

24.-25. aprill 2025 / 24-25 April 2025

22. rakenduslingvistika konverents  
**(DE)STANDARDISEERIMINE  
KEELES JA KASUTUSES**



**TEESID / ABSTRACTS**

EAAL 22th Annual Conference

**(DE)STANDARDIZATION IN  
LANGUAGE AND USAGE**

Konverents kuulub eestikeelse raamatu 500. aasta  
juubeliürituste sarja.



## Sisukord

.....	1
Plenaarettekanded .....	5
Vilets keel, vilets meel. ÕSiga või ÕSita.....	5
Margit Langemets (Eesti Keele Instituut, Eesti)	
Standardization and prescriptivism in Dutch: past, present and future .....	6
Rik Vosters (Vrije Universiteit Brussel, Belgium)	
Linguistic justice in schools: deficit thinking and anti-deficit resistance.....	7
Ian Cushing (Manchester Metropolitan University, United Kingdom)	
Suulised ettekanded.....	8
Tehisintellekti rakendamine riigikaitseterminoloogia arendamisel.....	8
Eleri Aedmaa <sup>1</sup> , Reet Hendrikson <sup>2</sup> , Sandra Eiche <sup>1</sup> , Esta Prangel <sup>1</sup> , Mari-Liis Tikerperi <sup>2</sup> (¹Eesti Keele Instituut, Eesti, ²Kaitseväe Akadeemia, Eesti)	
Noorteele sõnavara – kelle meelest <i>verinaiss</i> , kelle meelest <i>sus</i> ? .....	9
Kristel Algvere, Mari-Liis Korkus, Virve-Anneli Vihman (Tartu Ülikool, Eesti)	
If-you and interpersonal meaning in women’s instructive texts (1700-1899).....	10
Francisco Alonso-Almeida, Francisco J. Álvarez-Gil, and Elena Quintana-Toledo (Universidad de Las Palmas de Gran Canaria, Spain)	
Eestikeelse laste kõnetuvastuse arendamine: sektoritevaheline koostöö ning tulevikuvõimalused .....	11
Tanel Alumäe <sup>1</sup> , Kelly Lilles <sup>2,3</sup> , Maarja Männik <sup>3</sup> (¹Tallinna Tehnikaülikool, Eesti, ²Eesti Keele Instituut, Eesti, ³ALPA Kids, Eesti)	
Keelemudelite ja keeletehnoloogia vahendite kasutamine eesti kui emakeele kirjandite hindamisel .....	12
Merilin Aruvee, Kais Allkivi, Andres Karjus, Krister Kruusmaa, Katarin Leppik, Silvia Maine, Taavi Kamarik, Harli Kodasma (Tallinna Ülikool, Eesti)	
Eestikeelse subjektiivsuse andmestiku arendamine.....	13
Karl Gustav Gailit, Kairit Sirts, Kadri Muischnek (Tartu Ülikool, Eesti)	
Miks õppejõud arvab, et teeb TI-I ja üliõpilasel vahet? .....	14
Helen Hint, Ann Veismann, Miina Norvik (Tartu Ülikool, Eesti)	
Ajaloaliste tekstide tärgtuvastusest suurte keelemudelitega Anton Thor Helle 1732. a sõnastiku näitel.....	15
Madis Jürviste (Eesti Keele Instituut & Tartu Ülikool), Joonatan Jakobson	
Digikratid avavad vana kirjakeelt: suurtest keelemudelitest 17.–18. sajandi sõnaraamatute uurimisel .....	16
Madis Jürviste <sup>1,2</sup> , Tiina Paet <sup>1</sup> , Sven-Erik Soosaar <sup>1</sup> (¹Eesti Keele Instituut, Eesti, ² Tartu Ülikool, Eesti)	
Kes kardab virisevat plikat? Seksistlik keelekasutus eesti meedia- ja uue meedia keeles.17	
Elisabeth Kaukonen, Raili Marling (Tartu ülikool, Eesti)	
Frequency types in lexicography: A practical comparison .....	19
František Kovařík <sup>1</sup> , Vojtěch Kovář <sup>2</sup> , Marek Blahuš <sup>2</sup> (¹Masaryk University in Brno, Czechia, ²Lexical Computing CZ, Czechia)	



Tsitaat- ja laensõnad käändumatu täiendina .....	19
Tiina Leemets (Eesti Keele Instituut, Eesti)	
Põhilise kohakonstruktsiooni vormist õppijakeeles .....	20
Liina Lutsepp (Tallinna Ülikool)	
Eesti keele sõnavaratest LexEst .....	21
Kaidi Lõo <sup>1</sup> , Katrin Leppik <sup>1</sup> , Anton Malmi <sup>1,2</sup> , Agu Bleive <sup>1</sup> , Raymond Bertram <sup>3</sup> (1Tartu Ülikool, Eesti, 2Lancasteri Ülikool, Suurbritannia, 3Turu Ülikool, Soome)	
Eesti keele puudepanga semantilisest märgendusest .....	22
Kadri Muischnek, Kaili Müürisep, Andriela Rääbis (Tartu Ülikool, Eesti)	
Miks, mida ja kuidas (de)standardi(seeri)da: võõrsõnade ortograafiast ja morfoloogiast ...	23
Tiina Paet (Eesti Keele Instituut)	
Mõnda ÕSi selgitustest ja soovitustest õigekeelsusallikais täna .....	24
Tiina Paet, Tuuli Rehemaa, Sirli Zupping (Eesti Keele Instituut)	
Maintaining and Challenging the Standard: Grassroots Online Public Language Policing as a Biopower in Finnish Language Discussion Groups .....	25
Henni Pajunen (University of Turku, Finland)	
Üles ja ülesse varieerumine kõnes .....	26
Margetta Pajuste, Miia Malm, Brigitte Pärn, Taavi Terasmaa, Annika Kängsepp, Maarja-Liisa Pilvik, Liina Lindström (Tartu Ülikool)	
Sõnarõhk keeleinfos ja kõnesünteesis .....	27
Liis Piits, Heete Sahkai, Meelis Mihkla, Indrek Hein, Liis Ermus, Hille Pajupuu (Eesti Keele Instituut, Eesti)	
Milline keelekasutus eestlasi häirib? .....	28
Maarja-Liisa Pilvik, Liina Lindström, Kristel Algvere, Lydia Risberg, Helen Plado, Annika Kängsepp (Tartu Ülikool)	
Koanimed uues ÕSis: mida ja kuidas normida .....	29
Peeter Päll (Eesti Keele Instituut)	
Normid ja tegelikkus: terminikasutuse varieeruvus ja standardiseerimine COVID-19 näitel .....	30
Mai Raet (Tallinna Ülikool, Eesti)	
Ajamäärsõnadel <i>koheselt</i> ja <i>kohe</i> pole mingit vahet! Kas tõesti? .....	30
Lydia Risberg <sup>1,2</sup> , Maarja-Liisa Pilvik <sup>2</sup> , Liina Lindström <sup>2</sup> (1Eesti Keele Instituut, Eesti, 2Tartu Ülikool, Eesti)	
(De)standardiseerides kõnekeelt sõnaraamatus. Kuidas rakendada keelekorpus ja suuri keelemudeleid? .....	32
Lydia Risberg <sup>1,2</sup> , Maria Tuulik <sup>1</sup> , Margit Langemets <sup>1</sup> , Ene Vainik <sup>1</sup> , Eleri Aedmaa <sup>1</sup> , Esta Prangel <sup>1</sup> , Kristina Koppel <sup>1</sup> , Hanna Pook <sup>1</sup> , Sirli Zupping <sup>1</sup> (1Eesti Keele Instituut, Eesti, 2Tartu Ülikool, Eesti)	
Partneriparandused telesarja „Õnne 13“ telefonikõnedes .....	33
Andriela Rääbis (Tartu Ülikool)	
Rääkimistesti hindega korreleeruvad prosoodilised tunnused eesti keele kui teise keele kõnelejate kõnes .....	34
Heete Sahkai (Eesti Keele Instituut, Eesti)	



Tehisintellekti (TI) rakenduste kasutamine kirjutamisoskuse õpetamisel Eesti gümnaasiumi inglise keele õpetajate seas .....	36
Anneli Sigus, Liina Tammekänd (Tartu Ülikool, Eesti)	
Morphological ambiguity of Lithuanian uninflected parts of speech .....	38
Miglė Žemrietė (Lithuanian Language institute, Lithuania)	
Akadeemilise teksti tulevik .....	39
Ilona Tragel, Eleriin Miilman, Nele Karolin Teiva, Nele Nove), Marri-Mariska Tammepõld, Laura Kiik, Reena Roos (Tartu Ülikool, Eesti)	
Kui hästi seletavad suured keelemudelid leksikograafide hinnangul eesti sõnade tähendusi? .....	41
Maria Tuulik <sup>1</sup> , Ene Vainik <sup>1</sup> , Esta Prangel <sup>1</sup> , Margit Langemets <sup>1</sup> , Eleri Aedmaa <sup>1</sup> , Kristina Koppel <sup>1</sup> , Lydia Risberg <sup>1,2</sup> ( <sup>1</sup> Eesti Keele Instituut, Eesti, <sup>2</sup> Tartu Ülikool, Eesti)	
Sõnastik – INIMENE ja MAHUTI .....	42
Ene Vainik <sup>1</sup> , Geda Paulsen <sup>1,2</sup> ( <sup>1</sup> Eesti Keele Instituut, Eesti; <sup>2</sup> Uppsala ülikool, Rootsi)	
Workshop of multilingualism and multilingual acquisition .....	44
Estonian language proficiency of Ukrainian and Russian speaking 3rd grade students ....	44
Reili Argus, Piret Baird, Andra Kütt-Leedis, Evelyn Neudorf, Tiina Rütmaa, Anne Uusen (Tallinna Ülikool)	
Linguistic gender stereotypes in Russian and their impact on politeness perception.....	45
Ilenia Del Popolo Marchitto (Tallinn University, Estonia)	
' <i>Kak eto po-russki?</i> ' Transliteration in Russian-medium online communication across the Baltic states and formation of local standards .....	46
Kapitolina Fedorova, Natalia Tšuiкина and Alexandra Lapa (Tallinn University, Estonia)	
Written Language Characteristics in Dyslexia: A Case Study .....	47
Eglė Krivickaitė-Leišienė, Viktorija Kavaliauskaitė-Vilkinienė (Vytautas Magnus university, Lithuania)	
<i>Feuer engine, wrinkelig, and sheepe</i> – The Role of analogy in child bilingual morphological code-mixing .....	48
Antje Quick <sup>1</sup> , Stefan Hartmann <sup>2</sup> , Nikolas Koch <sup>3</sup> / <sup>1</sup> Leipzig University, Germany, <sup>2</sup> HHU Düsseldorf, Germany, <sup>3</sup> LMU Munich, Germany	
Morphosyntactic Complexity in Lithuanian Noun-Adjective Agreement: Insights from Bilingual Children and Non-Native Adults .....	49
Jonas Stepšys, Laura Kamandulytė-Merfeldienė (Vytautas Magnus University)	
The use of modal verbs and adverbs by Estonian school-leavers in their English state examination essays .....	50
Reeli Torn-Leesik, Liina Tammekänd (University of Tartu, Estonia)	
Towards a typology of mixed copies.....	51
Anna Verschik, Helin Kask (Tallinn University, Estonia)	



## Plenaarettekanded

### Vilets keel, vilets meel. ÕSiga või ÕSita

**Margit Langemets (Eesti Keele Instituut, Eesti)**

Eesti keele ja sõnaraamatute ajaloos on olnud mitmeid etappe. Esimene "Eesti keele õigekirjutuse-sõnaraamat" (ÕS 1918, 20 000 sõna) keskendus Dudeni "Ortographisches Wörterbuch'i" (1880, 9. tr. 1915) eeskujul õigekirjale, et aidata ühtlustada ja kinnistada keelenormi. Järgmine "Eesti õigekeelsussõnaraamat ÕS 2025" on kirjakeele normi alus alates 1. jaanuarist 2026.

Eesti keelekorraldus on olnud tugevasti mõjutatud Euroopas 19. sajandil valitsenud standardkeele ideoloogiast, mille juured ulatuvad 1635. a loodud Prantsuse Akadeemiani. Pooleteist sajandi vältel on keelekorralduse põhimõtted lainetanud koos oma ajaga: on rõhutatud süsteemi ja sihipärasust (1920. aastad), eelistatud rangeid norme (1950-1970-ndad), siis pigem suunamist ja soovitamist (alates 2000-ndatest). Tegelikku kasutust ja keele *uurimist* on ÕSide puhul eri aegadel esile tõstetud eri mahus (1930-ndatel vähem, 2020-ndatel rohkem).

Ettekandes käsitlen järgmise ÕSi koostamisega seotud küsimusi.



## Standardization and prescriptivism in Dutch: past, present and future

Rik Vosters (Vrije Universiteit Brussel, Belgium)

In this presentation, I will delve into the history of standardization in Dutch, and use the findings of our historical investigations as a basis for reflection on the present and future of standardization practice. I will begin with some theoretical reflections on standardization from a historical sociolinguistic perspective (Auer et al., 2015), drawing mostly on Joseph et al. (2020) and Rutten & Vosters (2021). Next, I will discuss several myths concerning Dutch language, particularly regarding the form and status of Dutch in the Southern Low Countries (present-day Belgium), and debunk these myths based on recent research (e.g. Rutten & Vosters, 2016, 2023; Vosters et al., 2010). In particular, I will focus on prescriptive efforts in language history, systematically comparing prescribed norms with linguistic practices, and concluding with some thoughts on the complex and layered relationship between norms and usage (cf. Lismont et al., 2023). Based on this historical overview, I will then end by offering some thoughts for the present and future of standardization, prescriptivism and language advice in Dutch and for languages with a similar sociolinguistic profile more generally.

### References

- Auer, A., Peersman, C., Pickl, S., Rutten, G., & Vosters, R. (2015). Historical sociolinguistics: The field and its future. *Journal of Historical Sociolinguistics*, 1(1), 1–12.
- Joseph, J. E., Rutten, G., & Vosters, R. (2020). Dialect, language, nation: Fifty years on. *Language Policy*, 19(2), 161–182.
- Lismont, E., Rutten, G., & Vosters, R. (2023). Dutch prescriptivism in a historical-sociolinguistic perspective. Measuring the effect of institutionalized prescriptivism. In J. Beal, M. Lukač, & R. Straaijer (Eds.), *The Routledge Handbook of Linguistic Prescriptivism* (pp. 447–463). Routledge. 10.4324/9781003095125-29
- Rutten, G., & Vosters, R. (2016). Prescriptivism between the devil and the deep blue sea: Competing language norms in the Southern Low Countries (1815–1830). In I. Tiekens-Boon van Ostade & C. Percy (Eds.), *Prescription and tradition in language. Establishing standards across time and space* (pp. 137–151). Multilingual matters.
- Rutten, G., & Vosters, R. (2021). Language standardization ‘from above.’ In W. Ayres-Bennett & J. Bellamy (Eds.), *The Cambridge Handbook of Language Standardization* (pp. 65–92). Cambridge University Press.
- Rutten, G., & Vosters, R. (2023). Testing Frenchification. A sociolinguistic analysis of French loan morphology in seventeenth- and eighteenth-century Dutch. *Roczniki Humanistyczne*, 71(6), 225–248.
- Vosters, R., Rutten, G., & Van der Wal, M. (2010). Mythes op de pijnbank. Naar een herwaardering van de taalsituatie in de Nederlanden in de achttiende en negentiende eeuw. *Verslagen en Mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde*, 120(1), 93–112.



## **Linguistic justice in schools: deficit thinking and anti-deficit resistance**

**Ian Cushing (Manchester Metropolitan University, United Kingdom)**

Deficit thinking is a victim-blaming ideology which frames marginalised communities as deficient and in need of corrective interventions –especially about language, and especially in schools. It locates alleged faults within the language marginalised communities and deflects attention away from broader socioeconomic structures of injustice. Whilst deficit thinking about language is pervasive in schools, it has always been met with anti-deficit resistance which has pushed for linguistic justice. In this keynote I document the long histories of deficit thinking about language in England’s schools, and counter this with anti-deficit efforts which have put issues of language stigma in dialogue with social justice struggles more broadly. Using these efforts and extensive fieldwork with teachers as inspiration, I conceptualise a theory of linguistic justice which seeks to dismantle oppressive ideologies about ‘standard’ languages and abolish inherently unjust systems.



## Suulised ettekanded

### Tehisintellekti rakendamine riigikaitseterminoloogia arendamisel

Eleri Aedmaa<sup>1</sup>, Reet Hendrikson<sup>2</sup>, Sandra Eiche<sup>1</sup>, Esta Prangel<sup>1</sup>, Mari-Liis Tikerperi<sup>2</sup>  
(<sup>1</sup>Eesti Keele Instituut, Eesti, <sup>2</sup>Kaitseväe Akadeemia, Eesti)

Terminoloogias kui teadusharus ja erialakeele arendamisel on tehisintellektil põhinevate rakenduste laiema kasutuselevõtuga seoses saanud alguse uus ajajärk (vt nt Massion 2024: 92). Terminitöö hõlmab mitmeid etappe, millest üks esimesi on andmemassiivide läbitöötamine ja analüüs. TI-lahendused võivad seda protsessi märkimisväärselt toetada, pakkudes tuge paljudes tööloikudes, sh definitsioonide-kontekstide otsimisel alg- ja/või sihtkeeles, mõistetunnuste selekteerimisel ja definitsioonide algversioonide loomisel sihtkeeles, võimalike sünonüümide leidmisel ja mõistesüsteemide visandamisel, ent ka juba terminibaasis sisalduvate probleemsete kirjete tuvastamisel. Üks võimalikke lahendusi terminitöö tõhustamiseks on välistoelisel genereerimisel (Retrieval-Augmented Generation, RAG) (Lewis jt 2020) põhinevad süsteemid, mis ühendavad generatiivsete keelemudelite võimekuse usaldusväärsete allikate põhjal tehtava teabeotsinguga, võimaldades mudelil anda faktitäpsemaid vastuseid ning vähendada väljamõeldud info hulka. Nende süsteemide kasutamine vähendab käsitööd ja parandab teabe kättesaadavust.

Meie uurimus keskendub riigikaitsevaldkonnale, mille eripära – tehnika ja erialase mõtte kiire arengu tingitud uute mõistete pealetung (Hendrikson 2018) – ning teisteski valdkondades aktuaalsete aspektide, nt keerukate mõistesuhete ja lähimõistete eristamise vajadus teevad perspektiivikaks prototüübi arendusvaldkonnaks. Ühtlasi annab riigikaitsevaldkonna kogemus hea pinnase prototüübi kasutamiseks teistegi erialakeelte puhul, ent miks mitte ka teist tüüpi ülesannete lahendamiseks nii valdkonna sees kui sellest väljaspool. Järgmises arendusetapis plaanime prototüüpi rakendada mõnel külgsuunal erialakeelele.

Ettekandes tutvustame RAG-il põhinevat prototüüpi, mis on välja töötatud riigikaitse terminiarenduse toetamiseks. Käsitleme süsteemi arhitektuuri ja arenduspõhimõtteid, sh andmete eeltötlust ja promptimist, ning valdkonna ekspertide hinnanguid prototüübi tööle. Nende kaudu on võimalik analüüsida süsteemi tugevusi ja piiranguid, sh võimet töödelda erinevat tüüpi dokumente, leida andmetest asjakohaseid definitsioonikandidaate, välja pakkuda seotud termineid ja võimalikke sünonüüme, esitada kontekste ja lisada allikaviited. Need aspektid iseloomustavad loodud süsteemi sobivust terminitöö tõhustamiseks, tuues esile metodoloogilisi puudujääke ning aidates kaardistada edasisi arendusvajadusi.

#### Viited

- Hendrikson, Reet 2018. Kas sõjasõna sünnib sõtta? Erialakeele tõhusus sõjandusterminoloogia näitel. *Dissertationes philologiae estonicae Universitatis Tartuensis* 42. Tartu: Tartu Ülikooli Kirjastus.
- Lewis, Patrick; Perez, Ethan; Piktus, Aleksandra; Petroni, Fabio; Karpukhin, Vladimir; Goyal, Naman;
- Küttler, Heinrich; Lewis, Mike; Yih, Wen-tau; Rocktäschel, Tim; Riedel, Sebastian; Kiela, Douwe 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP tasks. – *Advances in Neural Information Processing Systems* 33, 9459–9474.
- Massion, François 2024. Terminology in the Age of AI: The Transformation of Terminology Theory and Practice. – *Journal of Translation Studies* 4, 67–94.





## Noortekeele sõnavara – kelle meelest *verinaiss*, kelle meelest *sus*?

Kristel Algvere, Mari-Liis Korkus, Virve-Anneli Vihman (Tartu Ülikool, Eesti)

„Vähe sellest, et noored on eesti keele üle külvanud slängiga, me asendame oma igapäevases kõnes varsti juba iga teise sõna võõrkeelest laenatuga. [---] Eesti keele saatus ripub juuksekarva otsas.” See tsitaat pärineb küll 18 aastat tagasi ilmunud arvamusest (Nugis 2007), ent taolisi negatiivseid sõnavõtte kohtab Eesti meediaruumis tänini. Noorte kõnepruuki kirjeldatakse sageli problemaatiliseks, kasutades ühe kõige silmapaistvama mõõdikuna nende sõnavara. Ent nagu Annist ja Kaaristo (2023: 56) Michel Foucault’ile (2005 [1971]) viidates tõdevad, et „noored on ühiskonnas sageli „teine”, keda on vaja kontrollida ja ettekirjutatud raamidesse suruda. Sellepärast esitatakse noori tihtipeale seoses probleemidega.” Ettekandes näitlikustame, et noortekeele sõnavara on loomulik, mitte probleemne.

Stereotüübid on otseselt seotud hoiakutega, keelelised stereotüübid jagunevad staatusel (nt pädev, intelligentne) ja solidaarsusel (nt soe, sõbralik) põhinevate hinnangute vahel (Dragojevic jt 2018: 30). Tundub, et ühiskond ootab noortelt enamasti staatusel põhineva keele ehk standardkeele kasutamist, kuid seesama ebastandardne keelekasutus tähistab noorte omavahelises suhtluses hoopis solidaarsust.

Eesti noortekeele uurimise juured ulatuvad juba peaaegu sajandi tagusesse perioodi. Asjaolu, et noortekeele uurimisega on enim olnud seotud just noored ise, peegeldub arvukates õpilaste sõnavara kirjeldavates (üli)õpilastöödes. (Praakli 2022) Noorte „slängist” hakati rohkem rääkima 20. sajandi lõpus, mil ilmus näiteks Mai Loogi Tallinna keskkooliõpilaste sõnavara uuriv väitekiri, millest kasvas omakorda välja ka „Eesti Esimene Slängi Sõnaraamat” (EESS 1991). Järgnevatel kümnenditel on noorte sõnavara kirjeldusi ilmunud veelgi, seda nii raamatu kui (üli)õpilastöödena, aga ka näiteks mõne ajalehe või ajakirja veergudel (nt ESS 2003; Meister 2008; Libe 2023).

Oma ettekandes tahamegi heita pilgu vähe sisulist tähelepanu saavale noorte omavahelise suhtluse sõnavarale. Tugineme siinkohal õpilaste endi toodud näidetele, mida kogusime noortekeele-teemaliste külalistundide käigus viimase kolme aasta (2022–2025) vältel. Külalistunde viisime läbi eri Eesti koolides ning nendes löid kaasa nii põhikooli- kui gümnaasiumiõpilased. Ühe ülesandena koostasime interaktiivse sõnapilve, milles palusime õpilastel nimetada, *milliseid sõnu ja väljendeid nad kasutavad oma sõprade, aga mitte vanemate või õpetajatega*. Ülesanne kujunes külalistunni üheks menukaimaks osaks, mis tekitas õpilastes alati suurt elevust ja ergutas omavahelist arutelu. Tänu oma atraktiivsusele võimaldas meil koguda hulga eriilmelisi keelenäiteid. Kuigi olemuselt kujutab ülesanne endast sõnavara kogumist, siis selle interaktiivsus võimaldas astuda noortega dialoogi ja peegeldada nende sõnavara üle, asetades seda laiemasse konteksti.

Kogutud keelenäidete toel arutleme, kuidas õpilased kirjeldavad neile eripärast sõnavara (nt pejoratiivid, lühendid, eri keelte väljendid, sisenaljad) ja kuivõrd see erineb standardkeelest. Samuti räägime *slängi* kui mõiste problemaatilisusest ning mõtestame, milliseid piiranguid seab üksiksõnade korje noortekeele terviklikule kirjeldamisele.

### Viited

- EESS 1991 = Loog, Mai. *Esimene Eesti Slängi Sõnaraamat 1991*. Tallinn: Oma kulu ja kirjadega.
- ESS 2003 = Vainola, Kätlin, Lemmit Kaplinski. *Eesti slängi sõnaraamat 2003*. Tallinn: Aule Kirjastus.
- Annist, Aet, Maarja Kaaristo (toim) 2023. *Antropoloogia kujunemine ja uurimislugu: õpik kõrgkoolidele*. Tallinn: Tallinna Ülikooli Kirjastus.
- Dragojevic, Marko, Christofer Berglund, Timothy K. Blauvelt 2018. Figuring Out Who’s Who: The Role of Social Categorization in the Language Attitudes Process. *Journal of Language and Social Psychology* 37(1). 28–50. DOI: 10.1177/0261927X17706942.



- Foucault, Michel 2005 [1971]. *Diskursuse kord. Collège de France'i inauguratsiooniloeng 2. detsembril 1970*. Tlk Marek Tamm. Tallinn: Varrak.
- Libe, Katariina 2023. Klaaser, simpima, slamper - kas sina ikka saad oma teismelise jutust aru? Suur teismeliste sõnavara leksikon! (10.01.2023). *Pere ja Kodu*. Vaadatud 16.03.2025. <https://perekodu.delfi.ee/artikkel/120109418/klaaser-simpima-slamper-kas-sina-ikka-saad-oma-teismelise-jutust-aru-suur-teismeliste-sonavara-leksikon>.
- Meister, Hendrik 2008. Slängi seos 12–19aastaste noorte tundmustega. *Oma Keel* 2. 28–38.
- Nugis, Gerly 2007. Eesti keele saatus ripub juuksekarva otsas (10.03.2007). *Sakala*. Vaadatud 16.03.2025. <https://sakala.postimees.ee/2180971/eesti-keele-saatus-ripub-juuksekarva-otsas>.
- Praakli, Kristiina 2022. Sissevaade noortekeele uurimisse. Üldist ja üksikasjalikku. *Keel ja Kirjandus* 11. 1014–1030. DOI: 10.54013/kk779a3.

## If-you and interpersonal meaning in women's instructive texts (1700-1899)

**Francisco Alonso-Almeida, Francisco J. Álvarez-Gil, and Elena Quintana-Toledo (Universidad de Las Palmas de Gran Canaria, Spain)<sup>1</sup>**

This study investigates mitigation strategies in 18th- and 19th-century English instructive texts written by women, with a particular focus on conditional constructions such as *If you...* These structures, central to the attenuation of directives and recommendations, played a key role in shaping the pragmatic force of instructional discourse. By softening imperatives, they reflect contemporary norms of politeness and social hierarchy, providing insight into the ways female authors negotiated authority within their discursive constraints. The research is based on the COWITE18 (Alonso-Almeida et al. 2025a) and COWITE19 (Alonso-Almeida et al. 2025b) subcorpora, which were specifically compiled to analyze contemporary used of language as well as the rhetorical and pragmatic strategies employed in women's instructive writing. Despite the significance of these constructions in instructional genres, little research has systematically examined their diachronic evolution, particularly from a sociolinguistic and discourse-analytical perspective. Additionally, the broader influence of historical transformations, including industrialization, rising literacy levels, and shifting gender roles, on the frequency and function of mitigation strategies remains largely uncharted.

This study employs both quantitative and qualitative analyses to explore the frequency, distribution, and function of *If you...* constructions in women's instructive texts. The findings reveal a strong reliance on explicit politeness markers in the 18th century (Leech, 2014; Mills, 2003), exemplified by expressions such as *If you please*, which emphasize relational politeness and social deference. Over time, however, a shift towards more pragmatic and direct constructions, such as *If you want*, emerges in the 19th century. This transition aligns with broader sociocultural shifts that encouraged a more functionally driven and accessible communicative style, suited to an expanding and increasingly literate audience.

Beyond their pragmatic function in softening directives (Brown & Levinson, 1987; Alonso-Almeida, 2015), *If you...* constructions serve as indicators of evolving discourse norms and changing social expectations regarding women's textual authority. The comparison of 18th- and 19th-century texts suggests that while earlier constructions foregrounded politeness and indirectness, later examples increasingly reflect a shift toward practicality and reader

---

<sup>1</sup> The research conducted in this paper has been supported by the Agencia Estatal de Investigación, Plan Estatal de Investigación Científica, Técnica y de Innovación 2021–2023, under award number PID2021-125928NB-I00. I hereby express my thanks.



engagement. These patterns underscore how women's instructive writing adapted to ongoing historical conditions, revealing an intricate balance between politeness, persuasion, and rhetorical adaptability in instructional discourse.

## References

- Alonso-Almeida, F., Álvarez-Gil, F., & Ortega-Barrera, I. (2025a). *Corpus of Women's Instructive Texts in English (1700-1799)*. (CoWITE18). University of Las Palmas de Gran Canaria.
- Alonso-Almeida, F., Álvarez-Gil, F., Ortega-Barrera, I., Quintana-Toledo, E., Bator, M., de la Cruz Cabanillas, I., Sánchez-Cuervo, M.E., & Gómez, M.J. (2025b). *Corpus of Women's Instructive Texts in English (1800-1899)*. (CoWITE19). University of Las Palmas de Gran Canaria.
- Alonso-Almeida, F. (2015). On the mitigating function of modality and evidentiality. Evidence from English and Spanish medical research papers. *Intercultural Pragmatics*, 12(1), 33-57. doi:10.1515/ip-2015-0002.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
- Leech, G. N. (2014). *The Pragmatics of Politeness*. Oxford University Press.
- Mills, S. (2003). *Gender and politeness*. Cambridge University Press.

## Eestikeelse laste kõnetuvastuse arendamine: sektoritevaheline koostöö ning tulevikuvõimalused

Tanel Alumäe<sup>1</sup>, Kelly Lilles<sup>2,3</sup>, Maarja Männik<sup>3</sup> (<sup>1</sup>Tallinna Tehnikaülikool, Eesti, <sup>2</sup>Eesti Keele Instituut, Eesti, <sup>3</sup>ALPA Kids, Eesti)

Kõnetuvastus on kiiresti arenev tehnoloogia, kuid laste kõne töötlemine kujutab endast jätkuvalt väljakutset, kuna see erineb akustiliselt ja keeleliselt täiskasvanute kõnest. Käesolev uurimus keskendub eestikeelse laste kõne automaatse transkribeerimise parandamisele ALPA Kidsi digiõppemängudest kogutud poolsontaanse kõne baasil. Vahemikul juuni kuni detsember 2024 koguti üle 20 000 laste kõnega heilifaili. Kõnetuvastuse treenimiseks kasutati 2 745 salvestist 470 kõnelejal, mille kestus oli kokku 28 tundi. Transkribeerimiseks kasutati programmi ELAN (2024).

Eesmärgiks oli trennida OpenAI Whisper kõnetuvastusmudelit, et see suudaks paremini käsitleda laste häälde- ja grammatikavigu ning ignoreerida last assisteeriva lapsevanema kõnet. Üks potentsiaalne strateegia nende väljakutsete lahendamiseks on kasutada suure eeltreenitud mitmekeelse mudeli, näiteks OpenAI Whisper Radford et al. (2023), peenhäälestamist. Näiteks Jain et al. (2024) näitasid, et Whisperi peenhäälestamine laste kõnekorpustel nagu MyST, PFSTAR ja CMU Kids parandab oluliselt kõnetuvastuskvaliteeti võrreldes mudeli nullist treenimise tulemusega. Seda lähenemist on proovitud ka eesti keele puhul: Luhtaru et al. (2023).

Peamised katsed viidi läbi Whisper large-v3 mudeliga, mis on avalikult kättesaadav ja võimaldab seda peenhäälestada enda valitud andmetel. Varasemate tööde põhjal on selgunud, et see mudel sobib väga hästi ka eesti keele tuvastuseks, eriti pärast peenhäälestamist (Sildam et al., 2024). See sobib hästi ka sellisteks juhtudeks, kus eesmärgiks ei ole saavutada täht-täheline transkriptsioon, vaid sisendkõne interpretatsioon (Fedorchenko ja Alumäe, 2025). Mudelit treeniti mitmeastmeliselt: esmalt täiskasvanute kõneandmetega, seejärel eesti koolilaste kõnekorpusega ning lõpuks ALPA Kidsi mängust



pärit laste kõnega. Lisaks rakendati andmete augmenteerimist, mis aitas vähendada veamäära eelkõige nooremate laste puhul.

Eksperimentide tulemusel suudeti vähendada vanusegruppide üleselt sõnavigade määra (WER) 53%-lt 21%-le, kusjuures suurim areng saavutati 3–4aastaste laste kõne tuvastamisel. Eksperimendid näitavad, et uus peenhäälestatud mudel oskab üsna hästi ortograafiliselt transkribeerida lastele omaste hääldus- ja grammatikavigadega poolsponsaanset kõnet ning ignoreerida last assisteeriva lapsevanema kõnet.

Tulevikusuunad hõlmavad pooljuhendatud treenimise rakendamist, mis võimaldaks kasutada transkribeerimata kõnesalvestusi, parandades veelgi kõnetuvastuse täpsust. Samuti võiksid täiendavad eksperimendid keskenduda dialoogi konteksti arvestamisele ja spetsiifiliste vigade automaatsele korrigeerimisele.

Antud uuring panustab rakenduslingvistika valdkonda, pakkudes uusi teadmisi laste kõnetuvastuse optimeerimisest ja laiemalt kõnetehnoloogia kohandamisest vähelevinud keelte jaoks. Lastele suunatud digirakenduste kontekstis on täpne ja usaldusväärne kõnetuvastus oluline vahend keeleõppe toetamiseks ning suhtluspõhiste mängude arendamiseks.

## Viited

- ELAN (Version 6.8) [Computer software]. (2024). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>
- Fedorchenko, A., Alumäe, T. (2025). Optimizing Estonian TV subtitles with semi-supervised learning and LLMs. arXiv preprint arXiv:2501.05234.
- Jain, R., Barcovschi, A., Yiwere, M. Y., Corcoran, P., Cucu, H. (2024). Exploring native and non-native English child speech recognition with Whisper. IEEE Access, 12:41601–41610.
- Luhtaru, A., Jaaska, R., Kruusamäe, K., Fishel, M. (2023). Automatic transcription for Estonian children's speech. In Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa), lk 705–709.
- Meister, E., Meister, L. (2017). Eesti laste kõne I. Põhitooni akustiline analüüs. Keel ja Kirjandus, 60(7):518–533.
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In International conference on machine learning, lk 28492–28518. PMLR.
- Sildam, T., Velve, A., Alumäe, T. (2024). Finetuning end-to-end models for Estonian conversational spoken language translation. In Proceedings of the Seventh Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024), lk 166–174.

## Keelemudelite ja keeletehnoloogia vahendite kasutamine eesti kui emakeele kirjandite hindamisel

**Merilin Aruvee, Kais Allkivi, Andres Karjus, Krister Kruusmaa, Katarin Leppik, Silvia Maine, Taavi Kamarik, Harli Kodasma (Tallinna Ülikool, Eesti)**

2027. aastast rakenduvad Eestis e-eksamid, sh e-kirjand. See toob kaasa vajaduse katsetada kirjutamisoskuse e-hindamise tööriistu. Praegu hindavad eksperthindajad ja õpilaste õpetajad katseeksami kirjandeid kolme aspekti alusel: sisu ja ülesehitus, lausestus ja sõnastus ning korrektsus ja vormistus. Uurisime keeletehnoloogia tööriistade ja generatiivsete keelemudelite kasutusvõimalusi kirjandite hindamise kontekstis ja küsisime: kui võimekad on statistilise



keeletöötuse vahendid ja generatiivsed keelemudelid eesti keele kirjandite hindamisel? Lisaks soovisime teada, kuidas erinevad inimhindajate ja e-tööriistade hinded.

Haridus- ja Noorteametilt saadud ligipääs võimaldas analüüsida 2024. aasta e-eksami kirjandeid, valimis on 9. klassi tekste 795 ja 12. klassi tekste 764. Analüüsimisel kasutasime korpusanalüüsi, juhendatud masinõpet ja suuri keelemudeleid, mille abil hindasime tekste inimhindajatega samade kriteeriumite alusel.

Tulemused näitavad, et automatiseeritud hindamine oli ekspertide hinnetest järjepidevam. Keelemudelid andsid stabiilseid ja võrreldavaid hinnanguid eelneva spetsiaalse treeninguga. 100% GPT-põhistest hinnetest jäid ekspertide antud vahemikku. Kui keelemudelid osutasid tõhusamaks sisu ja struktuuri hindamisel, oli statistiline lähenemine täpsem õigekirja ja vormistuse hindamisel.

## Viited

- Allkivi, K., Eslon, P., Kamarik, T., Kert, K., Kippar, J., Kodasma, H., Maine, S., Norak, K. (2024). ELLE – Estonian Language Learning and Analysis Environment. *Baltic Journal of Modern Computing*, 12(4), 560–569.
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (lk 101–108). <https://doi.org/10.18653/v1/2020.acl-demos.14>
- Siiber, M. (2018). Tarkvara lugemiskiiruse arendamiseks. *Magistritöö*. Tartu Ülikool.
- Luhtaru, A., Purason, T., Vainikko, M., Del, M. & Fishel, M. (2024b). To Err Is Human, but Llamas Can Learn It Too. *Findings of the Association for Computational Linguistics: EMNLP 2024* (lk 12466–12481). <https://doi.org/10.18653/v1/2024.findings-emnlp.727>

## Eestikeelse subjektiivsuse andmestiku arendamine

**Karl Gustav Gailit, Kairit Sirts, Kadri Muischnek (Tartu Ülikool, Eesti)**

Subjektiivsus kui arvamuste ja hoiakute avaldamine on levinud teema keeleteaduses ning seda käsitletakse ka keeletehnoloogias, kus subjektiivsuse tajumine annab kasulikku informatsiooni edasiste ülesannete lahendamiseks. Näiteks, meelestatuse analüüsiks on esmalt tarvilik teada, kas tekstis on avaldatud arvamusi, mida saaks hinnata, või mitte. Subjektiivsuse hindamine võib olla ka iseseisev ülesanne, kus keeletehnoloogilisi vahendeid luuakse ja kasutatakse, et uurida subjektiivsust kui keelelist nähtust.

Sageli käsitletakse subjektiivsust kahe sildi piires: subjektiivne, ehk arvamusi ja hoiakuid sisaldav, või objektiivne, ehk fakte sisaldav. Kuid selline lähenemine annab vaid piiratud vaate sellele, kuidas subjektiivsust tajutakse. Peamine probleem seisneb asjaolus, et inimesed tajuvad subjektiivsust eri määradel — kõrvutades kahte subjektiivset teksti on sageli võimalik eristada, et üks neist on subjektiivsem kui teine. Näiteks saab tuua kaks nutitelefoni arvustust, kus üks kritiseerib telefoni akumahtu ning ekraani eristusvõimet ning teine ütleb, et telefon on kole ja ebamugav käes hoida.

Lähtume eeldusest, et tekstide subjektiivsuse määr on erinev ning tekste on võimalik võrrelda nende subjektiivsuse taseme järgi võrrelda. Sellest tulenevalt on võimalik kirjeldada subjektiivsuse taset arvuliselt: mida suurem arv, seda subjektiivsem on tekst. Selline lähenemine subjektiivsusele ei ole aga nii levinud, kui märgendamine kahe sildiga subjektiivseks ja objektiivseks. Kuigi Likerti skaalat on varasemalt kasutatud subjektiivsuse hindamiseks (Escoufflaire jt 2024, Wiebe 2000), käsitletakse nendes uuringutes vaid



ajakirjandust. Seega, et analüüsida subjektiivsust arvuliselt üle žanrite, on esmalt vaja luua sellist informatsiooni sisaldav andmestik.

Ettekandes annan ülevaate sellisest eestikeelsest subjektiivsuse andmestikust. Andmestik koosneb 1000 Ühendkorpusest (Ühendkorpus 2021) pärinevast tekstist, mis varieeruvad oma standardiseeritud keelekasutuse astmelt, koosnedes 300-st ajakirjandustekstist, millest 150 on uudised ja 150 arvamused, ning 700-st juhuslikult valitud veebitekstist, sealhulgas sotsiaalmeedia tekstid, tootekirjeldused ning arvustused. Iga tekst on saanud kolmelt märgendajalt arvulise hinnangu vahemikus null kuni sada, kus null tähistab täielikku objektiivsust ning sada täielikku subjektiivsust.

Lisaks kirjeldan ettekandes neid hinnanguid. Võrdlen, kuidas hinnangud varieeruvad märgendajate vahel olenevalt tekstiliigist ning teistest aspektidest, sealhulgas teksti pikkus ja kuidas märgendaja on tajunud oma kindlust teksti hinnates. Annan ka ülevaate kasutades hinnangute alusel arvutatud keskmist väärtust ning analüüsin, kuidas need eelmainitud tegurid mõjutavad teksti tajutud subjektiivsust üldisemalt. Lisaks võrdlen üksikute märgendajate hinnanguid keskmisega ning analüüsin, millised tegurid võivad olla põhjustanud erinevusi nende vahel.

## Viited

Escoufflaire, Louis, Antonin Descampe, Cédric Fairoon 2024. Unveiling Subjectivity in Press Discourse: A Statistical and Qualitative Study of Manually Annotated Articles. *Discours* 34. <https://doi.org/10.4000/12jg5>.

Wiebe, Janyce. 2000. Learning Subjective Adjectives from Corpora. In AAAI/IAAI. <https://aaai.org/papers/00735-AAAI00-113-learning-subjective-adjectives-from-corpora/>.

Ühendkorpus 2021 = Kallas, Jelena, Kristina Koppel 2022. Estonian National Corpus 2021. Center of Estonian Language Resources. <https://doi.org/10.15155/3-00-0000-0000-0000-08D17L>.

## Miks õppejõud arvab, et teeb TI-I ja üliõpilasel vahet?

Helen Hint, Ann Veismann, Miina Norvik (Tartu Ülikool, Eesti)

Tehisintellekt (TI) põhinevate rakenduste ulatuslik levik tavakasutajate seas alates 2022. a lõpust on toonud kaasa põhimõttelisi muutusi nii õpetamises kui ka õppimises. Viimase kahe aasta jooksul on ilmselt enamik õpetajaid ja õppejõude olnud olukorras, kus (üli)õpilase esitatud kirjalik töö on tundunud keeleliselt ja/või sisuliselt kummaline. Ehkki TI kasutust ei ole võimalik täie kindlusega mingil moel tuvastada (Weber-Wulff *et al.* 2023), ei pruugi tahta selliste tööde eest hinnet välja panna, sest TI-ga koostatud tööde esitamine ei ole kooskõlas hea akadeemilise tavaga (Cotton, Cotton, Shipway 2023).

Et õppejõudusid selles olukorras toetada, oleme alates 2024. a sügisest TÜ humanitaarteaduste ja kunstide valdkonnas (HV) kogunud TI-kahtlusega tööde näiteid. Algatusel on kaks eesmärki. Esiteks soovime välja selgitada, mis paneb õppejõudusid kahtlustama TI-kasutamist teksti koostamisel. Keskendume just õppejõudude arvamustele ja kogemustele, sest eesmärk on mõista, kas eri õppejõudude kahtlused tuginevad sarnastele alustele ning kas kerkib esile ka selliseid tunnetuslikke jooni, mis suhteliselt objektiivselt TI kasutamise võivad reeta.

Teine eesmärk on tuvastada esitatud tekstinäidete põhjal peamised keelelised tunnused, mis panevad inimesi kahtlustama TI kasutamist. Neist tunnustest lähtuvalt planeerime koostada



praktilisi õppematerjale akadeemilise eesti keele õppejõududele, kes õpetavad akadeemilist kirjutamist ja selle häid tavasid eri erialade üliõpilastele üle kogu ülikooli.

Koostasime veebipõhise küsitluse, kus palume vastajatel esitada konkreetseid katkendeid nende meelest TI-ga koostatud töödest ning lisada sinna juurde selgitusi, mis paneb nende näidete puhul TI kasutust kahtlustama. Küsitlusega ei koguta vastajate identifitseerimist võimaldavaid andmeid, st nii esitatud üliõpilastöö autor kui näiteid jaganud õppejõud jäävad anonüümseks.

Küsitlusele on seni esitatud 20 näidet koos põhjendustega. Vastajaid on peaaegu kõikidest HV instituutidest. Kuna küsitlus on jätkuvalt avatud, siis koguneb vastajaid pidevalt juurde. Praegu on vastustes kõige rohkem esindatud TÜ eesti ja üldkeeleteaduse instituut, kus õpetatakse mh akadeemilise väljendusoskusega seotud aineid.

Ettekandes analüüsime õppejõudude vastuseid, millest on esiteks näha, et TI kasutusele vihjavad probleemid allikate ja viidetega. Teiseks mainivad õppejõud sageli, et TI-ga koostatud tekstid on “ümmargused”. Selliste tekstide puhul ei esita üliõpilased konkreetseid näidetele või allikatele tuginevaid väiteid kõnealuse nähtuse kohta. Kolmandaks vihjeks TI kasutuse kohta tekstides on n-ö liigne enesekindlus. Õppejõud tajuvad, et alles oma õpingute alguses olevad üliõpilased on hakanud esitama “teadja positsioonilt” kirjutatud tekste, mis ei ole kooskõlas sellega, millise tasemeni on üliõpilased oma õpingute käigus tegelikult jõudnud. Ettekandes vaatame esitatud näidete põhjal lähemalt, millised keelelised tunnused nende kahtluste puhul esile tõusevad. Samuti arutleme, kuidas saaks õppejõud üliõpilasi suunata selliseid probleeme vältima.

## Viited

- Cotton, Debby R. E., Peter A. Cotton, J. Reuben Shipway 2023. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>.
- Weber-Wulff, Debora, Alla Anohina Naumeca, Sonja Bjelobaba, Tomáš Foltýnek, Jean Guerrero Dib, Olumide Popoola, Petr Šigut, Lorna Waddington 2023. Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(26). <https://doi.org/10.1007/s40979-023-00146-z>.

## Ajalooliste tekstide tärgtuvastusest suurte keelemudelitega Anton Thor Helle 1732. a sõnastiku näitel

**Madis Jürviste (Eesti Keele Instituut & Tartu Ülikool), Joonatan Jakobson**

Ajalooliste tekstide tärgtuvastust (OCRi) on seni enamasti tehtud spetsiaaltarkvaraga nagu Transkribus, eScriptorium, Kraken jms. Hea tärgtuvastuse eeldusena on vaja programmi põhjalikult eeltreenida ja luua nn puhas “alusnäidis” (*ground truth*). Mida ulatuslikumalt tuvastusmudelit eeltreenitakse, seda täpsem on tulemus. Suured keelemudelid töötavad selles valdkonnas potentsiaalset läbimurret, võimaldades kvaliteetset väljundit ilma eeltreenimiseta tänu nn *zero-shot* võimekusele. Keelemudelite üldist suutlikkust eesti vanema sõnavara analüüsimisel on käsitletud varasemates katsetes (Jürviste, Paet, Soosaar: ilmumas).

EKKD programmi “Suurte keelemudelite rakendamine leksikograafias: uued võimalused ja väljakutsed” raames oleme teinud katseid eeltreenimata keelemudelite rakendamiseks Anton Thor Helle 1732. aastal ilmunud keeleõpetuse sõnastikuosa tärgtuvastuseks ja andmete struktureerimiseks. Otsustavaks teguriks sai nn *vision*-võimekusega keelemudelite hiljutine kasutuselevõtt, mis võimaldab skaneeritud dokumente senisest märksa tõhusamalt töödelda.



Eelkatsed näitasid, et Anthropicu mudel Claude 3.5 Sonnet suudab lihtsa prompti põhjal luua gooti kirjas tekstiga raamatufailist struktureeritud tabeli, kus lisaks tekstituvastusele on märksõnaartiklite sisu paigutatud asjakohastesse tulpadesse. Võrdlesime erinevaid generatiivseid keelemudeleid (Anthropicu Claude, OpenAI GPT mudelid, Google'i Gemini 2.0, Mistral) ning leidsime, et Claude on eesti 17. ja 18. sajandi gooti kirjas (fraktuuris) trükitud tekstide käsitlemisel teistest mudelitest märkimisväärselt tõhusam. Meie eelkatsete järel avalikustas Anthropic Claude Sonneti versiooni 3.7, millega viisime läbi ulatuslikuma katse kogu Helle sõnastiku digiteerimiseks.

Ettekandes uurime, kui edukalt suudab keelemudel muuta skaneeritud sõnaraamatu teksti struktureeritud ja muudetavaks dokumendiks. Hindame eesti märksõnade, saksa vastete ja väljendite tärgtuvastuse täpsust tähemärkide ja sõnade tasandil (*resp.* CER ja WER), samuti andmete struktureerimise täpsust. Lisaks vaatleme, millistel puhkudel keelemudel kõige sagedamini eksib, millised faktorid mõjutavad tuvastustäpsust ja milliseid väljakutseid on keelemudelil talle antud prompti sisu järgimisega.

Tööprotsessi hõlbustamiseks lõime veebirakenduse Anthonius (Jakobson, Jürviste 2025), mis võimaldas sõnastiku digiteerida ja hallata selle käigus tehtavaid API-päringuid. Kasutaja saab rakenduses üles laadida originaalteksti pildiformaadis leheküljed, misjärel keelemudel tuvastab märksõnaartikli sisu ning soovib struktureeritud esitusviisi JSON-formaadis. Kasutaja saab seejärel optimeerida keelemudelile antavaid juhiseid, mida rakendatakse järjestikku igale sõnaraamatu leheküljele. Viimaks saab kasutaja tuvastatud andmed valideerida ning digiteeritud sõnaraamatu redigeeritavaks dokumendiks eksportida.

Meie lähenemisviisi eelis seni levinud OCR-meetodite ees seisneb eelkõige märkimisväärses aja kokkuhoius, arvestades lisaks tärgtuvastusele ka andmete automaatset järelstruktureerimise võimekust. Kui klassikaline meetod nõuab ulatusliku *ground truth*'i loomist ja vahel ka teksti käsitsi segmenteerimist, siis keelemudelipõhine lähenemine võimaldab saavutada suurepäraseid tulemusi oluliselt väiksema ettevalmistusega. Isegi tasuliste keelemudelite (nagu Claude 3.7 Sonnet) kasutamine on väga kuluefektiivne.

Sama metodoloogiat on võimalik rakendada ka teiste ajalooliste tekstide digiteerimisel, kus seni levinud meetodite kasutamine oleks ajamahukuse tõttu ebapraktiline. See avab uued väljavaated ajaloolise tekstipärandi digiteerimiseks ning loob eeldused vanade tekstiallikate märksa ulatuslikumaks uurimiseks.

## Viited

Helle, Anton Thor 1732. Kurtzgefaszte Anweisung Zur Ehstnischen Sprache. Halle: Stephan Orban. <http://www.digar.ee/id/nlib-digar:100071>

Jakobson, Joonatan; Jürviste, Madis 2025. Veebirakenduse Anthonius repositoorium. <https://github.com/joonatanjak/anthonius> (ligipääs avalik alates 24.04.2025).

Jürviste, Madis; Paet, Tiina; Soosaar, Sven-Erik (ilmumas). Eesti vanade sõnakujude tuvastamise võimalustest suurte keelemudelite abil. – Eesti Rakenduslingvistika Ühingu aastaraamat 21.

## Digikratid avavad vana kirjakeelt: suurtest keelemudelistest 17.–18. sajandi sõnaraamatute uurimisel

Madis Jürviste<sup>1,2</sup>, Tiina Paet<sup>1</sup>, Sven-Erik Soosaar<sup>1</sup> (<sup>1</sup>Eesti Keele Instituut, Eesti, <sup>2</sup> Tartu Ülikool, Eesti)

Ettekandes tutvustame uurimust, mille käigus uurisime OpenAI, Anthropicu ja Google'i suurte keelemudelite võimekust tuvastada eesti vana kirjakeele sõnakujusid. Võrdleva analüüsi





käigus hindasime mudelite tõhusust ajalooliste vastete sidumisel tänapäevaste sõnavormidega, eesmärgiga selgitada keelemudelite rakendatavust sõnastikutöös vajalike tööprotsesside automatiseerimisel. Keskendusime peamiselt küsimusele, kuivõrd aitaksid keelemudelid vähendada sõnaraamatute koostamisele kuluvat aega ja ressursse (vt ka Lew 2023: 2; Jakubíček, Rundell 2023).

Uurimuse tegime EKKD programmi „Suurte keelemudelite rakendamine leksikograafias: uued võimalused ja väljakutsed“ raames. Oma ettekandes tutvustame katseid, mille käigus testisime mudelite suutlikkust lahendada vana kirjakeele sõnavaraga seotud erinevaid ülesandeid. Esitasime mudelitele süstemaatiliselt koostatud API-päringuid, mis hõlmasid 17. ja 18. sajandi sõnastikes (H. Stahl, J. Gutschlaff, H. Göseken, S. H. Vestring, A. T. Helle, A. W. Hupel) esinevaid ametinimetusid, murdekeelendeid ja vaimulikes tekstides kasutatud laensõnu. Testitud mudeleid ei olnud eesti vana sõnavara osas eelnevalt treenitud.

Esiõlgused tulemused näitavad olulisi erinevusi mudelite võimekuses, kusjuures nende täpsus varieerub märkimisväärselt, sõltuvalt neile antud sisendmaterjali keerukusest. Uurimuse käigus tuvastasime mitmeid piiranguid, sealhulgas mudelite raskusi haruldasemate murdekeelendite ja arhailiste vormide käsitlemisel. Katsete tulemused näitavad keelemudelite suurt potentsiaali vanade sõnakujude ühendamisel nüüdiskujudega: seda saab rakendada märksõnade diakroonilisel kirjeldamisel koos viidetega varasematele esinemisaegadele ja -kohtadele.

Tutvustame ka edasisi katseid, mille käigus keskendume vanades sõnastikes sisalduva keelematerjali analüüsimisele ja võrdlevale kõrvutamisele laiendatud mahus. Peamine eesmärk on luua meetodika vana eesti keelt sisaldavate sõnastike varustamiseks nüüdisvastetega. See loob võimaluse integreerida ajaloolist leksikograafilist infot tõhusalt tänapäevastesse keeleressurssidesse. Meie senised uurimistulemused pakuvad väärtusliku sisendi nii keeletehnoloogia kui ka diakroonilise leksikograafia rakendustesse.

## Viited

Jakubíček, Miloš; Rundell, Michael 2023. The End of Lexicography? Can ChatGPT Outperform Current Tools for Post-Editing Lexicography?. – Proceedings of the eLex 2023 Conference: Electronic Lexicography in the 21st Century. Brno: Lexical Computing, lk 508–523.

Jürviste, Madis; Paet, Tiina; Soosaar, Sven-Erik (ilmumas). Eesti vanade sõnakujude tuvastamise võimalustest suurte keelemudelite abil. – Eesti Rakenduslingvistika Ühingu aastaraamat 21.

Lew, Robert 2023. ChatGPT as a COBUILD lexicographer. *Humanit Soc Sci Commun* 10, 704 (2023). <https://doi.org/10.1057/s41599-023-02119-6>.

## Kes kardab virisevat plikat? Seksistlik keelekasutus eesti meedia- ja uue meedia keeles

**Elisabeth Kaukonen, Raili Marling (Tartu ülikool, Eesti)**

Sooküsimuste kontekstis on viimaste aastakümnete jooksul ilmnunud kaks vastandlikku tendentsi: soolise võrdõiguslikkuse põhimõtete üha suurem omaksvõtt tüdrukute ja naiste seas ning konservatiivsemate väärtuste levimine noorte meeste seas (Ella Mebane et al. 2020; Ümarik & Tinn 2023; Burn-Murdoch 2024). Seda näitlikustavad veebikogukonnad ja sotsiaalmeediakanalid, milles keskendutakse traditsiooniliste soorollide tähtsustamisele ning naiste jt vähemusgruppide suhtes vaenu õhutamisele. Kõige selle oluline edasikandja on ka



keel, mis peegeldab, ehitab ja kinnistab ühiskonnas kehtivaid väärtushinnanguid (Talbot 1998; Mills 2008; Cameron 2024).

Ettekandes vaatleme, kuidas levib seksism eesti meedia- ja uue meedia keeles. Täpsemalt keskendume küsimustele, kas juhtivaid Eesti nais- ja meespoliitiku kujutatakse erinevalt; kuidas väljendub nende kujutamises soolistatus, sh seksism, ning kas naispoliitiku erinev kohtlemine ilmneb konkreetsete keelendite kaudu. Materjalina kasutame eesti keele ühendkorpust 2023 ja selle veebitekstide allkorpuseid, seega moodustavad materjali ühelt poolt meediaväljaanded ja teiselt poolt uue meedia tekstid, kommentaariumid ja blogid, mis esindavad kirjalikku argikeelt. Ettekandega tutvustame tulemusi koostöös mõttekoda Praxisega tehtud uuringust “Alfamehed ja lillekesed. Misogüünia ja seksismi normaliseerumine eesti meediapildis” (Kaukonen et al. 2024).

Tulemused näitavad, et kui meediatekstides ilmneb veidi varjatumat soolist diskrimineerimist, siis naispoliitiku otsesõnu halvustamist esineb rohkem uue meedia keeles. Sarnaselt teistes keeltes tehtud uurimustele (vt Wolfson & Manes 1980; Holmes & Sigley 2002; Cameron 2024) ilmnes, et naispoliitiku mahategemist väljendavad ironilised viisakusvormid, nagu kallis, deminutiivide kasutamine ning naise taandamine lapse staatusesse.

## Viited

- Burn-Murdoch, John. 2024. A new global gender divide is emerging. *Financial Times*, 26.01.2024. Veebiaadress: <https://www.ft.com/content/29fd9b5c-2f35-41bf-9d4c-994db4e12998>
- Cameron, Deborah. 2024. *Language, sexism and misogyny*. London/New York: Routledge
- Ella Mebane, Minou, Antonio Aiello & Donata Francescato. 2020. Political Gender Gap and Social Dominance Orientation. In Dhastagir Sultan Sheriff (ed.), *Psycho-Social Aspects of Human Sexuality and Ethics*. IntechOpen. <https://doi.org/10.5772/intechopen.92222>
- Holmes, Janet & Robert Sigley. 2002. What's a word like girl doing in a place like this? Occupational labels, sexist usages and corpus research. *New Frontiers of Corpus Research*, 247–263
- Kaukonen, Elisabeth, Kaarel Lott, Raili Marling, Maria Murumaa-Mengel, Mari-Liis Sepper, Maarja Tinn. 2024c. *Alfamehed ja lillekesed: misogüünia ja seksismi normaliseerumine Eesti meediapildis*. Tallinn: Mõttekoda Praxis. Veebiaadress: <https://www.praxis.ee/uploads/2024/12/Raport-lillekesed.pdf>
- Mills, Sara. 2008. *Language and Sexism*. Cambridge: Cambridge University Press
- Talbot, Mary. 1998. *Language and Gender. An Introduction*. Malden: Blackwell Publishers Inc.
- Wolfson, Nessa & Joan Manes. 1980. The compliment as a social strategy. *Research on Language and Social Interaction*, 13(3), 391–410. <http://doi.org/10.1080/08351818009370503>
- Ümarik, Meril and Maarja Tinn. 2023. ICCS 2022. aasta uuringu Eesti aruande põhisõnumid. Meril Ümarik & Maarja Tinn (toim.), *Noorte Hää! Demokraatia Kriisis: Kodanikuhariduse Väljakutsed*. IEA 2022. aasta rahvusvahelise kodanikuhariduse uuringu (ICCS 2022) Eesti tulemused, 7–14. Tallinn: Tallinna Ülikool



## Frequency types in lexicography: A practical comparison

František Kovařík<sup>1</sup>, Vojtěch Kovář<sup>2</sup>, Marek Blahuš<sup>2</sup> (<sup>1</sup>Masaryk University in Brno, Czechia, <sup>2</sup>Lexical Computing CZ, Czechia)

A series of ongoing projects called Dictionary Express (DE) focuses on creating dictionaries using corpus data and manual annotation of these data by native speakers. Each project divides the dictionary-making process into simple tasks like selecting proper headwords or proper word forms for each headword. Our research concentrates on the data from the headwords selection task from the Czech DE project.

In this task, the annotators go through a list of headwords collected from a document frequency wordlist from a web corpus and state if a word is correct, incorrect or has a specific linguistic problem (typo, incorrect lemmatization, etc.).

Our research focuses on types of headword frequency in a corpus and their relation to the "correctness" of the headwords. There are four main corpus frequency types (absolute frequency, relative frequency, ALDF and ARF), each with a unique strategy to overcome word burstiness. We have compared the wordlists of each of these frequency types with the list of words marked as correct by the Dictionary Express annotators. The wordlists consisted of the top X most frequent headwords, where X is 10,000, 50,000, 80,000 and 100,000. We have also examined the words after the 100,000 threshold. We have taken a look at the wordlists as a whole but also examined them part by part and determined whether the "frequency-correctness" ratios differ type by type. To back this with proper data, we look deeper into the "correctness" quality, the dangers of this simplified approach to lexicographic data, and how the DE projects handle these problematic micro-decisions.

We have found that the wordlists of different frequency types are similar in terms of the first 100,000 headwords, with the biggest difference between absolute frequency and the other types. For less frequent words however, selecting the right type of frequency could make a bigger impact on vocabulary building. In our presentation, we introduce the biggest differences and other examples and findings regarding the frequency types and their relation to the headword annotations in Czech Dictionary Express.

## Tsitaat- ja laensõnad käändumatu täiendina

Tiina Leemets (Eesti Keele Instituut, Eesti)

Rühm uuemaid tsitaatsõnu (nt *online*, *offshore*, *tax-free*, *unisex*, *halal*) ning laensõnu (*vegan*, *koššer*, *super*) võib kasutuses esineda adjektiivi funktsioonis nimetavakujulise lahku kirjutatud käändumatu täiendina. Osa neist käitub lisaks adverbi või substantiivina.

Eesti omadussõnad üldjuhul ühilduvad täiendina põhisõnaga, seetõttu on vaadeldavaid sõnu õigekirjas käsitletud prefiksisarnaste muutumatute tüvedena, mis tuleks põhisõnaga kokku kirjutada (tsitaatsõnad sidekriipsuga).

Varem on käändumatu täiendina olnud käibel *nud-* ja *tud-*kesksõnad ja *mata-*vormid ning omastava käände kujud: *katoliku*, *gooti*, *prantsuse*, *soome-ugri* jt. On ka erandlikke vaegomadussõnu, nt *päris*, *-võitu*, *eht*.

Ühendsõnastiku ja ÕSi koostajad on otsustanud 28.03.2025 anda ESi keeleteimkonnale arutada küsimuse, kas mõnda oletatavasti kasvava rühma tsitaat- ja võõrsõnade puhul ka lahkukirjutust nimetavakujulise käändumatu omadussõnana. Konverentsil saaks tutvustada selle arutluse tulemust.



## Viited

EKI teatmik. Eesti õigekeelsuskäsiraamat. Toimetanud Peeter Päll (peatoimetaja), Tiina Paet, Margit Langemets. Koostanud Peeter Päll, Maire Raadik, Tuuli Rehemaa, Margit Langemets, Tiina Leemets, Sirje Mäearu, Tiina Paet, Lydia Risberg, Arvi Tavast. Eesti Keele Instituut 2022–. eki.ee/teatmik/ (vaadatud 14.03.2025).

Kiik, Johanna. Nelja uusklassikalise keelendi kohanemisest eesti keeles. – Keel ja Kirjandus 2021, 4, 301–317. [https://keeljakirjandus.ee/wp-content/uploads/2021/03/J.KIIK\\_.pdf](https://keeljakirjandus.ee/wp-content/uploads/2021/03/J.KIIK_.pdf)

## Põhilise kohakonstruktsiooni vormist õppijakeeles

Liina Lutsepp (Tallinna Ülikool)

Ruumikeel (*space language*) hõlmab kolme põhilist valdkonda: topoloogilist kirjeldamist, liikumise kirjeldamist ja kirjeldamist taustsüsteemide abil (*frame of reference*). Stephen C. Levinson ja David Wilkins (2006) on erinevate keelkondade keeli uurides jõudnud järeldusele, et topoloogiliste stseenide kirjeldamiseks kasutatav põhilist kohafunktsiooni väljendav konstruktsioon (ingl *BLC – basic local construction*) võib olla vormistatud keeliti kas kaassõnade, nimisõnalise elemendi, käändelõppude või näiteks kohategasõnade abil.

Levinson ja Wilkins on pakkunud välja keelteülese hierarhia, mis järjestab topoloogilised stseenid, mille puhul põhi kohakonstruktsiooni kasutus ühes stseenis viitab sellele, et seda kasutatakse ka järgmises stseenis. Topoloogilise ruumi implikatiivne hierarhia näeb välja järgmine: elus taust (ing *ring on finger*; eesti *k sõrmus sõrmes*) > läbistatud figuur (*meat on skewer*; *liha vardas*) > läbistatud taust (*arrow in apple*; *nool läbi õuna*) > liimimine (*stamp on envelope*; *mark kirjal*) > tuumstseenid (*cup on table*, *fruit in bowl*, *lamp over table*, *ball under chair*; *tass laual*, *puuvili kausis*, *lamp laua kohal*, *pall tooli all*). Figuuri lähedus ja kontakt taustaga on omadused, mis määravad, kas kasutatakse põhikonstruktsiooni või mõnd muud konstruktsiooni. Tuumstseenides on suhteliselt väike liigutatav objekt suhteliselt suure püsiva tausta vahetus läheduses ja nende kirjeldamiseks kasutatakse suure tõenäosusega põhikonstruktsiooni. Kui neid omadusi muuta, suureneb tõenäosus, et kasutatakse mõnd muud konstruktsiooni, mitte põhikonstruktsiooni.

Kuna eesti keeles on põhilise kohakonstruktsiooni vormistamiseks kaks võimalust – taust võib selles konstruktsioonis olla nii NP kui ka PP –, on eesti keele õppijail võimalik valida üks või teine fraas. Agu Bleive (2023) on uurinud, kuidas hiina emakeelega eesti keele õppijad topoloogilisi stseene kirjeldavad, ja jõudnud selleni, et nad eelistavad kasutada kohakäändes nimisõnafraasi ja vähem kaassõnafraasi, kuigi hiina keele struktuuri silmas pidades võiks arvata vastupidist.

Eelnenule tuginedes annan ettekandes ülevaate, kuidas eesti keelt teise keelena omandavad 9–10aastased lapsed topoloogilisi stseene kirjeldavad: milliseid keelendeid nad valivad põhi kohakonstruktsiooni vormistamiseks, millisel juhul nad valivad nimisõnafraasi, millisel juhul kaassõnafraasi, millisel juhul mõne muu konstruktsiooni ning kuidas ilmnenud haakub ülal kirjeldatud topoloogilise ruumi hierarhiaga. Selleks analüüsin 2024. aastal projektis „Professionaalne eesti keele õpetaja haridusasutuses“ kogutud keelematerjali: täpsemalt 250 õpilase suulist pildikirjelduskatset, mille käigus õpetaja esitas küsimusi, et vastavat fraasitüüpi või grammatilist kategooriat esile kutsuda. Esialgse tulemusena võib välja tuua, et seesolemissuhteid väljendavad keeleõppijad sisekohakäändes nimisõnafraasiga, toetussuhete korral kalduvad kasutama kaassõnafraasi.



## Kirjandus

- Bleive, Agu (2023). How do Chinese L1 learners express Estonian static spatial relations? *Eesti Rakenduslingvistika Ühingu aastaraamat*, 19, 5–24. <https://doi.org/10.5128/ERYa19.01>
- Levinson, Stephen C., Wilkins, David (2006). *Grammars of Space: Explorations in Cognitive Diversity*. Cambridge University Press.

## Eesti keele sõnavaratest LexEst

**Kaidi Lõo<sup>1</sup>, Katrin Leppik<sup>1</sup>, Anton Malmi<sup>1,2</sup>, Agu Bleive<sup>1</sup>, Raymond Bertram<sup>3</sup>** (<sup>1</sup>Tartu Ülikool, Eesti, <sup>2</sup>Lancasteri Ülikool, Suurbritannia, <sup>3</sup>Turu Ülikool, Soome)

LexEst on 5-minutiline sõnavaratest täiskasvanud eesti keelt teise keelena kõnelejatele. Ettekandes tutvustame LexEsti koostamist ja valideerimist. LexEst on loodud ingliskeelse testi LexTALE (Lemhöfer & Broersma, 2012) eeskujul ning sarnaseid teste on loodud ka mitmete teiste keelte jaoks (nt soome, itaalia, hispaania, hiina). Testi koostamiseks viisime läbi katse, milles osales 485 inimest (274 rääkis eesti keelt emakeelena ja 211 olid vene, saksa, soome jne emakeelega eesti keele õppijad, vanus 18-77 aastat, 348 naist, 129 meest, 8 mittebinaarset).

Katses osalenud täitsid kõigepealt keeletausta küsimustiku ja siis pidid otsustama 135 stiimuli (90 sõna, 45 pseudosõna) puhul, kas tegemist on päris eesti keele sõnadega (Lõo jt 2018). Selleks et jõuda katse lõpliku versioonini, analüüsisime pärast andmete kogumist testsõnavara raskusastet ja eristusvõimet (Rizopoulos, 2006). Jätsime välja 45 kõige lihtsamat stiimulit. Lõplikus LexEsti testi versioonis on 90 stiimulit (60 sõna, 30 pseudosõna). Õigete ja valede vastuste põhjal arvatud punktisumma varieerub -20 ja 60 punkti vahel.

Analüüsi tulemused näitasid, et katses osalejad, kes andsid oma keeleoskusele kõrgema hinnangu, said sõnavaratestis ka kõrgema punktisumma. Samuti said kõrgema punktisumma need, kes olid Eestis kauem elanud, kes olid osalenud kõrgema tasemega eesti keele kursustel ja kes kasutasid igapäevaselt rohkem eesti keelt.

Meie uuringu tulemused näitavad, et LexEst sõnavaratest on hea alternatiiv tavapärasele enesehinnangu küsitlusele, eelkõige just teadusuuringute kontekstis, sest see on kiirem ja objektiivsem. Test, lähtekood, stiimulid ja hindamisjuhend on vabalt kättesaadaval aadressil [lexest.ut.ee](http://lexest.ut.ee).

## Viited

- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44, 325-343.
- Lõo, K., Järvikivi, J., & Baayen, R. H. (2018). Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing. *Cognition*, 175, 20-25.
- Rizopoulos, D. (2006). *ltm*: An R Package for Latent Variable Modeling and Item Response Analysis. *Journal of Statistical Software*, 17(5), 1–25.



## Eesti keele puudepanga semantilisest märgendusest

Kadri Muischnek, Kaili Müürisep, Andriela Rääbis (Tartu Ülikool, Eesti)

Eesti keele puudepanga on arendatud üle 20 aasta (Bick jt 2004, Muischnek jt 2014, Muischnek jt 2016) ja lõpuks on jõutud universaalsete sõltuvuste märgendussüsteemi juurde (de Marneffe jt 2021). Eesti keele puudepang on leitav universaalsete sõltuvuspuude repositooriumist [universaldependencies.org](https://universaldependencies.org).

Kuigi puudepang esitab peamiselt süntaktilist informatsiooni, ei saa kuidagi üle ega ümber semantikast. Paljud keeled on kaasanud märgendusse ka semantilise informatsiooni. Seda saab teha mitmel viisil ja erinevad abstraktsusastmega. Küllalt levinud on määruse liikide täpsustamine süntaktilises märgendis. Kui eesti keele puudepang kasutab nimisõnalise määruse tähistamiseks universaalset 'obl'-märgendit, siis paljude keelte märgendused<sup>2</sup> eristavad aja- ja kohamääruseid ('obl:tmod', 'obl:lmod'), aga ka vahendit, saajat jm. Seda liiki märgendust ja märgendamismeetodeid on eesti keele jaoks ka loodud (näiteks Orasmaa 2014). Suurem väljakutse on laiema tähenduse kaasamine: mis liiki tegevust kirjeldab lause ja kes/mis selles tegevuses osalevad.

Astumaks sammu lause semantilise esituse poole, märgendasime eesti kirjakeele universaalsete sõltuvuste (UD) korpuse EDT semantiliste rollidega vastavalt PropBanki<sup>3</sup> (Palmer jt 2005) märgendusskeemile. Valisime märgendamiseks sagedasemate lihtverbide (kõik sagedasemad kui 14 esinemist 440000-sõnalises korpuses ja suurem osa sagedasemad kui 9 esinemist) argumentidestruktuuri, selle käigus tuli luua ka minimaalne verbi tähenduste jaotus, mis sagedasemate verbide puhul (*saama*, *tulema*, *minema*, *panema* jne) võib koosneda kuni 15 erineva tähenduse eristamisest.

Märgendajate meeskond koosnes 4 inimesest, kuid märgendamine ei toimunud otseselt paralleelselt. Üks inimene koostas märgendamise skeemi, teine kandis märgenduse failidesse ja tegi vajadusel parandusi.

Märgendus on loodud Conllu+ formaadis, mis on teisendatav ka standardsesse Conllu formaati. Märgendus on osaliselt kaasatud 01.11.2024 avaldatud UD versioonis 2.15<sup>4</sup>.

Märgendati 570 erineva verbi 36760 esinemist, nende 848 erinevat tähendust ja nende 57792 argumenti.

Allpool on toodud näide Conllu+ formaadis märgendatud lausest ja tema argumentidestruktuurist. Morfoloogiline märgendus on ülevaatlikkuse huvides eemaldatud.

```
# sent_id = aja_sloleht20071217_941
# text = Olete ehk juhtunud kuulma, kuidas riigikogu nimetatakse pelgalt kummitempliks?
1 Olete          olema          AUX           ... 3 aux          3:aux          _
2 ehk            ehk            ADV           ... 3 advmod       3:advmod       _
3 juhtunud      juhtuma       VERB          ... 0 root         0:root         Verb_juhtuma
4 kuulma        kuulma        VERB          ... 3 xcomp        3:xcomp        juhtuma_Arg_5 Verb_kuulma
5 ,              ,              PUNCT        _ 8 punct        8:punct        _
6 kuidas        kuidas        ADV           _ 8 mark         8:mark         _
7 riigikogu     riigi_kogu    NOUN         ... 8 obj          8:obj          nimetama_Arg_1 _
8 nimetatakse   nimetama      VERB          ... 4 ccomp        4:ccomp        kuulma_Arg_1  Verb_nimetama
9 pelgalt       pelgalt       ADV           _ 8 advmod       8:advmod       _
10 kummitempliks kummi_tempel NOUN         ... 8 xcomp        8:xcomp        nimetama_Arg_2 _
11 ?            ?             PUNCT        _ 3 punct        3:punct        _
```

Tööd oleks veel palju teha: märgendamata jäid ühendverbid, ilmselt tuleb teha muutusi ka verbi tähenduste klassifitseerimisel ning ühendada tähendused PropBanki tähenduste indeksiga, samuti tuleb veel üle vaadata argumentide struktuur. Nii võib lugeda meie semantilist märgendust pigem eeltöökse eesti keele PropBanki loomisel.

<sup>2</sup> <https://universaldependencies.org/ext-dep-index.html>

<sup>3</sup> Proposition Bank, <https://propbank.github.io/>

<sup>4</sup> [https://github.com/UniversalDependencies/UD\\_Estonian-EDT/tree/master](https://github.com/UniversalDependencies/UD_Estonian-EDT/tree/master)



## Viited

- Bick, E., Uibo, H., Müürisep, K. (2004) Arborest—a VISL-style treebank derived from an Estonian Constraint Grammar corpus. *Proceedings of the Third Workshop on Treebanks and Linguistic Theories*, 9-20.
- De Marneffe, M. C., Manning, C. D., Nivre, J., Zeman, D. (2021) Universal dependencies. *Computational Linguistics*, 47(2), 255-308.
- Muischnek, K., Müürisep, K., Puolakainen, T., Aedmaa, E.; Kirt, R.; Särg, D. (2014) Estonian dependency treebank and its annotation scheme. *Proceedings of 13th Workshop on Treebanks and Linguistic Theories*, 285-291.
- Muischnek, K., Müürisep, K., Puolakainen, T. (2016) Estonian dependency treebank: from constraint grammar tagset to universal dependencies. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 1558-1565.
- Orasmaa, S. (2014) Towards an Integration of Syntactic and Temporal Annotations in Estonian. *LREC*, 1259-1266.
- Palmer, M., Gildea, D., Kingsbury, P. (2005) The Proposition Bank: A Corpus Annotated with Semantic Roles. *Computational Linguistics Journal*, 31(1), 2005.

## Miks, mida ja kuidas (de)standardi(seeri)da: võõrsõnade ortograafiast ja morfoloogiast

Tiina Paet (Eesti Keele Instituut)

Tänapäeval esitavad standardkeele ehk normi(ngu) kindla ajavahemiku järel ilmuvad ÕS-id. Olukorras, kus nii eesti kui ka teiste keelte korralduses on toimunud olulisi paradigmuuutusi – räägitakse suuremast varieerumisest ning keelenormide nõrgenemisest, standardite laienemisest ja destandardiseerimisest (vt nt Kristiansen 2021: 667–669), võtan ettekandes lähema vaatluse alla sellel aastal ilmuvas ÕSis võõrsõnade esitamisega seotud küsimused, peamiselt ortograafiat ja morfoloogiat, aga ka oma- ja võõrsõna vahekorda puudutavad. Võõrsõnade normimisega kaasneb spetsiifilisi küsimusi nagu nt ortograafias rööpkujude teke (teiste keelte mõjul) ja morfoloogias segunevad muutmisparadigmad.

Varasemate normingute ja soovitude üle vaatamiseks ja keelekorralduspõhimõtete uuendamiseks ning sõnastamiseks on olnud mitmeid põhjuseid: keele loomulik muutumine, eesti keelekorralduses toimunud pööre kasutuspõhisuse poole, aga ka digitehnoloogiate kasutuselevõtt, mis võimaldab analüüsida ja kasutada varasemast suuremates kogustes keeleandmeid, ent on muutnud ka sõnaraamatute koostamisprotsessi. Viimastel kümnenditel on muutunud dominantkeel ja keelekontaktide iseloom (ingliskeelestumine ja meediastumine) ning keelemuutuste kontekst, mis kõik on suurendanud eesti keelde tuleva võõrsõnavara hulka ja põhjustanud muutusi ka varem tulnud sõnade kujus (vt ka Paet 2023).

Tänapäevasel kujul fikseerus hulk eesti kirjakeele sõnu EKÕSis 1925–1937. Võõrsõnade mugandamise aluspõhimõtted (võõrsõnad kirjutatakse eesti kirjaväikes, sh vältesüsteemi arvestades) pärinevad 1930ndatest, kui need pani paika Elmar Muuk. Iga uus ÕS on siiski muutnud mingi osa võõrsõnade kirjakuju: nihked on toimunud tegelikult keelekasutuses või tehtud keelekorralduspõhimõtete najal kunstlikult. ÕS 2025s võõrsõnade kirjakuju esitamisel lähtutakse uutest EKIs väljatöötatud võõrsõnade normimise ja [mugandamise põhimõtetest](#), mis on läbi arutatud ja heaks kiidetud Emakeele Seltsi keeleteoimkonnas 2024. aastal. Võõrsõnade kirjakuju fikseerimisel arvestatakse senisest enam tegelikku keelekasutust (vt lähemalt "[Keelekorralduse põhimõtteid](#)"). ÕS 2025 üheks uuenduseks on eraldi infokihit (EKI Ühendsõnastikus rubriik "ÕS selgitab"). Eeskätt võõrainest puudutab selles infokihis erisugune



info normingute, uuemal ajal keelde tulnud või keelekasutajates küsimusi tekitavate rööpkujude kohta, sh põhjendused ja teave uute normingute kohta (nt gaala, tsunaami, karatee). Samuti leiab sealt selgitusi võõrsõnade vahenduskeelte (nt pomposne ~ pompöösne puhul), rööpkujude või sünonüümide kasutussageduse või mõistetavuse kohta (nt harvester), st püütakse käsitleda kõiki neid aspekte, mida tuleb eri kirjakeelade vahel otsustamiseks arvesse võtta.

## Viited

Kristiansen, Tore 2021. Destandardization. – The Cambridge Handbook of Language Standardization. (Cambridge Handbooks in Language and Linguistics.)

Wendy Ayres-Bennett, John Bellamy (toim). Cambridge: Cambridge University Press, 667–690. <https://doi.org/10.1017/9781108559249.026>

Paet, Tiina 2023. Võõrainese kinnistumine eesti keeles: keelekorralduslik ja leksikograafiline vaade. Doktoriväitekiri. Tartu Ülikooli Kirjastus.

## Mõnda ÕSi selgitustest ja soovitud õigekeelsusallikais täna

Tiina Paet, Tuuli Rehema, Sirli Zupping (Eesti Keele Instituut)

2025. aasta lõpus ilmub uus ÕS, mis üldjoontes järgib eelmise, 2018. aasta ÕSi eesmärke: esitab ortograafia, morfoloogia, sünonüümid ja hulga tähendusvihjeid, sidudes need soovitud ja selgitustega. Seoses muutustega keelekorralduse arengus ja põhimõtetes on uues ÕSis traditsioonilise kõrval ka mõningaid uuendusi. Näiteks ei esitata soovitud ÕS 2025-s enam lihtsalt näidetena, kus eespool on halvem ja selle järel parem variant, nagu eelmises ÕS-is, vaid selgitustena eraldi infokihis (Sõnaveebis rubriik “ÕS selgitab”). Selgitused sisaldavad sõna kasutusandmeid (korpusuuringute põhjal), suhteid varasemate normingute ja soovitudustega ning asjakohaseid viited teadustöödele ja EKI teatmikule. ÕS 2025 näitelausest ei leia enam lausetasandi sõnastussoovitud, nagu näiteks “linna {suunas} → poole minev rong” või “real inimestel, *parem* mitmel inimesel”, millesarnaseid ÕS 2018-s on üle saja.

Ettekandes tutvustame ÕS 2025 sõnastussoovitudustega tegelemise põhimõtteid ja lahendusi. Arutleme ka normingute ja soovitudustega vahekorra üle, peatudes keelekorralduse ja kirjakeele normimise probleemkohtadel. Nimelt on varasemad ÕS-id erisuguseid juhiseid andnud muu hulgas tähenduste ja sõnakasutuse kohta ega ole üheselt määratlenud, mis osa juhistest on soovitudust ja mis normingud. Praeguste põhimõtete järgi normitakse üldkeeles üksnes ortograafiat ja morfoloogiat, aga mitte tähendusi, sõnavalikut ega ka üldjuhul hääldust (vt ka Paet 2023). Seniste ÕS-ide soovitudust on käsitatud sageli rangete normingutena. Sõnavalik ja sõnakasutus lauses pole aga normitavad ning sõltuvad alati laiemast kontekstist ja sihtrühmast, mistõttu seda tasandit puudutavad soovitudust ja normingud ei peaks olema kirjakeele normi alusena sätestatud üldkeele sõnaraamatus. Osa seniseid ÕS-i sõnastussoovitudust võib edaspidi leida EKI teatmikust, kus on võimalik esitada pikemaid põhjendusi, tekstinäiteid ja üldistusi. Ettekandes selgitame näidete varal, miks pole otstarbekas anda sõnaraamatus lausetasandi soovitudust.

ÕS 2025 eesmärk on vähendada üldisi ettekirjutusi ja toetada keelekorralduse põhimõtet, et kirjakeelt ei saa lausetasandil normida, kuivõrd sõnavalik lausesse on tugevasti seotud kontekstiga. ÕS 2025 soovib kasutajat rohkem informeerida ja vähem õpetada, austada keelekasutaja valikuvabadusi ja toetada tema ise otsustamise julgust (vt ka EKI teatmik).





## Viited

Paet, Tiina 2023. Võõrainese kinnistumine eesti keeles: keelekorralduslik ja leksikograafiline vaade. Doktoriväitekiri. Tartu Ülikooli Kirjastus.

EKI teatmik. Eesti õigekeelsuskäsiraamat. Toimetanud Peeter Päll (peatoimetaja), Tiina Paet, Margit Langemets. Koostanud Peeter Päll, Maire Raadik, Tuuli Rehemaa, Margit Langemets, Tiina Leemets, Sirje Mäearu, Tiina Paet, Lydia Risberg, Arvi Tavast. Eesti Keele Instituut 2022–. eki.ee/teatmik/ (vaadatud 13.03.2025.)

Eesti õigekeelsussõnaraamat ÕS 2018. Toimetanud Maire Raadik. Koostanud Tiiu Ereht, Tiina Leemets, Sirje Mäearu, Maire Raadik. Eesti Keele Instituut. <https://arhiiv.eki.ee/dict/qs>

## Maintaining and Challenging the Standard: Grassroots Online Public Language Policing as a Biopower in Finnish Language Discussion Groups

Henni Pajunen (University of Turku, Finland)

Traditionally, standardization and language policy were seen as institution-led, but critiques of language policy and research on language ideologies have reshaped the understanding of language, power, and policy (e.g., Blommaert et al., 2009). This has led to a recognition that power in the context of language is distributed and consistently reproduced, as it is enacted within various societal structures and contexts. It invites alternative frameworks, such as Foucault's (e.g., 1980) theory of biopower, to examine the dynamics of power in language regulation. This concept of power operates through surveillance, discipline, and normalization with the goals of intervening, controlling, and regulating human life. It is constantly actualized through mundane interactions and activities.

In contemporary society, the internet and social media have become tools and platforms for such regulation. As such, regular language users can monitor and enforce language norms and use in accordance with language ideologies. *Online public language policing*, which can be defined as metalinguistic discourse that, e.g., laughs at, corrects, or mocks the language use of others in public online spaces, exemplifies this phenomenon. It often targets written language and evaluates it against the standard variety. By doing so, online public language policing functions as a grassroots mechanism that either reinforces or challenges standard language norms. In effect, it serves as a form of biopower.

This talk focuses on online public language policing practiced in Finnish Facebook language discussion groups by analyzing posts and comments made in the groups between December 2022 and February 2023 (Pajunen, 2024). Using a theory-driven content analysis approach which draws upon the concepts of shaming types developed by Murumaa-Mengel and Muuli (2018), it examines how norms of Standard Finnish are maintained but also how alternative normalities are suggested at a grassroots level. The analysis identifies six types of language policing: denunciatory, pedagogic, recreational, participative, reflective, and normalizing. These practices reveal two distinct techniques of biopower. First, normation occurs through denunciatory, recreational, pedagogic, and participative policing, which reinforce Standard Finnish and dominant language ideologies. Second, normalization emerges through reflective and normalizing practices, which critique established norms and accommodate norm-deviations, aligning with more critical language ideologies.

The findings demonstrate the ability of non-traditional actors within non-traditional domains, i.e., regular language users leveraging biopower in online spaces, to either maintain the standard or challenge it. The policing practices within these groups are more indicative of community empowerment than they are of civic digilantism.



## References

- Blommaert, J., Kelly-Holmes, H., Lane, P., Leppänen, S., Moriarty, M., Pietikäinen, S., & Piirainen-Marsh, A. (2009). Media, multilingualism and language policing: An introduction. *Language Policy*, 8, 203–207. <https://doi.org/10.1007/s10993-009-9138-7>
- Foucault, M. (1980). *Power/Knowledge: Selected interviews and other writings 1972–1977*. C. Gordon (Ed.). Harvester.
- Pajunen, H. (2024). Online public language policing as a biopower: Enforcing and challenging language norms and language ideologies in Finnish Facebook language discussion groups. *Apples - Journal of Applied Language Studies*, 18(3), 22–44. <https://doi.org/10.47862/apples.141099>
- Murumaa-Mengel, M., & Muuli, L. M. (2021). Misogynist content exposé pages on Instagram: Five types of shamings, moderators and audience members. *Participations: Journal of Audience and Reception Studies*, 18(2), 100–123.

## Üles ja ülesse varieerumine kõnes

**Margetta Pajuste, Miia Malm, Brigitte Pärn, Taavi Terasmaa, Annika Kängsepp, Maarja-Liisa Pilvik, Liina Lindström (Tartu Ülikool)**

Eesti keeles esineb varieerumist, mis on seotud keele standardiseerimisega, kus ühele norminguks kujundatud variandile lisaks kasutatakse ulatuslikult ka teist, norminguvastast varianti. Üks selline nähtus on *üles* ~ *ülesse* varieerumine, mille puhul keeleajalooliselt vanem vorm *üles* on saanud norminguks, ent Põhja-Eesti alal laialt kasutatud vorm *ülesse* on normingust välja jäänud. Varieerumine kajastus ka kirjutatud keeles (1-2). Eriti aktiivselt võttis *üles* normistamise osas sõna J. Aavik (1913, 1936: 1) ning *üles* oli ainsa variandina esitatud 1918. a ÕSis.

(1) *Laudad olid nii ilusaste ja mõnusaste kiwidest üles ehitud, nagu arwa mõisades olen näinud.* (<https://dea.digar.ee/page/eestipollomees/1870/09/01/5>)

(2) *Nüüd tõusis üks preester ülesse*  
(<https://dea.digar.ee/article/pernopostimees/1858/12/17/6>)

Ehkki normingut kajastab juba 1918. a ÕS, on *ülesse* siiski suhteliselt sage ka tänapäeval. Ettekandes uurime, missugune on normingupärase vormi *üles* ja kõnekeelse vormi *ülesse* vahetõrge suulises keeles ning millest sõltub vormi valik tänapäeva keeles.

Uurimuse materjal pärineb Eesti teismeliste keele korpusest (Vihman jt 2023), kus uuritavaid sõnavorme esines kokku 444 korda, ja eesti keele spontaanse kõne foneetilisest korpusest (Lippus jt 2023), kus sõnavorme esines 339 korda. Vaatlesime, kuidas kasutatakse osakaalud kahes korpuses sarnanevad, ning uurisime, kas ja kuidas seostub *üles* ja *ülesse* valik vormi asukohaga osalauses, vormile järgneva pausi olemasoluga, lause põhitegusõnaga ning kõneleja kõnetempo, soo ja vanusega.

Andmete ühetunnuseliseks analüüsiks kasutasime hii-ruut-teste ja t-teste; mitmetunnuselises analüüsis rakendasime segamõjudega logistilise regressiooni mudeleid.

Tulemused näitavad, et üldistatult umbes veerandil juhtudest esineb mõlemas korpuses norminguvastane vorm *ülesse*. Foneetikakorpuse materjali esialgsest analüüsist järeldus, et positsioon ja järgneva pausi olemasolu on peamised vormivalikut määravad tegurid, samas kui vanusel on väiksem, kuid siiski statistiliselt oluline mõju. Teismeliste keele korpuses kerkis esile vaid sõnavormi positsiooni roll. Samuti mängib vormivalikul olulist rolli kõneleja ise, viidates kõneleja on individuaalsetele eelistustele ühe või teise vormi kasutamisel.



## Viited

- Aavik, Johannes 1913. *Keele kaunima kõlavuse poole*. Tartu: Eesti Kirjastus-Ühisus Postimees.
- Aavik, Johannes 1936. *Eesti õigekeelsuse õpik ja grammatika*. Tartu: Noor-Eesti Kirjastus.
- Lippus, Pärtel, Kätlin Aare, Anton Malmi, Tuuli Tuisk & Pire Teras. 2023. *Phonetic Corpus of Estonian Spontaneous Speech v1.3*. Institute of Estonian and General Linguistics, University of Tartu. <https://doi.org/10.23673/re-438>
- Vihman, Virve-Anneli, Maarja-Liisa Pilvik, Aive Mandel, Annika Kängsepp, Mari Aigro, Kadri Koreinik, Kristiina Praakli & Liina Lindström. 2023. Estonian Teen Language Corpus. <https://doi.org/10.23673/RE-455>

## Sõnarõhk keeleinfos ja kõnesünteesis

**Liis Piits, Heete Sahkai, Meelis Mihkla, Indrek Hein, Liis Ermus, Hille Pajupuu (Eesti Keele Instituut, Eesti)**

Eesti keel on fikseeritud ehk kinnistunud sõnarõhuga keel, milles pearõhk on esimesel silbil. Hüüd- ja võõrsõnades, aga ka mõne tuletusliite korral võib pearõhuline olla muugi kui esisilp (Metslang jt 2023: 72). Eesti Keele Instituudi tänapäeva eesti keele sõnastikes on kasutusel ühtne hääldusmärkide süsteem, kus ebareeglipärasest rõhku märgitakse akuudiga silbituumaks oleva täishääliku ees. Põhimõtteid kirjeldatakse EKI Teatmikus (2025) (vt <https://eki.ee/teatmik/haaldusmargid-uhendsonastikus-us/>), Sõnaveebi lehel vormiparadigma all hääldusmärkide seletuses ja kõige põhjalikumalt ÕS 2018 eessõnas. Olemasolevates keeleinfoallikates on jäetud lahtiseks rõhu asukoht teatud tüüpi sõnades, kus järgsilp on tähistatud kolmanda välte märgiga, nt *adjut`an`n*, *mandol`iin*, *imperat`iiv* või kus järgsilbis on pikk vokaal, nt *mandariini* või *marmelaadi*. Kui EKI ühendsõnastiku (2024) hääldusmärkide seletustes öeldakse, et reeglipärane rõhk on sel juhul kolmandavärtelisel silbil või pikal täishäälikul, siis Teatmikus (2025) kirjutatakse, et pearõhk on neis sõnades pigem esimesel silbil, aga seda lihtsalt ei märgita: “Võõrsõnades, milles rõhk on esimesel silbil, aga järgsilp on III vältes, nt *presid`ent*, esisilbi rõhku harilikult ei märgita, sest tegelikus häälduses võib seda tüüpi sõnades esineda ka rõhk järgsilbil, nt *minarett* hääldub nii *minar`ett* kui ka *m`inar`ett*.” Kui emakeelekõnelejale ei pruugigi selline kahetine võimalus probleeme tekitada, siis keeleõppija ja kõnesüntesaator võivad segadusse sattuda.

Eesti Keele Instituudis loodi eelmisel aastal hääldusmärke arvestav kõnesüntesaatori prototüüp (<https://elo.eki.ee/yksiksona/>), et oleks võimalik sünteesida üksiksõnu, kuna tavasüntesaatoriga pole kontekstita võimalik homograafe eristada. Sellist süntesaatorit saab kasutada muuhulgas sõnaraamatute helindamisel. Süntesaatori treeningandmetena kasutati “Eesti keele põhisõnavara sõnastiku” (Kallas jt 2019) märksõnade kolme esimest sõnavormi, mis olid inimhäälega ette loetud. Hääldust mõjutavatest märkidest olid treeningandmestikus kasutusel liitsõnapiiri, rõhku, palatalisatsiooni ja kolmandat vältet tähistavad märgid nii nagu need sõnastikus esinesid. Katsed näitasid, et hääldusmärkidega kõnesüntesaator ei eksi palataliseeritud ja palataliseerimata ning II ja III vältes minimaalpaaride sünteesimisel (Kiissel jt 2025), aga ebareeglipärase rõhuga sõnade sünteesimisel paistab süntesaator sageli eksivat, interpreteerides rõhumärki kolmandavärtelise kõnetakti alguse märgistusena.

Ettekandes räägime võimalustest, kuidas saaks parandada rõhu märkimist keeleinfoallikates ja mis võimalusi oleks kõnesüntesaatori väljundi parandamiseks.

## Viited

*EKI ühendsõnastik 2024*. Eesti Keele Instituut, Sõnaveeb. <https://sonaveeb.ee/> (14.03.2025)



- Kallas, Jelena, Mai Tiits, Maria Tuulik, Kristina Koppel ja Madis Jürviste. 2019. *Eesti keele põhisõnavara sõnastik 2019*. 2., kohandatud veebiväljaanne. Eesti Keele Instituut. <https://sonaveeb.ee/>.
- Kiissel, Indrek, Liisi Piits, Heete Sahkai, Indrek Hein, Liis Ermus & Meelis Mihkla. 2025. Estonian isolated-word text-to-speech synthesiser. In Richard Johansson & Sara Stymne (eds.), *Proceedings of the Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies (NoDaLiDa/Baltic-HLT 2025)* (NEALT Proceedings Series No. 57), 302–306. University of Tartu Library.
- Metslang, Helle, Mati Ereht, Külli Habicht, Tiit Hennoste, Reet Kasik, Pire Teras, Annika Viht, et al. 2023. *Eesti grammatika*. Tartu Ülikooli Kirjastus.
- EKI teatmik. Eesti õigekeelsuskäsiraamat 2025*. Eesti Keele Instituut. Päll, Peeter (peatoimetaja), Tiina Paet, Margit Langemets. Koostanud Peeter Päll, Maire Raadik, Tuuli Rehemaa, Margit Langemets, Tiina Leemets, Sirje Mäearu, Tiina Paet, Lydia Risberg, Arvi Tavast. <https://eki.ee/teatmik/> (14.03.2025)
- ÕS 2018 = Raadik, Maire, Tiiu Ereht & Sirje Mäearu. 2018. *Eesti õigekeelsussõnaraamat ÕS 2018*. Tallinn: Eesti Keele Sihtasutus.

## Milline keelekasutus eestlasi häirib?

**Maarja-Liisa Pilvik, Liina Lindström, Kristel Algvere, Lydia Risberg, Helen Plado, Annika Kängsepp (Tartu Ülikool)**

Keeleideoloogiad kujundavad hoiakuid, mis peegeldavad inimeste uskumusi selle kohta, mida keel endast kujutab ja kuidas seda tuleks kasutada (Dragojevic, Giles & Watson 2013: 3). Standardkeele ideoloogia kallutab inimesi abstraktse idealiseeritud homogeense keele poole (Lippi-Green 2012: 67). Enamasti ei ole inimesed teadlikud ideoloogiate mõjust ja kalduvad näiteks standardkeele norme pidama talupojatarkuseks (Garrett 2010: 7), ehkki need on loodud kunstlikult. Üks standardkeele ideoloogia tahke on inimeste kindel usk keele korrektsusse (Milroy 2001: 535) ja variatiivsuse tülikaks pidamine (Davies 2000: 122). Ehkki keeleteadlased on aastakümneid rõhutanud, et keele varieerumine on loomulik ja vajalik, leidub hulgaliselt näiteid selle kohta, kuidas teiste keelekasutust peetakse valeks või halvaks (Cameron 1995; Lukač 2018).

Ettekandes räägime 2025. aasta veebruaris ja märtsis läbi viidud veebiküsitlusest, milles palusime täisealistel eesti keelt emakeelena kõnelejalatel nimetada kuni 10 nähtust, mis neile eesti keele kasutuses sageli silma või kõrva jäävad, ning hinnata skaalal 1–5, kas ja kuivõrd nimetatud nähtus neid häirib. Küsitluse põhieesmärk oli koguda praegu planeerimisjärgus olevate katsete jaoks sisendit, et saaksime uurida varieeruvate keelenähtuste töötlemist teksti lugemisel ja nendega seotud keelehoiakuid.

Lähtehüpoteesina eeldasime, et loetelukatses i) esilduvad enim kirjakeele normiga seotud ning keelekorraldusallikate soovitusel ja emakeeleõpetuses käsitletud nähtused ning ii) nimetatakse rohkem nähtusi, mis tugevalt häirivad.

Küsitlust jagati sotsiaalmeedia kanalites ning sellele vastas ühtekokku 252 vastajat vanuses 18–84. Nimetatud 1113 vastust kategoriseerisime vastavalt kirjeldatud nähtuse tasandile (nt morfoloogia, süntaks, tähendused, võõrkeelte mõju) ja kõnelejarühmale või keskkonnale, millega nähtust enim seostati. Ettekandes esitame esmase ülevaate küsitluse tulemustest ning arutleme neis ilmnevate keelehoiakute üle.

### Viited

Cameron, Deborah 1995. *Verbal Hygiene*. London: Routledge.



- Davies, Winifred V. 2000. Language Awareness amongst Teachers in a Central German Dialect Area. *Language Awareness* 9(3). 119–134. DOI: 10.1080/09658410008667141
- Dragojevic, Marko, Howard Giles & Bernadette M. Watson 2013. The Social Meanings of Language, Dialect and Accent: International Perspectives on Speech Styles. In Howard Giles & Bernadette M. Watson (eds.), *Language ideologies and language attitudes: a foundational framework*, 1–25. New York, NY, United States: Peter Lang Publishing.
- Garrett, Peter 2010. *Attitudes to Language (Key Topics in Sociolinguistics)*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511844713
- Lippi-Green, Rosina 2012. *English with an Accent: Language, Ideology, and Discrimination in the United States*. 2nd ed. Routledge. DOI: 10.4324/9780203348802
- Lukač, Morana 2018. Grassroots Prescriptivism: An analysis of individual speakers' efforts in maintaining the standard language ideology. *English Today* 34(4), 5–12. DOI: 10.1017/S0266078418000342
- Milroy, James 2001. Language Ideologies and the Consequences of Standardization. *Journal of Sociolinguistics* 5(4), 530–555. DOI: 10.1111/1467-9481.00163

## Kohanimed uues ÕSis: mida ja kuidas normida

Peeter Päll (Eesti Keele Instituut)

ÕS 2025 hakkab sisaldama u 7000 kohanime, mis on rohkem kui ÕS2018s, sest soov on hõlmata kõik Eesti asulanimed. Senine valimik vaadatakse üle, kärbitakse vähem aktuaalseid nimesid ja lisatakse uusi. Kohanimede esitus muutub lihtsamaks: antakse eesti teksti sobivad nimed, ära jäävad neile vastavad omakeelsed nimed, hääldus näidatakse eraldi, vrd ÕS 2018 **Ast`uuria**, hisp Asturias → ÕS 2025 **Astuuria** [ast`uuri`a].

Nimede kasutusvihjed püütakse anda selgelt. Esikohanimede märkimiseks on kolm võimalust: 1) esikohanime on märgendita, muud on märgendiga „mööndav“, „ajalooliselt“, „piiblis“ vms; 2) üks nimi on märgitud põhinimeks, teised jäävad märgendita; 3) eelistust ei märgita. 1. mall sobib juhul, kui esikohanime on endonüüm (kohalik omakeelne nimi), nt **Bremen**, mööndav Breemen; **Cēsis**, ajalooliselt Võnnu. 2. mall sobib muudel juhtudel, nt **Gruuksia** (põhinime), Georgia. Lõpuks on ka nimesid, mille puhul eelistuse märkimisel pole mõtet, tegu on sama nime eri vormistusega nt **Andamanid** ja **Andamani saared**.

Nimede häälduse põhiprobleem on selle kokkusobitamine nime käänamisega. Hääldust tuleb sageli mugandada, et nimi mahuks teatud käändtüüpi. Eriti puudutab see nime lõpposa. Nt kui osastavas on vorm Oxfordi [ˈɒksf`ordi], siis ei saa nimetavas olla hääldust [ˈɒksfəd], sest siis peaks osastav olema Oxfordit [ˈɒksfədit]. Omaette küsimus on levinud hääldused, mis erinevad kohalikust hääldusest, kuigi kirjalikult on reeglipärane. Eesti häälduses on levinud rõhu ülekandumine esisilbile, nt Amsterdam [ˈamsterdam], vrd hollandi [amsterd`am]. Ent kas Micheline hääldus [mišelin], mitte [mišl`ää(n)], on kultuurituse tunnus või Eestis siiski aktsepteeritav?

Kohanimede käänamisel on probleeme sellega, et kohakäänded ei ole ennustatavad, st automatiseeritavad. Nt kust tuleb vorm Riiga [r`iiga], kui nimetav on Riia? Kui *aasta*-tüübi puhul lühikest sisseütlevat ei ole, siis väliselt samasse tüüpi kuuluvate nimede puhul on see siiski võimalik (minnakse Pärnu ja Tartu). Näib, nagu oleks kohanimedel omaette käändtüübid.

Ettekandes tuleb juttu kohanimede vormistamise probleemidest ja võimalikest lahendustest.



## Viited

ÕS 2018 = Raadik, Maire, Tiina Leemets, Sirje Mäearu 2018. *Eesti õigekeelsussõnaraamat ÕS 2018*. Tallinn: EKSA.

## Normid ja tegelikkus: terminikasutuse varieeruvus ja standardiseerimine COVID-19 näitel

Mai Raet (Tallinna Ülikool, Eesti)

Keel ei ole vaid sõnaraamatutes ja keelekorralduse juhistes kirjas – see kujuneb igapäevases kasutuses, kus normidest kõrvalekaldumine ja varieeruvus on loomulikud nähtused. Käesolev ettekanne uurib COVID-19 pandeemia näitel, kuidas õigekeelsusallikate soovitusi järgiti ning kas ja kuidas tegelik keelekasutus standardeid järgis või mitte.

Ettekandes analüüsitakse, kuidas kasutati pandeemia ajal samade mõiste tähistamiseks erinevaid termineid ja kirjakujuksid (nt covid vs. COVID-19) ning milliseid eelistusi ja täpsusastmeid keelekasutuses järgiti. Samuti vaadeldakse, kui palju lähtuti õigekeelsusallikate soovitustest ning kuidas terminoloogiline mitmekesisus avalikkuses vastukaja leidis.

Uurimus põhineb mitmel allikal: nelja ministeeriumi ametlikel teadaannetel 2020. aastal, eesti keele ühendkorpusel (ÜK 2023) ning assotsiatsioonikatse käigus kogutud keeleandmetel 1454-lt eesti emakeelega kõnelejalt 2023. aasta alguses.

Analüüs näitab, et kuigi ametlikus suhtluses eelistati standarditud vorme (nt COVID-19), levisid igapäevases keelekasutuses rohkem mugandatud ja lühemad variandid (nt koroon, covid). Eriti pandeemia algusfaasis oli terminite kasutuses suur varieeruvus: esines sünonüüme ja erinevaid kirjaviise.

Ettekanne uurib, kuidas leida tasakaal keele reeglite järgimise ja loomuliku muutumise vahel ning millist mõju avaldab normide järgimine keele arusaadavusele ja suhtluse tõhususele. Samuti rõhutatakse, kuidas keel kohandub erinevates sotsiaalsetes ja ametlikes olukordades.

## Viited

ÜK 2023 = Eesti keele ühendkorpus 2023. Kristina Koppel, Jelena Kallas, Madis Jürviste, Helen Kaljumäe. Eesti Keele Instituut. <https://doi.org/10.15155/3-00-0000-0000-0000-08C04M>.

## Ajamäärsõnadel *koheselt* ja *kohe* pole mingit vahet! Kas tõesti?

Lydia Risberg<sup>1,2</sup>, Maarja-Liisa Pilvik<sup>2</sup>, Liina Lindström<sup>2</sup> (<sup>1</sup>Eesti Keele Instituut, Eesti, <sup>2</sup>Tartu Ülikool, Eesti)

Alates 1960. aasta õigekeelsussõnaraamatust on ajamäärsõna *koheselt* esitatud looksulgudes ehk sõnana, mida ei soovitata kasutada (ÕS 1960: 11; ka ÕS 2018). Põhjus seisneb selles, et *koheselt* on tuletatud omadussõnast *kohene*, mis omakorda tuleb määrsõnast *kohe*, ning keelekorralduses on selliseid *ne*-omadussõnadest tuletatud *lt*-tuletisi peetud liiasiks ületuletuseks, sest sama tähendus on juba kaetud lühema samast sõnaliigist sõnaga (nt Mäearu 2000; EKK 2020: 440). Lisaks on *lt*-tuletisi peetud omaseks kantseliitlikule keelekasutusele, mis ei sobi neutraalsesse keelde (EKG 1995: 600; EKK 2020: 372).



EKI ühendsõnastikus (2025) kajastub keelekorralduse värskem seisukoht, kus *lt*-tuletisi ei tõrjuta. See tugineb keeleteaduse tõdemusele, et liiasusel on keeles oluline roll ja see täidab eri funktsioone, näiteks kasutatakse seda rõhutamiseks ja intensiivistamiseks. Liiasus pole seega ülearune ega kasutu, nagu seda on preskriptiivsetes käsitlustes sageli esitatud (Wit, Gillette 1999; Lehmann 2005; Bazzanella 2011). Samuti on keeleteaduses osutatud, et *lt*-tuletised on produktiivne tuletustüüp ja nende puhul ei kehti alati piirang, et sarnase tähenduse jaoks pole vaja kahte samast sõnaliigist sõna (Kasik 2015: 47). Kas aga sõnadega *koheselt* ja *kohe* väljendatakse täpselt sama tähendust või on pikemat sõna vaja millekski muuks, nagu näiteks intensiivsuse lisamiseks ja ajalise vahetuse rõhutamiseks?

Ettekandes anname ülevaate ajamäärsõnade *koheselt* ja *kohe* võrdlevast uurimusest, mille eesmärk oli mõista nende tänapäevaks välja kujunenud kasutusmustreid ning teha järeldusi selle kohta, milliseid põhjuseid leidub tuletise *koheselt* valimise taga lühema *kohe* asemel. Püstitasime kaks hüpoteesi. Esiteks, et *koheselt* kasutuskontekstid rõhutavad kiireloomulisust ja ajalist vahetust enam kui lühema vormi *kohe* kasutuskontekstid. Hüpoteesi motiveerib *kohe* ajatähenduse ähmasus (näiteks 'selsamal hetkel' kõrval ka 'õige pea') ning tõdemus, et vahetu ajalise järgnevuse rõhutamiseks kasutatakse *kohe* asemel ka vormilt komplekssemaid sõnu *otsekohe* ja *kohemaid* ning kordust *kohe-kohe*. Teiseks eeldame, et *koheselt* ei ole levinud ainult kantseliitlikes, formaalsetes tekstides, vaid seda kasutatakse laialdasemalt, erinevates žanrites.

Uurimus toetub kasutuspõhisele keeleteooriale (Díaz-Campos & Balasch 2023). Analüüsisime ajamääruste *kohe* ja *koheselt* kasutuskontekste eesti keele ühendkorpusest (Koppel jt 2023; korpustööriistas Sketch Engine, Kilgarriff jt 2014) võetud juhuvalimite põhjal. Et juhuvalimid pärineksid nii autentsetest, toimetamata tekstidest kui ka toimetatud tekstidest, võtsime näited kolmest žanrist: blogid, foorumid ja perioodika. Sõnavaliku seoseid žanri, modaalsuse, lausestruktuuri ja kontekstuaalse kompleksuse teguritega analüüsisime keelevarieerumise levinud meetoditega (hii-ruut-testid ning tingimuslikud otsustuspuud). Selleks et mõista sõna *koheselt* kasutuse ja keelekorralduse soovitusetausta, uurisime DIGARi Eesti artiklite põhjal ka seda, kuidas arenes *koheselt* kasutus 20. sajandi esimeses pooles.

## Viited

- Bazzanella, C. (2011). Redundancy, repetition, and intensity in discourse. *Language Sciences* 33, 243–254.
- Díaz-Campos, M., Balasch, S. (2023). Introduction. *The Handbook of Usage-Based Linguistics*, 1–6. John Wiley & Sons, Ltd.
- DIGARi Eesti artiklid. Eesti Rahvusraamatukogu digitaalarhiiv. <http://data.digar.ee>
- EKG 1995 = *Eesti keele grammatika 1. Morfoloogia, sõnamoodustus*. Toim Mati Erelt. Tallinn: Eesti TA Keele ja Kirjanduse Instituut.
- EKI ühendsõnastik 2025. Eesti Keele Instituut, Sõnaveeb. <https://sonaveeb.ee>
- EKK 2020 = Erelt, M., Erelt, T., Ross, K. (2020). *Eesti keele käsiraamat*. Tallinn: Eesti Keele Sihtasutus.
- Kasik, R. (2015). *Sõnamoodustus*. Tartu: Tartu Ülikooli Kirjastus.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography* 1, 7–36.
- Koppel, K., Kallas, J., Jürviste, M., Kaljumäe, H. (2023). *Estonian National Corpus 2023*. Lexical Computing Ltd., Eesti Keele Instituut.
- Lehmann, C. (2005). Pleonasm and hypercharacterisation. *Yearbook of Morphology*, 119–154.
- Mäearu, S. (2000). *lt*-liitelised määrsõnad. *Keelenõuanne soovitab 2*. Tallinn: Eesti Keele Sihtasutus.
- Wit, E. C., Gillette, M. (1999). *What is linguistic redundancy*. The University of Chicago.



ÕS 1960 = *Õigekeelsuse sõnaraamat*. Toim Ernst Nurm, Erich Raiet, Magnus Kindlam. ENSV TA Keele ja Kirjanduse Instituut. Tallinn: Eesti Riiklik Kirjastus.

ÕS 2018 = *Eesti õigekeelsussõnaraamat ÕS 2018*. Toim Maire Raadik. Koost Tiiu Erelt, Tiina Leemets, Sirje Mäearu, Maire Raadik. Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus.

## **(De)standardiseerides kõnekeelt sõnaraamatus. Kuidas rakendada keelekorpus ja suuri keelemudeleid?**

**Lydia Risberg<sup>1,2</sup>, Maria Tuulik<sup>1</sup>, Margit Langemets<sup>1</sup>, Ene Vainik<sup>1</sup>, Eleri Aedmaa<sup>1</sup>, Esta Prangel<sup>1</sup>, Kristina Koppel<sup>1</sup>, Hanna Pook<sup>1</sup>, Sirli Zupping<sup>1</sup> (<sup>1</sup>Eesti Keele Instituut, Eesti, <sup>2</sup>Tartu Ülikool, Eesti)**

Sõnaraamatus kasutatakse erinevaid märgendeid, osutamaks, et sõna kaldutakse kasutama mujal kui neutraalses kontekstis. Informaalsele registrile osutab märgend *kõnekeelne* (nt *kärakas* tähenduses 'viin vm hrl kangem alkoholne jook'). Märgendite järele sõnaraamatus, eriti õigekeelsussõnaraamatus (ÕS), on ühiskondlikku nõudlust, muuhulgas on väljendatud soovi, et 2025. a ilmuv *ÕS* oleks kirjakeele kiht välja sõelatud.

Senised registrimärgendiga sõnad on olnud süsteemselt läbi uurimata, samas liigub osa sõnu aja jooksul näiteks kõnekeelest neutraalkeelde, mispärast on põhjust iga sõnaraamatu uustrüki eel ainek läbi töötada. Registrimärgendite määramist on alati mõjutanud ka leksikograafi intuitsioon või keelekorralduse valikud (nt Karelson 1990). Kuid registrianalüüs ei tohi põhineda üksikisiku eelistustel, vaid peab tuginema valimile, mis esindab registrit võimalikult terviklikult (Biber, Conrad 2009). Seega on olnud vajadus registrimärgendi määramisel laiema keelekasutuse süstemaatilise analüüsi järele.

EKI ühendsõnastikku koostades on subjektiivseid otsuseid vähendatud eesti keele ühendkorpuse (Koppel, Kallas 2022) analüüsimisega. Siiski on puudunud ühine suunis, orientiir, millele leksikograaf oleks saanud nt kõnekeelsuse üle otsustamisel tugineda – kõnekeel aga on piiripealsem ja subjektiivsem register kui näiteks vulgaarsused või lastekeel ning seega on leksikograafil seda keerulisem tabada. Ühtlasi on korpus tööriista võimalusi aja jooksul edasi arendatud: alles alates 2021. a versioonist saab uurida žanreid ja teemasid, kus sõna esineb, ja näha, kas ta on seal tüüpiline või mitte (Koppel, Kallas 2022). Suurte keelemudelite (SKMid) tulekuga on kerkinud küsimus, kas ja kuidas saab neidki sõnaraamatutöös rakendada, sh kõnekeelse märgendi määramisel.

ÕSi ajalooline eesmärk on olnud keelt standardiseerida. Standardiseerimine on protsess, kus norminguid tugevdades vähendatakse varieeruvust, mistõttu liigub keel mitmekesisusest ühtsuse suunas, samas kui destandardiseerimine viitab vastupidisele protsessile – normi järkjärgulisele nõrgenemisele varieeruvuse kasuks (Kristiansen 2021: 667). Kuna sõnaraamatu märgendeid on kipunud tõlgendada normatiivsena (vt ka Langemets jt 2024: 712; Trap-Jensen 2002), siis uustrükki ette valmistades tuleb üle vaadata ka see, kas mõne sõna (tähenduse) juurest tuleks (kõnekeele) märgend eemaldada, see n-õ destandardiseerida. Kas keelekorpus ja ehk ka SKMid võivad seda otsustusprotsessi toetada, pakkudes objektiivsemat alust kui subjektiivsed vaatlused, kajastades laiema keelekogukonna kasutusmustreid?

Ettekandes tutvustame, kuidas korpusanalüüs ja millised SKMid saavad aidata leksikograafil hinnata eesti sõnade kõnekeelsust sõnaraamatutöö jaoks ning pakume välja kõnekeelsuse orientiiri. Millised andmed ja mis juhul toetavad seda, et sõna või mõnd tema tähendust kaldutakse kasutama pigem kõnekeelses ehk informaalses ümbruses? Mil määral võib leksikograafi intuitsioon märgendamisotsuseid siiski mõjutada? Kas leksikograaf, keeleekspert, saab oma keeletaju välja lülitada ja kas ta üldse peaks seda tegema?





## Viited

- Biber, Douglas, Susan Conrad. 2009. *Register, Genre, and Style* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Karelson, Rudolf. 1990. „Eesti kirjakeele seletussõnaraamat” tegija pilgu läbi. *Keel ja Kirjandus* 1, 24–34.
- Koppel, Kristina, Jelena Kallas. 2022. Eesti keele ühendkorpuste sari 2013–2021: mahukaim eestikeelsete digitekstide kogu. *Eesti Rakenduslingvistika Ühingu aastaraamat* 18. 207–228.
- Kristiansen, Tore. 2021. Destandardization. *The Cambridge Handbook of Language Standardization*. Toim Wendy Ayres-Bennett, John Bellamy. Cambridge University Press, 667–690. <https://doi.org/10.1017/9781108559249.026>
- Langemets, Margit, Lydia Risberg, Kristel Algvere. 2024. To Dream or Not to Dream About ‘Correct’ Meanings? Insights into the User Experience Survey. *XXI EURALEX International Congress*, 700–719. Cavtat, Croatia.
- Trap-Jensen, Lars. 2002. Descriptive and Normative Aspects of Lexicographic Decision-Making: The Borderline Cases. *Proceedings of the Tenth EURALEX International Congress. Copenhagen*, 503–509.

## Partneriparandused telesarja „Õnne 13“ telefonikõnedes

### Andriela Rääbis (Tartu Ülikool)

Ettekandes vaadeldakse vestlusanalüüsi meetodit kasutades partneri parandusalgatusi telesarja „Õnne 13“ telefonikõnedes. Parandusmehhanismi abil lahendatakse vestluses ette tulevaid kõnelemis-, kuulumis- ja arusaamisprobleeme. Partneriparandus on selle üks alaliik, mille puhul partner (adressaat) osutab kõneleja voores olnud probleemile ning kõneleja viib paranduse läbi (Schegloff jt 1977; Hennoste 2023: 1143–1164).

Filmidiskursuse mudel on mitmekihiline: lisaks dialoogis vahetult suhtlejatele osalevad selles ka filmi vaatajad. Filmidialoogi luues tuleb lausungid moodustada mitte üksnes adressaati, vaid ka publikut silmas pidades, tuginedes nende arvatavale maailmateadmisele ja teadmistele, mis nad on filmist selleks hetkeks saanud. (Bubel 2008) Telefonikõned lisavad filmile realistlikkust, täites samal ajal lünki vaatajate teadmistes (Rossi 2011).

Uurimismaterjali moodustab 60 argitefonikõnet „Õnne 13“ eri hooaegadest (1994–2024). Nendes kõnedes näidatakse ainult üht osalejat, kes imiteerib dialoogi ja partneri juttu markeerivad pausid. Vaadeldavates kõnedes on 41 partneriparandust: 14 helistaja ja 27 vastaja algatatud. Uurimisküsimused on järgmised:

Missugune suhtlustegevus parandusalgatuse esile kutsub?

Mis tüüpi probleeme parandusega lahendatakse?

Missuguseid võtteid paranduse algatamiseks kasutatakse?

Vastaja parandusalgatused paiknevad enamasti vestluse alguses, sissejuhatuse järel, ning paranduse kutsub esile voores, milles helistaja esitab helistamise põhjuse. Enamasti on selle voores suhtlustegevuseks informeerimine, kahel juhul palve. Hilisemas vestluses tekitab probleemi helistaja küsimus, palve, info andmine või vastus küsimusele. Helistaja algatatud paranduste puhul on probleemlikuga voores suhtlustegevus samuti põhiliselt informeerimine, üksikutes näidetes palve, küsimus või vastus.

Parandusega lahendatakse peamiselt aktsepteerimis- (näide: rida 8) või arusaamisprobleemi, harvem kuulumisprobleemi (näide: rida 6) või soovitakse öeldule kinnitust (näide: rida 12). Kõige sagedasem paranduse tüüp on üleküsimine (näide: rida 8, 12), vähem kasutatakse avatud



algatust ja eriküsisõna (näide: rida 6). Parandusalgatus on valdavalt vormistatud ühe sõna või partikliga.

Parandusalgatused on sageli prosoodiliselt markeeritud: valjema häälega, üllatunud, ehmunud või naeruse tooniga, venitatud ja/või tugevalt rõhutatud. Vaadeldavates telefonikõnedes on sagedased mitmekordsed parandusjärjendid, mille puhul esimene parandus algataja seisukohast probleemi ei lahenda.

((telefonihelin))

01 V: firma Trans`vest, (.) sekretär. (0.5)

02 n tere `päevast, millega ma saan teile `kasulik olla?

03 (1.8)

04 V: sina `ema,

05 (2.3)

06 V: `KES,

07 (1.6)

08 V: @`ALLAN,@

09 (3.0)

10 V: ota `kus=se `juhtus.

11 (1.1)

12 V: `sadamas,

13 (1.6)

14 V: `ei=ei, `siin teda ei `ole,

### Viited

Bubel, Claudia M. 2008. Film audiences as overhearers. – *Journal of Pragmatics* 40 (1), 55–71.

Hennoste, Tiit 2023. Suuline keel. – *Eesti grammatika. Toim. Helle Metslang*. Tartu: Tartu Ülikooli Kirjastus, 997–1181.

Rossi, Fabio 2011. Discourse analysis of film dialogues: Italian comedy between linguistic realism and pragmatic non-realism. – *Telecinematic Discourse: Approaches to the Language of Films and Television Series*. Eds. Roberta Piazza, Monika Bednarek, Fabio Rossi. Amsterdam, Philadelphia: John Benjamins, 21–46.

Schegloff, Emanuel A., Gail Jefferson, Harvey Sacks 1977. The preference for self-correction in the organization of repair in conversation. – *Language* 52, 361–382.

## Rääkimistesti hindega korreleeruvad prosoodilised tunnused eesti keele kui teise keele kõnelejate kõnes

Heete Sahkai (Eesti Keele Instituut, Eesti)

Teise keele hääldusega tegelevates uurimustes pööratakse aina enam tähelepanu lauseprosoodiale (lauserõhk, intonatsioon, fraseerimine). On leitud, et lauseprosoodiaga seotud tunnused korreleeruvad hinnangutega sellistele kategooriatele nagu rääkimisoskus, arusaadavus (*comprehensibility*), mõistetavus (*intelligibility*), tajutav aktsent ja keeleoskustase



ning võivad korreleeruda nendega rohkem kui häälikutasandi tunnused ja samavõrd või rohkemgi kui sõnavara- ja grammatikapädevusega seotud tunnused (nt Jilka 2008, Mennen 2008, Kang 2010, De Jong jt 2012, Mennen & de Leeuw 2014, Isaacs & Thomson 2022, Munro & Derwing 2022). Lauseprosoodiaga seotud tunnustele on hakatud tähelepanu pöörama ka (automaatse) hindamise käsitlustes (Isaacs & Trofimovich 2017, Isaacs 2018, Kang & Johnson 2018, Kallio jt 2023) ja arvutipõhises hääldusõppes (Ballou 2018). Samuti on leitud, et prosoodia õpetamine parandab kõne arusaadavust rohkem kui häälikutasandi hääldusõpe (Levis 2023). Oluline on seega teha kindlaks lauseprosoodilised kategooriad, millele tuleks hääldusõppes tähelepanu pöörata, ning lauseprosoodiaga seotud tunnused, mida saab kasutada automaatsel hindamisel ja tagasiside andmisel.

Ettekandes tutvustan pilootuurimust, mis vaatles põhitooni taseme, ulatuse ja varieeruvusega seotud tunnuseid eesti keele kui teise keele kõnelejate kõnes. Varasemates uurimustes on leitud, et põhitooni tase võib eristada keeli (Mennen 2008, Mennen jt 2012) ja keeleoskustasemeid (Kang jt 2010); põhitooni ulatus võib eristada keeli (Mennen jt 2012), esimese ja teise keelena kõnelejate kõnet ja keeleoskustasemeid (Kallio jt 2023, Kuronen ja Kautonen 2018); ning põhitooni standardhälve ja keskmine kalle võivad eristada esimese ja teise keelena kõnelejate kõnet (Kallio jt 2023, Kuronen ja Kautonen 2018). Uurimuse andmestiku moodustasid 34 C1-taseme rääkimistesti suulist monoloogi aastast 2014. Monoloogid jagati kahte tingimusse: sooritusid hinnetega 7–10 ja sooritusid hinnetega 2–4; kummaski tingimuses oli 17 sooritust ning kõik kõnelejad olid naised. Olulisimaks tingimusi eristavaks tunnuseks osutus põhitooni kalle, mis oli suurem kõrgema hinde saanud sooritustes. See võib viidata sellele, et kõrgema hinde saanud sooritusi ehk parema rääkimisoskusega kõnelejaid iseloomustab varieeruvam intonatsioon. Kõrgema hinde saanud sooritusi eristasid ka keskmiselt madalam põhitooni miinimum ning suurem põhitooni ulatus ja standardhälve, kuid need erinevused ei olnud statistiliselt olulised.

Edasistes uurimustes tuleks lisaks uuritud tunnustele vaadelda konkreetseid lauseprosoodilisi kategooriaid nagu tooniaktsendid ja piirtoonid ning võrrelda lauseprosoodiaga seotud tunnuste olulisust tunnustega, mis on seotud muude prosoodiliste kategooriatega, nagu rütm, sõnavälde, sõnarõhk ja häälikupikkus, ja muude häälduse aspektidega, nagu häälikute artikulatsioon, samuti ladiususega ja muude keelepädevuse aspektidega, nagu sõnavara ulatus ja grammatika korrektsus.

## Viited

- Ballou, Nick 2018. Conceptualizing an Ideal Computer-Assisted Pronunciation Training System: Educational Technology for L2 Phonetic Features. *Cambridge occasional papers in linguistics*, Volume 11, Article 5: 120–130.
- De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34(1), 5–34. doi:10.1017/S0272263111000489
- Isaacs, T. (2018). Fully automated speaking assessment: Changes to proficiency testing and the role of pronunciation. In O. Kang, R. I. Thomson, & J. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 570–584). Abingdon, UK: Routledge
- Isaacs, T. & Thomson, R. I. 2022. Reactions to second language speech. In J. M. Levis, T. M. Derwing, M. J. Munro (eds.) *The Evolution of Pronunciation Teaching and Research. 25 years of intelligibility, comprehensibility, and accentedness*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 125-151.
- Isaacs, Talia and Trofimovich, Pavel (eds). *Second Language Pronunciation Assessment: Interdisciplinary Perspectives*, Bristol, Blue Ridge Summit: Multilingual Matters, 2017. <https://doi.org/10.21832/9781783096855>



- Jilka, Matthias 2008. Different manifestations and perceptions of foreign accent in intonation. In Trouvain, J., & Gut, U. (Eds.). (2008). *Non-native prosody : Phonetic description and teaching practice*. De Gruyter, Inc.. 77-96.
- Kallio, H., Kautonen, M., & Kuronen, M. (2023). Prosody and fluency of Finland Swedish as a second language: Investigating global parameters for automated speaking assessment. *Speech Communication*, 148, 66–80. <https://doi.org/10.1016/j.specom.2023.02.003>
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301–315. doi:10.1016/j.system.2010.01.005
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94(4), 554–566. <https://doi.org/10.1111/j.1540-4781.2010.01091.x>
- Kang, O., & Johnson, D. (2018). The roles of suprasegmental features in predicting English oral proficiency with an automated system. *Language Assessment Quarterly*, 15(2), 150–168. <https://doi.org/10.1080/15434303.2018.1451531>
- Kuronen, Mikko, Maria Kautonen 2018. Foneettisten piirteiden ja vieraan aksentin yhteydestä suomen kielessä. Lähivõrdlusi. *Lähivertailuja* 28, 207-241. doi:10.5128/LV28.06
- Levis, J. M. (2023). Teaching prosody in research studies: How researchers make decisions about pedagogy in research studies. *Journal of Second Language Pronunciation*, 9(3), 285–288. <https://doi.org/10.1075/jslp.23047.lev>
- Mennen, Ineke 2008. Phonological and phonetic influences in non-native intonation. In Trouvain, J., & Gut, U. (Eds.). (2008). *Non-native prosody : Phonetic description and teaching practice*. De Gruyter, Inc.. 53-76.
- Mennen I, de Leeuw E (2014) Beyond segments: Prosody in SLA. *Studies in Second Language Acquisition* 36: 183–94.
- Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *Journal of the Acoustical Society of America*, 131, 2249 – 2260
- Munro, M. J. & Derwing, T. M. 2022. Foreign accent, comprehensibility and intelligibility, redux. In J. M. Levis, T. M. Derwing, M. J. Munro (eds.) *The Evolution of Pronunciation Teaching and Research. 25 years of intelligibility, comprehensibility, and accentedness*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 7-32.

## **Tehisintellekti (TI) rakenduste kasutamise kirjutamisoskuse õpetamisel Eesti gümnaasiumi inglise keele õpetajate seas**

**Anneli Sigus, Liina Tammekänd (Tartu Ülikool, Eesti)**

**Märksõnad:** TI, inglise keele õpetamine, kirjutamisoskus, hindamine, tagasiside, Eesti gümnaasium

Tehisintellekti (TI) rakendamist kirjalike tekstide automaathindamisel (Xi, 2022; Mizumoto ja Eguchi, 2023; Xu ja Zhang, 2021) ja kirjalike tekstide põhjal õppija keeletaseme määramisel (Yancey jt, 2023; Gaillat jt, 2021) on küll uuritud, aga vähe on käsitletud TI kasutamist keeleõppes. Olemasolevad uuringud käsitlevad peamiselt piirkondi, kus domineerivad traditsioonilised õpetamismeetodid (Marzuki jt, 2023; Boudouaia jt, 2024; Li jt, 2023, Lei ja



Medwell, 2022), aga vähe infot on TI keeleõppesse integreerimisest piirkondades, kus õppekavade aluseks on Euroopa keeleõppe raamdokument (CEFR).

Inglise keel on Eestis kõige laialdasemalt õpetatav võõrkeel (Haridussilm, 2025) ja kuna riiklik õppekava on väljundipõhine, on õpetajatel autonoomia valida vahendeid, mis toetavad õpilaste jõudmist B2 tasemele. Kuna inglise keele eksam on enimvalitud keeleeksam gümnaasiumi lõpus, on inglise keele õpetajate suhtes ootus, et nad valmistavad õpilasi eksamiksi ette ja tagavad vajaliku keeletaseme saavutamise. Selleks peavad nad pakkuma õpilastele rohkelt võimalusi sihtkeele kasutamiseks ja tagasiside saamiseks. Kuna  $\frac{2}{3}$  gümnaasiumiõpilastest kasutab TI-d oma igapäevaelus ning paljud näevad selle suurimat kasutegurit just keeleõppes (Norstat, 2024), on õpetajatel vajadus end uuemate arengutega kurssi viia. Laak ja Aru (2024) uurisid Eesti üldhariduskoolide õpetajate kogemusi TI kasutamise kohta, aga TI kasutust inglise keele õpetamisel pole eraldi uuritud.

Käesoleva uuringu eesmärk on kaardistada, kas ja kuidas on inglise keele õpetajad TI-d oma töösse integreerinud. Otsime vastuseid järgmistele küsimustele: Kuidas kasutavad Eesti gümnaasiumi inglise keele õpetajad TI rakendusi oma oma õppetöös ja kirjutamisoskuse õpetamisel? Millised on TI õppetöösse integreerimise eelised ja väljakutsed inglise keele õpetamisel? Uuring kasutab kombineeritud metoodikat: küsitlus 540 õpetajale, kellest vastas 79 (15%), ning 7 õpetajaga viidi läbi intervjuud kvalitatiivseks analüüsiks.

Senised tulemused näitavad, et õpetajad kasutavad TI-d isikliku huvi tõttu ja vajadusest sammu pidada õpilaste TI kasutamise kohta. Intervjuudest tuleb välja, et kuigi TI on kasulik õppetöö planeerimiseks ja tunnitegevuste loomiseks, võib see olla aeganõudev. Toodi välja TI võimalikku kasulikkust kaasava hariduse seisukohalt, eriti rühmades, kus on erineva keeletasemega õpilased. Veel nähtub, et TI kasutamine õpilaste tööde hindamises on õpetajatele murekohaks, eriti hindamise läbipaistvuse, täpsuse, usaldusväärsuse ja eetiliste küsimuste osas, nagu õpilaste teadliku nõusolekuga arvestamine. Õpetajad, kellel on oskus TI-d ise trennida, on saanud õpilaste töödele kasulikku tagasisidet, aga trennimata TI-d kasutanud õpetajate kogemused õpilaste tekstide hindamisel on pigem negatiivsed. Uuring valgustab, kuidas Eesti gümnaasiumi inglise keele õpetajad praegu TI tööriistu kasutavad ja milliseid võimalusi nad TI õppetöös kasutamiseks näevad. Uurimuse tulemusi koos soovitusetega TI kasutamiseks on plaanis tutvustada inglise keele õpetajatele EATE suveseminaril ja ajakirjas *OPEN!*

## Viited

- Boudouaia, Azzedine, Samia Mouas, Bochra Kouider 2024. A Study on ChatGPT-4 as an Innovative Approach to Enhancing English as a Foreign Language Writing Learning. *Journal of Educational Computing Research*, 62(6), 1509-1537. <https://doi.org/10.1177/07356331241247465>
- Gaillat, Thomas, Andrew Simpkin, Nicolas Ballier, Bernardo Stearns, Annanda Sousa, Manon Bouyé, Manel Zarrouk 2022. Predicting CEFR levels in learners of English: The use of microsystem criterial features in a machine learning approach. *ReCALL* 34(2): 130–146. <https://doi.org/10.1017/S095834402100029X>
- Haridussilm 2025 = Haridussilm- Eesti hariduse teabevärv. [kuupäev puudub] Võõrkeelte õppimine. <https://haridussilm.ee/ee/tasemeharidus/oppetasemed/yldharidus/ektk-ja-voorkeeled> Viimati vaadatud 15.03.2025.
- Laak, Kristjan-Julius, Jaan Aru 2024. Generative AI in K-12: Opportunities for Learning and Utility for Teachers. *International Conference on Artificial Intelligence in Education* 502-509. Cham: Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-64315-6\\_49](https://doi.org/10.1007/978-3-031-64315-6_49)
- Lei, Man, Jane Medwell 2022. The Changing Role of Chinese English-as-Foreign-Language Teachers in the Context of Curriculum Reform: Teachers' Understanding of Their New Role. *Front Psychol.* <https://doi.org/10.3389/fpsyg.2022.904071>



- Mizumoto, Atsushi, Masaki Eguchi 2023. Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), Article 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
- Marzuki, Utami Widiati, Diyenti Rusdin, Darwin, Inda Indrawati 2023. The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*. <https://doi.org/10.1080/2331186X.2023.2236469>
- Norstat 2024 = Samsungi tellitud uurimus ideekonkursi *Solve for Tomorrow* raames. 2024. Kättesaadav: <https://www.solvefortomorrowbaltics.com/et/uuring-just-nendes-oppeainetes-saavad-noored-tehisintellektist-enim-abi/>. Viimati vaadatud 15.03.2025.
- Li, Yingxia, Wanyi Du, Bike Zhu 2023. Cognitive Engagement of Low Proficiency Learners: Evaluating the Impact of ChatGPT Feedback in EFL Writing. *Frontiers in Educational Research* Vol. 6, Issue 28: 69-73. <https://doi.org/10.25236/FER.2023.062809>
- Xi, Xiaoming 2022. Validity and the automated scoring of performance tests. Kogumikus G. Fulcher, and L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed.) 513-529 Routledge.
- Xu, Jinfen, Shanshan Zhang 2022. Understanding AWE Feedback and English Writing of Learners with Different Proficiency Levels in an EFL Classroom: A Sociocultural Perspective. *Asia-Pacific Edu Res* 31, 357–367. <https://doi.org/10.1007/s40299-021-00577-7>
- Yancey, Kevin P., Geoffrey Laflair, Anthony Verardi, Jill Burstein 2023. Rating Short L2 Essays on the CEFR Scale with GPT-4. *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*. 576–584. Toronto, Canada. Association for Computational Linguistics. DOI: 10.18653/v1/2023.bea-1.49

## Morphological ambiguity of Lithuanian uninflected parts of speech

Miglė Žemriūtė (Lithuanian Language Institute, Lithuania)

Around the year 2000, when the development of a morphologically annotated corpus began, morphological ambiguity became apparent (Rimkutė, 2003). This issue is not only a challenge for Lithuanian language researchers. In Czech, morphologically ambiguous forms account for 46% (Hajič 2004, cited in Rimkutė 2006), in Estonian and Latvian – nearly 50% (Puolakainen 2012, Paikens, Rituma, Pretkalnina 2013).

Initially, morphological taggers struggled with accuracy due to this ambiguity (for example, the Semantika.lt tagger annotated the Lithuanian Language Learner Corpus with 93.2% accuracy (Piečytė 2023)). This ambiguity occurs when a single word has multiple meanings or can be interpreted as different forms of the same word, resulting in incorrect grammatical tags. Despite improvements in taggers, the issue persists. Thus, solving the issue of annotating morphologically ambiguous forms remains significant (Žemriūtė, 2024).

Uninflected parts of speech also contribute to annotation inaccuracy, as they often overlap morphologically, accounting for 25% of ambiguous forms (Piečytė, 2023). For example, *arti* (English – near) can function as both adverb and preposition: *namas stovi arti* (“the house is near”) and *namas stovi arti miško* (“the house is near to the forest”). Taggers face challenges in determining the correct part of speech in such cases. While prepositions are typically followed by nouns, this distinction is not always clear. A review of Lithuanian dictionaries has revealed problematic cases where the distinction between adverbs and prepositions is theoretically unclear, and it indicates that distinguishing adverbs from prepositions is challenging. This also contributes to inaccuracies in automatic taggers. In the research, a total of 45 words are analyzed, which can be both adverbs and prepositions, e.g., *netoli* (near),



paraleliai (parallel), paskiau (afterwards), skersai (across), skradžiai (straight), statmenai (perpendicularly), toliau (further), žemiau (below).

Studying the morphological ambiguity of uninflected parts of speech in the Lithuanian language is important because uninflected parts of speech make up a large part of usage and are among the most common words in the Lithuanian language. In Lithuanian, not only adverbs and prepositions can overlap, but also conjunctions and particles, interjections and onomatopoeic words, conjunctions and prepositions, and so on. Inaccurate morphological annotation of these parts of speech hinders further word analysis, such as syntactic annotation.

The presentation analyzes cases of morphological ambiguity between uninflected parts of speech in the Lithuanian language, highlights the theoretical and practical difficulties in distinguishing them, and discusses potential opportunities for creating clearer rules to address this issue.

## References

- Paikens, P., Rituma, L. and Pretkalnina, L., 2013, Morphological analysis with limited resources: Latvian example. Proceedings of the 19th Nordic Conference of Computational Linguistics, pp. 267–277. Available at: <https://aclanthology.org/W13-5624.pdf> (Accessed: 17 January 2025).
- Piečytė, M., 2023. Morfologinio daugiareikšmiškumo analizė ir klasifikacija. Master's thesis. Vytautas Magnus University.
- Puolakainen, T. 2012. How Does the Choice of Morphological Analyser Influence the Quality of Syntactical Analysis? Human Language Technologies – The Baltic Perspective: Proceedings of the Fifth International Conference Baltic HLT 2012. Amsterdam, Berlin, Tokyo, Washington, DC: IOS Press, pp. 193–200. Available at: <https://dblp.org/rec/conf/hlt/Puolakainen12.html> (Accessed: 17 January 2025).
- Rimkutė, E., 2003. Morfologinio daugiareikšmiškumo tipologija. Lituaništika, [online] 4, pp. 60–78. Available at: <https://etalpykla.lituanistikadb.lt/object/LT-LDB-0001:J.04~2003~1367159925986/> (Accessed: 15 January 2025).
- Rimkutė, E., 2006. Morfologinio daugiareikšmiškumo ribojimas kompiuteriniame tekstyne. PhD dissertation. Vytautas Magnus University.
- Žemrietė, M., 2023. Morfologiškai daugiareikšmių formų atsiradimo priežastys. Bendrinė kalba, [online] 96, pp. 107–137. Available at: <https://www.lituanistika.lt/content/106422> (Accessed: 15 January 2025).

## Akadeemilise teksti tulevik

**Ilona Tragel, Eleriin Miilman, Nele Karolin Teiva, Nele Nove), Marri-Mariska Tammepõld, Laura Kiik, Reena Roos (Tartu Ülikool, Eesti)**

**Märksõnad:** akadeemiline tekst, juturobotid, kirjutamine, üliõpilased

2023. aastal lõime küsimustiku, millega uurisime gümnasistide ja üliõpilaste arvamusi tekstiloomeprotsessist ja sellega kaasnevatest murekohtadest (vt nt Tragel *et al.* ilmumas 2025). Kui toona olid Eestis juturobotid (ingl *chatbots*) alles populaarsust kogumas ja esimesi kasutajaid leidmas, on nüüdseks tehisaru rakenduste kasutamine tekstiloomes käivitanud suuremõõtmelisi ja vastuolulisi arutelusid (vt nt Riismaa 2024, Kull 2025, Laas 2025, Tragel & Hint 2025) ning muutumas hariduse, sh (akadeemilise) tekstiloomes, lahutamatuks osaks (vt nt Cardona *et al.* 2023; Tartu Ülikool 2024; TI-Hüpe 2025). Juturobotid jm uued tekstiloomes abivahendid on kaasa toonud tekstiloomes harjumuste muutumise, kirjutamisõpetuse



ümberkorraldamise ning kirjutamise (taas)mõtestamise. Meid huvitab, mida arvavad kirjutamistraditsiooni kandjad ja arendajad, praegused üliõpilased.

Kui tehniliste ülesannete täitmisel juturoboteid kiidetakse (vt nt Kull 2025, Krumsvik 2024), siis tekstiloomes on juturobotitel suured puudujäägid, eriti eesti keeles (vt tänavusi tööriistade katsetusi nt Teavik 2025). Muu hulgas on leitud, et juturobotite genereeritud akadeemilised tekstid on inimese kirjutatuga võrreldes lohisevad ja isikupäratud. Juturobotite genereeritud akadeemiliste tekstide esimene versioon on tihti paindumatu struktuuri, kindlate stiililiste eelistuste ja ühetaolise argumentide ehitusega. Kuna nende tekstide lähem uurimine on üldjuhul teadmata treeningandmete ja tööpõhimõtete tõttu keeruline, võiks pöörata pilgu inimtekstiloojatele juturobotite ajastul: kuidas ka akadeemilistes tekstides hoida loomungulisust ning mida teha, et lugejateni jõuaksid eesti keeles põnevad, julged ja vajalikud uurimused (vt teadusteksti igavusest Sand-Jensen 2010).

Soovime välja selgitada, milline on olnud praeguste üliõpilaste kogemus juturobotitega, mis motiveerib neid ise akadeemilisi tekste kirjutama ning millisena näevad nad akadeemilise teksti tulevikku (nt selle formaati, intertekstuaalsust, struktuuri). Selleks viime 2025. aasta märtsis ja aprillis läbi suulised paarisintervjuud kokku 30 bakalaureuse-, magistriastme üliõpilaste ja doktorantidega kuuest eesti avalik-õiguslikust ülikoolist (TÜ, TLÜ, EKA, EMTA, EMÜ, TalTech). Anonüümilised suulised intervjuud liitreeerime ning tõlgendame neid akadeemiliste tekstide kohta täna olemasolevate teadmiste põhjal ja keeleteaduslike vahendite abil (nt sõnaliikide proportsioonid, kasutatud adjektiivid ja verbid jne) ning ka metadiskursuse markereid otsides ning mõtestades.

## Viited

- Cardona *et al.* 2023 = Cardona, Miguel A., Roberto J. Rodríguez & Kristina Ishmael 2023. Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations. U.S. Department of Education, Office of Educational Technology 2023, <https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf> (15.03.2025).
- Krumsvik, Rune Johan 2024. Chatbots and academic writing for doctoral students. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-13177-x>
- Kull, Meelis 2025. Humans and AI: similarities, differences, and why it matters. Ettekanne konverentsil DHNB2025. Tartu 07.03.2025, <https://www.youtube.com/live/ioC0YxZqGbl> (15.03.2025).
- Laas, Oliver 2025. Juturobotid ja kriitiline mõtlemine. *ERRi arvamusportaal*, 03.02, <https://www.err.ee/1609594070/oliver-laas-juturobotid-ja-kriitiline-motlemine>
- Riismaa, Kaur 2024. Tehisarukus kirjaniku pilguga. *Sirp*, 22.03.
- Sand-Jensen, Kaj 2010. Kuidas kirjutada läbinisti igavat teadusteksti. Tlk. S. Rattasepp. *Acta Semiotica Estica VII*, 285–293. Eesti Semiootika Selts. Kättesaadav: [https://www.semiootika.ee/acta/wp-content/uploads/2017/02/Acta\\_VII.pdf](https://www.semiootika.ee/acta/wp-content/uploads/2017/02/Acta_VII.pdf)
- Tartu Ülikool 2024. Õppimine ja õpetamine tehisintellekti ajastul: kuidas kasutada tekstiroboteid õppetöös. <https://sisu.ut.ee/ti/> (15.03.2025).
- Teavik 2025. Akadeemiliste tekstide kirjutamise veebileht. Tehisintellektil põhinevate tööriistade kaasamine kirjutamisprotsessi, <https://www.teadustekst.ut.ee/tehisintellektil-pohinevate-tooriistade-kaasamine-kirjutamisprotsessi/> (15.03.2025).
- TI-Hüpe 2025 = TI-Hüppe veebileht. <https://www.tihupe.ee/> (15.03.2025)
- Tragel *et al.* (ilmumas 2025) = Tragel, Ilona, Liisa-Maria Komissarov, Eleriin Miilman, Nele Karolin Teiva & Marri-Mariska Tammepõld 2025. A Year of Generative AI: Observations from a Survey among University Students in Estonia. *Journal of Academic Writing*, vol. 15, no S2 2025.
- Tragel, Ilona & Helen Hint 2025. TI-hüpe üle õpetajate. *ERRi arvamusportaal*, 05.03, <https://www.err.ee/1609622519/ilona-tragel-ja-helen-hint-ti-hupe-ule-opetajate>





## Kui hästi seletavad suured keelemudelid leksikograafide hinnangul eesti sõnade tähendusi?

Maria Tuulik<sup>1</sup>, Ene Vainik<sup>1</sup>, Esta Prangel<sup>1</sup>, Margit Langemets<sup>1</sup>, Eleri Aedmaa<sup>1</sup>, Kristina Koppel<sup>1</sup>, Lydia Risberg<sup>1,2</sup> (<sup>1</sup>Eesti Keele Instituut, Eesti, <sup>2</sup>Tartu Ülikool, Eesti)

Korpuste kasutamine on (mh Eesti) leksikograafias igapäevane, aga kuivõrd korpusandmete analüüs ja korpusest automaatselt tuvastatud andmete järeltoimetamine on töömahukad, otsitakse ka leksikograafias agaralt võimalusi suurte keelemudelite (SKMid) kasutamiseks (nt Evert jt 2024; Kosem, Gantar jt 2024; Kosem, Kuhn jt 2024; Tadić 2024; Tiberius jt 2024). Ettekandes uurime, kas eesti keele kui väikese kõnelejaskonnaga ning mudelites tekstimahult alaesindatud keele puhul saavad keelemudelid olla leksikograafide abiks sellises olulises tööloigus nagu sõnatähenduste seletamine. Mil määral saavad suured keelemudelid (SKMid) leksikograafi aidata sõnastikuseletuste loomisel ja sõnatähendus(t)e esitusel?

Viisime 2024. a sügisel läbi küsitluse, mille tulemusi ettekandes tutvustame. Anname ülevaate sellest, kuidas hindavad leksikograafid kui kvaliteedi kontrollijad ja tagajad SKMide genereeritud sisu. Katsetasime, kuivõrd saavad kaks keelemudelit – Gemini 1.5 Pro ja GPT-4o –, hakkama eesti sõnade seletamisega. Kasutasime näitetu promptimise (*zero-shot*) meetodit, sest see võimaldas selgitada välja leksikograafide eelistusi ja nõudmisi seletustele ilma nende vormi ennetavalt piiramata. Genereerisime seletused 60 EKI ühendsõnastiku märksõnale, seejärel hindas loodud seletusi 21 leksikograafi. Vaatamaks, kuidas tulevad SKMid toime polüseemsete sõnadega, sõelusime välja märksõnad, mil oli 1–4 tähendust.

Ettekandes teeme esmalt ülevaate keelemudelitele antud skalaarsetest hinnangutest: leksikograafid hindasid tähenduste tabamist, seletuse keelelist korrektsust ja seletuse kasulikkust leksikograafiatöös. Seejärel kommenteerime hinnanguküsimuste vastuste varieerumist ja analüüsime uuritud tegurite (skalaarsed hinnangud, seletatava sõna/üksuse mitmesõnalisus, polüseemia, korpusesagedus ja võõrsõnalisus) korrelatsiooni, et selgitada välja tegureid, mis võivad SKMide väljundi kvaliteeti mõjutada. Viimaks analüüsime leksikograafide jäetud vabas vormis kommentaare küsimusele, kui kasulikuks hindavad nad pakutud seletust praktilise sõnaraamatutöö jaoks. Hinnangutele lisatud kommentaarid võimaldavad vaadelda täpsemalt, mis just tähendusseletuste puhul on need aspektid, mis mõjutavad leksikograafide rahulolu uue tööriistaga (nt puuduv või üleliigne tähendus, kirjelduse pikkus, toodud näidete prototüüpsus või ebaloomulikkus).

### Viited

- Evert, S., Ganslmayer, C., & Rink, C. 2024. Multi-Level Analysis as a Systematic Approach to Evaluating the Quality of AI-Generated Dictionary Entries. Proceedings of the XXI EURALEX International Congress, 317–334. <https://euralex.jezik.hr/wp-content/uploads/2021/09/Euralax-XXI-final-web.pdf>
- Kosem, I., Gantar, P., Arhar Holdt, Š., Gapsa, M., Zgaga, K., & Krek, S. 2024. AI in Lexicography at the University of Ljubljana: Case Studies. XXI EURALEX International Congress. Lexicography and Semantics, Cavtat, Croatia.
- Kosem, I., Kuhn, T. Z., Arhar Holdt, Š., Koppel, K., Tiberius, C., Zviel Girshin, R., Waszink, V., & Zgaga, K. 2024. Can AI Assist in Selecting Dictionary Examples? A Case Study in Four Languages. XXI EURALEX International Congress. Lexicography and Semantics, Cavtat, Croatia. [https://euralex.jezik.hr/wp-content/uploads/2021/09/EURALEX\\_2024\\_Full\\_Programme\\_chairs\\_FINAL.pdf](https://euralex.jezik.hr/wp-content/uploads/2021/09/EURALEX_2024_Full_Programme_chairs_FINAL.pdf)
- Tadić, M. 2024. Can LLMs Really Generate New Words? XXI EURALEX International Congress. Lexicography and Semantics, Cavtat, Croatia. [https://euralex.jezik.hr/wp-content/uploads/2021/09/EURALEX\\_2024\\_Full\\_Programme\\_chairs\\_FINAL.pdf](https://euralex.jezik.hr/wp-content/uploads/2021/09/EURALEX_2024_Full_Programme_chairs_FINAL.pdf)



Tiberius, C., Heylen, K., Vanroy, B., Vandeghinste, V., de Does, J., & van Doeselaar, J. 2024. LLMs and Evidence-Based Lexicography: Pilot Studies at INT. XXI EURALEX International Congress. Lexicography and Semantics, Cavtat, Croatia. [https://euralex.jezik.hr/wp-content/uploads/2021/09/EURALEX\\_2024\\_Full\\_Programme\\_chairs\\_FINAL.pdf](https://euralex.jezik.hr/wp-content/uploads/2021/09/EURALEX_2024_Full_Programme_chairs_FINAL.pdf)

## Sõnastik – INIMENE ja MAHUTI

Ene Vainik<sup>1</sup>, Geda Paulsen<sup>1,2</sup> (<sup>1</sup>Eesti Keele Instituut, Eesti; <sup>2</sup>Uppsala ülikool, Rootsi)

**Märksõnad:** tähendus, metafoor, metonüümia, kujundlik mõtlemine, diskursus

Uurimuse taustaks on nn keelekriisi diskursus Eesti avalikus meedias oma tõusude ja langustega (2020—2024), mille sisuks on olnud arutelud keele, kirjakeele normide, keelekorralduse ja eri sõnaraamatute ning nende ülesannete ja kuju üle. Oleme varem sedastanud, et diskursuses tõepoolest avaldusid kriisi tunnused: olukorra ebamäärasus, ajaraami piiratus ja põhiväärtuste ohustatus (Vainik ja Paulsen 2023a, 2023b).

Kuivõrd kriisidiskursuse üheks käivitajaks oli EKI väljendatud kavatsus muuta senise traditsioonilise ÕSi formaati (Langemets ja Päll 2021), siis on siinses ettekandes vaatluse all, kuidas mõeldi ja argumenteeriti sihtmõistetest SÕNASTIK ja SÕNARAAMAT. Ehkki loomult konkreetsemad kui varasemalt käsitletud sihtmõiste KEEL, on sõnastikestki mõtlemisel abiks kujundlikkus, seda paratamatult eriti siis, kui mõelda-kirjutada tuleb välise kujuta e-sõnastikest.

Ühest küljest on loomulik, et avalduvad samad kujundliku mõtlemise viisid, mis on leitud olevat käibel interneti kohta üldiselt: INTERNET ON KOHT; INFO HANKIMINE ON LIIKUMINE FÜÜSILISES RUUMIS (nt Matlock jt 2014); INTERNET ON RESSURSS (Dekavalla 2022). INTERNET ON MITTETOIMIV RAAMATUKOGU (Ratzan 2000). Teisest küljest on oodatav, et kriisidiskursuses avaldub kujundlikkus reedab argumenteerijate alusväärtusi ja nende ohustatuse tunnet (Vainik ja Paulsen 2023b) ning et esile tulevad identiteedi ja keelelise ideaaliga seotud projektsioonid kirjakeelt esindavale sõnaraamatule.

Uuritav diskursus, milles arvamusi vahetasid erinevad osapooled – praktikud, teadlased, ajakirjanikud, üldsus – koosneb 94 tekstist aastatest 2020—2024. Andmestikku analüüsiti käsitsi kvalitatiivselt kujundliku keelekasutuse mustrite tuvastamiseks. Tõlgendamise peamine teooria raamistik on kognitiivne metafoorteooria (Lakoff ja Johnson 2011, Kövecses 2020), metafoorne raamistamine (Burges jt 2016).

Tulemused osutavad, et sõnastikke eeskätt isikustatakse (st rakendub mõistemetafoor SÕNARAAMAT ON INIMENE), kasutatakse ka ruumilisi lähtevaldkondi nagu MAHUTI ja KOHT: spetsiifilisemalt veel ka EHITIS, ALLIKAS, SEADUS, TÕÕRIIST, PIIBEL jpt. Seejuures toome esile, millised kujundliku mõtlemise viisid olid väitlejatel kasutusel Ühendsõnastikust ja millised ÕSist mõeldes-kirjutades. Pöörame tähelepanu ka metonüümiatele nagu SÕNARAAMAT ESINDAB KEELEKORRALDUST; SÕNARAAMAT ESINDAB NORMI jpt.

### Viited

- Burgers, Christian; Konijn, Elly; Steen, Gerard 2016. Figurative framing: Shaping public discourse through metaphor, hyperbole, and irony. *Communication Theory* 26 (4), 410–430. <https://doi.org/10.1111/comt.12096>
- Dekavalla, Marina 2022. Metaphors of the virtual: how ordinary people frame what the internet is. *Social Semiotics*, 34 (4), 539–552. <https://doi.org/10.1080/10350330.2022.2144203>
- Dorst, A. G. 2011. Personification in discourse: Linguistic forms, conceptual structures and communicative functions. *Language and Literature*, 20(2), 113-135. <https://doi.org/10.1177/0963947010395522>



- Kövecses, Z. 2020. *Extended Conceptual Metaphor Theory*. Cambridge University Press.
- Lakoff, George; Johnson, Mark 2011. *Metafoorid, mille järgi me elame*. (Gigantum humeris.) Tlk Ene Vainik. Tallinn: Tallinna Ülikooli Kirjastus.
- Langemets, Margit; Päll, Peeter 2021. Muutuv ÕS. *Sirp* 4 (14), 29.01.2021. <https://www.sirp.ee/s1-artiklid/varamu/muutuv-os/>
- Matlock, Teenie; Castro Spencer C.; Fleming Morgan; Gann, Timothy M.; Maglio, Paul P. 2014. Spatial Metaphors of Web Use. *Spatial Cognition & Computation* 14, 306–320. <https://doi.org/10.1080/13875868.2014.945587>
- Ratzan, Lee 2000. Making sense of the Web: a metaphorical approach. *Information Research*, 6 (1), 85–95.
- Vainik, Ene; Paulsen, Geda 2023a. Kujundlik mõtlemine 2020–2022 keelekriisis. *Keel ja Kirjandus*, 66 (7), 651–677.
- Vainik, Ene; Paulsen, Geda 2023b. Values behind metaphors: figurative speech in the discourse of a “language crisis”. *Tertium Linguistic Journal*. *Cracow Tertium Society for the Promotion of Language Studies* 8 (1), 22–51.



## Workshop of multilingualism and multilingual acquisition

### Estonian language proficiency of Ukrainian and Russian speaking 3rd grade students

Reili Argus, Piret Baird, Andra Kütt-Leedis, Evelyn Neudorf, Tiina Rüütmaa, Anne Uusen (Tallinna Ülikool)

The study of second language acquisition among students with migration background is a significant area of research, particularly concerning the growing population of Ukrainian refugee children in Estonia. However, the relationship between Estonian language proficiency and the home language environment has received limited attention. Previous research indicates that refugees typically acquire a second language more slowly than other immigrant groups (e.g., Browder 2018; Stolk et al. 2022).

This study aims to assess: 1) the level of Estonian language proficiency among Ukrainian refugee students and Russian-speaking students at the beginning of the 3rd grade; and 2) the extent to which this proficiency is influenced by factors in the students' home language environment, including the timing and nature of their exposure to Estonian.

The sample includes 25 Ukrainian refugee students from various schools in Tallinn and 25 children with different mother tongues (not classified as refugees) who entered the 3rd grade in the fall of 2024. A picture description test (see Kütt-Leedis 2024) will evaluate the students' Estonian proficiency, while a parental questionnaire will gather data on socioeconomic factors. Vocabulary and constructions will be analyzed based on the test results.

Preliminary analysis of the language environment of children reveals that groups have had different amounts of exposure to Estonian. Among 21 children with Russian as their home language, four attended bilingual kindergartens, while 11 were in Estonian-speaking settings. All Ukrainian-speaking children relocated to Estonia in 2022 within six months of the onset of the war in Ukraine.

The initial vocabulary analysis shows that Ukrainian-origin children's vocabulary is more varied (longest description: 250 words; shortest: 6) compared to their Russian-speaking peers (191 and 27 words, respectively). Nonetheless, average vocabulary richness and diversity of word forms are higher among Russian-speaking students. While the average number of words in both groups' texts is similar, Russian-speaking students demonstrate a greater variety of vocabulary. Notably, among Ukrainian students, richer texts tend to contain more repetitions of words and forms.

In both groups, nouns dominate the vocabulary, followed by verbs and adverbs, with adjectives being less common. Pronouns are more frequently used by Russian-speaking children. Both Ukrainian and Russian-speaking narratives exhibit numerous hesitation markers, indicating frequent pauses in thought.

The analysis of constructions reveals that there are no significant differences in the types of constructions used by Ukrainian refugee children and Russian-speaking children. However, the proportion of different construction types varies between the observed groups.

### References

- Browder, Christopher T. 2018. Recently Resettled Refugee Students Learning English in US High Schools: The Impact of Students' Educational Backgrounds. *Educating Refugee-background Students*. Multilingual Matters.
- Kütt-Leedis, Andra 2024. Eesti keelt esimese ja teise keelena kõnelevate laste sõnavara pildikirjelduskatse põhjal. *Emakeele Seltsi aastaraamat*, 69, 107–131. Eesti Teaduste Akadeemia Kirjastus. DOI:10.3176/esa69.05.



Stolk, Yvonne; Kaplan, Ida; Szwarc, Josef 2022. Majority language acquisition by children of refugee background: a review. *International Journal of Inclusive Education* 1–24. <https://doi.org/10.1080/13603116.2023.2210593>.

## Linguistic gender stereotypes in Russian and their impact on politeness perception

Ilenia Del Popolo Marchitto (Tallinn University, Estonia)

This study explores gender-based stereotypes in Russian and their effect on individuals' perception of linguistic (im)politeness. The primary aim is to determine what are the common preconceptions about male and female speech in relation to politeness and whether certain linguistic choices tend to be associated with a particular gender. Moreover, the study aims to examine whether people evaluate politeness not just based on linguistic features but also on the presumed gender of the speaker.

Previous research suggests that Russian culture exhibits deeply rooted norms and expectations regarding male and female speech patterns (Mills 1999: IX). For instance, as a result of traditional societal roles, women are generally seen and expected to be more polite, cooperative and accommodating than men in interaction (Zemskaya et al 1993: 134). Since speakers' linguistic choices in request-making are frequently conditioned by established conventions and social expectations, this study specifically focuses on linguistic gender stereotypes in Russian requests.

Based on these premises, the working hypothesis is that direct, unmitigated requests are still nowadays typically associated with male speech, whereas indirect formulations containing mitigation markers, hesitations and hedges tend to be stereotypically associated with female speech. Moreover, we hypothesize that an identical request formulation is generally perceived as more polite when attributed to a woman.

These hypotheses were tested through a sociolinguistic experiment. Two nearly identical online questionnaires were administered to a total of 250 Russian-speaking respondents from 33 different countries. The majority (77.2%) has a multilingual profile, using up to five languages, including Russian, on a daily basis. A small percentage of the total (12%) consists of highly proficient users who consider Russian as a FL.

Both questionnaires included a set of 19 anonymised requests, derived from spontaneous interactions in Russian. In the first one, 125 participants had to evaluate the degree of politeness of each of the 19 requests. In the second questionnaire, participants were also requested to infer the speakers' gender. Thus, the sole difference between the questionnaires was that the first had no references to gender, allowing for a comparison between neutral evaluations on (im)politeness and potentially gender-biased ones.

The results confirmed our hypotheses. Participants' responses revealed systematic associations between linguistic formulations and gender, while also highlighting the strong impact of these stereotypes on politeness perception. Findings show that politeness is not only expected from women but also more readily perceived in their speech, reinforcing existing biases.

### References

Biernat, M., Sesko, A. K. 2018. Gender Stereotypes and Stereotyping: A Cognitive Perspective on Gender Bias. In N. K. Dess, J. Marecek, L. Bell (eds). *Gender, Sex, and Sexualities: Psychological Perspectives*, 171–194.



- Giles, H., Scholes, J., Young, L. 1983. Stereotypes of male and female speech: A British study. *Central States Speech Journal*, 34(4), 255–256.
- Lindvall-Östling, M., Deutschmann, M., Steinvall, A. 2020. An Exploratory Study on Linguistic Gender Stereotypes and their Effects on Perception. *Open Linguistics*, vol. 6(1), 567–583.
- Mills, M. H. 1999. *Slavic Gender Linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Mills, S. 2003. *Gender and politeness (Studies in Interactional Sociolinguistics 17)*. Cambridge: Cambridge University Press.
- Zemskaya, E. A., Kitaygorodskaya, M. A., Rozanova N. N. 1993. Osobennosti muzhskoy i zhenskoy rechi [The Peculiarities of Male and Female Speech]. In E. A. Zemskaya, D. N., Shmelev (eds). *Russkiy yazyk v ego funktsionirovanii: Kommunikativno-pragmaticheskiy aspekt [Russian Language in its Functioning: Communicative-pragmatic aspect]*. Moscow: Nauka. 90–136.

## **‘Kak eto po-russki?’ Transliteration in Russian-medium online communication across the Baltic states and formation of local standards**

**Kapitolina Fedorova, Natalia Tšuiкина and Alexandra Lapa (Tallinn University, Estonia)**

Multilingual practices, although extremely diverse and often perceived as chaotic, tend to follow certain trends and patterns sometimes resulting in emergence of local sub-standards and communication norms. These processes are supported by digital communication, which nowadays constitutes a significant part of human interaction. Participation in various online communities creates numerous opportunities for both maintaining and changing certain linguistic practices. Minority languages speakers, in particular, can connect with fellow speakers online and use their mother tongue even if they have to predominantly use the majority language in their everyday life.

The Baltic region, characterized by its linguistic diversity and historical intricacies, serves as a unique backdrop for investigating the ways in which Russian speakers navigate linguistic boundaries in the digital space and create new ways of employing strategies of linguistic hybridization. Considering the fact that all majority languages in the Baltic countries (Estonian, Latvian and Lithuanian) use alphabets based on Latin but with different modifications, Russian speakers in informal written communication, when not able or not willing to use Cyrillic alphabet, may apply different transliteration rules, both conventional and non-conventional.

The proposed paper presents the results of comparative analysis of the data on transliterating practices employed by Russian speakers from Estonia, Latvia and Lithuania from two main sources. First, we studied communication in local Facebook groups, creating a corpus of examples with transliteration from each respective country. Second, a survey was conducted among Russian speaking residents of Estonia, Latvia and Lithuania to collect their opinions on using transliteration, and also obtain experimental data from transliteration task. In the presentation, we will compare our data from real life communication and from metalinguistic experiment, and reveal differences between patterns typical for Russian speakers from each Baltic country.

By analyzing transliteration choices, the paper aims not only to attest to linguistic consequences of contact between Russian and three different languages but also to identify patterns and trends that reflect the negotiation of linguistic and cultural hybridity within the digital realm. Ultimately, this research contributes to the broader understanding of language



dynamics in online communities, offering insights into the ways multilingualism and transliteration practices shape the formation of new local versions of Russian turning it into polycentric language.

## Written Language Characteristics in Dyslexia: A Case Study

**Eglė Krivickaitė-Leišienė, Viktorija Kavaliauskaitė-Vilkinienė (Vytautas Magnus university, Lithuania)**

The ability to read and write is one of the most crucial factors in achieving academic success. To acquire knowledge, children must be able to read words accurately, comprehend what they read, write words correctly, and express their thoughts coherently. Reading and writing involve both decoding and comprehending printed materials. Word decoding requires the activation of multiple brain functions, including visual and auditory modalities, as well as orthographic, phonological, and semantic systems (Çaylak, 2010).

This presentation aims to analyze the written language characteristics of a Lithuanian child with dyslexia. The data was collected over two years and consists of the child's written work from the first and second grades.

Dyslexia is characterized by difficulties in learning to read, spell, and write despite adequate intellectual capacity, educational resources, and social background (Bree et al., 2010). Research indicates that children with dyslexia may also struggle with semantics, syntax, and discourse (Plaza et al., 2002). According to the Yale Center for Dyslexia, dyslexia is one of the most common specific learning disorders, accounting for approximately 80% of all cases of learning difficulties.

The study's findings reveal the written language challenges faced by children with dyslexia, including poor handwriting, letter substitution and omission (especially vowels and consonant clusters), spelling difficulties (such as errors in soft signs and distinguishing between short and long vowels), grammatical errors (such as incorrect case usage and application of singular accusative case endings), and sentence structure issues.

Recognizing and addressing the educational needs of children with dyslexia is essential. Parents, teachers, and other specialists must be educated to identify the signs and characteristics of dyslexia and intervene appropriately in cases related to specific learning disabilities (Thomas, Uthaman, 2019). The study highlights key written language characteristics of dyslexic children, which may serve as indicators for early identification and intervention in dyslexia.

### References

- de Bree, Elise, Wijnen, Frank, Gerrits, Ellen 2010. Non-word Repetition and Literacy in Dutch Children At-risk of Dyslexia and Children with SLI: Results of the Follow-up Study. *DYSLEXIA* 16. 36–44. DOI: 10.1002/dys.395
- Çaylak, Emrah 2010. The Studies about Phonological Deficit Theory in Children with Developmental Dyslexia: Review. *American Journal of Neuroscience* 1 (1). 1-12
- Plaza, Monique, Cohen, Henri, Chevrie-Muller, Claude 2002. Oral language deficits in dyslexic children: Weaknesses in working memory and verbal planning. *Brain and Cognition* 48(2-3). 505-512. DOI: 10.1006/brcg.2001.1407
- Thomas, Elizabeth K., Uthaman, Seema P. 2019. Knowledge and Attitude of Primary School Teachers Towards Inclusive Education of Children with Specific Learning Disabilities. *Journal of Social Work Education and Practice* 4(2). 23–32



## **Feuer engine, wrinkelig, and sheepe – The Role of analogy in child bilingual morphological code-mixing**

**Antje Quick<sup>1</sup>, Stefan Hartmann<sup>2</sup>, Nikolas Koch<sup>3</sup> / <sup>1</sup>Leipzig University, Germany, <sup>2</sup>HHU Düsseldorf, Germany, <sup>3</sup>LMU Munich, Germany**

Language contact phenomena such as code-mixing and transfer are ubiquitous in multilingual language use, especially in multilingual first language acquisition (e.g., Nicoladis 2006, Vihman 2018, Author et al. 2021). From a usage-based perspective, children's mixed utterances are often based on productive lexico-grammatical constructions in which open slots in a frame-and-slot pattern are filled with items from the other language (Authors 2022, 2023). Analogy is a key mechanism underlying many of those phenomena, therefore, in this study, we are interested in how far these transfer phenomena can be attributed to analogy.

In particular, we analyze cases of morphological code-mixing such as *sleepen* (*sleep* + German infinitive marker), *fliegig* (German 'fly' + English progressive marker), and *feete* (*feet* + German plural marker) drawing on the naturalistic speech of three German-English bilingual children, "Fion", "Silvie", and "Lily", between the ages of 2 and 3 (n=109,815 utterances). Our main research questions are a) which types of morphological mixes can be found in the children's data, b) how they relate to the children's overall use of the two languages, c) to what extent the mixes can be explained by direct analogy to phonologically similar words in the source language, and d) which individual differences we find between the children, and whether and how they can be related to different input situations. We extracted all code-mixed utterances (n=10,040) and checked whether they contained a morphological mix, e.g. an English verb with a German inflectional marker or vice versa. The mixes were then annotated for whether the word is a cognate of the corresponding word in the other language, and the phonological distance between the word form and the cognate word form was assessed by calculating the Levenshtein distances between the SAMPA (Wells 1997) transcriptions of both words using a combination of automatic annotation (BAS services, Reichel & Kisler 2014; PLD2flex, Wedig et al. 2024) and manual correction. We predict that forms with a close formal correspondence to a cognate or a word with similar meaning in the other language will be more prone to occur in morphologically mixed forms.

Preliminary results show that phonological similarity plays a role, especially in mixed compounds (e.g. *oster bunny* 'easter bunny') whereas inflectional mixes are mostly driven by selecting the most entrenched frame-and-slot patterns that mirror each child's language proficiencies (e.g. *I am X-ing* as in *I am kaufing* 'I am buying').

### **References**

- Nicoladis, Elena. 2006. Cross-linguistic transfer in adjective–noun strings by preschool bilingual children. *Bilingualism: Language and Cognition* 9(1). 15–32.
- Reichel, Uwe D. & Thomas Kisler. 2014. Language-independent grapheme-phoneme conversion and word stress assignment as a web service. In Rüdiger Hoffmann (ed.), *Elektronische Sprachverarbeitung 2014*, 42–49. Dresden, Germany: TUDpress.
- Reichel, Uwe D. & Thomas Kisler. 2014. Language-independent grapheme-phoneme conversion and word stress assignment as a web service. In Rüdiger Hoffmann (ed.), *Elektronische sprachverarbeitung 2014*, vol. 71, 42–49. Dresden, Germany: TUDpress. <http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-22780-1>.
- Vihman, Virve-Anneli. 2018. Language Interaction in Emergent Grammars: Morphology and Word Order in Bilingual Children's Code-Switching. *Languages* 3(4). 40.
- Wedig, Helena, Felix Theodor, Joshua Wieler & Eva Belke. 2024. PLD2flex: Establishing the Phonological Levenshtein Distance for Pairs or Groups of (Pseudo)Words. *Journal of Open Research Software* 12. 12. <https://doi.org/10.5334/jors.510>.





- Wedig, Helena, Felix Theodor, Joshua Wieler & Eva Belke. 2024. PLD2flex: Establishing the Phonological Levenshtein Distance for Pairs or Groups of (Pseudo)Words. *Journal of Open Research Software* 12. 12. <https://doi.org/10.5334/jors.510>.
- Wells, John C. 1997. SAMPA computer readable phonetic alphabet. In Dafydd Gibbon, Roger Moore & Richard Winski (eds.), *Handbook of standards and resources for spoken language systems*, 684–738. Berlin, New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110804508-008>.

## **Morphosyntactic Complexity in Lithuanian Noun-Adjective Agreement: Insights from Bilingual Children and Non-Native Adults**

**Jonas Stepšys, Laura Kamandulytė-Merfeldienė (Vytautas Magnus University)**

**Keywords:** morphosyntax, agreement, bilingualism, second language acquisition, Lithuanian language

Lithuanian is distinguished by its complex morphological system and extensive inflectional patterns. Adjectives follow three distinct paradigms for masculine and feminine forms, while nouns fall into twelve paradigms. Some adjective paradigms are iconic to noun paradigms (e.g., mėlyn-as ‘blue-ADJ’ and nam-as ‘house-NOUN’). Adjectives in Lithuanian must align with the grammatical categories of the nouns they modify (e.g., skan-us ‘delicious-MS-SG-NOM’ and maist-as ‘food-MS-SG-NOM’), making noun-adjective agreement a crucial morphosyntactic feature of the language.

Previous research has shown that Lithuanian-English bilingual children and adult learners of Lithuanian as a foreign language struggle with noun-adjective agreement, leading to frequent errors (Blažienė, 2015; Dabašinskienė & Čubajevaitė, 2009; Bružaitė-Liseckienė, 2017; Stepšys & Kamandulytė-Merfeldienė 2024). Studies indicate that bilingual children whose second language (L2) shares grammatical categories with their first language (L1) perform better in noun-adjective agreement than those whose L1 lacks such features (Schwartz et al., 2015).

This study aims to identify which morphosyntactic features of noun-adjective agreement pose the most significant difficulties for bilingual children and non-native adults.

The research was based on the Theory of Natural Morphology (Dressler et al., 1987) and the Usage-Based Model (Tomasello, 2000). An experimental task was designed using data from a corpus of children’s language and child-directed speech. The task included a diverse set of sentences representing different adjective and noun declension paradigms, as well as grammatical variations in number (singular/plural), gender (masculine/feminine), and case (six cases).

Participants were presented with multiple-choice questions in which they selected one answer from four options: the correct answer, a paradigm error (inflection from a different paradigm), a gender error (incorrect gender within the target paradigm), and a number error (incorrect number within the target paradigm). The order of the answer choices was randomized for each item to prevent response bias. This methodology allowed for an in-depth examination of how participants process inflections.

The study involved two participant groups: 35 bilingual children (aged 6–12) who are native speakers of Lithuanian and 25 adult learners (aged 18–72) studying Lithuanian as a foreign language.

The findings indicate that both bilingual children and non-native adults face significant challenges with noun-adjective agreement due to Lithuanian’s highly inflectional nature. Errors primarily stem from overgeneralizing productive paradigms, replicating noun inflections in



adjective forms, and favouring unmarked grammatical categories over marked ones. Furthermore, agreement difficulties are closely linked to the learner's first language, with challenges observed among those whose L1 is an analytic language.

## References

- Blažienė Agnė 2015. Dvikalbių vaikų rišliojo pasakojimo produktyvumas ir leksinė įvairovė [Narrative of bilingual children: general productivity and lexical diversity]. *Taikomoji kalbotyra / Applied Linguistics* 7. 1–24. <https://doi.org/10.15388/TK.2015.17492>
- Bružaitė-Liseckienė, Justina 2017. Lietuvių kalbos kaip svetimosios semantiškai motyvuotos daiktavardžių giminės atpažinimo ir skyrimo klausimai [On the assignment of natural gender in Lithuanian as a second language]. *Taikomoji kalbotyra / Applied Linguistics* 9. 1–30. <https://doi.org/10.15388/TK.2017.17445>
- Dabašinskienė, Ineta, Laura Čubajevaitė 2009. Acquisition of case in Lithuanian as L2: Error analysis. *Eesti Rakenduslingvistika Ühingu aastaraamat* 5. 47–65. <http://dx.doi.org/10.5128/ERYa5.04>
- Dressler, Wolfgang, Ullrich, Willi Mayerthaler, Oswald Pangl, Wolfgang Ullrich Wurzel 1987. *Leitmotifs in Natural Morphology*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Schwartz, Mila, Miriam Minkov, Elena Dieser, Ekaterina Protassova, Victor Moin, maria Polinsky 2015. Acquisition of Russian gender agreement by monolingual and bilingual children. *International journal of bilingualism* 19. 726–752. <https://doi.org/10.1177/1367006914544989>
- Stepšys, J., Laura Kamandulytė-Merfeldienė 2024. Būdvardiškųjų žodžių derinimo sunkumai: eksperimentinis dvikalbių vaikų ir negimtakalbių suaugusiųjų kalbos tyrimas [Challenges in adjective agreement: an experimental study on bilingual children and adults learning Lithuanian as a foreign language]. *Darbai ir dienos / Deeds and Days* 81. 117–136. <https://doi.org/10.7220/2335-8769.81.7>
- Tomasello, Michael 2000. First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* 11. 61–82. <https://doi.org/10.1515/cogl.2001.012>

## The use of modal verbs and adverbs by Estonian school-leavers in their English state examination essays

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

The Common European Framework of Reference for Languages (CEFR) defines six proficiency levels (A1–C2) in a language-neutral way but does not specify which linguistic elements learners use at each level. Despite its widespread adoption, the CEFR has been criticized for lacking empirical validation based on authentic learner language, weak links to SLA theories, and reliance on subjective descriptors (Wisniewski 2017; Hulstijn 2014; Hawkins and Filipović 2012; North 2007).

To refine CEFR for English, the English Profile Project (EPP) identifies grammatical features characteristic of each level using the 45-million-word Cambridge Learner Corpus (CLC). EPP provides a data-driven progression of grammatical development but concerns remain: the corpus is unbalanced, does not fully account for L1 influence, and has limited links to SLA theories (Chen and Baker 2014; Ilc and Stopar 2014; Juknevičienė and Šeškauskienė 2014). These issues raise the question of whether EPP-defined progressions apply to learners from understudied L1 backgrounds such as Estonian.



In Estonia, English proficiency of upper secondary school leavers is assessed by means of a state examination. Despite the examination's role in English proficiency certification, school-leavers' linguistic performance at it has not been systematically compared with CEFR descriptors or EPP findings. The extent to which Estonian learners' English conforms to EPP-defined grammatical progressions remains unexplored.

This study is part of a larger project that examines whether Estonian learners' grammatical features appear in a similar order and complexity to those in EPP findings. The study evaluates whether Estonian learner English follows EPP-identified progressions of modal verbs and adverbs across proficiency levels. Analysing a 230,000-word corpus of state examination essays, we address the following research questions: (1) How do Estonian learners' modal verb usage patterns align with EPP findings at levels A2, B1, and B2?; (2) How does modal verb and adverb usage improve with advancement to higher proficiency levels? (3) Which type of modality do Estonian learners use most?.

The study is based on a three-subcorpus dataset (A2, B1, B2), classified by examination scores. The corpus will be analysed using AntConc to examine modal verb/adverb frequency distributions, co-occurrence patterns to assess modality types, and comparisons with EPP descriptors to determine alignment or divergence. This study aims to provide the first empirical validation of EPP findings on modal verb usage in Estonian learner English, informing language assessment and CEFR-based instruction for Estonian learners.

## References

- Chen, Y-H., Baker, P. 2014. Investigating Criterial Discourse Features across Second Language Development: Lexical Bundles in Rated Learner Essays, CEFR B1, B2 and C1. *Applied Linguistics* 37: 6, 849–880.
- Hawkins, J.A., L. Filipovic. 2012. *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework*. English Profile Studies I. Cambridge: Cambridge University Press.
- Hulstijn, J. H. 2014. The Common European Framework of Reference for Languages. A challenge for applied linguistics. *International Journal of Applied Linguistics* 165:1, 3–18.
- Ilic, G.; A. Stopar 2014. Validating the Slovenian national alignment to CEFR: The case of the B2 reading comprehension examination in English. *Language Testing* 1, 1-20.
- Juknevičienė, R., I. Šeškauskienė 2014. The national examination of English in Lithuania: searching for evidence of CEFR criterial achievement levels. *Kalbų Studijos* 25, 88-96.
- North, B. 2007. The CEFR Illustrative Descriptor Scales. *The Modern Language Journal* 91: 4, 656-659.
- Wisniewski, K. 2017. Empirical learner language and the levels of the Common European Framework of Reference. *Language Learning* 67:S1, 232–253.

## Towards a typology of mixed copies

**Anna Verschik, Helin Kask (Tallinn University, Estonia)**

Among various models of codeswitching and contact-induced language change the Code-Copying Framework (CCF) developed by Johanson (2002) remains one of the less known. Yet CCF accurately describes a phenomenon that has been overlooked in other models. In CCF, the mental mechanism is not “borrowing” or “transfer” but copying. Properties of an item can be copied globally (form + meaning), selectively (meaning, combinational rules etc.) and in a mixed form (mixed copies, MC): English *time span* yields Estonian *aja spänn*, where the



combination of items and overall meaning comes from English, one component is a global (*spänn*) and the other a selective copy (*aeg* 'time', genitive *aja*).

MCs can be of various types: they can occur in compounds, analytical verbs but also in idiomatic expressions that may look like analytic verbs (cf. Estonian *eksmatti saama* – local Russian *получить эксма́т*). A typology of Estonian-Russian MCs has been suggested (Verschik 2016), where borderline cases exist depending on the (phonetic) realization of common internationalisms. Therefore, it would be desirable to look into MCs in other language pairs.

The aim of the study is to identify English-Estonian MCs. The data comes from the vlogs and podcasts of young Estonians (30 hours of footage, ~ 270,000 tokens). There are 3,618 copies altogether (13 copies per 1,000 words), 184 are MCs. All MCs in the data are compound nouns like *nude-roosa* 'nude pink', Estonian *heleroosa*), *reality-saade* 'reality show', Estonian *tõsielusaade*. Compounding is a productive mechanism in both languages. MCs of phrasal verbs like *kukkus off* 'fell off' (Estonian *kukkus maha*, English *fell off*) are absent in our data but attested elsewhere in Estonian-English bilingual speech (Verschik *et al.* 2024). No MCs within frame-and-slot constructions *X to be* (*ema to be* 'mother to be') have been attested, nor in constructions like *X-driven*, *X-related* (*raha-driven*, English *money-driven*, Sell 2024: 41). Other Estonian-English bilingualism data have to be analysed in order to obtain a full typology of MCs.

## References

- Johanson, L. 2002. Contact-Induced Change in a Code-Copying Framework. – Language Change: The Interplay of Internal, External and Extra-Linguistic Factors, vol 86, Contributions to the Sociology of Language. Ed. M. C. Jones, E. Esch. Berlin: de Gruyter Mouton, 285–314.
- Sell, M. 2024. Eesti-inglise koodivahetuse tüpoloogia kakskeelsetes vlogides. MA thesis. Tallinna Ülikool, humanitaarteaduste Instituut.
- Verschik, A. 2016. Mixed copying in blogs: Evidence from Estonian-Russian language contacts. – Journal of Language Contacts 9 (1), 186–209.
- Verschik, A., Baird, P., Argus, R. 2024. Contact phenomena in English-Estonian early bilingualism: Application of Triangle model. Presentation at LaCoLa network meeting, Leipzig.