

LATVIEŠU-IGAUNU VĀRDNĪCA

—
Igauņu valodas institūts

—
Projekta vadītājs
Arvi Tavasts

—
Vārdnīcas darba grupa
Merle Madisone
Aive Mandele
Tauno Neuliks
Anita Tannenberga

—
Igauņu valodas fonds
Tallina
2015

LÄTI-EESTI SÕNARAAMAT

Eesti Keele Instituut

Projektijuht
Arvi Tavast

Sõnaraamatu töörihm
Merle Madisson
Aive Mandel
Tauno Nõulik
Anita Tannenber

Eesti Keele Sihtasutus
Tallinn
2015

Projekti konsultant

Margit Langemets

Koanimede konsultant

Peeter Päll

Koostööpartner

Läti Keele Agentuur

Kaanekujundus

Anna Kaarma

Kujundus

Merle Moorlat

Andmestruktuur, koostamispõhimõtted ja automaatne küljendus

Arvi Tavast

Koostamisel on kasutatud terminibaasisüsteemi Termeki term.eki.ee

Sõnaraamatu koostamisele aitas kaasa tõlke- ja keeletehnoloogiafirma Tilde www.tilde.com

Sõnaraamatu väljaandmist on toetanud Eesti-Läti programm
“Development of Estonian-Latvian and Latvian-Estonian dictionary, 2012–2015”.

Tänu Jaak Villerile

Projekta konsultante

Margita Langemetsa

Vietvärdude konsultants

Pēters Pells

Sadarbības partneris

Latviešu valodas aģentūra

Vāka dizains

Anna Kārma

Dizains

Merle Morlata

Datu struktūra, sastādīšanas principi un automātiskā maketēšana

Arvi Tavasts

Vārdnīcas sastādīšanā tika izmantota terminu bāzu sistēma Termeki term.eki.ee

Vārdnīcas sastādīšanā palīdzēja tulkošanas un valodas tehnoloģiju firma „Tilde” www.tilde.com

Vārdnīcas sastādīšanu ir atbalstījusi Igaunijas-Latvijas programma
“Development of Estonian-Latvian and Latvian-Estonian dictionary, 2012–2015”.

Pateicamies Jākam Villeram



Eesti Keele Instituut

© Eesti Keele Instituut, 2015 / © Igaunü valodas institüts, 2015
© Eesti Keele Sihtasutus (väljaandja), 2015 / © Igaunü valodas fonds (izdevums), 2015

Trükk AS Pakett / Druka AS Pakett

ISBN 978-9985-79-627-6

SISUKORD

Eessõna	7
Lāti-eesti sõnaraamatu koostamisest	7
Märksõnad	8
Lāti tähestik	9
Sõnastik	13

SATURA RÄDITÄJS

Priekšvārds	10
Par Latviešu–igauņu vārdnīcas sastādīšanu	10
Šķirkļi	11
Latviešu alfabēts	12
Vārdnīca	13

EESSÕNA

Juba ligikaudu pool sajandit on eestlastele ja lätlastele, kes on soovinud naaberrahva keelt õppida, mõista ja kasutada, olnud pea ainsaks abivahendiks Karl Abeni läti-eesti ja eesti-läti sõnastikud. Aastakümnete jooksul on aga mõlemas ühiskonnas ja keeles paljugi muutunud ja uuenenud, seega on aeg uute ja märksa mahukamate sõnaraamatute ilmutamiseks ammu küps.

Kuigi lävimine naaberrahvaste vahel on eri elualadel tihe, on omavahelises suhtluses sageli harjutud kasutama pigem suuremate võõrkeelte vahendust. Selle keelebarjääri murendamiseks saigi kahe riigi koostöös ning Eesti-Läti programmi toetusel ellu kutsutud projekt „Eesti-läti ja läti-eesti sõnaraamatu koostamine“. Töö tulemusena valmis Eesti Keele Instituudis (EKI) ja Läti Keele Agentuuris (*Latviešu valodas aģentūra*) samaaegselt kaks sõnaraamatut, kummaski rohkem kui 40 000 märksõna. Mõlemad sõnaraamatud on kättesaadavad ka veebis, kus neid edaspidi võimaluse korral täiendatakse.

LÄTI-EESTI SÕNARAAMATU KOOSTAMISEST

Uus läti-eesti sõnaraamat erineb koostamismeetodi poolest mõneti teistest üldkeele sõnaraamatutest. Enne suurte keelekorpuste laiemat kättesaadavust alustati sõnastike koostamist n-ö tühjalt lehelt, kuhu koostajad hakkasid lisama märksõnu oma teadmiste ning kättesaadavate materjalide põhjal. Tänapäevased suured keelekorpused võimaldavad aga sõnaraamatu koostamist alustada hoopis teisest otsast, võttes aluseks automaatselt koostatud mustandi, mis juba sisaldab reaalsest keelekasutusest leitud andmeid iga sõna kohta ning mille kvaliteeti koostajad seejärel parandama asuvad. Erinevalt traditsioonilisest meetodist seisneb see töö ennekõike korpusest leitud ebatäpsete või liigsete vastete parandamises ja kõrvaldamises, mitte uute vastete otsimises ja lisamises.

Koostöös tõlke- ja keeletehnoloogiafirmaga Tilde täiendati kõigepealt nende olemasolevat rööpkorpust – tekstikogu, mis sisaldab tekste ja nende tõlkeid. Kuna otse läti keelest eesti keelde või vastupidi tõlgitud tekstide maht pole väga suur, lisati korpusele ka kolmanda keele kaudu leitud vastavused (näiteks olukordades, kui üks ja sama ingliskeelne tekst oli tõlgitud nii eesti kui ka läti keelde). Korpuse kogumahuks kujunes üle 120 miljoni sõna kummaski keeles. Tavaliselt on rööpkorpused omavahel vastavusse viidud (joondatud) dokumendi- või

lausetasandil; selle sõnastiku tarvis joondas Tilde korpuse ka sõnatasemel. Tulemuseks saadi põhjalik vastekandidaatide loetelu: 51 000 läti ning 307 000 eesti sõna ja väljendit, mille hulgas oli korpuse tekstide kaasaegsuse ja mitmekesisuse tõttu palju uuemat ning varasemates sõnaraamatustest kajastamata sõnavara. Mahult oli sõnastik sellega valmis, aga tegelik koostamistöö alles ees.

Tööks kasutati sõnastikusüsteemi Termeki (term.eki.ee), mida EKI pakub tasuta ka teiste eesti keelt sisaldavate sõnastike koostamiseks. Kõigepealt tuli üle kontrollida automaatse joondamise tulemused ja parandada võimalikud vead. Termeki baasstruktuur on mõistepõhine, st ühes andmebaasikirjes esitatakse üks mõiste (täendus) koos kõigi selle tähistamiseks kasutatavate sõnadega mõlemas keeles, mitte üks sõna ja selle tähendused, nagu sõnapõhistes sõnaraamatutes tavaks. Meile teadaolevalt ei ole suurtes üldkeelesõnastikes niisugust baasstruktuuri varem kasutatud.

Mõistepõhisuse eesmärk on sõnastiku sisemise ühtluse parandamine. Näiteks kui sõna A on sõna B sünonüüm, siis on B kindlasti ka A sünonüüm ja nende artiklid on omavahel kooskõlas, või kui C viitab D-le, siis tagab mõistepõhine koostamine, et D on ka kindlasti sõnastikus olemas.

Sünonüümid on ühes andmebaasikirjes ja koostajal korraga silme ees, viited nende vahele luuakse vajadusel automaatselt. Sellega välistab mõistepõhine koostamismeetod sünonüümi-vastuolud ja rippuvad viited, mis võiksid kasutaja eksikombel juhatada sõnadeni, mida sõnaraamatatus tegelikult märksõnana pole.

Järgmine etapp oligi sünonüümsete kirjete ühendamine. Termeki keskkond lihtsustas seda protsessi, pakkudes sarnaseid sõnu sisaldavate mõistete seast ühendamiskandidaate. Siiski polnud tegu lihtsa ülesandega. Absoluutseid sünonüüme ega tõlkevasteid pole teatavasti olemas, ikka on ka kõige täpsemate sünonüümidenäivate sõnadel mõni erinev nüans. Igasugune sünonüümide või vastete esitamine sõnastikus tähendab seetõttu materjali (kohati ulatuslikku) lihtsustamist. Mil määral täpselt lihtsustada või kuhu tõmmata samatähenduslikkuse piir, oleneb sõnastiku mahust ja eesmärgist ning on igal konkreetsel juhul koostaja otsustada. Loodame, et meie selles sõnastikus tehtud samatähenduslikkuse-otsused nii ühe keele sees kui ka keelte vahel on lugejale informatiivsed ning et suutsime leida sünonüümide ühendamise ja tähendusnüansside lahushoidmise vahel optimaalse tasakaalu. Samuti loodame, et kasutajale on abiks ühes märksõnaartiklis nii eesti kui ka läti sünonüümide nägemine.

Paberkujul avaldamiseks tuli materjal teisendada harjumuspärasele sõnapõhisele kujule. Selle käigus tegime ühtlasi ära ka küljenduse, väljastades Termeki baasist $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ i lähtekoodi ja kompileerides selle $\text{X}_{\text{E}}\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ iga. Ka see on meie teada esimene kord Eestis, kui õnnestub suur üldkeelesõnastik andmebaasist otse trükki saata, ilma et see vajaks käsitsi küljendamist.

MÄRKSSÕNAD

Sõnaraamatus on üle 40 000 läti märksõna, mille valikul on lähtunud põhimõttest, et kajastatud sõnavara oleks päevakohane ning et lisaks põhisõnavarale oleks esindatud ka sagedamini kasutatavad erialasõnad, kõnekeel ning släng. Praktiliselt kõik märksõnad on pärit eelmainitud rööptekstide korpuse põhjal automaatselt tekitatud loendist. Korpuse tekstid kajastasid erinevaid valdkondi, muuhulgas näiteks turismi, IT-d, teadust, ilukirjandust ja Euroopa Liidu temaatikat.

Sõnaraamat sisaldab ka olulisemaid ning tuntumaid Eesti, Läti ning muu maailma kohanimedid, kuna läti nimekirjutusreeglid erinevad eesti keele omadest oluliselt. Nimede valikul ja vastete leidmisel kasutati EKI kohanimede loendit, mida kohati parandati ja täiendati. Kohanimed pole esitatud eraldi loendina, vaid üldise sõnaloendi sees.

Tegijanimede ja ametite puhul on üksikute eranditega esitatud ainult meessoost variant, samamoodi on toimitud ka omadussõnadega.

Käändumatud ja genitiivi vormis esinevad ning omadussõna funktsioonis kasutatavad nimi-sõnad pole eraldi märgistatud.

Sõnaraamatu märksõnad on esitatud tähestikulises järjekorras, esineb ka mitmesõnalisi märksõnu. Homonüümsed ning samakujulised, kuid erinevasse sõnaliiki kuuluvad märksõnad on tähistatud punktiga araabia numbritega (nt *mērs 1.*, *mērs 2.* või *absurds 1.*, *absurds 2.*). Polüseemid on esitatud araabia numbritega 1), 2), 3) jne. Märksõnadele pole üldiselt grammatilist teavet lisatud, v.a VI käändkonna naissoost sõnad, mis on tähistatud lühendiga *f* (feminiin), ning üldsoost (läti *k kopdzimte*) sõnad märgendiga *c* (*substantiva communia*).

Kui märksõna esineb ainult mitmuses või harilikult mitmuses, siis järgneb märksõnale sellekohane märgend (*mitmuses* või *hrl mitmuses*). Võimalikule mitmusemärgendile võib järgneda valdkonna- või stiilmärgend. Stiilmärgendid (nt *kõnekeelne*, *poetiline*) on esitatud stiililiselt markeeritud sõnadel ning valdkonnamärgendit (nt *keel*, *bioloogia*) on kasutatud vajaduse korral märksõna sisu või kasutusvaldkonna täpsustamiseks.

Läti sünonüümid koos nende kohta käivate stiilmärgenditega on ära toodud sama sõnaartikli sees sulgudes.

LÄTI TÄHESTIK

Aa, Ää, Bb, Cc, Čč, Dd, Ee, Ēē, Ff, Gg, Ģģ, Hh, Ii, Īī, Jj, Kk, Ķķ, Ll, Ļļ, Mm, Nn, Ņņ, Oo, Pp, Rr, Ss, Šš, Tt, Uu, Ūū, Vv, Zz, Žž

Täname kõiki, kes on olnud abiks läti-eesti sõnaraamatu koostamisel ja väljaandmisel. Täname ka oma Läti kolleege, kes koostasid eesti-läti sõnaraamatu.

Loodame, et sõnaraamatu kasutajad leiavad siinsest palju vajalikku ja head!

Sõnaraamatu tööriühm

PRIEKŠVārds

Jau apmēram pusgadsimtu igauņiem un latviešiem, kas vēlējušies mācīties, saprast un lietot kaimiņtautas valodu, gandrīz vai vienīgais palīg līdzeklis bijušas Karla Abena Latviešu–igauņu un Igaņu–latviešu vārdnīcas. Gadu desmitos abu tautu sadzīvē un valodā daudz kas ir mainījies un parādījies no jauna, tāpēc jau sen nobriedusi nepieciešamība pēc aktuālākām un ievērojami apjomīgākām vārdnīcām.

Kaut gan sakari starp kaimiņtautām dažādās dzīves jomās ir intensīvi, komunikācijā kā starpniekvalodas ierasts izmantot lielās svešvalodas. Lai pārvarētu šo valodas barjeru, pateicoties abu valstu sadarbībai un Igaunijas–Latvijas programmas atbalstam, dzīvē tika ieviests projekts „Igaņu-latviešu un Latviešu-igauņu vārdnīcas izstrāde”. Tā rezultātā Igaņu valodas institūtā (*Eesti Keele Instituut*) un Latviešu valodas aģentūrā vienlaicīgi tika sagatavotas divas vārdnīcas, katrā no tām ir vairāk nekā 40 000 šķirķļu. Abas vārdnīcas ir pieejamas arī tīmeklī, kur tās iespēju robežās turpmāk tiks papildinātas.

PAR LATVIEŠU–IGAUNU VĀRDNĪCAS SASTĀDĪŠĀNU

Jaunā Latviešu–igauņu vārdnīca no citām vispārlietojamās leksikas vārdnīcām atšķiras ar sastādīšanas metodi. Pirms kļuva pieejami lielie valodas korpusi, vārdu krājuma veidošana tika sākota no baltas lapas. Vārdnīcas veidotāji savāca šķirķļus, balstoties uz savām zināšanām, kā arī pieejamajiem materiāliem. Mūsdienu lielie valodas korpusi pieļauj vārdnīcas sastādīšanu sākt no otra gala, par pamatu ņemot automātiski veidoto melnrakstu, kurš jau ietver datus par katra vārda reālo pielietojumu un kura kvalitāti pēc tam sastādītāji uzlabo. Atšķirībā no tradicionālās metodes, šis darbs vispirms nozīmē korpusā atrasto neprecīzo vai aptuveno atbilstību labošanu un atmešanu, nevis jaunu atbilstmju meklēšanu un pievienošanu.

Sadarbībā ar tulkošanas un valodas tehnoloģiju firmu *Tilde* sākumā tika papildināts jau esošais paralēlkorpus – tekstu krājums, kas sastāv no tekstiem un to tulkojumiem. Tā kā no latviešu valodas igauņu valodā vai otrādi tiešo tulkoto tekstu apjoms nav pietiekami liels, korpusam tika pievienotas arī caur trešo valodu atrastās atbilstības (piemēram, gadījumos, kad viens un tas pats angļu valodas teksts bija tulkots gan igauņu, gan latviešu valodā). Korpusa kopējais

apjoms sasniedza vairāk nekā 120 miljonu vārdu katrā no valodām. Parasti paralēlkorpusi savstarpēji ir atbilstoši pieskaņoti (pielīdzināti) dokumenta vai teikuma līmenī, šīs vārdnīcas gadījumā firmai *Tilde* korpusu nācās pieskaņot arī vārda līmenī. Rezultātā tika iegūts pamatīgs atbilstīgu kandidātu uzskaitījums: 51 000 latviešu un 307 000 igauņu vārdu un izteicienu, turklāt korpusa tekstu mūsdienīguma un daudzveidības dēļ tas bija jaunāks un agrāko laiku vārdnīcās neatspoguļots vārdu krājums. Pēc apjoma vārdu krājums bija savākts, taču reālais vārdnīcas sastādīšanas darbs tikai sākās.

Darbā tika izmantota vārdnīcu sistēma *Termeki* (term.eki.ee), kuru Igaunņu valodas institūts bez maksas piedāvā arī citu ar igauņu valodu saistīto vārdnīcu sastādīšanai. Vispirms bija jāpārbauda automātiskās pielīdzināšanas rezultāti un jāizlabo iespējamās kļūdas. *Termeki* bāzes struktūra balstās uz jēdzienisko principu, t.i., vienā datu bāzes ierakstā tiek uzrādīts viens jēdziens (nozīme) kopā ar visiem tā apzīmēšanai lietojamajiem vārdiem abās valodās, nevis viens vārds un tā nozīmes, kā tas ir parastajās vārdnīcās. Cik zināms, vispārlietojamās leksikas vārdnīcās šāda bāzes struktūra iepriekš nav tikusi izmantota.

Jēdzieniskā principa mērķis ir vārdu krājuma iekšējās vienmērības labošana. Piemēram, ja vārds A ir vārda B sinonīms, tad B noteikti ir arī A sinonīms un to šķirkļi savstarpēji ir saskaņoti, vai, ja C norāda uz D, tad jēdzieniskā sastādīšana nodrošina arī D esamību vārdnīcā.

Sinonīmi atrodas vienā datu bāzes ierakstā, un sastādītājam tie tūlīt ir acu priekšā, nepieciešamības gadījumā norādes starp tiem veidojas automātiski. Līdz ar to jēdzieniskā principa sastādīšanas metode izslēdz sinonīmu neatbilstības un norādes, kuras maldu gadījumā lietotāju varētu aizvest līdz vārdiem, kuri kā šķirkļi vārdnīcā nav atrodami.

Nākamais etaps bija sinonīmu ierakstu apvienošana. *Termeki* vide vienkāršoja šo procesu, piedāvājot savienojumu kandidātus no līdzīgiem vārdiem un jēdzieniem. Tomēr tas nebija vienkāršs uzdevums. Absolūtu sinonīmu un tulkojuma atbilstību, kā zināms, nav, vienmēr šķietami pat visprecīzākajiem sinonīmiem ir kāda atšķirīga nianse. Dažādu sinonīmu vai atbilstību pastāvēšana vārdu krājumā nozīmē materiāla (vietām plašu) vienkāršošanu. Kādā mērā precīzi vienkāršot vai kur novilkt vienādas nozīmības robežu, ir atkarīgs no vārdnīcas apjoma un mērķa, kā arī katrā konkrētajā gadījumā no sastādītāja lēmuma. Ceram, ka šajā vārdnīcā mūsu veiktie vienādas nozīmības risinājumi gan vienā valodā, gan starp abām valodām lasītājam būs pietiekami informatīvi un mēs esam spējuši atrast optimālu līdzsvaru starp sinonīmu apvienojumu un nozīmju nianšu atšķirtību. Tāpat ceram, ka lietotājam būs noderīgi vienā šķirklī saskaņot gan igauņu, gan arī latviešu sinonīmus.

Lai izdotu vārdnīcu papīra formātā, materiālu nācās modificēt uz ierasto vārda principu. Tā gaitā mēs vienlaikus veicām arī drukas aplauzumu, izsniedzot *Termeki* bāzes L^AT_EX_i pirmkodu un kompilējot to ar X_YL^AT_EX_i. Cik zināms, šī Igaunijā ir pirmā reize, kad lielā vispārlietojamās leksikas vārdnīca bez papildu aplauzuma veikšanas no datu bāzes tiešā veidā nonāk līdz drukai.

ŠĶIRKĻI

Vārdnīcā ir vairāk nekā 40 000 latviešu šķirkļu, kuru izvēle pamatojas uz principu, lai atspoguļotais vārdu krājums būtu ikdienā aktuāls un krājumā būtu pārstāvēti biežāk lietojamie speciālie termini, sarunvaloda un slengs. Praktiski visi šķirkļi ņemti no iepriekšminētā paralēltekstu

korpusa automātiski izveidotā saraksta. Korpusa teksti atspoguļo dažādas nozares, piemēram, tūrismu, IT, zinātni, daiļliteratūru un Eiropas Savienības tematiku.

Vārdnīca satur arī svarīgākos un zināmākos Igaunijas, Latvijas un pasaules vietvārdus, jo latviešu un igauņu vārdu rakstības likumi būtiski atšķiras. Vārdu izvēlei un atbilstmju atrašanai tika izmantots Igaņu valodas institūta vietvārdu saraksts, kas vietumis tika labots un papildināts. Vietvārdi vārdnīcā tiek uzrādīti nevis atsevišķā, bet vispārējā vārdu sarakstā.

Kas attiecas uz darītājevārdiem un amatu nosaukumiem, ar atsevišķiem izņēmumiem ir uzrādīti tikai vīriešu dzimtes varianti, līdzīgi ir arī ar īpašības vārdiem.

Nelokāmie un ģenitīva formā esošie, kā arī īpašības vārda funkcijā lietojamie lietvārdi nav atsevišķi atzīmēti.

Vārdnīcas šķirkļi ir uzrādīti alfabētiskā secībā, tur atrodas arī daudzvārdu šķirkļi. Homonīmi un vienādi rakstāmie, bet atšķirīgai vārdšķirai piederošie šķirkļi ir atzīmēti ar punktu un arābu ciparu (piemēram, *mērs 1.*, *mērs 2.* vai *absurds 1.*, *absurds 2.*). Polisēmas ir atzīmētas ar arābu ciparu 1), 2), 3) utt. Šķirkļiem nav pievienota gramatiskā norāde, izņemot VI deklinācijas sieviešu dzimtes lietvārdus, kas ir apzīmēti ar saīsinājumu *f* (femīns), kā arī kopdzimtes vārdus, kas atzīmēti ar *c* (*substantīva communia*).

Ja šķirklis ir tikai daudzskaitlī vai parasti daudzskaitlī, tad šķirklim pievienots atbilstošs apzīmējums (*mitmuses* vai *hrl mitmuses* – daudzskaitlī vai parasti daudzskaitlī). Iespējamam daudzskaitļa apzīmējumam var sekot arī jomas vai stila apzīmējums. Stila apzīmējumi (piemēram, *kõnekeelne* (sarunvaloda), *poetiline* (poētisks)) ir lietoti stilistiski marķētajiem vārdiem, bet jomu apzīmējumi (piemēram, *keel* (valoda), *bioloogia* (bioloģija)) vajadzības gadījumā ir lietoti šķirkļa satura vai minētās jomas precizēšanai.

Latviešu sinonīmi kopā ar stila apzīmējumiem, kas attiecas uz tiem, ir norādīti iekavās pie tā paša šķirkļa.

LATVIEŠU ALFABĒTS

Aa, Āā, Bb, Cc, Čč, Dd, Ee, Ēē, Ff, Gg, Ģģ, Hh, Ii, Īī, Jj, Kk, Ķķ, Ll, Ļļ, Mm, Nn, Ņņ, Oo, Pp, Rr, Ss, Šš, Tt, Uu, Ūū, Vv, Zz, Žž

Pateicamies visiem, kas palīdzējuši sastādīt un izdot Latviešu–igauņu vārdnīcu. Pateicība arī Latvijas kolēģiem, kas sastādījuši Igaņu–latviešu vārdnīcu.

Ceram, ka vārdnīcas lietotāji te atradīs daudz vajadzīga un derīga!

Vārdnīcas darba grupa

Tulkojis Guntars Godiņš