



Sentiment Analysis for the Humanities: the Case of Historical Texts

*Alessandro Marchetti, **Rachele Sprugnoli**, Sara Tonelli*
Digital Humanities Joint Research Project – <http://dh.fbk.eu>
Fondazione Bruno Kessler, Trento

Sentiment Analysis (SA)

“Computational treatment of opinion, sentiment and subjectivity in text” Pang and Lee (2008)

- A popular research topic in NLP, text mining, and Web mining in recent years

Social Media



News



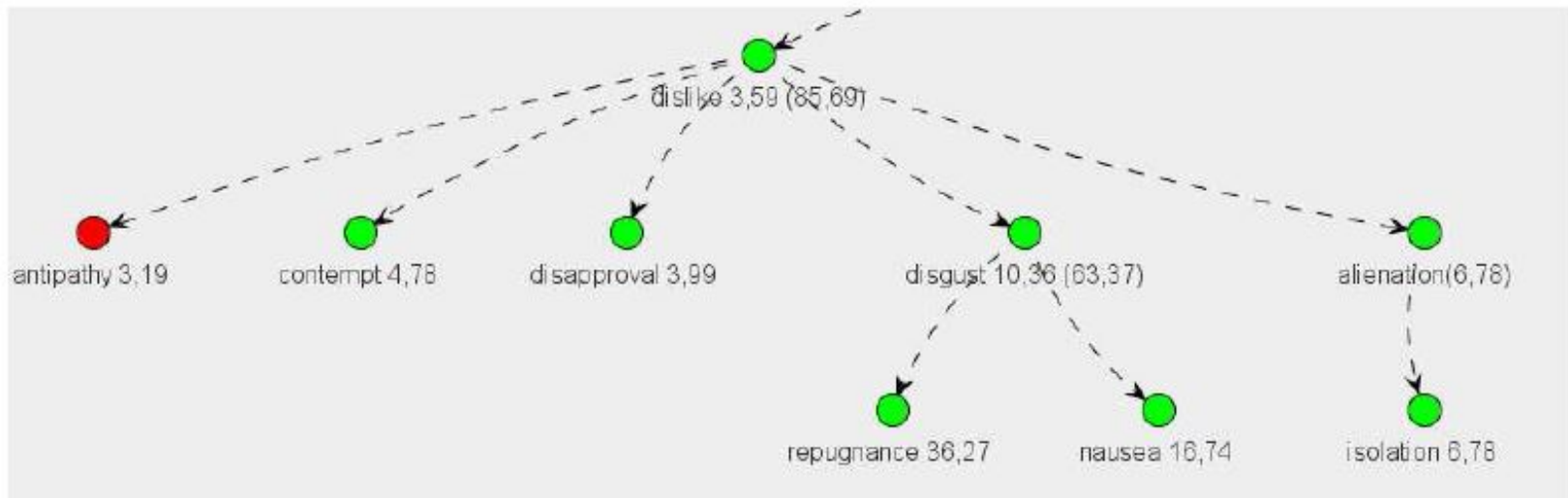
Customer Reviews



Sentiment Analysis in the Humanities

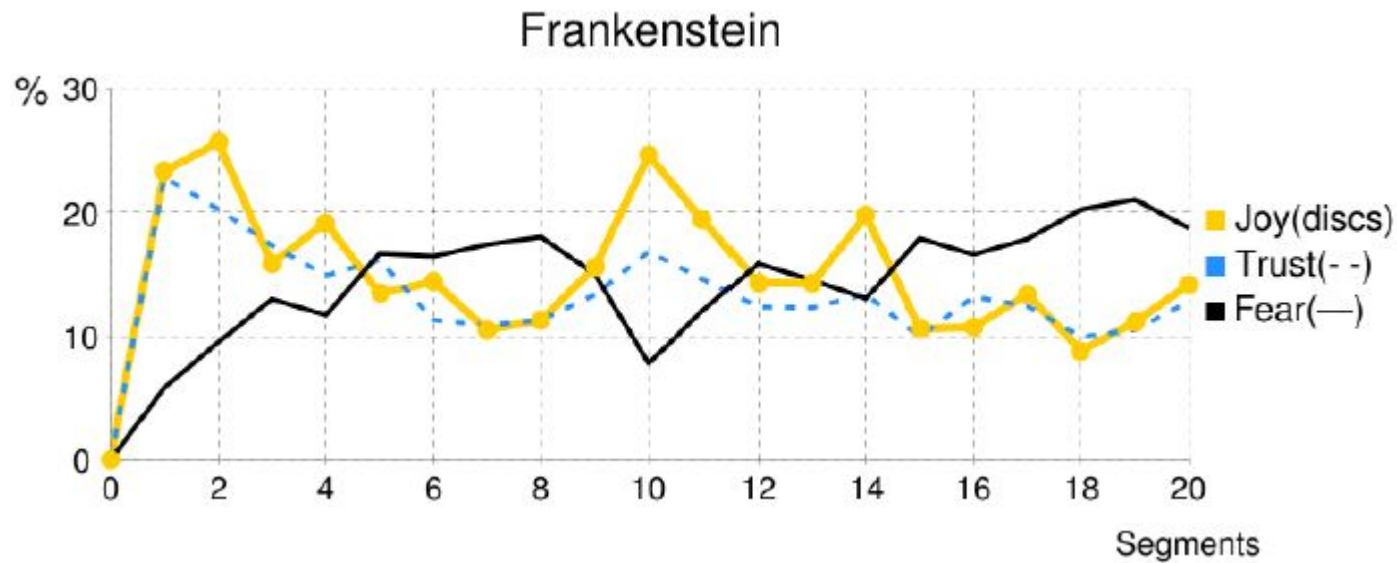
- Some applications on literary research:
 - Kakkonen and Kakkonen (2011)
 - Mohammad (2011)
 - Heuser and Le-Khac (2012)

SentiProfiler



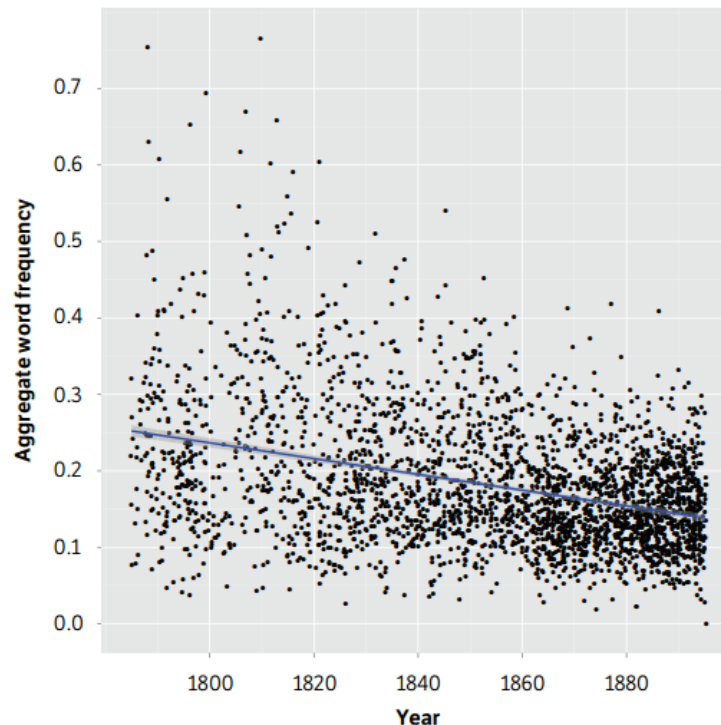
Sentiment Analysis in the Humanities

- Some applications on literary research:
 - Kakkonen and Kakkonen (2011)
 - Mohammad (2011)
 - Heuser and Le-Khac (2012)



Sentiment Analysis in the Humanities

- Some applications on literary research:
 - Kakkonen and Kakkonen (2011)
 - Mohammad (2011)
 - Heuser and Le-Khac (2012)



Prior vs. Contextual Polarity

- **Prior polarity:** the sentiment a term evokes out of context
 - Polarity lexica: each word associated with its polarity score
 - Positive: *beautiful, amazing*
 - Neutral: *Italian, general*
 - Negative: *bad, poor*
 - Key linguistic feature of ML approaches to SA
 - No available lexicon for Italian
- **Contextual Polarity:** the sentiment a term evokes according to its syntactic, semantic or pragmatic context
 - *they fought a **terrific** battle*
 - *I loved the film, it was **terrific***

Approaches to Polarity Assignment

1. Manual Annotation
2. (Semi-)Automatic Mapping
3. Crowdsourcing Annotation

“Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task”

Estellés-Arolas and González-Ladrón-De-Guevara (2012)

SA on Historical Texts at FBK

- Part of our research on the adaptation of **Human Language Resources and Technologies** to texts of late-modern and contemporary history
- Collaboration with the *Italian-German Historical Institute* in Trento
- SA has been identified as notably relevant to:
 - quantify the **general sentiment** of single document
 - allow **search** based on sentiment
 - track the attitude towards a **specific concept or entity over time**

SA on Historical Texts at FBK

- To be integrated in ALCIDE (Analysis of Language and Content In a Digital Environment)
- Case Study: Complete collection of Alcide De Gasperi's writings
 - 3K documents
 - 3million words
 - 1901 – 1954

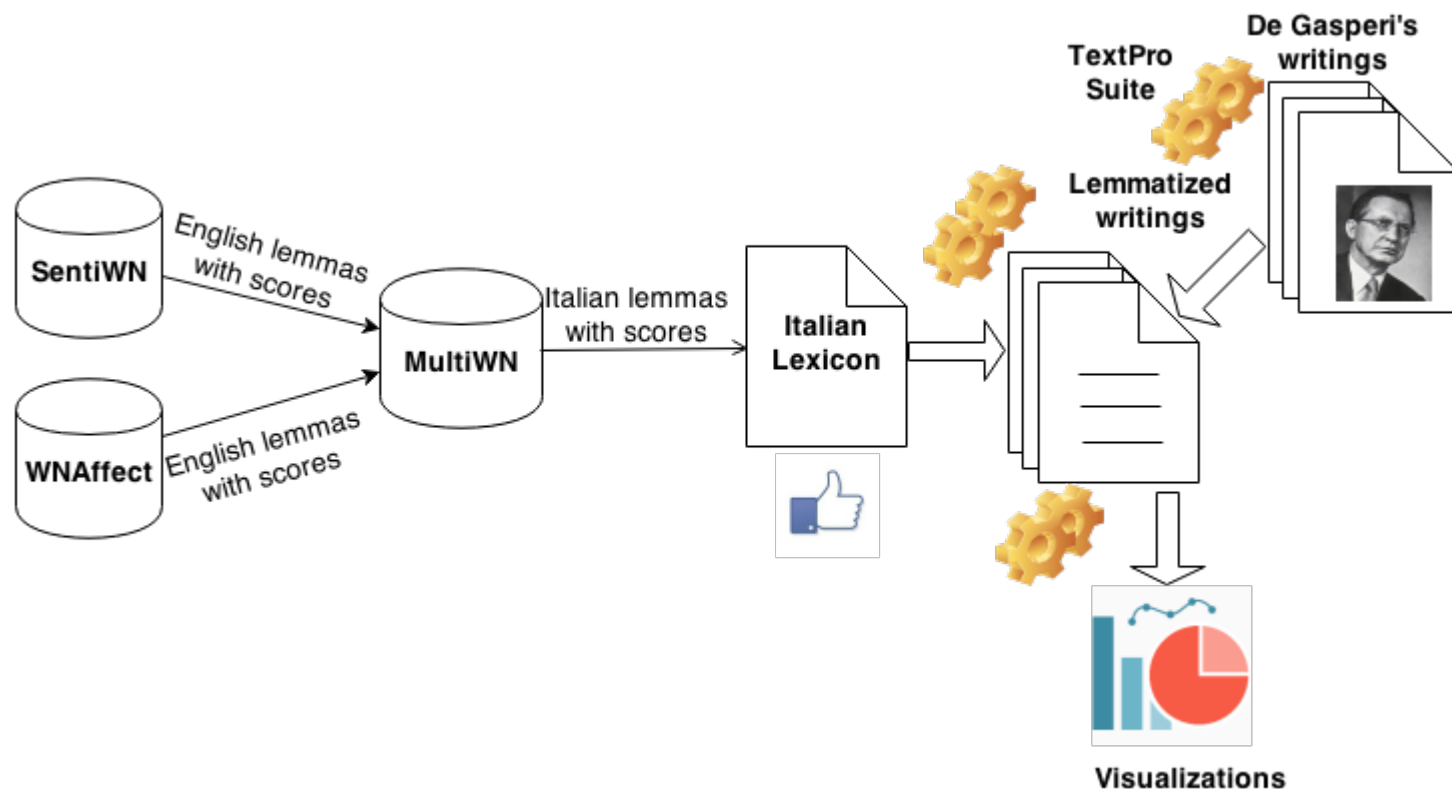
FIRST STEP: 2 experiments



Prior Polarity Experiment

RESEARCH QUESTIONS:

- how lexical resources built on contemporary languages can deal with historical texts?
- WordNetAffect, Strapparava and Valitutti (2004)
- SentiWordNet 3.0, Baccianella and Sebastiani (2010)



Prior Polarity Experiment: some Numbers

- Lemmas in De Gasperi's writings: 70,178
 - after excluding lemmas that can't have a polarity: 36,304
 - the lexicon covers 14,874 lemmas, i.e. **40.97%**
- 14,874 lemmas out of which
 - 9,650 are neutral (score = 0)
 - 5,224 lemmas have a polarity score:
 - 449 with an absolute positive score (score = 1)
e.g. 'eccellente' / *excellent*
 - 576 with an absolute negative score (score = -1)
e.g. 'affranto' / *broken-hearted*
 - the others with intermediate scores
e.g. 'intellettuale' / *intellectual* score = 0.875

Prior Polarity Experiment: visualization

Radiomessaggio

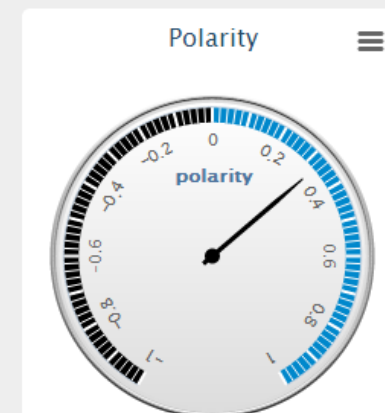
Roma, 1950-04-8

È una vigilia pasquale piena di serenità e di speranza che viviamo oggi a Roma. Popoli di ogni razza e di ogni colore vi si incontrano come fratelli e le antiche «meraviglie di Roma» (mirabilia Romae) sembrano quasi rinnovarsi nel colore dell'ammirazione universale ed esprimere più solennemente il senso della nostra civiltà millenaria. In questo clima il nostro spirito si dilata, al di sopra delle quotidiane contese; e si vorrebbe dire e pensare soltanto di cose buone e di opere di pace. Perché ci tormentiamo sempre nella nostra travagliata vita pubblica con le immagini del male, perché la cronaca della delinquenza ha tanto spazio nei giornali, perché la stampa illustrata esibisce, esaltandola, la vita lussuosa e degenerata di poche centinaia di dissipatori, quasi a sfida della povertà e della sofferenza, perché non ci si conforta invece almeno oggi col pensiero che questo popolo italiano, pur nello sforzo della ricostruzione, prova ancora ad accrescere notevolmente la sua opera di solidarietà verso i deboli, gli ammalati, i colpiti dalla sventura? È vero, la via verso la giustizia sociale cioè verso una più equa distribuzione della proprietà e del reddito è ancora lunga; leggi e costumi vi potranno provvedere solo per gradi; ma intanto, ove mancano e non è ancora raggiungibile la giustizia distributiva, supplisce in qualche misura l'assistenza collettiva e privata. Ricordiamo alcune cifre, per confortarci e fare meglio. Le spese sostenute dallo Stato per la beneficenza e per l'assistenza sociale nel 1938 ascendevano a 783 milioni, hanno raggiunto nel 1948 la cifra di 85 miliardi e 205 milioni, ossia pur tenendo conto della svalutazione le erogazioni dello Stato sono più che triplicate. Gli Istituti di ricovero (orfanotrofi, alberghi popolari, ecc.) al 31 maggio 1948 erano 6.607 con circa mezzo milione di ricoverati. E badate, mentre nel periodo dal 1920 al 1943 di nuovi istituti ne erano stati fondati solo 1.459, nei soli quattro anni dal 1944 al 31 maggio 1948 se ne fondarono altri 698. I refettori erano 17.811 con 2.021.893 assistiti: quasi il 90% dei Comuni risultavano forniti di refettori popolari. Di questi 17.811, dal 1944 al 1948 ne sono stati fondati ben 7.654. Quanti altri settori della nostra comunità nazionale (scuole, istituti economici, opere pie, associazioni religiose) costituiscono degli scompartimenti vitali in

Persons Keywords Polarity Tags



Text Classification



Prior Polarity Experiment: visualization

Graph

La marcia di...Sfruz

Trento, 1923-03-30

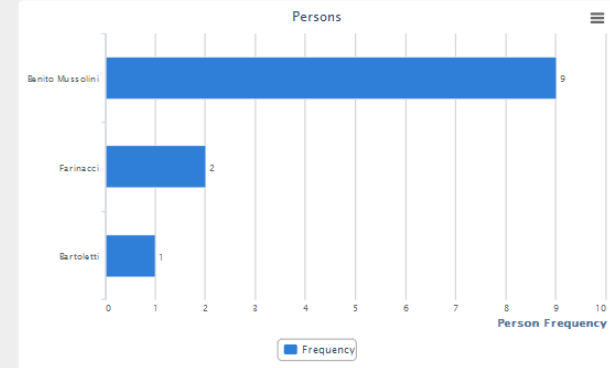
Quando Mussolini o i suoi generali richiamano alla disciplina e condannano con gradita precisione le violenze degli eroi della sesta giornata, nella stampa fascista e nelle riunioni di partito si risponde con un atto di sottomissione: qui e là però gli avvenimenti ed il linguaggio che riprende poco dopo ci dicono che la sottomissione è formale, ma non nello spirito. Tuttavia che ciò accada nelle provincie per le quali la marcia di Roma fu o doveva essere la fine di tutto un lungo periodo di conflitti sanguinosi, è ancora spiegabile, perché l'uso ripetuto e prolungato della violenza produce una perturbazione delle coscienze così profonda che gli stessi esaltatori d'un metodo di eccitazione da loro ritenuto necessario ma transitorio, né d'anzi prevedevano né oggi sanno dominare. Ma tale spiegazione e il compatimento che ne deriva non si confanno a regioni che del fascismo hanno ignorata la epoca cruenta. Nella Venezia Tridentina, se si eccettuano le due imprese di Bolzano e di Trento, le quali ebbero stile e sviluppo militare e carattere antistatale, ma non furono rivolte contro l'incolumità delle persone, o di un gruppo di persone, nessun episodio grave si ebbe a deplorare, né per difesa né per offesa. Le risse domenicali non insanguinarono né i centri né i villaggi della provincia, cosicché la tradizione d'ordine, di quiete tolleranza, di rispetto alla legge, non subì interruzioni notevoli. Lo spirito di violenza non è quindi nel Trentino né endemico né epidemico. Se tuttavia qui e là fa capolino anche fra noi, vuol dire che vi viene importato di proposito e forzando il naturale sviluppo delle cose. Vestigia tentent e, per conto nostro, ci preoccupiamo che anche le faville più leggere non suscitino qualche incendio. Onde ci chiediamo se sia lecito che un giornale locale che fa professione di fascismo pubblichi minacce e intimidazioni che (in una corrispondenza da Tione) polemizzando con un o.d.g. di una assemblea di capifamiglia stampi bravate consimili: «Illustri capi di famiglia, fate il piacere di tacere anche e soprattutto perché oltre al Partito fascista che vostro malgrado oggi governa energicamente e sagacemente potrebbe svegliarsi quel santo manganello che vi farebbe ingoiare, ecc.». Come? Oltre il Governo, del quale vantate l'energia e la sagacia, invocate ancora il manganello? Oltre la forza della legge minacciate la violenza del vostro bastone? Sentite questa, che viene a proposito. Nella «Cremona Nuova» (che è pur l'organo di quell'agnellino di Farinacci) si è svolto un dibattito a proposito delle recenti sentenze dei tribunali contro i somministratori di olio di ricino. Nella discussione intervenne anche l'avv. Bartoletti che, messo in rilievo la severità dell'art. 154 del C.p., secondo il quale, infatti, chiunque usa violenza o minaccia per costringere qualcuno a fare, tollerare od omettere qualche cosa è punito con la reclusione di anni da uno a tre e con la multa di lire da mille a cinque mila, ecc.; così conclude (e le sue parole scritte sul Cremona Nuova, dell'on. Farinacci hanno un valore anche maggiore) commentando le osservazioni di un fascista che trovava meno punite le manganellate dell'olio di ricino: «Io, tuttavia, ho osservato a quei giovani che non si possono misurare le conseguenze dei colpi di manganello e quindi si possono produrre anche lesioni che sono punite con un minimo di cinque anni di reclusione, o la morte, e che, ad ogni modo, non ha più ragione d'essere né manganello né olio di ricino, oggi consule Mussolini. Ed è giusto per questo, a mio credere, che i magistrati hanno messo da banda la benignità, con la quale interpretavano la legge per l'addietro». Noi non condividiamo naturalmente il criterio soggettivo del collaboratore della Cremona Nuova, il quale, se Mussolini, non fosse al Governo, troverebbe forse inapplicabile per simili casi l'art. 154; ma il criterio soggettivo o quello oggettivo, dato che al Governo c'è proprio Mussolini, conduce alla stessa conclusione. Invece i signori di Tione – e, pare, gli altri che raccolgono la voce loro – pensano che consule Mussolini si debba tuttavia far ricorso al manganello il quale dovrebbe fornire una sanzione suppletoria a quella delle leggi, pur applicate con energia e con sagacia! E magari fra chi statuisce così ossequenti principi si distingue qualche impiegato governativo! La verità è che anche nelle valli più alpestri una strana emulazione sembra rivelarsi. Se Mussolini ha fatto la marcia su Roma, perché non vorremo noi intraprendere la marcia di ... Sfruz? Sfruz è un paesello di 500 abitanti circa con 108 ettari di agro comunale. Ora che cosa può essere accaduto su codesti 108 ettari di china montana e che razza di delinquenti alberga codesto pacifico nido, se lo stesso giornale, ultimo numero, stampa da oltà una corrispondenza che termina così: «Lottiamo per una causa giusta ed umana, e se nella legge non troveremo, una pronta giustizia – i fascisti italiani la faranno da loro. Roma fulcro di civiltà e progresso fu conquistata dal fascismo come lo furono tutti i comuni d'Italia. Sfruz è un granello di sabbia in mezzo al mare che potrebbe essere travolta dalla bufera fascista che epurerà l'atmosfera satura di miasmi, d'ingiustizia e malvagità». Non sappiamo proprio di che si tratti, né d'interesse di sapere, poiché pensiamo che gli organi del governo Mussolini – carabinieri, sindaco, sottoprefetto, pretore – siano sufficienti a salvaguardare la causa della giustizia e del patriottismo, quando fossero in pericolo. Ma quei giovanotti di lassù pensano diversamente e proclamano che se la legge non fa giustizia essi faranno giustizia fuori della legge, minacciando di travolgere nella bufera fascista quel miserabile «granello di sabbia» che risponde al dolce nome di Sfruz. Come Roma..., tanto che dire: come Mussolini... Ehi, signori! Siete in arretrato. Mussolini ha detto e ripetuto che siamo già entrati nel secondo tempo, nel tempo cioè in cui, valendosi delle forze sociali e politiche dello Stato, si assimila, si rinvigorisce, si ricostruisce. Non si scatenano bufera, ma si lavora nell'ordine e nella disciplina, non si fa dello «stupido antierico illegalismo» (parole sue), ma si dà forza e obbedienza alle leggi. Senonché non intendiamo ricordare questa paternale proprio per voi e solo a voi. Voi siete un caso fra tanti che abbiamo colto fra gli ultimi, ad illustrazione degli altri. Il monito vale per tutti. Il Trentino non è ancora quanto dalle piaghe della guerra e dei martiri che, durante la guerra soprattutto, gli fece soffrire lo straniero; deve oltre a ciò fare opera di adattamento alle nuove condizioni economiche – l'adattamento anche se per il meglio comporta sempre uno sforzo –, deve assuefarsi ad un nuovo tipo di finanza e di amministrazione; è tutto un lavoro questo che esige spirito di solidarietà e di fattivo patriottismo fin nell'ultimo paesello di montagna, e a questo lavoro sono chiamati quei pochi che, superando l'attuale crisi economica nonché emigrare, possono dedicare qualche attività alla pubblica cosa. Ebbene questa è opera urgente di italianità ch'esige concordia, equità, tolleranza e presuppone un ambiente di pace. A tale opera noi diamo il nostro concorso e nessun altro ammonimento abbiamo ripetuto ai nostri, se non questo. Perché è a che pro devono risuonare ancora parole di odio e di minaccia? Oggi non si marcia, si lavora e si ricostruisce, per l'Italia, e per questa terra finalmente sua.

Persons Keywords Polarity Tags



Text Classification

Saggista/Giornalista X



Polarity



Prior Polarity Experiment: document aggregation

- Sentiment of De Gasperi's writings dated back to 1914 and related to the outbreak of WW1

Words with negative prior polarity



Prior Polarity Experiment: document aggregation

- Sentiment of De Gasperi's writings dated back to 1914 and related to the outbreak of WW1

Words with positive prior polarity



Crowdsourcing Experiment: Contextual Polarity

RESEARCH QUESTIONS:

- Is it possible to apply crowdsourcing methodologies to the assignment of contextual polarity in historical texts?

EXPERIMENT:

- 2 lemmas ‘sindacato’ (*trade-union*) and ‘sindacalismo’ (*trade-unionism*)
- 525 sentences
- 2 expert annotators judged the contextual polarity
- third judgment collected through a CrowdFlower job:
- quality control mechanisms:
 - regional qualifications
 - gold units
 - majority vote on 5 judgments



Crowdsourcing Experiment: Job Interface

Come se ne parla? Giudica una parola in un contesto [Italiano]

Instructions ▲

In questo task ti verranno presentati dei testi politici della prima metà del '900. Ti è richiesto di leggere delle frasi e giudicare se si parla della parola **evidenziata** in maniera *Positiva, Negativa o Neutra*.

Esempio:

- *Positiva*: Riconosciamo d' altra parte , continua la rivista cattolica , che , nella grande industria , i **sindacati** operai avevano compiuto efficaci sforzi al fine d' ingaggiare colle organizzazioni padronali delle conversazioni.
- *Negativa*: Le vostre saranno belle dottrine , ma rimangono parole . I **sindacati** fascisti non sono capaci di realizzare quello che c' è di buono nelle vostre e nelle altrui teorie .
- *Neutra*: I telegrammi sono qua . Io non entro in questa discussione , poiché non sono membro del **sindacato** ferrovieri . Pero è un fatto che se questo è possibile , qualche cosa

Contadini , fate un viso sorridente ! Il partito dei contadini è scomparso , il partito socialriformista è scomparso , i **sindacati** non esistono . Che resta ?

In che maniera si parla della parola evidenziata?

- ☐ Positiva
- ☐ Negativa
- ☐ Neutra
- ☐ Non so

Commenti?

Crowdsourcing Experiment: Results

- At the end:
 - 21 contributors, out of which only 12 were reliable
 - 5 days to complete the job
 - 36 \$ total cost of the experiment

ACCURACY

Accuracy	Crowd	Prior
Overall	68.30%	43%
Negative	55.50%	22%
Neutral	80%	31%
Positive	46.60%	86%
Unknown	0	n.a.

→ Prior polarity of the sentence based on the lexicon

Crowdsourcing Experiment: Results

INTER-ANNOTATOR AGREEMENT

IAA	Fleiss' kappa
annotator A vs annotator B vs crowd	0.39
annotator A vs annotator B	0.46
annotator A vs crowd	0.35
annotator B vs crowd	0.35

Conclusions

- new Italian **lexical resource** for SA

eccellente a#02232109 1 0 of the highest quality;

- measurement and visualization of polarity at **document level** integrated in ALCIDE
- standard **crowdsourcing methods** used in other domains cannot be straightforwardly adopted to historical texts

Future Works

- From document level to concept-based / entity-based SA
 - De Gasperi on *corporatism* before and after 1946
 - De Gasperi on *Togliatti* in propaganda vs Parliament speeches
- Extend SA to English texts
 - Next case study: 1960 USA Presidential campaign speeches
- Improve visualization:

*It's a rule in Digital Humanities:
you need an Italian designer in your project*

Bruno Latour



THANK YOU!

Email: sprugnoli@fbk.eu

Web Site: <http://dh.fbk.eu>

Twitter: https://twitter.com/DH_FBK