

SÕNALIIK LEKSIKOGRAAFI TÖÖLAUAL: SÕNALIIKIDE ROLL TÄNAPÄEVA LEKSIKOGRAAFIAS

Geda Paulsen, Ene Vainik, Maria Tuulik

Ülevaade. Sõnaliik on traditsiooniliselt keele kirjelduse osa, mis hõlmab sõna morfoloogilist ja süntaktilist käitumist, aga ka tähendust ning pragmaatilist funktsiooni. Vaatamata sellele, et sõnaliigid on tinglike piiridega ning keeleteaduslikult kaugel ühtsest või selgest nähtusest, on see kategooriana leksikograafias vajalik kui kompaktne viis andmaks ülevaadet sõna käitumisest kasutuses. Ka e-leksikograafias vältimatud automaatsed protseduurid nagu morfoloogiline ja süntaktiline märgendamine ning tähenduste ühestamine eeldavad kas konkreetselt või vaikimisi leksikaalsete üksuste eristamist. Käesolevas artiklis anname sissevaate sõnaliikide rolli leksikograafi töös. Vahendame tulemusi aastal 2019 tehtud uuringust, mille eesmärk oli välja selgitada eesti keelt kirjeldavate leksikograafide probleeme ning vajadusi seoses sõnaliikidega, aga ka ootusi keeletehnoloogilistele lahendustele sõnaliigi määramise hõlbustamiseks.*

Võtmesõnad: sõnaliigid, leksikograafia, eesti keel, keeletehnoloogia

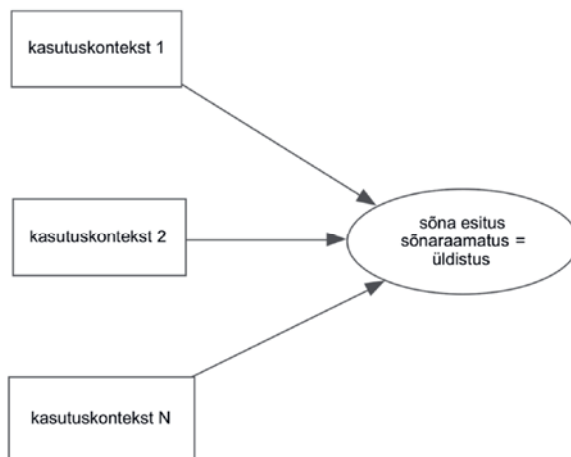
1. Sissejuhatus

Kuigi kategooriana on sõnaliik (st eristused nagu *nimisõna*, *tegusõna*, *omadussõna*, *määrsõna* jne) leidnud ohtralt kriitikat (nt Culicover 1999, Anward 2000, Croft 2001, Smith 2015, Kenesei 2018), ei saa me nende eristusteta läbi ei grammatikates ega sõnaraamatute koostamisel. Sõnaliik on leksikograafias paratamatult üks keskseid mõisteid, kuna sõnastikud ei esita ainult sõnade hääldust, tähendusi ja muutevorme, vaid avavad ka nende kasutusvõimalusi teksti tõlgendamisel ja loomisel.

Võrreldes loomuliku keelekasutusega on sõnaraamatu "tekst" kunstlik ja ebaloomulik; seda on nimetatud isegi äärmiseks veidruseks (*the extreme oddities*, Alexander 2015). Selline hinnang põhineb tõigal, et loomulikus keelekasutuses esinevad sõnad alati kontekstis, kuid sõnastikes on nad esitatud märksõnade loendina.

* Uurimust on toetanud Eesti Teadusagentuuri projekt PSG227 "Eesti keele sõnaliikide kategooriad korpusanalüüsi valguses". Täname artikli retsensente sisukate kommentaaride eest.

Seega kehtib eeldus, et sõnad on käsitletavad kontekstivabalt ja üksishaaval, neil on latentselt olemas kõik vajalikud karakteristikud (sh sõnaliik) ja leksikograafi ülesanne on need vaikimisi esinevad tõigad sõnaartiklis selgelt esile tuua. Just leksikograaf on traditsiooniliselt olnud see, kes vaatleb sõnade kasutusjuhtumeid kontekstis ning teeb selle põhjal üldistuse sõna omaduste, sh sõnaliigi kohta, kui sõnaraamatu struktuur selle esitamist ette näeb (joonis 1).



Joonis 1. Sõna esitus sõnaraamatus kui üldistus kasutusjuhtumitest

Leksikograaf kui ekspert suudab enamikul juhtudel vaevata teha üldistuse sõna morfoloogilistest, süntaktilistest ja semantilistest omadustest. Siiski on ka eesti keeles vahealaseid ja juhtumeid, kus sõna liik näib olevat mitmeti mõistetav, näiteks kui sõna kasutatakse kahes või enam eri sõnaliikidele tüüpiliseks peetavas kontekstitüübis või konstruktsioonis. Nii tekib näiteks küsimus, kas vormi *südamega* lauses *ta teeb tööd südamega* tuleks sõnastikus käsitleda nimisõna vormina või kasutusvalmis adverbina; või kuidas kirjeldada adjektiivi *hea*, kui ta esineb nimisõna funktsioonis: *tantsimine on hea* (adjektiiv) : *tantsimine teeb mulle head* (substantiiv)?

Mõneti teevad leksikograafide elu lihtsamaks sõnaraamatutöösse kaasatud mitmesugused elektroonilised andmebaasid ning võimalus tugineda laiadele tekstikorpustele. Teisest küljest on leitud, et üha suurenevate andmehulkade tingimustes pole sõnaraamatutegija peamine probleem enam mitte piisaval hulgal representatiivsete näidete leidmine, vaid pigem tohutu hulga materjaliga toimetulek (Alexander 2015). Korpuste ajastul küündib joonisel 1 näidatud kasutuskontekstide arv N nii mõnegi sõna puhul tuhandetesse ning sellise hulga kasutusnäidete käsitlemine oleks leksikograafi jaoks mõeldamatu ilma mingit laadi kasutustoeta. Kas infotehnoloogia saaks problemaatiliste sõnaliigijuhtumite korral abistada leksikograafi näiteks rakendusega, mis eeltöötleks suuremahulise korpuse andmeid selliselt, et leksikograafi töölauale ilmuks üksnes vajalikul määral ning piisavalt eristatud infot? Vastus eeldab selgust nii mõneski muus küsimuses. Kui palju puutub leksikograaf oma töös üldse sõnaliigiküsimustega kokku? Kuidas tulevad leksikograafid toime sõnaliikide piirialadele jäävate lekseemidega ning millised sõnaliigid valmistavad kõige suuremaid probleeme? Kas praegused sõnaliigijaotused on sõnastike

koostamise seisukohalt optimaalsed ning kuidas sõnaliigiinfot sõnaraamatus esitada? Kust saab leksikograaf probleemsete olukordade lahendamiseks abi? Kui laialdaselt kasutatakse olemasolevaid elektroonseid abivahendeid?

Neile küsimustele vastuste leidmiseks küsitlesime poolstruktureeritud intervjuu vormis kokku 23 leksikograafi. Artiklis anname ülevaate nii intervjuude käigus selgunud murekohtadest kui leksikograafide visioonidest võimalike lahenduste kohta. Esmalt tutvustame uurimuse teoreetilist ja sõnastikutööst võrsvat praktilist tausta ning anname infot küsitluse läbiviimise kohta. Järgneb kokkuvõtte intervjuudest välja koorunud hinnangutest ning sõnaliikide määramisega seotud probleemidest. Neljandas alajaotuses avame kõige probleemsemaks osutunud ja sõnaliigilt ambivalentsete keelendite olemust ning nende mitmetise käitumise tagamaid. Artikli lõpuosas kirjeldame lühidalt viise, kuidas leksikograafid sõnaliigiprobleeme lahendavad, ning küsitlevate ideid selle kohta, millist abi nad infotehnoloogialt probleemsete sõnaliikide käsitlemisel ootaksid, ning teeme kokkuvõtte sõnaliigiprobleemist leksikograafias.

2. Sõnaliikidest teoreetiliselt ja praktiliselt

Inimmeelele on omane paigutada nähtusi, sh ka sõnu, kategooriatesse, et luua üldistusi. Nagu kõik loomulikud kategooriad, ei ole ka sõnaliigid jäigalt üksteisest eristatavad. Sõnaliigipiiride hägususe tekkepõhjuseks võib pidada samu (optimaalsus)tegureid, mis mõjutavad keele arengut üldiselt – ökonoomsem on kasutada samu leksikaalseid ressursse nii paljudes üksteisest eristuvates funktsioonides kui võimalik, selle asemel et luua igale funktsioonile unikaalne leksikaalne tähistaja. Seejuures valitseb igas keeles ainuomane tasakaal paindlikkust (taaskasutus) ja kontrastsust (eristumine) suunavate hoobade vahel (Anward 2000) ning leitud tasakaal kajastub ka kehtivas sõnaliikide süsteemis. Nii on ka eesti keelele iseloomulik, et sõnaliikide vahel on sujuvaid üleminekuid ning leidub sõnu või sõnade muutevorme, mis vastavad rohkem kui ühe sõnaliigi tunnustele. Kuidas ja millistele tunnustele tuginedes sõnaliigid jaotuvad, sõltub niisiis keele ehitusest ja omapärast, aga lisaks ka liigituse eesmärkidest ning otstarbest.

Sõnaliik on kompleksne nähtus, mis hõlmab kõiki keele tasandeid. “Eesti keele süntaks” (Erelt 2017: 58) defineerib sõnaliigi mõistet nii: “Sõnaliigid on sõnade põhilised leksikaalgrammatilised klassid, mis määratakse sõnade semantiliste või pragmaatiliste, süntaktiliste ja morfoloogiliste tunnuste kombinatsioonide alusel.” Eesti keele puhul on sellisteks olemuslikeks ja kombineeruvateks tunnusteks morfoloogiline muudetavus või muutumatus, süntaktiline iseseisvus või sõltuvus, semantiline täis- või mitte-täistähenduslikkus. Sõnade kategoriseerimine ei ole aga ainult teoreetiline vahend keele kirjeldamiseks ning süstematiseerimiseks, vaid vältimatu ka erinevates rakenduslikes valdkondades. Üks selliseid alasid on leksikograafia, kus sõnavara liigendamise teoreetilised lahendused tuleb viia vastavusse n-ö tegelikkusega.

Kas, milliseid sõnaliike ja kui süstemaatiliselt märgendada, oleneb sõnaraamatu otstarbest ja sihtgrupist, mistõttu varieerubki eesti leksikograafias sõnaliikide esitus sõnaraamatuti. Märksõnad saavad reeglina sõnaliigimärgendi õppesõnastikes (näiteks naabersõnade ja põhisõnavara sõnastik), seletavates ning

sõnamoodustussõnaraamatutes (EKSS ja “Eesti keele sõnapered”). Kõige põhjalikumalt on sõnade liigid märgendatud eesti keele seletavas sõnaraamatus (EKSS 1988–2007, 2. trükk EKSS 2009) ning 2019 ilmunud eesti keele sõnaraamatus (EKS). Õigekeelsus-, nime-, termini-, murde- ning etümoloogilistes sõnaraamatutes märksõnade liike pigem ei esitata. Kakskeelsetes sõnaraamatutes on aga näha mõlemat tendentsi.

Lisaks sõnaraamatu otstarbele ja sihtgrupile mõjutab sõnaliikide esitamist traditsioon. Kahes eesti keele suures üldsõnaraamatus ehk seletavas sõnaraamatus ja ÕS-is on kujunenud välja omamoodi tööjaotus: esimeses on tava sõnaliike esitada, teises aga üldjuhul mitte. ÕS-is lisatakse siiski erandjuhtudel märksõnale sõnaliigimärgend, kui see seostub otseselt keelehooldejuhtnõoriga (nt *lemmik* nimis, omds, ÕS 2018). Samas ei ole ka seletavas sõnaraamatus sõnaliik kõigil juhtudel määratletud, näiteks allmärksõnade puhul. Need on põhilise märksõnaga tähenduslikult seotud üksused, mis pole (veel) tunnustatud täisväärtuslikuks märksõnaks, mistõttu ei kohaldata neile ka sõnaliigi määratlemise nõuet. EKSS 2009 ning EKS 2019 kasutavad allmärksõnana esitamise lahendust näiteks juhtudel, kus noomen on oma käändeparadigmast välja nihkunud ja esineb adverbilise või kaassõna funktsioonis. Selline on näiteks sageli määrusena esinev adessiivivorm *juhtimisel*, mille esitus Eesti Keele Instituudi uues keeleportaalis Sõnaveeb¹ (Sõnaveebi kohta vt Koppel jt 2019) viitab põhimärksõnale *juhtimine*:

see on ka vorm sõnast JUHTIMINE

juhtimisel

nii, et keegi korraldab ja suunab kellegi tööd või millegi toimimist
kelle

Arheoloogi juhtimisel kooti kangast ning õmmeldi nahast susse

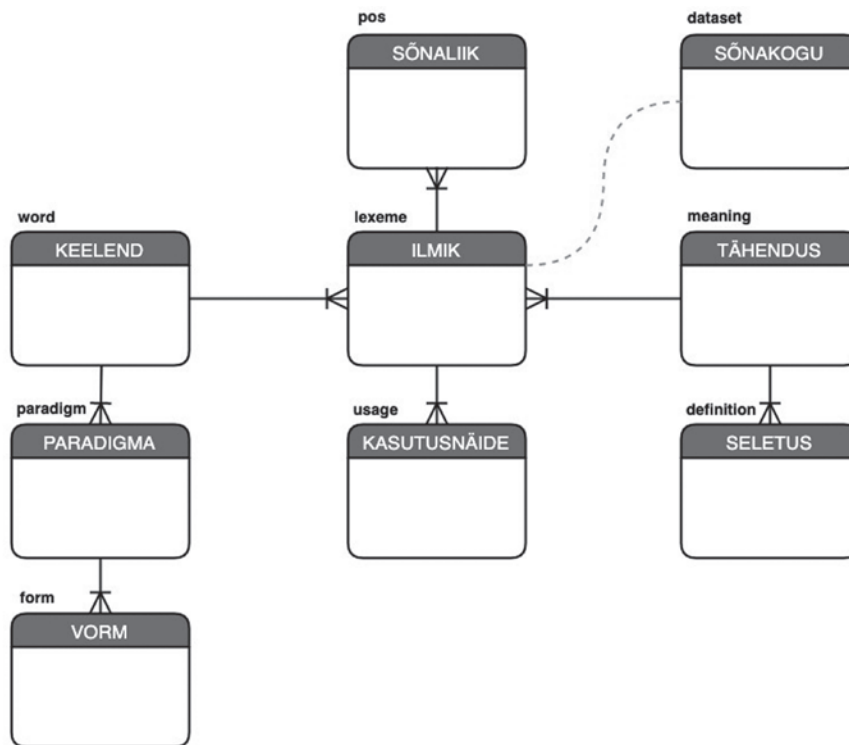
Selliste allmärksõnade ehk oma peamise tähendusüksusega veel seotud, kuid kasutuses juba teatud iseseisvumise saavutanud sõnavormide hulka on hiljutises seletavas sõnaraamatus tunduvalt suurendatud (Langemets jt 2018).

Eesti leksikograafias on keeletehnoloogilise arengu tuules toimumas muutused, mis mõjutavad sõnaliigi staatust leksikograafilise infoüksusena. Nimelt on üksiksõnastike koostamise asemel, milles teatavat infot (sh sõnade muutmise kohta) paratamatult süstemaatiliselt korratakse, võetud suund olemasoleva info taaskasutamisele. Eesti keele instituudi (EKI) uus sõnastikusüsteem Ekilex ühendab seni eraldi seisnud sõnastikud ja terminibaasid ühtsesse andmemudelisse. (Tavast jt 2018, Langemets jt 2018) Selles andmemudelis² (vt joonist 2) on sõnaliikidel oluline koht, kuivõrd nad kuuluvad leksikaalses andmebaasis kirjete ehk ilmikute³ jaoks välja töötatud oluliste/kohustuslike karakteristikute hulka. See tähendab, et igal ilmik-kirjel peaks ideaalis olema selle sõnaliigilise kuuluvuse märgend (või mitu), vastavalt sellele, kuidas teda on erinevates andmekogudes kirjeldatud. Andmemudelis eraldi allmärksõnu ette ei ole nähtud: kõik kirjelduse saanud sõnad on andmebaasis käsitletud keelendina (koos oma morfoloogiaga). Põhimärksõnaga või mõne teise keelendiga võib teda ühendada teatav seos (nt allmärksõna, kaashüponüümi seos).

¹ <https://sonaveeb.ee/search/est-est/detail/juhtimisel/1> (20.9.2019).

² <https://github.com/tripledev/ekilex/wiki/Andmemudel> (20.9.2019). Täname Arvi Tavastit nõuannete ja joonise eest.

³ Ilmik on määratletud kui keelendi ja tähenduse vaheline lüli, mis ühendab sõnakoguspetsiifilist infot, vt Tavast jt 2018.



Joonis 2. Väljavõte Ekilexi andmemudelist

Ekilexi andmemudelil on leksikograafilistele tööprotsessidele vaieldamatult suur mõju ja see puudutab ka sõnaliigi rolli. Sõnaliigist kui mudeli struktuuri osast kujuneb erinevate seniste üksiksõnaraamatute andmebaaside ühendamisel andmetöötamise jaoks vajalik üksus, mis hõlbustab näiteks tähenduste ühestamist. Kuna eri sõnaraamatutes on sõnaliikide märgendamisel olnud erinev kaal (nii liigitussüsteemides kui ka põhimõtetes, millal sõnaliigiinfot anda, millal mitte jne), on üks leksikograafide ülesandeid üksiksõnaraamatute ühisesse andmebaasi ehk ühendsõnastikku ületoomisel ka sõnaliikide kategoriseerimise läbimõtlemine. Sõnaliigiinfo jõuab ühendsõnastiku andmebaasist mitte ainult EKI Sõnaveebi portaali, kust ta on nähtav kasutajale, vaid saab sisendiks ka automaatsetele morfoanalüsaatoritele. Viimaste paremaks tööks peaks sõnaliigilised mitmesused olema kas minimeeritud või tuleks programmeerida süsteem ambivalentsete juhtumitega toime tulema. Mõlemal puhul on vaja rohkem teadmisi sellest, kus ja kuidas sõnaliigilised mitmetimõistetavused ilmnevad.

Ülalkirjeldatud teoreetilised orientiirid ning mõtteviisi muutus leksikograafiliste andmete kogumises, hoidmises ja töötlemises koos ulatuslike korpuste saamisega leksikograafi igapäevatöö lahutamatuks osaks moodustavad tausta, millele paigutub uurimus leksikograafide endi vaatenurga kohta sõnaliigiküsimustele.

3. Uuring sõnaliikide osast leksikograafi töös

Artikli analüüsiosa põhineb aastal 2019 toimunud poolstruktureeritud ekspertintervjuudel leksikograafidega. Intervjuude eesmärk oli kaardistada eesti keele sõnaliikide probleeme, mis valmistavad raskusi sõnaraamatute koostamisel ning rakenduslike lahenduste (tekstitöötuses kasutatavate morfoloogiliste ja süntaktiliste päringusüsteemide) väljatöötamisel. Intervjuud põhinesid avatud küsimustest koosneval küsimustikul ning kahel 5-pallisel skaalal (vt lisa 1). Küsitletavaid oli vestluse teemast (probleemid sõnaliikidega) ette teavitatud. Intervjuud toimusid isiklike kohtumiste käigus umbes poole tunniste vestluste vormis, mis lindistati ning mille sisu transkribeeriti.

Uuringus osales kokku 23 leksikograafi, neist 21 naist, 2 meest. Põhiosa küsitletavatest, kokku 20 leksikograafi, märkis töökohaks Eesti Keele Instituudi, ühe küsitletava leksikograafia-alane töö oli tehtud Tartu ülikoolis ning ühe MTÜ Eesti-Prantsuse leksikograafia ühingus. Töökogemus leksikograafina varieerus poolest aastast 48 aastani; ligi pooltel osalenutest oli üle kahekümne aasta pikkune sõnaraamatutöö kogemus⁴.

Küsitletute pädevuse ulatust näitab ka erinevate sõnaraamatute arv, mille koostamises või toimetamises nad on osalenud. Kokku 38 mainitud eri nimetuse seas on kolmandik üldsõnaraamatud (seletav sõnaraamat, naabersõnad, õigekeelsussõnaraamat jne), kolmandik kakskeelsed (eesti-soome, eesti-vene, eesti-prantsuse jne) ning ülejäänud mitmesugused spetsiifilisema iseloomuga sõnastikud (termini-, murde- ja etimoloogilised sõnastikud). Tähelepanuväärne on ka asjaolu, et pea pool küsitletutest (47%) on töötanud mitut tüüpi sõnastikega, veerandil piirdus kogemus mingit laadi spetsiifilise sõnastikuga, viiendikul oli kogemus vaid üldisest ükskeelsetest sõnastikest ning napp 4% oli senises töös kokku puutunud vaid kakskeelsete sõnastikega.

Valdav osa leksikograafidest (23-st 18) on töötanud elektrooniliste sõnastike koostamiseks mõeldud tarkvara abil (EELex), kahel on muu elektroonilise platvormi kogemus ning vaid kahel küsitletaval puudus elektroonilise sõnastikusüsteemiga töötamise kogemus. Suurem osa leksikograafidest (23-st 16), kasutab oma töös tekstikorpust – mainiti eesti keele ühendkorpust, veebikorpust, eesti keele õppekorpust, vana kirjakeele korpust, rahvaluulekorpust.

Sisu poolest koondusid intervjuud kolme põhiteema ümber: i) Milline on eesti keelt kirjeldavate leksikograafide nägemus sõnaliikide ja sõnaliigimärgendite vajalikkusest? ii) Millised probleemid kerkivad leksikograafi töös esile seoses sõnaliikidega? iii) Mida võtavad leksikograafid ette probleemide korral? Millised on nende endi visioonid võimalustest, kuidas infotehnoloogia saaks neid probleeme aidata lahendada?

Järgnevates peatükkides esitame sisulised kokkuvõtted aruteludes kõlanud ideedest ning arvandmed kahe uuringus kasutatud Likert-tüüpi hinnanguskaala kohta.

⁴ Võrdluseks: Euroopa leksikograafidest on sama pikk kogemus umbes kolmandikul, vt Kallas jt (2019: 524).

4. Kas sõnaliike on tarvis (märkida)?

Sõnaliigimärgendeid peavad leksikograafid üldiselt sõnaraamatu vajalikuks osaks. Sõnaliigiinfost kasusaajatena nimetati üksmeelselt kasutajat – esmajoones (võõr-keele)õppijat, aga ka õpetajat, keeleuurijat ja teisi igapäevaselt keelega tegelejaid, kuid ka sõnaraamatu tegijat ennast.

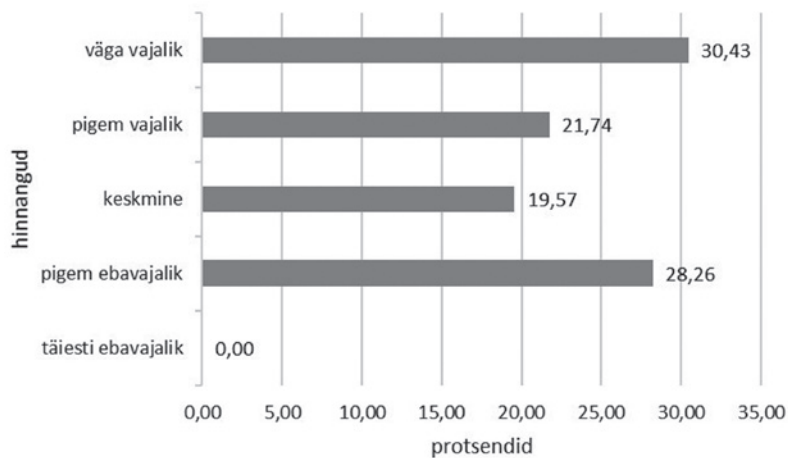
Leksikograafi töös mängib sõnaliigiline eristus tavalisest suuremat rolli näiteks siis, kui sõna kuulub mitmesse sõnaliiki ja sõnaartikkel liigendub selle järgi. Sõnaliigid on koostajale abiks ka tähenduste eristamisel ning üldisel lingvistilise materjali liigendamisel. Näiteks võib tuua EKSS-i esituse sõnast *võõras*, mis jaguneb nelja tähendusse, millest kahe puhul on ta adjektiiv ja kahel substantiivi tähisega. Sama sõna on eesti keele sõnaraamatus (Sõnaveeb 2019) liigendatud kolme tähendusse, millest kaks on adjektiivid ja üks nimisõna; samas esimesel, st peamisel omadussõnatähendusel ‘selline, kellest või millest mitte midagi ei teata, keda või mida ei tunta’ on nimisõnaline alltähendus ‘see, keda ei tunta’. On ilmne, et sõna üldtähenduslikku sarnasust lõhestab erinev sõnaliik.

Sõnaliigid on olulised ka selleks, et leksikograafi tööaluale jõuaksid adekvaatsed kokkuvõtted keelendi kasutamisest korpuses. Kuivõrd osa leksikograafidest kasutab sõnastike koostamisel automaatselt sõnavisandeid genereeriva tarkvara Sketch Engine (vt ka Kilgarriff jt 2004) abi, on tähtis, et morfoloogiliselt märgendatud korpuses, mida see tarkvara sisendina kasutab, oleks korrektne sõnaliigiinfo.

Üks traditsioonilisi viise sõnade iseloomustamiseks (eriti ükskeelsetes) sõnaraamatutes on olnud ka vormimoodustust kirjeldavad muuttüüpide numbrid, mis põhimõtteliselt võimaldavad eristada muutumatuid sõnu muutumatutest ning omavahel pöörd- ning käändsõnu. Leksikograafid leidsid siiski üksmeelselt, et üksnes muuttüübi number märksõna juures ei anna piisavalt infot selle sõnaliigi (ja seega kasutamise) kohta. Näiteks sõna *võõras* käändub ühtemoodi, olgu ta siis kasutatud adjektiiv või substantiivina. Ka avaldati arvamust, et muuttüübid on eesti keele õppijale keerukad ning et elektrooniliste sõnastike ajastul, mis lubavad tüübi numברי asemel kuvada kohe sõna põhivorme (nagu Sõnaveebis), on tüübinumbrite kui selliste aeg koguni ümber saamas.

Selgitamaks välja leksikograafide seisukohta sõnaliigiinfo vajalikkuse kohta parajasti käsil olevas sõnaraamatus, palusime intervjuueeritavatel anda oma hinnang skaalal: väga vajalik – pigem vajalik – keskmine – pigem ebavajalik – täiesti ebavajalik. Kokkuvõtte vastuste tulemustest annab joonis 3.

Joonisest nähtub, et keegi ei pidanud sõnaliikide märkimist sõnastikus täiesti ebavajalikuks, küll aga pidas ligi kolmandik neid sõnastikus “pigem ebavajalikuks”. Seega on üldpilt sõnaliikide (märgendite) vajalikkusest mõnevõrra polariseerunud, kuna teine kolmandik vastajatest pidas sõnaliikide märkimist “väga vajalikuks”. Siinkohal on sobiv meelde tuletada, et kolmandik sõnastikest, millega küsitletavad olid töötanud, olid ükskeelsed üldsõnaraamatud, samas kui teine kolmandik olid iseloomult spetsiifilised: etümoloogilised, termini- ja/või murdesõnastikud, kus seda liiki infot üldiselt ei ole pakutud (vrd p 3). Võib oletada, et vajalikkushinnanguid on mõjutanud just vastaja töökogemus eri tüüpi sõnastikega. Ka viitasid leksikograafid ise tõsiasi, et see, kas sõnaliigiinfo on vajalik või mitte, oleneb sõnaraamatu tüübist ning sihtgrupist (vt p 2).



Joonis 3. Sõnaliigi vajalikkushinnangud käsilolevas sõnastikus

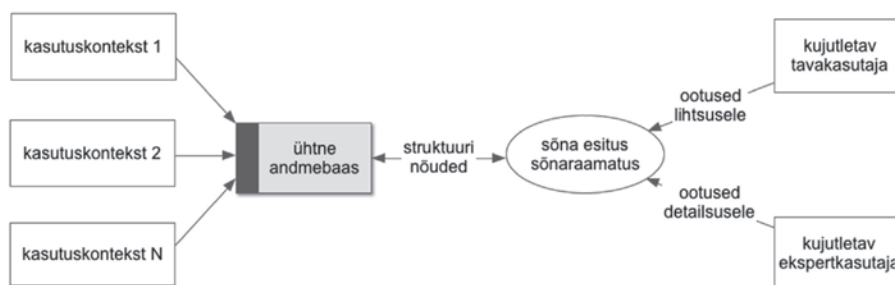
Osa leksikograafe leidis, et kuigi kõik sõnaraamatud ei pea sõnaliiki kajastama, oleks seda infoüksust täielikult kajastava andmekogu olemasolu siiski hea. Sõnaliigimärgendi puudumist sõnastikes, kus sõnaliigi näitamine on standardiks võetud, nimetati otsesõnu probleemist möödahiilimiseks. Väljendati seisukohta, et igal sõnal on põhimõtteliselt liik olemas, mõnikord on seda lihtsalt raske tuvastada ning vaja läheb konteksti. Märgendi puudumist peeti õigustatuks allmärksõnades, mille omadusi selgitatakse kasutusnäidetega. Toodi ka esile, et kõigile märksõnadele ei saa sõnaliigimärgendit parima tahtmisega juures lisada (nt mitmesõnalised väljendid, ühendverbide ja väljendverbide muutumatud osad, püsiühendid jne).

Rohkem kui pool leksikograafidest, kellel oli mitut tüüpi sõnaraamatu koostamise kogemus, pidas vältimatuks ja loomulikuks, et sõnastikus esineb märksõnu, millel on kaks sõnaliigimärgendit; tüüpilisim näide on siinkohal adjektiivide substantiivne kasutus (nt *kolm haiget*). Toodi esile, et kui sama väliskuju ja tähendusega sõna esinebki järjepidevalt kahesugustes konstruktsioonides, siis on ootuspärane tema esinemine kahesugustes suhetes ning kahe sõnaliigi märkimine on ka sõnastikus vajalik. Sedalaadi argumentatsioon tugineb eeldusele, et keele kirjeldus peab peegeldama keelt selle tegelikus kasutuses. Selge märgendus on kasutaja huvides: leiti, et parem on esitada kaks märgendit eksplitsiitselt, kui jätta kasutaja näitelausete põhjal mõistatama, mis liiki sõnaga on tegemist ja kuidas see lauses toimib. Esile toodi ka tähendusühtsuse argumenti: üks märksõna kahe sõnaliigimärgendiga on õigustatud ainult siis, kui tähendus jääb tõesti mõlemal samaks (näiteks muutumatute sõnade puhul määr- ja kaassõnalise kasutuse esitamine samas artiklis, nagu *all*).

Arvestatav hulk vastanuist leidis aga, vastupidi, et “ühel sõnal ei saa olla mitut liiki”, põhjendades oma arvamust erinevate argumentidega. Toodi esile, et sõna kasutamisega teises sõnaliigis varieerub paratamatult ka tähendus ja see on aluseks märksõna (või vähemalt allmärksõna) staatusele. Apelleeriti ka kasutajasõbralikkusele: kaks sõnaliigimärgendit ühe märksõna juures ajab vähemteadlikku kasutajat segadusse. Lähtuma peab siinkohal sõnastiku sihtgrupist – kellele (millise tasemega kasutajale) on sõnastik mõeldud.

Ekilexi sõnastikusüsteemiga kokku puutunud leksikograafid töid esile asjaolusid, mida dikteerib sõnavara selle täiuses ja mitmekesisuses kirjeldava andmebaasi struktuur. Sõnaliigiline kuuluvus on andmemudeli põhiüksuse ehk ilmiku (eristuva tähendusega seotud kirjeldus) süsteemne omadus. Üldjuhul kehtib seos: üks kirje andmebaasis kirjeldab märksõna kasutust eristavas tähenduses ja üheselt määratletavas sõnaliigis, kuigi vajadusel saab lisada ka mitu märgendit⁵. Praegu on võetud suund nii sõnaliigilise mitmesuse kui ka vaegmärkimise vähendamisele ning püütakse sõnaliigid selgelt eristada. Oluline parameeter on siin leksikograafide arvates tähendusnihe – tähenduse eraldumine põhitähendusest, tähenduse kitsenemine või abstraherumine. Märksõna kasutust iseloomustav näide jäägu intervjuueeritavate arvates sealjuures siiski illustratsiooniks ja sellega ei peaks asendama sõnaraamatu metakeelt.

Nii mitmese sõnaliigimärgendi poolt- kui ka vastuargumentide lähtekohad on niisiis ühelt poolt märksõna kasutuskontekst ning teiselt poolt kasutajasõbralikkus, aga ka tehnilised (andmebaas) ning sisulised (sõnaraamatu ülesehitus) struktuuri nõuded. Märksõnaartikli loomist mõjutavad aspektid võtab kokku joonis 4.



Joonis 4. Aspektid, millega leksikograaf peab sõnaartiklit luues arvestama

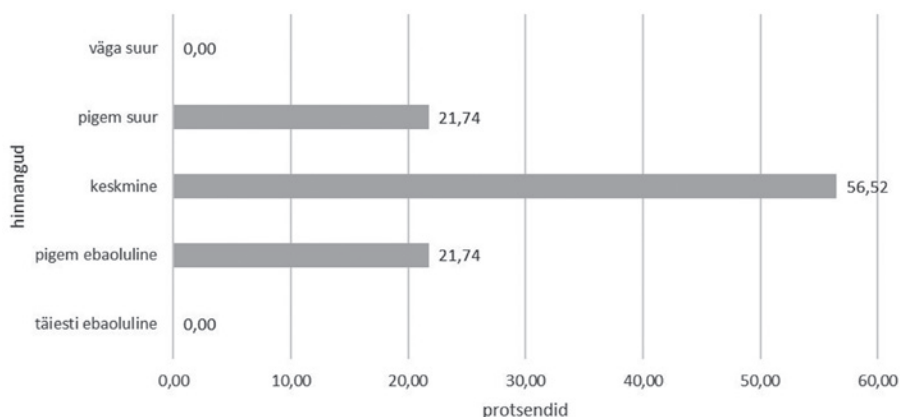
5. Leksikograafi probleemid sõnaliikidega

Sõnaraamatu kirjeldusobjektist, eesmärkidest ning sihtgrupist sõltuvad suuresti ka sõnaliikidega seotud probleemide iseärasused. Näiteks vana kirjakeele sõnastike ning korpuste loomise puhul kerkivad esile spetsiifilised sõnaliigiprobleemid, nagu vana kirjakeele sõnavara tänapäevasesest erinevad leksikaliseerumisastmed ning materjalinappus. Otsustustes ei saa lähtuda tänapäeva keeletunnetusest ning keeruline küsimus on näiteks, millal saab substantiivist afiksaaladverb (*andeks andma, tähele panema*) (vana kirjakeele sõnaliigiprobleemidest vt Habicht jt 2011). Ka murdesõnaraamatu koostamisel on sõnaliigi määramisel oma loogika, võrreldes vana kirjakeelega on aga materjali, mille põhjal otsust teha, rohkem. Etümoloogia-sõnaraamatus, kus olulisim on sõnatüve päritolu ning üldiselt sõnaliigi üle palju juurelda pole tarvis (ühesuguse tüvega vormid esitatakse ühes artiklis koos), on problemaatilisimad tuletised ning nende alussõnad; kahtluse korral eelistatakse kindlamat vormi ning sõnaliigivihje peitub sageli seletuses. Naabersõnade sõnaraamatu koostamisel tuleb aga lähtuda sellest, et kasutaja-keeleõppija leiaks vajalikud sõnad ning vormid võimalikult loogilisest kohast.

⁵ Üks-mitmele-vastavus on mudelis olemas kolmeharulise seosejoone näol, vrd joonis 2.

Leksikograafid, kes töötavad korpusetega (nii kasutajana kui ka ise korpusi luues), nendivad, et korpusest probleemkohti välja juurida ei ole lihtne. See muudab päringutegemise ja uurimise keerulisemaks, aga positiivsena nähakse, et nii saab jälile infole keelekategoriate ebaselgete piiride kohta. Sõnaliigid on näiteks eesti keele ühendkorpuses automaatselt morfoloogiliselt märgendatud, kuid märgendaja teeb vigu, näiteks lemmatiseerib valesti (tähejärjend *joon* (nimisõna) võib pakuda ühendi *joon õlut*). Sel juhul aitab märgendusest sõltumatute konkordantside vaatamine. Need leksikograafid, kes igapäevatoos sõnaliikidega tegelevad, peavad siiski ideaaliks, et sõnade liigitamisel valitseks selgus ning sõnadele saaks “lipiku” külge panna.

Palusime intervjuueeritavatel hinnata sõnaliikide määratlemise probleemi ulatust sõnaraamatute koostamisel võrreldes teiste protsessidega, märkides viiepallisel skaalal (väga suur – pigem suur – keskmine – pigem ebaoluline – täiesti ebaoluline) vastav aste. Vastajate hinnangud võtab kokku joonis 5:



Joonis 5. Sõnaliigi määramise suhteline probleemsus

Skaala mõlemad otspunktid ehk “väga suur” ja “täiesti ebaoluline” küsitluse tulemustes puuduvad. Küsitletavad hindavad sõnaliikide määratlemise probleemi teiste sõnaraamatu tööprotsesside seas pigem keskmise ning selle lähedaste astmete abil: 57% on valinud “keskmise” (13 vastajat) ning ülejäänud vastajad jaotuvad võrdselt “pigem suure” ning “pigem ebaolulise” vahel (kummalgi juhul 5 vastajat).

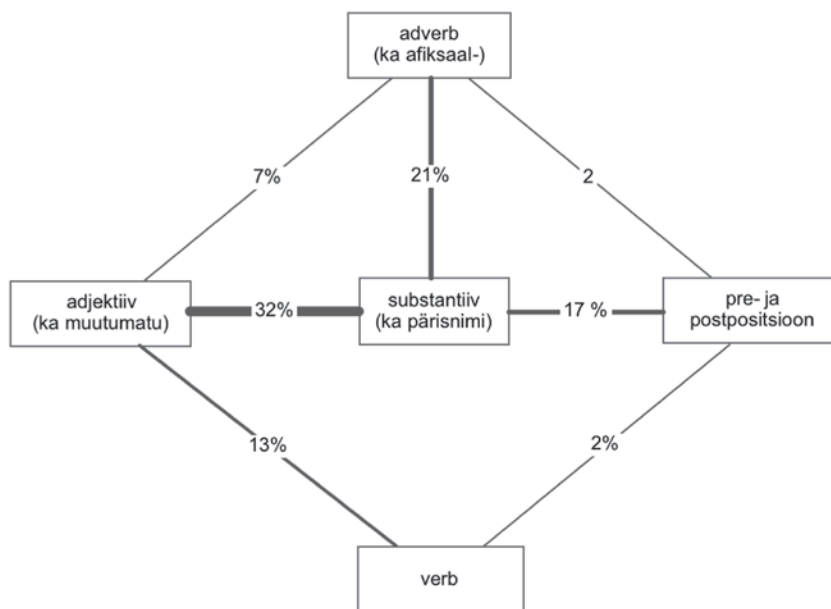
Tulemustest võib järeldada, et teiste tööülesannete seas moodustab sõnaliikide määramine leksikograafide jaoks küllaltki kaaluka osa tööst. Palusime leksikograafidel oma valikut suuliselt iseloomustada. Mitu leksikograafi (4) rõhutas, et probleem on pigem suur, kui sõnaraamatus tuleb sõnaliigid määrata; kui mitte, on see pigem ebaoluline. Teine vaatenurk lähtub selgetest ja ebaselgetest juhtumitest – üldjuhul suuri raskusi ei ole, kuid “keerulistel juhtudel pigem suur probleem”, ning üldhinnang kujuneb ümardatult keskmiseks. On ka arvamusi, et sõnaliigid on “[...] põhimõtteline probleem, sest sõnaliigitus ei lähtu ühtsest alusest”. Sõnaliikide määratlemiseks igatsetakse selgeid juhiseid (“Kui mingi asi on keeles olemas, siis tal võiks märk olla küljes.”). Ka nenditakse, et leksikograaf kohtab oma töös vahel sõnaliigiliselt ambivalentseid juhtumeid enne, kui grammatikad jõuavad neid käsitleda. Probleemide põhjuseks pakutakse muuhulgas eesti keele rikast morfoloogiat.

Küsitletavad peavad sõnaliikide määratlemist eriti keeruliseks sel juhul, kui on tarvis koostada sõnaliigipõhist märksõnastikku ning otsustada, kas tegu on näiteks märksõna käände- või pöördvormi või omaette üksusega.

Küsitluse tulemuste põhjal võib öelda, et keerulisim on sõnaraamatu koostamisel tabada ja määratleda kahes või enamais sõnaliigis esinevate või ühest sõnaliigist teise liikuvate sõna(vormi)de staatust. Suurimad murekohad võib kokku võtta järgmiselt.

- Millal kuuluvad ambivalentset, sõnaliigiliselt mitmeti mõistetavad sõnad juba selgelt kahte eri sõnaliiki? Millal teha sõnaraamatusse uus, iseseisev märksõna, muuta normingut (näiteks nimisõna *lemmik* adjektiivina aktsepteerimine)?
- Kuidas lahendada juhtumeid, mille kohta esineb erinevaid analüüsivõimalusi (nt kas *araabia* ühendis *araabia numbrid* on adjektiiv, muutumatu adjektiiv või pärisnime genitiivivorm)?
- Millal võib öelda, et sõna (vormi) tähendusnihe algse tähenduse suhtes on püsiv (nt *stardivalmis*, (*muutuste*) *lävel*)?
- Kui suur sagedus korpuses on piisavalt suur mõjutamiseks sõnaliigilist nihet?
- Kui sõnad kuuluvad mitmesse sõnaliiki ühe morfoloogilise klassi sees: milline neist on primaarne või "põhisõna" (nt nimisõna *ainurakne* ning omadussõnaline *ainurakne organism*)?

Küsitluse käigus püüdsime välja selgitada sõnaliike, mille vahel leksikograafidel on eriti keeruline otsust teha, julgustades vastajaid meenutama konkreetseid juhtumeid. Peaaegu kõik küsitletavad suutsid välja tuua näiteid, kokku 145, millest 127 olid erinevad. Näidete distributsiooni sõnaliigiti kujutab joonis 6 (arvestatud on peamisi sõnaliike, üksikjuhud, nagu interjektsiooni ja konjunktsiooni näited, on välja jäetud):



Joonis 6. Ambivalentsete sõnade distributsioon küsitluses toodud näidete põhjal

Leksikograafidele intervjuu hetkel meenunud keeleainese põhjal ilmneb, et mitmeti mõistetavate sõnaliikide seas on kesksel kohal substantiiv. Substantiividest lähtuvaid liigitusküsimusi tekitavad ühelt poolt suhe adjektiividega (*lemmik, vigur, koer, haige, uksealune*), teisalt nimisõnade adverbilised (*esirinnas, kõrini, eluga* 'kiiresti', *moega*) ja adpositsioonilised (*ajal, alusel, lävel, teel*) kasutusjuhud.

Verbikesksetest sõnaliikidest on olulisim adjektiivisuund (põhiliselt partitsiivivormid, aga on ka konverbivorme (*värvimata*), kokku 14 näidet), kuid leidub ka verbivormides adpositsioone (postpositsioon *arvates*) ning substantiive (*otsitav*), lisaks nendele vana kirjakeele adverbilised eitusverbide vormid (keelavad määr-sõnad *ära, ärgu* jne). Ka verbiklassisisesed liigitused valmistavad peavalu, kui tuleb teha otsustusi transitiivsete ja intransitiivsete verbide (kaaluda nt intransitiivsete verbide sisaldusobjektilisi kasutusi) ning aspektivõrude ja mitteaspektivõrude vahel. Tähelepanuväärselt ei ole intervjuueeritavatele meenunud verbi ja substantiivi suunalised probleemalad, nagu nominalisatsioonid, mis võivad samuti sageduse suurenedes ning tähendususerisuste tekkides tekitada vajaduse kaaluda iseseisva sõna staatust.

Alapeatükkides 5.1–5.3 kirjeldame täpsemalt kolme küsitluse materjali põhjal esilduvaima hädasate piiridega sõnaliigi ehk adjektiivi, adverbi ning adpositsiooniga seostuvaid aspekte.

5.1. Adjektiivi suhted teiste sõnaliikidega

Adjektiivide puhul on silmatorkav nende sõnaliigiline paindlikkus, mis enamasti väljendub sõnaliigilise transpositsiooni ehk sõnaliigimuutusena, mille puhul sõna liigub teise sõnaliiki nii, et lähtesõna ei kao, vaid säilib endisel kujul (Vare 2006: 199). See nähtus tuli tugevasti esile ka küsitluses, ning ilmselgelt ei piirdu see toodud näidetega, vaid peegeldab keeles laialdaselt toimivat malli: pea kõiki omadussõnu saab teistsuguses lausekonstruktsioonis (näiteks elliptilise väljajätu teel) kasutada ka nimisõnana. Leksikograafide ülesanne on nende juhtumite juures otsustada, kas ja kui võrd peaks sõnaraamatud neid paralleelseid kasutusvõimalusi eri sõnaliikidena fikseerima.

Nagu leksikograafide toodud näidete kokkuvõte (vt joonis 6) näitas, mainiti kõige sagedamini (59 korda) problemaatilisi sõnu, mille üks tõlgendusvõimalus on adjektiiv. Nendest omakorda kõige esileküündivama rühma (57, 6%) moodustavad sõnad, mille puhul toimub adjektiivide substantivatsioon ning adjektiiv toimib ka substantiivina, kuid ei kaota kasutusvõimalust adjektiivina. Nimetatute hulgas on struktuurilt lihtsaid tüvisõnu, nagu *hea, külm, noor, väike, haige*, aga ka tuletisi, näiteks *roheline, kasulik, kasutu, jurakas, kodutu*, ning selliseid, kus on kombineeritud liitmine ja tuletamine, näiteks *ainurakne, ahjualune, tuhvialune, kaitsealune*.

Neile "kahepalgelistele" sõnadele on iseloomulik osutada mingit omadust (nt *teismeline, jändrik*) ja paralleelselt ka omaduse kandjat – seega on tegemist sõnaliigi vaheldumisega seotud metonüümsel ülekandel põhineva süstemaatilise polüseemiaga (vt Langemets 2010: 159–161). Olenevalt süntaktilisest rollist lauses tõlgendame samu sõnu seega kord omadussõnana (täiendi või öeldistäitena, näiteks *teismeline tüdruk, see puu on jändrik*) või nimisõnana, kui nad esinevad kas sihitise funktsioonis (*tabati joobes teismeline*) või alusena (*vana jändrik peab südilt*

vastu). Leksikograafidel on adjektiivse substantiveerivate süntaktiliste mallide kohta hulgaliselt näiteid: *kasutu kasulikkusest; ma teen sulle head; kuidas ühendada kasulik meeldivaga; väiksed läksid magama, aga suured jäid üles*.

Enamatel juhtudel on suund adjektiivist (st omadusest) substantiivi (ehk omaduse kandja) poole, kuid ei puudu ka vastassuunalised näited, mispuhul traditsiooniliselt nimisõnana esinevad *koer, lemmik, vigur* ilmuvad fraasides, kus nad on nimisõna laiendid (*koer poiss, rahva lemmikud telesaated*). Sellisel juhul on nimisõnaga tähistatud nähtus mingi omaduse või omaduste komplekti iseloomulik kandja ning metonüümse ülekande najal (*totum pro parte*) kasutatakse sama sõna omaduste osutamiseks. Leksikograafid tõdesid, et mistahes substantiiv võib käituda adjektiivina (*kõige raamatum raamat*) ja mõni neist saabki lõpuks kasutussageduse tõustes adjektiivi staatuse (*lemmik*).

Veel üks probleeme valmistav rühm, mis piirneb ühelt poolt adjektiivide ja teiselt poolt substantiividega, on käändumatu omadussõna ehk täiendsõna, mispuhul pärisnime, näiteks kohanime või sellest kujunenud üldnime kasutatakse täiendi positsioonis iseloomustamiseks põhisõna vastava koha, kultuuri või ajastuga seotud nähtusi: *araabia, ladina, rooma, inglise, Soome, Rootsi, Balti*. Nende sõnadega kaasnevad komplitseeritud suure ja väikese algustähe ortograafia reeglid. Küsitluses kerkis esile ka nn tavaliste adjektiivide ja mõne muutmata adjektiivi eristamise kasutajasõbralikkuse küsimus, seda eriti (õppe)sõnaraamatu seisukohalt (*eri vs. erinev, eht vs. ehtne, väärt vs. väärtuslik*).

Adjektiivide sõnaliigilist transpositsiooni ilmestavad ka mitmed adverbiga piirnevad näited. Leksikograafid nimetasid sõnu (20,3% kõnealuselt rühmast), mis olenevalt kasutusest võivad funktsioneerida kas adjektiivi või adverbina, nt *kikkis, kumaras, kuplas, lokkis, pilves, sompus, purjus, alasti, abivalmis, stardivalmis*. Adjektiivses kasutuses esinevad need sõnad nimisõna laiendina (täiendi positsioonis, nt *purjus mees, sompus ilm, lokkis juuksed, abivalmis õpetaja*), adverbina tõlgenduvad aga juhul, kui nad funktsioneerivad lauses (seisundi)määrusena (sageli öeldistäite positsioonis, nt *meeskonnad on stardivalmis, rinnad on kikkis, juuksed on lokkis*) (seisundiadverbide ja adjektiivide suhtest vt Rätsep 1978: 50, Tiits 1982: 18). Lisaks võib adjektiiv (näiteks *kõva*) esineda ka kvantori täpsust või ligikaudsust näitava intensiivistava laiendina (*kõndisime kõva/tubli kilomeetri*) (vt Metslang 2017: 468).

Adjektiiv piirneb ka verbiga – selle rühma näidetest moodustavad lõviosa partitsiibid (14 näitest kokku 13), nagu *armastatud, armunud, tunnustatud, täiskasvanud, austav*. Partitsiipide adjektiivse kasutuse leksikaliseerumisastme määratlemine on siinkohal üks põhilisi probleeme (kas näiteks *masendav* ja *hävitav* on juba adjektiivid, nagu *meeldiv*?) ning väljakujunenud adjektiivsete partitsiibivormidegi puhul võib leksikograafide keeletaju olla erinev. Lisaks tuleb leksikograafil eristada kattuvaid vorme: *nu*-tuletiste käänduvad vormid partitsiibivormidest (*armunud noored, joobnud inimesed*) ja predikatiivi positsioonis partitsiibid liitjavormidest (*mees on surnud*). Samalaadseid küsimusi tekitavad ka adjektiivsed konverbiivormid (*maja on värvimata*). Adjektiivi ja verbi piirialadel on niisiis tegu nii transpositsiooni kui ka homonüümiaga.

5.2. Adverb

Eesti keele sõnaliigimuutusi iseloomustavad peamiselt sõnamoodustuslikud mehhanismid (liitelise tuletuse või konversiooni⁶ näol) (Vare 2006: 199). Kui leksikograafide sõnaliigiprobleeme ilmestama toodud adjektiivide puhul saab rääkida konversioonipõhistest protsessidest, siis adverbid ja adpositsioonid osutavad ka teistsugustele nähtustele sõnaliigimuutuse taga, mida saab iseloomustada käände- või pöördevormi grammatiseerumis- ja leksikaliseerumisprotsessidena. Semantilise käändevormi kivinemine ning täistähenduslike sõnade tähenduse abstraheerumine ongi eesti keeles põhiline adverbide ja adpositsioonide allikas (vt nt EKG II: 38, Habicht 2001, Grünthal 2003: 26, Erelt jt 2017: 448–452; leksikograafia vaate-nurgast Habicht jt 2011, Paulsen 2018 ja 2019), mistõttu eksisteerib pidevalt nn üleminekustaadiumis sõnu, mis esinevad nii täistähenduslike nimisõnade kui ka teatud kasutuses kinnistunud muutumatute sõnade ehk suhtesõnadena.

Adverb on küsitluses toodud näidete põhjal adjektiivi järel enim erinevat tüüpi kategoriseerimisprobleeme tekitavaid sõnaliike, millest leksikograafid on maininud kokku 42 sõna. Adverbil on seoseid substantiivide (*eluga, eesotsas*), adpositsioonide (*eesotsas, moel*), muutumatute (*alasti*) ja muutuvate (*stardivalmis*) adjektiivide, verbide (*eestvedamisel*) ja isegi interjektsioonide (*aamen*) ning konjunktsioonidega (rõhuline *kui*). Leksikograafidele tekitavad peamurdmist näiteks seisundi- ja viisidverbid, modaalsed mäarsõnad, afiksaaladverbid. Nii sõnaraamatu koostamise kui juba korpusanalüüsi staadiumis on küsitletavate hinnangul üks peamisi probleemkohti afiksaaladverb ning ühendverbide tuvas-tamine (kas on tegu vaba laiendi ehk adverbi või ühendverbi osaga (*ette, taha*)?). Küsimusi tekitavad seisundiadverbid ning nende sarjaline käitumine ehk vormide paljusus – milline kohakäändelistest variantidest on “lähtevorm”? Nagu eelmises alapeatükis mainitud, on mõnel seisundiadverbil lisaks kokkupuutekohti adjektiividega (*lokkis, alasti*).

Küsitluses esitati kokku 28 näidet adverbisarnaselt käituvast substantiivist (nt *tasuta*), millest 9 suhestuvad lisaks ka adpositsioonilise kasutusega (nt *eesotsas, viisil, (kellegi) kombel, (kellegi) küüsis*). Kõige sagedamini on mainitud sõnad inessiivis, kokku 16 seisundit või lisaks ka viisi ja/või ruumisuhet märkivat (*elus ‘iaal’, kuplas, suukorvis, eesotsas*). Järgnevad viisi (*kombel, moel, viisil*) ja ruumisuhet (*peal*) väljendavad adessiivvormid (4), viisi (*eluga, südamega*) ning seisundit (*moega*) väljendavad sõnad komitatiivis (3), kaks viisidverbilist abessiivivormi (*tasuta, viperusteta*), üks viisi väljendav elatiivivorm (*südamest*) ning üks seisundiadverb terminatiivis (*kõrini*). Lisaks viisi ja seisundit väljendavatele vormidele nimetati ka afiksaaladverbilist nimisõna illatiivis (*kraesse*) ning allatiivis (deverbaalne tegevusprotsessi väljendav substantiiv *tarvitusele*), samuti üht saavas käändes vormi (*vastukaaluks*).

Adjektiivi ja adverbi piirialast töid intervjueeritavad kokku 11 nende kahe sõnaliigi funktsioonis esinevat näidet: adjektiivide rühmas (p 5.1) juba käsitletud seisundiadverbide transpositsioonijuhtumid *kikkis, kumaras, kuplas, lokkis, pilves, purjus, sompus*. Mõned neist sõnadest on kujunenud nimisõnade nt *lokk, pilv* kohakäändelistest (inessiivi)vormidest ja neist esineb ka illatiivi meenutavaid kasutusi: *ilm läks pilve, juuksed tõmbusid lokki*. Lisaks mainiti sõnu *valmis, abi-valmis, stardivalmis, alasti*.

⁶ Termin *konversioon* all mõistetakse üldjuhul sõnade moodustamist tuletusliiteta sõnaliigi ja süntaktilise funktsiooni muutmise teel (EKG I: 419–420, Vare 2004: 39, Kasik 2012).

Küsimusi tekitab ka liitsõnade (*külavahel, stardivalmis*) või nende osade (nime-tavalise liitumisega intensiteedisõnad nagu *vesi-, tuli-*) liik. Viimane mall annab võrdlusel baseeruvaid värviliitsõnu ning moodustab avatud süsteemi; *tuli*-tüüpi nimisõnad esinevad mallis omadussõna täiendosana intensiivistavate partiklitenä (vt Oja 1995: 812). Sõnade liitmist ning liitsõnade leksikaliseerumisprotsesse võib pidada eesti keelele omaseks nähtuseks, milleks on keeles eeskujuna olemas liitmäärsõnade struktuurimall ning sellele lisaks ka liitkaassõnade kasutust toetav liitkaassõnakonstruktsioon (Jürine, Habicht 2013).

5.3. Kaassõnadega piirnevad juhtumid

Küsitletavad mainisid 24 korral sõnavorme, mis võivad ühe variandina olla tõlgendatavad adpositsioonina. Kirjanduses on osutatud, et kaassõnad (nagu ka adverbid) ongi sõnaliigina kujunenud täistähenduslike sõnade peamiselt kohakäändelistest vormidest grammatiseerumisprotsessi käigus. Nimisõnast lähtuvate juhtumite puhul toimub tähenduse abstraherumine, misjärel jääb fraasis leksikaalset tähendust kandma üksnes seostuv nimisõna, mida hakatakse tõlgendama vastse kaassõna kohustusliku laiendina. Kuna tegu on protsessiga, mis toimub keeles pidevalt, siis leidub kaassõnade (nagu ka adverbide) seas reliktna samatüvelisi sõnu, mis moodustavad kohakäänete sarju (nt *otsa, otsas, otsast*), ning terve hulk sõnu on alles kaassõnastumise protsessis.

Erelt jt (2017: 448–452) toovad kaassõnastumise näidetena valdavalt adessiivi vormitunnusega sõnu, millest arvestatav osa on funktsioonilt evidentsiaalsust ehk teate vahendatust osutavad vormid: (*kellegi*) *ütlusel, teatel, arvamusel, kinnitusel* (vt ka Sepper 2006: 54–56). Sahkai (2008: 183) on loetlenud veelgi samas funktsioonis kasutatavaid sõnu: *ennustusel, hoiatusel, järeldusel, kartusel, lubadusel, meenutusel, mäletamisel, rõhutusel, seletusel, selgitusel, veendumusel, viitel, viitel*. Leksikograafide küsitluses tulid sellest rühmast esile *vahendusel, väitel, sõnul* ning *andmetel*.

Mõneti sarnase rühmana võib märgata ka kaassõnastuvaid vorme, mis väljendavad koos nimisõnast laiendiga tahte vahendamist: (*kellegi*) *tahtel, soovil, palvel, loal, käsul*. Veel üks aktuaalne valdkond näib olevat (kollektiivse) tegevuse initsiatiivi kirjeldamine, mille tarbeks on kasutusele tulnud piltlikud liitsõnad (*kellegi*) *eestvõtmisel ja eestvedamisel*, aga miks mitte ka *taganttorkimisel, otsusel, sundimisel* (Sahkai 2008: 183). Samast motivatsioonist on ajendatud ka leksikograafide mainitud sisekohakäändeline (*kellegagi*) *eesotsas* (nt *Sportlased eesotsas treeneriga nõuavad paremaid tingimusi*). Enamik eespool loetletud “uutest” kaassõnadest on deverbaalse taustaga, st tähistavad tegevusi ning eeldavad seetõttu, et seostuv nimisõna tähistaks nende tegevuste subjektiks olevat teo- ja tahtevõimelist elusolendit.

Teine tekkivate kaassõnade allikas on verbid, peamiselt *des-* ja *-mata* vormid (konverbid) (Erelt jt 2017: 451, nt (*millegagi*) *võrreldes, (millestki) alates, (millegi) vaatamata, (millestki) hoolimata*). Nendegi puhul on ähmastuva tegevustähendusega verbitüvele liitunud käändelõpud (vastavalt inessiivi -s ning abessiivi -ta). Seega on verbid omandanud esmalt substantiivi ning seejärel kaassõna tunnuseid. Mõneti teistsugused on abstraktsemaid suhteid väljendavad näited (*millegi*)

alusel, ajal, lävel, teel. Pea kõigi puhul on aimatav ruumilis-ajalisele taustale osutav piltlikkus.

Kaassõnade kujunemine täistähenduslikest sõnaliikidest on protsess, mis on kindlasti keeles laialdane ja mis on ajendatud vajadusest väljendada varasemast rohkem teatud tüüpi suhteid, nagu teate või tahte vahendamine, initsiatiivi osutamine, sündmuste tausta ja juhtumisviisi ning tegevuslike asjaolude kirjeldamine. Näib, et kaassõnastumas on eeskätt uurivale ajakirjandusele omast täpsustaotlust teenivad väljendid.

6. Lahendused ja ideekaeve tulemused

6.1. Kuidas leksikograafid sõnaliigiprobleeme lahendavad

Küsitluse kolmas peamine küsimuseasetus on seotud viisidega, kuidas leksikograafid sõnaliikide kirjeldamise ja määratlemisega toime tulevad ning milliseid oma tööd lihtsustavaid lahendusi nad ootaksid keeletehnoloogialt. Käsitlеме esmalt samme, mida leksikograafid sõnaliigi määramisel esinevate keeruliste olukordade puhul ette võtavad.

Kui sõnaartikli loomisel tekib raskusi märksõna sõnaliigi mõistmisel ja määramisel, sirvivad leksikograafid kõigepealt läbi olemasolevad sõnaraamatud: esimesena pööratakse seletava sõnaraamatu kui põhjalikema sõnaliigiinfoallika poole, aga kontrollitakse ka ÕS-i, murdesõnaraamatute, “Väikese vormisõnastiku” (Viks 1992a–b) andmeid ning kasutatakse keeleportaali Sõnaveeb. Vastuseid otsitakse keeleteaduslikest uurimustest, mis käsitlevad üldteoreetiliselt sõnaliikide muutmist või teatud sõnade juhtumiuuringuid. Piiripealsete juhtumite käsitlusi otsitakse ka grammatikatest (“Eesti keele süntaks” 2017, EKG) ja käsiraamatutest, arvestusega, et need võivad siiski kasutus põhise sõnaraamatu vaatenurgast aegunud olla.

Oluline on võtta arvesse kolleegide tööd. Parasjagu koostatava sõnaraamatu andmebaasist otsitakse analoogseid juhtumeid, et näha eelnevaid lahendusi või koguda kokku analüüsimist vajavad juhtumid. Leksikograaf ei tööta enamasti üksi, nii on aruteludel oma töörühmaga suur kaal – erinevate arvamuste ja põhjenduste korral võetakse otsustuse aluseks rohkem hääli saanud arvamus. Parimal juhul on töörühmal argumendid, kriteeriumid ja kokkulepped selgelt fikseeritud. Sõnaliigi muutmise otsust kergekäeliselt ei tehta – iga otsustus tekitab juurde piire ja ebamääraseid alasid. Näiteks kui liigitada adverb *purjus* süntaktilise jaotuse põhjal lisaks ka adjektiiviks (*purjus mees* vs. *ta on purjus*), tekib sama küsimus terve hulga adjektiivide suhtes, mis käituvad lauses sarnaselt.

Sõnaliiki on kõige kindlam määrata (kon)tekstis – küsitluses kõlas arvamus, et kontekstivabalt ei saagi sõnal liiki olla ning sõnaraamat alati üldistab. Sõna käitumise tabamiseks kasutavad leksikograafid süntaktilisi/morfoloogilisi/semantilisi teste, näiteks adjektiivikandidaadile täiendi lisamine (adjektiivi ette saab lisada adverb *üksna särav*) ja moodustada võrdlusastmeid: *säravaim täht*; vrldl **väga olev*, **olevaim*) või sõna asendamine lauses sünonüümiga. Mõeldakse välja näiteid lausemallidest, millesse sõna võiks sobida. Kui enda keeletajus tekib kahtlusi, küsitakse arvamust sõpradelt-sugulastelt. Semantiliselt analüüsil aitavad leksikograafe ka entiteedi ja tähendussuhete kombinatsioone kokkuvõtavad semantilised tüübid (Langemets 2009).

Kõiki võimalikke kontekste ei suuda ega saagi leksikograaf ise välja mõelda. Sõnade kasutust vaadatakse korpustest, näiteks ühendkorpusest rakenduse SketchEngine'i kaudu, kus saab näha sõna(vormi) süntaktilist distributsiooni ning sagedusi, esinemist kollokatsioonides ja konstruktsioonides. Sõnaliigi seisukohalt on kasu nii sõnavisandi infost (näitab mh, kas sõna mingi vorm on teistega võrreldes sagenenud) kui ka konkordantsiridadest (näitab, kas sõna muutub või mitte). Spetsiaalselt leksikograafidele mõeldud keskkonnas SketchEngine on tööriist nimega "lempos", mis näitab vormile automaatselt omistatud lemmat ning sõnaliiki (Lem+PoS). Homonüümia tõttu tekkivate vigade pärast ei saa seda funktsiooni aga lõpuni usaldada. Tööriist "Most common collocation pairs" tuvastab naabrusõnade seast muuhulgas hägusaid ja üleminekufaasis üksusi. Neid vahendeid kasutab kokkuvõttes üsna vähe leksikograafe. Nenditakse, et sõnaliigikahtluste korral korpusest kindla liigi alusel otsingut teha ei saa (ei ole päringut "anna sõna *jultunud* adjektiivsed või substantiivsed kasutused"). Piiripealsete juhtumite sõelumisel suhtutakse korpusmaterjali reservatsiooniga: otsitav sõna(vorm) võib olla korpuses valesti märgendatud või vahel üldse puududa. Vaadatakse ka Google'i otsingutulemusi või mõne kindla väljaande keelekasutust ("kuna otsimootor annab väga palju prahti").

Leksikograafid kasutavad oma töös ka teisi keeletehnoloogilisi rakendusi. Lihtsamad ja selgemad sõnaliigilised vihjed saab kätte eesti keele morfoloogilisest analüsaatorist ja süntesaatorist ESTMORF, mis määrab sõna struktuuri (tüvi, järelliide, lõpp), sõnaliigi ja käände või pöörde. Mõni leksikograaf kasutab kahtluste korral ka EKI reeglipõhise morfoloogia demo⁷ abi, millel on "Väikese vormisõnastiku" (Viks 1992) klassifikatsioonidele vastav sõnaliigi- ja tüübituvastuse funktsioon (seab algvormis sisendsõna vastavusse tema muuttüübi ja sõnaliigiga). Kasutatakse ka eesti keele süntaksianalüsaatorit⁸. Esile võib tõsta Tartu ülikooli vana kirjakeele töörühma märgendamise tarkvara Vakker (vt Prillop 2004), mis võimaldab muuhulgas retrospektiivselt mõne näitelause põhjal vaadata, kuidas käsilolevaga sarnaseid juhtumeid on eelnevas töös tõlgendatud.

6.2. Kuidas sõnaliigitamist lihtsustada ja milline võiks olla tehnoloogia roll?

Intervjuu lahendusteosa koondus küsimuste ümber, kuidas leksikograafi tööd sõnaliigi määramisel hõlbustada ning millised on selle protsessi automatiseerimise võimalused. Püüdsime suunata leksikograafe kujutlema võimalikku keeletehnoloogilist abivahendit, mis aitaks määratleda sõnade liiki.

Osa intervjuueeritavaid reageeris potentsiaalse digitööriista mõttele skepsisega või ei osanud selget arvamust kujundada. Osalt võib selle taga olla mulje, et küsiti juba valmis lahendusi, aga teisalt väljendati ka teatud usaldamatust digitaalsete keeletööriistade suhtes üldiselt. Võib oletada, et skepsise taga on küsitletavate senine kogemus (ebatäiuslike) IT-lahendustega, mis sunnivad kasutajat end süsteemi järgi kohandama või koguni süsteemi tehtud vigu parandama, mida aga soovitakse vältida. Kõhklusi tekitas ülesande tehniline teostatavus: väljendati seisukohta, et sõnaliigi määramiseks vajalik automaatne morfoloogilise ja süntaktilise analüüsi

⁷ <http://artur.eki.ee/morf/> (20.9.2019).

⁸ <http://kodu.ut.ee/~kaili/parser/> (20.9.2019).

tulemus ei saa olla märkimisväärselt parem (praegusest) korpusemärgendusest. Kaheldi ka selles, kas korpusest (“sõnade hunnik”) põhimõtteliselt saaks vastused seni lahendamata teoreetilistele ja praktilistele probleemidele. Teisalt tundus mõne spetsiifilise sõnastiku ainestik (murded, vana kirjakeel) tegijaile liiga ebastandardne ning piiratud ulatusega, et sellel korpuspõhist tööriista isegi mõttes rakendada.

Teine osa küsitletavaid esitas küllaltki selgeid ootusi ja lootusi sõnaliigi määramise protsessi lihtsustamise kohta. Need, kes harjunud töötama korpusega SketchEngine'i vahendusel, oskasid unistada samalaadsest, kuid täiuslikumast tööriistast, mis aitaks problemaatilise üksuse distributsiooni korpusest välja võtta ja esinemust oma silmaga vaadata. Soovitakse, et tööriist teeks andmete otsimise töö leksikograafi eest ära, et leksikograaf saaks langetada põhjendatud otsuseid. Korpusmaterjali põhjal tahab leksikograaf näha statistikat: kui palju mingi vorm esineb ja protsentuaalselt millise sõnaliigi tüüpilises funktsioonis. Antakse endale aru, et keel on pidevas muutumises – leksikograafid näeks meelevdi, et abivahend näitaks, kui kaugel arengus on mingi küsitav sõnavorm, näiteks kas sõna *eluga* on rohkem nimisõna või adverb. Korpusmaterjali põhjal soovitakse näha üldistust, mida võiks visualiseerida näiteks skaala, kaalukausside või hierarhia kujul. Ideaalis hõlmaks korpusanalüüs lisaks morfoloogilisele märgendusele ka süntaktilist distributsiooni (näiteks teeks kindlaks, kas sõna laiendab verbi või esineb täiendi positsioonis, nagu *kõnnib kumaras* (adverb) : *kumaras* (adjektiiv) *selg*).

Samas tahavad leksikograafid olla veendunud automaatse abimehe argumentide vettpidavuses. Ühest küljest soovitakse kindlasti näha otsustuse aluseks olevat autentset algmaterjali – tekstinäiteid, tõendeid (näiteks konkordantside kujul) keelendi esinemusest kontekstis, teatavale sõnaliigile iseloomulikus süntaktilises ümbruses. Teisest küljest soovib osa leksikograafidest selgelt näha kriteeriume, millele otsustust rajada – küsimusi tekitava sõnavormi sagedus, tähendusnihe olemasolu, süntaktiline positsioon, millele otsitav sõnavorm tekstis vastab ja mil määral. See tähendab, et abi oodatakse nii kvalitatiivsel kui ka kvantitatiivsel analüüsil. Kokkuvõtvalt võib intervjueritavate mõtted jagada kahte ossa (vt ka Paulsen jt 2019).

1. **Kvalitatiivsed lahendused** ehk abistava info kättesaadavaks tegemine:
 - a) andmebaas, mis koondaks problemaatilised juhtumid ning kõik (ka erinevad) otsustused;
 - b) leksikaliseerumis- ja grammatiseerumisskaalad – selgelt määratletud kriteeriumid, millele vastavust saaks leksikograaf sõna liiki määrates hinnata. Skaalad võiks olla ühendatud moodulina leksikograafi töölauale;
 - c) süntaktilised, morfoloogilised ja semantilised testid, mille abil leksikograaf saaks kahtlase vormi kuuluvuse kindlaks teha.
2. **Kvantitatiivsed lahendused** ehk korpuspõhine nutikas “sõnaliigimääraja”:
 - d) toob kokku ainestiku küsitava vormi kohta, teeb statistilisi kokkuvõtteid selle morfosüntaktilise käitumise kohta;
 - e) teeb võrdlusi sõnaliikide “tüüpilise profiiliga” etteantud kriteeriumite põhjal;
 - f) on seseisev tööriist, mis raputab ise korpusest välja sõna(vormi)d, mis võiksid saada iseseisva sõna staatuse. Ekstraheeritud vormid võiksid olla esitatud koos tüüpiliste mustritega (käänded, milles vorm sagedasti esineb, seosed naabersõnadega). Sellele järgnevad toimingud (d) ja (e).

Sõnaliigimääraja võiks niisiis tulevikus parandada automaatse morfoloogilise analüsaatori tööd: see teeks vähem vigu ega pakuks välja ebakohaseid analüüse vormidele, mis on kinnistunud omaette sõnaks (nt *kohal* on tõenäolisemalt abisõna kui alalütleva vorm sõnast *koht*). Leksikograafi abivahend peaks põhimõtteliselt olema samasugune nagu SketchEngine, ainult perfektssem; rakendus peaks pakkuma soovitud, aga otsuse langetab alati leksikograaf.

7. Kokkuvõtteks

Uurimus püüdis selgitada sõnaliigi kui kategooria rolli tänapäeva leksikograafias ning kaardistada eesti keelt kirjeldavate leksikograafide probleeme ja vajadusi seoses sõnaliikidega, sealhulgas nende ootusi keeletehnoloogilistele lahendustele sõnaliigi määramise hõlbustamiseks. Analüüs põhineb 23 leksikograafiga poolstruktureeritud intervjuu vormis peetud vestlustel, mis andis intervjuueeritavatele võimaluse väljendada oma mõtteid sõnaliigiga seotud teemadel. Leksikograafide refleksioonidele toetudes võib öelda, et sõnaliik on mõtteline, abstraktne kategooria, mis on leksikograafi töös alati taustal. Sõnaliik annab leksikograafidele materjali süstematiseerimisel raamistikku, millele tugineda – seda ka siis, kui leksikograaf oma töös otseselt sõnaliike ei määra. Teisalt mõjutab sõna liik otseselt sõnaartikli ülesehitust sõnaraamatus.

Sõnaliikide määramisega tegelevatel leksikograafidel tuleb (sageli piiratud aja jooksul) leida praktiline lahendus teoreetiliselt mitmetahulistele küsimustele. Oma töös sõnaliikidega peab leksikograaf arvestama mitme aspektiga. Ühelt poolt peab sõnaartikkel kajastama sõna esinemist eri kasutuskontekstides, tehes üksikjuhtumite põhjal üldistuse, teisest küljest aga vastama vastandlike huvidega kasutajate vajadustele (lihtsus või detailsus). Kolmanda faktorina peab leksikograaf võtma arvesse ka ühtse andmebaasi (Ekilex) struktuurinõudeid. Lisaks neile aspektidele peab analüüs vastama tänapäevastele üldkeeleteoreetilistele arusaamadele.

Uuringu tulemustest selgus, et sõnaliikide määramisel on leksikograafide töös küllaltki kaalukas osa, eriti juhul, kui koostatavas sõnaraamatus on vajadus sõnaliike märgendada. Keerulisima küsimusena kerkis esile kahes või enam sõnaliigis esinevate või ühest sõnaliigist teise liikuvate sõna(vormi)de määratlemine. Üldistades võib öelda, et problemaatilisimad sõnaliigimuutuse näited ei esinda mitte üht eesti keelele tüüpilisimat sõnaklassi vahetajat ehk sufikstuletust, vaid käände- või pöördevormi juurdumist teise sõnaliigi funktsioonis grammatisatsiooni- ja leksikalisatsiooniprotsesside tagajärjel (tüüpiliselt adverbid ja adpositsioonid) või konversioonipõhist transpositsiooni (adjektiivid). Morfoloogilise sõnatuletuse põhist sõnaliigimuutust võib küsitluse põhjal seega pidada selgepiiriliseks nähtuseks kui süntaktilist, või ka käände- ja pöördevormide liikumist teise klassi semantilise sõnaliigimuutuse tagajärjel.

Leksikograafid teravitavad digitööriista, mis kiirendaks sõnaliigi määramise protsessi. Nn sõnaliigimäärajal oleks varasemate korpuspõhiste tööriistadega (nagu SketchEngine) võrreldes parem automaatse morfoloogilise analüüsi programm, mis aitaks korpusest statistilist infot paremini välja sõeluda. See esitaks korpuseotsingute põhjal kvantitatiivse kokkuvõtte ambivalentse vormi morfosüntaktilise käitumise distributsiooni kohta. Leksikograafid soovivad näha, milline sõnaliigiline funktsioon prevaleerib, ning ka andmeid, millele see põhineb.

Küsitluse põhjal võib järeldada, et kuigi keeletehnoloogilised rakendused võimaldavad leksikograafil töötada efektiivsemalt ja langetada märksõna sõnaliigi määratlemisel paremini põhjendatud otsuseid, on digitaalsetel tööriistadel omad piirangud, millega sõnaraamatutegija peab arvestama. Leksikograafias valitseb suund üha suuremale automatiseerimisele, kuid tõenäoliselt jäävad just kriitilised alad ka tulevikus nõudma leksikograafi ekspertotsustust.

Praktilises töös on ehk olulisim leksikograafidevaheline infovahetus ning sõnaliike puudutava teabe kättesaadavus. Intervjueeritavate hinnangul oleks kasu hajusate sõnaliigipiiridega sõnade ja vormide koondamisest andmebaasi (koos võimalike eksperthinnangute ning eelnevate otsustega), selgete ja üksikasjalike leksikaliseerumis- ja grammatiseerumiskriteeriumite esitusest (näiteks moodulina) ning süntaktiliste ja semantiliste testide kogumist. Kokkuvõttes tõdeme, et küsitledes avaldas muljet leksikograafide erakordne vastutustunne ja empaatia nii potentsiaalse kasutaja kui ka kolleegide vajaduste suhtes.

Viidatud kirjandus

- Alexander, Marc 2015. Words and dictionaries. – John R. Taylor (Ed.), *The Oxford Handbook of the Word*. Kindle Edition. Oxford: Oxford University Press, 37–53. <https://doi.org/10.1093/oxfordhb/9780199641604.013.021>
- Anward, Jan 2000. A dynamic model of part-of-speech differentiation. – Petra M. Vogel, Bernard Comrie (Eds.), *Approaches to the Typology of Word Classes*. Berlin: Mouton de Gruyter, 3–45.
- Croft, William 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Culicover, Peter W. 1999. *Syntactic Nuts*. Oxford: Oxford University Press.
- EKG I = Ereht, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1995. *Eesti keele grammatika I. Morfoloogia. Sõnamoodustus* [‘The Grammar of the Estonian Language I: Morphology. Word-formation’]. Tallinn: ETA Eesti Keele Instituut.
- EKG II = Ereht, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1993. *Eesti keele grammatika II. Süntaks. Lisa: kiri* [‘The Grammar of the Estonian Language II: Syntax’]. Tallinn: ETA Keele ja kirjanduse instituut.
- Eesti keele naabersõnad 2019 [‘The Estonian Collocations Dictionary, ECD’]. Jelena Kallas, Kristina Koppel, Geda Paulsen, Maria Tuulik (Toim.). Eesti Keele Instituut. Sõnaveeb 2019. <https://sonaveeb.ee> (14.2.2019).
- EKS = Eesti keele sõnaraamat 2019 [‘The Dictionary of Estonian, DicEst’]. Margit Langemets, Mai Tiits, Udo Uibo, Tiia Valdre, Piret Voll (Toim.). Eesti Keele Instituut. Sõnaveeb 2019. <https://sonaveeb.ee> (14.2.2019).
- EKSS 2009 = Eesti keele seletav sõnaraamat I–VI [‘The Explanatory Dictionary of Estonian’]. “Eesti kirjakeele seletussõnaraamatu” 2., täiendatud ja parandatud trükk. Margit Langemets, Mai Tiits, Tiia Valdre, Leidi Veski, Ülle Viks, Piret Voll (Toim.). Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus, 2009.
- Ereht, Mati 2017. Sissejuhatus süntaksisse [‘Introduction to syntax’]. – Mati Ereht, Helle Metslang (Toim.), *Eesti keele süntaks. Eesti keele varamu III*. Tartu: Tartu Ülikooli Kirjastus, 537–564.
- Grünthal, Riho 2003. *Finnic Adpositions and Cases in Change*. Suomalais-Ugrilaisen Seuran toimituksia 244. Helsinki: Finno-Ugrian Society.

- Habicht, Külli 2001. Lexicalization and grammaticalization – opposite phenomena? About some lexicalized verb forms in Old Written Estonian. – *Estonian Typological Studies* V. Tartu Ülikooli eesti keele õppetooli toimetised 18. Tartu: Tartu Ülikool, 26–76.
- Habicht, Külli; Penjam, Pille; Prillop, Külli 2011. Sõnaliik kui rakenduslik probleem: sõnaliikide märgendamine vana kirjakeele korpuses [‘Parts of speech as a functional and linguistic problem: Annotation of parts of speech in the corpus of Old Written Estonian’]. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 7, 19–41. <https://doi.org/10.5128/ERYa7.02>
- Jürine, Anni; Habicht, Külli 2013. Kaassõnade tsükliline areng [‘The cycle of postpositions’]. – *Keel ja Kirjandus*, 12, 736–749.
- Kallas, Jelena; Koeva, Svetla; Langemets, Margit; Tiberius, Carole; Kosem, Iztok 2019. Lexicographic practices in Europe: Results of the ELEX Survey on user needs. – I. Kosem, T. Zingano Kuhn, M. Correia, J. P. Ferreria, M. Jansen, I. Pereira, J. Kallas, M. Jakubiček, S. Krek, C. Tiberius (Eds.), *Electronic Lexicography in the 21st century. Proceedings of the eLex 2019 conference. 1–3 October 2019, Sintra, Portugal*. Brno: Lexical Computing CZ, s.r.o., 519–536.
- Kasik, Reet 2012. Nulltuletus ja konversioon eesti keeles [‘Zero-derivation and conversion in Estonian’]. – *Keel ja Kirjandus*, 11, 793–806.
- Koppel, Kristina; Tavast, Arvi; Langemets, Margit; Kallas, Jelena 2019. Aggregating dictionaries into the language portal Sõnaveeb: Issues with and without a solution. – I. Kosem, T. Zingano Kuhn, M. Correia, J. P. Ferreria, M. Jansen, I. Pereira, J. Kallas, M. Jakubiček, S. Krek, C. Tiberius (Eds.), *Electronic Lexicography in the 21st century. Proceedings of the eLex 2019 conference. 1–3 October 2019, Sintra, Portugal*. Brno: Lexical Computing CZ, s.r.o., 434–452.
- Kenesei, Istvan 2018. On the definition of word classes. Preprint. <https://doi.org/10.13140/rg.2.2.34114.27842>
- Kilgariff, Adam; Rychlý, Pavel; Smrž, Pavel; Tugwell, David 2004. The Sketch Engine. – Geoffrey Williams, Sandra Vessier (Eds.), *Proceedings of the 11th EURALEX International Congress*. Lorient, France: Université de Bretagne Sud, 105–116.
- Langemets, Margit 2010. Nimisõna süstemaatiline poliüseemia eesti keeles ja selle esitus keelevaras [‘Systematic Polysemy of Nouns and its Lexicographic Treatment in Estonian Language Resources’]. Tallinn: Eesti Keele Sihtasutus.
- Langemets, Margit; Uibo, Udo; Tiits, Mai; Valdre, Tiia; Voll, Piret 2018. Eesti keel uues kuues. Eesti keele sõnaraamat 2018 [‘Estonian lexis revisited: The Dictionary of Estonian 2018’]. – *Keel ja Kirjandus*, 12, 942–958.
- Metslang, Helle 2017. Kvantorifraas [‘Quantifier phrases’]. – Mati Ereht, Helle Metslang (Toim.), *Eesti keele süntaks. Eesti keele varamu III*. Tartu: Tartu Ülikooli Kirjastus, 463–478.
- Oja, Vilja 1995. Võrdlusel baseeruvaist värviliitsõnadest [‘Comparison-based compound color words in Estonian’]. – *Keel ja Kirjandus*, 12, 809–817.
- Paulsen, Geda; Vainik, Ene; Tuulik, Maria; Lohk, Ahti 2019. The lexicographer’s voice: Word classes in the digital era. – I. Kosem, T. Zingano Kuhn, M. Correia, J. P. Ferreria, M. Jansen, I. Pereira, J. Kallas, M. Jakubiček, S. Krek, C. Tiberius (Eds.), *Electronic Lexicography in the 21st century. Proceedings of the eLex 2019 conference. 1–3 October 2019, Sintra, Portugal*. Brno: Lexical Computing CZ, s.r.o., 319–337.
- Paulsen, Geda 2018. Manner and adverb: Fuzzy categorial boundaries in collocations. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 14, 117–135. <https://doi.org/10.5128/ERYa14.07>
- Paulsen, Geda 2019. Sõnaliigipiiridest kollokatsioonide vaatenurgast: erikäändelised noomen-adverbid [‘Word class boundaries and collocations: The Estonian nominal adverbs in special cases’]. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 15, 121–137. <https://doi.org/10.5128/ERYa15.07>

- Prillop, Külli 2004. Kuidas märksõnastada vanu eestikeelseid tekste? [‘How to lemmatize old Estonian texts’] – Keel ja Kirjandus, 2, 90–99.
- Rätsep, Huno 1978. Eesti keele lihtlausete tüübid [‘Types of Simple Sentences in Estonian’]. Valgus: Tallinn.
- Sahkai, Heete 2008. Konstruksioonipõhine keelemudel ja sõnaraamatumudel [‘Some lexicographic implications of a construction-based model of language’]. – Eesti Rakenduslingvistika Ühingu aastaraamat, 4, 177–186. <https://doi.org/10.5128/ERYa4.11>
- Sepper, Maria-Maren 2007. Indirectal in Literary Estonian. – Trames, 11 (3), 299–323.
- Smith, Mark C. 2015. Word categories. – John R. Taylor (Ed.), The Oxford Handbook of the Word. Kindle Edition. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199641604.013.019>
- Sõnaveeb 2019 [‘Dictionary portal WordWeb’]. Eesti Keele Instituudi sõnastikuportaal. Indrek Hein, Jelena Kallas, Kristina Koppel, Margit Langemets, Kaur Männiko, Tõnis Nurk, Ülle Viks (arendajad ja toimetajad); Martin Laubre, Raigo Ukkivi, Arvi Tavast, Sander Lastovets, Sander Rautam (arendajad). <https://sonaveeb.ee> (20.9.2019).
- Tavast, Arvi; Langemets, Margit; Kallas, Jelena; Koppel, Kristina 2018. Unified data modeling for presenting lexical data: The case of EKILEX. – J. Čibej, V. Gorjanc, I. Kosem, S. Krek (Eds.), Proceedings of the 18th EURALEX international congress. Ljubljana University Press, Faculty of Arts, 749–761.
- Tiits, Mai 1982. Seisundiadverbidest [‘About adverbs of state’]. – Keel ja Kirjandus, 1, 17–21.
- Vare, Silvi 2004. Eesti keele verbimoodustus: desubstantiivne konversioon [‘Verb formation in Estonian: Denominal conversion’]. – Emakeele Seltsi aastaraamat, 50, 39–65.
- Vare, Silvi 2006. Adjektiivide substantivatsioonist ühe tähendusrühma näitel [‘On the substantivation of adjectives in Estonian’]. – Ellen Niit. Keele ehe. Tartu Ülikooli eesti keele õppetooli toimetised 30. Tartu: Tartu Ülikool, 205–222.
- Viks, Ülle 1992. Väike vormisõnastik [‘Concise Morphological Dictionary of Estonian’]. I–II. Tallinn: Eesti Teaduste Akadeemia.
- ÕS = Eesti õigekeelsussõnaraamat 2018 [‘Dictionary of Standard Estonian’]. Maire Raadik, Tiiu Ereht, Tiina Leemets, Sirje Mäearu (Toim.). Tallinn: Eesti Keele Sihtasutus.

Võrguviited

- Eesti keele süntaksianalüsaator. <http://kodu.ut.ee/~kaili/parser/> (20.9.2019).
- EKI reegli põhise morfoloogia demo. <http://artur.eki.ee/morf/> (20.9.2019).
- ESTMORF. <http://eki.ee/keeletehnoloogia/projektid/estmorf/estmorf.html#2> (20.9.2019).
- Sketch Engine. <https://www.sketchengine.eu/> (20.9.2019).

Lisa 1. Poolstruktureeritud intervjuu küsimustik

Sõnaliigid leksikograafia vaatenurgast

Küsitlus

Küsimustik on mõeldud lahenduste leidmiseks eesti keele sõnaliikide probleemidele, mis valmistavad raskusi sõnaraamatute koostamisel ning ka rakenduslike lahenduste väljatöötamisel. Kasutame Teie andmeid uuringus, mida teostame projekti PSG277 “Eesti keele sõnaliikide kategooriad korpusanalüüsi valguses” raames. Vastuseid ei avaldata ega avalikustata ilma vastaja loata ja soovi korral tagatakse Teile anonüümsus.

I Vastaja andmed

1. Mitu aastat olete töötanud leksikograafina?

2. Millistes asutustes töötate ja/või olete töötanud leksikograafina?

3. Milliste sõnaraamatutega olete töötanud (tüüp, nimetused)?

4. Kas kasutate sõnaraamatute koostamisel tekstikorpust? Milliseid?

5. Kas Teil on kogemus elektroonilise sõnastikusüsteemiga töötamisega ja kui, siis millise süsteemiga?

II Üldised küsimused sõnaliikide kohta sõnaraamatutes

(arvamused-kogemused)

1. **Kellele ja milleks on Teie arvates sõnaraamatutes sõnaliigimärgendid?** Kas muutevormide või muutetüübi andmine ilma sõnaliiki eraldi märkimata on sõnaliigiinformatsiooni mõttes piisav?
2. **Kui vajalik on Teie arvates sõnaliigiinfo sõnaraamatus, millega praegu töötate?** Palun hinnake sõnaliigiinfo vajalikkust järgmisel skaalal: VÄGA VAJALIK – PIGEM VAJALIK – KESKMINE – PIGEM EBAVAJALIK – TÄIESTI EBAVAJALIK
3. **Kas Teil on olnud vajadus määratleda oma sõnaraamatutöös sõnaliike? Millised on Teie kogemused sõnaliikide esitusega erinevates sõnaraamatutes?** Mille poolest erinevad eri tüüpi sõnaraamatud selles suhtes? Palun lahti seletada.

III Leksikograafide probleemid sõnaliikide määratlemisel

4. **Milliseid probleeme ja küsitavusi Teile meenub oma leksikograafitööst seoses sõnaliikidega?** Kuidas näete ise sõnaliikide eristamise ja mitteeristamisega seotud küsimusi? Kas on mingid sõnaliigid, mille vahel on eriti keeruline otsust teha? Näiteid? Materjali?
5. **Kuidas tavaliselt toimite, kui märksõna sõnaliik pole üheselt selge? Mida ette võtate?** Mille põhjal otsustate sõnaliigi üle? Milliseid allikaid kasutate? Süntaktilised, semantilised, morfoloogilised tunnused? Kas kasutate teste? Kas sõnaraamatul, millega töötate, on koostamisjuhend sõnaliikide kohta? Palun esitage konkreetseid näiteid. Kas olete saanud abi korpusetest, grammatikatest, keeletehnoloogilistest rakendustest?
6. **Kui suur on Teie hinnangul sõnaliikide määratlemise probleem sõnaraamatute koostamisel võrreldes teiste protsessidega?** Palun hinnake järgmisel skaalal (ja kuidas iseloomustate oma valikut?): VÄGA SUUR – PIGEM SUUR – KESKMINE – PIGEM EBAOLULINE – TÄIESTI EBAOLULINE

IV Võimalikud lahendused

7. **Kuidas suhtute sellesse, kui ühel märksõnal on mitu sõnaliigimärgendit? Kuidas suhtute sellesse, kui märksõnal (ja alamärksõnadel) ei ole sõnaliigimärgendit?**
8. **Millised võiks olla lahendused, abistavad vahendid sõnaliigi määratlemises?**
9. **Kuidas suhtute sellesse, kui praegune sõnaliikide jaotuse süsteem eesti keeles muutuks?** Kui sõnaliike tuleb juurde (jaotus läheb peenemaks)? Kui sõnaliikide süsteemi lihtsustatakse (osa eristus läheb kaduma)?
10. **Kui sõnaliikide eristamiseks oleks olemas infotehnoloogiline abivahend, siis kuidas abivahend peaks seda infot esitama?** Mis laadi abi sõnaliikide osas Te sellelt ootaksite? (Statistilised, kvalitatiivsed kokkuvõtted vms.)

ON WORD CLASSES IN CONTEMPORARY LEXICOGRAPHY: THE LEXICOGRAPHERS' VIEW

Geda Paulsen, Ene Vainik, Maria Tuulik

Institute of the Estonian Language

This study examines the role of word classes in the contemporary lexicographic work. The analysis is based on a metalexigraphic semi-structured interview carried out among 23 Estonian lexicographers. We focus on three main research questions: i) How necessary are word classes in lexicographers' work? ii) What kinds of problems they do experience related to the word classes? iii) How do lexicographers solve these problems and how could digital tools facilitate the word class judgement processes?

The results indicate that in lexicography, word classes provide an (implicit or explicit) frame for organisation of data, even if word-class tagging does not directly belong among the tasks of the lexicographer. The word class determination process in unclear cases is a complicated one. In addition to the usage of a word and the expected needs of the user, the lexicographer must consider the structural requirements of the particular dictionary (s)he is working with. Along with the development of digital resources, the demands of the data model behind the electronic database have to be considered; not to mention the theoretical linguistic insights. Linguistically, the most problematic issues appear in the grammaticalisation and lexicalisation processes related to inflectional suffixes (mainly adverbs and adpositions) or conversional transposition (adjectives). The typical word-class altering devices in Estonian such as suffixal derivation (see, e.g. Vare 2006: 199) do not emerge as particularly problematic.

The interviewed lexicographers would welcome a digital application that would facilitate and quickened the word class determination process. The application would have an improved morphological tagger compared to the present tools and it would be more efficient in extracting statistical information from corpora. Lexicographers wish to acquire information about the prevailing word-class function together with the data this assessment is based on. Yet, the lexicographer makes the final decision.

Keywords: word classes, lexicography, semi-structured interview, Estonian

Geda Paulsen (Eesti Keele Instituut) tegeleb leksikaalse semantikaga ning sõnaliigipiiride uurimisega leksikograafia vaatenurgast.
Roosikrantsi 6, 10119 Tallinn, Estonia
geda.paulsen@eki.ee

Ene Vainik (Eesti Keele Instituut) tunneb huvi semantika ja keele psühholoogiaga piirnevate aspektide vastu (nt tundesõnad, emotsioonide väljendamine kirjalikus ja suulises tekstis, piltlik keelekasutus, sõna-assotsiatsioonid). Hetkel tegeleb sõnaliigipiiride uurimisega leksikograafia vaatenurgast.
Roosikrantsi 6, 10119 Tallinn, Estonia
ene.vainik@eki.ee

Maria Tuulik (Eesti Keele Instituut, Tartu Ülikool) uurib adjektiivide semantikat ja sõnaliigipiire leksikograafia vaatenurgast.
Roosikrantsi 6, 10119 Tallinn, Estonia
maria.tuulik@eki.ee