

Tartu Ülikool
Maailma keelte ja kultuuride instituut

KORPUSTE RAKENDAMINE TERMINILOOMES KAUGSEIRE TERMINIBAASI
NÄITEL

Magistritöö

Liisi Jakobson

Juhendajad:

Sirje Kupp-Sazonov, PhD

Jelena Kallas, PhD

Tartu 2023

Sisukord

Kesksed mõisted ja lühendid	3
Sissejuhatus.....	4
1. Kaugseire	6
1.1. Lühike ülevaade kaugseirest ja selle ajaloost.....	6
1.2. Kaugseire oskuskeele areng Eestis.....	7
2. Oskusleksikograafia.....	9
2.1. Korpused oskusleksikograafias	9
2.2. Terminitöö Eestis	10
3. Andmed ja meetodika.....	14
4. Tulemused ja analüüs	18
Kokkuvõte.....	21
Kirjanduse loetelu	22
Resümee	26
Lisad.....	28
<i>Lisa 1</i>	28
Kaugseire terminite loend	28
<i>Lisa 2</i>	42
Estonian Remote Sensing Termbase: data sources and implementation	42

Kesksed mõisted ja lühendid

definiitsioon ehk **määratlus** – mõiste esitus kirjeldava lausungiga, mille abil teda eristatakse sugulasmõistetest

GDEXi eesti keele moodul – eesti keele jaoks loodud GDEXi (Good Dictionary Examples) konfiguratsioonid, mis arvestavad keelespetsiifilisi parameetreid

kaugseire – eemal asuvate objektide kohta informatsiooni hankimine mittekontaktsete meetoditega

korpus – suur elektrooniline tekstikogu, mis on reeglina otsitav korpus-päringusüsteemi kaudu

fookuskorpus – korpus, millest soovitakse võtmesõnu või mitmesõnalisi termineid leida

referentskorpus – võimalikult suur üldkorpus vastavas keeles

korpuspäringusüsteem – tarkvara, mis võimaldab korpuse mitmekülgset analüüsi

kroolimine – veebilehtede süstemaatiline sirvimine ja tekstide kogumine spetsiaalse veebiroboti abil

mõiste – teadmusüksus, mille moodustab ühene tunnuste kombinatsioon

sõnastikusüsteem – haldussüsteem, mis võimaldab sõnastikke koostada, toimetada küljendada ja veebis avaldada ning teha lihtsaid ja keerulisi struktuuripõhiseid päringuid ning päringutulemusi sortida

termin – sõnaüksus, st sõna või sõnaühend, millega mõistet tähistatakse. Oskuskeelne sõna on termin. Termin on miski, mida saab välja öelda või kirja panna

termini ekstraheerimine – termini eraldamine ülejäänud sõnade kogust. Terminite ekstraheerimine on automaatne meetod teksti analüüsimiseks, et teha kindlaks terminikriteeriumidele vastavad üksused

Sissejuhatus

Vajadus terminoloogia järele tuleneb lihtsast inimese tarvidusest anda asjadele nimed ning neid tuvastada (Sageder 2010: 123). Terminoloogia ilmus sellisena nagu me seda täna mõistame 1930. aastatel. Kuid alles viimastel aastakümnetel on kasutatud tõeliselt süstemaatilist lähenemisviisi, mis hõlmab täielikult läbimõeldud põhimõtteid ja meetodikat (Bidnenko 2018: 213). Kaasaegsed keeleteadlased väidavad, et 18. ja 19. sajandil arendasid terminoloogiat teadlased; 20. sajandil insenerid ja tehnoloogid; 21. sajandil astusid lavale tõlkijad ja terminoloogid (Bidnenko 2018: 222).

Valisin magistritöö teemaks praktilise terminiloome valdkonna, kuna puutun kokku oma igapäevatoos Tü Tartu observatooriumis kaugseire terminitega. Alates 2008. aastast on Tartu observatooriumis korraldatud Eesti kaugseire päeva, mis on ainus eestikeelne kaugseire teaduskonverents maailmas. Koos konverentsiga antakse välja ka eestikeelset artiklite kogumikku. Need kaks sündmust on andnud suure panuse eestikeelsete kaugseire terminite arengusse. Kui varem rääkisid kaugseirest peamiselt ainult eksperdid ning enamasti inglise keeles, siis nüüd kasutavad kaugseire võimalusi ka eraettevõtted, avalike teenuste loojad ja lihtsalt huvilised. Ja seda juba eesti keeles. Sageli tõlgitakse termineid otse inglise keelest ja erinevates valdkondades veidi erinevalt. Siiani pole kaugseire eestikeelset terminoloogiat süstemaatiliselt arendatud.

Eestis tegeleb terminite väljatöötamisega peamiselt Eesti Keele Instituut (EKI)¹, mis koordineerib ja toetab ka oskuskeele terminibaaside loomist. Igal aastal toetatakse termini konkursi kaudu terminikomisjone, mis arendavad oskuskeelt väga erinevates valdkondades. Konkursil osutus 2022. aastal valituks ka kaugseire terminikomisjon. Kuna see oli esimene oskuskeele terminibaas Eestis, mida oli plaanis luua korpuste abiga, siis tegin sellest 2022. aasta terminikomisjonide teabepäeval ettekande.

Mõte kasutada kaugseire terminibaasi loomiseks korpuseid tuli sellest, et vajasin terminite analüüsimiseks kindlamat alust kui lihtsalt minu enda intuitsioon. Hanks (2012a: 62) toob välja, et korpused pakuvad piisavalt andmeid, et teha usaldusväärseid üldistusi ja on seetõttu terminibaaside loomiseks väga kasulikud. Stefanowitsch (2020: 15) selgitab, et

¹ eki.ee

corpuse andmete kvaliteet on intuiitivsete andmete kvaliteedist parem kahes aspektis: 1) andmete usaldusväärsus (kui kindlad saame olla, et teised inimesed saavad samade meetoditega sama tulemuse) 2) andmete tõlevastavus (kui hästi me mõistame, millisele reaalmaailma nähtusele andmed vastavad). Siiski ei plaaninud ma ekspertide teadmisi korpustega asendada, vaid kasutasin corpuseid ekspertide teadmiste kõrval. Kogu korpustest saadud materjali vaatasin koos ekspertidega üle.

Vajadus eestikeelsete terminite järele on sundinud paljusid pakkuma (otse)tõlkeid inglise keelest. Selle tulemusena ringleb samadest terminidest mitmeid erinevaid variante, mis muudab (eriti uustulnukate jaoks) arusaamise keerukaks. See omakorda pärsib kaugseire valdkonna arengut. Magistritööl on kaks peamist uurimisküsimust: 1) kas on otstarbekas kasutada eestikeelse kaugseire terminibaasi loomiseks corpuseid? 2) kas ja milliste tööloikude juures on vaja ekspertide abi?

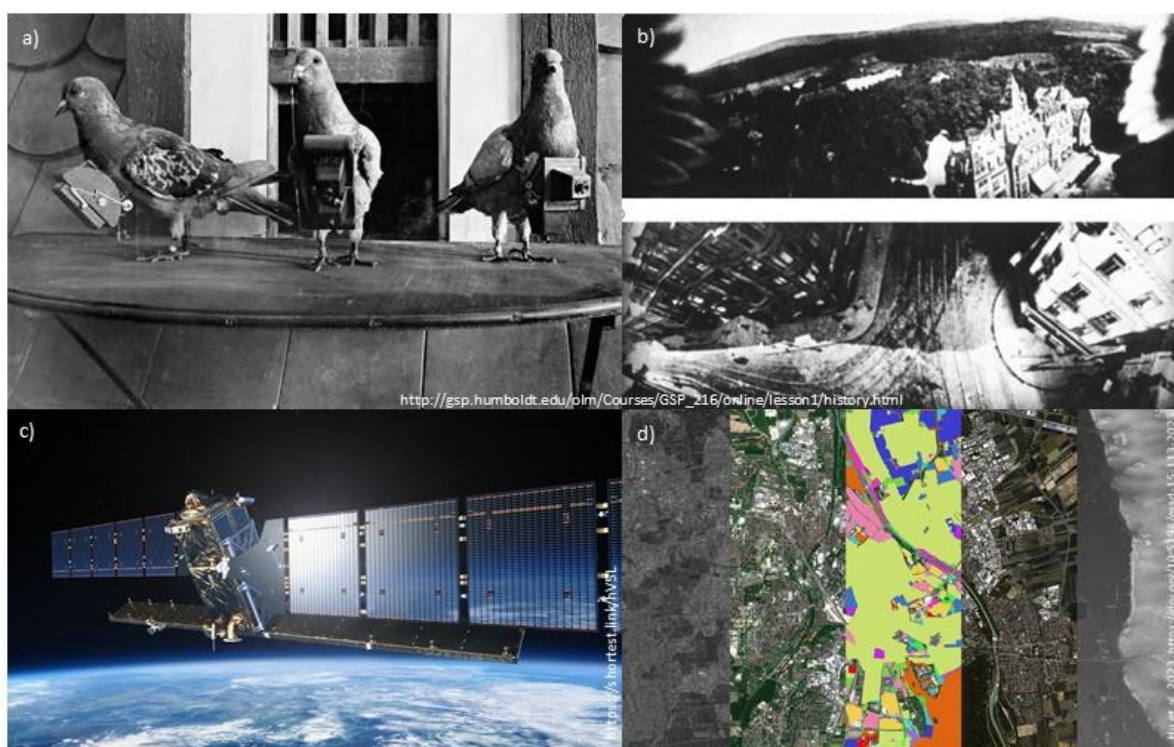
Magistritöö eesmärk on läbi kaugseire terminibaasi ning korpuste loomise süstematiseerida kaugseire oskuskeelt Eestis.

Magistritöö koosneb sissejuhatusest, neljast peatükist, kokkuvõttest ja lisadest. Esimeses peatükis käsitlen kaugseire valdkonda ning selle oskuskeele arengut Eestis. Teises peatükis annan lühikese ülevaate terminibaasidest Eestis ning tekstikorpuste kasutamise võimalustest terminibaasi loomisel. Kolmandas peatükis tutvustan magistritöö eesmärkide täitmiseks kasutatud meetodeid. Seejärel analüüsin neljandas peatükis korpuste kasutamise plusse ja miinuseid terminibaasi loomisel. Lisas 1 toon välja olulisemad kaugseire terminid ja nende variandid eesti ja inglise keeles. Uurimuse tulemusi on plaanis tutvustada konverentsil CLARIN Annual Conference 2023, teadusartikkel konverentsi kogumiku jaoks on valminud (lisa 2).

1. Kaugseire

1.1. Lühike ülevaade kaugseirest ja selle ajaloost

Kasutan selle magistritöö jaoks kaugseire kõige üldisemat definitsiooni ilma tehnilistesse detailidesse laskumata – „kaugseire on eemal asuvate objektide kohta informatsiooni hankimine mittekontaktsete meetoditega“ (IATE 2023). Kaugseire võeti terminina kasutusele 1960ndatel (Fussell ja Rundquist 1986: 1507). Siiski saab kaugseire sisuliseks alguseks pidada fotoaparaadi leiutamist. Kõige esimesed kaugseire pildid on võetud õhupallide pealt, aga selleks on kasutatud ka tuvisid (vt joonis 1). Suur hüpe kaugseire arengusse tuli II maailmasõjaga, mil arendati välja radar, sonar ning infrapuna tuvastuse süsteem. Alates 1960. aastatest töötavad sensorid peaaegu kogu elektromagnetilises spektris (Moore 1979: 477). Tänapäeval on lisaks hea lahutusvõimega, täpsetele ning perioodilistele andmetele pööratud suurt tähelepanu ka andmete kättesaadavusele ning võimalikult lihtsale käsitlemisele. Euroopa Liidu Copernicuse programmi raames on välja töötatud hulk Maa jälgimise teenuseid, millel on igapäevaselt kümneid tuhandeid kasutajaid (vt Peuch jt 2014: 11; Peuch jt 2022: 2650).



Joonis 1. Kaugseire areng: a) Julius Neubronneri kaameratega tuvid ja b) tuvide abil tehtud ülesvõtted

20. saj alguses Saksamaal. c) Tänapäevase satelliidi näitena Sentinel-1 ja d) tänapäevase kaugseire piltide näited (erinevate satelliitide, kanalite ja töötluste ribad).

Kaugseire on viimastel aastakümnetel muutunud mitmetes valdkondades väga oluliseks infoallikaks. Näiteks on kaugseire kaasa toonud tõelise revolutsiooni ökoloogias ning keskkonnauuringutes (Anderson ja Gaston 2013: 138; Chapman jt 2011: 337; Kerr ja Ostrovsky 2003: 299; Pettorelli jt 2014: 839; Zellweger jt 2019: 327), võimaldades automatiseeritud viisil koguda suurt hulka perioodilisi andmeid. Viimastel aastakümnetel on kaugseire andmed läinud nii täpseks, et võimaldavad isegi putukate uurimist (Rhodes jt 2022: 343). Tõuke tormiliseks kaugseire andmete kasutuselevõtuks Eestis on andnud märkimisväärne hulk tasuta kättesaadavaid andmeid, mida kogub ja töötleb Euroopa Liidu (EL) keskkonna programm Copernicus (Noorma jt 2020: 2). Lisaks teadlastele ning avaliku sektori esindajatele kasutavad kaugseire andmeid ka paljud eraettevõtted (nt Domnich jt 2021: 4100; Komisarenko jt 2022: 1; Voormansik jt 2020: 3784).

1.2. Kaugseire oskuskeele areng Eestis

Kaugseire valdkonna oskussõnavara ei ole eesti keeles siiani süstemaatiliselt arendatud. Palju kasutatakse toortõlkeid ja mugandusi inglise keelest. Käibel on mitmeid erinevaid variante sama sisuga mõistete jaoks. Kaugseire oskuskeele arendamine pole üksnes keeleilu küsimus, segane ja ebahütlane keelekasutus peletab eemale potentsiaalseid huvilisi, pärssides sellega kaugseire valdkonna ja Eesti teadmispõhise majanduse arengut.

Alates 2008. aastast on TÜ Tartu observatoorium korraldanud Eesti kaugseire päevi, kus antakse ülevaade Eesti kaugseire hetkeseisust. Viiel korral toimunud teaduskonverentsi järel on alati välja antud ka eestikeelsete kaugseire alaste teadusartiklite kogumik. Need on kättesaadavaks tehtud läbi kaugseire kogukonna lehe kaugseire.ee². Paljud eestikeelsed terminid on esmakordselt kasutusel just nendes kogumikes. Teine oluline allikas eestikeelsete terminite jaoks on üliõpilaste eestikeelsed bakalaureuse- ja magistritööd. Õppejõud on välja töötanud ka mõned eestikeelsed kaugseire loengukursused (vt kaugseirega seonduvad kursused

² kaugseire.ee

Eestis³). Samuti on kaugseire teemasid puudutatud eestikeelsetes õpikutes (nt Roosaare jt 2019: pt 8.2.2.).

Osaliselt kattub kaugseire valdkonna oskuskeel kosmosetehnoloogia, metroloogia (mõõteteadus) ja meteoroloogia valdkondade oskuskeelega. Nende valdkondade kohta on loodud terminibaasid Eesti Keele Instituudi sõnastiku- ja terminibaasisüsteemis Ekilex⁴ (vt Tavast jt 2018, 2021), mille väliskasutajaliides on keeleportaal Sõnaveeb⁵.

³ <https://kaugseire.ee/kaugseire-alane-haridus-eestis>

⁴ ekilex.ee

⁵ sonaveeb.ee

2. Oskusleksikograafia

2.1. Korpused oskusleksikograafias

Oskusleksikograafia on leksikograafia kui sõnastike tegemise õpetuse ja praktika osa, mille tegevusobjektiks on oskussõnastikud. See on just eesti oskuskeeletõpetusse kuuluv termin, teistes maades kasutatakse terminit terminograafia ja rõhutatakse mitte ühist, vaid erinevat – terminograafia versus leksikograafia (Erelt 2007: 360).

Korpused on mahukad elektroonilised tekstikogud, mis on koostatud keeleteaduse, arvutilingvistika ja leksikograafia vajadusi silmas pidades. Korpusi töödeldakse spetsiaalse tarkvaraga ehk korpuspäringusüsteemiga, mis võimaldab korpusandmete mitmekülget analüüsi (Koppel 2020: 11). Eesti Keele Instituudis kasutatakse alates 2011. aastast peamiselt korpuspäringusüsteemi Sketch Engine⁶ (Kilgarriff jt 2014; vt lisaks Fišer jt 2016; Jakubíček ja Šmerk 2016; Kostka 2022), mida kasutatakse laialdaselt ka Euroopa leksikograafide seas (Kallas jt 2019: 521). Vähemal määral kasutatakse Eesti Keele Instituudis ka korpuspäringusüsteemi KORP⁷.

Sõnastike koostamisel hakati korpuseid esmakordselt kasutama 1987. aastal (Hanks 2012b: 398), kui sõnaraamatu välja töötamise jaoks kasutati üldkorpust (Hanks 2012a: 61). Suurim korpuste mõju ilmnis võõrkeeletõppe sõnaraamatutele. Korpused andsid piisavalt andmeid, et teha usaldusväärseid üldistusi (Hanks 2012a: 62). Tänapäeval kasutatakse korpuseid väga erinevates valdkondades, näiteks isegi filosoofiliste probleemide analüüsimiseks (Caton 2020: 51).

Korpuspäringusüsteem Sketch Engine võimaldab automaatselt koostada sagedusloendeid, leida statistilisi kollokatsioone, genereerida sõnavisandeid (*word sketch*) (Kilgarriff jt 2004: 105), tuvastada definitsioone (Kovář jt 2016: 391) ja termineid (*term extraction*) (Jakubíček jt 2014: 53), koostada tesaurust (Rychlý ja Kilgarriff 2007: 41), ning tuvastada sõnastiku näitelauseks sobivaid korpuslauseid (Koppel 2020: 13). Näitelausete jaoks on Sketch Engine'isse integreeritud reeglipõhisel valemil töötav tööriist Good Dictionary Examples ehk GDEX (Kilgarriff jt 2008: 425). Eesti keele korpuspõhise analüüsi tarbeks on

⁶ sketchengine.eu

⁷ korp.keeleressursid.ee

välja töötatud spetsiaalne moodul sõnavisandite ja terminite jaoks (Kallas jt 2017b: 309; Koppel 2020: 13).

Sõnastiku andmebaasi täisautomaatne genereerimine tähendab, et kõik sõnastiku üksused (märksõnastik, kollokatsioonid, näitelaused jm) ekstraheeritakse automaatselt korpuspäringusüsteemist sõnastikusüsteemi. Andmebaasi genereerimisele järgneb üldjuhul järeltoimetamine, mille käigus leksikograaf kontrollib ja puhastab automaatselt loodud sisu (nö poolautomaatne koostamine). Täisautomaatse koostamise korral korpusest genereeritud sisu ei toimetata (Koppel 2020: 13). Eestis on poolautomaatselt loodud mitmeid sõnaraamatuid (nt „Eesti-läti sõnaraamat 2015“⁸, „Läti-eesti sõnaraamat 2015“⁹ ja „Eesti keele naabersõnad 2019“¹⁰).

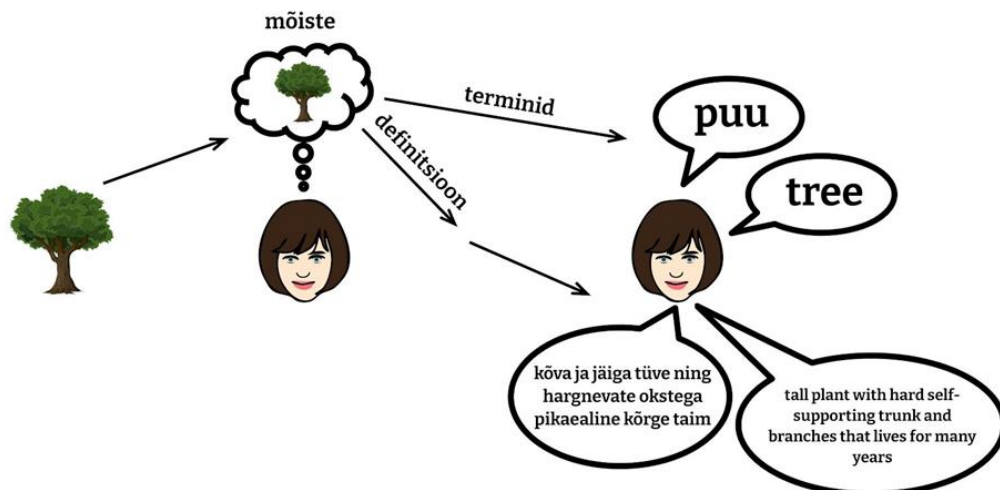
2.2. Terminitöö Eestis

Terminibaasid võivad olla väga sarnased sõnastikega, aga neil on üks olemuslik erinevus – terminibaasid on mõistepõhised, aga sõnaraamatud on üldjuhul sõnapõhised. Mõistepõhine tähendab, et kogu info on koondatud mõistete alla ning sõnapõhine, et kogu info on koondatud sõnade alla (EKI teatmik 2022). Mõiste on teadmüksus, st ettekujutus mingist objektist või nähtusest; termin on sõnauksus, st sõna või sõnaühend, millega mõistet tähistatakse (joonis 2). Ühele mõistele võib viidata mitme eri terminiga (variandid), samuti erikeelsete vastetega. Terminibaasi kirje on seega mõistekirje, mis sisaldab üht mõistet tähistavaid termineid (EKI teatmik 2022).

⁸ <https://www.eki.ee/dict/et-lv/>

⁹ <https://www.eki.ee/dict/lv-et/>

¹⁰ <http://www.eki.ee/dict/kol/index.cgi>



Joonis 2. Mõiste, termini, definitsiooni ja sõna mõistekaart (EKI teatmik 2022).

Eestis koordineerib oskuskeelee terminibaase ja nende loomist peamiselt EKI. Igal aastal toetavad nad eriala eksperte, kes kirjutavad oma eriala terminid, variandid, definitsioonid ning näited Eesti Keele Instituudi poolt välja arendatud sõnastiku- ja terminibaasisüsteemi Ekilex (vt Tavast jt 2018, 2021). Praegu on väga erinevates valdkondades (vt näiteid joonisel 3) kokku 120 andmebaasi, see arv aasta-aastalt kasvab. Instituut korraldab igal aastal konkursi uute terminoloogiliste andmebaaside toetamiseks või olemasolevatesse andmebaasidesse terminite lisamiseks. Samuti korraldab EKI erinevate valdkondade ekspertidele koolitusi ning vajadusel juhendavad eksperte.

Vali kõik	Eemalda valikud
<input type="checkbox"/>	Aianduse terminibaas (ait)
<input type="checkbox"/>	Akadeemilise väljendusoskuse terminibaas (avt)
<input type="checkbox"/>	Andmeanalüüsi ja statistika oskussõnastik (aso)
<input type="checkbox"/>	Arheoloogia terminibaas (arh)
<input type="checkbox"/>	Arhitektuuri oskussõnastik (aoss)
<input type="checkbox"/>	Betoonkonstruktsioonide terminibaas (EELBÜTK)
<input type="checkbox"/>	Biokeemiasõnastik (bks)
<input type="checkbox"/>	Botaanika terminibaas (bot)
<input type="checkbox"/>	Eesti E-tervise SA terminibaas (ett)
<input type="checkbox"/>	Eesti rahvatantsu oskussõnastik (ros)
<input type="checkbox"/>	Eesti-vene-eesti õigusterminoloogiabaas (õtb)
<input type="checkbox"/>	Eesti-vene-inglise spaaterminid (spaa)
<input type="checkbox"/>	Eesti-vene töötervishoiu ja -ohutuse terminibaas (tot)
<input type="checkbox"/>	Eesti viipekeele IT terminid (evkit)
<input type="checkbox"/>	Eesti viipekeele meditsiinterminid (vkm)
<input type="checkbox"/>	Ehitiste projekteerimise terminibaas (ehpr)
<input type="checkbox"/>	EKI ühendterminibaas Esterm 2 (est_test)
<input type="checkbox"/>	Elektrotehnika (eleth)
<input type="checkbox"/>	Entomoloogia terminibaas (ent)
<input type="checkbox"/>	Etenduskunstide terminibaas (eko)
<input type="checkbox"/>	Etümolooogia (ety)
<input type="checkbox"/>	Euroopa keeleõppe raamdokumendi terminid (cefr)
<input type="checkbox"/>	E-õppe terminid / E-learning terms (eõt)
<input type="checkbox"/>	Filmikunsti terminibaas Filmterm (fkm)
<input type="checkbox"/>	Filosoofia terminibaas (fil)
<input type="checkbox"/>	Folkloorsete uskumusolendite sõnastik (usk)
<input type="checkbox"/>	Foneetika sõnastik (fon)
<input type="checkbox"/>	Galeegi-eesti sõnaraamat (gal)
<input type="checkbox"/>	Geneetika terminibaas (gen)
<input type="checkbox"/>	Geoloogia terminibaas (get)
<input type="checkbox"/>	Geomorfoloogia terminibaas (gmt)
<input type="checkbox"/>	Geriaatria terminibaas (GER)
<input type="checkbox"/>	Hambatehnika terminibaas (den)
<input type="checkbox"/>	Hümnoloogia terminibaas (hym)
<input type="checkbox"/>	Ida mõtteleo leksikon (ida)
<input type="checkbox"/>	iht import 1 (iht_200915)
<input type="checkbox"/>	Ihtuoloogia terminibaas (iht)
<input type="checkbox"/>	Immunoloogia terminibaas (imm)
<input type="checkbox"/>	Inglise vasted (ing)
<input type="checkbox"/>	Katsebaas (kce)
<input checked="" type="checkbox"/>	Kaugseire terminibaas (kas)
<input type="checkbox"/>	Keemiaterminite baas (kem)
<input type="checkbox"/>	Kognitiivse keeleteaduse terminibaas (kkt)
<input type="checkbox"/>	Kokanduse terminibaas (kok)
<input type="checkbox"/>	Kooliinformaatika terminibaas (eiops)
<input type="checkbox"/>	Koolisõnastikud 2005–2010 (kool_KV)
<input type="checkbox"/>	Kosmetotehnoloogia (kth)
<input type="checkbox"/>	Kriisijuhtimise terminibaas (kjt)
<input type="checkbox"/>	Käsitööteaduse oskussõnad (kto)
<input type="checkbox"/>	Kõite ja konserveerimise terminibaas (LoB)
<input type="checkbox"/>	Küberfüüsikalise süsteemitehnika terminibaas (kfs)
<input type="checkbox"/>	Linnoloogia sõnastik (lim)
<input type="checkbox"/>	Linguae, maailma keeled, kirjad ja rahvad (linguae)
<input type="checkbox"/>	Loomakasvatuse terminibaas (lkt)
<input type="checkbox"/>	Loomanimetuste terminibaas (lon)
<input type="checkbox"/>	Loomaparasitide nimistu (pre)
<input type="checkbox"/>	Loomaparasitide terminibaas (lpr)
<input type="checkbox"/>	Loomi kaasvate organisatsioonide terminibaas (lko)
<input type="checkbox"/>	Läti-eesti sõnastik 2015 (les)
<input type="checkbox"/>	Materjalitehnika terminibaas (mtr)
<input type="checkbox"/>	Meditsiinifüüsika terminibaas (mef)
<input type="checkbox"/>	Mesindusleksikon (mes)

Joonis 3. Valik Ekilexi terminibaase. Linnukesega on märgitud magistritöö raames valminud kaugseire terminibaas.

Ajalooliselt esimene sõnastikusüsteem Eesti Keele Instituudis oli EELex (kasutusel alates 2003. aastast), selle kõrvale lisandusid teised ning ühel hetkel kasutati üht sõnastikusüsteemi ja kaht terminibaasi, igal oma head ja vead. EELex oli mõeldud üldkeele sõnaraamatute jaoks, terminoloogid kasutasid terminibaasisüsteeme Termeki (pole enam kasutusel) ja MultiTerm (Langemets jt 2021: 756). Sõnastikuinfo ühendamine algas EKI-s 2017. aastal. Nüüd jooksevad kasutajasõbralikuma lahendusena eri sõnakogude ja terminibaaside info kokku ühte (agregeeritud) leksikograafilisse andmebaasi, mis tavakasutajale avaldub keeleportaalis Sõnaveeb (Koppel jt 2019b: 434). Igal terminibaasil on oma kodulehekülg, nt kaugseire terminibaasil¹¹ (joonis 4).

¹¹ [Kaugseire terminibaasi avaleht Sõnaveebi portaalis](#)

Kaugseire terminibaas

Kaugseire terminibaas koosneb hetkel 100 avalikust terminist ning on loodud koostöös Eesti Keele Instituudiga kaugseire korpuste ning TÜ Tartu observatooriumi kaugseire ekspertide teadmiste abil. Kontakt: liisi.jakobson@ut.ee

Õtsi sõnade tähendusi ja tõlkeid



abDEefghIklmoprSstvõü

Joonis 4. Kaugseire terminibaasi avaleht Sõnaveebi portaalis. Internetis aadressil (<https://sonaveeb.ee/ds/kas/>)

3. Andmed ja metoodika

Kuna varasemalt eesti keeles puudus kaugseire korpus, alustasin kaugseire terminite süstematiseerimise tööd just korpusete loomisega. Töö käigus valmis neli korpus – üks eesti keeles, üks inglise keeles ning kaks paralleelkorpus (eesti-inglise ja inglise-eesti). Kallas jt (2017: 318) soovib terminitöö tegemise jaoks kasutada vähemalt 5 miljoni sõnalisi korpusete, et saaks usaldusväärset rakendada kogu korpusparengusüsteemi võimalusi. Polnud keeruline leida 5 miljonit sõna ingliskeelses kaugseire korpusesse. Samas eestikeelne korpus jäi ka pärast kõikide kättesaadavate eestikeelsete kaugseire teemaliste tekstide lisamist alla 3 miljoni. Veelgi keerulisem oli leida kaugseire alaseid paralleeltekste; peamine allikas selleks oli Eurlex¹². Paralleelkorpusete maht jäi napilt alla 0,5 miljoni sõna.

Kasutasin korpusparengusüsteemi Sketch Engine, mis toetab umbes 100 keelt (Kostka 2022: 123) ja pakub kasutajatele tööriistu oma korpusete loomiseks (Kilgarriff jt 2014: 7). Korpusete võib luua üleslaaditud tekstide najal ja/või kasutada kroolimisfunktsiooni ning lasta korpusparengusüsteemil ette antud võtmesõnade abil internetist tekste leida (Fišer jt 2016: 136). Rakendasin kaugseire korpusete loomisel mõlemat võimalust.

Kuna kasutan terminibaasi loomiseks peamiselt eestikeelset korpuset, siis kirjeldan selle sisu. Eestikeelne kaugseire korpus „Kaugseire korpus 2022“ koosneb magistritöö kirjutamise ajal 347 dokumendist, millest 58% laadisin ise üles ja ülejäänud tekstid on saadud kroolimise teel internetist. Üleslaaditud dokumendid on peamiselt teadusartiklid, õppematerjalid, ettekanded, õigustekstid jne. Veebitekstid pärinevad peamiselt populaarteaduslikest ajakirjadest, õigustekstidest, kaugseiretegevõtete tekstidest, Põllumajanduse Registrite ja Informatsiooni Ameti kodulehelt, Eesti kaugseire kogukonna kodulehe materjalidest, kaugseire teadusprojektide materjalidest jne. Sotsiaalmeedia allikatest, näiteks kasutajafoorumitest, blogidest, kommentaaridest jne on materjali väga vähe, sest kaugseire pole Eesti sotsiaalmeedias populaarne teema. Sealt leiab vaid mõne ametliku kaugseire teemalise teadaande ning kaugseire väga kitsa valdkonna – droonide – tehnilist infot.

Ingliskeelset korpuset ning paralleelkorpusete kasutasin peamiselt selleks, et leida eestikeelsetele terminitele ingliskeelseid vasteid (ning vastupidi), samuti ingliskeelseid definitsioone ning näitelauseid.

¹² eur-lex.europa.eu

Korpuspäringusüsteemis Sketch Engine on võtmesõnade ning mitmesõnaliste terminite ekstraheerimise (*Keyword and Term Extraction*) moodul. Selles kasutatakse referentskorpuse ning fookuskorpuse võrdlemist (Jakubíček ja Šmerk 2016: 143). Referentskorpusena kasutasin Eesti keele ühendkorpust 2021 (2,4 miljardit sõna) ning võtmesõnu ja mitmesõnalisi termineid otsisin korpusest „Kaugseire korpus 2022“ (3 miljonit sõna). Selleks, et saada kvaliteetsemaid vasteid terminite ekstraheerimisele, kasutab Sketch Engine terminite grammatikat (*Term Grammar*) (Fišer jt 2016: 136). See tähendab, et korpuspäringusüsteem kontrollib, kas otsitavad võtmesõnade või mitmesõnaliste terminite kandidaadid vastavad terminite jaoks vajalikule grammatikale vastavas keeles. Eesti keele jaoks välja töötatud terminite grammatika v2.0 põhineb tõenduspõhistel termini grammatikadisaini põhimõtetel. Versiooni v2.0 kirjutas Marek Blahuši koostöös EKI keeletehnoloogi Eleri Aedmaa ja vaneterminoloogi Merily Pladoga (Koppel ja Kallas 2022: 220) ning see põhineb IATE¹³ andmete analüüsil.

Peale grammatikale vastavuse kontrollimist kuvab terminite ekstraheerimise moodul võtmesõnade või mitmesõnaliste terminite kandidaadid vastavalt esinemissagedusele ning erinevusele referentskorpusest. Joonisel 5 on toodud korpuse „Kaugseire korpus 2022“ ühe- ja mitmesõnaliste terminite ekstraheerimise 10 esimest tulemust. Tabelist on näha, et mõned terminite kandidaadid on inglise keeles. Põhjuseks on see, et paljudes meie (teaduslikes) tekstides on kokkuvõte (või mõni muu teksti osa) inglise keeles. Kokku vaatasin läbi 1000 ühesõnalist ning 1000 mitmesõnalist termini kandidaati.

	Lemma	Frequency ²		Term	Frequency ²	
		Focus	Reference		Focus	Reference
1	kaugseire	4,598	892	1 euroopa liidu teataja	2,748	1,728
2	copernicus	4,122	1,719	2 remote sensing	861	9
3	sensing	1,474	111	3 liidu teataja	2,208	1,663
4	landsat	1,108	39	4 füüsiline kliimarisk	655	6
5	satelliidipilt	1,476	355	5 ruumiline lahutus	528	33
6	kliimarisk	958	163	6 tartu observatoorium	1,082	1,058
7	esthub	861	62	7 maa kaugseire	455	11
8	sentinel-2	765	39	8 kaugseire andmed	422	53
9	l174	733	8	9 turvalisuse akrediteerimine	588	466
10	spektraalne	766	191	10 copernicuse programm	494	343

¹³ iate.europa.eu

Joonis 5. Korpuspäringusüsteemi Sketch Engine terminite ekstraheerimise mooduli esimesed 10 termini kandidaati ühe- ja mitmesõnaliste terminite jaoks korpusest „Kaugseire korpus 2022“.

Selleks, et valida korpuspäringusüsteemi poolt pakutud terminite kandidaatide seast kõige sobivamad, tegin eelnevalt koostööd kaugseire ekspertidega (nn kaugseire terminikomisjonis) ning palusin neil välja pakkuda probleemseid kaugseire termineid eesti keeles. Probleemseteks defineerisime koos TÜ Tartu observatooriumi kaugseire ekspertidega järgmised olukorrad:

- ühele ingliskeelsele terminile vastab mitu eestikeelset terminit,
- eestikeelse termini tähendus on erinev vastava ingliskeelse termini omast,
- mitme erineva ingliskeelse termini jaoks on üks eestikeelne termin.

Otsisin korpuspäringusüsteemi poolt pakutud terminite kandidaatide seast kõigepealt kaugseire ekspertide poolt pakutud nõ probleemseid termineid ning nende variante. Metoodiliselt on see *corpus-based* lähenemine (vt Storjohann 2005: 9). Seejärel vaatlesin ilma eelneva ülesande püstituseta terminite kandidaate (*corpus-driven* lähenemine, vt Storjohann 2005: 5), nende esinemist võrreldes referentskorpusega ning sagedust miljoni sõna kohta; terminite esinemise kontekste ning näitelauseid.

Tabelis 1 on näitena toodud olukord, kus ühele ingliskeelsele terminile (*reflectance*) vastab eestikeelses kaugseire kirjanduses seitse erinevat terminit. See on teadaolevalt kõige rohkemate variantidega termin eesti kaugseire oskuskeeles.

Tabel 1

Inglise termini *reflectance* tõlkevasted.

Eestikeelne termin	Esinemiste arv korpuses „Kaugseire korpus 2022“	Sagedus miljoni kohta
heleduskordaja	242	61
heleduskoefitsient	158	40
heledustegur	12	3
peegeldustegur	246	62
peegeldumistegur	69	17
peegelduskoefitsient	15	4
peegeldumiskoefitsient	29	7

Kaugseire ekspertidega arutlesime ka terminite üle, millel pole veel eesti keeles vastet (nt *handheld lidar*, *backpack lidar*). Kui leidsime sobiva vaste, siis läks see otse terminibaasi. Eesti Keele Instituut valib iga kuu välja „Kuu termini“, 2023. aasta mais valiti selleks meie poolt välja pakutud täiesti uus termin „märsilidar“ (*backpack lidar*).

Näidete leidmiseks korpusest kasutasin korpuspäringusüsteemi Sketch Engine integreeritud tööriista GDEX (*Good Dictionary Examples*), mis tuvastab näitelauseks sobivaid korpuslauseid (Kilgarriff jt 2008: 426). GDEX reastab laused vastavalt eelnevalt kindlaksmääratud kriteeriumidele (Koppel jt 2019a: 764), näiteks vaadatakse, et kõik lauseliikmed on olemas ja lause pikkus oleks sobiv. Esimene GDEXi konfiguratsioon eesti keele jaoks töötati välja 2014. aastal (Kallas jt 2015: 1). Praegu kasutatakse Sketch Engines 2018. aastal välja töötatud versiooni GDEX 1.4 (Koppel jt 2019b: 446).

4. Tulemused ja analüüs

Kaugseire korpuste andmete ning kaugseire ekspertidelt saadud info tulemusena valmis sõnastiku- ja terminibaasisüsteemis Ekilex kaugseire terminibaas. Terminikirje Ekilexis on struktureeritud mõistepõhine üksus, mis sisaldab teavet eelistatud termini, variantide, lühendite, definitsioonide, selgituste, näitelause, kontekstide jm kohta (joonis 6). Hetkel on kaugseire terminibaasis 100 terminit (vt lisa1).

The screenshot shows the Ekilex terminibaas interface. At the top, there is a search bar with the number '629391' and buttons for 'Lisa termin', 'Ühenda', 'Dubleeri', and 'Kustuta'. The date '04.12.2022' is also visible. Below the search bar, there are two main sections: 'Seletus' and 'Tähenduse seos'. The 'Seletus' section shows the definition of 'tehisavaradar' in Estonian and English. The 'Tähenduse seos' section shows a link to '576026 sünteesavaradar (et) [kth] [eelistatud]' with a 'duplikaadikandidaat' label. Below this, there are three entries for 'tehisavaradar', 'synthetic-aperture radar', and 'SAR', each with its own definition and context.

Seletus	Tähenduse seos
<p>et</p> <ul style="list-style-type: none">radarisüsteem, mis korreleerib satelliidi orbiidi või õhusõiduki lennutee eri punktides emiteeritud signaalide kajasid <i>definitsioon</i> (Avalik) [The Dictionary of Space Technology]Tehisavaradari eripäraks on antenni ja uuritava piirkonna vahelise suhtelise liikumise kasutamine. Sel viisil tekitab lühikeste vaheaegade järel signaale välja saatev liikuv antenn efekti suuremast antennist – tekib nn virtuaalne antenn ehk tehisava. <i>selgitus</i> (Avalik) [Voormansk 2013]Eestikeelses kaugseire alases kirjanduses kasutatakse valdavalt terminit tehisavaradar (või tehisava-radar), kosmosetehnoloogia alases kirjanduses sünteesavaradar. <i>selgitus</i> (Avalik) [Kaugseire korpus 2022] <p>en</p> <ul style="list-style-type: none">a radar system that correlates the echoes of signals emitted at different points along a satellite's orbit or an aircraft's flight path <i>definitsioon</i> (Avalik) [The Dictionary of Space Technology]	<p>576026 sünteesavaradar (et) [kth] [eelistatud] duplikaadikandidaat</p>
<p>tehisavaradar</p> <p>Kontekst</p> <ul style="list-style-type: none">Kogu Soome poolt teostatav jääkaardistus Läänemeres on tehisavaradari piltide põhine, aerolaserskanneerimist jääseireks ei kasutata. (Avalik) [Innovatsiooniuringud 2013]	
<p>synthetic-aperture radar</p> <p>Kontekst</p> <ul style="list-style-type: none">Synthetic Aperture Radar (SAR) is an emerging technology in remote sensing. (Avalik) [GISGeography 2022]	
<p>SAR</p> <p>Kontekst</p> <ul style="list-style-type: none">Generally, SAR satellite imagery uses several deferent reflection, propagation, and scattering to collect information from the target on Earth. (Avalik) [Jefriza et al., 2020]	

Joonis 6. Terminikirje näidis sõnastiku- ja terminibaasisüsteemis Ekilex.

Lisaks sissejuhatuses nimetatud Stefanowitschi (2020) välja toodud korpuste eelistele terminibaasi loomise juures, lisan siia väga praktilise loetelu korpuste võimaluste kohta. Korpused võimaldavad:

- näha kõiki asjakohaseid variante,

- eristada kasutussageduse järgi vähem ja rohkem kasutuses olevaid variante,
- selgitada variantide erinevaid kontekste ja kasutusvaldkondi (vt näide allpool),
- saada infot variantide kasutuse muutuse kohta ajas,
- tuua välja variantide tähendus erinevate kontekstide (näitelausete) abil,
- tuua välja kasutussagedusel põhinev ning korrektse terminile vastava grammatikaga terminite kandidaadid,
- leida usaldusväärseid allikaid definitsioonide, selgituste ja näitelausete tarvis.

Kaugseire jaguneb alavaldkondadeks (taimkatte kaugseire, veekogude kaugseire, atmosfääri kaugseire). Erinevates alavaldkondades võib ette tulla erinevate variantide kasutamist. Lisaks on kaugseire oskuskeelel ühisosa kosmosetehnoloogia, metroloogia ja meteoroloogia valdkondade oskuskeelega. Näiteks kasutatakse metrooloogias tabelis 1 kirjeldatud variantide hulgast enim terminit *heledustegur*, mis on kaugseires kasutuses üsna harva (jätsin selle ka segaduse vähendamiseks kaugseire terminibaasist välja).

Tabel 1 on kujundlik näide selle kohta, mida korpuste abil saab teha ning milleks ükski ekspert ei ole võimeline. Ekspertidel on suurepärase teadmised valdkonnast ning ta teab terminite tähendusi ja kasutusvaldkondi, küllap ka enamus terminite variante. Aga ekspert ei suuda numbrilise täpsusega ära määrata teiste inimeste terminikasutust. Selleks on korpused suurepäraseks abilisid.

Lisaks korpuste võimalustele, tuleb silmas pidada ka mõningaid puudusi ja ohukohti. Kõigepealt asjaolu, et kui soovitakse luua erialast terminibaasi korpuste baasil, siis tõenäoliselt pole selliseid korpuseid eesti keeles veel olemas. Uue ja spetsiifilise valdkonna puhul võib eestikeelsete tekstide leidmine piisavas mahus (Kallas jt 2017: 318 soovivad korpuste suuruseks vähemalt 5 miljonit sõna) osutada võimatuks ülesandeks. Kaugseire terminibaasi loomine põhiliselt 3 miljoni sõna suuruse korpuse abiga näitas, et on võimalik toetuda ka väiksemale korpusele. Kaugseire korpusesse kaasasin kõikvõimalikud eestikeelsed kättesaadavad tekstid. Paljudes tekstides oli kaugseirest juttu üsna põgusalt, peateema oli hoopis miski muu. See on tõenäoliselt põhjust, miks asjasse puutuvaid ja mõttekaid termineid oli ekstraheerimise tulemustes üsna vähe (vt joonis 5). Samas oli terminite ekstraheerimise tulemus 20 korda suurem kui planeeritud terminite arv terminibaasis (vastavalt 1000 ühesõnalist + 1000 mitmesõnalist termini kandidaati ja 100 terminit terminibaasis), mis andis piisavalt suure varu terminite valimise juures. Suurem ja puhtalt kaugseire tekstidel põhinev korpus oleks võimaldanud näiteid ja definitsioone valida suuremast hulgast tekstidest.

Korpuseid kasutades tuleb meeles pidada, et seal on tekstid nii nagu inimesed need kirjutanud on. Tekste pole täiendavalt parandatud. Kuna meie üleslaaditud dokumendid ja enamik kroolitud tekstidest on peamiselt ametlikud või ajakirjanduslikud tekstid, võime eeldada, et vigade arv (müra) on väga madal. Nö müra on iseloomulik sotsiaalmeediale (Baldwin jt 2013: 356). Teine probleem kroolitud andmetega on sageli masintõlke tekstid (Jakubíček jt 2020: 1). Kaugseire eestikeelsest korpusest üle poole moodustasid üleslaaditud tekstid. Nende seas kindlasti masintõlke tekste ei olnud. Samuti ei õnnestunud mul tuvastada ühtegi masintõlke teksti kroolitud materjali seast. Arvan, et masintõlke probleem ei ole eestikeelse kaugseire korpuse puhul oluline. Siiski on probleemiks kolmas suur interneti baasil loodud korpuste puudus, nimelt pole kaugseire korpuses piisavalt metaandmed (Jakubíček jt 2020: 1). Metaandmeid on liiga vähe ka üleslaaditud tekstide juures. Kui edaspidi on plaanis teha korpus (näiteks läbi korpuspäringusüsteemi KORP) (tavakasutajale) kättesaadavaks, siis tuleb metaandmeid kindlasti lisada.

Kokkuvõte

Magistritöö eesmärk oli süstematiseerida kaugseire oskuskeelt Eestis. Selleks koostasin kaugseire korpused ning nende abil kaugseire terminibaasi, kus iga termini kohta tõin selle variandid, ingliskeelse(d) vaste(d), definitsiooni, selgitus(ed), kasutusnäited, allikad. Hetkel on kaugseire terminibaasis 100 terminit, need on toodud lisas nr 1. Kaugseire terminibaas on Sõnaveebi kasutajaliidese kaudu kõigile kättesaadav.

Magistritöös püstitatud esimesele uurimisküsimusele (kas on otstarbekas kasutada eestikeelse kaugseire terminibaasi loomiseks korpuseid?) vastan kindlasti jaatavalt. Kuigi mul ei õnnestunud eestikeelse kaugseire korpuse suuruseks saada Kallas jt (2017: 318) poolt soovitatud 5 miljonit sõna, andis korpus väga hea baasi terminibaasi loomiseks. Inglisekeelsete terminite jaoks kasutasin eelnevalt loodud ingliskeelset ning kahte paralleelkorpust.

Lisas 1 toodud kaugseire terminibaasi kirjete põhjal on näha, et umbes kolmandikul praeguse terminibaasi terminitest on kaks või enam varianti. Kõige suurema arvu variantidega termin on toodud tabelis 1. See näitab kujundlikult mida korpuste abil saab teha ning milleks ükski ekspert ei ole võimeline. Ekspertidel on valdkonna kohta suurepärased teadmised, aga ükski inimene ei suuda numbriliselt tuvastada teiste inimeste terminikasutust.

Korpused on terminibaasi loomisel väga kasulikud, aga kindlasti ei asenda nad ekspertide teadmisi. Kõige rohkem läks ekspertide abi tarvis definitsioonide koostamisel ning eeliterminite valimisel (vastus teisele uurimisküsimusele).

Korpuste baasil terminibaasi loomisel tuleb arvestada ka korpuste mõningate puudustega. Sõltuvalt terminibaasi teemast, võib kroolimisel saadud materjal olla rohkem või vähem mürane. Teaduslike teemadega on see probleem väiksem. Tähelepanu tuleb pöörata ka sellele, et kroolimise käigus ei võetaks korpusesse masintekste. Samuti on oluline korpuste varustamine metaandmetega.

Tulevikus tuleb kaugseire terminibaasi täiendada vastavalt kasutajate tagasisidele, ekspertide soovitudele ning uutele kaugseire tekstidele. Enne eestikeelse kaugseire korpuse lisamist korpusingüsteemi KORP tuleb metaandmeid kindlasti täiendada.

Kirjanduse loetelu

- Anderson, K., Gaston, K.J., 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*. 11, 138–146. <https://doi.org/10.1890/120150>
- Baldwin, T., Cook, P., Lui, M., MacKinlay, A., Wang, L., 2013. How Noisy Social Media Text, How Different Social Media Sources? *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, Japan, 356–364.
- Bidnenko, N., 2018. Practical and theoretical issues of modern terminology. ВІСНИК УНІВЕРСИТЕТУ ІМЕНІ АЛЬФРЕДА НОБЕЛЯ, Серія «ФІЛОЛОГІЧНІ НАУКИ» No 1, 213–223.
- Caton, J.N., 2020. Using Linguistic Corpora as a Philosophical Tool. *Metaphilosophy* 51, 51–70. <https://doi.org/10.1111/meta.12405>
- Chapman, J.W., Drake, V.A., Reynolds, D.R., 2011. Recent Insights from Radar Studies of Insect Flight. *Annual Review of Entomology*, 56, 337–356. <https://doi.org/10.1146/annurev-ento-120709-144820>
- Domnich, M., Sünter, I., Trofimov, H., Wold, O., Harun, F., Kostiukhin, A., Järveoja, M., Veske, M., Tamm, T., Voormansik, K., Olesk, A., Boccia, V., Longepe, N., Cadau, E.G., 2021. KappaMask: AI-Based Cloudmask Processor for Sentinel-2. *Remote Sensing* 13, 4100. <https://doi.org/10.3390/rs13204100>
- EKI teatmik, 2022. Terminibaas ja sõnaraamat - EKI teatmik. Termin. Ja Sõnaraamat. <https://eki.ee/teatmik/terminibaas-ja-sonaraamat/> (loetud 19.02.2023).
- Erelt, T., 2007. Terminõpetus, Tartu Ülikooli Kirjastus, Tartu, 470 lk.
- Fišer, D., Suchomel, V., Jakubíček, M., 2016. Terminology Extraction for Academic Slovene Using Sketch Engine, *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2016*, Czech Republic, 135–141.
- Fussell, J., Rundquist, D., 1986. On Defining Remote Sensing. *Photogrammetric Engineering & Remote Sensing* 52, 1507–1511.
- Hanks, P., 2012a. Corpus evidence and electronic lexicography, in: Granger, S., Paquot, M. (Eds.), *Electronic Lexicography*. Oxford University Press, 57–82. <https://doi.org/10.1093/acprof:oso/9780199654864.003.0004>
- Hanks, P., 2012b. The Corpus Revolution in Lexicography. *International Journal of Lexicography* 25, 398–436. <https://doi.org/10.1093/ijl/ecs026>
- IATE - otsing <https://iate.europa.eu/search/result/1674046067698/1> (loetud 18.01.2023).
- Jakubíček, M., Šmerk, P., 2016. Large Scale Keyword Extraction using a Finite State Backend, *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2016*, Czech Republic, 143–146.
- Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., Suchomel, V., 2014. Finding Terms in Corpora for Many Languages with the Sketch Engine, *Proceedings of the Demonstrations at*

- the 14th Conference of the European Chapter of the Association for Computational Linguistics, 53–56. <https://doi.org/10.3115/v1/E14-2014>
- Jakubíček, M., Kovář, V., Rychlý, P., Suchomel, V., 2020. Current Challenges in Web Corpus Building, *Proceedings of the 12th Web as Corpus Workshop*, Marseille, 1–4.
- Jüri Roosaare, Kiira Mõisja, Raivo Aunap, 2019. Geoinformaatika. Tartu Ülikooli Kirjastus, Tartu, 830 lk.
- Kallas, J., Kilgarriff, A., Koppel, K., Kudritski, E., Michelfeit, J., Viks, Ü., 2015. Automatic generation of the Estonian Collocation Dictionary database, *Proceedings of the ELex 2015 Conference. Herstmonceux Castle*, United Kingdom, 1–20.
- Kallas, J., Suchomel, V., Khokhlova, M., 2017. Automated Identification of Domain Preferences of Collocations, *Proceedings of ELex 2017 Conference*, 2017, 309–320.
- Kallas, J., Koeva, S., Langemets, M., Tiberius, C., Kosem, I., 2019. Lexicographic Practices in Europe: Results of the ELEXIS Survey on User Needs. Presented at the *eLex 2019 Conference*, Sintra, Portugal, 519–536.
- Kerr, J.T., Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution* 18, 299–305. [https://doi.org/10.1016/S0169-5347\(03\)00071-5](https://doi.org/10.1016/S0169-5347(03)00071-5)
- Kilgarriff, A., Rychlý, P., Smrz, P., Tugwell, D., 2004. The Sketch Engine, *Proceedings of the 11th EURALEX International Congress*, 105–115.
- Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., Rychlý, P., 2008. GDEX: Automatically Finding Good Dictionary Examples in a Corpus, *Proceedings of the XIII EURALEX International Congress*, Barcelona, 425–432.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V., 2014. The Sketch Engine: ten years on. *Lexicography* 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Komisarenko, V., Voormansik, K., Elshawi, R., Sakr, S., 2022. Exploiting time series of Sentinel-1 and Sentinel-2 to detect grassland mowing events using deep learning with reject region. *Scientific Reports* 12, 1–15. <https://doi.org/10.1038/s41598-022-04932-6>
- Koppel, K., 2020. Näitelausete korpuspõhine automaattuvastus eesti keele õppesõnastikele (doktoritöö), <https://dspace.ut.ee/handle/10062/67138>, 117 lk.
- Koppel ja Kallas, 2022. Eesti keele ühendkorpuste sari 2013-2021: mahukaim eestikeelsete digitekstide kogu. *Eesti Rakenduslingvistika Ühingu aastaraamat* 18, 207–228. <https://doi.org/doi:10.5128/ERYa18.12>
- Koppel, K., Kallas, J., Khokhlova, M., Suchomel, V., Baisa, V., Michelfeit, J., 2019a. SkELL Corpora as a Part of the Language Portal Sõnavieb: Problems and Perspectives, *Proceedings of the 6th Biennial Conference on Electronic Lexicography*, Czech, 763–782.
- Koppel, K., Tavast, A., Langemets, M., Kallas, J., 2019b. Aggregating Dictionaries into the Language Portal Sõnavieb: Issues With and Without Solutions. *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference*, Portugal, 434–452.

- Kostka, M., 2022. Pipeline Effectiveness in the Sketch Engine. *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN*, 123–130.
- Kovář, V., Moc̣iarikova, M., Rychly, P., 2016. Finding Definitions in Large Corpora with Sketch Engine. *Proceedings of the Tenth International Conference on Language Resources and Evaluation, Slovenia*, 391–394.
- Langemets, M., Koppel, K., Kallas, J., Tavast, A., 2021. Sõnastikukogust keeleportaaliks. *Keel ja Kirjandus* 64, 755–770. <https://doi.org/10.54013/kk764a6>
- Moore, G.K., 1979. What is a picture worth? A history of remote sensing. *Hydrological Sciences Bulletin* 24, 477–485. <https://doi.org/10.1080/02626667909491887>
- Noorma, A., Jakobson, L., Lang, M., Kutser, T., Oja, T., Uiboupin, R., Voormansik, K., Puust, R., Post, P., Sepp, K., Liibus, A., 2020. Kaugseire andmete kasutuselevõtt avalike teenuste väljatöötamisel ja arendamisel (projekti RITA1 KAUGSEIRE lõpparuanne), 46 lk. <https://doi.org/10.23673/re-255>
- Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H., Turner, W., 2014. Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology* 51, 839–848. <https://doi.org/10.1111/1365-2664.12261>
- Peuch, V.-H., and Coauthors, 2014. Monitoring Atmospheric Composition and Climate–Interim Implementation (MACC-II): Final report, 137 lk.
- Peuch, V.-H., Engelen, R., Rixen, M., Dee, D., Flemming, J., Suttie, M., Ades, M., Agustí-Panareda, A., Ananasso, C., Andersson, E., Armstrong, D., Barré, J., Bousserez, N., Dominguez, J.J., Garrigues, S., Inness, A., Jones, L., Kipling, Z., Letertre-Danczak, J., Parrington, M., Razinger, M., Ribas, R., Vermoote, S., Yang, X., Simmons, A., Marcilla, J.G. de, Thépaut, J.-N., 2022. The Copernicus Atmosphere Monitoring Service: From Research to Operations. *Bulletin of the American Meteorological Society* 103, E2650–E2668. <https://doi.org/10.1175/BAMS-D-21-0314.1>
- Rhodes, M.W., Bennie, J.J., Spalding, A., French-Constant, R.H., Maclean, I.M.D., 2022. Recent advances in the remote sensing of insects. *Biological Reviews* 97, 343–360. <https://doi.org/10.1111/brv.12802>
- Rychlý, P., Kilgarriff, A., 2007. An efficient algorithm for building a distributional thesaurus (and other Sketch Engine developments), *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, Czech Republic*, 41–44. <https://doi.org/10.3115/1557769.1557783>
- Sageder, D., 2010. Terminology today: a science, an art or a practice? : some aspects on terminology and its development. *Brno Studies in English* 36, 123–134.
- Stefanowitsch, A., 2020. Corpus linguistics A guide to the methodology, Textbooks in Language Sciences 7. Language Science Press, Berlin, 490 lk.
- Storjohann, D.P., 2005. Corpus-driven vs. corpus-based approach to the study of relational patterns. *Presented at the Corpus Linguistics Conference 2005, Birmingham*, p. 20.

Tavast, A., Langemets, M., Kallas, J., Koppel, K., 2018. Unified Data Modelling for Presenting Lexical Data: The Case of EKILEX. *Proceedings of the XVIII EURALEX International Congress EURALEX*, 749–761.

Tavast et al., 2021. Towards the Superdictionary: Layers, Tools and Unidirectional Meaning Relations, *Proceedings of the EURALEX XIX Lexicography for Inclusion Congress*, 215–223.

Voormansik, K., Zalite, K., Sünter, I., Tamm, T., Koppel, K., Verro, T., Brauns, A., Jakovels, D., Praks, J., 2020. Separability of Mowing and Ploughing Events on Short Temporal Baseline Sentinel-1 Coherence Time Series. *Remote Sensing* 12, 3784.
<https://doi.org/10.3390/rs12223784>

Zellweger, F., De Frenne, P., Lenoir, J., Rocchini, D., Coomes, D., 2019. Advances in Microclimate Ecology Arising from Remote Sensing. *Trends in Ecology & Evolution* 34, 327–341. <https://doi.org/10.1016/j.tree.2018.12.012>

Resümee

The last few decades have witnessed a significant increase in the use of remote sensing technology in Estonia. Remote sensing, which used to be a topic spoken mainly in English and mainly by scientists and experts, is now widely used also by Estonian-speaking non-experts. As a result, there is a need for remote sensing termbase for Estonian that can help make terms less ambiguous.

The aim of this thesis was to systematize remote sensing terms in Estonian. For this purpose, remote sensing corpora were built and used to compile the Estonian termbase for remote sensing. I created four corpora in the remote sensing field – in Estonian, English and two parallel English-Estonian-English corpora. For corpora creation from scratch, I used a Corpus Query System Sketch Engine. The term extraction module was used to find significant remote sensing terms and variants of preliminarily chosen terms. I used GDEX configuration for Estonian to find sentences for contexts from Estonian Remote Sensing Corpus 2022. Although our main corpus, Estonian Remote Sensing Corpus 2022, was smaller than suggested by Kallas et al., 2017, it gave a good base for compiling the Estonian Remote Sensing Termbase.

In order to get experts' assessment, ten remote sensing experts from UT Tartu Observatory were asked to fill in the questionnaire about remote sensing terms that have obscure meanings, variants and inconsistency in usage.

A dictionary writing system Ekilex was used to create an Estonian Remote Sensing Termbase, which includes 100 terms. It contains definitions in Estonian and English, explanations, contexts, variants, and occasionally the status of the term is indicated. All termbases created in Ekilex are publicly available through the Estonian dictionary portal Sõnaveeb¹⁴.

It is important to note that texts in corpora are as people have written them. No extra checking has been done. It means there may be some spelling mistakes. As our uploaded

¹⁴ <https://sonaveeb.ee/?lang=en>

documents and most results from web crawling are mainly official or journalistic documents, we can assume that the noisiness is not very high.

Finally, it is essential to emphasize that corpora are not replacing expert knowledge in the creation process of termbases. A corpora-based approach should supplement an expert-based approach. Most of the terms still need expert consulting. The future perspectives involve adding new terms to Remote Sensing Termbase in Ekilex and revising present terms according to the users' feedback. Also, we plan to make Estonian Remote Sensing Corpus 2022 publicly available using the Corpus Query System KORP¹⁵. It should be made available for end-users in the future through Sõnaveeb.

¹⁵ [KORP](#)

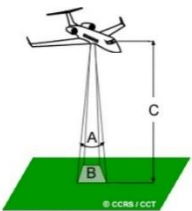
Kaugseire terminite loend

Kaugseire terminibaasi eestikeelsete ja ingliskeelsete terminite loetelu koos variantide, definitsioonide, seletuste ja näidetega. Mõistekirje koostamise nõuete järgi algavad definitsioonid väikse tähega ning lõpus ei ole punkti. Tabel on terminibaasi (25.05.23 seis) lühendatud versioon (pole näidatud rohkem kui ühte selgitust ega näidet; näidetel puuduvad viited; osad selgitused on lühendatud). Kõikide kirjete täisversioonid on kättesaadavad Sõnaveebis¹⁶.

ET termin	variandid	definitsioon	seletus	näited	EN terminid	definitsioon	seletus	näited
ajaline lahutusvõime	mõõdistamise ajasammu pikkus	aeg, mis jääb kahe kordusmõõdistamise vahele (Geoinformaatika 2019)	Tuleb silmas pidada, et kaugseire ajaline lahutusvõime ei pruugi olla sama, mis aegruumilise rea tikslis (aegruulise rea ühte rakslit nimetatakse tikslis) ajasamm, sest skaneeritud kujutise kvaliteedi parandamise nimel võidakse lõpptulemus, nt kommertskasutusse minev satelliidipilt, panna eeltöötuluse käigus kokku mitmest eri ajal tehtud kujutisest. (Geoinformaatika 2019)	Kuna paljud maakattealased uuringud lähtuvad satelliidiandmetest erineval viisil saadud (digitaal)kaartidest, siis kolmandaks oluliseks aspektiks on nende mõõtkava ja generalisatsiooni aste ning lähteandmete ruumiline ja ajaline lahutusvõime.	temporal resolution	time between the sensor can capture the images for a specific ground area (Research methods in RS, 2013)		The temporal resolution is an important consideration for a number of monitoring researches, especially when frequent imaging is required.
andmete ja teabele juurdepääsu teenus	DIAS	viis pilvepõhist platvormi, mis võimaldavad tsentraliseeritud juurdepääsu Copernicuse programmi andmetele ja teabele ning töötlusvahenditele (ELi terminibaas IATE, Copernicus DIAS et, kaugseire ekspert)	Andmetele juurdepääsu lihtsustamiseks ja ühtlustamiseks on Euroopa Komisjon rahastanud viie pilvepõhise platvormi kasutuselevõttu. Need platvormid võimaldavad tsentraliseeritud juurdepääsu Copernicuse programmi andmetele ja teabele ning töötlusvahenditele. Neid platvorme tuntakse nime all DIAS (Data and Information Access Services – andmetele ja teabele juurdepääsu teenus). (Copernicus DIAS et)		DIAS	Data and Information Access Services (Copernicus DIAS)	They provide centralised access to Copernicus data and information, as well as to processing tools. These platforms are known as the DIAS, or Data and Information Access Services. (Copernicus DIAS)	The five DIAS online platforms allow users to discover, manipulate, process and download Copernicus data and information.
atmosfääri-korreksioon		kaugseire meetodil mõõdetud aluspinna peegeldumisparameetritest atmosfääri mõju eemaldamine (kaugseire ekspert)	Atmosfääri mõju eemaldamisel võetakse arvesse atmosfääri gaasides ja aerosoolides toimuvaid hajumise ja neeldumise efekte. (kaugseire ekspert)	Maailmamere ookeaniuuringutes on taoline lähenemine eriti tavaline ja ka tulemused on head, sest sealne atmosfäärikorreksioon on täpsem kui ranniku- ja siseveekogudel.	atmospheric correction	a process in which the scattering and absorption effects from the atmosphere are removed to obtain the surface reflectance characterizing (surface properties) (GISGeography 2022)		By comparing the two brightness temperatures, the atmospheric correction can be estimated and eliminated.

¹⁶ <https://sonaveeb.ee/ds/kas/>

ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited
atmosfääri-mudel		dünaamikavõrranditel põhinev atmosfääri liikumise matemaatiline mudel (kaugseire ekspert)	Kasutatavad diferentsiaalvõrrandid on mittelineaarsed ning neid lahendatakse numbriliselt. (kaugseire ekspert)	Tugimõõtmiste süsteemi rakendamisel satelliidisensorite andmete on oluline osa atmosfäärimudelil, mis arvestab kiirguse levimist maapinnale ja tagasi satelliidisensorile.	atmospheric model	a mathematical model constructed around the full set of primitive dynamical equations which govern atmospheric motions (Wikipedia. Atmospheric model)		Atmospheric and meteorological data gathered at the time of the flight, at the surface, or using a radiosonde, can improve the atmospheric model.
DIAS	andmete ja teabe juurdepääsu teenus	viis pilvepõhist platvormi, mis võimaldavad tsentraliseeritud juurdepääsu Copernicuse programmi andmete ja teabe ning töötlusvahenditele (ELi terminibaas IATE, Copernicus DIAS et, kaugseire ekspert)	Andmete juurdepääsu lihtsustamiseks ja ühtlustamiseks on Euroopa Komisjon rahastanud viie pilvepõhise platvormi kasutuselevõttu. Need platvormid võimaldavad tsentraliseeritud juurdepääsu Copernicuse programmi andmete ja teabe ning töötlusvahenditele. Neid platvorme tuntakse nime all DIAS (Data and Information Access Services – andmete ja teabe juurdepääsu teenus). (Copernicus DIAS et)		DIAS	Data and Information Access Services (Copernicus DIAS)	To facilitate and standardise access to data, the European Commission has funded the deployment of five cloud-based platforms. They provide centralised access to Copernicus data and information, as well as to processing tools. These platforms are known as the DIAS, or Data and Information Access Services. (Copernicus DIAS)	The five DIAS online platforms allow users to discover, manipulate, process and download Copernicus data and information.
geoviidatud	geoviidetega varustatud	digitaalse kaardi või aerofoto koordinaatide seotus maapealsete geograafiliste koordinaatidega (kaugseire ekspert)			geo-referenced	the internal coordinate system of a map or aerial photo image is related to a geographic coordinate system (Wikipedia. Georeferencing)		One commonly used method to fuse co-registered images from multiple sources is to skip the third step and directly display the georeferenced images as a colour composite.
geoviidetega varustamine	geoviitamine	digitaalse kaardi või aerofoto koordinaatide sidumine maapealsete geograafiliste koordinaatidega (kaugseire ekspert)			geo-referencing	aligning geographic data to a known coordinate system so it can be viewed, queried, and analysed with other geographic data (Sommer & Wade 2006)		Direct georeferencing solves a large part of the image rectification problem, but not all of it.

ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited
geoviidetega varustatud	geoviidatud	digitaalse kaardi või aerofoto koordinaatide seotus maapealsete geograafiliste koordinaatidega (kaugseire ekspert)			geo-referenced	the internal coordinate system of a map or aerial photo image is related to a geographic coordinate system (Wikipedia. Georeferencing)		One commonly used method is to skip the third step and directly display the georeferenced images as a colour composite.
geoviitamine	geoviidetega varustamine	digitaalse kaardi või aerofoto koordinaatide sidumine maapealsete geograafiliste koordinaatidega (kaugseire ekspert)			geo-referencing	aligning geographic data to a known coordinate system so it can be viewed, queried, and analysed with other geographic data (Sommer & Wade 2006)		Direct georeferencing solves a large part of the image rectification problem, but not all of it.
heleduskoefitsient	heleduskordaja peegelduskoefitsient peegeldustegur	objektilt peegeldunud ja sellele langenud kiirgusvoo suhe (kaugseire ekspert)	Peegeldustegur kirjeldab objekti omadusi, kirkus sõltub nii objektist kui valgustustingimustest. (kaugseire ekspert)	Teise erinevusena on näha lokaalset maksimumi tsüanobakterite heleduskoefitsiendi spektris 650 nm ümbruses.	reflectance	proportion of perpendicularly incident light reflected from a component compared to that reflected from a standard (Encyclopedia of Analytical Science)		Invariant targets are features which have constant reflectance over time.
heleduskordaja	heleduskoefitsient peegelduskoefitsient peegeldustegur	objektilt peegeldunud ja sellele langenud kiirgusvoo suhe (kaugseire ekspert)	Peegeldustegur kirjeldab objekti omadusi, kirkus sõltub nii objektist kui valgustustingimustest. (kaugseire ekspert)	Samasse ajavahemikku jääb ka roostike heleduskordaja sesoonne maksimum spektri lähisinfra punases piirkonnas.	reflectance	proportion of perpendicularly incident light reflected from a component compared to that reflected from a standard (Encyclopedia of Analytical Science)		Invariant targets are features which have constant reflectance over time.
hetkvaateväli		koonuse ruuminurk, mida sensor näeb. Hetkvaateväli määrab Maa pinna osa, mida sensor „näeb“ etteantud kõrguselt antud ajahetkel (kaugseire ekspert, Fundamentals of remote sensing)		Linnas, kus aluspinnalt peegelduv kiirgus varieerub tunduvalt enamike satelliitide sensorite hetkvaatevälja piires, ei sarnane üksiku piksli spektraalne peegeldumine üksiku maakatteklassi peegeldumisega.	instantaneous field of view IFOV	the solid angle cone of visibility of the sensor. It determines the area on the Earth's surface which is "seen" from a given altitude at one particular moment in time (kaugseire ekspert, Fundamentals of remote sensing)		The signal level of the reflected energy increases if the signal is collected over a larger instantaneous field of view (IFOV).

ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited
hüper-spektraalne seade		suure spektraalse lahutusvõimega seade, mille eesmärk on saada valitud spektrivahemikus iga piksi kohta pidev spektraalne heleduskõver (Geoinformaatika 2019)		Eraettevõtetel on plaane lasta orbiidile terve parv mikrosatelliite, mille pardal on hüperspektraalsed kaamerad.	hyper-spectral instrument	high spectral resolution device designed to produce a continuous spectral luminance curve for each pixel in the selected spectrum range (Geoinformaatika 2019)		When selecting instrument for hyperspectral application there are many functional and non-functional quality attributes to consider.
kanal	spektrikanal spektraal-kanal riba spektririba spektraalriba	lainepikkuste vahemik elektromagnetkiirguse spektris (kosmosetehnoloogia ekspert)		Erinevad sensorid salvestavad kanalitesse veidi erineva lainepikkusega väärtusi, kuid alati on need kindlaks määratud vahemikud.	spectral band band	wavelength range in the electromagnetic spectrum (kosmosetehnoloogia ekspert)		The few spectral sensors from the earliest missions have been upgraded to hyperspectral sensors with hundreds of spectral bands.
keskkonna-satelliit	Maa kaugseire satelliit	satelliit, mis tuvastab ja vaatleb erinevaid Maa atmosfääri, maapinna ja ookeani karakteristikuid (kaugseire ekspert, Environmental Satellites)		Euroopa riigid töötasid kas Euroopa Kosmoseagentuuri (ESA) või riiklike kosmoseagentuuride kaudu välja ka keskkonnasatelliitide katsekemparid.	environ-mental satellite Earth observation satellite	satellites that detect and observe different characteristics and features of the Earth's atmosphere, land surface, and ocean (Environmental Satellites)		In aviation safety, measurements and predictions from our weather and environmental satellites are being integrated with other traditional aviation weather information.
kosmose-andmed		satelliitide kogutud andmed ja pildid (ELi terminibaas IATE)		Copernicus on maailma suurim kosmoseandmete pakkuja, mis toodab praegu 12 terabaiti päevas.	satellite data space data	information and images collected by satellites (ELi terminibaas IATE)		The water spread boundary captured by the satellite data provides water spread contour at that particular reservoir water level.
ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited

käsilidar		lidar käest mõõtmiseks (kaugseire ekspert)			handheld lidar	lidar for hand measurements (kaugseire ekspert)		PX-80 handheld lidar is harmless to the eyes.
läbi-paistmatus		elektromagnetilise või muu kiirguse, eriti nähtava valguse läbimatus mõõt (Wikipedia. Opacity)	Kui läbipaistmatus on 1 või 100%, siis ei paista objekt üldse läbi ja kui 0, siis on täiesti läbipaistev. (kaugseire ekspert)		opacity	measure of impenetrability to electromagnetic or other kinds of radiation, especially visible light (Wikipedia. Opacity)		Let's change the upper layer opacity to 50% and save the parameters.
maajaam		seadmestik satelliitidelt ja kosmosesondidelt informatsiooni vastuvõtuks (ka juhtimiseks) maapinnal (Kosmoseterminoloogia sõnastik)		Loodav taristu koosneb maajaamadest, mis asuvad Eestis või Eesti andmesaatkondade juures, et vahendada satelliitsideteenusega omavahelist strateegilist tagavarasidet.	Earth station	collection of equipment installed on the earth's surface that enables communications over one or more satellites (Gartner Glossary)		Earth stations are required to detect small signals so the control of the noise parameters is important.
multi-spektraalne seade		väikese spektraalse lahutusvõimega seade, mis jagab vaadeldava spektriosa erinevaks 3–7 kanaliks (Geoinformaatika 2019)	Multispektraalsete kanalite lainepikkused valitakse eesmärgiga esile tuua aluspindade kindlaid omadusi. (Geoinformaatika 2019)	Sentinel-2 on optiline multispektraalne spektromeeter, mille andmeid kasutatakse nii mere kui maismaa seireks.	multi-spectral instrument			
möödistamise ajasammu pikkus	ajaline lahutusvõime	aeg, mis jääb kahe kordusmöödistamise vahele (Geoinformaatika 2019)	Tuleb silmas pidada, et kaugseire ajaline lahutusvõime ei pruugi olla sama, mis aegruumilise rea tikslis (aegruulise rea ühte rakslit nimetatakse tikslis) ajasamm, sest skaneeritud kujutise kvaliteedi parandamise nimel võidakse lõpptulemus, nt kommertskasutusse minev satelliidipilt, panna eeltöötuse käigus kokku mitmest eri ajal tehtud kujutisest. (Geoinformaatika 2019)	Ajaline lahutusvõime (ingl temporal resolution) e möödistamise ajasammu pikkus näitab aega, mis jääb kahe kordusmöödistamise vahele.	temporal resolution	time between the sensor can capture the images for a specific ground area (Research methods in RS, 2013)		The temporal resolution is an important consideration for a number of monitoring researches, especially when frequent imaging is required.
märsilidar		kotis kaasas kantav lidar (kaugseire ekspert)	Märsilidar on lihtsalt kaasas kantav, kuid väiksema täpsusega kui statsionaarsed lidarid. Suurimaks vigade allikaks on kandja asukohta ja seadme enda kalde- ja pöördnurkade määramine. (kaugseire ekspert)		backpack lidar			The backpack LiDAR solution is a unique attempt for effective and accurate forest inventory and has important implications for forest inventory methods.

ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited
optiline kaugseire		kaugseire, mis kasutab nähtavat, lähisinfra-puna ja lühilaine infra-puna sensoreid, et saada pilte ja punktmõõtmisi maapinnalt peegelduva päikesekiirguse tuvastamise abil (kaugseire ekspert)		Optilise kaugseire meetodika väljatöötamiseks on loodud mitmesugust spetsiaalaparatuuri maapealseteks ja lennukimõõtmisteks.	optical remote sensing	remote sensing that uses visible, NIR and short-wave infrared sensors to form images of the earth's surface by detecting the solar radiation reflected from targets on the ground (Research methods in RS, 2013)	Optical remote sensing is performed within the optical region (0,3–3,0 µm), thermal remote sensing uses the thermal region (3,0–5,0 and 8,0–14,0 µm), and microwave remote sensing is conducted within the microwave region (1 mm–1 m). (Research methods in RS, 2013)	Optical remote sensing data are perhaps most widely used because of their simple nature and easier analysis techniques.
optiline spektri-piirkond		nähtava valguse, lähisinfra-puna ja lühilainelise infra-puna piirkond (0,3–3,0 µm) (Research methods in RS, 2013, kaugseire ekspert)		Copernicuse programmi Sentinel-seeria satelliitidest on Sentinel-2 ja Sentinel-3 pardal optilise piirkonna spektroradiomeetrid, mis seiravad maad ja vett nähtavas ja lähisinfra-puna spektripiirkonnas.	optical region optical spectrum	visible, NIR and short-wave infrared regions (Research methods in RS, 2013)		The low reflectance in the optical region is largely governed by the presence of pigments.
peegeldumisspekter	peegeldusspekter	objektilt peegeldunud ja sellele langenud kiirgusvoo suhe kindlal lainepikkusel (kaugseire ekspert)		Kui vee peegeldumisspekter ei olnud klassikaline Eesti veekogudes mõõdetav peegeldumisspekter ja võis kahtlustada kas taimestiku või kalda mõju, siis jäeti piksli veetüüp määramata.	spectral reflectance	the ratio of reflected surface radiation to received surface radiation at selected wavelengths (kaugseire ekspert)		The spectral reflectance of vegetation changes according to the structure and health of a plant.
peegelduskoefitsient	heleduskordaja heleduskoefitsient peegeldustegur	objektilt peegeldunud ja sellele langenud kiirgusvoo suhe (kaugseire ekspert)	Peegeldustegur kirjeldab objekti omadusi, kirkus sõltub nii objektist kui valgustustingimustest. (kaugseire ekspert)	Kui valgusallika ja tausta heledused oleksid kalibreeritavad (näiteks 100% peegelduskoefitsiendiga valge ketas 50% peegelduskoefitsiendiga hallil taustal), siis oleks võimalik korrigeeritud pildi järgi määrata PSF-i väärtust suurte kauguste korral.	reflectance	proportion of perpendicularly incident light reflected from a component compared to that reflected from a standard (Encyclopedia of Analytical Science)		Invariant targets are features which have constant reflectance over time.
ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited

peegeldus-spekter	peegeldumis-spekter	objektilt peegeldunud ja sellele langenud kiirgusvoo suhe kindlal lainepikkusel (kaugseire ekspert)		Tuleb leida baasfunktsioonide hulk, mis kirjeldab peegeldusspektrite 131 varieeruvust, ning leida iga puistu jaoks selle baasfunktsioonide kaalud.	spectral reflectance	the ratio of reflected surface radiation to received surface radiation at selected wavelengths (kaugseire ekspert)		The spectral reflectance of vegetation changes according to the structure and health of a plant.
peegeldus-tegur	heleduskordaja heleduskoefitsient peegelduskoefitsient	objektilt peegeldunud ja sellele langenud kiirgusvoo suhe (kaugseire ekspert)	Peegeldustegur kirjeldab objekti omadusi, kirkus sõltub nii objektist kui valgustustingimustest. (kaugseire ekspert)	Vees teostatud peegeldusteguri mõõtmiste ja optiliselt aktiivsete ainete vahel leitud algne korrelatsioon ning algortimidega saadud tulemused on samas suurusjärgus.	reflectance	proportion of perpendicularly incident light reflected from a component compared to that reflected from a standard (Encyclopedia of Analytical Science)		Invariant targets are features which have constant reflectance over time.
radar-kaugseire		kaugseire tüüp, kus radarsensorid saadavad välja kiirgust raadiokiirguse lainepikkustel (s.o umbes 1 cm kuni mitu meetrit). Väljasaadetud impulsi aluspinnalt peegeldunud kiirguse parameetreid kasutatakse Maapinna omaduste määramiseks (Encyclopedia of Life Support Systems)		Radarkaugseire sobib hästi hoonestatud alade globaalseks kaardistamiseks, sest võimaldab kiirelt uuendada andmeid maailma eri piirkondadest ning tagab andmete võrreldavuse.	radar remote sensing	remote sensing type where radar sensors transmit radiation at radio wavelengths (i.e. from around 1 cm to several meters) and use the measured return to infer properties of the earth's surface (kaugseire ekspert, Encyclopedia of Life Support Systems)		Another important feature of radar remote sensing is the penetration depth, respectively the penetration of loose sediments and vegetation in relation to the used wavelength.
radiomeetri-line kalibreerimine		kalibreerimine (elektromagnet-)kiirgust iseloomustavate radiomeetriliste suuruste suhtes (kosmosetehnoloogia ekspert)	Kalibreeritakse mõõtevahendit hierarhias kõrgemal asuva etaloni (mõõtevahendi) suhtes, iga suuruse jaoks oma spetsiifiline jälgitavuse skeem. (kosmosetehnoloogia ekspert)	Peale radiomeetrilise kalibreerimise tuleb spektraalse kirkuse ja kiiritustiheduse sensoreid iseloomustada ka teiste parameetrite osas.	radiometric calibration	calibration of radiometric parameters of electromagnetic radiation (kosmosetehnoloogia ekspert)		Image preprocessing includes restoration of bad lines; geometric rectification; radiometric calibration ; and atmospheric and topographic corrections.
ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited

radiomeetri- line lahutusvõime		väikseim heleduse erinevus, mida skaneerimisseade on suuteline mõõtma (Geoinformaatika 2019)	Seda näitajat kirjeldatakse eristatavate heldusastmete arvuga. (Geoinformaatika 2019)	Satelliidi sensoritel on erinev spektraalne ja radiomeetriline lahutusvõime.	radio- metrical resolution	the sensitivity of a remote sensing detector to differences in signal strength as it records the radiation flux reflected or emitted from the terrain (Research methods in RS, 2013)		However, higher radiometric resolution does not always mean a higher quality image.
riba	kanal spektrikanal spektraal- kanal spektririba spektraalriba	lainepikkuste vahemik elektromagnetkiirguse spektris (kosmosetehnoloogia ekspert)		Superspektraalne sensor suudab pildistada kümnetel erinevatel lainepikkuste ribadel, mis on kitsamad kui multispektraalse sensori lainepikkuste vahemikud.	spectral band band	wavelength range in the electromagnetic spectrum (kosmosetehnoloogia ekspert)		The few spectral sensors from the earliest missions have been upgraded to hyperspectral sensors with hundreds of spectral bands.
ruumiline lahutusvõime		väikseima eristatava objekti suurus (Fundamentals of remote sensing)	Rasterpildidel võib ruumilist lahutusvõimet üldjuhul vaadelda kui vastavat ala maapinnal, mida piksel katab. Geomeetrilise korrektsiooni käigus võib tulemuseks saadud andmetes ruumiline lahutusvõime muutuda. Kaugseires sõltub tulemuseks saadud andmestiku ruumiline lahutusvõime paljudest teguritest. (kaugseire ekspert)	Näidati, et nüüdisaegsete, hea ruumilise lahutusvõimega satelliidiandmete abil on võimalik teha kindlaks Läänemere puhtaveelisemad alad ning veeõitsengud ja lahustunud orgaanilise aine suurenenud kontsentratsioonid.	spatial resolution	size of the smallest possible feature that can be detected (Fundamentals of remote sensing)		If a sensor has a spatial resolution of 20 metres and an image from that sensor is displayed at full resolution, each pixel represents an area of 20m × 20m on the ground.
satelliidi ülelend		sündmus, mille käigus satelliidi instrumendi vaateväli läheb üle huvipakkuva ala (kaugseire ekspert, ESA 2022)		Selleks, et andmeid valideerida, tuli satelliidi ülelennu ajal teostada satelliitide trajektooridel paadi külge kinnitatud GNSS seadmega mõõtmised.	satellite pass satellite overpass	event in which the field of view of a satellite's instrument passes over an area of interest (ESA 2022, kaugseire ekspert)		If the orbit is also sunsynchronous, the ascending pass is most likely on the shadowed side of the Earth while the descending pass is on the sunlit side.
segupiksel		pikslid, milles on esindatud rohkem kui üks maakattetüüpidest (Kaidi 2013)		Väga soovitav on kasutada ka põllugeomeetriate sisse poole puhverdamist, et vältida segupikslite kasutamist statistikute arvutamisel.	mixed pixel	pixels that contain number of cover types (Kaidi 2013)		Normally a sizable area should avoid mixed pixels (mixed land cover).
ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited

spektraal-kanal	kanal spektrikanal riba spektririba spektraalriba	lainepikkuste vahemik elektromagnetkiirguse spektris (kosmosetehnoloogia ekspert)		Männikus on spektraalkanalite erinevate süstemaatiliste vigade tulemuseks kahemodaalne jaotus, kui kõigi spektraalkanalite suhtelised vead on koondatud ühte massiivi.	spectral band band	wavelength range in the electromagnetic spectrum (kosmosetehnoloogia ekspert)		The few spectral sensors from the earliest missions have been upgraded to hyperspectral sensors with hundreds of spectral bands.
spektraal-mõõtmine		kiirguse (kiiritustiheduse) mõõtmine erinevatel lainepikkustel (kaugseire ekspert)			spectral measurement	measurement of radiation (radiation density) at different wavelengths (kaugseire ekspert)		If spectral measurements have been made of ground materials in the scene, the images can be rescaled to represent actual reflectance values (spectral calibration).
spektraalne kirkus		kiirgustugevus kiirguri ühikuliselt pinnalt ruuminurga ja lainepikkuse kohta, ühikuks on $Wm^{-2} \cdot sr^{-1} nm^{-1}$ (kaugseire ekspert)		Peale radiomeetrilise kalibreerimise tuleb spektraalse kirkuse ja kiiritustiheduse sensoreid iseloomustada ka teiste parameetrite osas.	spectral radiance	the radiance of a unit area of the emitter surface per solid angle and wavelength, the unit is $Wm^{-2} \cdot sr^{-1} nm^{-1}$ (kaugseire ekspert)		At shorter and shorter wavelengths, the spectral radiance does not continue to increase as would be implied by the Rayleigh-Jeans approximation.
spektraalne lahutusvõime		elektromagnetkiirguse lainepikkuse väikseim erinevus, mida on võimalik mõttekalt eristada (kosmosetehnoloogia ekspert)	Lahutusvõime on aktuaalne iga spektrit kirjeldava ühiku korral (kvandi energia - eV, pöördsentimeeter, Hz, jne). (kosmosetehnoloogia ekspert)	Suure spektraalse lahutusvõimega (hyperspectral) satelliidi kasutamine on avanud unikaalse võimaluse uurida detailsemalt sinivetikate massesinemist Läänemeres.	spectral resolution	measure of the smallest object that can be resolved by a given sensor, and is commonly expressed as the most closely spaced line pairs per unit distance that can be distinguished in an image. (Lubin & Massom 2006)	Spectral resolution is the sensitivity of a sensor to respond to a specific frequency range, where the discrete frequency/wavelength range that a sensor is able to detect and measure is called a band. (Lubin & Massom 2006)	High spectral resolution is achieved by narrow bandwidths which, collectively, are likely to provide a more accurate spectral signature for discrete objects than broad bandwidths.
ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited

spektraalne peegeldumisvõime		pinnalt peegeldunud kiirguse ja pinnale langenud kiirguse suhe valitud lainepikkustel (kaugseire ekspert)			spectral reflectance	the ratio of reflected surface radiation to received surface radiation at selected wavelengths (kaugseire ekspert)		The spectral reflectance of vegetation changes according to the structure and health of a plant.
spektraalriba	kanal spektrikanal spektraal-kanal riba spektririba	lainepikkuste vahemik elektromagnetkiirguse spektris (kosmosetehnoloogia ekspert)		Optilise kaugseire süsteem jaguneb pildi koostamisel kasutatavate spektraalribade arvu alusel nelja suuremasse klassi.	spectral band band	wavelength range in the electromagnetic spectrum (kosmosetehnoloogia ekspert)		The few spectral sensors from the earliest missions have been upgraded to hyperspectral sensors with hundreds of spectral bands.
spektrikanal	kanal spektraal-kanal riba spektririba spektraalriba	lainepikkuste vahemik elektromagnetkiirguse spektris (kosmosetehnoloogia ekspert)		Tema ruumiline lahutus on 300 m, kuid suurem spektrikanalite hulk annab parema võimaluse fütoplanktoni rühmade tuvastamiseks.	spectral band band	wavelength range in the electromagnetic spectrum (kosmosetehnoloogia ekspert)		The few spectral sensors from the earliest missions have been upgraded to hyperspectral sensors with hundreds of spectral bands.
spektririba	kanal spektrikanal spektraal-kanal riba spektraalriba	lainepikkuste vahemik elektromagnetkiirguse spektris (kosmosetehnoloogia ekspert)		Kõige kõrgem resolutsioon on kolmel nähtaval spektriribal (sinine, roheline ja punane).	spectral band band	wavelength range in the electromagnetic spectrum (kosmosetehnoloogia ekspert)		The few spectral sensors from the earliest missions have been upgraded to hyperspectral sensors with hundreds of spectral bands.
ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited

spektrisegu		segupiksli spekter, kus on korraga esindatud kaks või enam maapealse objekti klassi (nt rukkipõld, lehtmets, veekogu, jms) (kaugseire ekspert)		Linnades toimunud muutuste leidmiseks lahutati 2011. aasta spektrisegu lahutamise pilt 1985. aastast.	spectral mixture			
spektrisegu lahutamine		tegevus, mille eesmärgiks on segupiksli spekter lahutada komponentideks ja anda komponentidele kaal (nt pikslis on 50% rukkipõldu, 20% lehtmetsa, 20% veekogu ja 10% mingit määratlemata objekti klassi) (kaugseire ekspert)		Linnades toimunud muutuste leidmiseks lahutati 2011. aasta spektrisegu lahutamise pilt 1985. aastast.	spectral unmixing	the procedure by which the measured spectrum of a mixed pixel is decomposed into a collection of constituent spectra, or endmembers, and a set of corresponding fractions or abundances that indicate the proportion of each endmember present in the pixel (Keshava 2003)		Spectral unmixing strives at finding the relative or absolute fractions (or abundance) of a set of spectral components that together contribute to the observed reflectance of ground resolution cells.
spektri-vahe		kiirguse lainepikkuste vahemik (Eesti Kiirguskliima Teatmik)		RGB kaamerad on tavalised fotokaamerad, mis salvestab kolme spektrivahe (R, red ehk punane, G, green ehk roheline, B blue ehk sinine).	spectral range			
summaarne kirkus		kiirgustugevus kiirguri ühikuliselt pinnalt kõikides suundades, s.t ruuminurga 2π sr (sterradiaan) piires (kaugseire ekspert)		Kuna atmosfääri kohal mõõdetava summaarse kirkuse signaalist umbes 90% on põhjustatud atmosfäärist ning veest väljuva kiirguse kirkuse osakaal on isegi alla 10%, siis suhe 1:10 iseloomustab ka mõõtemääramatusi.	total radiance	the radiance of a unit area of the emitter surface per all directions, i.e. including all solid angles of 2π sr (steradian) (kaugseire ekspert)		This shows that the fraction of the total radiance between the wavelength limits increases with brightness temperature over this range.
ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited

tehisavaradar		radarisüsteem, mis korreleerib satelliidi orbiidi või õhusõiduki lennutee eri punktides emiteeritud signaalide kajasid (The Dictionary of Space Technology)	Eestikeelses kaugseire alases kirjanduses kasutatakse valdavalt terminit tehisavaradar (või tehisava-radar), kosmosetehnoloogia alases kirjanduses sünteesavaradar. (Kaugseire korpus 2022)	Kogu Soome poolt teostatav jääkaardistus Läänemeres on tehisavaradari piltide põhine, aerolaserskanneerimist jääseireks ei kasutata.	synthetic-aperture radar SAR	a radar system that correlates the echoes of signals emitted at different points along a satellite's orbit or an aircraft's flight path (The Dictionary of Space Technology)		Synthetic Aperture Radar (SAR) is an emerging technology in remote sensing.
tehisavaradari interferomeetria		kaugseire rakendus, kus kasutatakse interferomeetrilist protsessi vähemalt kahe erineval ajal tehtud tehisavaradari kujutisega (ESA sõnastik)	Tehisavaradari interferomeetrias genereeritakse vähemalt kahe erineval ajal tehtud tehisavaradari (SAR) kujutise abil nt maapinna deformatsiooni- või reljeefikaarte, kasutades vastuvõtjale naasvate elektromagnetlainete faasierinevusi. (Wikipedia InSAR)	Londonis, Crossrailis, teostati aastatel 2011 kuni 2013 Crossrail Ltd poolt seosteuuring kohapealse mõõdistamise ning tehisavaradari interferomeetria vahelmittelineaarsest pinnaliikumistest.	Synthetic Aperture Radar interferometry InSAR	an application of remote sensing where the interferometry process is used with at least two SAR images (ESA sõnastik)		Synthetic Aperture Radar interferometry (InSAR) is a particularly interesting tool whenever aiming at assessing ground deformation phenomena.
tugi-mõõtmised		maapealsed mõõtmised, millega võrreldakse satelliidiandmeid (kaugseire ekspert)	Erinevalt satelliitsensorist saab maapealset tugimõõtmisaparatuuri laboris perioodiliselt kalibreerida. (kaugseire ekspert) Tugimõõtmiste süsteemi usaldusväärsus sõltub eelkõige lokaalsete sensorite metrooloogilisest tasemest ja sobivusest ettenähtud rakendusteks, tugiparameetrite mõõtemetoditest, sensorite kalibreerimiseks kasutatavate laborite võimekusest ja tegelikust kalibreerimiste ning võrdlusemõõtmiste praktikast. (Kaugseire Eestis 2016)	Rünkpilvede tõttu võis spektraalne valgustus maapealsete tugimõõtmiste kohal ja mõõdetava puistu kohal olla erinev.	reference measurements	terrestrial measurements to which satellite data are compared (kaugseire ekspert)	Unlike to satellite sensors, terrestrial measuring instruments can be regularly calibrated. (kaugseire ekspert)	Ground reference measurements are usually difficult to obtain due to the medium/coarse resolution of the imagery.
ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited

valevärvikomposiit	valevärvipilt	tehiskult genereeritud värvipilt, millel sinine, roheline ja punane pannakse vastavusse teistsuguste lainepikkustega, kui nad on loomulikult nähtavad inimese silmale (kaugseire ekspert, Introduction to RS 2022)		Joonisel 3a on näidatud valevärvides komposiit SWI (Shortwave Infrared) ja joonisel 3b protsessori pilvemask.	false color image false color composite	artificially generated colour image in which blue, green and red colours are assigned to the wavelength regions to which they do not belong in nature (Introduction to RS 2022)	A false color image is used to reveal or enhance features otherwise invisible or poorly visible to a human eye. In other words, a false color composite is a multispectral image interpretation using the standard visual RGB band range (red, green, and blue). (EOS data analytics)	NearIR is usually assigned to red on the image; thus, vegetation often appears bright red in false color images, rather than green, because healthy vegetation reflects a lot of nearIR radiation.
valevärvipilt	valevärvi-komposiit	tehiskult genereeritud värvipilt, millel sinine, roheline ja punane pannakse vastavusse teistsuguste lainepikkustega, kui nad on loomulikult nähtavad inimese silmale (kaugseire ekspert, Introduction to RS 2022)		Valevärvi pildi tegemiseks kasutasin IDRISI moodulit COMPOSITE, milleks on mul vajalikud kolme spektripiirkonna pildid: roheline, punase ja lähisinfrapunase piirkonna pildid.	false color image false color composite	artificially generated colour image in which blue, green and red colours are assigned to the wavelength regions to which they do not belong in nature (Introduction to RS 2022)	A false color image is used to reveal or enhance features otherwise invisible or poorly visible to a human eye. In other words, a false color composite is a multispectral image interpretation using the standard visual RGB band range (red, green, and blue). (EOS data analytics)	NearIR is usually assigned to red on the image; thus, vegetation often appears bright red in false color images, rather than green, because healthy vegetation reflects a lot of nearIR radiation.
vikaarkalibreerimine		atmosfääri mõju arvestavate mudelite ebatäiuslikkusest tulenevate parandite määramine ja rakendamine satelliitkaugseire mõõteandmete, kasutades maapealsete (<i>in situ</i>) tugimõõtmiste andmeid (kosmosetehnoloogia ekspert)		Satelliidi ja maapealsete mõõteandmete kooskõla tagatakse nn vikaarkalibreerimiste süsteemi abil, mis seab atmosfääripealsed mõõtmised vastavusse atmosfäärialuste mõõtmistega.	vicarious calibration	identifying and applying the corrections (that are needed due to the imperfection of the atmospheric models) to satellite remote sensing data using in situ ground measurements. (kosmosetehnoloogia ekspert)		Vicarious calibration provides an indirect means of QA of remotely sensed data and sensor performance that is independent of direct calibration methods.
ET termin	variandid	definiitsioon	seletus	näited	EN terminid	definiitsioon	seletus	näited

õhust tehtav kaugseire		kaugseire, kus Maa kujutised saadakse õhusõidukitel (nt lennukid) asetsevate sensorite abil, mis võimaldavad koguda detailseid andmeid suvalisel ajahetkel (kaugseire ekspert, Airborne RS)	Õhust tehtava kaugseire platvormiks on tavaliselt tiibadega õhusõidukid, kuigi mõnikord kasutatakse ka helikoptereid. (Lang 2006)	Õhust tehtav kaugseire võimaldab saada üsnagi detailset informatsiooni praktiliselt suvalisel ajal suvalise Maa koha jaoks.	airborne remote sensing	remote sensing where downward or sideward looking sensors are mounted on an aircraft to obtain images of the earth's surface (Airborne RS)	An advantage of airborne remote sensing, compared to satellite remote sensing, is the capability of offering very high spatial resolution images (Airborne and Spaceborne Remote Sensing)	Historic England uses airborne remote sensing methods to identify record, illustrate and monitor the condition of a wide variety of heritage assets.
ülelennu sagedus		ajavahemik, mis jääb mingi kindla piirkonna satelliidivaatluste vahele (kaugseire ekspert)		Kombineerides Sentinel-3A ja Sentinel-3B andmeid, saab satelliidi ülelennu sageduseks 13,5 päeva.	revisit time revisit period	the time elapsed between observations of the same point on Earth by a satellite (Wikipedia. Revisit period)		Revisit time or temporal approaches are still limited by satellite orbital constraints and/or data download capacity.

Estonian Remote Sensing Termbase: data sources and implementation

Liisi Jakobson¹¹University of Tartu, Tartu, Estonia

To be submitted to the CLARIN 2023 Conference Proceedings

ABSTRACT

This article explores the process of creating a corpus-based termbase in countries and regions where less than a million people use the official or dominant language and there are not enough resources to create termbases for particular domains from scratch. We describe the process of creating an Estonian Remote Sensing Corpus 2022, which was used as a basis for the term extraction to compile an Estonian Remote Sensing Termbase. The Estonian Remote Sensing Corpus was built using the Corpus Query System Sketch Engine (Kilgarriff et al., 2004). As a next step, we identified 100 terms that were added to Estonian Remote Sensing Termbase in the dictionary writing system Ekilex (Tavast et al., 2018). For each term, definitions, variants and contexts were added. In the last stage, terms, their varia and usage contexts were edited by remote sensing experts. Estonian Remote Sensing Termbase is publicly available as a part of the dictionary portal Sõnaveeb¹⁷.

KEYWORDS

termbase, domain corpora, remote sensing, Dictionary Writing System, Corpus Query System, Estonian

1. Introduction

The last few decades have witnessed a significant increase in the use of remote sensing technology in Estonia (Noorma et al., 2020a). Due to the considerable amount of high-quality data freely available through the European Union program Copernicus, the Estonian public sector would like to use this data to improve public services (Noorma et al., 2020b). Also, the private sector has started using remote sensing data to develop their products (e.g. Voormansik et al., 2020; Dõmnich et al., 2021; Khoshkhah et al., 2022; Komisarenko et al., 2022). Remote sensing, which used to be a topic spoken mainly in English and mainly by scientists and experts, is now widely used also by Estonian-speaking non-experts. As the use of the Estonian language in the remote sensing field has increased, there is a need for remote sensing termbase for Estonian that can help to make the use of terms less ambiguous.

Traditionally dictionaries and termbases were created based on expert knowledge; at best, specialists and linguists worked together (Erelt, 2007). It needed lots of expert effort and time and was mainly based on intuition (Stefanowitsch, 2020), not statistics. Stefanowitsch (2020) brings out that the quality of corpus data is higher compared to the quality of intuited 'data' in terms of two aspects that are considered crucial in methodological discussions outside of linguistics: 1) data reliability (roughly, how sure can we be that other people will arrive at the same set of data using the same procedures); 2) data validity or epistemological status of the data (roughly, how well do we understand what real-world phenomenon the data correspond to) (Stefanowitsch, 2020).

Nowadays, more and more corpora are used to create dictionaries or termbases. There are two different approaches to using corpora in creating a termbase. The corpus-driven approach is a methodology that serves as an empirical basis from which language specialists extract their data and detect linguistic phenomena

¹⁷ sonaveeb.ee/ds/kas (20.03.2023)

without prior assumptions and expectations (Tognini-Bonelli, 2001). In general, corpus-based research relies on established linguistic forms and theory to conduct investigations; the primary goal is to analyze the systematic patterns of variants using predefined linguistic features (Biber, 2012). Storjohann (2005) argued that both the corpus-driven and corpus-based approaches could be complementary.

In Estonia, the creation of termbases is coordinated by the Institute of the Estonian language (see more Estonian terminology homepage¹⁸). The most extensive termbase, ESTERM (Vaus et al., 2019), is created by the Institute of the Estonian language and consists mainly of European Union and Estonian Republic terms from legal acts. The Institute develops Dictionary Writing System Ekilex¹⁹ (Tavast et al., 2018) for the compilation of general language dictionaries and termbases, which are mainly created by experts. Currently, there are 120 databases in very different fields, which is growing over time. Every year the Institute holds a contest to support new termbases or add terms to existing databases.

A corpus-based approach to creating an Estonian termbase in remote sensing is a significant and noteworthy event because the Estonian language has a small number of native speakers, and remote sensing is quite a new topic in the Estonian language. As the number of native speakers is small, also the amount of texts and corpora size is comparatively small. The size difference can be seen if we compare the size of TenTen Web Corpus Family in Corpus Query System Sketch Engine²⁰ (Kilgarriff et al., 2004) in English (enTenTen20 – 36 billion words) and in Estonian (etTenTen21 – 0,725 billion words). Domain corpora are much smaller than general ones. To create a corpus-based termbase, we need a sufficiently extensive domain corpus to apply the necessary corpus analysis (e.g. term extraction). According to Kallas et al. (2017), the sufficient amount of words in corpora for term extraction is 5 million words. We managed to gather only about 3 million words for our Estonian Remote Sensing Corpus 2022 (*Kaugseire korpus 2022*), although we included (almost) all available texts in Estonian in the remote sensing field. Though it is a considerable effort to create a particular corpus in a language which has a small number of native speakers, it is worth the effort. It creates a solid ground for the termbase and represents a valuable linguistics resource.

The aim of this article is to show the possibility of creating a termbase using a corpus-based approach, even if the language has a small number of native speakers and the topic is relatively new. The paper is organized as follows: in Chapter 1, we show the need for the creation of Remote Sensing Termbase for Estonian; in Chapter 2, we describe the process of the compilation of Estonian Remote Sensing Corpus 2022, term extraction and GDEX using Sketch Engine, and the cooperation with remote sensing experts; in Chapter 3 we describe the compilation of Remote Sensing Termbase and the use of corpus as a source for the termbase. Finally, we discuss the advantages and disadvantages of the whole process.

2. Estonian Remote Sensing Corpus 2022 compilation and term extraction using Sketch Engine

2.1. Corpus compilation from files and from the web

We created four corpora in the remote sensing field – in Estonian, English and two parallel English-Estonian corpora. For Estonian Remote Sensing Termbase, mainly Estonian Remote Sensing Corpus 2022 was used, and English and parallel corpora were used as supporting materials. Although the goal was to gain the size of

¹⁸ terminologia.ee

¹⁹ ekilex.ee

²⁰ sketchengine.eu

5 million words in every corpus (as suggested by Kallas et al., 2017), we managed it only in English; the Estonian corpus is about three million and parallel corpora about 0.5 million words.

For corpora creation from scratch, we used a Corpus Query System Sketch Engine (Kilgarriff et al., 2004), which supports approximately 100 languages (Kostka, 2022) and offers tools for users to create their corpus (Kilgarriff et al., 2014) either by uploading their texts²¹ or by building the corpus semi-automatically from the web²² according to the keywords given by the user (Fišer et al., 2016); Sketch Engine also supports the compilation of parallel corpora²³.

As a result, our Estonian Remote Sensing Corpus 2022 consists of 347 documents, 58% of which are uploaded and the rest are crawled from the web by providing seed words such as here: *kaugseire* (remote sensing), *spektraalne lahutusvõime* (spectral resolution), *spektraalmõõtmine* (spectral measurement), *multispektraalne seade* (multispectral instrument), *keskkonnasatelliit* (Earth observation satellite, environmental satellite), *andmete ja teabe juurdepääsu teenus* (DIAS), *ülelennu sagedus* (revisit time, revisit period).

Uploaded documents were primarily scientific articles, BA and MA theses, textbooks, study materials, lecture notes, and legal texts. Among the essential sources were five remote sensing conference article collections published in Estonian and issued by the UT Tartu Observatory²⁴. Texts from the web are mainly from the leading popular science journals, legal texts, texts about remote sensing enterprises, Estonian Agricultural Registers and Information Board, materials from the Estonian remote sensing community homepage, and materials from remote sensing scientific projects. There is no material from social media sources like user forums, blogs, comments etc. because this topic hasn't been covered much in Estonian social media.

2.2. Term extraction

Sketch Engine contains a keyword and terminology extraction module using a contrastive approach to find term candidates. The input for this task consists of two arbitrary corpora: a focus corpus from which the keywords should be extracted and a reference corpus that the term candidates from the focus corpus are contrasted (Kilgarriff et al., 2014; Jakubíček and Šmerk, 2016). We used the Estonian Remote Sensing Corpus 2022 as a focus corpus and the Estonian National Corpus 2021 (2,4 billion) (Koppel and Kallas, 2022b) as a reference corpus. To improve the accuracy of terminology extraction, Sketch Engine selects only grammatically valid phrases using the so-called term grammar²⁵ (Fišer et al., 2016). The currently used Estonian term grammar v2.0 is based on evidence-based term grammar design principles. It was developed by Marek Blahuš in consultation with Eleri Aedmaa and Merily Plado (Koppel and Kallas, 2022a). The analysis allows the identification of the most distinctive term candidates in the focus corpus (in our case, Estonian Remote Sensing Corpus 2022), in contrast to the reference corpus (in our case, Estonian National Corpus 2021). Term candidates are arranged on the basis of frequency in focus corpus. The first ten single-word and multi-word term candidates are shown in Table 1. Some term candidates are in English because many input texts have a summary (or some other parts) in English. All in all, 1000 term candidates for single-word terms and 1000 for multi-word terms were looked through. Around 150 were selected and further developed with the assistance of remote sensing experts.

²¹ [Sketch Engine: create a corpus from files](#)

²² [Sketch Engine: create a corpus from web](#)

²³ [Sketch Engine: parallel corpora](#)

²⁴ <https://kosmos.ut.ee/en>

²⁵ [Sketch Engine: term grammar](#)

Table 1. First ten results of single-word terms and multi-word terms in Sketch Engine terminology extraction module.

	Lemma	Frequency ²		Term	Frequency ²		
		Focus	Reference		Focus	Reference	
1	kaugseire	4,598	892	1	euroopa liidu teataja	2,748	1,728
2	copernicus	4,122	1,719	2	remote sensing	861	9
3	sensing	1,474	111	3	liidu teataja	2,208	1,663
4	landsat	1,108	39	4	füüsiline kliimarisk	655	6
5	satelliidipilt	1,476	355	5	ruumiline lahutus	528	33
6	kliimarisk	958	163	6	tartu observatoorium	1,082	1,058
7	esthub	861	62	7	maa kaugseire	455	11
8	sentinel-2	765	39	8	kaugseire andmed	422	53
9	l174	733	8	9	turvalisuse akrediteerimine	588	466
10	spektraalne	766	191	10	copernicuse programm	494	343

2.3. Cooperation with remote sensing experts

The terminology commission was a leading form of vocabulary development characteristic of the 20th century. In Estonia, the terminology commissions usually consist of four or five to ten domain experts and one or two language specialists who have expertise in the field of terminology creation (Erelt, 2007).

Currently, there is no remote sensing terminology commission in Estonia. In order to get experts' assessment, ten remote sensing experts from UT Tartu Observatory were asked to fill in the questionnaire about remote sensing terms that have obscure meanings, variants and inconsistency in usage. The questionnaire consisted of several parts. First, the experts were asked to assess the candidates for specific English terms identified as a result of the corpus analysis, and second, to propose which terms should be added to the base. The questionnaire results were discussed in the seminar held on 2 June 2022 in Tartu Observatory.

These terms, their use and variants were again checked in Remote Sensing Corpus 2022 in Sketch Engine. The example in Table 2 shows how diverse variants can be, e.g. for English term *reflectance*; we detected seven different Estonian terms.

Table 2. An example of how one term (*reflectance*) in English has been translated into Estonian in seven ways (in the remote sensing field).

The term in Estonian	Frequency in Estonian	Frequency per million
	Remote Sensing Corpus 2022	
heleduskordaja	242	61
heleduskoefitsient	158	40
heledustegur	12	3
peegeldustegur	246	62
peegeldumistegur	69	17
peegelduskoefitsient	15	4
peegeldumiskoefitsient	29	7

Remote sensing experts also suggested some new term candidates for terms that do not have equivalents in Estonian yet. These were not present in the corpus and were added directly to the termbase. For example, the English term 'handheld lidar' was suggested to translate as *käsilidar* and the term 'backpack lidar' as *märsilidar*.

The final list of terms was used as a headword list for the compilation of the Estonian Remote Sensing Termbase using the Dictionary Writing System Ekilex.

2.4. Finding usage examples/contexts using Good Dictionary Examples (GDEX)

In addition to term extraction, corpora can be used to find usage examples in different contexts. Contexts are crucial; a well-selected example of the use of the relevant term helps to understand the meaning of the term (Tavast and Taukar, 2013). Contexts definitely must be provided with references, which is the main difference compared to examples. According to Erelt (2007), the context has a narrow meaning – a part of the text which makes it possible to understand the meaning of a word (term); and a broader meaning – other units neighboring the language unit, i.e. phoneme, morpheme, word, a compound of words, sentence. The guidelines for Ekilex²⁶ state that context clarifies or complements the definition if necessary; the context field also contains the term itself.

Selecting examples is a great challenge, not only because they need to find examples that meet the criteria of a good dictionary example (criteria may differ depending on the target users) but also because the sources of examples are getting larger and larger (Kosem et al., 2019). This is why during the last decade, much research and development has been put into language technologies that can assist in finding good dictionary examples.

One such technology is the GDEX tool in Sketch Engine (Kilgarriff et al., 2008), which ranks corpus sentences according to predefined criteria, assigning a numerical score to each sentence. This mechanism can be seen as a filter, as it helps lexicographers work with more relevant citations even though they have not been manually annotated (Koppel et al., 2019b). GDEX configurations have been developed for several languages (see, e.g. Kosem et al., 2019). The first version of the GDEX configuration for Estonia was developed in 2014. It was used for extracting example sentences into the Estonian Collocations Dictionary (ECD) database (Kallas et al., 2015). The latest version of the GDEX for Estonian (GDEX 1.4) was used to compile the Estonian Corpus for Learners 2018 (etSkELL), which is used in both the etSkELL interface and in the language portal Sõnaveeb (Koppel et al., 2019b). In 2019, the Estonian GDEX configuration was evaluated by Estonian language learners at the B2-C1 proficiency levels, and 85% of the GDEX-selected examples were actually rated as suitable dictionary examples (Koppel et al., 2019a).

We used GDEX configuration for Estonian to find sentences for contexts from Estonian Remote Sensing Corpus 2022. Most of our contexts in the Remote Sensing Termbase in Ekilex are from the Estonian Remote Sensing Corpus 2022.

3. The compilation of Remote Sensing Termbase using Ekilex

Estonian Remote Sensing Corpus 2022 and Sketch Engine functionality described in Chapter 2 were used to create the Estonian Remote Sensing Termbase (*Kaugseire terminibaas*). We used a Dictionary Writing System Ekilex (Tavast et al., 2018, 2021) developed by the Institute of the Estonian Language to create the Estonian Remote Sensing Termbase. All termbases created in Ekilex are publicly available through the Estonian dictionary portal Sõnaveeb; every termbase has its homepage, e.g. Remote Sensing Termbase²⁷. A term entry in Ekilex is a structured unit that contains information about one term, e.g. Estonian and foreign language terms, variants, definitions, contexts, source references, notes and domain (Figure 1). The term entry is concept-based – terms in one item (e.g. preferred term, variant, abbreviation) are identical and represent only

²⁶ [Ekilex guidelines](#)

²⁷ [Remote Sensing Termbase in Sõnaveeb](#)

one term; one item = one term = one meaning (Eesti Keele Instituut²⁸). The Remote Sensing Termbase contains 100 terms.

Figure 1. A concept entry in a Dictionary Writing System Ekilex. Definitions in Estonian and English are in the upper box. There is also an explanation in English. An entry can have many explanations in both languages. The variants with contexts are shown in lower boxes. Each context has a reference source.

Definitions were one theme where expert help was needed most. Some of the definitions were still found in handbooks, but most of the definitions evolved based on definitions in English. Based on the principles of Vezzani et al. (2018), we tried to make definitions understandable not only to remote sensing experts but also to translators and ordinary end-users.

Variants in Ekilex can be either preferred, admitted, former, deprecated or new (Vaus, 2022). As the topic of our termbase is relatively new, only very few terms can be marked as former (e.g. a term *peegeldustegur* (reflectance) is substituting the term *heleduskordaja*; soon *peegeldustegur* will be probably marked as a former term). We do not have any deprecated terms yet.

4. Discussion and results

²⁸ [Eesti Keele Instituut](https://www.eki.ee/)

4.1. Advantages of a corpus-based approach

The practical advantages that corpora gave us while creating the Estonian Remote Sensing Termbase are the possibility:

- 1) to choose relevant terms (through term extraction) on the basis of frequency;
- 2) to detect all possible variants of the same term;
- 3) to analyze the usage context of variants. It helps to see in which domains the term is used. For example, for the English term 'reflectance', the water remote sensing domain may prefer the term *heleduskordaja* and remote sensing of vegetation, the term *peegeldustegur*. Based on our corpus, the distinction is actually not clear right now. Most often, the preference is strongly connected with the domains outside the remote sensing (in our case, e.g. metrology, meteorology and climatology etc.). E.g. the term *heledustegur* (reflectance) is described and marked as a preferred term in the metrology (the scientific study of measurement) termbase, but in the remote sensing field, it is not the preferred term (Table 2).
- 4) to distinguish old variants based on data from different times (Kilgarriff et al. 2015). We divided the data into two periods: 1993-2010; 2011-2022 and found that the term *peegeldustegur* (see Tabel 2) is significantly more used after 2010 and *heleduskordaja* (see Tabel 2) is significantly less used after 2010. We can say that *peegeldustegur* is substituting the term *heleduskordaja*.
- 5) to clarify the meaning based on contexts analyses;
- 6) to find contexts and definitions from trustworthy sources.

4.2. Challenges and possible solutions

It is important to note that texts in corpora are as people have written them. No extra checking has been done. It means there may be some spelling mistakes. As our uploaded documents and most results from web crawling are mainly official or journalistic documents, we can assume that the noisiness is not very high. Noisiness is characteristic of web texts, especially social media (Baldwin et al., 2013). Our web crawling results also don't have a big problem with another typical problem with web texts (Jakubíček et al., 2020) – machine translation texts. Still, we are struggling with the minimal set of metadata, which is also characteristic of web corpora (Jakubíček et al., 2020). In addition, the metadata of uploaded texts likewise needs more metadata. We may say that the texts in our corpus are quite reliable and with small noisiness, but the corpus lacks metadata. We also have to bear in mind that corpora are not static. The results show us the current situation with the given texts. It can slightly change if new texts are added.

Although our main corpus, Estonian Remote Sensing Corpus 2022, was smaller than suggested by Kallas et al., 2017, the list of term candidates extracted using the term extraction function in Sketch Engine was a valuable base for the compilation of Estonian Remote Sensing Termbase. Because of the small size of our Estonian Remote Sensing Corpus 2022, we considered the results with caution and relied more on results from remote sensing experts. The most valuable help from the corpus was the statistics about variants (see Table 2). It is vital to describe all the variants in the termbase at the same time to identify the preferred one. The preference is mainly based on statistics and also on experts' assessments. It helps users to pick the correct term (and they know what term can be replaced by the preferred one), as our aim is to have unambiguous terms.

5. Conclusions and future perspectives

Creating a corpus-based termbase for languages with a small number of native speakers is an important topic to be addressed. The Estonian Remote Sensing Corpus 2022 was created using the Corpus Query System Sketch Engine. The term extraction module was used to find significant remote sensing terms and variants of preliminarily chosen terms. One of the main advantages of corpus data is that corpus data enables one to find different variants of a term with their frequencies of occurrence, which is not possible with the experts' intuited data.

A dictionary writing system Ekilex was used to create an Estonian Remote Sensing Termbase, which includes 100 terms. It contains definitions in Estonian and English, explanations, contexts, variants, and occasionally the status of the term is indicated. Our work has proved that creating a termbase using a corpus-based approach is possible even if the language has a small number of native speakers and the topic is quite new.

Finally, it is essential to emphasize that corpora are not replacing expert knowledge in the creation process of termbases. A corpora-based approach should supplement an expert-based approach. Most of the terms still need expert consulting.

The future perspectives involve adding new terms to Remote Sensing Termbase in Ekilex and revising present terms according to the users' feedback. Also, we plan to make Estonian Remote Sensing Corpus 2022 publicly available using the Corpus Query System KORP²⁹. It should be made available for end-users in the future through Sõnaveeb.

²⁹ [KORP](#)

References

- Baldwin, T., Cook, P., Lui, M., MacKinlay, A., Wang, L. (2013). How Noisy Social Media Text, How Diffrent Social Media Sources? In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 356–364, Nagoya, Japan. Asian Federation of Natural Language Processing.
- Biber, D. (2012). Corpus-Based and Corpus-driven Analyses of Language Variation and Use, in: *The Oxford Handbook of Linguistic Analysis*. pp. 159–192. <https://doi.org/10.1093/oxfordhb/9780199544004.013.0008>
- Domnich, M., Sünter, I., Trofimov, H., Wold, O., Harun, F., Kostiukhin, A., Järveoja, M., Veske, M., Tamm, T., Voormansik, K., Olesk, A., Boccia, V., Longepe, N., Cadau, E.G. (2021). KappaMask: AI-Based Cloudmask Processor for Sentinel-2. *Remote Sens.* 13, 4100. <https://doi.org/10.3390/rs13204100>
- Erelt, T. (2007). *Terminiõpetus*. Tartu Ülikooli Kirjastus; Tartu. 978-9949-11-723-9, 469p.
- Fišer, D., Suchomel, V., Jakubíček, M. (2016). Terminology Extraction for Academic Slovene Using Sketch Engine.
- Jakubíček, M., Šmerk, P. (2016). Large Scale Keyword Extraction using a Finite State Backend.
- Jakubíček, M., Kovář, V., Rychlý, P., Suchomel, V. (2020). Current Challenges in Web Corpus Building.
- Kallas, J., Kilgarriff, A., Koppel, K., Kudritski, E., Michelfeit, J., Viks, Ü. (2015). Automatic generation of the Estonian Collocation Dictionary database, in: *Proceedings of the ELex 2015 Conference*. Herstmonceux Castle, United Kingdom, pp. 1–20.
- Kallas, J., Suchomel, V., Khokhlova, M. (2017). Automated Identification of Domain Preferences of Collocations [WWW Document]. URL <https://www.semanticscholar.org/paper/Automated-Identification-of-Domain-Preferences-of-Kallas-Suchomel/fd910903a6d19a9270998c236ab6ff402834a3af> (accessed 1.18.23).
- Khoshkhah, K., Medianovskiy, K., Kolesnykov, D., Hadachi, A., Voormansik, K., 2022. A Hidden Markov Model Method for Non-Stationary Noise Reduction, Case study on Sentinel Data for Mowing Detection.
- Kilgarriff, A.; Rychly, P.; Smrž, P. and Tugwell, D. (2004). The Sketch Engine. In: G. Williams, S. Vessier (eds.) *Proceedings of the XI Euralex International Congress*. Lorient: Université de Bretagne Sud, pp. 105–116.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V., 2014. The Sketch Engine: ten years on. *Lexicography* 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., Rychlý, P., 2008. GDEX: Automatically Finding Good Dictionary Examples in a Corpus. *Proc Euralex*.
- Komisarenko, V., Voormansik, K., Elshawi, R., Sakr, S., 2022. Exploiting time series of Sentinel-1 and Sentinel-2 to detect grassland mowing events using deep learning with reject region. *Sci. Rep.* 12, 1–15. <https://doi.org/10.1038/s41598-022-04932-6>
- Koppel, K., Kallas, J. 2022a. Eesti keele ühendkorpuste sari 2013-2021: mahukaim eestikeelsete digitekstide kogu. *Eesti Raken. Ühin. Aastaraam.* 18, 207–228. <https://doi.org/doi:10.5128/ERYa18.12>
- Koppel, K., Kallas, J. 2022b. Eesti keele ühendkorpus 2021. DOI: 10.15155/3-00-0000-0000-0000-08E60L
- Koppel, K., Kallas, J., Khokhlova, M., Suchomel, V., Baisa, V., Michelfeit, J., 2019a. SkELL Corpora as a Part of the Language Portal Sõnaveeb: Problems and Perspectives.

- Koppel, K., Tavast, A., Langemets, M., Kallas, J., 2019b. Aggregating Dictionaries into the Language Portal Sõnavveeb: Issues With and Without Solutions. Proc. ELex 2019.
- Kosem, I., Koppel, K., Zingano Kuhn, T., Michelfeit, J., Tiberius, C., 2019. Identification and automatic extraction of good dictionary examples: the case(s) of GDEX. *Int. J. Lexicogr.* 32, 119–137. <https://doi.org/10.1093/ijl/icy014>
- Kostka, M., 2022. Pipeline Effectiveness in the Sketch Engine. Proc. Sixt. Workshop Recent Adv. Slavon. Nat. Lang. Process.
- Noorma, A., Jakobson, L., Lang, M., Kutser, T., Oja, T., Uiboupin, R., Voormansik, K., Puust, R., Post, P., Sepp, K., Liibus, A., 2020a. Kaugseire andmete kasutuselevõtt avalike teenuste väljatöötamisel ja arendamisel (projekti RITA1 KAUGSEIRE lõpparuanne). <https://doi.org/10.23673/RE-255>
- Noorma, A., Lang, M., Alikas, K., Jakobson, L., Olesk, A., Kutser, T., Post, P., Toll, V., Uiboupin, R., Sipelgas, L., Raudsepp, U., Rikka, S., Oja, T., Sagris, V., Puust, R., Liibus, A., Sepp, K., Järveoja, M., Voormansik, K., 2020b. Kaugseire jätkusuutliku rakendamise kava. <https://doi.org/10.23673/RE-250>
- Stefanowitsch, A., 2020. *Corpus linguistics A guide to the methodology*, Textbooks in Language Sciences 7. Language Science Press, Berlin.
- Storjohann, D.P., 2005. Corpus-driven vs. corpus-based approach to the study of relational patterns. Presented at the Corpus Linguistics Conference 2005, Birmingham, p. 20.
- Tavast, A., Taurkar M. 2013. *Mitmekeelne oskussuhtlus*. Kirjastus Valgus; Tallinn. ISBN 978-9985-68-287-6 272 p.
- Tavast, A., Langemets, M., Kallas, J., Koppel, K., 2018. Unified Data Modelling for Presenting Lexical Data: The Case of EKILEX. In J. Čibej, V. Gorjanc, I. Kosem and S. Krek (eds.) *Proceedings of the XVIII EURALEX International Congress: EURALEX: Lexicography in Global Contexts*. Ljubljana, Slovenia, pp. 749–761.
- Tavast et al., 2021. Towards the Superdictionary: Layers, Tools and Unidirectional Meaning Relations, in: *EURALEX XIX Lexicography for Inclusion Congress of the European Association for Lexicograph*. pp. 215–223.
- Tognini-Bonelli, E., 2001. *Corpus Linguistics at Work*, scl.6. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Vaus, M., Saar, S., Plado, M. 2019. *Mitmekeelne terminibaas Esterm v20190101*. DOI: 10.15155/3-00-0000-0000-0000-07BDAL.
- Vaus, M. 2022. Ekilexis kirjete koostamise töötuba 25. mai 2022. https://terminoloogia.ee/wp-content/uploads/2022/05/Terminitoo-alused_Mari-Vaus_EKI_25.05.2022.pdf
- Vezzani, F., Nunzio, G.M.D., Henrot, G., 2018. TriMED: A Multilingual Terminological Database. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, p. 5.
- Voormansik, K., Zalite, K., Sünter, I., Tamm, T., Koppel, K., Verro, T., Brauns, A., Jakovels, D., Praks, J., 2020. Separability of Mowing and Ploughing Events on Short Temporal Baseline Sentinel-1 Coherence Time Series. *Remote Sens.* 12, 3784. <https://doi.org/10.3390/rs12223784>

Lõputöö autori kinnitus

Olen lõputöö kirjutanud iseseisvalt. Kõigile töös kasutatud teiste autorite töödele, põhimõtteliste seisukohtadele ning muudest allikaist pärinevatele andmetele on viidatud.

Autor: Liisi Jakobson

27.05.2023

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Liisi Jakobson,

(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose
KORPUSTE RAKENDAMINE TERMINILOOMES KAUGSEIRE TERMINIBAASI NÄITEL,

(lõputöö pealkiri)

mille juhendajad on Sirje Kupp-Sazonov ja Jelena Kallas,

(juhendaja nimi)

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Liisi Jakobson

27.05.2023