



# Eesti Keele Instituut



Euroopa Liit  
Euroopa Sotsiaalfond



Eesti tuleviku heaks



Eesti Keele Instituut



Euroopa Liit  
Euroopa Sotsiaalfond



Eesti tuleviku heaks



# Extending the functions of the EELex dictionary writing system using the example of the Basic Estonian Dictionary

**Madis Jürviste, Jelena Kallas,**

**Margit Langemets, Andres Loopmann, Maria Tuulik, Ülle Viks**

**Institute of the Estonian Language**

**Tallinn, Estonia**

madis.jyrviste@eki.ee, jelena.kallas@eki.ee, margit.langemets@eki.ee,  
andres.loopmann@eki.ee, maria.tuulik@eki.ee, ylle.viks@eki.ee



# Dictionary writing system EELEX: general features

- **Unicode support**
- **XML databases**
- **XSD schemas**
- **XSL transformations for generating different views**
- **click-to-edit**
- **structural queries**
- **sorting of query results**



# **New functions of EElex (I)**

## **- dictionary compilation and customization -**

- **schema editor**
- **article preview generator**
- **cross-reference checker**
- **spell check in input text boxes**



## **New functions of EELex (II) - dictionary compilation and customization -**

- **morphological analysis, synthesis**
  - **automatic generation of morphological information**
- **bulk corrections interface**
- **list data manager / menu compiler**
- **use of multimedia data**



## **New functions of EELEX (III) - dictionary publishing -**

- **create different versions of print dictionaries from one database**
  - **selecting different structure elements to display**
- **export entries to print layout format**
- **step-by-step entry content displaying**  
**in e-dictionaries**



## Global search function

- **allows the user to search for a headword in other EELex dictionaries or common headwords in different dictionaries**



# The Basic Estonian Dictionary XML-database

1. pronunciation
  2. morphology
  - + pronunciation in mp3
  3. meaning description
  4. word formation
  5. government patterns
  6. collocation patterns
  7. multiword expressions
  8. sense relationships
    - synonymy & antonymy
  9. illustrations
  10. comments
- **example: *usk* 'faith'**





# Metadata

- **semantic type** of the word sense(s)
- **entry templates**
- **perspective: possibility to generate onomasiological dictionaries out of semasiological**



# Government and collocation patterns in the BED

- Sketch Engine for Estonian -

- **patterns are presented in separate blocks and grouped according to collocation and government codes**
- **patterns are provided by word sketches**



## Conclusions and perspectives (I)

- **integration of EELEX and Estonian Sketch Engine**
- **integration of EELEX and Estonian speech synthesizer**
- **compilation methods allow to select display elements in the end-user interface: headwords, domain labels, usage examples, grammatical information, etc.**



## Conclusions and perspectives (II)

- **direct links to entry content word-forms to find according headwords**
- **direct links to EELex global search, as well as other resources (e.g. Wikipedia)**
- **language learning environment based on EELex data**



# References

- **Kallas, J., Tuulik, M. (2011).** Eesti keele põhisõnavara sõnastik: ajalooline kontekst ja koostamispõhimõtted. In *Eesti Rakenduslingvistika Ühingu aastaraamat 7. Tallinn: Eesti Rakenduslingvistika Ühing, pp 59–75.*
- **Langemets, M., Loopmann, A. & Viks, Ü. (2010).** Dictionary management system for bilingual dictionaries. In S. Granger, M. Paquot (eds.) *eLexicography in the 21st century: New challenges, new applications. Proceedings of eLex 2009. Louvain-la-Neuve: Presses universitaires de Louvain, Cahiers du CENTAL, pp 425–430.*

**Thank you for Your attention!**



# **Appendix: screenshots**

Skeem -- Webpage Dialog

Otsi: @c:aT    1/0

- [0, 1] - <c:sls> = sag-S sõnaliik
- [0, 1] - <c:vk> = vormikood
- @c:g = genereeritud
- [0, -1] - <c:mfp> = morfoloogia plokk
- @c:aT = toimetaja**
- @c:aTA = toimetamise aeg
- @c:g = genereeritud
- @c:kl = kirje liik
- [0, -1] - <c:gkg> = gram.kategooria grupp
  - @c:gkn = kategooria nimi
  - [0, -1] - <c:mvg> = vormigrupp
    - @c:vn = vorminimi
    - [0, -1] - <c:mvgp> = paralleelvormi grupp
      - [0, 1] - <c:mv> = muutevorm
      - [0, -1] - <c:hldf> = häälduse failid
        - @c:mT [obl] = meedia tüüp
        - @c:aK = ploki koostaja
        - @c:aKL = ploki koostamise lõpp
- [0, 1] - <c:mt> = muuttüüp
- [0, -1] - <c:sl> = sõnaliik
- [0, 1] - <c:gki> = gr.kasutusinfo
- [0, -1] - <c:par> = paronüüm
- [0, -1] - <c:mvt> = märksõnaviide
- @c:i = hom-number

Nim ek.	toimetaja	Nim ik.	new attribute
Nim vk.	noovõli atribut		
Liik	ainult tekst	Tüüp	xs:string
Min.		Maks.	

B    I    U    ↑aA  
Font: [ ]   Suurus: [ ]  
Värv: [ ]   Taust: [ ]  
Algv. [ ]    Luua   Põhielement

c:aT    27: c:sr /c:A /c:P /c:mg /c:mfp /@c:aT    Tüübid

xs:boolean  
xs:dateTime  
xs:decimal  
xs:language  
xs:positiveInteger  
xs:string  
xs:unsignedByte

http://eelex.dyn.eki.ee/\_shs/html/gen\_skeem.htm    Trusted sites

# Entry schema editor



Artikli vaate kujundus: Eesti keele põhivõnava sõnastik, 'Kallas' -- Webpage Dialog

@xml:lang> - sõnaraamat  
A @c:KF @c:AS[obl]> - artikkel  
:c:P> - päis  
- <c:mg> - märksõna grupp  
,1] - <c:sag> - sagedusrühm  
,1] - <c:eelrek> - eelreksioon  
∞] - <c:m @c:i @c:liik @c:ps @c:u @  
,1] - <c:järelrek> - järelreksioon  
,1] - <c:hldf @c:mT[obl] @c:aK @c:aK  
,1] - <c:lyh> - lühend  
,1] - <c:sls> - sag-S sõnaliik  
,1] - <c:vk @c:g> - vormikood  
∞] - <c:mfp @c:aT @c:aTA @c:g @c  
[0, ∞] - <c:gkg @c:gkn> - gram.kategor  
[0, ∞] - <c:mvg @c:vn> - vormigrupp  
[0, ∞] - <c:mvgp> - paralleelvormi  
[0,1] - <c:mv> - muutevorm  
[0, ∞] - <c:hldf @c:mT[obl] @c:  
[0,1] - <c:mt> - muuttüüp  
[0, ∞] - <c:sl> - sõnaliik  
[0,1] - <c:gki> - gr.kasutusinfo  
- <c:par> - paronüüm  
- <c:mvt @c:i @c:t @c:mvtl> - märksõ  
<c:S> - sisu  
- <c:tyg1> - ühisinfo\_1  
,1] - <c:grg> - grammatikagrupp  
[0, ∞] - <c:sl> - sõnaliik  
[0,1] - <c:gki> - gr.kasutusinfo  
[0,1] - <c:gkom> - gr. kommentaar  
- <c:tp @c:tnr[obl]> - tähendusnumbri  
,1] - <c:grg> - grammatikagrupp  
[0, ∞] - <c:sl> - sõnaliik

**<c:S> - sisu (esineb c:A)**  Glob.  Näidata

**B**  *I*  U  ↑a  Link  Uus rida  Allelem  Ei trüki

Lähtekeel:

Loend  Äär  Stiiil  Paigutus

Font  Suurus

Värv  Taust  Asendus

Süm. eelm:  ümber:  A ees:  tag

xx-small  
x-small  
small  
**medium**  
large  
x-large  
xx-large

**usk**  
1. (*ainsuses*) kindel sisemine veendumus, milles ei kahelda *Ainult ravimid ei aita, ka usku peab olema.*  
- **kellesse-millesse** *Mul on tema võimetusse usku. Ta leidis uuesti usu jumalasse.*  
- **usku kaotama** *Ma olen kaotanud usu inimestesse.*  
- **kindel, suur; pime usk**  
2. religioon *Mis usku sa oled?*  
- **katoliku, luteri usk**

art ... tingArt Salvesta

OK

# Entry preview generator

http://eelex.dyn.eki.ee/ - PSV - Windows Internet Explorer

Otsitav piirkond [0,∞] - <c:A> - artikkel

Otsitav element [0,∞] - <c:kn> - käändereksioon

A - Y     tõstutundetu    \*.\*   

märkidega     globaalselt

AB: A - Y

muuda väärtusi (c:kn)  
 muuda väärtusi (c:kn) **muuda väärtusi (c:kn)**  
 kustuta (c:kn)  
 asenda element (c:reg) grupis  
 lisa element (c:reg) grupis  
 kustuta element (c:reg) grupis  
 muuda gruppe (c:reg)  
 kustuta artikkel (c:A)

<input checked="" type="checkbox"/> 24	koosnema	<c:reg>	kellest-millest Ansambel koosneb viiest lauljast., Romaan ko osast.		
<input checked="" type="checkbox"/> 25	kuuluma	<c:reg>	kellele-millele See raamat kuulub Jaanile., See maa kuulub riigile.		kellele v millele
<input checked="" type="checkbox"/> 26	kõlbama	<c:reg>	kelleks-milleks Ta ei kõlba õpetajaks.		kelleks v milleks
<input checked="" type="checkbox"/> 27	käima	<c:reg>	kelle-mille kohta See märkus ei käi sinu kohta.		kelle v mille kohta
<input checked="" type="checkbox"/> 28	üumber käima	<c:reg>	kellega-millega Ta oskab kõigiga hästi ümber käia.; Ta ei oska rahaga ümber käia. (ta ei oska raha mõistlikul viisil kasutada)		kellega v millega
<input checked="" type="checkbox"/> 29	lugu pidama	<c:reg>	kellest-millest Temast peetakse lugu.		kellest v millest
<input checked="" type="checkbox"/> 30	lähinema	<c:reg>	kellele-millele Rong läheneb jaamale.		kellele v millele
<input checked="" type="checkbox"/> 31	mööda minema	<c:reg>	kellest-millest Läheneb iga päev sellest kauplusest mööda.		kellest v millest
<input checked="" type="checkbox"/> 32	mõjuma	<c:reg>	kellele-millele See jutt talle ei mõju.		kellele v millele
<input checked="" type="checkbox"/> 33	mõjutama	<c:reg>	mida-keda Toit mõjutab tervist., See otsus mõjutab kõiki.		mida v keda
<input checked="" type="checkbox"/> 34	mööduma	<c:reg>	kellest-millest Auto möödus bussist.		kellest v millest
<input checked="" type="checkbox"/> 35	olenema	<c:reg>	kellest-millest Temast oleneb palju.		kellest v millest

     MySQL Otsing: 0m, 1s; 40 objekti, 34 art.

# Bulk corrections

# [SP] "Eesti keele sõnapered"

Tööversioon!

Kasutajate tagasiside on oodatud: [Silvi Vare](#)



Päring:

artikli osas

Leitud 6 artiklit

**kool** s *õppe- ja kasvatusasutus; haridus- ja koolisüsteem; koolitus; kogemused, õppetund, koolkond*

[P\_TUL]

kool=kond s *suund kunstis, kirjanduses, teaduses vm.*

kooli=ta#ma v

kooli|t=us s

täie|nd|us+kooli|t|us s

muusiku+kooli|t|us s

õpi#poisi+kooli|t|us s

kooli|t|us+telli|mus s

harju|t=us+kool s

**muusika** s *(kunstiliik); PILT miski muusikat meenutav v. muusikana mõjuv*

[P\_TUL]

muusik s *helilooja, (instrumentaalmuusika) interpret jm. (kutseliselt) muusikaga tegelev isik*



muusiku+kooli|t|us s

**Example of step-by-step  
entry content displaying in EELex**

<input checked="" type="checkbox"/>	Väärtus	Seletus eesti	Seletus ingl	Parandaja	Parandamise aeg
	aiand	aiandus	gardening	-	-
	aj	ajalugu	history	-	-
	anat	anatoomia	anatomy	-	-
	antr	antropoloogia	anthropology	-	-
	arheol	arheoloogia	archeology	-	-
	arhit	arhitektuur	architecture	-	-
	astr	astronoomia	astronomy	-	-
	auto	autondus	automobile engineering	-	-
	bibl	bibliograafia	bibliography	-	-
	biol	bioloogia	biology	-	-
	bot	botaanika	botany	-	-
	ehit	ehitusala	building	-	-
	el	elekter ja elektrotehnika	electricity	-	-
	etn	etnograafia	ethnography	-	-
	farm	farmaatsia	pharmacy	-	-

Uue väärtuse lisamine

Sisesta uus väärtus!

OK

Cancel

aut|

OK

# List data manager

http://eelex.dyn.eki.ee/run/SL/Koond/Koond.html#/Algus indeks

**sõnastike koondindeks** indeks sõnastikud rakendusest

usk

Nr.	Märksõna	Hom. nr	Liik	Sõnastik
1	usk			Eesti-vene sõnaraamat
2	usk			Eesti-X andmebaas
3	usk			Eesti-Soome sõnaraamat
4	usk			Põhisõnavara sõnastik
5	usk			Õigekeelsussõnaraamat ÕS 2006
6	usk			SõnaPered
7	usk			Eesti keele seletav sõnaraamat
8	usk			Ühekõiteline seletav sõnaraamat
9	usk			Eesti-udmurdi sõnaraamat

9 kirjet / 0m, 0s.

Otsiti märksõnu 'usk' järgmistest sõnastikest: EVS, ETY, EX\_, FIN, GEO, HAR, IES, IEF, LAT, LDW, LEX, LIT, OD\_, PSV, QQV, QS\_, SP\_, SS\_, SS1, UDM, UKR, USV, VSL.

**usk** <20:usu, .usku>. Ta on luteri usku *ta on luteriusuline, luterlane*. Rahvas pööراتи risti\_usku, vene usku, muhamedi usku. Usk enesesse, tulevikku = enese\_, tuleviku\_usk. Ema usk oma poja andekusse. Läksime kindlas usus, heas usus v hea usu peale. Elu\_, võidu\_, õnne\_usk. Eba\_, lahk\_, riigi\_, risti\_, umb\_, väär\_, õige\_usk. Usu\_elu, filosoofia, kannataja *märter*, komme, kultus, liikumine, lõhe, salgaja, sallivus, taganeja = vääraja *ketser*, teaduskond, vabadus, vend, õde, ühing

## EELEX global search function

## usk **S**

♪ <'usk, usu, 'usku, 'usku; *mitmus*: usud, 'uskude, 'uske = 'uskusid  
22 S>

1. (*ainsuses*) kindel sisemine veendumus, milles ei kahelda  
*Ainult ravimid ei aita, ka usku peab olema.*

▪ *kellesse-millesse* *Mul on tema võimetusse usku.*

● **usku kaotama** *Ma olen kaotanud usu inimestesse.*

● **usku leidma** *Ta leidis uuesti usu jumalasse.*

● **kindel, suur; pime usk**

2. religioon *Mis usku sa oled?*

● **katoliku, luteri usk**

**BED entry for noun *usk* ‘faith’**

http://eelix.dyn.eki.ee/\_shs/art.cgi Eesti keele põhisõnava sõ...

Kõide: Kõik kõited

semantiline tüüp loom

Otsi xml

**Päring: [semantiline tüüp (./c:st) [ 'loom' (<->: 4)], tt-tu, m-deta, glob.; 25 leidu, 23 artiklit.**  
(näidatakse esimesed 5000 artiklit)

**Lex**

Jnr: KB	Kd.	Märksõna(d)	Leid	K.	K. lõpp	T.	T. aeg	T. lõpp	Pt.
1: 1	I	<a href="#">hiir</a>	<st> ▪loom	MTuulik	2011-03-16 16:52:50	JKallas	2011-05-04 15:47:28	JKallas, 2011-05-04 15:47:32	
2: 1	I	<a href="#">hobune</a>	<st> ▪loom	MTuulik	2011-03-16 11:21:58	JKallas	2011-06-16 13:12:44	JKallas, 2011-06-16 13:13:04	
3: 1	I	<a href="#">hunt</a>	<st> ▪loom	MTuulik	2011-03-16 11:31:12	JKallas	2011-04-19 15:20:13	JKallas, 2011-04-19 15:20:46	
4: 1	I	<a href="#">jänes</a>	<st> ▪loom	EKI		LHollman	2011-08-10 00:41:56		
5: 1	I	<a href="#">kala</a>	<st> ▪loom	EKI		JKallas	2011-10-11 11:42:18		
6: 1	I	<a href="#">karu</a>	<st> ▪loom	MTuulik	2011-03-16 16:22:04	JKallas	2011-04-20 10:45:30	JKallas, 2011-04-20 10:46:05	
7: 1	I	<a href="#">kass</a>	<st> ▪loom	MTuulik	2011-03-25 17:04:02	JKallas	2011-10-25 11:41:01	JKallas, 2011-10-25 11:41:01	
8: 1	I	<a href="#">kits</a>	<st> ▪loom :: <st> ▪loom	MTuulik	2011-03-25 17:02:23	JKallas	2011-05-05 11:52:45	JKallas, 2011-05-05 11:52:50	
9: 1	I	<a href="#">koer</a>	<st> ▪loom	MTuulik	2011-03-25 14:32:42	JKallas	2011-04-20 15:44:32	JKallas, 2011-04-20 15:45:06	
10: 1	I	<a href="#">konn</a>	<st> ▪loom	MTuulik		LHollman	2011-09-01 00:38:32		

Version : 13.06.2011

## Semantic types in the BED

	<b>Kollokatsiooniplokk</b>	
	<u>Kollokatsioonigrupp</u>	
	kollokatsioonikood	Adj+S
	<u>Kollokatsiooni rühm</u>	S+S
↓	kollokatsiooni reksioon	Adj+S
	kollokatsioon	Adj+Sk
	kollokatsioon	Ss+V
	kollokatsioon	So+V
	kollokatsioon	Sk+V
	kollokatsioon	Adj+V
	<u>Kollokatsiooni rühm</u>	Adv+Adj
↓	kollokatsiooni reksioon	Adv+V
	kollokatsioon	Adv+Adv
	kollokatsioon	kange kohv
	kollokatsioon	lahja kohv
↓	kasutusnäited	

## Collocation patterns in the BED



**usk** () EstonianRC freq = 11245

<b>object_of</b> 481 5.6	<b>subject_of</b> 588 1.7	<b>a_modifier</b> 1625 2.1	<b>noun_sisseütlev</b> 324 32.7	<b>adj_modifier_käändumatu</b> 237 8.5
sisendama 47 9.84	puuduma 42 5.49	hea 319 5.84	jumal 48 6.84	katoliku 81 10.33
andma 47 2.61	kaduma 41 5.58	kindel 124 6.45	õiglus 13 7.16	luteri 74 11.33
lisama 34 4.99	aitama 35 4.98	uus 113 3.36	<b>olema_noun</b> 79 8.7	vene 18 3.37
kaotama 28 4.87	lubama 23 3.72	suur 74 2.92	uskmatust 11 11.27	eesti 18 2.93
väljendama 21 6.12	tulema 20 1.14	pime 45 7.37	oopium 10 9.94	muhamedi 17 10.62
kinnitama 19 3.37	jääma 19 1.57	eriline 45 5.69		buda 10 9.72
avaldama 16 3.51	andma 19 1.31	kristlik 39 7.4		
<b>ja/või</b> 953 4.9	<b>adverbial_of_seestütlev</b> 65 3.0	<b>gen_modifier</b> 1053 1.1	<b>a_modifier_comp</b> 49 1.0	<b>gen_modifies</b> 821 0.8
lootus 34 6.1	rääkima 16 2.6	islam 153 10.22	suurem 16 2.42	põhimõte 35 5.33
keel 33 4.52	<b>adverbial_of_seesütlev</b> 68 2.3	inimene 135 3.91		jumal 17 5.29
rahvus 31 6.98	elama 20 3.46	rahvas 50 4.73		esindaja 16 2.74
rass 30 8.73	<b>a_modifier_ordinal</b> 72 1.1	eestlane 24 3.89		kirik 12 3.73
enesekindlus 24 7.92	teine 71 2.83	juut 19 5.87		puudumine 11 3.88
armastus 23 5.83		moslem 16 7.23		küsimus 11 1.5
teadus 18 5.82		kodanik 15 4.16		

**Word sketch for noun *usk* 'faith'**