

# EESTI KEEL UUES KUUES

## Eesti keele sõnaraamat 2018

MARGIT LANGEMETS, MAI TIITS, UDO UIBO,  
TIIA VALDRE, PIRET VOLL

**K**irjeldasime üsna põhjalikult 2010. aastal ilmunud artiklis „*In spe*: ühekõiteline eesti keele sõnaraamat” (Langemets jt 2010), kuidas kavatseme koostada ühekõitelist, nn lauasõnaraamatu mõõtu teost. Selline sõnaraamat kujutab endast tavaliselt järgmist leksikograafilist korrust võimsal vundamendil, milleks sageli on pika aja jooksul paljudes köidetes ilmunud akadeemiline sõnaraamat – meil on selleks mahukas eesti keele seletav sõnaraamat (EKSS 1988–2007, 2. trükk EKSS 2009). Uue ühekõitelise sõnaraamatu ettevalmistustööd alustasime 2009. aastal, rööbiti sõnaraamatu 2. trüki (EKSS 2009) toimetamisega.

Ometi on asjad läinud teisiti. Kümne aastaga on ühiskond muutunud niivõrd, et oleme korduvalt pidanud häälestama ja täpsustama nii oma sõnaraamatu olemust kui ka selle meediumit ehk vahendit, mille abil infot kasutajani tuua. Uus sõnaraamat ilmub veebis, Eesti Keele Instituudi uues Sõnaveebi portaalis, moodustades ühtlasi selle portaali selgroo. Enamgi veel, koos uue seletava sõnaraamatuga ilmub Sõnaveebis ka kollokatsioonide ehk naaber-sõnade sõnastik (vt Kallas jt 2015).

Järgnevas kirjeldame, mis vahepeal on juhtunud. Uue sõnaraamatu sisust käsitleme vaid mõningaid aspekte: uusi sõnu, allmärksõnu ja leksikaliseerunud/grammatiseerunud vorme, kasutusnäiteid ja etümoloogilist infot.

## Kümne aastaga tulevikku

Kui 2009–2010 hakkasime kavandama ühekõitelist pabersõnaraamatut, olid suured muutused juba alanud. Umbes 2004.–2005. aastal sündinud uue põlvkonna veebidisaini Veeb 2.0 seostatakse eeskätt interneti plahvatusliku levikuga, osalusveebi ja kollektiivse intelligentsi mõistetega (Kalda 2009). Veeb 2.0 tõi kaasa uued tehnoloogiad ja standardid, läbimõeldud kasutusmugavuse, üldise avatuse. Veebist sai üldine äriplatvorm. Sõnastikumaaailm ei saanud kuidagi jääda sellest puutumata.

2003. aastal valmis instituudi veebipõhine sõnastikusüsteem EELex (Langemets jt 2006), mida arendati kuni 2015. aastani ning mis on meid teeninud siiani. EELexi keskkonnas on valminud või tegemisel üle 70 sõnastiku. ASTRA programmi rahastusel loome praegu uut sõnastike ja terminibaaside süsteemi Ekilex. Vahepealse aja jooksul on leksikograafia tugevasti suunda muutnud, liikudes paberilt elektrooniliseks, introspektiivsest empiiriliseks, käsitsi koostamisest (pool)automaatseks korpuspõhiseks koostamiseks, nor-

matiiivsest kirjeldavaks, inimloetavast masinloetavaks jm (Tavast jt 2018). Rõhutatakse, et nüüdse aja digisõnastikes on oluline lahus hoida andmete esituskuju (arendajate terminina ingl *front-end*) ja andmed ise (ingl *back-end*) (Tavast jt 2018; Gorjanc jt 2017: 11). Kolmas komponent on neid kihte ühendav rakendusliides ehk API.

Selline ühtsel andmemudelil põhinev sõnastikusüsteem võimaldab integreerida eri andmeallikaid (sõnastikke, korpusi jm), näidates kasutajale ühes kohas võimalikult mitmekesisist infot, samuti filtreerida leksikaalset infot vastavalt kasutaja oskustele ja vajadustele (nt lihtne/detailne vaade). Põhimõte on, et kasutaja näeks infot sõna, mitte sõnastike kohta. Euroopa leksikograafias liigutakse praegu just nimelt selles suunas (vt nt Horizon 2020 programmi projekti ELEXIS 2018–2021, kus Eesti Keele Instituut osaleb partnerina, <https://elex.is/>). Sama toimub ka Oxfordi ja Macmillani kirjastustes: Oxfordi sõnaraamatu ühekõiteline versioon trükiti küll veel 2010 (Jamieson 2010), aga samal ajal avati (tasulisena) ka veebileht Oxford Dictionaries Online, kus sõnaraamatut täiendavad muud materjalid, sh teistest sõnastikest, ning mis alates 2017. aastast kannab brändinime Oxford Living Dictionaries. Macmillan ühendab samamoodi olemasolevate sõnastike materjale, nt kollokatsioonid püsiühenditega, et teha baasid ühest päringuaknast hõlpsasti otsitavaks (Rundell 2014).

Kui varasem tava on olnud koostada aina uusi üksiksõnastikke (milles teatav hulk infot kordub), siis moodne leksikograafia keskendub pigem küsimusele, mil viisil täiendada või kuidas hõlpsalt taaskasutada juba olemas olevat infot. Leksikograafil peab tähelepanu jaguma lisaks kvaliteetsele sisutööle ka sõnaraamatutöö tehnilisele poolele ehk andmebaasile. Nii olemegi endalt kümne aasta jooksul korduvalt küsinud: mida me teeme, kas (paber)sõnaraamatut või andmebaasi või mõlemat? Aastal 2018 on vastus ühene: me teeme andmebaasi. 10 aasta jooksul on meie uuest sõnaraamatust saanud aina vähem jutustus ning aina rohkem struktureeritud andmebaas. Leksikograafi töö lähitulevik on järeldoimetamine: arvuti töötleb korpuse keelematerjali ning genereerib algse info, mida inimene seejärel toimetab.

Ekilexi andmebaasi soovime lõppkokkuvõttes mahutada rikkalikult infot eesti keele sõnade kohta: iga sõnakogu annab sinna juurde oma panuse. Uuest eesti keele sõnaraamatust on pärit esmane märksõnastik (120 000 sõna), tähendusjaotus, seletused ja kasutusnäited, samuti etümoloogiline info umbes 20 000 sõna juures. EKI morfoloogilisest andmebaasist tuleb kogu sõnamuutmisinfo (koos morfofonoloogilise transkriptsiooniga). Kollokatsiooni-sõnastikust (Kallas jt 2015) tulevad naabersõnad (kollokatsioonid). Edaspidi soovime liita muud olemasolevat või uut keeleinfot (nt keelenõu soovitusi jm ÕS-i infot, keeleoskustasemete sõnavara ja muutevormide kasutussageduse infot). Ekilex ühendab seni eraldi seisnud sõnastikud ja terminibaasid ühtsesse andmemudelisse (Tavast jt 2018). Kasutaja jaoks loome kaks uut portaali: Sõnaveeb koondab üldsõnastike infot (avame 2019. aasta veebruaris), Terminiveeb koondab terminibaaside infot (avame 2020). Kaks portaali sünnib seetõttu, et andmete esituskuju (ingl *front end*) on erinev: Sõnaveebis on info sõnade kohta, Terminiveebis mõiste kohta.

## Paberilt veebi, veebist mobiili

Kolimine paberilt veebi tekitab alati vastakaid tundeid: see, mis ühtele on kurb uudis, paistab teistele suure vabanemisena (Rundell 2002). Tõsiasi on, et ligi pooled (44,7%) praegu Euroopas töös olevatest sõnaraamatutest ilmuvad ainult veebis. Kolmandik (35,8%) sõnaraamatutest ilmub nii paberil kui ka veebis. Ainult raamatuna ilmub viiendik (19,5%) sõnastikest (ELEXIS Survey 2018). Suund on ilmne – korruga paberil ja veebis ilmuv sõnastik on üldjuhul veebisõnastikule üleminekut mahendav eelaste. Nii on juba juhtunud Macmillani kirjastuses (Rundell 2002, 2009, 2014), samuti Oxfordi kirjastuses (Jamieson 2010), Eesti Keele Instituudis, küllap mujalgi.

Veebi kiituseks on sõna võtnud näiteks lingvist Martin Haspelmath (2017). Tema leiab (nagu meiegi), et sõnaraamatud on loomult struktureeritud andmebaasid. Lineaarsed on nad üksnes sellepärast, et raamatus ruumi kokku hoida. Teda kui tüpoloogii huvitavad väikesed keeled, mille sõnastike sisu ta soovib võrrelda. Ta on ise algatanud elektroonilise andmebaasi Dictionaria (<https://dictionaria.cldd.org/>), kus eri keelte sõnastikud on esitatud relatsioonilise, tabelitest koosneva andmebaasina ning mille veebilehel saab teha päringuid iga infoüksuse kaupa (nt sõnaliigi, semantilise tüübi, sõnatähenduste võrdluse järgi). Dictionaria on küll veel lapsekingades, aga Haspelmath selgitab meelasti, mis tas on head võrreldes lineaarsete sõnastikega (Haspelmath 2017).

Raamatut seevastu on kiitnud leksikograaf ja arheoloog Barbara Ann Kipfer (2013). Tema meelest kingib raamat õnnelikke juhuseid: seda lehitsetes, leheküljel ringi uidates võib kohata palju muud huvitavat. Kipfer hindab ka aju treenimist: tuleb lahti mõtestada kasutatud tähised, lühendid ja sümbolid, otsida õiget kohta nii tähestikust kui ka pikast sõnakirjeldusest. Sõnaraamatus pole veebi virvarri, häirivaid reklaame ega muud. Käte vahel on puhas tarkus, mis tekitab imelise tunde.

Nüüdseks ollakse kolimas (või juba kolitud) arvutist mobiili, mis on mõjutanud ka sõnastikukasutajate harjumusi ja ootusi. Elektroonilise leksikograafia konverentsidel, kus „mõeldakse väljaspool paberit” (Langemets 2014), on sõnaraamatut nähtud kui rikkalikku ja funktsionaalset andmekogu: väga hea, kui keeleinfot on palju, samas pole seda kõike mõistlik korruga välja käia. Veebisõnastike tegijad on üksmeelsed ses osas, et pealiskiht tuleks hoida võimalikult lihtne. Spetsiifilise info juurde peaks kasutaja jõudma sammhaaval, vastavalt vajadusele, avades järjest detailsemaid infokihte.

Veebisõnastikud kasutavad üldjuhul kahesugust taktikat: sõnastikke uuendatakse perioodiliselt (nt kord kvartalis, kord aastas) või avalikustatakse iga toimetatud muutus kohe. Inglise veebisõnastikud (Oxford English Dictionary, Macmillan English Dictionary) uuendavad oma versioone kord kvartalis, iga kord lisandub umbes 120–150 uut sõna (Rundell 2014). Leksikograafid jälgivad ka tähenduste muutumist ja ajakohastavad pidevalt tähenduste seletusi (nt telefon pole enam laual seisev kettaga seade, sõnaraamat pole raamat). Poola ja Hollandi suured ükskeelsed veebisõnaraamatud avalikustavad oma täiendused kohe, kui nad on tehtud (Žmigrodzki 2014; Tiberius, Schoonheim 2016). Uue eesti keele seletava sõnaraamatu täiendusi hakkame samuti avaldama kohe, aga uue versiooni arhiveerime kord aastas. (Sama puudutab Sõnaveebi teisi allikaid.)

## Uued sõnad ja tähendused

Uues eesti keele sõnaraamatus leidub umbes 17 000 uut märksõna ja 1300 uut tähendust. Need suured arvud vajavad selgitust, sest neologisme ehk uudiskeelendeid, mis on kasutusse tulnud viimase paarikümne aasta jooksul ning mida keelekasutajad tajuvad uuena, leidub selles hulgas palju vähem – ehk umbes 5000.

Esiteks on uute sõnade hulka suurendanud see, et oleme lisanud ca 1300 üldkeeles sageli esinevat pärisnime, toomaks välja nende sisu, kirjakuju, muutmisviisi, kasutuskonteksti, paralleelvorme jm keelekasutaja jaoks olulist infot. Pärisnimede valikus esineb enim riigi-, koha-, organisatsiooni-, sündmusenimesid ja peamiselt müütilisi või kirjanduslikke isikunimesid, nt *Allah, Amor, Beneluxi maad ja Benelux, Bundestag ja Liidupäev, Buratino ja Pinocchio, Checkpoint Charlie, Hendriksoni küür, Kahejõemaa, Kaitseliit, Kopli lüürid, Kotšinihiina, Kreenholm, Kristalliöö, Liiva-Annus ja Liiva-Hannus, Loodearmee, Lubjanka, Läänekallas, Magrib, Maksameri, Merevaigutee, Nutumüür, Paks Margareeta, Päikesekuningas, Päästearmee, Räniorg ja Silicon Valley, Sossi mägi, Stromka, Taliban, Tšornobõl ja Tšernobõl, Valhalla, Ämblikmees*. Näiteks:

### **Elu Sõna** <pärisnimi>

1980. aastail Eestis levima hakanud uusususundiline liikumine

▫ *Elu Sõna kogudused*.

Teiseks oleme märksõnadena esitanud ühendeid, mida varem pole seletavas sõnaraamatus iseseisvate üksustena käsitletud, ent mis keelekasutuses sellisena sageli ilmnevad. Need ühendid on üldtuntud ja „vanad”, ent varem on nad ära toodud kas näidete hulgas (koos seletusega või ilma) või jäänud sõnaraamatust üldse välja. Oleme lisanud ja ära seletanud nt järgmised märksõnad: *alkohoolne jook, alumine vererõhk, baaside leping, elu ja surma küsimus, geograafiline laius, helkivad ööpilved, huvi pakkuma, häälest ära, identsed kaksikud, iga rakuga, inetu pardipoeg, juua täis, karva ajama, klassikaline filoloogia, kodanlik natsionalism, kohut mõistma, kuiv kõha, lohku tõmbama, löömaks minema, metsalise märk, mine metsa!, naiste paradüüs, pidu panema, pähe õppima, tähelepanu juhtima, viina viskama* jpm. Näiteks:

### **inimene tänavalt**

suvaline inimene, tavaline inimene (nt vastandina professionaalile, asjatundjale)

▫ *Mida teab inimene tänavalt kvantmehhaanikast!*

Kolmandaks oleme märksõnadena välja toonud iseseisva leksikaalse üksuse moodustanud verbi- ja noomenivorme. Näiteks:

### **kadunud**

**1** hukkamisele või läbikukkumisele määratud

▫ *Kui ma vahele jään, olen kadunud mees.*

**2** surnud

▫ *Kadunud onu teadis arvutiasjandusest kõike.*

### **kaelani**

täiesti, üleni, läbinisti

▫ *Firmaomanik sipleb kaelani võlgades.*

Neljandaks, kuna eesti keele sõnaraamatu sihiks on kajastada tegelikku keelekasutust, oleme sisse võtnud juba pikemat aega kasutuses olnud, harilikult kõnekeelseid või normatiivses keelekasutuses taunitud sõnu, mis esinevad tänapäeva eesti keele korpuses, ent seletavas üldkeele sõnaraamatus varem fikseeritud pole olnud, nt *eputrilla, tatsipau, essa, tessa, pursuika, jorss, jummel, ludri, minett, parlanks, peika, pruta, piider, kakama, põhulõug, põiekas, mankurt, sitsiprintsess, sasku, munapiiks, glasnost, samizdat, stange*, aga ka *tupiktee, tupiktänav, maoism, turundama ja kehand*.

Ima kahtluseta on uute sõnade seas kirsiks tordil n-õ päris uued, hiljuti keelde tulnud sõnad ja väljendid (nt väljend *kirsiks tordil* ise). Meie leksikograafid monitoorivad nii paber- kui ka *online*-kirjasõna ning panevad uute sõnade baasi kirja leitud uued keelendid: nende kirjakuju, tähenduse, kasutusnäited, allika(d) jm. Sel viisil uute sõnade fikseerimine tugineb eeskätt leksikograafide kompetentsile. Leidude hindamisel on abiks korpusest saadav info sõnade esinemissageduse, kasutusjuhtude ja kollokatsioonide kohta. Automaatset uute sõnade tuvastajat meil veel kasutada ei ole, see on tuleviku-soov.

Uue sõna tavapäraseim teekond üldkeelde on tema väljumine mõne kitsama eriala või subkultuuri keelekasutuse nišist sageda kasutuse tõttu. Uusi sõnu on tulnud enam-vähem igalt elualalt, ootuspäraselt torkab silma infotehnoloogia valdkond, aga ka poliitika, majandus, ühiskonnaelu, meditsiin, kokandus, vaba aja harrastused ja vaimsed praktikad. Enamiku uutest sõnadest moodustavad liitsõnad, palju on võõrsõnu ja laene. Näiteks paistavad kulinaarsed uudissõnad iseäranis silma sellega, et lisavad eesti keelde ohtralt võõr- ja tsitaatsõnu, kuna kannavad meieni toitude ja toiduainete nimetusi maailma eri paigust.

On loomulik, et suurem osa uusi sõnu tähistab hiljaaegu meie ellu astunud nähtusi, asju, omadusi või tegevusi, nt *erisoodustusmaks, eelhääletus, eeslinnastumine, euroskeptitsism, e-valimised, vastustaja, makroanalüütik, asümmeetriline sõda, elektrooniline võru ja elektronvõru, veebimajutus, managerima, puuküürnik, putinism, hapu laen, infoväli, islamimaailm, islami-seeruma, islamifoobia; küüslauguvõi, maitsool, kabotša, flamm; epiduraal- anesteesia, kompuutertomograaf, mammograafiabuss ja mammobuss, unearst, hapnikuvõlg; maalilaager, kelguhoki, kendo, enduuro, eeterkeha ja eeterlik keha, sisekosmos, energiaväli, üliteadvus, sekstiil*. Näiteks:

## **maailma | koristus**

prügi koristamine talgute vormis ühisaktsioonina paljudes riikides

▫ *Järgmisel aastal toimuval maailmakoristusel on kavas prügi kokku korjata vähemalt sajas riigis.*

Hea teada: Ülemaailmsed prügikoristustalgud on eestlaste algatatud.

## **kroksid<sup>1</sup> pl**

firma Crocs kerged vabaajajalatsid, mis pehmenevad vastavalt kehatemperatuurile

▫ *Kroksidega ei tahtnud kohvikusse minna.*

Teada-tuntud nähtusi, asju, omadusi või tegevusi tähistavad neologismid on kasutusele võetud tahtmisest asju näha ja neist rääkida uut moodi, teisiti kui varem. Sõnu on laenatud – põhiliselt inglise keelest –, tegeldud on sõnaloomega. Nii on sünonüümidena üldkeelde tulnud sõnause laureaadid *taristu* ('infrastruktuur') ja *vabakond* ('kolmas sektor'), samuti *intellektipuue*, *tapjahai*, *tüpo* jpt. Kasutusse on tulnud nt uus sõna *vedajafirma* vaatamata sellele, et leidub kuus samatähenduslikku sünonüümi (*transpordiettevõte*, *transpordifirma*, *veoettevõte*, *veofirma*, *veondusettevõte*, *veondusfirma*). Eesti keele sõnaraamat sünonüümseid sõnu paremusjärjekorda ei aseta.

Eesti keele sõnaraamat esitab märkimisväärselt kõnekeelseid sõnu ja väljendeid. Seda joont on jätkatud ka uudissõnade osas. Nii leiduvad sõnaraamatus nt *moššima*, *moppima*, *pepupoiss*, *pohhuism*.

Uute lisandunud tähenduste puhul on hakatud sõna kasutama uues tähenduses harilikult mingi sarnasuse alusel või on võõrsõna tähendus teise keele mõjul avardunud. Näiteks:

## **lõim**

**2** INFO programmi haru, mis võimaldab mitut tegevust täita samaaegselt või vaheldumisi

▫ *Protsessid ja lõimed.*

▫ *Programmilõim.*

## **kodakondne**

**2** mingi riigi kodanik

▫ *Naine ja poeg on Eesti kodakondsed, mina mitte.*

▫ *Aleksandr on nüüd Hispaania kodakondne.*

## **opereerima**

**2** mingit tehnilist või ettevõtete süsteemi (või seadet) juhtima, vastavaid töid teostama ja korraldama

▫ *Ettevõtte opereerib seitsmel erineval laevaliiniil.*

▫ *Praegu opereerib firma tanklaid Tallinnas ja Kiviõlis.*

▫ *Servereid opereeris 30-liikmeline sakslastest grupp.*

Ent on ka palju keelendeid, mis veel pole uute sõnade baasist lõplikku valikusse – eesti üldkeelt esindavasse eesti keele sõnaraamatusse – mahtunud, nt *vebinar, valvefeminist, varpajam, juntima, sõjateater, mhmitama, hullumaja puhvet, e-nina, energiapõld, maasikablond, ökokana, saagi käima tõmbama, tuubitama*. Soovime neile tuult tiibadesse!

## Sõna ja sõnavormi vahel: märksõnad ja allmärksõnad

Korpusleksikograafia ajastul tekib paljudest sõnadest sootuks uutmoodi pilt. Oluliselt on meie tööd muutnud uute korpusetöötlusvahendite kasutuselevõtt: alates 2012. aastast analüüsime sõnavara Sketch Engine'i eesti keele mooduli tööriista abil (Kallas 2013: 31–87). Sageli selgub Sketch Engine'i sõnavisan-dist ühe või teise sõnavormi esiletõusev, jõuliselt domineeriv sagedus (nt *reba* vormid *rebadel, rebadele*), on näha eripärase tähendusega sõnavormi grammatiline käitumine: kas ta seostub vabalt mitmesuguste substantiividega, nt */sõidu, võistluse, puhkuse, valimiskampaania, lennu, magamise, festivali, haiguse, kontserdi, äikese, trenni, remondi jne/ ajal*; kas tal on kinnistunud kollokatsioone, nt */telefoni, posti, meili, interneti/ teel*; kas ta esineb vahel koos täiendiga või mitte. Korpusprogramm Sketch Engine pakub mõnesugust tuge otsustamiseks, kuidas ja kus mingi eripärase tähendusega sõnavorm esitada: kas iseseisva märksõnana või teatava tähendusüksusega seotud allmärksõnana.

Järgnevate näidete puhul tuleb silmas pidada, et nende esitus (pilt) võib sõnaveebis veel muutuda.

Korpuse toel on iseseisvate märksõnadena esitatud nt sõnavormid *eestvõtmisel, eestvõttel* ja *eestvõtul* (*eestvõtmine, eestvõtte* ja *eestvõtt* asemel), toeks nende vormide valdav kasutussagedus korpuses. Mõistagi võib leida ka üksikuid teisi kasutusjuhte (*Minister tervitas erasektori eestvõtmist*), aga esituses joondume valdavast, tüüpilisest. Näiteks:

### **eestvõtul**

[kelle] algatusel, korraldamisel, juhtimisel

= eestvõtmisel, eestvõttel

▫ *Asi sai teoks vallavalitsuse eestvõtul.*

Uute kaassõnadena, mõnel puhul ka adverbina, leiab eesti keele sõnaraamatust nt *alusel, baasil, jõul, kahjuks, kokkuvõttes, käigus, mõõdus ja mõõtu, ootel, osas, pinnal, pinnale ja pinnalt, seoses, suhtes, sunnil, suunas, süles, tarbeks, tükkis*. Verbivormidest pärinevate kaassõnadena on iseseisvate märksõnadena esitatud *võrreldes* (*Võrreldes kohtuga on pressinõukogu menetlus kiire ja tasuta, Kell on seal Eestiga võrreldes viis tundi ees*), *lõpetades, sõltumata, olenemata, rääkimata, kõnelemata* (viimase kasutussagedus on korpuses küll oluliselt väiksem kui kaassõnal *rääkimata*, ent muu poolest on nad sarnased) ning ajasuhet väljendav ja evidentsiaalsuse markerina toimiv *arvates*.

Paljud sõnavormid on seotud teatava tähendusüksusega ning jäetud allmärksõna staatusse. Ka EKSS 2009 sisaldab selliseid allmärksõnu, ent oleme nende hulka tunduvalt suurendanud. Allmärksõnadena on esitatud eripärase

käitumisega sõnavormid (nt kui sõna on teatavas tähenduses kasutatud mitmuslikuna, adverbi- või kaassõnalaadsena). Näiteks:

### **kuup**

**2** arvu või avaldise kolmas aste

▪ **kuubis** KÕNEKEELNE (suurt määra rõhutavalt:) äärmiselt suur, väga suur

= ruudus

▫ *See on lausa sigadus kuubis!*

Adverbi- või kaassõnalaadse kasutuse korral võib tähendusüksuse juurde kuuluda mitu kohakäändelist allmärksõna, mis moodustavad allmärksõnasarja:

### **rool**

kettakujuline seade auto, laeva vm sõiduki või masina juhtimiseks

▪ **roolis, rooli** (sõiduki juhtimise kohta)

▫ *Kes oli avarii hetkel roolis?*

▫ *Purjuspäi ta rooli ei lähe.*

### **haigusleht**

töövõimetusleht haiguse korral

▪ **haiguslehel, haiguslehele, haiguslehel** (olukorra kohta, kus töötaja on haiguse tõttu ajutiselt töölt ära)

Allmärksõnaga tähendusüksus võib võtta kokku nimisõnana (adjektiiv-tribuudiga) ja kaassõnana (genitiivse või elatiivse substantiiviga) käsitatava sõnavormi. Näiteks:

### **komme**

**3 kombel** teatud moel, viisil; kellegi sarnaselt

▫ *Läksin hirmsal kombel närvi.*

▫ *Ameeriklase kombel tööd rabav mees.*

Uute allmärksõnadena on esitatud nimisõna teatud tähendusrühmadega seotud adessiivseid sõnavorme. Konstruksioonigrammatika kohaselt kujutavad need endast poolproduktiivset konstruksiooni, mis toimib lauses määrusena ja koosneb adessiivis tegevusnimisõnast ja selle tegevussubjekti vormistavast genitiivsest nimisõnafraasist (Sahkai 2008: 181): *juhtimisel, kaasabil, tähelepanekul, väitel, teatel, sõnul* jms. Näiteks:

### **teade**

**2** [kelle/mille] **teatel** kellegi poolt või kuskil teatatuna

▫ *Politsei teatel on osa kannatanutest kriitilises seisundis.*

▫ *Tuumajaama võimsuseks tuleb ajalehe teatel üle 1000 megavati.*

Suurem allmärksõnade või ka uute iseseisvunud (endiste) sõnavormide hulk on seotud asjaoluga, et oleme pööranud rohkem tähelepanu grammatiseerumisele, nt kaassõnade kujunemisele noomeni- või harvemini verbivormidest.



Grammatiseerumisele viitab tähenduse nihkumine abstraktsemaks ja kasutuskontekstide lisandumine, grammatiseerunud üksused muutuvad produktiivsemaks ning nende esinemissagedus kasvab (Metslang 2002; Velsker 2010). Keeruliseks teeb olukorra see, et grammatiseerumine on järkjärguline protsess. Pole sugugi lihtne määrata, kas mingi sõnavorm on oma grammatilises tähenduses juba iseseisvunud või alles teel sinna. Leksikograafile tähendabki see otsustamist, kas esitada sellised sõnavormid/sõnad iseseisva artikli märksõnana või allmärksõnana vastava noomeni, harvemini verbi artiklis.

Tegelikult võib õhku visata küsimuse, kui oluline üldse on, kuidas eripärased sõnavormid sõnu seletavas sõnaraamatus esitatakse, kas allmärksõnana lähtesõna artiklis või iseseisva sõnana. Ja et kas otstarbekam on lähtuda otsustamisel sõnavormi tähendusest (täenduslikust seosest lähtesõnaga) või grammatilisest käitumisest.

Vana kirjakeele uurijad on eristanud primaarseid ja sekundaarseid kaassõnu (Habicht, Penjam 2007: 54). Primaarsete kaassõnade seos lähtesõnaga on hämarunud, nad on keeles juurdunud, sageli ka foneetiliselt kulunud. Sekundaarsete kaassõnade seos lähtesubstantiiviga on veel läbipaistev, nad on keeles kujunemisstaadiumis. Uues „Eesti keele süntaksis” (EKS) on äsja kaassõnastunud või kaassõnaks kujunemas olevaid sõnavorme nimetatud perifeerseteks kaassõnadeks (EKS: 450). Seega kaassõnad kõik, olgu mõned teatud mõõndusega. Ei oska ju grammatikudki alati ühe või teise sõna grammatiseerumisastmete vahel selget piiri tõmmata. Võib öelda, et sõnaraamatu puhul on oluline, et sekundaarsed või perifeersed kaassõnad seal leiduksid ja seletuse saaksid, olgu siis iseseisva sõnana või tähendusüksuse juures. Praegu ei ole eesti keele sõnaraamatu pilt selliste sõna(vormi)de asjus päris ühene, nende esitus nõuab kohati veel otsuste langetamist ja vastavat toimetamist. Näiteks pleekunud tähendusega *teel* (‘millegi abil, vahendusel; mingil kombel, moel’) on esialgu esitatud allmärksõnana *tee* artiklis, kuigi võiks ka eraldi olla. Samuti on sõnartiklis *aeg* allmärksõna *ajal* (‘millegi kestel, millegi vältates algusest lõpuni; siis, kui miski parasjagu kestab või toimub’), kuigi selle sünonüümid *jooksul*, *kestel* ja *vältel* on esitatud väljakujunenud kaassõnadena (vrd ka ingl vaste *during*). (Siinkohal olgu osutatud, et EKI reeglipärase morfoloogia demo (<http://artur.eki.ee/morf/mf.aspx>) käsitleb vormi *teel* nii adverbina kui ka kaassõnana ja vormi *ajal* kaassõnana.) Erinevalt on praegu esitatud ka *arust* (allmärksõnana, postpositsiooniks saamas) ja *meelest* (postpositsioonina), nii on ka EKI reeglipärase morfoloogia demos. Otsustamist nõuab seegi, kas tuua eraldi (all)märksõnadena välja *kaugusel*, *kaugusele*, *kauguselt* ja *kõrgusel*, *kõrgusele*, *kõrguselt*, mis EKS-i järgi on ekvatiivsed kohaadverbid (EKS: 439), aga uues sõnaraamatus praegu lahustunud märksõnade *kaugus* ja *kõrgus* näidete hulka. Seega võib öelda, et kuigi eesti keele sõnaraamat saab Sõnaveebis peagi kõigile kasutatavaks, jätkub selle toimetamise töö ka edaspidi.

## Kasutusnäited

Eesti keele sõnaraamatu stardipositsioon (teadmine, et valmib ühte köitesse mahtuv pabersõnastik) seadis piirid nii mahule kui ka esitusele. Otsustasime tunduvalt piirata näitelausete hulka, muuhulgas jätta terminid hoopis näiteta

(Langemets jt 2010: 806). Tekstikorpuste maht on alates Sketch Engine'i kasutuselevõtust (2012) suurenenud umbes kaks korda: kui 2013. aastal veebitekstidega täiendatud koondkorpus sisaldas 563 miljonit sõnet, siis eesti keele ühendkorpus 2017 sisaldab juba 1,3 miljardit sõnet. Kasutusnäidete valikul oleme eeskätt toetunud korpuse andmetele: näide esindab sagedamini esilduvate näidete rühma. Et uue sõnaraamatuga rõõpselt valmib kollokatsioonisõnastik (vt Kallas jt 2015), siis seda infot me eksplitsiitselt oma sõnaraamatuse ei esita: Sõnaveebis on kogu info ühest kohast kättesaadav.

Näidete valiku põhimõtted ise muutunud ei ole: jälgime ja hindame kasutust, süntagmaatilisi funktsioone (konstruktsioone, rektsioone, kollokatsioonide), püüame säilitada neutraalsust, vältida hinnangulisust. Eraldi tuleb märkida, et näidete funktsioon on toetada seletust (mitte seda asendada), aidata sõna paremini mõista. Kõik need näited ei sobi automaatselt teist tüüpi sõnastikesse, näiteks kakskeelsetesse või õppesõnastikesse, mille ootustarve on teine. Näidete sobivus/mittesobivus on probleemina esile tulnud nii eesti-soome sõnaraamatu kui ka eesti-vene uute sõnade koostamisel: paljusid näitelauseid pole tervikuna mõtet tõlkida, pigem tuleks neis vahendada kollokatsioonide.

Mõne sõna kasutuskonteksti on kõige parem illustreerida autoritsitaadiga. Eriti hästi sobib tsitaat näitlikustama luulekeelset, naljatlevat või muul moel tavapärasest erinevat sõna või kasutust, kus õige konteksti ja terviklikult sobiva registri leiutamine mõjuks kunstlikult. Näiteks:

### **sübariit**

luksuslikku eluviisi ja meelelisi mõnuseid harrastav elunautija

*Mu hinges koos on munk ja sübariit.* Artur Alliksaar

### **sips**

▪ (vähese koguse kohta)

*.. sips erootikat mamsli magamistoas, sips nutulaulu sotsiaalabi kabinetis ..*

Lehte Hainsalu

Harva esinevate sõnade puhul aitab näide sõnast paremini aru saada. Näiteks:

### **eksempeel**

HARVAESINEV näide

▫ *Eks ma ole üks paras eksempeel küll!*

### **taie**

HARVAESINEV kunst

▫ *Galeriis eksponeeritakse tait 1960. aastatest.*

Sõnaveebis võib sõnaraamatuse hoolikalt valitud näidete kõrval näha ka automaatselt leitud veebilauseid. See on ühtaegu hea ja halb. Hea on see, et kasutaja näeb kasutusnäiteid ka siis, kui leksikograaf neid lisanud ei ole, eriti hea on, kui need on sobivad näited. Halb on see, et automaatselt leitud laused võivad olla kummalised, arusaamatud või lausa vigased. Et veidruse-vigade

kordumist vähendada, näidatakse veebilauseid vahelduvas järjekorras. Lähitulevikus see olukord kindlasti paraneb, kuna senist tööriista, heade näitelauseste valiku eesti moodulit arendatakse edasi (Koppel 2017, 2019 ilmumas). Praegu ei suuda programm eristada polüseemseid tähendusi ega (esialgu ka) homonüüme. See tähendab seda, et kui Sõnaveebi otsing „tamm” annab meile nii *tamme* (‘puu’) kui *tammi* (‘ehitis’), siis veebilauseste näidetes domineerivad *tammed*, sest kasutus on ilmselt selle sõna kasuks, *tammi* tuleb nende näidete hulgast ise otsida. Samuti ei ole arvestatud otsisõna vormihomonüümiaga. Nii annab nimisõnapäring *kalla* vastuseks lauseid, kus tegemist on tegusõna *kal-lama* käskiva kõneviisi 2. pöörde vormiga (Koppel 2019 ilmumas). Meie sõnaraamatus on probleem ka see, et kuna heade näitelauseste moodul on loodud keeleõppijale (õppesõnastiku jaoks), siis on programmist eemaldatud vulgariidid, tundlikud sõnad ja sõimusõnad, nende mugandused, aga ka kirjakeele normist erinevalt kirjutatud sõnad (Koppel 2017). Nii peab Sõnaveebi kasutaja (esialgu) leppima, et nn mustas nimekirjas olevate sõnade kohta veebilauseid ei näidata.

## Sõnade päritolu

Uues eesti keele sõnaraamatus on lühidalt kommenteeritud ka sõnade päritolu, nagu seda teevad teisedki suured seletussõnaraamatud (Webster, Dudeni võrguversioon jt). Seejuures pole eesmärk käsitluse ammendavuse mõttes etimoloogiasõnaraamatuga konkureerida, vaid anda sõnadele mõnesugust ajaloolist tausta, et kasutaja saaks teiste allikate poole pöördumata vajaduse korral esialgse vastuse. Me ei esita alternatiivseid etimoloogiaid, hüpoteese ega oletusi; käsitlus on paratamatult lihtsustav ning pigem konservatiivne ja maalähedane kui lennukalt uljas. „Eesti etimoloogiasõnaraamatuga” (2012) võrreldes on sõnavalik suurem, kuid käsitlus napim; „Võõrsõnade leksikoniga” (2016) võrreldes on sõnavalik väiksem, aga käsitlus põhjalikum. Esialgu sai oletatud, et kommenteerida jõuab umbes 20 000 sõna, ja nii see tegelikult kujuneski. Loodetavasti on võimalik tööd hiljem jätkata ja seda arvu suurendada.

Sõnade päritolu selgitus võib olla erinev. Enamasti on kõigepealt esile toodud päritolu liik ehk see, kas tegu on põlissõnaga („omasõnaga”, genuiitse sõnaga) või laensõnaga. Mis on laensõna, on seletamatagi enam-vähem selge; põlissõna mõiste alla on paigutatud vanemad sõnad, mille kohta pole teada, et nad oleksid laenatud, aga ka sellised meie sõnavara ürgseimasse kihti kuuluvad sõnad, millel on küll selge vaste mõnes teises keelkonnas (nt *mesi*, *muna*, *tooma*, *vask*, *vesi*), kuid mille laenulisus pole ikkagi ühemõtteliselt kindel. Esiteks pole sageli selge laenu suund (ehkki millegipärast väidetakse sellistel juhtudel alati, et sõna on laenatud indoeuroopa keeltest uurali resp. soome-ugri keeltesse) ja teiseks on olemas alternatiivne teooria, mis seletab nende sõnade sarnasust mitme keelkonna ürgse sugulusega. Põlissõnadel on enamasti olemas vaste vähemalt ühes sugulaskeeles, kuid mitte alati: näiteks sõnal *põis* see puudub, kuid käänamismalli järgi otsustades (*põis* : *põie* : *põit*) on tegemist ürgvana sõnaga. Laensõnade puhul on märgitud laenuallikas või laenuallikad, uuemate laenude puhul enamasti konkreetne keel (nt alamsaksa, saksa, vene, soome), vanemate puhul keelerühm (nt aarja, balti, germaani); kui lähtekeele

sõna on omakorda laenatud, on võimaluse korral püütud esitada kogu teekond, kuidas ta on mitmeid keeli läbides eesti keelde jõudnud. Põlissõnade puhul on esile toodud nende levik sugulaskeeltes ja esitatud maksimaalselt kolm näidet. Leviku iseloomustamisel on kasutatud kategooriaid *uurali sõna* (vaste vähemalt ühes samojeedi keeles), *soome-ugri sõna* (vaste vähemalt ühes uugri keeles), *läänemeresoome-permi sõna* (vaste vähemalt ühes permi keeles), *läänemeresoome-volga sõna* (vaste mari keeles), *läänemeresoome-mordva sõna* (vaste mordva keeltes), *läänemeresoome-saami sõna* (vaste saami keeltes), *läänemeresoome sõna* (levinud mitmes läänemeresoome keeles, kusjuures vaste on olemas nii vepsa kui ka liivi keeles), *piiratud levikuga läänemeresoome sõna* (levinud mitmes läänemeresoome keeles), *läänemeresoome lõunarühma sõna* (levinud eesti, liivi ja vadja keeles), *eesti-soome sõna*, *eesti-liivi sõna* ja *eesti-vadja sõna*. Tavaliselt kiputakse sõna geograafilist levikut pidama ka tema vanuse näitajaks, kuid päris sajarprotsendilise tõena seda võtta ei saa.

Mõned näited selle kohta, kuidas see kõik tegelikult välja näeb.

**aju** PÕLISSÕNA läänemeresoome-saami *v* soome-ugri: soome *aivo(t)* 'aju', saami *vuoiḡâš* 'aju', ?ungari *agy* 'aju, üdi'

**akvalang** LAENSÕNA vene *акваланг* 'akvalang' < inglise *aqualung* 'akvalang' (1950). Ladina sõnast *aqua* 'vesi' ja inglise sõnast *lung* 'kops'

**matma** PÕLISSÕNA läänemeresoome (lõunarühm): vadja *mattaa* 'kinni katta; (kõvasti) sulgeda', liivi *mattō* 'maasse kaevata, matta'. Tuletis sõnast \**manta* ~ \**mante* 'maa', mis pole eesti keeles iseseisvana säilinud

**mets** LAENSÕNA balti: leedu *medis* 'puu', läti *mežs* 'mets'

**misjon** LAENSÕNA saksa *Mission* 'misjon; missioon' < ladina *missio* 'saatmine, lähetamine' (sõnast *mittere* 'saatma, lähetama')

**taldrik** LAENSÕNA rootsi *tallrik* 'taldrik' < alamsaksa *talloreken* 'taldrikuke' (sõnast *tallör* 'taldrik')

**vesi** PÕLISSÕNA soome *vesi* 'vesi', ungari *víz* 'vesi', sõlkupi *üt* 'vesi'. Sama tüvi on levinud ka indoeuroopa keeltes (vrd kreeka *hydōr*, inglise *water*, läti *ūdens* jt)

Kolmanda sõnade päritolu liigina on sõnaraamatus välja toodud kõlasõnad. Nende alla on koondatud niihästi otsesed kõlajäljendused (onomatopoeetilised sõnad) kui ka kõlajäljenduslike sõnade ülekandeline kasutamine (nn onomatopoeetilis-deskriptiivsed sõnad). Kõlasõnad pakuvad etümoloogilises mõttes vähe, kuid nende sõnaraamatu tasemel käsitlemine on paras pätkel, sest tähenduse lahknemise korral pole selge, kas sugulaskeeltes on tegemist ühe ja sama iidse sõna vastetega või hiljem eri keeltes iseseisvalt tekkinud sõnadega. Niisugustel segastel juhtudel on sõnaraamatus piirdu- tud lihtsalt kõlasõna märgendiga ega ole püütud kaheldavaid sugulaskeelte vasteid esile tuua.

Üksnes mõnekümne sõnaga on esindatud nimi ja tootenimetus. Omaette liigina sunnib neid esitama asjaolu, et osal juhtudest, näiteks ravimite kaubanduslike nimetuste puhul, on raske kindlaks teha, kus ja millal need on moodustatud, osal juhtudest aga on selge, et vastavad sõnad pärinevadki otsest tootenimetusest, mida ei saa siduda konkreetse keelega. Näiteks 1950. aastate teisel poolel eesti keeles käibele tulnud sõna *ketsid* on peetud vene laenuks, mis aga jätab seletamata sõnas sisalduva *s*-hääliku (vrd vn *кеды*). Ilmselt on nii eesti kui ka vene sõna rööpselt üle võetud US Rubber Company kummitallaga spordijalatsite tootenimest *Keds*, kusjuures vene keeles on *s*-i peetud mitmuse tunnuseks ja sõnast välja jäetud.

Sõnaraamatus on püütud mõnevõrra märgendada ka tõlkelaene, s.o niisuguseid liitsõnu ja väljendeid, mis on eesti keeles loodud mõne teise keele liitsõna või väljendi tõlkena (nt sks *Plattfuß* > ee *lampjalg*, vn *летающая тарелка* > ee *lendav taldrik*, ingl *chain smoker* > ee *ahelsuitsetaja*), ja pooltõlkelaene, milles sõna või väljendi üks osa on laenatud, teine tõlgitud (nt sks *Planwagen* > ee *plaanvanker*, vrd sks *Plane* 'present, presentkatus' ja *Wagen* 'vanker'; sks *Bubikragen* > ee *bubikrae*, vrd sks *Bubi* 'poiss, poisu, poisike' ja *Kragen* 'krae'). Eesti keeles on tõlkelaene tohtul hulgal, aga uurimise seisukohalt on ala täiesti söötis ning mõistagi pole uues eesti keele sõnaraamatus neid vähegi ammendavalt esile toodud. Siingi tuleks tööd jätkata.

Nagu eelnevast näha, ei ole päritolu käsitlus pelgalt tüvekeskne nagu „Eesti etümoloogiasõnaraamatus”, vaid võimalust mööda on kaasa haaratud ka tuletisi, liitsõnu ja väljendeid. Suur osa selliseid keelendeid on eesti keeles üpris uued, kirjakeele arendamiseks teadlikult loodud. Tuletiste ja väliseskujuta liitsõnade niisuguseid juhtumeid katab märgend *kirjakeele sõna*, kusjuures vahel on võimalik esile tuua ka sõna autor ja loomisaeg. Sadakonna kunstlikult loodud tüvisõna juures on kasutatud märgendit *tehissõna*. Nende autor ja loomisaeg on enamasti teada.

## Kokkuvõtteks: miljon kasutajat kuus

Korpuseajastu (pool)automaatne sõnastike koostamine mõjutab väga tugevalt sõnastike tüübi teisenemist normatiivsest kirjeldavaks ehk preskriptiivsest deskriptiivseks. Enamik infot tuleb otse tekstikorpustest: märksõnastik, kollokatsioonid, näited, info muutevormide kohta, paralleelkorpusest tõlkevasted jne. Sloveenlased teatavadki näiteks, et nad täiendavad oma sõnaraamatut üksnes deskriptiivselt olemasolevate ja uute korpuste põhjal, ilma normatiivsete ettekirjutusteta (Gorjanc jt 2017: 159).

Meil on juba praegu mitu sõnaraamatut, kus tehakse miljon päringut kuus. Need on eesti keele seletav sõnaraamat (EKSS 2009) ja õigekeelsussõnaraamat ÕS 2013. Eesti-vene sõnaraamatut uuritakse lausa 48 000 korda päevas ehk umbes 1,4 miljonit korda kuus. Need arvud kasvavad aasta-aastalt. (Tõsi, osalt tegutsevad ka robotid.) Kuulsa Oxford English Dictionary veebilehel, mis oli (lisaks raamatule) veebis raha eest kasutatav umbes 10 aastat (aastamaks 240 £), tegid selle tellijad samuti rohkem kui miljon päringut kuus (Jamieson 2010).

2017. aasta elektroonilise leksikograafia konverentsil eLex 2017 toodi välja (Google'i andmete põhjal), mida inimesed sõnastikust kõige rohkem otsivad. Esikolmik on: 1) tähendused, 2) õigekiri ja 3) grammatika. Järgnevad: 4) sünonüümid, antonüümid jm sõnadevahelised suhted (ehk teaurus), 5) sõna päritolu, 6) homonüümid, 7) poolitus, silbitus jm info, 8) hääldus (transkriptsioonina).

Vojko Gorjanc jt (2017: 94) on omakorda uurinud, millest kasutajad sõnaraamatus kõige rohkem puudust tunnevad. Tuleb välja, et kõigest: kirjanikud soovivad rohkem erilisi, näiteks vähekasutatud sõnu, ajakirjanikud rohkem uusi sõnu, reklaamikirjutajad rohkem ja selgemaid tähenduste seletusi, teadlased rohkem oskussõnu ja seletusi.

Sõnaraamatutöö ei lõpe kunagi.

## Kirjandus

- Dictionaria. Toim Martin Haspelmath, Barbara Stiebels. Jena: Max Planck Institute for the Science of Human History. <https://dictionaria.cld.org/> (25. IX 2018).
- Eesti etimoloogiasõnaraamat. Koost ja toim Iris Metsmägi, Meeli Sedrik, Sven-Erik Soosaar. Peatoim I. Metsmägi. Tallinn: Eesti Keele Sihtasutus, 2012.
- Eesti keele ühendkorpus 2017. Eesti Keele Instituut ja Lexical Computing Ltd. Korpus on lemmatiseeritud, märgendatud ja ühestatud analüsaatori EstNLTK abil. <https://doi.org/10.1515/3-00-0000-0000-071e71> (25. IX 2018).
- EKS = Eesti keele süntaks. (Eesti keele varamu III.) Toim Mati Ereht, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus.
- EKSS 1988–2007 = Eesti kirjakeele seletussõnaraamat (vihikud 1–26). Peatoim Rudolf Karelson, Valve Kullus (Põlma), Erich Raiet, Mai Tiits, Tiia Valdre, Leidi Veskis. Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus.
- EKSS 2009 = Eesti keele seletav sõnaraamat I–VI. „Eesti kirjakeele seletussõnaraamatu” 2., täiendatud ja parandatud trükk. Toim Margit Langemets, Mai Tiits, Tiia Valdre, Leidi Veskis, Ülle Viks, Piret Voll. Tallinn: Eesti Keele Sihtasutus. <http://www.eki.ee/dict/ekss/> (25. IX 2018).
- ELEXIS Survey 2018 = Projekti „European Lexicographic Infrastructure” (Horizon 2020) raames korraldatud küsitluse esialgsed andmed. (Eesti Keele Instituut on projekti üks partnereid.) <http://elex.is> (25. IX 2018).
- Gorjanc, Vojko, Gantar, Polona, Kossem, Iztok, Krek, Simon (toim) 2017. Dictionary of Modern Slovene: Problems and Solutions. Ljubljana: University of Ljubljana, Faculty of Arts. [http://www.ff.uni-lj.si/sites/default/files/Dokument/ Knjige/e-books/dictionary\\_of\\_modern\\_slo.pdf](http://www.ff.uni-lj.si/sites/default/files/Dokument/ Knjige/e-books/dictionary_of_modern_slo.pdf) (25. IX 2018).
- Habicht, Külli, Penjam, Pille 2007. Kaassõna keeleuurija ja -kasutaja käsitus. – Emakeele Seltsi aastaraamat 52 (2006). Tallinn: Teaduste Akadeemia Kirjastus, lk 51–68.
- Haspelmath, Martin 2017. Dictionaria: Farewell to linear dictionaries. – Diversity Linguistics Comment blog. <https://dlc.hypotheses.org/971> (25. IX 2018).
- Jamieson, Alastair 2010. Oxford English Dictionary ‘will not be printed again’. – The Telegraph, 29. VIII. <https://www.telegraph.co.uk/culture/books/book-news/7970391/Oxford-English-Dictionary-will-not-be-printed-again.html> (25. IX 2018).

- Kaalep, Heiki-Jaan 2018. Statistika koht keelemudelis. – Keel ja Kirjandus, nr 8–9, lk 713–727.
- Kalda, Kaupo 2009. Mis on Web 2.0. <https://okia.ee/mis-on-web-20/> (25. IX 2018).
- Kallas, Jelena 2013. Eesti keele sisusõnade süntagmaatilised suhted korpus- ja õppeleksikograafias. (Tallinna Ülikooli humanitaarteaduste dissertatsioonid 32.) Tallinn: Tallinna Ülikool. <http://e-ait.tlulib.ee/id/eprint/303> (25. IX 2018).
- Kallas, Jelena, Koppel, Kristina, Tuulik, Maria 2015. Korpusleksikograafia uued võimalused eesti keele kollokatsioonisõnastiku näitel. – Eesti Rakenduslingvistika Ühingu aastaraamat, nr 11, lk 75–94.
- Kipfer, Barbara Ann 2013. 9 reasons why print dictionaries are better than online dictionaries. – The Week, 3 VII. <http://theweek.com/articles/462575/9-reasons-why-print-dictionaries-are-better-than-online-dictionaries> (25. IX 2018).
- Koppel, Kristina 2017. Heade näitelausete automaattuvastamine eesti keele õppesõnastike jaoks. – Eesti Rakenduslingvistika Ühingu aastaraamat, nr 13, lk 53–71.
- Koppel, Kristina 2019 (ilmumas). Eesti keele A-, B- ja C-keeleoskustaseme õpikulausete parameetrid ja nende arvestamine näitelausete automaatsel valikul. – Eesti Rakenduslingvistika Ühingu aastaraamat, nr 15.
- Langemets, Margit, Loopmann, Andres, Viks, Ülle 2006. The IEL dictionary management system of Estonian. – DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writing Systems. Pre-EURALEX workshop: Fourth International Workshop on Dictionary Writing System. Turin, 5th September 2006. Toim Gilles-Maurice de Schryver. Turin: University of Turin, lk 11–16.
- Langemets, Margit, Tiits, Mai, Valdre, Tiia, Voll, Piret 2010. *In spe*: üheköiteline eesti keele sõnaraamat. – Keel ja Kirjandus, nr 11, lk 793–810.
- Langemets, Margit 2014. Sõnaraamat pärast Gutenbergi ehk mõeldes väljaspool paberit. – Sirp 13. III 2014.
- Metslang, Helle 2002. Grammatikalisatsiooniteooria eesti keele taustal. – Teoreetiline keeleteadus Eestis. (Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 4.) Tartu: Tartu Ülikooli Kirjastus, lk 164–178.
- Rundell, Michael 2002. No more print dictionaries: a ‘sad day’ or a ‘day of liberation’? – Macmillan Dictionary blog. <http://www.macmillandictionaryblog.com/no-more-print-dictionaries> (25. IX 2018).
- Rundell, Michael 2009. The future has arrived: A new era in electronic dictionaries. – MED Magazine, nr 54. <http://www.macmillandictionaries.com/MED-Magazine/August2009/54-Feature-Electronic-Dictionaries.htm> (25. IX 2018).
- Rundell, Michael 2014. Macmillan English Dictionary: The End of Print? – Slovenščina 2.0, kd 2, nr 2, lk 1–14. <https://www.dlib.si/stream/URN:NBN:SI:DOC-ZQZQ8CMC/4a002dbb-ec00-45f1-a3f0-31b39bdc8df/PDF> (25. IX 2018).
- Sahkai, Heete 2008. Konstruksioonipõhine keelemudel ja sõnaraamatud. – Eesti Rakenduslingvistika Ühingu aastaraamat, nr 4, lk 171–185.
- Żmigrodzki, Piotr 2014. Polish Academy of Sciences Great Dictionary of Polish. [Wielki słownik języka polskiego PAN.] – Slovenščina 2.0, kd 2, nr 2, lk 37–52. [http://slovenscina2.0.trojina.si/arhiv/2014/2/Slo2.0\\_2014\\_2\\_04.pdf](http://slovenscina2.0.trojina.si/arhiv/2014/2/Slo2.0_2014_2_04.pdf) (25. IX 2018).
- Tavast, Arvi, Langemets, Margit, Kallas, Jelena, Koppel, Kristina 2018. Unified data modelling for presenting lexical data: The Case of EKILEX. – Proceedings of the XVIII EURALEX International Congress. EURALEX:

Lexicography in Global Contexts, Ljubljana, 17–21 July 2018. Toim Jaka Čibej, Vojko Gorjanc, Iztok Kosem, Simon Krek. Ljubljana: Ljubljana University Press, Faculty of Arts, lk 749–761.

Tiberius, Carole, Schoonheim, Tanneke 2016. The Algemeen Nederlands Woordenboek (ANW) and its lexicographical process. – Der lexikografische Prozess bei Internetwörterbüchern. (OPAL 1/2016). Toim Vera Hildenbrandt. Mannheim: Institut für Deutsche Sprache, lk 20–28.

Velsker, Eva 2010. Leksikaliseerumine keelemuutuses. – Keel ja Kirjandus, nr 7, lk 521–536.

Velsker, Eva 2013. Aeg eesti murretes. – Emakeele Seltsi aastaraamat 58 (2012). Tallinn: Teaduste Akadeemia Kirjastus, lk 265–295.

Vääri, Eduard, Kleis, Richard, Silvet, Johannes 2016. Võõrsõnade leksikon. Peatoim Tiina Paet. Tallinn: Eesti Keele Instituut, Kirjastus Valgus, 2016.

*Margit Langemets (sünd 1961), PhD, Eesti Keele Instituut, sõnaraamatute peatoimetaja, margit.langemets@eki.ee*

*Mai Tiits (sünd 1954), Eesti Keele Instituut, vanemleksikograaf, mai.tiits@eki.ee*

*Udo Uibo (sünd 1956), Eesti Keele Instituut, vanemleksikograaf, udo.uibo@eki.ee*

*Tiia Valdre (sünd 1956), Eesti Keele Instituut, vanemleksikograaf, tiia.valdre@eki.ee*

*Piret Voll (sünd 1967), MA, Eesti Keele Instituut, vanemleksikograaf, piret.voll@eki.ee*

## **Estonian lexis revisited: The Dictionary of Estonian 2018**

Keywords: electronic lexicography, online publishing, monolingual dictionary, new words, grammatization and lexicalization, usage examples, etymological information

The article discusses the circumstances of the compilation and publication of a new explanatory Dictionary of Estonian (2018). Compared to the early days of its compilation in 2010 European lexicography has undergone several major changes, such as electronic vs paper dictionaries, empirical vs introspective approach, semi-automatic corpus-based vs manual compilation, integration of various data sources, while online dictionaries have made their way to mobile apps. The semi-automatic compilation launched by the advance of the corpus era has made a strong impact on the type of dictionaries, which has changed from normative or prescriptive to descriptive. The major challenge facing modern lexicography is how to update existing information or to easily reuse it.

In addition, some substantive aspects are addressed, including new words and word senses, grammatization and lexicalization, sentence examples and etymological information. The new Dictionary of Estonian includes about 17 000 new entries and 1300 new senses. The decisions whether a word with a specific meaning should get an entry of its own or be presented as a subentry of a semantic unit have been made using the Sketch Engine corpus tool. Sentence examples invented by lexicog-



raphers are supplemented by ones picked automatically from the corpus. Also, the presentation of etymological information in the new dictionary is described.

*Margit Langemets (b. 1961), PhD, Institute of the Estonian Language, Executive Editor of Dictionaries, margit.langemets@eki.ee*

*Mai Tiits (b. 1954), Institute of the Estonian Language, Senior Lexicographer, mai.tiits@eki.ee*

*Udo Uiibo (b. 1956), Institute of the Estonian Language, Senior Lexicographer, udo.uiibo@eki.ee*

*Tiia Valdre (b. 1956), Institute of the Estonian Language, Senior Lexicographer, tiia.valdre@eki.ee*

*Piret Voll (b. 1967), MA, Institute of the Estonian Language, Senior Lexicographer, piret.voll@eki.ee*