Computational Linguistics + Data Visualization
Towards the Interactive Exploration of Newspaper Data

Rachele Sprugnoli
Digital Humanities Research Group
Fondazione Bruno Kessler (Trento, Italy)
Computational Linguistics
Computational Linguistics and Newspaper Data

- News: the most extensively investigated domain in CL
- Many annotation standards, many annotated corpora
- Example: temporal information

### Contemporary News

Sprugnoli and Tonelli, *One, No One and One Hundred Thousand Events: Defining and Processing Events in an Inter-Disciplinary Perspective*, NLE Journal, forthcoming

<table>
<thead>
<tr>
<th>Domain</th>
<th>Corpus</th>
<th>Lang</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE 2005 (training)*</td>
<td></td>
<td>EN</td>
</tr>
<tr>
<td>French TimeBank (Rittar et al., 2011)</td>
<td></td>
<td>FR</td>
</tr>
<tr>
<td>Romanian TimeBank (Forascu and Tufi, 2012)</td>
<td></td>
<td>RO</td>
</tr>
<tr>
<td>TimeBankPT (Forascu and Tufi, 2012)</td>
<td></td>
<td>PT</td>
</tr>
<tr>
<td>Persian TimeBank (Yaghoobzadeh et al., 2012)</td>
<td></td>
<td>FA</td>
</tr>
<tr>
<td>Catalan TimeBank 1.0*</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>Spanish TimeBank 1.0*</td>
<td></td>
<td>ES</td>
</tr>
<tr>
<td>BCCWJ-TimeBank (Asakawa, 2015)</td>
<td></td>
<td>JA</td>
</tr>
<tr>
<td>EVENTI corpus (Casali et al., 2017)</td>
<td></td>
<td>IT</td>
</tr>
<tr>
<td>TempEval 1 (training)</td>
<td></td>
<td>EN</td>
</tr>
<tr>
<td>TempEval 1 (test)</td>
<td></td>
<td>EN</td>
</tr>
<tr>
<td>TempEval-3 (AQUAINT+TimeBank+Platinum) (Sauri and Pustejovsky, 2009)</td>
<td>EN</td>
<td></td>
</tr>
<tr>
<td>2012 EventCorefBank (ECB) (Lee et al., 2012)</td>
<td>EN</td>
<td></td>
</tr>
<tr>
<td>ECB+ (Cybulskia and Vossen, 2014)</td>
<td></td>
<td>EN</td>
</tr>
<tr>
<td>Light ERE (Mott et al., 2016)</td>
<td></td>
<td>EN</td>
</tr>
<tr>
<td>Rich ERE (Mott et al., 2016)</td>
<td></td>
<td>EN</td>
</tr>
<tr>
<td>Event Nugget (training+test) (Mitamura et al., 2015)</td>
<td>EN</td>
<td></td>
</tr>
<tr>
<td>TimeLine (Minard et al., 2015)</td>
<td></td>
<td>EN</td>
</tr>
<tr>
<td>MEANTIME (Minard et al., 2016)</td>
<td></td>
<td>IT, ES, NL</td>
</tr>
</tbody>
</table>
Data Visualization

Minard, 1861
Data Visualization and Newspaper Data

- Explore phenomena through visual models
- Example: “Nuovi Linguaggi” in Corriere della Sera
Why CL + DV?

- Computational linguistics (CL) and data visualizations (DV) together can help explore and analyze linguistic data.
- In particular, they can improve:
  - the navigation through digital collections
  - the display of complex statistical and semantic analyses
The Challenge

- Develop online platforms to search and visualize (newspaper) data in an interactive and intuitive way
- Many issues to deal with:
  - Scalable reading
The Challenge

- Develop online platforms to search and visualize (newspaper) data in an interactive and intuitive way
- Many issues to deal with:
  - Flexibility
The Challenge

- Develop online platforms to search and visualize (newspaper) data in an interactive and intuitive way

- Many issues to deal with:
  - Interactivity
The Challenge

- Develop online platforms to search and visualize (newspaper) data in an interactive and intuitive way

- Many issues to deal with:
  - The time dimension
The Challenge

▪ Develop online platforms to search and visualize (newspaper) data in an interactive and intuitive way

▪ Many issues to deal with:
  ○ Users’ needs
The Challenge

- Develop online platforms to search and visualize (newspaper) data in an interactive and intuitive way

- Many issues to deal with:
  - Ease of use
The Challenge

- Develop online platforms to search and visualize (newspaper) data in an interactive and intuitive way

- Many issues to deal with:
  - Adaptation of CL techniques
ALCIDE architecture

http://celct.fbk.eu:8080/Alcide_Demo/
Two projects
One (last???) issue: evaluation

Three types of evaluation:

1. of the methodology: include expert users in the development loop
2. of the computational linguistics tools: participate to evaluation campaigns
3. of the platform: usability tests (*System Usability Scale*), task-based evaluations
THANK YOU!

Email: sprugnoli@fbk.eu
Web Site: http://dh.fbk.eu
Twitter: https://twitter.com/DH_FBK