



# TÜ eesti keele tesaauruse (Eesti Wordneti) täiendamine

**Projektijuht:** *Heili Orav*

**Täitjad:** *Kadri Kerner, Sirli Parm, Lauri Eesmaa, Piia Taremaa, Nele Salveste, Ingmar Jaska, Liisi Pool jt*

25.- 26. november 2010 Tartu

# Ajalugu

- A. Saareste mõisteline sõnaraamat, Filosoofi teaurus
- 1998-1999 osalesime EuroWordNet-2 projektis koos inglise, hollandi, itaalia, hispaania, saksa, prantsuse, tšehhi *wordnet*'idega. Praeguseks maailmas üle 50 keele wordneti.
- Ülesanne, mis on fikseeritud eesti keele arendamise strateegias (2004-2010): arendada välja eesti keele leksikaal-semantiline andmebaas (teaurus)

# Ühte mõistet väljendavad sünonüümsed sõnad (ja sõnaühendid) moodustavad sünohulga.

Näiteks:

nimisõnaline mõiste: *pang, ämber*;

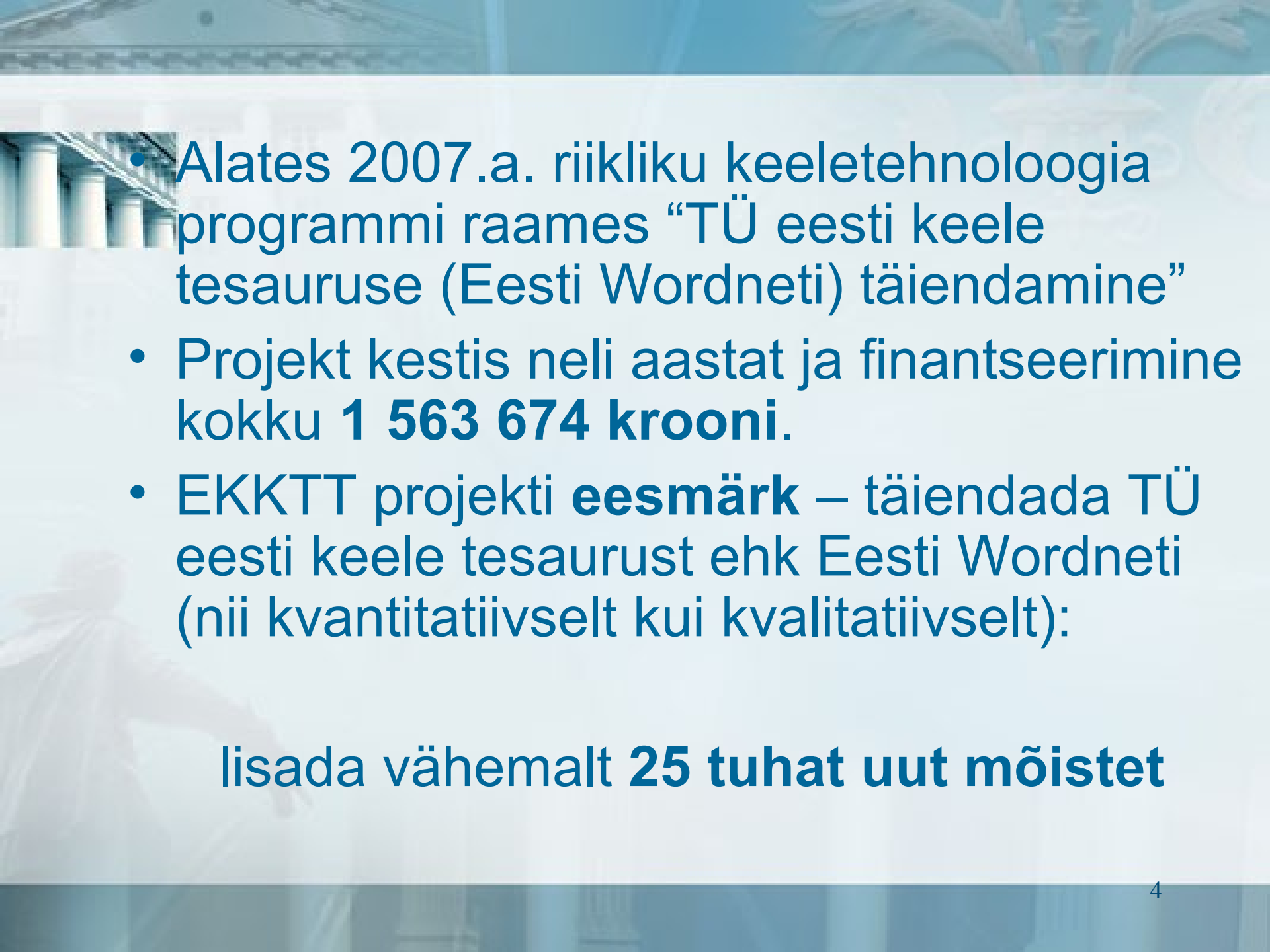
tegusõnaline: *õhutama, agiteerima, kihutama, ässitama, kehitama*;

omadussõnaline: *rõõmsameelne, rõõmus, rõõmsatujuline, päikseline*


## Mõistetevahelised semantilised suhted:

- hierarhilised suhted
- osa-terviku suhted
- rollisuhted
- põhjussuhted jm

**Seotud inglise keelega (ILI, InterLingualIndex)**

- 
- Alates 2007.a. riikliku keeletehnoloogia programmi raames “TÜ eesti keele tesauruse (Eesti Wordneti) täiendamine”
  - Projekt kestis neli aastat ja finantseerimine kokku **1 563 674 krooni**.
  - EKKTT projekti **eesmärk** – täiendada TÜ eesti keele tesaurust ehk Eesti Wordneti (nii kvantitatiivselt kui kvalitatiivselt):

lisada vähemalt **25 tuhat uut mõistet**



Projekti käigus 4 aasta jooksul  
oleme teinud:

Sünohulki/mõisteid **27476**,

milles on erinevaid sõnu **36408**,

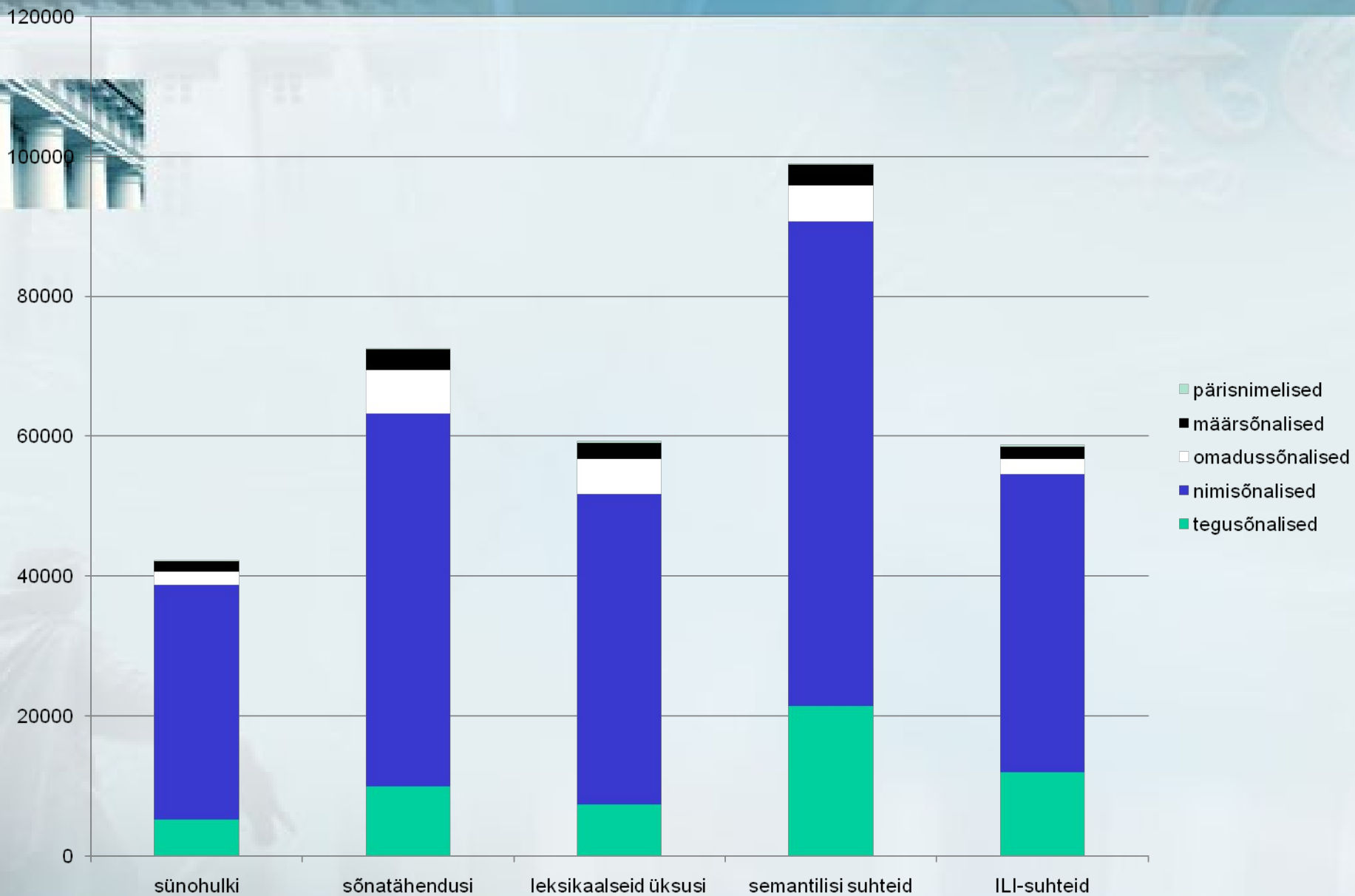
Semantilisi suhteid: **38926**

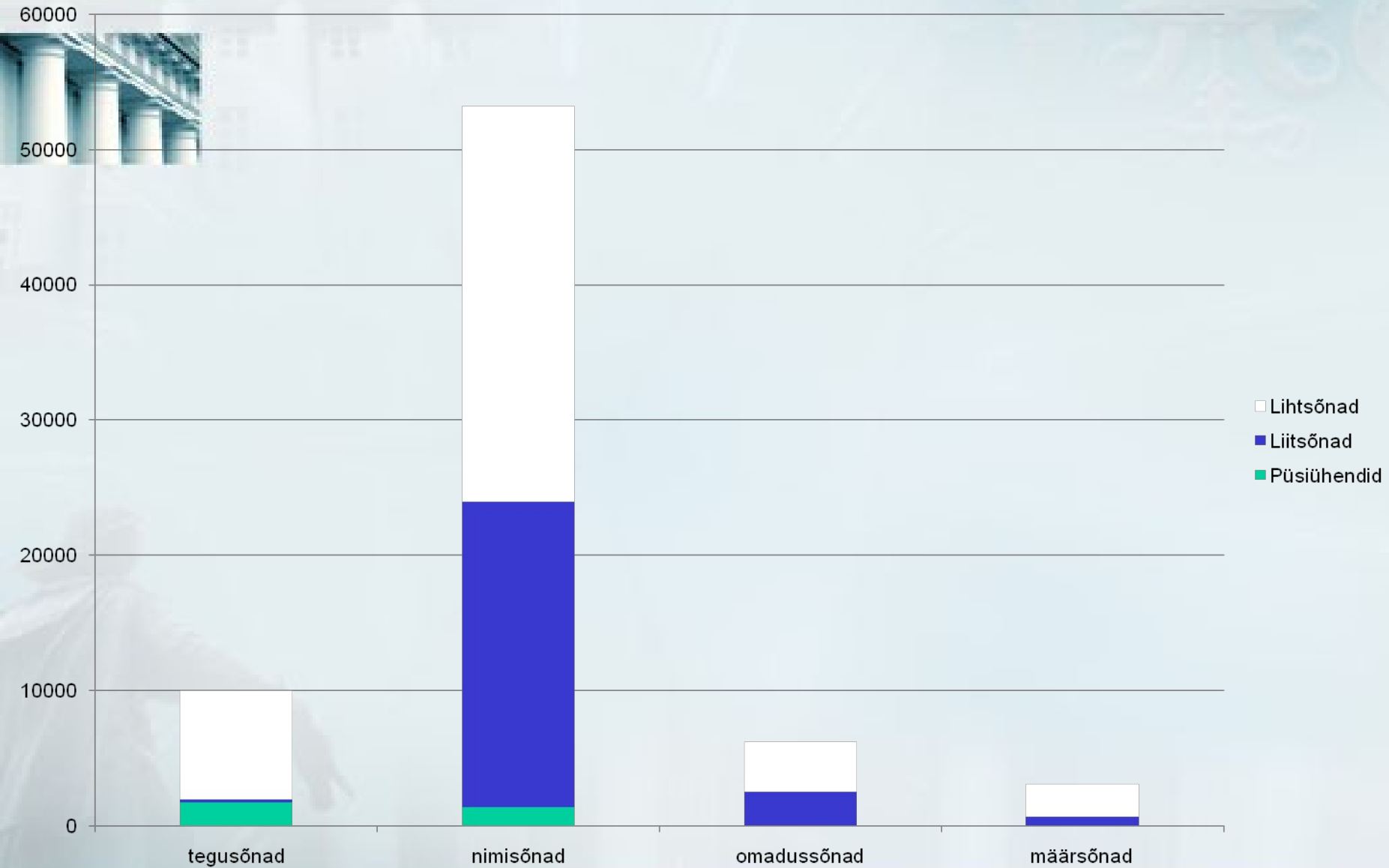
ILI suhteid: **32086**

# Statistika

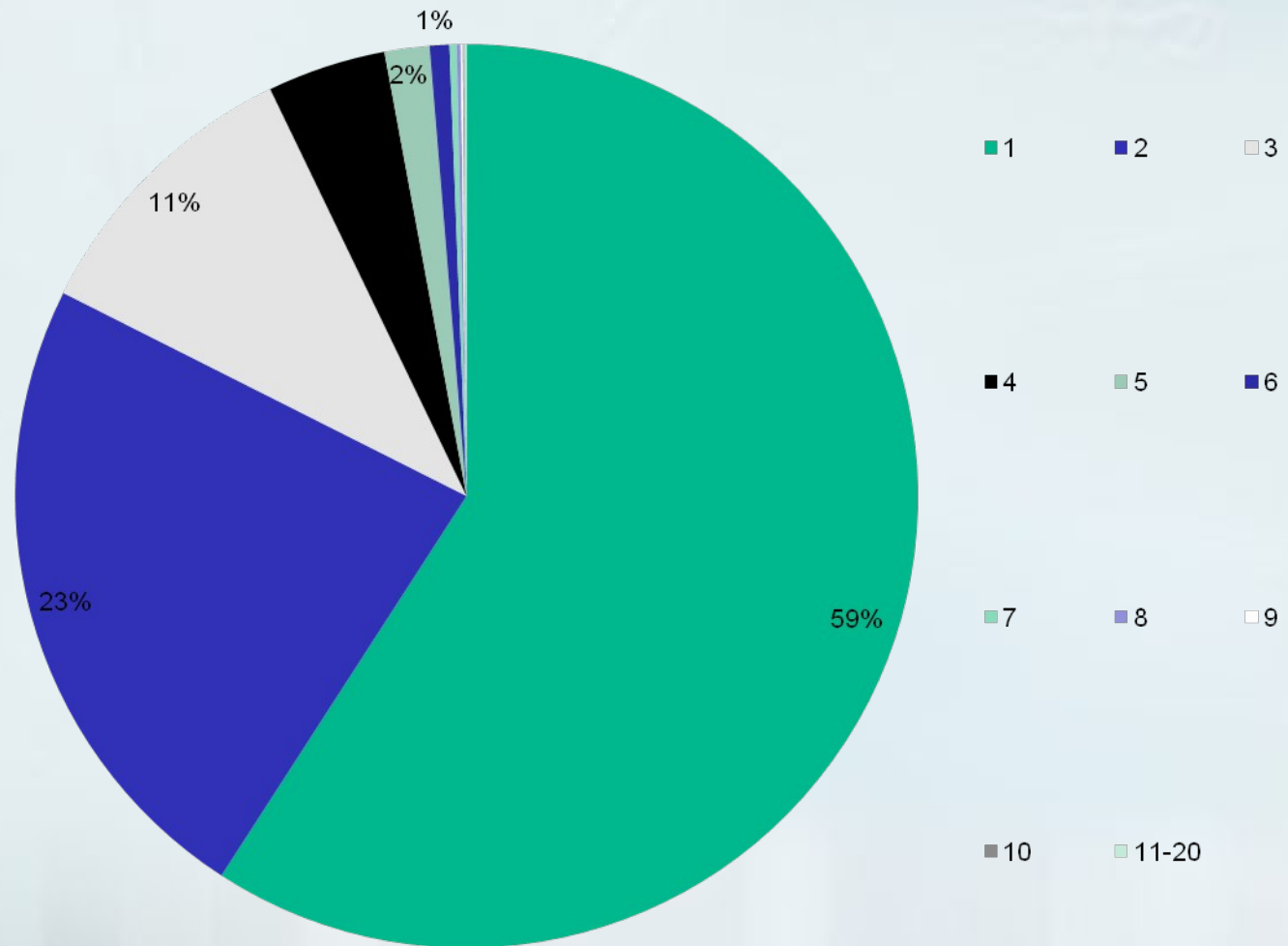


	Enne EKKTT projekti	EKKTT 1. konverents (nov. 2007)	EKKTT 2. konverents (aprill 2009)	Praegu – EKKTT 3. konverents
Nimisõnad	10371	11052	13744	33441
Verbid	3255	3685	4935	5263
Adjektiivid	361	881	1902	1916
Adverbid	0	410	1353	1570
Pärisnimed	446	446	446	151
<b>KOKKU</b>	<b>14433</b>	<b>15828</b>	<b>22380</b>	<b>42341</b>





# Sünohulga liikmete arv



# Kuidas oleme töötanud:

- Uued sõnaliigid
- Valdkondade kaupa
- Automaatselt
  - sünonüümiread Filosoofi tesaurusest
  - *mine*-liited
  - *ja*-liited

## Miinused: kontrollimine

- Sõnatähenduste ühestamise andmed
- Osalemine Global Wordnet Association'i korraldatud konverentsidel

## Edasised ülesanded:

- ❖ Tesaauruse suurendamine – u 10-12 % puudu!
- ❖ Hierarhiapuude kontrollimine – teisiti!
- ❖ Ingliskeelsete vastete kontrollimine
- ❖ Valdkonnamärgendiga sidumine:

Nt

*kool -> õppeastus*

*traktor -> põllumajandus*

- ❖ Tuletiste, liidete automaatne lisamine
- ❖ Tesaauruse litsents



<http://www.cl.ut.ee/ressursid/teksaurus/>

**AITÄH!**