

## Eesti Keele Instituudi 2020. aasta TEGEVUSARUANNE

Töötajaid: 85, neist akadeemilisi töötajaid 64

Direktor: Arvi Tavast, tel 617 7500, arvi.tavast@eki.ee

Teadusjuht: Tiina Laansalu, tel 617 7502, tiina.laansalu@eki.ee

Alates 1. detsembrist 2020 on Eesti Keele Instituudis viis struktuuriüksust:

- keeleajaloo, murrete ja soome-ugri keelte osakond (juhataja Tiina Laansalu)
- keeletehnoloogia kompetentsikeskus (al 01.12.2020 juht Susanna Oja)
- koolituskeskus
- kõneuurimise ja kõnetehnoloogia osakond (juhataja Meelis Mihkla)
- tänapäeva eesti keele osakond (juhataja Margit Langemets)
- üldosakond

Kuni 1. detsembrini 2020 oli Eesti Keele Instituudis seitse struktuuriüksust:

- eesti keele ajaloo, murrete ja soome-ugri keelte osakond (juhataja Tiina Laansalu)
- keeleressursside ja keeletehnoloogia osakond (juhataja kt Tõnis Nurk)
- kõneuurimise ja kõnetehnoloogia osakond (juhataja Meelis Mihkla)
- keelekorraldusosakond (juhataja Peeter Päll)
- sõnaraamatute osakond (juhataja Margit Langemets)
- terminoloogiaosakond (kuni 11.10.2020 juhataja Tiina Soon)
- personali- ja haldusosakond (juhataja Kai Oro)

Aruanne järgib 1. detsembrile 2020 järgnenud struktuuri (v.a tegevust alles alustanud keskused).

### Keeleajaloo, murrete ja soome-ugri keelte osakond

**1. Personaalne rühmagrant PRG34 „Võimusuhted ja identiteet 16.–18. sajandi eesti kirjakeeles“ (2018–2022).** Juht Kristiina Ross, põhitäitjad Kai Tafenau, Piret Lotman, Annika Viht, täitjad Heiki Reila, Ahti Lohk. Aastal 2020 ilmus üks 3.2 artikkel (Kristiina Ross „Aja tõlkimisest“). Jätkus tekstikriitilise kogumiku „Eesti keelemõte 1632–1732“ koostamine ja toimetamine (Kristiina Ross, Kai Tafenau, lepinguliselt ka Aivar Põldvee), seoses ulatuslike täiendustega lükkus ilmumine 2021. aastasse; kogumiku ainetel avaldati üks 3.2 artikkel (Põldvee, Reila, Ross, Tafenau 2020). Alustati 16.–18. sajandi eesti asjaajamisekeele uurimist. Grandiga on seotud veel kolm suuremat allteemat: ajalooline piiblisõnastik, esimese täispiibli tõlkekäsikirja lähivaatlus, kirikulaulude konkordants.

**2. Eesti piiblitõlke ajalooline sõnastik (paberväljaanne).** Eelmisel aastal valminud digitaalse konkordantsi (<http://www.eki.ee/piibel/>) põhjal valmistatakse ette liigendatud märksõnastikuga paberversiooni. Töörühma kuuluvad Kristiina Ross (juht), Inge Käsi, Maeve Leivo, Ahti Lohk, Heiki Reila, Anu Pedaja-Ansen. Märksõnastikuga jõuti märksõnani *seitsmekesi*. Kuivõrd kõik töörühma liikmed on väljaande koostamisse kaasatud vaid osalise koormusega, on töö kulgenud plaanitud aeglasemalt. Praeguse seisuga loodetakse sõnastik kirjastusele üle anda 2022. aasta alguses.

**3. Eesti esimese piiblitõlke käsikirja tutvustava publikatsiooni ettevalmistustööd.** Töörühma kuuluvad Kristiina Ross (juht), Annika Kilgi, Maimu Liiv, Tiiu Lagle. 2020. aastal alustati Esimese Moosese raamatu ja Psalmide raamatu mustandkäsikirja alg- ja lõppversiooni lemmatiseerimist ning morfoloogilist märgendamist (Maimu Liiv, Tiiu Lagle). Lemmatiseeriti ja märgendati 65% tekstist.

Alustati eri tõlkijate murdeliste iseärasuste lähivaatlust (Annika Kilgi). Töö valmib 2022. aasta lõpuks.

**4. Ajalooliste kirikulaulutõlgete konkordants ja sõnaraamat.** Töörühma kuuluvad Kristiina Ross (juht), Heiki Reila, Piret Lotman, Sigrid Saagpakk, Ahti Lohk. Jätkus ajalooliste kirikulaulutõlgete süstematiseerimine ja nende tõlkekeele analüüsimine. Alustati saksa originaalide lisamist konkordantsi juurde (sisestatud, konkordantsi ehitusele vastavalt märgendatud ja toimetatud on kolm 16.–17. sajandi saksakeelset lauluraamatut).

**5. Personaalne järeldoktori grant PUTJD842 „Protoeuroopa substraat lõunaläänemeresoome keelte sõnavaras“ (2019–2021).** Uuriti substraatkeelte kindlakstegemise metodoloogiat ning rakendati seda liivi keele ja eesti murrete tüvevara uurimisel. Valmimas on kaks artiklit lõunaläänemeresoome keelte võimaliku substraadse tüvevara iseloomulikest joontest. (Sven-Erik Soosaar)

**6. Eesti murrete sõnaraamat.** Valmis sõnaraamatu 31. vihik (*osatlema–pari*), mis ühtlasi avas VII köite. Kogu seni (alates 1994. aastast) ilmunud materjal (*a–pari*) on kasutatav ka sõnaraamatu veebiversioonis (<http://www.eki.ee/dict/ems/>), mis täieneb uute osade valmisaamisega. Jätkus sõnaraamatu koostamine ja toimetamine: toimetati eesti murrete sõnaraamatut vahemikus (*parigas–piiastama*). Toimetatud osa põhjal ilmub aastal 2021 sõnaraamatu 32. vihik. Sõnaraamatu toimetajad on Mari Kendla (töörühma juht) ja Mari-Liis Kalvik. Lisaks koostati 101 lk sõnaraamatu käsikirja (*soonevai–suruma*), koostamisega tegeleb Hanna Pook.

**7. Etümoloogiasõnaraamat.** Jätkati akadeemilise etümoloogiasõnaraamatu koostamist ja toimetamist (peatoimetaja Iris Metsmägi, koostajad Meeli Sedrik, Vilja Oja, Udo Uibo, lepinguline koostaja Lembit Vaba, assistent Anna Leontjeva). Koostamise lõpu märges lisati umbes 1300 märksõnale. 2020. aasta lõpuks on koostatud umbes 5950 märksõnaartiklit. Täiendati etümoloogilise kirjanduse elektroonilist kogu, sh viidete kogu, täiendati ja toimetati kirjandusallikate üldnimekirja

**8. Piirkondlikud murdesõnaraamatud.** Koostöös teiste asutustega (Tartu Ülikool, Mulgi Kultuuri Instituut, Seto Instituut) koostati järgmisi piirkondlikke murdesõnaraamatuid:

- 1) Ilmus (nii paberraamatuna kui ka veebis) „Seto eripäraste sõnade sõnaraamat“ (<http://www.eki.ee/dict/setoeri/>), koostajad Eva Saar, Paul Hagu, Inge Käsi, Maeve Leivo, Hanna Pook ja Karl Pajusalu, toimetajad Eva Saar ja Jüri Viikberg. Sõnaraamat sisaldab üle 3600 märksõnaartikli.
- 2) Seto suursõnaraamat (koostamine EELEXIS; EKI-poolsed koostajad Inge Käsi, Maeve Leivo, Hanna Pook). Koostati märksõnavahemik *puian–pääl* ja ühistoimetati „Seto eripäraste sõnade sõnaraamatu“ märksõnade osa (kattuvad EELEXIS andmebaasis).
- 3) Suur Mulgi sõnastik
- 4) Saaremaa sõnastik
- 5) Lüganuse sõnastik

Sõnaraamatuid koostati EELEXIS, töid koordineeris Jüri Viikberg.

**9. Eesti-komi sõnaraamat.** Koostati 2800 märksõnaartiklit (Nikolay Kuznetsov, Sven-Erik Soosaar), toimetati 4500 märksõnaartiklit (Anu-Reet Hausenberg, Sven-Erik Soosaar). Sõnastik valmib jaanuaris 2021 ja avaldatakse EKI veebilehel (<http://www.eki.ee/dict/komi/>). Koostamist kaasrahastab riiklik hõimurahvaste programm.

**10. Ingerisooime keskmurde sõnaraamat.** Koostati käsikirja 4000 märksõnaartiklit (lepinguliselt Taisto-Kalevi Raudalainen). Käsikirja valmistati ette Ekilexi viimiseks. Koordineeris Sven-Erik Soosaar.

**11. Saksa laenude leksikon.** Jätkus saksa laenude leksikoni koostamine, EELEXIS baasi „Ülemsaksa laenud“ sisestati 610 märksõnaartiklit. Käsikiri on plaanis valmis saada 2021. aasta lõpuks. (Jüri Viikberg)

**12. Eesti slängi andmebaas.** Vangla-, sõduri- ja vangilaagrite slängisõnavara kohta koostati EELexis 2950 märksõnaartiklit. (Marit Alas)

### 13. EMSUKA

- Korrastati eesti murrete ja soome-ugri keelte arhiivi (EMSUKA) andmebaasi (<http://emsuka.eki.ee>) 2293 kirjet, loodi 77 uut kirjet, lisati 543 faili.
- Korrastati EMSUKA heliarhiivi digitaalfaile 248 tunni ulatuses, tehti kordusdigiteeringuid 40 tunni ulatuses. Digiteeriti käsikirjakogusid 2442 lk ulatuses.
- Arhiivi lisandus suur kogus Henn Saari pärandit: umbes 50 000 sedelit sõnavara ja umbes 100 tunni ulatuses helisalvestisi. Samuti lisandus 612 fotot Eesti Keele Instituudi üritustest.
- Koostöös Eesti Teaduste Akadeemiaga ja Venemaa Teaduste Akadeemia Karjala kirjanduse ja ajaloo uurimiskeskuse keeleteaduse instituudiga kontrolliti ja tänapäevastati EMSUKA arhiiviviited teose „Народные песни Ингерманландии“ (1974) uusväljaande jaoks.
- EMSUKA materjale kasutati näitustel ja üritustel Kodavere Pärimuskeskuses ning Ludza ja Kolka muuseumis.
- Registreeriti 35 külastust väljastpoolt instituuti. Uurijaid nõustati umbes 60 tunni ulatuses telefoni ja veebi teel (sh Läti, USA, Venemaa) ning vahendati ligipääs umbes 150 failile.
- Veebiandmebaasi külastati 1348 korda.
- Inventeeriti ja sünkroniseeriti servereid.
- EMSUKAt haldab Liis Ermus.

**14. Soome-ugri keeleteaduse seminari korraldamine.** Korraldati rahvusvaheline seminar „Soome-ugri keelekogud, korpused ja andmebaasid meie kasutuses“ (12.11.2020). Seminar toimus veebi vahendusel, peeti 8 ettekannet. Seminari kava leiab [siit](#). (Liis Ermus, Tiina Laansalu)

**15. Teaduspublikatsioonid ja -ettekanded.** Ilmus 8 kõrgetasemelist teadusartiklit (1.1 – 7; 3.1 – 1), 2 monograafiat, 1 dissertatsioon, 2 kõrgkooliõpikut/käsiraamatut, 7 muud kogumikuartiklit (3.2 – 6, 3.4 – 1), toimetati 3 väljaannet ning esineti 13 ettekandega. Täpsemad andmed ETISes.

### 16. Välis- ja ühisprojektid

- Osaleti Eesti-uuringute Tippkeskuse töös. (Kristiina Ross)
- Osaleti Eesti Keele Instituudi, Võru Instituudi ja Eesti Kirjandusmuuseumi ühisprojektis „Eesti perekonnanimeraamat“. (Marit Alas, Marja Kallasmaa, Tiina Laansalu, Peeter Päll)
- Koostöös Kultuuriministeeriumi ja ERRiga koostati „Digikultuuri aasta 2020“ raames emakeelenädala veebiviktoriin „Kas tunned Eesti keelepärandit?“. (Tiina Laansalu)
- Etümoloogiaalane koostöö Eesti Rahva Muuseumiga virtuaalnäituse „Moekad retropühadekaardid“ jaoks. (Iris Metsmägi, Meeli Sedrik)
- Osaleti rahvusvahelises maastikusõnavara töörühmas Landlex. (Iris Metsmägi, Vilja Oja)
- Osaleti rahvusvahelises ühisprojektis Euroopa keelettas Atlas Linguarum Europae (ALE). (Vilja Oja)
- Osaleti Euroopa Nõukogu keeleoskustasemetest lähtuvate eesti keele tasemetestide Innove C1-taseme arenduskomisjoni töös. (Tiina Laansalu)
- Tehti Fenno-Ugria Asutusega koostööd hõimuklubi seminaride korraldamisel. (Tiina Laansalu, Kristiina Ross, Sven-Erik Soosaar)

### 17. Muu teadus- ja arendustegevus

- Peeti loengukursusi (Liis Ermus, Nikolay Kuznetsov, Fred Puss, Annika Viht, Jüri Viikberg), üksikloenguid (Mari-Liis Kalvik, Tiina Laansalu, Hanna Pook, Fred Puss, Kristiina Ross, Jüri Viikberg) ja koolitusi (Fred Puss, Meeli Sedrik).
- Juhendamisel kaitsti 1 doktoritöö (Marja Kallasmaa) ja 1 magistratöö (Annika Viht).
- Juhendati 2 doktoritööd (Marja Kallasmaa, Nikolay Kuznetsov, Jüri Viikberg), 2 magistratööd (Liis Ermus, Nikolay Kuznetsov) ja 1 bakalaureusetööd (Fred Puss).
- Oponeeriti 1 magistratööd (Nikolay Kuznetsov).

- Eelretsenseeriti teadusartikleid (Marja Kallasmaa, Nikolay Kuznetsov, Tiina Laansalu, Fred Puss, Kristiina Ross) ja monograafiat (Kristiina Ross). Retsenseeriti „Otto Wilhelm Masingu „Marahwa Näddala-Lehhe“ sõnastikku“ (Annika Viht).
- Korraldati osakonna väljaannete esitlusüritus 26.08.2020 (Tiina Laansalu), osaleti EKI suveseminari ja EKI doktorikooli sügisseminari korraldamisel (Liis Ermus, Tiina Laansalu).
- Nõustati murdekeele, etümoloogia jm keeleküsimuste korral (Mari-Liis Kalvik, Mari Kendla, Tiina Laansalu, Iris Metsmägi, Vilja Oja, Meeli Sedrik, Sven-Erik Soosaar, Jüri Viikberg) ja EMSUKA kasutajaid ning külastajaid (Liis Ermus).
- Osaleti mitmetes (nii Eesti kui ka rahvusvaheliste institutsioonide) komisjonides, nõukogudes, tööruhmades.
- Tegeleti erinevate teadusadministratiivsete ülesannetega (akadeemiliste töötajate atesteerimine, direktorivalimiste korraldamine, ETISE administreerimine jne). (Tiina Laansalu)

**18. Kaitstud doktoritööd.** Tiina Laansalu „Kesk-Harjumaa kohanimedede kujunemine ja nimesiire“, juhendajad Marja Kallasmaa ja Karl Pajusalu, Tartu Ülikool, humanitaarteaduste ja kunstide valdkond, eesti ja üldkeeleteaduse instituut.

**19. Doktorandid.** Osakonnas on 4 doktoranti: Marit Alas (Tallinna Ülikool), Liis Ermus (Tartu Ülikool), Hanna Pook (Tartu Ülikool) ja Fred Puss (Tartu Ülikool).

## **20. Teaduse populariseerimine**

- Avaldati 16 populaarteaduslikku artiklit (Liis Ermus, Mari-Liis Kalvik, Mari Kendla, Nikolay Kuznetsov, Tiina Laansalu, Iris Metsmägi, Vilja Oja, Fred Puss, Sven-Erik Soosaar), 8 lühikirjutist teadusajakirjades/-kogumikes (Liis Ermus, Tiina Laansalu, Fred Puss, Kristiina Ross, Jüri Viikberg) ja 3 raamatuarvustust (Kristiina Ross, Jüri Viikberg).
- Osaleti (EKI ja ERRi koostööna) populaarteaduslike teleklippide (<http://portaal.eki.ee/videod.html>) valmistamisel. Valmis 5 keelekildu (Tiina Laansalu, Iris Metsmägi, Fred Puss, Meeli Sedrik, Jüri Viikberg).
- Esineti raadio- ja telesaadetes (Mari Kendla, Fred Puss, Jüri Viikberg), ning peeti mitmeid populariseerivaid ettekandeid ja loenguid (Kristiina Ross).
- Osaleti teaduste akadeemia konkursil „Teadus kolme minutiga“ (Fred Puss).

## **21. Teaduspreemiad ja tunnustused**

- Ajakirja Keel ja Kirjandus 2019. aasta parim lühitekst: Meeli Sedrik, Udo Uiho „Trips, traps, trull, sina oled kull“
- HTMi konkursi „Keeletegu 2019“ nominendid:
  - Murdeportaali „Murdekiiker“ (koost. Jüri Viikberg, toim. Mari-Liis Kalvik ja Tiina Laansalu, tehniline teostus Marko Petron ja Indrek Hein)
  - „Kuusalu rannakeele sõnaraamat“ (koost. Heli Kendra ja Piret Norvik, toim. Vilja Oja)
  - Eesti Keele Instituudi keelekooliartiklite sari

## **Kõneuurimise ja kõnetehnoloogia osakond**

Osakond keskendus oma viiendal tegutsemisaastal eelkõige kõne uurimisele ning eestikeelse kõnesünteesi arendamisele ja selle rakendustele. 2020. aastal täitsime kahe uurimisprojekti, kahe arendusprojekti, kahe baasfinantseerimisprojekti ning ka EKI tegevuskava ülesandeid.

**1. Eestikeelset kõnet uuriti institutsionaalse uurimistoetuse projekti ja Eesti uuringute tippkeskuse raames. Kõnestiilid, lauseprosoodia ja fonoloogiline varieerumine: kirjeldus, teooria ja modelleerimine (IUT35-1) 2015-2020, Eesti-uuringute tippkeskus – CEES (TK145) 2016–2023,**

teemajuht dr Meelis Mihkla; põhitäitjad: Rene Altrov, Mari-Liis Kalvik, Hille Pajupuu, Liisi Piits, Heete Sahkai; täitjad: Liis Ermus, Kairi Tamuri.

#### **Uurimistulemused ja teadustegevus:**

- 1) Jätkus kõnestiilide uurimine eesmärgiga rakendada tulemusi sünteeskõne väljendusrikkuse suurendamiseks stiilisiirdamise teel. Keskenduti kahele põhilisele eesti reklaamistiilile – energiline ja rahulik. Stiilikorpust täiendati nais- ja meeshääle loetud reklaamidega. Akustiliselt analüüsiti meeshääle loetud reklaame ja tehti kindlaks energilist ja rahulikkude kõnestiili eristavad tunnused. Valmistati ette tajutestide reklaamistiilide eristuvuse kindlakstegemiseks.
- 2) Lauseprosoodia valdkonnas leiti, et lauserõhk mõjutab eesti keele sõnajärge ning et lauserõhkude paiknemine ja lauseprosoodia foneetilised aspektid (kestus, põhitooni kõrgus ja ulatus, häälekvaliteet) eristavad eesti keeles kõneakte.
- 3) Rakendati uut meetodit eesti sõnaprosoodias – väldete analüüs sünteesi teel.
- 4) Fonoloogiliselt varieeruvatest nähtustest uuriti edasi II ja III välte vaheldumist vastavates sihtsõnades. Analüüs põhines 101 keelejuhi materjalil ning selle põhjal selgus mitmes uuritavas sõnatüübis peamine välte-eelistus. Korrastati varieeruvate nähtuse uurimiseks salvestatud häälduskorpust ning valmistati seda ette muudes teadusuuringutes kasutamiseks.

**2. Kõnesünteesi arendustöö jätkus riikliku programmi „Eesti keeletehnoloogia (2018-2027)“ projektide raames: Väljendusrikas ja mitmekesine eestikeelne kõnesüntees (EKTB21) 2018-2022, projektijuht Liisi Piits; põhitäitjad: Rene Altrov, Indrek Hein, Indrek Kiissel, Ahti Lohk, Meelis Mihkla, Hille Pajupuu, Heete Sahkai, ja Kairi Tamuri.**

#### **Põhitulemused:**

- 1) Korpused: Salvestati kolm meeshääle korpust à 3 tundi. Naishääle ilukirjandustekstide korpuses Külli (u 10 tundi) on numbrid, lühendid ja võõrnimed viidud häälduspärasele kujule, kontrollitud heli ja teksti vastavust, otsekõne eraldatud jutustavast kõnest ja märgendatud.
- 2) HTS-häälest on valminud uued versioonid, millele on lisatud numbrite ja lühendite käänamine ning parandatud sulghäälikute grafeem-foneem teisendust.
- 3) Pythonil põhinevad arendussüsteemid (Ossian, Merlin) kohandati numbrite ja lühendite töötluse ja EstNLTK teekidega.
- 4) Hindamine: loodi eeltöötluskriptide võrdluseks ja hindamiseks vajalik keskkond. Välja töötati sünteesihääle ühtsed hindamis põhimõtted.
- 5) Treenitud on kaks stiilimudelit erinevate u 100-lauseliste korpuste peal: rahulik ja energiline reklaamistiil.

**3. Kõnesünteesi jaoks vajaliku kõnekorpuse suurendamine ja kõnesünteesi mudelite arendamine ning hindamine.** Koostööprojekt TÜ arvutiteaduse instituudiga. EKI põhitäitjad: Rene Altrov, Indrek Hein, Hille Pajupuu ja Liisi Piits.

#### **Põhitulemused:**

- 1) Teksti eeltöötluskripti täiendati ja uuenenud tekstitöötlus on paranenud 15 protsendipunkti võrra.
- 2) Hinnati 13 sünteesihääle sobivust uudiste ja ilukirjanduse lugemiseks.
- 3) EKI naishäälekorpuse Külli põhjal TÜ treenitud sünteesihääle osutus ilukirjanduse lugemisel kõige sobivamaks.

#### **4. Baasfinantseerimisprojektid**

- 1) **Häälekorpuse märgendamine, IV etapp.** Projektijuht Hille Pajupuu. Häälekorpust täiendati raadioreklaamidega (lisandus üle 500 reklaami). Korpuse märgendamise hõlbustamiseks loodi korpuste toimetamise süsteem, mis on kasutatav mis tahes kõnematerjali ja teksti vastavuse kontrollimiseks lausete kaupa (Indrek Hein). Et üks osa märgendusest toetub tajutestidele, siis viidi läbi kaks tajutesti vaegnägijate seas hääle ja stiili vastuvõetavuse määramiseks (Artur Rapp).

- 2) **Instituudi institutsionaalse uurimisprojekti „Kõnestiilid, lauseprosoodia ja fonoloogiline varieerumine: kirjeldus, teooria ja modelleerimine“ kaasfinantseerimine.** Projektijuht Meelis Mihkla; põhitäitjad: Rene Altroy, Mari-Liis Kalvik, Hille Pajupuu, Liisi Piits, Heete Sakhai. Projekti raames said kõik teadustöötajad suuremal määral uurimistööle keskenduda, osaleda rahvusvahelistel konverentsidel, tellida teadustöö tarvis kõnematerjali märgendustöid ja artiklite toimetamist ning osta riist- ja tarkvara teadustöö tarvis.

**5. Teaduspublikatsioonid ja -ettekanded.** Ilmus 4 kõrgetasemelist teadusartiklit (1.1 – 2; 3.1 – 2) ning esineti 4 ettekandega.

## 6. Teaduse populariseerimine

- Avaldati 2 populaarteaduslikku artiklit (Mari-Liis Kalvik, Heete Sakhai).
- Osaleti (EKI ja ERRi koostööna) populaarteaduslike teleklippide (<http://portaal.eki.ee/videod.html>) valmistamisel. Valmis 1 keeleklipp (Heete Sakhai).

## 7. Keele- ja kõnetehnoloogiline töö

- Tajutestide loomine, statistiline analüüs:
  - sünteeshääle sobivuse hindamiseks eri stiilis materjali lugemisel (uudis ja audioraamat) nägijatele ja vaegnägijatele. Tulemuste statistiline analüüs
  - reklaamilugemisstiilide vastuvõetavuse/meeldivuse määramiseks
- Reklaamistiilide akustiline analüüs (100 eGeMAPSi parameetrit)
- Stiilisiirdamise katsed (koos Jaan Pajupuuga):
  - audioraamatu mees- ja naistegelase stiili siirdamine (mudelid treenitud ühe meeslugeja hääle põhjal)
  - rahuliku ja enrgilise reklaamistiili siirdamine (mudelid treenitud 120 meeslugeja hääle põhjal) vt <https://github.com/pajupuujh/CycleGAN>
- Uute kõnemudelite treenimine HTS- kõnesünteesiks, uute kõnebaaside kasutamine.
- Üksuste valikul põhineva süntesaatori arendamine.
- Hääldeharjutuste tehnilise lahenduse väljamõtlemine ja näidisharjutuste koostamine.
- Põhitooni parameetrite optimeerimine sünteesiks.
- Leibelite valik närvivõrkudel ja Markovi peitmodelitel kõnesünteesile.
- Lemma- ja vormisagedused koondkorpusele, ports keelekogusid impordiks ette valmistatud.

## 8. Korpuste/andmebaaside ning kogude loomine ja täiendamine

- Kõnekorpuste loomine stiilisiirdamise jaoks (44.1 kHz, 16 Bit, mono):
  - naishääle poolt loetud tegelaskõnekorpused (meestegelane, naistegelane)
  - nais- ja meeshääle poolt loetud reklaamikorpused (rahulik ja energiline kõnestiil)
  - naishääle emotsioonikorpused (kurbus, viha, rõõm)
  - Kõnestiilikorpuse märgendamine tajukatsetega (spontaanne, poolsontaanne ja ettevalmistatud kõne)
- Kõnesünteesi korpuste teksti ja heli sünkroonsuse toimetamine.
- Salvestanud kolme näitlejaga (Tambet Tuisk, Peeter Tammearu ja Indrek Sammul) representatiivse üksiklausete korpuse ja üksiksõnu hääldeharjutuste jaoks.
- Hääldebaasi korrashoid: <http://heliraamat.eki.ee/mate/vaata.php>
- Neologismide baasi korrashoid [Néoveille \(eki.ee\)](http://neoveille.eki.ee) (parool vajadusel) ja koos sellega [morfoloogiabaasi](http://morfoloogiabaasi) täiendamine (2020 umbes 3500 kirjet).
- ÕS1918 Digari linkide parandamine <https://dea.digar.ee/cgi-bin/dea> jääb enamasti 2019 aastas.
- Kogutud, segmenteeritud ning ülekuulatud 30 tundi naishääle ilukirjanduskorpusi.

## 9. Välis- ja ühisprojektid

- Osalemine G. F. Parrot projektis „Surprise Questions from a Comparative Perspective“ 2019–2021 (Anu Treikelder, Tartu Ülikool, Agnès Celle, Pariisi Denis Diderot ülikool).
- Osalemine projektis EKKD10 „Üllatusküsimuste prosoodia ja infostruktuur eesti keeles võrreldes muude keeltega (1.01.2019–31.12.2021)“ (Eva Liina Asu Garcia, Tartu Ülikool)

- Kõneveebi raames koostöö Industry62 OÜ-ga ja Eesti Digiraamatute Keskusega.
- Osalemine EKT koostööprojektis „Eestikeelne kõnesüntees närvivõrkudega“ (partnerid TÜ arvutiteaduse instituut ja EKI), hinnanud erinevate eesti sünteeshäälte sobivust uudiste ja ilukirjanduse lugemisel.
- Osalemine TK145 Eesti-uuringute tippkeskuses kõne ja muusika ning keeletehnoloogia ja digitaalhumanitaaria töörühmas.

## 10. Muu teadus- ja rakendustegevus

Kirjutati järgmiste teadus- ja arendusprojektide taotlused:

- PRG1062 „Kõnesüntees keeleteaduse teenistuses ja *vice versa*“ (mitterahuldatud, kuid hinnatud kõrgete hinnetega)
- ASTRA alategevuse 7 „Koostöö tugevdamine teadus- ja arendusasutuste, kõrgkoolide ja ettevõtete vahel“ projekt „Kõneveeb“ (rahuldatud)
- EKT koostööprojekt „Eestikeelne kõnesüntees närvivõrkudega“ TÜ arvutiteaduse instituudiga (rahuldatud).
- EKTB21 „Väljendusrikas ja mitmekesine eestikeelne kõnesüntees“ (rahuldatud)

## Tänapäeva eesti keele osakond

Osakond moodustati 1. detsembril 2020. Üheks osakonnaks ühendati kolm senist struktuuriüksust:

- keelekorraldusosakond (juhataja Peeter Päll)
- sõnaraamatute osakond (juhataja Margit Langemets)
- terminoloogiaosakond (juhataja Tiina Soon)

**1. Ekilex 2020.** Eesti Keele Instituudi sõnastiku- ja terminibaas. Hein, Indrek; Männiko, Kaur; Kallas, Jelena; Koppel, Kristina; Langemets, Margit; Nurk, Tõnis; Plado, Merily; Vaus, Mari; Viks, Ülle; Tavast, Arvi; Laubre, Martin; Sharma, Yogesh; Niilo, Hardi. Arendaja OÜ TripleDev.

**2. Sõnaveeb 2020.** Sõnaveeb <https://sonaveeb.ee/> on Eesti Keele Instituudi keeleportaal, kuhu on koondatud info instituudi paljudest sõnakogudest ja andmebaasidest. Alates 2020. aastast esitab üldkeele infot „EKI ühendsõnastik 2020“, oskuskeele osas täiendavad ühendsõnastikku kümned oskussõnastikud ja terminibaasid. Eesti keele õppijale on mõeldud Keeleõppija Sõnaveeb. Sõnaveebis kuvatav info pärineb Ekilexist, uuest Eesti Keele Instituudi hallatavast ja arendatavast sõnastiku- ja terminibaasisüsteemist. Sõnaveebiga on seotud ka „Eesti keele veebilauseite korpus 2020“. Sellest korpusest päritakse veebilauseid korpuspäringusüsteemi **KORP API** kaudu. Arendaja OÜ TripleDev. Hein, Indrek; Männiko, Kaur; Kallas, Jelena; Koppel, Kristina; Langemets, Margit; Nurk, Tõnis; Plado, Merily; Vaus, Mari; Viks, Ülle; Tavast, Arvi; Laubre, Martin; Sharma, Yogesh; Niilo, Hardi.

**3. Keeleõppija Sõnaveeb 2020.** Keeleõppijatele mõeldud keeleportaal <https://sonaveeb.ee/lite>. Keeleõppija Sõnaveebis on umbes 5000 sõna, mida on seletatud lihtsas keeles ning mis aitavad nii algajal kui edasijõudnud keeleõppijal (eeskätt A2 ja B1 keeleoskustase) eesti keeles õigesti rääkida ja kirjutada. Kui Keeleõppija Sõnaveebi valikus küsitud sõna puudub, suunatakse kasutaja Sõnaveebi info juurde. Arendaja OÜ TripleDev. Jelena Kallas, Kristina Koppel, Katrin Tsepelina.

**4. EKI ühendsõnastik 2020.** Eesti Keele Instituudi uusim sõnaraamat, mille tuumiku moodustavad 2019. aastal Sõnaveebis ilmunud Eesti Keele Instituudi üldkeele sõnakogud: „Eesti keele sõnaraamat 2019“, „Eesti keele naabersõnad 2019“, „Eesti keele põhisõnavara sõnastik 2019“ (1. trükk 2014), „Eesti-vene sõnaraamat 2019“ (1. trükk 1997–2009) ja „Eesti Keele Instituudi eesti keele morfoloogiline andmebaas 2019“. Ühendsõnastikku täiendatakse jooksvalt, sõnastiku versioon ilmub kord aastas (versiooni näitab aastaarv).

- 1) Jätkus sünonüümide infokihi lisamine (kokku 3062 sõna) (Kristina Koppel, Maria Tuulik, Tiina Leemets).
- 2) Alustati ÕSi info esitamist Sõnaveebis (rubriigis „ÕS soovitab“, kokku ca 400 soovitust). Järgmine ÕS ilmub ühendsõnastiku osana. Alustati keelekorralduse põhimõtete uuendamise seminaridega (teemasid käsitletakse valdkondade kaupa) (Maire Raadik, Sirje Mäearu, Tiina Paet, Tiina Leemets, Peeter Päll, Tiia Valdre, Mai Tiits, Margit Langemets).
- 3) Jätkus eesti-vene infokihi koostamine/korrastamine Ekilexis (Katrin Tsepelina, Hanna Tammik, Alina Smirnova, Edgar Sai, Olga Kiisla, Kristina Matt, Valentina Tubin, Maia Melts).
- 4) Alustati kohanimede baasi info ületoomist (Peeter Päll).

**5. Oskussõnastikud.** Alates 2020. aastast said Sõnaveebis kasutatavaks kümned oskussõnastikud ja terminibaasid (kokku umbes 70). Enamasti on need sõnakogud, mis on koostatud varasemas, EKI Termeki andmebaasis, kus aga termineid polnud võimalik ühisotsinguga leida. Juurde on tulnud 2019. aastal toetust saanud terminikomisjonide sõnakogud. Ekilexi ei ole veel jõudnud EKI mitmekeelne terminibaas Esterm.

**6. Terminibaas Esterm.** EKI mitmekeelne terminibaas Esterm koondab üle 50 valdkonna terminoloogiat (umbes 53 000 mõistet, eestikeelseid termineid ligi 66 000, ingliskeelseid umbes 72 000). Uusi kirjeid: 400; muudetud kirjeid: 1100; lisatud termineid: 1500. Ühtlustati IATE ja Estermi terminibaasi kirjeid koostöös Euroopa Liidu tõlkeasutustega. 2021. aastal alustatakse Estermi auditiga, et otsustada, mis info Ekilexi baasi üle võtta. (Merily Plado, Signe Saar, Mari Vaus)

**7. Terminitöö.** Osalemine erialakomisjonides ja nõukodades: merekeel (Maire Raadik), lennundus (Tiina Leemets, Signe Saar, Mari Vaus), käsitööteadus (Tiina Leemets), nimekomisjonid (Peeter Päll), botaanika (Sirje Mäearu), aiandus (Sirje Mäearu), laborimeditsiin (Sirje Mäearu), imetajad (Tiina Paet), tõlkijate nõustamine (Katrin Hallik, Katre Kasemets), religioon, usundilugu (Küllu Kuusk), teenuste disain (Ene Vainik), põhiseaduse tõlge (Mari Vaus), infokorraldused (Peeter Päll, Tiina Soon).

**8. EKI vene keele morfoloogiline andmebaas ruMorf** <http://www.eki.ee/keeletase/admin/rumorf>. Loodud on vene keelendite morfoloogiline andmebaas. Muuttüüpide eristamisel on aluseks „Vene keele grammatilise sõnaraamatu“ veebiversioon (<http://gramdict.ru>). Keelendite morfoloogiliste kategooriate märgendamiseks on kasutatud avatud koodiga OpenCorpora projektis (<http://opencorpora.org/>) välja töötatud põhimõtteid ja nomenklatuuri. Paradigmade genereerimisel on kasutatud M. Hageni andmebaasi „Täisparadigma. Morfologia. Sagedussõnastik. Ühendsõnastik“ veebiversiooni (<http://www.speakrus.ru/dict/hagen-details.txt>). 05. jaanuari seisuga on andmebaasis 96 821 märksõna. Käsisti on kontrollitud umbes 3500 sõna paradigmat.

**9. Õpime eesti keelt** (koostöös TripleDev OÜga) <https://sonaveeb.ee/learn>. On valminud Sõnaveebi alamleht „Õpime eesti keelt“ (aluseks põhisõnavara sõnastiku õppe- ja pildilehed).

**10. Õpetaja tööriistad** <https://sonaveeb.ee/teacher-tools/#/home> (meeskond: Tiiu Üksik, Katrin Tsepelina, Kristina Koppel, Raili Pool, Anu Epner, Pilvi Alp, Mai Tiits). Projekt koosneb mitmest moodulist:

- 1) Sõnavara moodul: on koostatud eelA1-B2 noore keeleõppija sõnavara loendid (6278 sõna), lisaks ka temaatilised loendid (ca 20 teemat).
- 2) Grammatika moodul: on koostatud noore õppija grammatilise pädevuse kirjeldus tasemetele eelA1-B2 (Raili Pool).
- 3) Kasutusolukordade moodul: on koostatud noore keeleõppija CEFR-ile vastav kasutusolukordade moodul (avalikustame jaanuaris 2021).
- 4) Häälduharjutuste projekt: vt esialgsed tulemused Sõnaveebi testis <https://ekitest.tripledev.ee/ww/pronunciation-exercises/#/>



**11. Korpusalane töö** (Kristina Koppel, Katrin Tsepelina, Tiiu Üksik). Koostöös Lexical Computing Ltd ja Eesti Keeleressursside Keskusega (Kadri Vider, Neeme Kahusk) on valminud:

- 1) Eesti keele ühendkorpus 2019.
- 2) Eesti keele A1–C1 õpikute lausete korpus 2021.
- 3) Eesti keele kui teise keele kooliõpikute lausete korpus 2021.
- 4) Eesti keele kui teise keele kooliõpikute sisu korpus 2021.
- 5) Töös on Keeleõppija Sõnaveebi lausete korpus. Valmib veebruaris 2021.
- 6) Täiendati EMMA korpust (Tiiu Üksik, Sven Aller, Kertu Niilus, Kristjan Suluste). EMMA korpuse materjalid on tõhus ressurss eesti keele kui teise keele uurimiseks. Korraldati pilootuuring lasteaedades ja algkoolis, mis võimaldab senisest täpsemalt kirjeldada koolieelikute ja 3. klassi õpilaste eesti keele oskust.
- 7) Eesti uudisvoogude korpus 2020.
- 8) Oskuskeelekorpusete loomine ja täiendamine (Merily Plado): täiendati keskkonnakorpusi, loodi minikorpusi väiksemate terminiprojektide jaoks.

**12. Kohanimed ja perekonnanimed.** „Eesti perekonnanimeraamat“: korraldustöö, seminaride korraldamine (09.–10.06.2020 Ojakol ja 22.–23.10.2020 Võrus), autorite kaastöö koordineerimine: koostatud või koostamisel on 4940 artiklit eelduslikust üldmahust 7000 artiklit (Peeter Päll). Eesti kohanimeraamatu veebiversiooni uuendamine (üksikute uute artiklite lisamine-toimetamine). Kohanimeandmebaasi KNAB edasiarendamine ja täiendamine. Tööversioonis on 31.12.2020 seisuga 163 406 nimeobjekti (531 286 nime), sh Eesti osas 47 186 nimeobjekti (133 224 nime), välisriikide osas 116 220 nimeobjekti (398 062 nime). Juurdekasv tööversioonis 2020. aasta jooksul 1280 nimeobjekti (8759 nime). Interneti-versiooni uuendati viimati 2020. a septembris, selles on 161 299 nimeobjekti (516 260 nime). KNABi väljavõttena avaldati 13.10.2020 Mordva kohanimeloend (1924 kirjet) ja 15.10.2020 Udmurtia kohanimeloend (4027 kirjet) koos Indrek Heina programmeeritud kaardiliidestega. Koostöös Võru Instituudiga ilmus Eesti ja soome-ugri keeleteaduse ajakirja ESUKA nimeteaduslik erinumber „Nimede kiiluvees. In the wake of names“, mis oli ühtlasi pühendusteos EKI nimeuurijale Marja Kallasmaale tema 70. sünnipäeva puhul.

**13. Keelekorraldus, sh oskuskeele- ja nimekorraldus ja selge keel.** Ilmus „Keelenõuanne soovitab 6“ (koostanud ja toimetanud Maire Raadik). Õigekeelsuskäsiraamatu ja e-teatmiku töö: a) analüüsitud õigekeelsuskäsiraamatu tagasisidet ning täiendatud ja parandatud käsiraamatut laekunud ettepanekute põhjal; b) valmistatud Emakeele Seltsi keeleteoimkonnale ette arutamisele tulevaid materjale (avaldatud on kaks toimkonna otsuse kavandit: soome-ugri nimede transkriptsioon, ÕSi 7. tüübi omadussõnade rööpkäänamine); c) koostatud e-teatmiku esialgne kava. Korraldati keelehooldekoolitusi (Sirje Mäearu jt) ja selge keele koolitusi (Katrin Hallik, Katre Kasemets). Selge keele auditid (Katrin Hallik, Katre Kasemets). Korraldati 15 tasuta koolitust riigi ja kohaliku omavalitsuse asutuste töötajatele (Marit Alas).

**14. Keelenõuanne, sh termini- ja nimenõuanne.** Telefonitsi keelenõu anti 3340 küsijale (vrd 2019. a – 4441 küsijale), vastati 4033 meilile (2019 – 3215), seega kokku **7373** pöördumisele. Telefoninõuannete arvu kahanemist tingis paus kevadise koroonakriisi ajal. Keelenõu andmebaasis on 184 700 kirjet, neist 7500 on avalikud (Keelenõuvakk). Keelenõulehel on 13 lühikest kirjutist.

**15. Teadusprojekt PSG227 „Eesti keele sõnaliikide kategooriad korpusanalüüsi valguses“.** Projekti [koduleht](#). Selgitati, kuidas jõuda korpuspõhise sõnaliigimäärajani, mis lihtsustab leksikograafilist tööd. Plaanitakse arendada välja mitmest moodulist (statistiline, morfosüntaktiline, võimalikult ka semantiline moodul) koosnev sõnaliigiliides, mis töötleks korpuse tekste nii, et leksikograafi töölauale tekiks kokkuvõtte sõna(vormi) sõnaliigilisest käitumisest.

- Liidese statistilise vormidistributsiooni mooduli idee vormistati artiklik (esitatud avaldamiseks) ning aasta lõpuks jõuti selle mooduli prototüübi väljaarendamiseni (olemas on algoritm).
- Teine, morfosüntaktilise distributsiooni moodul (adjektiviseerumismoodul), on jõudnud testimisjärku.

- Selgitati välja sõnaliigiliselt ambivalentsete sõnade ja vormide tüpologia.
- Pandi paika kesksete sõnaliikide profiilid, nende prototüüpsete tunnuste põhjal tuletatakse teste, mida plaanitakse edaspidi rakendada sõnaliigilidese moodulites.

**16. Teadusprojekt EKKD64 „Eesti keele sõnavara ja korraldus: deskriptiivne ja preskriptiivne vaatenurk (1.01.2019–31.12.2022)“.** Projekti eesmärk on leksikoloogilised alusuuringud, mis keskenduvad sõnavarale, mis vajaksid keelekorraldusliku hinnangu uuendamist. Sõnavarakorralduse põhimõtteid uuendatakse, arvestades tegelikku keelekasutust ja uusi digivõimalusi info analüüsimisel. Uue infoga täiendatakse sõnastikusüsteemi Ekilex ja keeleportaali Sõnaveeb (rubriigis „ÕS soovitab“). Ilmunud või avaldamiseks üle antud on 4 artiklit. Alustati keelekorralduse põhimõtete uuendamise seminaridega (teemasid käsitletakse valdkondade kaupa).

**17. Teaduspublikatsioonid ja -ettekanded.** Täpsem info ETISes ja individuaalsetes aastaaruannetes.

**18. Kaitstud doktoritööd.** Kristina Koppel „Näitelause korpuspõhine automaattuvastus eesti keele õppesõnastikele“, juhendajad Raili Pool ja Jelena Kallas, Tartu Ülikool, humanitaarteaduste ja kunstide valdkond, eesti ja üldkeeleteaduse instituut.

**19. Doktorandid.** Osakonnas on neli doktoranti: Maria Tuulik (Tartu Ülikool), Lydia Raadik (Tartu Ülikool), Natalia Vaiss (Tallinna Ülikool) ja Tiina Paet (Tartu Ülikool).

## **20. Välis- ja ühisprojektid**

- Osalemine rahvusvahelises leksikograafide ühingu EURALEX juhtkomitees (Jelena Kallas).
- Osalemine rahvusvahelise elektroonilise leksikograafia konverentside (eLex 2013, 2015, 2017) korralduskomitees ja teaduskomitees (Jelena Kallas, Margit Langemets).
- Projekt ELEXIS „Euroopa leksikograafilise taristu (2018–2022)“ (European Lexicographic Infrastructure, Horizon 2020) (Jelena Kallas, Margit Langemets, Kristina Koppel).
- Euroopa elektroonilise leksikograafia võrgustik ENeL (European Network of e-Lexicography) (Maria Tuulik, Jelena Kallas, Kristina Koppel).
- Keeleõppe ja rahvahanke ühendamise Euroopa võrgustik (European Network for Combining Language Learning with Crowdsourcing Techniques) (Jelena Kallas).
- Crowdsourcing for Language Examples, enetCollectist välja sündinud koostööprojekt. (Töörühma kuulub Kristina Koppel.)
- Koostöö ettevõttega Lexical Computing Ltd. (Inglismaa) ja Trojina Instituudiga (Sloveenia) (Jelena Kallas, Kristina Koppel).
- Selge keele projekt „Fair Trials“ koostöös Eesti Advokatuuri ja ELi partneritega (Katrín Hallik).
- Projekt ELIPS (European Languages and their Intelligibility in the Public Space) (Katrín Hallik).
- Selge keele uuringute ettevalmistamine koostöös Kantar Emoriga (Katre Kasemets).
- Alates 01.06.2020 jätkuprojekt „Federated eTranslation Termbank Network“ (tegevuse number: 2019-EU-IA-0049) koostöös projekti koordineerija Tilde ja teiste projekti partneritega (Institut „Jožef Stefan“, Sloveenia; Institutet för språk och folkminnen, Rootsi; Internationales Terminologienetz TermNet, Austria; Københavns Universitet, Taani; Kultūras informācijas sistēmu centrs, Läti; Lietuvių kalbos institutas, Leedu; Stofnun Árna Magnússonar í íslenskum fræðum, Island; Svenska institutet för standarder, Rootsi) (Kelly Lilles, Mari Vaus).

**21. Loengukursused.** Tiina Paet: Eesti keele korraldamine ja seisund. 2020/2021 sügissemestri bakalaureuse ja magistri õppekava aine Tartu Ülikoolis. Loenguid 24 tundi, iseseisva töö juhendamine 54 tundi; Lydia Raadik: Ülikoolis õpetamise praktika (156 tundi), assisteerinud Tartu Ülikooli õppejõudu Andra Rummi aines „Eesti keele suulise ja kirjaliku väljenduse õpetus“.

## 22. Teaduse populariseerimine

- Ajakirjas Horisont ilmus 6 Udo Uiibo populaarteaduslikku artiklit (sõnade päritolust).
- Avaldati lühikirjutisi Postimehe EKI keelekoolis (Ene Vainik, Margit Langemets, Geda Paulsen, Tiia Valdre, Kristina Koppel, Jelena Kallas, Peeter Päll, Maire Raadik, Lydia Raadik).
- Osaleti koostöös ERRiga populaarteaduslike teleklippide (<http://portaal.eki.ee/videod.html>) valmistamisel, valmis 4 keeleklippi (Tiina Paet, Ene Vainik).

## 23. Teaduspreemiad ja tunnustused

- HTMi konkursi „Keeletegu 2019“ nominendid:
  - Sõnastikuportaal Sõnaveeb (Margit Langemets, Jelena Kallas, Kristina Koppel, Ülle Viks, Tõnis Nurk ja Indrek Hein)
  - Eesti-soome veebisõnaraamat (toim. Heikki Hurta ja Margit Langemets, koost. Sofia Björklöf, Anu Haak, Külli Kuusk, Saarni Laitinen, Ilona Paajanen ja Ulla Vanhatalo, grammatikaekspert Ülle Viks)
  - Eesti-hiina põhisõnavara sõnaraamat (Gao Jingyi)
  - Eesti Keele Instituudi keelekooliartiklite sari