

# Il Rilevamento Automatico dell'Inganno nel Linguaggio Parlato



Frank Enos  
Columbia University  
Department of Computer Science  
22 Maggio 2006



## Columbia University Speech Lab

Emozione  
Carisma  
Inganno

Espressioni  
discorsive:  
"Okay"




AuTOBI



Broadcast news  
summarization,  
in inglese ed in arabo.




## Definire l'Inganno

- La Menzogna:
    - ◆ Una persona intende trarre in inganno un'altra deliberatamente, senza avvertire delle sue intenzioni e senza che il destinatario dell'inganno gliel'abbia esplicitamente chiesto. (Ekman '01)
  - Spesso per ottenere il vantaggio del destinatario.
  - Esclude:
    - ◆ Autoinganno
    - ◆ Teatro
    - ◆ Menzogne dovute all'ignoranza / all'errore
    - ◆ Comportamento patologico
- 




## Perchè studiare l'inganno?

- Giurisprudenza / ordine pubblico
  - Servizi segreti / sicurezza
  - Commercio
  - Politica
  - Psichiatria
  - Rapporti interpersonali
    - ◆ E mai giusto mentire?
- 




## Perchè studiare l'inganno?

- Cosa rende il linguaggio parlato “credibile”?
  - Com'è che le persone individuano le bugie?
  - Qual è la relazione fra l'inganno ed altri fenomeni ‘soggettivi’ nel linguaggio parlato: l'emozione, il carisma, ecc.?
- 



## Difficoltà nello studiare l'inganno

- La maggior parte della gente fa fiasco nell'individuare le bugie — accuratezza di ~50% (Ekman & O'sullivan 1991, Aamodt 2006, etc.)
  - Le persone usano impressioni soggettive — l'emozione, ecc.
  - Riconoscere l'emozione è difficile.
  - Le persone sono diverse.
- 

## La maggior parte della gente fa fiasco...

Group	#Studies	#Subjects	Accuracy %
Criminals	1	52	65.40
Secret service	1	34	64.12
Psychologists	4	508	61.56
Judges	2	194	59.01
Police	8	511	55.16
Federal officers	4	341	54.54
Students	122	8,876	54.20
Detectives	5	341	51.16
Parole officers	1	32	40.42

## Difficoltà nello studiare l'inganno

- E difficile trovare dati utilizzabili
  - ◆ Dal mondo reale
  - ◆ Dal laboratorio
- Questioni etiche
  - ◆ Privacy
  - ◆ Diritti del soggetto
  - ◆ Pretese di successo
- Ma ci sono anche imperativi etici:
  - ◆ Necessità di metodi affidabili
  - ◆ Svelare metodi difettosi
  - ◆ False confessioni



## Frank Dice delle Bugie

*Maria:* Vado a comprare biglietti per *Messiah* di Händel per me e qualche amico — vuoi venire?

*Frank:* Quando è?

*Maria:* Il 19 dicembre.


*Frank:* Beh... il 19...

*Maria:* Vengono le mie due amiche di scuola, e la Robin...

*Frank:* Con piacere!



## Come mentire (Ekman '01)

- Dissimulare
  - Falsificare
  - Celare la vera causa dell'emozione
  - Dire la verità per sviare i sospetti
  - Dissimulazione a metà
  - Suggestire una conclusione sbagliata
- 



## Frank Dice delle Bugie

*Maria:* Vado a comprare biglietti per *Messiah* di Händel per me e qualche amico — vuoi venire?

*Frank:* Quando è?

*Maria:* Il 19 dicembre.

*Frank:* Beh... il 19...

*Maria:* Vengono le mie due amiche di scuola, e la Robin...

*Frank:* Con piacere!

- *Dissimulare*
- *Falsificare*
- *Celare la causa dell'emozione*
- *Dire la verità per sviare i sospetti*
- *Dissimulazione a metà*
- *Suggerire una conclusione sbagliata*



## Motivi per mentire (Frank '92 )

- Autoconservazione
- Autopresentazione
- Vantaggio, guadagno
- Bugie altruistiche (sociali)





## Perchè le Bugie Fanno Fiasco (Ekman '01)

- Errori strategici
  - ◆ Il mentitore mostra sentimento inconsistente
  - ◆ Il mentitore dice qualcosa di inconsistente con la bugia
- 'Dimenticare la parte'
  - ◆ Falsificazione vuol dire mantenere l'intreccio
  - ◆ Dissimulare vuol dire ricordare che cosa è stato omesso.
  - ◆ Tutto questo crea il carico cognitivo
    - più difficile nascondere l'emozione.



## Perchè le Bugie Fanno Fiasco (Ekman '01)

- La paura di essere scoperti
  - ◆ Il destinatario è sospettoso
  - ◆ La posta in gioco è alta
  - ◆ Sono in gioco grandi punizioni e/o grandi ricompense
- Il senso di colpa
  - ◆ L'inganno non è **autorizzato**
  - ◆ Chi mente non mente spesso
- Il piacere della beffa
  - ◆ La vittima pone una sfida
  - ◆ L'inganno stesso è una sfida
  - ◆ Altri osservano ed apprezzano l'abilità del mentitore



## Tratti dell'Inganno

- Cognitivi
  - ◆ Coerenza, scorrevolezza
- Pragmatici
  - ◆ Discorso, atti locutori, l'avvicendamento dei turni, ecc.
- Emozione




## Emozione e Inganno

- L'errore di Otello
  - ◆ La gente ha paura della polizia
  - ◆ Le persone s'arrabbiano quando sono ingiustamente imputate.
  - ◆ (Certe) persone trovano comica la pizza
- L'effetto-Brokov
  - ◆ Non tener conto delle differenze individuali






## La pragmatica nell'inganno

- Le massime di Grice non sono operative: soprattutto il principio di cooperazione!
  - Il successo della mezzogna dipende dalla conoscenza del mondo (riguardo al contesto della menzogna) del destinatario.
  - Chi mente bene ha la capacità (forse inconsapevole) di falsificare gli atti linguistici.
- 



## La maggior parte della ricerca sull'inganno...

- non cerca di verificare i metodi del 20° secolo.
    - ◆ 'Macchina della verità'
    - ◆ Voice Stress Analysis
    - ◆ Reid
  - Lavori di psicologia
  - Esamina soprattutto gli indizi facciali e corporali
  - Poco lavoro sul rilevamento automatico
- 



## Elementi d'un modello dell'inganno nel linguaggio parlato

- Lessicale
- Prosodico/Acustico
- Pragmatico/Discorso



## L'Inganno nel linguaggio parlato (Depaulo '03)

- Termini di correlazione positiva
  - ◆ Parole interrotte/ripetute
  - ◆ Riferimento ad avvenimenti fuori tema
  - ◆ Incertezza verbale/vocale
  - ◆ Tensione vocale
  - ◆ F0



## L'Inganno nel linguaggio parlato (Depaulo '03)

- Termini di correlazione negativa
  - ◆ Il soggetto mantiene il tema di conversazione
  - ◆ Il soggetto ammette incertezza
  - ◆ Immediatezza verbale/vocale
  - ◆ Correzioni spontanee




## Columbia/SRI/Colorado Corpus

- Un lavoro con Julia Hirschberg, Stefan Benus, e colleghi di SRI/ICSI e U. C. Boulder
- Scopi:
  - ◆ Esaminare la fattibilità del rilevamento automatico dell'inganno nel linguaggio parlato
  - ◆ Scoprire o verificare i termini di correlazione acustici/prosodici, lessicali e pragmatici dell'inganno




## Il nostro paradigma

- Come elicitare la menzogna dai soggetti?
    - ◆ In quantità sufficiente
    - ◆ In modo che sappiamo il ‘ground truth’, cioè la veridicità di ogni enunciato
  - Vogliamo che la ‘posta in gioco’ abbia una qualità emozionale.
  - Vogliamo uno scenario ‘non-colpa’
  - Vogliamo un corpus ‘pulito’
- 

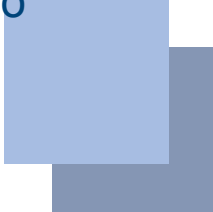


## Columbia/SRI/Colorado Corpus

- Scenario di “Inflated-performance”
  - Motivi: guadagno e autopresentazione
  - 32 soggetti: 16 donne, 16 uomini
  - Madrilingua di inglese americano
  - I soggetti credevano che si cercassero persone con le stesse qualità dei “25 Imprenditori migliori americani” per un esperimento di comunicazione
- 



## Columbia/SRI/Colorado Corpus

- I soggetti danno un esame per determinare se hanno il profilo degli imprenditori immaginari.
  - Diciamo ai soggetti che, secondo l'esame, non hanno il profilo giusto.
  - Seconda opportunità: possono cercare di far credere ad un intervistatore che abbiano il profilo giusto. **Devono mentire.**
- 

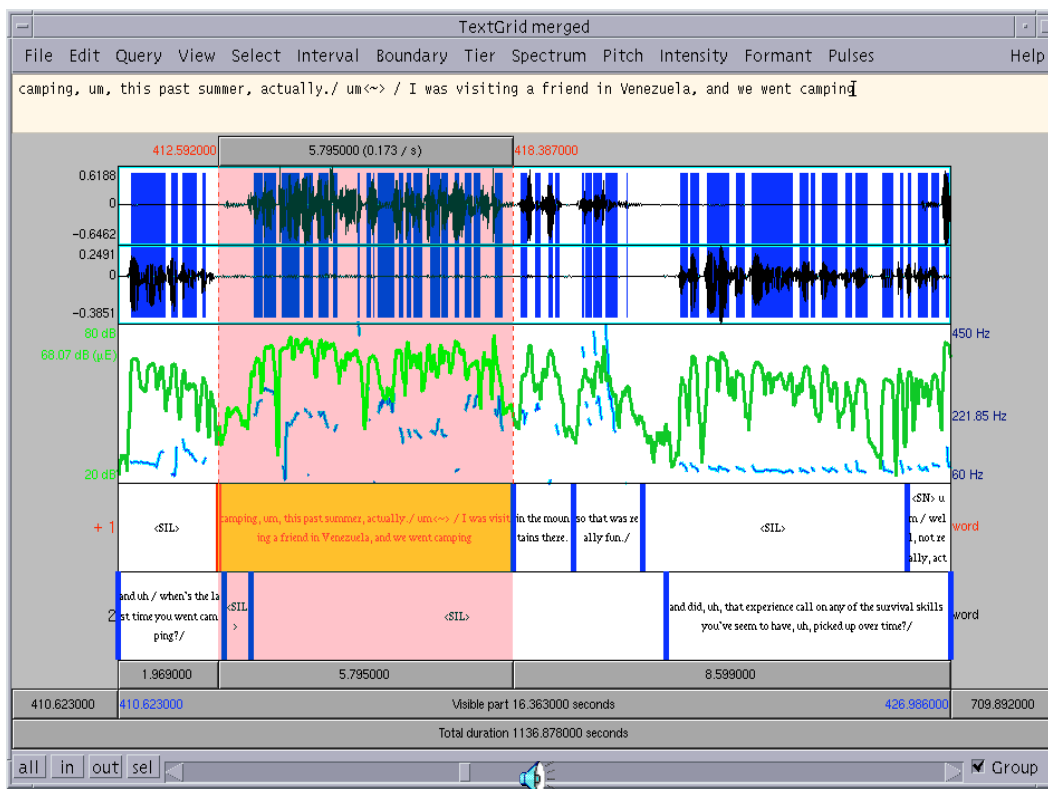


## Columbia/SRI/Colorado Corpus

- **Lo scopo è di far mentire i soggetti in una intervista.**
  - Prima dell'intervista, i soggetti danno un esame su 6 campi:
    - ◆ Interactive, music, survival, food, NYC geography, civics
  - La difficoltà delle domande è manipolata →
    - ◆ Ogni soggetto ha 6 punteggi: 2 meglio degli "imprenditori"; 2 peggio degli "imprenditori"; 2 uguali
  - Diciamo ai soggetti che cerchiamo anche persone che possano far credere ad un intervistatore che abbiano il profilo giusto.
    - ◆ Autopresentazione + guadagno
  - I soggetti sostengono un'intervista registrata
    - ◆ Indicano la veridicità di ogni enunciato usando pedali nascosti sotto la tavola.
- 

# CSC Corpus: Dati

- 15.2 ore d'interviste;  
7 ore di linguaggio parlato dei soggetti
- Trascritte lessicalmente & allineate automaticamente  
→ tratti lessicali/discorsi
- Due livelli di veridicità : globale / locale
- Segmentations (LT/LL):  
'slash units' — delimitati di punteggiatura (5709/3782),  
frasi (11,612/7108), turni (2230/1573)
- Tratti acustici (± output del riconoscimento vocale)



**Columbia University– SRI/ICSI – University of Colorado  
Deception Corpus: An Example Segment**



## CSC Corpus: Risultati

- Classificazione automatica  
(Induzione di regole Ripper, randomizzato 5-fold)
  - ◆ Slash Units / bugie locali — base di 60.2%
    - lessicali & acustici: 62.8 %; + 'subject dependent': 66.4%
  - ◆ Phrases / bugie locali — base di 59.9%
    - lessicali & acustici : 61.1%; + 'subject dependent': 67.1%
- Other findings
  - ◆ Parole d'emozione positiva → inganno (LIWC)
  - ◆ 'Pleasantness' → inganno (DAL)
  - ◆ Pause riempite → verità
  - ◆ Correlazione con frequenza fondamentale — **varia con soggetto**

## Regole esemplari JRIP :

```
(cueLieToCueTruths >= 2) and (TOPIC = topic_newyork) and  
(numSUwithFPtoNumSU <= 0) and  
(wu_ENERGY_NO_UV_STY_MAX__EG_ZNORM-  
ENERGY_NO_UV_STY_MIN__EG_ZNORM-D <= 5.846)  
=> PEDAL=L (231.0/61.0)
```

```
(cueLieToCueTruths >= 2) and (numSUwithFPtoNumSU <= 1) and  
(wu_ENERGY_NO_UV_STY_MAX__EG_ZNORM-  
ENERGY_NO_UV_STY_MIN__EG_ZNORM-D <= 5.68314) and  
(wu_ENERGY_NO_UV_RAW_MAX-ENERGY_NO_UV_RAW_MIN-D >= 8.41605)  
and (wu_F0_SLOPES_NOHD__LAST >= -2.004)  
=> PEDAL=L (284.0/117.0)
```

```
(cueLieToCueTruths >= 2) and (wu_F0_RAW_MAX >= 5.706379)  
and (wu_DUR_PHONE_SPNN_AV <= 1.0661)  
=> PEDAL=L (262.0/115.0)
```



## CSC Corpus: A Perception Study


- Con Julia Hirschberg, Stefan Benus, Robin Cautin e colleghi di SRI/ICSI
  - 32 soggetti che giudicano
  - Ogni soggetto etichetta due interviste.
  - Metodi:
    - ♦ Bugie locali usando Praat
    - ♦ Bugie globali usando penna e carta
  - Riempie questionari pre- e post-etichettare
  - Inventario di Personalità
  - Il soggetto riceve 'istruzione' riguardo a un parlante
- 



Table 1: *Judges' aggregate performance classifying TRUTH / LIE.*

Lie Category	Chance		Median	Std.		
	Baseline	Mean <sup>a</sup>		Dev.	Min.	Max.
Local	63.87 <sup>b</sup>	58.23	57.42	7.51	40.64	71.48
Global	63.64 <sup>c</sup>	47.76	50.00	14.82	16.67	75.00

Aggregati per 'guidice'



<sup>a</sup>Each judge's score is his or her average over two interviews; as percentages.

<sup>b</sup>Guessing **TRUTH** each time.

<sup>c</sup>Guessing **LIE** each time.

Aggregati per parlante



Table 1: *Aggregate performance by interviewee.*

Lie Type	Chance		Std.		
	Mean <sup>a</sup>	Median	Dev.	Min.	Max.
Local	58.23	58.58	9.44	35.86	87.79
Global	44.83	45.58	17.40	10.00	81.67

<sup>a</sup>Each interviewee's score is the average over two judges; as percentages.

👉 Accuratezza del sistema automatico: ~67%

## Inventario di Personalità: NEO-FFI

- Costa & McCrae (1992) Five-factor model
  - ◆ Openness to Experience
  - ◆ Conscientiousness
  - ◆ Extraversion
  - ◆ Agreeability
  - ◆ Neuroticism
- Ampiamente usato nella psicologia

## Neuroticism, Openness & Agreeableness sono correlati della precisione dei giudici

Con rispetto alle bugie globali.

Table 1: Correlations between personality factors and judge performance at labeling global lies.


Factor	Measure	Pearson's corr. coef.	p-value
Neuroticism	Proportion of segments judged LIE	-0.44	0.012
Openness	Accuracy	0.51	0.003
Agreeableness		0.41	0.021
Neuroticism	F-measure for TRUTH	0.37	0.035
Agreeableness		0.41	0.019
Openness	F-measure for LIE	0.52	0.003

## Modelli di regressione (semplificati) per accuratezza e Misura F

<b>Accuratezza</b>	<b>R<sup>2</sup> = 0.48</b>	<b>p &lt; 0.0001</b>
agreeableness	0.0056	p = 0.0016
openness	0.0079	p = 0.0002
<b>Misura F (lie)</b>	<b>R<sup>2</sup> = 0.27</b>	<b>p = 0.0026</b>
openness	0.0101	p = .0026




## Questionari pre- e postcompito

- Hai avuto un lavoro per cui la capacità di individuare le menzogne era importante?
  - Quante volte al giorno dice una menzogna una persona normale?
  - Con quale frequenza puoi individuare una menzogna?
  - Come hai fatto al compito?
- 




## Altre scoperte

- Non c'è stato effetto dovuto ad istruzione.
  - Non abbiamo trovato nessuna relazione tra 'confidence' precompito e postcompito dei giudici.
  - I soggetti che avevano avuto esperienza avevano una sicurezza di giudizio pre-compito significativamente più alta.
    - ◆ Ma non la accuratezza più alta!
  - Molti soggetti hanno usato le pause riempite come indizi dell'inganno.
    - ◆ Ma in questo corpus, le pause riempite sono correlati della VERITA! (Benus et al. '06)
- 



## Il nostro lavoro futuro.

- Differenze individuali
    - ◆ 'I Magi dell'Inganno'
    - ◆ Direzione dei termini di correlazione
  - Classificare bugie globali
    - ◆ Bugie locali come 'punti caldi'.
  - Nuovo paradigma
    - ◆ Più corto
    - ◆ Inventario di personalità per i parlanti
    - ◆ Carico cognitivo
- 



### OTTAVIO:

*...le bugie rendono l'uomo ridicolo,  
infedele, odiato da tutti; e che per non  
esser bugiardi, conviene parlar poco,  
apprezzare il vero, e pensare al fine.*

*Il Bugiardo di Carlo Goldoni*





# Rivelamento dell'inganno nel 20° secolo

- 'Macchina della verità'
    - ◆ <http://antipolygraph.org>
    - ◆ The Polygraph and Lie Detection (N.A.P. 2003)
  - Voice Stress Analysis
    - ◆ Microtremori: 8-12Hz
    - ◆ Universal Lie response
    - ◆ <http://www.love-detector.com/>
    - ◆ <http://news-info.wustl.edu/news/page/normal/669.html>
  - Reid
    - ◆ Behavioral Analysis Interview
    - ◆ Interrogazione
- 

Table 1: *Regression models of performance on global lies.*

Proportion of Segments Judged LIE				
	Value	Std. Err.	t-value	p-value
(Int.)	0.7092	0.1065	6.6606	0.0000
Neurot.	-0.0056	0.0021	-2.6749	0.0120
Multiple $R^2$ : 0.19		p-value: 0.0120		
F-statistic: 7.16, 1 and 30 deg. of freedom				
Classification Accuracy				
	Value	Std. Err.	t-value	p-value
(Int.)	-0.2508	0.1427	-1.7572	0.0894
Agree.	0.0056	0.0016	3.4713	0.0016
Open.	0.0079	0.0019	4.1929	0.0002
Multiple $R^2$ : 0.48		p-value: < 0.0001		
F-statistic: 13.39, 2 and 29 deg. of freedom				
F-measure for TRUTH				
	Value	Std. Err.	t-value	p-value
(Int.)	-0.0029	0.1224	-0.0237	0.9813
Neurot.	0.0044	0.0018	2.4251	0.0218
Agree.	0.0047	0.0018	2.6686	0.0123
Multiple $R^2$ : 0.31		p-value: < 0.0046		
F-statistic: 6.50, 2 and 29 deg. of freedom				
F-measure for Lie				
	Value	Std. Err.	t-value	p-value
(Int.)	-0.1469	0.1896	-0.7747	0.4446
Open.	0.0101	0.0031	3.2906	0.0026
Multiple $R^2$ : 0.27		p-value: < 0.0026		
F-statistic: 10.83, 1 and 30 deg. of freedom				