

## Mis saab teoriast? Mõtteid lingvistilise klasteranalüüsi näitel

*Pille Eslon, Kais Allkivi, Jekaterina Trainis (Tallinna Ülikool)*

Mis saab teoriast kvantitatiivse pöörde tulemusel? Vastuses küsimusele lähtume 1) üldisest arenguloogikast (eitades eelnevat, jõutakse ringiga selle juurde tagasi, kuid kvalitatiivselt uuel, kõrgemal tasemel) ja 2) teooria elujõulisusest, mida kontrollivad rakendusvaldkonnad – kas teooria suudab lahendada tänaseid probleeme, kas teooria on valmis lahendama uusi probleeme ning kas teooria on piisavalt nutikas, et **lahendada vanu probleeme uutmoodi**. Kvantitatiivne pööre keeleteaduses on reaktsioon nendele küsimustele. Miks muidu on fookuses uurija individuaalsusest, konkreetsest keelest ja teoriast sõltumatu algoritmidele tuginev analüüs ning kardetakse takerduda teoriasse. Järelikult ei anna teooriad piisavalt ainet lingvistika rakendusvaldkondadele (nt keeleõpe, keeletehnoloogia). Siis haaravad rakendusvaldkonnad ohjad oma kätte, tulemuseks on nt korpuslingvistika kujunemine iseseisvaks uurimissuunaks, mis koos talle omaste meetoditega võib olla teiste valdkondade uurimuste metodoloogiliseks aluseks.

Ettekandes peatume lingvistilisel klasteranalüüsil, mis aitab vanu probleeme – klassifikatsioonid ja reeglid – lahendada uutmoodi. Tutvustame oma uurimusi, milles oleme kirjeldanud, kuidas andmekaeve Klastrileidja programmiga toob alt üles meetodil esile loomuliku objektilähedase ehk ontoloogilise taksonoomia, mida oleme rakendanud ülalt alla põhimõttel, rajades keelekasutusmustrite süsteemse esituse ning kirjelduse klasside ja alamklasside, klasterite ja n-grammide hierarhiale ning nende üksuste statistilisele olulisusele tekstiloomes. Kuna Klastrileidja andmekaeve põhineb eesti keele süntaksianalüsaatori metakeele formalismil, siis nimetame seda lingvistiliseks klasteranalüüsiks, mis kujundab ka kontseptuaalselt sidusa kasutusgrammatilise mõistevara. Lingvistilise klasteranalüüsi tulemused annavad rikkalikku ainet nii aktiivse grammatika koostamiseks (keeles elementide loomulikud kombineerimisvõimalused) kui ka avastuste mõttes, mis pakuvad huvi teoreetilisele keeleteadusele.