





Text Enhancement, Text Mining, Text Analysis  
for Digital Humanities

## Programme

Pisa, Area della Ricerca CNR  
AULA 27  
Via G. Moruzzi, 1 - 56124 PISA  
(online: **Microsoft Teams**)  
Oct 28<sup>th</sup>, 2022, 11:00 - 11:40

### Introduction

*Marcello Ferro*, ILC-CNR (1')

### Studio dello spostamento semantico nell'evoluzione storica della lingua italiana condotto attraverso modelli di linguaggio neuronali

*Eva Sassolini*, Istituto di Linguistica Computazionale  
(ILC-CNR) (30')

### Discussion

(15')

## Studio dello spostamento semantico nell'evoluzione storica della lingua italiana condotto attraverso modelli di linguaggio neuronali

### *Abstract*

Il seminario illustrerà lo studio condotto per un progetto di tesi sperimentale, finalizzato alla creazione di un metodo computazionale in grado di trattare il cambiamento di significato in diacronia, basato sui modelli di linguaggio di tipo BERT (*Bidirectional Encoder Representations from Transformers*), con particolare riguardo al trattamento di varietà storiche della lingua italiana.

L'obiettivo dello studio riguarda l'ipotesi che il clustering, applicato agli embeddings contestuali prodotti da modelli BERT, sia lo strumento più adatto per uno studio di questo tipo. La recente letteratura di riferimento lo indica infatti come possibile proxy, ma non lo ha dimostrato definitivamente.

L'approccio sperimentale adottato è volto a comprendere quali sono le più recenti tecniche informatiche d'indagine, quali sono i modelli di riferimento in letteratura per questo compito, quali gli strumenti più efficaci, come si implementano e con quali risultati. La presentazione ripercorre le fasi di lavoro che hanno portato alla costruzione del metodo e tutte le sperimentazioni condotte, provando a rispondere a domande quali: come si configura l'operazione di clustering rispetto alla rappresentazione dei sensi di una parola? Ovvero, la distanza tra cluster indica necessariamente che le parole che appartengono ad uno, siano 'distanti' semanticamente da quelle di un altro?

Spostando poi l'attenzione dallo strumento di indagine all'oggetto della stessa, si è voluto analizzare cosa realmente aggregano questi strumenti. In particolare provare a comprendere quanto il contesto di una frase contribuisca all'interpretazione del significato delle parole contenute.