

EESTI KEELE KUI TEISE KEELE ÕPETAJA TÖÖRIISTAD EESTI KEELE INSTITUUDI KEELEPORTAALIS SÕNAVEEB

Jelena Kallas, Kristina Koppel,
Raili Pool, Katrin Tsepelina, Tiiu Üksik,
Pilvi Alp, Anu Epner

Ülevaade. Artiklis tutvustatakse Eesti Keele Instituudis (EKI) arendatavaid eesti keele kui teise keele õpet toetavaid ressursse, mis on koondatud keeleportaali Sõnaveeb veebirakendusse Õpetaja Tööriistad. Selle moodulid – sõnavara, grammatika, kasutusolukorrad ja teksti hindamine – moodustavad terviku, pakkudes keeleõpetajatele ja keeleõppega seotud spetsialistidele abi kursuste kavandamisel ning õppematerjalide, harjutusvara ja testide koostamisel.

Metodoloogilise raamistiku loomisel ja keeleoskustasemete määramisel on lähtekohaks Euroopa keeleõppe raamdokument (CEFR 2001, eestikeelne versioon Raamdokument 2007), selle sõsarväljaanded (CEFR/CV 2018, CEFR/CV 2020), Euroopa Nõukogu noorte õppijate tasemekirjeldused vanustele 7–10 (Szabo 2018a) ja 11–15 (Szabo 2018b) ning Eesti-sisesed keeleoskustasemeid puudutavad õigusaktid. Keeleliste andmete allikad on peamiselt aastatel 2018–2020 EKI-s loodud eesti keele kui teise keele õpikute ja õppijakeele korpused.

Artiklis kirjeldatakse tööriistamoodulite loomise ja esituse põhimõtteid, tuuakse välja esile kerkinud probleemid ning võimalikud edasiarendused.

Võtmesõnad: korpusleksikograafia, Euroopa keeleõppe raamdokument, keeleoskustasemed, suhtluse keeletoimingud ja -strateegiad, keelepädevus, L2, eesti keel

1. Sissejuhatus

Euroopas on nii täiskasvanute kui ka noorte (kuni 15-aastaste) õppijate keeleõppe korraldamine, planeerimine ja keeleoskuse hindamine seotud Euroopa Nõukogu (EN) väljatöötatud keeleoskustasemete jaotuse ja kirjeldustega, eelkõige 2001. aastal välja antud Euroopa keeleõppe raamdokumendi (CEFR 2001, edaspidi CEFR) ja selle 2018. aastal ja 2020. aastal ilmunud sõsarväljaannetega (edaspidi CEFR/CV).

CEFR-is eristatakse kuut keeleoskustaset (algtasemel keelekasutaja: A1 – läbimurre, A2 – esmane keeleoskus; iseseisev keelekasutaja: B1 – suhtluslävi, B2 – edasijõudnu; vilunud keelekasutaja: C1 – vaba suhtluse pädevus, C2 – haritud emakeelekõneleja tase), CEFR/CV-s on eristatud lisaks eelA1-tase ja C2-taset käsitletakse kui eriti edukatele õppijatele omast täpset, kergest ja asjakohast keelekasutust, mille kirjeldamisel ei kasutata enam võrdluspunktina emakeelekõneleja mõistet. Lubatakse ka hargkäsitlust, mis tähendab, et põhitasemeid tükeldatakse praktilistel kaalutlustel ja/või kohalikest vajadustest lähtuvalt väiksemateks osadeks ehk alltasemeteks (nt B2 > B2.1 ja B2.2), tähistamaks keeleoskuse edenemist ühe taseme sees.

Eestis on CEFR-i keeleoskustasemete käsitlust rakendatud riiklikul tasandil nii täiskasvanute keeleõppe korralduses (lähtekoht on keeleseadus (Riigi Teataja (edaspidi RT) I, 16.06.2020, 5) ning ametniku, töötaja ning füüsilisest isikust ettevõtja eesti keele oskuse ja kasutamise nõuded (RT I, 27.06.2011, 1)) kui ka koolihariduse kontekstis (lähtekoht on põhikooli riiklik õppekava (RT I, 14.07.2020, 24) ja gümnaasiumi riiklik õppekava (RT I, 29.08.2014, 21)). Kehtivate õppekavade järgi on vene õppekeelega kooli 3. klassi lõpetaja oodatav eesti keele oskuse tase A1.2, 6. klassi lõpetajal A2.2, 9. klassi lõpetajal B1.2. ja gümnaasiumi lõpetajal B2.

CEFR ja CEFR/CV-d on funktsionaalse sisuga ja kehtivad kõigi keelte kohta. Igat taset iseloomustavad aga ka keelespetsiifilised lingvistilised omadused ning neid tuleb kirjeldada iga keele jaoks eraldi. Seetõttu on koostatud keelespetsiifilisi CEFR-ile ja CEFR/CV-dele tuginevaid väljaandeid, nt English Profile (Hawkins, Filipović 2012, Capel 2010, 2012, O’Keeffe, Geraldine 2017), The Core Inventory for General English (North jt 2010) ja The Global Scale of English (Benigno, Jong 2019). Prantsuse, rootsi, inglise, hollandi ja hispaania keeleoskustasemetega märgendatud leksikaalseid ressursse koondab CEFRLex¹ andmebaas, mis sisaldab kokku üle 80 000 leksikaalse üksuse. Rootsi keelele on korpusanalüüsi ja masinõppe meetodil loodud keeleõppe rakenduste platvorm Lärka (Alfter jt 2019). Eesti keele oskust eri tasemetel on CEFR-põhiselt põhjalikumalt käsitletud täiskasvanute jaoks Laur 1998, Ilves 2008, 2010, Hausenberg jt 2008, Ehala jt 1997, Kerge 2008, Allkivi-Metsoja, Eslon 2019, Eslon, Kaivapalu 2020, Eslon jt 2010, Kitsnik 2018. Noore eesti keele õppija keeleoskust pole seni CEFR-põhiselt uuritud.

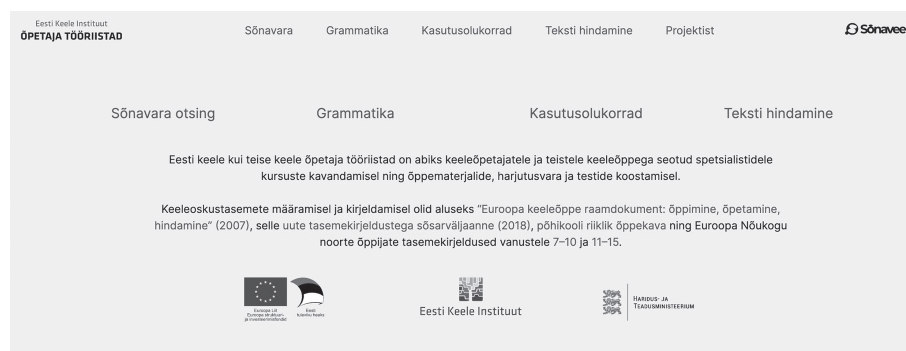
EKI-s on ilmunud mitu õppesõnastikku, mille sihtgrupp on määratud lähtuvalt CEFR-i tasemekirjeldustest. Aastal 2014 ilmus eeskätt A2- ja B1-keeleoskustasemetele suunatud “Eesti keele põhisõnavara sõnastik” (Kallas jt 2014), seejärel aastal 2019 “Sihitisesõnastik” (Vaiss jt 2019) ja B2–C1-keeleoskustasemel õppijale suunatud “Eesti keele naabersõnade sõnastik 2019” (Kallas jt 2015). EKI uues sõnastikusüsteemis Ekilex (Tavast jt 2018) on õppesõnastike info ühendatud ühtsesse andmebaasi, mida kasutajale kuvatakse keeleportaali Sõnaveeb (Koppel jt 2019)

62 ¹ CEFRLex = CEFR-i keeleoskustasemetega märgendatud leksikaalsed ressursid võõrkeele (L2) õppimiseks, õpetamiseks ja võõrkeeleõppe uurimiseks.

kaudu. Aastal 2020 loodi alamportaal Keeleõppija Sõnaveeb (sisaldab u 5000 sõna), mille sihtgrupp on eesti keelt teise keelena õppija (kuni B1-keeleoskustase).

Koostöös tarkvarafirmaga Lexical Computing Ltd. on EKI-l valminud B2–C1-keeleoskustasemele sobiv keeleõppekeskkond SkELL (varasemalt etSkELL), mille abil saab kasutaja lugeda autentseid näitelauseid, vaadata sagedasemaid naabersõnu ning teisi sarnaseid sõnu, sh sünonüüme (loe lähemalt Koppel 2020).

Siinses artiklis kirjeldatav veebirakendus Õpetaja Tööriistad jätkab EKI senist tööd CEFR-põhiste keeleõpet toetavate rakenduste loomisel. Projekti eesmärk on luua ühtne raamistik eesti keele kui teise keele sõnavara- ja grammatikapädevuse käsitlemiseks. Veebirakendus sisaldab nelja moodulit (joonis 1): sõnavara², grammatika, kasutusolukorrad ja teksti hindamine. Tööriistamoodulites on kirjeldatud, milliseid leksikaalseid ja grammatilisi vahendeid peab keelekasutaja valdama ning millistes olukordades toime tulema, et saavutada teatud keeleoskustase. Keeleõppijad võivad olla väga erineva taustaga: noored vene õppekeelega koolides, uusimmigrandid ja nende lapsed, Eestis alaliselt elavad muu emakeelega inimesed jne. Hetkel on loodud kaks üldistavat profiili: täiskasvanu ja noor keeleõppija.



Joonis 1. Tööriistamoodulite kasutajavaade

2. Sõnavara- ja grammatikapädevuse kirjelduse allikandmed

Selleks et välja selgitada, millist sõnavara ja grammatikat teatud keeleoskustasemetel kasutatakse, on esmalt vaja luua sobilikud ressursid, mis võimaldavad sõnavara- ja grammatikapädevuse kujunemise analüüsi. Keeleoskustasemetega seotud tekstid võib jagada kaheks: 1) sisend ehk õpikutekstid või õpetaja valitud tekstid õppeprotsessis kasutamiseks. Nende tekstide analüüs võimaldab välja selgitada, milliseid sõnu ja millist grammatikat on õpikute autorid ja/või õpetajad pidanud vajalikuks teatud keeleoskustaseme pädevuste arendamisel; 2) väljund ehk õppija loodud tekstid, mis peegeldavad õppija oskusi. Kuna sisend mõjutab väljundit, siis on keeleõppe protsesside mõistmiseks vajalik uurida nii õpikutes leiduvaid kui ka keeleõppijate endi loodud tekste. (Volodina, Kokkinakis 2013)

Allikandmete (sisendi) saamiseks loodi õpikute korpused (ptk 2.1).

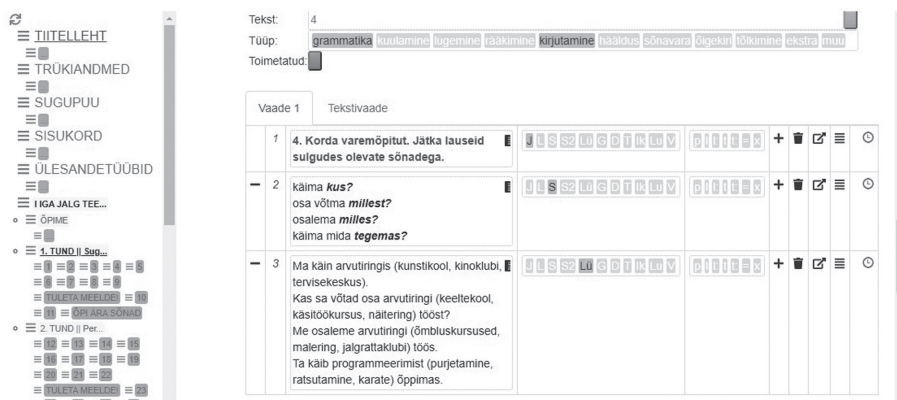
² Enne 2020. aastat nimetati sõnavara moodulit etLex-iks (<https://doi.org/10.15155/3-00-0000-0000-071F3L>).

Õppijakeele (väljundi) analüüsi tarbeks täiendati eesti keele õppija korpust EMMA³ (ptk 2.2.)⁴.

2.1. Eesti keele õpikute korpused

Õpikute sisu korpusete loomine koosnes mitmest etapist. Esiteks koostati loend viimase 20 aasta jooksul ilmunud õpikutest. Õpikute valikul oli oluline, et esindatud oleksid kõik keeleoskustasemed (kooliõpikute puhul klassid 1–12, täiskasvanute õpikute puhul tasemed A1–C1) ning et need sisaldaksid võimalikult palju tekstilist materjali. Lõpliku valimi moodustasid kaheksa täiskasvanute õpikut ja 27 kooliõpikut. Õpikud, mis ei olnud masinloetavas formaadis kättesaadavad, digiteeriti Eesti Rahvusraamatukogu digiteerimiskeskuses.⁵

Seejärel imporditi õpikute failid sisu ja lausete märgendamiseks selleks otstarbeks spetsiaalselt loodud andmebaasi etLex⁶ (vt joonist 2).



Joonis 2. Õpikute märgenduskeskkond andmebaasis etLex

Märgendamisel säilitati võimalikult ligilähedaselt õpikute esialgne struktuur, mille alusel jagati õpik väiksemateks osadeks (tiitelleht, sisukord, peatükid jmt). Automaatsel tekstivastusel tekkinud vead parandati käsitsi. Iga peatüki sees määrati harjutuste piirid ja tüübid (grammatika, kuulamine, lugemine, rääkimine, kirjutamine, hääldus, sõnavara, õigekiri, tõlkimine, muu). Iga harjutuse tekst liigendati lõikudeks vastavalt tekstitüübile ja/või funktsioonile (juhend/saate tekst, täislaused, sõnavara, sõnavara vormiinfoga, lünk tekst, grammatika, dialoog, tarbetekst, ilukirjandustekst, luule). Lõikudele lisati ka vormindusinfo (pealkiri, loend, saatelausega loend, tabel, saatelausega tabel, paralleeltekst).

Selline märgendamine võimaldas luua mitu eraldi korpust. Esiteks korpused, mis sisaldavad märgendatud kujul kogu õpikutes sisalduvat materjali:

- 1) “Eesti keele A1–C1 õpikute sisu korpus 2017” (u 500 000 sõna);
- 2) “Eesti keele kui teise keele kooliõpikute sisu korpus 2020” (u 1 000 000 sõna).

³ EMMA arendajad on TÜ arvutiteaduse instituudi informaatika nooremlektor Sven Aller ja Kadri Sõrmus. Täname arendajaid koostöö eest.

⁴ Muud eesti keele kui teise keele õppijate kirjalike tekstide kogud on Tallinna Ülikoolis loodud “Eesti vahekeele korpus” (EVKK) ja Tartu Ülikoolis kogutud “Õppijakeele korpus” (<https://www.keel.ut.ee/et/keelekogud>).

⁵ Täname õpikute autoreid ja kirjastusi, kes andsid loa kasutada õpikuid korpuse loomiseks.

⁶ etLex on EKI-sisene andmebaas. etLexi arendaja on Katrin Tsepelina.

Teiseks korpused, mis sisaldavad õpikutes sisalduvaid terviklauseid⁷:

- 3) "Eesti keele A1–C1 õpikute lausete korpus 2020" (ligi 16 600 lauset);
- 4) "Eesti keele kui teise keele kooliõpikute lausete korpus 2020" (ligi 30 000 lauset).

2.2. Eesti keele õppija korpus EMMA

Eesti keele õppija korpus EMMA on Tartu Ülikoolis loodud sisestus- ja väliskasutajaliidesega korpuspäringusüsteem, mis sisaldas algselt eesti keelt emakeelena valdavate 9. ja 12. klassi õpilaste loodud eestikeelseid tekste (sh eksamid) aastatest 1999–2017. Alates 2019. aastast on lisatud korpusesse EKI ja Haridus- ja Noorteameti koostööna ka eesti keele kui teise keele õppijate töid, siinse projekti raames kokku 6778 L2 eksami- ja tasemetöö kirjaliku osa teksti (tabel 1). (Haridus- ja Noorteamet: Eesti keele tasemeeksamid)

Tabelis 1 on toodud andmed EMMA korpusesse sisestatud taseme- ja eksamitöödest. L1 tähistab eesti keelt kui emakeelt rääkivate õpilaste tekste, L2 eesti keelt kui teist keelt rääkivate õpilaste tekste.

Tabel 1. 2019. ja 2020. aastal lisatud taseme- ja eksamitööde arv EMMA korpuses

Klass/aasta	2016	2017	2018	2019	2020	Kokku
3. klass	–	299 (L1)	–	–	132 (L2)	299 (L1) 132 (L2)
6. klass	613 (L2)	487 (L2)	514 (L2)	530 (L2)	(tasemetöid ei tehtud)	2144 (L2)
9. klass	301 (L1) 300 (L2)	301 (L1) 301 (L2)	297 (L1) 301 (L2)	600 (L2)	600 (L2)	899 (L1) 2102 (L2)
12. klass	–	600 (L2)	600 (L2)	600 (L2)	600 (L2)	2400 (L2)
					Kokku	1198 (L1) 6778 (L2) 7976 tööd

Valimi koostamise esmane tingimus oli positiivne sooritustulemus, mis oli tasemetööde puhul vähemalt 75 punkti ja riigieksamitel vähemalt 60 punkti sajust. Lisaks jälgiti, et valitud tööd oleks sooliselt ja geograafiliselt (testkoht) võimalikult ühtlaselt jaotatud. Korpusesse on lisatud töö originaalfailide pdf-formaadis skaneeringud, mis on omakorda transkribeeritud. Skaneeringud laekusid Haridus- ja Noorteametist pseudonümiseeritud kujul, teksti autori nime asendab unikaalne kood. Iga teksti juurde kuuluvad autori unikaalse sooritustähisega seotud metaandmed: kirjutamise aasta, eesti keel emakeelena või teise keelena, kirjutaja sugu, tulemus, testkoht (kohalik omavalitsus ja maakond). Tekstides esinevaid vigu ja kõrvalekaldeid standardkeelest ei parandatud ega märgendatud.

Eelnimetatud korpuseid kasutati Õpetaja Tööriistade veebirakenduses sõnavaraloendite koostamisel ja grammatilise pädevuse kirjeldamisel, aga ka näitelausete allikana grammatika moodulis.

⁷ Kättesaadavad korpuspäringusüsteemis KORP.

3. Tööriistamoodulid

Veebirakendus Õpetaja Tööriistad koosneb neljast moodulist: sõnavara, grammatika, kasutusolukorrad ja teksti hindamine. Järgnevalt kirjeldatakse neid lähemalt.

3.1. Sõnavara moodul

Sõnavara moodul sisaldab noore keeleõppija ja täiskasvanu sõnavaraloendeid, mis on abiks õppematerjalide koostamisel, tekstide valikul ja tasemekohasuse hindamisel.

Sõnade tasemetesse jagamise üldpõhimõtted on järgmised:

- 1) sagedus korpustes (ühendkorpus, õpikute korpus ja õppijakeele korpus);
- 2) tuumsus (vt lähemalt Pajusalu jt 2004);
- 3) kasutatavus suulises ja kirjalikus tekstiloomes (vt lähemalt Pajupuu jt 2009);
- 4) keeletasemele iseloomulikud kasutusolukorrad ja vajalikkus igapäeva elu valdkondades;
- 5) sõnavara ajakohasus.

Eri tasemete sõnavara suuruse määramise lähtekoht on Krista Kerge koostatud käsiraamat “Funktsionaalse keeleoskuse taseme määratlemine” (1996), mille kohaselt ulatub eesti keelt algtasemel valdava inimese sõnavara 1500–2000 sõnani, kesktaseme korral aga umbes 5000 sõnani.

Täiskasvanute sõnaloendid sisaldavad ligikaudu 12 500 keeleoskustasemega A1–C1 märgendatud sõna, noore keeleõppija (eelA1–B2) sõnaloendid umbes 9000 sõna.

Loendites on sõnad esitatud algvormis, mis aga ei tähenda, et keeleõppija oskab moodustada ja kasutada nende sõnade kõiki vorme, näiteks omandatakse tegusõna *tohtima* esimesena kindla kõneviisi eitavas vormis *ei tohi*. Sõnavormide esinemust keeleoskustasemeti saab korpuspõhiselt tuvastada ja selle info näitamine kuulub tööriistade edasiarenduste hulka.

3.1.1. Noore keeleõppija sõnavaraloendid

Noore keeleõppija loendite koostamisel olid aluseks lastekeelekorpus CHILDES, eesti keele õppija korpuse EMMA L2 tekstikogud, “Eesti keele kui teise keele kooliõpikute sisu korpus 2020” ja “Eesti keele ühendkorpus 2017”, samuti seni koostatud sõnaloendid (Argus jt 2019) ja õppematerjalid (Haridus- ja Noorteamet: Materjalid lasteaedadele ja õpetajatele).

Noore keeleõppija A1-, A2-, B1- ja B2-taseme sõnaloendite koostamisel on kasutatud järgmisi allikmaterjale:

- 1) EMMA korpus, millest kasutati 6. klassi eesti keele kui teise keele tasemetööde kirjalikku osa (2016–2020; 2144 tööd), põhikooli eesti keele kui teise keele lõpueksami kirjalikku osa (2016–2020; 2102 tööd) ning gümnaasiumi eesti keele kui teise keele riigieksami kirjalikku osa (2017–2020; 2400 tööd);
- 2) “Eesti keele kui teise keele kooliõpikute sisu korpus 2020”.

Võrdlusmaterjalina kasutati täiskasvanute sõnavaraloendeid. Erinevused täiskasvanud ja noore keeleõppija sõnavaravalikus on tingitud eelkõige sihtrühmade kognitiivse, sotsiaalse ja kogemusliku võimekuse ning keelekasutuse põhivaldkondade erinevusest. Noore keeleõppija puhul on lisaks arvestatud ka eesti keele kui teise keele ainekava eesmärgi ja oodatavaid tulemusi.

Koolieelikute (eelA1-tase) sõnavaraloend on koostatud alusharidust omandava keeleõppija vajadusi silmas pidades. Sõnavaraloend sisaldab koolieelse lasteasutuse riikliku õppekava (RT I 2008, 23, 152) kõnearendusteemadest tulenevalt 6–7-aastase keeleõppija sõnavara. Loendi koostamisel on kasutatud järgmisi allikmaterjale:

- 1) lastekeelekorpus CHILDES kolme kõige süsteemsemalt kogutud eesti laste suulise kõne alamkorpus Viija, Kapanen ja Zupping;
- 2) Haridus- ja Teadusministeeriumi ekspertide Tiina Petersoni ja Viivian Jõemetsa koostatud algset sõnavaraloendit⁸;
- 3) eesti keele kui teise keele sõnavaraloendit koolieelikutele (Argus jt 2019);
- 4) eesti keelest erineva kodukeelega laste keeleõppeks välja töötatud materjale.

Koolieelikute sõnavaraloendite koostamisel analüüsiti süstemaatiliselt ka temaatilist sõnavara (kehaosad, loomad, linnud, toit, transport jms) ning lasteaiarühmaruumiga seotud sõnavara, lisati Eesti kultuuri tundmise seisukohalt oluline sõnavara, nt riiklike ja rahvakalendri tähtpäevade ning riigipühade nimetused.

3.1.2. Täiskasvanute sõnavaraloendid

Täiskasvanute sõnavaraloendite (Kallas, Koppel 2018a, 2018b, 2018c) koostamisel olid aluseks seni ilmunud eesti keele tasemete kirjeldused (Ilves 2010, 2008, Hausenberg jt 2008, Ehala, Liiv 1997, Kerge 2008), korpused (“Eesti keele ühendkorpus 2017” ja “Eesti keele A1–C1 õpikute sisu korpus 2017”), EKI-s koostatud õppesõnastikud (Kallas jt 2014, 2019) ja sõnavaraloendid (Pärn, Simm 1989).

Loendid koostati mitmes etapis. Esmalt loodi seni ilmunud eesti keele tasemete kirjelduste sõnavaraloenditest ühtne elektrooniline andmebaas, mis võimaldas välja selgitada loenditevahelisi kattuvusi ja erinevusi. Analüüsi käigus kõrvaldati vanaenenud sõnavara, näiteks *ühiskorter*, *kooperatiivkorter*, *komposteerima*, *panderoll*, *telefonikabiin*, *videomagnetofon*, *kassettmagnetofon*.

Teiseks loodi sõnavaraloend “Eesti keele A1–C1 õpikute sisu korpus 2017” (u 13 000 sõna). Andmebaasi jäeti vaid sõnad, mis esinesid õpikutes vähemalt kaks korda.

Kolmandaks liideti need loendid omavahel ning võrreldi “Eesti keele ühendkorpus 2017” sagedusandmetega. Ühendkorpusel alusel genereeritud sagedusloend (u 60 000 sõna) võimaldas kontrollida senistes loendites ja õpikutes kasutatud sõnavara ajakohasust ning lisada korpuses sagedasti esinevaid sõnu, mis olid varasematest loenditest ja ka õpikute tekstist välja jäänud (nt *nutikell*, *nutitelefon*).

⁸ Sõnavaraloend on kättesaadav käsikirjana EKI arhiivis.

3.2. Grammatika moodul

Grammatika moodulis on seni esitatud vaid noore keeleõppija grammatikapädevuse kirjeldus. Nagu kõigis moodulites, kirjeldatakse ka siin 6–7-aastaste koolieelikute (eelA1), 3. klassi lõpetajate (A1), 6. klassi lõpetajate (A2), 9. klassi lõpetajate (B1) ja 12. klassi lõpetajate (B2) oodatavat grammatikapädevust. Grammatikapädevus tähendab keele grammatiliste vahendite tundmist koos oskusega neid kasutada (CEFR 2007: 131). Kirjelduse eesmärk on anda ülevaade, mida ja kui hästi keelekasutaja või keeleõppija vormi-, sõna-, fraasi- ja lausemoodustuse tasandil teha oskab. Grammatilised kategooriad on kirjeldatud kõigil keeleoskustasemetel, kuid nende kasutusfunktsioone ja piiranguid (kui tihti õppija eksib ja milles eksimus seisneb) on täpsustatud vastavalt keeleoskuse arengule. Selline kirjeldus aitab ühelt poolt keeleoskustasemeid eristada ning teiselt poolt seada edasisi eesmärgid.

Koolieelikute (eelA1) grammatikapädevuse kirjeldamisel⁹ olid lähtekohaks CHILDES-i eesti keele alamkorpused (nimedega Vija, Kapanen ja Zupping) ning EKI korraldatud koolieelikute ja I kooliastme grammatikapädevuse valideerimise ja tekstikorje pilootprojektis (2020 sügis) õpetajatelt kogutud grammatika omandamise hindamise ankeetide andmed. Pilootprojekti käigus koguti keelenäiteid kolme lasteaia erinevates õppevormides (eesti keel tegevusena, keelekümbelus ja kolmas õpetaja rühmas ehk osaline keelekümbelus) õppivatelt koolieelikutelt ja kahe kooli erinevate õppevormide (keelekümbelus, varane keelekümbelus, osaline keelekümbelus, tavaklass) kolmandate klasside õpilastelt.

A1–B2-tasemete grammatikapädevuse kirjeldus tugineb eelkõige eesti keele kui teise keele õpikutekstide ja õppijakeele tekstide korpuste analüüsile (vt ptk 2).

Koolieelikute grammatikapädevuse kirjeldamisel on lähtutud eeldusest, et grammatilised kategooriad, mis on esimese keele omandajate varases repertuaaris, omandatakse varakult ka siis, kui laps omandab teist keelt. Kasutus põhiste keeleomandamisteooriate järgi omandatakse varakult need kategooriad, mis on sagedased ja produktiivsed (vt nt Dressler jt 2002, Dressler 2012). Kasutus põhise lähenemise ideestikuga on seotud ka arusaam tüüpilisest keeleomandamise keskonnast ja keelendite sagedusest. Eeldatakse, et koolieelik omandab keele mitteformaalses situatsioonis lapsekesksele keskkonnale omaste igapäevategevuste ja suhtluse käigus (nt lasteaia mängulistest tegevustes, mitte niivõrd tunnis). Seetõttu vajab koolieelik eelkõige neid kategooriaid ja vorme, mis on igapäevasisuolulistes olukordades kõige sagedasemad. Koolieelikute grammatikapädevuse kirjeldusse kategooriate lisamisel lähtuti kolmest tingimusest: 1) L1 korpuses esines teatud grammatiline kategooria juba varases eas (enne kolmeaastaseks saamist), 2) grammatiline kategooria oli L1 laste kõnes produktiivselt omandatud (st sama vormi kasutas laps mitmest sõnast ning vähemalt ühest sõnast oli kasutatud ka teisi vorme) enne kolmeaastaseks saamist, ning 3) grammatiline kategooria oli L1 laste kõnes sage (st vorme, mida lastekeelekorpuses leidis vaid 2–3 korral, L2 grammatilises profiilis ei esitatud).

Kooliõpilaste grammatikapädevuse uurimiseks analüüsiti esmalt eesti keele kui teise keele kooliõpikute sisu korpuse põhjal, mida on õpikute autorid seni pidanud oluliseks õpetada. Veendumaks, et kirjeldused vastavad õppija tegelikele saavutustele, võrreldi grammatiliste kategooriate esinemist õpikukorpuses EMMA korpuse

teise keele õppija tekstidega. Kui kategooria õppijakeele korpuses esines, kirjeldati see vastaval keeleoskustasemel koos tasemekohaste piirangutega.

Grammatikapädevuse kirjeldus on üles ehitatud laienevalt: lihtsamast keerulisemani. Kirjeldus on esitatud lähtuvalt peakategooriatest (vormi-, sõna-, fraasi- ja lausemoodustus), millel on kokku 19 alamkategooriat (nt vormimoodustuse alamkategooriad on nimisõna, omadussõna, tegusõna, määrsõna ja asesõna).

Igal alamkategoorial on omakorda grammatilised tunnused (nt vormimoodustuse peakategoorias on nimisõnal tunnused arv ja kääne; vt joonist 3), igal tunnusel omakorda tunnuse liigid (nt tunnusel arv on liigid ainsus ja mitmus).

Joonis 3. Grammatika mooduli otsinguliides lähtuvalt alamkategooria tunnustest

Iga kirjelduse juurde kuuluvad ka näitelaused kategooria kasutamise kohta ja lisainfo vaade. Lauseid on kolme tüüpi: projekti grammatikaekspertide koostatud laused (põhikoostajad on Raili Pool, Reili Argus ja Annika Bauer), õpikute laused “Eesti keele kui teise keele kooliõpikute lausete korpusest 2021” ja õppijakeele laused EMMA korpusest. Tasemetel eelA1–B1 kuvatakse õpikute ja õppijakeele laused ainult lisainfo vaates, otsingutulemuse all kuvatakse grammatikaekspertide koostatud laused. B2-tasemel kuvatakse otsingutulemuste all õpikute laused ja lisainfo all õppijakeele laused. Vormimoodustuse kirjelduste lisainfo vaates on antud ka kirjeldatava grammatilise kategooria süntaktiline funktsioon (vt joonist 4).

	Näide	Allikas
1.	Tassid ja taldrikud on laual. Neiljapäeval on meil viis tundi. Jullal on uus telefon. Kristiinai valutab hammas.	projekti grammatikaekspertid
2.	Piia istub diivanil ja siltab Miisut.	Kooliõpikute lausete korpus 2020
3.	Piial on väga head sõbrad.	Kooliõpikute lausete korpus 2020
4.	Leal valutab kõht ja ta läheb koju.	Kooliõpikute lausete korpus 2020
5.	Pärast kooli on Jukui kodus igav.	Kooliõpikute lausete korpus 2020

Joonis 4. Grammatika mooduli lisainfo vaade

Grammatikapädevuse kirjeldus võimaldab korraga välja otsida sama grammatilise kategooria kohta käiva info eri keeleoskustasemete kohta, pakkudes eesti keele õpetajatele ja õppematerjalide koostajatele tervikpildi teatud kategooria esinemusest õppijate keelekasutuses ning kasutamisel esinevatest piirangutest (mis tingimustel õppija vorme kasutada oskab). Nii saab kasutajaliidese kaudu ülevaate teatud

grammatilise kategooria ilmumisest õppijakeelde ning selle kasutamisel toimuvast arengust keeleoskustasemete kaupa. Joonis 5 illustreerib ühe kategooria eri tasemete info otsingu tulemust umbisikulise tegumoe lihtmineviku vormide näitel.

Keeleõppija: @ noor keeleõppija

Keeletase: eelA1 A1 A2 B1

Peakategooria: VORMIMOODUSTUS Alamkategooriad: TEGUSÕNA

tegumood umbisikuline aeg lihtminevik

OTSI

+ Lisa peakategooria

Otsingu tulemused (3) Näita 20

Peakategooria	Alamkategooria	Tase	Kirjeldus	Lisainfo
VORMIMOODUSTUS	TEGUSÕNA	A2	umbisikulise tegumoe lihtminevik jaatavas kõnes Mõistab umbisikulise tegumoe lihtmineviku jaatavate vormide (nt <i>kirjutati</i>) tähendust. Suudab ise neid vorme moodustada üksikute õpitud tegusõnadest ning kasutada peamiselt tervikuna omandatud struktuurides, et väljendada minevikus toimunud tegevust, kui lauses ei ole tegijat mainitud. Umbisikuliste verbivormide moodustamisel ja terviklike lausestruktuuride vormistamisel eksib sageli, nt võib umbisikulisi ja isikulisi verbivorme omavahel segi ajada. Näide: <i>Matemaatikas anti palju õppida.</i> Tunnused: kõneviis kindel tegumood umbisikuline aeg lihtminevik kõnelik jaatav	Lisainfo
VORMIMOODUSTUS	TEGUSÕNA	B1	umbisikulise tegumoe lihtminevik jaatavas kõnes Oskab õpitud tegusõnadest moodustada umbisikulise tegumoe lihtmineviku jaatava kõne vorme (nt <i>kirjutati, räägiti, toodi</i>), suudab neid kasutada, väljendamaks minevikus toimunud tegevust, kui lauses ei ole tegijat mainitud. Suudab neid vorme moodustada nii ilma astmevahelduseta kui ka astmevahelduslikest tegusõnadest. Eksib umbisikuliste vormide moodustamisel, eriti astmevahelduslikest tegusõnadest, ning võib umbisikulisi ja isikulisi verbivorme lauses segi ajada. Näide: <i>Turniiri parimatele anti medalid ja auhinnad. Karjäärpäeval räägiti meile mitmetest elukutsetest. Mis sulle õeldi?</i> Tunnused: kõneviis kindel tegumood umbisikuline aeg lihtminevik kõnelik jaatav	Lisainfo
VORMIMOODUSTUS	TEGUSÕNA	B1	umbisikulise tegumoe lihtminevik eitavas kõnes Mõistab umbisikulise tegumoe lihtmineviku eitava kõne vormide tähendust (nt <i>ei räägitud, ei tehtud</i>) ja suudab neid üksikute õpitud juhtudel kasutada minevikus toimunud tegevuse eitamiseks lauses, kus tegijat pole mainitud. Eksib <i>tud</i> -kesksõna moodustamisel ning umbisikulise lause vormistamisel tervikuna (nt kasutab <i>tud</i> -kesksõna asemel mõnd muud verbivormi). Näide: <i>Miks mulle midagi ei räägitud? Eesti keele tunnis seekord hindeid ei pandud.</i> Tunnused: kõneviis kindel tegumood umbisikuline aeg lihtminevik kõnelik eitav	Lisainfo

Joonis 5. Umbisikulise tegumoe lihtmineviku kirjeldus eri keeleoskustasemetel

Jooniselt 5 nähtub, et umbisikulise tegumoe lihtmineviku jaatavate vormide kasutamine on korpusmaterjalidest lähtuvalt A2-tasemel kirjeldatud peamiselt mõistmise ja tervikstruktuuride tasandil, B1-tasemel aga oskab õppija neid õpitud verbidest aktiivselt moodustada ja kasutada, piirangutena on kirjeldatud eksimine vormide moodustamisel ning isikuliste ja umbisikuliste vormide segijamine.

3.3. Kasutusolukorrad

Kasutusolukordade moodul pakub võimaluse tasemeskaalade ja kirjelduskriteeriumide kasutamiseks keeleõppes.

Keeleõppija peab omandama oskused erinevates keelekasutusolukordades toimetulekuks – ta peab igapäevaolukorras suutma ennast väljendada ja täita erinevaid ülesandeid. Mida rohkem keeleoskus areneb, seda nõudlikumate

keelekasutusolukordade ja suhtlusülesannetega keelekasutaja toime tuleb. Mooduli kasutajale pakuvad tuge CEFR-i suhtlustoimingute näidisskaalad, kus kirjeldatakse keeleliste toimingute ja suhtlusstrateegiatega erinevaid tahke. CEFR-is on selliseid skaalasisid 44, need hõlmavad suulise ja kirjaliku teksti vastuvõtutoiminguid ja -strateegiaid (kuulamine ja lugemine), suulist ja kirjalikku tekstiloomet ja loomestrateegiaid ning suulist ja kirjalikku suhtlust ning interaktiivseid strateegiaid.

CEFR-i skaalad on koostatud täiskasvanud keeleõppijat silmas pidades ja noore keeleõppija vajadusi kõiges ei peegelda. Sõsarväljaandes (CEFR/CV) on skaalade valikut ja sõnastusi oluliselt täiendatud ja ajakohastatud. Paralleelprojektina valminud noore keeleõppija tasemekirjelduste koonddokumentide (vanusele 7–10 ja 11–15) aluseks võeti eri riikide Euroopa Keelemappide kirjeldused, mis on mugandatud CEFR-i põhikirjeldustest (tasemeni B2+) ja on suunatud õpilastele (CEFR/CV 2018: 42). Näiteks A2-tasemelt on võetud kõik vanuserühmale 7–10 sobivad kirjelduskriteeriumid ja lisatud sobivad näited keelemappidest. Nii saab ülevaate, millised keelekasutusolukorrad (koos näidetega) on konkreetsele vanuserühmale jõukohased.

Keelekasutusolukordade mooduli aluseks on noore keeleõppija keeleoskuse kirjeldused ja eakohasele kontekstile sobivad kasutusolukordade näidised (Szabo 2018a, 2018b). Moodulis on kirjeldatud A1–B2-taseme keelekasutusolukordi.

Moodul peaks hõlbustama ka CEFR-i ja CEFR/CV-de kasutamist, kuna skaalade detailne esitus toob välja tasemekohased kriteeriumid, mis on õppekontekstis asjakohased ning vajaduse korral kohandatavad.

Nii nagu grammatika moodulis on ka kasutusolukordade moodulis eristatud peakategooriad ja alamkategooriad, millele järgneb iga kategooria detailne kirjeldus (joonis 6).

The screenshot shows the etLex application interface. At the top, there are search filters: 'otsisõna' (search term) set to 'TEKSTI VASTUVÖTT', 'üldine kuulamisoskus' (general listening skill), 'tuumus' (age), 'tuumuse väärtus' (age value), 'projekt' (project), 'A2' (level), 'olek' (status), 'id' (ID), 'kasutaja' (user), and 'Näidatakse' (display) set to '001'. Below the filters, the title 'Kasutusolukorrad' (Usage Examples) is shown, followed by 'Lettud 5 rida' (Showing 5 rows). A pagination bar includes 'First', 'Previous', '1', 'Next', and 'Last'. A 'Lisaa uus' (Add new) button is also present. The main content is a table with columns: '#', 'id', 'olek', 'peakategooria', 'alamkategooria', 'projekt', 'tase', and 'kirjeldus'. Two rows are visible:

#	id	olek	peakategooria	alamkategooria	projekt	tase	kirjeldus
1	108		TEKSTI VASTUVÖTT	üldine kuulamisoskus	noor	A2	<p>sõnadest ja fraasidest arusaamine</p> <p>selge hääldus, aeglane kõnetempo, esmatähtsate ekuvaldkondadega seotud fraasid ja väljendid</p> <p><i>N:</i> Saan aru sõnadest ja fraasidest, mis on seotud minu perega, hobidega ja igapäevaeluga.</p> <p>tekstis näited (tekst) teemade näited (teema) (teema) (teema) suhtluse vorm (vorm)</p>
2	109		TEKSTI VASTUVÖTT	üldine kuulamisoskus	noor	A2	<p>infost arusaamine</p> <p>selge hääldus, aeglane kõnetempo, esmatähtsate ekuvaldkondadega seotud fraasid ja väljendid, teemade ringi piirub kooli, pere ja kooliga</p> <p><i>N:</i> Saan aru tunnipisarnist ja infost koolis toimuvate tegevuste kohta. Saan aru, kui mulle öeldakse, millised riided ma pean kooli kaasa võtma. Saan aru infost, millist toitu on lubatud kooli kaasa võtta. Saan aru, kui mulle öeldakse, kellele võin abi paluda. Saan aru, kui keegi räägib, kuidas ta oma sünnipäeva tahab tähistada. Saan aru, kui mulle öeldakse, mida kass sööb.</p> <p>tekstis näited (tekst) teemade näited (teema) (teema) (teema) (teema) (teema) (teema) (teema) (teema) suhtluse vorm (vorm)</p>

Joonis 6. Kasutusolukordade sisestusliidese vaade andmebaasis etLex

Peakategooria liigitusel on lähtunud CEFR/CV 2018 suhtlustegevuste jaotusest: kasutaja saab valida teksti vastuvõtu, tekstiloomet, suhtluse või vahendamise vahel.

Alamkategooria viitab CEFR/CV 2018 suhtlustoimingute näidisskaaladele.

Kirjelduse rubriigis kuvatakse esimesel real keelekasutusolukord. Sellele järgneb piirang ehk millistel tingimustel keelekasutaja suhtlusülesandega toime tuleb või mida tal on toime tulemiseks veel vaja. Kasutusolukorrad on illustreeritud näidetega

konkreetsetest olukordadest (põhinevad noore keeleõppija koonddokumendis esitatud olukorrajeldustel). Viimasena näeb kasutaja suhtlusvormi, tekstiliigi, teksti alaliigi ja teemade näiteid.

Näiteks kui valida peakategooriaks teksti vastuvõtt ja alamkategooriaks üldine kuulamisoskus, on B1-tasemele jõukohased järgmised keelekasutusukorrad: vestlusest arusaamine; jutu peamisest mõttest arusaamine, selgitustest arusaamine. Kasutusolukordi iseloomustavad näited: *saan aru igapäevasest vestlusest koolis; saan aru, kui räägitakse vaba aja veetmisest; saan aru õpetaja selgitustest koolitunnis*. Tasemele omased märksõnad (piirangud või tingimused) nende kasutusolukordade juures on järgmised: lühike, tuttav ja igapäevane teema, normaalne kõnetempo. Suhtlusvormi, tekstiliigi näidete ja teemade näidete alusel on võimalik otsingutulemusi täiendavalt filtreerida.

Kasutusolukordade moodul annab tasemekohaste oskuste üldise kirjelduse, mida omakorda on võimalik täpsustada, panna konteksti või kohandada vanusegrupile vastavaks. Moodulis kuvatakse valitud tasemele tüüpilised ja tasemekohased keelekasutusolukorrad koos kontekstile sobivate näidistega ja keeleoskustasemele iseloomulike märksõnadega.

3.4. Teksti hindamine

Teksti hindamise moodul aitab määrata sõnade keerukust ja on abiks teksti tasemekohasuse hindamisel. Teksti hindamise mooduli aluseks on noore keeleõppija ja täiskasvanute sõnavaraloendid.

Kuigi teksti tasemesobivuse automaathindamine on suhteliselt uus praktika, on katsetusi selles valdkonnas tehtud näiteks inglise (Xia jt 2016), rootsi (Pilán jt 2015, 2016, Östling jt 2013), saksa (Hancke, Meurers 2013, Zesch jt 2015), hiina (Sung jt 2015), portugali (Branco jt 2014), prantsuse (François, Fairon 2012) ja eesti (Vajjala, Lõo 2013, Hallik 2016) keele põhjal. Teksti keerukuse ja loetavuse hindamiseks on loodud ka erinevaid lause ja sõna pikkust arvestavaid valemeid, eelkõige inglise keelele¹⁰. Eesti keelele kohandatud autentse teksti tasemesobivuse hindamise tööriista, mis arvestaks nii leksikat kui grammatikat ja töötleks lauseid või kogu teksti tervikuna, pole seni loodud. Õppija koostatud teksti (ingl *learner writing*) vastavust konkreetsele keeleoskustasemele hindav tööriist on arendamisjärgus Tallinna Ülikoolis¹¹.

Teksti sõnavara hindamise tööriistad on loodud näiteks rootsi (Lärka¹²) ja inglise keelele (ETS TextEvaluator Tool¹³, Text Inspector¹⁴). Kõigi aluseks on andmebaas valitud sõnatüvedega, mille tase on määratud. Sõnadega, mida andmebaas ei sisalda, ei oska programmid midagi peale hakata. Tööriistad ei anna ka tagasisidet teksti kui terviku kohta.

Teksti hindamise moodul töötab eesti keele töötlusprogrammi estNLTK v 1.6 teekidel, mis pakuvad tekstide töötlemiseks vajalikke funktsioone (lemmatiseerimist ehk eestikeelsete sõnade algvormide leidmist ja morfoloogilist analüüsi ehk sõna algvormi ja selle grammatilise vormi, nt käände või pöörde, tuvastamist).

¹⁰ <https://readabilityformulas.com/free-readability-formula-tests.php> (1.10.2020).

¹¹ <https://github.com/centre-for-educational-technology/evkk/wiki/Demos> (2.2.2021). Tekstihindaja võtab arvesse teksti üldist keerukust (teksti, sõnade ja lausete pikkust), sõnavara (sõnavara mitmekesisus, ulatus, tihedus ja abstraktsus) ja morfoloogiat (sõnaliikide ja muutevormide osakaale ning rohkust) ning nendele parameetritele toetudes prognoosib õppija keeleoskustaset.

¹² <https://spraakbanken.gu.se/larka/texteval> (1.10.2020).

¹³ <https://www.ets.org/c/23491/> (1.10.2020).

¹⁴ <http://www.englishprofile.org/wordlists/text-inspector> (1.10.2020).

Kuna tegemist on automaattöötusega, siis võib esineda vigu. Hindamine toimub lemmapõhiselt ehk sõna algvormi alusel, arvestamata sõnade eri tähendusi ja homonüümide tasemekohasust. Homonüümide puhul (nt *tamm* ‘puu’ ja *tamm* ‘veevoolu tõkestav rajatis’) annab programm madalama keeletaseme hinnangu. Samuti ei analüüsi programm eraldi mitmesõnalisi üksusi, nagu näiteks ühend- (nt *lahku minema*) ja väljendverbid (nt *silma paistma*).

Keeleõppija:
 noor keeleõppija täiskasvanu

Sisesta tekst: 667 / 2000

Keeleoskustase: A1 A2 B1 B2 C1

Sõnu kokku: 98

Tase	Sõnu	%
A1	55	56.12%
A2	18	18.37%
B1	11	11.22%
B2	5	5.10%
C1	4	4.08%
määramata	2	2.04%

Joonis 7. Teksti sõnavara hindamise moodul

Enne teksti hindamist tuleb määrata, kas seda soovitakse teha noore keeleõppija või täiskasvanu sõnavaraloendite põhjal. Hinnatav tekst tuleb sisestada vasakpoolsesse kasti. Keskmisse kasti ilmub hinnatud tekst, kus iga sõna on värvitud vastavalt keeleoskustasemele (määramata sõnad jäävad hindamata). Hinnatud teksti kasti kursoriga liikudes kuvatakse ka iga sõna sõnaliik. Parempoolses kastis on näha statistika kogu teksti kohta – mitu sõna tekstis on ning kuidas need sõnad protsentuaalselt keeleoskustasemete vahel jaotuvad. See aitab õppematerjalide koostajatel paremini hinnata, kas tekst on konkreetsele keeleoskustasemele jõukohane või vajab see lihtsustamist.

4. Kokkuvõte

Õpetaja Tööriistade veebirakendus keeleportaal Sõnaveeb on mõeldud abivahendiks eesti keele kui teise keele õpetajatele ja keeleõppega seotud spetsialistidele kursuste kavandamisel ning õppematerjalide, harjutusvara ja testide koostamisel. Tööriistast koosneb neljast moodulist: sõnavara, grammatika, kasutusolukorrad ja teksti hindamine. Loodud on ühtne raamistik ja metakeel eesti keele kui teise keele sõnavara- ja grammatikapädevuse käsitlemiseks.

Kuna keeleõppija sõnavara- ja grammatikapädevuse arengu uurimiseks puudusid varasemalt sobivad ressursid, loodi neli korpust: “Eesti keele kui teise keele kooliõpikute lausete korpus 2020”, “Eesti keele kui teise keele kooliõpikute sisu korpus 2020”, “Eesti keele A1–C1 õpikute lausete korpus 2020” ja “Eesti keele A1–C1 õpikute sisu korpus 2017”. Eesti keele õppija korpusesse EMMA lisati projekti raames 6778 eesti keele kui teise keele taseme- ja eksamitööd.

Sõnavara mooduli noore keeleõppija (eelA1–B2) sõnavaraloendid sisaldavad kokku umbes 9000 sõna, täiskasvanute (A1–C1) omad ligikaudu 12 000 sõna. Lisaks eelmainitud õppijakeele- ja õpikute korpustele on loendite koostamisel kasutatud

“Eesti keele ühendkorpust 2017” ja koolieelikute (eelA1) puhul ka CHILDES-i eesti keele alamkorpusi.

Grammatika moodul annab ülevaate noore keeleõppija (eelA1–B2) grammatikapädevuse arengust, st mida keeleteadaja või -õppija vormi-, sõna-, fraasi- ja lause moodustuse tasandil teha oskab ja milliste piirangutega ta seda teeb (nt kasutab mõnd vormi vaid tervikuna omandatud struktuurides, eksib vormide moodustamisel jne). Grammatikapädevuse kirjeldamisel olid aluseks õpiku- ja õppijakeelekorpused.

Kasutusolukordade kirjeldus aitab paigutada keelt konteksti – millisel tasemel keeleõppija millises olukorras ja millistel tingimustel toime võiks tulla. Mooduli aluseks on CEFR-i tasemekirjeldused ja noore keeleõppija keeleoskuskirjelduste koonddokumendid, millest on valitud eakohasele kontekstile sobivad näidised.

Teksti hindamise moodul määrab sõnade keerukust tekstis lemma tasandil. Rakendus töötab eesti keele töötlusprogrammi estNLTK v 1.6 teekidel.

Õpetaja veebipõhise tööriistakasti täiustamiseks ja edasiarendamiseks on palju võimalusi. Eraldi tuleks tegelda õigekirjapädevuse kirjeldusega. Oluline uurimisküsimus on ka keelepagasi sõnavormipõhine ulatus. Teksti hindamise tööriista oluliseks ja vajalikuks edasiarenduseks on grammatika keerukuse hindamine.

Kuna tänase seisuga on nii sõnavaraloendid kui ka grammatikapädevuse kirjeldus endiselt valideerimise ja täiendamise tööjärgus, siis on oluline rõhutada, et neid ei tohiks käsitleda kui keeleõppe normi, mille täitmist õpilastelt nõuda. Eelkõige on need mõeldud õpetajatele ja õpivara koostajatele suuniseks, mille abil õppetööd planeerida ja õppematerjali koostada. Eesti keelt õpivad väga erineva tausta, eelteadmiste, võimaluste ja võimekusega õppurid ning kuigi kõiki neid erisusi ei ole võimalik kirjelduste koostamisel arvesse võtta, tuleks õpetajal neid märgata ja õppeprotsessi vastavalt kohandada.

Viidatud kirjandus

- Alfter, David; Borin, Lars; Pilán, Ildikó; Lindström Tiedemann, Therese; Volodina, Elena 2019. Lärka: From Language Learning Platform to Infrastructure for Research on Language Learning. – Inguna Skadiņa, Maria Eskevich (Eds.), Selected Papers from the CLARIN Annual Conference 2018. Linköping Electronic Conference Proceedings 159, 1–14.
- Allkivi-Metsoja, Kais; Eslon, Pille 2019. Estonian interlanguage corpus and automated assessment of proficiency in learner texts. – Abstract Book: The 4th International Conference Thinking, Doing, Learning: Usage Based Perspectives on Second Language Learning, Finland, June 17–19, 2019. Jyväskylä: University of Jyväskylä, 63.
- Argus, Reili; Maiberg, Lea; Koho, Mariana; Kallas, Jelena; Koppel, Kristina 2019. Eesti keele kui teise keele sõnavara koolieelikutele [“Vocabulary of Estonian as a second language for preschoolers”]. Tallinn: Eesti Keele Instituut.
- Benigno, Veronica; Jong, John de 2019. Linking vocabulary to the CEFR and the Global Scale of English: A psychometric model. – Ari Hutha, Gudrun Erickson, Neus Figueras (Eds.), Developments in Language Education. A memorial volume in honour of Sauli Takala. Jyväskylä: University Printing House, 8–29.
- Branco, António; Rodrigues, Jose; Costa, Francois; Silva, Jaime; Vaz, Richard 2014. Assessing automatic text classification for interactive language learning. – International Conference on Information Society (i-Society), 70–78. <https://doi.org/10.1109/i-Society.2014.7009014>

- Capel, Annette 2010. A1-B2 vocabulary: Insights and issues arising from the English Profile Wordlists project. – *English Profile Journal*, 1(1), 1–11. <https://doi.org/10.1017/S2041536210000048>
- Capel, Annette 2012. Completing the English Vocabulary Profile: C1 and C2 vocabulary. – *English Profile Journal*, 3 (1), 1–14. <https://doi.org/10.1017/S2041536212000013>
- CEFR = The Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Strasbourg: Council of Europe.
- CEFR/CV 2018 = Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors. Strasbourg: Council of Europe. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989> (04.02.2021)
- CEFR/CV 2020 = Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion Volume. Strasbourg: Council of Europe. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4> (4.2.2021).
- Dressler, Wolfgang U.; Kilani-Schoch, Marianne; Klampfer, Sabine 2002. How does a child detect morphology? Evidence from production. – R. Harald Baayen, Robert Schreuder (Eds.), *Morphological Structure in Language Processing*. Trends in Linguistics. Studies and Monographs 151. Berlin: Mouton de Gruyter, 391–425. <https://doi.org/10.1515/9783110910186.391>
- Dressler, Wolfgang U. 2012. On the acquisition of inflectional morphology: Introduction. – *Morphology*, 22, 1–8. <https://doi.org/10.1007/s11525-011-9198-1>
- Ehala, Martin; Saarso, Kristi; Vare, Silvi; Õispuu, Jaan 1997. Eesti keele suhtluslävi [‘Threshold Level for Estonia’]. Strasbourg: Council of Europe Publishing.
- Eslon, Pille; Kaivapalu Annekatrin 2020. Teel sihtkeelepärase keelekasutuse poole: vene- ja soomekeelsete eesti keele õppijate kirjaliku keelekasutuse dünaamika A2- ja B1-tasemel [‘Towards target-like language use: Russian and Finnish learners’ dynamics of written Estonian on A2- and B1-level’]. – *Lähivõrdlusi. Lähivertailuja*, 30, 57–88. <https://doi.org/10.5128/LV30.01>
- Eslon, Pille; Õim, Katre; Kaivapalu, Annekatrin; Argus, Reili; Matsak, Erika 2010. Kuidas uurida esimese ja teise keele omandamist? [‘Perspectives of the research of the first and of the second language acquisition’] – *Lähivõrdlusi. Lähivertailuja*, 20, 11–48. <https://doi.org/10.5128/LV20.01>
- François, Thomas; Fairon, Cédric 2012. An ‘AI readability’ formula for French as a foreign language. – *Proceedings of the EMNLP and CoNLL 2012*.
- Hallik, Virgo 2016. Eesti vahekeele korpuse klasteranalüüsi vahendite kasutamine teksti keeletaseme prognoosimisel [‘Using Estonian Interlanguage Corpus Cluster Analysis Tools to Predict Language Level of Text’]. *Bakalaureusetöö*. Tallinn: Tallinna Ülikool.
- Hancke, Julia; Meurers, Detmar 2013. Exploring CEFR classification for German based on rich linguistic modeling. – *Proceedings of the Learner Corpus Research (LCR) conference*.
- Hausenberg, Anu-Reet; Ilves, Marju; Kaivapalu, Annekatrin; Kerge, Krista; Kern, Katrin; Kitsnik, Mare; Krall, Ingrid; Rummo, Karin; Rüütmaa, Tiina 2008. Iseseisev keelekasutaja. B1- ja B2-taseme eesti keele oskus [‘Independent user: B1- and B2-level proficiency in Estonian’]. Tallinn: Riiklik Eksami- ja Kvalifikatsioonikeskus.
- Hawkins, John A.; Filipović, Luna 2012. Criterial features in L2 English: Specifying the Reference Levels of the Common European Framework. *English Profile Studies* 1. Cambridge: Cambridge University Press.
- Ilves, Marju 2008. Algaja keelekasutaja. A2-taseme eesti keele oskus [‘Estonian for beginners: A2-level Proficiency in Estonian’]. Tallinn: Eesti Keele Sihtasutus.
- Ilves, Marju 2010. Läbimurre. A1-taseme eesti keele oskus [‘A1-level Proficiency in Estonian’]. Tallinn: Eesti Keele Sihtasutus.

- Kallas, Jelena; Koppel, Kristina; Tuulik, Maria 2015. Korpusleksikograafia uued võimalused eesti keele kollokatsioonisõnastiku näitel [‘New possibilities in corpus lexicography based on the example of the Estonian Collocations Dictionary’]. – Eesti Rakenduslingvistika Ühingu aastaraamat, 11, 75–94. <https://doi.org/10.5128/ERYa11.05>
- Kallas, Jelena; Koppel, Kristina; Paulsen, Geda; Tuulik, Maria 2015. Eesti keele naabersõnade sõnastik [‘Estonian Collocations Dictionary’]. Eesti Keele Instituut. <http://www.eki.ee/dict/kol> (1.10.2020).
- Kallas, Jelena; Tiits, Mai; Tuulik, Maria 2014. Eesti keele põhisõnavara sõnastik [‘Basic Estonian Dictionary’]. Madis Jürviste, Kristina Koppel, Maria Tuulik (Koost). Tallinn: Eesti Keele Sihtasutus. <https://www.eki.ee/dict/psv/> (1.10.2020).
- Kerge, Krista 1996. Funktsionaalse keeleoskuse taseme määratlemine [‘Defining the Level of Functional Language’]. Tallinn: Haridusministeerium.
- Kerge, Krista 2008. Vilunud keelekasutaja. C1-taseme eesti keele oskus [‘C1-level Proficiency in Estonian’]. Tallinn: Eesti Keele Sihtasutus.
- Kitsnik, Mare 2018. Iga asi omal ajal: eesti keele B1- ja B2-taseme verbikonstruktsioonid keeleoskuse arengu näitajana [‘All in Good Time: Estonian B1- and B2-level Verbal Constructions as Indicators of the Development of Language Proficiency’]. Humanitaarteaduste dissertatsioonid 43. Tallinn: Tallinna Ülikooli Kirjastus.
- Koppel, Kristina 2019. Eesti keele kui teise keele õpikute lausete analüüs ja selle rakendamine eri keeleoskustasemetega sõnastike näitelause automaatsel valikul [‘Analysis of CEFR-graded coursebook sentences and their use for automatic detection of good dictionary examples’]. – Eesti Rakenduslingvistika Ühingu aastaraamat, 15, 99–119. [10.5128/ERYa15.06](https://doi.org/10.5128/ERYa15.06).
- Koppel, Kristina 2020. Näitelause korpuspõhine automaattuvastus eesti keele õppe-sõnastikele [‘Corpus-based Automatic Detection of Example Sentences for Dictionaries for Estonian Learners’]. *Dissertationes linguisticae Universitatis Tartuensis* 38. Tartu: Tartu Ülikooli Kirjastus.
- Koppel, Kristina; Tavast, Arvi; Langemets, Margit; Kallas, Jelena 2019. Aggregating dictionaries into the language portal Sõnaveeb: Issues with and without a solution. – Iztok Kosem, Tanara Zingano Kuhn, Margarita Correia, José Pedro Ferreira, Maarten Jansen, Isabel Pereira, Jelena Kallas, Miloš Jakubiček, Simon Krek, Carole Tiberius (Eds.), *Proceedings of the eLex 2019 conference*. 1–3 October, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o, lk 434–452.
- Laur, Mall 1998. Tuumiksõnade loend [‘List of core words’]. – M. Laur, Esimene verstaapost. Eesti keele suhtluse algtaase. Tallinn: REKK.
- North, Brian; Ortega, Angeles; Sheehan, Susan 2010. The British Council-EAQUALS Core Inventory for General English. https://www.equals.org/wp-content/uploads/EAQUALS_British_Council_Core_Curriculum_April2011.pdf (1.10.2020).
- O’Keeffe, Anne; Geraldine, Mark 2017. The English Grammar Profile of learner competence: Methodology and key findings. – *International Journal of Corpus Linguistics*, 22 (4), 457–489. <https://doi.org/10.1075/ijcl.14086.oke>
- Pajupuu, Hille; Kerge, Krista; Alp, Pilvi 2009. Sõnavara loomulik rikkus haritud keeleoskaja tekstides [‘Natural lexical richness in educated language use’]. – Eesti Rakenduslingvistika Ühingu aastaraamat, 5, 187–196. <https://doi.org/10.5128/ERYa5.12>
- Pajusalu, Renate; Tragel, Ilona; Veismann, Ann; Vija, Maigi 2004. Tuumsõnade semantikat ja pragmaatikat [‘Semantics and Pragmatics of Core Words’]. Tartu: Tartu Ülikool.
- Pilán, Ildikó; Vajjala, Sowmya; Volodina, Elena 2015. A readable read: Automatic assessment of language learning materials based on linguistic Complexity. – *Journal of Computational Linguistics and Applications*, 7, 143–159. <http://arxiv.org/abs/1603.08868> (17.9.2020).
- Pilán, Ildikó; Vajjala, Sowmya; Zesch, Torsten 2016. Predicting proficiency levels in learner writings by transferring a linguistic complexity model from expert-written

- coursebooks. – Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. Osaka, Japan, December 11–17, 2101–2111.
- Pärn, Hele; Simm, Leeni 1989. Eesti keele baassõnastik = Базовый словарь эстонского языка [‘Basic Dictionary of Estonian’]. Tallinn: Keele ja Kirjanduse Instituut.
- Szabo, Tunde 2018a. Collated Representative Samples of Descriptors of Language Competences Developed for Young Learners Aged 7–10 Years. Council of Europe. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680697fca> (1.10.2020).
- Szabo, Tunde 2018b. Collated Representative Samples of Descriptors of Language Competences Developed for Young Learners Aged 11–15 Years. Council of Europe. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680697fc9> (1.10.2020).
- Sung, Yao-Ting; Lin, Wei-Chun; Dyson, Scott Benjamin; Chang, Kuo-En; Chen, Yu-Chia 2015. Leveling L2 texts through readability: Combining multilevel linguistic features with the CEFR. – *The Modern Language Journal*, 99 (2), 371–391. <https://doi.org/10.1111/modl.12213>
- Zesch, Torsten; Wojatzki, Michael; Scholten-Akoun, Dirk 2015. Task-independent features for automated essay grading. – Joel Tetreault, Jill Burstein, Claudia Leacock (Eds.), *Proceedings of the Building Educational Applications Workshop at NAACL. Association for Computational Linguistics*, 224–232. <https://doi.org/10.3115/v1/W15-0626>
- Tavast, Arvi; Langemets, Margit; Kallas, Jelena; Koppel, Kristina 2018. Unified data modeling for presenting lexical data: The case of EKILEX. – Jaka Čibej, Vojko Gorjanc, Iztok Kosem, Simon Krek (Eds.), *Proceedings of the XVIII EURALEX International Congress: EURALEX: Lexicography in Global Contexts*, Ljubljana, 17-21 July 2018. Ljubljana University Press, Faculty of Arts, 749–761.
- Vaiss, Natalia; Ogren, David; Metslang, Helena; Pool, Raili 2019. Sihitisesõnastik. Eesti Keele Instituut. <http://www.eki.ee/dict/asp/> (1.10.2020).
- Vajjala, Sowmya; Lõo, Kaidi 2013. Role of morpho-syntactic features in Estonian proficiency classification. – *Proceedings of the 8th Workshop on Innovative Use of NLP for Building Educational Applications (BEA8)*. Atlanta, Georgia, USA: Association for Computational Linguistics, 63–72.
- Volodina, Elena; Kokkinakis, Sofie Johansson 2013. Compiling a corpus of CEFR-related texts. – *Proceedings of the Language Testing and CEFR conference*, Antwerpen, Belgium, May 27–29, 2013. <https://gup.ub.gu.se/file/117087> (30.9.2020).
- Östling, Robert; Smolentzov, André; Tyrefors, Björn; Höglin, Erik 2013. Automated essay scoring for Swedish. – *The 8th Workshop on Innovative Use of NLP for Building Educational Applications*.
- Xia, Menglin; Kochmar, Ekaterina; Briscoe, Ted 2016. Text readability assessment for second language learners. – *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*. San Diego, CA. Association for Computational Linguistics, 12–22. <https://doi.org/10.18653/v1/W16-0502>

Võrgumaterjalid

- Ametniku, töötaja ning füüsilisest isikust ettevõtja eesti keele oskuse ja kasutamise nõuded. RT I, 27.06.2011, 1. <https://www.riigiteataja.ee/akt/129122011169> (1.10.2020).
- CEFRLex. <http://cental.uclouvain.be/cefrlex/> (1.10.2020).
- CHILDES. <https://childes.talkbank.org/browser/index.php?url=Other/Estonian/> (1.10.2020).
- Eesti keele A1–C1 õpikute sisu korpus 2017. <https://doi.org/10.15155/3-00-0000-0000-0000-06ADEL>

Eesti keele A1–C1 õpikute lausete korpus 2020. <https://doi.org/10.15155/3-00-0000-0000-0000-0848BL>

Eesti keele kui teise keele kooliõpikute lausete korpus 2020. <https://doi.org/10.15155/3-00-0000-0000-0000-08236L>

Eesti keele kui teise keele kooliõpikute sisu korpus 2020. <https://doi.org/10.15155/3-00-0000-0000-0000-083AEL>

Eesti keele õppija korpus EMMA. <https://korpused.keeleressursid.ee/emma/index.php> (1.10.2020). <https://doi.org/10.15155/3-00-0000-0000-0000-083AEL>

Ekilex. <https://ekilex.eki.ee/> (1.10.2020).

EstNLTK teek Pythoni jaoks (ver.1.6.2b). <https://github.com/estnltk/estnltk> (1.10.2020). <https://doi.org/10.15155/1-00-0000-0000-0000-0015EL>

EVKK = Eesti vahekeele korpus. <http://evkk.tlu.ee/> (2.2.2021).

Gümnaasiumi riiklik õppekava. RT I, 29.08.2014, 21. <https://www.riigiteataja.ee/akt/129082014021> (1.10.2020).

Haridus- ja Noorteamet: Materjalid lasteaedadele ja õpetajatele. <https://www.innove.ee/oppevara-ja-metoodikad/muu-kodukeelega-laps/abimaterjalid-lasteaedadele/> (1.10.2020).

Haridus- ja Noorteamet: Eesti keele tasemeeksamid. <https://harno.ee/eksamid-testid-ja-uuringud/eksamid-testid-ja-lopudokumendid/eesti-keele-tasemeeksamid> (1.10.2020).

Kallas, Jelena, Kristina Koppel 26.03.2018. Eesti keele ühendkorpus 2017. Center of Estonian Language Resources. <https://doi.org/10.15155/3-00-0000-0000-0000-071E7L>

Kallas, Jelena; Koppel, Kristina 2018a. Eesti keele B1-taseme sõnavara. Tallinn: Eesti Keele Instituut. <http://www.eki.ee/keeletase/lists/B1.pdf> (1.10.2020).

Kallas, Jelena; Koppel, Kristina 2018b. Eesti keele A2-taseme sõnavara. Tallinn: Eesti Keele Instituut. <http://www.eki.ee/keeletase/lists/A2.pdf> (1.10.2020).

Kallas, Jelena; Koppel, Kristina 2018c. Eesti keele A1-taseme sõnavara. Tallinn: Eesti Keele Instituut. <http://www.eki.ee/keeletase/lists/A1.pdf> (1.10.2020).

Kallas, Jelena; Koppel, Kristina 2018d. etLex täiskasvanutele: eesti keele kui teise keele eri keeleoskustasemetega sõnavara ja teksti tasemesobivuse hindamine. <https://doi.org/10.15155/3-00-0000-0000-0000-071F3L> (1.10.2020).

Keeleseadus. RT I, 16.06.2020, 5. <https://www.riigiteataja.ee/akt/116062020005> (1.10.2020).

Keeleõppija Sõnaveeb. <https://sonaveeb.ee/lite> (1.10.2020). <https://doi.org/10.15155/3-00-0000-0000-0000-085B6L>

Koolieelse lasteasutuse riiklik õppekava. RT I 2008, 23, 152. <https://www.riigiteataja.ee/akt/12970917> (1.10.2020).

KORP. <https://korp.keeleressursid.ee/> (1.10.2020).

Lärka. <https://spraakbanken.gu.se/larka/> (1.10.2020).

Põhikooli riiklik õppekava. RT I, 14.07.2020, 24. <https://www.riigiteataja.ee/akt/129082014020> (1.10.2020).

SkELL. <https://skell.sketchengine.eu/> (1.10.2020).

Sõnaveeb. Eesti Keele Instituut, 2020. <https://sonaveeb.ee/> (1.10.2020). <https://doi.org/10.15155/3-00-0000-0000-0000-0823EL>

Õpetaja tööriistad. Eesti Keele Instituut, 2020. <https://sonaveeb.ee/teacher-tools> (1.10.2020). <https://doi.org/10.15155/3-00-0000-0000-0000-08357L>

ESTONIAN AS A SECOND LANGUAGE TEACHER'S TOOLS IN THE INSTITUTE OF ESTONIAN LANGUAGE'S LANGUAGE PORTAL SÕNAVEEB

**Jelena Kallas¹, Kristina Koppel¹,
Raili Pool^{1,2}, Katrin Tsepelina¹,
Tiiu Üksik¹, Pilvi Alp³, Anu Epner³**

Institute of the Estonian Language¹, University of Tartu²,
Education and Youth Board³

The paper presents the interim results of the project Teacher's Tools (Õpetaja tööriistad) published as a subpage of the language portal Sõnaveeb. The toolbox includes four modules: vocabulary, grammar, language use situations and text evaluation. The tools are aimed to help second language teachers and specialists plan courses and create new educational materials, exercises and tests.

The methodological framework and CEFR level evaluation for Teacher's Tools is based on the Common European Framework of Reference for Languages: Learning, teaching, assessment (2001), its Companion Volume with New Descriptors (2018), Collated Representative Samples of Descriptors of Language Competences Developed for Young Learners for Ages 7–10 (Szabo 2018a) and 11–15 (Szabo 2018b) and Estonian legislation on the topic. The methodology is adapted from similar projects for other languages (e.g. Capel 2010, 2012, O'Keeffe, Geraldine 2017, Alfter et al. 2019).

In order to gather linguistic data the Institute of Estonian Language compiled Estonian language coursebook and learner's language corpuses in 2018–2020. First, the textbooks were studied for creating wordlists and analysis of explicit grammar teaching. Second, the results were validated by experts and compared to the wordlists created on the basis of learners' texts.

The vocabulary and grammar modules represent CEFR-based lexical and grammar profiles for learners of Estonian as a Second Language. The lexical profile covers both young (preA1–B2) and adult (A1–C1) learners, the grammar profile the young learners (pre A1–B2). The text evaluation module runs on morphological analyser estNLTK v 1.6 and marks lemmas in texts according to their CEFR-assignment in vocabulary profile. The language use situation module is mainly based on the descriptors for young learners (Szabo 2018a, 2018b) and is to offer information about the typical situations where the learner should be able to communicate.

Keywords: Corpus Lexicography, Linguistic Competence, Grammar Competence, CEFR, Second Language acquisition, Estonian as a Second Language

Jelena Kallas (Eesti Keele Instituut) on "Eesti keele ühendsõnastiku" töörühma liige ja Keeleõppija Sõnaveebi infokihi töörühma juht. Põhilised uurimisvaldkonnad: korpusleksikograafia, õppeleksikograafia.
Roosikrantsi 6, 10119 Tallinn, Estonia
jelena.kallas@eki.ee

Kristina Koppel (Eesti Keele Instituut) on "Eesti keele ühendsõnastiku" töörühma liige ja sünonüümide infokihi töörühma juht. Põhilised uurimisvaldkonnad: korpuslingvistika, e-leksikograafia.
Roosikrantsi 6, 10119 Tallinn, Estonia
kristina.koppel@eki.ee

Raili Pooli (Tartu Ülikool, Eesti Keele Instituut) peamised uurimisvaldkonnad on eesti keele kui teise keele omandamine ja kasutus.
Jakobi 2-437, 51005 Tartu, Estonia
raili.pool@ut.ee

Katrin Tsepelina (Eesti Keele Instituut) on vanemtarkvaraarendaja. Põhilised uurimisvaldkonnad on korpuslingvistika, korpusleksikograafia.
Roosikrantsi 6, 10119 Tallinn, Estonia
katrin.tsepelina@eki.ee

Tiiu Üksik (Eesti Keele Instituut) töötab arvutilingvistina. On uurinud eesti keele kui teise keele grammatika õpetamist õpikutes.
Roosikrantsi 6, 10119 Tallinn, Estonia
tiiu.yksik@eki.ee

Pilvi Alp (Haridus- ja Noorteamet) tegeleb keeleoskuse mõõtmisega.
Lõotsa 4, 11415 Tallinn, Estonia
pilvi.alp@harno.ee

Anu Epneri (Haridus- ja Noorteamet) uurimisvaldkond on keeleoskuse mõõtmine.
Lõotsa 4, 11415 Tallinn, Estonia
anu.epner@harno.ee