

KAITSTUD DOKTORITÖÖD

23. mail kaitstes Andre Tättar Tartu Ülikoolis doktoritöö „Multilingual machine translation for under-resourced languages” („Mitmekeelne masintõlge ressursivaestele keeltele”). Juhendaja oli professor Mark Fišel (TÜ), oponendid prof Jörg Tiedemann (Helsingi Ülikool) ja David Vilar Torres (Google Berlin, Saksamaa).

Väitekiri käsitleb masintõlkesüsteemide arendamist piiratud digitaalsete keeleressurssidega keelte jaoks. Põhifookus on soome-ugri keeltele, mida räägib kokku üle 20 miljoni inimese Euroopas ja Põhja-Aasias. Nii riigikeeled eesti, soome ja ungari kui ka väiksemad kohalikud keeled, nagu võru, liivi ja komi, kannavad endas rikkalikku kultuuripärandit, kuid paljud neist seisavad silmitsi digitaalse mahajäämusega.

Uurimistöö eesmärk oli vähendada lõhet suurema ja väiksema kõnelejakonnaga keelte vahel, arendades välja

tehisnärvivõrkudel põhinevad masintõlkesüsteemid ressursivaeste keelte jaoks. Doktoritöös on kasutatud uusimaid teaduslikult tõestatud meetodeid, et lahendada andmete vähesuse ning tõlke kvaliteedi ja efektiivsusega seotud probleeme. Töös välja toodud masintõlkesüsteemid toetavad 23 soome-ugri keelt, tõhustades seeläbi avalikku juurdepääsu teabele ja teenustele väikestes kogukondades.

Doktoritöö raames on arendatud eesti keele jaoks tõlkimisvahendeid, mis suudavad konkureerida Google'i ja DeepL-i tõlkesüsteemidega. Eesti avalik sektor on need süsteemid kasutusele võtnud, näidates nende tõhusust igapäevaolukordades. Kõik väljatöötatud mudelid on avatud litsentsiga, mis võimaldab teistelgi neid kasutada ja edasi arendada. Uurimuses rõhutatakse, et tehnoloogilise edu tagab keelelise mitmekesisuse väärtustamine. See on samm digiajastu poole, kus keelte võimalused on võrdsemad kui seni. (<https://dspace.ut.ee/items/bf25ee4e-074a-4c3f-8cec-565b39a3c818>)