

Non dire málvísindi se non  
l'hai nel sacco

# Contenuti:



1. Che cos'è la Linguistica (computazionale)?
2. Da dove viene, la linguistica computazionale?
3. Ma i computer sentono?
4. Ma i computer sanno leggere?
5. Ma i computer sanno la grammatica?
6. Vabbè, ma funziona?
7. E Siri?
8. Diventare un(a) linguista computazionale

# Che cos'è la **Linguistica** (Computazionale)?

Un(a) linguista:



- corregge i difetti di pronuncia
- parla generalmente molte lingue
- studia la grammatica
- sa per certo cosa è **giusto** e cosa è **sbagliato** in una lingua, mi correggerà quando parlo!
- cerca di scoprire come fanno i bambini ad imparare la loro lingua
- studia la lingua come oggetto biologico, psicologico e culturale
- esamina e descrive le lingue del mondo
- inventa nuove lingue
- spiega agli altri come parlare in modo corretto
- ricostruisce lingue che nessuno parla più

# Che cos'è la **Linguistica** (Computazionale)?

Un(a) linguista:



- ~~corregge i difetti di pronuncia~~
- ~~parla generalmente molte lingue~~
- studia la grammatica
- ~~sa per certo cosa è giusto e cosa è sbagliato in una lingua, mi correggerà quando parlo!~~
- cerca di scoprire come fanno i bambini ad imparare la loro lingua
- studia la lingua come oggetto biologico, psicologico e culturale
- esamina e descrive le lingue del mondo
- ~~inventa nuove lingue~~
- ~~spiega agli altri come parlare in modo corretto~~
- ricostruisce lingue che nessuno parla più

# Che cos'è la **Linguistica** (Computazionale)?



Nella slide precedente abbiamo visto come alcune comuni affermazioni sui linguisti e le linguiste siano in realtà solamente luoghi comuni.

Non è vero, ad esempio, che un(a) linguista conosca a menadito le convenzioni *canoniche* dell'italiano standard, cioè quello che c'è scritto sulle grammatiche scolastiche.

E anzi ricercatori e ricercatrici sono interessati/e proprio a studiare il comportamento linguistico dei parlanti oltre quelle convenzioni, osservando la lingua dal punto di vista **biologico** (come strumento del pensiero), **psicologico** (come manifestazione delle nostre individualità e mezzo per esprimerci) e **culturale** (come prodotto della nostra evoluzione).

In quanti modi  
posso  
pronunciare la  
*n* in italiano?

s. capogruppo  
p. capigruppo

s. capolavoro  
p. capolavori

*"Ho mangiato  
un panino con  
le mani".*

Se Lucia *ha  
tagliato la corda*, è  
fuggita o ha diviso  
una fune in due  
parti con le  
forbici?

*Il Lonfo non  
vaterca né gluisce  
e molto  
raramente  
barigatta...*

*Maschile e  
femminile  
esistono in  
tutte le lingue?*

In quanti modi  
posso  
pronunciare la  
*n* in italiano?

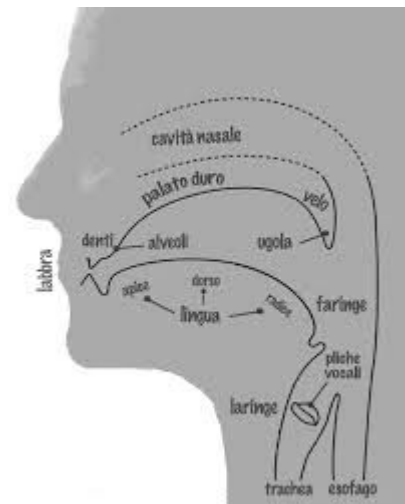
Prova a pronunciare, senza spezzare il suono ma facendo particolare attenzione a dove è posizionata la lingua quando pronunciate la lettera “n”, le parole:

cane - ancora - anfora - Giannpaolo

Mentre la grafia non cambia, il suono che pronunciamo è leggermente diverso.

In questo caso, la variazione dipende dal contesto in cui si trova la lettera:

- in posizione intervocalica (cioè tra due vocali), come in cane, pronunciamo una nasale alveolare
- in associazione con occlusive velari come la “c” di ancora o la “g” di valanga, otteniamo una nasale velare
- in associazione a suoni labiodentali come la “f” di anfibio o la “v” di Monviso, otteniamo una nasale labiodentale
- in associazione a consonanti bilabiali come la “p” in Giannpaolo, otteniamo invece una nasale bilabiale, meglio conosciuta come “m”





s. capogruppo  
p. capigruppo

s. capolavoro  
p. capolavori

A prima vista, sembrerebbe che *capogruppo* e *capolavoro* siano parole composte del tutto simili (CAPO + GRUPPO, CAPO + LAVORO).

Un(a) parlante di una lingua diversa dall'italiano potrebbe quindi pensare di formare in entrambi i casi il plurale come per i nomi maschili italiani in -o (ovvero, albero - alberi, mondo - mondi, tavolo - tavoli etc.).

**Come mai questi due nomi si comportano diversamente?**

Il comportamento è dovuto alla posizione del componente **testa**, cioè il componente fondamentale senza il quale la parola non sussisterebbe. Sia *capogruppo* che *capolavoro* sono infatti nomi composti, ovvero formati da due componenti parzialmente indipendenti tra loro.


Nel caso di “*capogruppo*” la testa è *capo* (ovvero: il capogruppo è un capo), ed è questa che viene posta al plurale in *capigruppo*.

Nel caso di “*capolavoro*”, invece, la testa è *lavoro* (ovvero: il capolavoro è un lavoro), e il plurale risulta quindi *capolavori*.



Questa frase sembrerebbe non presentare alcun tipo di problema, leggendola immaginiamo un succulento panino addentato durante un picnic sul mare.

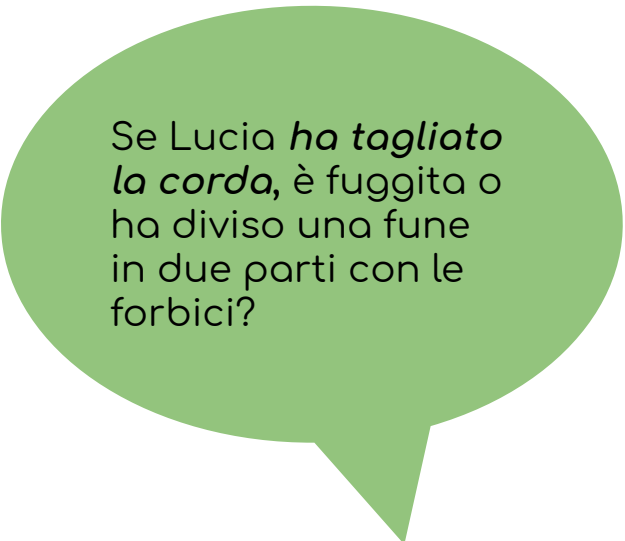
**E se la frase fosse stata pronunciata da Hannibal Lecter?**



*“Ho mangiato  
un panino con  
le mani”.*

Siamo davanti a un classico problema di *attachment* (cioè di “collegamento”): dobbiamo decidere se il sintagma *con le mani* modifica il verbo (*“ho mangiato con le mani”*), o se fa parte del complemento oggetto (*“un panino con le mani”*).

Siamo in grado di preferire una interpretazione (*“ho mangiato con le mani”*) rispetto all'altra (*“panino con le mani”*) grazie a **conoscenze extralinguistiche**: se la frase fosse stata *“ho mangiato un panino con le olive”* avremmo fatto scelte diverse, anche se non c'è nulla nella struttura della frase che ci faccia propendere per l'una o per l'altra analisi!



Se Lucia *ha tagliato la corda*, è fuggita o ha diviso una fune in due parti con le forbici?

Come facciamo a capire le frasi che ci vengono dette? Per esempio, come sappiamo cosa vuol dire un'espressione come "*Lucia compra il pane*"?

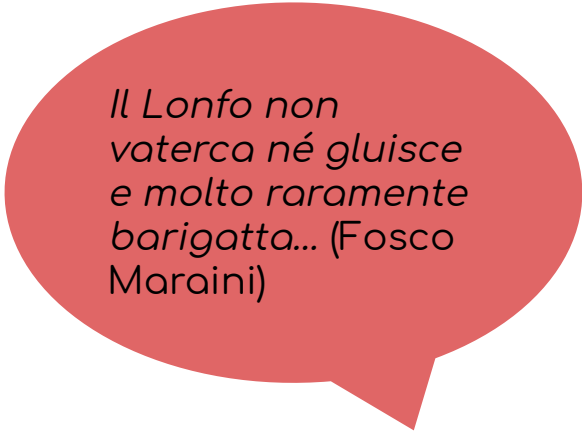
"Sommiamo" i significati delle singole parole (Lucia + comprare + pane) ed otteniamo il senso generale della frase.

E allora perchè in questo caso non funziona?

Fraasi come "*tagliare la corda*", "*ingoiare un rospo*", "*avere la testa tra le nuvole*", ecc. si chiamano **espressioni idiomatiche**, e si riconoscono perchè hanno un significato **non compositazionale**, cioè che non si può dedurre dai significati delle parole che le compongono.

Infatti, per capire la frase "*Lucia ha tagliato la corda*" dobbiamo conoscere il significato dell'espressione generale, e non possiamo derivare il significato da quello delle sue parti.

Spesso i linguisti e le linguiste costruiscono esperimenti in cui usano parole simili a quelle della famosa [poesia di Fosco Maraini](#): parole che sembrano italiane per come sono formate, ma che non vogliono dire nulla.



*Il Lonfo non  
vaterca né gluisce  
e molto raramente  
barigatta... (Fosco  
Maraini)*

**Ma sono veramente *senza senso*?**

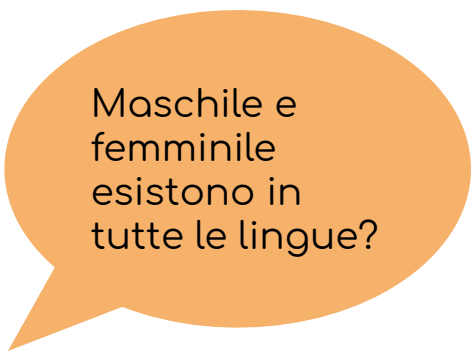
Il *Lonfo* sembrerebbe essere un animale, che non ha alcuni comportamenti (*vatercare*, *gluire*) e ne ha un altro sporadicamente (*barigattare*).

Riusciamo a immaginare un possibile significato perché - oltre al significato lessicale - anche la **struttura sintattica** della frase e **morfologica** delle parole contribuisce al significato dell'enunciato. Il *Lonfo* potrebbe essere un animale perché, oltre ad avere un aspetto morfologico compatibile con i nomi di animale italiani, è il soggetto (*agente*) delle azioni predicate da verbi, e questo è spesso un essere animato. I verbi poi sono tutti intransitivi, quindi simili ai verbi usati per descrivere il verso o i comportamenti degli animali (*grugnire*, *galoppare*...).

Usando parole così fatte nei loro esperimenti, i linguisti riescono a studiare la sintassi senza possibili "distrazioni" da parte del lessico.

Esiste un'ovvia relazione tra le **parole** della nostra lingua e i **concetti** che utilizziamo: vi sarà capitato di incontrare parole difficilmente traducibili, o di trovarvi in situazioni in cui avreste proprio voluto usare una parola in dialetto, perché sembra esprimere la vostra idea in modo più calzante.

**Ma esistono dei concetti universali, comuni a tutta l'umanità?**



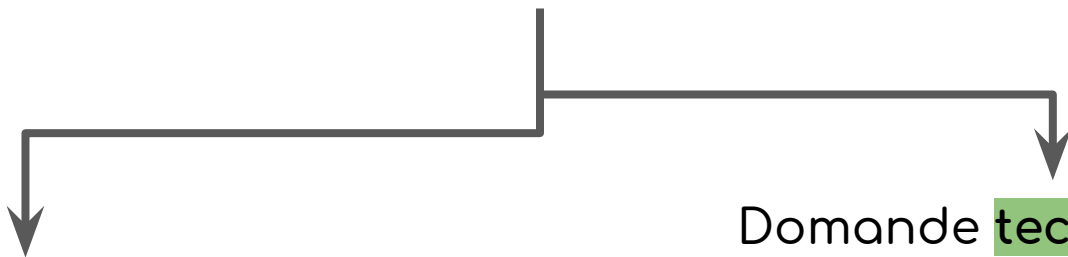
Maschile e femminile esistono in tutte le lingue?

L'analisi di una larghissima parte delle lingue del mondo (di cui si occupa la **linguistica tipologica**) ha rivelato che la relazione tra lessico e mondo concettuale è più complessa di ciò che intuitivamente potremmo pensare.

Certo, esistono delle tendenze *universali*, ma ogni lingua "sceglie" modi leggermente diversi di codificare l'esperienza, e molte similarità sono spesso dovute al contatto culturale tra i popoli. Pensa ad esempio alla parola "*aperitivo*", o al "*brunch*" inglese: esprimono concetti culturalmente specifici ed è spesso necessaria una perifrasi per tradurli adeguatamente.

E così, se noi italiani siamo portati a dividere il mondo in *maschile* e *femminile*, **esistono lingue** che non evidenziano alcun genere a livello grammaticale e lingue che ne hanno anche più di cinque. È evidente qui come il numero o il tipo di generi grammaticali sia quindi indipendente dalla nostra capacità di categorizzare i concetti in uno, due o più gruppi.

# E la Linguistica Computazionale, allora?



## Domande linguistiche:

Il nostro cervello è come un computer:

Come possiamo capire la mente e il linguaggio usando simulazioni computazionali?

Cosa possiamo capire dei nostri processi mentali studiando il linguaggio?

## Domande tecnologiche:

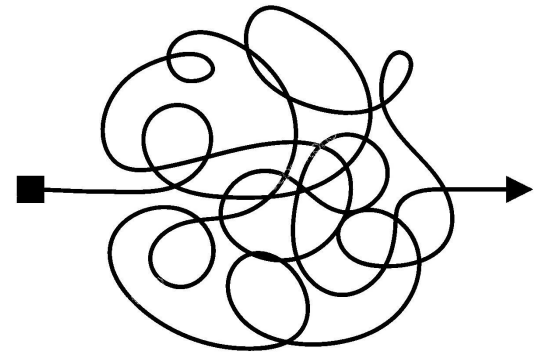
Siri, che tempo farà domani?

Alexa, sai tradurre un romanzo?

In Cina non riuscirei a capire neanche i segnali stradali, il mio smartphone mi aiuterà a orientarmi?



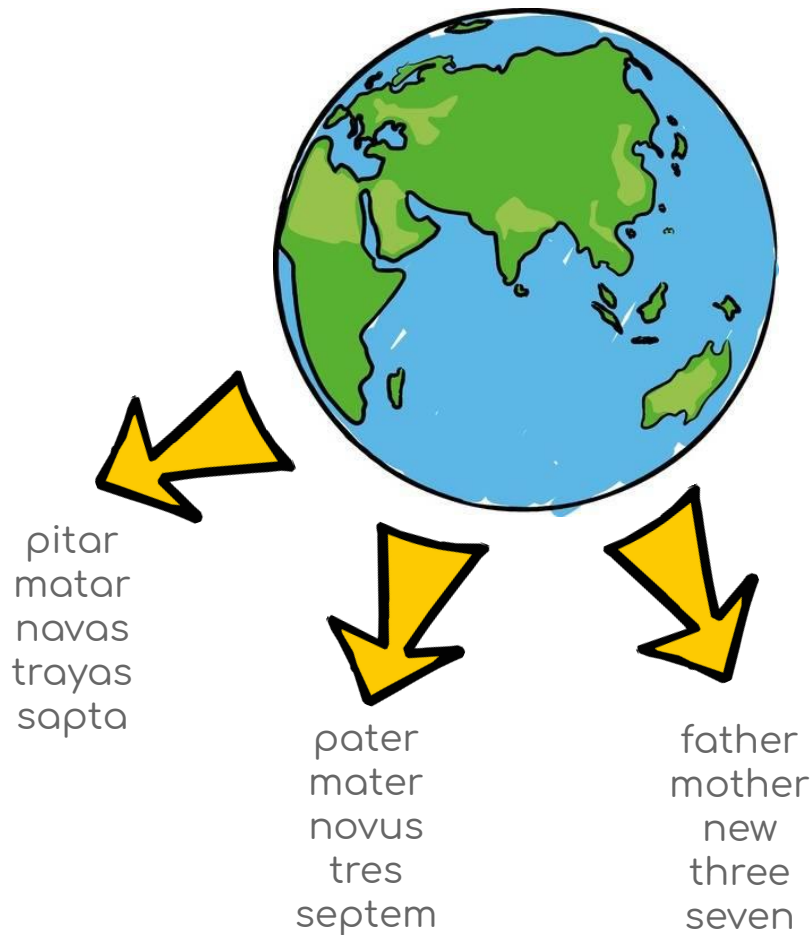
Da **dove** viene,  
la Linguistica  
Computazionale?



# Parentele linguistiche

Nel 1786 sir William Jones, un giudice inglese di servizio in India, notò somiglianze tra i suoni di molte lingue parlate in Europa, India e Persia.

Da questa osservazione si arrivò a scoprire, grazie allo studio sistematico delle lingue euro-asiatiche, che queste lingue derivano tutte da un antenato linguistico comune, il Proto-Indoeuropeo.

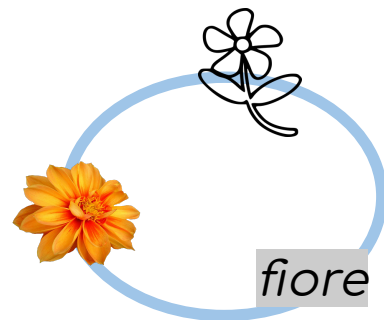


# Tante linguistiche per tanti linguisti



Negli anni '50 Noam Chomsky fonda la **SCUOLA GENERATIVISTA**: il linguaggio è innato e regolato da una **Grammatica Universale**.

Ferdinand de Saussure e gli **STRUTTURALISTI** studiano la differenza tra **lingua** (sistema di regole) e **linguaggio** (uso che i parlanti fanno della lingua)



Ogni lingua "usa" un sottoinsieme di **regole possibili**, che sono definite dalla **Grammatica Universale**: ognuno di noi le impara da piccolo e le usa automaticamente.





Negli anni '60, William Labov contribuì alla diffusione della **SOCIOLINGUISTICA**.

Il **gruppo sociale** di appartenenza, il **genere**, la **provenienza** geografica influiscono sulla lingua.

Negli anni '70 nasce l'idea che la grammatica emerga dall'**uso** della lingua: il significato dipende da **frequenza** e **distribuzione** delle parole.

Una linguista importantissima in questo campo è Joan Bybee.



Possiamo distinguere due tipi di approccio allo studio del linguaggio:

- esperimenti e test su **soggetti umani**.

Una famosa studiosa italiana è Cristina Cacciari

- analisi di grandi collezioni di linguaggio (parlato o scritto): i **corpora**.

Un famoso linguista in quest'ambito è Stephan Gries.

# E la linguistica computazionale?

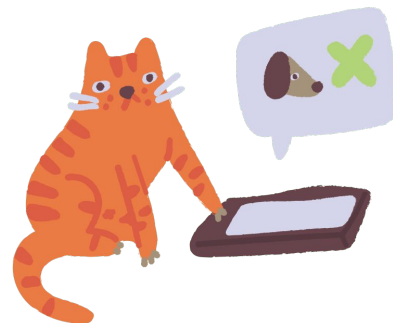
Negli anni '40 e '50, con l'avvento dei primi **calcolatori elettronici**, alcuni videro nei neonati computer una grande possibilità per risolvere problemi di **traduzione automatica** e per l'analisi di grandi quantità di testo.

**Padre Roberto Busa** fu uno dei primi ad usare strumenti informatici per l'analisi degli scritti di San Tommaso d'Aquino.



Grazie a computer sempre più potenti, la linguistica computazionale acquista con il tempo sempre più importanza e realizza sempre più applicazioni, come il **riconoscimento vocale**.

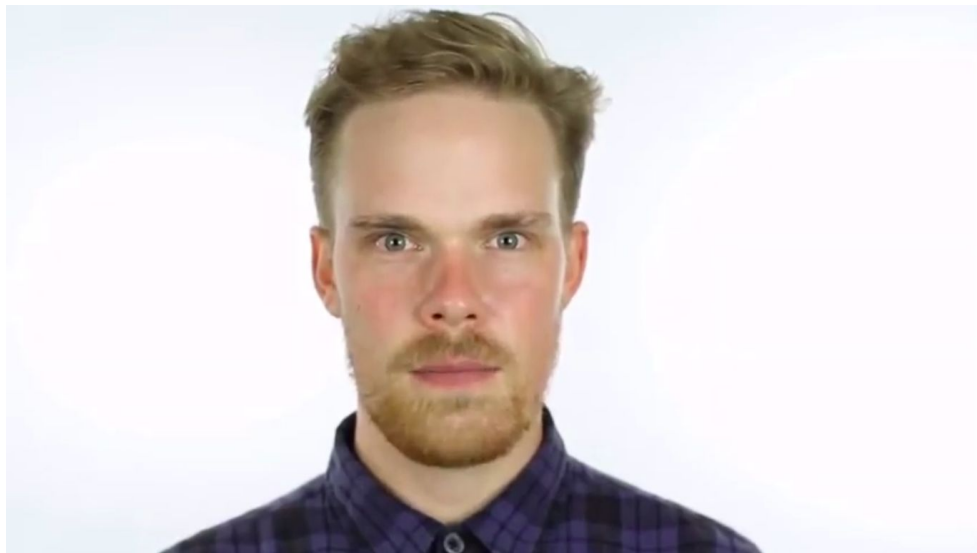
Le ultime sfide per la linguistica computazionale la avvicinano all'**intelligenza artificiale**: la possibilità di simulare con un computer la risoluzione di problemi che richiedono intelligenza umana.



# 1. Ma i computer, sentono?



# E voi, ci sentite bene?



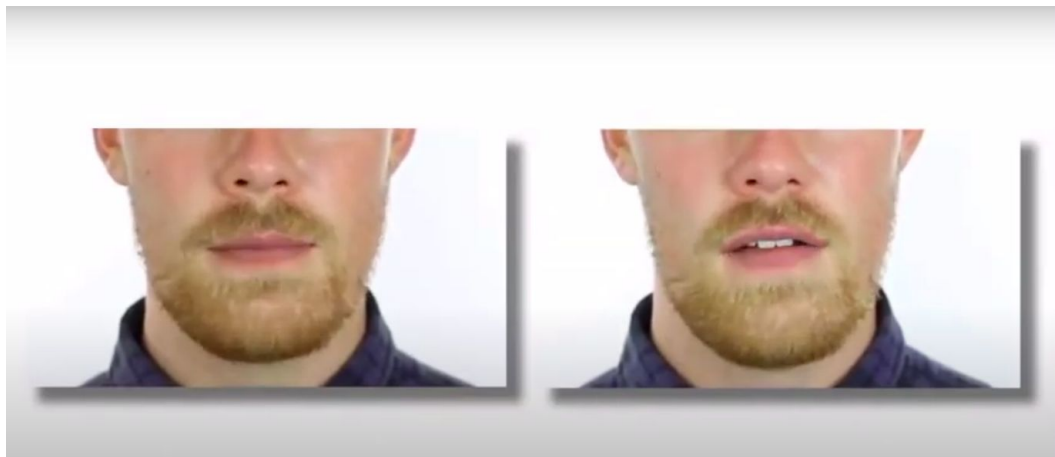
Che suono sta  
pronunciando il  
ragazzo nel video?

# E voi, ci sentite bene?



Che suono sta  
pronunciando il  
ragazzo nel video?

# E voi, ci sentite bene?



Questa volta guardate  
il video più volte,  
concentrati prima sulla  
porzione a sinistra e  
poi su quella a destra.

Cosa succede?

# L'effetto McGurk

Con ogni probabilità, avrai pensato che il ragazzo del primo video stesse pronunciando la parola *par* o *bar*, mentre il ragazzo del secondo video stesse pronunciando *var* o *far*.

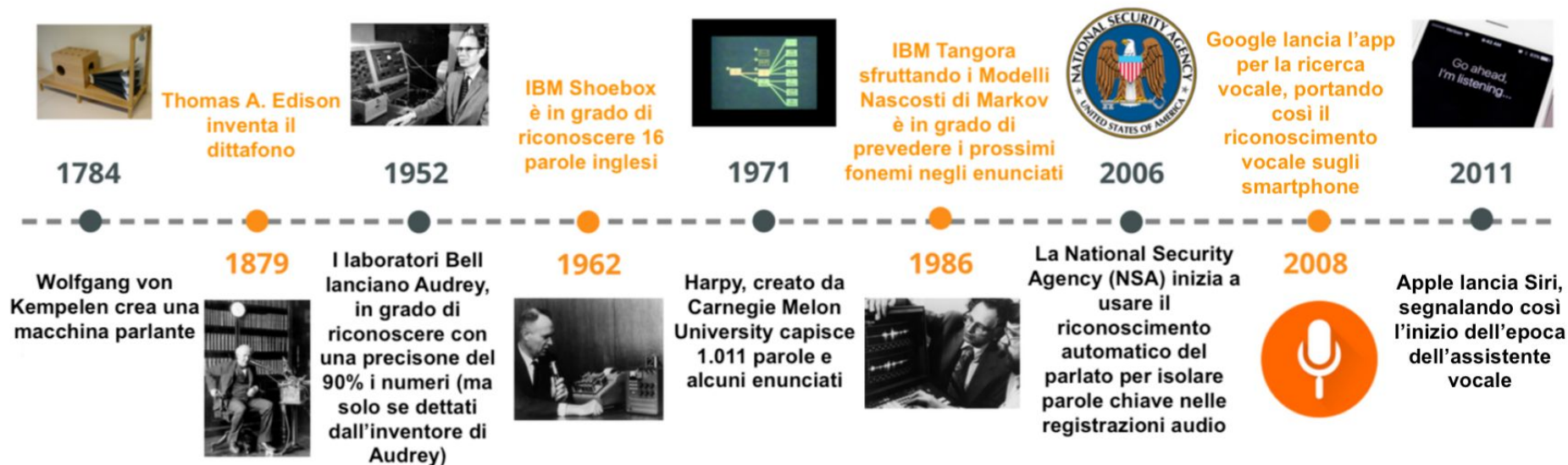
Guardando il terzo video, ciò che sentiamo cambia a seconda di dove rivolgiamo la nostra attenzione: se guardiamo le labbra sulla sinistra percepiamo *par*, guardando a destra percepiamo invece *far*.

L'effetto è dovuto a un'illusione acustica meglio nota come effetto McGurk: un fenomeno percettivo descritto per la prima volta da Harry McGurk e John MacDonald nel 1976.

L'effetto McGurk dimostra la multimodalità del processo di riconoscimento di un suono o di una parola: c'è un'interazione tra l'udito e la vista e l'informazione ricevuta dalla fonte visiva può cambiare la percezione dei suoni o delle parole che sentiamo.

In altre parole, è stato il movimento delle labbra del ragazzo nel video a modificare la nostra percezione: il suono riprodotto è *par* (o *bar*) in tutti e tre i video proposti. Abbiamo sentito *far* solo quando il movimento delle labbra ci ha tratto in inganno.

# Tentativi ed errori...



Basato su Boyd (2018), Speech Recognition Technology: The Past, Present, and Future. **ClickZ**



Come fanno i computer a riconoscere e a capire il parlato?

Un compito molto complesso, composto da diversi passaggi...



# Come fanno i computer a riconoscere e a capire il parlato?

Il primo passo è  
il  
riconoscimento  
vocale



si passa dal  
parlato al testo



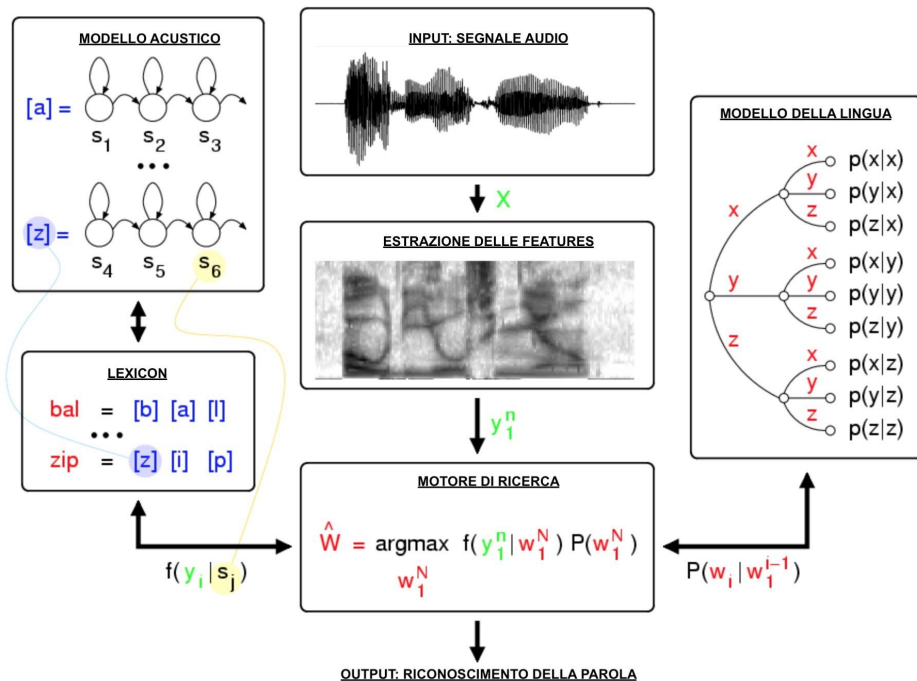
# Come fanno i computer a riconoscere e a capire il parlato?

Un riconoscitore vocale cerca di **trascrivere** un segnale acustico.

Il calcolatore ha già a disposizione molte associazioni tra segnali e parole e cerca quella più simile al suono percepito.

È una delle tante applicazioni del teorema di Bayes sulle probabilità condizionate: cerchiamo, tra tutte quelle a disposizione, la parola che ha probabilità più alta di essere associata al segnale acustico che abbiamo a disposizione

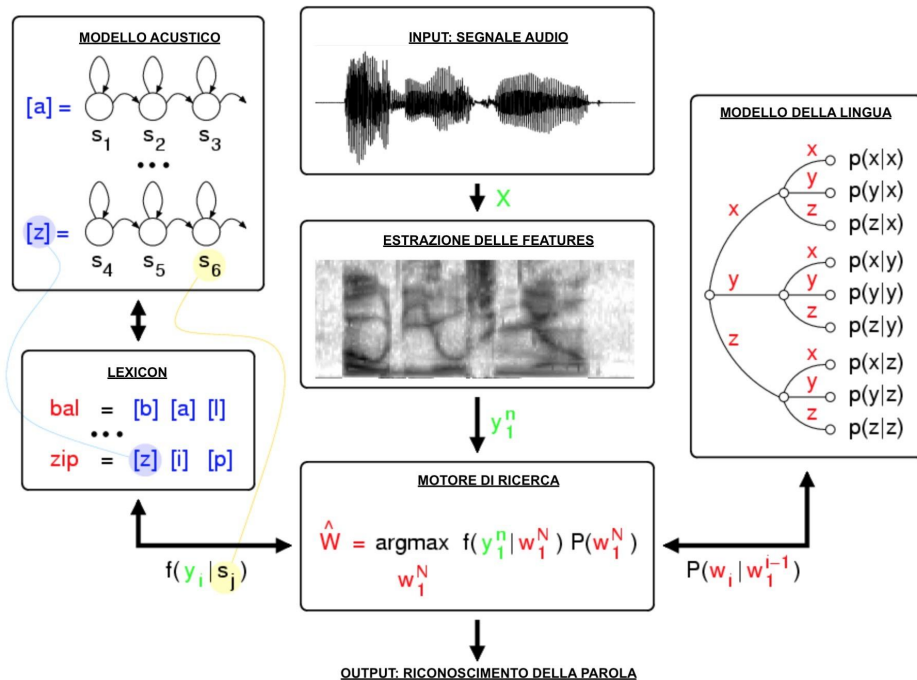
$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W/X) = \underset{W}{\operatorname{argmax}} \frac{f(X/W) P(W)}{f(X)}$$



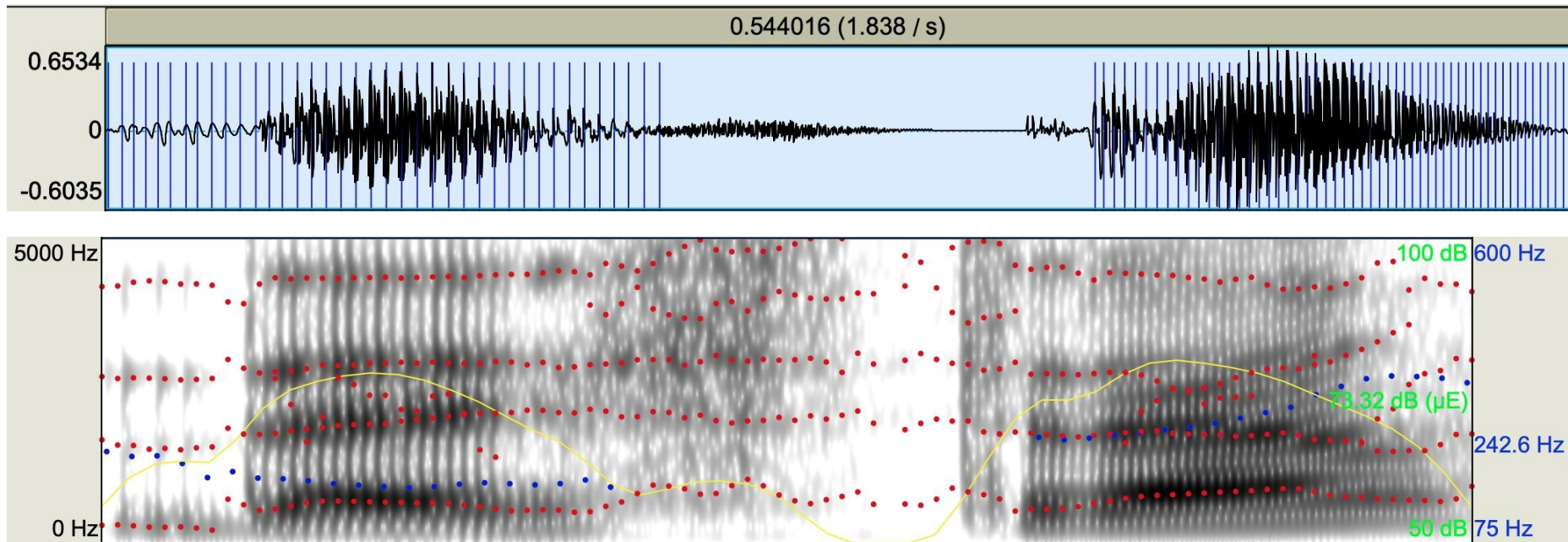
# Come fanno i computer a riconoscere e a capire il parlato?

Le vibrazioni dell'audio vengono convertite in impulsi elettrici e poi in segnale digitale.

Così individuiamo suoni e intonazione, riuscendo a riconoscere, per esempio, domande o esclamazioni.



Voi riuscireste a riconoscere una parola solo dalla forma d'onda e dallo spettrogramma?

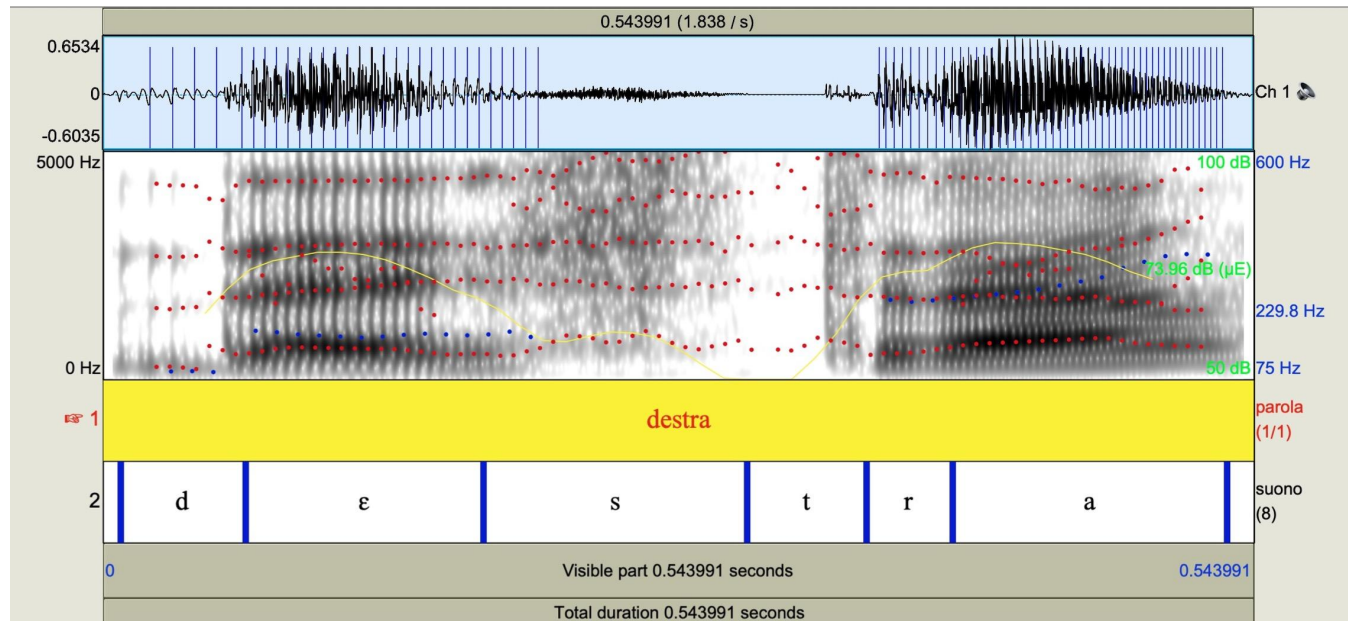


# Per voi è stato difficile, vero?

forma d'onda



spettrogramma



# Parole chiave: forma d'onda e spettrogramma

**Forma d'onda** = rappresentazione grafica sul piano cartesiano dell'andamento di un segnale audio nel tempo.

**Spettrogramma** = rappresentazione grafica dell'intensità di un suono in funzione del tempo e della frequenza.

Ci permettono di:

- “vedere” le **caratteristiche fisiche** dei suoni (intensità, frequenza)
- **isolare** (*segmentare*) più facilmente i suoni (in particolare le vocali)
- osservare i fenomeni di **coarticolazione** (i suoni subiscono l'influenza del contesto nel quale vengono articolati, ovvero dei suoni che li precedono o che li seguono)

Per un calcolatore non è difficile  
come lo è per noi...

MA

solo se chi parla usa una varietà  
standard

A differenza nostra, i sistemi  
automatici non sono ancora in  
grado di comprendere allo stesso  
modo gli **accenti regionali** (o  
stranieri), o i discorsi prodotti da  
**pazienti** con qualche disturbo del  
linguaggio...

Question 4: Which writer wrote <i>Westworld</i> ?			
	Siri	Alexa	Google Home
American	✓	✓	✓
British	✓	✗	✓
Scottish	✗	✗	✗
Irish	✓	✗	✗
Australian	✓	✗	✓
Japanese	✗	✓	✓
Italian	✗	✓	✓
German	✓	✗	✓

Basato su Wood & Kirshen (2017), Esperimento sulla comprensione del parlato accentato (<https://www.youtube.com/watch?v=GZnUibN6m4A&frags=pl%2Cwn>)

un [video](#) sullo stesso esperimento



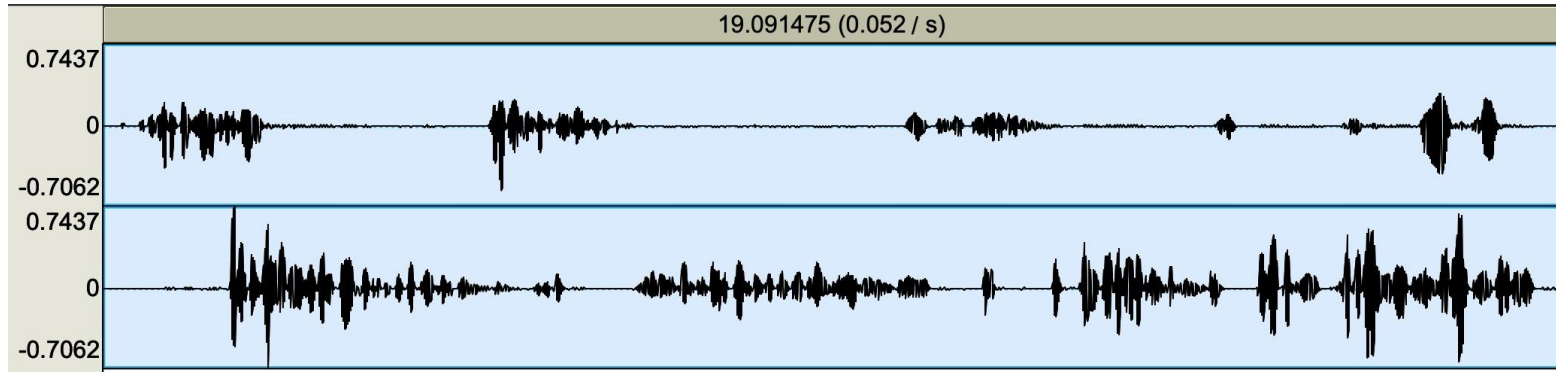
E voi come ve la  
cavate con gli  
accenti regionali?

Riuscite a  
riconoscere la  
provenienza delle  
persone sentendole  
parlare?



# Di dove sono le due ragazze?

Indizio: vengono entrambe dalla stessa città

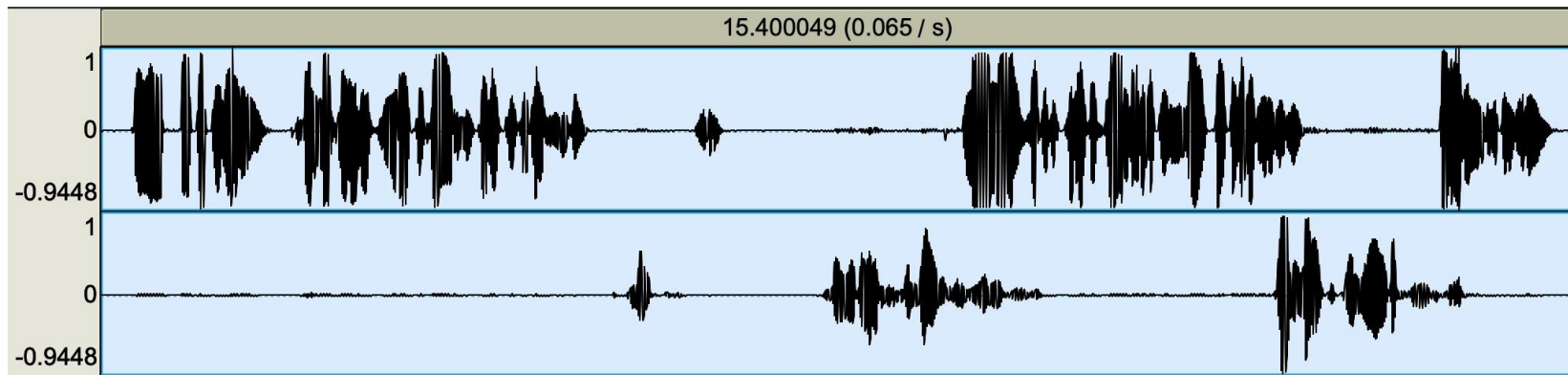


# Roma!



# Di dove sono i due ragazzi?

Indizio: vengono entrambi dalla stessa città



Materiale audio scaricato dal Corpus CLIPS: (<http://www.clips.unina.it/it/corpus.jsp>)

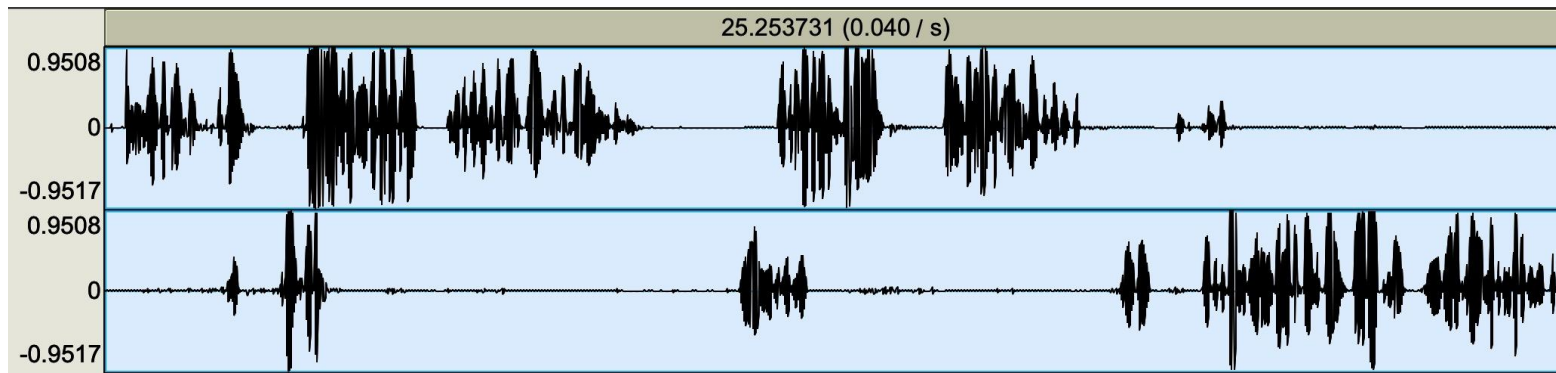
# Bari!





# Di dove sono i due ragazzi?

Indizio: vengono entrambi dalla stessa città



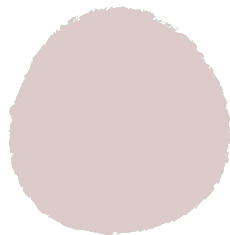
# Bergamo!



Riconoscere gli *accenti* e comprendere il parlato *non-standard* (regionali, straniere, patologiche) per noi è abbastanza facile.

Lo facciamo in automatico, senza riflettere su *prosodia* (*intonazione, ritmo, intensità*) oppure sulla pronuncia dei singoli suoni.

Ma per i calcolatori, questi tratti rappresentano forme di *variabilità*: se un riconoscitore vocale è addestrato sulla lingua standard, le deviazioni da questa pronuncia creano ostacoli alla decodifica.





## 2. Ma i computer, sanno leggere?



giardino

è

il

nel

mio

cane

Quante frasi possibili?



# Quante frasi possibili?

il mio cane è nel giardino

il cane è nel mio giardino

il giardino è nel mio cane

il cane nel giardino è mio



# Quante frasi possibili?

il mio cane è nel giardino

il cane è nel mio giardino

il giardino è nel mio cane

il cane nel giardino è mio

nel cane giardino mio è il

giardino il mio è cane nel

cane nel è il giardino mio

# Quante frasi possibili?

il mio cane è nel giardino

il cane è nel mio giardino

il giardino è nel mio cane

il cane nel giardino è mio

nel cane giardino mio è il

giardino il mio è cane nel

cane nel è il giardino mio

$$6*5*4*3*2*1 =$$

**720** frasi possibili!

# Quante frasi probabili?

il mio cane è nel giardino

il cane è nel mio giardino

il giardino è nel mio cane

il cane nel giardino è mio

nel cane giardino mio è il

giardino il mio è cane nel

cane nel è il giardino mio

# Quante frasi probabili?

il mio cane è nel giardino

il cane è nel mio giardino

il giardino è nel mio cane

il cane nel giardino è mio

nel cane giardino mio è il

giardino il mio è cane nel

cane nel è il giardino mio

## Perché?



# Una definizione intuitiva di **probabilità**

Assegnare un valore di probabilità a un **evento** significa fornire una misura dell'**aspettativa** che quell'evento si realizzi.



Ma come facciamo a **misurare** la nostra aspettativa?

Facciamo un esempio:

*Andrea e Lucia oggi giocheranno a scacchi.*

*Chi ci aspettiamo che vinca?*

Per poter fare una stima affidabile, ci occorrono almeno due informazioni:

1. Quante partite hanno **vinto** a testa?
2. Quante partite hanno **giocato** a testa?

Noi parlanti basiamo i nostri giudizi sulla nostra **storia linguistica**: tutte le frasi che abbiamo sentito, letto e pronunciato durante la nostra vita.

In maniera simile, i calcolatori hanno bisogno di un grande numero di **dati linguistici** per calcolare gli eventi più probabili.

Una raccolta rappresentativa di testi viene chiamata

**CORPUS**

piace

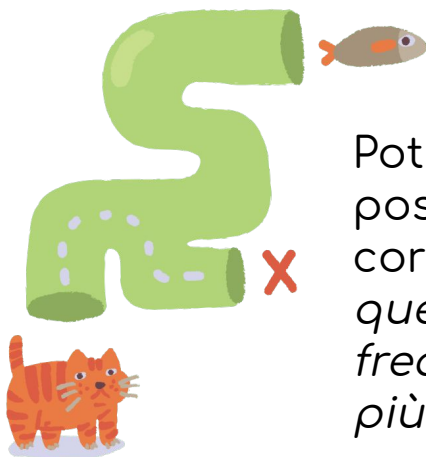
il

mio

gatto

bambino

al



Potremmo scrivere tutte le 720 possibili frasi e cercarle nel corpus:

*quelle che ricorrono più frequentemente saranno le più probabili...*

MA:

le frasi sono potenzialmente infinite, e ci sono frasi che vorremmo considerare possibili anche se non le abbiamo mai viste (per esempio: “Il gatto stava leggendo un papiro sull’orlo di un burrone”).

Allora, facciamo finta che la frase sia un  
braccialetto:

il

è

mio

giardino

nel

cane

Quale parola è più  
probabile trovare all'inizio  
di una frase in italiano?

Allora, facciamo finta che la frase sia un  
braccialetto:

il

è

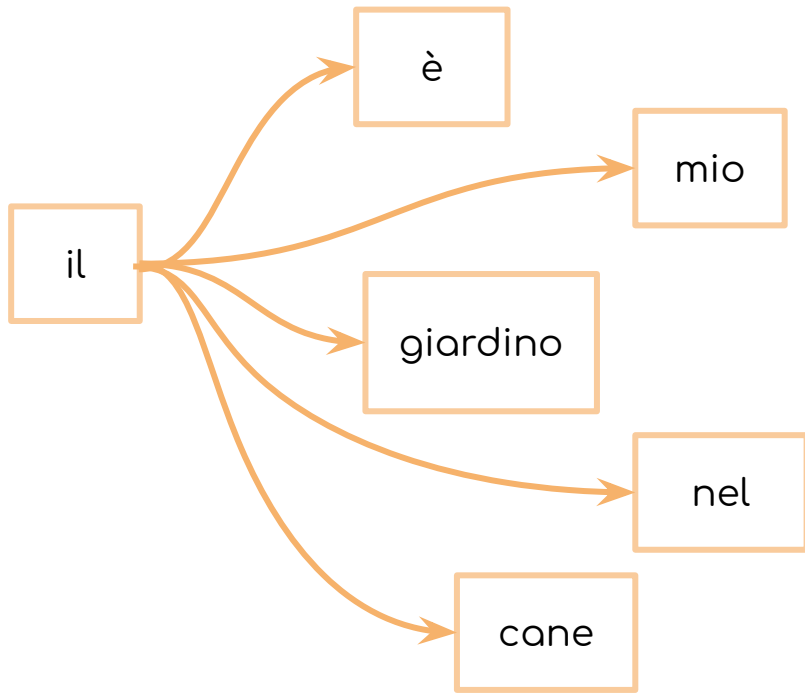
mio

giardino

nel

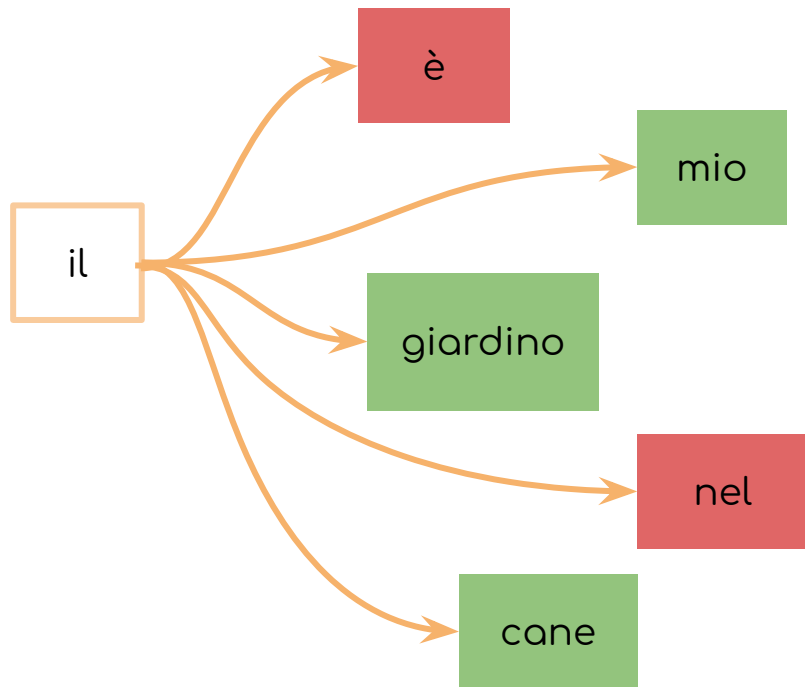
cane

Allora, facciamo finta che la frase sia un  
braccialetto:

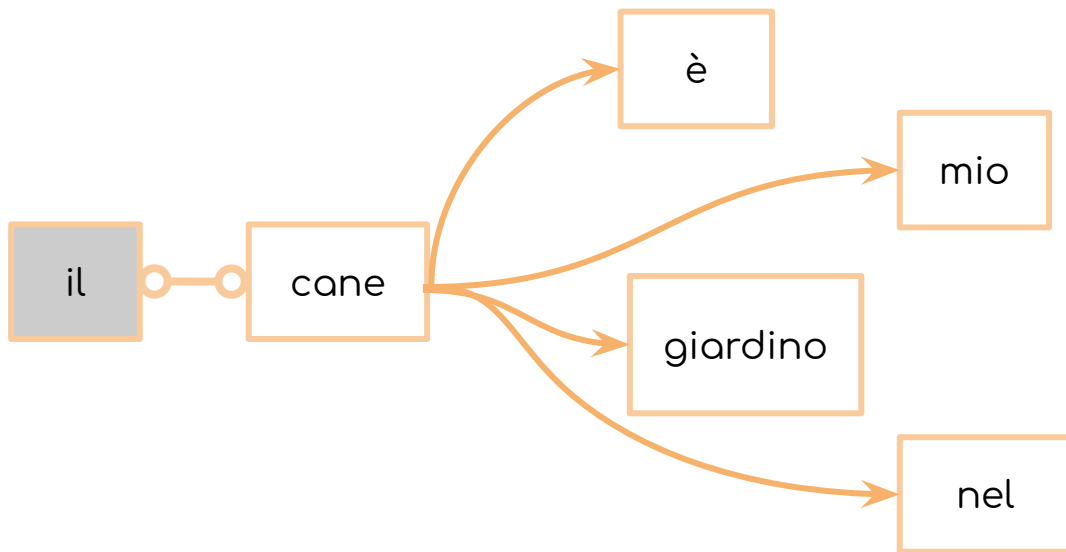


Abbiamo scelto di iniziare con "il", cosa è più probabile trovare subito dopo?

Allora, facciamo finta che la frase sia un  
braccialetto:



Allora, facciamo finta che la frase sia un braccialetto:

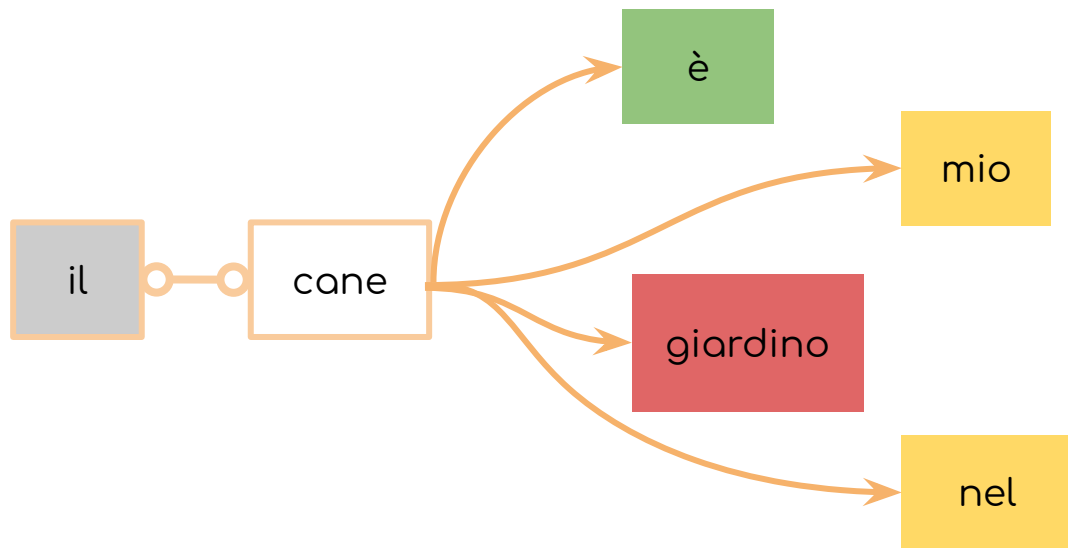


Abbiamo scelto “cane”.  
Dobbiamo ora chiederci cosa è probabile trovare dopo “cane” nelle frasi italiane.

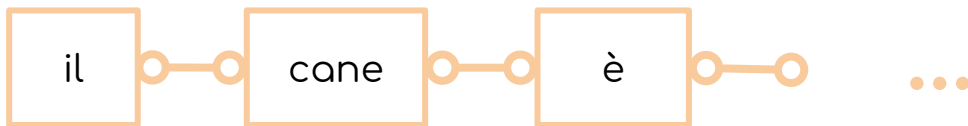
Nel modello più semplice, possiamo dimenticarci di “il”.



Allora, facciamo finta che la frase sia un braccialeto:



Allora, facciamo finta che la frase sia un  
braccialeto:



# Una procedura per riordinare la frase:

1. Scegliamo la parola più probabile per iniziare e posizioniamola
2. Finché abbiamo ancora parole a disposizione:
  - a. guardiamo l'ultima parola posizionata P
  - b. cerchiamo nel corpus tutte le occorrenze di P seguita da un'altra parola
  - c. scegliamo la parola più probabile, e posizioniamola nella frase

# Una procedura per riordinare la frase:

1. Scegliamo la parola più probabile per iniziare e posizioniamola
2. Finché abbiamo ancora parole a disposizione:
  - a. guardiamo l'ultima parola posizionata P
  - b. cerchiamo nel corpus tutte le occorrenze di P seguita da un'altra parola
  - c. scegliamo la parola più probabile, e posizioniamola nella frase

E quindi, che frase  
ti aspetti?

# Una **procedura** per riordinare la frase:

1. Scegliamo la parola più probabile per iniziare e posizioniamola
2. Finché abbiamo ancora parole a disposizione:
  - a. guardiamo l'**ultima** parola posizionata P
  - b. cerchiamo nel corpus tutte le occorrenze di P seguita da un'altra parola
  - c. scegliamo la parola più probabile, e posizioniamola nella frase

Ma perché tenere conto solo dell'ultima parola?

Cosa cambierebbe se, per esempio, tenessimo conto delle ultime due?

# Regole del Gioco

*Noam*, del popolo degli **Hapax**, e *Zellig*, del popolo **Token**, stavano chattando su *LinguApp* quando un cortocircuito ha interrotto la loro comunicazione.

Gli ultimi messaggi scritti da *Noam* e *Zellig* si sono spezzettati. Il vostro ruolo, di calcolatori, è di fare sì che il messaggio venga ricostruito e arrivi a destinazione.

Mentre ricostruiamo il messaggio, cerchiamo di scoprire cosa i calcolatori fanno o non fanno con il linguaggio.

3. Ma i computer,  
sanno la  
grammatica?

# Alla **taverna** di Siri...

fusilli al pomodoro  
pollo agli asparagi  
pannacotta

penne agli asparagi  
pollo con funghi  
tiramisù

fusilli ai funghi  
cotoletta con patate  
frutta

penne ai funghi  
pollo agli asparagi  
tiramisù



penne agli asparagi  
pollo con funghi  
tiramisù

fusilli al pomodoro  
pollo agli asparagi  
pannacotta

penne ai funghi  
pollo agli asparagi  
tiramisù

fusilli ai funghi  
cotoletta con patate  
frutta

I pasti possibili sono tanti,  
ma sembra siano stati  
scritti tutti seguendo una  
regola generale...

Pasto  
=  
Primo + Secondo + Dessert

penne agli asparagi  
pollo con funghi  
tiramisù

fusilli al pomodoro  
pollo agli asparagi  
pannacotta

penne ai funghi  
pollo agli asparagi  
tiramisù

fusilli ai funghi  
cotoletta con patate  
frutta

I pasti possibili sono tanti,  
ma sembra siano stati  
scritti tutti seguendo una  
regola generale...

Pasto  
=  
Primo + Secondo + Dessert

In questo modo abbiamo creato tre nuove **categorie**:

## Primo

penne agli asparagi  
fusilli al pomodoro  
fusilli ai funghi  
penne ai funghi

## Secondo

pollo con funghi  
cotoletta con patate  
pollo agli asparagi

## Dessert

tiramisù  
frutta  
pannacotta

E una **regola**:

Pasto = Primo + Secondo + Dessert

Osserviamo meglio la categoria:

## Primo

penne agli asparagi  
fusilli al pomodoro  
fusilli ai funghi  
penne ai funghi

## Secondo

pollo con funghi  
cotoletta con patate  
pollo agli asparagi

## Dessert

tiramisù  
frutta  
pannacotta

Osserviamo meglio la categoria:

## Primo

penne agli asparagi

fusilli al pomodoro

fusilli ai funghi

penne ai funghi

**Primo** = **Formato** + **Condimento**

Abbiamo creato altre 2 categorie e  
nuove regole:

Pasto = Primo + Secondo + Dessert

**Primo = Formato + Condimento**

Osserviamo meglio la categoria:

## Primo

penne agli asparagi

fusilli al pomodoro

fusilli ai funghi

penne ai funghi

**Primo** = **Formato** + **Condimento**

Abbiamo creato altre 2 categorie e  
nuove regole:

Pasto = Primo + Secondo + Dessert

Primo = Formato + Condimento

Formato = penne *oppure* fusilli *oppure* ...

Condimento = funghi *oppure* asparagi *oppure* ...

**Primo** = Condimento + Formato

asparagi alle penne

pomodoro ai fusilli

funghi ai fusilli

funghi alle penne

**Attenzione!**

L'operazione di concatenazione ("+") non è commutativa!

L'ordine in cui compaiono gli operandi non può essere modificato



Proviamo a scrivere le regole per la categoria “Secondo”

## Primo

penne agli asparagi  
fusilli al pomodoro  
fusilli ai funghi  
penne ai funghi

## Secondo

pollo con funghi  
cotoletta con patate  
pollo agli asparagi

## Dessert

tiramisù  
frutta  
pannacotta



# Ricapitoliamo:

Pasto = Primo + Secondo + Dessert

Primo = Formato + Condimento

Secondo = Pietanza + Condimento

Dessert = tiramisù *oppure* pannacotta *oppure* frutta

Formato = penne *oppure* fusilli *oppure* ...

Pietanza = pollo *oppure* cotoletta *oppure* ...

Condimento = funghi *oppure* asparagi *oppure* pomodoro *oppure* ...

# Formalmente, una grammatica è definita da...

Un alfabeto di simboli terminali → {fusilli; pannacotta; funghi; frutta; pomodoro; asparagi; patate... }

Un insieme di categorie → {pasto; primo; secondo; dessert; formato; condimento; pietanza}

Una categoria iniziale → pasto

Un insieme di regole → {  
Pasto = Primo + Secondo + Dessert;  
Primo = Formato + Condimento;  
Secondo = Pietanza + Condimento;  
...  
}



Possiamo scrivere una grammatica per qualsiasi cosa che mostra regolarità!

Usando queste regole,  
riusciamo ad analizzare  
il menu?

Pasto = Primo + Secondo + Dessert

Primo = Formato + Condimento

Secondo = Pietanza + Condimento

Dessert = tiramisù oppure pannacotta oppure frutta  
oppure ...

Formato = penne oppure fusilli oppure ...

Pietanza = pollo oppure cotoletta oppure ...

Condimento = funghi oppure asparagi oppure  
pomodoro oppure ...

# Usando queste regole, riusciamo ad analizzare il menu?

penne ai funghi  
pollo agli asparagi  
tiramisù

Pasto = Primo + Secondo + Dessert

Primo = Formato + Condimento

Secondo = Pietanza + Condimento

Dessert = tiramisù oppure panna cotta oppure frutta  
oppure ...

Formato = penne oppure fusilli oppure ...

Pietanza = pollo oppure cotoletta oppure ...

Condimento = funghi oppure asparagi oppure  
pomodoro oppure ...

# Usando queste regole, riusciamo ad analizzare il menu?

penne ai funghi  
pollo agli asparagi  
tiramisù



Formato Condimento  
Pietanza Condimento  
Dessert

Pasto = Primo + Secondo + Dessert

Primo = Formato + Condimento

Secondo = Pietanza + Condimento

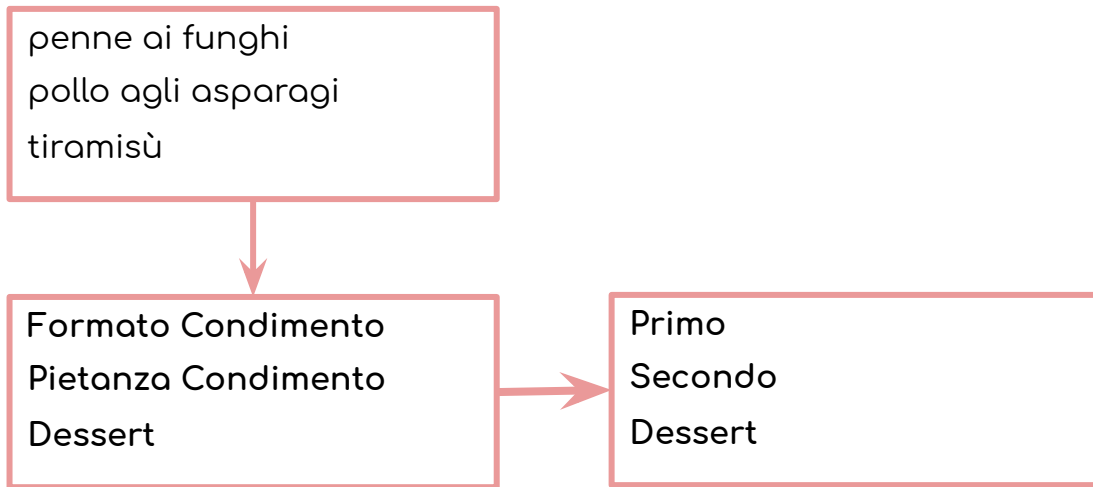
Dessert = tiramisù oppure panna cotta oppure frutta  
oppure ...

Formato = penne oppure fusilli oppure ...

Pietanza = pollo oppure cotoletta oppure ...

Condimento = funghi oppure asparagi oppure  
pomodoro oppure ...

# Usando queste regole, riusciamo ad analizzare il menu?



Pasto = Primo + Secondo + Dessert

Primo = Formato + Condimento

Secondo = Pietanza + Condimento

Dessert = tiramisù oppure pannacotta oppure frutta  
oppure ...

Formato = penne oppure fusilli oppure ...

Pietanza = pollo oppure cotoletta oppure ...

Condimento = funghi oppure asparagi oppure  
pomodoro oppure ...

# Usando queste regole, riusciamo ad **analizzare** il menu?

**Pasto = Primo + Secondo + Dessert**

Primo = Formato + Condimento

Secondo = Pietanza + Condimento

Dessert = tiramisù oppure panna cotta oppure frutta  
oppure ...

Formato = penne oppure fusilli oppure ...

Pietanza = pollo oppure cotoletta oppure ...

Condimento = funghi oppure asparagi oppure  
pomodoro oppure ...



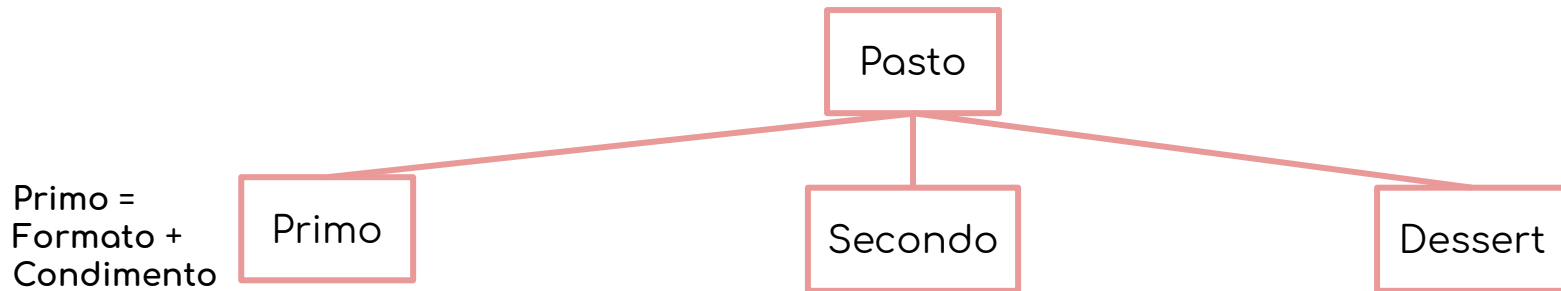
Ma possiamo fare di più: generiamo un nuovo Pasto

Pasto

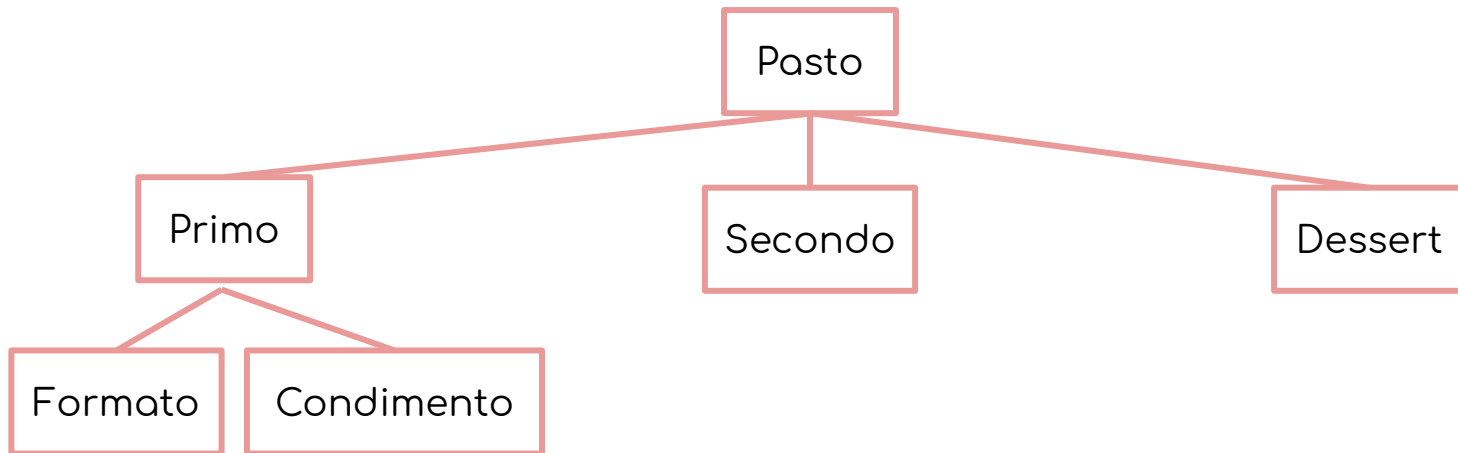
Pasto = Primo + Secondo + Dessert



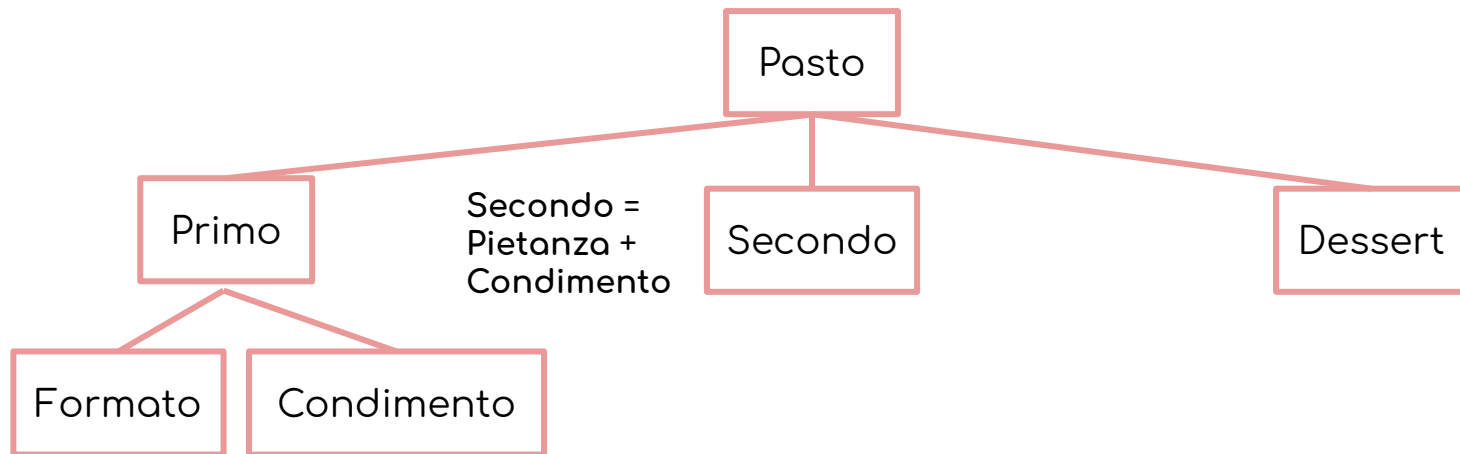
Ma possiamo fare di più: generiamo un nuovo Pasto



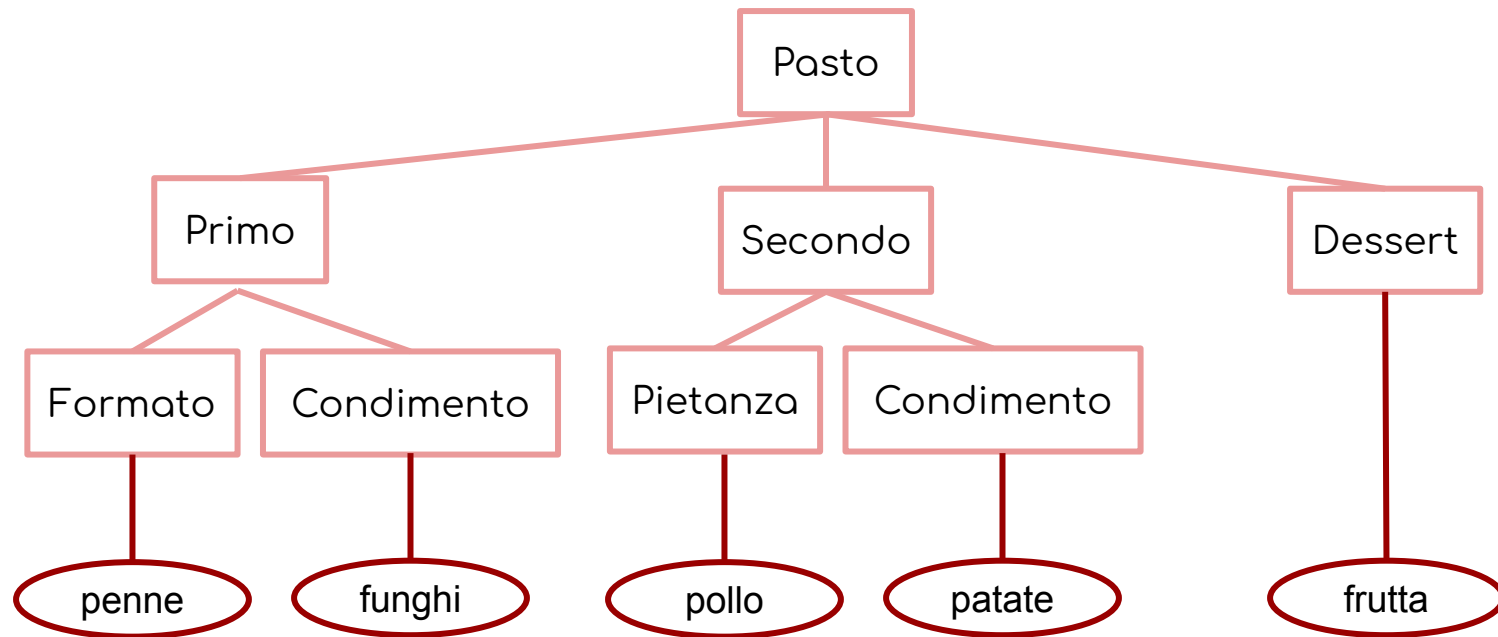
Ma possiamo fare di più: generiamo un nuovo Pasto



Ma possiamo fare di più: generiamo un nuovo Pasto

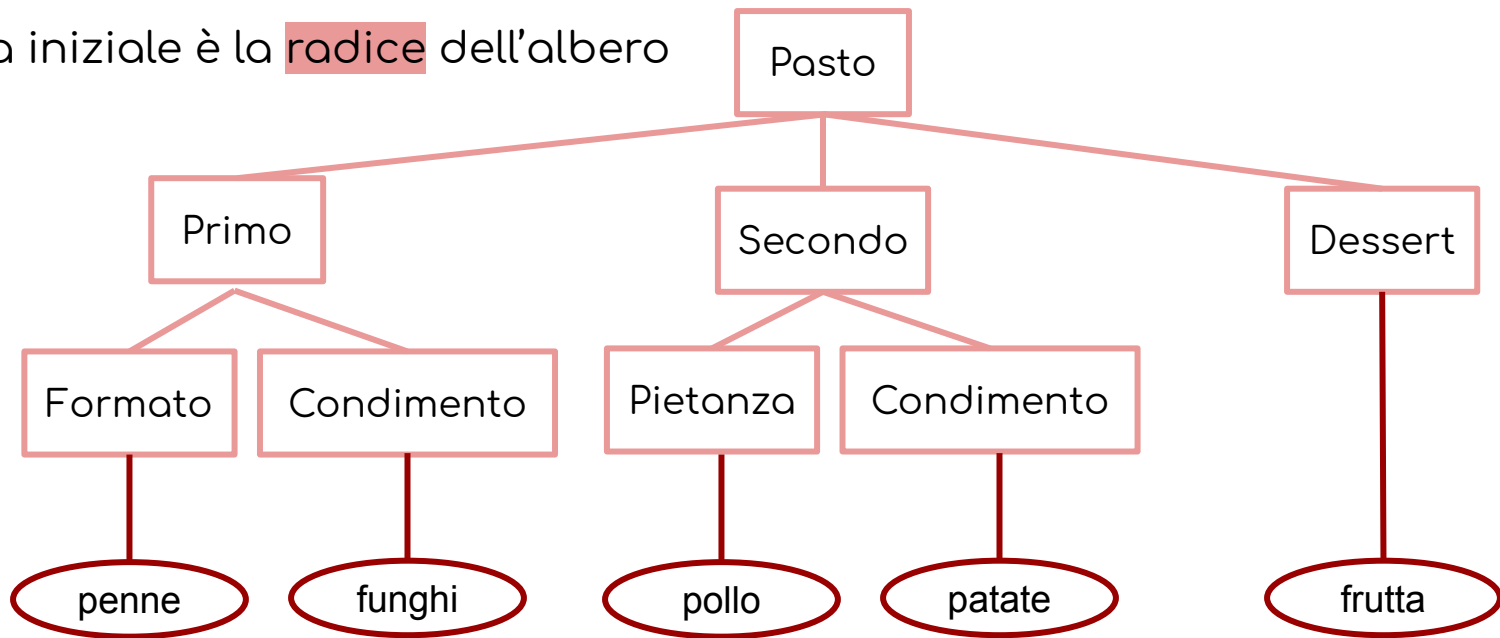


Ma possiamo fare di più: generiamo un nuovo Pasto



La struttura dati formata dall'espansione delle regole si chiama **albero**

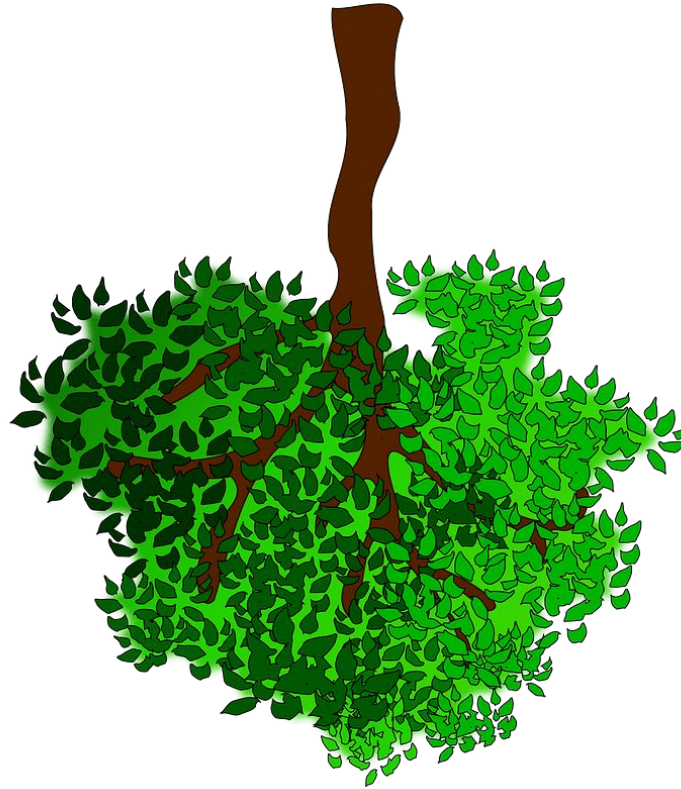
La categoria iniziale è la **radice** dell'albero



I simboli terminali sono le **foglie** dell'albero. Leggendoli uno dopo l'altro si ottiene il nuovo pasto.

La struttura dati formata dall'espansione delle regole si chiama **albero**

Gli informatici, stando tutto il giorno al computer e non uscendo mai di casa, si sono dimenticati che gli alberi veri crescono dal basso verso l'alto



# E nella lingua?

Luca mangia un panino

Luisa viaggia in treno

Marisa vince una medaglia

Luigi corre la maratona

Eva dorme sulla spiaggia

# Cerchiamo regolarità...

Luca mangia un panino

Luisa viaggia in treno

Marisa vince una medaglia

Luigi corre la maratona

Eva dorme sulla spiaggia

Frase  
=  
Soggetto + Verbo + Complemento



# Cerchiamo **regolarità**...

Luca mangia un panino

Luisa viaggia in treno

Marisa vince una medaglia

Luigi corre la maratona

Eva dorme sulla spiaggia

**Soggetto** = Luca oppure Luisa oppure  
Marisa oppure Luigi oppure Eva oppure ...

**Verbo** = mangia oppure viaggia oppure  
vince oppure corre oppure dorme  
oppure ...

**Complemento** = Complemento Oggetto  
oppure **Complemento Di Luogo** oppure ...

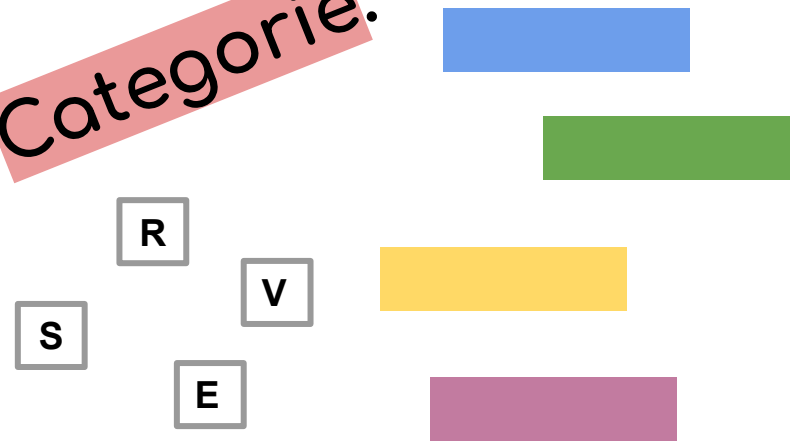
Ci sono tante grammatiche  
possibili!

Definiamo **categorie** e **regole**  
sulla base dei dati che  
abbiamo raccolto.

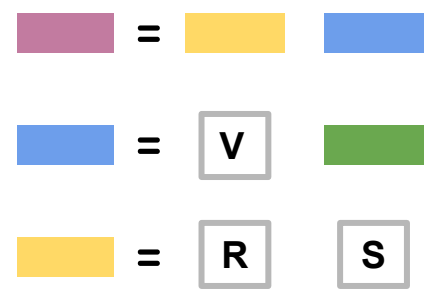




Categorie:

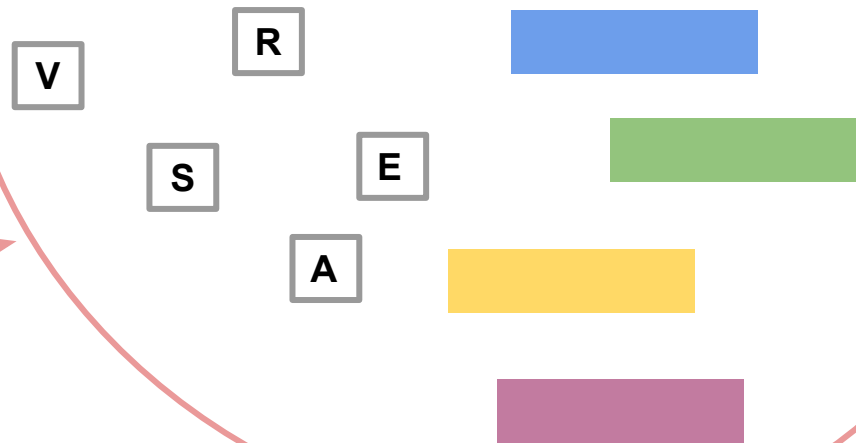


Regole:



e tante altre...

Alle **categorie** ci  
abbiamo pensato  
noi:

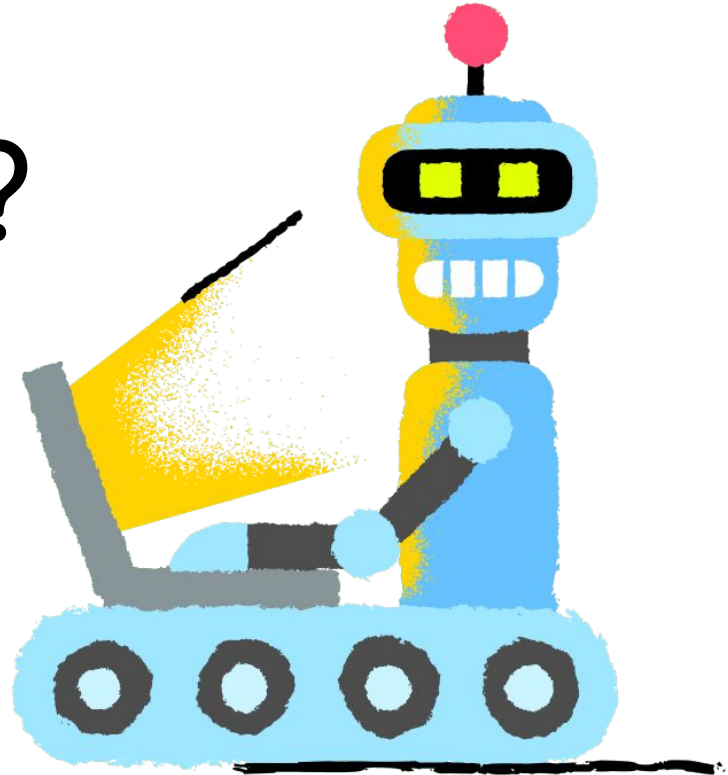


Alle **regole** dovete  
pensarci voi:

A diagram illustrating a rule. A red arrow points from the text 'Alle regole dovete pensarci voi:' to a dashed box. To the right of the dashed box is an equals sign, followed by another dashed box, a plus sign, and a third dashed box.

$$\boxed{\phantom{00}} = \boxed{\phantom{00}} + \boxed{\phantom{00}}$$

Vabbè,  
ma funziona?



Sì

I computer sono in grado di risolvere molti problemi linguistici, e lo fanno grazie a tecniche simili (ma un po' più avanzate!) a quelle che abbiamo adoperato.

e

No

Le applicazioni che conosciamo risolvono spesso un problema specifico, e lo fanno molto bene, ma si adattano difficilmente a risolvere problemi diversi.

La lingua è un oggetto **complesso**,  
e ci sono ancora tante domande da affrontare...

# E Siri?

Apple iPhone 11 (256GB) - Viola [More product info](#)

From APPLE



[View product on Amazon.it](#)

amazon.it Rating:

4.8



3,659 Ratings

708 Reviews

("Ratings" vs. "Reviews")



78% of potentially  
unnatural reviews removed  
(why?)

[View/Edit Adjustment](#)

[Jump to Report Card](#)



Adjusted Rating:

4.0



157 Reviews

[See Product Details](#)



Hi, how can I help?



a young boy is holding a  
baseball bat  
logprob: -7.61

quanti anni ha barack obama

[All](#) [News](#) [Images](#) [Videos](#) [Maps](#) [More](#)

[Settings](#) [Tools](#)

About 137,000 results (0.83 seconds)

Tip: Search for **English** results only. You can specify your search language in [Preferences](#)

Barack Obama / Età

59 anni

4 agosto 1961



Ricerche correlate



Donald Trump  
74 anni



Michelle  
Obama  
56 anni



Joe Biden  
78 anni

[Feedback](#)

## Barack Obama

44° presidente degli Stati Uniti

Barack Hussein Obama II è un politico statunitense, 44° presidente degli Stati Uniti d'America dal 2009 al 2017, prima persona di origini afroamericane a ricoprire tale carica. [Wikipedia](#)

**Vicepresidente:** Joe Biden (2009–2017) [In risalto](#)

**Nascita:** 4 agosto 1961 (età 59 anni), Kapi'olani Medical Center for Women & Children, Honolulu, Hawaii, Stati Uniti

**Altezza:** 1,85 m

**Coniuge:** Michelle Obama (s. 1992)

**Mandato presidenziale:** 20 gennaio 2009 – 20 gennaio 2017

**Figli:** Natasha Obama, Malia Ann Obama



**Donald J. Trump** ✓  
[@realDonaldTrump](#)

I will be speaking before a very important Pennsylvania State Senate hearing which is in process now. Talking about the massive voter fraud which took place in the 2020 Election!

[Traduci il Tweet](#)



This claim about election fraud is disputed

# E Siri?

I problemi che fanno uso di conoscenza linguistica sono vari, molto diversi tra loro, e sicuramente ben più difficili rispetto al riordinare una frase.

Qualsiasi applicazione ha però bisogno di un modello del linguaggio, ovvero un *modo* per poter decidere se una certa frase fa o meno parte della lingua.

Ed è questo il problema che abbiamo risolto oggi.





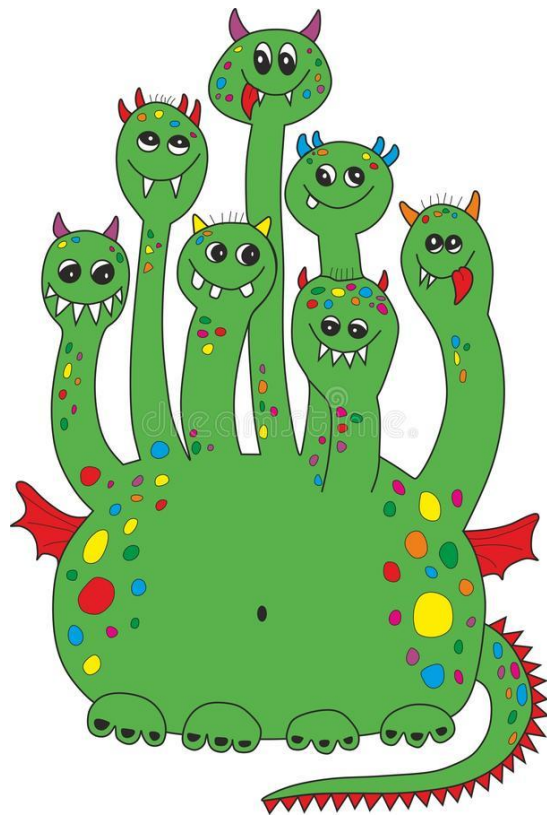
# Diventare un(a) Linguista Computazionale

Curiosità

Coding

Logica

Spirito di  
osservazione



Lettere

Filosofia

Informatica

Psicologia

Scienze cognitive

Dove?



Associazione Italiana di  
Linguistica Computazionale

# Link per curiosi/e....



CAISSA ITALIA EDITORE

LINGUISTICAMENTE