

# Towards Latvian WordNet

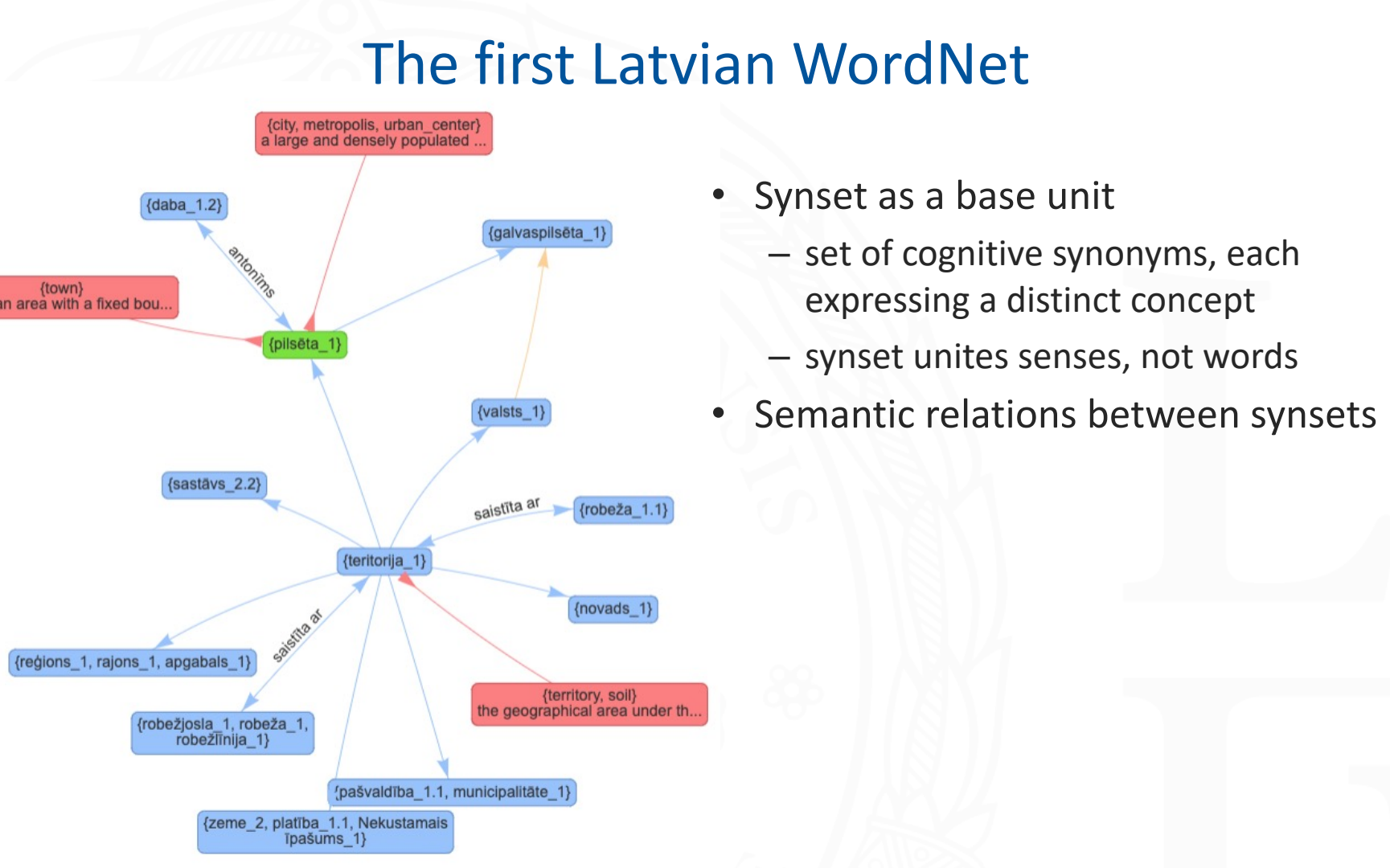
Pēteris Paikens, Mikus Grasmanis, Agute Klints, Ilze Lokmane,  
Lauma Pretkalniņa, Laura Rituma, Madara Stāde, Laine Strankale

[[peteris@ailab.lv](mailto:peteris@ailab.lv)]



# The first Latvian WordNet

- Synset as a base unit
  - set of cognitive synonyms, each expressing a distinct concept
  - synset unites senses, not words
- Semantic relations between synsets



- # The first Latvian WordNet
- 
- Synset as a base unit
    - set of cognitive synonyms, each expressing a distinct concept
    - synset unites senses, not words
  - Semantic relations between synsets

# Project goals

- High quality core WordNet
  - as a basis for bootstrapping for larger coverage
- Based on corpus evidence to reflect the lexical system of Latvian
- Mapping to other WordNets
  - Princeton WordNet
  - Open Multilingual WordNet
- Resource for NLP systems, word sense disambiguation

# Data sources

- Tēzaur.lv online dictionary
  - Largest dictionary for Latvian, 384 000 entries
  - Fully electronic, structured entries
  - Summarized many dictionaries, so a lack of consistency
- Latvian National Corpora collection korpus.lv
  - Balanced contemporary corpus of Latvian (10M tokens)
  - Common Crawl web corpus (400M tokens)
  - Blogs and parliament debates

# Methodology

- Word sense inventory is revisited and adjusted as needed
  - Criteria for distinguishing senses include semantic and syntactic distribution, lexical decomposition, ability to substitute with a synonym
  - Consistent granularity
  - Bottom-up approach based on language usage
- Relations used
  - synsets and gradation sets
  - similarity, hyponymy, antonymy, meronymy

# Public availability - Tezaurs.lv



[T]:

vārds vai frāze



Līdzīgi šķirkļi:

cits burtu lielums: [Baznīca](#)

Apkaime

bazītemporāls

bazīti

bazīties

bāzītis

bāzlinija

bāzlis

bazn.

**baznīca**

Baznīca

baznīcājs

baznīcānis

baznīcāns

baznīcaspriekša

baznīcējš

baznīcdiena

## baznīca

**baznīca** sieviešu dzimtes 4. deklinācijas lietvārds ▼ [Locīšana](#)

1. Kristīgo konfesiju reliģiskajam kultam paredzēta celtne.

▼ Piemēri *Bille ieskatās — cauri kokiem saredzama balta baznīca ar ziliem un zaļiem torņiem.*

▲ Saistītās nozīmes

Sinonīmi ⓘ

***dievnams*** <sub>1</sub> — Baznīca, lūgšanas nams.

Hiponīmi ⓘ

***svētnīca*** <sub>1</sub> — Reliģiska kulta celtne, telpa, arī vieta; arī templis, baznīca, dievnams.

***templis*** <sub>1</sub> — Sakrāla celtne reliģiskajiem rituāliem; arī baznīca, mošeja, pagoda, sinagoga u. tml.

Hiperonīmi ⓘ

***celtne*** <sub>1</sub> — Celtniecības procesā izveidots objekts vai objektu komplekss; arī ēka.

***ēka*** <sub>1</sub> — Celtne, ko parasti izmanto dzīvošanai, saimnieciskām, ražošanas vai sabiedriskām vajadzībām; atsevišķs objekts celtnu kompleksā.

▼ Stabili vārdu savienojumi *iet baznīcā.*

▲ Tulkojumi

*church, church building*

a place for public (especially Christian) worship; "the church was empty"

[Princeton WordNet 3.0]

2. Vienas konfesijas ticīgo organizācija ar noteiktu dogmatiku un kultu.

▼ Piemēri *Viņi iekasēja desmito daļu no visa mužiķu nopelnītā un algas dienā saņemtā — kā katoļu baznīca viduslaikos.*

# Annotation tool

☒ Show all

Synset



**bērnelis<sub>1</sub>, bērņuks<sub>1</sub>, ķipars<sub>2</sub>, ute<sub>3</sub>, bērns<sub>1</sub>, knīpa<sub>1</sub>, kverpis<sub>1</sub>**

bērnelis<sub>1</sub> Bērns.

bērņuks<sub>1</sub> Bērns.

ķipars<sub>2</sub> humoristiska eksprestīvā nokrāsa Bērns.

ute<sub>3</sub> sarunvaloda Bērns.

bērns<sub>1</sub> zēns vai meitene (aptuveni līdz 14 gadu vecumam).

knīpa<sub>1</sub> Maza meitene, mazs zēns.

kverpis<sub>1</sub> Bērns.

## SYNONYM DICTIONARY

**bērns** - mazulis, mazais, bērņuks, bērnelis, ķipars

## TRANSLATIONS

**bērns** - preadolescent, wean, bairn, youngling, babe, child, children, tad, kid, trick, baby, infant, fruit of the womb

## LINKS

### EXTERNAL LINKS:

**(n) child, kid, youngster, minor, shaver, nipper, small\_fry, tiddler,**

**tike, tyke, fry, nestling**

a young person of either sex; "she writes books for children"; "they're just kids"; "tiddler" is a British term for youngster"

### HYPONYMS:

**knīveris<sub>1</sub>, mazpuika<sub>1</sub>, puika<sub>1</sub>, puisāns<sub>1</sub>, puīškins<sub>1</sub>, zēns<sub>1</sub>, zeņķis<sub>1</sub>,**

**zeperis<sub>1</sub>, puisis<sub>1,2</sub>**

knīveris<sub>1</sub> Zēns, puisēns.

mazpuika<sub>1</sub> Zēns.

puika<sub>1</sub> Zēns.

puisāns<sub>1</sub> Zēns.

puīškins<sub>1</sub> Zēns.

zēns<sub>1</sub> Vīriešu dzimuma bērns (aptuveni līdz 11 gadiem); arī pusaudzis.

zeņķis<sub>1</sub> Zēns, aptuveni skolas vecumā; arī pusaudzis.

☒ Only senses ☐ With subsenses ☐ English ☐ English\*

## SENSES

**jaunmeitene<sub>1</sub>**

Jauniete.

**Ganu meitene<sub>1</sub>**

ganumeita.

**zemniekmeitene<sub>1</sub>**

Meitene, kas aug zemnieku ģimenē; arī

## SYNSETS

**jaunekle<sub>1</sub>, jauniete<sub>1</sub>, meitene<sub>2</sub>, skuķis<sub>1,1</sub>,**

**meitēns<sub>2</sub>, jaunmeita<sub>1</sub>, mamzele<sub>1</sub>**

jaunekle<sub>1</sub> Jauniete.

jauniete<sub>1</sub> Sieviete vecumā starp pusaudzes un brieduma gadiem.

meitene<sub>2</sub> Jauniete.

skuķis<sub>1,1</sub> Nepieredzējusi, arī nenopietna jauniete.

## LINKS

### EXTERNAL LINKS:

**(n) girl, miss, missy, young\_lady,**

**young\_woman, fille**

a young woman; "a young lady of 18"

### HYPERONYMS:

**sieva<sub>2</sub>, sieviete<sub>1</sub>**

sieva<sub>2</sub> Sieviete.

# Results

- 5201 words, 4430 synsets, 2936 links
- Served to general public via Tezaurs.lv dictionary platform
- Available as open data via Clarin and on website [wordnet.ailab.lv](http://wordnet.ailab.lv)
- Basis for ongoing WSD tool development for Latvian
- Work on transferring data from other wordnets



# Thank you!

More information on [wordnet.ailab.lv](http://wordnet.ailab.lv)



This work was supported by the Latvian Council of Science, project “Latvian WordNet and word sense disambiguation”, project No. LZP-2019/1-0464 and National Research Programme “Letonika – Fostering a Latvian and European Society” project “Research on Modern Latvian Language and Development of Language Technology” (grant agreement No. VPP-LETONIKA-2021/1-0006).