

KEELESTATISTILISEST TERMINOLOOGIAST

KRISTEL UIBOAED

kristel.uiboaed@ut.ee

Tartu Ülikool

Eesti ja üldkeeleteaduse instituut

Oskuskeelepäev

Eesti Rahvusraamatukogus

10.05.2013

Ülevaade

- Taust
 - Doktoritöö
 - EstDiaSyn
 - Õppetöö
- Korpuslingvistika
- Dialektomeetria
- Statistika

- Verbiühendite tuvastamine ja analüüs eesti murretes
- Dialektomeetriline töö
- Oluline osa erinevatel kvantitatiivsetel meetoditel

- Eesti murdesüntaksi projekt 2013–2016 (Eesti Teadusagentuuri projekt PUT90)
- Projekti täitjad
 - Projektijuht vanemteadur Liina Lindström
 - Doktorandid Helen Plado ja Kristel Uiboaed
 - Magistrant Maarja-Liisa Pilvik

- Esimene murdesüntaksi projekt Eestis
- Keskne osa statistilistel meetoditel
- Rõhk dialektomeetrial
- Sagedaste süntaktiliste nähtuste varieerumismustrite kaardistamine

- Kursus „Kvantitatiivsed meetodid keeleteaduses ja statistikaprogramm R“ koos Pärtel Lippuse ja Jane Klavaniga
- Esimene statistikakursus ainult keeleteadlastele

- Statistikaalaseid eestikeelseid publikatsioone vähe
- Viimased eestikeelsed doktoritööd TÜ-s 2000 aastate alguses
- Statistika seisukohalt on keeleteaduslikud andmestikud problemaatilised.
- Keeleteaduslike andmestike jaoks on sageli sobilikud vaid väga keerulised statistilised meetodid.

Taust (Lippus 2007)

- Parim viis terminite loomiseks: tõlkimine, mugandamine või tuletamine?
- Eelistada rahvusvahelisi termineid?

- „Eesti-inglise keeleteaduse sõnastik“
(Erelt jt 2003)
- Mitmekeelne Rahvusvahelise
Statistika Instituudi terminibaas
<<http://isi.cbs.nl/glossary/>>

Sõnadevahelise seose mõõtmine

- Sõnadevahelise seose tugevuse mõõtmise statistikud (mõõdikud?) (Uiboaed 2010; 2013; *ilmumas*)
 - **odds ratio** – *suhteline risk; šansisuhe; riskisuhe (ISI)*
 - **Fisher-Yates test** – Fisher-Yatesi test; Fisher-Irwini test; täpne hii-ruut-test (ISI)
 - **Minimum Sensitivity** – *minimaalne tundlikkus?*

Konstruksioonigrammatika ja statistika

- ***collostructional analysis*** (Stefanowitsch & Gries 2003) – *kollostruktsiooniline analüüs* (Uiboaed 2013)
kollostruktsioon = kollokatsioon + konstruksioon
- ***covarying collexemes*** – *koosvarieeruvad kolekseemid*
- ***distinctive collexemes*** – *eristavad kolekseemid*

Dialektomeetria

- ***aggregate dialectology*** –
kobardialektoloogia
- ***local inchoherence*** (Nerbonne &
Kleiweg 2007) – *kohalik (lokaalne?)*
seosetus

Klassifitseerimismeetodid

- ***fuzzy clustering*** – *hägus klasterdamine (sm.k ryvästys / ryhmittely)*
- ***conditional inference trees*** – *tingimuslikud otsustuspuud?*

Supervised – unsupervised

- *õpetaja abiga*
- *näidete alusel*
- *juhendatud*
- *suunatud*
- *õpetamine*
- *ise(õppimine)*
- *iseseisev*
- *juhendamata*
- *suunamata*
- *ise(õppimine)*

- **jitter** – tavakeeles *virvendus, võdin*
 - Foneetikas: keskmine erinevus, mille võrra järjestikuste häälekurruperioodide sagedus hälbib mõõdetud aja jooksul (***hääle sagedushälve?***) (Lippus 2007)
 - „*moving each point a small distance in a random direction, in order to avoid banding due to the integer quantisation of 0 and to avoid overplotting word pairs with identical frequency signatures*“ (Evert 2008)

Kokkuvõte

- Statistiline terminoloogia on varieeruv nii ühe valdkonna sees kui ka valdkondade vahel.
- Erialati kasutatakse erinevaid meetodeid.
- Mittestatistikul on raske leida sobivaimat terminit.

- Evert, Stefan. 2008. **Corpora and Collocations**. Merja Kytö and Anke Lüdeling, eds. Corpus Linguistics. An International Handbook. Berlin: Mouton de Gruyter.
- ISI = **Rahvusvahelise Statistika Instituudi terminibaas**. <<http://isi.cbs.nl/glossary/>> [08.05.2013]
- Erelt, Mati, Tiiu Erelt, Enn Veldi. 2003. **Eesti-inglise keeleteaduse sõnastik**. Tartu Ülikool, Eesti Keele Instituut. Eesti Keele Sihtasutus. Tallinn.
- Lippus, Pärtel. 2007. **Foneetikaterminoloogiast**. Terminoloogiaseminari ettekanne Türi-Allikul 10.-11. dets.
- Nerbonne, John, Peter Kleiweg. 2007. **Toward a Dialectological Yardstick**. Journal of Quantitative Linguistics 14(2): 148–167.

- Stefanowitsch, Anatol; Stefan Th. Gries. 2003. **Collostructions: Investigating the Interaction of Words and Constructions.** International Journal of Corpus Linguistics 8(2): 209–243.
- Uiboaed, Kristel. 2010. **Statistilised meetodid murdekorpuse ühendverbide tuvastamisel.** Eesti Rakenduslingvistika Ühingu aastaraamat 6. 307–326.
- Uiboaed, Kristel. 2013. **Kollostruktsioonilised meetodid ja konstruktsioonilise varieerumise tuvastamine.** Eesti ja soomeugri keeleteaduse ajakiri ESUKA / Journal of Estonian and Finno-Ugric Linguistics JEF, 4(1). 185 - 204.
- Uiboaed, Kristel. *Ilmumas.* **Verbiühendite tuvastamine ja analüüs eesti murretes.** Doktoritöö. Tartu Ülikool: eesti ja üldkeeleteaduse instituut.