

L'italiano generato dai *Large Language Model*: forma, frequenza e funzione delle interferenze con la lingua inglese

Anna-Maria De Cesare (TU Dresden)

Come è stato notato da più parti, ma finora solo in base a un numero piuttosto esiguo di testi, l'italiano generato dai *Large Language Model* (LLM) presenta numerose interferenze con l'inglese, spiegabili con il fatto che i testi scritti in inglese sono sovrarappresentati nei dati di addestramento. L'obiettivo dell'intervento consiste nel presentare una prima rassegna di interferenze riscontrate in diversi campioni di biografie generate in italiano da LLM multilingui (GPT-3, 4 e 4o). Ci soffermeremo sulla forma, frequenza e funzione delle interferenze di natura grafico-interpuntiva, lessicale e sintattica, mettendo anche in luce che gli anglicismi individuati non hanno tutti lo stesso statuto: alcuni sono documentati anche nella scrittura naturale, in particolare quella giornalistica, mentre altri si riscontrano solo o prevalentemente nell'output degli LLM.