Linguistic Atlases as a Source for Dialectal Dictionaries

Jožica Škofic
Fran Ramovš Institute of Slovenian Language, ZRC SAZU, Novi trg 4, 1000 Ljubljana
E-mail: guzej@zrc-sazu.si

COST ENeL meeting
Online Dictionaries, Dictionary Portals and Enhanced Search
Dictionary Portals and Interlinking content
Budapest, 24th–25th February 2017
Keywords:

• dialectology
• geolinguistics
• dialectal lexicography

• Slovene Linguistic Atlas
• European Linguistic Atlas

• e-atlas
• e-dictionary
Linguistic atlases and their data-bases

- **linguistic atlas** is a rich source of dialectal lexics
  - geolinguistical presentation
  - lexicographical presentation

- geographically edited dialectal **dictionary**

- **database**
  - Slovene linguistic atlas (SLA) with SlovarRed database, connected with GIS
  - European linguistic atlas (ALE) with no common database (so far).
Organization of dialectal lexics

• in linguistic atlas
  - the signified ➔ the signifier
  - »How do you say/call /…/?«
  - one map – one meaning – more words

• in dictionary
  - the signifier ➔ the signified
  - »What does it mean?«
  - one word – one/more meanings
How to create multi-dialectal e-dictionary - on the base of linguistic atlase(s)?

• a database with carefully morphologically analyzed dialectal lexics

• careful semantic explanation

• properly chosen grammatical and other markers
  – net-points and their geographical coordinates
  – dialect/language name
  – time of recording etc.

• links to commentaries, maps and indexes (in the atlase(s))
Example from SLA (239 standard lexems – cca. 4500 dialectal lexemes)

**fogolar** -ja m < *(fogolár)-b ← Friul. *[fogolár]* ‘fireplace’


2) **kitchen** (Slov. *kuhinja*), SLA 2/23 – *[fogolár]* T080 Špeter (LSM, TL, 1989)

3) **living room** (Slov. *dnevna soba*), SLA 2/24 – *[fogolár, fogolár]* T078 Podbonesec (TL, 1963), *[fogolár]* T081 Ošnije (TL, 1963), *[fuγoˈlar]* T082 Mirnik (DZK, 2006)

- see also [http://www.fran.si/iskanje?View=1&Query=fogolar](http://www.fran.si/iskanje?View=1&Query=fogolar)
First step – recording dialectal material

Jožica Škofic, COST ENeL, Budapest, 24th February 2017
Povprečna starost informantov v zapisih za SLA

Jožica Škofic, COST ENeL, Budapest, 24th February 2017
Jožica Škofic, COST ENeL, Budapest, 24th February 2017
Jožica Škofic, COST ENEl, Budapest, 24th February 2017
Second step – database (SlovarRed, prepared by ing. Tomaž Seliškar)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>€</td>
<td>Euro</td>
</tr>
<tr>
<td>£</td>
<td>Pound</td>
</tr>
<tr>
<td>¥</td>
<td>Yen</td>
</tr>
<tr>
<td>₹</td>
<td>Rupee</td>
</tr>
<tr>
<td>₪</td>
<td>Lira</td>
</tr>
<tr>
<td>$</td>
<td>Dollar</td>
</tr>
<tr>
<td>₿</td>
<td>Bitcoin</td>
</tr>
<tr>
<td>₣</td>
<td>Franc</td>
</tr>
<tr>
<td>¥</td>
<td>Yen</td>
</tr>
<tr>
<td>₪</td>
<td>Lira</td>
</tr>
<tr>
<td>$</td>
<td>Dollar</td>
</tr>
<tr>
<td>₿</td>
<td>Bitcoin</td>
</tr>
<tr>
<td>₣</td>
<td>Franc</td>
</tr>
</tbody>
</table>

Jožica Škofic, COST ENeL, Budapest, 24th February 2017
Komentar in karta: 2/23

KOMENTAR

1. Gradivo

2. Morfološka analiza

3. Posebnosti kartiranja

Jožica Škofic, COST ENeL, Budapest, 24th February 2017

4. Uporabljena dodatna literatura

5. Primerjaj

6. Etimološka osvežitev

Kot tretji odgovor v posameznih točki so z znakovom za komentar kartirani leseci *ata* in *hata* v T049 ter *kuhinja* v T087.

Jožica Škofic, COST ENeL, Budapest, 24th February 2017
Fourth step – cartography
Team

- linguists – dialectologists and comparative linguists
- ethnologist
- IT specialist – language technology
- others: data-base, geography

Jožica Škofic, COST ENeL, Budapest, 24th February 2017
Fifth step – connecting with dictionary portal

Jožica Škofic, COST ENeL, Budapest, 24th February 2017

G
• **fogolar** 2/23 (V146B.01 ‘*kitchen*’), 2/24 (V131a.01 ‘*living room*’), 2/41 (V147B.01 ‘*fireplace*’) ... prepared by J. Škofic

• **XML** ... prepared by dr. Jernej Vičič (language technologist) ➔ automatically generate **HTML** presentation from XML data (server implemented by AMEBIS for www.fran.si)

```xml
<geslo>
  <iztocnica>fogolar</iztocnica>
  <komentar_in_karta>
    <oznaka>2/23</oznaka>
    <vprasanje>V146B.01</vprasanje>
    <opis>kuhinja</opis>
  </komentar_in_karta>
  <komentar_in_karta>
    <oznaka>2/24</oznaka>
    <vprasanje>V131a.01</vprasanje>
    <opis>dnevna soba (hiša)</opis>
  </komentar_in_karta>
  <komentar_in_karta>
    <oznaka>2/41</oznaka>
    <vprasanje>V147B.01</vprasanje>
    <opis>ognjišče</opis>
  </komentar_in_karta>
</geslo>
```
Sixth step – e-atlas

Cooperation:

dialectologist (idea) → language technologist (technological realisation)

interactive e-linguistic atlas

- crowd-sourcing (citizen science)
- linking with different humanistic open sources (digital humanities)

- exportation of data layers from individual maps to ArcGIS server
- Adobe Flash web application presents data on a set of web maps
- the application as a portal for additional data about the location and presented linguistic data
Seventh step – e-dictionary

• How to prepare a dictionary?

  – Which information are missing?
  – How to rearrange the data-base, prepared for linguistic atlas? How to connect it with „dictionary tools“?
  – How to prepare linguistic data?
  – e-dictionary or (just) dictionary or maybe only glossary?
Example from SLA

• fogolar -ja m < *(fogolar)-b ← Friul. [fogolâr] ‘fireplace’

1) fireplace (Slov. ognjišče), SLA 2/41

2) kitchen (Slov. kuhinja), SLA 2/23


• prepared in Word
• in data-base (Access) / e-SLA
• missing in database

complete the data-base with data from archive (when exist!)
Example - words for „spruce“ in European languages and their dialects

Data-base, linguistic analysis and commentary:

- Gradivo\ALE\ALE QI-050a_gradivo.pdf
- Gradivo\ALE\Spruce database_15 6 2016.xlsx
- Gradivo\ALE\ALE_050_Epicea JS 2016 2.ppt
- Gradivo\ALE\Clanek ALE050 2016 JS 4.pdf

Cartography:

- symbols: Gradivo\ALE\Legends and symbols.pdf
- matrix:
  - Gradivo\ALE\Map Digitalisation Software Instructions.pdf
  - Gradivo\ALE\Albania.xls
  - Gradivo\ALE\5VATER.pdf

Jožica Škofic, COST ENeL, Budapest, 24th February 2017
ALE: e-atlas, e-dictionary?

• How to prepare e-atlas?
  – How to prepare a simple but useful data-base?

• How to prepare e-dictionary?
  – Which information are missing? Is it possible to get additional linguistic data?
  – How to rearrange the existing data-base, prepared for ALE (Excel)? How to connect it with „dictionary tools (ILEX ...“?
  – How to arrange the whole entry, head-word and grammatical information about it?
  – What about the context of the word and its use (no data about it)?
  – Is it possible to connect dialectal dictionary, prepared from atlas-database, with other standard or dialectal dictionaries ...?
  – ALE dictionary or „just“ ALE glossary?

• EU-project (ENeL, eLex ...)?!
Hvala!
Thank you!
Köszönöm!