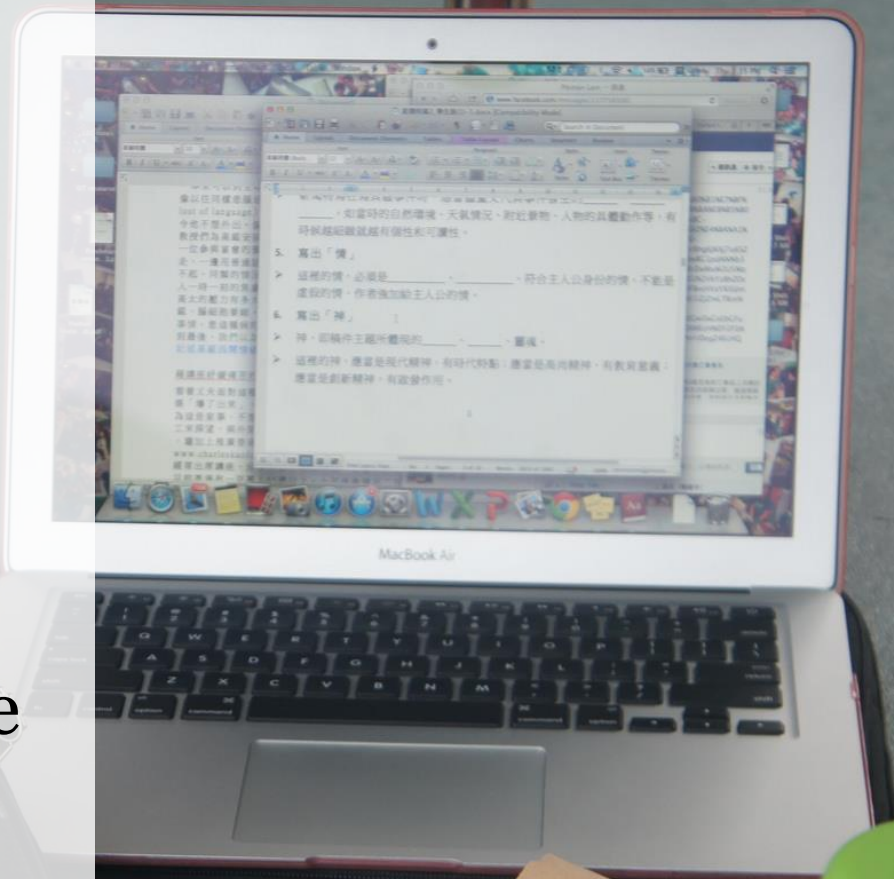


Modulo B

Valutare i sistemi automatici di interazione linguistica

1. Punto della situazione

23 febbraio 2017



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

Calendario

Date possibili:

- 23 febbraio
- 1, 8, 9, 15, 16, 22, 29, 30 marzo
- 5, 12, 13, 19, 20, 26, 27 aprile
- 3 e 4 maggio

(Alcune date, come il 2 marzo, sono bloccate per sovrapposizione con il corso intensivo di Knowledge management)

Probabilmente salteremo una o due lezioni tra marzo e aprile e le recupereremo il 10 e l'11 maggio

Dove stiamo andando?

- Tiriamo le fila (oggi):
 - prospettive imminenti
 - comportamenti nuovi notati in questi mesi
 - alcuni dati utili sul riconoscimento
- Altri tipi di interazione: traduzione automatica
- Elaborazione del testo
- Generazione automatica di testi

Prospettive imminenti



Prospettive imminenti

- Probabilmente nei prossimi mesi ci sarà il lancio anche in Italia di alcuni sistemi innovativi
- Riusciremo a parlarne? Vedremo
- Intanto, una rapida rassegna dei prodotti di ampio consumo che potrebbero essere interessati, da Amazon Echo a Google Home

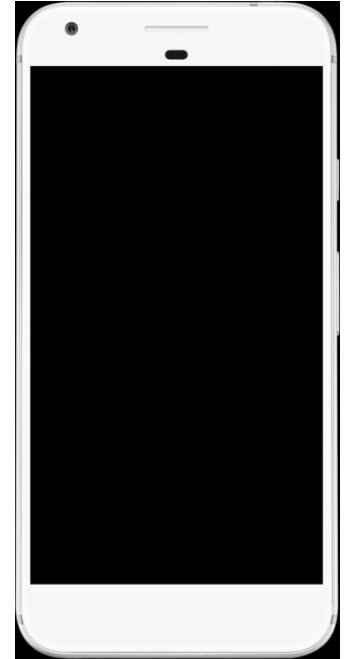
Possibili: Amazon Echo e Alexa

- Dall'autunno scorso esiste già una versione in tedesco (prima versione in una lingua diversa dall'inglese)
 - ... anche se le poche persone che hanno provato il prodotto dichiarano che la funzionalità è molto inferiore, anche a livello di riconoscimento del parlato
<https://earlymoves.com/2016/11/24/alexas-german-version-still-in-kindergarten/>
- I rappresentanti dei termostati intelligenti Nest in Italia hanno dichiarato che l'assenza del controllo vocale è solo momentanea perché Alexa arriverà presto anche in Italia... anche se sarebbe più naturale usare Google Home
<http://www.dday.it/redazione/22359/amazon-echo-italia-uscita>
- Un centro di ricerca Amazon è stato aperto a Torino
<http://www.lastampa.it/2016/10/26/tecnologia/news/amazon-sviluppa-echo-per-litalia-alex-parler-anche-in-italiano-00sL1XrBnHxUBu2jRv6YIK/pagina.html>

Possibili: Google Assistant

Disponibile al momento in tre sfumature diverse (informazioni ufficiali del sito Google):

- **Con Google Allo**
In Google Allo, the Google Assistant is available in English, German, Hindi, Japanese, and Portuguese. More languages are coming soon.
- **Su telefoni Pixel (non ordinabili in Italia)**
On Pixel phones, the Google Assistant is available in English and German. More languages are coming soon.
- **Con Google Home (non ordinabili in Italia)**
With Google Home, the Google Assistant is available in English only.



Possibili: Bixby / Viv

- Samsung ha annunciato l'integrazione di un assistente vocale, Bixby, per il nuovo telefono di punta, Galaxy S8
- Non ci sono ancora informazioni precise e indicazioni sulle lingue supportate (una notizia non confermata parla di «otto lingue», ma non specifica quali)
- Non si sa neanche se sarà basato su Viv o su S-Voice (di basso profilo)
<http://www.theverge.com/circuitbreaker/2017/2/10/14573954/samsung-ai-assistant-bixby-viv-s-voice>
<http://www.theverge.com/circuitbreaker/2017/2/8/14545908/samsung-new-ai-assistant-bixby-galaxy-s8>
- Comunque, l'inizio della commercializzazione del Galaxy S8 è previsto per aprile



Comportamenti notati in questi mesi

Comportamenti notati in questi mesi

- In alcuni casi, gli assistenti mostrano una limitata capacità di gestire una conversazione con più di un turno
- Google now può rispondere a una serie di domande sul meteo («che tempo farà domani a Pisa?»; risposta; «e dopodomani?»; risposta; ecc.)
- Cortana fornisce uno stimolo a domande ulteriori dopo aver risposto ad alcune domande... però questo non sono ancora riuscito a verificarlo
- Forse queste funzioni erano già presenti l'anno scorso: è molto difficile documentare evoluzioni di questo genere

Rumore e riconoscimento del parlato

- Le prestazioni di sistemi come Google e Dragon sono pesantemente collegate alle caratteristiche audio
- Il passaggio dal messaggio interno agli altoparlanti esterni ha prodotto in alcuni casi (esaminati da Beatrice Caputo su DGmtB01L) un raddoppio del WER:
 - Da 32,5 a **70,7** per Google
 - Da 34,9 a **67,8** per Dragon
- Ovviamente, qui dipende moltissimo dalla qualità degli altoparlanti e dei microfoni! E in un caso (solo) la qualità del riconoscimento tramite microfono è stata leggermente superiore a quella tramite messaggio interno
- In alcuni casi forse le differenze che ho notato si spiegano in questo modo
- Inoltre, il risultato migliora se si separano le voci: da 19 omissioni a 7 nel caso di Google

Dato generale: gli accenti regionali non sono un problema

- Il riconoscimento di Dragon e Google è buono anche su accenti molto marcati; quando ci sono errori, è raro che la causa si possa ricondurre a una questione di pronuncia
- Chiaramente, in molti casi questa prestazione è dovuta alla semplice mancanza di alternative
- Una pronuncia come *automobbile* (con rafforzamento della consonante occlusiva sonora) è foneticamente molto diversa dalla pronuncia dell'italiano standard, tuttavia non è facile confonderla con altre parole: sia gli esseri umani sia le macchine la riconducono facilmente ad *automobile*
 - In alcuni, rarissimi, casi, Google fornisce trascrizioni che non esistono nel vocabolario italiano: per LFp2A02N, turno 20, si ha un *progeva* al posto di *porgeva*
- In altri casi, però, sequenze di suoni foneticamente simili *potrebbero* essere confuse

Sequenze plausibili: Napoli

- Per esempio, nella registrazione LMp2A02N (Napoli) Alessandro Giosi aveva notato che la sequenza “camion rosso”, pronunciato con pesante accento per chiusura metafonetica, veniva interpretata da Google come “camion russo” (da 0:25)
- Tuttavia, quando ho provato a controllare, né Google né Dragon hanno dato la stessa risposta: in entrambi i casi la trascrizione è stata “camion rosso”



Sequenze plausibili: Parma (1)

- Nella registrazione DGmtA04E (Parma) Sofia Ghisellini ha notato che alla pronuncia “all’esterno” diventava “ha le stanno” in evidente rapporto col fatto che la -r- veniva pronunciata con la caratteristica pronuncia uvulare emiliana, portata quasi alla cancellazione (da 4:11)
- Neanche qui sono stato in grado di riprodurre il problema: tanto Google quanto Dragon mi hanno fornito la trascrizione “all’esterno”



Sequenze plausibili: Parma (2)

- Nella registrazione DGmtA04E (Parma) Sofia Ghisellini ha notato che alla pronuncia con *e* aperta di *tre* nella sequenza “tre dita” corrispondevano trascrizioni sbagliate
- Google: “tradita”
- Dragon: “eredità”
- Qui la situazione è un po’ diversa perché un altro fattore è sicuramente la velocità di enunciazione

Inoltre...

- È difficile precisare quanto debba essere marcata la pronuncia per produrre problemi: non esistono sistemi per misurare sinteticamente quanto un accento regionale sia diverso dallo standard
- Le differenze possono riguardare piani diversi: ritmo, intonazione, pronuncia delle vocali o delle consonanti, e così via
- Due accenti locali possono sembrare ugualmente lontani dallo standard a un ascoltatore, e al tempo stesso esserlo per ragioni molto diverse
- Quanto contano i diversi piani, o la loro interazione, per i sistemi di riconoscimento del parlato? Non esistono informazioni pubblicate, e sospetto che nemmeno le aziende che li sviluppano abbiano dati del genere.

Risultati dimostrabili

- I problemi dovuti agli accenti regionali *in quanto tali* sono veramente pochi
- Molto più consistenti sono i problemi dovuti alle caratteristiche della lingua parlata; non abbiamo fatto un controllo mirato, ma l'opinione di molti (e mia) è che il singolo fattore più importante sia la velocità del parlato
- I sistemi di riconoscimento hanno molte difficoltà con le espressioni olofrastiche (sì) e con sequenze come «ce l'ho» (Caputo)
- In ogni caso, Google mostra prestazioni leggermente superiori a Dragon con quasi tutte le registrazioni, e quindi con quasi tutti gli accenti – le eccezioni sono poche

Ipotesi

- Dal lavoro sono venute fuori diverse ipotesi sul funzionamento dei sistemi di riconoscimento del parlato
- Per esempio, potrebbero esistere «gruppi semantici»: una parola si porta dietro parole connesse... qualcosa che ha a che fare con le collocazioni, probabilmente
 - In molti casi, io ho la sensazione che non esista niente del genere e che le parole siano sbagliate in modo casuale
- Potrebbero esistere anche «gruppi sintattici»: tentativi di fornire un output regolare cambiando le parole in modo da avere una sintassi regolare
- Altra ipotesi: la presenza di duplicazioni («devo devo») produce un numero di errori più alto di quello che si ha normalmente

Applicazioni: le API

- Vincenzo Caproni ha preparato un'interfaccia per sottoporre file audio alle API di diversi servizi di riconoscimento vocale
- La cosa è utile per valutare le prestazioni astratte: si saltano i problemi di registrazione visti con Google
- Tuttavia, in questo caso non si simula più il comportamento di un prodotto reale rivolto all'utente finale
- Sono accessibili in questo modo servizi cloud come Microsoft Azure, Google Speech API e così via