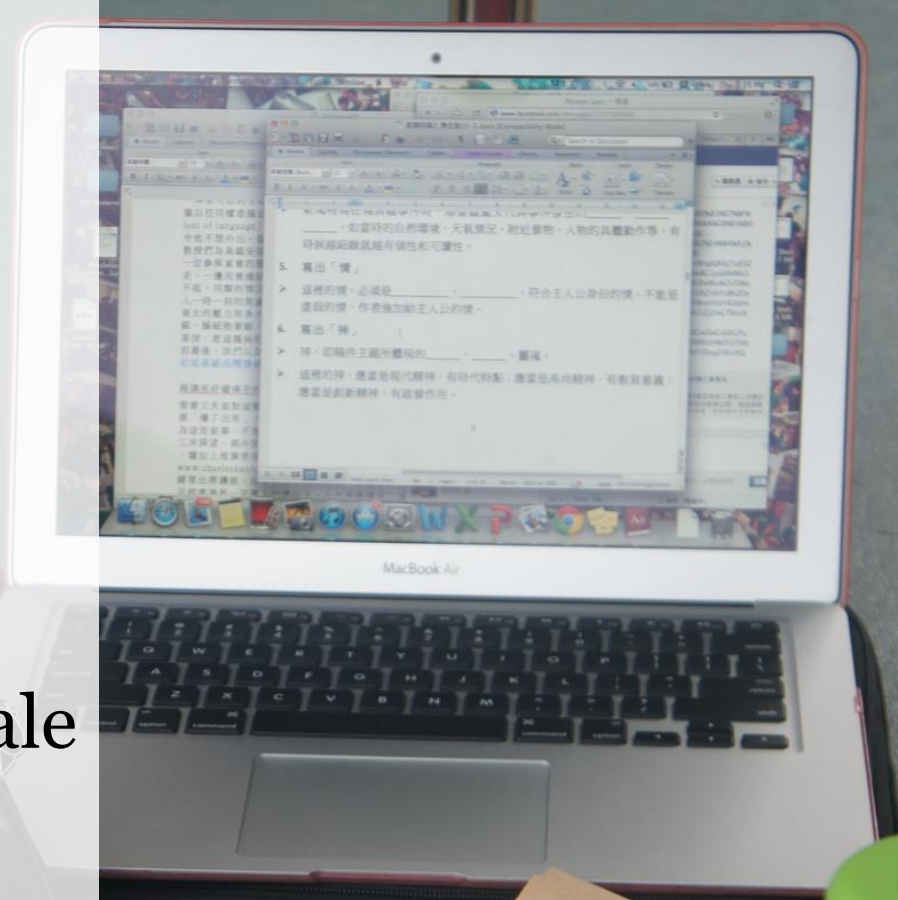


Modulo A

Valutare i sistemi automatici di interazione linguistica

10. Variabilità individuale

2 novembre 2016



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

Diversità del parlato

- Vediamo che anche una frase identica può essere realizzata in modi molto diversi
- La diversità può corrispondere:
 - ai tratti indicali (correlati all'età, al sesso, all'individuo)
 - ai tratti tipici di una comunità (accenti locali)
- Non mi è ancora chiaro quali siano i limiti del riconoscimento del parlato in questa direzione... è un buon progetto di ricerca
- Passeremo poi a vedere la codifica

Analisi degli spettrogrammi

- Salvo sorprese, non ce ne occuperemo, ma...
- Le informazioni fornite dagli spettrogrammi sono molto più complete rispetto a quelle fornite dalla forma d'onda (che in sostanza misura solo l'intensità del suono)
- Un programma molto semplice per vedere gli spettrogrammi: Spectrogram
<http://www.softpedia.com/get/Multimedia/Audio/Other-AUDIO-Tools/Spectrogram.shtml>
- La differenza dei parlanti appare più evidente

Scheda di intestazione CLIPS

TEXT_inf.

MAT: mt Tipo di corpus: Map Task
MAP: A Codice della mappa (ci sono due mappe: A e B)
TRM: Trasmissione
Ndl: 02 Numero dialogo
Ntr: Numero trasmissione
Nls: Numero lista di lettura
REG: G Varietà regionale = codice città = Genova

TRANSCRIPTION_inf.

DAT: 09/04/01
CMT:
Nst: 121 Numero di stringhe

p1G#1: allora Ste', <breath> la partenza alla sinistra della televisione <lp> quindi questo è il punto di partenza

Turno 1 del parlante 1 con ruolo di Instruction Giver

SPEAKERS_inf.

Iniziali, sesso, età, luogo di nascita

INp1: L. C., M, 22, Genova, spontaneo, fluente, G>F (Instruction Giver / Instruction Follower)
INp2: S. P., M, 21, Genova, spontaneo, fluente, F>G

RECORDING_inf.

TYP: DAT Tipo di registrazione
LOC: Genova
DAT: 13/03/01
DUR: 4.46,991
CON: buone Condizioni

Indicazioni

- Per il parlato dialogico, l'unità di trascrizione è il turno dialogico
 - Le sovrapposizioni sono annotate tra parentesi graffe {...}
 - Le pause vengono indicate tra parentesi uncinate <> che possono essere riempite in vario modo
 - <sp> = pausa breve
 - <lp> = pausa lunga
 - <P> = pausa lunga che interrompe un enunciato
 - <ehm> e simili
- Per il parlato televisivo, è l'enunciato
- Per i testi di lettura, sono gli elementi previsti (parole o frasi)

Altri progetti

- Corpus ADAM: ho ricevuto le trascrizioni in formato XML
- Mancano i file audio, che potrebbero essere molto interessanti
- Diamo un'occhiata alle trascrizioni

Sezione di file

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<?xml-stylesheet type="text/xsl" href="conceptual.xsl"?>
<!--          -->
<!-- Project SI-TAL (Italy), Mar 2001          -->
<!-- Provider ITC-irst, ref: Roldano Cattoni, cattoni@itc.it          -->
<!--          -->
<!-- $Id: dial_001_con.xml,v 4.1 2000/06/28 20:35:45 cattoni Exp $ -->

<!DOCTYPE conceptualannotation SYSTEM "concetal.dtd">

<conceptualannotation id="dial_001">

<turn id="t_001" href="dial_001_tra.xml#id(t_001)">
globetrotter viaggi buongiorno
<concept id="c_001" ctype="affiliation">
<slot sname="value" svalue="globetrotter_viaggi"/>
</concept>
</turn>
```

<turn id="t_002" href="dial_001_tra.xml#id(t_002)">

buongiorno senta io vorrei prenotare un viaggio in treno da roma a verona

<concept id="c_002" ctype="trip">

<slot sname="transportation-type" svalue="train"/>

<slot sname="origin" svalue="roma"/>

<slot sname="destination" svalue="verona"/>

</concept>

</turn>

<turn id="t_003" href="dial_001_tra.xml#id(t_003)">

un treno da roma a verona

<concept id="c_003" ctype="trip">

<slot sname="transportation-type" svalue="train"/>

<slot sname="origin" svalue="roma"/>

<slot sname="destination" svalue="verona"/>

</concept>

</turn>