



20. rakenduslingvistika konverents

27.–28. aprill 2023, Tallinn

KEEL JA KEELEKASUTAJAD

TEESID / ABSTRACTS

EAAL 20th Annual Conference

April 27-28, 2023, Tallinn, Estonia

LANGUAGE: THE USER IN FOCUS



EESTI
RAKENDUSLINGVISTIKA
ÜHING ESTONIAN ASSOCIATION OF
APPLIED LINGUISTICS





Sisukord

Kutsutud ettekanded / Invited speeches	6
Keeletehnoloog teoretiseerib korpuse üle	6
Heiki-Jaan Kaalep (Tartu Ülikool, Eesti)	
Digital Dictionary Database for Slovenian: unstructured, semi-structured and structured data in modern lexicography	8
Simon Krek (Jožef Stefan Institute, Slovenia)	
Linking Second Language Acquisition Research and Digital Language Learning	9
Detmar Meurers (University of Tübingen, Germany)	
Suulised ettekanded / Oral presentations	10
„Ei saa ju raske olla, kui eesti keel on emakeel.“ – Kas eestlasel on raske eesti keeles kirjutada? Mõningaid küsitluse tulemusi	10
Kristel Algvere (Tartu Ülikool, Eesti)	
Veamärgendusega korpus grammatikakontrollija arenduseks ja testimiseks	11
Kais Allkivi-Metsoja ¹ , Krista Liin ² , Karina Kert ¹ , Silvia Maine ¹ (¹ Tallinna Ülikool, Eesti. ² Tartu Ülikool, Eesti)	
Keelekasutusmustrid & kontekstid	12
Pille Eslon, Jaagup Kippar (Tallinna Ülikool, Eesti)	
<i>Kuna külasthan tihti Tartu kaubamaja, siis on juba enamvähem meelde jäänud kassatibid/kassatädid.</i> Sooliselt markeeritud ametinimetused eesti keeles	14
Elisabeth Kaukonen (Tartu ülikool, Eesti)	
Perspectives from the academic writing community on the impact of Generative AI: a discussion	14
Djuddah A. J. Leijen, Helen Hint, Helena Lemendik (University of Tartu, Estonia)	
Kõnetempo varieerumine eesti keeles	15
Pärtel Lippus, Maarja-Liisa Pilvik, Kaidi Lõo, Anton Malmi, Liina Lindström (Tartu Ülikool, Eesti)	
Eesti keele õigekirja- ja grammatikakontroll: mudelite võrdlus ja kombineerimine	16
Agnes Luhtaru ¹ , Mark Fišel ¹ , Krista Liin ¹ , Kais Allkivi-Metsoja ² , Jaagup Kippar ² (¹ Tartu Ülikool, Eesti. ² Tallinna Ülikool, Eesti)	
Eesti keele sõnavara testi LexEsti loomise taustast ja koostamisest	18
Kaidi Lõo, Agu Bleive, Katrin Leppik, Anton Malmi (Tartu Ülikool, Eesti)	
Teooria ja rakendus: keelekorraldus.....	19
Peep Nemvalts (Tallinna Ülikool, Eesti)	
Keelekasutaja valikud rööpsete kirjakuju puhul: <i>ekleer või ekläär?</i>	20
Tiina Paet, Madis Jürviste (Eesti Keele Instituut, Eesti)	
Alzheimeri tõvele viitavate märkide automaatne tuvastamine kõnest: kuidas, milleks, kellele ja kas?	21
Ulla Petti (Cambridge'i Ülikool, Suurbritannia)	
Spontaanses stiilis kõnesüntees	23



Liisi Piits ¹ , Heete Sahkai ¹ , Liis Ermus ¹ , Indrek Hein ¹ , Meelis Mihkla ¹ , Indrek Kiissel ¹ , Kristjan Suluste ¹ , Egert Männisalu ¹ , Rene Altrov ¹ , Hille Pajupuu ¹ , Jaan Pajupuu ² (¹ Eesti Keele Instituut. ² Vabakutseline tarkvaraarendaja)	
Võro ja seto keele kasutamine ja keelevahetus 20. sajandi jooksul.....	24
Helen Plado ^{1,2} , Liina Lindström ¹ , Maarja-Liisa Pilvik ¹ , Triin Todesk ¹ , Annika Kängsepp ¹ (¹ Tartu Ülikool, Eesti. ² Võro Instituut, Eesti)	
4. klassi õpilaste lugemis- ja kuulamisoskus	25
Helin Puksand, Erika Klettenberg, Kerli Haav (Tallinna Ülikool, Eesti)	
Automatic Part-of-Speech Tagging with CLAWS7: Impact of Learner Errors	26
Liina Tammekänd, Reeli Torn-Leesik (University of Tartu, Estonia)	
The treatment of idioms and conversational routines in <i>Woordcombinaties</i>	26
Carole Tiberius, Lut Colman (Instituut voor de Nederlandse Taal, Netherlands)	
Eesti kirjakeele kihiline korpus 1800-1940	27
Peeter Tinitis (Tartu Ülikool, Eesti. Eesti Rahvusraamatukogu, Eesti)	
TNR 12, 1.5, rööpjoendus: akadeemiliste tekstide kirjutamise juhendite tavade murdmine.....	28
Ilona Tragel, Helena Lemendik, Liisa-Maria Komissarov, Eleriin Miilman, Külli Habicht, Nele Novek, Nele Karolin Teiva, Betti Marie Peterson, Reena Roos (Tartu Ülikool, Eesti)	
Kujundlik mõtlemine 2020—2022 keelekriisis	30
Ene Vainik ¹ , Geda Paulsen ^{1,2} (¹ Eesti Keele Instituut, Eesti. ² Uppsala Ülikool, Rootsi)	
Vikipeedia kasutamine isiklikel eesmärkidel ja õppetöös: õpetajate kogemused	31
Maigi Vija (Tartu Ülikool, Eesti)	
Mitmekeelsuse töötuba / Workshop on Multilingualism and Multilingual Acquisition	33
Building a language: applying the traceback method to a 2-year-old Estonian-English bilingual's data	33
Piret Baird (Tallinn University, Estonia)	
Narrative Macrostructure in Lithuanian Monolingual and Bilingual Typically Developing 6-Year-Olds	34
Ingrida Balčiūnienė, Ineta Dabašinskienė, Laura Kamandulytė-Merfeldienė (Vytautas Magnus University, Lithuania)	
Local Russian in Kalmykia and mixed Kalmyk	35
Vlada Baranova (The Helsinki Collegium for Advanced Studies, Finland)	
The production of narratives in Lithuanian-speaking monolingual and bilingual typically-developing children and children with a language disorder.....	36
Ineta Dabašinskienė, Laura Kamandulytė-Merfeldienė (Vytautas Magnus University, Lithuania)	
Linguistic and pragmatic choices in different sociolinguistic contexts: a corpus-based analysis of Russian spoken discourse.	37
Ilenia Del Popolo Marchitto (Tallinn University School of Humanities, Estonia)	
Multilingual Tallinn: a pilot study on soundscape and linguistic landscape	38
Kapitolina Fedorova, Natalia Tshuikina (Tallinn University, Estonia)	
A constructionist approach to language attrition: The case of dative loss in Blumenau German ...	38



Stefan Hartmann¹, Antje Quick^{2,3}, Nikolas Koch⁴, Claudia Maria Riehl⁴ (¹University of Düsseldorf, Germany. ²University of Leipzig, Germany. ³FAU Erlangen-Nürnberg, Germany. ⁴Ludwig-Maximilians-Universität Munich, Germany)

Language alternation in a multilingual community: How Roma in Estonian use their rich language repertoire.....	39
Anette Ross (University of Helsinki, Finland)	
Egyptian Arabic-English Code-Switching in Egyptian female vlogs.....	40
Natalia Siereda (Tallinn University, Estonia)	
How multilingualism research can contribute to translation studies.....	41
Anna Verschik, Maris Saagpakk (Tallinn University, Estonia)	
Töötuba. Registrid keeles: uurimine ja rakendused	43
<i>"Küllap kirjanik märkas, et hakkas tüdinema". (Inter)subjektiivsuse markeerimine ilukirjandustekstides</i>	
Külli Habicht, Külli Prillop (Tartu Ülikool, Eesti)	
Üld- ja oskuskeele vahetuse kriiside keerises	44
Reet Hendrikson (Kaitseväe Akadeemia, Eesti)	
Sagedased pragmaatilised partiklid ja markerid registrite eristajana	45
Tiit Hennoste, Külli Prillop, Helle Metslang, Külli Habicht, Kirsi Laanesoo, Andriela Rääbis, Carl Eric Simmul (Tartu ülikool, Eesti)	
Veendumusartikkel ju retoorilistes küsimustes ja väidetes teadus- ja ajakirjandustekstis.	46
Marge Käsper (Tartu Ülikool, Eesti)	
Partiklid tõesti ja tõepoolest kolmes registris.....	47
Kirsi Laanesoo (Tartu Ülikool, Eesti)	
KirjakeelEEST.....	48
Margit Langemets ¹ , Lydia Risberg ^{1,2} (¹ Eesti Keele Instituut, Eesti. ² Tartu Ülikool, Eesti)	
<i>Kui maksusüsteemi muresid ära ei lahendata, tõuseb ümbrikupalkade osakaal kindlasti edasi. (Inter)subjektiivsusest ajakirjandustekstides.....</i>	
Helle Metslang (Tartu Ülikool, Eesti)	
<i>Kes te üldse olete?</i> Partikkel <i>üldse</i> suuliste argivestluste küsilausungites.....	51
Andriela Rääbis (Tartu Ülikool, Eesti)	
Veebitekstide liigitamine keeleliste tunnuste alusel	52
Kristiina Vaik, Kairit Sirts, Kadri Muischnek (Univeristy of Tartu, Eesti)	
"Ofc its eesti keel": inglise keel teismeliste tähtsustelustes	54
Virve Vihman, Kristiina Praakli, Maarja-Liisa Pilvik, Mari-Liis Korkus (Tartu ülikool, Eesti)	
Posterettekanded / Poster presentations	55
Lithuanian Heritage Language: Acquisition of Grammar.....	55
Skirmantė Gribauskiene (Vytautas Magnus University, Lithuania)	
Spoken language resources: learning Lithuanian as a second language	56
Laura Kamandulytė-Merfeldienė, Ineta Dabašinskienė (Vytautas Magnus University, Lithuania)	
Refusals strategies in the role-play of Lithuanian children.....	57
Viktorija Kavaliauskaite-Vilkinienė (Vytautas Magnus University, Lithuania)	



Production of consonant clusters of 4–8-year-old Lithuanian children	58
Eglė Krivickaitė-Leišienė (Vytautas Magnus University, Lithuania)	
Development of writing skills: task types and feedback.....	59
Sille Midt (Tartu Ülikool, Estonia)	
The role of input and output in developing second language speaking skills	60
Katrín Mikk (Tartu Ülikool, Estonia)	
Näita lapsele (võõr)keelt: Show the child a (foreign) language on a smart device or how smart devices enrich children's language acquisition process.....	61
Merje Miliste, Natalja Zagura (Tartu Ülikool, Estonia)	
Eesti keele kui teise keele grammatikapädevuse arendamise võimalusi ja vajadusi	62
Raili Pool ^{1,2} , Jelena Kallas ² (¹ Tartu Ülikool, Eesti. ² Eesti Keele Instituut, Eesti)	
„Parempoolse vindikeeraja elukutse omandanu jääb tööta” ehk elukestva õppe diskursuse konstrueerimine Eesti meediatekstides	64
Triin Rattiste, Halliki Põlda (Tallinna Ülikool, Eesti)	
The use of markers of stance among English L1 speaker and Estonian L1 learners of English	65
Denys Savchenko (University of Tartu, Estonia)	
Kirjalikud tööd Tartu Ülikoolis 2009 ja 2023 - mis on muutunud ja miks?	66
Ilona Trigel, Kati Käpp, Marri-Mariska Tammepõld, Eleriin Miilman, Nele Karolin Teiva (Tartu Ülikool, Eesti)	
Development of the Cross-Linguistic Lexical Tasks in Estonian	67
Adele Vaks, Virve Vihman, Marika Padrik (University of Tartu, Estonia)	
Parlamendikorpused ja mis neist leida võib	68
Kadri Vider ¹ , Neeme Kahusk ² , Martin Mölder ² (¹ Eesti Kirjandusmuuseum, Eesti. ² Tartu Ülikool, Eesti)	



Kutsutud ettekanded / Invited speeches

Keeletehnoloog teoretiseerib korpuse üle

Heiki-Jaan Kaalep (Tartu Ülikool, Eesti)

Ettekanne käsitleb teoreetilisi ja ehk isegi filosoofilisi küsimusi, millele eesti keeletehnoloogia praktikuna olen pidanud oma töös vastama.

Praktikuna oli minu algne eesmärk luua programm (praeguseks on sellest kujunenud vabavaraline Vabamorf), mis oskaks suvalise sõnavormi puhul öelda, milline on selle algvorm, kääne, pööre jm grammatilised tunnused, ja muidugi kas see on üldse mingi eesti keeles oleva sõna vorm. Ideaaljuhul oleks tegu programmiga, mis käitub nagu keeletunnetusega inimene, või vähemalt käitub nii küllalt sageli ja teeb etteaimatavaid vigu, mida päris inimesel on kerge ette näha ja parandada.

Rakendusena on tegu õigekirja kontrollija e. spelleriga ja morfoloogilise analüsaatoriga; viimast kasutatakse m.h selleks, et tekstis olev sõnavorm algvormiks teisendada ja seeläbi teksti sõnavara kergemini käsitleda.

Programmi loomise etapid olid: 1) liitsõnade analüüs, tuginedes Ü. Viksi „Väikesele vormisõnastikule“, 2) liitsõnade ja tuletiste analüüs, tuginedes mitmete Eesti lingvistide teoreetilistele töödele, ja 3) sõnastikust puudevate tüvedega sõnade analüüs (e. oletamine), mille jaoks sai tugineda lingvistide ideedele ja tekstikorpustele. Etapi kaupa muutus töö programmeerimisest järjest rohkem lingvistiliseks uurimistööks.

Kas Vabamorf on kõige tõenäolisem kandidaat „Eesti keele morfoloogia teooria“ nimetusele? Programmil on olemas teadmised selle kohta, millised on morfeemid, tuletusliited, liitsõnade moodustamise mehhanism jpm lingvistilisi teadmisi. Tõsi, nad on esitatud programmeerimiskeeles, mitte lingvistile mõeldud proosatekstina, kuid see on ju ainult esitusviisi valik. Tal on ka üks oluline omadus: ta oskab öelda, kuidas mistahes sõnast vorme moodustatakse, s.t. ta on ennustusvõimeline. Milline keelekirjeldus oleks veel sama põhjalik?

Kahjuks tuleb tunnistada, et kui mõni lingvist tõesti loeks Vabamorf lähtekoodi, siis oleks ta vist vapustatud, et kas tõesti selline ongi „Eesti keele morfoloogia teooria“ tase... Nimelt on seal palju loendeid ja reegleid, mille ainus õigustus on, et „nii saab palju ettetulevaid sõnu analüüsitud“, s.t programmi rumalus saab ära varjatud, kuid sügavam põhjendus puudub.



Programm kannab tema arendamisel kasutatud tekstikorpuse pitsarit. See viib meid küsimuse juurde: kas keele kui süsteemi seaduspärasused on üldse keelekorpusest kui kasutusnäidete hulgast tuletatavad? Või peame piirduma tõdemusega, et korpuse põhjal tehtud järeldus kehtibki kindlalt ainult sama korpuse raames?

Tekstikorpus on elektroonilisel kujul olevate tekstide kogu, mis esindab mingit valdkonda ning seal kasutatud keelt (nt ametikirjad, lastehoid) ja mida uurides saab uusi teadmisi nii valdkonna enese kui seal kasutatud keele kohta. Heal juhul võimaldab tekstikorpuse uurimine anda teadmisi, mis kehtivad laiemalt kui korpuse tekstidel; mis kehtivad võib-olla isegi terves keeles ehk kõigi selles keeles loodud juttude-tekstide puhul.

Viimane on võimalik juhul, kui korpus esindab üldkogumit – olles viimasest palju väiksem, on ta sellega mingis olulises mõttes sarnane e. üldkogumi suhtes esinduslik e. representatiivne.

Kumbki aspekt – mis on tekstide ja/või keele „üldkogum“ ja kuidas määratleda „sarnasus“ – ei ole lihtsalt defineeritav. „Mittesarnase“ korpuse peal tehtav statistika oleks aga eksitav.

Üks lahendus oleks vaadata, kas korpuse põhjal leitavad arvulised parameetrid on kooskõlas mujalt pärit andmetega, milleks võib olla nt. keele omandamisprotsess laste poolt, keele muutumine ajaloos, Zipfi seadus vms. Kui on kooskõlas, siis korpus on teatavas mõttes esinduslik.

Mõttekäigu aluseks on eeldus, et keele (all)süsteem – olgu selleks morfoloogia, sõnavara koosseis vms – kujuneb keelekasutuse alusel. Mitte igasugune korpus ei ole selline, et selle põhjal saaks kujuneda keele süsteem sellisena, nagu me seda tunneme. Illustreerimiseks vaadeldakse korpuse esinduslikkust eesti verbimorfoloogia näitel.

Sõna kõik muutevormid ei ole inimesel tervikuna mälus, vaid osa neist moodustatakse sama sõna mõne teise vormi alusel. Alusvormiks sobib ainult selline, mis on juba varem teada, s.t. mida on varem kohatud. Tekstikorpuse sõnavarale rakendatuna tähendab see, et alusvormi leksikaalne kasutusulatus – vormi levimus sõnavara lõikes – ei saa olla väiksem kui moodustatava vormi oma.

Käändsõnade puhul on kõige suurema kasutusulatusega vormiks ainsuse nimetav (s.t. kõige rohkem erinevaid sõnu esineb ainsuse nimetava kujulisena), sellele järgnevad ainsuse omastav, osastav ja seejärel muud käänded. Mitmuse vormid on harvemad kui ainsuse omad.



Vormide kasutusulatuse hierarhia on kooskõlas ka Zipfi seadusega, mille kohaselt sõnavormi sagedus ja pikkus on pöördvõrdelises seoses, s.t. mida sagedasem vorm, seda lühem see on. Nimetav kääne on ilma lõputa, omastava käände puhul lisandub türevokaal, osastava puhul veel lõpp d/t jne.

Pöörsõna vormide kasutusulatuse hierarhia on aga eri korpustes erinev. Osutub, et verbiparadigmade puhul on esinduslikuks korpuseks ainult väikelaste hoidjakeele ja interneti jututubade korpus.

Näib, et mitmed korpused sisaldavad liiga palju lugusid, narratiive, võrreldes selle päris loomuliku keelekasutusega, mis morfoloogia süsteemi kujundab ja mille iseloomule keelesüsteem kohandub. Selles loomulikus keelekasutuses räägitakse pigem soovidest, käskudest ja keeldumistest. Need on iseloomulikud olukordadele, kus inimesed teevad midagi koos, on need siis ema ja laps või ehitusmehed koos maja ehitamas („anna haamer“, „võta ise“, „nii ei saa“, „ära siia astu“).

Keelesüsteemi toimimise teatud aspekte, nt. üleminekukiirust sõna vanalt muutmisviisilt uuele saab jälgida siiski ka mujal kui ainult ülalnimetatud korpustel.

Ettekujutus tekstikorpusest kui keelesüsteemi kujundavast jõust tõstab esile ka küsimuse tiražeerimisest kui olulisest tegurist korpuse koostamisel ja hindamisel. Joh. Aavik ei piirdunud oma uute sõnade puhul nende avaldamisega leksikonides, vaid kasutas neid just laialt leviva meelelahutuskirjanduse tõlgetes.

Digital Dictionary Database for Slovenian: unstructured, semi-structured and structured data in modern lexicography

Simon Krek (Jožef Stefan Institute, Slovenia)

Lexicography in the 21st century is faced with many challenges, from the opportunities arising from the increasingly versed Artificial Intelligence, to the democratisation of knowledge through open collaboration and citizen science projects. However, it seems that the very basic *raison d'être* of lexicography persists - to offer as much knowledge about the vocabulary of a particular language, and its interaction with other languages, as possible. The real challenge in the beginning of the 21st century is how to organise this knowledge which comes in different forms, as unstructured data (text, corpora), semi-structured data (XML, JSON etc.) or structured data (relational databases). Traditionally, from the beginnings of their use in 1980s,



lexicography considered corpora as detached from lexicographic description, as a source consulted at a given point in time, not intrinsically connected with lexicographic resources, with such a connection ultimately enabling dynamic (semi-)automatic monitoring of language use in lexicographic fashion (identification of lexical units, sense division, explanation, exemplification etc.). Now, with the disappearance of book-driven organisation of lexicographic data, there is no need for compartmentalisation of data types. The new challenge is how to monitor, what to monitor, how to store lexicographically organised information, and how to visualise it. The organisation of Digital Dictionary Database (DDDS) for Slovenian tries to provide an answer for some of these challenges. In the talk, we will present DDDS data model, our tools for monitoring language use in real time, and visualisation of various types of data: monolingual data, bi- or multilingual data, corpus data, multi-modal data etc.

Linking Second Language Acquisition Research and Digital Language Learning

Detmar Meurers (University of Tübingen, Germany)

With digital devices permeating many aspects of our daily lives, there are also high hopes that digital tools will give education a big boost. Focusing on second language education, we will argue that in order to fulfil those hopes, digital tools should be more systematically grounded in the general learning and language acquisition mechanisms established in psychology and second language acquisition (SLA) research.

We will sketch a broad range of opportunities for adaptive digital tools grounded in SLA research, from providing motivating, meaningful input that is rich in developmentally proximal forms, via facilitating noticing through input enhancement, to supporting practice and pushed output with scaffolded feedback on form and meaning in a context that functionally links such practice to genuine language tasks. For each of those opportunities, we will showcase digital tools integrating artificial intelligence methods that successfully foster language learning - as confirmed by large-scale field studies in authentic learning settings. By adaptively taking the prevalent individual differences into account, such digital tools also readily complement traditional classroom-based instruction.



Suulised ettekanded / Oral presentations

„Ei saa ju raske olla, kui eesti keel on emakeel.“ – Kas eestlasel on raske eesti keeles kirjutada? Mõningaid küsitluse tulemusi

Kristel Algvere (Tartu Ülikool, Eesti)

Standardkeele ideoloogiast Euroopas on kirjutatud juba mitu aastakümnet. On leitud, et keele standardiseerimine on mõjutanud kogukonna keelekasutust (Milroy 2001: 535). Eestis tekkis 20. sajandi alguses vajadus eesti keelt standardiseerida, sest enamik eestlasi rääkis oma kohalikku murret (Plado 2022: 1092). Lisaks riiklikule keelekorraldusele hoiavad standardkeele ideoloogiaid üleval ka keelekasutajad ise. Seda nähtust on nimetatud „rohujuuretasandi preskriptivismiks“. (Lukač 2018) Ehkki sotsiolingvistid on näinud vaeva selgitamiseks keele varieerumise olulisust, arvavad mõned inimesed jätkuvalt, et peaks olema ainuüksi üks õige viis midagi öelda/kirjutada, varieerumist ei peeta väärtuslikuks (Davies 2000: 122). Keeleliste hoiakute uurimine sotsiolingvistikas on oluline, sest hoiakud on seotud sotsiaalsete normidega, ja sotsiaalsed normid võivad ideoloogilistest väärtustest erineda (Garrett 210: 226). Üks viis uurida keelelisi hoiakuid on korraldada küsitlus.

Selle ettekande baasiks on 2022. aasta detsembris läbi viidud veebiküsitlus, milles osales 668 inimest. Vastajategrupi moodustasid 88% ulatuses naised ja 71%-l vastajatest oli kõrgharidus. Uuringu üheks teemaks oli hinnangu andmine eesti keele õigekirja kasutamisele enda avalikes (nt sotsiaalmeedia postitus) ja eraviisilistes (nt tsätivestlus) tekstides. Sealjuures oli soov heita valgust vigade tegemise hirmu olemasolule kummaski žanris. Teiseks teemaks oli abivahendite kasutamine eestikeelse teksti kirjutamisel – kas, kes, millal ja millist abi kasutanud on?

Küsitluses osalenutest 46% vastas, et eesti keeles kirjutamine ei ole neile üldse raske, samas 67% vastanutest on otsinud eestikeelse teksti kirjutamisel abi (58% on kasutanud veebis leiduvaid õigekeelsusallikaid). Uuringust selgus, et nii avaliku kui ka eraviisilise teksti õigekeelsusele pööratakse tähelepanu, aga rohkem siiski avaliku teksti omale, eravestluses on hirm kirjavigu teha väiksem. Kommentaarid nagu „Eesti keel ju emakeel ja armas pealegi“, „Ei saa ju raske olla, kui eesti keel on emakeel“ vaheldusid kommentaaridega nagu „Kahtlen korrektuses lause-ehituses ja komade paigutuses“ ja „Teatud nüansid ja keelereeglid rasked“. Seega joonistusid välja erinevad vastajate grupid: ühed suhtuvad eesti keeles kirjutamisse kui



millessegi lihtsasse, teistele see nii aga ei ole. Ettekandes tutvustan mõningaid küsitluse tulemusi nüansseeritumalt sotsiolingvistilisest vaatenurgast.

Kirjandus

- Davies, Winifred V. 2000. Language Awareness amongst Teachers in a Central German Dialect Area. *Language Awareness* 9(3), 119–134.
- Garrett, Peter. 2010. *Attitudes to Language (Key Topics in Sociolinguistics)*. Cambridge: Cambridge University Press.
- Lukač, Morana. 2018. Grassroots Prescriptivism: An analysis of individual speakers' efforts in maintaining the standard language ideology. *English Today* 34(4), 5–12.
- Milroy, James. 2001. Language Ideologies and the Consequences of Standardization. *Journal of Sociolinguistics* 5(4), 530–555.
- Plado, Helen. 2022. Rahvakeelsus XX sajandi alguskümnendite Eesti keelekorralduses. (Vaateid eesti keelekorralduse arenguloole.) *Keel ja Kirjandus* 12, 1075–1092.

Veamärgendusega korpus grammatikakontrollija arenduseks ja testimiseks

Kais Allkivi-Metsoja¹, Krista Liin², Karina Kert¹, Silvia Maine¹ (¹Tallinna Ülikool, Eesti. ²Tartu Ülikool, Eesti)

Automaatkorrektuuri tulemuslikkuse hindamiseks on vaja tekstimaterjali, mis sisaldab eri liiki keelevigade näiteid ja parandusi. Eesti keele grammatikakontrollija arendamiseks ja testimiseks koostasime olemasolevate tekstikogude põhjal ühtse veamärgendusega korpuse, mis sisaldab suhteliselt võrdsel määral eesti keele kui teise keele õppijate ja eesti keelt emakeelena kõnelejate toimetamata kirjutisi. Kokku hõlmab korpus ligi 8000 lauset, olles mahult võrreldav rahvusvahelise veaparandussüsteemide võistluse BEA-2019 arendus- ja testandmestikuga (Bryant et al., 2019).

Vigade asukohta, liigi ja paranduse tähistamiseks lausete kaupa kasutasime M2-märgendusformaati (Dahlmeier & Ng, 2012), mis on kujunenud standardiks automaatse veaparanduse hindamisel (Ng et al., 2014; Bryant et al., 2019). Vealiigituse aluseks võtsime ERRANT-i klassifikatsiooni (Bryant et al., 2017), mida kohandasime eesti keele jaoks. Üldisem liigitus tugineb sellele, kas sõna või kirjavahemärk on liigne, puudub või vajab asendamist muu sõnakuju või kirjavahemärgiga. Sõnaasendused jagunevad omakorda täheortograafia-, algustähe-, kokku-lahkukirjutuse, käänd- ja tegusõna vormivaliku, sõnavaliku- ja sõnajärjevigadeks. Viimase puhul asendatakse mitmesõnaline üksus.



Juhtudel, kus samas sõnas on mitu viga, võtsime kasutusele liitveamärgendid. Näiteks võib vormi- või sõnavalikuviga langeda kokku eksimusega täheortograafias või kokku- ja lahkukirjutuses. Kui ebaloomuliku sõnajärgjega lauseosa hõlmab ka mõnda sõnatasandi viga, siis märkisime need eraldiseisvate vigadena, et nende parandamise üle saaks eraldi arvestust pidada. Ühele lausele lisasime kuni kolm märgendusversiooni. Selleks teisendasime eelneva veamärgenduse ja täiendasime seda.

Eesti keele kui teise keele õppijate tekstid pärinevad Eesti vahekeele korpusest ja jaotuvad nelja riiklikult hinnatava keeleoskustaseme (A2, B1, B2 ja C1) vahel. Need olid varem märgendatud CoNLL-U formaadis, kus morfosüntaktilise info kõrval oli välja toodud veainfo koos parandustega – üks märgendusversioon lause kohta. Selles formaadis testmaterjali on kasutatud eesti keele õigekirjakontrolli lahenduste võrdlemiseks, kuid grammatikakontrolli tulemusi on kasulik hinnata mitmekülgsema märgenduse alusel.

Emakeelekõnelejate tekstid on võetud eesti keele õppija korpuse EMMA eesti emakeelega keskkooliõpilaste eksamikirjutiste alamhulgast, kus on märgendatud õigekirja-, kirjavahemärgi-, stiili- ja faktivead. Kuna EMMA märgenduses on vealiigitus üldisem ning seal pole välja pakutud parandusi, siis täpsustasime vealiigi oma märgenduskeemis ning lisasime võimaliku paranduse. Sealjuures on ühe vea kohta eri märgendajatelt eri parandusversioonid. Osad algselt märgendatud vead jäid vaatluse alt välja, nt faktivead.

Ettekandes tutvustame loodud keeleressurssi ja selle kasutusvõimalusi ning arutleme märgenduskeemi väljatöötamise sõlmküsimuste üle.

Keelekasutusmustrid & kontekstid

Pille Eslon, Jaagup Kippar (Tallinna Ülikool, Eesti)

Keelekasutusele omased lingvistilised nähtused tulevad esile mustritena. Nende otsimiseks oleme siin ja varasemalt kasutanud andmekaeve põhimõttel töötavat programmi Klastreidja, mis seatud eraldama 3-komponentseid sõnaliigijärgendeid, kus iga sõna varieerub teatud piires leksikaalselt, morfosüntaktiliselt ja funktsionaalselt. Seda meetodit oleme kasutanud 1890 – 1990 ilukirjanduse, eesti õppijakeele ja Postimehe-Õhtulehe arvamusaluste keelekasutust analüüsides ja võrreldes (nt J. Trainis, K. Allkivi, V. Hallik, P. Eslon).

Kirjeldatud materjali oleme käsitlenud lingvistiliselt, et leida ainek keelekasutuse automaatse analüüsi vahendi arendamiseks. Et treenida masin keelekasutust hindama, on vaja teada,



milliseid sõnaliigijärgnevusi keeletekasutaja eelistab, missugused sõnaliigid esinevad tüüpiliselt muustrile eelnevas, missugused järgnevas kontekstis. Seni on analüüsitud üksiksõnade ümbrust, ent automaatset tekstitötlusvahendit arendades on perspektiivne tuvastada muustrite kui struktuurselt, grammatiliselt ja semantiliselt terviklike üksuste kontekstid.

Ettekandes käsitleme formaaltunnuse alusel eristuvate adverbi ja konjunktsiooni sisaldavate muustrite ümbrust. Selleks on Klustrileidjapõhiselt koostatud piisavalt mugav, kiire ja paindlik abivahend, mis kasutab vislcg3 parseriväljundit, loeb laused ja sõnad koos keeleandmetega (sõnaliik, morfosüntaks) andmepuusse. Otsing algab enamasti ühe sõnaliigi sageduse tuvastamisega muustri ees ja järel. Seejärel leitakse, milliseid sõnaliigikooslusi need moodustavad.

Näiteks kahest adverbist ja sidesõnast koosnev muster DDJ.

1) Lingvistiline kirjeldus

1. lõpeb tavaliselt rinnastava sidesõnaga ja (ju ikka ja), harva ega, või, siis (ei seest ega)
2. paarissidend nii .. kui ka (nii kultuuriliselt kui ka majanduslikult), samuti nii palju kui (nii palju kui vaja (tarvis) 'määr' → nii palju kui võib (saab, tohib) 'lubatud')
3. paarissidendi nii .. kui (ka)~ei .. ega korral sisaldab muster antonüümseid adverbe, millest kujunevad stereotüüpse vormistusega fraasid nagu nii suuliselt kui kirjalikult, nii kultuuriliselt kui majanduslikult, nii tänavu kui eelmisel aastal, nii füüsiliselt kui vaimselt; ei rohkem ega vähem, ei seest ega väljast, ei ühiskondlikult ega isiklikult
4. analoogia noomenite vastandusele rajaneva väljendiga eitavas kõnes, nt ei üks ega teine / ei see ega teine

2) Muustri DDJ ümbrus

A) Sõnaliigid ja sõnaliigikooslused muustri DDJ ees

1. verb DDJ ees (49%) + substantiiv, verb, adverb, pronoomen
2. substantiiv DDJ ees (23%) + substantiiv, pronoomen, verb
3. adverb DDJ ees (17%) + verb, pronoomen
4. pronoomen DDJ ees (11%) + verb, substantiiv

B) Sõnaliigid ja sõnaliigikooslused muustri DDJ järel

1. adverb DDJ järel (39%) + verb, substantiiv, adverb, adjektiiv
2. substantiiv DDJ järel (16%) + verb, konjunktsioon
3. adjektiiv DDJ järel (15%) + substantiiv
4. verb DDJ järel (14%) + pronoomen, adjektiiv, adverb
5. pronoomen DDJ järel (9%) + substantiiv, verb
6. numeraal (7%) + substantiiv



Kuna külastan tihti Tartu kaubamaja, siis on juba enamvähem meelde jäänud kassatibid/kassatädid. Sooliselt markeeritud ametinimetused eesti keeles

Elisabeth Kaukonen (Tartu ülikool, Eesti)

Keele ja soo uurimise üks oluline fookus on olnud sooliselt markeeritud ametinimetuste kasutusel. Seda nimelt põhjusel, et ametivaldkonda puudutav sõnavara annab märku sugude traditsioonilisest tööjaotusest ja ebavõrdsest kohtlemisest ametitel ning näitab, et soorollid on ühiskonna mõtteruumi sügavalt kinnistunud. (Holmes et al. 2009: 192; Penelope 1988: 121) Ametivaldkond on ka üks peamine sotsiaalsete rollide ja soostereotüüpide omistamise pinnas (Koenig, Eagly 2014).

Selles ettekandes vaatlen ametit ja tegevusi väljendavaid sooliselt markeeritud eesti liitsõnu. Täpsemalt on vaatluse all lemmaga *-tädi*, *-onu*, *-tibi* ja *-kutt* lõppevad liitsõnad. Materjalina kasutan eesti keele ühendkorpuse 2021. aasta veebitekstide alamkorpust, mille põhjal vaatlen, millised on nimetatud liitsõnade sõne- ja tüübisagedused korpuses ning sealhulgas naistele ja meestele osutavate liitsõnade osakaalud. Selgitan välja, missuguseid ameteid ja tegevusi nimetatud sõnad väljendavad ning kas naistele ja meestele osutavate sõnadega kerkib esile traditsioonilisi soorolle ja soostereotüüpe. Samuti keskendun küsimusele, milleks kasutatakse sõnu *-tädi*, *-onu*, *-tibi* ja *-kutt* ametit tähistatavates liitsõnades. Need sõnapaarid esindavad pisut värvikamat ja vähem neutraalset, eri vanuses naistele ja meestele osutavat sõnavara.

Uurimused teistest keeltest on näidanud, et soolistatud sõnavara puhul on just naissoole osutavad sõnad need, mis läbivad tähendusliku muutuse ja millest saavad madalamale staatusele osutavad, kohati halvustavad terminid (Schulz 1990; Baker 2014). Ettekandes näitan, et nii ka eesti keeles – näiteks tädid töötavad madalama staatuse ja palgaga ametitel kui onud.

Perspectives from the academic writing community on the impact of Generative AI: a discussion

Djuddah A. J. Leijen, Helen Hint, Helena Lemendik (University of Tartu, Estonia)

This session will provide a conversation about the introduction of generative AI writing technologies (such as ChatGPT) and the impact it has on student writing and the teaching of writing. Concerning student writing, in recent discussions at the conference Writing Research Across Borders (WRAB23 Trondheim), which gathered writing researchers from around the



globe, personal examples were provided by participants how students are able to make clever use of ChatGPT to complete badly formulated course assignments. However, the practitioners (writing instructors) were generally enthusiastic such generative AI tools can offer both to instructors and students. Some concerns were raised related to plagiarism; however, these concerns also provided opportunities to create better learning assignments for our students when it comes to writing.

Within the context of Estonia, the debates around and applications of the teaching of writing skills has increased in secondary and higher education. As we know, teaching writing includes addressing questions related to what (i.e. what does a text look like and, what is considered a good text?) and how (i.e. how does a text come to be, and how can we evaluate a text?), just to name a few. Generative AI writing technologies, such as ChatGPT, on the surface, seems to provide easy access to both the form of a text (what should a text look like), and the how (which is through inserted generative prompts). The question is, whether Estonian, as a smaller language model, is able to generate “good” texts. More importantly, whether the texts are biased and modelled according to Anglo-American traditions.

In this session, we invite a more general discussion amongst the participants to share their personal experiences regarding student use of ChatGPT for their writing assignments, and also share their concerns or excitement about generative AI in general. Furthermore, we will initiate the discussion providing some evidence from writing studies which highlight the importance of a more common approach to the teaching of writing skills to our students, focusing on issues such as assessment, form, and knowledge construction. Furthermore, we invite discussions how we, practitioners and researchers, can use generative AI to help improve our own practices.

Kõnetempo varieerumine eesti keeles

Pärtel Lippus, Maarja-Liisa Pilvik, Kaidi Lõo, Anton Malmi, Liina Lindström (Tartu Ülikool, Eesti)

Kõne on isikuti varieeruv ja suur osa sellest, kui jutukas keegi on või kui kiiresti ta tavaliselt räägib, sõltub tema isikuomadustest: on erineva kõnemaneeeriga jutukamaid ja vähemjutukaid, kiiremaid ja aeglasemaid inimesi. Individuaalse varieeruvus kõrval on aga leitud ka sotsiodemograafilisi mustreid, näiteks naised räägivad meestega võrreldes aeglasemalt, selgemalt ja ilmekamalt, samuti vananedes kõnetempo aeglustub (nt Simpson 2009).



Meister ja Meister (2022a, 2022b) on uurinud eesti noori vanuses 9-18 aastat ning vanemaealisi vanuses 60-80 aastat ja leidnud, et noortel kõnetempo kasvas 10. ja 14. eluaasta vahel, vanemaealistel see aga aeglustus ja olulised olid ka erinevused sugude vahel. Aare ja Lippus (2017) seevastu ei leidnud täiskasvanud kõnelejate põhjal, et sugu oleks tempot oluliselt mõjutanud, küll aga leiti, et terve täiskasvanuea jooksul kõnetempo vanusega aeglustus.

Ettekandes vaatame erinevate eesti keele kõnekorpuste põhjal, kuidas muutub kõnetempo sõltuvalt vanusest, soost, vestluspartnerist spontaanses suhtlussituatsioonis ning kõrvutame tulemusi ka loetud kõnest saadud andmetega. Esialsed tulemused on näidanud, et spontaankõnes on kõnetempo loetuga võrreldes oluliselt kiirem ning vanusest tingitud erinevused väiksemad, mis võib tuleneda erinevustest lugemisoskuses ja -harjumustes.

Kirjandus

- Aare, Kätlin & Pärtel Lippus. 2017. Some gender patterns in Estonian dyadic conversations. In Jardar Eggesbø Abrahamsen, Jacques Koreman & Wim van Dommelen (eds.), *Nordic Prosody. Proceedings of the XIth Conference*, Trondheim 2016, 29–38. Frankfurt am Main: Peter Lang. <https://doi.org/10.3726/b11152>.
- Meister, Einar & Lya Meister. 2022a. Estonian Elderly Speech Corpus – Design, Collection and Preliminary Acoustic Analysis. *Baltic Journal of Modern Computing* 10(3). <https://doi.org/10.22364/bjmc.2022.10.3.09>.
- Meister, Einar & Lya Meister. 2022b. Eesti laste kõne III. Kõnetempo ja silbikestuste analüüs. *Keel ja Kirjandus* 65(3). 226–245. <https://doi.org/10.54013/kk771a3>.
- Simpson, Adrian P. 2009. Phonetic differences between male and female speech. *Language and Linguistics Compass* 3(2). 621–640. <https://doi.org/10.1111/j.1749-818X.2009.00125.x>.

Eesti keele õigekirja- ja grammatikakontroll: mudelite võrdlus ja kombineerimine

Agnes Luhtaru¹, Mark Fišel¹, Krista Liin¹, Kais Allkivi-Metsoja², Jaagup Kippar² (¹Tartu Ülikool, Eesti. ²Tallinna Ülikool, Eesti)

Eestikeelse teksti automaatkorrekatuur on seni koosnenud mitmest omavahel sidumata komponendist, nagu reegli põhine õigekirjakontroll, komavigade tuvastamine ja eksperimentaalne masintõlkepõhine grammatikakorrekter. Käimasolevas projektis arendame edasi nii õigekirja- kui ka grammatilist kontrolli, mis kätkeb näiteks vormivaliku ja sõnajärje



vigade parandamist. Selleks võrdleme erinevaid lähenemisi, kasutades mitmesuguseid keelevigu sisaldavaid testimishulki.

Oleme Vabamorfi reeglipõhise spelleriga võrrelnud statistilisi õigekirjaparanduse tööriistu, mis arvestavad rohkemal või vähemal määral sõna kontekstiga: Symspelli ja Peter Norvigi bigrammidele ehk sõnapaaridele tuginevat algoritmi ja Jamspelli algoritmi, mis põhineb trigrammide ehk sõnakolmikute statistikal. Parimaid tulemusi oleme saavutanud Jamspelliga, mis võimaldab sageli parandada ka vigu, mis on olemasoleva sõna homonüümid (nt ilm oli vaga ~ väga ilus), ja mõnikord korrigeerida vormi- või sõnavalikut (nt endi ~ enda, enne ~ enne). Oleme Jamspelli katsetanud erineva treeningmaterjaliga, mis pärit 2019. aasta ühendkorpusest. Suurima parandustäpsuse saime veebikorpuse materjaliga (40,9 mln lauset), samas kui vigade leidmise saagis osutus suurimaks eesti keele koondkorpuse valimiga, mis moodustab umbes poole kirjakeelsest tekstikogust (6 mln lauset).

Grammatikakontrollija loomiseks kasutame lähenemist, mis põhineb tehisnärvivõrkudele tugineval masintõlkel ning seisneb selles, et vigane lause teisendatakse sama keele normipärasemaks lauseks. Võrdleme grammatiliste vigade parandamisel kahte meetodit: valdkonnas levinud sünteetilistel andmetel treenimist (Náplava & Straka, WNUT 2019) ja mitmekeelsel neuromasintõlkel põhinevat lahendust (Korotkova, et al., 2019). Esimene põhineb ühekeelsesse korpusesse sageduse põhjal vigade lisamisel, teine ühekeelsel zero-shot tõlkel, mille saame mitmekeelse masintõlke mudeli ühekeelsel rakendamisel. Mõlemal meetodil treenitud mudeleid treenisime ka edasi veaparanduskorpusega.

Automaatsetete meetrikate alusel on mudelite paranduste profiil erinev. Mitmekeelsel neuromasintõlkel põhinev meetod on kõrge saagisega, kuid asendab tihti ka korrektseid sõnu, mistõttu täpsus on madal. Sünteetilistel andmetel eeltreenimine seevastu jätab rohkem vigu parandamata, kuid lisab vähem tarbetuid parandusi. Veaparandusandmetega häälestamine tõstab sünteetilistel andmetel treenitud mudeli saagist ja mitmekeelse masintõlke täpsust, kuid mudelid erinevad siiski käitumiselt. Kombineerisime neid mudeleid ka spelleritega.

Ettekandes tutvustame grammatika- ja õigekirjakontrolli mudelitega eraldi ja kombineeritult saadud parandustulemusi vealiikide kaupa.

Kirjandus

- Náplava, J., & Straka, M. (2019). Grammatical Error Correction in Low-Resource Scenarios. Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019), Hong Kong, China (lk 346–356). Association for Computational Linguistics.



- Korotkova, E., Luhtaru, A., Del, M., Liin, K., Deksnē, D., & Fišel, M. (2019). Grammatical error correction and style transfer via zero-shot monolingual translation. <https://doi.org/10.48550/arXiv.1903.11283>

Eesti keele sõnavara testi LexEsti loomise taustast ja koostamisest

Kaidi Lõo, Agu Bleive, Katrin Leppik, Anton Malmi (Tartu Ülikool, Eesti)

Senini on eesti keele taset teise keelena, sh sõnavara tundmist, mõõdetud peamiselt eesti keele tasemeeksamitega Euroopa keeleõppe raamdokumendi (CEFR) raames kuuel tasemel A1-C2. Tasemeeksamid on vajalikud oma keeleoskuse tõendamiseks elamisloa või kodakondsuse taotlemiseks, need on suhteliselt ajamahukad ning ei ole avalikult kättesaadavad. Eesti Keele Instituudis on loodud eesti keele kui teise keele õpetaja tööriistad (Kallas et al. 2021), mis aitavad eesti sõnavara mõõta CEFR-i keeletasemete järgi. Niisiis on hetkel olemasolevad sõnavara hindamise vahendid konkreetse probleemi lahendamiseks ning nende kasutamine on ajamahukas. Puudu on kiiretest ja automaatsetest sõnavara mõõtmise vahenditest.

Võõrkeele uuringutes on sõnavara mõõtmiseks sageli kasutatud sõnavara testi LexTALE erinevaid variante (ingl. Lexical Test for Advanced Learners of English, Lemhöfer ja Broersma, 2012). Testis tuleb otsustada umbes saja stiimuli puhul kas sõna on vastavas keeles esinev või mitte. Sellise katse tegemine on kiire (tavaliselt 5-10 minutit) ning tulemuseks on õigete vastuste protsent, mis on annab aimu keeleõppija sõnavarast. Kui algset inglise keele põhjal koostatud testi kasutati vaid edasijõudnud keele õppijatega, siis nüüdseks on testi erinevate keelte edasiarendused (nt hispaania keele LexTale-Esp, itaalia keele LexIta soome keele Lexize) näidanud, et sellist testi saab edukalt kasutada eri keeletasemega keeleõppijate puhul (Izura jt 2014, Amenta jt 2021, Salmela jt 2021).

Eesti keele sõnavara test LexEst on hetkel koostamisel. Eeskujuna oleme kasutanud soome keele sõnavara testi Lexize (Salmela jt 2021). LexEst on lühem variant leksikaalsest otsustuskatses (inglise keeles lexical decision task), mida on varem eesti keeles kasutatud eesti emakeele rääkijatega (Lõo jt 2018). Testi lõppvariandis on 90 stiimulit, millest 45 on sõnad (nt maja) ja 45 pseudosõnad (nt naja). Osalema värbame 200 erineva taustaga eesti keele õppijat ja testi viime läbi veebis. Mõõdame nii vastuste õigsust kui ka vastamise aega. Eelnevate sarnaste uuringute põhjal eeldame, et kõrgema keeleoskusega eesti keele oskajad teevad katse täpsemalt ja kiiremini kui madalama keeleoskusega eesti keele oskajad. Eeldame ka, et suurim erinevus kõrgema ja madalama keeleoskusega eesti keele oskajate



vahel tuleb välja just madala sagedusega sõnade puhul. Samuti eeldame, et sarnaselt korea keele kõnelejatele (Lemhöfer ja Broersma 2012), on ka teistel Ida-Aasia kultuuritaustaga testi tegijatel suurem erinevus iseenda antud eesti keele oskuse hinnangu ja LexEsti skoori vahel (viimane annab adekvaatsema pildi), kui Euroopa kultuuritaustaga testi tegijatel.

Kokkuvõtteks, LexEst sõnavara testiga saab kiiresti ja objektiivselt hinnata keeleõppijate sõnavara suurust eelkõige keeleteaduslike katsete kontekstis. Samas kuna plaanime valideeritud testi tulevikus ka avalikult kättesaadavaks teha, on huvi korral testi võimalik kasutada ka näiteks eesti keele õppijatel ja õpetajatel.

Teooria ja rakendus: keelekorraldus

Peep Nemvalts (Tallinna Ülikool, Eesti)

ÕSi ja EKI teatmiku koostamise [tegevuskavas 2022–2025](#) mainitakse, et ÕS 2025 on „koostatud vastavalt eesti keelekorralduses alates 1980 kehtinud põhimõtetele“. Neid põhimõtteid selgelt koondatuna pole EKI allikaist leida. Küll on EKI teatmikus peatükk [„Keelekorralduse teisenemine“](#), mille märksõnadena esitatud *keelearendus*, *keelekorraldus*, *õigekeelsus*, *õigekiri*, ning kus nenditud: „*keelekorraldus* võib tähendada vähemalt kolme eri asja“. Sealsamas väidetakse: „Tänapäeva eesti keelekorraldus on jõudmas uude arengujärku. Käimas on intensiivsed arutelud – koostamisel on keelearenduse käsiraamat ja mõnedki küsimused on veel lahtised –, aga olulisema võib kokku võtta järgmiste punktidega.“

1. „Terminina on eelistatud **keelearendus**, mis on võrreldav keelekorraldusega traditsioonilises tähenduses.“ Seda „traditsioonilist“ *keelekorralduse* tähendust on ühena kolmest esitletud Tiiu Erelti (Eesti keelekorraldus. Tallinn 2002, lk 15) sõnastuses: „Kirjakeele teadlik arendamine, rikastamine, stabiliseerimine ja ajakohastamine.“ Puudub põhjendus, miks eelistada uut terminit, kui selle mõistesisu „on võrreldav“ senisega. Mida võrreldavus õigupoolest tähendab? Kas osundatud määratlusest jääb vaid esimene neljast sisutunnusest? Selgust ei lisa määratlus „Keelearendus on teaduspõhine kaasaaitamine keele loomulikule arengule, et edendada mõtteselgust ja väljendusrikkust.“ Mida kätkeb „kaasaaitamine“? Mis on „keele loomulik areng“?

2. ja 4. räägivad *normingust* ning *objektiivsest normist*. Nagu osutatud, tekitavad needki sarnassõnad terminitena rohkem segadust kui selgust, mistap neid ei tarvitseks nõnda tarvitada ([Nemvalts 2015, 2016](#)).

3. „Uurimisobjektiks on kogu kirjakeel, ühtne, kõnes ja kirjas üldrahvalikult kasutatav keel. Standardkeel on üks kirjakeele osa. Kirjakeeles ehk ühiskeeles korrastame seda osa, mida on vaja avaliku, haridus- ja töösfääri ladusaks asjalikuks kommunikatsiooniks.“ Needki väited tekitavad uusi küsimusi. Kas *korrastama* tähendab sedasama mis *korraldama*? Millist osa siis



„korrastame“, kas standardkeeleks nimetatut või midagi muud? Mis on teised kirjakeele osad? Mida ikkagi tähistab *kirjakeel*? Kas on seegi mitmetähenduslikult tarvitatud termin liiga eksitav?

EKI teatmiku peatükis „Mis on kirjakeel“ määratletakse: „**Kirjakeel** ehk **ühiskeel** (ka: **ühiskirjakeel**) on ühtne, kõnes ja kirjas üldrahvalikult kasutatav keel, keele tuumosa, mis käibib suhtluses, kirjanduses, ajakirjanduses ja mujal.“ Peatükis „Üldkeel ja oskuskeel“ teisiti: „**Üldkeel** on ühtne, üldtarvitav keel, mis käibib suhtluses, kirjanduses, ajakirjanduses ja mujal. Seda kasutab ühisvarana kogu rahvuskeele kasutajaskond vanusest, ametist, haridusest ja elupaigast olenemata.“ Rein Kull (2000: 11) on defineerinud: „Kogu rahvusele kuuluvat, teadlikult ja sihipäraselt väljaarendatud ning korraldatud keelekuju nimetatakse **kirjakeeleks**. Kirjakeeles toimub kooliõpetus ja ametlik suhtlemine, seda kasutavad teabevahendid.“ ja „On veel kasutatud nimetust **ühiskeel**, millega on mõeldud kirjakeelelähedast üldkasutatavat kõnekeelt.“ Mis siis on kirjakeel, ühiskeel, üldkeel? Neid ja seonduvaid sasipuntraid (Nemvalts 2021) harutades pürgigem selgusele.

Keelekasutaja valikud rööpsete kirjakuju puhul: *ekleer* või *ekläär*?

Tiina Paet, Madis Jürviste (Eesti Keele Instituut, Eesti)

Laen- ja võõrsõnade osakaal eesti keeles on märkimisväärne. Uusimas võõrsõnade leksikonis (VSL 2012, veebis 2015) on ligi 10% märksõnadest kas prantsuse lähte- v vahenduskeelega. Keel laenab sõnu enamasti mingis mõttes prestiižsemast, mingis eluvaldkonnas teatud aspektis kõrgema staatusega keelest, nt merendussõnad on tulnud hollandi ja inglise keelest, prantsuse keelest pärinevad diplomaatia- (*atašee, diplomaat*), moe- (*mansett, puuder*) ja sõjandussõnad (*kürassiir, seersant*) jne. Prantsusmaa on tooni andnud ka kokanduse vallas: koos toitudega on üle võetud nende toitude nimetused (nt *baguette* ja *bagett, croissant, mousse* ja *muss; ratatouille*) (vt ka Jürviste, Paet 2023).

Teiste seas on prantsuse keelest saadud sõna *ekleer* (*ekläär*) (Jürviste, Paet 2023). Laenamine ja mугanemine on keerukad protsessid ja eri põhjustel tuleb laenavasse keelde rööpseid kirjakujusid.

Oma ettekandes keskendumine võõrsõnade rööpsete kirjakuju esitusele: vaatleme, kuidas on kohanenud eesti keeles prantsuse laenud (*ekleer, ekläär*), ning paigutame selle küsimuse laiemasse konteksti: kui ÕS andis varem suuniseid ühe või teise kirjakuju eelistamiseks, siis kuidas esitada neid sõnu järgmises ÕSis (ilmub 2025 Sõnaveebis ÕSi vaatenäol) kasutajale kõige arusaadavamalt? Millised tegurid ajendavad keelekasutajat eelistama üht või teist



sõnakuju? Kas kasutaja valikuid rööpkujude vahel saab (või tuleb) suunata? Kuidas seda teha?

Vaatleme *ekleeri* (*eklääri*) näitel, kuidas võõrsõnad keelde tulevad ja millised on nende kuju fikseerumisel kõige olulisemad ajaloolised verstapostid. Kes seda protsessi kõige enam mõjutas(id)? Paljude prantsuse keelest üle võetud kokandussõnade teekond sai alguse esimese (rootsi keelest) eesti keelde tõlgitud kokaraamatuga 18. sajandi lõpus (Warg 1781). 19. sajandil hakkasid algselt mõisaköökidele mõeldud retseptiraamatud laiemalt levima ja koos retseptidega levisid ka kokandussõnad. Käibeale läks hulk võõrsõnu ja eesti keeles algas nende mугanemise protsess.

Seega leidub just kokandusvaldkonnas võõrsõnade rööpvormide kohta palju põnevat uurimisainest, kuna kokandussõnad on ühtaegu nii üld- kui ka oskuskeele sõnad ja enamus neist on suure kasutussagedusega. Seega illustreerivad arengud selles valdkonnas hästi keelekasutaja tehtavaid valikuid laiemalt.

Kirjandus

- VSL 2012 = Võõrsõnade leksikon. Koost Eduard Vääri, Richard Kleis, Johannes Silvet. Peatoim. Tiina Paet. Toim. T. Paet, Tuuli Rehemaa, Katrin Kuusik, Argo Mund, Kaspar Kolk, Ülle Viks. 8., põhjalikult ümber töötatud tr. Eesti Keele Instituut. Tallinn: Valgus.
- ÜK 2021 = Eesti keele ühendkorpus 2021. DOI: doi.org/10.15155/3-00-0000-0000-0000-08D17L
- Jürviste, Madis; Paet, Tiina 2023. Prantsuse päritolu võõrsõnade kirjakuju varieerumisest eesti keeles. Ekläär ja ekleer. (Ilmumas.)
- Paet, Tiina 2022. Võõrsõnade kuju normimise probleeme tänapäeva eesti keeles. Keel ja Kirjandus, 10, 923–947. DOI: 10.54013/kk778a3
- Warg, Christina 1781. Köki ja Kokka ramat, mis Rootsi kelest Eesti-ma Kele üllespandud on. Tõlk. Johann Lithander. Tallinn: A. H. Lindfors.

Alzheimeri tõvele viitavate märkide automaatne tuvastamine kõnest: kuidas, milleks, kellele ja kas?

Ulla Petti (Cambridge'i Ülikool, Suurbritannia)

Keel ja kõne kannavad olulist informatsiooni inimese tervisliku seisundi kohta. Kõne- ja keeleandmete automaatne töötlemine võimaldab tabada varaseid kognitiivsete võimete taandumisele viitavaid märke, mis võiksid aidata tuvastada neurodegeneratiivseid haiguseid,



näiteks Alzheimeri tõbe. Muutused keeles võivad ilmneda näiteks sõnavaras, lauseehituses, semantikas, ladususes, kõne tempos ja akustikas. Praeguseks on keelekasutusel põhinev automaatne klassifikatsioon juba Alzheimeri tõve diagnoosi saanute ja tervete inimeste vahel üsna täpne, seega keskendutakse viimasel ajal pigem väga varaste potentsiaalsete haiguse märkide ja pikaajaliste muutuste uurimisele. Ideaalis oleks selliste muutuste uurimiseks vaja pikaajalist, üle aastakümnete lindistatud kõnekorpusi paljude kõnelejadega, kellest osadel areneb lõpuks välja Alzheimeri tõbi ja osadel mitte. Kuna sellise korpusi koostamine on kulukas ja aeganõudev, otsitakse erinevaid alternatiivseid lahendusi, näiteks juba avalike kirjalike tekstide või lindistuste uurimine.

Kuigi Alzheimeri tõvele praegu ravi puudub, aitaks varaste märkide tuvastamine inimestel paremini tulevikku planeerida ning iseseisvalt otsuseid langetada. Samuti võimaldaks varane haiguse tuvastamine otsida abi ja tuge välistest allikatest. Kognitiivsete võimete langust aeglustavate ravimite arenedes kasvab ka varaste haigusele viitavate märkide tabamise olulisus. Kuigi keele- ja kõnepõhistel tööriistadel oleks potentsiaali pakkuda mitte-invasiivseid, odavaid, kiireid ja kergesti kättesaadavaid lahendusi varaste Alzheimeri märkide tuvastamiseks, tõstatab loodav tehnoloogia ka mitmeid eetilisi küsimusi, mis on seotud näiteks autonoomia, privaatsuse ja andmekaitse, heaolu, läbipaistvuse, ja õiglusega. Autonoomia küsimused seostuvad näiteks informeeritud nõusolekuga dementsuses, ebaisikustamise ja tulemuste teatamisega. Privaatsuse ja andmekaitse küsimused kerkivad nii seoses isikliku sisuga kõneandmete kui ka delikaatsete terviseandmete töötlemise, hoiustamise ja jagamisega. Heaolu alla kuuluvad näiteks testimisega kaasneva stressi ja tulemuste usaldusväärse küsimused; samuti on eriti oluline kaaluda, kas arendatavat tehnoloogiat võiksid väärtarvitada tööandjad või kindlustusfirmad, mis tooks kaasa ulatusliku diskrimineerimise riski. Läbipaistvus puudutab nii analüüsitavaid keelelisi parameetreid kui ka masinõppel põhinevaid otsuseid ning seda, kas need on arusaadavad ja tõlgendatavad nii arendajatele kui praktiseerivatele arstidele. Õigluse kohapealt on paslik küsida, kes lõpuks sellisest tehnoloogiast võidavad ning kuidas mõjutavad potentsiaalsete kasude ja riskide jaotust keeleline, kultuuriline, sooline, või vanuseline kuuluvus.

Konverentsil soovin rääkida kõne muutustest Alzheimeri tõves ja nende muutuste tuvastamisest arvutilingvistika abil, tutvustada osana doktoritööst valminud keelekorpusi, mis koosneb YouTube'is avalikult saadaolevate ja üle aastakümnete lindistatud kuulsate inimeste intervjuude transkriptsioonidest, kellest pooltel tuvastatakse hilisemas eas Alzheimeri tõbi ja pooltel mitte, ning juhtida tähelepanu ka arendatava tehnoloogiaga kaasnevatele eetilistele küsimustele.



Spontaanses stiilis kõnesüntees

Liisi Piits¹, Heete Sahkai¹, Liis Ermus¹, Indrek Hein¹, Meelis Mihkla¹, Indrek Kiissel¹, Kristjan Suluste¹, Egert Männisalu¹, Rene Altrov¹, Hille Pajupuu¹, Jaan Pajupuu² (¹Eesti Keele Instituut, ²Vabakutseline tarkvaraarendaja)

Ettekandes tutvustame esimeste eestikeelsete spontaanses stiilis kõnelevate sünteeshääle loomist. Kõik varasemad eestikeelsed sünteeshääled olid treenitud etteloetud kõnel ja sobisid hästi uudiste või õppematerjali ettelugemiseks. Kõnesünteesi vajatakse aga ka rakendustes, kus väljundkõne peaks kõlama nii, nagu oleks tegu loomuliku vestlusega, mitte etteloetud kirjaliku tekstiga. Sellised rakendused on näiteks virtuaalsed assistendid ja vestlusrobotid, kõnetute suhtlusabivahendid, sõnumivahetus, subtiitrite helindamine, suuline masintõlge.

Eduka kõnesünteesi alus on hoolikalt komponeeritud sobivas kõnestiilis ja mahus hea salvestuskvaliteediga kõnekorpused ja neile vastavad tekstikorpused. Spontaanses stiilis kõnesünteesi arendamiseks on varem kasutatud studios salvestatud etteloetud dialooge või repliike (Zandie jt 2021, Adigwe ja Klabbers 2022), studios salvestatud spontaansid dialooge (Huang jt 2017, Adigwe ja Klabbers 2022) ja monolooge (Székely jt 2020, Lameris jt 2022) ning olemasolevaid podcast'e (Székely jt 2019, Yan jt 2021). Kuna nii spetsiaalse treeningandmestiku salvestamine kui ka olemasolevast spontaansest kõnest treeningkorpuse koostamine on töö- ja ajamahukas protsess, proovisime eestikeelsete spontaanses stiilis sünteeshääle loomise käigus treeningandmestikuna kasutada ka audioraamatute otsekõnet, mis on juba nii heli kui tekstina olemas ja etteloetud žanridest eeldatavasti vestlusstiilile kõige lähedasem, kuid see ei osutunud siiski võrdlusaluselt paremaks. Töömahukuselt järgmise variandina kasutasime treeningkorpuse loomiseks olemasolevat taskuhäälingut, mille põhjal lõime nii nais- kui meeshääle korpuse.

Loodud korpuste põhjal sünteeshääle treenimiseks kasutasime kolme eri meetodit, millest edukaks osutusid kaks. Hindamistulemused näitasid, et spontaansel andmestikul treenitud hääletega sünteesitud lausungeid tajuti oluliselt spontaansemana kui loetud kõnel põhinevate hääletega sünteesitud lausungeid.

Kirjandus

- Adigwe, A. O., Klabbers, E. 2022. Strategies for developing a Conversational Speech Dataset for Text-To-Speech Synthesis. Interspeech 2022, 2318–2322.
- Huang, Y.-C., Wu, C.-H., Chen, Y.-Y., Shie, M.-G., Wang, J.-F. 2017. Personalized Spontaneous Speech Synthesis Using a Small-Sized Unsegmented



Semispontaneous Speech. IEEE/ACM Trans. Audio Speech Lang. Process, 25(5), 1048–1060.

- Lameris, H., Mehta, S., Henter, G. E., Gustafson, J., Székely, É. 2022. Prosody-controllable spontaneous TTS with neural HMMs (arXiv:2211.13533).
- Székely, É., Henter, G. E., Beskow, J., Gustafson, J. 2019. Spontaneous Conversational Speech Synthesis from Found Data. Interspeech 2019, 4435–4439.
- Székely, E., Edlund, J., Gustafson, J. 2020. Augmented Prompt Selection for Evaluation of Spontaneous Speech Synthesis. Proceedings of the Twelfth LREC Conference, 6368–6374.
- Zandie, R., Mahoor, M. H., Madsen, J., Emamian, E. S. 2021. RyanSpeech: A Corpus for Conversational Text-to-Speech Synthesis. Interspeech 2021, 2751–2755.
- Yan, Y., Tan, X., Li, B., Zhang, G., Qin, T., Zhao, S., Shen, Y., Zhang, W.-Q., Liu, T.-Y. 2021. Adaptive Text to Speech for Spontaneous Style. Interspeech 2021, 4668–4672.

Võro ja seto keele kasutamine ja keelevahetus 20. sajandi jooksul

Helen Plado^{1,2}, Liina Lindström¹, Maarja-Liisa Pilvik¹, Triin Todesk¹, Annika Kängsepp¹ (¹Tartu Ülikool, Eesti. ²Võru Instituut, Eesti)

Kagu-Eestis kõneldakse piirkondlike keeltena võro ja seto keelt, mille oskajate arv on 2021. aasta rahvaloenduse järgi kokku 97 300. Siiski näitavad uurimused, et lõunaeesti keele oskuses on tugev põlvkondlik varieerumine: vanemad inimesed hindavad oma keeleoskust tunduvalt paremaks kui noored inimesed (vt nt Koreinik, Praakli 2013; Koreinik 2015). Olukorda ilmestab ka see, et võro ja seto keelt kasutatakse vanemate sugulastega rääkimiseks tunduvalt enam kui lastega rääkimiseks (vt nt Koreinik 2015), st on katkemas või juba katkenud põlvkondadevaheline keele edasiandmine. Seda peetakse aga peamiseks keele ohustatuse näitajaks (Fishman 1991: 92–95). Varem ongi Võrumaa Sute küla põhjal näidatud, et intensiivne keelevahetus on toimunud 1960.–1980. aastail (Org jt 1994).

Suhtumist lõunaeesti keelde ja selle kasutamisse on kindlasti mõjutanud koolikeskkond. Kuigi lõunaeesti keel ei ole kunagi olnud ametlikult õppekeel, kasutati seda koolides 1920.–1930. aastatel, mil oli tavapärane olukord, kus loeti ja kirjutati eesti keeles, kuid loetu üle arutleti lõunaeesti keeles; samuti oli siis ainumõeldav see, et väljaspool tunde suhtlesid lapsed omavahel lõunaeesti keeles. Samas aga nägid emakeeleõpetajad ja ka mitmed koolinõunikud kohaliku keele kasutuses ohtu eesti kirjakeelele. Olukord hakkaski muutuma 1930. aastate lõpus, mil üha enam sai koolis valitsevaks eesti ühiskeel; nõukogude ajal häbimärgistati koolis



lõunaeesti keele kasutamist. (Lõbu 2005: 133–138) Ka meie 2021.-2022. aastal tehtud Setomaa välitööde intervjuudest võib järeldada, et kuigi veel 1940. aastatel ja 1950. aastate alguses oli Setomaa valdavalt setokeelne, oli koolielu peamiselt eestikeelne. Vahetundides aga ei keelatud lastel omavahel seto keeles rääkida. Olukord on teine 1950. aastatel sündinud keelejuhtide puhul. Nemad on öelnud, et koolis keelati seto keeles rääkida (ka vahetundide ajal) ja et seto keele kasutamise eest võis ka karistada saada.

Seega on keelevahetus toimunud 20. sajandi jooksul. Ettekandes püüame keelevahetuse perioodi ja põhjuseid täpsemalt kaardistada, otsides vastuseid järgmistele küsimustele:

- Kellega on viimase 70–80 aasta jooksul eri põlvkonnad lõunaeesti keeles rääkinud? Millal on jäänud lastega lõunaeesti keele kasutamine marginaalseks?
- Kas lõunaeesti keelt on koolis kasutatud ja kuidas on olnud suhtumine sellesse?
- Kuidas suhtutakse lõunaeesti keelde ja selle õpetamisse, st kas viimaste aastakümnete pingutused lõunaeesti keele prestiiži tõstmiseks on vilja kandnud?

Küsimustele vastame kahe andmestiku põhjal: 1) 2021. ja 2022. aastal Setomaal tehtud välitööde intervjuude põhjal ja 2) 2022. a lõpus ja 2023. a alguses tehtud veebiküsitluse põhjal. Välitööde andmestik hõlmab intervjuusid 26 keelejuhiga, veebiküsitlusele vastas 660 võro ja seto juurtega inimest.

4. klassi õpilaste lugemis- ja kuulamisoskus

Helin Puksand, Erika Klettenberg, Kerli Haav (Tallinna Ülikool, Eesti)

Me elame ümbritsetuna mitmesugustest tekstidest, mida me loeme või kuuleme. Elus hakkama saamiseks tuleb tekstidest aru saada, olenemata nende esitusviisist, nii et kuulamis- ja lugemisoskus on igapäevaelus võrdset olulised. Koolis pööratakse aga enamasti rohkem tähelepanu lugemisoskuse arendamisele, kuulamisoskusega tegeletakse vähem. Võõrkeeleeõppes peetakse kuulamisoskust üheks keerulisemaks oskuseks, aga kuidas on lood emakeeles kuulamisega, on seni suhteliselt vähe uuritud. Veelgi vähem on võrdlevalt uuritud emakeelse teksti mõistmist lugemisel ja kuulamisel.

Lapsel areneb kuulamisoskus emakeeles ebateadlikult juba enne kooli, kuid lugemisoskuse arendamisega tegeletakse peamiselt koolis. Douglas Fisher ja Nancy Fray (2014) toovad, et õpilased mõistavad kuni kolmanda kooliastmeni teksti paremini kuulates ja alles seejärel mõlemad oskused ühtlustuvad. See on ka üks põhjus, miks lugemiskustega laste õpetamisel soovitatakse pikkade teksti lugemise asemel lasta neid tekste kuulata (Wagner et al., 2020). Seega tuleb koolis pöörata tähelepanu mõlema oskuse arendamisele.



Oma ettekandes anname ülevaate uurimusest, mille eesmärk oli võrrelda 4. klassi õpilastel teksti mõistmist lugemisel ja kuulamisel.

Kirjandus

- Fisher, D., & Frey, N. (2014). Speaking and listening in content area learning. *The Reading Teacher*, 68(1), 64–69.
- Wagner, R. K., Zirps, F. A., Edwards, A. A., Wood, S. G., Joyner, R. E., Becker, B. J., ... & Beal, B. (2020). The prevalence of dyslexia: A new approach to its estimation. *Journal of learning disabilities*, 53(5), 354–365.

Automatic Part-of-Speech Tagging with CLAWS7: Impact of Learner Errors

Liina Tammekänd, Reeli Torn-Leesik (University of Tartu, Estonia)

The paper gives an overview of how learner errors affect the output of the CLAWS7 tagger's automated assignment of part-of-speech (POS) tags. A sample of 24,812 words of the Tartu Corpus of Estonian Learner English (TCELE) was tagged using the CLAWS7 tag set. Learner errors in the sample were identified and an error taxonomy was created by dividing errors in the TCELE sample into two broad categories (spelling errors and language errors), which were subcategorized. The POS-tagged and error-tagged samples were collated and compared to identify the error taxons that likely led to tagger errors. Hypotheses were formulated to account for the impact of learner errors on the tagger's performance. The CLAWS7 tagger misanalysed only 2.8% of forms representing learners' language errors but assigned wrong tags to every fifth spelling error (22%), which means that spelling errors appear to be considerably more problematic for the tagger. Therefore, fully correcting learner spelling errors would increase the tagging accuracy of the CLAWS7 automatic POS-tagging system.

The treatment of idioms and conversational routines in *Woordcombinaties*

Carole Tiberius, Lut Colman (Instituut voor de Nederlandse Taal, Netherlands)

Woordcombinaties (Word Combinations) is a relatively new online lexicographic resource for advanced learners of Dutch as a second or foreign language combining access to collocations, idioms, conversational routines and constructions in one tool.

For collocations, the example of Sketch Engine for Language Learning ([SkeLL](#)) is followed, whereas the constructions are encoded according to Patrick Hanks' Corpus Pattern Analysis (CPA) technique (Hanks 2004, 2013). In *Woordcombinaties* this technique has been tailored



to the needs of the target audience. Sentences are sorted before annotation, using a specially developed **GDEX** configuration to enable the output of short, comprehensible and yet informative sentences. Furthermore, arguments are represented by indefinite pronouns *iemand* 'someone', *iets* 'something', *ergens* 'somewhere', *zo* 'such' (or combinations thereof) in the pattern slots where this is possible for increased readability and the corresponding semantic types and syntactic functions are accessible through a tool tip. The slots are also enriched with typical realisations offering a kind of advanced word sketch in the constructions.

Idioms and conversational routines are also included in *Woordcombinaties*. Currently they are encoded as special instances among the collocations and the patterns. However, separate access with specific search options for idioms and conversational routines is planned and currently being designed. For instance, it will be possible to search for idioms based on image categories, such as 'body parts' and 'food' for *een vinger in de pap hebben* 'have a finger in the pie' and less specific sense categories, such as 'have a property'. Conversational routines will be linked to speech acts, such as 'greeting' or 'apologizing'.

In this presentation, we will look at the implications of treating idioms and conversational routines at the macrostructural level and we will discuss how consistency can be ensured in the editing environment.

References

- Hanks, P. (2004). Corpus pattern analysis. In *Proceedings of the 11th EURALEX International Congress*, pp. 87-98.
- Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. Cambridge, MA: The MIT Press.

Eesti kirjakeele kihiline korpus 1800-1940

Peeter Tinit (Tartu Ülikool, Eesti. Eesti Rahvusraamatukogu, Eesti)

Ettekandes esitlen mineviku eesti keele kogukonna uurimiseks loodud keelekorpus. Korpus koondab kokku ajavahemikus 1800-1940 digiteeritud vabalt kättesaadavaid tekste ja nende tekstide ja teksti autorite kohta koondatud metainformatsiooni. Korpuse eesmärk on võimaldada uuringuid ajaloolises sotsiolingvistikas, kus on fookuses keeleline variatiivsus eri ühiskonnakihtides mineviku kogukondades.

Korpus on loodud koondades digiteeritud tekste avalikest allikatest ja digiarhiividest (nt Rahvusraamatukogu, Kirjandusmuuseum, Tartu Ülikool, Vikitekstid). Metainformatsiooni teoste kohta on kogutud Eesti rahvusbibliograafiast ja sellega ühendatud rahvusvahelistest



andmebaasidest (VIAF, Wikidata, Saksa rahvusbibliograafia) ning isikuloolistest andmebaasidest ISIK ja VEPER. Metaandmeid on analüüsi hõlbustamiseks puhastatud, harmoniseeritud ja struktureeritud.

Korpus sisaldab, kus võimalik, andmeid tekstide ilmumise kohta (nt kirjastus, ilmumiskoht, tiraaž) ja tekstide autorite kohta (nt sünniaasta, sünnikoht, kodumurre, haridustee). Korpusesse on hetkel kaasatud 4412 trükist 1096 erinevalt autorilt, mis hõlmab 11.2% sel perioodil Eesti rahvusbibliograafias registreeritud teostest.

Ettekandes esitlen korpuse sisu ja kolme näidisjuhtu, kus korpusest võib uurimistöös kasu olla: 1) pikaajaliste trendide analüüsiks keeles, kasutades ka kirjutajate taustainformatsioon; 2) autorite keelekasutuse muutumist eluea jooksul, sidudes neid väliste muutustega nagu näiteks avaldatud õigekeelsussoovitused; 3) keelemuutuste mehhanismide uurimine autorite taustainformatsiooni kaudu, näiteks tuues esile põlvkondlikke kihistusi keelekasutuses.

Toetudes suurele tekstikogule ja tekstidega seotud metaandmetele on võimalik detailsemalt uurida toonast keelekogukonda ning selle rolli eesti keele kujunemises. Korpus on avaldatud veebis, et võimaldada selle uuendamist metaandmete täiendamisel ja uute tekstide digiteerimisel. Korpuse edasisel arendamisel on teretulnud nõu ja jõud, mille abil luua tulevikus juba paremad vahendid möödunud aegade eesti kirjakeele mitmekesisuse uurimiseks.

Korpus on kättesaadav aadressil <https://osf.io/zbup2/>.

TNR 12, 1.5, rööpjoendus: akadeemiliste tekstide kirjutamise juhendite tavade murdmine

Ilona Tragel, Helena Lemendik, Liisa-Maria Komissarov, Eleriin Miilman, Külli Habicht, Nele Novek, Nele Karolin Teiva, Betti Marie Peterson, Reena Roos (Tartu Ülikool, Eesti)

Ehkki kirjutamise protsessi võib käsitleda erinevalt, on enim levinud jaotus, mis eristab kavandamist (lähteülesande mõtestamine, ideede kogumine ja struktureerimine, allikatega töötamine), mustandite kirjutamist (esialgsete variantide kirjutamine ning nende tagasiside põhjal täiendamine) ja viimistlemist (stiil, õigekeelsus, vormistus) (vt nt Prior, Thorne 2014). Ehkki mõtteviis kirjutamisest kui mitmeosalisest protsessist ulatub juba 1970. aastatesse (Murray 1972), ei kajastu see enamasti Eestis ei gümnaasiumi ega kõrgkooli praktikas.

Analüüsisime 60 praegu eesti gümnaasiumites ja ülikoolides kasutatava akadeemilise kirjutamise abimaterjali struktuuri ning nägime, et juhendid on orienteeritud peamiselt vaid protsessi viimasele etapile, vormistusele: neis on esile toodud teksti tehnilised parameetrid



(kirjatüüp, kirja suurus, reavahe), töö kohustuslikud osad, viitamisnõuded jne. Osa juhendeid sisaldab lisaks mõne lause ka töö keele ja stiili kohta.

See tendents on kooskõlas uuringutega, mis näitavad, et sageli puudutab kirjutamise õpetamine ja kirjutistele antav tagasiside pigem teksti pindmisi kihte - õigekeelsust, stiili, vormistust -, sest neid on sisuliste aspektidega võrreldes lihtsam märgata ja hinnata (Brannon, Knoblauch 1982). Selline pealiskaudsus tähendab aga seda, et õpilane või üliõpilane, kellel seisab ees mõni kirjalik töö, ei pruugi töö kirjutamiseks saada piisavalt sisulist tuge (Kruse 2003). See omakorda viib olukorrani, et (üli)õpilane ei mõista kirjutamisprotsessi olemust ning ei oska oma tööd planeerida, mistõttu tundub kirjutamine vaevalise ja ebameeldivana ning tulemus ei pruugi õnnestuda.

Seetõttu koostame Tartu ülikooli eesti ja üldkeeleteaduse instituudi projekti "Protsessikeskne tekstiloome gümnaasiumis ja ülikoolis: tõenduspõhine teadmus ja praktikad" raames uut tüüpi üliõpilastööde teavikut, mis kataks kogu kirjutamisprotsessi kavandamisest viimistluseni. Lähtume teavikus ideest, et kirjutamine on arendatav oskus, mis algab sellest, et mõista erinevaid komponente, millest kirjutamine tegevusena koosneb. Ettekandes esitame esiteks läbi vaadatud juhendite analüüsi ning seejärel kirjeldame teaviku loomise põhimõtteid ning protsessi.

Kirjandus

- Brannon, Lil; Knoblauch, C.H. 1982. On Students' Rights to Their Own Texts: A Model of Teacher Response – *College Composition and Communication*, Vol. 33, No. 2, 157–166.
- Kruse, Otto 2003. *Getting Started: Academic Writing in the First Year of a University Education*. – Björk, L., Bräuer, G., Rienecker, L., Jörgensen, P.S. (Eds.). *Teaching Academic Writing in European Higher Education. Studies in Writing*, vol 12. Springer, Dordrecht. https://doi.org/10.1007/0-306-48195-2_2
- Murray, Donald M. 1972. Teach Writing as a Process Not Product. – *The Leaflett* 71(3), 11–14.
- Prior, P.; Thorne, S. L. 2014. 3 Research paradigms: Beyond product, process, and social activity. – J. Eva-Maria; D. Perrin (Eds.). *Handbook of Writing and Text Production, The Mouton de Gruyter Handbooks of Applied Linguistics Series, Volume 10*. Berlin, Boston: De Gruyter Mouton, 31–54.



Kujundlik mõtlemine 2020—2022 keelekriisis

Ene Vainik¹, Geda Paulsen^{1,2} (¹Eesti Keele Instituut, Eesti. ²Uppsala Ülikool, Rootsi)

Keelekriisi nime all võtame kokku Eesti meedias aastatel 2020—2022 aset leidnud debati keelelise vabaduse, keelenormide, keelekorralduse ning EKI sõnastikureformi üle. Suurte kriiside (nagu sõda, pandeemia või globaalne kliimasoojenemine) taustal mõjub kõnealune kriis tormina veeklaasis. Ometi on ühiskonna üldisem polariseerumine ja vastandumistung mõjutamas ka seda valdkonda.

Usume, et nn keelekriisi küllaltki kompaktses diskursuses (62 teksti) analüüs sellest seisukohast, kuidas kombineeruvad omavahel eri autoripositsioonide väärtused, emotsioonid ning kujundlik eneseväljendus, võiks olla õpetlik selles suhtes, kuidas näeb välja kriisikommunikatsioon üldisemalt ja kas ning kuidas panustab diskursuse pingestumisse kujundlik keelekasutus. Loodetavasti on tulemustest midagi õppida ka paremaks kriisikommunikatsiooni juhtimiseks tulevikus.

Uurimuse lähemas fookuses on kujundid, mida kasutatakse keele enda kohta. Nimelt on keel oma erinevates avaldumisvormides sedavõrd mitmetahuline ja abstraktne nähtus, et sellest on pea võimatu üldse rääkida või mõelda, tuginemata konkreetsematele ja argikogemusest tuttavatele valdkondadele. Mõelgem kasvõi tavalistele keeleterminitele, mis kasutavad lähtevaldkonnana loodust (sõnatüved, juursõnad) või kaubavahetust (keele omandamine, laensõnad). Kujundlikkus on äratuntavam, kui räägitakse elavast keelest, keele elujõust või keele heaolust või sellest, kas keel on vaba või vangis jne. Kujundliku mõtlemise alusvaldkonnad võivad reeta midagi selle kohta, kuidas keelega suhestutakse—kas ollakse selle uurijad, tõlkijad või keelekasutuse korrigeerijad, ning samuti midagi selle kohta, milliseid alusväärtusi peetakse esmatähtsateks ja kuidas püütakse lugejat mõjutada. Samuti avaldub kujundlikkuses autorite loovus ning emotsionaalne seotus teemaga.

Metodoloogiliselt toetume kriitilisele diskursuseanalüüsile; kujundite analüüsile aga kognitiivsele metafooriteooriale ning selle edasiarendusele, mis käsitleb tahtlikke metafoore ning diskursuse raamistamist.

Kirjandus

- Burgers C., Konijn E.A., Steen G.J. 2016. Figurative Framing: Shaping Public Discourse Through Metaphor, Hyperbole, and Irony, *Communication Theory*, Volume 26, Issue 4, 410–430, <https://doi.org/10.1111/comt.12096>



- Fairclough, N. 2010. Critical Discourse Analysis: The Critical Study of Language (2nd ed.). Routledge. <https://doi.org/10.4324/9781315834368>
- Halliday, M.A.K. 2003. Introduction: On the "architecture" of human language. In On Language and Linguistics. Volume 3 in the Collected Works of M.A.K. Halliday. Edited by Jonathan Webster. London and New York: Continuum.
- Kövecses, Z. 2018. Metaphor in media language and cognition: A perspective from conceptual metaphor theory. In: Lege artis. Language yesterday, today, tomorrow. The journal of University of SS Cyril and Methodius in Trnava. Warsaw: De Gruyter Open, 2018, III (1), June 2018, p. 124-141. DOI: 10.2478/lart2018-0004 ISSN 2453-8035
- Lakoff, G.; Johnson, M. 1980. Metaphors we live by. Chicago: The University of Chicago Press.
- Steen, G. 2015. Developing, testing and interpreting deliberate metaphor theory. Journal of pragmatics, 90, 67-72.

Vikipeedia kasutamine isiklikel eesmärkidel ja õppetöös: õpetajate kogemused

Maigi Vija (Tartu Ülikool, Eesti)

Tänapäevase õpikäsituse põhimõtete järgi on õppimine sotsiokultuuriline nähtus (Lankshear & Knobel, 2011) ja õpikogemus peaks olema õppija jaoks tähenduslik (Willinsky, 2017; Kutsestandard, 2020). Tänu tehnoloogilistele vahenditele on teabele ligipääs võimalik igal hetkel ja mis tahes kohas. Info- ja allikarikas (veebi)maailmas kasutatakse teabe hankimisel esmase allikana sagedasti just Vikipeediat, mis on koostöise veebikeskkonnana kõigile kättesaadav ja millest on saadaval palju keeleversioone. Õpetaja digipädevusmudelis (s.a.) on välja toodud õpetaja digikompetentsid ning õpilaste digipädevuse kujundamise aspektid: digivahendite ja -keskkondade kasutamine, koostöö tegemine ja digisisuloome eri keskkondades. Info otsimine ja selle analüüsimine on info tarbimise näide, kuid digipädevuse arendamise raamistikus on oluline ka sisuloome ehk ise digitaalse sisu loomisse panustamine, seda nii õpetajate kui ka õpilaste tegevuses.

Ettekandes analüüsitakse Vikipeedia keskkonna kasutamist nii isiklikuks tarbeks kui ka õppetöö jaoks Eesti üldhariduskoolide õpetajatega tehtud intervjuude põhjal. Uurimismaterjali moodustavad 11 fookusrühma intervjuud 49 õpetajaga. Intervjuudest saadud tulemused annavad ülevaate Eesti üldhariduskoolide õpetajate Vikipeedia kasutamise kogemusest isiklikus elus ehk õppetööväliselt ning tööelus ehk õppetöös.



Kirjandus

- Kutsestandard (2020). Õpetaja, tase 7. Sihtasutus Kutsekoda. <https://www.kutseregister.ee/ctrl/et/Standardid/vaata/10824233>
- Lankshear, C., & Knobel, M. (2011). *New Literacies: Everyday Practices and Social Learning*. Maidenhead: McGraw-Hill/Open University Press.
- Willinsky, J. (2017). *The new literacy: Redefining reading and writing in the schools*. Routledge.
- Õpetaja digipädevusmudel (s.a.). <https://digipadevus.ee/opetaja-digipadevusmudel/>



Mitmekeelsuse töötuba / Workshop on Multilingualism and Multilingual Acquisition

Building a language: applying the traceback method to a 2-year-old Estonian-English bilingual's data

Piret Baird (Tallinn University, Estonia)

Recent years have seen an increase in the application of usage-based theory in bilingual child language acquisition research. Usage-based theory postulates that children's language acquisition is item-based and formulaic (Tomasello, 2003). In the light of usage-based theory, several studies have found that caregiver speech to be repetitive, there are input-output connections between mother's and their children's speech and that children's utterances are composed of recurring patterns (Cameron-Faulkner et al., 2003; Lieven et al., 2009). In order to investigate patterns in children's speech, the traceback method has been developed. The method helps to identify recurrent patterns in monolingual as well as multilingual data.

The current study investigates the spontaneous speech (45h) of a 2-year-old Estonian-English simultaneous bilingual (2;3-2;11). Using a computational implementation of the traceback method (similarly to Quick & Hartmann, 2021) it is shown that about one third of the child's code-mixed utterances can be traced back to her own previous language use as well as to the input of her caregivers'. The results also indicate that the child relies on fixed chunks and partially schematic units in her speech. Entrenchment and input effects are discussed.

References

- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27(6), 6. https://doi.org/10.1207/s15516709cog2706_2
- Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20(3), 3. <https://doi.org/10.1515/COGL.2009.022>
- Quick, A. E., & Hartmann, S. (2021). The Building Blocks of Child Bilingual Code-Mixing: A Cross-Corpus Traceback Approach. *Frontiers in Psychology*, 12, 682838. <https://doi.org/10.3389/fpsyg.2021.682838>
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.



Narrative Macrostructure in Lithuanian Monolingual and Bilingual Typically Developing 6-Year-Olds

Ingrida Balčiūnienė, Ineta Dabašinskienė, Laura Kamandulytė-Merfeldienė (Vytautas Magnus University, Lithuania)

Narrative macrostructure has been widely investigated in various populations (children and adults; typically- and atypically-developing; monolingual and bilingual subjects) and languages. However, majority of the studies has focused on narrative macrostructure as a pure structural variable and, thus, its relations to linguistic competence and skills remain unclear. From one hand, following psycholinguistic serial models of speech production (Bock & Levelt, 1994), a speaker first builds a macrostructural frame and then applies linguistic competence for selecting proper syntactic constructions, morphological rules, and lexical items. On the other hand, a speaker with limited linguistic competence may be prevented from building well-developed macrostructure.

In our study, we aimed to analyze and compare narrative macrostructure in two populations of Lithuanian 6-year-olds. The sample consisted of a group of typically-developing bilingual children speaking Lithuanian as L1 and English as L2 (N = 12) and a group of typically-developing monolingual Lithuanian peers (N = 12). The bilingual children lived in the UK and, in addition to school in English, attended a Lithuanian language tutor centre; the monolingual children lived in Lithuania and attended a state kindergarten. During individual session, each child was asked to tell a story according to a picture sequence consisting of six wordless pictures; the stories were audio-recorded and transcribed. To carry out a macrostructural analysis, we employed the NAP methodology (Peterson & McCabe, 1983). This methodology comprises six variables of discourse coherence, such as topic maintenance, event sequencing, informativeness, referencing, conjunctive cohesion, and fluency. During the analysis, we built individual narrative profiles according to the given variables and compared them within and between the groups.

The study evidenced, that topic maintenance, event sequencing, and informativeness were quite similar between the groups despite their different linguistic background and educational experience. The other variables (referencing, conjunctive cohesion, and fluency) were developed better in the monolingual group. Since the latter group of variables is closely related with the linguistic competence, we may speculate that incomplete language acquisition in bilinguals (Balčiūnienė & Dabašinskienė, 2019) prevent them from building appropriate macrostructural frame.



References

- Balčiūnienė, I., & Dabašinskienė, I. (2019). Language dominance in bilingual acquisition: A case study of narrative production in Lithuanian. *Eesti Rakenduslingvistika Uingu Aastaraamat*, 15, 5–19.
- Bock, K., & Levelt, W. J. M. (1994). Language production. Grammatical encoding. In M.A. Gernsbacher (Ed.). *Handbook of Psycholinguistics* (pp.741–779). New York: Academic Press.
- McCabe, A., & Bliss, L. S. (2003). *Patterns of Narrative Discourse. A Multicultural, Lifespan Approach*. Boston: Pearson Education, Inc.

Local Russian in Kalmykia and mixed Kalmyk

Vlada Baranova (The Helsinki Collegium for Advanced Studies, Finland)

Although many studies on language contact have focused on the transfer of linguistic features from a dominant language to a minority one, mutual influence between languages has received comparatively little attention. However, it's important to note that stable and prolonged multilingualism can affect both languages involved, not just in convergent areas or mixed language situations. Code-switching and code-mixing in a bilingual context can lead to a relatively stable system that Auer (1999) called 'fused lect' and the process of conventionalization is language 'fusion' (Auer and Hakimov 2021).

This paper aims to describe the continuum of bilingual Kalmyk-Russian speech and, in particular, to define some mixed varieties among different generations of Kalmyk. The main grammatical interference from Kalmyk to local Russian is the way of expressing the evidentiality which is frequent in local Kalmyk Russian analogically an obligatory distinction between direct and indirect sources of knowledge in Kalmyk. There are also some other features like reduplication and the word order with constructions with proper names. In Kalmyk, there are also many borrowings from Russian including the fundamental transformation of the morphological system and the syntactic structure. All cases of contact-induced features are, however, the developments of internal devices or markers, so they can be considered both as a contact-induced change and independent grammaticalization.

Data show that the result of long period of language contact in a stable bilingual community is the variety of mixing codes or contact-induced varieties. The language ideologies impact the process of language fusion: the members of the community consider both languages as a part of their heritage and such attitudes oppose the ideas about values of 'pure language' that are also held by many within the Kalmyk ethnic community.



References

- Auer, Peter. 1999. From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism* 3(4). 309–332.
- Auer, Peter, Hakimov, Nikolay. 2021. From language mixing to fused lects: The process and its outcomes. *International Journal of Bilingualism* 25 (2). 361–368.

The production of narratives in Lithuanian-speaking monolingual and bilingual typically-developing children and children with a language disorder

Ineta Dabašinskienė, Laura Kamandulytė-Merfeldienė (Vytautas Magnus University, Lithuania)

Although the importance of narrative comprehension and production tasks is highlighted by a large number of scholars (e.g., Bohnacker & Gagarina, 2020), narrative studies in Lithuania can still be characterized as very scarce. Oral narrative considered to be essential at the transition from preschool to school age. Children with sufficient skills in story (re-)telling at the preschool age demonstrate rather easy reading acquisition at school; this, consequently, lead to high academic skills and social life. In contrast, low skills in oral narrative production (due to incomplete language acquisition caused by different reasons, such as developmental language disorder or bilingualism / multilingualism), at the preschool age may prevent from acquiring written language at the sufficient level at school, and, thus may cause learning difficulties (Babayigit et al., 2021). Our study aims to examine the general linguistic performance (microstructure) in Lithuanian using the narrative elicitation (telling) procedure in a group of Lithuanian-speaking 5-7 years old subjects (N=360). In our study, we aimed at comparative analysis between 3 groups: 1) a group of typically-developing (TD) children, 2) a group of children with developmental language disorder (DLD), 3) a bilingual group. The study investigated a number of linguistic features such as general productivity, lexical diversity, and syntactic complexity. The results demonstrated that the mean length of utterance, the total number of words, the type/token ratio and the syntactic complexity increased with the age in the TD group. The analysis of the microstructure displayed differences between the 3 groups (TD children, children with DLD, and bilingual children) regarding general productivity, lexical diversity, and syntactic complexity. The possible causes will be discussed further in the presentation.



References

- Babayigit, S., Roulstone S., & Wren, Y. (2021). Linguistic comprehension and narrative skills predict reading ability: A 9-year longitudinal study. *British Journal of Educational Psychology*, 94, 148–168.
- Bohnacker, U., & Gagarina, N. (2020). Introduction to MAIN–Revised, how to use the instrument and adapt it to further languages. *ZAS Papers in Linguistics*, 64, xiii–xxi.

Linguistic and pragmatic choices in different sociolinguistic contexts: a corpus-based analysis of Russian spoken discourse.

Ilenia Del Popolo Marchitto (Tallinn University School of Humanities, Estonia)

Communication is regulated by cultural-specific principles and norms that are expected in conversations not only to successfully achieve a certain communicative purpose, but also to reduce social friction and avoid offence. This means that, at least at a prescriptive level, interlocutors should employ a given set of linguistic means in the attempt of being, at the same time, truthful, informative, relevant, clear and polite.

Nevertheless, language in use is far from being a predictable convention, which makes it crucial to analyse spontaneous speech in different natural communicative situations.

This contribution aims to propose a corpus-based analysis of Russian speakers' linguistic choices when performing the speech act of request, known as one of the most complex “face-threatening acts” because of its high estimated risk of face loss. Given that interlocutors, consciously or not, employ certain grammatical forms and lexical items in accordance with personal values, pragmalinguistic and sociocultural factors, the goal is to analyse their spontaneous linguistic behaviour through a qualitative usage-based approach. In particular, the paper deals with the use of indirectness, politeness strategies, mitigation markers and hedges from a pragmatic viewpoint.

The study will be carried out by examining the written transcriptions of the corpus of Russian everyday speech *Odin rečevoj den'* (ORD), which was created by the Faculty of Philology of Saint Petersburg State University by means of the continuous 24-hour recording method. This corpus constitutes a relevant source of linguistic material, because it fixes authentic Russian spoken discourse in various naturally-occurring situations from a representative sample of the targeted population. Furthermore, it provides detailed information about the sociolinguistic context of each communicative episode and the informants' social, professional and educational backgrounds. This allows a more precise pragmatic investigation, which takes into



account sociolinguistic factors that are often overlooked in speech act research, such as social roles, socioeconomic status, horizontal and vertical distance and power relations.

Multilingual Tallinn: a pilot study on soundscape and linguistic landscape

Kapitolina Fedorova, Natalia Tshuikina (Tallinn University, Estonia)

Multilingualism in Post-Soviet countries where as a result of the Soviet time migration and Russification, Russian speakers still constitute a significant part of population, presents an interesting case for analysis. Estonia, being officially monolingual country with quite strict language policy, is extremely linguistically diverse in terms of actual everyday communication, with both Russian and English playing a significant role. Moreover, as Estonia attracts many foreign visitors, expats and refugees, other languages, including Ukrainian, also become more and more noticeable in both visual and acoustic form in public spaces. However, spatial distribution of different languages in urban landscapes vary significantly, and by analysing it one can reveal not only the patterns of language use but also power relations between different linguistic, ethnic, and social groups. The paper reports on preliminary results of a study conducted in Tallinn in 2021-2022 and focused on collecting and comparing data on both linguistic landscapes and soundscapes in several locations in the city (two shopping centers in different districts, concert halls, theatres and cinemas, and at and around Tallinn University's premises). The study shows a mismatch between visual linguistic representation and actual linguistic diversity and suggests methodological procedures that can be employed in further researches on multilingualism and language policy.

A constructionist approach to language attrition: The case of dative loss in Blumenau German

Stefan Hartmann¹, Antje Quick^{2,3}, Nikolas Koch⁴, Claudia Maria Riehl⁴ (1University of Düsseldorf, Germany. 2University of Leipzig, Germany. 3FAU Erlangen-Nürnberg, Germany. 4Ludwig-Maximilians-Universität Munich, Germany)

Minority language enclaves are an ideal “laboratory” for studying the dynamics of language taking social, cultural, and cognitive factors into account. In particular, the phenomenon of language attrition – both in individuals and in entire communities – has been studied extensively including data from German language islands abroad (see e.g. Riehl 2019, Boas 2009). The loss of case marking is a particularly salient example of “simplification” processes often found in island varieties. In this paper, we draw on data collected by Riehl (e.g. 2018) in Blumenau, Brazil, across three generations of heritage speakers of German focusing



specifically on dative case marking. Previous research has already shown that dative case marking becomes less frequent in subsequent generations, e.g. *wir haben Ferien in die Schule* ‘we have holidays at school-NOM/ACC’ instead of *in der Schule* ‘at school-DAT’. But which factors lead to the loss or maintenance of dative case marking, and what are the commonalities and differences between patterns of dative marking across individual speakers? This paper addresses these questions based on a systematic quantitative analysis of the Blumenau data, taking the token frequency and inflection class of individual lexemes into account while also considering individual variation. A first analysis of the data using CART Trees and Random Forests (see e.g. Levshina 2020) indicates that the generational effects are negligible, while other factors, such as the morphological features of the lexeme, play a much larger role. A qualitative follow-up analysis of the data suggests that the phenomena of loss and maintenance of case marking that can be observed in the data can partly be explained via frequency and chunking effects: For instance, the frequent use of so-called mixed prepositions like *in* (which can take both accusative and dative, depending on the context) with accusative complements, along with the fact that there is already a considerable amount of case syncretism between accusative and dative in German, gives rise to patterns like [in der/die/das N] that are used irrespective of whether the accusative or the dative would be used in this context in standard German. In cliticized preposition-article combinations like *beim* (< *bei dem*), by contrast, we often see that the dative is retained, which points to the conserving effect of chunking. We argue that an account of case loss in the framework of Construction Grammar (e.g. Hoffmann 2022) can prove highly insightful for explaining the complex phenomenon of language attrition.

Language alternation in a multilingual community: How Roma in Estonian use their rich language repertoire

Anette Ross (University of Helsinki, Finland)

Estonia is a linguistically diverse area with many Estonian, Russian or English bilingual and multilingual speakers. In such a linguistic environment, the members of the Roma minority of around 1000-1500 people form their language repertoires of the local languages and their own language Romani. This presentation explores the complexity of multilingual speech in a Roma community, where multilingualism and simultaneous use of several languages have a long tradition.

Individuals with knowledge of several languages create a reality, where more than 2-3 languages are spoken. Romani speakers in Estonia are multilingual and thus are prone to use



other languages as linguistic resources for Romani and alternate between languages. The research provides examples from recorded narratives and spontaneous conversations between the members of the Roma community. While the narratives are dominantly in Romani with code-switches to Russian and Estonian, the language alternation in conversations is frequent between Romani, Russian and Estonian, and occasional switches to Latvian, English, Finnish and other languages take place.

Use of several languages in one conversation can be described as code-switching between codes or languages that have clear boundaries. The use of Romani that is intensively influenced by contact languages complicates defining these codes or languages. The study shows a variety of ways languages are alternated in different settings and between speakers with various individual language repertoires and questions the ways in which language alternation in such a multilingual setting is described. The aim is to reach a description that would show the versatility of languages and their alternation patterns in the most accurate way.

Better understanding of a multilingual communities' and multilinguals' speech could help to overcome the ignorance towards less common language use strategies of multilingual individuals and communities and encourage inclusion of divergent speech communities in formal education. In Estonia, minority language speakers and multilingual speakers do not gain much attention in formal education and this presentation intends to show language use of the Roma community as an example for providing better understanding of the actual language use.

Egyptian Arabic-English Code-Switching in Egyptian female vlogs

Natalia Siereda (Tallinn University, Estonia)

Code-Switching as a result of language contact has been studied since a few decades already, however applying a usage-based method in this context is still a novelty. In this qualitative study concerning code-switching between Egyptian Arabic and English in Egyptian female vlogs, the video content from five different lifestyle bloggers was analysed with the help of usage-based method, combined with the typology of code-switching proposed by Muysken. The material collected for this study has been published by the authors on YouTube in 2022.

Vlogging, seen as a form of semi-performance, contains elements from spontaneous speech but most probably also carefully thought out phrases. Because of earlier preparation, the



language choices may be not as natural and spontaneous, as they would be in the case of an informal conversation. Arabic and English being two typologically different languages, their merging may possibly result in new forms being a mix of stem from L1 and ending from L2. The study focuses specifically on Egyptian dialect of Arabic, a dialect that evolves very quickly and has a history of Ottoman and English influences which have affected its contemporary shape. Therefore, the question of whether a word is an existing in the dialect borrowing or code-switching is not always to be answered easily here.

How multilingualism research can contribute to translation studies

Anna Verschik, Maris Saagpakk (Tallinn University, Estonia)

Multilingualism is a term that literary and translation studies share with many other disciplines such as linguistics, sociology, language teaching and learning. However, multilingualism is par excellence the central topic in contact linguistics. Yet, it appears that the disciplines are not concerned with each other and use different metalanguage. Recently, some scholars in contact linguistics started addressing multilingualism in writing and in fiction in particular but without any reference to translation studies. However, multilingualism in literature also expands and tests language borders, plays with language and its symbolic value. It uses linguistic encounters as a fictional tool to create additional layers of meaning, that may be based on reality or not. The research in translation of literary multilingualism appears to address the desire of the academic community to challenge the national and monolingual paradigm that has been prevalent in literary history. The need to refrain from interpreting multilingualism merely as portraying situations of language contacts in real life (Grutman 2006) might have led to the disregard of the scholarly work in contact linguistics, a problem that is targeted in this paper. To date, only few studies in contact linguistics address translation of multilingualism in fiction, concentrating especially on code-switching (CS). To avoid inventing new labels for well-known phenomena, we strongly suggest that contact linguistics, multilingualism research and, partly, sociolinguistics provide necessary terms for the analysis of literary multilingualism and its translation, which can lead to a strengthened interdisciplinary understanding of multilingualism. In this contribution we expand and combine contact linguistics terminology (lexical impact: CS/borrowing; non-lexical impact: loan translation and structural change) with terms concerning metalinguistic information in literary texts and with the translation model proposed by Delabastita (1993). We are going to show that the study of literary multilingualism and translation of multilingualism could gain insights from sociolinguistics and contact linguistics. Both have developed meta-language and instruments for description and classification of the effects of multilingualism. In order to demonstrate the



synergetic effects of this kind of interdisciplinary work, the paper will refer to the case of translation of Baltic German literature into Estonian.

References

- Delabastita, Dirk 1993. *There's a Double Tongue: An Investigation Into the Translation of Shakespeare's Wordplay, with Special Reference to Hamlet*. Amsterdam, Atlanta: Rodopi.
- Grutman, Rainier. 2006. "Refraction and Recognition." *Target. International Journal of Translation Studies* 18, no. 1: 17–47. <https://doi.org/10.1075/target.18.1.03gru>.



Töötuba. Registrid keeles: uurimine ja rakendused

"Küllap kirjanik märkas, et hakkas tüdinema". (Inter)subjektiivsuse markeerimine ilukirjandustekstides

Külli Habicht, Külli Prillop (Tartu Ülikool, Eesti)

Kuidas paigutub ilukirjandus kui tekstiliik (inter)subjektiivsete markerite kasutuse poolest võrdluses teiste tekstiliikidega? Mis on iseloomulik ilukirjandustekstide (inter)subjektiivsuse markeritele? Missuguseid funktsioone täidavad ilukirjandustekstides esilduvad (inter)subjektiivsed partiklid? Nendele küsimustele otsime oma ettekandes vastust. Lähtekohaks on (inter)subjektiivsuse markerite võrdlev korpuspragmaatiline ja tekstiliigiline analüüs (Biber, Conrad 2009; Aijmer 2013).

Uurimus on seotud projektiga "Pragmaatika grammatika kohal: subjektiivsus ja intersubjektiivsus eesti keele registrites ja tekstiliikides (PRG 341)", milles analüüsitakse (inter)subjektiivsuse markereid suulise, neti- ja trükiregistri tekstiliikides. Ettekande alusmaterjal pärineb TÜ eesti ja üldkeeleteaduse instituudis loodud tasakaalus „Pragmaatika“ korpusest, mis annab ülevaate 11 tekstiliigist (Prillop jt 2021).

Ilukirjandustekstide materjal moodustab selles korpuses 500 312-sõnelise alamkorpuse. Ettekandes keskendume (inter)subjektiivsuse markerite kasutuse eripäradele ilukirjandustekstides. Ilukirjandus kui sünteesiv tekstiliik koondab endas mitmekesisest keeleainest. Mitmekesisus tuleneb esiteks ilukirjanduse žanrilisest varieeruvusest, teiseks eri allkeelte kombineerumisest selles ning kolmandaks autori-, tegelas- ja siirdkõne esinemisest proosatekstides. Lisanduvad autoristiiliga seostatavad erijooned. (Hennoste 2023) Nii võiks eeldada, et ilukirjandustekstid paistavad teiste tekstiliikide hulgas silma suurema hulga tekstiliigile spetsiifiliste partiklitega.

Analüüsist selgub, et ilukirjandustekstide (inter)subjektiivsuse markerite kasutus ei ole silmatorkavalt rikkalik: võrreldes suulise keele ja netikeele registritega on seal partikleid vähem; võrreldes teadustekstidega siiski rohkem (Hennoste jt 2023). Sagedusvõrdluse põhjal võib ilukirjandustekstidele eriomaseks pidada erineva funktsiooniga üksusi: suurt tõenäosusastet väljendavat oletavat partiklit küllap (1), suurt tõenäosusastet väljendavat rõhutavat partiklit muidugi, vastanduvat väidet pehmendavat partiklit öieti ja väikest tõenäosust väljendavat oletavat partiklit vahest (2).



(1) „Kas sellest pildist annab koopia teha?“ „Küllap annab. Kui mul nooremad koju käima tulevad, küll need korraldavad, kui nimetan.“

(2) Vahest ma näin naisena, kuid tegelikult olen olemuselt mees.

Ettekandes toome näiteid ilukirjandustekstides esilduvate partiklite funktsioonide kohta ning põhjendame nende kasutuse tagamaid. Võrdleme „Pragmaatika“ korpuse andmeid ka eesti keele ühendkorpuse (2021) ilukirjandustekstidest saadud tulemustega. Analüüsis lähtume kahest tasandist: ühel eristame ilukirjanduse autori- ja tegelaskõnes kasutatud pragmaatilisi üksusi, teisel teeme kokkuvõtte tekstiliigi kui terviku kohta, mille partiklikasutuse eripära tuleneb tekstiliigile omasest heterogeensusest.

Kirjandus

- Aijmer, Karin 2013. Understanding Pragmatic Markers: A Variational Pragmatic Approach. Edinburgh: Edinburgh University Press.
- Biber, Douglas, Susan Conrad 2009. Register, Genre, and Style. (= Cambridge Textbooks in Linguistics.) Cambridge: Cambridge University Press.
- Hennoste, Tiit 2023. Kirjanduskeel. Käsikiri. Tartu.
- Hennoste, Tiit, Külli Prillop, Külli Habicht, Helle Metslang, Kirsi Laanesoo, Liina Pärismaa, Elen Pärt, Andra Rumm, Andriela Rääbis, Carl Eric Simmul (ilmumas), Pragmatic particles, roles, and registers: the case of Estonian. Open Linguistics.
- Prillop, Külli, Tiit Hennoste, Külli Habicht, Helle Metslang 2021. Ei saa me läbi „Pragmaatika“ korpusega. Korpuspragmaatika ja pragmaatikakorpus. – Mäetagused 81, 161–176.

Üld- ja oskuskeele vahekorras kriiside keerises

Reet Hendrikson (Kaitseväe Akadeemia, Eesti)

Keel on muutuste peegel. Üks valdkondi, kus maailma muutumine hästi peegeldub, on ajakirjanduskeel: on ju ajakirjandus esmane žanr, mille kaudu jõuab maailmas toimuv tavakodanikuni. Sel moel saabki ajakirjanduskeelega seoses ehedalt ilmsiks üld- ja oskuskeele vahekorras dilemma. Piiri hägusust üld- ja oskuskeele vahel on esile toodud varemgi, ent julgeoleku- ja tervishoiukriisi taustal omandab piiride hägusus uue mõõtme. Nii on just ajakirjanduse vahendusel jõudnud Eesti inimese teadvusse ja ehk ka kõnepruuki teiste seas sellised keelendid nagu *üksuste siirdamine*, *rahvusvahelised üksused*, *sõjaline võimekus*, *kamikaze-droon*, *hulkurdroon*, *loiterdav moon*, *luusiv lahingumoon*, *kaaberдав laskemoon*, *treeninglend*, *kobarpomm*, samuti *kaitsetahe*, *laiapindne riigikaitse*, *keskmaa-*



õhutõrje jt. Ettekandes eritlengi üld- ja oskuskeelee vahekorda, tuginedes riigikaitsest pärit ainesele. Näidetest ilmneb, et osa ajakirjanduses kasutatavast riigikaitse sõnavarast lahkneb erialasest tavast ja erialakeele arendamisel kokkulepitust. Just selline nn pidžin-militaaria annab (üldkeeles) sõna ja termini vahekorra mõtestamisel tänuväärset ainet. Kes peaks hea seisma selle eest, et ajakirjanduses kasutataks erialakeeles kokkulepituid termineid? Kas ja kuivõrd peaks üldse eeldama, et ajakirjandus- ja üldkeeles järgitaks erialakeele tava ja normingut? Kuivõrd on ärevad ajad mõjutanud riigikaitse oskuskeelee arendamist ning kuidas peegeldub julgeolekupoliitiline olukord riigikaitse terminiarenduses ja -päringutes? Kuidas saab toimetaja, tõlkija, ajakirjanik või kes tahes riigikaitsehuviline teavet valdkonna mõistete-terminite kohta? Ettekanne on mõtteline jätk 2022. aasta detsembris toimunud teaduskeelee konverentsil kõneldule ja 2023. aasta jaanuaris aset leidnud terminipäevaku lõpuarutelule.

Sagedased pragmaatilised partiklid ja markerid registrite eristajana

Tiit Hennoste, Külli Prillop, Helle Metslang, Külli Habicht, Kirsi Laanesoo, Andriela Rääbis, Carl Eric Simmul (Tartu ülikool, Eesti)

Meie ettekande teemaks on eesti keele eriti sagedaste episteemiliste partiklite ja pragmaatiliste markerite kasutus registrite eristajana. Toome välja eriti sagedased partiklid seitsmes registris (suuline argivestlus, suuline ametisuhtlus, netivestlus, netikommentaariid, ilukirjandusproosa, trükiajakirjandus, teadus). Materjal pärineb Pragmaatikakorpusest, mis sisaldab 3,5 miljonit tekstisõna, igas registris 500 000 sõna aastatest 2010-2020 (Prillop jt 2021). Analüüsis kombineerisime kirjeldavat statistikat partiklite kvalitatiivse semantilise ja pragmaatilise analüüsiga.

Analüüs näitas, et eriti sagedased partiklid ja markerid jagunevad kolmeks: pehmendajad vist, lihtsalt, võib-olla, äkki, ilmselt, ehk; intensiivistajad tegelikult, üldse, muidugi, kindlasti, tõesti ning markerid ma ei tea ja ma arvan et (vt ka Hennoste jt 2021, 2022).

Ettekandes uurime, kuidas saab partiklite ja markerite sageduserinevusi kasutada registrite eristamiseks. Esmane analüüs näitab, et vist jagab registrid kolme rühma: argiregistrid, teadus + ajakirjandus, muud. Suuliste registrite omavahel eristamiseks sobivad vist ja ma ei tea. Ajakirjanduse ja teaduse eristamiseks sobib rida partikleid. Kommentaaride ja ilukirjanduse eristamiseks on vaja kasutada erinevate partiklite kombinatsioone, mida vaatame täpsemalt ettekandes.



Kirjandus

- Hennoste, Tiit, Metslang, Helle; Habicht, Külli; Prillop, Külli. Kuue (inter)subjektiivsuspartikli kasutus eesti keele registrites. – Emakeele Seltsi aastaraamat, 66 (1), 2021, lk 91–123.
- Hennoste, Tiit, Prillop, Külli; Habicht, Külli; Metslang, Helle; Laanesoo, Kirsi; Pärismaa, Liina ; Pärt, Elen; Rumm, Andra; Rääbis, Andriela; Simmul, Carl Eric. Komplementausega predikaatidel põhinevate diskursusemarkerite kasutus eri registrites. – Keel ja Kirjandus, 2022, nr 1–2, lk 130–150.
- Prillop, Külli; Hennoste, Tiit; Habicht, Külli; Metslang, Helle. Ei saa me läbi „Pragmaatika“ korpuseta. Korpuspragmaatika ja pragmaatikakorpus. – Mäetagused, 81, 2021, lk 161–176.

Veendumuspartikkel ju retoorilistes küsimustes ja väidetes teadus- ja ajakirjandustekstis.

Marge Käsper (Tartu Ülikool, Eesti)

Lingvistiliselt evidentsiaalsuse ja episteemilise modaalsuse sõlmpunkti asetuv „värvingupartikkel“ *ju* võimaldab eesti keeles kergelt rõhutada ning tugevdada lauses esitatut, vahel ka seetõttu, et öeldut peetakse endastmõistetavaks või tuntuks (EKI Ühendsõnastik 2022). Kui varasemalt on seda partiklit käsitletud pigem suulise kõne või seda kujutava teatriteksti kontekstis (Hennoste 2000, Kärk 2010, 2012), siis argumentatsioonianalüüsi seisukohast pakub selle partikli funktsioon markeerida lausungis ilmselgusele apelleerimist huvitava sisendi ka kirjalike argumenteerivate tekstide uurimiseks – ka akadeemilisest tekstist toob see marker välja iseenesestmõistetavusele apelleerimist (vt Käsper 2017, 2019). Täpsemalt toetun Marie-Anne Paveau (2006) eeldiskursuste teooriale, mis on kirjeldanud paljusid ilmselgusele apelleerimise keelelisi vahendeid. Partikkel *ju* esineb selles funktsioonis nii retoorilistes küsimustes kui eeldustel põhinevates väidetes. Funktsiooni kasutusmustrite täpsemaks uurimiseks kasutan ettekandes Eesti Keele Ühendkorpuse 2021 akadeemilise tekstide ja perioodikaväljaannete alamkorpusi.

Kirjandus

- Dendale, Patrick, et Julie Van Bogaert. 2012. « Réflexions sur les critères de définition et les problèmes d'identification des marqueurs évidentiels en français ». *Langue française* 173 (1): 13–29.
- Erelt, Mati. 2002. « Evidentiality in Estonian and some other languages. Introductory remarks ». *Linguistica Uralica* 2: 93–97.



- Hennoste, Tiit. 2000. « Suulise kõne erisõnavara III. Partiklid ». Akadeemia 8 : 1773–1806.
- Jaakola, Minna. 2012. « Displaying knowledge in journalistic texts : A contrastive analysis of an evidential particle in Estonian and Finnish ». Lähivõrdlusi. Lähivertailuja. 22 : 43–70. <http://dx.doi.org/10.5128/LV22.02>.
- Kärk, Janika. 2010. « Värvingupartiklid eesti- ja saksakeelsetes küsilausestes ». Eesti Rakenduslingvistika Ühingu aastaraamat 6 : 121–134. <http://dx.doi.org/10.5128/ERYa6.08>
- Kärk, Janika. 2012. « Värvingupartiklite kujunemisloost: saksakeelsete partiklite doch, wohl ning ja eestikeelsed vasted ». Lähivõrdlusi. Lähivertailuja 22 : 131–156. <http://dx.doi.org/10.5128/LV22.05>.
- Käsper, Marge. 2017. Présenter des ouvrages académiques en sciences humaines et sociales en français et en estonien. Pistes contrastives pour étudier les cultures discursives dans les comptes rendus de lecture français et estoniens. PhD diss., Tartu: Tartu University.
- Käsper, Marge. 2019. Une particule évidentielle dialogale et les discours académiques. In Juhani Härmä, Eva Havu, Hartmut Lenk, Begoña Sanroman, Elina Suomela-Härmä. Studies in Comparative Pragmatics. (105–121). Cambridge Scholars Publishing.
- Paveau, Marie-Anne. 2006. Les prédiscours. Sens, mémoire, cognition. Paris : Presses Sorbonne Nouvelle.

Partiklid tõesti ja tõepoolest kolmes registris

Kirsi Laanesoo (Tartu Ülikool, Eesti)

Ettekandes analüüsin partiklite *tõesti* ja *tõepoolest* kasutust ning võrdlen nende funktsioone netikommentaaries, ametlikes suulistes vestlustes ja teadustekstides, mis pärinevad kolmest registrist – neti-, trüki- ja suulisest registrist. Uurimisküsimus on, mis eesmärgil kasutatakse eri registrites üht või teist partiklit. Materjal on kogutud Pragmaatikakorpuse (Prillop jt 2021) netikommentaaries, teadustekstide ja ametlike suuliste vestluste alamkorpustest. Iga alamkorpuse suurus on 500 000 sõna.

Partikleid *tõesti* ja *tõepoolest* peetakse sünonüümideks, kuid uurimuse tulemused näitavad, et nende vahel on erinevused. Esiteks, eri registrid kalduvad eelistama üht partiklit teisele. Ametlikes suulistes vestlustes prevaleerib *tõesti* 252 esinemisega, *tõepoolest* esineb seal 49 juhul. Netikommentaaries eelistatakse samuti kasutada partiklit *tõesti* (375 esinemist).



Tõepool esineb seal 46 korda. Teadustekstides kasutatakse pigem partiklit *tõepool* (40 esinemist), *tõesti* kasutamise kohta leidub seal 18 näidet. Teiseks, esmane analüüs näitab, et kuigi nii *tõesti* kui *tõepool* peamine funktsioon on enda või kellegi teise öeldu paikapidavuse kinnitamine, siis partiklit *tõesti* kasutatakse enam hämmelduse, vabandamise või ironia rõhutamiseks.

Ettekandes analüüsin täpsemalt, millised erinevused tulevad kahe partikli kasutusel eri registrites esile ning mis põhjusel eelistavad registrid üht partiklit teisele.

Kirjandus

- Prillop, Külli, Hennoste, Tiit, Habicht, Külli, Metslang, Helle (2021). Ei saa me läbi „Pragmaatika“ korpuseta. Korpuspragmaatika ja pragmaatikakorpus. – Mäetagused, 81, 161–176.

KirjakeeleEST

Margit Langemets¹, Lydia Risberg² (¹Eesti Keele Instituut, Eesti. ²Tartu Ülikool, Eesti)

Järgmise, eesti keele sõnaraamatu ÕS 2025 koostamise käigus on avalikkuses teravalt üles kerkinud kirjakeele teema ja õigekeelsussõnaraamatu (ÕS) suhe sellega. Kirjakeele norm on selle rakendamise korra (RT I, 14.06.2011, 3) järgi määratud Eesti Keele Instituudi (EKI) uusima ÕSiga, samas pole mõiste *kirjakeele norm* üheski seaduses defineeritud.

Kuigi eesti kirjakeele algus ulatub juba mitme sajandi taha, siis rahvuslikult hakati seda korraldama (keelt üles ehitama ja reeglistama) 19. sajandil, eriti intensiivselt 20. sajandi alguses (vt Viht & Habicht 2022; Plado 2022). Aastakümnete jooksul on kirjakeelt kujundatud kord rangemalt, kord leebemalt, ent ideaaliks on olnud järgida norminguid maksimaalselt ka suulises kõnes, vähemalt avalikus elus (Hennoste 2019). Umbes 20. sajandi teisest poolest alates on kirjakeel muutunud põlvkondade kaupa järjest enam eestlaste emakeeleks (Kerge 2012), st kirjakeel on saanud stabiilseks eesti keeleks, mille reegleid vajadusel ajakohastatakse.

Mõistel *kirjakeel* ongi mitu sisu. Ühtpidi tähistab ta kogu tänapäevast ühiskeelt, keelt üldse, teisalt selle üht allkeelt, normikirjakeelt ehk standardkeelt (vt nt Hennoste 2013). Sageli arvatakse, et idealiseeritud standardkeel ongi keel ise (Milroy 2001: 539). Keelekorralduselt on nii Eestis kui ka mujal Euroopas oodatud ideaalkeele poole pürgimist (vt nt Vogl 2012) ning vähemalt eesti keelekorraldus ongi läbi 20. sajandi selle poole püüelnud (vt nt Erelt 2002;



Hennoste 2019). Tänapäeval, kui (eesti) keelt on võimalus uurida mahukate keelekorpuste põhjal, tekib järjest enam küsimus, kus saab olla ideaalkeel. Kus, millistes registrites seda kasutatakse?

Tihti arvatakse, et ideaalkeelt sisaldab ÕS. Ent senised ÕSid on esindanud eesti keelt tema paljudes registrites: nii üldrahvalikult kasutatavat tänapäeva keelt kui ka normatiivset keelt, arhailisemat rahvakeelt ja oskuskeelt. Vaid üks osa kogu ÕSis esitatud keelest on standardkeel ehk normikirjakeel. See hõlmab ametlikku suhtlust, eeskätt ameti- ja halduskeelt, ka teaduskeelt ja kooliõpetust. Riigi seisukohast on standardkeel üks olulisemaid allkeeli.

Ettekandes arutleme probleemide üle, mis on EKI töölaual seoses järgmise ÕSi koostamisega.

1. Kuidas määrata ühiskeelee ja normikirjakeele piiri? Kus vajatakse ranget normikirjakeelt ja kus võiks loomulikult keelel „ametlikult” rohkem eluõigust olla? Mis keeletasandeid normikirjakeel hõlmab? Mis on kirjakeele tuum (eliidi keel vs. tegelik kasutajaskonna ühisosa)? Analüüsime keelekasutust korpusest valitud tekstide põhjal.
2. Kuidas esitada ÕS 2025-s keeleinfot kasutajale arusaadavalt? Kuidas määratleda registreid? Kas registritähis märgib automaatselt mitte-normikirjakeelt? Kuivõrd sallida kirjakeeles varieeruvust? Mis võiks olla 2025. aastal ilmuva ÕSi paberversiooni sisu? Analüüsime ÕS 2018 registreid ja variantide esitust.

Kirjandus

- Erelt, Tiiu. 2002. *Eesti keelekorraldus*. Tallinn: Eesti Keele Sihtasutus.
- Hennoste, Tiit. 2013. *Eesti keele allkeeled: õpik gümnaasiumile*. Tallinn: Eesti Keele Sihtasutus.
- Hennoste, Tiit. 2019. Eeden ja võlad. *Vikerkaar*, märts. <https://www.vikerkaar.ee/archives/24259>
- Kerge, Krista. 2012. Keelekorralduse põhimõtted. *Sirp*, 27.09.
- Milroy, James. 2001. Language ideologies and the consequences of standardization. *Journal of Sociolinguistics* 5/4, 530–555.
- Plado, Helen. 2022. Rahvakeelsus XX sajandi alguskümnendite Eesti keelekorralduses (Vaateid eesti keelekorralduse arenguloole). *Keel ja Kirjandus* 12. 1075–1092.
- RT I, 14.06.2011, 3 = *Eesti kirjakeele normi rakendamise kord*, <https://www.riigiteataja.ee/akt/114062011003>.



- Viht, Annika & Külli Habicht. 2022. Baltisaksa periood eesti keele korralduses. *Keel ja Kirjandus* 11. 1031–1049.
- Vogl, Ulrike. 2012. Multilingualism in standard language culture. Matthias Hüning, Ulrike Vogl, Olivier Moliner (toim). *Standard Languages and Multilingualism in European History*. Amsterdam; Philadelphia: John Benjamins Publishing Company, 1–42.

Kui maksusüsteemi muresid ära ei lahendata, tõuseb ümbrikupalkade osakaal kindlasti edasi. (Inter)subjektiivsusest ajakirjandustekstides

Helle Metslang (Tartu Ülikool, Eesti)

Vaatlen ettekandes (inter)subjektiivsuse markerite kasutamist ajakirjandustekstides muude tekstiliikide taustal. Jälgin nelja rühma kuuluvate markerite kasutust: 1) episteemilised ja hinnangupartiklid (nt *vist*, *kindlasti*, *ilmselt*, *kahjuks*), 2) komplementlauselise laiendiga predikaatidel baseeruvad diskursusemarkerid (KP-markerid, nt *ma arvan*, *et*), 3) mitmesugused kinnisüksused (*minu meelest*, *minu arvates*), 4) modaalverbid (*võima*, *pidama*, *pruukima*). Eesmärk on selgitada, milliste markerite kasutamine ja millistes funktsioonides on ajakirjandustekstidele omane võrreldes muude tekstiliikidega ja mis tegurid seda võivad tingida (vt nt Arrese 2007). Raamistikuks on (inter)subjektiivsuse käsitlus (Narrog 2017) ning korpuspragmaatiline analüüs (Rühlemann 2019); toetun projekti PRG341 korpusele „Pragmaatika“ (vt Prillop jt 2021) ja senistele uurimistulemustele (nt Hennoste jt ilmumas), Esialgsed tulemused näitavad, et ajakirjandustekstides ilmnevad tendentsid esitada 1) väiteid veenvalt ja põhjendatusele vihjavalt; 2) hinnanguid, oletusi ja kahtlusi esiletõstvalt ja irooniliselt, 3) üldisi ja tulevikku suunatud oletusi ja soovitusi ebaisikuliselt, objektiivsust taotlevalt.

Kasutatakse suurt tõenäosust väljendavaid sõnu nagu *kindlasti*, *tegelikult*, *muidugi*, *ilmselt*, *tõesti*, mis toetavad püüdu lugejat veenda. Sagedamad kui mujal on ka negatiivse hinnangu partiklid *paraku* ja *kahjuks*. Negatiivse hinnangu partiklil ja ka KP-markeril (*ma*) *kardan* (*et*) ei ole seejuures mitte pehmendav ja kaasav funktsioon nagu muudes tekstiliikides, vaid pigem kriitiliselt esiletõstev. Modaalverbide eelistamist partiklitele on peetud omaseks teadustekstile (Pic, Furmaniak 2012), sama tendentsi võib täheldada ka eesti ajakirjandustekstides.



Kirjandus

- Arrese, Juana I. Marín 2007. Commitment and subjectivity in the discourse of opinion columns and leading articles. A corpus study. RAEL: revista electrónica de lingüística aplicada, N°. Extra 1, 2007.
- Hennoste, Tiit, Külli Prillop, Külli Habicht, Helle Metslang, Kirsi Laanesoo, Liina Pärismaa, Elen Pärt, Andra Rumm, Andriela Rääbis, Carl Eric Simmul (ilmumas), Pragmatic particles, roles, and registers: the case of Estonian. Open Linguistics.
- Narrog, Heiko 2017. Three types of subjectivity, three types of intersubjectivity, their dynamicization and a synthesis. – D. Van Olmen, H. Cuyckens, & L. Ghesquière (eds.), Aspects of Grammaticalization: (Inter)subjectification and directionality. Berlin/Boston: De Gruyter Mouton, 19–46.
- Pic, Elsa, Grégory Furmaniak 2012. A study of epistemic modality in academic and popularised discourse: the case of possibility adverbs perhaps, maybe and possibly. – Revista de Lenguas para Fines Especificos 18. Universite Sorbonne Nouvelle. PRES Sorbonne Pris Cite – Primes EA 4398, 13–44.
- Prillop, Külli, Tiit Hennoste, Külli Habicht, Helle Metslang, 2021. Ei saa me läbi „Pragmaatika“ korpuseta. Korpuspragmaatika ja pragmaatikakorpus. Mäetagused 81, 161–176.
- Rühlemann, Christoph 2019. Corpus linguistics for Pragmatics. A Guide for Research. Routledge Corpus Linguistics Guides. Taylor and Francis. Kindle Edition.

Kes te üldse olete? Partikkel üldse suuliste argivestluste küsilausungites

Andriela Rääbis (Tartu Ülikool, Eesti)

Üldse on eelkõige suulistele ja netivestlustele iseloomulik partikkel. Suulistes argivestlustes kasutatakse seda paljudes suhtlustegevustes, millest kesksed on küsimus (*millega te tulite üldse.*), informeerimine (*neljapäeva õhtuks meil üldse lund ei=olnd enam.*), kurtmine (*ma=i maganud üldse ma=i jäänd magama.*), hinnang/arvamus/väide (*ta on nii tüütu inimene et toda ei taha üldse nähagi.*).

Ettekandes käsitletakse partikli *üldse* kasutust küsilausungites. Uurimus on tehtud projekti “Pragmaatika grammatika kohal: subjektiivsus ja intersubjektiivsus eesti keele registrites ja tekstiliikides (PRG 341)” raames. Analüüsitav materjal pärineb TÜ suulise eesti keele korpuse 500 000-sõnalisest argivestluste alamkorpusest, milles partiklit *üldse* esines 587 korda. Analüüsiti 300 esinemisjuhtu, mille hulgas on 72 küsilausungit.



Uurimisküsimus on, missuguseid funktsioone *üldse* küsilausungites täidab. Esiadne analüüs näitab, et vaadeldavad küsilauseid jagunevad üldjoontes kolme rühma:

1) pööratud polaarsusega küsimused, mis väljendavad küsilausungi jaatusele-eitusele vastupidiseid väiteid (seisukohti, hinnanguid, vaidlustusi, etteheiteid, kahtlust) (Koshik 2005; Laanesoo 2018: 34–35) (*no `miks=seda `apu`koort `pannakse üldse.*). Partikliga *üldse* tugevdatakse väidet.

2) vestluskaaslase jutus olnud üllatavale, ebaootuspärasele vms infole reageerivad hoiakulised küsimused, millega soovitakse ühtaegu infot ja osutatakse normi rikkumisele (*ma ei `saa temaga ((perearstiga)) `rääkida ta on `venelane. /---/ – kas ta `ültse eesti keelt ei `oska=võ.*). Vestluskaaslase eelneva vooruga võrreldes lisatakse või intensiivistatakse partikliga *üldse* hinnangulisust (Pomerantz 1984).

3) infoküsimused. Sageli moodustavad need kõrvalepõike vestluse pealiinist ning *üldse* avardab küsimust, osutades üldisema info puudumisele (*a=kust=te `tunnete üldse.*). Ka teemat alustavates infoküsimustes näitab *üldse* täielikku mitteteadmist (*no: `räägi mulle algusest `seda=et (.) et=ee (0.8) millal see `ärasõitmine ültse on ja mis `kuupäevadel sa oled `mõelnud ja=nii=edasi ja=nii=`edasi.*). Infoküsimused on enamasti neutraalsed, kuid mõnikord muudab *üldse* küsimuse veidi hoiakuliseks.

Kirjandus

- Koshik, Irene 2005. Beyond Rhetorical Questions: Assertive Questions in Everyday Interaction. Amsterdam, Philadelphia: John Benjamins.
- Laanesoo, Kirsi 2018. Polüfunktsionaalsed küsilauseid eesti argivestluses. Dissertationes linguisticae Universitatis Tartuensis 33. Tartu: Tartu Ülikooli Kirjastus.
- Pomerantz, Anita 1984. Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. – M. Atkinson, & J. Heritage (Eds.), Structures of Social Action: Studies in Conversation Analysis. Cambridge: Cambridge University Press, 57–101.

Veebitekstide liigitamine keeleliste tunnuste alusel

Kristiina Vaik, Kairit Sirts, Kadri Muischnek (Univeristy of Tartu, Eesti)

Tänapäeval toetatakse tekstikorpuste loomisel enamasti automaatsele veebikraapimisele, kus tekstide nõ žanrilist kuuluvust on raske määratleda. Isegi, kui eksisteeriks sobiv üldlevinud taksonoomia, siis automaatne liigitamine eeldaks hästi anoteeritud treeningandmestikku. Probleem on aga selles, et inimesed pole võimelised järjekindlalt tekste liigitama (Crowston jt 2010, Meyer Zu Eissen jt 2004, Suchomel 2020), rääkimata sellest, et keel areneb tekstides



alatasa, st on tekkinud/tekivad hübriidžanrid. Meie pakume välja uue lahenduse: dimensionaalne tekstimudel (Vaik jt 2020). Selle põhiidee on kategoriseerida tekste nii, et tekstid, milles esinevad samasugused keelelised tunnused, on ka funktsiooni järgi sarnased. See on süntees Douglas Biberi multidimensionaalsest analüüsist (Biber 1986, 1988) ning Serge Sharoffi funktsionaalsetest tekstidimensioonidest (Sharoff 2018). Kuna sobilikku taksonoomiat veebitekstide klassifitseerimiseks hetkel ei eksisteeri, siis dimensionaalne tekstimudel läheb ühe sammu võrra sügavamale: oleme postuleerinud hulga latentseid dimensioone (ehk tekstides esinevaid tuumiknähtuseid või -omadusi), mida kirjeldame ja mõõdame koosinevate keeleliste tunnuste kaudu. Idee järgi tekstid, milles esinevad samasugused keelelised tunnused, on ka tuumiknähtuse osas sarnased ja omakorda koosinevad tuumiknähtused peegeldaksid mingisugust suuremat konstrukti, nt register või žanr. Ettekandes vaatleme neid latentseid dimensioone lähemalt ning jõuame lõpuks selleni, mida eelnevalt dimensioonide suhtes annoteeritud veebitekstide minikorpus tegelikult paljastab.

Kirjandus

- Biber, Douglas 1986. Spoken and written textual dimensions in English: Resolving the contradictory findings. – Language (62), nr 2, lk 384–414. <https://doi.org/10.2307/414678>
- Biber, Douglas 1988. Variation across Speech and Writing. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511621024>
- Crowston, Kevin, Kwasnik, Barbara, Rubleske, Joseph 2010. Problems in the use-centered development of a taxonomy of web genres. Genres on the Web, lk 69–84. Springer.
- Meyer Zu Eissen, Sven, Stein, Benno 2004. Genre classification of web pages. Annual Conference on Artificial Intelligence, lk 256–269. Springer.
- Sharoff, Serge 2018. Functional text dimensions for the annotation of web corpora. – Corpora (13), nr 1, lk 65-95. <https://doi.org/10.3366/cor.2018.0136>
- Suchomel, Vít 2020. Better Web Corpora For Corpus Linguistics And NLP. Brno, Masaryk University. https://is.muni.cz/th/u4rmz/Better_Web_Corpora_For_Corpus_Linguistics_And_NLP.pdf
- Vaik, Kristiina, Sirts, Kairit, Muischnek, Kadri 2020. Dimensionaalne tekstimudel: teoreetiline ülevaade. Keel ja Kirjandus (10), lk 875–898. <https://keeljakirjandus.ee/ee/archives/27097>



“Ofc its eesti keel”: inglise keel teismeliste tsätivestlustes

Virve Vihman, Kristiina Praakli, Maarja-Liisa Pilvik, Mari-Liis Korkus (Tartu ülikool, Eesti)

Ettekandes analüüsime inglise keele kasutust tsätivestlustes ning tutvustame loomisel olevat teismeliste tsätikeelekorpus. Teadusprojekti “Teismeliste keel Eestis” (TeKE, 2019–2022) raames kogutud teismeliste tsätikeele andmestik (hetkel kokku veidi alla 60 000 sõna) sisaldab autentset keeleainest 102 kõnelejal vanuses 10–18 eluaastat. Tsätikeel on üks veebisuhtluse variante, millest siinses ettekandes võtame vaatluse alla kiirsuhtlusrakendustest (Messenger, Discord) pärit keeleandmed.

Inglise keel on oluline sotsiaalne ressurss ning tähendusloome vahend teismeliste tsätisuhtluses, kelle väga hea inglise keele oskus on ühtlasi osa nende suhtuspädevusest. TeKE tsätikeele korpuse sõnadest on 7,3% ingliskeelsed; see on oluliselt rohkem, kui suulises keeles esinenud 3,3% (Vihman jt 2022). Ettekande kitsam fookus lasub ingliskeelsete sõnade ja väljendite kasutuse kirjeldamisel ning nende ortograafilise (nt *tänkjuu*, *dunt võrri*) ja morfoloogilise (nt ‘jalad *shavemata*’, ‘miks sa *depressionis*’) integreerimise ja mitteintegreerimise (nt ‘mingid *fancy* veinipokaalid’) analüüsimisel (vt ka Deuchar 2020). Ortograafiline integratsioon on harvem nähtus kui morfoloogiline ning TeKE andmed näitavad, et ortograafilise integreerimise kasutus ja funktsioon erineb oluliselt fonoloogilisest integratsioonist suulises keeles. Inglise keele määr varieerub kõnelejati, kuid inglise keel kuulub pea kõigi keeleandmeid TeKE korpusele vahendanud teismeliste keelepruuki. Eesti noorte tsätikeeles mängib inglise keel erilist rolli, mille analüüsimiseks on vaja selget metodoloogilist ja teoreetilist raamistikku.

Kirjandus

- Deuchar, M. (2020). Code-switching in linguistics: A position paper. *Languages*, 5(2), doi:10.3390/languages5020022.
- TeKE = Teismeliste keel Eestis. www.sisu.ut.ee/teke/ (27.02.2023).
- Vihman, Virve, Kristiina Praakli, Maarja-Liisa Pilvik, Mari-Liis Korkus (2022). Kas noored on inglise keelega obsessed? Millest räägivad korpusandmed? – *Philologia Estonica Tallinnensis* 7, 292–231. DOI: 10.22601/PET.2022.07.11



Posterettekanded / Poster presentations

Lithuanian Heritage Language: Acquisition of Grammar

Skirmantė Gribauskienė (Vytautas Magnus University, Lithuania)

Ireland is one of the European countries with the highest density of speakers of Lithuanian as a heritage language. Heritage language user is often defined as a person who acquired the heritage language as L1 from one's parents, but later adopted the dominant language of the country as L1 and lost (partially or completely) heritage language skills (Rothman 2009: 156; Benmamoun et al. 2013; Scontras et al. 2015). Heritage language research is a relatively new topic within sociolinguistics and migration studies. Therefore, new large-scale empirical studies are necessary for linguists, educators, and language minorities.

The goal of this research is to analyse the acquisition of Lithuanian grammar (case, gender, number, tense) in children living in Ireland. This study has also considered the influence of the dominant English language, language contacts and other multilingual factors on Lithuanian grammar acquisition and maintenance.

The participants of this study are young Lithuanian heritage speakers between the ages of 4 to 16. All the participants were born and raised in Ireland and at least one of their parents is a native Lithuanian speaker. The study is based on the data of a sociolinguistic questionnaires (for parents and for children) and semi-experimental (elicited production) tasks.

The findings demonstrate that the development of Lithuanian grammar is affected by the dominant English language. Elicitation tasks showed that some of the English language features could be found in participants' Lithuanian language. Children are prone to making mistakes in noun-verb case and gender agreement in Lithuanian, linguistic code-switching is also prevalent in children's spontaneous speech. It is interesting to note that older children tend to have slightly better Lithuanian language skills comparing to the younger age groups. Questionnaire analysis demonstrated that a family's linguistic practices and preferences also have a huge impact on a child's language skills and development.

Since the sociolinguistic context of the Lithuanian diaspora is dynamic, more in-depth research should be conducted in the future.



Spoken language resources: learning Lithuanian as a second language

Laura Kamandulytė-Merfeldienė, Ineta Dabašinskienė (Vytautas Magnus University, Lithuania)

The aim of this presentation is to display the Pedagogic Corpus of Spoken Lithuanian (<https://kalbu.vdu.lt/en/>), that contains authentic Lithuanian language texts, selected according to criteria that are relevant to language learners of different proficiency levels. All the texts are classified into levels A1-A2, and B1-B2 according to the Common European Framework of Reference for Languages. The corpus can be useful for both Lithuanian language learners and teachers. By examining a search word in the corpus, users will be able to understand how this word is used in current Lithuanian and what it means. The authentic examples of language use obtained from the corpus can be used when preparing teaching materials targeting different learner competencies in the areas of vocabulary and grammar (Kovalevskaitė et al. 2022). The spoken part of the corpus will be discussed in more detail. This part consists of 15 000 words for levels A1-A2 and 35 000 words for levels B1-B2. The data consist of natural conversations recorded in different settings, covering different communicative situations and various social roles of the participants. The spoken data includes face-to-face and telephone service calls in catering establishments (restaurants, cafes, etc.), trading venues (shops, kiosks, bookstores, pharmacies, theatres, etc.), establishments of different service providers (beauty salons, sewing shops, banks, etc.), as well as conversations at home and in the workplace environment. The development of the spoken part is composed of three main stages: collecting the data, the transcription of the recorded data, and the grammatical annotation. It should be noted that the creation of the corpus was inspired by the existing Corpus of Spoken Lithuanian (see <http://sakytinistekstynas.vdu.lt/>), as well as by the Lithuanian corpus of child speech and child-directed speech (Dabašinskienė, Kamandulytė 2009, Kamandulytė-Merfeldienė 2017). The second part of the talk presents a corpus-based analysis and provides preliminary results. The data of spoken language are analysed to reveal the frequency distribution of parts of speech, as well as the frequency of grammar forms of the noun, verb, and adjective. The main differences and similarities within the two corpus parts (A1-A2 and B1-B2) are discussed as well.

References

- Dabašinskienė I., Kamandulytė L. 2009. Corpora of Spoken Lithuanian. Estonian Papers in Applied Linguistics 5, 67-77.
- Kamandulytė-Merfeldienė L. 2017. Sakytinės lietuvių kalbos tekstynas – natūralios vartosenos tyrimų šaltinis. Taikomoji kalbotyra 9, 176-198.



- Kovalevskaitė J., Rimkutė E., Ruzaitė J. 2022. Tekstynai ir jų išvestiniai produktai lietuvių kalbos mokymui(si) bei tyrimams. Kaunas.

Refusals strategies in the role-play of Lithuanian children

Viktorija Kavaliauskaite-Vilkiniene (Vytautas Magnus University, Lithuania)

This study investigates refusals strategies in the speech of Lithuanian children. Refusals are face-threatening speech acts. In order to maintain good social relationships, it is typical to mitigate their expression by using politeness markers (Katinaitė and Hilbig 2021). Thus, the acquisition of refusals is an integral part of communicative competence as well as an essential part of the language socialisation process. So far, plenty of studies in children's language have been conducted on the expression and development of other speech acts that require the employment of politeness markers (e.g., requests, apologies), while the acquisition of refusals strategies has received only a few researchers' attention (e.g., Yang and Chang 2008; Chang and Ren 2020). Thus, the present investigation, using the data on Lithuanian children, contributes to scarce studies on the acquisition of refusals and children's communicative competence in general.

In this study, refusals are analysed in children's role-play conversations in which they were imitating formal (e.g., appointments at the doctor's, visits to cafés and shops) and informal (e.g., birthday celebrations and a family's everyday life) social situations. In the role-plays participated 83 children of two age groups: 3-4 years old (average age 4;3) and 5-6 years old (average age 6;0). The main finding suggests that both age groups prefer to express refusals directly. However, older children employ more indirect refusal strategies than younger children. The results are discussed not only in relation to the children's age but also in relation to the imitated social situation (formal vs informal) and utterance type (produced acting in the role vs in metacommunication).

References

- Chang, Y. F. and Ren, W., 2020. Sociopragmatic competence in American and Chinese children's realisation of apology and refusal. *Journal of Pragmatics*, 164, pp.27-39.
- Katinaitė, D. and Hilbig, I., 2021. Lithuanian Refusals and Politeness. *Respectus Philologicus*, (39 (44)), pp.58-72.
- Yang, L.C. and Chang, Y.F., 2011. *Young Children's Refusal in Mandarin Chinese: Their Production and Perception*. VDM Publishing.



Production of consonant clusters of 4–8-year-old Lithuanian children

Eglė Krivickaitė-Leišienė (Vytautas Magnus University, Lithuania)

In the present research, the analysis of production of consonant clusters was based on the ideas of the Naturalness Theory (Dressler 1993, 2005; Dressler et al. 2001;). Naturalness in a language refers to a form which is cognitively simple, easy to pronounce, and to perceive; therefore, it is used frequently. Naturally, in language acquisition, children make more errors in using marked forms, i.e. the elements which are not very natural are frequently substituted by more natural and more usual elements and/or the elements that are easier to pronounce (Dziubalska-Kołodziejczyk 2007, 2014; Marecka, Dziubalska-Kołodziejczyk 2014).

The aim of the study is to analyze the production of consonant clusters of 4–8-year-old Lithuanian children. Applying the Non-word repetition test, the data of 480 typically developing children of five age groups from four to eight years old were collected.

The results showed that participants repeated consonant clusters with the accuracy of 83%. The pronunciation results of consonant clusters in the age groups 4;00-8;11 are distributed in the interval 74-87%. The children aged 4;00-4;11 repeated the clusters statistically significant ($p=0.000$) worse than other research groups.

The results of different age groups have revealed that younger children have difficulty in repeating the clusters in the medial position (the clusters in the initial position were repeated by four-year-old children with the accuracy of 82% and five-year-old children – 89%, while the clusters in the medial position were repeated by four-year-old children with the accuracy of 68% and by five-year-old children – 78%). The children in the age groups from six to eight repeated the clusters in the initial position with the accuracy of 92-95%, and in the medial position with the accuracy of 78-79%.

The qualitative analysis shown that children apply these universal pronunciation simplification strategies to pronounce consonant clusters in the word: substitution, omission, consonant assimilation, metathesis, sound migration to another syllable, and sound addition.

References

- Dressler, W. U. 1993: What is Natural in Natural Morphology (NM)? Prague Linguistic Circle Papers. Prague: Charles University, 135–145.



- Dressler W. U., Dziubalska-Kořaczyk K., Rossella S. 2001: Sources of Markedness in Language Structures. *Folia Linguistica Historica* 22 (1-2). 103–135.
- Dressler W. U. 2005: Word-formation in Natural Morphology. *Handbook of Word-Formation*. Springer Netherlands, 267—284.
- Dziubalska-Kořaczyk K. 2007: Natural Phonology: Universal Principles for the Study of Language (Insiders Meet Outsiders). *Saarbrücken* 6 (10), 71–75.
- Dziubalska-Kořaczyk K. 2014: Explaining Phonotactics Using NAD. *Language Sciences*, 6–12.
- Marecka M., Dziubalska-Kořaczyk K. 2014: Evaluating Models of Phonotactic Constraints on the Basis of sC Cluster Acquisition Data. *Language Science*, 37–47

Development of writing skills: task types and feedback

Sille Midt (Tartu Ülikool, Estonia)

Writing is a cognitively complex activity that generally requires a long-term effort from the student. If at A2 level, students write short letters according to the information or model provided, then the writing skills of B1 level students must expand significantly, from being able to convey their experiences and impressions, describe events, make summaries and present factual information to justifying actions (EKR 2007). In order to acquire writing confidence, it is necessary to deal with the development of writing skills in a diverse and systematic manner, following communicative principles and applying active learning methods. In the past, it has been found that too little development of writing skills is dealt with in Estonian language classes: the writing assignments do not take into account all the applications of the learner's developing language and writing is often practiced in the same types of assignments. (Metslang et al. 2013)

Feedback often follows the finished piece of writing. Feedback can be given on different aspects of the text, such as content, structure, processes, accuracy. In communicative learning, it is important to give feedback on the content to inform on communicative success on the task (Larsen-Freeman, Anderson 2016), but positive effect of corrective feedback on the acquisition of language structures has also been shown (Ferris, Kurzer 2019). Nevertheless, the feedback effect depends on many factors, both methodological (including correction techniques, focus and tone of feedback) and contextual (language learner's language proficiency level, type of writing task) (Biber et al. 2014; Kang, Han 2015).

In the presentation, I will give an overview of the most frequently used task types in the 7th-8th grades of Russian-language general education schools, based on texts collected through



a pilot study. I also describe the corrective feedback given to these texts, compare the feedback to texts written at different levels and analyze the results obtained. Based on my master's thesis (Kitsnik, Midt 2022), I also compare how different the results of the analyzed papers are from teachers' opinions on how they give feedback.

Convenience sampling has been used to collect data for the pilot study. The texts requested from the teachers have been analyzed both qualitatively and quantitatively.

The role of input and output in developing second language speaking skills

Katrin Mikk (Tartu Ülikool, Estonia)

When it comes to language acquisition, it is precisely the achievement of speaking skills that has caused the greatest difficulties for learners (Masuram, Sripada 2020). Teaching speaking skills, it is important to ensure that the learner has enough authentic input (Krashen 1985), i.e. situations and learning material in which he hears and sees the language being studied. However, not any linguistic text is suitable for input. It is most effective if the input is of interest to the learner, corresponds to the learner's level of language proficiency or is a little more complex. (Long 1996) At the same time, it is not always easy to find authentic, accessible and interesting input. (Krashen 2015) Krashen considers input to be more important than output, and even argues that good language proficiency is possible without any output (Krashen 1998). However, in order to achieve the ability to speak, the output is also very important, i.e. a situation where the learner can use the language himself. According to Swain (1985), the output has three roles: the language testing function, the meta-linguistic function, and the learner becoming aware of what he already knows and where shortcomings are still present. When using language, linguistic fluency, understanding of real meanings and automaticity also develops, without which it is impossible to speak (Thornbury 2005).

In the presentation, I will discuss the importance of input and output in the development of speaking skills. I will speak about the quality and quantity of input and output based on the lesson observations of the ongoing study on teaching Estonian as a second language. The study is part of my doctoral thesis, which studies the impact of language learning techniques on the development of adult speaking skills and motivation to speak in B1 level courses.

References

- Krashen, Stephen. 1998. Comprehensible Output? System 175–182.
- Krashen, Stephen 2015. Controversies and issues in language teaching. <https://www.youtube.com/watch?v=TdoDTXwZVi0>.



- Long, Michael 1996. The Role of the Linguistic Environment in Second Language Acquisition. In the Handbook of Second Language Acquisition. by W. and Bhatia, T. Ritchie, 413–468. San Diego: Academic Press.
- Masuram, Jyothi; Sripada, Pushpa Nagini 2020. Developing spoken fluency through task-based teaching. Procedia Computer Science. 6 Language: English. DOI: 10.1016/j.procs.2020.05.080, 172: 623–630.
- Swain, Merrill 1985. Communicative competence: some roles of comprehensible input and comprehensible output in its development. In Input in Second Language Acquisition, by In S. Gass & C. Madden (Eds.), 235–253. Rowley, MA: Newbury House.
- Thornbury, Scott. 2005. How to Teach Speaking. Harlow: Longman.

Näita lapsele (võõr)keelt: Show the child a (foreign) language on a smart device or how smart devices enrich children's language acquisition process

Merje Miliste, Natalja Zagura (Tartu Ülikool, Estonia)

Smart devices (phones, tablets) are nowadays children's companions from an early age on. Moving one's thumb may lead to a sense of pleasure (Aru 2022) and parents can put this 'addiction' to good use before the child reaches school age: given the synaptogenesis of the hippocampus and the development of the brain in relation to hearing and vision, children can acquire a foreign language naturally at home without even the parents speaking the target language.

On the basis of the optimal input hypothesis (Krashen/Terrell 1983), we analyse two videos in which Paula (5:4) and Harald (5:10), both from a monolingual Estonian family, use English while playing alone with different toys, without having been physically present in the corresponding cultural space or formally learning English in a pre-school institution. Semi-structured, non-standardised paired interview with parents is planned to explore the role of parents in foreign language acquisition through the use of a smart device. We aim to explore whether the toddler is a self-directed language learner, i.e. how much or how little the parent should intervene in the toddler's activities in acquiring a foreign language online. We also address the question whether the results are generalisable or transferable to early E2 (Estonian as the second language). The interview will be analysed through qualitative content analysis.



References

- Aru, Jaan. 2022. Loovusest ja logelemisest. Nutineedusest, mõtteväljatustest ja laste aregust. Tallinn: Raamatutrükikoda.
- Kirch, Michael. 2009. Englisch lernen mit dem Fernsehen. Eine Studie über die Eignung des Fernsehens im Rahmen des frühen Fremdsprachenerwerbs am Beispiel der Sendung Something Special. Münster.
- Kirch, Michael. 2008. Sprachenlernen mit dem Fernsehen. Können Programme das Lernen einer zusätzlichen Sprache fördern? In: TelevIZion. Hrsg. von Internationales Zentralinstitut für das Jugend- und Bildungsfernsehen (IZI) beim Bayerischen Rundfunk. München : http://www.br-online.de/jugend/izi/deutsch/publikation/televizion/21_2008_2/kirch.pdf (20.02.2023)
- Krashen, Stephan. D., Tracy Terrell. 1983. The Natural Approach: Language Acquisition in the Classroom. Oxford: Pergamon Press.
- Kübler, Hans Dieter, Wolfgang H. Swoboda. 1998. Wenn die Kleinen fernsehen. Forschungsprojekt über die Bedeutung des Fernsehens in der Lebenswelt von Vorschulkindern. Hrsg. von den Landesmedienanstalten. Berlin 1998.
- Saks, Kristiina. 2022. Laste elumaailma meediastumine. Dissertationes de mediis et communicationibus Universitatis Tartuensis. <https://dspace.ut.ee/handle/10062/84228>
- Tracy, Rosemarie. 2007. Wie Kinder Sprachen lernen. Und wie wir sie dabei unterstützen können. Tübingen: Franke Verlag.

Eesti keele kui teise keele grammatikapädevuse arendamise võimalusi ja vajadusi

Raili Pool^{1,2}, Jelena Kallas² (¹Tartu Ülikool, Eesti. ²Eesti Keele Instituut, Eesti)

Euroopas on nii täiskasvanute kui ka noorte (kuni 15-aastaste) õppijate keeleõppe korraldamine, planeerimine ja keeleoskuse hindamine seotud Euroopa Nõukogu väljatöötatud keeleoskustasemetega jaotuse ja kirjeldustega, eelkõige 2001. aastal välja antud Euroopa keeleõppe raamdokumendi (CEFR 2001) ja selle 2020. aastal ilmunud sõsarväljaandega (CEFR/CV 2020). Ettekandes keskendume nendes dokumentides käsitletud grammatikapädevuse arendamisega seotud aspektidele.

Grammatikapädevus tähendab keele grammatiliste vahendite tundmist koos oskusega neid kasutada (CEFR 2007: 131). Eesti keele kui teise keele eri keeleoskustasemetega (A1-C1)



grammatikapädevuse kirjeldus on esitatud Eesti Keele Instituudi keeleportaali Sõnaveeb alamlehel Õpetaja tööriistad (vt lähemalt Kallas jt 2021). Kirjeldus annab ülevaate noore ja täiskasvanud keeleõppija grammatikapädevuse arengust, st mida keelekasutaja või -õppija vormi-, sõna-, fraasi- ja lausemoodustuse tasandil teha oskab ja milliste piirangutega ta seda teeb (nt kasutab mõnd vormi vaid tervikuna omandatud struktuurides, eksib vormide moodustamisel jne). Grammatikapädevuse kirjeldamisel olid aluseks õpiku- ja õppijakeelekorpused.

Ettekande eesmärk on vaadelda, kuidas suhestub eelmainitud tasemepõhise grammatikapädevuse kirjeldus seni koostatud eesti keele kui teise keele grammatika käsiraamatute, grammatika teemasid sisaldavate veebikeskkondade, e-harjutusvara ja e-kursustega Moodle'is. Analüüsime ka nende materjalide sihtgruppide ja taseme määratlemist, kasutatavat metakeelt (eelkõige grammatiliste kategooriate nimetamist), vahenduskeele kasutamist ja harjutuste tüüpe. Veebikeskkondade ja e-kursuste juures vaatleme ka videomaterjalide kasutust ja avalikku kättesaadavust. Fookuses on täiskasvanud õppijale mõeldud materjalid.

Esialgne analüüs näitab, et materjalide mitmekesisus on suur, kuid puudu on raamdokumenti keelespetsiifiliselt arvestav, keeleõppijat tasemelt tasemele suunav grammatika teemasid käsitlev veebikeskkond koos harjutusvaraga. Osa olemasolevaid materjale, eriti käsiraamatud, on tasemeülesed, pakkudes terviklikku ülevaadet kogu eesti keele grammatilisest struktuurist. Veebikeskkondades ja e-kursustel esitletakse grammatika teemasid kursuse õppetükkide osana, seletades teemasid kas video vahendusel või eraldi failis oleva teksti abil ning juures on ka harjutused. E-kursuste puhul on probleemiks nende kättesaadavus: kursused on seotud kindla kõrgkooli kasutajatunnustega ja need on koolivälistele õppijatele tasulised. E-harjutusvara on veebis laiali, linkimata mingi konkreetse grammatika käsiraamatu, veebikeskkonnaga või kursusega. Täheldatav on ka metakeele variatiivsus õppematerjalide lõikes, varieerudes rahvusvahelisest või omakeelsest lingvistilisest terminoloogiast kuni autorite endi loodud süsteemini.

Kirjandus

- Euroopa keeleõppe raamdokument: õppimine, õpetamine, hindamine. Tartu: Haridus- ja Teadusministeerium, 2007.
- Euroopa keeleõppe raamdokument: õppimine, õpetamine ja hindamine. Uute tasemekirjeldustega sõsarväljaanne (2020) Intrenetis aadressil: https://hm.ee/sites/default/files/documents/2023-01/S%C3%B5sarv%C3%A4ljaanne_Jaanuar%202023.pdf



- Kallas, Jelena; Koppel, Kristina; Pool, Raili; Tsepelina, Katrin; Üksik, Tiiu; Alp, Pilvi; Epner, Anu 2021. Eesti keele kui teise keele õpetaja tööriistad Eesti Keele Instituudi keeleportaalis Sõnaveeb. – Eesti Rakenduslingvistika Ühingu aastaraamat 17, 61–80.

„Parempoolse vindikeeraja elukutse omandanu jääb tööta” ehk elukestva õppe diskursuse konstrueerimine Eesti meediatekstides

Triin Rattiste, Halliki Põlda (Tallinna Ülikool, Eesti)

Ettekanne keskendub elukestva õppe diskursuse keelelisele konstrueerimisele Eesti meedias ja kirjeldab selle diskursuse muutust viimase kümnendi jooksul, s.o ajavahemikus, mil hakkas kehtima „Eesti elukestva õppe strateegia 2020” ning jätkustrateegia „Haridusvaldkonna arengukava 2021–2035”. Ühiskondlike väärtuste kujundamisel on meedial oluline roll ja võim algatada avalikke diskussioone ja arutelusid, peegeldades ning kajastades sotsiaalseid arusaamasid ja arvamusi (Kõuts-Klemm & Seppel, 2018). Sellest tulenes ka meie uurimishuvi.

Püstitame uurimisküsimused: 1) Kuidas konstrueeritakse Eesti meedias elukestva õppe diskursust? ja 2) Kuidas on elukestva õppe diskursus viimasel kümnendil Eesti meedias muutunud?

Uurimisobjektiks on 2012. ja 2022. aasta elukestva õppe meediakajastused valdkondlikes ja üleriigilistes väljaannetes. Valimi moodustavad elukestva õppe kajastused Delfis (n=58) ja Õpetajate Lehes (n=24). Tekstide analüüsimisel juhendusime Norman Fairclough’ kriitilise diskursuseanalüüsi mudelist (2001). Tulemusena kirjeldame kuut peamist elukestva õppe diskursust ja nende keelelisi konstruktsioone: 1) *rahvusvahelistumise diskursus*, mida konstrueeriti õppereiside ja rahvusvaheliste suhete ring võõrkeelte õppimise kaudu; 2) kiirelt muutuva maailma ning selle vajadusi iseloomustav *keskkonna diskursus*, mis avaldus digitaliseerumise ja tehnoloogia arengu, rohepöörde, vaimse tervise ning õpikeskkonna kujundamise kaudu; 3) elukestva õppe peamisi mõjutegureid ja takistusi iseloomustav *õppimise mõjutamise diskursus*; 4) enesetäiendamise ja õppimise vajadust iseloomustav *edudiskursus* kui uute teadmiste ja oskuste pidev täiendamine, konkurents ja töö ning tunnustus; 5) *nüüdisaegse õpikäsituse diskursus*, mida konstrueeriti personaliseeritud õppe ja kaasava hariduse kaudu, osutades ühelt poolt haridusuuenduste vajalikkusele, samas väljendades muutustele vastuseisu ning 6) *täiskasvanu hariduse diskursus*, mis põhineb kavatsuslikul ja mitte-kavatsuslikul õppel.



Uuringust järeldub, et elukestva õppimise diskursus on viimase kümnendiga muutunud õppijakessemaks ja suhtlust ning kogemust väärtustavamaks. Kiirelt muutuv maailmas on olulised tulevikuoskused, nagu paindlikkus, loovus, ettevõtlikkus, analüütilisus, koostöövõime jms, ning ka “parempoolse vindikeeraja elukutse omandanu” peab konkurentsivõime säilitamiseks jätkama elukestva õppimisega.

Kirjandus

- Eesti elukestva õppe strateegia 2020. (2014). Tallinn
- Fairclough, N. 2001. Language and power. (2.ed). New York: Longman Fairclough, N. 2010. Critical Discourse Analysis: The Critical Study of Language. (2.ed). Harlow: Longman
- Haridusvaldkonna arengukava 2021-2035. (2021). Haridus- ja Teadusministeerium.
- Kõuts-Klemm, R., Seppel, K. (2019). Juhatus meedia- ja kommunikatsiooni teooriasse. Tartu Ülikooli Kirjastus

The use of markers of stance among English L1 speaker and Estonian L1 learners of English

Denys Savchenko (University of Tartu, Estonia)

Markers of stance are used to express “personal feelings, attitudes, value judgments, or assessments” (Biber et al. 1999: 966). There are multiple ways in which speakers express their viewpoint regarding the propositions. In this presentation, following Larsson (2017), I will discuss the use of small set of stance markers in the written and spoken language of English L1 speakers and Estonian L1 learners of English. Larsson (2017, 2019) show that lexis play a role in the distribution of the following realisations of stance: stance adverbials (*obviously, unfortunately* etc.), stance complement clause constructions (*I'm sure that* etc.), stance noun plus prepositional phrase (*the importance of* etc.). Gilquin et al. (2007) and Larsson (2019) reported that learners of English and novice academic writers rely on markers of stance associated with spoken language, stressing that register unawareness is a factor influencing the distribution of markers of stance.

Using Multiple Correspondence Analysis I will outline the main trends in the use of stance markers by Estonian L1 learners of English and English L1 speakers. To examine the distribution of morphologically related sets of stance markers, I will use Estonian component of the Louvain International Database of Spoken English Interlanguage (LINDSEI, Gilquin et al. 2010) and the Louvain Corpus of Native English Conversation (LOCNEC, De Cock 2004),



The Louvain Corpus of Native English Essays (LOCNESS, Granger 1998) and a collection of BA thesis.

References

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. Longman Grammar of Spoken and Written English. Harlow: Longman.
- De Cock, Sylvie. 2004. Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures (BELL)*, 2:1, 225-246
- Gilquin, Gaëtanelle, Sylviane Granger, Magali Paquot. 2007. Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6:4, 319-335.
- Gilquin, Gaëtanelle, Sylvie De Cock, Sylviane Granger. 2010. The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM. Presses universitaires de Louvain : Louvain-La-Neuve
- Granger, Sylviane. 1998. The computer learner corpus: A versatile new source of data for SLA research. In Sylviane Granger (ed.). *Learner English on Computer*, 3-18. London & New York: Addison Wesley Longman.
- Larsson, Tove. 2017. The Importance of , It Is Important That or Importantly ? : The Use of Morphologically Related Stance Markers in Learner and Expert Writing'. *International Journal of Corpus Linguistics*. 22:1. 57–84
- Larsson, Tove. 2019. Grammatical Stance Marking across Registers: Revisiting the Formal-Informal Dichotomy. *Register Studies*, 1:2, 243–268.

Kirjalikud tööd Tartu Ülikoolis 2009 ja 2023 - mis on muutunud ja miks?

Ilona Tragel, Kati Käpp, Marri-Mariska Tammepõld, Eleriin Miilman, Nele Karolin Teiva (Tartu Ülikool, Eesti)

Kirjutamine on kõrghariduse omandamisel üks olulisi aspekte, millega üliõpilased pea iga päev õppetöös kokku puutuvad. Samas tunnevad Euroopa ülikoolide üliõpilased, et neid pole akadeemiliste tekstide kirjutamise jaoks piisavalt ette valmistatud ja neid ei toetata akadeemilise kirjaoskuse omandamise teekonnal piisavalt (Kruse 2003: 21). Seda tõdemust kinnitasid ka Tartu Ülikoolis 2009. aastal läbiviidud küsitluse tulemused (Leijen et al. 2015). Selleks et, mõista paremini üliõpilaste raskusi kirjutamisel ja leida nende murekohtade võimalikke lahendusi, viime 2009. aasta küsitluse eeskujul 2023. aasta kevadel läbi küsitlused Tartu ülikooli õppejõudude ja üliõpilaste seas.



2009. aasta küsitlusega uuriti Tartu Ülikooli õppejõudude ja üliõpilaste kogemusi akadeemiliste tekstide kirjutamisega. Küsitluse eesmärk oli teada saada, milliseid kirjutamisülesandeid õpetamisel kasutati ja kuidas üliõpilased nendega toime tulid. Sealjuures oli oluline ka õppejõudude ja üliõpilaste vastuste võrdlus, mis tõi esile huvitavaid tendentse: näiteks hindasid üliõpilased enda kirjutamisoskust heaks (52% vastustest) või väga heaks (12%), samas kui õppejõud pidasid neid oskusi rahuldavaks (51%) või nõrgaks (29%). (Leijen et al. 2015)

2023. aasta küsitlus on 2009. aasta uuringu järg, mis võimaldab võrrelda, mis on ülikoolis kirjutamisega viimase 14 aasta jooksul muutunud. Arvestades mh vahepealset pandeemiast tingitud distantsõpet ja selle kaasa toodud muutusi hariduses laiemalt, eeldame, et näeme muutusi nii kirjutatavate ülesannete liikides kui ka mahus. Ettekandes anname ülevaate õppejõudude küsitluse tulemustest ning tendentside muutumisest võrreldes 2009. aastaga.

Kirjandus

- Leijen, Djuddah; Jürine, Anni; Tragel, Ilona 2015. University Teachers and Students' Perspectives on Academic Writing: a Case From a University in Estonia. EDULEARN15 Proceedings. 7768–7776.
- Kruse, O. 2003. Getting Started: Academic Writing in the First Year of a University Education. In Björk, L., Bräuer, G., Rienecker, L., Jörgensen, P.S. (eds) Teaching Academic Writing in European Higher Education. Studies in Writing, vol 12. Springer, Dordrecht. https://doi.org/10.1007/0-306-48195-2_2

Development of the Cross-Linguistic Lexical Tasks in Estonian

Adele Vaks, Virve Vihman, Marika Padrik (University of Tartu, Estonia)

The assessment of bilingual children's language skills is often held back by a lack of suitable assessment tools. Norms based on monolingual development are not applicable to bilinguals; whenever possible, both of a bilingual child's languages should be assessed. However, speech and language therapists commonly do not have access to tests that would yield comparable data on both languages. Lack of tests and relevant knowledge contribute to both under- and over-diagnosis of developmental language disorder among bilingual children. To address these gaps, we are adapting the LITMUS test battery developed by the international COST action IS0804 (bi-sli.org) for the assessment of Estonian-speaking bi- and multilingual children's language skills. Several tests in the battery have been adapted for use in Estonian, including the Sentence Repetition Task (Padrik, Vihman, Fil 2022), Nonword Repetition Task



(Pree 2021) and Multilingual Assessment Instrument for Narratives (Argus, Kütt 2020). This talk presents an overview of the vocabulary assessment currently being adapted for Estonian.

Vocabulary size is an important indicator of linguistic development, with effects on reading comprehension, and consequently, academic success (Biemiller 2003). Comparable assessment tools in both languages are needed to get a comprehensive picture of bilingual children's vocabulary knowledge (Hoff et al. 2012). To this end, an Estonian version of the Cross-Linguistic Lexical Tasks (CLTs) is being developed for the KaLaKe (*Kakskeelsete Laste Keeletest*) toolbox for use by Estonian speech and language therapists and others. The tasks, assessing noun and verb comprehension and production, are developed for use with children aged 5 to 6.

This presentation will introduce the principles underlying the CLTs, the process of creating an Estonian version, and the next steps in norming the tasks. To ensure the tasks are cross-linguistically comparable, the target words in each task are chosen according to the same criteria (Haman, Łuniewska, Pomiechowska 2015). Three preliminary studies were conducted with adult speakers to identify the most suitable target words in Estonian: a picture naming study, an age of acquisition study and an assessment of morphological complexity. The resulting tasks have been tested in a pilot study with four groups: typically developing mono- and bilinguals and mono- and bilinguals with developmental language disorder. We will give an overview of the results of this pilot study, and describe the next steps in norming the tasks and making them available for speech and language therapists and researchers alike.

Parlamendikorpused ja mis neist leida võib

Kadri Vider¹, Neeme Kahusk², Martin Mölder² (¹Eesti Kirjandusmuuseum, Eesti. ²Tartu Ülikool, Eesti)

Eesti parlamendikorpuste ajalugu on suhteliselt pikk. Juba eesti keele segakorpuse koosseisus KORPis leidub alamkorpusedena parlamendikorpus aastatest 1995-2001, kokku 14 milj sõnet.

ParlaMint on CLARINi projekt, mille eesmärk on luua ühtse märgendusega mitmekeelsed parlamendiistungite korpused. Eesti meeskond liitus projektiga teises etapis, kus XML-skeem ja enamused tööriistu olid juba välja töötatud. ParlaMint-2 projekt on kokku kogunud 31 parlamendi istungite stenogrammid aastatest 2015-2022.



Eesti ParlaMinti korpus koosneb praegu täielikust riigikogu stenogrammide kogust aastatest 2011-2022 (8 koosseisu, 1283 erinevat kõnelejat, rohkem kui 50 milj sõnet)

Parlamendikorpuse meta-infos sisaldub kõneleja staatus (saadik, valitsuse liige, külaline), sugu, sünnikuupäev, fraktsioon, valimisinfo (nimekiri, mandaat, valimisringkond, häälte arv).

Stenogrammide tekstid on märgendatud automaatselt mitmes kihis: lemma, sõnaliik, vormianalüüs, lauseanalüüs, nimeüksused (isik, koht, organisatsioon). Iga tekst sisaldab ka linki Riigikogu kodulehel olevale originaalstenogrammile.

Korpusi säilitatakse ja levitatakse TEI vormingus. TEI võimaldab märgendada väga erinevaid ilminguid. Selleks, et andmestik oleks teisendatav muudele kujudele (conllu, vert failid otsingusüsteemide tarvis) ja paremini võrreldav, kohandati juba varem parlamendikõnede märgendamiseks kasutatud Parla-CLARINi soovitusi, muutes skeemi rangemaks.

Sel moel meta-andmetega rikastatud ja analüüsitud korpusel on mitmekesised kasutusvõimalused.

Tüüpiliselt on parlamendikorpused huvitanud lingviste (vormianalüüs, lauseanalüüs, kõnekujundid, anafoorid) ja diskursuse-uurijaid (debattide struktuur, teema ja kommentaar, üleminekud, meelsusanalüüs). Kasu võib olla ka keeletehnoloogiliste vahendite treenimisel ja info ammutamisel (nimeüksused, teemade kategoriseerimine, võtmesõnade noppimine jne). Valdcondlikus võtmes on parlamendikorpused kui andmekogud aga huvipakkuvad nii politoloogidele kui laiemalt ühiskonnateadlastele.

Oma posterettekandes tutvustame mõningaid selliste korpuste stiili ja sisu analüüsimise võimalusi, näiteks esimesi tulemusi meelsuse ja populismi klassifitseerimismudelite kohta, mis on spetsiaalselt treenitud poliitiliste tekstide jaoks, ning uurime, kuidas teemamudeleid saab kasutada parteisüsteemi arengu jälgimiseks parlamendikõnede analüüsimise teel.



Konverentsi korraldavad:

Eesti Rakenduslingvistika Ühing ja Eesti Keele Instituut

Konverentsi toetab:

Haridus- ja Teadusministeerium