

Ciao a tutte e tutti!

Che bello scrivere questa lettera! Spero di scoprire tanto su di voi durante questi scambi di corrispondenza, e di aiutarvi a conoscere qualcosa del mondo della ricerca e dell'affascinante materia che studio.

Ma prima mi presento: mi chiamo Ludovica, ho 27 anni e sono una **linguista computazionale**.

Per capire come lo sono diventata, devo raccontarvi qualcosa in più di me: sono nata e cresciuta a Campobasso (eh già, esistiamo anche noi! 🇮🇹), ma dopo il liceo mi sono spostata a **Pisa** per studiare. Lì mi sono iscritta al corso di laurea in [Informatica Umanistica](#): ero indecisa, a scuola mi piacevano molto sia la **matematica** che le **lingue**, e Informatica Umanistica mi permetteva di continuare in un certo senso a studiare entrambe le cose.

Questa sono io, durante una vacanza in Bulgaria



La mia città,
Campobasso



Pisa, dove ho studiato e ora vivo con il mio compagno

Così ho scoperto la linguistica prima, e la linguistica computazionale subito dopo: in particolare, ho scoperto che **la grammatica di una lingua non è un libro pieno di regole** da imparare a memoria, ma un oggetto vivo e complesso, e che si può studiare con un **approccio scientifico**.

Il **dottorato di ricerca** è un percorso di studi che si intraprende dopo la laurea e che dura 3 o 4 anni. A differenza dei percorsi precedenti, durante il dottorato ogni studente o studentessa fa attivamente ricerca sotto la supervisione di un/una docente, scegliendo autonomamente un argomento su cui concentrarsi.

Questo argomento, di cui vi parlerò meglio dopo, e tutte le domande ad esso collegate, mi ha appassionato talmente tanto che ho deciso di continuare a studiarlo anche dopo la laurea: così oggi sono una **dottoranda** del [Centro Interdipartimentale Mente e Cervello](#) dell'Università di Trento.

In questa prima lettera ho pensato di descrivervi brevemente cos'è la linguistica computazionale: così, se vi interessa, potrò parlarvi del mio progetto di ricerca nelle prossime lettere e provare a rispondere alle vostre domande.

Probabilmente vi è capitato di interagire con un **assistente vocale**, per esempio Siri o Alexa, oppure di utilizzare un **traduttore automatico**: queste tecnologie sono possibili grazie all'esistenza di un **modello** della **lingua naturale**.



Le **lingue naturali**, a differenza di quelle artificiali, sono quelle parlate attivamente da una comunità di persone, e sono l'oggetto di studio della linguistica. Queste solitamente sono nate spontaneamente, e vengono principalmente usate per la comunicazione all'interno della comunità.



Un modello (disegno) del Sistema Solare

Potete pensare a un modello un po' come pensate al Sistema Solare: ci immaginiamo una grossa palla incandescente più o meno al centro (il Sole), e i pianeti che si muovono intorno, ognuno sulla propria orbita. Se guardiamo alle foto dell'universo, però, non somiglia affatto al nostro modello: da un lato ci sono **molti oggetti in più**, ma nel nostro modello di sistema solare abbiamo riportato solo quelli che ci interessano (il sole e i pianeti), dall'altro **mancano alcuni elementi** che invece abbiamo inserito nel nostro modello, ad esempio le orbite.

È semplice intuire perché disegniamo “meno” oggetti: perché quando costruiamo un modello vogliamo concentrarci su alcune cose e non altre, è un modo per **semplificare ed isolare** alcuni fenomeni. Quello che è più interessante, però, secondo me, è perché disegniamo le orbite. Sono oggetti di nostra “invenzione”, ma sono ciò che rende il nostro disegno del sistema solare un modello: ci aiutano a **spiegare** perché un pianeta si trova in una certa posizione e a **predire** dove andrà dopo.



Una foto dallo spazio, catturata dal telescopio Hubble

Le linguiste e i linguisti cercano di fare lo stesso per le lingue naturali: costruirne un modello, che spieghi come noi parlanti le usiamo. In questo senso, **la grammatica è un modello** di quello che noi parlanti, inconsciamente, sappiamo della nostra lingua madre. Come per il sistema solare, c'è qualcosa che decidiamo di non includere, perché complicherebbe troppo la nostra descrizione e sarebbe forse poco utile a studiare quello che ci interessa: per esempio, le frasi che troviamo nel libro di grammatica non somigliano a quelle che pronunciamo davvero nella vita di tutti i giorni, mentre dialoghiamo.

G: Allora metti la matita sulla partenza va ben ?
 F: [eh] , non ce l'ho la partenza !
 G: Allora okay , allora
 F: [eeh] dimmi , [me] mettila sul puntino nero , tu ce l'hai il puntino nero ?
 G: [eh] ? va ben , dal puntino nero allora
 F: [eh]
 G: allora scendi e passi in mezzo alle due macchine
 F: [SUSSURRANDO] il puntino nero scendi e passi in mezzo alle due macchine

Un esempio di un dialogo tra due ragazzi veneziani, che è stato trascritto da alcuni ricercatori per studiare cosa succede durante la conversazione

Come erano le orbite del sistema solare, anche per la lingua è utile introdurre dei concetti nuovi, perché ci aiutano a **descrivere quello che sappiamo o “vediamo”** della lingua, e quindi a scoprire nuove cose che non sarebbero evidenti altrimenti.

Per esempio, raggruppiamo i verbi italiani in tre gruppi (le tre coniugazioni), ma di queste solo la prima (-are) si dice **produttiva**, ovvero è la coniugazione che seguono tutti i nuovi verbi che vengono creati ogni giorno dai parlanti, per esempio adattando in italiano parole inglesi.

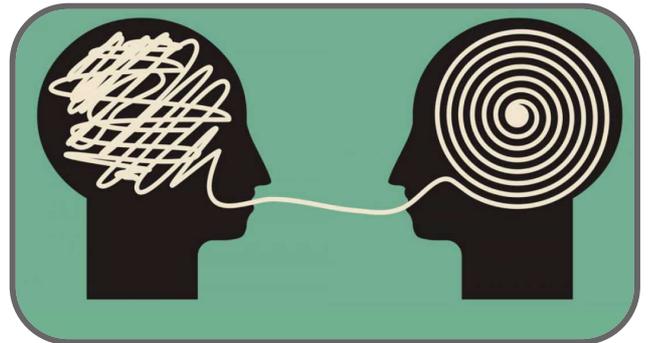
Raggruppare i verbi nelle tre coniugazioni, come siamo abituati a fare, ci aiuta così a spiegare perché *twittare*, *trollare*, *postare*, *googlare*, *tamponare*... siano tutti verbi di prima coniugazione, e anche a predire come si comporteranno i nuovi verbi che introdurremo nella lingua italiana.

Ci sono vari motivi per cui a volte è utile costruire questo modello con **strumenti computazionali**, ovvero usando un computer: come dicevo prima, questo modello può essere usato poi per costruire assistenti vocali o traduttori o motori di ricerca... tutte **applicazioni che simulano le nostre competenze linguistiche** e ci aiutano a svolgere alcuni compiti quotidiani, ma non solo. Avendo a disposizione un modello computazionale, possiamo facilmente simulare **scenari che sarebbe impossibile riprodurre nella realtà**, e porci domande a cui sarebbe difficile rispondere se dovessimo fare affidamento solo sui dati che possiamo raccogliere ed analizzare a mano.

Pensate ad esempio di voler studiare come cambia la lingua attraverso le **generazioni**: quanto diversamente parliamo rispetto ai nostri nonni e nonni, e quanto diversamente parleranno i e le nostre nipoti.

Sarebbe difficile raccogliere sufficienti dati reali, ma con delle simulazioni computazionali possiamo simulare centinaia di migliaia di parlanti e creare dati fittizi per **simulare il processo che vogliamo studiare**.

E sono proprio modelli del genere, che "si fingono" parlanti, ad essere usati da Siri per interagire con noi.



Spero di avervi incuriosito!
Non vedo l'ora di leggere la vostra risposta,
A presto,
Ludovica