

algoritmi e comunicazione

il software si sta letteralmente *mangiando*
le nostre parole, che si fa?

stefania iannizzotto

università degli studi di firenze – accademia della crusca



algoritmo = linguistica computazionale



- analisi ed elaborazione del **linguaggio naturale** attraverso l'uso di **metodologie informatiche**
- dare ai computer la capacità di elaborare il **linguaggio naturale**
- **C-3PO** = robot protocollare: si occupa delle comunicazioni tra umani e robot (Lucas, *Star Wars*, 1977)

conoscenze linguistiche

che dovrebbero possedere i computer

- articolare e decodificare i suoni di una lingua [fonetica articolatoria e acustica, fonologia, prosodia ecc.]
- conoscere le parole di una lingua, la loro struttura e la loro organizzazione [lessico e morfologia]
- comporre le parole in espressioni linguistiche complesse (sintagmi, frasi ecc.) [sintassi]
- assegnare significati alle espressioni linguistiche semplici e complesse [semantica lessicale e compositiva]
- usare le frasi nei contesti, situazioni e modi appropriati agli scopi comunicativi [pragmatica]

trattamento automatico dell'informazione applicato al linguaggio naturale

- esplorazione e analisi dei **dati linguistici** con strumenti informatici
- sviluppo di sistemi informatici **dotati di conoscenze linguistiche** e in grado di esibire capacità linguistiche comparabili a quelle umane
- elaborazioni di **modelli computazionali e simulazioni** della competenza linguistica umana, della sua acquisizione e del suo uso

Natural Language Processing (NLP) o Trattamento Automatico del Linguaggio (TAL) – alcune applicazioni

- correttori ortografici, grammaticali ecc.
- recupero “intelligente” di documenti [Information Retrieval]
- riconoscimento automatico del parlato [Automatic Speech Recognition]
- sintesi automatica della voce [Text-To-Speech]
- estrazione automatica di informazione da testi [Information Extraction]
- interrogare documenti attraverso domande in linguaggio naturale [Question Answering]
- traduzione (semi)-automatica di testi [Machine translation]
- interazione (conversazione) uomo-macchina multimodale [agenti conversazionali complessi]

algoritmo o provocazione?

Buongiorno Stefania,

come ti avevo accennato l'altro giorno al telefono ecco un pò più in esteso il tema del wrkshp di digit17 a cui ti ho chiesto di partecipare:

spazio linguistico e CMT

Lo **spazio linguistico italiano** ha delle coordinate:

- **tempo** (diacronìa)
- **luogo** (diatopìa)
- **strato sociale** (diastratìa)
- **situazione comunicativa** (diafasìa)
- **mezzo di comunicazione** (diamesìa)
- [• **tecnologia a supporto della comunicazione** (diatecnìa)]

L'ingresso delle **parole nella Rete** ha provocato un terremoto nello spazio linguistico.

SCRITTO / PARLATO / TRASMESSO **parlato** / TRASMESSO **scritto**

la lingua fruita attraverso **uno schermo** (più o meno grande)

caratteristica tecnica specifica del mezzo
(**gruppo di mezzi**) che determina delle scelte
linguistiche particolari, ad esempio:

- uso di una **sintassi** meno pesante che nella lingua scritta
- **organizzazione del testo** funzionale alla fruizione per blocchi

scrittura “faccia a faccia”

il **parlato** per costrizione del mezzo si trova a essere **intrappolato nella scrittura**:

scrivere un sms, chattare, postare qualcosa su Facebook, twittare o comunicare su WhatsApp è **come parlare** anche se lo si fa su una tastiera (di computer, di telefono, di tablet...)

dimensione dell'oralità

(scrittura volatile, visibile parlare, italiano inviato)

- **punteggiatura intonativa** (sequenze di punti esclamativi, interrogativi, puntini di sospensione)
- **emoticon** (faccine) ed **emoji** (ideogrammi digitali standard, supportati da quasi tutti i dispositivi elettronici)
- **maiuscole** con valore funzionale (per indicare l'urlato o lo scandito)

italiano digitato

- **brevità**
- **frammentarietà sintattica**
- **scarsa importanza data agli errori ortografici**

- **velocità con cui avviene la trasmissione
(anche esigenze tecniche)**

ipotesi

[praticamente **frammenti di testo**]

duepuntozerità: interazione in rete

- scrivono persone che **non scrivevano**
- **tratti linguistici** dell'oralità, dialettali, regionali, gergali e si affiancano ai tratti tecnico-scientifici, letterari, aziendali, burocratici che già pervadevano la rete
- **nuove parole tecniche** legate alle nuove modalità di interazione e agli strumenti per realizzarle (fanno parte del vocabolario quotidiano di moltissimi italiani)

la lingua della rete

- **non è una lingua omogenea**: molteplici tipologie testuali e comunicative
- **tempo**: scrittura rapida, non pianificata, spesso non revisionata (o velocemente revisionata)
- **spazio fisico**: il testo è digitato in coda in qualche ufficio, in sala d'attesa dal medico ecc.
- **espressività grafica**: parole sincopate e abbreviate, punteggiatura smodata, sequenze di cifre e lettere in forma di rebus (**tachigrafie**)

meccanismi di alterazione

- **SIMBOLI**: 6 per **sei** (seconda persona del presente indicativo del verbo *essere*), **x** per **per** (anche all'interno di parola), **+** per **più**; (inglese) 4 per **four/for** o 2 per **two/too** e anche **to**
- **ABBREVIAZIONI**: (per contrazione) **nn** per **non**, **qlk** per **qualche**, **qnd** per **quando**; (per troncamento): **asp** per **aspetta**, **cell** per **cellulare**, **doma** per **domani**, **pome** per **pomeriggio**, **uni** per **università**; (antroponimi) **Ale**, **Cate**, **Edo** ecc.
- **SIGLE**: **tat** (ti amo tanto) , **tvb** (ti voglio bene) , **tvtb** (ti voglio tanto bene) , **tv1mdb** (ti voglio un mondo di bene); **ASAP** (as soon as possible 'appena possibile'), **LOL** (laughing out loud 'ridendo sonoramente')

altri tipi di alterazione

- **USO DI J PER GL** romanesco ma di uso diffuso ormai senza connotazione locale, ad esempio **daje** per **dagli**
- **USI TABUISTICI** per ingannare gli strumenti di controllo, ad esempio **sexo** per **Sesso**
- **LEETSPEAK** caratteri rimpiazzati da altri di forma simile: ad esempio **c4sa** per **casa**, **s3greto** per **segreto**
- **CAMELCASE** scrittura ‘a gobba di cammello’ con lettere maiuscole nel corpo della parola (anche nei marchionimi: **iPhone**, **iPad**, **WhatsApp**)

5 PAZIALE.

NUOVA MINI. ORA CON 5 PORTE.
TUA A 18.900 €*.



5 e S
graficamente
somiglianti
per mettere in
evidenza la
novità delle
5 porte

ipergrafia o graforrea

la confidenza con il mezzo scritto non è priva di ricadute linguistiche e ha modificato il rapporto tra lo scrivente e la **norma**

deproblematizzazione dell'atto scrittorio

diffusione di **usi grafici aberranti** che diventano **consuetudini linguistiche**

testi “a perdere”

l'uso di **forme linguistiche correnti** (diafasicamente basse o addirittura substandard) in virtù del loro impiego frequente appaiono via via meno marcate

il minor controllo sui testi facilita la presenza di **errori** e innalza la **soglia di tolleranza** nei loro confronti, anzi quasi non si riconoscono più!

qualcosa non va?

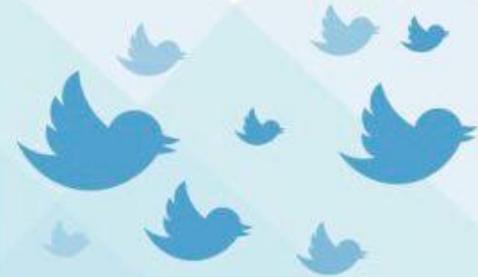
Ore 9,00

apertura saluto autorità



Panel ore 9,30 -12,30

“Il bene comune non sta dentro gli algoritmi, o forse si?”
con Michele Mezza, Derrick de Kerckhove,
Piero Dominici, Mario Rasetti, Matteo Biffoni, Pier Luca Santoro
modera Marco Renzi



Workshop aula 1

ore 13,00 – 15,00

“L’algoritmo della libertà”
con Benedetto Ponti e Claudia Dani

ore 15,00 – 17,00

“L’algoritmo dell’algoritmo”
con Luca Corsato e Vittorio Pasteris

ore 17,00 – 19,00

“Algoritmi e comunicazione”
con Stefania Iannizzotto e Daniele Chieffi

Workshop aula 2

“L’algoritmo della politica”
con Pier Luca Santoro e Lorenzo Pregliasco

“L’algoritmo della notizia iperlocale”
con Marco Giovannelli e Andrea Boscaro

“5G e algoritmi”
con Benedetta Squittieri e Maurizio Decina

tendenza alla semplificazione in certi specifici contesti grafici: le **lettere accentate**

- apostrofo al posto dell'accento (**e'** per **è**, **ne'** per **né**) soprattutto per le lettere maiuscole (**E'** per **È**) prima per iniziali limiti tecnici, adesso frequente anche per imperizia!
- una scelta “economica”: scrivere **e'** al posto della lettera correttamente accentata richiede comunque (oltre che una competenza) anche un tempo maggiore

generi testuali e variabilità nella rete

i diversi servizi CMT mostrano notevole **variabilità**

- posta elettronica e newsgroup sono i **più scritti**
- chat, messaggistica istantanea, microblogging hanno una **struttura semplice** [vincoli tecnici]
- blog e Facebook hanno una **progettualità medio-bassa**

forme di testo tipizzate

[anche per esigenze tecniche] per esempio la struttura dei messaggi di **posta elettronica**: mittente, destinatario, oggetto

- formula di apertura in alto a sinistra
- corpo del testo
- formula di congedo in basso a sinistra

raggruppare le email in conversazioni le rende vicine alla messaggistica istantanea e alle chat [**convergenza delle piattaforme**]

innovazioni tecnologiche e sistema comunicativo

i fatti tecnici modificano sia le **forme della scrittura** sia quelle dell'interazione

l'italiano digitato tenta o di recuperare, con strategie che appartengono a vari livelli linguistici, ciò che viene perso nel **passaggio dal parlato allo scritto**

italiano dell'uso medio o neostandard

è un italiano corretto, ma semplificato che corrisponde più o meno alla lingua che impieghiamo tutti i giorni:

- domani vado **invece di** domani andrò
- adesso mi bevo un caffè **invece di** adesso bevo un caffè
- se lo sapevo non venivo **invece di** se lo avessi saputo non sarei venuto

alcuni tratti dell'italiano dell'uso medio

- uso della forma obliqua del pronome personale di terza persona in funzione di soggetto **lui**
- uso della forma obliqua del pronome personale di seconda persona in funzione di soggetto **te**
- **gli** unificato (lui, loro, lei)
- **che** polivalente
- **dove** polivalente [*sono iscritto in un gruppo dove ci sentiamo*]
- **imperfetto** controfattuale (doppio imperfetto)
- **presente pro futuro**
- forme aferetiche del dimostrativo **sto, sta**
- strutture pleonastiche con **pronomi doppi** [*vi trovo a voi, non mi va a me*]
- forme verbali con pronome lessicalizzato graficamente scorrette [*cosa centra, non cio soldi*]
- sintassi improntata alla **brevità**, periodi **monoproposizionali**, predominio della **paratassi** sull'ipotassi
- abbondano le **frasi nominali**
- strutture sintattiche tematizzanti: **frasi scisse** (*è che, non è che*) **dislocazioni a sinistra**

eteroglossia in rete

inserti alloglotti (inglese e dialetti) e
neoformazioni: contaminazioni tra forme,
strutture e codici anche a fini ludici

sperimentazione che finisce per
fissare convenzioni e canoni

xenismi tecnici

settore lessicale informatico (internet e social)
diffusi anche nella lingua comune

● *software* ● *hardware* ● *reset* ● *username* ●
password ● *screenshot* (fotografia della
schermata di un computer o un cellulare) ●
tag (etichetta)...

neoformazioni collegate ai servizi telematici

verbi formati con base inglese e morfema flessivo della prima coniugazione dell'italiano

● *mailare* ● *spammare* ● *forwardare* ● *trashare* (posta elettronica) ● *chattare* ● *postare* ● *fiammare* ● *lurkare* ● *quotare* ● *spoilerare* ● *trollare* ● *lollare* ● *bannare* ● *taggare*, *friendare* ● *unfriendare* ● *followare* ● *googlare* ● *whatsappare* ● *screeshottare*...

inglese sì, ma consapevolmente

sarebbe saggio seguire i consigli di

Francesco Sabatini:

- essere padroni del significato della parola
- conoscere la pronuncia e la grafia corretta
- assicurarsi che l'interlocutore capisca

rideterminazioni semantiche

ignorare	non rispondere più in chat
esserci	essere collegati
cadere	perdere il collegamento
entrare/uscire	in/da una chat
girare	inoltrare un messaggio di posta elettronica
registrarsi	iscriversi a un servizio
accedere	fare il login
seguire	nei social media

sostituzione del **ch** con la **k**

- gli adolescenti la considerano una caratteristica da **bimbominkia**
- usata in maniera anche ideologicamente connotata [**maskio** o **komunista**]

fooo fooco kcolla coffee per kcolla p^{na} que fo. Concene
afonawo conni Le paffewas profce fo benedico

gli spazi e la loro assenza

(quasi) una rivoluzione della convenzione ortografica

errori

- appart**e** **per** a parte
- appost**o** **per** a posto
- avvol**t**e **per** a volte
- aldil**à** ('oltretomba')

per al di là

- allat**o** **per** a lato (di qualcosa)
- a fatt**o** **per** affatto



L'italiano sta bene, gli italiani meno

- popolazione di **lettori deboli**
- **analfabetismo funzionale** difficoltà nella comprensione dei testi scritti
- incapacità di muoversi tra **registri linguistici diversi** e di gestire la **testualità tipografica [scrittura solida]** e la **testualità digitale [scrittura liquida]**

- dov'è il problema?

Biperio per “Nino Bixio”

Malcolm decimo per “Malcolm X”