

Acesta este un document demonstrativ, cu scopul de a ilustra funcționalitatea și capacitățile unui model lingvistic. Acesta este, în esență, o funcție matematică antrenată cu obiectivul să poată prezice un cuvânt pe baza cuvintelor din jur. Practic, la fiecare pas, modelul lingvistic asignează o probabilitate fiecărui cuvânt din dicționar. Cuvintele “potrivite”, fie că vorbim din punct de vedere tipografic, sintactic ori semantic, vor avea probabilități mari, în timp ce cuvintele nepotrivite vor avea probabilități foarte mici. Astfel, comparând un text dat bucată cu bucată, pot fi identificate eventuale nepotriviri: acelea sunt locuri în care ar trebui să începem să căutăm posibile greșeli. Mai jos ilustrăm câteva situații:

Scenariul greșelii de tipar:

Folosim un text generic, în care introducem greșeli, inclusiv nefolosirea diacriticelor: Primul Război mondial a fost un război global declanșat în Europa, la care au participat peste 70 de milioane de militari, inclusiv 60 de milioane de europeni, mobilizați într-unul dintre cele mai mari războaie din istorie.

Scenariul greșelii sintactice

În timp ce primul scenariu este rezolvat și cu un simplu dicționar de verificări, un caz mai subtil îl reprezintă greșelile sintactice, de exemplu, un dezacord. Cei doi prieteni buni merge spre școală. Puțin mai subtil, luăm alt caz. Nu mai știu cât timp a trecut dar Anca și Andrei a fost mereu în gândul meu.

Scenariul greșelii semantice

Cea mai dificilă eroare, pe care am zice că doar un om o poate repera, presupune înțelegerea textului și folosirea cunoștințelor exterioare textului. Primul Război Mondial a început în anul 1914 și a fost un conflict extrem de sângeros. Ziua Națională a României este sărbătorită în dată de 7 August. Toate acestea, dar ce se întâmplă când efectiv scriitorul nu reușește să se exprime fluent? Ce avem este că s-a produs o extorsiune a unui degajament ceea ce face ca conflagrația să se propage recursiv, în paralel cu curgerea temporală.