

3. rakenduslingvistika konverents

Tallinn, 25.–26. märts 2004

Teesid

Eesti Keele Instituut
Tallinna Pedagoogikaülikool

Konverentsi pearahastajad: Eesti Keelenõukogu, Haridus- ja Teadusministeerium
Toetajad: Kirjastus TEA, Eesti Keele Sihtasutus
Logo autor ja kaanekujundaja: Martin Metslang
Kodulehekül: <http://www.eki.ee/rakenduslingvistika.html>
Toimetajad: Jane Lepasaar, Maria-Maren Sepper

NELJAPÄEV, 25. MÄRTS

Eesti Keele Instituut (Roosikrantsi 6)

	EKI suur saal (III korrus)	
10.00	Konverentsi avamine	
10.15–10.45	Urmas Sutrop (EKI)	Keel ja murre
10.45–11.00	Kohvipaus	
	Lingvistilise klassifitseerimise eri aspekte	
11.00–11.30	Pille Eslon (TPÜ)	Analoogiast keelte kõrvutamisel
11.30–12.00	Katrin Karu (TÜ)	Mööndlausete klassifitseerimisvõimalustest eesti ja vene keele näitel
12.00–12.30	Ilona Tragel (TÜ)	Mille alusel liigitada, kui vormiline liigitusalus puudub? (modaalverbide näitel)
12.30–13.00	Neeme Kahusk, Kadri Vider (TÜ)	Sõnatähenduste ühestamine kui klassifitseerimisülesanne
13.00–13.30	Toomas Kirt (TTÜ), Ene Vainik (EKI)	Iseorganiseeruvad keeleelemendid eesti keele emotsioonisõnavara näitel
13.30–14.00	Marja Kallasmaa (EKI)	Eesti kohanimede liigitamisest
14.00–15.00	Lõuna	
	Keelestrateegia	
15.00–15.30	Jüri Valge (HTM)	„Eesti keele arendamise strateegia 2004–2010“: kaasnevad publikatsioonid
15.30–16.30	Publikatsioonide tutvustused	
16.30–17.30	Rakenduslingvistika Ühingu asutamine	
17.30	EKI direktori vastuvõtt (EKI väikeses saalis, II k)	

NELJAPÄEV, 25. MÄRTS

Eesti Keele Instituut (Roosikrantsi 6)

EKI väike saal (II korrus)

Tööpada “Keel, etnos, rahvus”

- | | | |
|-------------|---|--|
| 11.00–11.20 | John M. Kirk
(Queen’s University
Belfast) | The present-day Situation of
Ulster-Scots |
| 11.20–11.40 | Deirdre Davitt | Irish: some remarks |
| 11.40–12.00 | Ana Pascaru | Some ethnolinguistic issues
in Moldova |
| 12.00–12.20 | Adina Friedman | Jews, Arabs and coexistence |
| 12.20–12.40 | Klara Hallik (TPÜ) | About ethnic processes in
Estonia |
| 12.40–13.00 | Kadri Korenik,
Sulev Iva
(Võru Instituut) | The Võro language –
wannabe language? |

Klassifitseerimine tõlketeoorias, keeleõppes ja testimises

- | | | |
|-------------|----------------------------------|---|
| 13.00–13.30 | Arvi Tavast (TPÜ) | Tellijä kriteeriumid tõlke
hindamisel ja nende sobivus
koolitööde hindamissüsteemi
alusena |
| 13.30–14.00 | Hille Pajupuu (EKI) | Kuidas hinnata hindajaid |
| 14.00–14.30 | Andres Valdre (TÜ) | Inglise-eesti tehnikatõlke
vigade liigid ja põhjused |
| 14.30–15.00 | Raili Pool,
Elle Vaimann (TÜ) | Kõrgtasemel eesti keele kõ-
nelejate keelevigade klassi-
fitseerimise võimalusi ja
probleeme |
| 15.00–15.30 | Regina Toom (TÜ) | Eesti viipekeele õpetamise
eripära |

REEDE, 26. MÄRTS

Tallinna Pedagoogikaülikool (Narva mnt 25, 29)

TPÜ keeltemaja, ruum 240

(Narva mnt 29, II korrus)

Keeletehnoloogia ja arvutirakendused

9.00–9.30	Kari Sajavaara (University of Jyväskylä)	DIALANG: Learner-Centred Assesment Using Information Technology
9.30–10.00	Einar Meister, Jürgen Lasn, Lya Meister (TTÜ)	Eestikeelse kõnetuvastuse andmebaas: lõpparuanne
10.00–10.30	Tanel Alumäe (TTÜ)	Pseudomorfeemidel põhi- neva eestikeelse kõne- tuvastussüsteemi loomine
10.30–11.00	Kohvipaus	
11.00–11.30	Einar Kraut (Eesti Raadio)	Foneemiklassid ja hääldus- praktika
11.30–12.00	Meelis Mihkla (EKI)	Pausid ja pauseelsed taktipikendused kõnes
12.00–12.30	O. Gerassimenko, T. Hennoste, M. Koit, A. Rääbis, M. Valdisoo (TÜ)	Dialogiaktide klassifitseerimine: eestikeelsete infodi- loogide analüüsi probleeme
12.30–13.00	Jaak Henno (TTÜ)	Sõnavara ja nimede tekkimisest
13.00–14.00	Lõuna	
14.00–14.30	Rein Kuusik, Leo Võhandu (TTÜ)	Ühest objektsüsteemide struktureerimise efektiivsest meetodist
14.30–15.00	Rein Kuusik, Leo Võhandu (TTÜ)	Monotoonsed süsteemid ja tekst

REEDE, 26. MÄRTS

Tallinna Pedagoogikaülikool (Narva mnt 25, 29)

TPÜ keeltemaja, ruum 232

(Narva mnt 29, II korrus)

Leksikograafia ja sõnaliigid

9.00–9.30	Valentina Štšadneva (TÜ)	Sõnaliikide klassifitseerimise probleemidest võrdlevas aspektis
9.30–10.00	Anni Silk (TÜ)	Klassifitseerimisprobleeme saksa-eesti verbirektsiooni- sõnaraamatu koostamisel
10.00–10.30	Anne Arold (TÜ)	Sõnaliikidest seisundi ja sei- sundimuutuse väljendamisel eesti ja saksa keeles
10.30–11.00	Kohvipaus	
11.00–11.30	Heete Sahkai (EKI, TPÜ)	Kategoriseerimisstrateegiaid teoreetilises keeleteaduses
11.30–12.00	Margit Langemets, Merike Mägedi, Ülle Viks (EKI)	Süntaktiline info sõnastikes, eriti EXS-s
12.00–12.30	Heili Orav (TÜ)	Ühe sõnaliigi – omadussõna – eripärast
12.30–13.00	Silvi Vare (EKI)	Ühest sõnaliigiprobleemist sõnamoodustuse nurga alt vaadatuna
13.00–14.00	Lõuna	
14.00–14.30	Kadri Muischnek, Kadri Vider (TÜ)	Hoolimata normide kohaselt püstitatud ülesandest näivad tulemused teisel pool määratust
14.30–15.00	L. Lindström, M.-L. Kalvik, M. Mets, R. Läänemets, K. Paju- salu, E. Velsker, A. Veis- mann, P. Teras (TÜ)	Sõnaklasside liigituse küsimusi eesti murrete korpuses

REEDE, 26. MÄRTS

Tallinna Pedagoogikaülikool (Narva mnt 25, 29)

TPÜ keeltemaja, ruum 240

(Narva mnt 29, II korrus)

Klassifitseerimine ja lingvistika naaberalad

- | | | |
|-------------|--|--|
| 15.00–15.30 | Silvi Tenjes (TÜ) | Kategoriseerimise
kognitiivsed ja
sotsiokultuurilised alused |
| 15.30–16.00 | Suliko Liiv (TPÜ) | Politically correct terms: pros
and cons |
| 16.00–16.30 | Silvi Salupere (TÜ) | Märkide klassifitseerimine.
Ajalooline ülevaade |
| 16.30–17.00 | Andres Soosaar (TÜ) | Hjelmslevi valguses
Lahterdused füsioloogias |
| 17.15 | TPÜ rektori vastuvõtt
(TPÜ peamaja kohvikus, Narva mnt 25, I k) | |

REEDE, 26. MÄRTS

Tallinna Pedagoogikaülikool (Narva mnt 25, 29)

TPÜ keeltemaja, ruum 232

(Narva mnt 29, II korrus)

Leksikograafia ja sõnaliigid (jätkub)

- | | | |
|-------------|---|---|
| 15.00–15.30 | Reili Argus (TPÜ) | Onomatopoeetiliste sõnade kohast lastekeeles: sõnaliigilisest ambivalentsusest ja morfoloogiast |
| 15.30–16.00 | Karin Kaljumägi (KVÜÕA) | Sõjandusterminite klassifitseerimise probleeme |
| 16.00–17.15 | Sõnaliikide diskussioon | |
| 17.15 | TPÜ rektori vastuvõtt (TPÜ peamaja kohvikus, Narva mnt 25, I k) | |

Sisukord

<u>Pseudomorfeemidel põhineva eestikeelse</u> <u>kõnetuvastussüsteemi loomine</u> <i>Tanel Alumäe</i>	11
<u>Onomatopoeetiliste sõnade kohast lastekeeles: sõnaliigilisest</u> <u>ambivalentisusest ja morfoloogiast</u> <i>Reili Argus</i>	13
<u>Sõnaliikidest seisundi ja seisundimuutuse väljendamisel eesti ja</u> <u>saksa keeles</u> <i>Anne Arold</i>	15
<u>Analoogiast keelte kõrvutamisel</u> <i>Pille Eslon</i>	16
<u>Dialogiaktide klassifitseerimine: eestikeelsete infodialoogide</u> <u>analüüsi probleeme</u> <i>Olga Gerassimenko, Tiit Hennoste, Mare Koit, Andriela Rääbis,</i> <i>Maret Valdisoo</i>	18
<u>Sõnavara ja nimede tekkimisest</u> <i>Jaak Henno</i>	21
<u>Sõnatähenduste ühestamine kui klassifitseerimisülesanne</u> <i>Neeme Kahusk, Kadri Vider</i>	24
<u>Eesti kohanimedega liigitamisest</u> <i>Marja Kallasmaa</i>	25
<u>Sõjandusterminite klassifitseerimise probleeme</u> <i>Karin Kaljumägi</i>	26
<u>Mööndlausete klassifitseerimisvõimalustest eesti ja vene keele näitel</u> <i>Katrin Karu</i>	27
<u>The present-day Situation of Ulster-Scots</u> <i>John M. Kirk</i>	29
<u>Iseorganiseeruvad keelelemendid eesti keele emotsioonisõnavara</u> <u>näitel</u> <i>Toomas Kirt, Ene Vainik</i>	30
<u>Foneemiklassid ja häälduspraktika</u> <i>Einar Kraut</i>	32
<u>Ühest objektsüsteemide struktureerimise efektiivsest meetodist</u> <i>Rein Kuusik, Leo Vöhandu</i>	35

<u>Monotoonsed süsteemid ja tekst</u>	
<i>Rein Kuusik, Leo Võhandu</i>	37
<u>Süntaktiline info sõnastikes, eriti EXS-s</u>	
<i>Margit Langemets, Merike Mägedi, Ülle Viks</i>	39
<u>Politically correct terms: pros and cons</u>	
<i>Suliko Liiv</i>	42
<u>Sõnaklasside liigituse küsimusi eesti murrete korpus</u>	
<i>Liina Lindström, Mari-Liis Kalvik, Rutt Läänemets, Mari Mets, Karl Pajusalu, Pire Teras, Ann Veismann, Eva Velsker</i>	43
<u>Eestikeelse kõnetuvastuse andmebaas: lõpparuanne</u>	
<i>Einar Meister, Jürgen Lasn, Lya Meister</i>	45
<u>Pausid ja pauseelsed taktipikendused kõnes</u>	
<i>Meelis Mihkla</i>	48
<u>Hoolimata normide kohaselt püstitatud ülesandest näivad tulemused teisel pool määratud</u>	
<i>Kadri Muischnek, Kadri Vider</i>	49
<u>Ühe sõnaliigi – omadussõna – eripärast</u>	
<i>Heili Orav</i>	50
<u>Kuidas hinnata hindajaid</u>	
<i>Hille Pajupuu</i>	51
<u>Kõrgtasemel eesti keele kõnelejade keelevigade klassifitseerimise võimalusi ja probleeme</u>	
<i>Raili Pool, Elle Vaimann</i>	53
<u>Sõnaliigikimbatusi arvutuslingvistikas</u>	
<i>Tiina Puolakainen</i>	55
<u>Kategoriseerimisstrateegiaid teoreetilises keeleteaduses</u>	
<i>Heete Sahkai</i>	57
<u>DIALANG: Learner-Centred Assessment Using Information Technology</u>	
<i>Kari Sajavaara</i>	58
<u>Märkide klassifitseerimine. Ajalooline ülevaade Hjeltslevi valguses</u>	
<i>Silvi Salupere</i>	60
<u>Klassifitseerimisprobleeme saksa-eesti verbirektsioonisõnaraamatu koostamisel</u>	
<i>Anni Silk</i>	61
<u>Lahterdused füsioloogias</u>	
<i>Andres Soosaar</i>	63

<u>Keel ja murre</u>	
<u>Urmas Sutrop</u>	65
<u>Sõnaliikide klassifitseerimise probleemidest võrdlevas aspektis</u>	
<u>Valentina Štšadneva</u>	67
<u>Tellija kriteeriumid tõlke hindamisel ja nende sobivus koolitööde</u>	
<u>hindamissüsteemi alusena</u>	
<u>Arvi Tavast</u>	68
<u>Kategoriseerimise kognitiivsed ja sotsiokultuurilised alused</u>	
<u>Silvi Tenjes</u>	70
<u>Eesti viipekeele õpetamise eripära</u>	
<u>Regina Toom</u>	73
<u>Mille alusel liigitada, kui vormiline liigitusalus puudub?</u>	
<u>(modaalverbide näitel)</u>	
<u>Ilona Tragel</u>	75
<u>Inglise-eesti tehnikatõlke vigade liigid ja põhjused</u>	
<u>Andres Valdre</u>	77
<u>„Eesti keele arendamise strateegia 2004–2010“:</u>	
<u>kaasnevad publikatsioonid</u>	
<u>Jüri Valge</u>	79
<u>Ühest sõnaliigiprobleemist sõnamoodustuse nurga alt vaadatuna</u>	
<u>Silvi Vare</u>	81

Pseudomorfeemidel põhineva eestikeelse kõnetuvastussüsteemi loomine

Tanel Alumäe

TTÜ Küberneetika Instituudi foneetika ja kõnetehnoloogia labor
tanel.alumae@aqris.com

Ettekandes tutvustatakse eksperimente suure sõnavaraga kõnelejust sõltumatu eesti keele tuvastussüsteemi loomisel, kus keelemudelina kasutatakse pseudomorfeemidel põhinevat statistilist trigramm-mudelit.

Suure sõnavaraga kõne tuvastamise eesmärgiks on sidusa ja tehnilike piiranguteta inimkõne teisendamine tekstiks, mille rakendusteks on näiteks dikteerimissüsteemid ning automaatsed kõne transkriptsioonisüsteemid. Sellise tuvastussüsteemi üheks osaks on keelemudel, mis peaks võimalikult hästi modelleerima kõiki antud keeles esineda võivaid lauseid. Tihti kasutatakse selleks sõnadel põhinevat statistilist keelemudelit, mis annab paljude keelte puhul häid tulemusi. Erinevate sõnajärgnevuste tõenäosust hinnatakse siis suurte tekstikorpuste põhjal, kasutades tavaliselt kuni 64K suurust sõnastikku. Sõnastik koostatakse korpuses kõige sagedamini esinenud sõnadest. Eesti ja teiste aglutinatiivsete keelte puhul selline lähenemine ei toimi, sest erinevate sõnavormide arv on väga suur. Näiteks annab 60 000 ühiku suurune sõnastik eesti keele puhul ainult 85% korpuse katvuse (Alumäe 2004), võrdluseks: inglise keeles on katvus 99%. Selline katvus tagab aga hinnanguliselt ainult 76% maksimaalsest sõnade tuvastustäpsusest, arvestades, et iga kõnes esinev tundmatu sõnavorm põhjustab kontekstuaalselt keskmiselt 1,6 valesti tuvastatud sõnavormi.

Kirjeldatud probleemist ülesaamiseks kasutatakse antud süsteemis keelemudeli põhiühikutena morfeeme. Baasandmetena kasutatakse Tartu Ülikooli Eesti Kirjakeele Korpuse 1990-ndate ilukirjandus-

tekste ning suuremahulisi ajalehtede Eesti Ekspress ja Postimees korpusi. Korpusi töödeldi eesti keele morfoloogilise analüüsi vahenditega (Kaalep jt 2000), mille tagajärjel eraldati teatud piiranguid järgides sõnavormide tüvedest lõpud, samuti tükeldati liit-sõnad. Nii saadud uue korpuse põhjal koostati 60 000 sagedamini esinenud morfeemist koosnev statistiline trigramm-mudel, mida rakendati kõne dekodeerimisel talle vastavaks morfeemide jadaks. Tuvastusele järgnevas järeltöötuses liideti tuvastatud morfeemid sõnavormideks.

Eksperimentides võrreldi esialgset sõnavormidel põhinevat keele-mudelit saadud morfosüntaktilise mudeliga. Hinnati sõnastike katvust, entroopiat ning bigrammide ja trigrammide katvust. Samuti võrreldi foneemide, morfeemide ja sõnavormide tuvastustäpsust. Selgus, et morfeemidel põhineva keelemudeli sõnastiku katvus on 97,1 ning mudeli kasutamisel saadud sõnavormide tuvastustäpsus 69,2%, mis on tunduvalt parem kui sõnavormidel põhineva mudeli kasutamisel.

Kirjandus

- Alumäe, T.** 2003. Eksperimendid eesti keele tuvastussüsteemi loomisel. – Toimiv keel II. TPÜ eesti keele õppetooli toimetised 3. EKI, TPÜ. Toim Jane Lepasaar, Maria-Maren Sepper. Tallinn: TPÜ Kirjastus.
- Kaalep, H.-J.; Vaino, T.** 2000. Teksti täielik morfoloogiline analüüs lingvisti töövahendite komplektis. – Arvutuslingvistilt inimesele. Tartu. 87–99.

Onomatopoeetiliste sõnade kohast lastekeeles: sõnaliigilisest ambivalentsusest ja morfoloogiast

Reili Argus
argusr@tpu.ee

Onomatopoeetiliste sõnade rohkust on traditsiooniliselt peetud soome-ugri keeltele omaseks jooneks, eriti rohkesti on sedalaadi sõnu laste- ja hoidjakeeles. Varaste onomatopoeetiliste sõnade uurimise teeb aga keerukaks nende sõnaliigiline ambivalentsus, eriti selgelt ilmneb see protosõnaperioodil, mil hääliitsuse sõnaliigilist kuuluvust ei õnnestu täie kindlusega määrata isegi mitte konteksti abil.

Onomatopoeetilised sõnad on lapse keekekasutuses väga varased, põhjused võivad olla nii prosoodilised kui ka semantilised. Kõnealused sõnad on kõnejadast kergemini segmenteeritavad, leitavad, täiskasvanu toob neid kuuldavale prosoodiliselt eristuvatena, segmenteerimist hõlbustab sageli reduplikatsioon. Tähelepanuta ei saa aga jätta ka semantilisi tegureid: onomatopoeetiliste sõnade häälikusümboolika on motiveeritud, helijäljendus viitab otse referendile, vähemalt osale sellest, st helile, mida ese või olend kuuldavale toob.

Helijäljenduslikud sõnad saavad tuge hoidjakeelest. Kuigi eesti keeles on pea igale koduloomale üldaktsepteeritav onomatopoeetiline vaste, pole ometi päris selge, kui suure osa onomatopoeetilised sõnad hoidjakeele sõnavarast moodustavad, kas ja mil määral nende sõnade arv lapse vanemaks saades muutub.

Onomatopoeetiliste sõnade moodustamine on produktiivne: peale konventsionaalsete sõnade esineb lapse kõnes sageli ka hulk iseloodud helijäljenduslikke sõnu, samuti moodustab laps onomatopoeetilise täiendosaga liitsõnu.

Üldise leksikaalse arengu ja onomatopoeetiliste sõnade kasutamise vaheline seos ei ole nii otsene kui esmapilgul paista võiks: sõnavara arengu kiirel perioodil onomatopoeetilisi sõnu oluliselt vähemaks ei jää, küll aga väheneb nende hulk järsult aktiivse vormimoodustuse alguses ehk keele arengu protomorfoloogilisel perioodil. Seega võib konventsionaalsete onomatopoeetiliste sõnade olulisus seisneda just morfoloogilise keerukuse vältimises – hoidjakeel pakub võimaluse kasutada astmevaheldusliku muuttüübi asemel astmevahelduseta muuttüüpi.

Sõnaliikidest seisundi ja seisundimuutuse väljendamisel eesti ja saksa keeles

Anne Arold
Tartu Ülikool
anne.arold@ut.ee

Kavas oleks lähemalt vaadelda eesti ja (eelkõige) saksa keele kõrvutamisel ilmnevat erinevust, mis seisneb selles, et seisundeid ja seisundimuutusi tähistatakse saksa keeles valdavalt verbilisel kujul, sh (pseudo-)partitsiipidega, eesti keel seevastu võimaldab samu mõisteid tähistada osaliselt käänduvate seisundiadverbidega. Tähelepanu väärib asjaolu, et kui saksakeelsed verbid on enamasti ka eesti keelde verbidena tõlgitavad, siis teises suunas võib tekkida raskusi adekvaatsete tõlkevastete leidmisel. See on muu hulgas tingitud eesti keelele omasest deskriptiivsete tüvede arvukusest. Igapäevakeeles ja ilukirjanduses võib selle tüpoloogilise iseärasusega seoses ette tulla isegi tõlkimatuid sõnamänge (Lapsesuu: *Mis Päike siis teeb, kui tahab looja minna, aga Kuu on loojas ees?*). Probleemi on küll varemgi käsitletud nii keeleteaduslikel diskussioonidel kui ka publitsistikas (nt J. Kaplinski), kuid praeguse euroopastumise taustal ei teeks paha sellest meie keele pahatihti unustatud rikkusest veel kord juttu teha. Oma ettekandes kasutaksin eelkõige näiteid, mis on kogutud inimese välimust kirjeldavate omadussõnade võrdlemisel saksa ja eesti keeles.

Analoogiast keelte kõrvutamisel

Pille Eslon
Tallinna Pedagoogikaülikool
peslon@tpu.ee

Probleemiasetuse tingib tõsiasi, et igasugune tüpoloogiline klassifikatsioon põhineb tegelikult analoogiaprotseduuril, mis on uurija peas kui omamoodi võrdlusalus, et seostada erinevaid keelendeid nende vormi, tähenduse või funktsioonide samasuse-läheduse alusel. Samased funktsioonid esinevad ühe suurusjärguga objektidel (näiteks analoogia vene keele imperfetiivse verbi ja eesti sünteetilise verbi oleviku vormi või vene perfektivse verbi tuleviku vormi ja eesti analüütilise verbi oleviku vormi aspektitähenduse vahel), lähedased funktsioonid on aga erineva suurusjärguga objektidel (näiteks vene verbi aspektisemantika grammatiline väljendus ning eesti *ma*-infinitiivi käändeliste vormide kasutamine aspektitähenduse edastajana). Kuna mõisted, mida seejuures keelendite ja keeltevaheliste suhete võrdlemisel kasutatakse, on olemuselt tunnetuslikud, siis võib neid ühthästi kritiseerida ja samas ka aktsepteerida. Kõik oleneb sellest, kui võrd analoogiaprintsiibi rakendamist on suudetud põhjendada, kui võrd analoogial rajanevate suhete tuvastamise (kas ühe või mitme keele piires) aluseks on võetud tegelikud keelesisesed või keeltevahelised põhjuslikud seosed, mis reaalselt toimivad erinevate objektide vahel. Seejuures on oluline, et uurija mõistaks ja tunnetaks vajadust järjekindlalt keskenduda paralleelnähtuste (variatiivsuse) uurimisele. Eelkõige neis peegelduvad keelendite sekundaarfunktsioonid, mis keeleõppija ning tõlkija (ka tõsiteaduri) silmis raskemini mõistetavad, kuid samas äärmiselt vajalikud, et ennetada libainterferentsi, mis võib tekkida, kui keeleõppija hakkab omaalgatuslikult ühtlustama erinevaid paradigmasid. Just veal põhinev analoogia on aldis tekkima, eriti kui see aitab hõlbustada õpiraskusi. Seetõttu ongi oluline, et loogiline ja loomulik klassifikatsioon, mis peegeldub sõnade leksikaalsemantilises tähenduses, mõistete ja kategooriate eristamise hierarhias, leiaks adekvaatse

väljundi ka keeleteaduslikes klassifikatsioonides. Erilise tähenduse saab see aga seoses keelekasutuse pragmaatiliste vajadustega (võõrkeeleõpikud, sõnastikud, kõrvutavad grammatikad, metoodilised vahendid), kus on oluline, et analoogiat osataks rakendada keelendite ja keelekategoriate üldise uurimis- ja klassifitseerimisprintsipiina.

Dialoogiaktide klassifitseerimine: eestikeelsete infodialoogide analüüsi probleeme

Olga Gerassimenko, Tiit Hennoste, Mare Koit, Andriela Rääbis,
Maret Valdisoo
Tartu Ülikool
{gerro, Tiit.Hennoste, Mare.Koit, Andriela.Raabis, maret}@ut.ee

1. Eesti dialoogikorpused sisaldab kaht liiki materjali: suulisi ja kirjalikke dialooge. Suulised dialoogid on inimeste vahel tegelikult asetleidnud vestlused, kirjalikud dialoogid aga on saadud arvutisimulatsioonides, kus arvuti rolli etendab inimene. Meie kaugem eesmärk on luua arvutiprogramm, mis eestikeelses telefonidialoogis kasutajaga suudaks automaatselt ära tunda dialoogiakte ja nendele adekvaatselt reageerida. Lähemaks eesmärgiks on detailselt uurida suulisi vestlusi.

2. Dialoogikorpuses dialoogiakte märgendama asudes tuli meil esmalt määratleda ja liigitada dialoogiaktid. Olemasolevad liigitused (hea ülevaate annab nt Klein, Soria 1998) osutasid enamasti olevat orienteeritud kitsale ainevaldkonnale või kasutasid liiga üldisi või ebatäpseid kategooriaid. Seetõttu otsustasime – olemasolevaid liigitusi eeskujuks võttes – välja töötada uue dialoogiaktide klassifikatsiooni, mis lähtuks eestikeelsetest suulistest dialoogidest ja vastaks täpselt meie eesmärkidele. Algul koostasime dialoogiaktide esialgse loendi, mida täiendasime ja korrigeerisime tegelike dialoogide märgendamisel. Seejuures oleme oma dialoogiakte järkjärgult süstematiseerinud, saamaks võimalikult loogilist ja arusaadavat tüpoloogiat. Seega toetub meie tüpoloogia tegelike vestluste analüüsil leitud dialoogi universaalsetele alusmehhanismidele, aga samas vastab ta ka klassifitseerimise üldistele printsiipidele.

3. Põhiliseks küsimuseks tüpologia tegemisel on aktide piiride määramine. Infodialoogides on oluline naaberpaar küsimus – vastus. Küsimuste märgendamisel tekib kaks probleemirühma:

- Kuidas eristada küsimust ja direktiivi? Nende keskne erinevus on vormis: küsimused moodustavad keeleliselt selgelt piiritletava rühma, direktiivid aga ei oma eesti keeles selget keelelist eripära (otsesed käskivas kõneviisis direktiivid on analüüsitud dialoogides väga haruldased). Probleme tekitavad kaudsed kõneaktid, nt küsimuse vormis esitatud soov. Siin on oluline, millist tüüpi vastust naaberpaari esiliikmele oodatakse, nt küsimusele *Kas te teate silmaarsti vastuvõtuaegu?* ei oodata vastust *jah*.
- Küsimuste tüpoloogiad on enamasti loogiliselt üles ehitatud (Metslang 1981). Meie oleme aga oma tüpologia välja töötanud empiirilise materjali põhjal.

Meie aktitüpologia sisaldab naaberpaaride esiliikmetena suletud ja jutustava *kas*-küsimuse, avatud küsimuse, vastust pakkuva küsimuse, alternatiivküsimuse ning järelliikmetena *jah*, *ei*, nõustuva *ei*, info andmise või puudumise, mitmed alternatiivvastuse variandid.

Kui avatud küsimusega (nt *kes*, *millal*, *kuidas*, *miks*) ja alternatiivküsimusega (nt *Kas haiglas või polikliinikus?*) pole analüüsil enamasti probleeme, siis praegu on meie jaoks alles ebamäärased piirid suletud ja jutustava *kas*-küsimuse ning vastust pakkuva küsimuse vahel. See ebamäärasus alandab muuhulgas ka dialoogikorpusse märgendamise usaldusväärsust (mida mõõdetakse kapakoefitsiendiga). Infodialoogide märgendamisel oleme saavutanud kapa väärtuse 0,74, heaks loetakse väärtust 0,8.

Suletud ja jutustav *kas*-küsimus ning vastust pakkuv küsimus on vormilt sarnased, realiseerudes enamasti *kas* või *vä* abil. Erinevus seisneb selles, missugust vastust oodatakse. Suletud *kas*-küsimus ootab *jah/ei*-vastust (nt *Kas te olete ka talvel avatud?*). Jutustava *kas*-küsimusega soovitakse aga rohkem infot saada. Näiteks küsimusele *Kas mingi buss jõuab Tallinnasse kella kaheksaks?* ei oodata *jah/ei*-vastust, vaid küsija soovib, et vastaja loetleks nende busside väljumisajad. Vastust pakkuva küsimuse puhul on küsijal mingi hüpotees ning ta ootab sellele kinnitust (*pluss maksud jah*). Olukorra

teeb keerulisemaks asjaolu, et vastaja võib küsimust tõlgendada teisiti, kui see oli mõeldud.

4. Analüüsisime küsimuste ja vastuste naaberpaare 101 suulises infodialoogis eesmärgiga leida formaalseid tunnuseid, mis võimaldaksid neid dialoogiakte täpselt klassifitseerida. Ettekandes käsitleme lähemalt selle analüüsi tulemusi.

Viited

- Klein, M.; Soria, C.** 1998. MATE Deliverable D1.1. Supported Coding Schemes. Dialogue Acts. – <http://www.dfki.de/mate/d11/chap4.html>
- Metslang, H.** 1981. Küsilause eesti keeles. Tallinn: Valgus.

Sõnavara ja nimede tekkimisest

Jaak Henno
Tallinna Tehnikaülikool
jaak@cc.ttu.ee

Siin on uuritud ühise sõnavara ja (üld)nimede tekkimist agentide ühiskonnas. Algselt ei ole agentidel mingit sõnavara; vajadusel (kui agent tahab kirjeldada mingi välismaailma objekti omadust või anda objektile nimi) „leiutavad” agendid uusi sõnu või kasutavad sõna, mida (rääkija arvates) on varem kõige sagedamini kasutatud selle omaduse või nime tähistusena. Sõnade esinemine kontekstis (agentide poolt vahetatavad sõnumid on alati välismaailma objektide kirjeldused) võimaldab protsessil koonduda – agendid hakkavad kasutama omaduste ja objektide nimedena samu sõnu, tekib ühine sõnavara. Peamised siinkasutatud mudeli erinevused varasematest (Batali 1998, Hurford 1998, Zuidema (<http>), Steels 1997, 2002, Kirby 2002, Komarova, Nowak 2003 *etc*) on järgnevad:

- pole kasutatud vertikaalset arengut (generatsioonid, loomulik valik) – osutub, et agendid suudavad omandada ühise sõnavara ilma suremise ja loomuliku valikutagi;

- välismaailma objektid on määratud oma omadustega (atribuutidega) ja nad võivad olla struktuursed (st objektid võivad koosneda teistest objektidest); objektide hierarhiline struktuur aitab uurida keele kompositsioonilisusega seotud probleeme; kompositsioonilisus, vastavus süvastruktuuri (Chomsky *deep structure*) ja pindstruktuuri, süntaksi (*surface structure*) vahel on loomuliku keele kõige keerukamaid probleeme, sest erinevate keelte süntaksis kasutatakse süvastruktuuri kirjeldamiseks väga erinevaid vahendeid – sõnade järjekord, muutelõpud, propositioonid jne;

- agendid annavad välismaailma objektidele nimesid; nimede andmisest ja objektide struktuursusest arusaamisest näib algavat „sisemise” välismaailma mudeli loomine;

- algselt vahetavad agendid vaid nn põhjustatud (*grounded*) sõnumeid, mis koosnevad viitest kommunikatsiooniobjektile (paljud

uurijad arvavad, et loomulik keel tekkis zestide abil toimuvast kommunikatsioonist) ja kommunikatsiooniobjekti omadusi kirjeldavast sõnumist; pärastpoole on aga agendid võimelised õppima uusi sõnu ka põhjustamata (ainult sõnaliste) sõnumite abil; ka inimesed õpivad suurema osa oma sõnavarast põhjustamata sõnumitest (st kontekstist) – vestlustest, loetud tekstidest jne.

Keelt vaadeldakse vastavusena sümbolite (sõnade ja sõnumite) ruumi ja tähenduste ruumi (atribuutidega määratud lihtobjektid ja lihtobjektide ja atribuutide abil määratud lihtobjektid) vahel. On kirjeldatud lihtne kommunikatsioonialgoritm (kõik agendid kasutavad sama algoritmi) ja tulemused, mis saadi selle kasutamisel arvutisimulatsioonides. Mudelil on mitmeid ühiseid jooni laste keeleõppimise protsessiga – arusaamine tekib tunduvalt varem kui sõnade kasutamisoskus, esineb nn sõnavarahüpe (*vocabulary spurt*), nimed omandatakse teiste sõnaliikidega võrreldes kiiremini jne. Kasutatud mudel on funktsionaalne (Prolog-programm) ja võimaldab uurida ka protsessi „sisemisi” mehhanisme (närvivõrkudel põhinevad mudelid seda ei võimalda).

Kirjandus

- Batali, J.** 1998. Computational simulations of the emergence of grammar. – Approaches to the Evolution of Language: Social and Cognitive bases. J. Hutford, C. Knight, M. Studdert-Kennedy (Eds). Cambridge University Press, 405–426.
- Hurford, J.** 1998. Social transmission favours linguistic generalisation. – The emergence of language. C. Knight, J. Hurford, M. Studdert-Kennedy (Eds.).
- Kirby, S.** 2002. Natural language from artificial life. – Artificial Life, 8:2, 185-215.
- Komarova, N. L.; Nowak, M. A.** 2003. Language, learning and evolution. – Language Evolution. M.H. Christiansen, S. Kirby (Eds.). Oxford University Press, 317–337.
- Steels, L.** 1997. The synthetic modeling of language origins. – Evolution of Communication, 1:1, 1-34.
- Steels, L.; Kaplan, F.** 2002. Bootstrapping grounded word semantics. – Linguistic Evolution through Language Acquisition: Formal and Computational Models. T. Briscoe (Ed). Cambridge University Press.

Zuidema, W. H. Emergent syntax: the unremitting value of computational modelling for understanding the origins of complex language.
http://citeseer.nj.nec.com/cache/papers/cs/22259/http:zSzzSzarti.vub.ac.bezSz~jellezSzresearchzSzecal_zuidema.pdf/zuidema01emergent.pdf

Sõnatähenduste ühestamine kui klassifitseerimisülesanne

Neeme Kahusk, Kadri Vider
Tartu Ülikool
Neeme.Kahusk@ut.ee, Kadri.Vider@ut.ee

Selleks, et sõnade tähendusi ühestada, on vaja lahendada kaks alamülesannet: esiteks, teha kindlaks, mis tähendustes sõna (antud teema puhul) üldse esineda võib ja teiseks, leida, millises nendest tähendustest antud sõna (tekstis sõne) on kasutatud. Triviaalse lahenduse puhul leitakse vajalikud tähenduse võimalused leksikonist (tesaurusest) ja omistatakse sõnale tähendus, mis antud konteksti puhul kõige paremini sobib.

Kui hakata hindama sõna tähenduse ühestamise sisulist täpsust, tekivad probleemid, mis puudutavad mõlemat eelpool mainitud alamülesannet: esiteks, ei pruugi tesauruses vajalikku tähendust olla ja teiseks ei pruugi ükski olemasolevatest tähendustest sobida seetõttu, et antud kontekstis pole vaja nii täpset eristamist kui tesaurusekirjed seda eeldaks. Tesaurust muutmata oleks üks lahendus selline, kus ühestamisprogramm lisab ise tesauruses puuduva tähenduse (muidugi mõista formaalselt, näiteks kujul „oleks vaja veel ühte tähendust“) või märgib, millised tähendused oleksid võrdselt sobivad.

Selline klassifitseerimine ei mahu vormi „üks ja ainult üks etteantud valikust“. Kuna taolised probleemid ilmnevad ka teiste ühestamisülesannete (näiteks morfoloogilise ja süntaktilise) puhul, siis võib oletada, et tegemist ei ole mitte leksikograafide otsustamisraskustega, vaid keelele üldomase nähtusega.

Eesti kohanimedede liigitamisest

Marja Kallasmaa
Eesti Keele Instituut
marja@eki.ee

1. Kohanimedede liigitamine nime objekti järgi: loodusnimed, kultuurnimed, artefaktide nimed. Probleemid piiride tõmbamisega objektide vahel. Keeleline piir eesti kohanimedede objektide järgi liigitamisel ja selle nihe.

2. Kohanimedede liigitamine nimeelementide (nimes sisalduvate sõnatüvede) järgi: lihtnimed kui ühest tüvest koosnevad kohanimed, liitnimed kui mitmest sõnatüvest koosnevad kohanimed. Sufiksiliste kohanimedede kahetine käsitlemine. Topoformantidesse suhtumise erinevus diakroonias ja sünkroonias. Keeleline nimemudel K1Nom/Gen + K2Nom/Gen eesti kohanimedede puhul. Nime-tüpoloogia ja semantilise kriteeriumi sissetoomine kohanimedede edasisel liigitamisel.

3. Nn süntaktiline liigitus, selle positiivsed ja negatiivsed küljed. Primaar- ja sekundaarnimed – ebaõnnestunult oluline jaotus eesti ortograafia jaoks.

4. Mõned visioonid, kuidas seletada süntaktiliste ja kohanimes sisalduvate sõnatüvede järgi liigitades saadud erinevaid tulemusi. Rahvaetümolooogia kui oluline tegur kohanime kujunemisel: kohanime arengut võib käsitleda kui rahvaetümolooogiate rida. Võimatus teha kindlaks eesti kohanimedede algset kuju.

5. Kaks kohanimekoosluse tüüpi tänapäeva eesti keeles. Kohanimekooslus kui mingil konkreetsel territooriumil elava inimhulga poolt tarvitavad kohanimed, mis moodustavad sel alal terviku. Otstarbekas on Eestis võtta niisuguseks territooriumiks ühe küla maa ja selliseks inimhulgaks ühe küla rahvas. Iseseisvate (onomastiliselt tasandilt tuletamata) loodus- ja viljelusnimedega tüüp – vanem. Onomastiliselt tasandilt tuletatud tüüp – uuem. Nimevara käsitlemine homogeensena või heterogeensena.

Sõjandusterminite klassifitseerimise probleeme

Karin Kaljumägi
Kaitseväe Ühendatud Õppeasutused
karinka@ut.ee

Ettekanne käsitleb terminite klassifitseerimist Kaitseväe Ühendatud Õppeasutustes koostatava võrgusõnastiku näitel.

Ettekandes:

- tutvustatakse klassifikaatorit ja
- kriteeriume, mille alusel otsustada sõna „klassikuuluvuse” üle,
- arutletakse üleminekuala üle.

Mööndlausete klassifitseerimisvõimalustest eesti ja vene keele näitel

Katrin Karu
Tartu Ülikool
katrin.karu@ut.ee

Nii eesti kui ka vene keeles eksisteerib lausetüüp, mida vastava keele grammatikas nimetatakse mööndlauseks. Mööndlausel on ühelt poolt spetsiifiline tähendus, teiselt poolt spetsiifiline grammatiline markeering. Ettekandes tulevad vaatluse alla eelkõige semantikal põhinevad klassifikatsiooni võimalused.

Eesti keeles on mööndlauseid kirjeldatud EKGs, kus neid liigitatakse *vaatamata-* ja *olenemata*-tüübiks (EKG 1993: 308–309). Mööndlauset eesti keeles on uurinud ka nt Lehte Rannut (1978).

Vene grammatikas on olnud mitmeid katseid mööndlausete liigitamiseks (vt kirjandust). Kõigis viidatud allikates on mööndlausetele lähenetud erinevast vaatevinklist. Ettekandes püütakse analüüsida olemasolevaid klassifikatsioone nii formaalsest kui ka semantilistest aspektist lähtudes ning välja pakkuda mööndlausete liigitusvõimalus, mis oleks kasutatav erineva struktuuriga keelte kirjeldamisel.

Kirjandus

EKG = Eesti keele grammatika. Tallinn, 1993.

Rannut, L. 1978. Eesti keele mööndlausest. *ETA Toimetised*, 27, ? 1, Tallinn, 47–53.

??????, ?. ?. 1999. ?????????????? ? ??? ?? ? ?????? ?? ?????????? ????????. ? ?????? ?? ??????????. ? 5, ? 24–45.

????? ?????? ?????? ?????? ?????????????? ??? ?? ., 1970.

???????, ?. ?. 1997. ?????????? ?????????????????? ? ?????????????? ??????? ???? ? ??????? ?????? ?????????????????? ??????????. ? ? ?.

????, ? . ? . 1986. ?????? ????????? ???? ???? ?????? ????: ?
???????? ????????????????? ???? ???? .
?????? ???? ?????. ? ., 1982. ? . 2.
???????, ? . ? . 1986. ?????? ????????????????? ? ?? ?? ??? ???? ?
????? ?????? ?????? ?? ?? . ? .

The present-day Situation of Ulster-Scots

John M. Kirk
Queen's University Belfast
<http://www.qub.ac.uk/en/johnmkirk/>

In the context of a project devoted to minority languages in countries emerging from conflict, I will set out the case of Ulster-Scots very briefly under the following headings:

The status of Ulster-Scots
Legislation for Ulster-Scots
Practicalities for Ulster-Scots
Tensions surrounding Ulster-Scots

References

- Kirk, J. M.** 1998. Ulster-Scots: realities and Myths. – *Ulster Folklife* 44, 69–93
- Kirk, J. M.** Two Ullans Texts. – *Language and Politics: Northern Ireland, the Republic of Ireland, and Scotland*. Kirk, J. M. and Ó Baoill, D. P. (eds.) Belfast: Cló Ollscoil na Banríona. 33–44
- Kirk, J. M.** 2003. Archipelagic Glotto-Politics: the Scotstacht. – *The Celtic Englishes III*. Tristram, H. L. C. (ed.). Heidelberg: Carl Winter. 339–356

Iseorganiseeruvad keelelemendid eesti keele emotsioonisõnavara näitel

Toomas Kirt, Ene Vainik
Eesti Keele Instituut, Tallinna Tehnikaülikool
Ene.Vainik@eki.ee, Toomas.Kirt@mail.ee

Keel on protsess, milles kaose ja korra elemendid dünaamiliselt põimuvad. Ei saa väita, nagu oleks keel läbini korrastatud ja alluks ainult rangele struktuurile, samuti ei saa ka väita, nagu oleks keel läbini kaootiline ja ennustamatu.

Strukturalistlikust taustast välja kasvanud põlvkonnana ning ajendatuna teaduse kõrgematest tõe tunnetamise ideaalidest on meil tung kõike liigitada ja klassifitseerida, usume nähtuste põhimõttelisse ennustatavusse, millest on saanud ka teooria teaduslikkuse kriteerium. Küsimust nähakse vaid kategoriseerimise protseduurides ning põhimõtetes ja teooriate täiustamises. Keeleelementide tuska-toovad kategoriseerimisprobleemid usutakse lahenduvat, kui leitakse õigemad ja paremad liigitamise ja süstematiseerimise viisid. Keegi ei küsi, kas keel üldse vajab mingisugust välist klassifitseerimist, keegi ei küsi, kas keelenähtused vajavad ennustamist.

Kuidas küsida keelelt, kas ta soovib sisemiselt liigenduda, kuidas küsida keeleelementidelt, kas nad soovivad omavahel grupeeruda ja allianssi astuda? Üks võimalik vastus on proovida, mis juhtub, kui lasta keeleelementidel ise organiseeruda. Iseorganiseerumise ja mittelineaarse dünaamika mõiste pärineb üldisest süsteemiteooriast (nt Bertalanffy 1968) ning kaose filosoofiast (nt Bohm 1980), neid ideesid on edukalt rakendatud loodusteadustes, nüüd aga ka juba ka sotsiaalteadustes, nt organisatsioonikäitumise käsitlustes (Wheatley 2002), psühhoterapiakäsitlustes (nt Butz, Carlson & Carlson 1998) ja sellesuunalisi lähenemisi on keeleteaduseski (nt Skousen 1998). Kahekümnnes sajand on arvutitehnoloogia näol andnud uued võimalused iseorganiseerumisprotsesside modelleerimiseks, üks variantidest on nn iseorganiseeruvate kaartide meetod (Kohonen

2000), andmete reduktsioonimeetod, mis kasutab närvivõrgustiku analoogiat.

Ettekandes kirjeldatakse, mis juhtus, kui autorid lasid osal eesti keele emotsioonisõnavarast iseorganiseeruda, kasutades iseorganiseeruvate kaartide meetodit ja lähtudes tavakõnelejustest informantide hinnangutest sõnade tähendustele. Aine spetsiifikast johtuvalt osutus valdavaks jaotus positiivseteks ja negatiivseteks emotsioonisõnadeks, kuid autoreid huvitas, mis tegurid veel peale „hea ja kurja tundmise” mõjutavad emotsioonikategooria iseorganiseerumist.

On põhjust oletada, et iseorganiseermisprotsessidel on keelenähtuste tekkes, kujunemises ja dünaamikas suurem roll, kui seni on üldiselt osatud või tahetud mõelda. Käesolev ettekanne on ainult väike osutus alternatiivsele viisile, kuidas keelenähtustestki mõelda võiks. Ja keegi ei tea, millega see kõik veel ükskord lõpeb.

Viited

- Bohm, D.** 1980. Wholeness and the implicate order. New York: Ark.
- Butz, M. R.; Carlson, J. M.; Carlson, Jon** 1998. Chaos theory: Self-organization and symbolic representation in family systems. – Family Journal 6 (2), 106–126.
- Kohonen, T.** 2000. Self-organizing maps. 3rd ed. Berlin: Springer.
- Skousen, R.** 1989. Analogical modelling of language. Kluwer Academic Publishers: Dordrecht.
- Wheatley, M. J.** 2002. Juhtimine ja loodusteadus: korra otsingul kaootilises maailmas. Tartu: Fontese Kirjastus.

Foneemiklassid ja häälduspraktika

Einar Kraut
Eesti Raadio
einar.kraut@mail.ee

Ettekanne vaatleb foneeme ja nende liigitamist kõnetehnika ja praktilise hääldusõppe seisukohast. Töö neis valdkondades annab tulemusi juhul, kui õpetust kannab pädev arusaam keele hääldusstruktuurist. Autor selgitab mõningaid probleemsemate artikuleerimisvõtete üldistamisvajadusest väljakasvanud ideid eesti foneemiklasside tõlgendamisel.

Vajadus hääldusliigutusi kontrollida-korrigeerida toob fonoloogiatasandilgi välja siduva ehk konnektiivse aspekti. Olgugi keelemärkide häälduslik ehitus olemuslikult diskreetne, ei kujuta tegelik hääldamine endast kunagi isoleeritud üksiksegmentide jada – igas konkreetse hääldussüsteemis on vähemalt sama tähtsal kohal häälduselundite see töö, mis eri häälikute moodustusasenditest kujundab ühtse terviku. Häälduse keeleomase kvaliteedi kujundajana (vrd aktsent) on konnektiivne pool määrav.

Eesti keele kõnekorrektsiooni ja hääldusõppe kõige nõudlikuma ülesandena näeb autor liigse häälduspinge mahavõtmist, sest ainult sel moel õnnestub selgelt artikuleeritud foneemivariantidest kujundada ladus, vormikindel kõnetervik. Vajadus lõdvestada häälduselund rõhujärgses hääldusfaasis on osutanud kahe, foneetikuil seni vaidlusaluse või siis puht koartikulatiivseks arvatud hääldusüksuse tegelikku kuulumist fonoloogilisse struktuuri: need on vokaalieelsed [j], [w] sõnades, nagu *käia*, *siia*, *õue*, *tuues*. Nagu mõõtmised kinnitavad, ei erine nende siirdeosiste moodustusasend tõepoolest vahetult eelneva kõrge vokaali asendist ($i=j$, $u=w$). Silbipiiri dünaamiline selgus ühelt poolt ja teiseks eesti vokaalide üldine naabrusekeeld sõnasisesel silbipiiril veenavad, et tegemist saab olla vaid poolvokaalsete geminaatidega /jj ww/.

Nõue näha keelesüsteemi tervikuna tingib eesti keele pikkade vokaalsegmentide analüüsis järgmised loogikasammud. 1) Fonooloogiliselt kujutavad *i-*, *u-lõpulised* difongid endast vokaali ühendit poolvokaalse konsonandiga ning seda kõigil juhtudel (mitte ainult vokaali ees): /[^]äjja äjj [˘]õww õws õjs kawss/ – vrd rõhujärgse ehitusega sõnades *ämm nõmm õrs varss*. Pikad avahäälikud *ii*, *uu* esindavad sellesama struktuuri erijuhte. 2) Kui järelliige on konsonant kõrgetes pikkvokaalides *ii*, *uu*, on ta seda ka kõigis ülejäänutes: /a: ä: e: o:/ jt. Viimaste puhul seisneb erinevus vaid jätkufoneemi /:/ enda võimetuses gemineeruda – vrd homorgaansete teisendustega /si:n – s`ijja/, /u:s – uwwe/, geneetiliselt ka /h`ü:ta – hüjjap <*hüyyap/ – vrd *künda – künnab*. 3) Kui järgosis on konsonant *i-*, *u-lõpulistes* diftongides, on ta seda ka *e-*, *a-* ja *o-lõpulistes* – faktiliselt häälduvad kõik diftongid tänapäeva eesti keeles ühtse dünaamikamalli alusel.

Eesti häälduse tegelik korrastatus osutub seni eeldatust mitmeti põhjanevamaks. 1) Geminatsioon kui üks süsteemi kandetalasid hõlmab läbivalt mõlemad peamised konsonandiklassid: tõkkhäälikud /ptks/ ja kõlahäälikud /mnlrjw/ (senikasutatud terminid: obstruendid, sonorandid). Gemineerumistingimuste lahknevusest johtub kummagi häälikuklassi erinev koht süsteemis. Foneemi /w/ erandlikkus ilmneb positsioonilises lahknemises [v]/[w] (komplementaarne distributsioon); /j/ ei saa esineda sõna lõpus pika vokaalosis järel. 2) Vokaalosis põhimõtteline (fonoloogiline) lühidus laieneb kõigile silbipositsioonidele – pearõhusilp on eriline vaid üliulatuslike konsonandikombinatsioonide poolest, võrreldes järgsilpidega. Igasuguse, ka pikkvokaali või diftongi sisaldava pikksilbi kogu see osa, mis järgneb esimesele vokaalsegmentile (silbisõlm – teistes käsitlustes ka silbilõpp või -kooda), kujutab endast dünaamilises mõttes konsonantset hääldusfaasi. 3) Silbisõlmes kui dünaamilises konsonantruumis esineva foneetilise variatiivsuse taandamisel on otsustav roll eesti keele keskse hääldusfenomeni, vältesüsteemi ratsionaalsel mõistmisel ja tulemuslikul õpetamisel.

Ettekandes tuuakse näiteid põhilistest hääldus- ja diktsiooni-
vigadest, mille vältimist ja parandamist kirjeldatud lähenemisviis
hõlbustab. Esitatakse konkreetsed artikulatsioonifoneetilised kritee-
riumid, mis lasevad avahäälikuid konsonantidena tajuda ning
määratleda.

Ühest objektsüsteemide struktureerimise efektiivsest meetodist

Rein Kuusik, Leo Võhandu
TTÜ Informaatikainstituut
kuusik@cc.ttu.ee, leov@staff.ttu.ee

Klassikaline statistika on üha kiiremini kuhjuva infotulva ees vägisi jänni jäämas. Suhteliselt siledalt kulgenud või silutud nähtuste modelleerimise asemele on aiva ägedamalt tunginud NP-keerulised kombinatoorsed meetodid (nn avastusteadus (*discovery science*)). Klassikalised optimeerimismetodid on asendunud bioloogiliste ja kultuurialgoritmidega, mille võimsus on sadu kordi suurem. Peamine on seejuures asjaolu, et keeruliste nähtuste süvastruktuuri korrapärasid on nende meetodite abil hoopis kergem avastada.

Muude rahvusvahelises kirjanduses esinevate uute meetodite hulgas on kindla koha võitnud ka meie instituudis loodud nn monotoonsete süsteemide meetod. (Mullat, Võhandu, Kuusik jpt).

Selle meetodi põhiidee on ülimalt lihtne:

1. Objektsüsteemi iga objekti jaoks leidub mõõt, mis on nõrgalt monotoonne, st kui muuta süsteemi ühe objekti mõõtu pluss- või miinussuunas, siis teiste objektide mõõdud muutuvad samas suunas või jäävad konstantseks.
2. Süsteemist kõrvaldatakse selle mõõdu järgi nõrgim objekt ja süsteemi mõõdud arvutatakse uuesti. Seda tegevust korratakse seni, kuni kõik objektid on elimineeritud.

Sama protseduuri võib korrata ka objekte kirjeldavate tunnuste puhul. Lõpuks korrastatakse nii objektid kui tunnused elimineerimisjärjekorda ümber. Süsteemi struktuur avaneb lokaalsete maksimumide kohal tükeldusi tehes. Meetodi olemust kirjeldatakse lihtsa näitega. Demonstreeritakse korrastuse seost Hammingi meetrikaga ja saavutatavat ajavõitu.

Meetodi tööaeg on alati polünomiaalne. Saadud struktuuri-tükeldus on kasutatud mõõdu suhtes alati optimaalne.

Esitatud lihtsast põhiideest on arendatud terve meetodite pere, mille abil saab edukalt uurida mitmesuguseid objekt-tunnus- andme-tabeleid, genereerida automaatselt hüpoteese, leida graafide klikke, üldistada tesauruse päringuid katvuse parandamiseks, uurida tekste mitmel eri moel jne. (Meenutame, et graaf on igasuguste suhete kirjeldamiseks universaalselt sobiv mudel)

Ettekandes tuuakse ka tuutornäide hüpoteeside genereerimise kohta.

Lõpuks visandatakse üldistatud töökeskkond ja näidatakse paari kena demopilti keerulise struktuuriga objektsüsteemide olemuse avamise kohta.

Monotoonsed süsteemid ja tekst

Rein Kuusik, Leo Vöhandu
TTÜ Informaatikainstituut
kuusik@cc.ttu.ee, leov@staff.ttu.ee

Tekst on üpris raske objekt analüüsimiseks. Võimalike vaadete hulk tekstile on vist küll piiramatu. Käesoleva konverentsi korraldajad oma kenas käivituskutses viitasid hämmastavalt laiale probleemiringile, mis ootab toetust ja loodab edeneda mitmesuguste klassifitseerimismeetodite kasutamise abil.

Me ei ole teksti uurimisele spetsiaalselt pühendunud, kuid paljude konsultatsioonide käigus ja juhendamisel on siiski mõningaid kontaktkohti tekkinud. Üldised meetodid on spetsiaaluurimustes korraka nii head kui halvad. Head selles mõttes, et meetodika ei sõltu konkreetsest uurimuseesmärgist ning saadud struktuuripilt annab autorite mitmekümneaastase kogemuse põhjal uurijale alati midagi uut. Häda on aga selles, et see uus tuleb alles lahti mõtestada, sest hoolimata tuttavate kirjelduselementide kasutamisest peitub saadud struktuuripiltides alati midagi hämmastavat, ebatraditsioonilist. Iga uus nähtus aga tekitab alati probleeme.

Kõige esimeseks monotoonsete süsteemide (MS) rakenduseks teksti analüüsimisel võib pidada K. Ääremaa tööd juriidilise info-süsteemi JURIOS loomisel. Teise autori aspirandina kasutas ta juriidilise tesaurusega seotud päringusüsteemis stardipäringu automaatseks laiendamiseks süsteemi terminite temaatilise seostatuse mõõdu infot. Näiteks „lahutamine” laieneb loomulikul viisil „mees, naine, laps, abielupool” jne). Sellist laiendamist võiks teha kuitahes kaugele, loomuliku laienduspiiri andis MS-is kasutatav tuuma mõiste. (Mõiste piir).

Huvitav oli ka V. Sarve doktoritöö rahvalaulust, kus MS abil saadud muustrite skeemid tõid selgelt välja laulustiili ajalised muutused.

Teksti uurimise juures on raskendavaks veel asjaolu, et me ei saa nõuda sõnade ega lausete kirjelduse homogeensust (ühetähendus-

likkust ega sama pikkust). See tähendab, et teksti ei saa esitada andmetabelina. Meie õnneks selgub, et seda polegi tarvis. Peaasi, et on täidetud monotoonsuse nõue ja on määratletud täheühendi kui mustri olemus. Kasutades MS-del baseeruvat mustrite avastamise meetodit, saame eraldada kõikvõimalikud tekstis esinevad täheühendid, nn N-grammid nii sõna alguse suhtes kui asukohast sõltumatult. Algoritmid töötavad üpris kiiresti, keskmiselt „musterdatakse” 1 MB teksti (ca 400 lk) 3–4 sekundiga.

Ettekandes esitatakse näiteid MS kasutamise kohta teksti analüüsimisel. Huvitav on näiteks uurida ka autori kirjutamisstiili, millele otsa vaatamiseks kasutame ainult kirjavahemärkide ja muude sümbolite rivi. Visandatakse võimalus eesti keele sõnaperede süsteemi loomise toetamiseks jne.

Sõnede jada asemel võime teksti käsitleda lauseehituslikult, sõnatüüpide jadana. Sellele jadale saab otse rakendada monotoonseid mustriuurimise meetodeid. Tuleb luua mustrite skeem, uurida seda lähemalt ja teha vajalikud ja võimalikud üldistused.

Loodud meetoodika on rakendatav igasuguste tekstiteisenduste järel.

Süntaktiline info sõnastikes, eriti EXS-s

Margit Langemets, Merike Mägedi, Ülle Viks
Eesti Keele Instituut
margit@eki.ee, merike@eki.ee, Ylle.Viks@eki.ee

Sissejuhatus

Emakeelekõneleja valdab tuhandeid sõnu. Sõna kasutus eeldab, et inimene peab teadma nii tolle hääldust, õigekirja, muutmist, süntaktilist käitumist, koosinemist mingite kindlate sõnadega, põhitähendusi, tähendusvarjundeid, selle sõna semantilisi suhteid muude sõnadega jms. Sõnastik on koht, kus sellist tarkust on võimalik iga sõna kohta kõige paremini esitada. Sõnastiku otstarbest ja oletatava kasutaja vajadustest sõltub, kui palju andmeid ja millisel kujul tegelikult sõnastiklisse pannakse.

1. EXS

EKI-s on käivitunud projekt Eesti–X-keele sõnastik (EXS), mille käigus valmistatakse ette kakskeelse sõnaraamatu eesti pool, mis saab edaspidi aluseks konkreetsetele eesti lähtekeelele tõlkesõnaraamatutele. EXS tahab olla universaalne, st ta ei sõltu vastete keelest. Iga uue Eesti–X-keele sõnaraamatu koostamisel on võimalik sõnastikubaasis olevat infot vastavalt vajadusele muuta: võtta osa tähendusi kokku (või lisada uusi), jätta osa grammatilist infot (või kõik) üle võtmata, jätta kasutamata (või lisada) tekstinäiteid. Jne. See tähendab seda, et EXS peab märksõnu kirjeldama väga põhjalikult, et sealt saaks vastavalt koostatava sõnastiku iseloomule valikuid teha.

EXS maht on esialgu ca 40 000 märksõna, kuid edaspidi luuakse ka suurema ja väiksema mahuga versioonid. EXS-i koostatakse andmestruktuuride kaupa (mitte alfabeetiliselt), sest see annab võimaluse keskenduda eraldi igale probleemiringile. Meie ettekanne keskendub grammatilisele infole, peamiselt süntaktilisele.

2. Grammatiline info

Eri sõnastikes võib leida grammatilist teavet mitmelt keelekirjeldustasandilt:

- ortograafia: sõna tähtkoostis, kokku-/lahkukirjutamine, suur/väike algustäht
- foneetika: hääldus (tsitaatsõnadel), palatalisatsioon
- morfofonoloogia: välde, rõhk
- morfoloogia: märksõna vorm, vormimoodustus (muuttüüp, põhivormid, erandid), vormikasutus
- derivatsioon: liitsõna-liigendus, tuletusalus, tuletustüüp
- süntaks: sõnaliik, sõltuvusnõuded
- semantika: semantilised tunnused, teaurusinfo

3. Süntakiline info

Süntakiline info on seotud sõnade käitumisega lauses. Lause struktuuriline keskus, põhielement, millest sõltub (liht)lause ehitus, on verb. Seetõttu mõtlemeegi sõnastiku süntakilise info all eelkõige verbi sõltuvusseoseid ehk seda, milliseid laiendeid ja millisel kujul verb oma tähendusega nõuab. Jõudumööda kavatseme hiljem lisada ka teiste sõnaliikide sõltuvusseosed.

Ettekandes räägime lühidalt süntakilise info senisest esitamisest eesti sõnastikes ja mujal sõnastikumaailmas. Lähemalt tuleb juttu sellest, mida ja kuidas plaanitakse esitada EXS-s ja millised on suuremad probleemid.

3.1. Sõnaliikide puhul on tegemist peamiselt klassifitseerimisraskustega, sageli on raske määrata, millisesse liiki sõna kuulub (nt *haige* adj/subst), kas tegemist on sõna muutevormiga või iseseisva sõnaga (nt *jalga*, *pähe*). Mõnikord oleks vaja sõnaliike detailsemalt liigitada, nt pronoomenid ja numeraalid käituvad substantiivina (*see*, *kolm*) või adjektiivina (*selline*, *kolmas*).

3.2. Sõltuvussuhete esitamine sõnastikus on iseenesest tülikas. Verb määrab oma laiendid mitmel eri tasandil: süntakiline funktsioon (lauseliikmed), temaatiline roll (aktandid), morfoloogiline vorm (rektsioon). Adekvaatse pildi loomiseks on vaja arvestada kõiki tasandeid, isegi kui neid alati kirja ei panda (nt subjekti vorm on vaikumisi nominatiiv).

Üks väga esialgne ja lünklik näide (XML-märgendusega):

<msona>**kirjutama**</msona> <sliik>**V**</sliik>
<txh_nr>**1.**</txh_nr> <def>**kirja panema**</def>
<subj_roll>**agent**</subj_roll>
<obj_roll>**teema**</obj_roll> <obj_rekts>**NomGenPart /
kõrvallause**</obj_rekts>
<advb_roll>**instrument**</advb_roll>
<advb_rekts>**Kom**</advb_rekts>
<advb_roll>**koht:kuhu**</advb_roll>
<advb_rekts>**vormiklass:kuhu**</advb_rekts>

Probleeme on palju, nt on keeruline otsustada, kas mingi laiend on kohustuslik või mitte, kas objekt esineb ainult partitiivis (osaobjekt) või ka nominatiivis ja genitiivis (täisobjekt) jne.

4. Semantiline info

Lõpetuseks pisut semantikast, niivõrd kui see on seotud süntaksiga. Sageli määrab verb ka selle, millised tähenduselemendid (semantilised tunnused) peavad olema teda laiendaval noomenil, nt verb *kärisema* eeldab elutut subjekti, verb *kirjutama* aga nõuab subjektiks inimest. Kui verbi kirjes temaatilised rollid on olemas, siis ilmselt pole semantilisi tunnuseid tarvis eraldi märkida, sest temaatilise rolli mõistes on need juba sees (*agent on elus*). Küll aga tuleb tunnused esitada noomenite sõnaartiklites, nt *kirjanik* [inimene], *riie* [elutu] jne.

Politically correct terms: pros and cons

Suliko Liiv
Tallinn Pedagogical University
liiv@tpu.ee

The use of „politically correct” terms and „appropriate” phrases has been in the centre of attention for some time already. How to choose one phrase over another? Is the expression you use the „right” one? Is it possible to insult someone unintentionally while using the „wrong” term? We must be careful and not use maliciously offensive terms to offend some groups of people.

While using the politically correct terms we are not at all sure whether this new language causes more negative or positive attitudes. „Should I call you black or African American? Do you want to be known as vertically challenged or short? Does it hurt your feelings when I call things gay?”, asks Jennifer Mertens.

The paper will give a short overview of national stereotyped expression (*Dutch treat, French sickness, Irishman’s dinner*, etc), the cases where the word „man” is substituted for another word (*man-made – manufactured, mankind – people, humanity, chairman – chairperson, mailman – mail carrier, policeman – police officer*, etc), the use of personal pronouns *he* and *she* and many other cases.

The samples of „politically correct” new year cards and bedside stories will be introduced.

Sõnaklasside liigituse küsimusi eesti murrete korpuses

Liina Lindström, Mari-Liis Kalvik, Rutt Läänemets, Mari Mets, Karl Pajusalu, Pire Teras, Ann Veismann, Eva Velsker
Tartu Ülikool
liina@murre.ut.ee

Eesti murrete korpus on elektrooniline andmekogu, mis sisaldab helilindistatud ja litereeritud murdetekste (vt ülevaated Lindström, Pajusalu 2003; Lindström jt 2001). Viimaste aastate jooksul oleme tegelenud ka murdetekstide morfoloogilise märgendamisega, mille üheks oluliseks küsimuseks on sõnaliikide või sõnaklasside määratlemine ja piiritlemine.

Morfoloogilise märgendamise aluste väljatöötamisel oleme osalt lähtunud EKIs tehtud Hargla murraku morfoloogilise andmebaasi eeskujust (vt <http://www.eki.ee/dict/hargla/>). Murdekorpusse jaoks oleme seal kasutatud liigitust detailsemaks muutnud ning lisanud suulise kõne eripärast tingitud sõnaklasse (diskursuspartikkel, suhtlussõna).

Sõnaklasside piiritlemisel on esile kerkinud hulk probleeme. Esitame ettekandes mõned neist:

1) nn klassikalised sõnaliigiprobleemid:

- adverb või substantiivivorm ajamäärusena: *ja `kykkõ tet'ti ja talvõl siss rabatti linnu ja* (HAR); *uus p- `piibli+lugu tõse talve ja usu+õppõttuss ja tõstameint* (RÕN). Lihtsam on juhtudel, kus ajamäärusena funktsioneeriv sõnavorm eristub selgelt muust paradigmat, sel juhul võime seda pidada adverbistumiseks;

- substantiiv või kaassõna: kaassõna on suletud sõnaklass, millel on selge grammatiline funktsioon ning mis kuulub käändsõna juurde. Siiski ei ole kaassõnade hulk mingi väga kindlate piiridega sõnadekogum, sest grammatikaliseerumisprotsessi tulemusena võib sinna pidevalt uusi sõnu lisanduda. Samuti võib kaassõnade loend eri murretes olla erinev. Kaassõnade ja substantiivide piir ei ole alati

selge, nt *sääl tuu tõse tii veeren miss jõe puul't lätt* (HAR), *lei pin'i jala `perrä* 'löi peni jala pihta' (SE), selgelt kaassõnaline kasutus: *lät's'ivä siss leevä `perrä* (SE).

2) Pro-klassidega seotud küsimused:

- Oleme eristanud prosubstantiivi, proadjektiivi, proadverbi. Vaja oleks ka kategooriat pronumeraal (peamiselt sõna *mitu* jaoks).

- Prosubstantiivi ja proadjektiivi piirimaile jääb nii tänapäeva suulises kõnes kui murretes laialt kasutatav artiklilaadne *see* ja *tuu: tu `tütrükkõnõ `kas'v'i siss suurõst* (SE).

- Eraldi sõnaklassina oleme võtnud kasutusele kategooria „küsisõna”. Sellesse klassi kuulub nii küsisõna *kas*, relatiiv-interrogatiivpronoomenid *kes* ja *mis* kui ka küsiv-siduvad adverbid *kus*, *millal*, *kuidas*, *mitu* jne. Ka siin on piirialad, kus neid sõnu kasutatakse pigem pronoomenina: *e no tuust midä* 'midagi' *inämb õs saaq*.

3) Sõnapüüri määratlemine:

- kuidas kategoriseerida sõnu, mis on hääldatud lahku – kas märgendada ühe liitsõnana või eri sõnadena: *kõevu* (.) *kõevu puu* (VÕN);

- eriti Setus on palju lühenenud vorme, mis muudes murretes võivad olla sõnaühendid: nt *nakkavat tõist `käppä imemä nikka omm kevväi jahh niij* (SE) 'nii kaua'.

Viited

- Lindström, Liina; Varje Lonn, Mari Mets, Karl Pajusalu, Pire Teras, Ann Veismann, Eva Velsker, Jüri Viikberg** 2001. Eesti murrete korpus ja kolme murde sagedasema sõnavara võrdlus. – Keele kannul. Pühendusteos Mati Ereli 60. sünnipäevaks. TÜ eesti keele õppetooli toimetised 17. Toim. R. Kasik, Tartu, 186–211.
- Lindström, Liina; Karl Pajusalu** 2003. Corpus of Estonian Dialects and the Estonian Vowel System. – *Linguistica Uralica* nr 4, 241–257.

Eestikeelse kõnetuvastuse andmebaas: lõpparuanne

Einar Meister, Jürgen Lasn, Lya Meister
TTÜ Küberneetika Instituut
einar@ioc.ee, jyrgen@phon.ioc.ee, lya@phon.ioc.ee

Resümee

Kõne- ja kõnelejatuvastuse uuringuteks ning tuvastussüsteemide arendamiseks koostatava andmebaasi loomine on jõudnud lõpusirgele. Projekti planeeritud kestuseks oli 24 kuud ja see jagunes neljaks põhiliseks etapiks:

1. ettevalmistus – arendati välja salvestuste tehniline keskkond, fikseeriti andmebaasi struktuur, koostati tekstikorpused, loodi kõnelejate registreerimissüsteem;
2. salvestused – kõnelejate värbamine, salvestuste teostamine ja nende kvaliteedikontroll;
3. märgendamine – kõigi kvaliteetsete salvestuste ortograafiline märgendamine;
4. lõpetamine – signaalide varukoopiate tegemine, dokumentatsiooni koostamine.

Tehniline keskkond

Salvestused teostati ISDN-telefonikanali kaudu otse arvuti kõvakettale automaatse salvestusprogrammi abil. EMT, Elioni ja Radiolinja klientidele oli helistamine tasuta.

Salvestusteks koostatud tekstikorpus sisaldab mitmekesist materjali, sealhulgas:

- isoleeritud numbrid;
- numbrijadad (PIN-koodid, pangaarve numbrid, jne);
- kellaajad, nädalapäevad, kuud;
- rahasummad ja -ühikud;
- IT-valdkonnas sagedamini kasutatavad sõnad;
- linnade, asulate, firmade nimed;

- foneetiliselt tasakaalustatud laused;
- foneetiliselt rikkad sõnad.

Andmebaasi struktuur

Kui kõnetuvastuse tarvis on vajalik võimalikult palju erinevatelt inimestelt salvestatud kõnenäiteid, siis kõneleja verifitseerimise jaoks on vajalikud samadelt isikutelt eri aegadel kogutud kõnenäited. Seetõttu koosnebki andmebaas kahest alamkorpusest: **tuvastuskorpusest** ja **verifitseerimiskorpusest**. Neist esimene sisaldab 1335 inimese kõnenäiteid, igalt inimeselt üks salvestus; teine hõlmab 160 inimese korduvsalvestusi, igahelilt vähemalt 5 eri aegadel tehtud salvestust. Iga täielik salvestus koosneb 60 helifailist, mille kogukestus on keskmiselt 5 minutit.

Kõnelejate valik

Projektis püüti tagada, et salvestuses osalevate isikute valim oleks lähedane eestikeelse elanikkonna soolisele, vanuselisele ja piirkondlikule jaotusele. Salvestustes osalenutest olid 49% mehed ja 51% naised; vanuseline jaotus vahemikus 12 kuni 70 aastat on kooskõlas elanikkonna vastava jaotusega. Piirkondadest on enim esindatud Harjumaa (koos Tallinnaga) – 46% ja Tartumaa (koos Tartuga) – 24% osalejatest, teiste maakondade esindatus on 1–3%.

Kõnelejate värbamiseks kasutati mitmeid meetodeid – artiklid ja intervjuud ajalehtedes, raadios, televisioonis; informatsiooni levitamine avalikes internetipunktides, ülikoolides ja asutustes. Kõige efektiivsemaks osutusid reklaamikampaaniad Postimehe internetiväljaandes. Salvestusteks registreerus kokku 2512 inimest, neist kvaliteetse salvestuse tegid 1335.

Märgendamine

Kõik kvaliteetseks tunnustatud salvestused segmenteeritakse lausungi (lause või sõna) tasemel ja märgendatakse ortograafiliselt (lisatakse helisignaalile vastav transkriptsioon).

Ühe salvestuse (60 signaali) märgendamiseks kulub keskmiselt 20 minutit, seega kulub kõigi salvestuste märgendamiseks $20 \times 2356 = 47120$ minutit ehk ligikaudu 785 töötundi.

Märgendamisel luuakse igale signaalifailile vastav fikseeritud struktuuriga märgendfail, mis sisaldab informatsiooni salvestuse aja ja tehniliste parameetrite kohta, kõneleja-spetsiifilist informatsiooni ning märgendite andmeid.

Kokkuvõtteks

Projekt on kulgenud plaanipäraselt ja projekti alguses püstitatud eesmärk – koguda 1000 inimese kõnenäiteid – on saavutatud ja ületatud. Töö käigus seatud uue eesmärgi – 2000 inimese salvestused – täitmine osutus siiski ebareaalseks.

Pausid ja pauseeelsed taktipikendused kõnes

Meelis Mihkla
Eesti Keele Instituut
meelis@eki.ee

Pause ja pauseeelseid taktipikendusi eestikeelses kõnes on foneetikaalastes töödes käsitletud suhteliselt põgusalt või riivamisi üldisemate hüpoteeside tõestustega seoses (Lehiste 1982; Krull 1997; Eek, Meister 2003), süstemaatiline lähenemine seni puudub. Ettekandes uuritakse pause ning fraasi- ja lauselõpupikendusi kõnesünteesi kontekstis. On ju pausid ja kõnetempo muutused olulised sünteesikõne rütmi loomulikkuse tagamisel (Campell 2000).

Lähtematerjaliks on näitleja ette loetud ilukirjandusliku teksti ja diktorite loetud raadiouudiste salvestised. Ettekandes käsitletakse, kas ja kuidas saab pause ja pauseeelseid taktipikendusi klassifitseerida sõltuvalt ettelooitava teksti grammatilisest ja süntaktilisest liigendusest, kas pausi kõnes võib käsitleda tavahäälikuna, mille kestust saab prognoosida lähteteksti, põhjal ning taktipikenduste formaliseerimist.

Viited

- Campell, Nick** 2000. Timing in speech: a multilevel process. – Prosody: theory and experiment. Ed. by M. Horne. Dordrecht/Boston/London: Kluwer Academic Publishers, 281–334.
- Eek, Arvo; Meister, Einar** 2003. Foneetilisi katseid ja arutlusi kvantiteedi alalt (I). – Keel ja Kirjandus 11, 815–837; 12, 904–918.
- Krull, Diana** 1997. Prepausal lengthening in Estonian: Evidence from conversational speech. – Estonian Prosody: Papers from a Symposium. Ed. by I. Lehiste, J. Ross. Tallinn, 136–148.
- Lehiste, Ilse** 1982. Sentence and paragraph boundaries in Estonian. – Estonian Papers in Phonetics. Tallinn, 9–13.

Hoolimata normide kohaselt püstitatud ülesandest näivad tulemused teisel pool määratust

Kadri Muischnek, Kadri Vider
Tartu Ülikool
Kadri.Muischnek@ut.ee, Kadri.Vider@ut.ee

Sõnaliigid ja sõnade jaotumine nende vahel on eesti keeleteaduses ikka probleemiks olnud. Nii on näiteks 1977. aasta Keeles ja Kirjanduse 9. numbris publitseeritud dissertatsiooni materjale teemal „Kuidas käsitada sõnaliigi mõistet”. Selle raames on Mati Erelt kirjutanud: „Kõige realistlikum on pidada keeleüksuste klasse, sealhulgas sõnaliike, avatud klassideks, s. o. klassideks, mis ei sisalda mitte ainult selliseid elemente, mis kuuluvad neisse kahtlemata, vaid ka selliseid, mis kuuluvad ühte klassi rohkem kui teise. Mõttekas on kujutleda tervet keelesüsteemi ning selle allsüsteeme organiseerituna /.../ tsentri-perifeeria põhimõttel.”

Oma ettekandes räägime sõnaliikide täpse määratlemise vajadusest ja sellega seotud tegelikest praktilistest probleemidest tekstide morfoloogilise ühestamise kogemuse põhjal. Morfoloogiline ühestamine tähendab teksti igale sõnale antud kontekstis õige morfoloogilise analüüsi lisamist. Mitmese analüüsi jätmine pole võimalik, sest viiks automaatanalüüsi kurnava liiasuseni. Nii tulevad perifeeria määratlematuse ja hajusa ülemineku kitsaskohad ilmsiks hoopis arvutilingvistika ja keeletehnoloogia uurimistöodes. Varasemat sellelaadset tööd on kirjeldatud artiklis „Kas tegelik tekst allub eesti keele morfoloogilistele kirjeldustele” (Kaalep jt 2000).

Aastatel 2002–2003 ühestati morfoloogiliselt TÜ arvutuslingvistika uurimisrühmas 500 000 sõna eestikeelseid tekste. 100 000 sõna sellest moodustas suuline kõne, millega autorid ise ei tegeleenud, nii et meie kogemus põhineb 400 000 sõna morfoloogilisel ühestamisel.

Ühe sõnaliigi – omadussõna – eripärast

Heili Orav
Tartu Ülikool
Heili.Orav@ut.ee

1977. a kirjutas M. Erelt Keeles ja Kirjanduses kokkuvõtte diskussioonist sõnaliikide teemal. Selle diskussiooniga sõnaliigiprobleemide lahendamiseni ei jõutudki. Tõdeti ainult, et primaarne on süntaktiline kriteerium.

Sõnaliikide (eelkõige nimi- ja tegusõnade) semantilistes eripärastest on saadud mingi selgus alles suuremate semantiliste andmebaaside – tesauruste – koostamisel. Omadussõnade kohta on aga erinevatel keeltel erinevad lähenemised, sest selle sõnaliigi leksi-kaalne positsioon on unikaalne ja erineb teistest kategooriatest.

Oma ettekandes keskendun inimest iseloomustavatele omadussõnadele eesti keeles. Analüüsin küsitluse (100-l eestlasel paluti öelda inimest iseloomustavaid omadussõnu) tulemusi, mis näitavad omadussõna liigituskriteeriumite suvalisust.

Kuidas hinnata hindajaid

Hille Pajupuu
Eesti Keele Instituut
hille@eki.ee

2003. aastal eksamineeriti riiklikult 46 339 inimese keeleoskust (riigieksamid, eesti keele tasemeeksamid). Valdavalt koosnevad eksamitööd 4–5 osast, millest kahte – kirjutamist ja rääkimist – hinnatakse hindamisjuhendi järgi. Subjektiivsuse vähendamiseks soovitatakse igat tööd hinnata kahel teineteisest sõltumatul hindajal, suure hindevahe korral kolmandalgi. Seda soovitust järgitakse tase-meeksamite puhul ning riigieksamite kirjutamisosa puhul. Riigi-eksamite suulist osa hindab reeglina üks hindaja.

Et eksamitöid tuleb hinnata suhteliselt lühikese aja jooksul, siis võetakse neid hindama mitukümmend hindajat (nt 2003. a inglise keele riigieksamit hindas 70 hindajat).

Pikaajalise eksamitraditsiooniga maades, nt Suurbritannias, pannakse suurt rõhku hindajakoolitusele. Hindajad hindavad harjutamiseks sadu töid, hindeid võrreldakse ekspertide pandud hinnetega. Hindajate oskust hinnata jälgitakse pidevalt, nt lastakse teatud aja pärast hindajal hinnata uuesti töid, mida ta kord varem on juba hin- nanud ning uuritakse, kas hindaja hindab järjekindlalt või juhuslikult. Juhuslike hindajate asemele koolitatakse uued. Eksamitööde hin- damisel jälgitakse hindajapaari tööd, arvutatakse nende pandud hinnete vahelised korrelatsioonid. Tugev positiivne korrelatsioon näitab, et paar saab ühtemoodi aru sellest, kes on tugev keeleoskaja, kes nõrk. Nõrga ja negatiivse korrelatsiooni puhul mõistavad hin- dajad keeleoskuse all erinevaid asju, seega hinde õigsuses võib kahelda. Teiseks oluliseks näitajaks on hindaja pandud kõigi hinnete keskmine ning standardhälve. Keskmise põhjal saab otsustada, kas hindaja on range, leebe või hindab adekvaatselt. Standardhälve näitab, kuidas hindaja kasutab hindedeskaalat. Liiga väikse standard- hällbega hindaja puhul võib kahtluse alla seada tema oskuse hinde- skaalat interpreteerida.

Eestis piirdub hindajakoolitus tihtipeale mõne(kümne) töö üheskoos hindamisega vahetult enne eksamit. Hindajate töö kohta antakse hinnang muljetele tuginedes – see ei põhine statistilisel analüüsil.

Ettekandes käsitletakse eesti keele tasemeeksami hindajaanalüüsile tuginedes meetodeid hindamise kvaliteedi uurimiseks ja parandamiseks.

Kõrgtasemel eesti keele kõnelejate keelevigade klassifitseerimise võimalusi ja probleeme

Raili Pool, Elle Vaimann
Tartu Ülikool
Raili.Pool@ut.ee, Elle.Vaimann@ut.ee

Analüüsimiseks kasutatav materjal on kogutud Tartu ülikooli eesti keele (võõrkeelena) eriala lõpetajate tänapäeva eesti keele lõpuksamite esseedest. Informantideks on valitud 28 vene rahvusest üliõpilast, kes on enne lõpueksamit sooritanud eesti keele kõrgtaseme eksami.

Klassifikatsiooni aluseks on 665 näitest koosnev korpus, mis sisaldab kõiki juhtumeid, mis meile kui emakeelsetele kõneleajatele ei tundunud vastuvõetavana. Meie eesmärk ei ole uurida vigade ja eksimuste erinevusi ja olemust, vaatleme igasuguseid kõrvalekaldeid emakeelekõneleja keelest. Mõistet *viga* kasutame oma uurimuses tinglikult, selle all mõistame nii kehtestatud keelenormide rikkumist kui intuiitiivselt tajutavaid kõrvalekaldeid.

Kuna analüüsitav materjal on kogutud seotud tekstist, mitte üksiklausetest, on kontekstil oluline roll selles, kas mõnda juhtumit veaks pidada või mitte. Nii võib üksiklause tunduda vastuvõetavana, kuid laiemas kontekstis võib häirivana mõjuda näiteks ajavormi valik.

Esitame omapoolse vigade klassifitseerimise võimaluse, lähtudes nende esinemissagedusest.

- | | |
|-----------------------------|---------------------|
| 1. Sõnajärg 117 | 6. Rektsioon 33 |
| 2. Sõnavalik 94 | 7. Liigne sõna 25 |
| 3. Objekt 66 | 8. Ajavorm 22 |
| 4. Sõnastus 44 | 9. Morfoloogia 21 |
| 5. Pronoomenivormi valik 43 | 10. Käände valik 18 |

- | | |
|-----------------------------------|-------------------------------|
| 11. Kokku- ja lahkukirjutamine 17 | 20. Kõneviis 5 |
| 12. Ühildumine 17 | 21. Eitav konstruktsioon 4 |
| 13. Puuduv sõna 15 | 22. Võrdlusastmed 4 |
| 14. Liigne täht 11 | 23. Puuduv täht 4 |
| 15. Lausete sidumine 7 | 24. Sõnakordus 4 |
| 16. Sõnamoodustus 6 | 25. Poolitamine 3 |
| 17. Arv 5 | 26. Häälikuortograafia 3 |
| 18. Vale täht 5 | 27. Olematu sõna 2 |
| 19. Verbivormi valik 5 | 28. Tähed vales järjekorras 1 |
| | 29. Subjekt 1 |

Klassifitseerimisel tekkisid järgmised probleemid:

1. Mida teha mitmeti analüüsitavate juhtumitega?
2. Mida teha vigadega, mida üks uurijatest on veaks pidanud ja teine mitte?
3. Mille alusel järjestada vigade klassid klassifikatsioonis?
4. Kui detailne peaks olema klassifikatsioon?
5. Millised vead jätta klassifikatsioonist välja?
6. Kas ja kuidas on klassifitseerimisel võimalik arvesse võtta vigade võimalikke põhjusi, näiteks interferentsi?
7. Kas ja kuidas on klassifitseerimisel võimalik arvestada, kas tegemist on kommunikatiivse veaga või mitte?

Ettekanne esitleb ainult üht vigade klassifitseerimise võimalust ning toob esile vigade määramise ja liigitamisega seotud probleeme.

Sõnaliigikimbatusi arvutuslingvistikas

Tiina Puolakainen
Eesti Keele Instituut
tiinap@eki.ee

Sõnaliigi määramise rakenduslikud raskused tulevad esile eelkõige teksti automaatse morfoloogilise analüüsi ja järgneva morfoloogilise ühestamise etappidel, seda nii automaatselt, pool-automaaitselt kui käsitsi toimetades (nt korpuse morfoloogilisel märgendamisel). Morfoloogilise analüüsi valda kuulub näiteks otsustamine, kui paljudele ja millistele omadussõnadele pakkuda ka nimisõna tõlgendamise võimalust. Ühest küljest, praktiliselt iga omadussõna võib teatud kontekstis esineda süntaktilises mõttes nimisõna rollis, kuid teisest küljest, kõigile omadussõnadele teise sõnaliigi tõlgenduse andmine pole ka mõistlik – siis tekib küsimus, mille poolest need kaks sõnaliiki (formaalselt) erinevad, eriti arvestades asjaolu, et nimisõna võib vastupidi esineda süntaktiliselt omadussõna rollis (nt *koer laps*). Kus peaks jooksuma piir, millised sõnad on *haige*-tüüpi ja millised enam mitte? Ja kuidas lahendada olukord teisele poole piiri määratud omadussõnadega, mis esinevad siiski aeg-ajalt nimisõna rollis?

Morfoloogilise ühestamise eesmärk on valida, milline morfoloogilise analüüsi antud tõlgendustest on õige konkreetses kontekstis. Täiendavalt lisandub sellel etapil analüüsi etapist pärit olevatele veel teisi raskusi valmistavaid sõnaliigirühmitusi. Kõige raskemaks neist võib pidada mineviku partitsiipide rühma, mille puhul on valida verbi, omadussõna ja nimisõna tõlgenduse vahel, kusjuures valiku kriteerium on sisuliselt süntaktilist laadi: *nud-tud*-partitsiip on verb, kui ta kuulub öeldise koosseisu; omadussõna, kui täidab täiendi või öeldistäite funktsiooni ning *nu-tu*-nimisõna mitmuse nimetava vorm juhul, kui tegemist on alusega. Rasken-

davaks asjaoluks automaatsel ühestamisel on see, et programm ei saa oma otsustes toetuda tähendusele; lause teiste sõnade morfoloogiline info, tihti samuti mitteühene, on kõik, mida on võimalik kasutada.

Ettekandes tuuakse arvutimorfoloogias raskusi tekitavad sõnaliigirühmad lingvistide ette arutelu algatamiseks ning lahenduste otsimiseks. Automaatanalüüsi puhul on küll võimalik lihtsustada morfoloogilist kirjeldust, mis teatud ülesannete puhul (nt infootsimine või sagedussõnastiku tegemine) võib ka kõige mõistlikum olla. Kuid igal juhul on üheks keeletehnoloogia vahendite tarbijaks lingvistid ise, kes kindlasti soovivad programmidelt oma uuringute tarbeks võimalikult täpset ja detailset informatsiooni.

Kategoriseerimisstrateegiaid teoreetilises keeleteaduses

Heete Sahkai

Eesti Keele Instituut, Tallinna Pedagoogikaülikool
heete@hot.ee

Igasuguse kategooriate süsteemi proovikiviks on nähtused, mis süsteemi ei sobitu – näiteks seeläbi, et ilmutavad üheaegselt mitme kategooria tunnuseid. Ka teoreetilises keeleteaduses on pööratud palju tähelepanu üksustele, mis näivad kuuluvat korraga mitmesse traditsioonilisse kategooriasse. Sellised üksused on näiteks nominaalisatsioonid, partitsiibid, infiniitivid, liitpredikaadid, kliitikud. Ettekandes vaadeldakse mõningaid võimalusi, mida selliste nähtuste kategoriseerimiseks on kasutatud.

DIALANG: Learner-Centred Assessment Using Information Technology

Kari Sajavaara
Centre for Applied Language Studies
University of Jyväskylä
sajavaar@campus.jyu.fi

DIALANG is a new European system for on-line diagnostic language assessment. The system covers 14 European languages (all official EU languages plus Icelandic, Irish, and Norwegian). It delivers tests across the Internet. A prominent feature in DIALANG philosophy is its diagnostic nature: it supports language learners by giving them information about ways in which they can enhance their language acquisition process. It can be used for diagnostic purposes by private persons and institutions.

The aim of the DIALANG system, which reaches from the elementary level up to the near-native, is to provide diagnostic information about their strengths and weaknesses for language learners and users and to help them find ways of improving their language proficiency. The system also is targeted at raising learners' awareness of the level of their language proficiency, of language learning and proficiency in general, and the role of language assessment and tests in the learning process. DIALANG also serves the objectives of life-long learning.

DIALANG is an application of the Common European Framework of Reference for Language Learning, Teaching and Assessment. The assessment scales have been built upon those of the Framework, and the self-assessment statements of the Framework have been modified for the purposes of the DIALANG system. There

are clear indications now of DIALANG providing evidence of the validity of the Framework scales.

DIALANG is the first large-scale assessment system which is available for diagnosis and not certification. It is available for all kinds of learners free of charge without any restrictions of time or place. The Internet makes it easily accessible, and the fact that it is centralised makes the availability of a large item bank possible; the updating is easy; data can be easily gathered for calibration.

The system involves separate sections for reading, writing, listening, structures and vocabulary. The assessment procedure consists of a vocabulary size placement test, selection of the section, self-assessment, responding to tests, and feedback.

The presentation at the Conference will also give examples of DIALANG test items.

The system was developed under the auspices of the SOCRATES Lingua Programme.

The system can be accessed through www.dialang.org.

Märkide klassifitseerimine. Ajalooline ülevaade Hjeltslevi valguses

Silvi Salupere
Tartu Ülikool
silvi11@ut.ee

Vaadeldakse erinevaid märkide klassifikatsioone antiikajast kuni tänapäevani. Põhitähelepanu pööratakse erinevate jaotuste aluste eksplitseerimisele.

Eraldi vaadeldakse L. Hjeltslevi semiootika klassifikatsiooni ja tema poolt välja toodud analüüsiühikute eristamise põhimõtteid.

Kavas on ka püüda vaadelda olemasolevaid märgiklassifikatsioone Hjeltslevi eristuse sisu/ vorm/ väljendus/ substants foonil.

Klassifitseerimisprobleeme saksa- eesti verbireksioonisõnaraamatu koostamisel

Anni Silk
Tartu Ülikool
anni.silk@ut.ee

Sõnaraamatu koostamine, nagu paljud muud „koostamised”, nõuab teatud reeglite järgimist. Et neid reegleid järgida, tuleb need kõigepealt koostada. Saksa-eesti verbireksioonisõnastiku koostamiseks, millega Tartu Ülikooli saksa filoloogia õppetooli töörühm tegeleb, oli vaja paika panna terve hulk põhimõtteid ja kriteeriume, millest lähtudes edasine töö kulgeks. Kõige muu hulgas puudutasid küsimused, mida oli vaja lahendada, otsesemas või kaudsemas mõttes ka liigitamist.

Kõige esimene ja tähtsaim küsimus hõlmas verbide valikut, mida sõnaraamatus käsitleda. Selge oli see, et kõiki „Dudenis” või ka „Saksa-eesti sõnaraamatus” leiduvaid tegusõnu ei ole mõtet ega saagi reksioonisõnaraamatusse võtta, samas tundus ka „Ungari-eesti reksioonisõnaraamatus” verbide valik liiga piiratud.

Teiseks tuli otsustada, kuivõrd peensustesse minna tähenduste v tähendusvariantide eristamisega. Teadupoolest on paljudel saksakeelsetel tegusõnadel eesti keeles mitu erinevat tähendusvastet. Probleemiks oligi, millised tähendused ja kui palju neid sisse võtta. Variant käsitleda ainult erineva tõlkevastega tähendusi osutus kõlbmatuks, kuna paljudel juhtudel võib ka eestikeelne vaste olla mitmetähenduslik (nt *führen* – *autot juhtima*; *ettevõtet juhtima*; *4 punktiga juhtima*; *mees peab tantsu ajal juhtima*) või olla erineva reksiooniga (nt *sich sehnen* – *igatsema midagi*; *igatsema millegi järele*).

Või mida teha siis, kui tegusõna mitu erinevat tähendust on sama reksiooniga (nt *aufführen* – *(midagi) teatris mängima, etendama*;

(*midagi*) püstitama, rajama, ehitama)? Kuidas tähistada juhtumeid, kus ühe ja sama tähendusvarianadi puhul on võimalik mitu erinevat käänat (nt *Martin führte Karin als seine Frau heim. – Martin abiellus Kariniga. Martin võttis Karini naiseks.*)? Siin on küll tegemist juba pigem märgistusprobleemidega, ent omamoodi liigitamist on vaja siingi.

Ettekandes leiaksid üksikasjalikumalt käsitlemist ülalmainitud probleemid ja nende võimalikud lahendused.

Lahterdused füsioloogias

Andres Soosaar

Tartu Ülikool

<http://biomedicum.ut.ee/~andress>

Lahterdamine paistab olevat teaduste pärisosa, ja niisamuti on lood ka füsioloogiaga. Füsioloogia soovib teada saada ja ära seletada seda, mis toimub elavas organismis. Kui rääkida inimese organismist, siis füsioloogia piiritleb ennast tüüpiliselt siiski inimeksistenti bioloogilise poolega või pigem bioloogilistest keskkondadest tulenevate interaktsioonidega; hingeline ning sotsiaalne jäävad teistele teadustele ja kunstidele lahterdamiseks. Et elus olla, peavad organismis toimuma pidevalt mitmesugused protsessid, ja mis veelgi segavam, see organismis toimuv paistab olevat veel ka mitut laadi, sest me ei kipu rahul olema lihtsalt sellega, et elavas organismis toimub elu. Bioloogia, sealhulgas ka füsioloogia, kipub elu ikkagi põhiliselt uudishimust kimbutama, isegi teda ohjeldama. Samas on füsioloogia empiiriline ja eksperimentaalne teadus, mis olemuslikult tähendab seda, et korraga suudetakse uurida vaid organismi suuremaid või väiksemaid fragmente ning mitte organismi tervikuna. Füsioloogia uurimisobjektiks on küll sageli kogu organism, kuid konkreetsetel vaadeldakse ikkagi vaid üht või mõnda tema omadust. Füsioloogia ihkab väga terviku järele, kuid enamasti jäävad sedalaadi püüdlused ikkagi meetodiülesteks mõttelisteks ponnistusteks. Tegelikkus on see, et terviklikke organismikäsitlusi on füsioloogias ikkagi vaid üksikuid. Ehk ongi vaid homöostaas see suur printsiip, mis püüab erinevaid füsioloogia lahterdusi taas tervikuks kokku siduda?

Järgnevalt sooviks lähemalt vaadelda füsioloogilise lahterdamise kolme põhjendust:

Lahterdamise vajadus kui märk kimbatusest: me ei suuda haarata organismi tervikuna (loomulikult mitte niivõrd mõõtmetelt, vaid potentsiaalsetelt omadustelt ja võimetelt) ja see tuleb uurimise jaoks tükeldada sobivamateks fragmentideks. Seejärel katsutakse aru saada

tükkidest ja kuskil bioloogide ja eriti meedikute peades peaks organismist taas tervik saama ja sellisel moel tuleks seda ka käsitleda.

Siit on vaid samm edasi, et vaadelda füsioloogia lahterdamist vahendi ja meetodi kontekstis. Füsioloogia ise armastab end jaotada lähtudes organismi asukohast bioloogilise taksonoomia avarustes. Kindlad jaotused on taimefüsioloogia ja loomafüsioloogia, millest omakorda kipub inimese erilise tõttu eralduma inimese füsioloogia. Küllalt eriline valdkond on ka võrdlev füsioloogia, mis püüab täpsustada bioloogiliste funktsioonide realiseerumist eri liiki organismides. Inimese füsioloogia sees eristuvad selgesti ealine füsioloogia, inimtegevuste füsioloogia, normaalse ja haigestunud organismi füsioloogia. Väga levinud on füsioloogia vaatlemine ja jaotamine organismi struktuuralse organisatsiooni tasemetest lähtuvalt: nii eristuvad väga selgesti erinevate molekulide ja rakkude füsioloogiad, organite ja organsüsteemide füsioloogiad. Vahendilise jaotamise väga heaks näiteks on elektrofüsioloogia, mille raames uuritakse teatud universaalsete meetoditega bioelektrilisi protsesse kõikjal organismis. Nii teoreetiliste kui materiaalsete vahendite kontekstis püütakse füsioloogias (nagu mujalgi bioloogias) taandada organismis toimuvat füüsikalistele ja keemilistele nähtustele. Selline reduktsionism on bioloogia arengus olnud väga edukas, kuid samas loonud hulga n-ö lahtiseid otsi, mis vajavad kuidagi tervikpiltidega sidumist. Nii nagu teisteski teadustes, on kogutud teadmiste ja võimalike metodoloogiliste lähenemiste hulk muutnud juba ammu nõnda suureks, et inimesel on lootust saada asjatundjaks vaid õige kitsas füsioloogia ja teiste teaduste sektoris, mis omakorda soosib edasisi lahterdamisi ja lahterdusi.

Oluliseks lahterdamise aspektiks on selle uurijatest lähtuvad eesmärgid, olgu siis tegemist uurimise vahe- või lõppeesmärkidega. Et eesmärgid ja vahendid muutuvad, siis kipuvad aegajalt ka lahterdamised muutuma või vahetuma ja mõnikord juhtub sedagi, et muutunud jaotus võib koguni uuritavale endale saatuslikuks saada.

Keel ja murre

Urmas Sutrop
Eesti Keele Instituut
Urmas.Sutrop@eki.ee

Ettekandes küsitakse, kuidas eristada ja defineerida murret ja keelt. Käsitletakse murrete kontiinumit ja laineteooriat aluseks võttes sellest keelte teket. Murrete pidevat üleminekut katkestavad jõukeskused (armee, laevastik, kirik), looduslikud tõkked (jõed, mäed, orud, mered, liigendatud rannajoon), inimgruppide liikumine (rahvapärandid, sõjad) ja/või inimasustuse hõrenemine (pudelikaelaefekt, katkestatud tasakaal) haiguste, nälja, sõdade, vulkaanipursete, põua, üleujutuste vms tõttu.

Kõige lihtsam lingvistiline definitsioon ütleb, et keeli saab eristada ainult vastastikuse mittemõistmise alusel. Kui on pidev murrete üleminek (kontiinum), mille piires kõnelejad saavad üksteise kõnест aru, siis on tegemist ühe keelega. Kui see kontiinum katkeb, siis inimesed ei saa vastastikku üksteise kõnест aru (nt eesti ja läti keel).

Praktikas ei defineerita keeli mitte lingvistiliselt, vaid poliitiliselt. Kõige lihtsama poliitilise definitsiooni järgi on keelt määratletud armee, laevastiku ja kirikuga. Armee ja laevastiku põhjal on piirid tõmmatud näiteks norra, rootsi ja taani keele eristamiseks. Samuti eristab armee ja laevastik saksa ja hollandi keelt. Hollandi ja flaami keele eristajaks on aga kirik. Kirik eristab ka serbia keelt horvaadi keelest. Ka nt bulgaaria ja makedoonia keele eristus on loodud kunstlikult poliitilistel põhjustel. Islandi keel seevastu on piiritletud geograafiliselt (üksik kauge saar).

On olemas veel keele psühholoogiline definitsioon. Kui mingi (nt territoriaalse) grupi, murde- või murrakuala valdav enamus tunneb, et nad räägivad oma naabritest erinevat keelt (siin ei ole mingit tähtsust lingvistilistel keele tunnustel), siis nii see ongi. See ei saa aga olla paljalt üksikute juhtide arvamus. Lihtsaim viis psühholoo-

gilise kriteeriumi rakendamiseks on korraldada vastavates piirkondades (maakondades ja/või valdades) rahvahääletus.

Ettekande lõpus küsitakse: mitu läänemeresoome keelt on olemas? Kas neid tehakse 21. sajandil juurde ka eesti keele arvel? 19. sajandi keskpaigani oletati ainult ühte läänemeresoome keelt, mille murreteks olid nii eesti kui soome keel. 20. sajandil eristati viit kuni kümnet läänemeresoome keelt (korruga mitte rohkem kui üheksat), enamasti seitset.

Viimasel ajal püütakse juurde luua kolme uut läänemeresoome regionaalkeelt – lõunaeesti keelt ja selle sees omakorda võru ja setu keelt. Eesti keele lõunaeestilised murded on: Mulgi, Tartu ja Võru. Mingist eesti keelest erinevast keelest nende murrete puhul rääkida on võimatu.

Võru murre (inglise keeles *Võru dialect*) on eesti keele murre. Ka murretel võib olla normeeritud kirjalik kuju. Võru Instituudi poolt propageeritav niinimetatud Võru keel on kunstlik murdekeele kirjalik väljendus, mis hävitab elavat murret, ehedat rahvakeelt ja nivelleerib eheda Võru murde (murrakute) erinevusi.

Setu idioomiks nimetatav ei ole isegi mitte murre, vaid Võru murde murrakute rühm (inglise keeles *a group of Võru sub-dialects*). Setu murrakute rühm ei eristu ülejäänud Võru murde murrakutest mitte keeleliselt, vaid selle poolest, et nende murrakute kõnelejad on teist usku.

Regionaalkeel ei ole definitsiooni järgi murre. Kui on kahtlust, kas murre on ikka murre või hoopis iseseisev (regionaal)keel, siis otsustab seda valitsus (jõukeskus). Uue keele kõnelejad ei oleks enam eestlased ja nende emakeel ei oleks enam eesti keel. Kas me oleme selleks valmis, et Tartu-, Võru- ja Mulgimaal ei oleks enam eestlasi ega eesti keele kõnelejaid?

Selliseks destruktiivseks ja separatistlikuks tegevuseks ei ole ka ajaloolist õigustust. Henrik ei kahelnud *Liivimaa kroonikat* kirjutades, et Sakala ja Ugandi on Eesti osad ning sakalased ja ugandilased on eestlased.

Sõnaliikide klassifitseerimise probleemidest võrdlevas aspektis

Valentina Štšadneva
Tartu Ülikool
valent@ut.ee

Teadus kui maailma tunnetamise vahend on mõeldamatu ilma uurimisobjektide klassifitseerimiseta. Klassifikatsioon (ld *?lassis* 'rühm' + *facere* 'tegema') eeldab nähtuste **süsteematiseerimist**, st objektide korrastamist nende klassidesse e gruppidesse jaotamise teel. Uuritavate fenomenide süsteematiseerimisel kujutab iga klassifikatsioon endast teadmiste ühes või teises valdkonnas **rööpseoses olevate kaasallutatud** objektide, mõistete ja liikide süsteemi.

Sarnaselt teistele humanitaarteadustele klassifitseeritakse lingvistikas objekte – keelenähtusi – mõningal määral **relatiivselt**. Vaatamata keelefaktide **rööpseoses olevatesse hierarhilistesse rühmadesse** jaotamise suhtelisusele, võimaldab selline rühmitamine orienteeruda objektide mitmekülgsuses. Rahvuskeele teoreetiline grammatika on iseenesestmõistetavalt suunatud eeskätt seda keelt emakeelena valdavale inimesele. Taoline lingvistiline klassifikatsioon võib osutada ebapiisavaks/puudulikuks allikaks keelefenomenide kõrvutava keeleuurimise seisukohalt.

Sõnaliikide klassifitseerimisel arvestatakse lisaks **rahvuskeele eripärale** ka olemasolevaid **traditsioone** ja **kontseptsioone** uurija eelistustest lähtuvalt. Sellest tingituna puutuvad keeleteadlased võrdluse käigus paratamatult kokku terminoloogiliste erinevustega, millest kõige tõsisemaid probleeme tekitavad mittetäistähenduslike sõnade, näiteks partiklite, liigitamine. Seetõttu, pidades silmas võrdleva analüüsi praktilist eesmärki, oleks tarvilik välja töötada sõnaliikide klassifikatsioon nn tööskeemi näol. Sellise tööskeemi aluseks võiksid tõenäoliselt olla sõnade olulisemad leksikaalgrammatilised funktsioonid. Viidatud seisukohtadel peatutakse pikemalt ettekandes.

Tellija kriteeriumid tõlke hindamisel ja nende sobivus koolitööde hindamissüsteemi alusena

Arvi Tavast
Tallinna Pedagoogikaülikool
arvi@imprimaatur.ee

Mitmesugustel majanduslikel ja poliitilistel põhjustel on ka eesti keel nüüd jõudnud nende umbes poolesaja maailma vägeva keele hulka, milles oma tekstide avaldamist peavad vajalikuks hargmaised ettevõtted ja rahvusvahelised organisatsioonid ning millesse tõlkimine on seetõttu omandanud tööstuslikud mõõtmed. Üks massilise ja rutiinse tõlkimise tulemusi on eraldi tegevusala – tõlgete kvaliteedikontrolli – tekkimine, teine aga tõlkijakoolituse kiire areng. Ettekandes käsitlen esimeses kasutatavate vealligistuste sobivust teise jaoks.

Kuna tõlke tellijal endal puudub reeglina igasugune pädevus sihtkeele alal, aga tõlkekvaliteedist sõltub palju nii tõlkekulude kui ka tulevase tarbijaedu mõttes, peab kvaliteedikontroll olema võimalikult keelesõltumatu, seega täpselt reeglistatud. Selleks on iga tellija välja töötanud oma hindamissüsteemi, millest mõnda ettekandes tutvustan. Need on esmapilgul küll üsna erinevad, kuid põhinevad siiski teatud ühistel eeldustel ja valikutel.

Tarbetekstidele orienteeritud tõlkekoolituse eesmärk on valmistada tõlkijaid ette just selliseid kriteeriume kasutavale tõlketurule. Sellest võiks järelduda samalaadse hindamissüsteemi kasutamise ka koolitööde puhul, kuid leidub olulisi vastuargumente.

- Vead on nii tüüpide kui ka raskusastmete mõttes liiga vähe diferentseeritud, mis teeb hindamise töömahukaks, nõudes iga vea puhul käsitsi selgituse lisamist; sobiva peenusega eristatud ja asjakohaste veatüüpide puhul piisaks tüübi määramisest.

- Vigade jaotus on kas teooriaeelne või põhineb teooriatel, mille sobivuses tõlkekoolitusse on põhjust kahelda.

- Positiivsed nähtused on kategoriseerimata.

Tõlgete hindamine eeldab seisukohta tõlke olemuse teemal. Soovides jõuda tööstuslikelt hindamissüsteemidelt koolitusse sobiva juurde, kirjeldan Frege'st ja klassikalisest terminoloogiateooriast lähtuvat ning arvutivõrkude protokollivirna sarnasusest inspireeritud mudelit, mille headus tõlkekoolituse jaoks on seni ümber lükkamata hüpotees. Lähtuvalt sellest mudelist esitan tõlkenähtuste liigituse, mille eesmärk on maksimaalselt toetada koolitusprotsessi, lihtsustades õppejõu rutiinset tööd ja andes samal ajal võimalikult konstruktiivset tagasisidet õppijale.

Kategoriseerimise kognitiivsed ja sotsiokultuurilised alused

Silvi Tenjes
Tartu Ülikool
stenjes@madli.ut.ee

Kognitiivse vaateviisi kohaselt kuuluvad kategooriad *olemuslikult* aruka *tegevuse* ja *mõtlemise* juurde. Kui inimene seisab silmitsi uue objekti, situatsiooni või probleemiga, siis ta seostab seda oma *teadmistega*, selleks et seda mingisse tüüpi liigitada e klassifitseerida (Ross, Makin 1999). Vaatame kategooriate arenguid inimese üldise kognitiivsuse raames, kuhu kuuluvad nii evolutsiooniline kui ka sotsiokultuuriline areng.

Üldiselt uurib kognitiivteadus, *kuidas kategooriateadmisi* (ingl *category knowledge*) *representeeritakse*, konkreetselt aga uurib, *kuidas representeeritakse* neid *teadmisi*, mida kasutatakse *liigitamiseks* (Ross, Makin 1999). Selleks on arvukalt võimalusi, kuid siin vaatleme mõnda kõige levinumat moodust, näiteks: 1) mudelid; 2) klassikaline vaade; 3) prototüübiteooria ja ähmased hulgad; 4) näidisvaade (ingl *exemplar view*) ja 5) dünaamilised kategooriad (Rocha 1997).

Teadmised on tekkinud ja arenenud kategoriseerimisprotsessis, kui arenev organism hakkas kasutama *mälu*, mis sai püsijäämise üheks oluliseks vahendiks. Mälu abil sai hakata „välja sortima” eluks vajalikku infot keskkonna „kontekstist”. Seega on kategoriseerimine evolutsiooni käigus tekkinud *suhe* või *seos* mälu ja keskkonna vahel. Kultuurimehhanismi aluseks võib olla *emotsionaalne mälu*. Vaatame, kuidas inimene on *kasutanud* mälu sotsiokultuurilises ja keele kaudu vahendatavas keskkonnas. Teadmiste võrdlevas filosoofilises teoorias on *representatsioonid* kindlad *mentaalsed struktuurid*, mis sisaldavad intentsionaalseid kategooriaid (usu et p, soovi et q),

mis koosnevad lingvistiliste (semantiliste) struktuuride sisust süvatasandil. Psühholoogias on representatsioonid tüüpiliselt kirjeldatud kui *kontseptuaalsed struktuurid*, mis on oma sisudega individualiseeritud. Kontseptid või teadmiskonstruktuurid on krunditud intentionaalse kategoriseerimise sisse.

Keelelise suhtlemise maailmas on peale ruumitaju ja ruumi kategoriseerimise oluline just *teisele osutamine* ja teiseks *tõlkimise* võimalus – see olemuslik, mis teeb märgisüsteemi selleks, mis ta on. Eksperimentaaluuringud näitavad, et oluline on *kogemuse* roll tajumises, kategoriseerimises ja kontseptualiseerimises. Näiteks primaarne episteemiline kontrast *mina-sina* peegeldab esimest ja kõige olulisemat sammu kognitiivses kategoriseerimises, mis on keele kui representatsioonisüsteemi fenomenoloogiline tuum (Kravchenko 2002). On uuritud ja leitud, et näiteks sookategoriad peegeldavad üsna suvalisi klassifikatsioone ning seega pole leitud semantilist õigustust noomentite klassifikatsiooniks erinevatesse soolistesse klassidesse.

Loodusrahvastel on klassifikatsioonid sageli inspireeritud looma- ja taimeliikide omapärast ja individualiseeritusest. Lévi-Strauss (1962/2001) leiab, et loodusrahvaste klassifikatsioonisüsteemide uurimine ei kirjelda mitte erijuhtumeid, vaid teadmiste ja mõtlemise vorme nende rahvaste juures. Põlisrahvaste klassifikatsioonide vooruseks on metoodilisus, tugev teoreetiline vundament, ja vahel on nad vormiltki võrreldavad nende klassifikatsioonidega, mis on jätkuvalt kasutusel zooloogias ja botaanikas.

Kognitiivsuse konkreetne panus lingvistikasse võiks olla *kaardistamine* (ingl *mapping*). Täendus on kaardistav suhe või seos lingvistiliselt kontseptualiseeritud osutava situatsiooni ja kontseptuaalselt motiveeritud väljendi vahel (Sinha 1999: 231). Lõpetuseks esitatakse näiteid kultuurispetsiifilistest klassifikatsioonidest.

Kirjandus

- Kravchenko, Alexander V.** 2002. The cognitive roots of gender in Russian. – *Glossos*, The Slavic and East European Language Resource Center 3, 1–13. <http://new.seelrc.org/glossos/issues/3/kravchenko.pdf>
- Lévi-Strauss, Claude** 2001. *Metsik mõtlemine*. Tallinn: Vagabund. [Tõlgitud väljaandest Claude Lévi-Strauss 1962. *La pensée sauvage*. Plon.]
- Rocha, Luis Mateus** 1997. Evidence Sets and Contextual Genetic Algorithms. Exploring Uncertainty, Context, and Embodiment in Cognitive and Biological Systems. – PhD Dissertation in Systems Science. State University of New York at Binghamton. <http://www.c3.lanl.gov/~rocha/dissert.html>
- Ross, Brian H.; Makin, Valerie S.** 1999. Prototype versus Exemplar Models in Cognition. – Robert J. Sternberg (Ed.), *The Nature of Cognition*. Cambridge, Massachusetts, London, England: A Bradford Book, The MIT Press, 205–241.
- Sinha, Chris** 1999. Grounding, mapping and acts of meaning. – Theo Janssen and Gisela Redeker (Eds.), *Cognitive Linguistics: Foundations, Scope and Methodology*. Berlin & New York: Mouton de Gruyter, 223–255.

Eesti viipekeele õpetamise eripära

Regina Toom
Tartu Ülikool
regina@psych.ut.ee

Viipekeele kui võõrkeele õpetamisele Eestis hakati suuremat tähelepanu pöörama alles 1990. aastate lõpul, kuigi juba esimesel omariiklusperioodil räägiti selle keele (ehk tollase nimetusega *kurtide keele*) õpetamise vajalikkusest kuuljatele (Eesti Kurttummade Selts 1924. a). Arhiivimaterjalidest paraku ei selgu, kas jõuti ka reaalsete tegudeni. Teadaolevalt aga toimusid esimesed nn „miimikat ja daktüloloogiat tutvustavad” õppepäevad Kurtide Ühingu süsteemis töötavatele inimestele aastatel 1970–1980.

Huvi tärkamist viipekeele õppimise vastu eri sihtgruppide poolt mõjutasid mitmed tegurid, kuid olulisemateks neist olid kurtide kui keelelis-kultuurilise vähemusgrupi käsitlemine ning viipekeele propageerimine kurtide hariduskeelena. Vähemtähtsaks ei saa pidada ka ETV saadet „Silmside” aastatel 1989.–1990. a, mil laiem üldsus tutvus teleekraanil viipekeele ja selle eripäraga esmakordselt. Saated tõlgiti viipekeelde ning õpetati, kuidas üksikute viibete abil hõlbustada suhtlemist kurtidega.

Tänaseks on läbi viidud hulganisti viipekeele lühi- ja pikemaajalisi kursusi peamiselt Eesti Viipekeele Tõlkide Ühingu ja Eesti Kuulmispuudega Laste Vanemate Liidu eestvedamisel. Keeleõpe on vabaainena kättesaadav mitmetes õppeasutustes sh alates 1992. aastast valikkursusena Tartu Ülikooli eripedagoogika osakonnas. Viipekeel kurtide emakeelena on lülitatud ka kuulmispuudega laste erikoolides põhikooli programmi.

Kuigi eesti viipekeel vähemusgrupi keelena on võitmas üha suuremat kõlapinda vabariigi keelepoliitikas, on selle õpetamine tervikuna veel ebaühtlane ja lapsekingades. Näiteks, kui eelmise sajandi lõpus oli keeleõppe eesmärgiks eelkõige suhtluskeele või üksikute viibete õpetamine kindlatele sihtgruppidele (viipekeele-tõlgid, lapsevanemad, huvilised, eripedagoogid jt), siis tänapäeval on

keele õpetamise sisu ja vorm oluliselt muutumas enam keeleteaduslikumaks, ühtlasi kasvab vajadus keelevaldajate (viipekeeletõlkide) järele.

Ettekandes vaadeldakse lähemalt, missugused on eesti viipekeele kui võõrkeele õpetamise ja õppimisega seonduvad probleemid ning kuidas leida nimetatud probleemidele võimalikke lahendusviise, et tagada eesti viipekeele valdamise kõrge tase.

Mille alusel liigitada, kui vormiline liigitusalus puudub? (modaalverbide näitel)

Ilona Tragel
Tartu Ülikool
ilona.tragel@ut.ee

Ettekandes esitan mõningaid semantilise liigitusega seotud probleeme. Materjaliks on modaalverbid (näitematerjal põhiliselt Tartu ülikooli kirjakeele korpuse morfo-märgendatud osast).

Kuidas eristada sama modaalverbi eri modaaltähendusi? Kas on võimalik kindlaks teha, et mingi kindel modaaltähendus esineb kindlas ümbruses: modaalse finiiiverbi komplement ja argumendid on mingis kindlas vormis (nt *saama* nn õnnestumismodaalsus, mis esineb tavaliselt *ma*-infinitivse komplementiga: *sai sealt tulema kell kuus*)? Kas mõnel juhul on teatud modaalne funktsioon vaid verbi ühel kindlal vormil (nt imperfektivormid *jäi, sai, pidi, lasi*)? Kas mõni modaalne funktsioon esineb valdavalt eitavates kasutustes (nt *ei saa teha*) ja kuivõrd selliste andmete põhjal saab teha järeldusi modaaltähenduste arenguskeemide kohta (vrd nt Traugott 1989)? Kindlas konstruktsioonis modaaltähenduse saavate verbide (*tulema, andma, ka tahtma*) puhul jälgin konstruktsiooni osaliste omadusi (elusus, aktiivsus) ja vormistusi (nt adessiivse omajana vormistatud kohustav modaalsus, vrd .. *nii et emadel tuli teinekord väga otsustavalt vahele astuda.*). Modaalse funktsiooni kasutuseelistusi saab vaadelda ka osaliste olemasolu ja puudumise põhjal (nt *tulema*-konstruktsioonimodaal esineb enamasti ilma eksplitsiitselt vormistatud modaalsuse agendita, vrd .. *et [Ø] tuleb minna selle akna taha piiluma.*, vrd Talmy 2000)

Modaaltähenduste ja laiemalt semantiliste liigituste puhul on enamasti tegemist suhteliselt hägusate piiridega rühmadega, liigituskriteeriumid täienevad enamasti liigituse käigus.

Kirjandus

- Talmy, Leonard** 2000. *Toward a Cognitive Semantics, Vol. I: Concept Structuring Systems*. Cambridge: The MIT Press.
- Tragel, Iona** 2004. Tuumverbi ja *da*-infinitiivi ühendid. Ettekanne Tartu ülikooli eesti keele õpetooli korraldatud teisel süntaksipäeval 27.02.2004 (ilmumas preprintina).
- Traugott, Elizabeth Closs** 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. – *Language* 65:1, lk 31–55.

Inglise-eesti tehnikatõlke vigade liigid ja põhjused

Andres Valdre

Tartu Ülikool

andres@valdre.com, <http://www.valdre.com/>

Viimase kümnendi jooksul on oluliselt suurenenud inglise-eesti tehnikatõlke maht. Üpris sageli on nendes tõlgetes aga vigu, mis vähendavad tõlke kvaliteeti ja kahjuks seega ka tõlkija usaldusväärsust. Vigu teevad peale asjaarmastajate ka kogenud tõlkijad. Käesolev uurimus, mis on käsil Tartu Ülikooli germaani-romaani filoloogia osakonnas kirjaliku tõlke magistriprojektina, püüab selgitada, kuidas ja miks – ning mida tuleks vigade vältimiseks teha. Uurimus liigitab vead ja põhjused eraldi ning esitab ka nende vahelised seosed.

Tehnikatõlke vead jagunevad üldjoontes üheksaks liigiks, sisaldades nii vormi- kui sisuvigu. Täpne piiritus ei ole võimalik, sest samad veanäited võivad korraga esindada eri vealiike. Tehnikatekstide olemusest tingituna on pearõhk pööratud sisuvigadele. Vigade põhiliigid on järgmised: keelevead (vorm, õigekiri, ühilduvus), teema-reema vead, kategooriavead, stiilivead, tähendusvead (mitu taset), terminivead (sh konteksti- ja valdkonnavead, poliitsemiaivead), sõnastusvead (kui mõte ei muutu) ning eriti otsetõlked ja loogikavead (sõnastusvead, kus mõte muutub).

Vigade, eelkõige sisuvigade tekkepõhjused jagunevad üldjoontes kaheks: tõlkijavälisteks ja tõlkijaomasteks põhjusteks. Tõlkijavälised põhjused on nt tõlkeülesande puudused, lähteteksti vähene sidusus, lähteteksti vead ja sõnastikuvead; hea tõlkija oskab neid põhjuseid vältida. Tõlkijaomased põhjused on nt terminipiiri tajumata jätmine, sarnane võõrsõna, oma väljamõeldise kasutamine, faktiviga, terminimoodustuse puudulik oskus, väike silmaring, vähene lähtekeele oskus, sõnastike sihipäratu kasutamine, tõlkimise abitarkvara puudulik kasutusoskus, psühholoogilised tegurid (nt kiirustamine ja

väsimus) ning tavalisimad põhjused – loogilise mõtlemise ja ettekujutusvõime puudumine. Loomulikult võivad ühe vea põhjustada mitu põhjust koos.

Uurimus tugineb autori isiklikule tehnikatõlgete toimetamise kogemusele.

„Eesti keele arendamise strateegia 2004–2010”: kaasnevad publikatsioonid

Jüri Valge
Haridusministeerium
jyri.valge@hm.ee

„Eesti keele arendamise strateegia 2004–2010” eeltööde käigus korraldas Eesti Keelenõukogu aastail 2001–2002 eesti keele seisundi uuringuid, mille tulemusena on ilmunud keelestrateegia ametlike lisadena käsitletavat teosed:

- Eesti keele uurimise analüüs. – Emakeele Seltsi aastaraamat 48 (2002). Toim Mati Ereht. Tallinn 2003.
- Eesti kirjakeele kasutusvaldkondade seisundi uuringud. – Tallinna Pedagoogikakooli eesti filoloogia osakonna toimetised 4. Toim Maire Raadik, koostaja Helle Metslang. Tallinn 2003.
- Tiiu Ereht, Arvi Tavast. Eesti oskuskeelekorralduse seisund. – Tallinn: Eesti Keele Sihtasutus, 2003.
- Eve Alender, Kairit Henno, Annika Hussar, Peeter Päll, Evar Saar. Nimekorralduse analüüs. – Tallinn: Eesti Keele Sihtasutus, 2003.
- Kadri Muischnek, Heili Orav, Heiki-Jaan Kaalep, Haldur Õim. Eesti keele tehnoloogilised ressursid ja vahendid. Arvutikorpused, arvutisõnastikud, keeletehnoloogiline tarkvara. – Tallinn: Eesti Keele Sihtasutus, 2003.
- Vahur Laiapea, Merilin Miljan, Urmas Sutrop, Regina Toom. Eesti viipekeel. – Tallinn: Eesti Keele Sihtasutus, 2003.

Lisandumas on veel:

- Birute Klaas, Eesti keel ja teised keeled. Tartu Ülikool. Tartu 2004.

- Mart Rannut. Eesti keel arvudes. Tallinna Pedagoogikaülikool. Tallinn 2004.
- Mart Rannut. Eesti keeleõiguse analüüs. Tallinna Pedagoogikaülikool. Tallinn 2004.

Peale keelestrateegia otsuste lisade on strateegia eeltööde käigus välja antud mitu strateegia seisukohti kujundavat ja väljendavat teost:

- Tiiu Erelt. Eesti keelekorraldus. Tallinn 2002.
- Eesti murded ja kohanimed. Toim Tiit Hennoste. Tallinn 2002.
- Estonian language. Toim Mati Erelt. Tallinn 2003.
- Euroopa keelteaasta Eestis. Koostaja Jüri Valge. Tartu 2001.
- 100 aastat akadeemilist eesti keele õpet Uppsala ülikoolis. Eesti keele ja keelepoliitika ettekanded. Toim Raimo Raag, Jüri Valge. Uppsala, Tartu 2001.

Konverentsi keelestrateegia sektsioonis annab Eesti Keelenõukogu esimees Jüri Valge ülevaate keelestrateegiaga seoses avaldatust ja avaldatavast ning autorid, koostajad ja toimetajad tutvustavad publikatsioone lähemalt.

Ühest sõnaliigiprobleemist sõnamoodustuse nurga alt vaadatuna

Silvi Vare
Eesti Keele Instituut
silvi.vare@eki.ee

Käsitlen ettekandes küsimusi ja küsitavusi, mis puudutavad tegevust väljendavate tüvede omavahelisi moodustusseoseid ja sõnaliigilist vahekorda. Täpsemalt on vaatluse all *e*-liitelised teonimed ja nende suhe teiste sama tüve sisaldavate üksustega. Eeskätt kuuluvad siia kohakäändelised seisundivormid ja liitsõna moodustusosana esinevad seotud tegevustüved.

Käsitlus põhineb „Eesti keele sõnaraamatu ÕS 1999” materjalil, mida on võrreldud varasemate õigekeelsussõnaraamatutega (ÕS 1960, 1976) ja „Eesti kirjakeele seletussõnaraamatu” ilmunud vihikutega.

Sõnaraamatutes on *a priori* lähtunud üldjuhul kolmest eeldusest:

1) struktuurilise sobivuse korral on verbist võimalik moodustada sellega rööpne tegevust väljendav *e*-liiteline nimisõna ehk *e*-liiteline teonimi; nt puhka/ma → puh-e 1. <.puhke, puhet> *puhkamine*; puhu/ma → puh-e 2. <.puhke, puhet> *puhumine*;

2) kohakäändeline seisundivorm on teonime muutevorm; nt Talvine mets on puhkel. ← puhe 1. *puhkamine*; *puhkel* õied ← puhe 2. *puhumine*;

3) liitsõna moodustusosana esinev *e*-tüvi on *e*-liitelise teonime tüvevorm; nt liitsõnad puhke/ala, /asend, /hetk, /kodu, /koht, /maja, /paus, /ruum, /seisund, /õhtu ← puhe 1. *puhkamine*; sisse/puhe ← puhe 2. *puhumine*.

Toon ettekandes esile sellise tõlgitusviisi probleemsed küljed ja pakun uusi lahendusi, sh sõnaliigiga seonduvaid hüpoteese.