Language Resources for
Natural Language Processing (Part II)

Viviana Patti, Dipartimento di Informatica, University of Turin

viviana.patti@unito.it

a.y. 2024/2025 - Language Technologies and Digital Humanities

6 CFU - 36 ore - 6 weeks

Teacher: Viviana Patti

Dipartimento di Informatica

Università di Torino

E-mail: viviana.patti@unito.it

Tel.: 011 6706804

Homepage: https://www.unito.it/persone/vpatti

❖ Ufficio: Dipartimento di Informatica, 3rd floor, new building, n. 6.

   ❖ WebEx Virtual Room: https://unito.webex.com/meet/viviana.patti

❖ Available during the reception hours and by appointment (e-mail) for explanations: Tue, h 16-17 (after lesson, send an email before).

Who am I?
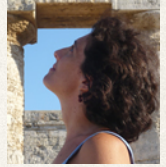
- Msc Computer Science in Philosophy

    - Multidisciplinary curriculum (analysis of language and computer science): philosophy of language, computational linguistics, cognitive science, programming, artificial intelligence.

- PhD in Computer Science

- Associate Professor at Computer Science Department, University of Turin

- co-founding member of the Logic, Language and Cognition center, University of Turin (Philosophy, Computer Science, Psychology)

- Past member of Guarantee Committee at UniTo

- https://www.unito.it/persone/vpatti

Research areas

✤ AI@Dipinfo research group: Content-Centered Computing (CCC)

✤ Emotion and sentiment in Social Media

  ✤ AI: Natural Language Processing (NLP) & Computational Linguistics

    ✤ Sentiment Analysis and Opinion Mining (OM&SA) in Social Media

      ✤ Developing corpora and linguistic resources: sentiment, stance, hate speech, i

      ✤ Developing tools for automatic sentiment/stance/hate/irony detection

      ✤ Hate speech detection in social media

        ✤ Multidisciplinary approach

        ✤ Multilingual perspective

    ✤ NLP &Visualisation

    ✤ Language and social structure in online debates
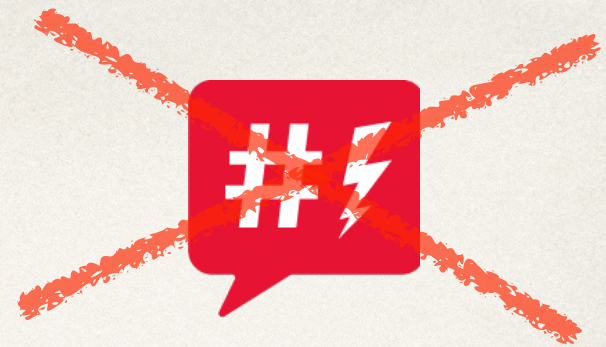
      ✤ Political sentiment and stance detection

The multi-disciplinary team

Valerio BASILE

Viviana PATTI

Cristina BOSCO

di.unito.it

DEPARTMENT RESEARCH EDUCATION THIRD MISSION

HOME › RESEARCH › RESEARCH GROUPS › CONTENT CENTERED COMPUTING › ABOUT
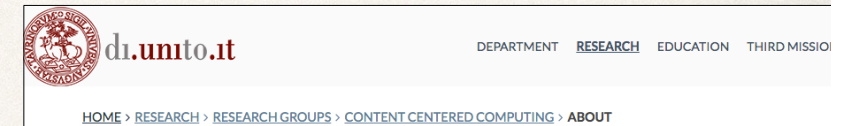
Content Centered Computing

PEOPLE PROJECTS PUBLICATIONS MISCELLANEOUS

Simona Frenda

Elisa Di Nuovo

Silvia Casola

Tom Bourgeade

Manuela SANGUINETTI

Delia Irazu HERNANDEZ FARIAS

Komal Florio

Sohail Aktar

Soda Marem Lo

Endang W. Pamungkas

Alessandra T. Cignarella

ARC²S Group

Applied Research on Computational Complex Sy

Home About ∨ People Projects Publica

Workshop

Giancarlo RUFFO

Fabio POLETTO

Marco STRANISCI

Arthur CAPOZZI
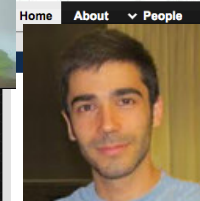
Mirko LAI

Rossano SCHIFANELLA

Web pages

✤ Web pages of the course:

    ✤ Moodle: https://elearning.unito.it/scienzeumanistiche/course/view.php?id=9728

    ✤ CampusNet: https://www.digitalhumanities.unito.it/do/corsi.pl/Show?_id=nxwv

UNIVERSITÀ DI TORINO

🔔 💬 ♿ Default Font ⌄   🌐 Italiano (it) ⌄   Viviana Patti 👤 ⌄

Scuola di Scienze Umanistiche   UNITO ⌄   HELPDESK   MY MEDIA   🔍

🏠 Home   🎨 Dashboard   💼 i miei corsi   🗂 Questo corso   ⚙ ⌄   ↗ Schermo intero

📁 › I miei corsi › STU0674_24_25_BOSCO/PATTI

Linguistic Resources for Natural Language Processing (STU0674) - Bosco/Patti - 2024/2025

⌄ **Information about classes and exams**   Minimizza tutto

Schedule

---

* **When?**
  Mon, Tue, Wed: 14-16

* **Where?**

* Aula 3.06 Thin Client
  Via Sant'Ottavio - 54
  Torino

Final exam

The evaluation for the <u>second part</u> of the course will consist of a **Project Work** on the development of annotated corpora and analysis, to be implemented by following the instructions described in a document that will be made available on Moodle.

Love and Hate in the Time of Big Data

Text analysis + (big) social media data = new challenges and opportunities
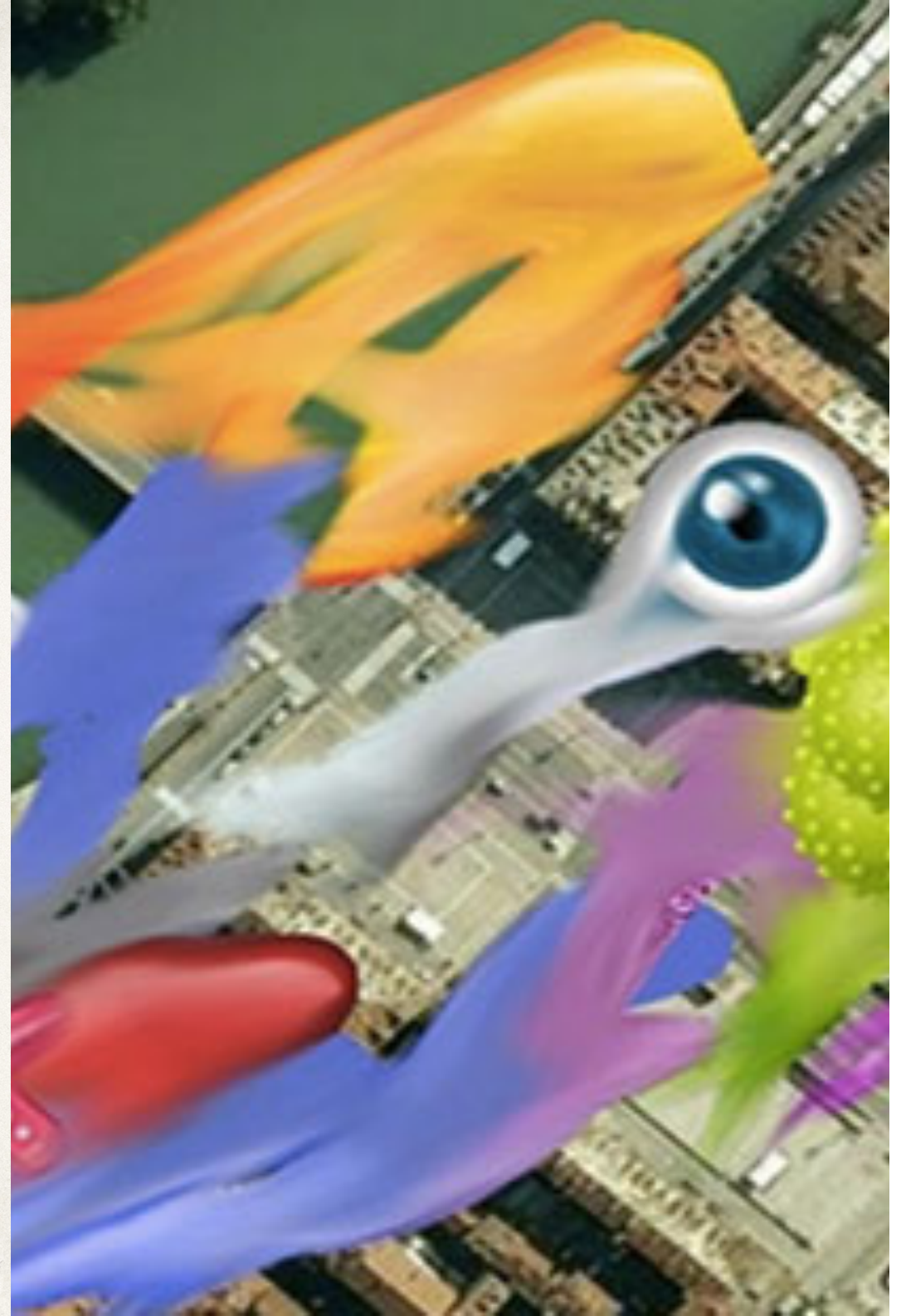
NLP and Machine Learning

✣ Natural Language Processing (NLP) & Computational Linguistics (CL)

    ✤ Morphology → how words are

    ✤ Phonetics → how words sound

    ✤ Syntax → how words relate

    ✤ Semantics → what words means

    ✤ Pragmatics → what is the intention of the speaker (context)

✣ Different sources of texts

✣ Different tasks: sentiment analysis, stance detection, emotion detection, hate speech detection, misogyny identification, conversational agents

✣ Computational models, Deep Learning

PRAGMATICS
SEMANTICS
SYNTAX
MORPHOLOGY
PHONOLOGY
PHONETICS
speech sounds
phonemes
words
phrases and sentences
literal meaning of phrases and sentences
meaning in context of discourse

Artificial Intelligence
Deep Learning
Natural Language Processing
Machine Learning

from quora.com
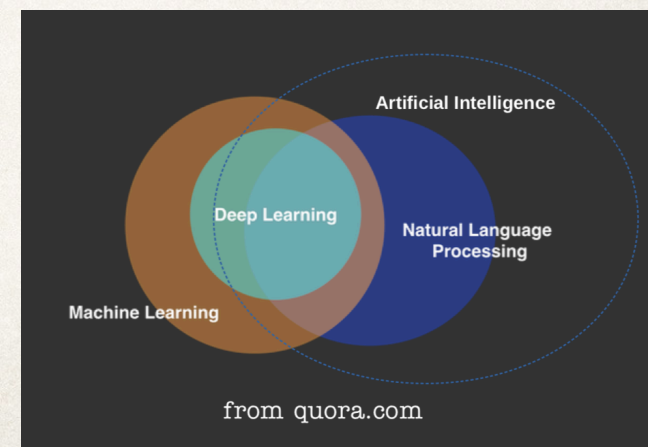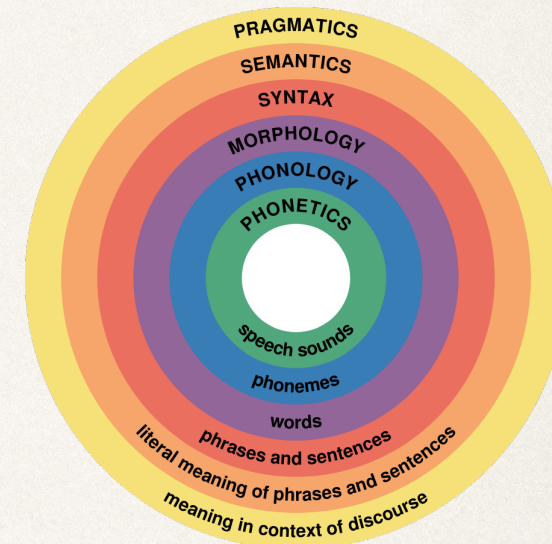
Base Models + Further Training + Fine-Tuning

MLM: masked language model

Pre-training — ELMo ULMFiT GPT BERT

MLM on unlabelled data

LOTS!

Adaptive/ Behavioural Fine-tuning

Loss on relevant data

Model specialised to target data/task

Fine-tuning

Cross-entropy on task labels

classification sequence labeling Q&A ....

Task-specific data
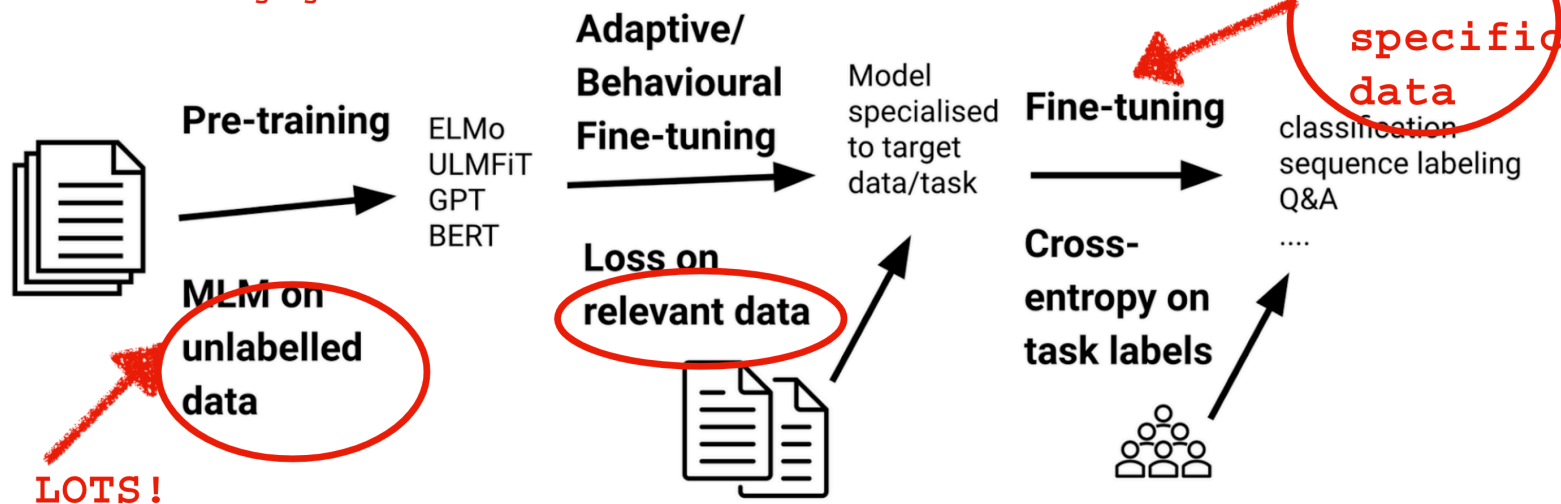
Revolutionised everything, from generation to MT, and lots of other tasks BUT, bias in data   https://ruder.io/recent-advances-lm-fine-tuning/

Language Models are few-shot learners

Zero-shot
The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

    Translate English to French: task description
    cheese =>   prompt

One-shot
In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

    Translate English to French: task description
    sea otter => loutre de mer   example
    cheese =>   prompt

Few-shot
In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

    Translate English to French: task description
    sea otter => loutre de mer   examples
    peppermint => menthe poivrée
    plush girafe => girafe peluche
    cheese =>   prompt

Computational linguistics

✣ https://www.aclweb.org/portal/what-is-cl

Study language with computational methods
→ Computational Linguistics

Elaborate language-based data
→ Natural Language Processing

On the gap between
theoretical and computational linguistics

Marco Baroni
EACL
April 21th, 2021

ICREA

upf. Universitat Pompeu Fabra Barcelona

Computational Linguistics and Linguistic Theory

Facebook AI Research

What is computational linguistics?

Computational linguistics is the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistic phenomena. These models may be "knowledge-based" ("hand-crafted") or "data-driven" ("statistical" or "empirical"). Work in computational linguistics is in some cases motivated from a scientific perspective in that one is trying to provide a computational explanation for a particular linguistic or psycholinguistic phenomenon; and in other cases the motivation may be more purely technological in that one wants to provide a working component of a speech or natural language system. Indeed, the work of computational linguists is incorporated into many working systems today, including speech recognition systems, text-to-speech synthesizers, automated voice response systems, web search engines, text editors, language instruction materials, to name just a few.

Popular computational linguistics textbooks include:

- Christopher Manning and Hinrich Schütze (1999) *Foundations of Statistical Natural Language Processing*, Cambridge, Massachusetts, USA. MIT Press.
  Also see the book's supplemental materials website at Stanford.

- Daniel Jurafsky and James Martin (2008) *An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Second Edition. Prentice Hall.

https://www.virtual2021.eacl.org/plenary_session_keynote_by_marco_baroni.html

Suggested texts
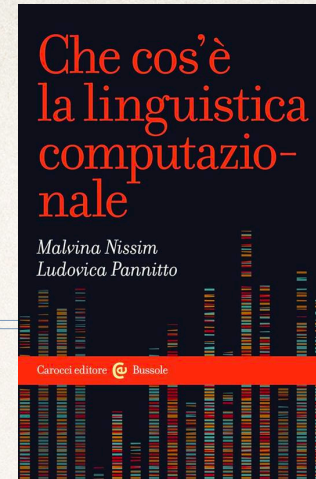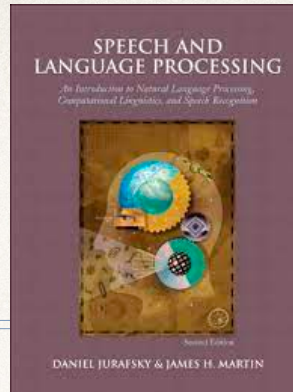
✤ D. Jurafsky & J.H. Martin
Speech and Language Processing:
https://web.stanford.edu/~jurafsky/slp3/ (third edition),
Prentice Hall, 2008. August 20, 2024 last release!.

✤ Malvina Nissim, Ludovica Pannitto. Che cos'è la
Linguistica Computazionale. Carocci nella serie
Bussole. 2022.

✤ Elisabetta Jezek, Rachele Sprugnoli. Linguistica
computazionale. Introduzione all'analisi automatica dei
testi. Il Mulino. 2022.

The NLP Pipeline
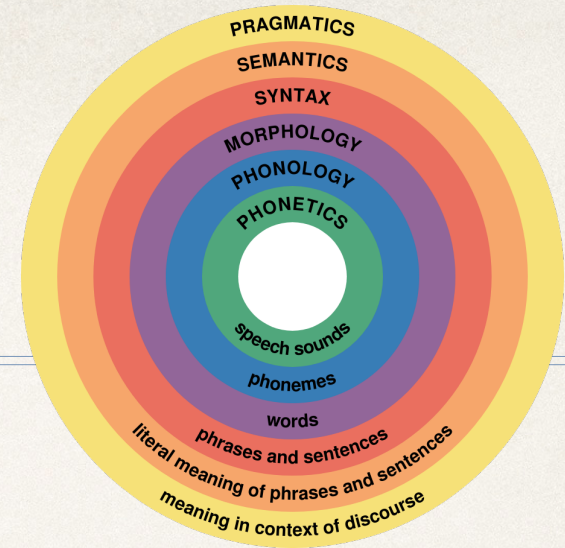
* Many tasks in NLP

* Aligned with traditional levels of analysis in Linguistics
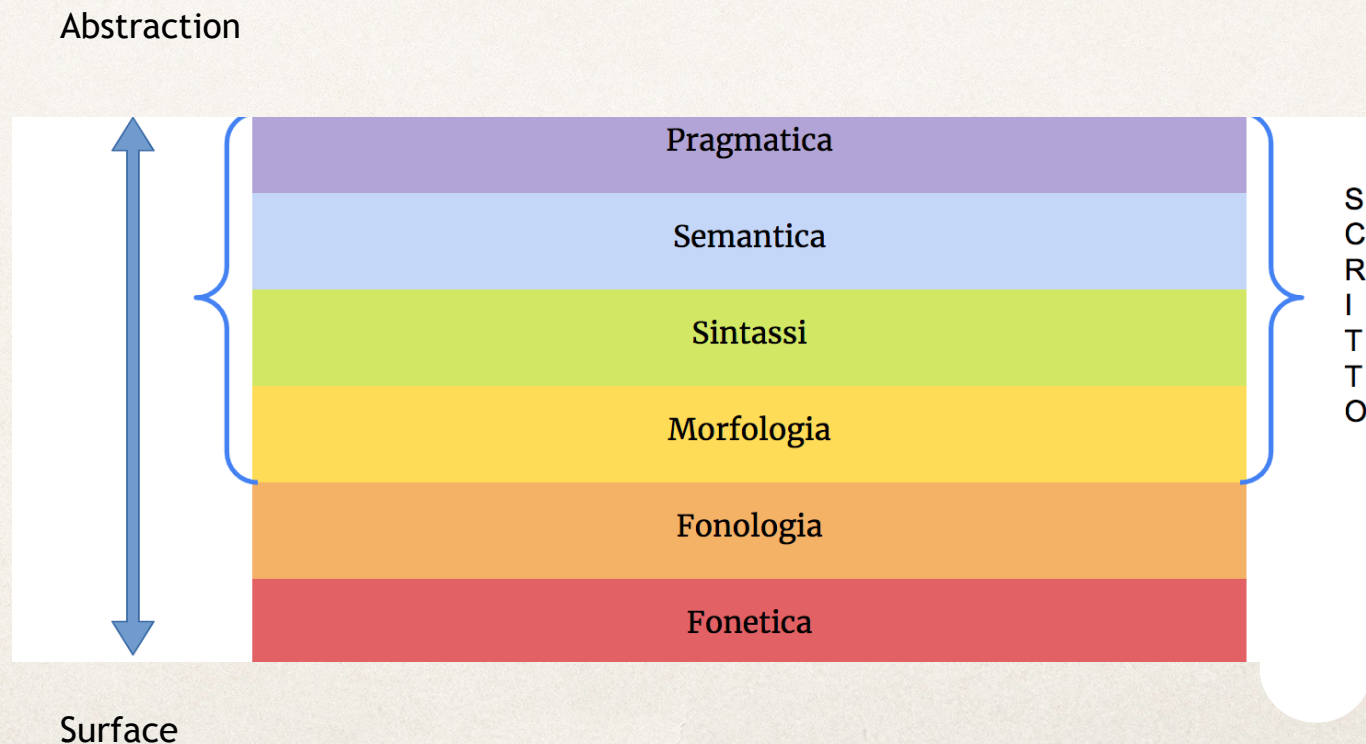
The NLP Pipeline
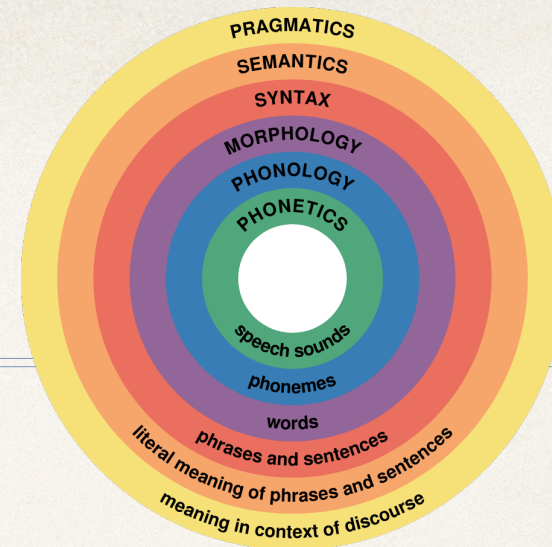
✤ Many tasks in NLP

✤ Aligned with traditional levels of analysis in Linguistics
...

Sentiment Analysis

Named Entity Recognition

Role Labeling

Syntax Parsing

POS-tagging

Lemmatization

Tokenization

Abstraction

Surface

Pragmatica

Semantica

Sintassi

Morfologia

Fonologia

Fonetica

S C R I T T O
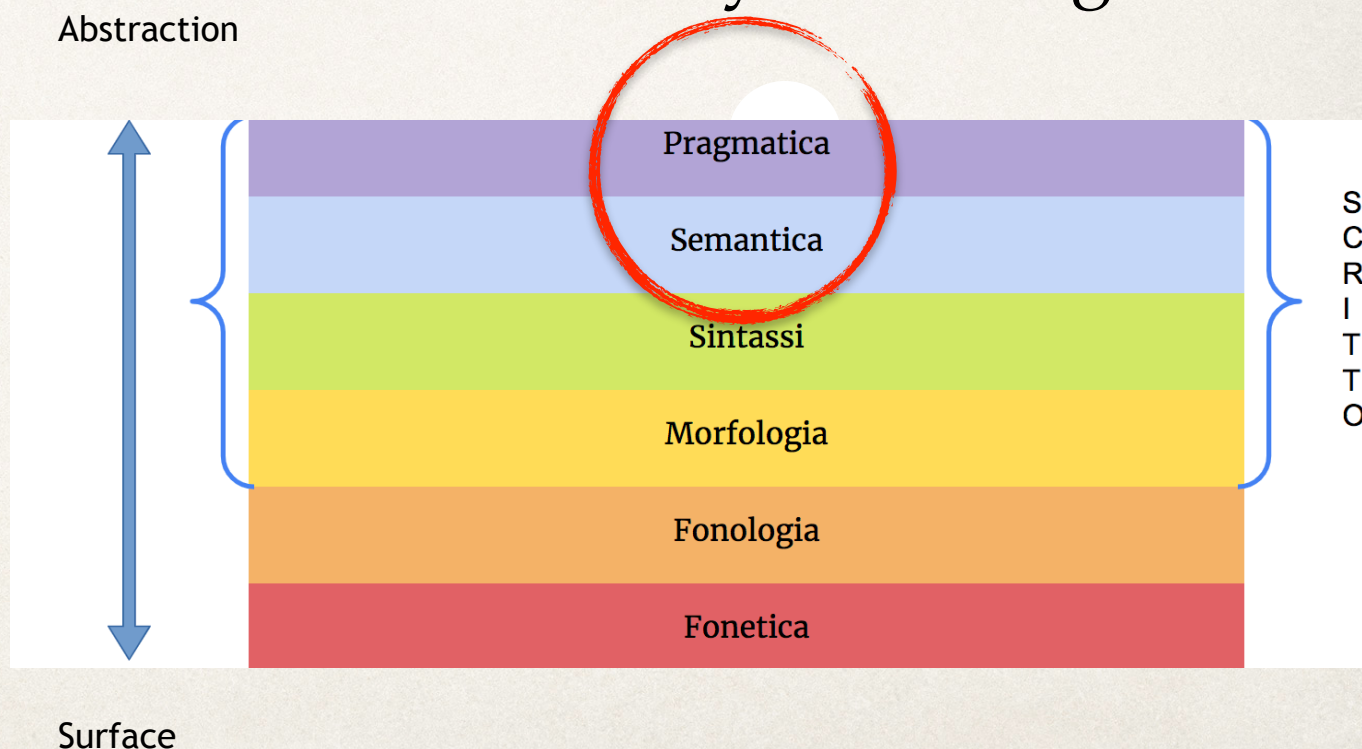
Motivation
Social Media

- Pervasive and immersive technology

- On social media users nowadays freely express what is on their mind at any moment in time, at any location, about virtually anything.

- Social media changed the way we share, produce and acquire information

  - Share and shout!

- Studying the media ecosystem:

  - Social media posts as reactions to news from newspapers or television news

    - Multimodality
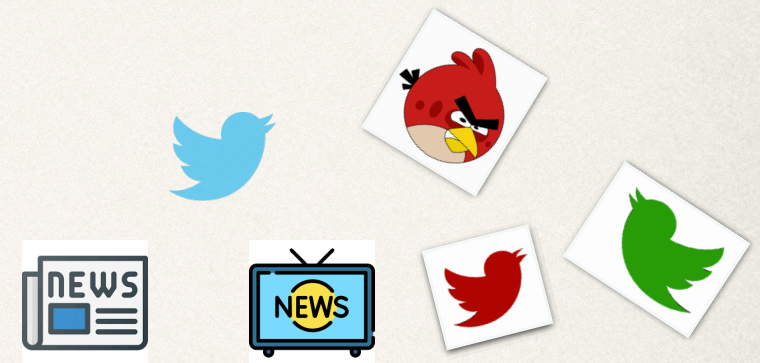
Credits: C. Musto

Motivation
Social Media

- Twitter (now X)

- Millions tweets every day:

    - A really huge amount of user-generated texts

- Millions users

- Relatively easy access to data (policies)

- Easy access to user generated contents in many languages

    - Multilingual data

- Contents + Metadata:

    - Geo-time coordinates: who says what, where, when

    - Followers and friends, re-tweets, (hash)tags:
        - Context of the conversation
        - Social graphs of users: community detection

http://mappa.controlodio.it/
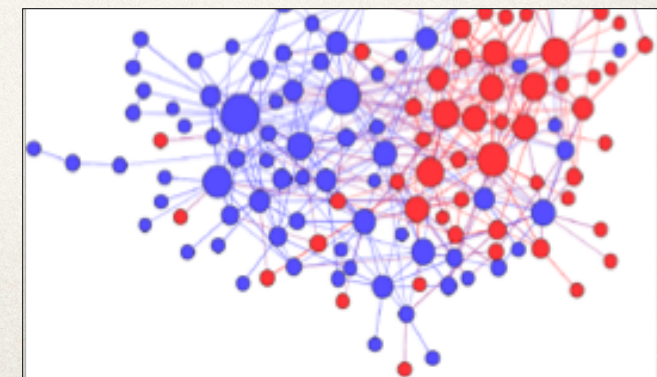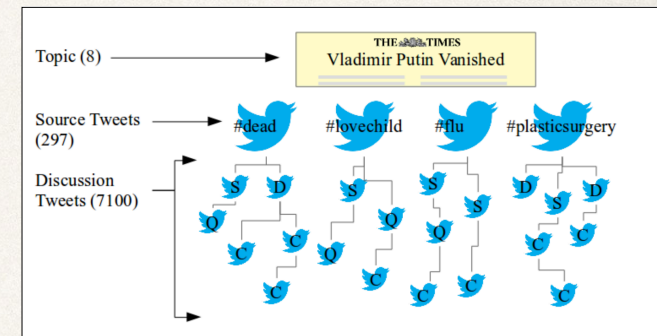
Motivation
Social Media

✤ Large amounts of spontaneously produced texts - users digital traces open up unique opportunity to learn more about such users:

  ✤ Can be used to understand social dynamics

  ✤ Can be a valuable source to know the world we live in and the communication dynamics we are daily involved with

  ✤ Ecological setting for monitoring, understanding and counteracting: hate speech, misogynistic behaviours on line, political polarisation, spreading of rumours and fake news,…

  ✤ Ecological setting for analysing linguistic phenomena (e.g., irony)

Easy access to Data and Metadata in Tweets

- A sample of tweets is available to researchers/practitioners through public APIs (Application Programming Interface) (until March 2023 for free)

- Format: JSON (JavaScript Object Notation), format for data exchange in client-server applications

- Roughly, you can extract the following types of information from Twitter:

  - Tweets published by a user (status message): the text!

  - Information about a user profile

  - User's network consisting of his connections

  - Search/filter results on Twitter by using keywords or hashtags: e.g. #silviaromano, #sischerza, #brexit, :

    - e.g., you can get a continuous stream of Tweets matching some search criteria.

Easy access to Data and Metadata in Tweets

* A sample of tweets is available to researchers/practitioners through public APIs (free)

* Format: JSON (JavaScript Object Notation), format for data exchange in client-server applications

* The rate limitations of Twitter APIs can be too restrictive for certain types of applications.

  * To satisfy such requirements, Twitter Firehose provides access to 100% of the public Tweets on Twitter at a price.

    * Obtaining Data via third-party resellers: e.g. http://gnip.com

Metadata on Tweets
Contextual features

- Rich structured information about the users involved in the communication
- Tweets are associated to several kinds of meta-data:
    - geographical coordinates of where the tweet was sent from
    - the id of the sender
    - the time of the day…

- Twitter maintains information of
    - who follows whom: followers and friends
    - re-tweets
    - (hash)tags inside of tweets provide information about the context of the conversation

    information that can be exploited to yield a more accurate picture on the social graphs of users involved in communications on given topics

Important: this contextual information can become interesting and powerful signal for a machine learning model.

Experiments on stance detection both for English and in a multilingual setting

Twitter's language

- **280 characters: tweets are short!**
  - a sentence or a headline rather than a document.
- **Language used very informal/spontaneous speech**
  - creative spelling and punctuation, misspellings, slang, new words, URLs, and genre-specific terminology, abbreviations, such as RT for "re-tweet" and #hashtags
  - New `language', new challenges

Disgustibus #Sanremo2019

11:01 AM · Feb 10, 2019 · Twitter for iPhone

Ma solo a me sta cosa che l'Istat misura il capitale umano - il valore in € delle nostre vite mette addosso un ansia assurda?
#buonlunedì

11:22 AM · 18 dic 2016 · Twitter for Android

Non guardando mai @Striscia, mi ero persa la polemica su #GiovannaBotteri: provo disgusto e tristezza, aumentata dal fatto che proprio una donna, @m_hunziker si sia prestata alla squallida denigrazione di una grande professionista.

3:35 PM · 2 mag 2020 · Twitter for iPhone

Facebook, Instagram data, …

* Not only Twitter (X)

    * Facebook data

    * Instagram data (multimodality issues)

    * Tik Tok

    * Whatsapp data / Telegram data (groups)

    * YouTube, BeReal (comments)

    * Reddit data

    * VK.com (mainly Russian, but also other language). Similar to FB

    * Ekşisözlük (only Turkish)

    * Weibo (Chinese) / Xiaohongshu (Chinesee - similar to Instagram - e-commerce)

    * WeChat (Chinese)

    * Stackoverflow / Quora / Tumblr

    * Access policies, GDPR,
      new owners…from Twitter to X

SOCIAL MEDIA ICONS
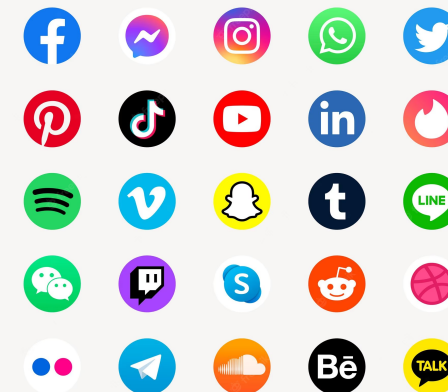
Sentiment Analysis

✤ Sentiment analysis and opinion mining

✤ Sentiment analysis (SA) is

  ✤ "The computational study of opinions, sentiments and emotions expressed in text" (Bing Liu, 2012)

  ✤ A relevant topic of research in NLP, especially with respect to the study of new forms of digital and social communication.

✤ Sentiment analysis or opinion mining? Often used interchangeably to denote the same field of study

  ✤ Sub- areas of subjectivity analysis (Bo Pang and Lillian Lee, 2007)

✤ Sentiment analysis is usually interpreted as a text classification task:

  ✤ The task of classifying text (e.g. a tweet) by assigning it a label drawn from some set of labels.

NLP tasks
Sentiment Analysis et al.

- **Text classification tasks:** classifying text (e.g. a tweet) by assigning it a label drawn from some set of labels

- **Different facets** of the **affective content ->** different NLP tasks:

    - sentiment polarity classification

    - emotion detection

    - stance detection

    - hate speech detection

    - misogyny identification …

    - transversal tasks: irony detection

Malvina Nissim and Viviana Patti. 2017. Semantic aspects in sentiment analysis. In Federico Alberto Pozzi, Elisabetta Fersini, Enza Messina, and Bing Liu, editors, Sentiment Analysis in Social Networks, pages 31–48. Morgan Kaufmann, Boston.
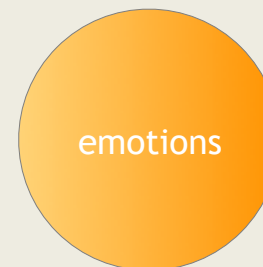
Sentiment Analysis et al.

opinions   mood

✤ **(Classical task)** sentiment polarity detection task: positive vs negative texts

``decide whether a given message is of positive, negative, neutral (or mixed*) sentiment.'' *optional

Given a text systems need to classify/predict the positive or negative sentiments in the text, when present.

opinions
#lamicageniale

✤ Positive vs negative opinions about a target

Mannaggia
M hai fatto ricordare #lamicageniale

Come siamo passati da quel piccolo capolavoro a sta roba 😢😢😢

Manuela 🍌 @ManuCia0 · 21 mag
Ecco adesso il signor Cerullo se li gioca tutti quei soldi #vivielasciavivere

✤ Positive vs negative mood

♡ 1

Sentiment Analysis et al.
Stance and Target

✤ Finer-grained tasks!

   ✤ Online political debates: a large source of informal and opinion-sharing dialogue on current socio-political issues

      ✤ Dual-sided debates: 2 possible polarizing sides can be taken by participants

   ✤ Stance detection: detection of positions pro or con a particular target entity (person, organization, movement, policy, etc.) that users assume within debates

"decide whether the tweeter is
in favor of the target, against the given target,
or whether neither inference is likely"

      ✤ Controversial topics

   ✤ Aspect-based sentiment analysis: products

Ibereval 2017 / Stance and Gender Detection in Tweets on Catalan Independence

HOME   DATA AND RESULTS   CONTACT

Stance and Gender Detection in Tweets
on Catalan Independence@Ibereval 2017

Stance and Gender Detection in Tweets on Catalan Independence will take place as part of IberEval 2017, the 2nd
Workshop on the Evaluation of Human Language Technologies for Iberian languages, at SEPLN 2017 at University of
Murcia, Murcia, Spain, on September 19th, 2017.

Applications:
Making Sense of Political debates

Multilingual perspective
    different socio-political debates

Selection:

    hashtags largely exploited in the debate. Accepted
    within the dialogical and social context growing around the
    controversial topic

French: #mariagepourtous

Debate on the homosexual wedding
    in France (Bosco et al. @LREC2016)
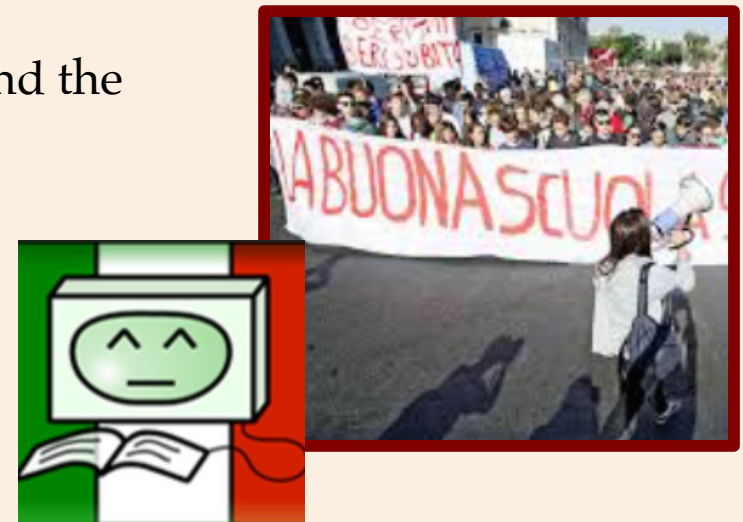
Italian: #labuonascuola

Debate on the reform of the education sector
    in Italy (Stranisci et al. @LREC2016)

Multilingual perspective

Multilingual perspective
different socio-political debates

English: #brexit

Debate on British Exit from EU
(Lai et al. @CLEF2017)

Spanish/Catalan: #independencia #27S

Debate on Catalan Independence

27S: September 27, 2015
Regional elections in Catalonia
de facto referendum on independence

(Taule et al. @Ibereval2017)

Sentiment analysis et al.
Aspect-based Sentiment Analysis

✤ Finer-grained tasks!

  ✤ Aspect-based sentiment Analysis,

  ✤ Products, services

  ✤ User's opinions

Booking.com

«Posizione e comodità ai mezzi.Personale cortesissimo.»

R **Rita**
🇮🇹 Italia   7,0

Location and Staff Positive

Arredo camere, un po' vetusto. Assenza di mini-bar interno o anche semplice macchinetta caffè/bevande calde, abbastanza anomalo in una struttura simile!

L **Luca**
🇮🇹 Italia   4,3

Comfort Negative

I croissant erano buoni tipici di Roma! Grande varietà di cibo, cappuccino ottimo! Grande varietà di cioccolata Perugina, marmellate, succhi di frutta.

R **Roserica**
🇨🇭 Svizzera   7,6

Service Positive

Sentiment analysis et al.
Aspect-based Sentiment Analysis

✤ Finer-grained tasks!

    ✤ Aspect-based sentiment Analysis

At a glance

Participants are asked to **detect** within sentences (expressing opinions about accommodation services) some of the **aspects considered by the writer**.

For each detected aspect, participants are asked to detect a specific **polarity class**

The set of considered aspects is: PULIZIA (cleanliness), COMFORT, SERVIZI (amenties), STAFF, QUALITA-PREZZO (value), WIFI and POSIZIONE (location).

"La posizione è molto comoda per il treno e la metro"

Aspect Category Detection (ACD)

In the ACD task, one or more **"aspect categories"** evoked in a sentence have to be identified, e.g. the posizione (location).

COMFORT ✗ STAFF ✗ LOCATION ✓ VALUE ✗ ...

Aspect Polarity Detection (ACP)

Each **category aspect** detected in the ACD task have to be annotated with **polarity label**: **POS** (positive) , **NEG** (negative), also in a not exclusive way (**Mixed**)

LOCATION POS ✓   LOCATION NEG ✗

cool! ABSITA
Aspect-based Sentiment Analysis at EVALITA 2018
http://sag.art.uniroma2.it/absita/

Pierpaolo Basile
University of Bari
pierpaolo.basile@uniba.it

Valerio Basile
University of Turin
basile@di.unito.it
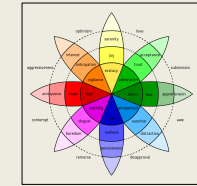
Danilo Croce
University of Rome "Tor Vergata"
danilo.croce@uniroma2.it

Marco Polignano
University of Bari
marco.polignano@uniba.it

Sentiment Analysis et al.
Emotions and target

- Finer-grained task: beyond the simple sentiment

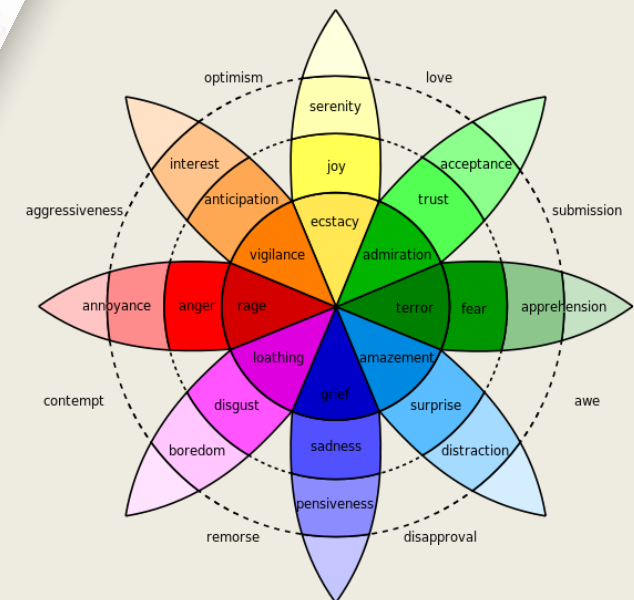 - Emotion detection: sadness, joy, anger, disgust, vigilance, fear, surprise, trust +

  - ♥

    ``decide whether a given message expresses joy, sadness, …,or anger"

  - Model: Robert Plutchik

    - Basic emotions

    - Complex/Mixed emotions: composition of basic emotions

    - Target

è normale che mi venga da piangere? #LAmicaGeniale
A me già viene da piangere annamo bene #lamicageniale
Sono già a pezzi. #lamicageniale

[negativo, ma è tristezza/commozione per il personaggio, non significa che non piaccia la serie o la puntata]
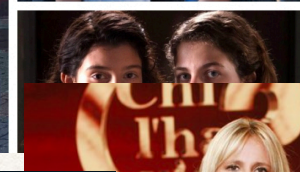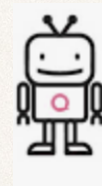
Donato Sarratore schifoso #LAmicaGeniale

[negativo ma è disgusto per il personaggio, non significa che non mi piaccia la serie o la puntata]

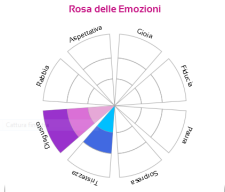# Applications: Art and Culture Domain

✤ **M.EMO.RAI:** with RAI - Radiotelevisione Italiana, the national public broadcasting company of Italy, on measuring the emotional engagement of audiences of TV programs through social media

✤ **ArsEmotica**
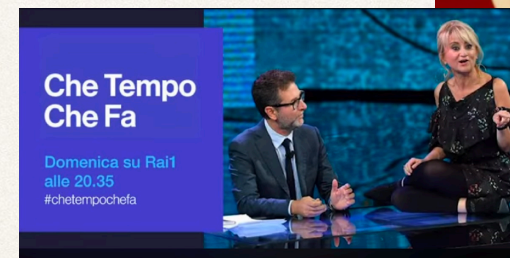
Hate speech detection

- Hate speech is commonly defined as any communication that that is abusive, insulting, intimidating, harassing, and/or incites to violence, hatred, or discrimination, and it is directed against a person or a group on the basis of some characteristics:

  - race, color, ethnicity, gender, sexual orientation, nationality, religion…

- Binary classification task:

  ``decide whether a given message is a hateful speech utterance or a harmless one against a given target''

- Online hate is expressed in different forms depending on the subject against whom it is targeted:

  - Hate speech against women, misogyny… target is important!

in giro per l'Italia
@GuerriJimmy   Follow

Anche oggi sono in arrivo 2000migranti dalla Libia avanti in italia ce posto per tutti vero @lauraboldrini ? Li puoi accogliere a casa tua

12:27 am - 16 Feb 2015 From Reggio nell'Emilia, Emilia Romagna

1 Like

Abusive language detection

- Abusive language detection

  - Hate speech detection

  - Offensive language detection

  - Misogyny identification

- Binary classification, but also finer grained tasks

  - E.g., misogynistic behaviours, …
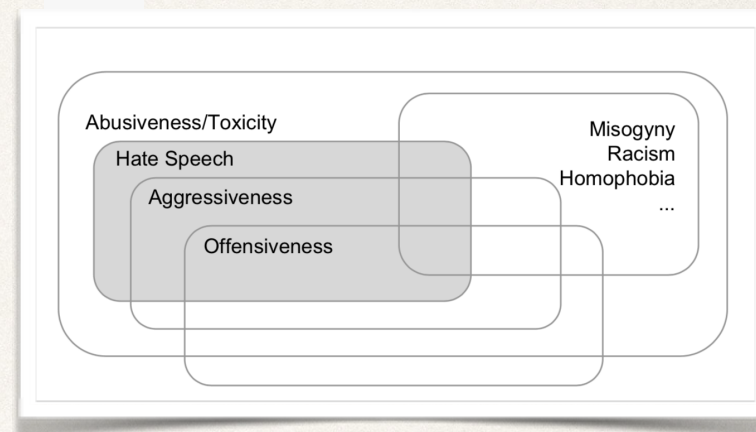
Fabio Poletto, Valerio Basile, Manuela Sanguinetti, Cristina Bosco, and Viviana Patti. 2020. Resources and Benchmark Corpora for Hate Speech Detection: a Systematic Review. Journal of Language Resources and Evaluation. Springer.

Irony and Sarcasm 🙃 😈 #⚡

- **Implicit vs** explicit, direct vs **indirect** ways to express sentiment / stance / emotions / hate

  - Expressive signals:

    - Emoji, emoticons, reactions:

      > So I just colored with Ava for an hour. Yeah my summer so far has been so fun #not 🙃

  - **Irony and sarcasm detection**

    - **Transversal task**

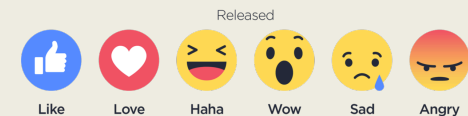# Irony and Sarcasm

🙃 😈

- ❖ Implicit vs explicit, direct vs indirect ways to express sentiment/stance/emotions/hate
  - ❖ Figurative language: Irony & Sarcasm

- ❖ Interpretation of irony involves linguistic patterns related to syntactic (how word relate) and semantic aspects (what words means), and pragmatic issues (what is the intention of the speaker / context, cultural background)

- ❖ Variety of of ironic devices and functions / role of affect 🙃 😈

- ❖ **Transversal task**

- ❖ Irony & sentiment: irony as polarity reverser:
  - ❖ One says something `good' to mean something `bad'

- ❖ Irony & Hate: sarcasm can be a linguistic expedient to (mitigate or) indirectly convey a hateful content

- ❖ Irony is a pervasive and viral phenomenon
  - ❖ Can be an ingredient to make the hateful posts viral
  - ❖ Important ingredient to develop counternarratives

So I just colored with Ava for an hour. Yeah my summer so far has been so fun 🙃 #not

I just love it when I speak to folk and they totally ignore me!!!
#Sarcasm! 🇬🇧

I carabinieri hanno individuato come possibile spacciatore un 27enne del Marocco. La tipica #risorsa straniera, ammiro la madre!

*The Italian police have identified a 27-year-old from Morocco as a possible drug dealer. The typical foreign #resource, I admire the mother! URL 🇬🇧

Variety of ironic devices and functions

✤ A variety of typologies of figurative messages can be recognised in social media:

  ✤ Twitter communications include a high variety of ironic devices and samples of use of figurative language

  ✤ From irony to sarcastic posts, and to facetious tweets that can be playful, aimed at amusing or at strengthening ties with other user

  ✤ Easy access to user generated contents in many languages

So I just colored with Ava for an hour. Yeah my summer so far has been so fun   #not 🇬🇧

I just love it when I speak to folk and they totally ignore me!!! #Sarcasm! 🇬🇧

Fun fact of the day: No one knows who invented the fire hydrant because its patent was destroyed in a fire. #irony 🇬🇧

I carabinieri hanno individuato come possibile spacciatore un 27enne del Marocco. La tipica#risorsa straniera, ammiro la madre! URL* 🇮🇹

* The Italian police have identified a 27-year-old from Morocco as a possible drug dealer.

The typical foreign #resource, I admire the mother! URL 🇬🇧

Variety of ironic devices and functions

🙃 😈 💬

❧ **Irony detection as a conceptual challenge in itself:**

 ❧ Creative and complex linguistic phenomenon, different functions

  ❧ reverse sentiment
  reinforce / mitigate hate /
  reinforce / contrast stereotypes …

 ❧ Can help to shed some light on how human beings use irony as a communicative tool

---

So I just colored with Ava for an hour. Yeah my summer so far has been so fun   #not 🇬🇧

I just love it when I speak to folk and they totally ignore me!!! #Sarcasm! 🇬🇧

Fun fact of the day: No one knows who invented the fire hydrant because its patent was destroyed in a fire. #irony 🇬🇧

I carabinieri hanno individuato come possibile spacciatore un 27enne del Marocco. La tipica#risorsa straniera, ammiro la madre! URL* 🇮🇹

---

* The Italian police have identified a 27-year-old from Morocco as a possible drug dealer.

The typical foreign #resource, I admire the mother! URL 🇬🇧

Irony and sentiment /
Irony as polarity reverser 🙃 😈

✤ Irony and sentiment

✤ The presence of ironic devices in a text can flip the polarity of an opinion expressed with positive words

to the intended negative meaning (or rarely vice versa) working as an unexpected polarity reverser

✤ One says something `good' to mean something `bad'

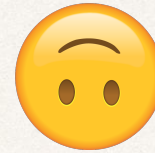I just love it when I speak to folk and they totally ignore me!!! #Sarcasm!

✤ This can undermine sentiment analysis systems' accuracy

✤ Crucial issue in NLP for the development of irony-aware sentiment analysis systems

Irony and sentiment
Irony as polarity reverser /
Variety of ironic devices

🙃 😈

I just love it when I speak to folk and they totally ignore me!!! #Sarcasm! 🙃

So I just colored with Ava for an hour. Yeah my summer so far has been so fun [smiling face emoji] #not

Fun fact of the day: No one knows who invented the fire hydrant because its patent was destroyed in a fire. #irony

Emilio Sulis, Delia Irazú Hernández Farías, Paolo Rosso, Viviana Patti, Giancarlo Ruffo: Figurative messages and affect in Twitter: Differences between #irony, #sarcasm and #not. Knowl. Based Syst. 108: 132-143 (2016)

Irony and hate /
Abusive or not? Implicitness 🙃 😈

✤ Irony / Sarcasm can be a linguistic expedient to (mitigate or) indirectly convey a hateful content

    ✤ reinforce / mitigate hate /

        reinforce / contrast stereotypes …

✤ Irony is a pervasive and viral phenomenon

  ✤ Can be an ingredient to make the hateful posts viral

    ✤ The presence of ironic devices often increases the viral load

      of the hate message (and its dangerousness or the possibility

      of fuelling a hate campaign): users sharing such contents do

      so with more levity when the message does not contain

      explicit insults.

  ✤ Important ingredient to generate counternarratives

😈

Context?

✤ A huge amount of ironic samples

✤ Context?

Exploring the Impact of Pragmatic Phenomena on Irony Detection in Tweets: A Multilingual Corpus Study
Jihen Karoui, Benamara Farah, Véronique Moriceau, Viviana Patti and Cristina Bosco, EACL 2017.