

# EESTI-X-KEELE SÕNASTIK JA GRAMMATIKA

Ülle Viks

**Ülevaade.** Artikkel käsitleb probleeme, mis on seotud grammatilise info esitamisega kakskeelses sõnaraamatus. Lähemalt tutvustatakse Eesti-X-keele sõnastikku (EXS), tema kohta sõnastike tüpoloogias ja sellest tulenevaid nõudeid. EXS on universaalne leksikaal-grammatiline andmebaas, millest on võimalik genereerida suvalise kakskeelse sõnaraamatu elektrooniline põhi, kus eesti lähtekeel on standardkujul olemas. Grammatiline info esitatakse EXS-is struktureeritult, hõlmates erinevaid keelekirjelduse tasandeid fonoloogiast semantikani; keeleandmete valik ja esitusviis tuginevad sõnastiku ja grammatika vahekorra analüüsile. Vormistuselt peab EXS kui metatasandi andmebaas olema paindlik ning sõltumatu tulemsõnastiku sihtkeelest, tüübist ja kujundusest. Lõpuks tutvustatakse lühidalt töövahendeid, mille abil EXS-i luua ja rakendada: sõnastike haldussüsteemi EELex (leksikograafi töökeskkond) ja grammatilist kirjegeraatorit.\*

**Võtmesõnad:** kakskeelne leksikograafia, grammatiline informatsioon, keeletöötlus, XML, sõnastike haldussüsteem, eesti keel

## 1. Eesti-X-keele sõnastik

“Eesti-X-keele sõnastik” on keeletehnoloogia projekti “Leksikograafi töökeskkond” allprojekt (2006–2010). Eesti Keele Instituudis (EKI) algatatud Eesti-X-keele sõnastik (EXS) on kavandatud kui universaalne leksikaal-grammatiline andmebaas, mis sobiks aluseks eelkõige uute eesti lähtekeele kakskeelsete sõnaraamatute koostamisel (inimese jaoks), aga ühtlasi peaks ta olema kasutatav mitmesugustes keeletehnoloogia rakendustes (arvuti jaoks).

EXS pole valmis sõnastik, vaid pigem metatasandi nähtus. See on paindlik polüfunktsionaalne andmebaas, millest on võimalik genereerida põhimõtteliselt ükskõik millise kakskeelse (eesti lähtekeele) sõnaraamatu elektrooniline põhi, kus eesti pool on standardkujul juba olemas. Selles on ära tehtud esialgne sõnavalik,

\* Tööd on toetanud riikliku programmi “Eesti keele keeletehnoloogiline tugi” projekt “Leksikograafi töökeskkond” ja Eesti Teadusfondi grant nr 5969.

esitatud grammatilised andmed, näidatud sõna tähendusliigendus koos vajalike seletustega, lisatud sõna kasutusinfo, olulised sõnaühendid jne.

EXS-i olemuse ja koostamisaluste mõistmiseks püüan määratleda, milline on tema koht sõnastike väga kirevas maailmas.

## 1.1. Sõnastikutüpoloogiast

Sõnaraamat (laiemas tähenduses sõnastik) on teatavasti korrastatud sõnavarakogu, mis koosneb kindla struktuuriga sõnaartiklitest. Sõnaartiklit alustab märksõna, millele järgnevad teatud skeemi kohaselt mitmesugused andmed märksõna kohta, nt hääldus, sõnaliik, reksioon, stiilimärgend, seletus, sünonüümid, kasutusnäited, vasted teises keeles jne.

Sõnaartikli ülesehitus, andmete hulk ja esitusviis sõltuvad sõnastiku tüübist. Sõnaraamatuid võib liigitada paljude eri tunnuste järgi, kõige üldisemad neist on:

- **keelte hulk**, mis jagab sõnastikud ükskeelseteks ja kaks- või mitmekeelseteks;
- **sihtgrupp**, mille alusel eristuvad üldsõnastikud (laiale kasutajaskonnale) ja mitmesugused erisõnastikud (kitsale ringile) – igal kasutajagrupil on oma vajadused ja nõuded sõnaraamatule, millest sõltub eelkõige märksõnavalik, aga ka info valik, maht ja esitusviis sõnaartiklis;
- **otstarve**, mis liigitab sõnastikud funktsiooni järgi: nad kirjeldavad keelenähtusi, annavad õigekeelsuse soovitusi (normatiivne), abistavad keeleõppijat (õppesõnastik) või tõlkijat (tõlkesõnastik), toetavad teksti mõistmist (passiivne sõnastik) või moodustamist (aktiivne sõnastik) jne – ka otstarve mõjutab märksõnavalikut ning sõnaartikli koostist ja sisu;
- **maht**, mis seostub nii märksõnade arvuga kui ka artiklis esitatava info hulgaga – tavaliselt eristatakse kolme mahuklassi märksõnade arvu järgi: väike (10 000–20 000), keskmine (30 000–50 000), suur (60 000 ja rohkem) –, kuid mahu hindamisel tuleb arvesse ka sõnaartikli suurus.

Tunnused võivad kombineeruda, andes mitmeid eri- ja alaliike. Kakskeelsete (v mitmekeelsete) sõnastike ülekaalukas põhitüüp on **üldsõnastik**. See on universaalne, mõeldud eelkõige tõlkimiseks ühest keelest teise. Peamised liigitusalused on:

- sihtgrupp, nt tõlkija, keeleõppija, turist, spetsialist, teadlane, lingvist, keeletehnoloog jne;
- otstarve: kirjeldav e tõlkesõnastik, õppesõnastik, vestmik, formaalne leksikon jne; otstarbega seostub ka jaotus passiivseks v aktiivseks sõnastikuks (vt lähemalt 1.2.);
- maht: suurem osa tõlke- ja õppesõnaraamatuid on keskmise mahuga, suure näiteks võiks olla soome-estis suursõnaraamat (SUVI: 90 000 märksõna, 2 köidet), tüüpilised väikesed on nt turistsõnastikud.

Kakskeelseid **erisõnastikke** leidub suhteliselt vähe, erinevalt ükskeelsetest, mille seas on mitut liiki erisõnastikke.<sup>1</sup> Kakskeelsete erisõnastike liigitusalustest võiks välja tuua:

<sup>1</sup> Ükskeelsete erisõnastike liigitusalused on spetsiifilised, olulisemad neist on:

- märksõnatüüp, nt lühendid, võõrsõnad, (fraseoloogilised) väljendid, kohanimed jne;
- infotüüp e mille kohta sõnaartiklis teavet antakse, nt ortograafia, hääldus, morfoloogia, tuletus, valents/reksioon, sünonüümid/antonüümid, riimid, etümoloogia jne;
- tekstitüüp e allkeel, nt släng, erialane tekst (terminid), murded, vana kirjakeel, piibel, autoritekstid jne;
- makrostruktuur e sõnastiku organiseerimise viis, nt pöörsõnastik, konkordants, mõisteline sõnastik, tuletuspepede sõnastik jne.

- märksõnatüüp, nt fraseoloogilised väljendid, käibefraasid;
- tekstitüüp, nt släng, erialane tekst (terminoloogiasõnastikud on sagedamini mitmekeelsed).

## 1.2. Aktiivne ja passiivne sõnaraamat

Kakskeelsete sõnastike üks olemuslikumaid liigitusi on jaotus passiivseks ja aktiivseks, mis seostub kasutaja teadvuses toimuvate erisuunaliste keeletöötlusprotsessidega: analüüsiga ja sünteesiga.

**Passiivse** sõnastiku eesmärgiks on toetada kasutajat võõrkeelse teksti mõistmisel e analüüsil. Selleks et tundmatu sõna sõnastikust üles leida (ja selle tähendust vaadata), peab kasutaja võõrkeelses tekstis tegema kõigepealt morfoloogilise analüüsi, st ta peab suutma kindlaks teha, milline on lauses oleva muutevormi algvorm, sest sõnastikes on märksõnad algvormis. Passiivse sõnastiku puhul on sõnastiku lähtekeel (märksõnade keel L1) kasutaja jaoks võõrkeel ning sõnastiku sihtkeel (vastete keel L2) on kasutaja emakeel.

**Aktiivse** sõnastiku eesmärgiks on toetada kasutajat võõrkeelse teksti loomisel e sünteesil. Selleks, et oma lauset grammatiliselt õigesti kokku panna, peab kasutaja suutma moodustada (sünteesida) sõnastikus esitatud algvormist kõik lauses vajalikud muutevormid. Aktiivse sõnastiku puhul on sõnastiku lähtekeel (märksõnade keel L1) kasutaja emakeel ning sõnastiku sihtkeel (vastete keel L2) on võõrkeel.

Neist eesmärkidest tulenevad erinevused sõnaraamatute ülesehituses e makro- ja mikrostruktuuris, eelkõige puudutab see andmete valikut ja esitusviisi sõnaartiklis. Makro- ja mikrostruktuuri erinevusi aktiivses ja passiivses sõnastikus on käsitlenud nt Soome keeleteadlane Marja Järventausta (2004) seoses õppesõnastikega (soome ja saksa keele näitel). Võtsin tema poolt esile toodud põhipunktid kokku tabelisse 1.

**Tabel 1.** Sõnastikuinfo passiivses ja aktiivses sõnastikus

	<b>passiivne sõnastik (analüüs)</b>	<b>aktiivne sõnastik (süntees)</b>
<b>märksõnavalik</b>	palju sõnu, sh haruldased	keskne sõnavara
<b>andmete valik</b>	märksõna tähendus	märksõna kasutusinfo
<b>metakeel</b>	sihtkeel (L2)	lähtekeel (L1)
<b>grammatiline info</b>	märksõna juures	vaste juures
<b>vormid märksõnana</b>	jah: ebareeglipärased	ei

Järgnevad kommentaarid on tabeli 1 vaba tõlgendus.

1. Sõnavalik passiivses sõnastikus peab olema suur, et see kataks võimalikult suure osa loetavates tekstides leiduvast sõnavarast. Aktiivne sõnastik seevastu võib piirduda ainult olulise sõnavaraga, mis võimaldab kasutajal ennast väljendada. Mida paremini kasutaja keelt valdab, seda suuremat sõnavara (ja sõnaraamatut) ta oma tekstide loomiseks vajab.
2. Andmete valik passiivse sõnastiku sõnaartiklis peab arvestama eelkõige märksõna tähendusi, et kasutaja saaks sõnastikust teada, mida tekstis esinev sõna tähendab. Aktiivne sõnastik peaks keskenduma märksõna kasutusele

- erinevates kontekstides (nii grammatilistes kui ka semantilistes), et kasutaja suudaks valida sõnastikus pakutud variantidest tema lausesse sobivaima.
3. Metakeel (seletuste ja märgendite keel) peaks olema kasutaja emakeel, et ta seletustest ja kommentaaridest saaks paremini kätte kogu sisu. Passiivses sõnastikus on emakeel sihtkeele (L2) rollis, aktiivses aga lähtekeele (L1) rollis.
  4. Grammatiline info peaks passiivses sõnastikus olema esitatud märksõna juures, et võõras keeles oleva sõna vormidest ja grammatilisest käitumisest tekiks ülevaade. Aktiivses sõnastikus vajavad grammatilist infot aga vasted, et oleks võimalik moodustada vajalikud vormid ja kasutada neid lauses vastavalt sihtkeele (L2) reeglitele.
  5. Sõna (ebareeglipärased) muutevormid märksõna positsioonis (koos viitega algvormile) on passiivses sõnastikus vajalikud selleks, et kasutaja saaks võimaluse siduda kõik tekstis esinevad ühe sõna muutevormid vastava märksõnaga, olgu nad omavahel nii erinevad kui tahes. Eesti keeles, mida iseloomustab suur tüvemuutuste hulk (vrd *sugu : soo, uba : oa, saag : sae, raba : rappa, siduma : seob, ütleva : öelda, minema : läinud* jne), on see väga oluline. Aktiivses sõnastikus pole vormide esitamine märksõnana vajalik, sest märksõna on kasutaja emakeeles.

M. Järventausta (2004: 27) sõnul on uurijad üsna üksmeelsed selles, et ideaaljuhul peaks iga keelepaari jaoks olema neli sõnastikku: kummagi suuna jaoks nii aktiivne kui ka passiivne, sõltuvalt kasutaja emakeelest.<sup>2</sup> Praktikas on sellist luksust muidugi raske leida, enamasti üritatakse kummagi suuna sõnastikku kokku koondada nii passiivse kui aktiivse sõnastiku omadused. Sedasama üritame ka meie Eesti-X-keele sõnastikus, aga me püüame seda teha teadlikult ja kindla eesmärgiga.

### 1.3. Tüpoloogilised nõuded EXS-ile

Sõnastikutüpoloogia seisukohalt on EXS-i andmebaas **kakskeelne üldsõnastik**, st universaalne sõnastik, mis on mõeldud eelkõige tõlkimise abivahendiks. Et aga EXS on metatasandi andmebaas, siis peab ta mingil moel hõlmama ka nende sõnastikutüüpide omadusi, mille aluseks ta saab. Selleks peab ta vastama järgmistele nõuetele.

1. EXS-i andmebaasi **maht** on maksimaalne, nii märksõnade arvu kui ka info hulga osas, kuid sellest peab saama genereerida väiksemaid tulemsõnastikke.
2. EXS on otseselt suunatud kahele **sihtgrupile**: a) leksikograafidele erinevate kakskeelsete sõnaraamatute koostamiseks, b) keeletehnoloogidele mitmesuguste sõnastikurakenduste loomiseks (sh masintõlge). Kaudselt – leksikograafide vahendusel – võivad EXS-i sihtgruppideks saada kõik tavapärased kakskeelse sõnaraamatu kasutajad (e lõppkasutajad), nt tõlkijad, keeleõppijad, turistid jt.
3. EXS-i **otstarbe** tunnussõnad on: kirjeldav, passiivne (ja aktiivne), formaalne.

<sup>2</sup> Sama rõhutab eesti ja inglise keelepaarist rääkides ka Enn Veldi (2000: 178).

Keelekirjeldus EXS-is peab olema põhjalik ja hästi struktureeritud, et sealt saaks teha erinevaid väljavõtteid, arvestades konkreetse tulemsõnastiku vajadusi. EXS on eelkõige passiivne sõnastik, mis siiski ei välista selle hilisemat kasutamist aktiivse sõnastikuna. Vastavalt otseste sihtgruppide vajadustele pööratakse EXS-is formalismidele suuremat tähelepanu, kui see tavasõnastikes kombeks on. Sõnastiku formaalsus ei tähenda sisu kaotamist, vaid sisu esitamist kujul, mis on kättesaadav eksplitsiitselt, tuginemata kasutaja oletatavatele taustateadmistele.

## 2. Grammatilise info esitus EXS-is

### 2.1. Keelekirjeldustasandid

EXS-i universaalset iseloomu ja võõrkeelset kasutajat arvestades on seal kavas esitada grammatilist infot võimalikult paljudelt keelekirjeldustasanditelt: 1) morfofonoloogiast välde ja ebareeglipärane rõhk, 2) morfoloogiast põhjalik vormimoodustuse kirjeldus ja jõudumööda andmed ka vormikasutuse kohta, 3) tuletusest kindlasti liitsõna-liigendus ja komparatsioon, võib-olla ka tuletustüübid, 4) (morfo)süntaksist sõnaliik ja arv, reksioon ja püsiühendid, 5) semantikast võimaluse korral semantilised rollid, edaspidi ehk ka semantilised tüübid.

Kõige põhjalikumalt on välja töötatud morfoloogia esitussüsteem, mille aluseks on avatud morfoloogiamudel (Viks 2000a) ja sellele vastav morfoloogiline klassifikatsioon (VVS ja Viks 2001). See erineb mõnevõrra ÕS2006 tüübistikust (vt 2.3.), kuid sobib kakskeelse sõnastiku jaoks paremini, sest tema aluseks on formaalne mudel. VVS-i tüübistik on testitud ja leidnud praktilise rakenduse eesti keele morfoloogiatarkvaras (nt EKI automaatne morfoloogiasüsteem<sup>3</sup>, Filosoofi morfoloogiline analüsaator (Kaalep 1996) ja sellel tuginev eesti keele speller) ning mitmes kakskeelses sõnaraamatus, nt EVS, NEEN, Viks 2003 jt. Morfoloogiasüsteemiga koos on lahendatud ka morfofonoloogia probleemid, sest need on omavahel vastastikuselt sõltuvused. Süntaktilise info kohta on ilmunud ülevaade (Langemets jt 2005). Palju konkreetseid eeltöid on tehtud tuletusinfo osas, mis suures osas toetub töös olevale sõnapere materjalidele<sup>4</sup> (Vare 2002).

Grammatilise info esitamise üksikasjadeni ma siin ei lähe, vaid peatun sõnastiku ja grammatika vahekorral, millest otseselt sõltub sõnastiku keeleline funktsionaalsus.

### 2.2. Sõnastiku ja grammatika vahekord

Milleks üldse ja kui palju vajab sõnaraamat grammatilist infot, kui selle jaoks on olemas spetsiaalsed grammatikaraamatud? Inimene haarab sõnastiku järele tavaliselt siis, kui ta tahab teada sõna tähendust, õigekirja või vastet teises keeles. Kuid sellest ei piisa aktiivse keelekasutuse puhul – vaja on ka teada, kuidas konkreetset sõna ühel või teisel puhul pruukida: kuidas teda hääldada, kuidas käänata-pöörata, millisesse positsiooni teda lauses saab paigutada, milliste teiste sõnadega või muutevormidega sobib teda kokku panna jne, jne. Seda tarkust on raske grammatikast üles leida, kuigi ta on seal olemas.

<sup>3</sup> <http://www.eki.ee/tarkvara/> (28.09.2007).

<sup>4</sup> Kasutatav EKI sisevõrgus: <http://www.eki.ee/dict/sp/> (28.09.2007).

Sõnastik ja grammatika sisaldavad suures osas sama grammatilist infot, kuid nad on organiseeritud erinevalt ja ideaalis peaksid teineteist vastastikku täiendama. Grammatika esitab keelenähtusi üldistatult, tema struktuuri põhiüksus on grammatiline seaduspärasus (keelereeglite kujul), mida illustreeritakse üksikute näidetega (hrl sõnadega), milles see või teine grammatiline omadus avaldub. Ükski grammatika ei saa loetleda kõiki iga reegli alla kuuluvaid sõnu. Sõnastik seob needsamad keelenähtused konkreetsete sõnadega: sõnastiku struktuuri põhiüksus on sõna, mille juures sõnaartiklis esitatakse temale iseloomulikud grammatilised omadused.

Kuid ka sõnastikus ei saa grammatiline info olla ammendav, nt puudub seal ruum näitamaks sõna kõiki muutevorme või sõna võimalust esineda lauses teatud lauseliikmena jne. Seetõttu esitatakse grammatiline sõnastikuinfo tavaliselt üldistatud kujul, nt klassitähisena või reegliviidana, mis juhib kasutaja reeglite endi juurde. Üldisema kategooria näide võiks olla sõnaliik: sõnad kirjes varustatakse sõnaliigimärgenditega (nt *ehtne Adj*), millele vastavad reeglid (põhimõtted sõnade üldise morfoloogilise ja süntaktilise ja semantilise käitumise kohta) tuleb leida grammatikaraamatust. Morfoloogiliste reegliviitadena kasutatakse eesti leksikograafilises traditsioonis sõna juures tüübinumbreid (nt *krae S 26*), need juhivad sõnaraamatu lisaosas esitatud tüübikirjelduste (e reeglite) juurde, mille alusel saab vajalikud vormid moodustada (vt Viks 2001). Samal kombel saab esitada sõnastikus ka regulaarse sõnamoodustuse reeglid, millele viitavad sõnaartiklis tuletustüüpide tähised.

Kui mõni sõna käitub teisiti, kui üldine grammatikareegel ette näeb, siis sellekohane info täiendab sõnaartiklis esitatud klassitähist. Nt sõna *eht* puhul tuleb adjektiivimärgendile lisada märkus, et see sõna ei käändu (erinevalt tüüpilistest omadussõnadest). Erandliku morfoloogiaga sõna kirjes tuuakse lisaks tüübinumbri-le ära konkreetsed erandlikud muutevormid (nt *pea S 26; pl p päid, adt pähe*).

Grammatilise info valik ja esitusviis sõnaartiklis sõltuvad suurel määral sõnastiku tüübist: sellest, milline on sõnastiku sihtgrupp ja otstarve, kas sõnastik on suur või väike, üks- või kakskeelne. Ükskeelne sõnastik lähtub emakeelse kasutaja vajadustest, kes tavaliselt otsib sõnastikust lisa oma olemasolevatele teadmistele. Muidugi võivad ükskeelset sõnastikku kasutada ka muukeelsed inimesed, aga nende jaoks on ta mingis mõttes liiane (info on kohati liiga põhjalik ja detailne), aga mingis mõttes ka puudulik, sest kõike pole otse välja kirjutatud. Ükskeelses sõnastikku on vaikimisi sisse kodeeritud eeldus, et kasutaja valdab keelt ja seega ei pea seal kõike eksplitsiitselt näitama – seda nii sisulise poole esitamisel kui ka sõnaartikli vormistamisel (nt vormide või näidete lühendamise puhul). Seevastu kakskeelse sõnastiku adressaadilt ei saa head teise keele oskust nõuda (selle saavutamiseks aitab sõnastik alles kaasa). Selles suhtes on võõrkeelne kasutaja sarnane arvutiga – mõlema jaoks kehtib eksplitsiitsuse ja formaalsuse nõue. Oluline on, et sõnastiku kasutaja saaks talle vajaliku info ja selle rakendamise juhtnöörid kätte ammendavalt ja ühemõttelise töökindlusega.

### 2.3. Ükskeelne kõrvalepõige

Toon kõrvalepõikena ühe näite ükskeelsest ÕS2006-st, kus morfoloogiline info on esitatud küll põhjalikult, kuid üksnes sellist kasutajat silmas pidades, kes eesti keelt valdab ja otsib sõnastikust tuge peamiselt kahtluse korral. Võõrkeelse kasutaja jaoks pole vormide moodustusalgoritm piisavalt selge ning on mõnikord lausa vasturääkiv.

Näiteks noomeni üks keerulisemaid vorme – pluurali partitiiv – on sõnaartiklis harva esitatud ja kasutaja peab selle ise tüübikirjelduse põhjal moodustama. Tüübid 20, 21 ja 22 erinevadki omavahel ainult mitmusevokaali poolest. Tüübikirjeldused sõnaraamatu lisas (ÕS2006: 20) on sellised (siin lühendatult):

20. mitmuse osastavas *-e*  
paks, paksu, .paksu, .paksu ja paksusse, .paksude, .pakse ja .paksusid,  
.paksudesse ja paksesse
21. mitmuse osastavas *-i*  
külma, külma, .külma, .külma ja külmasse, .külmade, .külmi ja .külma-  
sid, .külmadesse ja külmisses  
täht, tähe, .tähte, .tähte ja tähesse, .tähtede, .tähti ja .tähtesid, .tähte-  
desse ja tähisses
22. mitmuse osastavas *-u*  
õrn, õrna, .õrna, .õrna ja õrnasse, .õrnade, .õrnu ja .õrnasid, .õrnadesse  
ja õrnusse  
leib, leiva, .leiba, .leiba ja leivasse, .leibade, .leibu ja .leibasid, .leibades-  
se ja leivusse

Neisse tüüpidesse kuuluvate sõnade kirjetes pluurali partitiivi vormi pole, õige vormi saamiseks tuleb uurida tüübikirjeldust ja toimida analoogia põhjal tüüpsõna eeskujul. See on kasutajale ebamugav, aga siiski võimalik. Nt

haug <20: haugi, .haugi>	→ .hauge ja .haugisid	(nagu paks)
halg <20: halu, .halgu>	→ .halge ja .halgusid	(nagu paks)
õng <21: õnge, .õnge>	→ .õngi ja .õngesid	(nagu täht)
pung <21: punga, .punga>	→ .pungi ja .pungasid	(nagu külm)
lõug <22: lõua, .lõuga>	→ .lõugu ja .lõugasid	(nagu leib)

ÕS2006-s on aga ka tüüpe, mis sisaldavad erineva mitmusevokaaliga sõnu, kusjuures tüübikirjeldus näitab ainult üht vokaali, nt 24 (ainus tüüpsõna *pood*). Tüübikirjeldus:

24. pood', poe, .poodi, .poodi ja .poesse, .poodide, .pood'e ja .poodisid, .poo-  
didesse

Sel juhul on pluurali partitiiv kõigi sõnade kirjetes antud ja seda EI TOHI moodus-  
tada tüübikirjelduse järgi.

laug 1. <24: lau, .laugu; .lauge ja .laugusid>	(vrd halg <20>)
saag <24: sae, .saagi; .saage ja .saagisid>	(vrd haug <20>)
roog <24: roa, .rooga; .roogi ja .roogasid>	(vrd pung <21>)
lõög <24: lõa, .lõõga; .lõõgu ja .lõõgasid>	(vrd lõug <22>)

Eelnevast järeldub (kirjutamata) algoritm kasutaja jaoks:

- 1) Kui pluurali partitiiv on sõnastikukirjes olemas, võta sealt.
- 2) Kui kirjes pluurali partitiiv puudub, moodusta see tüübikirjelduses oleva tüüpsõna eeskujul.

Segadus tekib neis tüüpides, kus tüübikirjelduses on pluurali partitiivi vorm antud, aga kõigist selle tüübinumbriga sõnadest teda ei moodustata või ei tarvitata (kusujuures eri sõnadel on partitiiv ka erineva vokaaliga), nt tüüp 16: tüüpsõnad *rida* ja *sõda*. Tüübikirjeldus:

16. rida, rea, rida, .ritta *ja* .reasse, ridade, ridu *ja* ridasid, ridadesse  
sõda, sõja, sõda, .sõtta *ja* sõjasse, sõdade, sõdu *ja* sõdasid, sõdadesse

Algoritmi järgides juhtub nii, et sõna *tuba* saab oma pluurali partitiivi vormi *tube* kirjest kätte:

tuba <16: toa, tuba, .tuppa *ja* .toasse; tube *ja* tubasid>

Aga sõnade *koda*, *õde*, *tegu* puhul satub kasutaja eksiteele, sest sõnaartiklis pluurali partitiivi vormi pole ja järelikult tuleks see teha tüübikirjelduse järgi, aga tüübikirjeldus näitab ainult *u*-lõpulist vormi koos *sid*-lõpulisega. Nii võib keeleoskamatu kasutaja moodustada vigased muutevormid:

koda <16: koja, koda, .kotta <i>ja</i> kojasse>	→ *kodu <i>ja</i> kodasid	( <i>vrđ</i> sõda)
õde <16: õe, õde>	→ *õdu <i>ja</i> õdesid	
tegu <16: teo, tegu, .teosse <i>ja</i> .tekku>	→ *tegu <i>ja</i> tegusid	

Kui ükskeelne sõnastik ehk võib endale lubada mõningaid lünki keelekirjelduses, lootes kasutaja keeleoskusele, siis kakskeelses sõnastikus peab grammatika esitus olema täpsem ja algoritmilisem. Seetõttu pole EXS-i grammatika esitamisel võetud eeskujuks kirjakeele normi alust ÕS2006.

### 3. EXS kui universaalne andmebaas

EXS-ile seatud põhinõudest (alus erinevatele kakskeelsetele sõnaraamatutele) lähtudes postuleeritakse kolm sõltumatusenõuet: EXS peab olema sõltumatu tulemsõnastiku sihtkeelest, sõnastiku tüübist (mahust, sihtgrupist, otstarbest) ja kujundusest.

#### 3.1. Sõltumatus sihtkeelest

EXS-i põhinõue on sõltumatus koostatava kakskeelse sõnaraamatu sihtkeelest – ta peab sobima aluseks paljude eesti-X keelepaaride sõnaraamatutele. See tähendab, et ta peab olema eelkõige eesti keele andmebaas, mis kirjeldab eesti sõnade leksikaal-grammatilisi omadusi sel kombel, et nendega saaks kõrvuti panna mistahes teise keele sõnade leksikaal-grammatilised omadused.

Eelkõige puudutab see sõnade tähenduse avamist, ja eriti **tähendusliigendust**, mis peaks olema hästi detailne ja lähtuma eesti keelest. Kui eesti sõna mitu



eri tähendust saavad sihtkeeles sama vaste, siis need võetakse kokku alles tulemsõnastiku (X-keele sõnaraamatu) koostamisel, sest mõne teise keele vasted võivad eesti sõna eri tähenduste vahel teisiti jaotuda.

Teine oluline valdkond on **grammatiline info**, mis peab olema esitatud formaalselt ja algoritmiliselt, st arusaadaval kujul inimesele, kes eesti keelt piisavalt ei oska ja kelle emakeel ei paku tuge eesti grammatilise süsteemi mõistmisel. Universaalsete keelenähtuste esitamisel on mõttekas järgida üldist leksikograafist traditsiooni (nt sõnaliigimärgendid sõnastikukirjes märksõna juures ja tähendusnumbrite juures). Suuremat tähelepanu nõuavad grammatilised nähtused, mis on eesti keelele spetsiifilised, nagu välde, morfoloogia, rektsioon jne. Keelespetsiifilist infot saab esitada kas a) tervikuna sõnaartiklis, nt välde, rektsioon, erandlikud keelendid, või b) täiendades kirje infot vastava grammatikaosaga, nt tüübnumber ja mõned vormid kirjes, kogu tüübikirjeldus ja muutevormide moodustusalgoritm sõnastiku lisas. Kui mõni väiksem tulemsõnastik ei vaja nii põhjalikku grammatilist infot, siis seal on võimalik osa andmeelemente esitamata jätta.

Tulemsõnastiku sihtkeelest sõltub otseselt **metakeel** (st seletuste ja märgendite keel), mis peaks olema lõppkasutaja emakeel. Kuigi EXS-i enda metakeel on eesti keel, peab ta arvestama teistega: tulemsõnastiku jaoks peab olema loodud võimalus vahetada teatud elementide keelt või lisada teises keeles olev üksus.

### 3.2. Sõltumatus sõnastiku tüübist

EXS on sõltumatu tema baasil koostatava kakskeelse tulemsõnastiku mahust ja otstarbest ning lõppkasutaja sihtgrupist. See tähendab, et ühest andmebaasist peab saama automaatselt genereerida erineva märksõnade arvuga ja erineva andmevalikuga standardseid sõnastikupõhju, mida konkreetse tulemsõnastiku koostaja saab oma töö käigus modifitseerida.

**3.2.1. Märksõnade arvu** reguleerimiseks on esialgu plaanis ette valmistada kolm erimahulist standardit. Seda saaks teha näiteks nii, et andmebaas sisaldab maksimaalse hulga märksõnu (ja kirjeid), kusjuures iga märksõna saab külge tunnuse, millisesse mahukategooriasse ta kuulub: kas suurde, keskmissse või väiksesse sõnastikku. Otsustava tähtsusega parameetrid mahukategooria määramisel on kuuluvus põhisõnavarasse, sagedus tekstides ja mõisteline kuuluvus (nt kui sageduse põhjal on ühte mahukategooriasse valitud mõni nädalapäeva nimetus, siis peab vaatama, et kaasa tuleks kõik nädalapäevad).

Selleks, et sõnu saaks paindlikult erimahulistest sõnastikes esitada ja et vajalikku infot kaotsi ei läheks, on oluline, et iga leksikaalne üksus oleks andmebaasis eraldi täiskirjena oma märksõnaga, mitte paigutatud mõne teise märksõna alla. Peamised kriitilised üksused on:

- **homograafilised** sõnakujud, nt *hall* (3 homonüümi);
- **liitüksused**: tuletised, liitsõnad, sõnaühendid (mille tähendus ei tulene komponentide tähendustest);
- **sõna muutevormid** leksikaalse üksusena (oma tähendusega), nt *õues*, *tasuta*, *elagu*, *avatud*;
- **üksuste variandid**: ortograafilised, häälduslikud, morfoloogilised jne.

Oma lõpliku koha sõnastiku struktuuris saavad sellised üksused tulemsõnastikus, ja neid lahendusi võib olla mitmeid. Vastavalt mahukategooria tähisele saab EXS-i märksõna teatud suurusega sõnastikupõhjas kas märksõnaks või artikli mingi allstruktuuri põhiüksuseks (allmärksõnaks) või mingi (näite)loendi osaks või jääb hoopis välja. Tulemsõnastiku toimetajal on võimalus sõnastiku ülesehitust ka ise kujundada.

**3.2.2.** Mahuvalikuga kuulub kokku **andmevalik**: suurem sõnastik esitab hari-likult ka põhjalikumalt infot märksõna kohta. Põhiliselt sõltub andmevalik aga tulemsõnastiku sihtgrupist ja otstarbest. Sobiva valiku tegemiseks on mõeldav kasutada filtrite süsteemi. Näiteks nii, et andmebaas sisaldab märksõna kohta maksimaalset infot, kuid iga konkreetse tulemsõnastiku jaoks valmistatakse ette filter, mis osa infot välja praagib. Filtriga määratakse kirje koostis ja ülesehitus: milline on andmeelementide valik ja asukoht sõnaartikli struktuuriskeemis. Välja võivad jääda nt hääldus, etümoloogia, rektsioon; morfoloogia võib olla antud erineva põhjalikkusega jne.

Enamasti on grammatilised andmed üksteisest küllaltki sõltumatud, ja ka nende valik on suhteliselt vaba. Eriolukorras on morfoloogiline info, mis on tihedalt seotud muutevormide analüüsi ja sünteesi algoritmidega – muidu pole kindlustatud ei märksõna leidmine ega sõna kasutamine lauses. Välja on töötatud kolm morfoloogiakirjete standardit, mis sisaldavad mitmeid väiksemaid sisulisi elemente, mis peaksid filtrisse kuuluma ühe tervikliku kompleksina. Neid on eri sõnastikes ka juba kasutatud:

- 1) vormipõhine kirje (suur) – sõna kõik põhivormid ja muuttüüp (nt EVS):  
tuba <tuba t`oa tuba t`uppa, tuba[de tuba[sid & tub/e S 18]>  
lastetuba <+tuba t`oa tuba t`uppa, tuba[de tuba[sid & tub/e S 18]>  
andma <`and[ma `and[a anna[b `an[tud, `and[is `and[ke V 34\*]>
- 2) tüvepõhine kirje (keskmine) – kõik erinevad tüvevariandid ja muuttüüp (nt NEEN eesti pool):  
tuba [toa, *adt.* tuppa 18e]  
laste-tuba [toa, *adt.* tuppa 18e]  
and**ma** [annab, *ips.* antud 34] ..
- 3) tüvemuutuste-põhine kirje (väike) – olulisi tüvemuutusi näitavad vormid ilma muuttüübita (nt ENNET eesti pool):  
tuba <toa, *adt* tuppa>  
laste-tuba<sup>˘</sup>  
an|**dma** <-nab -tud>

Kaks esimest – vormi- ja tüvepõhine kirje – annavad mõlemad piisava info selleks, et oleks võimalik nii morfoloogiline analüüs kui ka süntees, kuid tüvepõhine kirje eeldab lõppkasutajalt suuremat valmidust teha tööd tüübikirjeldustega. Vormipõhine kirje on kasutajale mugavam ja sobib paremini nõrgema keeleoskuse puhul, sest kirjes on kõik sõnavormide moodustamiseks vajalikud põhivormid välja kirjutatud. Kolmas – tüvemuutuste-põhine kirje – näitab ainult põhilisi tüvevariante ja annab enamasti võimaluse märksõna leidmiseks (lemmatiseerimiseks), aga täielik morfoloogiline analüüs ja vormide süntees jäävad selle info põhjal kättesaamatuks.

**3.2.3.** Sõnastiku otstarbega ja kasutaja emakeelega on seotud tema **aktiivsuse ja passiivsuse** määratlus. Milline peaks olema EXS sellest aspektist? Kuna andmebaasi suund on E–X (eesti on lähtekeel), siis järelikult: X-keelsele kasutajale peaks see olema passiivne ja võimaldama E-keelsest tekstist arusaamist (E-analüüsi); E-keelsele kasutajale peaks see olema aktiivne ja võimaldama X-keelse teksti loomist (X-sünteesi). Et aga tegemist on pooliku sõnastikuga E–(X), kus eesti keele andmed on põhjalikult esitatud, samas kui X-keele asemel on ainult tühjad struktuuripositsioonid, siis langeb teine variant ära: X-keele sünteesi toe peab tagama iga konkreetse keelepaari jaoks koostatav sõnaraamat ise. EXS on järelikult orienteeritud X-keelsele kasutajale ja on passiivne sõnastik (võimaldab E-analüüsi).

Kaugemas perspektiivis hoiame sihikul siiski ka sõnastiku pööramise võimalust, nii et sama andmebaasi saaks kasutada X–E sõnastike loomisel. (X)–E puhul on eesti sihtkeeleks ning järelikult: E-keelsele lõppkasutajale peaks see olema passiivne ja võimaldama X-keelsest tekstist arusaamist (X-analüüsi); X-keelsele kasutajale peaks see olema aktiivne ja võimaldama E-keelse teksti loomist (E-sünteesi). Meie poolikus sõnastikus jääb sellest paarist alles teine variant, mis on samuti orienteeritud X-keelsele kasutajale, kuid on aktiivne sõnastik (võimaldab E-sünteesi).

Suurim muutus sõnastiku pööramisel on see, et endine märksõna liigub vaste positsiooni ja koos sellega ka märksõna juurde kuulunud info, kusjuures algselt passiivse sõnastiku lähtekeele nõuete kohaselt esitatud andmed peavad hakkama sobima aktiivse sõnastiku sihtkeele nõuetega (vrd tabel 1). Aktiivse sõnastiku jaoks vajalik grammatiline (sünteesi)info on EXS-is olemas, samuti ka sõna kasutamisega seotud andmed, nt valdkonna- ja stiilmärgendid, rektsioon jms, tarvis on kogu see andmestik transformeerida vastainfole sobivasse vormingusse. Sõna tähendusega seotud teave vajab põhjalikumalt revideerimist – see on vaste juures esitamiseks liiane.

Sõnastiku pööramine saab võimalikuks tänu EXS-i andmebaasi universaalsusele: kuna keeleandmete valik eesti märksõna juures on maksimaalne ja andmed ise põhjalikult struktureeritud ning struktuuritähistega varustatud, siis saab nende positsiooni ja staatust sõnastiku struktuuris vastavalt vajadusele muuta.

Kokkuvõttes peaks loodav EXS-i andmebaas olema orienteeritud X-keelsele lõppkasutajale, kusjuures andmete esitusviis peab võimaldama genereerida sellest passiivsete (suund E–X) ja edaspidi ka aktiivsete (suund X–E) tulemsõnastike põhjasid.

### 3.3. Sõltumatus kujundusest

Lõpuks peab EXS olema sõltumatu ka konkreetse tulemsõnaraamatu kujundusest (vormistusest). Selle (nagu üldse sõnastikuandmete töötlemise) tagatiseks on andmete esitamine XML (ingl *Extended Markup Language*) vormingus (vt ka 4. ptk). Kogu koostamis- ja toimetamistöö toimub XML märgendusega andmebaasis, kus iga sõnaartikli element on varustatud XML sildiga, mis näitab üksnes elemendi sisulist staatust. Kuidas üks või teine sõnaartikli komponent küljendatud tekstis välja näeb (fondid, kirjastiilid, taanded jms), määratakse eraldi reeglitega, mis pole seotud sisulise koostamistööga. Iga tulemsõnastiku toimetaja saab need reeglid ise määrata ja oma sõnastikku sel viisil kujundada.

Teine kujundusega seotud probleem on teksti lühendamine, mida tavasõnastikes kasutatakse paberi kokkuhoiu huvides küllalt palju. Muutevormidel asendatakse korduv (märksõnaga identne) sõnaosa sidekriipsuga, nt ÕS2006

**kajut** <g: -i> (*loe*: kajuti) ..

**alfa.beet** <20: -beedi, -.beeti> (*loe*: alfabeedi, alfa.beeti) ..

Näidetes asendab märksõna või tema osa sageli tilde, nt EVS:

**algupära|ne** .. ~ne kirjandus (*loe*: algupärane kirjandus) ..; ~sed tekstid (*loe*: algupärased tekstid) ..

**algus** .. julge ~ (*loe*: julge algus) ..; .. ~te ~ (*loe*: alguste algus) ..

ÕS2006 teeb lühendusi ka liitsõnaloendi piires, nii et lühendusmärgi asemele tuleb panna mitte märksõna, vaid loendi esimene või viimane sõna, nt

**komme** .. Elu<sub>,</sub> laua<sub>,</sub>komme. Eksi<sub>,</sub> ime<sub>,</sub> õnne<sub>,</sub>kombel. Kombe<sub>,</sub>lõtv,  
<sub>,</sub>vastane, <sub>,</sub>vabadus, <sub>,</sub>õpetus

EXS-i andmebaasis pole ruumi kokkuhoid oluline, tähtsam on funktsionaalsuse tagamine. Seetõttu on kõik keelendid lühendamata välja kirjutatud, nii et neid saab andmebaasis otsida ja töö käigus ümber paigutada ilma vormistuse pärast muretsemata. Kõik lühendamised toimuvad automaatselt vastavate reeglite alusel küljendusvaate genereerimise ajal koos muu kujundusega. See tagab ka lühenduste töökindluse, nii et võõrkeelne kasutaja (v arvuti) saab lihtsa formaalse algoritmi alusel lühendatud kohad taastada.<sup>5</sup>

**3.4.** Ülaltoodud sõltumatusnõuete (sihtkeelest, sõnastikutüübist, kujundusest) täitmine EXS-i andmebaasis annab uue X-sõnaraamatu toimetajale üsna vabad käed. Ta saab valida sobiva suurusega sõnastikupõhja ja häälestada selle vastavalt oma tulemsõnaraamatu vajadustele nii andmete hulga ja struktuuri kui ka kujunduse ja muu vormistuse osas.

## 4. EKI sõnastike haldussüsteem ja kirjegeneraator

EXS-i universaalsuse ja paindlikkuse kindlustamisel mängib suurt rolli EKI keeletehnoloogide poolt loodud töövahend – sõnastike haldussüsteem EELex –, mis kujutab endast leksikograafi elektroonilist töökeskkonda ning toetab ja abistab sõnaraamatu koostajat, võimaldades ka rühmatööd veebis (lähemalt vt Langemets jt 2006, Loopmann 2007). Seda kasutatakse nii EXS-i enda koostamisel kui ka kakskeelsete tulemsõnaraamatute tegemiseks.

EELex-i tööpõhimõtted toetuvad Nancy Ide juhtimisel sõnastike jaoks välja töötatud mudelile (Ide jt 2000), mille alustala on sõnastiku vormi ja sisu sõltumatus ning vormingute variatiivsus. Sõnastikku käsitletakse kui XML vormingus kodeeritud dokumenti, mida saab transformeerimiskeele XSL (ingl *Extensible Stylesheet Language*) abil teisendada muudeks XML dokumentideks, mille ele-

<sup>5</sup> Üldjuhul kehtib sõnastikes lühendamisel kokkulepe, et muutevormides või näidetes asendab sidekriips v tilde tervet märksõna (nt EVS: algus .. ~te ~). Kui märksõnast kasutatakse ainult algusosa, siis tähistatakse lühendamiskoht mingi märgiga (nt EVS: algupära|ne .. ~sed). ÕS2006-s pole aga muutevormide lühendamisel formaalset algoritmi järgitud ja on jäetud lühendamiskohad märksõnas tähistamata. Seetõttu on osa ÕS2006 lühendusi arusaadavad ainult emakeelsele kasutajale, kes sõnavorme teab. Vrd nt

**lihas** <13: -e> ja **lihane** <13: -se>

**kirmas** <13: -e> ja **ainumas** <8: -ma, -mat>

**.lööke[ne]** <12 ja 13: -kese> ja **.maakene** <12 ja 13: -se>

**järguline** <12: -lise> ja **järgutine** <12: -se>

**püra.miid** <20: -miidi, -.miidi> ja **püra.miidjas** <8: -ja, -jat>

mentide konfiguratsioon (struktuur) võib olla hoopis teistsugune. See mudel annab võimaluse sõnastike struktuurielemente soovitud moel esitada, valida ja omavahel kombineerida, samuti ka sõnastikest infot otsida ja sobival kujul väljastada.

Kõik sõnastikubaasid EELex-is on XML vormingus (lähemalt vt Loopmann jt 2006: 248 jj), mis on kujunenud leksikograafias üldtunnustatud standardiks. XML on standardiseeritud andmekirjelduskeel, mis võimaldab andmeid struktureerida sisu järgi, sõltumata nende vormistusest. Iga sisuline sõnaartikli element (märksõna, tähendus, stiilimärgend, tüübinumber jne) on märgistatud tähisega (XML sildiga) ja tal on oma kindel koht sõnastiku struktuuriskeemis. Kuigi kogu sõnastiku koostamis- ja toimetamistöö toimub XML vormingu baasil, pakutakse kasutajale võimalust töötada ka visuaalselt mugavamate vaadetega, nt täita struktuurilahtreid sõnastikutabelis või teha väiksemaid parandusi küljendusvaates. XML vormingule toetub ka tulemsõnastike genereerimine EXS-i andmebaasist.

Haldussüsteemiga on plaanis siduda EKI grammatiline kirjegeneraator, mis dialoogis sõnaraamatu koostajaga lisab sõnaartiklisse morfoloogilise info, vastavalt valitud kirjestandardile. Kirjegeneraatori varasemat versiooni on juba mitme sõnastiku puhul kasutatud (vt Viks 2000b). Praegu käib selle edasiarendamine, et temast saaks igale leksikograafile käepärane töövahend. Kirjegeneraatorit ei hakata kasutama kõigi märksõnade puhul, sest EXS-i andmebaas on valmiskujul salvestatud teatud hulk standardseid kirjeid. Kui aga koostaja lisab sõnaraamatusse mõne uue sõna, mida andmebaas pole, siis käivitub dialoogis töötav kirjegeneraator. Dialoogiliidese abil lahendatakse muuhulgas ka probleemid, mis pole arvuti poolt üheselt lahendatavad, nt kirjete jaotumine homonüümsete märksõnade vahel. Sõnastiku koostajal on võimalus teha sõna sisust sõltuvaid valikuid ja vajaduse korral korrigeerida generaatori väljundit.

## Kirjandus

- Ide, Nancy; Kilgarriff, Adam; Romary, Laurent 2000. A formal model of dictionary structure and content. – U. Heid, S. Evert, E. Lehmann, C. Rohrer (eds). Proceedings of the ninth EURALEX International Congress, EURALEX 2000. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 113–126.
- Järventausta, Marja 2004. Sanakirjat suomenoppijan avuna. – H. Sulkala, H. Laanekask (toim.). VIRSU II. Suomi ja viro kohdekielinä. Lähivertailuja 15. Suomen ja saamen kielen ja logopedian laitoksen julkaisuja 24. Oulu: Oulun Yliopisto, 24–44.
- Kaalep, Heiki-Jaan 1996. ESTMORF: A morphological analyzer for Estonian. – H. Õim (ed.). Estonian in the Changing World. Tartu: University of Tartu, 43–98.
- Langemets, Margit; Loopmann, Andres; Viks, Ülle 2006. The IEL dictionary management system of Estonian. – G-M. de Schryver (ed.). DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writing Systems. Pre-EURALEX workshop, Turin, 5th September. Turin: University of Turin, 11–16.
- Langemets, Margit; Mägedi, Merike; Viks, Ülle 2005. Süntakiline info sõnastikus: probleeme ja väljavaateid. – M. Langemets (koost.), M-M. Sepper (toim.). Eesti Rakenduslingvistika Ühingu aastaraamat 1 (2004). Tallinn: Eesti Keele Sihtasutus, 71–98.
- Loopmann, Andres 2007. Sõnastike haldussüsteem EELex. Magistritöö. Käsikiri Eesti Keele Instituudis. Tartu: Tartu Ülikool.
- Loopmann, Andres; Sein, Kati; Viks, Ülle 2006. Sõnastike haldussüsteem Eesti Keele Instituudis. – M. Koit, R. Pajusalu, H. Õim (toim.). Keel ja arvuti. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 6. Tartu: Tartu Ülikooli Kirjastus, 246–258.

- Vare, Silvi 2002. Eesti keele sõnaperede sõnaraamatu koostamise põhimõtetest. – Keel ja Kirjandus 12, 855–866.
- Veldi, Enn 2000. Estonian and English: A lexicographer's point of view. I. – M. Erelt (ed.). Estonian: Typological Studies IV. Publications of the Department of Estonian of the University of Tartu 14. Tartu: Tartu Ülikooli Kirjastus, 173–197.
- Viks, Ülle 2000a. Eesti keele avatud morfoloogiamudel. – T. Hennoste (toim.). Arvutuslingvistikalt inimesele. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 1. Tartu: Tartu Ülikooli Kirjastus, 9–36. [http://www.eki.ee/teemad/avatud\\_mrf.html](http://www.eki.ee/teemad/avatud_mrf.html) (28.09.2007).
- Viks, Ülle 2000b. Kuidas tekib sõnastikukirjese grammatika. – Keel ja Kirjandus 7, 486–495. <http://www.eki.ee/teemad/kirjegeraator.html> (28.09.2007).
- Viks, Ülle 2001. Muutüübid eesti sõnastikes. – M. Langemets (toim.). Leksikograafiaseminar: Sõna tänapäeva maailmas. Leksikografinen seminaari: Sanat nykymaailmassa. Eesti Keele Instituudi toimetised 9. Tallinn: Eesti Keele Sihtasutus, 157–185. <http://www.eki.ee/teemad/tyybijutt.html> (28.09.2007).
- Viks, Ülle 2003. Viron muoto-oppia / Eesti morfoloogia. Viron vastineiden taivutussanasto / Eesti vastete morfoloogiasõnastik. – Soome-eesti suursõnaraamat 2. Tallinn: Eesti Keele Instituut; Helsinki: Kotimaisten kielten tutkimuskeskus, 929–1301.

### **Sõnaraamatud ja võrgumaterjalid**

- ENNET = Farbregd, Turid; Lepp, Hille; Viks, Ülle 1999. Estisk Lommeordbok. Eesti-norra norra-eesti. Oslo: Kunnskapsforlaget.
- EVS = Eesti-vene sõnaraamat. 1997–2006 (1–4). Tallinn: Eesti Keele Sihtasutus.
- NEEN = Farbregd, Turid; Kangur, Sigrid; Viks, Ülle 2006. Norra-eesti eesti-norra sõnaraamat. 2. trükk. Tallinn: Eesti Keele Sihtasutus.
- SUVI = Soome-eesti suursõnaraamat. 1-2. Anu Haak, Paul Kokla, Külli Kuusk, Helga Laanpere (koost.), Valdek Pall (toim.). Tallinn: Eesti Keele Sihtasutus, 2003.
- VVS = Viks, Ülle 1992. Väike vormisõnastik I, II. Tallinn: Keele ja Kirjanduse Instituut.
- ÕS2006 = Eesti õigekeelsussõnaraamat ÕS 2006. Tiiu Erelt (toim.), Tiiu Erelt, Tiina Leemets, Sirje Mäearu, Maire Raadik (koost.). Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus.
- EKI automaatne morfoloogiasüsteem. <http://www.eki.ee/tarkvara/> (28.09.2007).

**Ülle Viks** (Eesti Keele Instituut) on lõpetanud Tartu Riikliku Ülikooli eesti keele ja kirjanduse erialal. Kaitses samas ülikoolis 1978. a kandidaadiväitekirja eesti verbide morfoloogilisest klassifikatsioonist ja 1994. a doktoriväitekirja eesti keele klassifikatoorsest morfoloogiast. Uurimisvaldkonnad: avatud morfoloogiamudel ja selle realisatsioon arvutis (sõnavormide reeglipõhine süntees ja analüüs), grammatilise informatsiooni esitamine sõnastikes.  
[ylle@eki.ee](mailto:ylle@eki.ee)

## ESTONIAN-X DICTIONARY AND GRAMMAR

**Ülle Viks**

Institute of the Estonian Language

Problems connected with representation of grammatical information in a bilingual dictionary are discussed. The focus lies on the Estonian-X dictionary (EXS), its place in dictionary typology and consequent requirements. The EXS is a universal lexico-grammatical database enabling to generate the electronic basis for an arbitrary bilingual dictionary, providing Estonian as the source language ready-made in a standard format. It is designed to present grammatical information as well structured as possible, tapping various levels of language description from phonology to semantics; data selection and presentation are based on relations of dictionary vs. grammar. The EXS being a metalevel database its format should be flexible enough to be independent of the target language, type and design of the output dictionary. A brief introduction is made of the available tools to develop and implement the EXS: the dictionary management system EELEX (lexicographer's working environment) and a grammatical entry generator.

**Keywords:** bilingual lexicography, grammatical information, language processing, XML, dictionary management system, Estonian