

SÜNTAKTILINE INFO SÕNASTIKUS: PROBLEEME JA VÄLJAVAATEID

MARGIT LANGEMETS, MERIKE MÄGEDI, ÜLLE VIKS

Ülevaade. Artiklis käsitletakse süntaktilise info esitust elektroonilistes sõnastikes ja suurtes leksikaalsetes andmebaasides. Esiteks tehakse ülevaade seda tüüpi infost muude keelte elektroonilistes allikates, seejärel eesti üldsõnastikes. Neile teadmiste toetusel pakutakse välja süntaktilise info valikukriteeriumid ja esitusviisid Eesti Keele Instituudi uute sõnastikuprojektide, ennekõike eesti–X-keele sõnastiku jaoks.¹

Võtmesõnad: leksikograafia, arvutisõnastikud, sõnaliigid, sõltuvussuhted, semantilised rollid, XML

Sissejuhatus

Keeles on palju tuhandeid sõnu. Sõna õige kasutamine eeldab, et inimene valdab selle hääldust, õigekirja, muutmist, süntaktilist käitumist, koosenemist teiste üksustega, põhitähendusi, tähendusvarjundeid, semantilisi suhteid muude sõnadega jms. Üks allikaid, kus sõnad ja kogu nende kohta käiv tarkus kokku saavad, on sõnaraamat. Siin aga kerkib põline semantikat ja grammatikat puudutav praktiline küsimus: milline info kuulub sõnastikku ja milline grammatikasse, kui on ju teada, et teataval määral on mõlemas allikas mõlemat laadi infot?

Morfoloogiline info on vormirikaste keelte sõnastikes üsna tavaline, aga süntaks on traditsiooniliselt rohkem kuulunud grammatikasse. Kuid mida aeg edasi, seda suuremal määral ilmuvad süntaktilised andmed ka sõnastikesse. Oma osa on siin leksikaalse semantika arengul, mille uuemad suunad lähenevad sõnatähendusele ainult ühenduses süntaksiga,

¹ Artikli valmimist on toetanud riikliku sihtprogrammi “Eesti keel ja rahvuskultuur (2004–2008)” allprogramm “Eesti–X-keele sõnastik” ja ETF grant nr 5969.

kuid vähetähtis pole ka keeletehnoloogia roll. Keeletehnoloogilised rakendused vajavad lisaks tarkvarale suurt hulka formaalseid keeleandmeid (sh süntaktilisi), mida ammutatakse kahel põhimõtteliselt vastandlikul viisil: harvemini suuri tekstikorpuse analüüsides ja sealt grammatilisi teadmisi tuletades (vt Kaalep jt 2000), sagedamini elektroonilisest sõnastikust otse üle võttes seda, mis seal olemas on. Grammatikaraamatute info pole arvutile ilma inimese vahenduseta kättesaadav. Teoreetilist põhjendust hoida sõnaraamatu ja grammatika infot lahus pole leitud ja seda toetab teadmine, et nendevahelised piirid puuduvad ka inimese enda meeles (Wakely, Béjoint 2002: 67). Sõnastikus või leksikaalses andmebaasis on väga hästi võimalik näidata, kuidas leksikaalsed üksused moodustavad eri konstruktsioonimalle – ja see on ülimalt väärtuslik info nii kasutaja kui ka keelekirjelduse kogupildi jaoks. Robert Ilsonit tsiteerides:

Kui oleks olemas teos, mis sisaldaks *kõiki* keelefakte, siis meenutaks see pigem sõnaraamatut kui grammatikat. (Ilson 1992: 277)

Kui morfoloogilise info esitamisel on eesti leksikograafias pikk traditsioon, siis eksplitsiitne süntaktiline info meie sõnaraamatutes sama hästi kui puudub. Vajadust selle järele ja otsust sellega tegelema hakata on mõjutanud ka aina detailsema grammatilise info paigutamine leksikaalsetesse andmebaasidesse mujal maailmas. Käesolevas artiklis püüame põhjendada ja kirjeldada, mis laadi süntaktilist infot peaks Eesti Keele Instituudi leksikaal-grammatiline andmebaas sisaldama ja kuidas seda formaalselt esitada. Artiklis vaatleme süntaktilise info esitust nii muude keelte sõnastikes ja semantilistes andmebaasides kui ka senistes eesti üldsõnastikes. Seejärel kirjeldame eesti–X-keele sõnastiku jaoks välja töötatud põhimõtteid ja võimalikke esitusviise. Aga kõigepealt mõni sõna projektist endast.

1. Eesti–X-keele sõnastik

Projekt “Eesti–X-keele sõnastik” käivitus Eesti Keele Instituudis 2002. aastal. Kakskeelse sõnaraamatu üks probleeme on, kuidas vältida lähteele semantilise ja grammatilise struktuuri moonutamist sihtkeele vastete poolt. Uuemas leksikograafiateoorias soovitatakse enne konkreetse kakskeelse sõnaraamatu koostamist luua kaks ükskeelset leksikaalset

andmebaasi (vt lähemalt Atkins 1996). Tänapäeva arvutiajastu tehnilisi võimalusi ära kasutades saab need andmebaasid metateksti, hüperlinkide jm abil ühendada dünaamiliseks kakskeelseks andmekoguks, millel oleks mitmeid perspektiivseid kasutusvõimalusi. Näiteks saab niisuguse andmebaasi põhjal luua keelte sarnasuste või kontrastide sõnastiku, samuti arvestada konkreetse sõnaraamatu koostamisel seda, kas kasutaja emakeel on lähtekeel või sihtkeel. Seda teooriat järgib ka projekt “Eesti–X-keele sõnastik”, mille esimeses etapis peaks sündima (universaalne) ükskeelne eesti keele leksikaal-grammatiline andmebaas, mis pikemas perspektiivis võiks saada mis tahes eesti lähtekeelega kakskeelse sõnaraamatu eesti poole aluseks. Iga uue tõlkesõnaraamatu koostamisel on võimalik sellises sõnastikubaasis olevat infot vastavalt vajadusele modifitseerida: võtta osa tähendusi kokku (või lisada uusi), jätta osa grammatilist infot (või kõik) üle võtmata, jätta kasutamata (või lisada) tekstinäiteid jne. See tähendab seda, et loodav ükskeelne andmebaas peab märksõnu kirjeldama väga põhjalikult, et sealt saaks vastavalt koostatava sõnastiku iseloomule valikuid ja muudatusi teha.

Eesti–X-keele sõnastiku (EXS) maht on esialgu *ca* 40 000 märksõna, kuid edaspidi luuakse ka suurema ja väiksema mahuga versioonid. Andmebaasi koostatakse andmestruktuuride kaupa (mitte alfabeetiliselt), mis annab võimaluse keskenduda eraldi igale probleemiringile. Eraldi töö on sõnavalik, tähenduste liigendamine ja defineerimine, stiilimärgendus jne. Samuti ka grammatilise info esitus, mida järgnevalt käsitleme.

2. Grammatiline info muude keelte leksikaalses keelevaras

Tuntumaid ja suuremaid kakskeelseid leksikaalseid keelevarasid kirjeldavas põhjalikus ülevaates (Calzolari jt 2001) on sõnakogud rühmitatud kolme suurde põhitüüpi: a) elektroonilises versioonis sõnaraamatud, b) arvutileksikonid ja c) masintõlkesüsteemide keelevara. (Tüüpide “keeletehnoloogilisuus” kasvab arusaadavalt eest tahupoole.) Kuivõrd meie huvi piirdub sõnastikega, siis oleme masintõlkesüsteemide temaatika täiesti kõrvale jätnud ja lähemalt uurime vaid kaht esimest leksikonitüüpi. Olgu need põgusalt üle korratud. Elektroonilised sõnastikud kujutavad endast pabersõnaraamatute arvutisse sisestatud versioone, sisaldades rikkalikku, aga traditsioonilist leksikograafilist infot. Arvutileksikonid on

ulatuslikud, algusest peale elektrooniliselt loodud, üldkasutuseks mõeldud, aga tihti mingile aspektile keskendatud keelekogud (nt FrameNet, WordNet, Parole-Simple), kus leksikaal-grammatiline info on võimalikult eksplitsiitselt esile toodud.

Grammatikasüsteemi kaks põhikomponenti, morfoloogia ja süntaks, on eri keeltes erineva osakaaluga, mis määrab ka grammatilise info liikide vahekorra sõnastikus. Näiteks soome-ugri keeltes, kus muutelõpud täidavad olulisi grammatilisi funktsioone, on ka sõnaraamatutes esikohal morfoloogiline info. Seevastu keeltes, kus morfoloogia ei ole nii ulatuslik, on sõnajärjel täita rohkem funktsioone ning ka sõnastikud sisaldavad enam süntaksiteavet. Siinkohal jälgime lühidalt, kuidas ja millist süntaktilist ning sellega seotud infot pakutakse mujal sõnastikumaailmas. Andmed pärinevad N. Calzolari jt ülevaatest (2001).² Keeled, mis seal esile tulevad, on inglise (alati üks paarilistest) ja teise paarilisena prantsuse, itaalia, saksa ja hispaania – seega rohkem süntaksi- kui morfoloogiakeeled. Vaatleme lähemalt kolme tüüpi süntaksiga seostuvaid andmeid: morfosüntaktilist, süntaktilist ja semantilist infot.

2.1. Morfosüntaktiline info

Käsitletavates elektroonilistes sõnastikes ja arvutileksikonides on alati (või enam-vähem alati) esitatud a) sõnaliik, b) arv, c) nimisõnadel asjasõnade ja ainesõnade vastandus ning d) adverbidel ja adjektiividel komparatsioon. Sõltuvalt keelest on esitatud e) sugu. Üsna üldiselt puudub f) vormimoodustuse info, ainult (ükskeelsetes) arvutileksikonides on esitatud g) tuletusinfo.

Sõnaliiki kui leksikaal-grammatilist kategooriat kvalifitseeritakse kui minimaalset süntaktilist infot – inglise keeles on oluline märkida seda juba sel põhjusel, et seal esineb palju noomeni ja verbi homonüümiat. Lisaks tavapäraselt esitatavatele sõnaliikidele (nimisõna, omadussõna,

² N. Calzolari jt (2001) on lähemalt uurinud järgmisi kakskeelseid sõnaraamatuid: itaalia-inglise / inglise-itaalia (Collins 1995), prantsuse-inglise (Collins Gem 1997, 40 000 märksõna), saksa-inglise (Collins Gem 1997, 40 000 märksõna), inglise-prantsuse (Oxford-Hachett 1997), hispaania-inglise (Oxford 1998). Tuntumatest arvutileksikonidest on vaatluse all FrameNet, WordNet ja Parole-Simple. Omalt poolt oleme lisanud näiteid ükskeelsest Longmani sõnaraamatust (LDOCE), mis on tuntud kui üks enam tarvitatud sõnaraamatuid arvutirakendustes.

asesõna, verb, adverb, prepositsioon, sidesõna, hüüdsõna) on tihti eristatud ka muid, vähem tavapäraseid sõnaliike, nagu määratlev asesõna, modaalverb, hulgasõna (kogunimi), lauselaiend jm. Inglise keeles märgitakse verbi juures enamasti ka selle transitiivsust (Vt) või intransitiivsust (Vi). Eriti hoolikalt on sõnaliigiinfo kirjas ükskeelsetes õppesõnaraamatutes, mis õigupoolest tähendab seda, et niisama detailselt on see kirjas ka emakeele seletussõnaraamatus (inglise praktikas on õppesõnaraamat tihti seesama mis seletussõnaraamat).³

Arvu tähistatakse, kui märksõna vorm viitab ühele arvukategooriale, aga märksõna ise funktsioneerib teise kategooria esindajana, samuti siis, kui tegemist on ainsus- või mitmussõnaga (mis lauses ühildub ainult sama kategooriat esindava verbiga ja teises kategoorias kunagi ei esine). Eri sõnaraamatutes kirjeldus varieerub: kasutatud on lühiseletusi (“funktsioneerib ainsuses”, “funktsioneerib nii ainsuses kui ka mitmuses”), samuti ainsusele-mitmusele viitavaid lühendeid või lühendamata sõnu (“pl”, “singular”).

Asja- ja ainesõnade eristust⁴ märgitakse eelkõige morfosüntaktilistel (mitte semantilistel) kaalutlustel, eri uurijad on küll teinud ka semantilisi järeldusi, aga sõnaraamatutesse need üldjuhul veel jõudnud ei ole.⁵

Komparatsiooni märgitakse tavaliselt erijuhtudel, näiteks kui tüvi muutub kas täielikult (*good–better–best* ‘hea–parem–parim’) või osaliselt (*fat–fatter–fattest* ‘rasvane–rasvasem–rasvaseim’).

Vormimoodustuse infost, mis läbivana tavaliselt puudub, piirdatakse vaid ebareeglipäraste muutevormide esitusega⁶ (ega inglise keel rohkem ei vajagi).

Tuletusinfot on eksplitsiitselt esitatud uuemates (ükskeelsetes) arvuti-leksikonides (nagu ükskeelsetes sõnaraamatutes üldiselt), aga näidatud on ainult esimese astme seosed (nt *snobbish* adj – *snobbishly* adv – *snobbishness* n), mitte kogu sõnapere (eraldi märksõnad on *snob*, *snobbery* ja *snobbish*). Kakskeelsetes sõnaraamatutes jääb moodustussuhe tavaliselt esile toomata.

³ Sõnaraamatutüüpide erinevuste ja sarnasuste kohta vt Langemets 2003.

⁴ Inglise keeles *count/mass noun*, ka *countable/uncountable noun* ‘loendatav/loendamatu nimisõna’.

⁵ Lähemalt vt Langemets 2004.

⁶ Mõnes sõnaraamatus näidatakse ebareeglipäraste verbide pööramist muuttüübi vahendusel, nii ka näiteks prantsuse-eeesti sõnaraamatus (Kann 1995).

2.2. Süntakiline info

Küllalt tihti on muude keelte leksikaalses keelevaras sõna või selle ühe tähenduse juures esitatud a) valents ehk sõna süntaktiliselt seotud laiendid, b) perifrastilised (ehk ümberütlevad) konstruktsioonid ja püsiühendid, samuti c) muud kollokatsioonid (ehk sõnade koosinemised) ja d) süntaktilised variandid. Harva on tähistatud e) laiendite kohustuslikkus.

Valentsi osas piirduaksegi sageli üksnes verbi laienditega – esitatakse verbi sõltuvusseosed ja vastavad rektsioonid –, aga esineb ka muude sõnade, näiteks adjektiivide ja nimisõnade laiendite kirjeldusi. Üldjuhul esitatakse seda tüüpi info malli abil, aga esineb ka sõnalist kirjeldust, mõnikord on laiendite arv ka eraldi välja toodud. Andmebaasis on alati rohkem ja detailsemat infot kui trükitud sõnaraamatus, kus tavaliselt loetakse rektsioonimallid lihtsalt järjest üles, nt Longmani sõnaraamatus (LDOCE)

hope *n* .. **5** [C, U]⁷ .. [+ **of**] *there was no hope of escape* | **hope that** *There is some hope that we'll find ..* (LDOCE)

Perifrastiliste konstruktsioonide ja püsiühendite puhul püütakse sageli märkida ka fraasi sõnajärge (üksuste positsiooni, nt LDOCE-s “ainult nimisõna ees”, “ainult nimisõna järel”, “mitte nimisõna ees” – inglise keeles puudutab see eriti adjektiive ja adverbe), konstruktsioonid järjestatakse sageli sageduse järgi, nt

increasingly *adv* **1** *more and more ..* [+**adj/adv**] *The classes at college have become increasingly full over the past five years* (LDOCE)

decide **1** [I, T]⁸ .. **decide to do sth** *Tina's decided to go ..* **4** **decide in favour of/decide against** .. (LDOCE)

Lisaks püsiühenditele tähistatakse eraldi üksusena ka muid leksikaalseid kollokatsioone ja nende koostisosi, näiteks tugiverbe ja kollokaate. Tugiverbil justkui oma tähendus puudub, tema panus kogu konstruktsiooni tähendusse on üsna väike – ühendi semantiline sisu lähtub eelkõige nimisõnafrasist: nii on fraasis *pay smb. a compliment* (‘kellelegi komplimenti tegema’) tugiverbi staatuses *pay*. Eesti keele levinuimaks tugiverbiks on *tegema* (konstruktsioonid *tööd/nalja tegema*), tuntud on ka *aja-*

⁷ C on asjasõna (*countable noun*), U on ainesõna (*uncountable noun*).

⁸ I on intransitiivne verb, T on transitiivne verb.

ma- (*asja ajama*) ja *lööma*-ühendid (*tantsu lööma*).⁹ Kollokaat võib olla verbi tüüpiline subjekt/objekt, adjektiivi poolt modifitseeritud nimisõna jm. Alati pole sugugi üheselt selge, kumb kumma määrab, kas verb kollokaadi või vastupidi.

Uuemad arvutileksikonid kirjeldavad semantilis-süntaktilise infona ka argumendistruktuuri variante (eesti keeles nt *Jüri vaidles Mariga, kas .. ~ Jüri ja Mari vaidlesid, kas ..*).

Infot selle kohta, kas laiend on kohustuslik või vaba, esitatakse harilikult mingi koodi (sõnaraamatus tihti sulgude) abil. Euroopa sõnastike standardid (EAGLES¹⁰, samuti selle edasiarendus ISLE¹¹) soovivad kodeerida nii laiendite arvu (kas siis otse numbri abil või kaudselt, laiendeid loetledes) kui ka laiendite obligatoorsust-fakultatiivsust (näiteks kahe eri (liht)kirjena või ühe (liit)kirjena).

Kogu verbilaiendeid puudutav info on tihedalt seotud semantiliste rollide infoga.

2.3. Semantiline info

Kui tavainimese jaoks on semantiline info sama mis seletus sõnaraamatus, siis tegelikus keekekasutuses on asi palju kordi keerulisem: mis tahes lauses põimuvad alati moodustajate süntaktiline ja semantiline funktsioon. Järgnevalt vaatamegi lühidalt vaid niisugust semantilist infot, millel on otseseos süntaksiga.

Grammatikaga seotud semantilist infot on rohkem kodeeritud uuemates arvutileksikonides, trükitud sõnaraamatutes on seda eksplitsiitselt esile toodud suhteliselt vähe (see iseenesest ei tähenda, et see seal puudub). Esitatud võivad olla: a) semantilised tüübid, b) semantilised rollid.

Semantilise tüübi abil klassifitseeritakse eelkõige sõnatähendusi (*resp.* asjade omadusi, mida need sõnad tähistavad). Näiteks on LDOCE andmebaasis nimisõnade jaoks 32 semantilist koodi, kodeeritud on: konkreetne asi, abstraktne asi, elutu konkreetne asi, inimene, isasloom või meessoost inimene, kollektiivsõna looma või inimese jaoks jne. Niisugust infot saab osaliselt kasutada semantilise ja süntaktilise analüüsi abivahen-

⁹ Lähemalt on nendest kirjutanud Muischnek 2004.

¹⁰ Vt <http://www.ilc.cnr.it/EAGLES96/home.html> (1.11.2004).

¹¹ Vt http://www.ilc.cnr.it/EAGLES96/isle/ISLE_Home_Page.htm (1.11.2004).

dina: lauset analüüsitakse fraasi põhisõna semantilisele tüübile toetudes (nt lause **kivi armastab Marti* ei ole vastuvõetav, sest *kivi* kui konkreetne elutu asi ei saa kedagi armastada). Tuntuim arvutileksikon, mis hierarhiline infot esitab, on WordNet (mille eeskujul koostatakse ka eesti keele tesaurust), ja mitmed automaatsed semantilise analüüsi süsteemid toetuvad just WordNetist üle võetud semantilistele tüüpidele.

Lause semantiliseks keskmeks peetakse üldjuhul öeldist, mille sõnatähendusest sõltuvad teised lause situatsioonis osalejad ehk aktandid. Süvasuhet öeldise (verbi) ja aktandi vahel on nimetatatud semantiliseks rolliks¹². Semantiline roll tähistab tegelikku osa, mida mingis situatsioonis osaleja etendab. Rollid on olemuslikud kõigile sündmustele, olulisemaid semantilisi rolle on agent, teema, siht, instrument, tulemus, kaasosaline jne. Semantilised rollid ja seotud laiendite süntaktiline funktsioon (subjekt, objekt jne) on omavahel tihedalt seotud, mis võimaldab teha teataval astmel lingvistilisi üldistusi. Nii on agent näiteks paljudes keeltes harilikult subjekt, teema harilikult objekt jne. Kuid üks ja sama roll võib olla süntaktiliselt väljendatud ka mitmel eri moel, st päris üksühest vastavust süntaktilise funktsiooni ja semantilise rolli vahel ei ole. Näiteks subjekt võib eri lausetes täita agendi, instrumendi või kogeja rolli: *mees haaras võtme* (agent); *võti keerab halvasti* / *võti ei keeranud lukku lahti* (?instrument); *üks avanes* (patsient). Sillaks leksikaalse semantika ja süntaksi vahel on argumendistruktuur, st argumendifreimid koos argumendi tüüpi tuvastava semantilise infoga.

Tuntuim arvutileksikon, mis sedalaadi infot esitab, on 1997. aastast peale koostatav FrameNet, ja sellel on tuntav mõju paljude teiste semantiliste andmebaaside ülesehitusele. FrameNeti andmebaasis on semantilised rollid (ehk freimielemendid) ja nende süntaktilised realisatsioonid tähistatud sõna (*resp.* leksikaalse üksuse) ühe tähenduse kaupa. (Tähenduse seletus ise on üldiselt üle võetud valmis sõnastikust.) Sõnad on korraldatud freimidesse, näiteks kuuluvad freimi *asetamine* (*placing*) verbid *bottle* 'pudelisse panema', *park* 'parkima', *plant* 'istutama', *archive* 'arhiveerima' jm, freim ise on defineeritud järgnevalt:

Asetamine (*placing*)

Üldiselt enam-vähem paigal olev Agent asetab Teema teatavasse kohta,

¹² Kasutatakse ka terminit *temaatiline roll*, uuemas generatiivse grammatika teoorias on kasutusel teeta-roll (*theta role*).

mis on väljendatud kui Siht. Agent/Põhjus kontrollib selles freimis Teemat ajahetkel, mil see jõuab Sihti.

Generally without overall motion, an Agent places a Theme at a location, the Goal, which is profiled. In this frame, the Theme is under the control of the Agent/Cause at the time of its arrival at the Goal. (FrameNet¹³)

Iga leksikaalse üksuse kohta esitatakse tema valentsimall koos sagedusandmetega FrameNeti aluseks olevas tekstikorpuses, nt verbi *bottle* sagedam mall on Agent-Siht (22 lauset), vähem esineb malli Agent-Siht-Teema (3 lauset), aga verb *park* esineb alati mallis Agent-Siht-Teema (32 näitelause). Rollide täpset hulka ei ole FrameNetis lõplikult piiritletud, vajadusel moodustatakse neid lisaks. Ootuspärane on, et verbide ja deverbhaalsete nimisõnadega kaasneb oluliselt rohkem infot kui lihtsate osutavate nimisõnadega.

3. Grammatiline info eesti sõnastikes

Et eesti keel on väga morfoloogiarikas, siis on grammatiline info sõnaraamatutes läbi aegade olnud valdavalt seotud vormimoodustusega. Tüübinumbrid märksõna järel ja tüüpsõna põhivormide näited eraldi osas on kasutusel juba Wiedemannist (1869) alates.¹⁴ Õigekeelsussõnaraamatute traditsiooni kuulub ka morfofonoloogiline info: kolmas välde ja ebareeglipärane rõhk, millel on oluline roll sõnade käänamisel ja pööramisel. Teave sõnamoodustuse kohta piirdub peamiselt liitsõnaosade piiri märkimisega, vahel näidatakse ka tulenemisseoseid (nt *tusaselt* < *tusane*).

Järgnevalt vaatame süntaksi ümber koonduvat infot, mida leidub hoopis vähem ja ebajärjekindlamalt. Mis laadi grammatilist infot esitatakse tänapäeva eesti sõnastikes, sellest annab kokkuvõtlikult aimu tabel 1, mis jätab täpsustamata morfoloogilise poole (koos tuletusega). Vaatluse all on “Eesti keele sõnaraamat ÕS 1999” (ÕS 1999), “Soome-eesti suursõnaraamat” (SM-EE), “Eesti-vene sõnaraamat” (EE-VN), “Eesti kirja-keele seletussõnaraamat” (EKSS), “Tänapäeva eesti-inglise sõnaraamat” (AULE) ja “Eesti keele verbirektsioone” (POOL). (Kakskeelsetes sõnastikes käsitletakse eesti poole infot.)

¹³ Vt <http://www.icsi.berkeley.edu/~framenet/> (30.10.2004).

¹⁴ Morfoloogilise info esitamisest eesti sõnastikes vt lähemalt Viks 2001.

Tabel 1. Grammatiline info eesti sõnastikes

	ÕS 1999	SM-EE	EE-VN	EKSS	AULE	POOL
morfoloogia	+	+	+	+		(+)
sõnaliik	(+)	+	+	+	(+)	
verbi sõltuvusnõuded	(+)	(+)	((+))	(+)	+	+
muude sõnaliikide sõltuvusnõuded	(+)	((+))	((+))	(+)	+	
fraasi sõnajärg			(+)	(+)	+	(+)
fakultatiivsed laiendid				(+)		(+)

Ühekordsed sulud märgivad ebajärjekindlat tähistamist, kahekordsed sulud näitavad, et tähistatud on vaid paar juhuslikku üksikjuhtu – sama ka edaspidi.

Nagu tabelist näha, esitavad morfoloogiat kõik sõnaraamatud peale AULE, mis keskendub eelkõige aktiivsele inglise keele kasutamisele. Ka sõnaliik on üsna hästi esindatud, otsest süntaksiteavet pakutakse napimalt. Sõltuvusnõuded on eesti sõnastikes tähistatud kas juhuslikult (verbide juures) või haruharva (muude sõnaliikide juures). Üldsõnastike seas tõuseb esile AULE, kus sõltuvusnõuded on üldiselt tähistatud. Loomulikult teeb seda ka spetsiaalne verbide rektsioonisõnastik – POOL. Muude sõnaliikide sõltuvust märgib üsna järjekindlalt EKSS. Fraasi sõnajärg ja fakultatiivsed laiendid on eri allikates tähistatud üksnes juhuslikult, ainsana püüab fraasi sõnajärge süsteemselt jälgida AULE.

Lähemalt tuleb juttu sõnaliikidest ja sõltuvusnõuetest.

3.1. Morfosüntaktiline info: sõnaliigid

Sõnastikud märgivad sõnaliiki erinevalt. On mitmeid sõnaraamatuid ja andmebaase, kus antakse sõnaliik järjekindlalt iga märksõna kohta, nt EKSS, EE-VN, Ülle Viksi “Väike vormisõnastik” (VVS), Eesti Keele Instituudi liitsõnade andmebaas, SM-EE eesti vastete morfoloogiasõnastik (SM-EE-vasted), Paul Saagpaku “Inglise-eesti sõnaraamat” (1982). Tavalisem on siiski märkimine “vastavalt vajadusele”: ÕS 1999, “Norra-eesti eesti-norra sõnaraamat” (Farbregd jt 1998), uuemad eesti-inglise sõnastikud (AULE, Veldi 2002). Terminoloogiasõnastikud tavaliselt sõnaliiki ei esita.

Siinkohal jälgime veidi lähemalt sõnaliikide esitust viies elektroonilises sõnastikus: VVS, EE-VN (3 köidet: A–P), SM-EE-vasted, ÕS 1999 ja EKSS (23 vihikut: A–unelus). Otsisime problemaatilisi juhtumeid: 1) artikleid, kus on mitu sõnaliigimärgendit, ja 2) leksikaliseerunud muutevorme, mis võiksid olla märksõna staatuses.

3.1.1. Mitu sõnaliigimärgendit

Kui sõna võib eri tähendustes või kontekstides kuuluda erinevatesse sõnaliikidesse, siis on mitu võimalikku esitusviisi. Esiteks võib sõnaartiklis olla loetletud mitu sõnaliiki koos, sõnaliigiinfo (näidetes alla joonitud) paikneb sel juhul märksõna (ja morfoloogilise info) järel (näide 1). See on iseloomulik juhtudel, kui sõnaliigierinevusega ei kaasne tähenduserinevusi (või kui sõnaraamat tähendusi ei erista).

- (1a) **h'ull** –u hullu **AS 22e** (VVS)
- (1b) **aamen** <'aamen 'aameni 'aameni[t -, 'aameni[te 'aamene[id S 02 && Interj> (EE-VN)
- (1c) **'ette DK** (SM-EE)
- (1d) **blond** <20: blondi, .blondi> omds, nimis (ÕS 1999)
- (1e) **hopsti adv. interj.** (EKSS)

Teiseks võivad eri sõnaliigid olla jagunenud eri artiklite vahel – sõna on esitatud mitme homonüümia ja sõnaliigiinfo paikneb vastava märksõna (ja morfoloogilise info) järel (2).

- (2a) **alam 1.** <8: -a> keskvõrdeline omds
alam 2. <8: -a> nimis (ÕS 1999)
- (2b) **haabjas**, -ja, -jat² **I adj.** haavakoore värvi ..
haabjas, -ja, -jat² **II s.** etn. Põhja-Eesti vana paadinimetus .. (EKSS)

Kolmandaks – polüsemsete sõnade puhul antakse sõnaliik tähendusnumbrite järel (3), väiksema tähendusliigenduse (täendusvarjundi või kasutusjuhtumi) puhul vastava tähise järel (4).

- (3a) **algul 1.** adv. esialgu; alguses .. **2. postp.** [gen.] millegi alguses (EKSS)
- (3b) **algul** <'algul Adv & Postp> **1.** <Adv> (esialgu) .. **2.** <Postp gen> (millegi alguses) (EE-VN)

(3c) **emane** .. 1. adj. ...; ant. isane .. 2. s. emasloom (EKSS)

(4) **juhm** adj. raske taibuga .. | (substantiivselt) .. (EKSS)

Märksõna ja numbrite järel kasutatakse üldiselt sõnaliigilühendeid, väiksema liigenduse järel on sageli kasutusel sõnaline väljend, nt *substantiivi / hüüdsõna funktsioonis, adjektiivselt* jms.

Kolme sõnastiku põhjal (VVS, EE-VN, EKSS) tehtud statistika näitas, et eesti keeles on mitmesse sõnaliiki kuuluvaid sõnu 2,2–2,5%. Kuivõrd topeltsõnaliikide märkimisviis on sõnastikes erinev ja sõnaliigilühendid võivad sõnaartiklis paikneda hajutatult (eri tähenduste all), siis eriti usaldusväärne see statistika pole, aga suurusjärgu siiski ehk annab. Kõige parema ülevaate saime EE-VN põhjal, kus sõnaliigiangmed on alati ka kirje päises olemas. Rohkem kui kolmandiku (36%) kõigist mitme sõnaliigiga sõnadest moodustavad adjektiivid, mis lauses võivad esineda ka substantiivina, ligi viiendik (19%) on adverbe, mis võivad käituda adjektiivina, peaaegu sama palju (18,5%) on substantiive, mis funktsioneerivad ka adjektiivina (vt tabel 2).

Tabel 2. Topeltsõnaliigid eesti-vene sõnaraamatus (EE-VN)

Sõnaliigid	Arv	Osakaal topelt-sõnaliikides	Näited
Adj Subst	338	36%	blond, eilne, emane
Adv Adj	182	19%	eraldi, harkis, ise+moodi
Subst Adj	176	18,5%	hauduja, jooksva+haige, kalda+pealne
Adv Interj	80	8,4%	kõlksti, justament
Interj Adv	72	7,6%	kõlks
Adv Postp	29	3%	algul, jälile
Num Subst	12	1,3%	kaks, kuuendik
Num Pron	8	0,8%	mitu+sada
Adv Prep	6	0,6%	paitsi, peal+pool
Subst Interj	6	0,6%	amps, hurraa
Adj Adv	5	0,5%	jube
Konj Adv	4	0,4%	just+kui
Subst Adv	4	0,4%	inkognito
Pron Subst	2	0,2%	paar
muud (10)	1	0,1%	nt Konj Prep: kuni

Kui liidame kokku samade liikmete eri järjestusega paarid, siis näeme, et eesti keeles on kõige sagedamad adjektiiv-substantiivi kombinatsioonid (kokku 54,5%), järgnevad adverbi-adjektiiv (19,5%) ning adverbi-interjektsiooni (16%) kombinatsioonid. Kõige hägusam on adjektiivide klass, mille liikmed osalevad ühtekokku rohkem kui kolmveerandis topeltsõnaliigiga paaris. Muid variante – numeraal-substantiiv, numeraal-pronoomen, substantiiv-interjektsioon jt – on hoopis vähem (kokku 6%).

3.1.2. Leksikaliseerunud sõnavormid

Vahel on raske otsustada, kas tekstis kasutatav sõnavorm kuulub mingi sõna muutmisparadigmasse, olles noomeni või verbi muutevormi staatuses, või on ta iseseisev muutumatu sõna, mis peaks saama hoopis teise sõnaliigimääratluse. Leksikaliseerunud sõnavormide esitus on hästi kirju, lahknevusi on nii sõnastike vahel kui nende sees. Niisugune vorm võib esineda märksõnana, millele järgneb terve sõnaartikkel (5) või ainult viide algvormi artiklile (6).

- (5a) **.õue** (*majast välja*). Läksin õue. Õues on külm. Õuest tulnu (ÕS 1999)
- (5b) **levinud**, -nu² partits. laialdase levikuga, üldiselt tuntud v. tarvitatud .. (EKSS)
- (5c) **lõpuks** <lõpuks Adv> **1.** (*viimaks*) .. **2.** (*tegelikult, päriselt*) .. (EE-VN)
- (6a) **alates** vt .algama (ÕS 1999)
- (6b) **jõkke** → jõgi (EE-VN)
- (6c) **'austusega** sg kom → 'austus (SM-EE-vasted)

Artikli sees võib leksikaliseerunud vormi kohata mitmel pool (siin alla joonitud). See võib olla vormistatud allmärksõnana tähendusvarjundi või tarvitusjuhtumi tähistaja järel (7), paradigmavälise vormina muutmisinfor sees (8) või pelga kasutusnäitena mõnes väljendis (9).

- (7a) **algama**, alata⁴⁸ **1.** algust saama .. **2.** (midagi) alustama .. || **alates** millestki peale, alustades .. (EKSS)
- (7b) **lõpp**, lõpu²² s. **1.** .. **2.** .. || **lõpuni** täielikult, täiesti, viimase piirini .. (EKSS)
- (8) **.hark** <20: hargi, .harki; seisundivorm .harkis> .. (ÕS 1999)
- (9a) **levima** <50>. Lööve levib .. **Levinud** eesnimi .. (ÕS 1999)

- (9b) **.algama** <68: alata, .algan>. Pidu algab .. Töötab meil sügisest, esimesest märtsist alates (*peale*). Alates lastest ja lõpetades vanakestega *lastest vanakesteni* (ÕS 1999)
- (9c) **selg**, selja²³ s. **1.** inimese keha tagaosas õlgadest tuharateni .. || (sisekohakäänetes adverbilaadselt ..) ülle, üll, ült. *Pani kleidi selga* ..*Aita mul mantel seljast!* .. (EKSS)

Sama vorm võib olla tõlgendatud ka kaheti ning korduda eri rollides: nii märksõnana (10a) kui ka näitena (10b).

- (10a) **elagu** <8>. Olümpiavõitjat tervitati elaguga, elagu hüüetega (ÕS 1999)
- (10b) **elama** <50> .. Elagu juubilar! (ÕS 1999)

3.2. Süntakiline info: sõltuvussuhted

Süntakiline info on rohkem seotud sõnade omavaheliste suhetega lauses, mitte üksiksõnadega, seetõttu ei ole seda mugav sõnaraamatus kirjeldada. Põhiline, mida eesti sõnastikes süntaksi alal üldse eksplitsiitselt esitatakse, on verbi sõltuvusnõuded laiendi vormile ehk rektsioon; vahel on märgitud ka teiste sõnaliikide (substantiivi, adjektiivi, adverbi, kaassõna) rektsiooni. Suuremates sõnaraamatutes on tavaliselt välja toodud püsiühendid ja muud kollokatsioonid, aga üksnes kasutusnäidete kujul, mitte süstemaatilise grammatilise info osana.

Ka rektsiooni märgivad eesti sõnastikud üldiselt ainult “vajaduse korral”, nagu eessõnades tavaliselt öeldakse. Üldsõnastikest tähistab sõltuvusnõudeid järjekindlalt üksnes AULE (POOL on spetsiaalne rektsioonisõnastik), teistes sõnastikes on seda esitatud rohkem või vähem juhuslikult (vt tabel 3).

Tabel 3. Sõltuvusnõuded eesti sõnastikes

	ÕS 1999	SM-EE	EE-VN	EKSS	AULE	POOL
verbi sõltuvusnõuded	(+) küsimus	(+) umbmäärane asesõna	((+)) küsimus	(+) kommentaar	+ umbmäärane asesõna	+ küsimus / vormilühend
muude sõnaliikide sõltuvusnõuded	(+)	((+))	((+))	(+) vormilühend	+	

Nii nagu sõnaliikidega juba nägime, on ka rektsiooni märkimiseks kasutusel mitmeid eri viise. Enamasti tehakse seda rektsiooniküsimuse abil (11a–f) või umbmäärase asesõnaga (11g, 12), vahel vormilühendiga (13), vahel lisatakse sulgudes sõnaline kommentaar (14). Oma asukoha poolest võib rektsiooniinfo olla iseseisev struktuurielement sõnaartiklis (12–14), märksõnarektsioon võib olla paigutatud morfoloogiakirje sisse (11a–c), vahel ka märksõna ette (11d). Sagedamini esineb rektsioonisõna näitefraasi juures (selle ees, sees või järel) (11e–g).

- (11a) **edestama** <52: keda> (ÕS 1999)
- (11b) **impo.neerima** <55: -.neerida, -neerin; millega kellele> (ÕS 1999)
- (11c) **samane** <13: -se; millega> (ÕS 1999)
- (11d) <mille> **jooksul** *vältel, kestel* (ÕS 1999)
- (11e) **arve** .. arveid klaarima v õiendama kellega (EE-VN)
- (11f) **käsi** .. *mis* on kellele käe järgi (EE-VN)
- (11g) **järgima** .. [kellegi] eeskuju järgima .. (AULE)
- (12a) **haarduma** .. [millegi külge v. millesegi] .. [millegi ümber].. [millega-gi] .. (AULE)
- (12b) **aprikoida**⁶⁸ arutleda {*jtak midagi, millegi üle*}; mõtiskleda {*jtak millest-ki, millegi üle*}; juurelda {*jtak millegi üle, millegi kallal*} (SM-EE)
- (13) **tänu** prep. [allat.] (EKSS)
- (14a) **harjutama 2.** (koos ma-infinitiiviga). *Ema harjutab last kõndima.* (EKSS)
- (14b) **ümber käima** kõnek. **1.** (komitatiiviga.) kohtlema .. (EKSS)
- (14c) **korda minema 1.** (sageli adessiiviga.) õnnestuma .. (EKSS)

Kõige tavalisem on aga rektsiooni edastamine üksnes näidete kaudu, millest sõnaraamatu kasutaja peab ise välja lugema verbi poolt nõutava vormi (15):

- (15a) **tulenema** <50> *lähtuma, põhjustatud olema*. Segadus tulenes eksiarvamustest. Sellest tulenevalt peame .. (ÕS 1999)
- (15b) **tulenema 1.** põhjustatud, tingitud olema. *Kehv saak tulenes maa viletsast väetamisest*. .. **2.** lähtuma, pärit olema, millestki saadud olema. .. *Talude nimed tulenevad sageli nende kunagiste peremeeste nimedest*. (EKSS)

Andres Aule suur “Tänapäeva eesti-inglise sõnaraamat” (= AULE), millest seni on ilmunud kaks köidet (A–K), on meeldiv erand üldkeele sõnastike hulgas. Rektsiooniteave (nurksulgudes) on seal antud tõlkevastete järel eraldi rektsiooniplokis, kus märksõna ennast üldiselt ei korrata.

Sõnajärge näidatakse vajadusel tilde (~) abil. Rektsiooni kirjeldatakse umbmääraste asesõnadega ja sõnaühenditega või tüüp kõrvallause tähisega ([et ...], [kas ...] jt – tõsi, siin on küll ka kõige rohkem ebaühtlusi näha). AULE sõnaraamatus on kirjeldatud nii verbi- (16), substantiivi- (17), adjektiivi- (18), adverbi- (19) kui ka kvantorirektsiooni (20) (kuigi sõnaliik ise ei ole märgitud).

(16a) **edvistama** .. [millegagi] ..; [kellegagi v. kellegi ees] ..

(16b) **kaaluma 3** .. [midagi] ~ ..; ~ [kas ...] ..

(17) **himu** .. [millegi] ~ v. ~ [millegi järele] ..; ~ [midagi teha] ..

(18) **analoogiline** .. [millegagi v. millelegi] ~ ..

(19) **analoogiliselt** .. [millegagi] ~ ..

(20) **hulk** .. [millegi] ~ ..; suur hulk [midagi v. kedagi] ..

Ka näited on varustatud rektsiooniga (21).

(21a) **hoop** .. [kellelegi] hoopi andma ..

(21b) **jagama** .. [midagi] pooleks jagama ..; [kellegagi] korterit jagama ..; [kellegi] seisukohta jagama ..

Eesti keele ainus spetsiaalne rektsioonisõnastik (kahjuks väike) on Raili Pooli “Eesti keele verbirektsioone” (= POOL). Raamatu põhiosa moodustab lihtverbide loend (kokku 563 verbi) koos tavalisemate rektsioonikirjelduste ja näitelausetega, eraldi on esitatud tavalisemad ühend- ja väljendverbid ning nende rektsioonid. Rektsioonikirjeldustena on esitatud käändeküsimused, lisaks neile ka kaassõnad, *ma-*, *da-* või *vat-*infinitiivi tähised ning *ma-*infinitiivi käändelised vormid. Kolmekäändeline objekt on tähistatud lühendiga O. Välja on toodud ka verbi fakultatiivseid laiendeid, mis on pandud sulgudesse. Kui verbil on mitu “täiesti erinevat tähendust” (POOL: 6), siis on rektsioonikirjeldused jagatud verbi tähenduse järgi rühmadeks ning eristatud araabia numbriga. Näitelause on paksu kirjaga osutatud sõnadele, mis näidatud rektsiooni esindavad (kolmekäändelise objekti sõna ei ole esile tõstetud). Kuna raamat on mõeldud eelkõige neile, kes eesti keelt võõrkeelena õpivad või õpetavad, siis on näitelauseid tõlgitud ka inglise, soome ja vene keelde. (Vt tabel 4.)

Tabel 4. Näiteid verbide reksioonisõnastikust (POOL)

Märksõna	Reksioon	Näide
lugema	kust? + O	Lugesin seda õpikust .
astuma	1. kuhu? 2. kuhu?	Rein astus ülikooli . Astusin tänavale/poriloiku .
hoiduma	millest? <i>-mast</i> mille eest?	Hoidu suitsetamisest! Hoidu külmetamast . Hoidu rongi eest!
keelitama	keda? + <i>ma</i> -inf	Keelitasin oma meest teatrisse tulema .
eelistama	keda? mida? (+ kellele? millele?)	Eelistan teed (kohvile) .
võrdlema	keda? mida? + kellega? millega?	Võrdlesin üht pilti teisega .
teadlik olema	millest?	Ta on oma haigusest teadlik.

3.3. Semantiline info

Semantilist infot (grammatilise info liigina) esindab eesti sõnastikes hierarhiline info Tartu Ülikoolis koostatava eesti üldkeelesaurus¹⁵ näol. 2004. a seisuga sisaldab tesaurus küll üksnes 10 000 mõistet, ent autorite väitel “näitavad katsed tekstisõnade tähenduste ühestamisega, et eesti põhisõnavara tähenduste hulk peaks olema enamuses tesaurus kirjetega kaetud”¹⁶.

Semantiliste rollide (argumendistruktuuri) märkimine tuleb muus maailmaski esile ainult (spetsiaalsetes) suuremates leksikaalsetes andmebaasides, nagu eespool ülevaates nägime. Eesti sõnaraamatutes ega andmebaasides seda senini tehtud ei ole.

¹⁵ Vt <http://test.cl.ut.ee/ressursid/teksaurus/> (1.11.2004).

¹⁶ Samas.

4. Grammatiline info eesti–X-keele sõnastikus

EXS on kavandatud kui universaalne ükskeelne leksikaal-grammatiline andmebaas, mis sobiks aluseks eelkõige uute eesti lähtekeeleka kakskeelsete sõnaraamatute koostamisel (inimese jaoks), aga ühtlasi peaks ta olema kasutatav mitmesugustes keeletehnoloogia rakendustes (arvuti jaoks). Sellest on tingitud tavapärasest suurem tähelepanu ja kõrgemad nõuded formalismile. Käesolev peatükk ei paku veel süstemaatiliselt läbimõeldud esituskava, vaid mõningaid põhimõtteid ja üksikuid detaile.

EXS-is kavatseme anda grammatilist infot mitmelt keelekirjeldustasandilt: 1) morfofonoloogiast välde ja ebareeglipärane rõhk, 2) morfoloogiast põhjalik vormimoodustuse kirjeldus¹⁷ ja jõudumööda andmed ka vormikasutuse kohta, 3) tuletusest kindlasti liitsõna-liigendus ja komparatsioon¹⁸, võib-olla ka tuletustüübid¹⁹, 4) morfosüntaksist sõnaliik ja arv, 5) süntaksist sõltuvusnõuded ja püsiühendid, 6) semantikast esialgu semantilised rollid, võimaluse korral edaspidi ka semantilised tüübid. Siin vaatame lähemalt süntaksiga seotud infot: sõnaliike, sõltuvusnõudeid ja semantilisi rolle.

4.1. Morfosüntaktiline info: sõnaliigid

4.1.1. Sõnaliigi hägusus

Sõnaliik on mõneti ebaselge kategooria, sest tema määratlemise aluseks on kolme eri tasandi tunnused (morfoloogia, süntaks ja semantika) ja need ei ole kõigil sõnadel võrdselt esindatud. On suur hulk sõnu, mis eristuvad teistest selgesti kõigi kolme tunnuse alusel – need kuuluvad sõnaliikide tuumossa, aga ka perifeeriaosa pole väga väike.

Segadused sõnaliikide piiridega on esiteks tingitud sellest, et klassid ise on määratletud puudulikult, kõiki liigitusaluseid arvestamata. Näiteks ei tule pronoomenite ja numeraalide liigitamisel arvesse süntaktilised

¹⁷ Morfoloogilisest infost sõnastikukirjes ja selle lisamise viisist vt Viks 2000a.

¹⁸ Komparatsiooni koht keeletasandite jaotuses on käsitlustes erinev: kord kuulub ta kokku vormimoodustusega, kord moodustab omaette tasandi, EXS-is on komparatsioon arvatud sõnamoodustuse juurde, sest nii tuletamise kui ka võrdlemise protsessi tulemuseks on morfoloogiliselt iseseisev sõna oma muutevormidega.

¹⁹ Tuge pakub siin Silvi Vare koostatav sõnamoodustussõnaraamat, selle kohta lähemalt vt Vare 2002.

tunnused ja mõlemad sõnaliigid võiks omakorda jagada allklassideks vastavalt sellele, kas nad käituvad lauses substantiivi või adjektiivi kombel. Nii jagunevad pronomeniid omadussõnalisteks asesõnadeks (*selline, kogu*) ja nimisõnalisteks asesõnadeks (*tema, keegi*), numeraalid aga (nimisõnalisteks) põhiarvudeks (*neli*) ja (omadussõnalisteks) järgarvudeks (*neljas*).

Teiseks võib üks sõna sõltuvalt kontekstist kuuluda kord ühte, kord teise sõnaliiki. Osa sõnu kuulubki kahe tunnuse alusel ühte sõnaliiki, aga kolmanda alusel teise. Näiteks seisundisõnad väljendavad küll omadust ja esinevad täiendina, nagu adjektiivid (*kõssis* poisike), aga nad ei käändu ja võivad väljendada ka viisi, nagu adverbid (*istus kõssis*). Hulk sõnu on aga mingi tunnuse suhtes ambivalentset. Näiteks on eesti keeles adjektiivi ja substantiivi erinevus üsna hajus ja omadussõnad võivad sageli lauses käituda nimisõnadena (*blondi* naljad). Omadusest saab rääkida kui asjast (daam *kollases*, mure *homse* pärast), vastupidist (asi kui omadus) esineb eesti keeles vähem, aga siiski (*koer* poiss, *lõpp* lugu).

Kolmandaks toimub keeles pidev leksikaliseerumise protsess – mõne sõna mõni muutevorm saab iseseisva tähenduse ja süntaktilise funktsiooni, muutudes pikkamisi sõnavormist (muutumatuks) sõnaks. Näiteks on paljudel juhtudel ilma kontekstita võimatu öelda, kas tegemist on sõnavormi või sõnaga: *selg* – *selga*, *seljas*, *seljast*, *tasu* – *tasuta* (*vrđ tasuline*), *lõpp* – *lõpuks*, *lõpuni*, *elama* – *elagu*, *elades*, *algama* – *alates*.

Nagu näha, pole mõnede sõnade sõnaliigilist kuuluvust väljaspool konteksti võimalik üheselt kindlaks teha. Sõnastikuinfo esitamisel tuleb arvesse võtta vastavate sõnade konkreetseid esinemisjuhte ja ühe või teise funktsiooni sagedust tekstis, nii on teatud sõnade korral võimalik (ja ka vajalik) rääkida sõnaliigilisest primaarsusest.

4.1.2. Sõnaliigi esitus

EXS-i üldine põhimõte on, et kõik sõnaliigiandmed kajastuvad kindlasti artikli päises, kus (koos morfoloogiaga) esitatakse kogu sõnaliigiinfo. Kui märksõnal tähendusi ei eristata, siis sellega asi piirdubki (22). Kui sõnal on mitu tähendust, on sõnaliik ära toodud iga tähenduse juures (23, 24). On juhtumeid, kus ka tähendusnumbri järel on mitu sõnaliiki (23), aga enamasti on iga üksiktähendus seotud ühe sõnaliigiga (24).

- (22a) **ese** <eseme 4> *Subst*
 (22b) **kössis** *Adv Adj* kõveras, kühmus
 (23) **lupsti** *Adv Interj*
 1 *Adv Interj* lupsuga (1), lupsu tehes
 2 *Adv Interj* kiiresti ning kergesti
 (24) **hall_2** <halli 22> *Adj S*
 1 *Adj* värvuselt musta ja valge vahepealne
 2 *Adj* valge ja pimedaga vahepealne, hämar; sombune
 3 *Adj* kaugesse minevikku ulatuv, ajaloo hämarusse kaduv
 4 *Adj* piltl. üksluine, ühetooniline, vaheldusetu, tuim, igav
 5 *Subst* see, kes v mis on hall (1)

Leksikaliseerunud vormid, mis on piisavalt iseseisvad, saavad EXS-is oma sõnartikli, kus on antud ka sõnaliik. Lisaks tuleb vormivii-
 de algvormile (25).

- (25a) **alates** *Postp Prep* → algama ..
 (25a) **tasuta** *Adj Adv* → tasuta ..
 (25c) **selga** *Adv* → selg ..
 (25d) **harkis** *Adj Adv* → hark ..
 (25e) **elagu** *Interj* → elama ..
 (25f) **austusega** *Adv* → austus ..
 (25g) **austatud** *Adj* → austama ..
 (25h) **levinud** *Adj* → levima ..

Sõnavormidele märksõna staatuse andmisel järgime “pigem rohkem”-
 põhimõtet. Eelkõige sellepärast, et teistes keeltes võib sellistel puhkudel
 olla vasteks iseseisev sõna (nagu näiteks SM-EE-s: *austusega* – sm
kunnioittavasti; alguses – sm *alkuaan, alulla, alumpana, ensimmältä*).

Sõnaliigiinfo täpsustamiseks on EXS-is kavas jagada osa sõnaliike
 alaliikidesse, nt pronoomenid nimi- ja omadussõnalisteks, samuti nume-
 raalid. Selleks lisatakse primaarsele sõnaliigitähisele juurde teine, täpsus-
 tav sõnaliigitunnus.

- Pron_Subst – substantiivne pronoomen (*tema, see*)
 Pron_Adj – adjektiivne pronoomen (*selline*)
 Num_Adj – adjektiivne numeraal ehk järgarvsõna (*kolmas*)

Alaliigitamisel on mõttekas eristada hulgasõnu (ehk kvantoreid), mis
 süntaksi seisukohalt on lähedased arvsõnadele. Kvantori tunnuse (_q)

saaksid lisaks arvsõnadele ka mõõtu või rühma märkivad nimisõnad (*meeter, kimp*), asesõna (*mitu*), määrsõnad (*palju*).

- Num_q – kvantor-numeraal (*kaks*)
- S_q – kvantor-substantsiiv (*liiter, kilo, hunnik*)
- Pron_q – kvantor-pronoomen (*mitu*)
- Adv_q – kvantor-adverb (*palju, natuke*)

Ka verbi alaliigitamine võiks kasulik olla: eraldi abiverbid, modaalverbid ja tugiverbid.

- V_abi – abiverb (*olema* osa tähendusi)
- V_mod – modaalverb (*näima, võima* jt)
- V_tugi – tugiverb (*tegema, ajama* jt)

4.2. Süntaktiline info: sõltuvussuhted

Sõnastiku süntaktilise info hulka kuuluvad sõltuvusseosed, laiendite kohustuslikkus, tüüpilised konstruktsioonid, sõnajärg jms. Eesti sõnastikes on seni pööratud suurimat tähelepanu verbi rektsioonidele, eriti neile, milles sagedamini eksitakse. Rektsioon on aga ainult üks aspekt verbi sõltuvusnõuetest – see, mis määrab verbi laiendi morfoloogilise vormistuse. Adekvaatse pildi loomiseks on vaja arvestada ka lause moodustajate (argumentide) teisi omadusi: laiendite arvu ja nende süntaktilisi funktsioone.

Sõltumine ehk rektsioon on teatavasti niisugune grammatiline suhe lause moodustajate vahel, mille puhul “alistava moodustaja” ehk põhja semantilised omadused määravad “alistuja” ehk laiendi grammatilise vormi. Sõltuvalt sellest, mis laadi laiendivormi põhisõna tingib, võime rektsiooni jagada (Rätsep 1978: 64):

- 1) käänderektsioon, nt *austame vanemaid* (partitiiv);
- 2) kaassõnarektsioon, nt *vastutab tagajärgede eest*;
- 3) infiniitvormirektsioon, nt *proovib laulda*;
- 4) lause- ja lausungirektsioon, nt *teab, mida teeb; rääkisime, et olete kirjanik*;
- 5) substitutsiooniklassi-rektsioon (nt suunaklass), nt *riputa jope KUUHU* (nt *nagi otsa / nagisse / kappi / nõõrile*). (Oleme seda nimetanud vormiklassiks.)

Paljude keelte (eriti inglise) sõnaraamatutes on sageli juba verbi sõnaliigitähise juures kirjas objekti olemasolu või puudumine, st transitiivsus / intransitiivsus. Erinevalt inglise keelest eesti verbi (in)transitiivsust sõnaliigi tasemel ei märgita. Sellest oleks ka vähe abi, sest eesti keeles on oluline eristada sihitise käände vaheldumise võimalust, aga see on juba rektsiooni tasandi probleem (millist muutevormi valida) ja sõltub lausekontekstist.

Lisaks objektile on vaja märkida ka muude lauseliikmete olemasolu. Näiteks mõnedel ilma- või füsioloogianähtustega seotud verbidel võib subjekt täiesti puududa (*Sajab. Mul iiveldab.*). See pole sama nähtus kui subjekti puudumine lauses, mis väljendavad tegijat ehk agenti implitsiitselt: pöördelõpu kaudu (*Jää-n magama.*) või verbi impersonaalse vormi vahendusel (*Joodi õlut.*). Süntaktiliste funktsioonide esitamine võimaldab näidata ka, et predikatiiv seostub ainult *olema*-verbiga (*Ilm on ilus.*) või et adverbiaale võib olla rohkem kui üks (*liikus klassist klassi*). Jne.

Objektiivseid andmeid eesti verbi laiendite hulga ja funktsioonide kohta sõnaraamatutest ei leia. Ainus allikas, mis sedalaadi infot süstemaatiliselt pakub, on Huno Rätsepa verbikesksete lausemallide käsitus (Rätsep 1978). Moodustajate arvu pole seal otseselt antud, see kajastub elementaarlausete jagamises obligatoorse laiendi järgi 0–4-kohalisteks. Lauseliikmete olemasolu verbinõudena H. Rätsep küll märgib, aga teeb seda kaudselt – positsiooni ja vormide vahendusel, ilma lauseliikmeid nimetamata. Fakultatiivseid laiendeid eristavad kohustuslikest sulud.

Kuigi verb on lause keskpunkt, ei koonu sõltuvussuhted lauses ainult verbi ümber. Ka muudes fraasides sõltub laiendite olemasolu ja grammatiline vorm fraasi põhjaks oleva sõna leksikaalsest tähendusest. Nii võib rektsiooni jagada ka põhja sõnaklassi järgi (vt ka Rätsep 1978: 65, EKG II: 137–141):

- 1) verbirektsioon, nt *sõltub tingimustest, tahab magada*;
- 2) substantiivirektsioon, nt *sõltuvus perekonnast, huvi muusika vastu*;
- 3) adjektiivirektsioon, nt *kuri sõbra peale, sarnane õega, kolme meetri laiune*;
- 4) adverbirektsioon, nt *käbi ei kuku kannust kaugele*;
- 5) kaassõnarektsioon, nt *laua all, päev otsa, trepist üleest, enne õhtut, läbi akna, ilma rahata*;
- 6) kvantorirektsioon, nt *mitu meetrit, kaks ahju, hunnik raamatuid, kamp noori*.

EXS-is keskendume esialgu ainult verbi sõltuvussuhetele, mille kohta on kõige rohkem lähtematerjali. Info muude sõnaliikide kohta tuleb hiljem juurde, aga see vajab enne täiendavat uurimistööd. Iga verbi jaoks näidatakse tema võimalikud laiendid koos süntaktilise funktsiooniga ja reksioonivõimalustega.

4.3. Semantiline info: semantilised rollid

Aktantide (lausesituatsioonis osalejate) semantilisi rolle (toona käände-grammatika nime all) on 1970-ndate eestikeelses teaduskirjanduses korraldvalt arutatud.²⁰ Kõige tähtsamad maailma kirjeldavad mõisted grammatikaliseeruvad, vastava protsessi käigus saavad semantilised rollid endale käänete, partiklite või sõnajärjepiirangute kujul kindla morfosüntaktilise vormistuse (Karlsson 2002: 181–182, 258 jm). Rollide hulka ei ole kuskil lõplikult määratletud, üldiselt nimetatakse umbes tosinat rolli. Kõige olulisemad neist on agent ja objekt. Kui süntaktilised funktsioonid on keeliti erinevamad, siis semantilised rollid on universaalsed (Rajandi 1976: 82). Rollide näiteid:

- 1) agent ehk tegija – tahtlik tegevuse põhjustaja, tegevuse aktiivne sooritaja (*Mart loeb*);
- 2) põhjus ehk jõud ehk mõjutegur – mittetahtlik tegevuse põhjustaja (*üleujutus hävitas vana templi*);
- 3) kogeja – psüühilises protsessis olija (*Mart oli pahane*);
- 4) patsient – kellele tegevus rakendub (*Mart lõi Tõnist*);
- 5) objekt ehk teema – millele tegevus rakendub (*Mardile meeldib õlu*; *Mart kirjutas kirja*);
- 6) benefitsient ehk saaja – see, kelle kasuks tegevus sooritatakse (*Mart andis sajalise Tõnisele*);
- 7) suund, siht – sihtkoht, kuhu toimub liikumine (*Mart sõitis Tartusse*);
- 8) instrument – tegevuses kasutatav vahend (*Mart viskas Tõnist kiviga*);
- 9) tulemus – tegevuse lõppsaadus (*Mart küpsetas piruka*);
- 10) allikas – kust tegevus lähtub (*rakett lasti õhku kindrali käsu peale*);

²⁰ Vt näiteks Õim 1976, Rajandi 1976, Ereht 1979 jt.

11) kaasosaline/partner – mitu (vastastikku suhtlevat) kaasosalist (*Mart ja Tõnis vaidlesid kogu öö*).

Keelesüsteemis on olemas mitmesuguste erinevate omadustega verbe, mis aitavad vormistada erinevaid aktantseid rolle, näiteks kogejarolli vormistamisel alusena on vaja kasutada verbi *kartma*, sihitisena verbi *hirmutama* (Rajandi 1976: 80). Jõudumööda proovimegi EXS-is verbi argumentide semantilisi rolle tähistada, kuid kogu see valdkond vajab eelnevalt põhjalikku uurimist. Rollid antakse kõigile argumentidele, mitte ainult neile, mis seostuvad rektsiooniga (26).

(26a) Mart küpsetas piruka. [agent] V [tulemus]

(26b) Mart küpsetas piruka Tõnisele. [agent] V [tulemus] [benefitsient]

4.4. Süntaksiga seotud info formaalne esitus

Keeletehnoloogia rakendusi silmas pidades püüame oma baasis esitada kõik andmed formaalsel kujul. Mil moel üht või teist asja konkreetsetes eesti lähtekeele kakskeelses sõnaraamatus esitada, see on eraldi küsimus. Üht ja sama formaalset kirjeldust saab teisendada väga mitmele eri kujule, nii et kaks sõnaraamatut võivad selles osas tundmatuseni erineda, nii trükiteksti väljanägemise poolest kui ka andmete valiku poolest. Projekti “Eesti–X-keele sõnastik” eesmärk on talletada ühe või teise eesti keele sõna kohta võimalikult palju kasulikku infot, et see oleks küllaldane ja teaberohke alus tulevaste instituudis tehtavate kakskeelsete – ja miks ka mitte ükskeelsete – sõnastike tarvis.

Oleme välja töötanud leksikaalse andmebaasi XML-põhise struktuuri-kirjelduse, kus igale infoüksusele vastab omaette struktuuriüksus. Sõnastiku koostamiseks ja toimetamiseks on loodud spetsiaalne oma-maja-tarkvara.

Järgnevas näites (27) on toodud verbi *kirjutama* ühe tähenduse ‘kirja panema’ juurde kuuluva süntaktilise info võimalik esitus EXS-is (lihtsus-tatud kujul). Süntaktiline info on koondatud omaette üksusesse (<SYNT>), mis omakorda hõlmab eri argumentide kirjeldusi (<ARG1>, <ARG2> jne). Iga argumendiüksuse sisus esitatakse kõigi kolme tasandi – morfoloogia, süntaksi ja semantika – nõuded:

1) semantilist funktsiooni märgib semantiline roll (Agent, Teema, Instrument, Koht);

- 2) süntaktilist fraasistruktuuri märgib lauseliige (Subj (alus), Obj (sihitis), Advb (määrus));
- 3) morfoloogilist vormistust märgib rektsioon (NF_Nom (noomeni-fraas nominatiivis), NF_NomGenPart (noomenifraas nominatiivis / genitiivis / partitiivis ehk kolmekäändeline objekt), NF_Vormiklass:kuhu (noomeni kuhu-suunaline vormiklass: illatiiv, allatiiv, kaassõnaühendid). Kolme eri tasandi info on näites omavahel eristatud tärnidega.

(27) **kirjutama 1** kirja panema

<SYNT>

<ARG1>Agent * Subj * NF_Nom</ARG1>

<ARG2>Teema * Obj * NF_NomGenPart / osalause</ARG2>

<ARG3>Instrument * Advb * NF_Kom</ARG3>

<ARG4>Koht:kuhu * Advb * NF_Vormiklass:kuhu</ARG4></SYNT>

Lausenäited esitatud argumentide kohta:

ARG1: Mart kirjutas.

ARG2: Mart kirjutas kirja. Mart kirjutas, et käis ujumas.

ARG3: Mart kirjutas pliiatsiga.

ARG4: Mart kirjutas paberile / raamatusse / tahvli peale.

Süntaktilise info esitamisel sõnastikus tekib probleeme palju, näiteks on keeruline otsustada, kas mingi laiend on kohustuslik või mitte; kas objekt esineb ainult partitiivis (osaobjekt) või ka nominatiivis ja genitiivis (täisobjekt); milline semantiline roll on verbi ühel või teisel argumendil. Sisuliste probleemide kõrval (ja nendega koos) tuleb lahendada ka otse-selt formalismiga seotud küsimused: kuidas ja kui palju märkida fakultatiivseid laiendeid; kuidas panna kirja süntaktiline variatiivsus (nt *vend sarnanes õega ~ vend ja õde sarnanesid*) jne.

5. Kokkuvõte

Usutavasti on lugejalegi selge, et süntaktilise info esitus, mis oleks üht-aegu nii ammendav (andmebaasis, rakenduse ja keeleuurimise jaoks) kui ka hõlpsalt loetav ja mõistetav (sõnaraamatus, inimese jaoks), on üksjagu keerukas ülesanne. Kui arvuti vajab peensusteni detailset kirjeldust, siis inimest liigne detailsus pigem segab. Süntaktilise, st aktiivse info pooldest on üldiselt olnud suhteliselt vaesemad just eesti kakskeelsed sõnaraama-

tud, ükskeelsed sõnaraamatud sisaldavad küll rikkalikku ainestikku, aga tihti implitsiitsel moel. EXS-i andmebaasis kirjeldatav süntaktiline info on neutraalne alus, millest lähtuda konkreetsetes leksikograafiatöös. Loodame, et sõnastiku koostamiseks ja toimetamiseks loodud spetsiaalne oma-maja-tarkvara võimaldab meil eesti sõnade kohta salvestada infot, millest saavad kasu võimalikult paljud. Soovitud eesmärgini jõudmine soovitud kujul võib aga võtta veel palju aega ja tööd.

Kirjandus

- AULE = Aule, Andres 2001–2003. Tänapäeva eesti-inglise sõnaraamat I–II. Tallinn: Eesti Keele Sihtasutus.
- Atkins, Beryl T. Sue 1996. Bilingual dictionaries: Past, present and future. – Euralex '96. Proceedings I–II. Ed. by Martin Gellerstam, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström, Catarina Røjder Papmehl. Göteborg University, Department of Swedish, 515–546.
- Calzolari, Nicoletta & Grishman, Ralph & Palmer, Martha (responsible authors) 2001. Survey of Major Approaches Towards Bilingual/Multilingual Lexicons. ISLE Computational Lexicons Working Group. Deliverable D2.1–D3.1. February 2001.
- EAGLES = <http://www.ilc.cnr.it/EAGLES96/home.html> (1.11.2004).
- EE-VN = Eesti-vene sõnaraamat 1–3 [5]. 1997–2003–. Eesti Keele Instituut. Tallinn.
- EKG II = Erelt, Mati & Kasik, Reet & Metslang, Helle & Rajandi, Henno & Ross, Kristiina & Saari, Henn & Tael, Kaja & Vare, Silvi 1993. Eesti keele grammatika II. Süntaks. Lisa: kiri. Eesti TA Keele ja Kirjanduse Instituut. Tallinn.
- EKSS = Eesti kirjakeele seletussõnaraamat I– (1–24). 1988–2004–. Eesti Keele Instituut. Tallinn.
- Erelt, Mati 1979. Eesti lihtlause probleeme. Tallinn: Eesti Raamat.
- Farbregd, Turid & Kangur, Sigrid & Viks, Ülle 1998. Norra-eesti eesti-norra sõnaraamat. Tallinn: Eesti Keele Sihtasutus.
- FrameNet = <http://www.icsi.berkeley.edu/~framenet/> (30.10.2004).
- Ibson, Robert 1992. BudaLEX Presidential Debate 1988: Part 4. – International Journal of Lexicography Vol. 5, No. 4, 275–277.
- ISLE = http://www.ilc.cnr.it/EAGLES96/isle/ISLE_Home_Page.htm (1.11.2004).
- Kaalep, Heiki-Jaan & Muischnek, Kadri & Müürisep, Kaili & Rääbis, Andriela & Habicht, Külli 2000. Kas tegelik tekst allub eesti keele morfoloogilistele kirjeldustele? Eesti kirjakeele testkorpuse morfosüntaktilise märgendamise kogemusest. – Keel ja Kirjandus 9, 623–633.
- Kann, Kallista 1995. Prantsuse-eesti sõnaraamat. 3., täiendatud trükk. Toim ja täiend Lauri Leesi. Tallinn: Valgus.
- Karlsson, Fred 2002. Üldkeeleteadus. Tõlk ja kohand Renate Pajusalu, Jüri Valge ja Ilona Tragel. Tallinn: Eesti Keele Sihtasutus.

- Langemets, Margit 2003. Kas ükskeelne või kakskeelne sõnaraamat? – Toimiv keel I. Töid rakenduslingvistika alalt. Eesti Keele Instituudi toimetised 12. Toim Heete Sahkai, Margit Langemets, Maria-Maren Sepper. Tallinna Pedagoogikaülikool, Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus, 151–177.
- Langemets, Margit 2004. Mõnda nimisõnade semantikast. – Keel ja Kirjandus 10, 736–754.
- LDOCE = Procter, Paul (ed.) 1995 (3rd edition). Longman Dictionary of Contemporary English. Harlow and London: Longman Group Ltd, 1978.
- Muischnek, Kadri 2004. Verbi ja noomeni püsiühenditest eesti keeles. – Keel ja Kirjandus 8, 574–589.
- POOL = Pool, Raili 1999. Eesti keele verbirektsioonid. Tartu: Tartu Ülikooli Kirjastus.
- Rajandi, Henno 1976. Keel: protsess ja süsteem. – Keel ja Kirjandus 2, 75–87, 3, 133–142.
- Rätsep, Huno 1978. Eesti keele lihtlausete tüübid. Emakeele Seltsi toimetised 12. Tallinn: Valgus.
- Saagpakk, Paul 1982. Eesti-inglise sõnaraamat / Estonian-English Dictionary. New Haven and London: Yale University Press.
- SM-EE = Soome-eesti suursõnaraamat 1–2. 2003. Eesti Keele Instituut, Kotimaisten kielten tutkimuskeskus. Tallinn-Helsinki: Eesti Keele Sihtasutus, Kotimaisten kielten tutkimuskeskus.
- Vare, Silvi 2002. Eesti keele sõnapere sõnaraamatu koostamise põhimõtetest. – Keel ja Kirjandus 12, 855–866.
- Veldi, Enn 2002. Eesti-inglise sõnaraamat. Tallinn: Koolibri.
- Viks, Ülle 2000a. Kuidas tekib sõnastikukirjese grammatika. – Keel ja Kirjandus 7, 486–495.
- Viks, Ülle 2000b. Tools for the generation of morphological entries in dictionaries. – Second International Conference on Language Resources and Evaluation. Proceedings Volume I. Athens, Greece, 31 May – 2 June 2000. Athens, 383–388.
- Viks, Ülle 2001. Muuttüübid eesti sõnastikes. – Leksikograafiaseminar: Sõna tänapäeva maailmas. Leksikograafinen seminaari: Sanat nykymaailmassa. Eesti Keele Instituudi toimetised 9. Tallinn, 157–185.
- VVS = Viks, Ülle 1992. Väike vormisõnastik 1–2. Keele ja Kirjanduse Instituut. Tallinn.
- Wakely, Richard & Béjoint, Henri 2002. Word groups in bilingual dictionaries: OHFD and after. – Lexicography and Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins. Ed. by Corréard, Marie-Hélène. United Kingdom: Euralex, 59–72.
- Wiedemann, Ferdinand Johann 1869. Ehstnisch-deutsches Wörterbuch. St. Peterburg: Kaiserliche Akademie der Wissenschaften.
- Õim, Haldur 1976. Elementaartähendused keele semantilises struktuuris. – Keel ja Kirjandus 10, 598–605, 11, 675–682.
- ÕS 1999 = Eesti keele sõnaraamat ÕS 1999. Toim Tiiu Erelt, koost Tiina Leemets, Sirje Mäearu, Maire Raadik ja Tiiu Erelt. Tallinn: Eesti Keele Sihtasutus.

Syntactic information in dictionaries: problems and solutions

MARGIT LANGEMETS, MERIKE MÄGEDI, ÜLLE VIKS

The article offers a brief overview of the kind and manner of syntactic information represented in typical machine-readable dictionaries and computational lexicons, focusing on the representation of this kind of information in Estonian dictionaries. Syntactic constructions have been described in several Estonian dictionaries, but mostly implicitly, by means of case examples, as no general syntactic encoding system has yet been elaborated. The ongoing national project *Estonian-X Dictionary*, started in 2002, aims to fulfil the needs of the source (L1) for different bilingual dictionaries and a new one-volume monolingual dictionary, as well as a variety of computer applications.

A closer look will be taken at morphosyntactic information, dealing mostly with parts of speech, and dependency information. The representation of dependency relations in the database of the Estonian-X Dictionary differs from the traditional practice in that the syntactic description of each verb consists of three components: syntactic function, thematic role and government.

The Estonian-X Dictionary is strictly formalized, using XML markup, in order to facilitate access by computers.

MARGIT LANGEMETS (1961) on lõpetanud Tartu Ülikooli eesti keele ja kirjanduse erialal. 2000. a kaitses Tartu Ülikoolis magistritöö arvutileksikograafiast, alates 2001. a Tallinna Pedagoogikaülikooli doktorant. Töötab Eesti Keele Instituudi leksikoloogiasektori juhatajana. Eesti Rakenduslingvistika Ühingu juhatuse liige. Uurimisvaldkonnad: leksikoloogia ja leksikograafia üldiselt, arvutileksikoloogia, leksikaalne semantika, nimisõnade polüseemia.

margit@eki.ee

MERIKE MÄGEDI (1975) on lõpetanud Tallinna Pedagoogikaülikooli eesti keele ja kirjanduse erialal. 2003. a kaitses Tallinna Pedagoogikaülikoolis magistritöö sõnavara võõrusest ja võõrsõnadest eesti tekstis. Töötab Eesti Keele Instituudi leksikoloogiasektoris teadurina. Uurimisvaldkonnad: sõnavara struktuur, semantilised rollid.

merike@eki.ee

ÜLLE VIKS (1948) on lõpetanud Tartu Ülikooli eesti keele ja kirjanduse erialal. 1978. a kaitses Tartu Ülikoolis kandidaadiväitekirja eesti verbide morfoloogilisest klassifikatsioonist ja 1994. a Tartu Ülikoolis doktoriväitekirja eesti keele klassifikatoorsest morfoloogiast. Töötab Eesti Keele Instituudi grammatikasektoris vanemteadurina. Uurimisvaldkonnad: avatud morfoloogiamudel ja selle realisatsioon arvutis (sõnavormide reeglipõhine süntees ja analüüs), grammatilise informatsiooni esitamine sõnastikes.

ylle@eki.ee