

**NETS**

**IUSS** Laboratory

for **N**euro**L**inguistics,  
C**o**mputational  
L**inguistics** &  
T**h**eoretical **S**yntax



**IUSS**

Scuola Universitaria Superiore Pavia

CRISTIANO CHESI  
NETS @ IUSS PAVIA

# *What do “we need”, in the end?*

Intelligenza Linguistica, Sovranità AI  
e Futuro Responsabile

Villanova.ai - Cagliari, 23 Settembre 2025

# Il problema di Platone



Che cosa ci serve, alla fine?

Com'è possibile che gli esseri umani, i cui contatti con il mondo sono **brevi, personali e limitati**, siano comunque in grado di **sapere così tanto**?



○ Un **esempio linguistico**:

○ **Quale macchina** pensi che il meccanico abbia riparato \_ ?

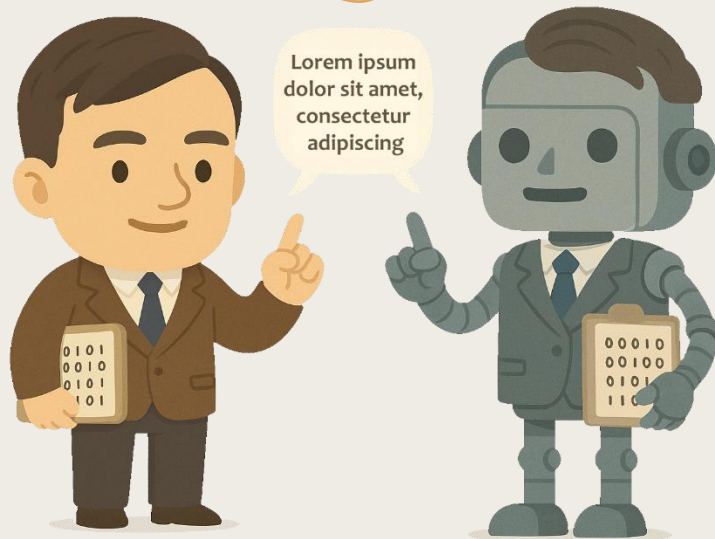
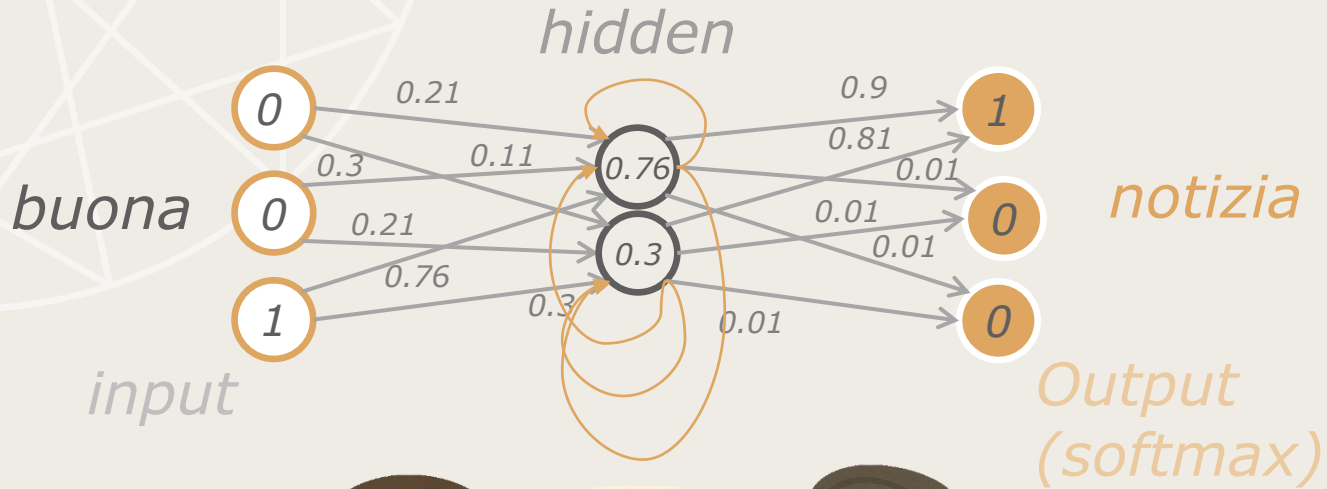
○ \***Quale macchina** pensi che il meccanico abbia riparato **il motore** ?

○ Una **conclusione logica**:

Veniamo al mondo con una **predisposizione innata** ad assegnare una **precisa analisi strutturale** all'**input** (linguistico) che riceviamo (e non **infinite altre possibili analisi**)

- ... della **parola successiva** (Elman 1990)

Questa è una buona ... **notizia**



C. CHESI

Che cosa ci serve, alla fine?

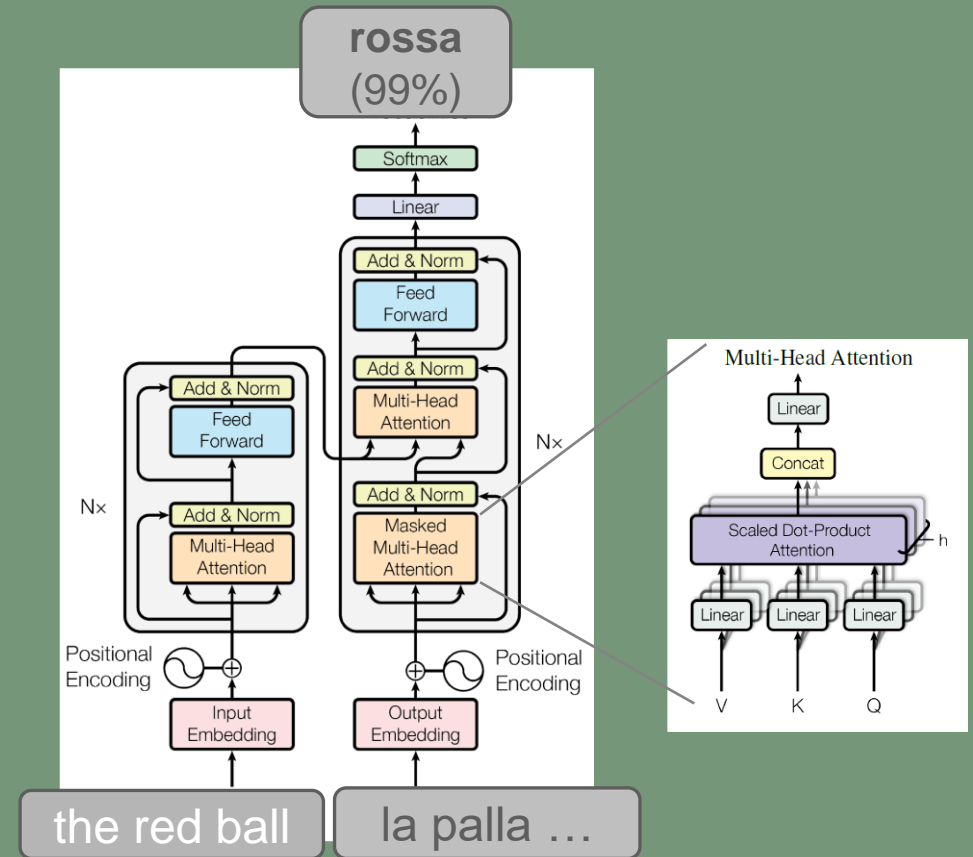
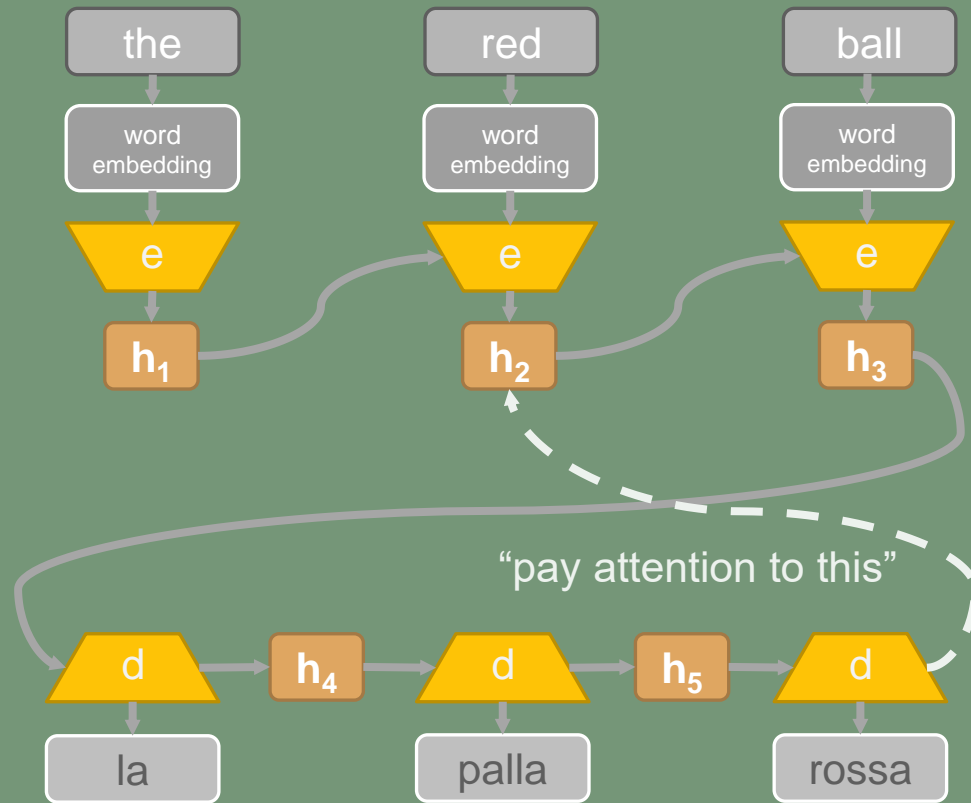
35 anni di previsioni...  
(word embedding e reti ricorrenti)



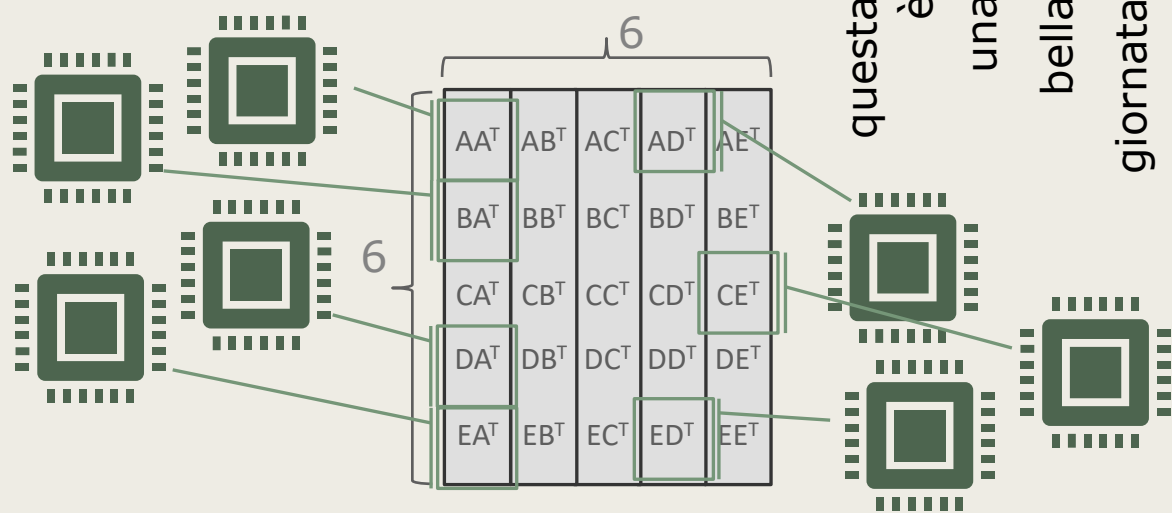
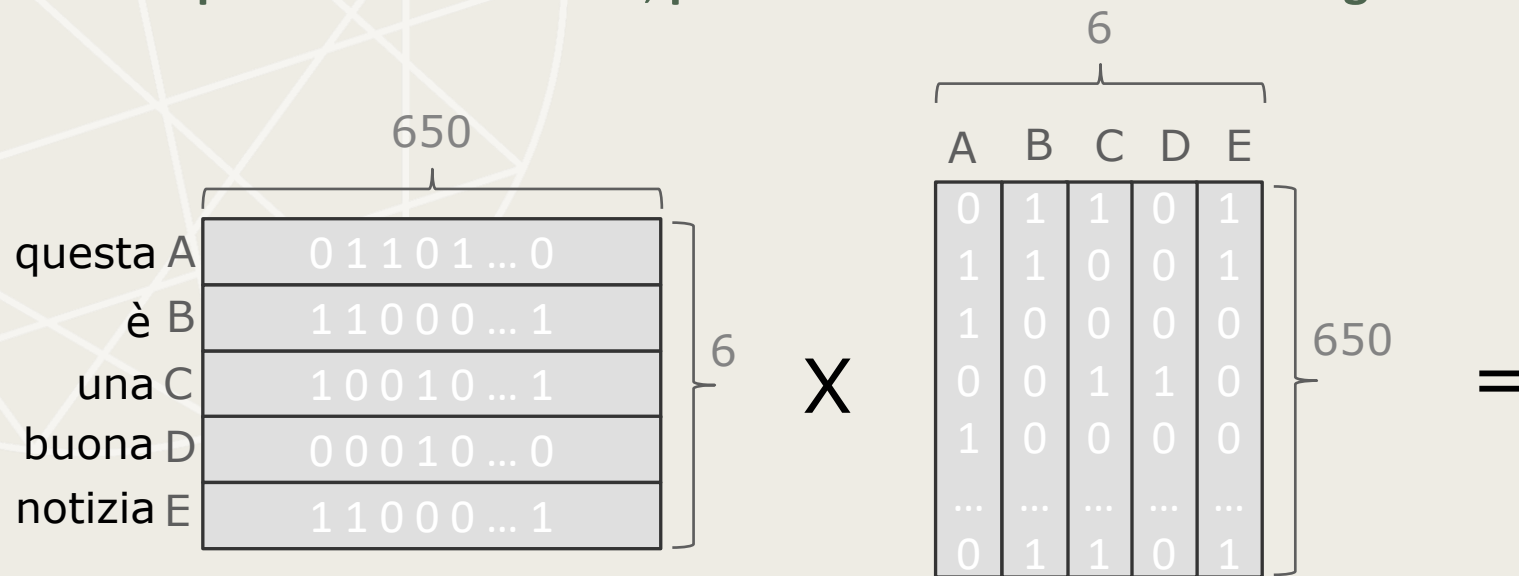
# L'era dei Transformers:

“Attention is all you need” Vaswani et al. (2017)

L'esempio della traduzione automatica

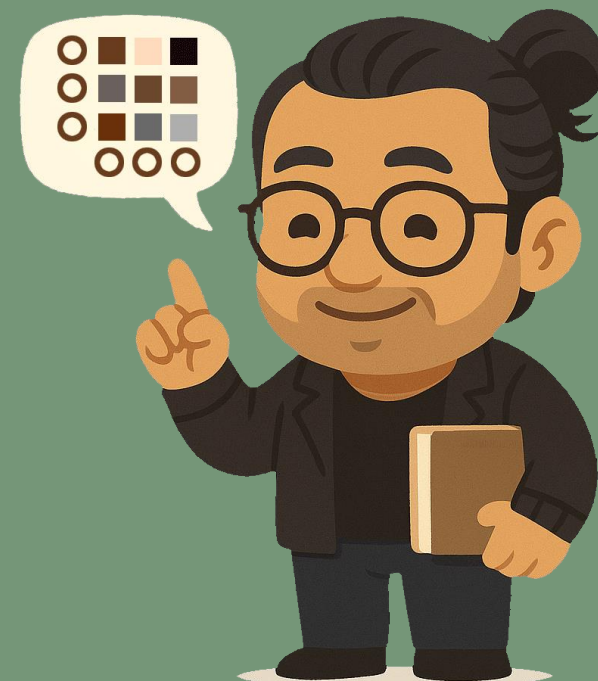


⦿ **Moltiplicazioni tra matrici, parallelismo e word embedding:**



questa è una bella giornata

Attention is ALL you need



# BabyLM Challenge

Sample-efficient pretraining on a developmentally plausible corpus



Che cosa ci serve, alla fine?

- ◎ **3 competizioni**
  - ◎ **Small-strict:** corpus da 10M di parole
  - ◎ **Small:** corpus da 100M di parole
  - ◎ **Multimodale:** corpus da 50M di parole + 50 milioni di parole accoppiate ad immagini
- ◎ La **valutazione** (piattaforma di valutazione condivisa basata su task accessibili a tutti)
  - ◎ **BLiMP** (Benchmark for Linguistic Minimal Pairs)
    - ◎ E.g. questa frase è corretta vs. \*questa frase è corretto
  - ◎ **(Super) GLUE**
  - ◎ ...

C. CHESI

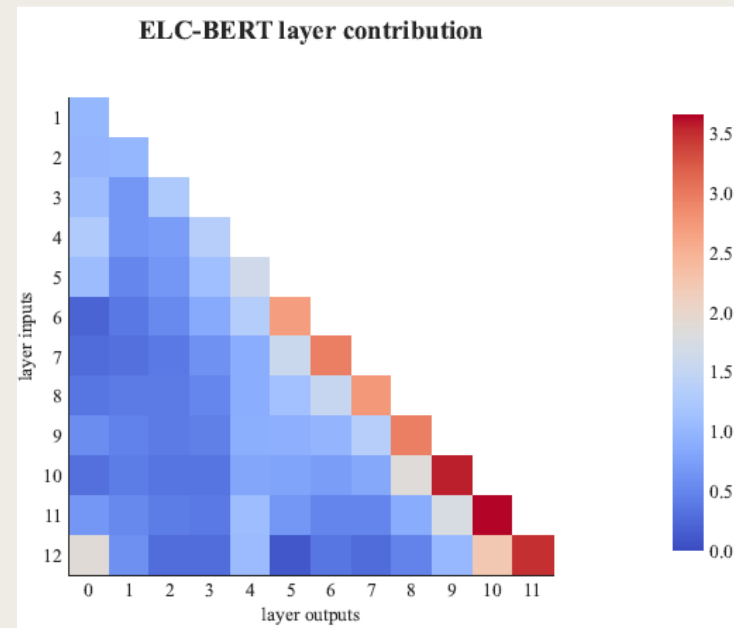
# BabyLM Challenge

Sample-efficient pretraining on a developmentally plausible corpus



Che cosa ci serve, alla fine?

- Il modello vincitore della competizione dello scorso anno: **Charpentier and Samuel (2024)**  
Not all layers are equally as important: Every Layer Counts BERT



- Quanto tempo ci vuole a trovare i coefficient migliori?

**Molto!**

In effetti quest'anno la competizione richiedeva di utilizzare non più di **100M di passi** (training sweep) per non più di **10 epoche** di apprendimento.

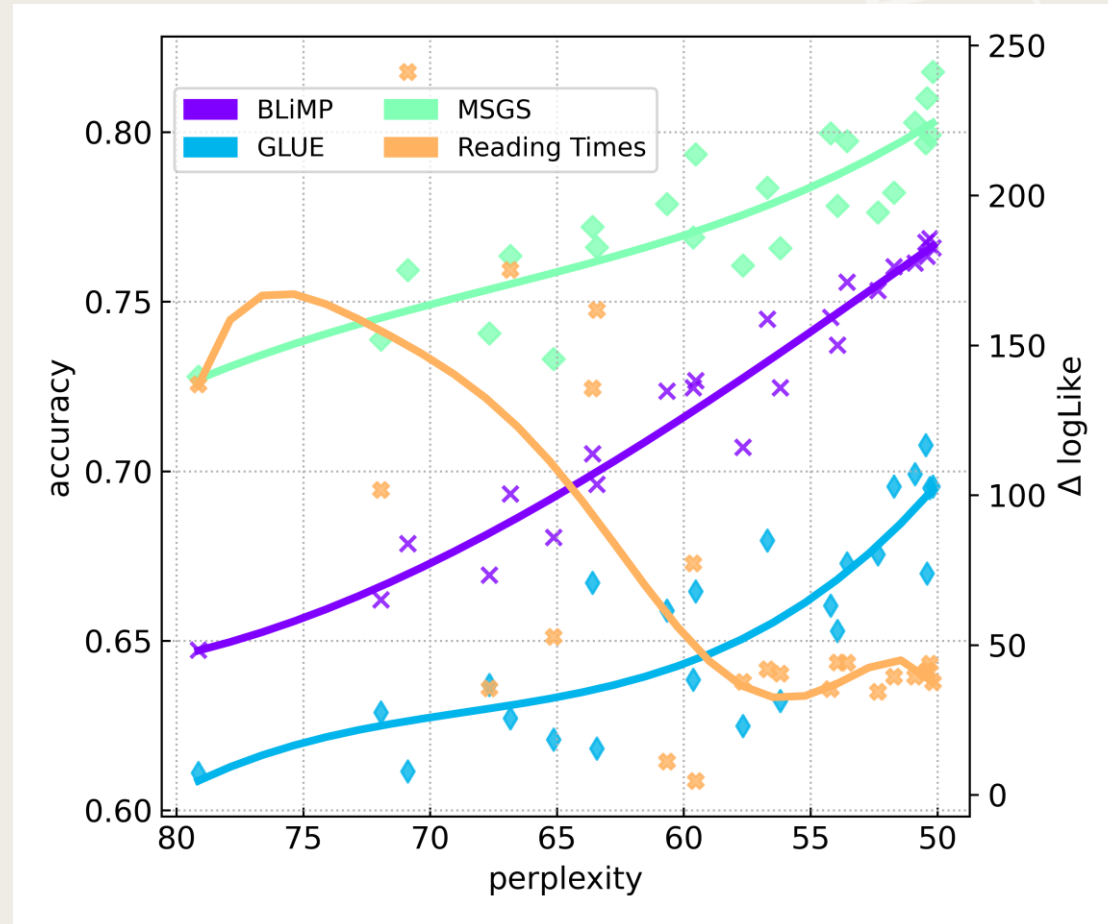
# Steuer et al. (2023)

Large GPT-like Models are Bad Babies:  
A Closer Look at the Relationship  
between Linguistic Competence and  
Psycholinguistic Measures



Che cosa ci serve, alla fine?

- Il **best-paper award** è andato al lavoro:  
I modelli **GPT** sono dei «cattivi bambini»



C. CHESI

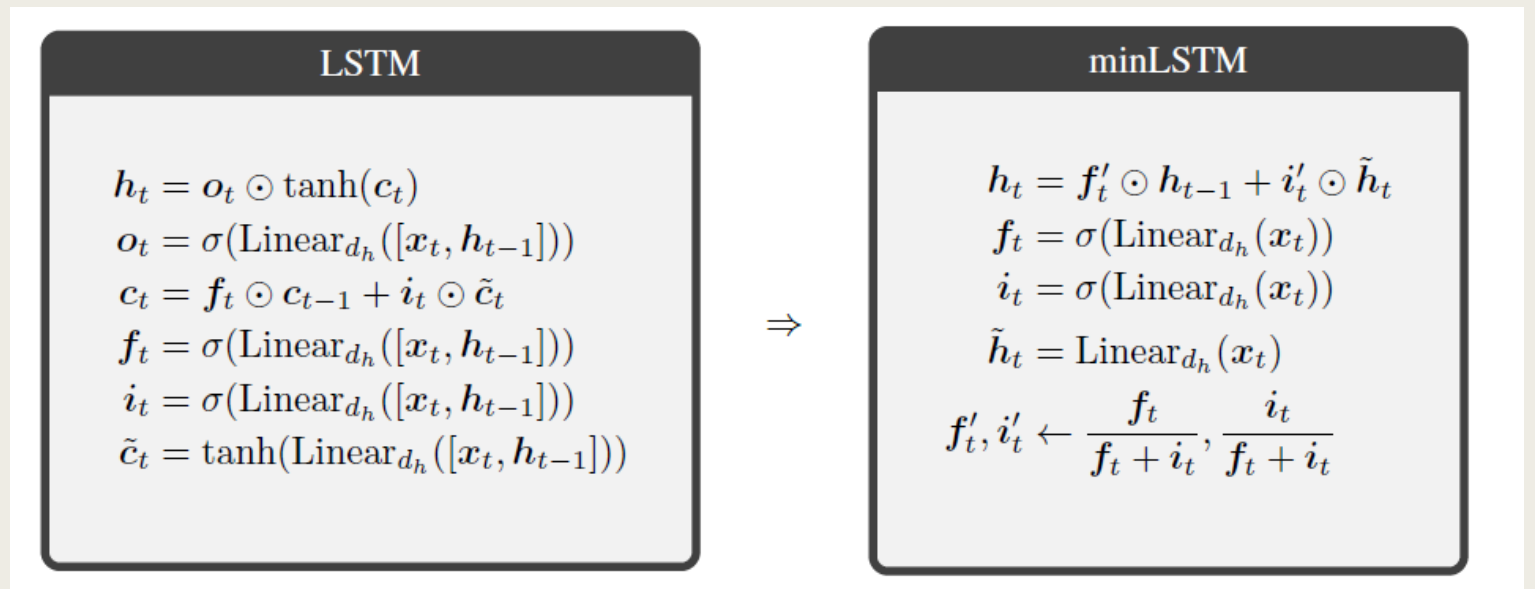
# Were RNNs All We Needed?



Che cosa ci serve, alla fine?

Feng, Tung, Ahmed, **Bengio (2024)**

- Il meccanismo di **gating** delle RNN può essere semplificato in modo da poter ricorrere ad una scansione parallela (**parallel prefix scan algorithm**, Bletloch, 1990).
- Bisogna garantire che gli output siano indipendenti su scala temporale (rimuovere i collegamenti ai precedenti hidden layer per cui non sia disponibile il prefix scan: e.g., le funzioni *tanh*)



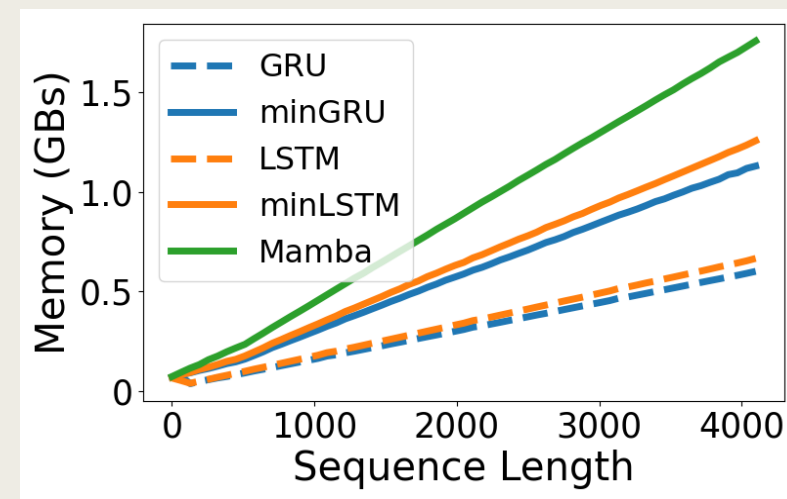
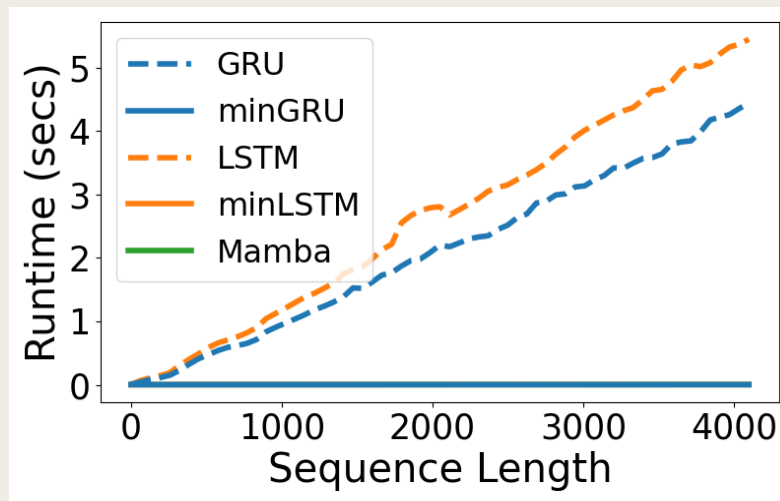
C. CHESI

# Were RNNs All We Needed?



Che cosa ci serve, alla fine?

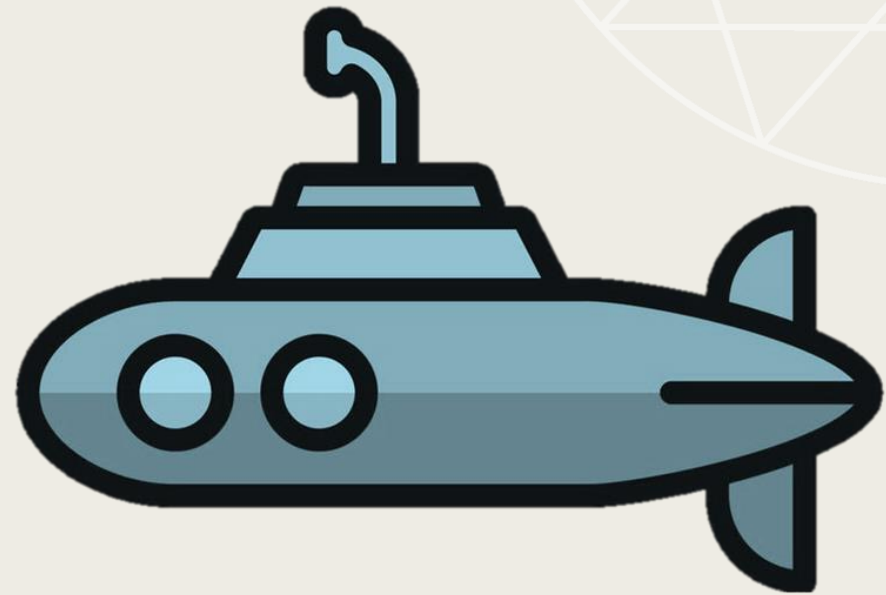
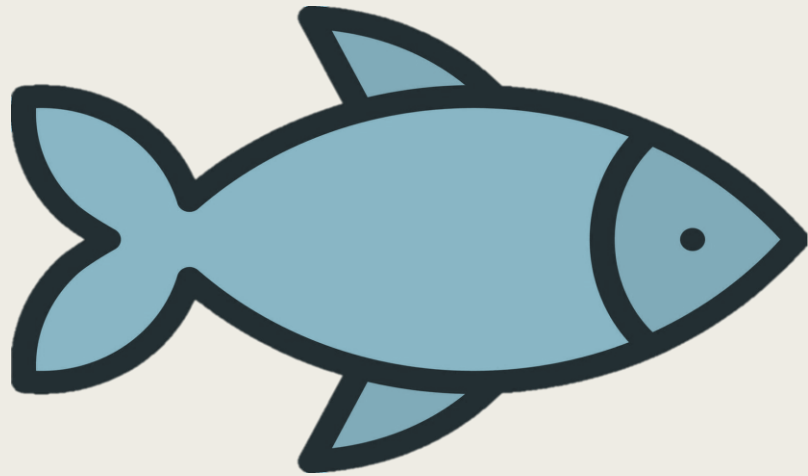
Feng, Tung, Ahmed, Bengio (2024)



C. CHESI

# Il reale e l'artificiale

---



THE ALGEBRAIC THEORY OF CONTEXT-FREE LANGUAGES\*

N. CHOMSKY

*Massachusetts Institute of Technology*

AND

M. P. SCHÜTZENBERGER

*Harvard University*

# La Gerarchia di Chomsky



Che cosa ci serve, alla fine?

## Recursive enumerable (Turing Machines)

Context-sensitive

Context-free

Regular



$ab^*$



$a^n b^n$



$a^n b^n c^n$

# Cosa sanno i bambini?



Che cosa ci serve, alla fine?

Due **vincoli fondamentali** (Rizzi 1990):

- ⊙ **C-command** - una volta “in scatolato” un elemento non può entrare in relazioni strutturali con elementi fuori dalla scatola
  - ⊙ [Gli amici [del nonno]] **salutano**\\*saluta tutti
- ⊙ **Località** - le relazioni sintattiche sono **rispettose** e **pigre**: non scavalcano elementi che potrebbero entrare a far parte della relazione e, se elementi strutturalmente più vicini (locali) sono presenti, devono soddisfare per primi la relazione
  - ⊙ **Cosa<sub>i</sub>** credi che abbia mangiato **\_i** ?
  - ⊙ **\*Cosa<sub>i</sub>** credi **chi** abbia mangiato **\_i** ?
  - ⊙ **Chi<sub>i</sub>** credi abbia **\_i** mangiato **cosa** ?

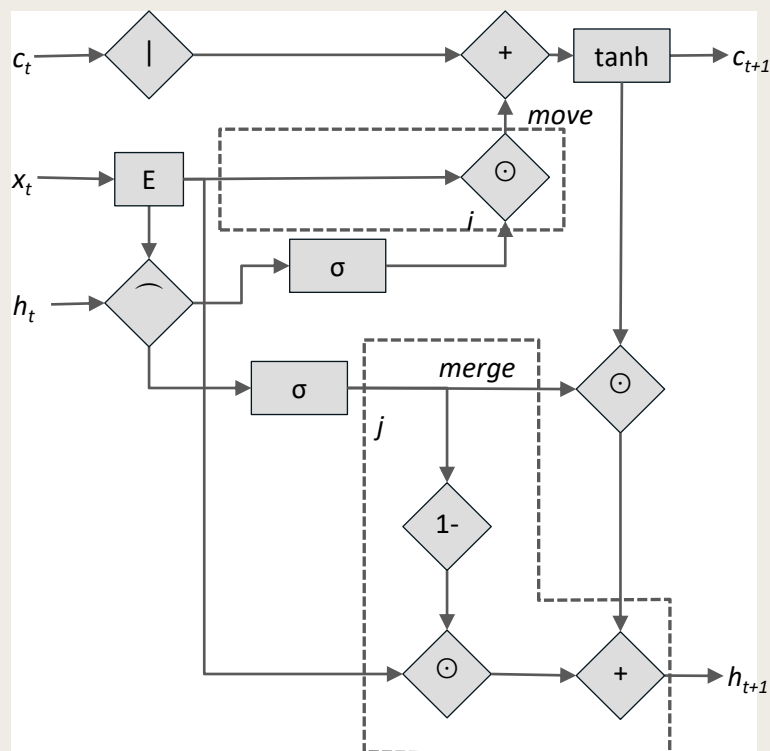
# L'architettura della nostra rete

(expectation-based Minimalist Grammar Recursive Neural Network - eMG-RNN)

Una rete  
"speciale"

Fusco et al. 2024 @ Clic-IT

## Architettura



## BLiMP performance

	LSTM	eMG-RNN				
		1	2	3	F-M	F-N
<i>Ana. agr</i>	0.67	0.82	0.76	0.77	0.88	0.81
<i>Arg. str</i>	0.56	0.65	0.64	0.63	0.64	0.66
<i>Binding</i>	0.54	0.69	0.66	0.63	0.57	0.65
<i>Ctrl. / Rais.</i>	0.59	0.58	0.59	0.60	0.58	0.60
<i>D-N agr</i>	0.57	0.67	0.63	0.67	0.68	0.68
<i>Ellipsis</i>	0.41	0.24	0.30	0.21	0.42	0.39
<i>Filler. gap</i>	0.55	0.64	0.60	0.47	0.48	0.65
<i>Irregular</i>	0.54	0.58	0.69	0.60	0.60	0.58
<i>Island</i>	0.54	0.58	0.54	0.53	0.50	0.62
<i>Npi</i>	0.45	0.33	0.50	0.55	0.32	0.31
<i>Quantifiers</i>	0.57	0.55	0.53	0.53	0.53	0.57
<i>S-V agr</i>	0.50	0.52	0.52	0.52	0.55	0.53
<b>Overall</b>	0.54	0.58	0.58	0.57	0.55	0.59

# La “tokenizzazione”

eMG-RNN

- Byte Pair Encoding (**BPE**)

- low (5 times), newer (4 times), higher (9 times)*

- e, g, h, i, l, n, o, r, w, \_*

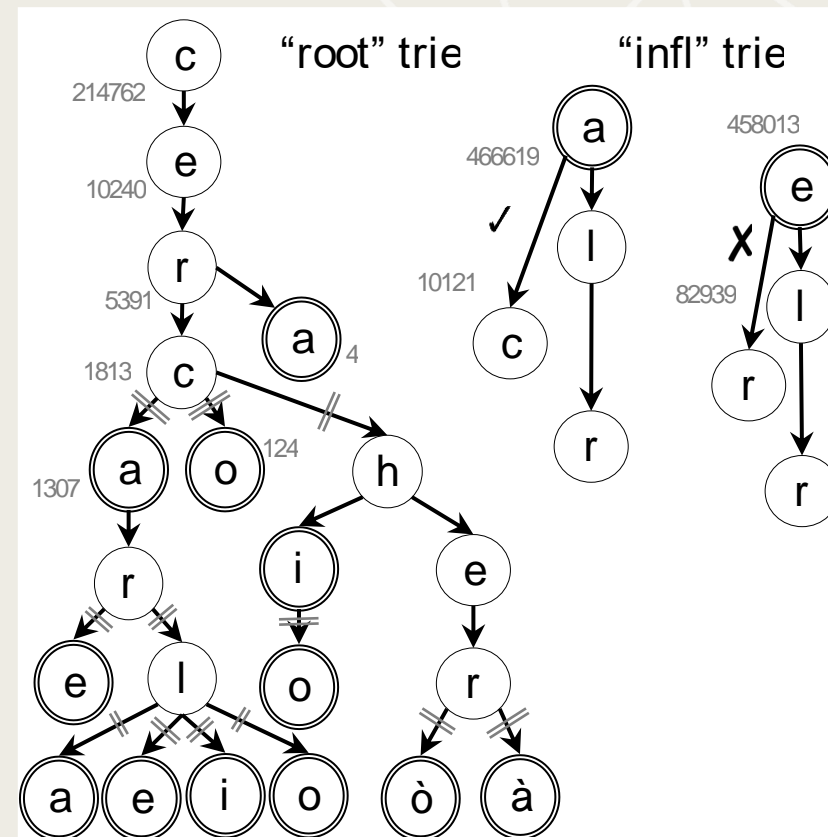
- er (13)      r\_ (13)*

- e, g, h, l, l, n, o, r, w, \_, er*

*ew (4)*

...

- MorPiece (**MoP**)  
basato sul Tolerance Principle (Yang 2016):  
una regola è produttiva  
se applicata  
 $n/\ln(n)$  volte



# NETS

IUSS Laboratory

for **N**euro**L**inguistics,  
**C**omputational  
**L**inguistics &  
**T**heoretical **S**yntax



# IUSS

Scuola Universitaria Superiore Pavia



## CRISTIANO CHESI

### NETS @ IUSS PAVIA

# Che cosa ci serve, alla fine?



S. Rossi



M. Barbini



V. Bressan



A. Zanollo



G. Formichi



S. Neri



A. Fusco



M. L. Piccini Bianchessi



T. Sgrizzi

# Come si valuta la Competenza Linguistica di un LLM?

Il test COnVERSA  
(Chesi et al. 2025)

## ○ A1. D-N

Il **giorno**.  
*the.SG.M day.SG.M*

\***I** **giorno**.  
*the.PL.M day.SG.M*

## ○ A2. Subject-Adjectival Predidate

Il **muro** è **rosso**.  
*the wall.SG.M is red.SG.M*

\*Il **muro** è **rossa**.  
*the wall.SG.M is red.SG.F*

## ○ A3. Subject-Verb

**La maestra** **corre**  
*the teacher runs*

\***La maestra** **corr~~o~~no** (Unergative)  
*the teacher run*

Arriva **la maestra**  
(there) arrives the teacher

\*Arriva **le maestre** (Unaccusative)  
(there) arrives the teachers

Il **maestro** **corregge** i compiti  
*The teacher corrects the homework.PL*

\*Il **maestro** **corregg~~o~~no** i compiti (Trans.)  
*The teacher correct the homework.PL*

## ○ A4. Attraction

Il **muro** della casa è **rosso**.  
*the wall.SG.M of the house.SG.F is red.SG.M*

\*Il **muro** della casa è **rossa**.  
*the wall.SG.M of the house.SG.F is red.SG.F*

**La maestra** degli alunni **corre**  
*The teacher of the students runs*

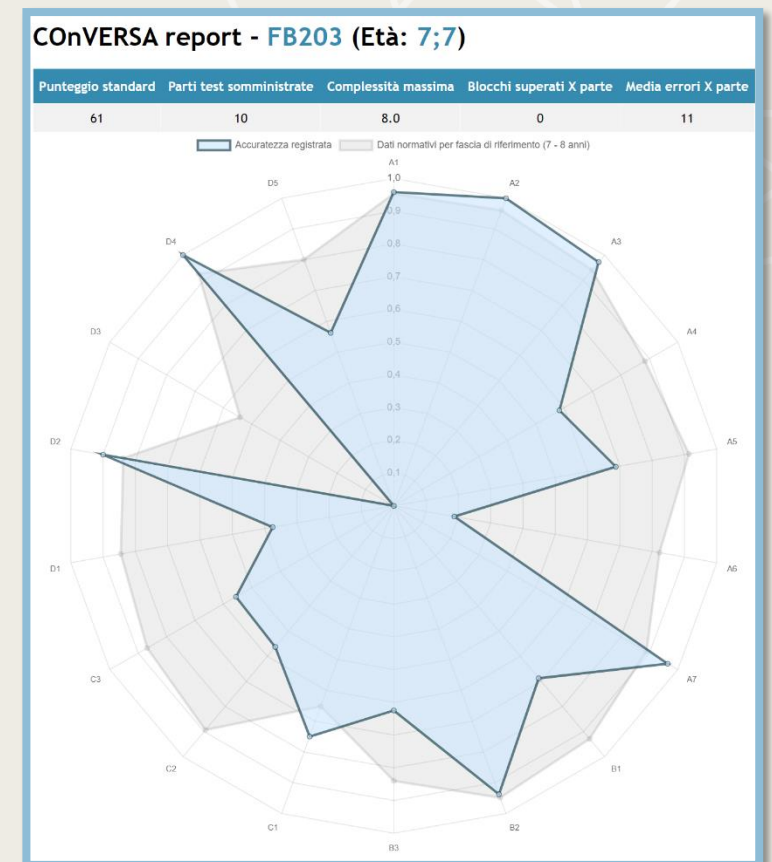
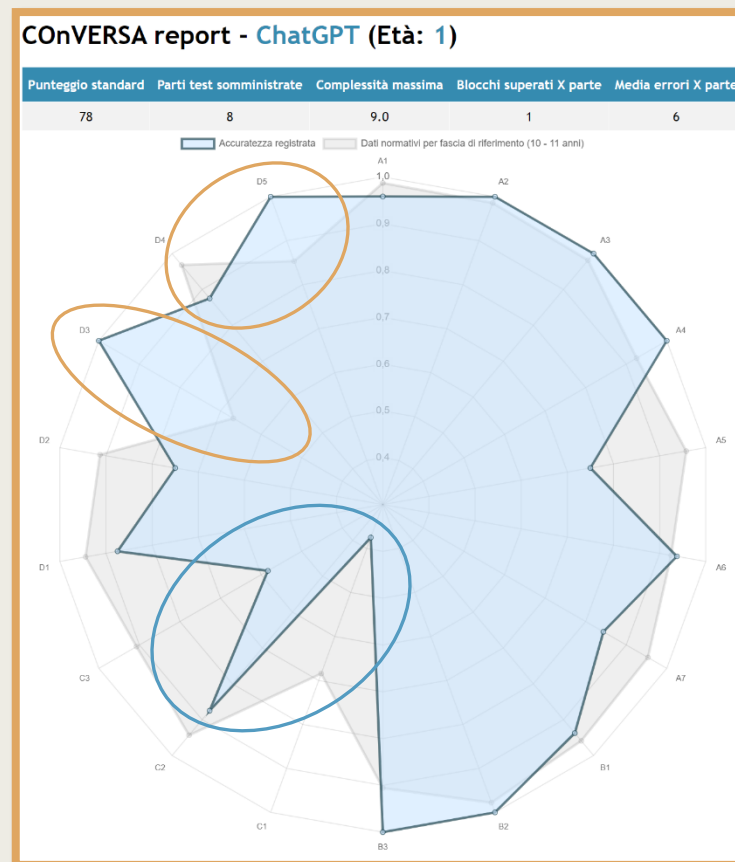
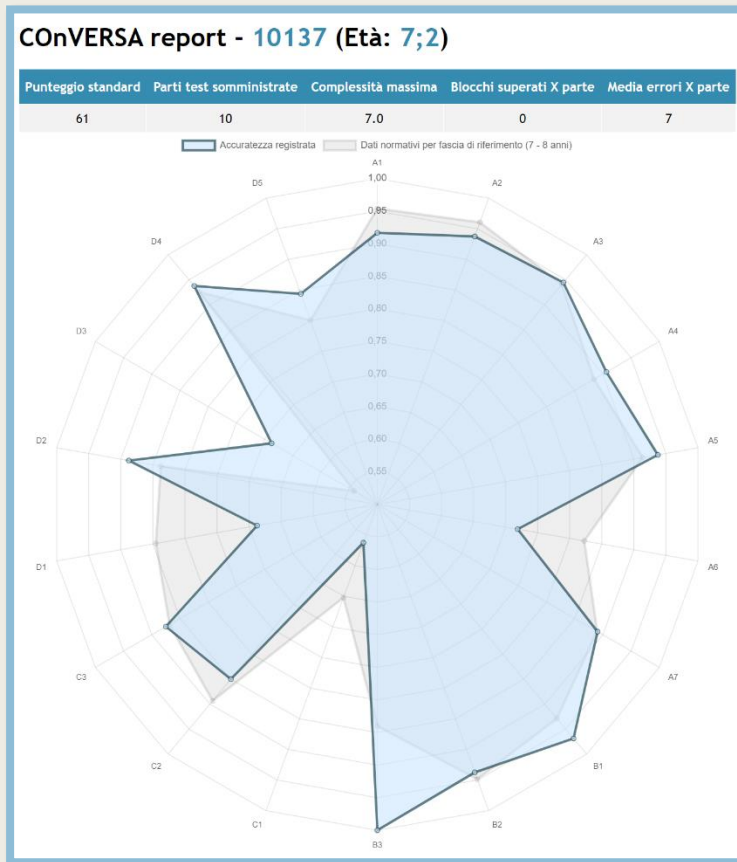
\***La maestra** **degli alunni** **corr~~o~~no**  
*The teacher of the students run*

## ○ A5. Past Participle

**La foglia** è **caduta**  
*the leaf.SG.F is fallen.SG.F*

\***La foglia** è **cadute**  
*the leaf.SG.F is fallen.PL.F*

# Confrontare le performance sulle coppie minime

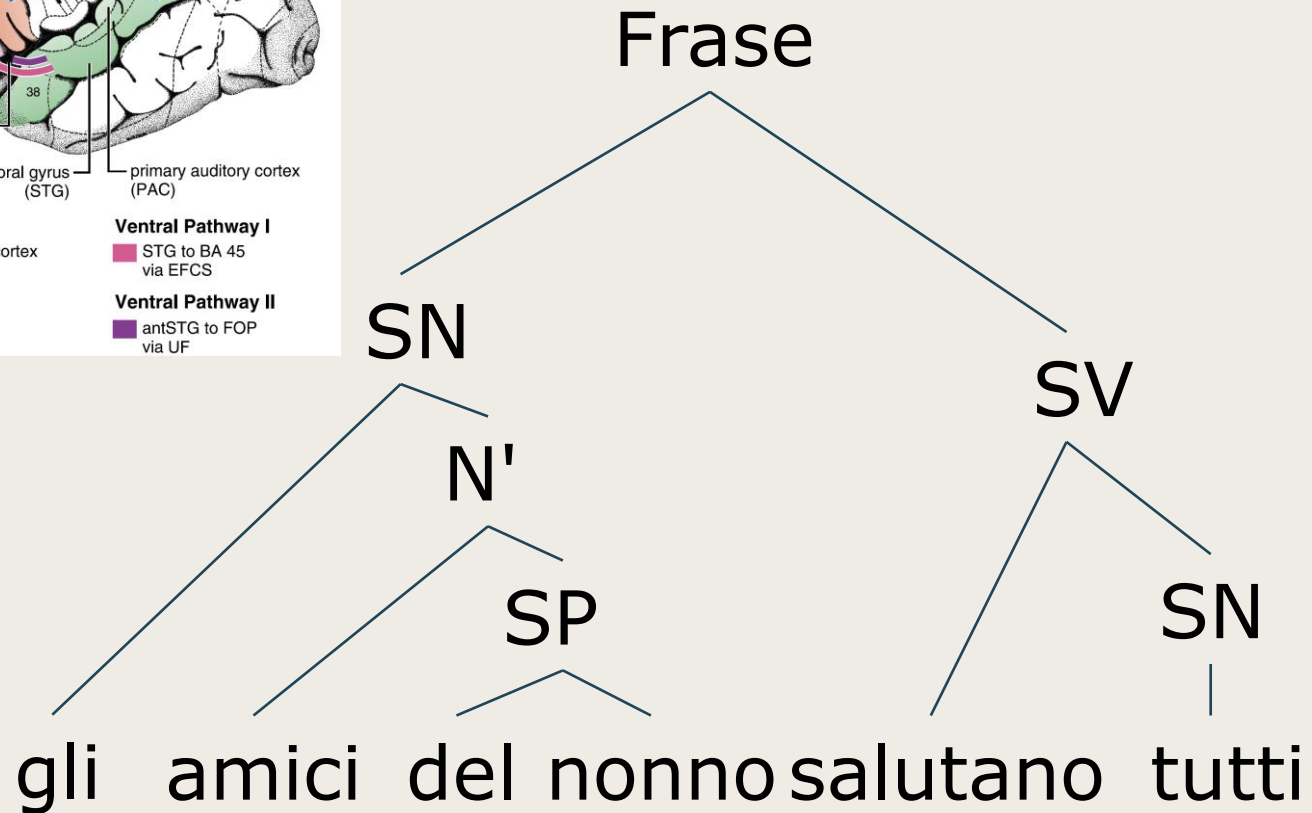
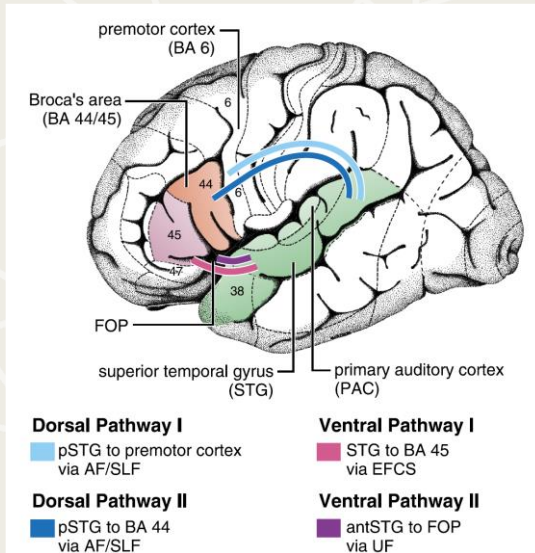


- ⦿ Com'è possibile che di tutta l'informazione che riceviamo in **input** riusciamo a processarne così **poca** e in modo così **distorto**?
- ⦿ Un **esempio linguistico**:
  - ⦿ La macchina rossa che il carrozziere ha portato in garage per riparare aveva il parafrangente ammaccato, una ruota bucata e un graffio sulla tappezzeria.
  - ⦿ Il meccanico ha riparato una macchina che aveva una ruota bucata, il parafrangente ammaccato e un graffio sul parabrezza.
- ⦿ Una **conclusione logica**:
  - ⦿ Il nostro processamento di un **input** (linguistico) **trascura numerosi dettagli** che vengono **ricostruiti** facendo affidamento alla nostra **competenza** (linguistica) e alle **conoscenze del mondo acquisite** (bias)

## Il problema di Orwell



◎ Friederici (2011)



Dove e come processiamo il linguaggio?

