

Modulo A

Valutare i sistemi automatici di interazione linguistica

4. Sistemi online

6 ottobre 2016



Linguistica italiana II
Mirko Tavosanis
A. a. 2016-2017

Oggi

- Sistemi: dal locale all'online
- Il caso di Google
 - Ricerca vocale
 - Dettatura all'interno di documenti
 - Traduzione
- Da notare: le complicazioni e le variabili (di cui dovremo tener conto nella pratica)

In locale

- Dragon e il riconoscimento vocale di Windows elaborano in locale – su computer
- Su smartphone, Google sta introducendo su Android capacità di riconoscimento offline, ma non ho ancora fatto controlli
- Nel 2015 Google aveva presentato integrazioni a Google Voice come programma offline per l'inglese dichiarando un tasso di errore del 13%; il programma pesava solo 20 MB e girava su un Nexus 5 (quad-core, 2.26GHz, 2GB RAM):
<https://9to5google.com/2016/03/11/google-accurate-offline-voice-recognition/>
- Sono comuni le capacità di riconoscimento di una singola frase di attivazione («Ehi, Siri», «OK, Google») basate su un chip dedicato e a basso consumo

Online

- I sistemi di maggior successo oggi sono online
 - Vantaggio: non richiedono elaborazione sul posto e quindi possono funzionare anche da tablet o da telefono
 - Svantaggio: per funzionare hanno bisogno di un collegamento Internet attivo e veloce (per telefono, di solito almeno 3G)
- Naturalmente, creano enormi problemi di privacy: le conversazioni vengono registrate, raccolte ed elaborate
- Raramente vengono usati per dettatura: gli usi più significativi sembrano la traduzione in tempo reale e l'interazione con assistenti come Siri e Cortana

Tecnologie di Google

- Google presenta diverse funzioni in modo un po' nascosto – a volte usando l'etichetta «Google Voice»
- Ha iniziato a supportare i servizi vocali nel 2009, usando come modello acustico il Gaussian Mixture Model (GMM), rafforzato con altre tecniche
- Dal 2012, per il riconoscimento del parlato ha iniziato a usare le Recurrent Neural Networks (RNNs), e in particolare un'architettura proposta nel 1997 e chiamata LSTM RNNs (“Long Short-term Memory Recurrent Neural Networks”)
- Credo che questo lo strumento usato ancora oggi per tutte le lingue, incluso l'italiano (non ho trovato informazioni più specifiche)
- Le reti neurali richiedono allenamento, e i corpora usati per altre attività di elaborazione del linguaggio non aiutano molto, perché sono composti soprattutto di testi scritti, molto lontani dall'uso parlato normale: ne parleremo più avanti

Google Voice Search

- Usa il riconoscimento del parlato per fare ricerche
- Nella home di Google, aperta con Chrome, compare l'icona del microfono al termine della casella di ricerca



Cerca con Google

Mi sento fortunato



Cerca su Google o digita l'URL



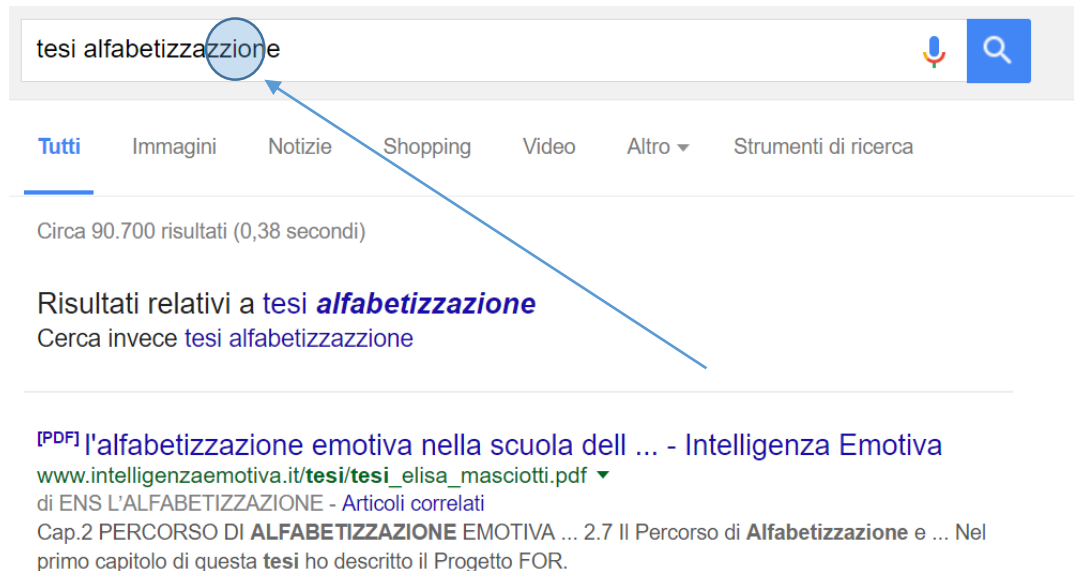
In alcuni casi c'è anche una risposta a voce... e non mi è ancora chiaro quale sia il criterio per cui a volte c'è e a volte no

Google Voice Search

- Nella mia esperienza, il livello di riconoscimento è molto alto
- Un sistema del genere è utile, ovviamente, soprattutto su dispositivi **mobili**
- Sundar Pichai ha dichiarato (Google I/O 2016) che il 20% delle ricerche da mobile adesso è vocale
- Sembra molto probabile che le percentuali possano aumentare rapidamente
- Una panoramica e qualche dato per il mercato americano:
<http://searchengineland.com/voice-search-explosion-will-change-local-search-251776>

Google Voice Search

- Non fa una trascrizione «pura»: ottimizza il lavoro per la ricerca e fornisce una versione già ottimizzata (evitando molti errori)
- Il meccanismo è diverso da quello usato per la ricerca testuale:



- A volte la trascrizione «pura» compare nella casella di ricerca, ma viene immediatamente sostituita dalla versione ottimizzata
- Qualche mese fa Alberto Guaita ha provato diversi esempi di dettatura

Google Voice Search: forzature

Individuate da Alberto Guaita a settembre, in alcuni casi si replicano e in altri no (dipende dalla voce?)

- Replicabile: «alchemico» > Alkemico (commerciale)
- Replicabile: «conto cocente» > conto corrente
- Replicabile: «Il pesce non **ha bocca**» > Il pesce non **abbocca**
- A settembre, «Tanto va la gatta al **largo**» > Tanto va la gatta al **lar**do
- Oggi, «Tanto va la gatta al **largo**» > Tanto va la gatta al **largo**

Quali sono le ragioni linguistiche (e no) dietro a questi comportamenti?

Google Voice Search: idiosincrasie

Individuate da Alberto Guaita e verificate da me a metà settembre, ma oggi non più replicabili:

- Ricercando numeri a 5 cifre che terminano con 991, 992, 993, e così via, le migliaia venivano interpretate come il numero di una legge italiana e il “99x” diventava l’anno di promulgazione: “23.994” diventava così la legge 23 del 1994. Per la corretta ricerca occorreva pronunciare le due componenti del numero come se fossero due numeri distinti, ossia prima 23 e poi 994; il sistema poi le univa in un numero solo.
- Sempre con i numeri, non era possibile chiedere di dividere 5 per un numero (per esempio, “5/73” o “5/40”), poiché il programma cambiava il 5 in una V (numero romano) e modificava la ricerca
- Frasi come “togli la giacca” o “spediscili via” venivano riportate senza il pronome enclitico

Il sistema, evidentemente, è aggiornato con frequenza

L'importanza del contesto

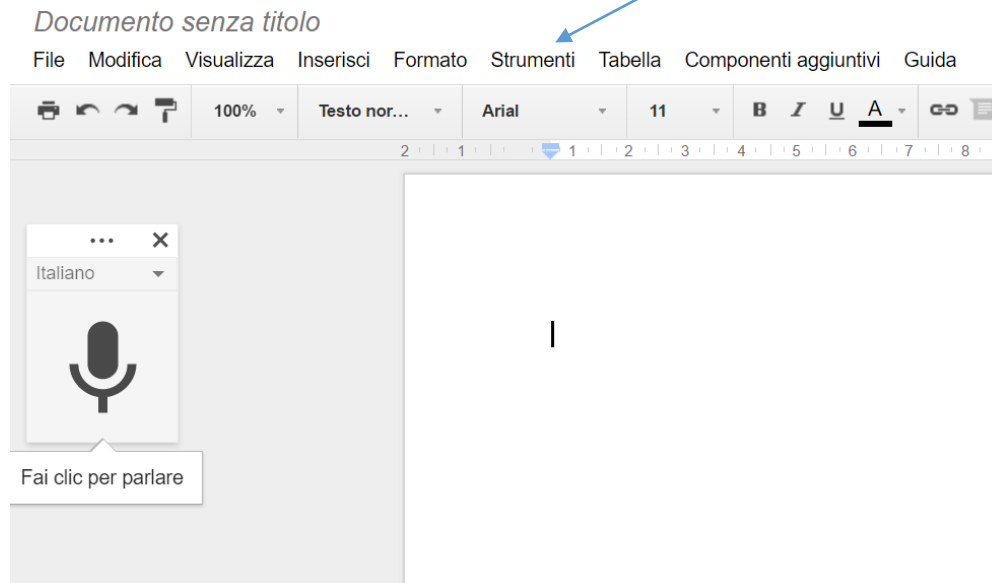
Pronunciando la stessa parola in contesti diversi si ottengono trascrizioni diverse

- «Gnu» > new
- «Gnu animale» > GNU animale (e risultati relativi agli animali)
- «Gnu sistema operativo» > new sistema operativo

Naturalmente si ottengono anche risultati diversi

Dettatura nei documenti

- Google permette di dettare all'interno dei file di Google Docs
- Si attiva (in modo non troppo intuitivo) da: Strumenti > Digitazione vocale oppure da Ctrl + Maiusc + S



- Non è richiesto addestramento, e riconosce voci molto diverse
- Al momento supporta 43 lingue
- Istruzioni: Guida di editor di documenti, registrazione vocale <https://support.google.com/docs/answer/4492226>

Inserimento di parole straniere

- In linguistica si distingue a volte tra:
 - **Mescolanza**, o *code mixing*: mescolo le due lingue all'interno di una **frase**
 - **Commutazione**, o *code switching*: all'interno di un **discorso**, inserisco frasi intere in lingue diverse
- In entrambi i casi, i sistemi di dettatura hanno grossi problemi a gestire il cambiamento – anche tra lingue supportate
- In buona parte è un problema di pronuncia: la pronuncia di uno straniero è spesso comprensibile, ma molto diversa rispetto a quella di un madrelingua
- Nella mia esperienza, Google gestisce bene la mescolanza con l'inglese, anche quando si tratta di parole inglesi non inserite nei dizionari italiani; le frasi in inglese invece rappresentano un problema, anche impostando la lingua a «inglese»
- Qui sarebbe interessante un confronto linguistico dettagliato!

Parole inventate

Il lonfo non vaterca né gluisce
e molto raramente barigatta,
ma quando soffia il bego a bisce bisce,
sdilenca un poco e gnagio s'archipatta.

(Fosco Maraini, *Il lonfo*, in *La gnosi delle fanfole*)

Il riconoscimento viene evidentemente accoppiato all'esame dei testi
sul web

Punteggiatura

Le indicazioni dicono che è possibile «aggiungere punteggiatura al testo» dando a voce questi comandi:

- Punto
- Virgola
- Punto esclamativo
- Punto interrogativo
- Nuova riga
- Nuovo paragrafo

Nota. La punteggiatura funziona in tedesco, inglese, spagnolo, francese, italiano e russo

Comandi

Sono disponibili comandi vocali per il testo, ma solo in inglese: sia l'account che il documento devono essere impostati sulla lingua inglese

«Una volta avviata la digitazione vocale, puoi utilizzare i comandi per modificare e formattare il documento. Ad esempio, puoi dire "Select paragraph" (Seleziona paragrafo), "italics" (corsivo) o "Go to the end of the line" (Vai alla fine della riga).»

- Selezionare testo
- Formattare il documento
- Modificare il documento
- Aggiungere e modificare tabelle
- Spostarsi all'interno del documento
- Riattivare la digitazione vocale
- Comandi per aprire la guida
- Lettura (per l'accessibilità)

Google Translate

- Supporta circa 100 lingue (10.000 combinazioni di lingue)
- La qualità del supporto è molto variabile: in alcuni casi è possibile la dettatura, in altri no; in alcuni casi c'è la lettura ad alta voce, ma la qualità delle voci varia molto
- Verifica: traduzione in cinese e in swahili

Google Translate

- La qualità è sorprendente
- L'integrazione tra STT, TTS e traduzione automatica produce effetti sorprendenti

The screenshot displays the Google Translate interface. At the top left is the Google logo. To the right are a grid icon, a notification badge with the number '3', and a user profile picture. Below the logo, the word 'Traduttore' is written in red. To its right is the text 'Disattiva traduzione istantanea' and a star icon. The main interface features a language selection bar with 'Swahili', 'Italiano', 'Inglese', and 'Trovato Italiano' (with a dropdown arrow). A double-headed arrow icon is positioned between the source and target language boxes. The source language box is set to 'Inglese' and the target language box is set to 'Swahili'. A blue 'Traduci' button is located to the right of the target language box. Below the language boxes are two text input areas. The left area contains the text 'Prova di traduzione con Google Translate' and has a small 'x' icon in the top right corner. Below this text are icons for a speaker and a pencil. The right area contains the text 'Test translation with Google Translate' and has icons for a star, a document, a speaker, and a share icon, along with a pencil icon in the bottom right corner. At the bottom of the page, there is a footer with links for 'Informazioni su Google Traduttore', 'Community', and 'Per cellulari', followed by the Google+ and Blogger logos.

Microsoft Skype Translator

- Consente la traduzione in tempo reale tra diverse coppie di lingue attraverso Skype
- Il risultato è sorprendentemente funzionale: nel maggio di quest'anno, al termine del corso di Linguistica italiana II abbiamo fatto prove di traduzione italiano-cinese (e viceversa), riuscendo a scambiare informazioni pratiche in modo funzionale
- Non permette di interagire direttamente con il computer o di produrre testi scritti: agisce solo come intermediario per interazione tra esseri umani
- Al momento le lingue supportate sono: inglese, francese, tedesco, cinese mandarino, italiano, spagnolo, portoghese
- Attivo sperimentalmente da fine 2014 (l'italiano è stato aggiunto nel maggio 2015)

