WG4 report: 08.2015-03.2016

Lexicography and Lexicology from a Pan-European Perspective

Eveline Wandl-Vogt, Krzysztof Nowak
Members

- 81 members | 23 countries
Members

- 81 members | 23 countries
- 25% of the members come from COST Inclusiveness Countries
- Near Neighbouring Countries (Russia) newly in the COST ENeL group
  - already contributing and collaborating
- Gender ratio: 1,5:1 (females:males)
  similar like 2015.08
Members

WG 4 Membership: Bolzano-Vienna-Herstmonceux-Barcelona Progress

© Nowak
Collaborations

H2020-Project PARTHENOS
→ BCN meeting: Technical heterogeneity @ Session 2

DARIAH-ERIC WG Lexical Resources, WG Standards
→ BCN meeting: Technical heterogeneity @ Session 2

DARIAH-ERIC WG Defining Cloud Infrastructure Services for Digital Humanities
→ Case Study eLexicography: Congress Participation

European Citizen Science Association (ECSA)
„Citizen Science Infrastructure“, former H2020-project SOCIENTIZE
→ BCN meeting: Social heterogeneity @ Session 3
Short Term Scientific Missions

September 2015:

Nathalie Mederake: „Lexicographic Hybrids“
Host: Institute of Polish Language, Polish Academy of Sciences

Followed by visit in Paris to discuss Joint Proposal
Short Term Scientific Missions

planned 2016:

Survey on Citizen Science and Open Science for Lexicography

Visualisation Tools supporting Citizen Science in Lexicography
Results

Events 2015

Workshop on Biodiversity and Linguistic Diversity: Linked Open Data 4 Living Organisms

funded by DARIAH-EU grant „Open Humanities“

coop-organised by COST-ENeL partners, DARIAH-members and NHM; supported by EUROPEANA.AT, open knowledge AT

about 30 persons from about 15 nations

2 student grants
Results

Events 2016

Workshop on Collaboration and Computing for Under-Resourced Languages. „Towards an Alliance for Digital Language Diversity“ (CCURL)
1 day Pre-LREC Workshop 2016: Portoroz (SI)

Joint Meeting with Wg on Atlas Linguarum Europae (ALE)
Results

Events 2016

Workshop on Biodiversity and Linguistic Diversity – Follow up

Workshop on eInfrastructures for Lexicography @ EURALEX 2016
Congress Papers

- 2015.12: AT: Digital Humanities Austria. Conference 2015 (3 Posters, 1 just about COST ENeL)
- 2016.03: DE: Digital Humanities deutschsprachig (1 Paper, 2 Poster)
- 2016.04: CZ: Slavic Lexicography at the beginning of the 21st century (1 Paper)
- 2016.05: PL: Russian Language in the language and cultural space of Europe and the world: man, consciousness, communication, internet (1 Paper)
- 2016.05: DE: European Citizen Science Congress (submitted)
- 2016.06: Journées internationales d´Analyse statistique des Données Textuelles (1 Poster)
- 2016.07: PL: Digital Humanities (2 Posters)
- 2016.08: LT: EI Manuscript 2016 (1 Paper)
- 2016.09: HU: International Conference on Sociolinguistics (1 Paper)
- 2016.09: GE:Euralex2016 (4 Paper, 1 Workshop)
- 2016.09: UK. Progress in Colour Studies 2016 (1 Paper submitted)
- 2016.11: ES: Teem´2016 (1 Track)

/../

in 4 languages (EN, GE, RU, FR)
in 9 European countries (AT, DE, CZ, PL, LT, HU, GE,, UK, ES)
Congress Papers

- 2015.12: AT: Digital Humanities Austria. Conference 2015 (3 Posters, 1 just about COST ENeL)
- 2016.03: DE: Digital Humanities deutschsprachig (1 Paper, 2 Poster)
- 2016.04: CZ: Slavic Lexicography at the beginning of the 21st century (1 Paper)
- 2016.05: PL: Russian Language in the language and cultural space of Europe and the world: man, consciousness, communication, internet (1 Paper)
- 2016.05: DE: European Citizen Science Congress (submitted)
- 2016.06: Journées internationales d’Analyse statistique des Données Textuelles (1 Poster)
- 2016.07: PL: Digital Humanities (2 Posters)
- 2016.08: LT: El Manuscript 2016 (1 Paper)
- 2016.09: HU: International Conference on Sociolinguistics (1 Paper)
- 2016.09: GE:Euralex2016 (4 Paper, 1 Workshop)
- 2016.09: UK. Progress in Colour Studies 2016 (1 Paper submitted)
- 2016.11: ES: Teem’2016 (1 Track)

/...

Pan European Heritage
Data modeling, standards for dictionaries
Tools, Services for eLexicography
eInfrastructures, grid- and cloud services for eLexicography
Citizen science and Open science for eLexicography
New dictionary projects, New dictionary publications
New Workflows; interdisciplinary, open-online collaboration in eLexicography
Results

Dictionary Projects

1. Russian Dialectal Dictionary
   ILE, Russian Academy of Sciences (RU)
2. Russian Manuscripts
   PHd Project Kira Kovelenko (RU)
3. Witchi – Spanisch Dictionary
   Institut für Linguistik, University of Vienna (AT)
4. Dagaare – Cantonese – English Dictionary
   Institut für Afrikawissenschaften, University of Vienna (AT)
5. Hausa – Cantonese – English Dictionary
   Institut für Afrikawissenschaften, University of Vienna (AT)
6. Historical Viennese Dictionaries
Results

Dictionary Projects

Virtual Research Environment

1. Russian Dialectal Dictionary
   ILE, Russian Academy of Sciences (RU)
2. Russian Manuscripts
   PHd Project Kira Kovelenko (RU)
3. Witchi – Spanisch Dictionary
   Institut für Linguistik, University of Vienna (AT)
4. Dagaare – Cantonese – English Dictionary
   Institut für Afrikawissenschaften, University of Vienna (AT)
5. Hausa – Cantonese – English Dictionary
   Institut für Afrikawissenschaften, University of Vienna (AT)
6. Historical Viennese Dictionaries
Results

Other Dictionary Projects

piloting case studies

- Example entries „red“ & „anger“
  9 European languages

- Reconciliation service to access plant names of Common Names Service (Beta)
Reconciliation service

The service was developed at the Austrian Centre for Digital Humanities as a first step towards Biodiversity and Linguistic diversity research portal. It works on top of Common Names Service by Natural History Museum of Vienna.

<table>
<thead>
<tr>
<th>Reconcile one scientific name</th>
</tr>
</thead>
<tbody>
<tr>
<td>bells perennis</td>
</tr>
<tr>
<td>e.g. platalea leucorodia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reconcile list of names (csv import)</th>
</tr>
</thead>
</table>

| Get RDF |
Reconciliation service

The service was developed at the Austrian Centre for Digital Humanities as a first step towards Biodiversity and Linguistic diversity research portal. It works on top of Common Names Service by Natural History Museum of Vienna.

### Reconcile one scientific name

<table>
<thead>
<tr>
<th>Common name</th>
<th>Id</th>
<th>Language</th>
<th>Geography</th>
<th>Taxon</th>
<th>Taxon id</th>
<th>Score</th>
<th>Match</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>bellis perennis</td>
<td>88</td>
<td>Bolmål</td>
<td>None</td>
<td>Bellis perennis</td>
<td>3</td>
<td>100</td>
<td>True</td>
<td>artsdatabanken.no</td>
</tr>
<tr>
<td>tussilgd</td>
<td>89</td>
<td>Nynorsk</td>
<td>None</td>
<td>Bellis perennis</td>
<td>3</td>
<td>100</td>
<td>True</td>
<td>artsdatabanken.no</td>
</tr>
<tr>
<td>Daisy</td>
<td>422675</td>
<td>English</td>
<td>None</td>
<td>Bellis perennis</td>
<td>3</td>
<td>100</td>
<td>True</td>
<td>BIS, Marine Species, ZooBank</td>
</tr>
<tr>
<td>Glansblomchen</td>
<td>421194</td>
<td>None</td>
<td>None</td>
<td>Bellis perennis L.</td>
<td>23597</td>
<td>100</td>
<td>True</td>
<td>Dalla Torre, C. W. von (1895) Die volkstümlichen Pflanzenamen in Tirol und Vorarlberg: [1-3] 4-7 (8) 9-76.</td>
</tr>
<tr>
<td>Monate</td>
<td>421195</td>
<td>None</td>
<td>None</td>
<td>Bellis perennis L.</td>
<td>23597</td>
<td>100</td>
<td>True</td>
<td>Dalla Torre, C. W. von (1895) Die volkstümlichen Pflanzenamen in Tirol und Vorarlberg: [1-3] 4-7 (8) 9-76.</td>
</tr>
<tr>
<td>Monate</td>
<td>421196</td>
<td>None</td>
<td>None</td>
<td>Bellis perennis L.</td>
<td>23597</td>
<td>100</td>
<td>True</td>
<td>Dalla Torre, C. W. von (1895) Die volkstümlichen Pflanzenamen in Tirol und Vorarlberg: [1-3] 4-7 (8) 9-76.</td>
</tr>
<tr>
<td>Münne</td>
<td>421197</td>
<td>None</td>
<td>None</td>
<td>Bellis perennis L.</td>
<td>23597</td>
<td>100</td>
<td>True</td>
<td>Dalla Torre, C. W. von (1895) Die volkstümlichen Pflanzenamen in Tirol und Vorarlberg: [1-3] 4-7 (8) 9-76.</td>
</tr>
</tbody>
</table>
Reconciliation service

The service was developed at the Austrian Centre for Digital Humanities as a first step towards Biodiversity and Linguistic diversity research portal. It works on top of Common Names Service by Natural History Museum of Vienna.

<table>
<thead>
<tr>
<th>Reconcile one scientific name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>bellis perennis</td>
<td>Submit</td>
</tr>
<tr>
<td>e.g. platalca leucorodia</td>
<td></td>
</tr>
</tbody>
</table>

| Reconcile list of names (csv import) |  |
| Get RDF |  |
Results

Other Dictionary related Projects

piloting case studies

- Europeana.AT exhibition: Dictionaries as cultural heritage furthering national | European identity
- EGI ENGAGE – DARIAH Competence Centre: eInfrastructure for Cloud-Storage, publishing, editing, accessing of Historical Lexicographical Collections
Results

Citizen Science in Lexicography

Pilot Study
Exploring semantic networks via Mind Path (Socientize)
case study: bread
Results

Citizen Science in Lexicography

WHAT IS THE PATH FROM 'SMART' TO 'CITY'?
Results

Visual Analysis in Lexicography

Pilot Study
Exploring corpora and lexical resources via visual Tools
Results

Visual Analysis in Lexicography
## Results

### ExploreAT!

Type any string to generate its context treemap. The 10 contexts with the most lemma occurrences will show up in the treemap.

[Generate TreeMap]

<table>
<thead>
<tr>
<th>Contexts where lemmas containing &quot;per&quot; appear in</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorbus aucuparia; Eberesche, Vogelbeerbaum (25)</td>
<td>Vaccinium uliginosum; Moor-Heidelbeere, Rauschbeere (25)</td>
<td>Vaccinium myrtillus; Heidelbeere, Schwarzbeere, Blaubeere (25)</td>
</tr>
<tr>
<td>Vaccinium vitis-idaea; Preiselbeere (25)</td>
<td>Rubus idaeus;(Gew.) Himbeere (22)</td>
<td>Rosa canina; Hunds-Rose, Hecken-Rose (20)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Viburnum lantana; Wolliger Schneeball (18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Echt-Traubenkirsche (15)</td>
</tr>
</tbody>
</table>
Results

Visual Analysis
Results

Publications

Upcoming 2016:
Congress Papers Colours and Emotions
Alina Villalva & Team
Results

Proposals (H2020):

Contribution to ELEXIS proposal
Lead: Simon Krek (SI)
European Lexicographical Infrastructure

Contribution to ICARUS proposal
Lead: Emiliano Degl´Innocenti (IT)
Sources for European History
Results

Proposals:

ANR-DFG project on French-German Medieval Latin Dictionary

ERC Starting Grant project on Research Environment for Lexical Study of Latin Texts
Results

Proposals:

COST ENeL, DARIAH-EU & EGI
Public Humanities Grant
April, 7th

Citizen Science
Workplan

2016

1. Improving interlinking in the EDP
2. Citizen Science and Open Science Survey
3. Investigation on Pan-European information representation;
   investigation on cultural representation in dictionaries