

ÕIGUSLIKUD VÄLJAKUTSED JA VÕIMALIKUD LAHENDUSED KEELERESSURSSIDE ARENDAMISEL

Arvi Tavast, Heiki Pisuke, Aleksei Kelli

Ülevaade. Keeleressursside arendamisel ei osata praktikas sageli arvestada lähtematerjali puudutavate õiguslike piirangutega, mis tekitab tarbetuid õigusriske ja võib takistada loodud ressursside sihipärasest kasutamist ja levitamist. Artikli eesmärk on kaardistada levinumad probleemid ja pakkuda kehtivale õigusele tuginevaid lahendeid, viimaste puudumisel aga teoreetilisi ja praktilisi tõlgendusi.

Autoriõiguslikult kaitstavate teoste kasutamiseks keeleressurssides esitatakse lisaks litsentsipõhisele meetodile kaks teoste vaba kasutamise viisi, tsiteerimine ja nn teadustöö erand, koos vastavate täpsemate juhistega. Lisaks kirjeldatakse täiendavaid tingimusi, mis puudutavad suulist ja eriti spontaanses olukorras salvestatud kõnet, varem avaldamata teoseid, tuletatud teoste loomist ja keeleressursside avalikustamist.*

Võtmesõnad: intellektuaalne omand, autoriõigus, litsentsid, isikuandmete kaitse, teose vaba kasutamine, teadustöö erand, tsiteerimine

1. Sissejuhatus

Eesti Keeleressursside Keskuse (EKRK) praktilises tegevuses on päevakorda kerkinud mitmed õiguslikud küsimused: millist keelematerjali tohib keele uurimiseks ja keeletehnoloogiliste lahenduste loomiseks kasutada, millistel tingimustel, kuidas need tingimused mõjutavad uute keeleressursside hilisemat levitamist jt. Võimalikke vastuseid nendele küsimustele tuleb otsida intellektuaalse omandi¹ jts õigusaktidest, teoriast ja praktikast. Just nimelt otsida, sest puuduvad valmislahendused ja kuni viimase ajani oli see valdkond jäänud välja nii filoloogide kui ka juristide uurimisalast. Autoriõigus on keeleressursside arendamisel ja kasutamisel

* Artikkel on valminud Euroopa Sotsiaalfondi kaasabil "Teadus- ja innovatsioonipoliitika seire" programmi raames. Kõik selles artiklis avaldatud seisukohad kuuluvad autoritele ning ei väljenda mingil viisil Euroopa Liidu Nõukogu või muude Euroopa Liidu institutsioonide ametlikke seisukohti.

¹ Intellektuaalse omani legaalmääratlus on antud Ülemaailmse Intellektuaalse Omandi Organisatsiooni (WIPO) asutamise konventsiooni artiklis 2, mis sätestab, et intellektuaalne omand sisaldab õigusi, mis tulenevad intellektuaalsest tegevusest tööstuse, teaduse, kirjanduse ja kunsti alal (RT II 1993, 25, 55).

keskse tähendusega (autoriõiguse kohta üldiselt vt Pisuke 2006). Kas selguse huvi- des peaks autoriõigus nägema ette eraldi keeleressursside loomise regulatsiooni ja vastava teoste kasutamise erandi või võimaldab ka kehtiv autoriõiguse süsteem keeleressursside loovat kasutamist ja arendamist, see on väljakutse, millele vastuse otsimise on allakirjutanud endale võtnud (vt ka Kelli jt 2012).

Keeleressursside koostamine, eriti eesti keele puhul, ei pea reeglina silmas ärilisi eesmärke. Ka õiguste omajad – autorid ja kirjastajad – on seni seda protsessi toetanud, andes lube oma teoste kasutamiseks. Õiguslikud probleemid tõusetuvad eelkõige seoses nende lubade puuduliku juriidilise vormistamisega seadusele vastavate litsentsidena. Praktikas tekivad probleemid samuti seoses loa taotlemata jätmisega, isikuandmete kaitsega ja kõne salvestamisega spontaanses olukorras.

Käesoleva artikli esimene osa liigitab keeleressurssid õiguslike probleemide alusel, mis võivad nende puhul tekkida ja raskendada nende kättesaadavaks tegemist EKRR ja CLARIN ERICu² kaudu. Teises osas analüüsivad autorid, milliste õiguslike aspektidega tuleb keeleressursside koostamisel arvestada, keskendudes keeletehnoloogias seni raskusi valmistanud teemadele. Kolmandas osas antakse autorite soovitusel keeleressursside edasise kogumise jaoks, millega arvestamine võimaldaks vältida õiguslike probleemide tekkimist kogu protsessi jooksul.

Käesoleva artikli kontekstis määratletakse keeleressurssi andmebaasina, mis sisaldab mitmeid kirjalikke ja suulisi tekste. Tekstide valimine on loomuline tegevus, mistõttu on keeleressurss kui andmebaas autoriõiguslikult kaitstav teosena.³

2. Keeleressursside liigitus õiguslike väljakutsete seisukohast

2.1. Leksikaalsed andmebaasid

Leksikaalsete andmebaaside hulka kuuluvad sõnastikud, tesaurused, leksikonid, terminibaasid ja muud inim- ja/või masinloetavad andmebaasid, mis lihtsamal juhul sisaldavad üksiksõnu ja levinud väljendeid ühes või mitmes keeles. Lisaks võivad leksikaalsed andmebaasid sisaldada näitelauseid, mille puhul võib olla oluline nende autentsus, st pärinemine mõne teise autori teostest vastandina leksikograafi omaloomingule. Samuti võivad need baasid sisaldada enam või vähem formaliseeritud infot keelendite päritolu, kasutusviisi (muuttüüp, rektsioon vms), tähenduse (tähenduseseosed, mõisteseosed) või muude sõna või mõiste parameetrite kohta ning tähenduse verbaalseid kirjeldusi (seletused, määratlused).

Siinses artiklis käsitletavad õiguslikud väljakutsed leksikaalsete andmebaaside juures piirduvad autentsete näitelausete kasutamisega teiste autorite teostest

² CLARIN ERIC www.clarin.eu (22.09.2012).

³ AutÕS § 4 lg 3 p 22 sätestab, et teosed, millele tekib autoriõigus, on mh "teoste kogumikud ja informatsiooni kogumikud (sealhulgas andmebaasid). Andmebaas käesoleva seaduse tähenduses on süstemaatiliselt või meetoodiliselt korraldatud iseseisvate teoste, andmete või muu materjali kogu, mis on individuaalselt kättesaadav elektrooniliste või muude vahendite abil. [...] Käesoleva seaduse alusel kaitstakse autoriõigusega andmebaasi, mis oma sisu valiku ja korralduse tõttu on autori enda intellektuaalse loomingu tulemus, ning ei kohaldata mingit muud kriteeriumi."

Andmebaas võib olla kaitstav ka AutÕS 8¹ peatüki alusel kui *sui generis* andmebaas. See tähendab, et õiguste saamiseks andmebaasile piisab investeeringu tegemisest andmebaasi loomisesse. Autoriõiguse seadus määrab *sui generis* andmebaasi tegijana isikut, kes on "teinud kas laadilt, väärtuselt või suuruselt olulise investeeringu selle andmebaasi sisuks olevate andmete kogumiseks, omandamiseks, kontrollimiseks, süstematiseerimiseks või kättesaadavaks tegemiseks" (AutÕS § 75³ lg 1). Iseenesest on keeleressurssid kaitstavad *sui generis* andmebaasina, kuid sellel puudub oluline tähendus, sest teosena kaitstava andmebaasi õiguskaitse on ulatuslikum. Põhiline eelis *sui generis* kaitsega seoses on see, et investeeringute jätkumisel pikeneb *sui generis* 15-aastane kaitsetähtaeg *ad infinitum* (vt AutÕS § 75⁷).

(vt 3.1). Lisaks on valdkonnas üldlevinud praktika koostada uusi leksikaalseid andmebaase varasemate põhjal, kuid selle kohta on hea tava suhteliselt hästi välja kujunenud ja võimalikud lahkarvamused püsivad valdkonna ekspertide siseringsis, mistõttu seda teemat siin ei käsitleta.

2.2. Korpused märgendatuse järgi

Tekstikorpused kui suured struktureeritud ja eesmärgipärasel viisil koostatud tekstikogumid koosnevad tervenisti teiste autorite teostest või nende osadest. Lihtsamal juhul ehk märgendamata korpuse juures on koostaja töö piirdunud tekstide valimise ja süstematiseerimisega. Selliseid korpuseid on võimalik (ja keeletehnoloogiliste rakenduste seisukohalt vajalik) koostada poolautomaatselt ja väga suurtes mahtudes, mis tõstab päevakorda sisend litsentside hankimise töömahukuse ja tekitab vajaduse otsida alternatiivseid viise ressursside seaduslikuks kogumiseks (vt 4.2–4.3, vt ka Vider jt 2013 siinses kogumikus).

Teine variant ehk märgendatud korpus (sh puudepank kui teatud viisil süntaktiliselt märgendatud korpus) sisaldab lisaks teiste autorite tekstidele ka lingvistilist ulatuslikku panust nende märgendamisel ehk eksplitsiitse info lisamisel tekstis esinevate keelendite sõnaliigi, häälduse, lauses täidetava funktsiooni, tähenduse vms kohta. Teatud piirini on märgendada võimalik poolautomaatselt, kuid tõeliselt väärtuslikud on just käsitsi märgendatud korpused, sest need annavad keeletehnoloogiale juurde sellist infot keele kohta, mida automaatselt tekitada ei ole tehnika selle hetke arengutaseme juures võimalik. Märgendatud korpuste täiendav õiguslik väljakutse seisnebki teiste autorite teoste muutmises märgenduse lisamise teel ehk tuletatud teose loomises. Harjumuspärane viis seda vormistada on kirjutada sõna kohta käiv metainfo iga märgendatud sõna juurde, näiteks selle järele mingite koodide kujul, kuid tehniliselt peaks olema teostatav ka märgenduse hoidmine eraldi failis, et teise autori teksti sõnade vahele midagi kirjutama ei peaks.

2.3. Korpused keelematerjali liigi järgi

Ajalooliselt esimesed ja praegu levinuimad tekstikorpused koosnevad kirjalikest õiguspäraselt avaldatud tekstidest, mille puhul väljakutsed piirduvad intellektuaalomandi temaatikaga.

Mõned korpused lisavad tekstid võivad olla ka avaldamata, näiteks koolitööd. Sellise materjali kasutusvõimalused on piiratumad kui avaldatud tekstidel, sest tsiteerimise ja teaduskasutuse luba laieneb vaid avaldatud tekstidele (vt 3.1).

Lisaks kirjutatud tekstidele läheb tarvis ka suulise kõne korpuseid, mis koosnevad helisalvestusest, videosalvestusest ja/või nende transkriptsioonidest, millega lisaks intellektuaalomandile kaasneb ka isikuandmete kaitse ja salvestamise õiguspärasuse temaatika (vt 3.2). Salvestamiseks tuleb küsida luba, kuid ehtsa suulise keelekasutuse uurimiseks vajalikku tõeliselt loomulikku kõnet saab ainult nii, et uuritav ei tea, et tema kõnet salvestatakse. Kui aga salvestatav kõneleja esitab kellegi teise teksti, siis tuleb võtta arvesse ka tolle teksti autoriõigustest tulenevaid piiranguid.

2.4. Keelemudelid, grammatikad

Seni nimetatud ressursid ise ei ole vajalikud millekski muuks kui keeleteadusliku või -tehnoloogilise uurimis- ja arendustöö jätkamiseks järgmistel astmetel. Korpuste ja sõnastike põhjal ning nende kasutaja loomingu abil saab hankida infot selle kohta, milline uuritav keel on või kuidas seda kasutatakse ning luua keelemudeleid, grammatikaid ja muid keele formaalseid kirjeldusi. Mõned neist ei sisalda üldse nende aluseks olevaid korpusetekste sõnast või kinnisväljendist pike-mate fragmentidena, näiteks kui lingvist (või tema projekteeritud arvutisüsteem) vaatab läbi suure koguse autentseid tekste ja jõuab järeldusele, et “dokumente” on eesti keeles oluliselt levinum mitmuse osastava vorm kui “dokumentisid”. Sellisel juhul on tuletuslik seos uuritud tekstide ja saadud järelduse vahel üsna vaieldav. Mõnel teisel juhul võivad tekstide äratuntavad osad keelekirjelduses siiski sisalduda, näiteks kui statistiline trigramm-mudel loetleb kõik uuritavates tekstides leiduvad kolmesõnalised järjendid ja nende tõenäosused, siis needsamad kolmesõnalised järjendid on mudelist ka leitavad.

2.5. Tarkvara

Keeleressursside hulka võidakse mõnikord lugeda ka keeletehnoloogiline tarkvara: analüsaatorid, süntesaatorid, sõnastikusüsteemid, otsingu- ja päringusüsteemid, masintõlkesüsteemid, dialoogsüsteemid, märgendusvahendid jms. Nende toodete endiga on küll seotud õiguslikke küsimusi, nt vajavad keeletehnoloogia riiklikust programmist finantseeritud tarkvara litsentsitingimused mingi kompromissi saavutamist finantseerija nõudmiste (et kõik riikliku programmi tulemused oleks kõigile vabalt kasutatavad) ja programmi autorite tööandjate soovide (et tööülesannete täitmise käigus loodud teoste varalised õigused kuuluvad tööandjale ja need tuleks rakendada tööandjale tulu genereerimise teenistusse) vahel. Kuna aga tarkvara eraldi vaadelduna ei sisalda keelematerjali, vaid üksnes kasutab seda, siis tarkvaraga seotud küsimused jäävad siinse artikli käsitluselast välja.⁴

3. Õiguslikke probleeme seoses keeleressurssidega

3.1. Kasutatavate teoste autoriõiguslik kaitse

Keeleressursside loomine tugineb suuliste ja kirjalike tekstide kasutamisele, millest suur osa on autoriõiguslikult kaitstavad teosed. Autoriõiguse seadus määratleb teost kui “mis tahes originaalset tulemust kirjanduse, kunsti või teaduse valdkonnas, mis on väljendatud mingisuguses objektiivses vormis ja on selle vormi kaudu tajutav ning reprodutseeritav kas vahetult või mingi tehnilise vahendi abil”. Teos on originaalne, kui see on “autori enda intellektuaalse loomingu tulemus” (AutÕS § 4 lg 2). Eeltoodud regulatsiooni tuleb mõista selliselt, et kui isik on midagi loonud kirjanduse, kunsti või teaduse valdkonnas ning see looming on väljendatud mingisuguses tajumist võimaldavas objektiivses vormis, siis on see autoriõiguslikult kaitstav teos. Teose objektiivseks vormiks võib olla näiteks salvestatud suuline

⁴ Ka autoriõiguse seadusest tuleneb, et andmebaasi mõiste ei hõlma selle tegemiseks ega käivitamiseks vajaminevat arvutiprogrammi (AutÕS § 4 lg 3 p 22).

kõne, käsikirjaline tekst, trükitud artikkel või raamat, arvutifail jms. Keskseks teose tunnuseks on seejuures originaalsus. Autoriõigusega kaitstavalt teoselt ei nõuta midagi ennenägematut või seniolematut, nagu see on patendiõiguses. Nõutakse lihtsalt autori enda isiklikku panust, mis ei tohi üks-üheselt kopeerida juba olemasolevaid teoseid. Originaalsus ei eelda seega uudsust, vaid on piisav, kui autoril on vähemalt minimaalne enda loomingiline panus. Autoriõigus põhineb kahel eeldusel ehk presumptsioonil: 1) teose originaalsust ja seetõttu ka kaitstust autoriõigusega eeldatakse ning vastupidise tõendamise kohustus lasub vaidlustajal (AutÕS § 4 lg 6) ja 2) isiku, kes avaldab teose oma nime all või oma üldtuntud pseudonüümi või autorimärgi all, autorsust eeldatakse, kuni pole tõendatud vastupidist. Tõendamiskohustus lasub autorsuse vaidlustajal (AutÕS § 29 lg 1). Kõik seniõeldu kehtib nii ainuisikuliselt kui ka kollektiivselt (AutÕS § 30) loodud teoste puhul.

Samas eksisteerib teatud hulk loomingut, mille autoriõiguslik kaitse on *expressis verbis* välistatud. Nii välistab autoriõiguse seadus autoriõigusliku kaitse alt ideed, mõisted, teooriad ja meetodid, rahvaloominguteosed, õigusaktid, haldusdokumendid, kohtulahendid ja nende ametlikud tõlked, faktid ja andmed ning muud seaduses otseselt nimetatud inimloomingu tulemused (AutÕS § 5).

Väga oluline on veel pidada silmas kahte printsiipi: 1) autoriõigusliku kaitse tähtajalisust ja 2) territoriaalsust.

- 1) Autoriõiguse üldine täheaeg on autori kogu elu ja 70 aastat pärast tema surma (70 aastat *post mortem auctoris* ehk *pma*). Seda tähtaega hakatakse lugema autori surma-aastale järgneva aasta 1. jaanuarist (AutÕS § 38 lg 1 ja § 43). On mõningad erisätted, mis kehtestavad kas lühema või isegi pikema tähtaja⁵, kuid kõige sagedamini esinev üldreegel on 70 a *pma*. Pärast teose kaitsetähtaja lõppu võivad kõik isikud seda vabalt kasutada ehk teos on nn ühises kasutuses (ingl *public domain*). Kasutaja põhiliseks kohustuseks on üksnes viitamine teose autorile, sest autoriõiguse seaduse kohaselt “[i] siku konkreetse teose autoriks olemist (teose autorsust), autori nime ning autori au ja väärikut kaitstakse tähtajatult” (AutÕS § 44 lg 1).

Vanade tekstide kasutamisel, mille puhul on möödunud 70 aastat autori surmast, on EL liikmesmaades kehtestatud üks erand juhuks, kui autori teos jäi tema eluajal või hiljem avaldamata.⁶ Autoriõiguse seaduse kohaselt saab isik, kes pärast autoriõigusliku kaitse lõppu avaldab varem avaldamata teose, sellele teosele autori varalised õigused 25 aastaks (AutÕS § 74¹ lg 1). Regulatsiooni mõte on motiveerida varem avaldamata teoste uurimist ja avaldamist ja kompenseerida sellega kaasnevad võimalikud kulud. (Täiendavaks analüüsiks vt Walter ja Lewinski 2010: 568–578.)

- 2) Territoriaalsuse printsiip tähendab seda, et igal maal kehtib oma autoriõiguse normistik ja põhimõtteliselt ei pea kaitsma välismaiste autorite välismaal avaldatud teoseid. Autoriõigusliku kaitse territoriaalsuse printsiipi laiendatakse rahvusvaheliste lepingute sõlmimise ja nendega ühinemise teel. Rahvusvaheliste konventsioonidega ühinemine toob kaasa kohustused

⁵ Näiteks anonüümselt või pseudonüümi all avaldatud teoseid, kollektiivseid teoseid (AutÕS § 31) või töökohustuste täitmise korras loodud teoseid (§ 32) kaitstakse 70 aastat pärast selle teose õiguspärast avalikustamist, kuid see tähtaeg võib teatud tingimustel pikeneda ka üldtähtajani (§ 40 ja 41). Samas autoriõiguse teosele, mille on loonud üks või enam isikut oma ühise loomingulise tegevusega, kehtib teised autorid üleelanud ühise autori eluajal ja 70 aastat pärast tema surma (§ 39).

⁶ Avaldamiseks loetakse teose kättesaadavaks tegemist määramata isikute ringile ehk seaduse sõnastuses: “Teos loetakse avaldatuks, kui teos või teose mis tahes vormis reprodutseeritud koopiad on autori nõusolekul antud üldsusele kasutamiseks koguses, mis võimaldab üldsusel sellega tutvuda või seda omandada. Teose avaldamiseks loetakse muu hulgas teose trükkis väljaandmist, teose eksemplaride panemist müügile, jaotamist, laenutamist, rentimist ja muul viisil tasuta või tasu eest kasutada andmist” (AutÕS § 9).

ühtlustada teatud konventsioonis sisalduvad üldprintsiibid (näiteks mida tuleb kaitsta ja kui pika tähtaja jooksul) ja kaitsta oma territooriumil kõigi konventsiooniga ühinenud liikmesriikide autorite teoseid⁷. EL liikmesmaade puhul tulenevad teatud täiendava ühtlustamise ja vastastikuse kaitse kohustused EL direktiividest. Konventsiooniga mitteühinenud riikide autorite teoseid (näiteks nendes riikides väljaantud sõnaraamatuid või muid keeleressursse) Eestis kaitsma ei pea. Selliseid riike on aga võrreldes konventsioonidega ühinenud riikidega tühine vähemus.

Keeleressursside loomise seisukohalt on eelneval väga oluline tähendus. Nimelt, enamus tekste, mida keeleressursside loomisel kasutatakse, on autoriõiguslikult kaitstavad. Sisuliselt saab nende kasutamine tugineda kas õiguste omaja nõusolekule või siis teose vaba kasutamise eranditele.

Keeleressursside kasutamisel omab teatud juhtudel tähendust autoriõiguse eristamine ühest teisest intellektuaalse omandi instituudist – autoriõigusega kaasnevatest õigustest. Näiteks teose esitaja, fonogrammitootja ja ringhäälinguorganisatsiooni (ehk seaduse sõnastuses televisiooni- ja raadioteenuse osutaja) õigused on iseseisvad õigused ja autoriõiguse sätted nende suhtes ei kehti (AutÕS 8. peatükk).

3.2. Suulise ja spontaanses olukorras salvestatud kõne kasutamine

Suulise ja spontaanses olukorras salvestatud kõne kasutamine keeleressurssides on juriidiliselt komplitseeritud valdkond. Üldjoontes võib praktikas sagedamini esineda probleemid jagada järgmistesse valdkondadesse kuuluvaiks: üldsusele avaldamata tekstide kasutamine, autori ja esitaja õigused ning isiku privaatsuse riive.

Autoriõiguslik kaitse laieneb ka suulistele teostele, mis tähendab, et kaitse saamiseks sellistele teostele ei ole vaja teksti eelnevalt kirja panna. Teksti suulist esitamist loetakse teosele objektiivse vormi andmiseks, kuid see peab olema kuulajate poolt tajutav ja reprodutseeritav (näiteks salvestamise või üleskirjutamise teel). Kaitse alguspunktiks on teose loomine ehk selle väljendamine suulises vormis.⁸ Suuliste teoste puhul kehtivad kõik autoriõiguse üldprintsiibid, sealhulgas ei nõuta loodud teose deponeerimist või registreerimist (AutÕS § 7). Kaitstavad on seejuures nii avalikustatud kui ka avalikustamata teosed. Avalikustamine tähendab teose avaldamist või üldsusele suunamist (AutÕS § 8). Seega pole tähtsust, kas suuline tekst loodi intervjuu käigus ühele ajakirjanikule või rahvusvahelise konverentsi ettekande esitamise ajal.

Keeleressursside kontekstis kasutatav suuline kõne on tihti avalikustamata. Avalikustamata teoste puhul on aga seadusega kehtestatud üks väga oluline lisapiirang. Teose vaba kasutamine ilma autori nõusolekuta võib toimuda ainult avalikustatud ehk autori enda selgel nõusolekul üldsuse ette viidud teoste puhul. Teose tsiteerimine ning kasutamine teaduslikel ja õppe-eesmärkidel eeldab, et see teos on avaldatud, ja lisatingimusena – avaldatud õiguspäraselt, seega autori selgelt väljendatud soovil ja tahtel. Asjaolu, et teos on avaldamata, ei tähenda, et seda ei võiks üldse kasutada. Kui teose autor ehk isik, kelle suulist kõnet salvestatakse või

⁷ Tähtsaim autoriõiguse rahvusvaheline leping, millega ka Eesti ühines, on Berni kirjandus- ja kunstiteoste kaitse konventsioon (Berni konventsioon; RT II, 1994, 16, 49). Teatud kohustused Eesti jaoks tulenevad ka Maailma Kaubandusorganisatsiooni asutamislepingu lisast 1C ehk Intellektuaalomandi õiguste kaubandusaspektide lepingust (TRIPS lepingust; RT II, 1999, 22, 123) ning Maailma Intellektuaalse Omandi Organisatsiooni autoriõiguse lepingust (RT II, 2006, 14, 39).

⁸ AutÕS § 4 lg 3 p 4 on autoriõigusega kaitstavate teoste näidisloetelus fikseeritud "kõned, loengud, ettekanded, jutlused jt teosed, mis koosnevad sõnadest ja on väljendatud suuliselt (suulised teosed)".

üles kirjutatakse, annab sellise salvestise või kirjalikult fikseeritud teksti kasutamiseks oma nõusoleku, siis on kasutamine õiguspärane.

Suulise kõne autoriõiguslikud küsimused võivad olla seotud mitme isiku loome-tegevusega. Näiteks kui esitatava teksti autoriks on üks isik ja see esitatakse teise isiku poolt, siis kehtivad teksti autori suhtes autoriõiguse normid, esitaja⁹ suhtes aga autoriõigusega kaasnevate õiguste kohta käivad normid. Teksti kui autoriõiguslikult kaitstava teose kasutamisel (esitamisel, salvestamisel või muul reprodutseerimisel) saab olla kaks õiguslikku alust: 1) õigustatud isiku ehk õiguste omaja¹⁰ nõusolek, mis on selgelt väljendatud loa (ehk litsentsi) või õiguste loovutamise vormis, või 2) seaduses ammendavalt sätestatud teose vaba kasutamise sätted.

Suulise ja spontaanses olukorras salvestatud kõne lülitamine keeleressursside hulka omab lisaks autoriõigusele tihedat kokkupuudet isikuandmete kaitsega, kuna kõneleja võib olla tuvastatav ning tema kõne kasutamine tähendab ka isikuandmete töötlust. Isikuandmete kaitse seaduse kohaselt on isikuandmed “mis tahes andmed tuvastatud või tuvastatava füüsilise isiku kohta, sõltumata sellest, millisel kujul või millises vormis need andmed on” (IKS § 4 lg 1). Vastavalt seadusele on isikuandmete töötlemine “iga isikuandmetega tehtav toiming, sealhulgas isikuandmete kogumine, salvestamine, korrastamine, säilitamine, muutmine ja avalikustamine, juurdepääsu võimaldamine isikuandmetele, päringute teostamine ja väljavõtete tegemine, isikuandmete kasutamine, edastamine, riskasutamine, ühendamine, sulgemine, kustutamine või hävitamine, või mitu eelnimetatud toimingut, sõltumata toimingute teostamise viisist ja kasutatavatest vahenditest” (IKS § 5). Isikuandmete töötlemine on lubatud üksnes puudutatud isiku nõusolekul või seaduse alusel (IKS § 10).

Keeleressursside seisukohalt taanduvad isikuandmete kaitse küsimused eelkõige sellele, et lisaks autoriõigusest tulenevale nõusolekule teose kasutamiseks on vajalik ka nõusolek isikuandmete töötlemiseks. Nõusolekuta isikuandmete töötlemine on karistatav (IKS § 42 ja 43).

Spontaanses olukorras salvestatud kõne kogumisega seonduvad probleemid on juriidiliselt kõige komplitseeritumad. Spontaanse olukorraga on tegemist siis, kui rääkija ei tea, et tema kõnet salvestatakse, mis tagab kõnelemisel täieliku vabaduse. Salaja lindistamisega võidakse samas riivata isikute privaatsusõigust. Jälitustegevuse seadus loeb andmete varjatud kogumist jälitustoiminguks, mida tohib läbi viia jälitusasutus (JTS § 12 lg 1 p 1) ning karistusseadustik näeb koguni ette kuni kolmeaastase vangistuse, juhul kui jälitustegevuseks seadusliku õigusega isik jälgib teist isikut tema kohta andmete kogumise eesmärgil (KarS § 137). Karistusseadustiku kommentaarides on rõhutatud, et nimetatud koosseisuga kaitstakse nii elulisi kui ka tööalaseid andmeid, kusjuures ei ole oluline nende kvaliteet. Samuti on isiku vabadus piiratud, kui ta peab pidevalt kartma, et tema tegevust salaja talletatakse (Sootak, Pikamäe 2009: 405).

⁹ Autoriõiguse seadus loeb teose esitajaks “näitlejat, lauljat, muusikut, tantsijat, muud isikut või kollektiivi, kes näitleb, laulab, deklameerib, mängib muusikariistal või muul viisil esitab kirjandus-, kunsti- või rahvaloominguteoseid või juhendab teisi isikuid teoste esitamisel, aga samuti isikut, kes esitab estraadi-, tsirkuse-, nukuteatri- jms numbreid” (AutÕS § 64).

¹⁰ Õiguste omajaks ei pruugi olla tingimata autor, vaid selleks võib olla ka näiteks kirjastaja, ülikool, uurimisinstituut või muu tööandja.

4. Võimalikud lahendused

4.1. Korrektsed litsentsid

Kui keeleressursside arendamine toimub ärielistel eesmärkidel, siis on vajalik õiguste omajaga kirjaliku litsentsilepingu sõlmimine. See on nn litsentsipõhine keeleressursside kasutamine. Litsentsileping on autorilepingu liik ja selles peavad kajastuma kõik seaduses nõutavad elemendid.¹¹ Eriti tuleks tähelepanu pöörata selle fikseerimisele, millisel eesmärgil, ulatuses ja vormis litsentseeritavat teost kavatsetakse kasutada. Oluline on kohe välja selgitada, kellele kuuluvad autoriõigused. Keeleressursside kasutamisel tuleb arvestada asjaoluga, et autor ei pruugi olla tingimata õiguste omaja ning suure tõenäosusega võivad autoriõigused kuuluda teose kirjastajale või tööandjale, mistõttu tuleb just nendega sõlmida litsentsileping. Kirjastajaga litsentsilepingu sõlmimise eeliseks keeleressursside looja seisukohalt on asjaolu, et kirjastajale kuuluvad reeglina ka paljude teiste teoste autoriõigused ning läbivõtmise objektiks saab olla korrigeeritud suurema hulga teoste kasutamine. Üldjuhul ei saa eeldada, et taoline litsentsilepingu sõlmimine riivaks kuidagi kirjastaja ärihuvisid, sest teoste kasutus keeleressursside kontekstis on erinev kui see, kuidas kirjastaja neid kommertsialiseerib (nt müüb raamatuna, e-raamatuna, annab filmi või teleõiguseid jne). Keeleressursside arendamisel kasutatakse teoseid eelkõige "toormaterjalina" ning teoste kirjanduslik või majanduslik väärtus ei ole oluline.

Litsentsil põhineva keeleressursside loomise suurim puudus ilmneb eelkõige massikasutuse tingimustes. Kuna keeletehnoloogia kasutab üha enam statistilisi meetodeid, siis läheb keeleressurssse vaja väga suures mahus. See omakorda muudab lepingute sõlmimise kõigi autoriõiguste omajatega väga töö- ja ressursimahukaks. Seetõttu, kui teoste kasutamine plaanitakse viisil ei ole ilma autori loata lubatud, ja autori loa hankimine on liiga keeruline, tasuks uurida, millisel viisil tohib teost kasutada ilma autori nõusolekuta. Keeleressursside puhul tuleb niisugustest viisidest praegu kehtiva autoriõiguse süsteemi raames kõne alla peamiselt kaks: tsiteerimine ja nn teadustöö erand.

4.2. Tsiteerimine

Teadustöös on tsiteerimine väga levinud teiste autorite teoste kasutamise viis. Ka keeleressurssid võivad sisaldada tsitaate teiste autorite teostest. Tsiteerimiseks on seadusega kehtestatud väga konkreetsed reeglid¹². Tsiteerida tohib autori nõusolekuta ja autoritasu maksmata (AutÕS § 19 p 1), kuid ainult avaldatud teoseid. Vajalik on autorile viitamine, eeldusel, et autor on teosele märgitud.

Tsiteerimise tingimused on seaduse üldnormi alusel samad kõigile teose liikidele, järelkult ka nii keeleressursside või näiteks teadusartiklite puhul. Konspektiivselt on need järgmised.

¹¹ Vastavalt AutÕS § 48¹ "Autorilepingu sisu" tuleb autorilepingus fikseerida: 1) teose kirjeldus (vorm, maht, nimetus jms); 2) üleantavad õigused, õigused, mille osas antakse luba, litsentsilepingu liik (liht- või ainulitsentsileping) ja all-litsentsi andmise õigus; 3) teose kasutamise viis ja territoorium; 4) autorilepingu kehtivuse tähtaeg ja teose kasutamise algustähtaeg. Samuti määratakse poolte kokkuleppel, kas leping on tasuta või tasuline. Autoritasu maksmise viis (protsent teose müügihinnast, konkreetselt kindlaksmääratud summa, protsent teose kasutamisel saadavast kasumist vms), tasu suurus, selle väljamaksmise tähtaeg ja kord määratakse autorilepingus poolte kokkuleppel.

¹² AutÕS § 19 sätestab: "Autori nõusolekuta ja autoritasu maksmiseta, kuid kasutatud teose autori nime, kui see on teosel näidatud, teose nimetuse ning avaldamisallika kohustusliku äranäitamiseega on lubatud:

1) õiguspäraselt avaldatud teose tsiteerimine ja refereerimine motiveeritud mahus, järgides refereeritava või tsiteeritava teose kui terviku mõtte õige edasiandmise kohustust."

- 1) Tsiteeritav teos peab olema õiguspäraselt avaldatud.
- 2) Iga tsitaadi juures peab reeglina olema viide autori nimele, teose pealkirjale ja avaldamisallikale, juhul kui need rekvisiidid on teosel näidatud. Samas tuleb arvestada ka sellega, et konkreetsetes valdkonnas on oma viitamise spetsiifika. Viitamise kohustus võib olla täidetud ka siis, kui allikate loetelu antakse lõpus ning iga tsitaadi juures viidet ei tehta. Näiteks on avaldatud õpikuid, kus kasutatud teosed viidatakse iga peatüki ees ning teksti sees puuduvad igasugused viited.
- 3) Tsitaat on teksti reprodutseerimine täpselt samal kujul (samas sõnastuses) nagu see on algses teoses; reeglina kasutatakse jutumärke.
- 4) Tsitaadiga ei tohi algse teose mõtet valesti edasi anda (ei tohi moonutada algse teose mõtet, sest tegemist on autori au ja väärikuse kaitsega). Keeleressursside kontekstis on selle nõude vastu eksimine suhteliselt vähe tõenäoline.
- 5) Tsiteerida tohib motiveeritud mahus.

Seaduses ei ole sätestatud, mis on motiveeritud maht. Seadusega polegi seda võimalik teha, sest erinevate teose liikide puhul on motiveeritud maht erinev. Tsiteerida saab kõiki teose liike – kirjandusteoseid, muusikateoseid, arhitektuuriteoseid jne. Tsiteerimine põhineb väljakujunenud heal taval, mis on kohati väljendatud ka näiteks teadusajakirjade, ajakirjandus- ja ringhäälinguorganisatsioonide või autorite ühingu poolt väljatöötatud reeglites. Sellised tsiteerimise hea tava reeglid võiksid olla kehtestatud ka keeleressursside kasutamisel. Nähtavasti tuleks need välja töötada EKRRK ja kirjastajate organisatsioonide poolt ühiselt ja kehtestada vastava hea tava kokkuleppena.

Ühtse lähenemise tagamiseks ja üldise mahuorientiiri andmiseks teeme siinkohal ettepaneku lugeda keeleressursside puhul motiveeritud mahuks järgmist:

- tsiteerida tohib kuni 5–10% tsiteeritava teose mahust, kuid mitte rohkem kui 10 lk (18 000 täheruumi),
- üks järjestikune tsitaat ei tohi ületada 1 lk (1800 täheruumi) ega olla terveviklik alamteos (nt üks terve novell novellikogust).

Keeleressursside loomisel tsiteerimise motiveeritud mahu väljatöötamisel tuleks arvestada nende ressursside loomise eesmärke. Kuna korpusi luuakse eelkõige rahvuskultuuri säilitamiseks, mitte esmaeesmärgiga koguda materjali ärilisi huvisid ja kommertsprojekte silmas pidades, siis võivad tsiteerimise motiveeritud mahud olla proportsionaalselt suuremad kui muude kasutusliikide puhul (näiteks ajakirjanduses või ringhäälingus). Samas peaksid nad jääma ühtsesse proportsiooni teadustöös heaks tavaks peetavate mahtudega¹³.

Teos, mis sisaldab teiste autorite teoste õiguspäraseid tsitaate, on uus iseseisev teos. Selle loomiseks ja edasiseks kasutamiseks ei ole tarvis kellelki luba küsida ega autoritasu maksta. Uut teost võib selle autor oma äranägemisel reprodutseerida, levitada ja litsentsida, seda ka ärilistel eesmärkidel. Ka avalikustada tohib autor sellist uut teost oma äranägemisel.

¹³ Järgnevas alaosas käsitletakse nn kolmeastmelist testi, mis seab teatud piirid motiveeritud mahu määramiseks teoste tsiteerimisel.

- 3) Teose reprodutseerimine toimub õppe- ja teaduslikel eesmärkidel, seega ärilistel eesmärkidel on reprodutseerimine keelatud. Ka siin tekib küsimus, kas õppe- ja teaduslikud eesmärgid peavad esinema samaaegselt. Analoogiliselt eelmisele punktile võib seega eeldada, et õppe ja teaduslikud eesmärgid võivad esineda ka eraldi. Eesti keele säilitamine ja arendamine on eeldatavasti vaadeldav teadustööna.

Vaadeldava sätte tõlgendamisel omab tähendust veel see, kas nimetatud teadusasutus taotleb ärilisi eesmärke (nagu see on näiteks erakõrgkoolide või tasuliste koolitusfirmade puhul) või mitte. EKI puhul võib väita, et selle tegevuse eesmärgiks ei ole äriliste eesmärkide teenimine.

Kuna keeleressursid ei ole üldiselt eesmärk iseeneses, vaid neid kogutakse keele uurimise ja kirjeldamise eesmärgil, siis võib püstitada küsimuse, kas kopeeritud tekstide põhjal loodud statistiline mudel, grammatika, puudepank või muu keelekirjeldus on lahutatav nendest tekstidest ning kasutatav ärilistel eesmärkidel. Antud küsimusele vastamisel on oluline, kuivõrd reprodutseeritud teosed sisalduvad nende põhjal loodud kirjelduses. Kui teosed ei sisaldu keelekirjelduses, siis ei peaks olema takistatud ka selle ärieesmärkidel kasutamine.

- 4) Teose reprodutseerimine motiveeritud mahus. “Motiveeritud mahu” nõue tähendab, et ei reprodutseerita rohkem kui on minimaalselt vaja algse teose autori originaalse idee, seisukoha, stiili, kujundi vms edasiandmiseks uues loodavas teoses. Võib tekkida küsimus, mille suhtes peab maht olema motiveeritud. Infoühiskonna direktiivi art 5 (3) a aitab nimetatud küsimust selgitada: “Liikmesriigid võivad ... näha ette erandeid või piiranguid ... tingimusel, et märgitakse ka allikas, sh autori nimi, kui see ei ole võimatu, ning *ulatuses, mis on õigustatult vajalik taotletava mitteärilise eesmärgi saavutamiseks*”. Seega on võimalik tõlgendus, et motiveeritud mahu määrab kasutuseesmärk. Siin peab uue teose autor olema valmis vajadusel selgitama, miks reprodutseeriti just selles mahus, nagu seda tehti. Argumendid saavad tulla selle konkreetse valdkonna teadlastelt. Isegi kui maht on motiveeritud, võib kasutus olla vastuolus kolmeastmelise testi tingimustega, st olla vastuolus teose tavapärase kasutamisega ja riivata põhjendamatult autori huve.
- 5) Reprodutseerimine ei ole vastuolus teose tavapärase kasutamisega. Ekspertid on soovitanud tavapärase kasutamise hindamisel lähtuda küsimusest, kas konkreetne kasutus on selline, millega õiguste omaja loodab tulu teenida. Arvestada tuleb nii juba tuntud kui ka veel kasutamise ajal tundmatute teose kasutusviisidega. Kui teose vaba kasutus ei konkureeri majanduslikult õiguste omaja huvidega, siis on see lubatud (Ricketson, Ginsburg 2006: 769–770). Erialakirjanduse järgi “tavapärase kasutamine” eeldab, et hinnatakse nii võimalikke tulevasi kui ka praeguseid kasutusviise, mis võimaldavad kasu saamist teosest” (samas: 769).

Antud küsimust on käsitletud ka intellektuaalomandi õiguste kaubandus- aspektide lepingu (TRIPS) kontekstis¹⁶ Maailma Kaubandusorganisatsiooni (WTO) aruandes ning selgitatud, et piirang on vastuolus teose tavapärase kasutamisega juhul, kui piiranguga lubatud kasutus hakkab konkureerima

¹⁶ TRIPS lepingus nähakse samuti ette kolmeastmeline test ning argumendid selle sisustamisel võivad olla abiks ka Eesti autoriõiguse seaduse tõlgendamisel.

sellega, kuidas õiguste omaja harilikult oma teost eksploateerib, jättes see läbi õiguste omaja ilma sissetulekuta. WTO aruande kohaselt omab tähtsust ulatus, milles autoriõiguste omaja oma õigusi eksploateerib. Sõltuvalt sellest saab määrata, mis on teose tavakasutus vastaval turul. Teatud tingimustes ei ole olemasolev teose litsentseerimispraktika siiski küllaldaseks juhiseks hindamaks vaba kasutuse konkreetset mõju teose tavakasutusele (näiteks juhul, kui puudub tõhus õiguskaitse). Seega lisaks sellele, kuidas õiguste omaja kasutab hetkel oma teost tulu saamiseks, tuleb hinnata ka selliseid teose kasutamise viise, mis teatud tõenäosusega hakkavad tulu looma tulevikus. (WTO 2000: 48, 50; Correa 2007: 134–155)

- 6) Reprodutseerimine ei kahjusta põhjendamatult autori seaduslikke huve. Õiguskirjanduses on juhitud tähelepanu asjaolule, et sõnastus “ega kahjusta põhjendamatult autori seaduslikke huve”¹⁷ viitab asjaolule, et piirang võib kahjustada autori seaduslikke huve tingimusel, et see vastab kahele eelnevale tingimusele (vaba kasutus on otseselt seaduses ette nähtud ja see ei ole vastuolus teose tavapärase kasutamisega) ning see on proportsionaalne (Ricketson, Ginsburg 2006: 776). Huvide kaalumisel tuleb võtta arvesse, millistest eesmärkidest on kantud teadustöö erandile tuginemine. Keeleressursside kontekstis on selleks eesti keele kaitse ja arendamine, mis on oluliseks argumendiks erandi kasutamise seaduspärasuse üle otsustamisel.

4.4. Isikuandmete töötlemine teadusuuringu eesmärgil ja salvestamise nõusolek

Isikuandmete töötlemisel keeleressursside kontekstis on risk kõige enam maandatud, kui puudutatud isik on andnud teadva nõusoleku. Samas näeb isikuandmete kaitse seadus ette ka võimaluse isikuandmete töötlemiseks ilma nõusolekuta. Teadusuuringute eesmärgil võib andmeid ilma puudutatud isiku nõusolekuta töödelda eelkõige kodeeritud kujul. Kui isik on tuvastatav, siis tema isikuandmete töötlemine teadustöö eesmärgil peab põhinema ülekaalukal avalikul huvil. Andmete töötleja peab lisaks võtma organisatsioonilisi, füüsilisi ja infotehnoloogilisi meetmeid isikuandmete kaitseks (IKS § 16).

Kuigi salajane lindistamine on üldreeglina keelatud, on siiski mitmeid argumente, mis seda teatud kontekstis lubavad. Kui igasugune isikute eelnevalt teavitamata audio- või videosalvestamine oleks keelatud, siis ei saaks ka teha “Varjatud kaamera” laadseid huumorisaateid ja ka dokumentaalfilme. Antud temaatika on käsitlust leidnud Riigikogu istungil karistusõiguse kontekstis, kus M. Soosaar juhib tähelepanu asjaolule, et dokumentalisti jaoks on varjatud kaameraga töötamine möödapääsmatu. R. Lepiksoni vastus tõstatatud probleemile oli järgmine: “[p]alun väga, filmi terviseks, aga kasutada saad seda filmitut siis, kui selleks on antud nõusolek. Küsimus on selles, et neid toiminguid ei tehtaks salaja, vaid teavitataks neid osapooli, keda fikseeritakse”. (Riigikogu 2006) Tegemist on seega komplitseeritud õiguslike küsimustega, mis kuuluvad nii eraõiguse (autoriõigus, isiku kujutise kaitse) ja avaliku õiguse (põhiseaduslikud õigused, isikuandmete kaitse, jälitustegevus) valdkondadesse.

Minemata selle temaatikaga käesolevas artiklis sügavuti, võib keeleteadlastele teha järgmised soovitusel. Esiteks, alati kui on vähegi võimalik, tuleb isikut, kelle kõnet salvestatakse, eelnevalt teavitada ning küsida temalt teadev nõusolek. See tähendab, et isik saab aru, millistel eesmärkidel ja kuidas tema kõne salvestist kasutatakse. Teiseks, kui ikkagi on vajadus täiesti loomuliku kõne lindistamiseks, siis tuleb alati pärast salvestuse tegemist teavitada puudutatud isikut ja küsida temalt nõusolekut lindistuse kasutamiseks keeleressursside kontekstis. Kui isik ei ole nõus, siis tuleb kogu materjal kohe kustutada. Kolmandaks, tuleb vältida lindistamist olukorras, kus juba eelnevalt on väga üheselt selge, et isik ei taha, et tema juttu lindistatakse ning, neljandaks, peab lindistust läbi viima lindistaja enda juuresolekul. Äärmiselt probleemne on olukord, kui keeleteadlane varustab erinevaid ruume, kohti ja telefone pealtkuulamise seadmetega eesmärgiga teha hiljem keelelist analüüsi. Riigikohtu kriminaalkolleegium on mõnevõrra teises kontekstis rõhutanud, et “[e]nda telefonivestluse kuulamine ja lindistamine ei vasta ühelegi seaduses kirjeldatud järeletoimingule, sest enda telefonikõnet ei saa isik ise salaja ega varjatult pealt kuulata ega sellel eesmärgil lindistada. Neil põhjustel ei teki eraelulise vestluse salvestise kui tõendi lubatavuse otsustamisel küsimust võimaliku eraviisilise järeletoimegevuse kohta KarS § 137 tähenduses”. (Riigikohus 2009: punkt 10) Sellest otsusest võib kaudselt järeldada, et Riigikohus ei loe kuriteoks olukorda, kui lindistatakse vestlust, milles ka lindistaja ise osaleb.

4.5. Keeleressursside avalikustamise vajadus

Keeletehnoloogia riikliku programmi finantseerimistingimustes on nõutud, et saadud tulemused oleksid kõigile vabalt kättesaadavad. See on läbimõeldud nõue, eesmärgiga võimaldada kõigile potentsiaalsetele uurimis- ja arendustöö jätkajatele juurdepääs seniloodud ressurssidele, edendada seeläbi eesti keele tehnoloogilist tuge ja lõpuks suurendada keele ellujäämise tõenäosust, mis ju ongi riigi sellealase finantseerimistegevuse eesmärk. Samas on pea kõik keeleressursid nii nõudlikud oma kasutamiseks vajalike eelteadmiste ja oskuste suhtes, et “inimesel tänavalt” ei ole nendega mitte midagi peale hakata ja ressursside täiesti piiranguteta avalikustamine ei ole tegelikult hädavajalik nimetatud eesmärkide saavutamiseks. Kuna aga avalikustamine sisaldab autoriõiguste omajate huvide kahjustamise riski, tasub kaaluda ressursside levitamist üksnes keeletehnoloogide siseriigis korralikult autenditud kasutajatele ja korralikult vormistatud lepingute kaudu.

5. Kokkuvõte

Keeleressursside arendamine seondub komplitseeritud õiguslike küsimustega, mis kuuluvad nii eraõiguse (autoriõiguse) kui ka avaliku õiguse (põhiseaduslike õiguste, isikuandmete kaitse, järeletoimegevuse) valdkondadesse. Tegemist on temaatikaga, mille uurimise vajadus tuleneb otseselt praktikast. Praktika ootab probleemide kaardistamist, kehtivale õigusele tuginevaid lahendeid, aga juhul kui viimased puuduvad, siis teoreetilisi ja praktilisi tõlgendusi ning ettepanekuid õigusaktide täiendamiseks.

Autoriõiguslikud probleemid on keeleressursside arendamisel ühed kesksemad. Kuigi keeleressursside loomisel kasutatakse autoriõiguslikult kaitstavaid teoseid või nende osi, on nende kasutus erinev teose tavapärasest kasutusest. Erinevate teoste abil analüüsitakse eesti keelt ning luuakse keelekirjeldusi, mis on sisendiks keeletehnoloogiale. Teos kui selline on keeleressursside seisukohalt eelkõige “toormaterjal”.

Praegune autoriõiguse süsteem ei arvesta keeleressursside spetsiifikaga. Ei ole välistatud, et tulevikus on vajalik kehtestada eriregulatsioon ehk täiendav vaba kasutuse alus, mis soodustab enam keeleressursside loomist ja levitamist. Samas on ka kehtiva autoriõiguse süsteemi tingimustes võimalik keeleressursside arendada. Eelkõige omavad keeleressursside puhul tähtsust kaks teose vaba kasutuse juhtu: tsiteerimine ja nn teadustöö erand. Nende üldtunnused on kohaldatavad ka keeleressursside puhul. Alati tuleb silmas pidada nn kolmeastmelist testi ja hinnata vaba kasutamise riske.

Suulise ja spontaanses olukorras salvestatud kõne puhul tõusetuvad päevakorda isikuandmete ja privaatsuse küsimused. Selleks, et suulise kõne kogumine ja töötlemine oleks õiguspärased, on vaja küsida nõusolek isikult, kelle kõnet salvestatakse. Kuna loomulik kõne eeldab, et isik ei tea, kui tema kõnet lindistatakse, siis tuleb nõusolek küsida kohe pärast salvestuse lõppu. Nii isikuandmete töötlemise kui ka spontaanses olukorras salvestatud kõne kasutamise nõusoleku andmisel peab isik aru saama, millistel eesmärkidel ja kuidas tema kõnet kasutatakse. Kogu keeleressursside loomise protsessi käigus tuleb austada ja tagada isikute õigused.

Viidatud kirjandus

- AutÕS = Autoriõiguse seadus. RT I 1992, 49, 615 ... RT I, 28.12.2011, 1.
Berni kirjandus- ja kunstiteoste kaitse konventsioon. RT II, 1994, 16, 49.
Consultation on copyright 2012. HM Government; <http://www.ipo.gov.uk/consult-2011-copyright.pdf> (24.09.2012).
Correa, Carlos María 2007. Trade Related Aspects of Intellectual Property Rights: A Commentary on the TRIPS Agreement. Oxford: Oxford University Press.
IKS = Isikuandmete kaitse seadus. RT I 2007, 24, 127 ... RT I, 30.12.2010, 2.
Infoühiskonna direktiiv = Euroopa Parlamendi ja nõukogu direktiiv 2001/29/EU 22.05.2001 autoriõiguse ja sellega kaasnevate õiguste teatavate aspektide ühtlustamise kohta infoühiskonnas. – ELT L 167, 22.06.2001.
Intellektuaalomandi õiguste kaubandusaspektide leping. RT II, 1999, 22, 123.
JTS = Jälitustegevuse seadus. RT I 1994, 16, 290 ... RT I, 29.06.2012, 2.
KarS = Karistuseseadustik. RT I 2001, 61, 364 ... RT I, 04.04.2012, 1.
Kelli, Aleksei; Tavast, Arvi; Pisuke, Heiki 2012. Copyright and Constitutional aspects of digital language resources: The Estonian approach. – Juridica International: Law Review University of Tartu, XIX, 40–48.
Maailma Intellektuaalse Omandi Organisatsiooni autoriõiguse leping. RT II, 2006, 14, 39.
Pisuke, Heiki 2006. Autoriõiguse alused. Tallinn: KPMS & Partnerid.
Rickatson, Sam; Ginsburg, Jane C. 2006. International Copyright And Neighbouring Rights: The Berne Convention And Beyond. Oxford: Oxford University Press.
Riigikogu = X Riigikogu stenogramm. VII istungjärk. Neljapäev, 18. mai 2006, kell 10:00; <http://www.riigikogu.ee/?op=steno&stcommand=stenogramm&date=1147935600> (23.09.2012).

- Riigikohus = Riigikohtu kriminaalkolleegiumi otsus nr 3-1-1-5-09 (26. märts 2009); <https://www.riigiteataja.ee/akt/13167445> (23.09.2012).
- Sootak, Jaan; Pikamäe, Priit 2009. Karistusseadustik: kommenteeritud väljaanne. 3., täiendatud ja ümbertöötatud väljaanne. Tallinn: Juura.
- Walter, Michel M.; von Lewinski, Silke (Eds.) 2010. European Copyright Law: A Commentary. Oxford, New York: Oxford University Press.
- Vider, Kadri; Kelli, Aleksei; Pisuke, Heiki; Mõtsküla, Peeter P.; Tuulik, Triin 2013. Keeleressursside litsentsimise õiguslikud kaalutlused. – Eesti Rakenduslingvistika Ühingu aastaraamat, 9, 33–347. <http://dx.doi.org/10.5128/ERYa9.21>
- WTO = WTO. Report of the Panel. WT/DS160/R (15 June 2000); http://www.wto.org/english/news_e/news00_e/1234da.pdf (28.09.2011).
- Ülemaailmse Intellektuaalse Omandi Organisatsiooni asutamise konventsioon. RT II, 1993, 25, 55.

Arvi Tavast (Eesti Keele Instituut, Tartu Ülikool, Tallinna Ülikool) juhatab EKI keeletehnoloogia osakonda; uurimisvaldkond on oskussuhtlus ja selle keelefilosoofilised eeldused.
arvi.tavast@eki.ee

Heiki Pisuke (Euroopa Liidu Nõukogu, Tartu Ülikool), põhilised uurimissuunad on intellektuaalne omand (autoriõigus ja autoriõigusega kaasnevad õigused), tõlkimise õiguslikud küsimused, ülikoolide intellektuaalse omandi küsimused.
heiki.pisuke@ut.ee

Aleksei Kelli (Tartu Ülikool) põhilised uurimissuunad on intellektuaalne omand, innovatsioon, teadmispõhine majandus, digitaalsete keeleressursside õiguslikud küsimused.
aleksei.kelli@ut.ee

LEGAL CHALLENGES AND POSSIBLE SOLUTIONS IN DEVELOPING LANGUAGE RESOURCES

Arvi Tavast^{1,2,3}, Heiki Pisuke^{4,1}, Aleksei Kelli²

Institute of the Estonian Language¹, University of Tartu², Tallinn University³,
Council of the European Union⁴

The development of language resources entails a number of legal challenges in both private (copyright) and public law (constitutional rights, personal data protection, privacy). The topic is motivated by practical issues. In creating language resources and sharing them through international research frameworks, there is a pressing need for raised awareness of the challenges as well as solutions based on current legislation or, in their absence, theoretical and practical analysis and proposals for updated legislation.

A central issue in developing language resources is copyright. Although copyrighted works or parts thereof are used in language resources, this differs from their normal use. The works are only used to analyse language and compile various language descriptions to serve as input for language technology, while their creative and expressive purposes are irrelevant.

While the current system of copyright law does not specifically regulate language resources, their development is still possible to a certain degree. In addition to explicit contracts, applicable solutions include citing and the research exception. It is still always necessary to observe the three-step test and to assess the risks of free use. In the long-term perspective, it may be necessary to establish an additional basis for free use that would facilitate the creation and sharing of language resources.

Resources of spoken language and especially spontaneous language usage bring about the additional issues of personal data protection and privacy. To gather and process samples of spoken language legally, permission from the subject is needed. For spontaneous speech, a prerequisite is that the subject is unaware of being recorded – in that case, permission must be sought immediately after completing the recording. In both cases, subjects need to understand how and why their speech will be used. Rights of other persons must be respected and protected during the whole process of creating language resources.

Keywords: intellectual property, copyright, licenses, personal data protection, free use, research exception, citing