

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

Domain	Corpus	Lang
NEWS	ACE 2005 (training) ^a	EN
	French TimeBank (Bittar et al., 2011)	ZH
	Romanian TimeBank (Forascu and Tufi, 2012)	RO
	TimeBankPT (Forascu and Tufi, 2012)	PT
	Persian TimeBank (Yaghoobzadeh et al., 2008)	FA
	Catalan TimeBank 1.0 ^b	CA
	Spanish TimeBank 1.0 ^c	ES
	BCCWJ-TimeBank (Asahara et al., 2007)	JA
	EVENTI corpus (Caselli et al., 2007)	IT
	TempEval 1 (training)	EN
		ZH
		EN
		IT
	TempEval 1 (training+test) ^c	FR
		KO
		ES
	TempEval-3 (AQUAINT+TimeBank+Platinum) ^f	EN
	EventBank (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	
	EN	
	IT	
MEANTIME ⁱ (Minard et al., 2016)	ES	
	NL	

CONTEMPORARY NEWS

Data Visualization

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.
 Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largueurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Légar, de Fezensac, de Chambrey et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps de Prince Jérôme et du Maréchal Davoust, qui avaient été détachés sur Minsk et Mohilew et qui se joignent vers Orscha à Wittelsk, avaient toujours marché avec l'armée.

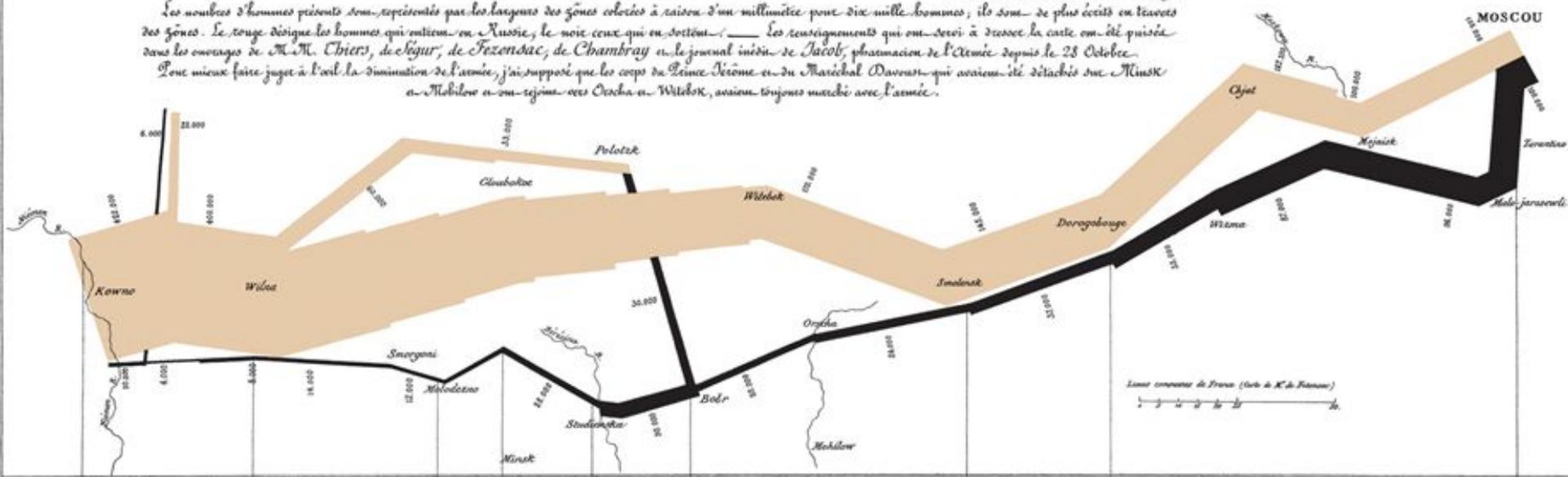
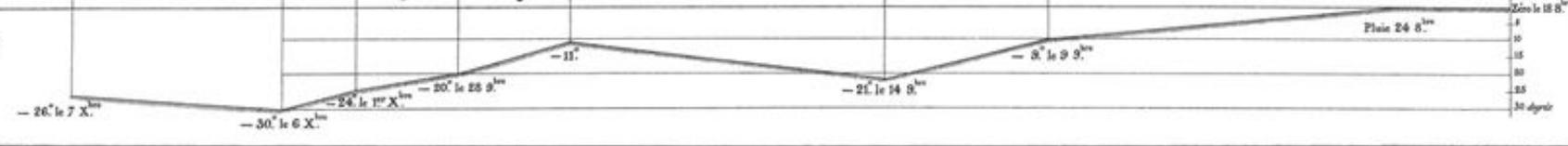


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop le Nilman gelé.



Antiq. par Regnier, 8. Par. 3^e Mars 27 69^e à Paris.

Imp. Lith. Regnier et Bretonnet.

Minard, 1861

Data Visualization and Newspaper Data

- Explore phenomena through visual models
- Example: “Nuovi Linguaggi” in Corriere della Sera



di LUIGI ACCATTOLI

La globalizzazione mescola le fedi sul pianeta. Migranti e rifugiati portano fedi in Europa, India e Indonesia vanno per il mondo con le schiere non abbondonano dei cinesi e degli indiani, opera e così vengono dalle Filippine e dall'Asia. In Europa il cristianesimo indiano e la sua influenza è in crescita. L'Arabia Saudita e nel Paese del Golfo da sempre è esso islamico. Il cristianesimo indiano è dalla secolarizzazione ma favorito da Internet e dalla televisione — l'Islam è ancora domi-

no delle immagini — penetra come brezza in ogni cultura.

Ne viene una miscelazione delle religioni sconosciute a ogni altra epoca. Un'idea mistica delle fedi spinge in tutto il mondo verso il modello di un'immersione che è tolleranza e di via comune. È evidente che anche il riconoscimento che esse vivono oggi in Europa — l'islamismo di Indonesia, il buddismo pubblico, le etnie cristiane che non vengono nominate — sta per finire.

La visibilità della preghiera musulmana provoca quella cristiana a tornare sulla scena pubblica. Più specificamente provoca il cristianesimo europeo a ritorni missionario, se sarà la forza di scuotere dalla secolarizzazione e dal declino, per usare le parole con le quali The papagallo ultimamente papa Ratzinger che ha invece indicato la spinta espansiva dei cristiani del XXI secolo che sono quelli che più vanno nelle chiese la domenica.

È evidente che il meticcio favorirà la diffusione di un cristianesimo non più coloniale in Asia. Segue dalla globalizzazione dell'economia, l'Asia e Cina affermeranno la loro avvezione alle religioni orientali.

L'Asia che nel secolo ha trasformato l'Islam — gliandone la matrice sebica — domani modificherà il cristianesimo facendolo meno dottrinale e organizzativo.

Nel competitivo meticcio planetario che si affaccia all'orizzonte la sfida per l'Europa — che non è una feda religiosa — potrebbe essere quella dell'irrigamento identitario. Il cristianesimo è riconosciuto dalla deriva settaria i movimenti neo pentecostali stanno già meticciosamente parlando nella Chiesa protestante. Per l'Asia la prova delle pene viene dall'immagine della donna che in campo aperto non sarà più violata.

DensityDesign Lab

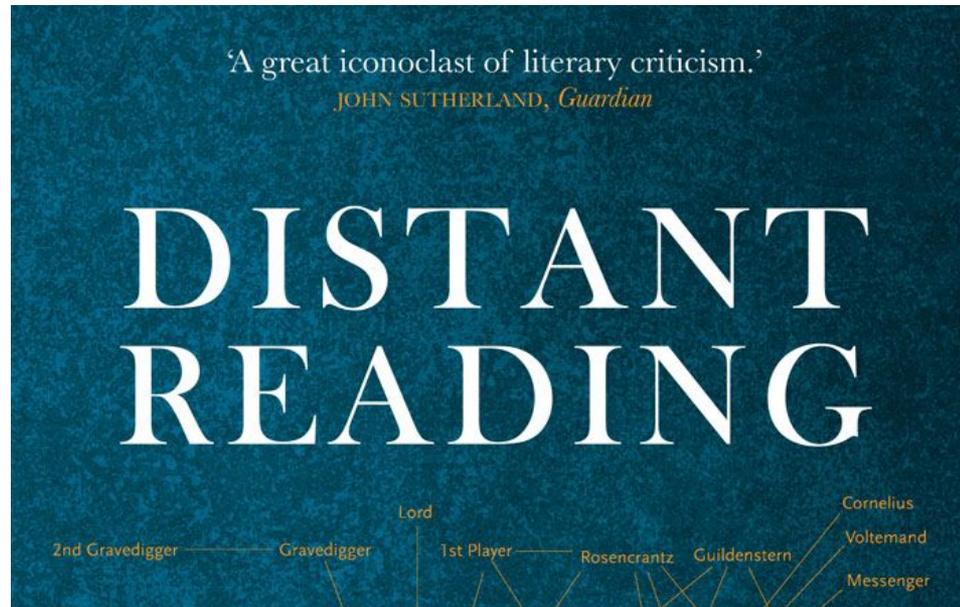
La visualizzazione dati è a cura del DensityDesign Lab del Politecnico di Milano guidato da Paolo Cuccarini. La realizzazione di questa settimana è di Giorgio Gagliardi, Federico Bartolotti, Gabriele Colombo, Carlo De Gaetano.

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**

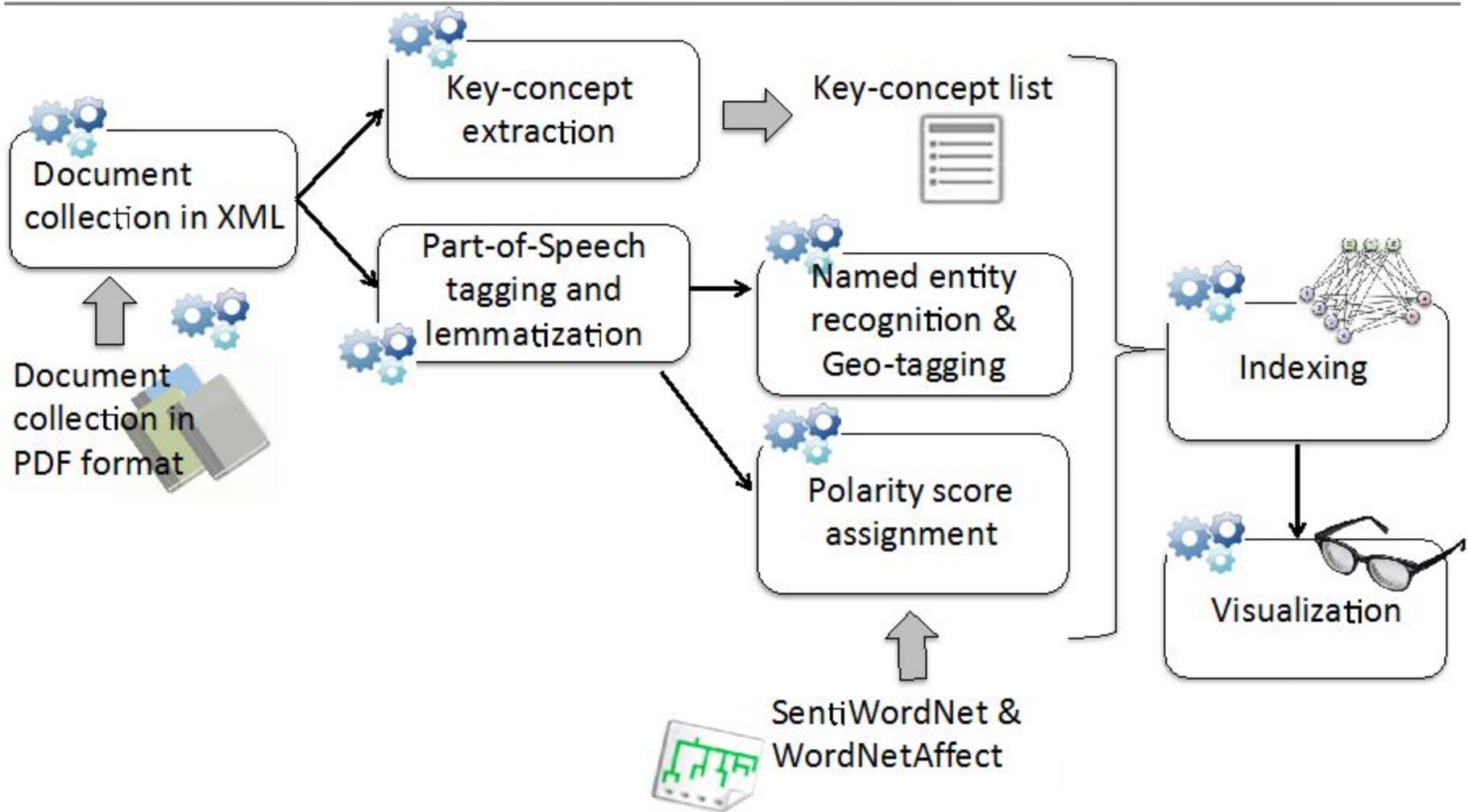


The Challenge

- Develop online platforms to search and visualize (newspaper) data in an interactive and intuitive way
- Many issues to deal with:
 - **Adaptation of CL techniques**

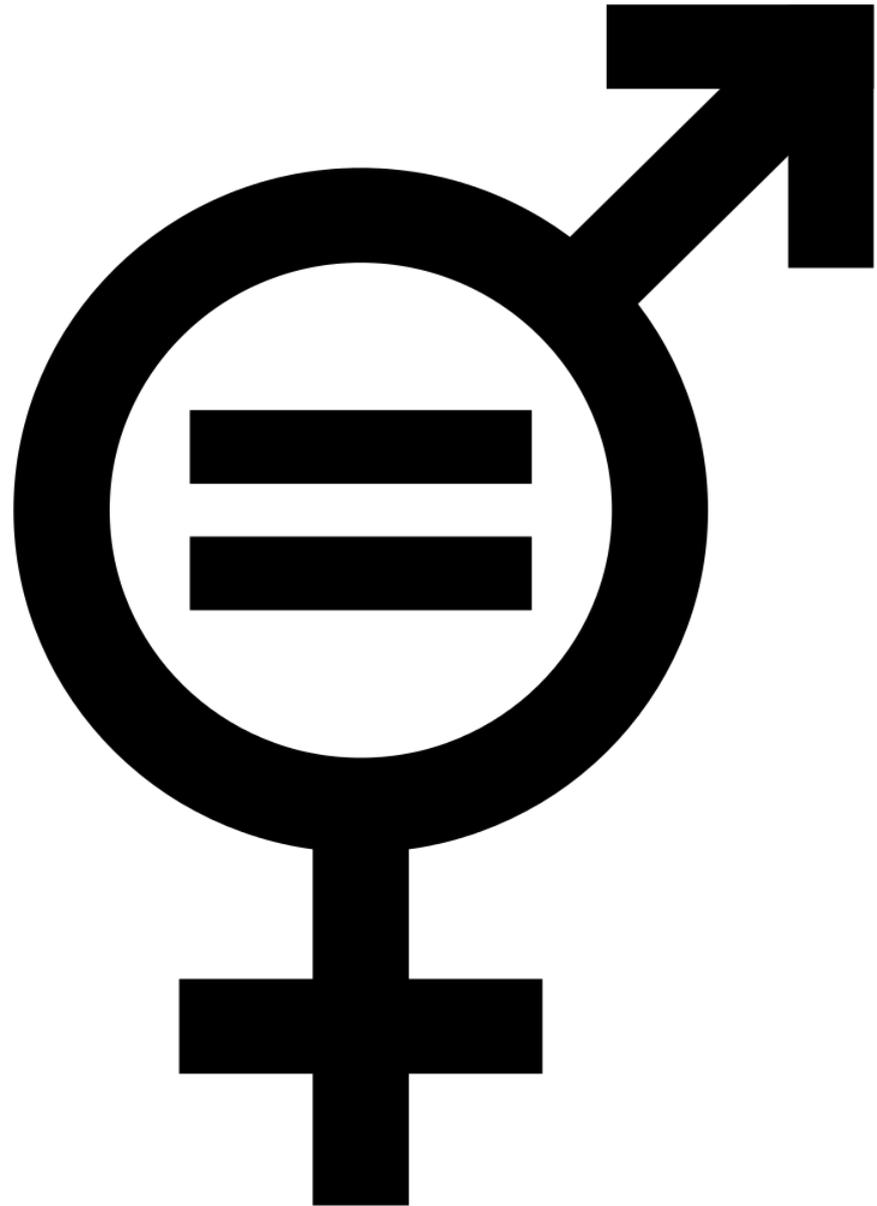


ALCIDE architecture



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