia (ingl *principled polysemy*) meetod, mis loodi inglise keele ruumisuhteid väljendavate eessõnde semantika süsteemseks seletamiseks. Meetod pakub kriteeriumid, mida saab kasutada käände eri tähenduste määratlemiseks. Uurimus põhineb tänapäeva udmurdi kirjakeele internetikorpuse andmetel, millest on kogutud iga produktiivse kohakäände kasutuse kohta 500 näidet ja 100 näidet aproksimatiivi kasutusest, kuna see kääne ei ole enam täielikult produktiivne. Valim sisaldas kokku 3100 märgendatud näidet.

Udmurdi kohakäänete väljendatakse eri hulka tähendusi: töö tulemuste põhjal väljendab inessiiv seitset, elatiiv kaheksat, illatiiv kümmet, prolatiiv seitset, egressiiv kuut, terminatiiv seitset ja aproksimatiiv viit tähendust. Väitekirjas jõutakse järeldusele, et kognitiivse lingvistika väited peavad paika ka rakendatuna udmurdi keelele. Kuigi töö ei võimalda hinnata kogu raamistiku rakendatavust, järeldub, et uuritud nelja võtmenähtust saab kasutada udmurdi keele käsitlemisel. Samuti osutus võimalikuks kohandada põhimõttelise polüseemia lähenemist kohakäänete semantika uurimiseks.

(<http://hdl.handle.net/10138/630195>)