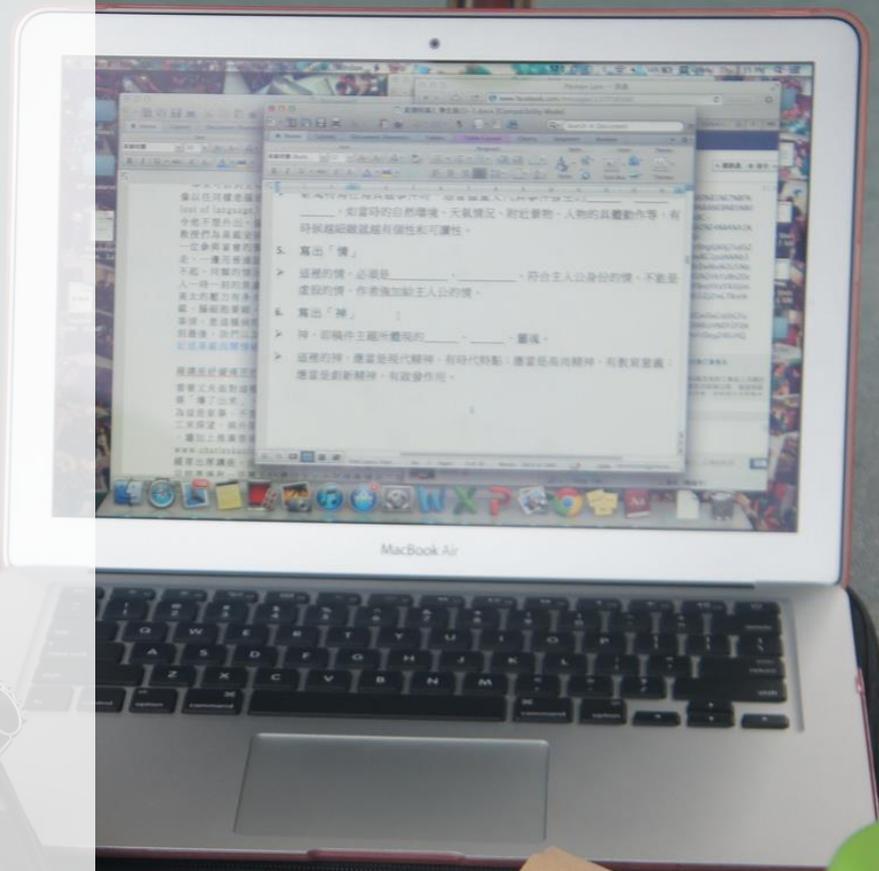


Modulo A

Valutare i sistemi automatici di interazione linguistica

13. I dialoghi di CLIPS

16 novembre 2016



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

Questionari per la didattica

- Sono disponibili dal 31 ottobre al link <https://esami.unipi.it>
- Siete invitati a compilarli già adesso («durante lo svolgimento del corso, quando siano stati svolti i 2/3 delle lezioni»)

Proseguendo

- Vediamo una pronuncia completamente diversa:
Firenze
- Un file da Dialogico – Firenze – Map Task:
DGmtA01F.wav
- Con i dialoghi è possibile prendere in esame anche i fenomeni del «parlato», il neostandard («ce n’hai tre, te») e così via
- Tuttavia il dialogo produce molti problemi, anche a livello alto: i sistemi di dettatura non sono pensati per avere a che fare con più interlocutori contemporaneamente

Problemi con i dialoghi

- Le registrazioni di dialogo si sono rivelate più complesse da gestire rispetto al previsto
- Per ora cancelliamo le indicazioni di lunghezza per gli elaborati fornite la settimana scorsa
- Spero di poter dare quelle definitive la prossima settimana

Dialogo

- File di riferimento, da Dialogico – Bari – Map Task:
DGmtA02B

SPEAKERS_inf.

INp1: S. R., M, 26, Bari, fluente, G>F

INp2: M. P., F, 24, Bari, fluente, F>G

RECORDING_inf.

**Trascrizione ortografica (da verificare):
767 parole**

TYP: DAT

LOC: Bari

DAT: 26/03/01

DUR: 4.29,507

CON: buone

Dal turno 63 i parlanti si invertono

Accorgimenti pratici e criteri

- Codifica: è ANSI e a seconda dell'editor va prima convertita in UTF-8 (ma non mi sembra ci sia nessuna perdita di caratteri)
- Ripulitura del file: meglio farla prima di verificare la trascrizione
- Nella versione definitiva ho tolto tutti gli <eeh>, <ehm>, <mbè> e così via: non sono parole normalmente usate in un testo scritto, quindi non dovrebbero essere trascritte
 - Possibili **eccezioni**, per interiezioni che hanno un uso scritto abbastanza riconosciuto:
 - <ah>
 - <eh>
 - forse anche <ehm>?
 - Occorre controllare l'uso scritto! Anche su questo darò linee guida
- Lo stesso vale per gli allungamenti di vocale una <aa> e simili: le parole devono andare in un testo scritto e quindi devono essere trascritte nella forma ortografica normale

Sovrapposizioni tra turni

Se due turni di locutori diversi si sovrappongono del tutto in parte, nella trascrizione:

- a) La sequenza cui si sovrappone la produzione dell'altro locutore è annotata tra cancelletti #... ..#,
- b) Essa è preceduta dalla notazione tra < > del codice di turno cui è sovrapposta;
- c) Lo stesso avviene nel turno corrispondente dell'altro informatore:

Esempio:

p1G#7: #<F#8> <lp> sulla# sinistra <sp> c'è scritto fiume

p2F#8: #<G#7> no # <lp> non c'è

Questo crea problemi, anche se nella realtà le sovrapposizioni sono meno estese di quanto risulta dalle trascrizioni

Audio e trascrizione

Un minuto e 20 secondi:

p1G#37: <vocal> giri completamente intorno alla #<F#38> macchina
rossa#

p2F#38: #<G#37> <ah> devo girare intorno# alla macchina

Qui le voci si sovrappongono in effetti solo su <ah>

Due minuti:

p1G#53: <tongue-click> e / e ci so+ / a me c'è anche una macchina blu
#<F#54> sopra <sp> devi passare <sp> tra questa macchina rossa e la
macchina blu di <inspiration> <laugh> {<NOISE> <lp>} <vocal>#

p2F#54: #<G#53> {[whispering] <laugh>} e a me non c'è <inspiration>
cioè a me c'è ma è quella che tu hai indicato come# rossa <lp> #<G#55>
<laugh>#

Qui le voci si sovrappongono solo su «e a me non c'è»

Rese dei dialoghi

- Sono comunque un grosso problema
- Intanto, i programmi di trascrizione non distinguono i diversi parlanti
- In alcuni casi chiaramente mettono in sequenza parole provenienti da enunciati diversi
- Alla radice, i programmi fanno una trascrizione molto ridotta:
 - Le parole originali sono **767**
 - Dragon ne trascrive solo **530**
 - Google arriva al massimo a **481**, nel modo che vedremo tra poco
- Causa apparente di alcuni problemi: la riduzione automatica delle duplicazioni

Duplicazioni eliminate

Originale:

u+ un po' più su sulla sinistra (a_3)

un po' più su sulla sinistra sì (b_4)

sì sì è è lì la partenza ora devi fare un giro intorno al televisore in
in basso (a_5)

Dragon:

si (b_2)

un po' più su sulla sinistra (a_3)

può più su sinistra (b_4)

e la partenza ora di fare un giro intorno al televisore in bosso (a_5)

Google:

si (b_2)

È un po' più su sulla sinistra (a_3)

(b_4)

Della partenza ora devi fare un giro intorno al televisore In basso
(a_5)

Problemi pratici

- Con Google la trascrizione ogni tanto si blocca: sottoponendo il file audio originale ne sono uscite solo **165** parole su **767**
- La situazione migliora inserendo delle pause all'interno del file: **202** parole
- Migliora ancora di più quando durante le pause si interrompe manualmente la trascrizione: **481** parole
- La separazione delle due tracce stereo non risolve alla radice il problema:
 - Mono traccia a (maschio): **214** parole
 - Mono traccia b (femmina): **204** parole
- All'interno di ogni traccia è però possibile sentire la voce dell'altro partecipante, che ogni tanto viene interpretata da Google (questo è un limite delle registrazioni CLIPS)

Da un canale all'altro

- Registrazione con pause mono a

La partenza e viene il televisore giù in basso è un po' più su sulla sinistra si va di fare un giro intorno al televisore in basso verso il basso del foglio

- Registrazione con pause mono b

La partenza è *un po' più sulla sinistra* c'è un cerchio sotto il televisore una linea che parte dalla macchina ok quindi quindi verso di nuovo il televisore argento

- Le parole in grassetto sono pronunciate da a, non da b; quelle in corsivo sono pronunciate (in momenti diversi) dai due parlanti

Trascrizioni

- Anche con Dragon, la distorsione è tale che in alcuni casi non è chiaro come devono essere segmentati i turni
- Per esempio: a17, b18, a19
- Con Google si può (forse) avere un'idea guardando lo sviluppo, ma con Dragon è impossibile
- Quindi: consapevolezza del fatto che la classificazione è difficile

WER

Dragon

- Parlante a: 53,9 (22,7 sostituzioni, 30 cancellazioni, 1,3 inserimenti)
- Parlante b: 71,4 (28,6 sostituzioni, 39,3 cancellazioni, 3,4 inserimenti)
- Totale: 58,6

Google

- Parlante a: 45,7 (12,1 sostituzioni, 32,3 cancellazioni, 1,3 inserimenti)
- Parlante b: 74,3 (15 sostituzioni, 56,8 cancellazioni, 2,4 inserimenti)
- Totale: 53,4
- Evidentemente: Google riconosce poco, ma bene