Computational Linguistics
+
Data Visualization

Towards the Interactive Exploration of Newspaper Data

Rachele Sprugnoli

Digital Humanities Research Group
Fondazione Bruno Kessler (Trento, Italy)

FONDAZIONE
BRUNO KESSLER

DH
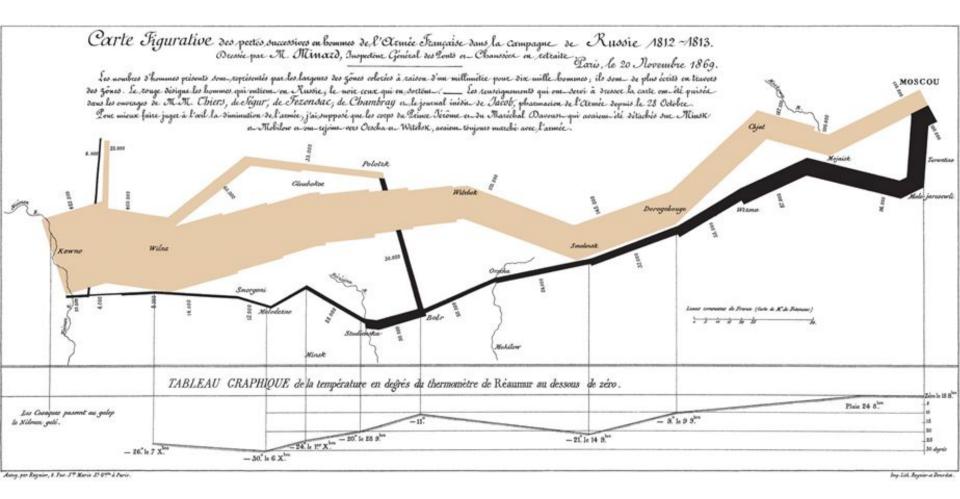DIGITAL
HUMANITIES

Computational Linguistics

Computational Linguistics and Newspaper Data

- News: the most extensively investigated domain in CL

- Many annotation standards, many annotated corpora

- Example: temporal information

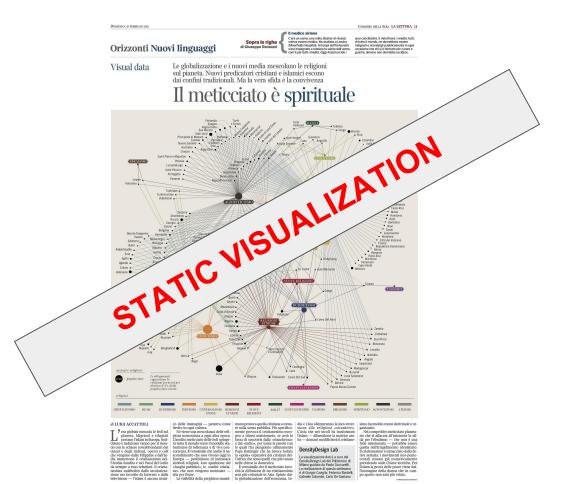| Domain	Corpus	Lang
	ACE 2005 (training)[a]	EN ZH
	French TimeBank (Bittar et al., 2011)	
	Romanian TimeBank (Forascu and Tufi, 2012)	
	TimeBankPT (Forascu and Tufi, 2012)	
	Persian TimeBank (Yaghoobzadeh et al., 20)	FA
	Catalan TimeBank 1.0[b]	CA
	Spanish TimeBank 1.0[c]	ES
	BCCWJ-TimeBank (Asaha)	JA
NEWS	EVENTI corpus (Casel)	IT
	TempEval 1 (traini)	EN
		ZH EN
	TempE g+test)[e]	IT FR KO ES
	val-3 (AQUAINT+TimeBank+Platinum)[f]	EN
	ctBank (Saurí and Pustejovsky, 2009)	EN
	2012 EventCorefBank (ECB) (Lee et al., 2012)	EN
	ECB+ (Cybulska and Vossen, 2014)	EN
	Light ERE[g](Mott et al., 2016)	ZH[h] EN
	Rich ERE[g](Mott et al., 2016)	ZH[h] ES
	Event Nugget (training+test) (Mitamura et al., 2015)	EN
	TimeLine (Minard et al., 2015)	EN
	MEANTIME[i](Minard et al., 2016)	EN IT ES NL

CONTEMPORARY NEWS

Data Visualization

Minard, 1861

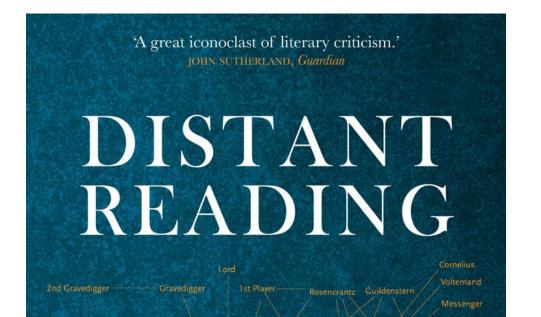# Data Visualization and Newspaper Data

- 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:
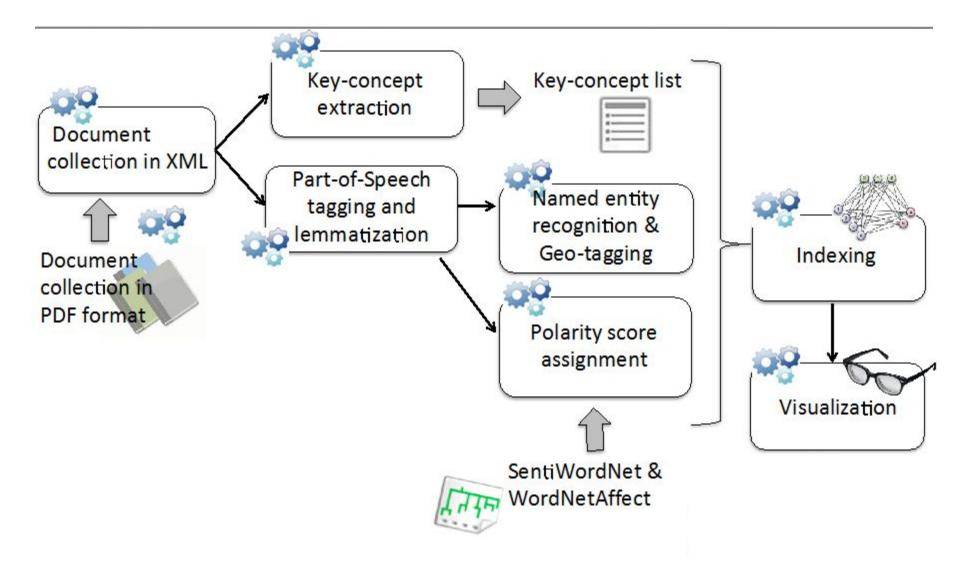
  - **Ease of use**

The Challenge

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

- Many issues to deal with:
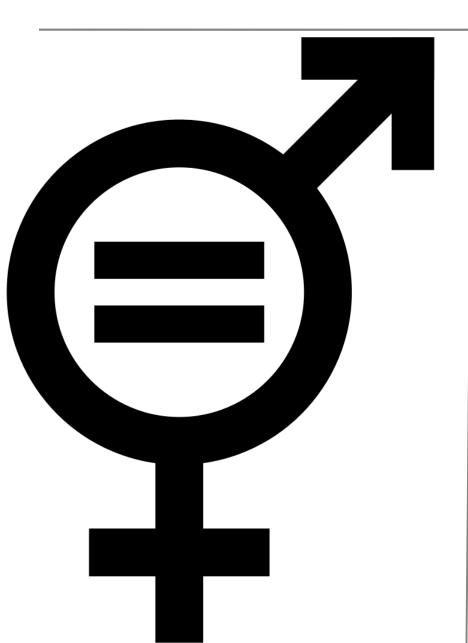
  - **Adaptation of CL techniques**

ALCIDE architecture
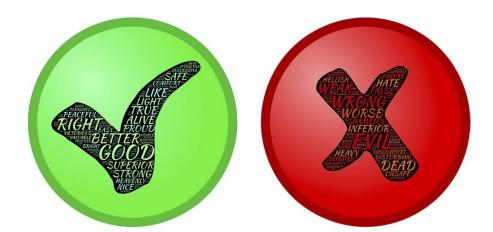

http://celct.fbk.eu:8080/Alcide_Demo/

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