

## Morfoloogilise muuttüübi automaatne tuvastamine

Juhendaja: Siim Orasmaa ( *siim . orasmaa @ ut . ee* )

Morfoloogiline muuttüüp määrab, kuidas toimub sõna käänamine/pööramine ning võimaldab ühtlasi grupeerida samasuguse vormimoodustusega sõnu. Näiteks käänduvad samamoodi sõnad *auto* (*auto, auto, autot*) ja *aku* (*aku, aku, akut*), aga neist erinevalt sõnad *habe* (*habe, habeme, habet*) ja *pide* (*pide, pideme, pidet*). Eesti keele jaoks on välja pakutud mitmeid muuttüübiga (Erelt jt 2020), aga automaatses morfoloogilises analüüsis on valdavalt kasutusel leidnud „Väikese vormisõnastiku” tüübigirjeldused (Viks 1992, 2001), millele toetub ka EstNLTK [morfoloogiline analüsaator](#).

Käesoleva töö eesmärk on luua masinõppel põhinev mudel, mis ennustab eestikeelse sõna muuttüüpi, toetudes „Väikese vormisõnastiku” tüübigirjeldustele. Loodavast mudelist võiks olla kasu automaatse morfoloogilise analüüsi kohandamisel uutele sõnadele, aga ka murdekeele ja vana kirjakeele sõnade analüüsimisel.

Detailsemalt:

- Treeningandmed (sõnade algvormid ja neile vastavad muutkonnad) saab eraldada Vabamorfii põhileksikonist (vt leksikoni kirjeldust lähemalt [siit](#));
- [Morfoloogilise sünteesi](#) abil saab laiendada treeningandmeid, genereerides sõnade kõik vormid ning sidudes need muuttüübiga;
- Meetoditena võib katsetada nii traditsioonilisel masinõppel kui ka tehisnärvivõrkudel põhinevaid lähenemisi;
- Hindamisel tuleks arvestada muutkondade sagedusi ning hinnata meetodeid nii sõnadel üksikunavõttes kui ka sõnadel teksti kontekstis (kattuvuse väljaselgitamiseks);

Töö eeldab huvi eesti keele morfoloogia ja masinõppe vastu ning head Pythoni programmeerimisoskust.

### Kirjandus

Eesti keele käsiraamat, 2020. Mati Erelt, Tiiu Erelt, Kristiina Ross. Eesti Keele Instituut.

<http://www.eki.ee/books/ekkr20/ekkr20.pdf>

Väike vormisõnastik, 1992. Ülle Viks. Eesti Keele Instituut.

Tüübitabelid, 2001. Ülle Viks. Eesti Keele Instituut. [http://www.eki.ee/teemad/tabel\\_jutt.html](http://www.eki.ee/teemad/tabel_jutt.html)