



Tipologia token-based: un laboratorio con “(mini)ciep+ - a sharable parallel corpus of prose”

L'analisi token-based è una metodologia di ricerca della linguistica tipologica che prevede l'utilizzo di istanze (token) di costruzioni, così come osservate nel reale utilizzo linguistico e approssimate attraverso risultati di esperimenti psicolinguistici, dati raccolti da questionari e corpora, meglio se paralleli.

Il laboratorio presenterà una selezione di dati da Miniciep+, la versione pubblica del *Corpus of Indo-European Prose and more*, un corpus parallelo di oltre trenta lingue sviluppato dal 2019 all'Università dello Saarland a partire dalle traduzioni di libri di narrativa. Grazie alla sua annotazione secondo il framework delle Universal Dependencies, il corpus si presta ad analisi quantitative di tipo morfologico e morfo-sintattico.

Durante il seminario si esploreranno una serie di script in Python ed R per l'estrazione dal corpus di dati relativi all'ordinamento di parola, concetti comparativi (ad es., quantificatori, aggettivi, ...), categorie verbali come la finitezza e pronomi personali.

About

Luigi Talamo è post doc all'Università dello Saarland, dove svolge attività di ricerca sulla linguistica tipologica con metodi quantitativi e computazionali. È parte del progetto 'Cross-linguistic Information-Theoretic Modelling of Communicative Efficiency' (PI: Annemarie Verkerk), finanziato dall'associazione tedesca per la ricerca (DFG) e volto ad investigare attraverso i metodi della teoria dell'informazione la variazione inter-linguistica, in particolare la struttura informativa e l'ordine delle parole. Si occupa anche dello sviluppo di corpora e treebank per lingue *low-resourced*, morfologia derivazionale e categorie miste.

08.04.2025 - h.15.30



https://site.unibo.it/laboratorio-sperimentale/it/agenda/seminario-mini_ciep

Dr. Luigi Talamo
Università dello Saarland

Studio 76, LILEC, via Cartoleria 5