

Rezumat

Coroborat cu popularitatea în creștere a tehnologiilor avansate de procesare a limbajului, necesitatea unei înțelegeri și evaluări automate a evoluției limbii și limbajului a crescut semnificativ începând cu a doua jumătate a secolului al XIX-lea. Această cerință este strâns legată de activitatea de cercetare în lingvistica computațională și de activitățile didactice în domeniul tehnologiilor de procesare a limbajului natural (TLN) și al învățării colaborative. Evaluările efectuate în cadrul acestui demers au subliniat imperativul unei înțelegeri profunde a textului, chiar dacă analizele superficiale sunt, în mod inerent, mai facile de realizat.

În acest context, lucrarea de față introduce și rafinează metode de detecție a formelor specifice de limbaj în texte scurte, cu un accent deosebit pe îmbunătățirea modelelor de analiză a sentimentului. Această arie tematică, abordată și în cadrul unor proiecte notabile ale universităților românești (e.g., produsul software pentru analiza sentimentelor din textele în limba Română al Universității “Babeș-Bolyai” din Cluj-Napoca - SENTITEXT), se inspiră din studiile fundamentale ale lui Lyons (1981) și Langacker (1985) privind subiectivismul lingvistic. Abordarea propusă permite procesarea unitară atât a textelor generale, cât și a conversațiilor online, facilitând integrarea activităților specifice de înțelegere și colaborare într-o platformă unică, precum cea a Laboratorului de Prelucrare a Limbajului Natural la Institutul de Informatică Teoretică al Academiei Române, Filiala Iași (LSPLR@ARFI-IIT).

Numeroase instrumente de procesare a textului scris prezentate au fost dezvoltate prin colaborări fructuoase cu Institutul de Cercetări pentru Inteligență Artificială “Mihai Drăgănescu” (ICIA), exemplificate de platforma <https://relate.racai.ro/>, și cu Facultatea de Informatică a Universității “Alexandru Ioan Cuza” din Iași (UAIC), vizibile pe platforma <https://nlptools.info.uaic.ro>.

1 Contribuții științifice și experiență academică

Prezenta teză de abilitare sintetizează realizările mele științifice, profesionale și academice, obținute în cadrul unui grup de cercetare extins, caracterizat mai ales de o abordare interdisciplinară. Rezultatele acumulate acoperă domenii variate, incluzând informatica, filosofia, lingvistica computațională, psihologia mulțimii și jurnalismul.

Parcursul meu educațional și profesional a început cu o licență în fizică în 1997, urmată de un masterat în comunicare și relații publice la Școala Națională de Studii Politice și Administrative (SNSPA). Dețin un dublu doctorat cu distincția *Magna Cum Laude* în Filosofie (UAIC, 2010), respectiv Informatică (în cotutelă UAIC – Academia Română, 2016). Prima teză de doctorat a fost

PROCESARE AVANSATĂ A TEXTELOR: DE LA DIACRONIE LA DETECȚIA SEMNALELOR LINGVISTICE ÎN TEXTELE SCURTE

publicată sub forma cărții “Temeliile Turnului Babel. O perspectivă integratoare asupra discursului politic” (Editura Academiei Române, 2013), iar a doua teză a generat numeroase lucrări științifice apreciate în mediul academic.

Până în prezent, am acumulat o semnificativă experiență în proiecte de cercetare naționale și internaționale (e.g., POSDRU, ERASMUS+, COST Action, POCA), concretizată în peste 230 de publicații. Acestea includ articole în cadrul unor conferințe de top (JCDL, ISD, TPDF, FUZZ-IEEE, SemEval-NAACL, SemEval-ACL), conferințe internaționale renumite în domeniul prelucrării limbajului natural (LREC, CICLing, KES) și articole în reviste prestigioase (Future Generation Computer Systems, Semantic Web, Revue Roumaine de Linguistique). Cumulat, aceste lucrări mi-au adus peste 1300 de citări.

În prezent, dețin funcția de cercetător științific gr. II și conferențiar asociat la UAIC și UMFST, unde sunt responsabilă pentru cursuri precum “Semantica și Pragmatica Limbajului Natural”, “Inteligență Artificială” și “Tehnologii Afective și de Persuasiune”. Din martie 2025, sunt și membru asociat al Academiei Oamenilor de Știință din România.

Pe lângă competențele mele în prelucrarea limbajului natural, analiza de text și dezvoltarea de corpusuri, dețin multiple certificări profesionale (e.g., formator, antreprenor) și o experiență semnificativă în proiecte strategice bazate pe fonduri nerambursabile de la Uniunea Europeană, SWISS National Science Foundation sau MEC. În anul 2024, am primit distincția “Premiul de Excelență” la Zilele Academice Iașene, ediția a 39-a, pentru activitatea de cercetare științifică în domeniul Științe Exacte la Institutul de Informatică Teoretică al Academiei Române, Filiala Iași. Această recunoaștere a fost precedată de alte premii pentru activitatea jurnalistică (Uniunea Ziariștilor Profesioniști din România, 2015, 2016) și pentru activitatea culturală (Societatea DORUL din New York, 2024), care au fundamentat noi colaborări de lungă durată în cercetarea interdisciplinară, în special în analiza presei scrise, domeniu în care am început să activez la sfârșitul anilor '90.

2 Direcții de cercetare și inovație

Specificitatea analizelor efectuate, raportate la procesul de detecție a formelor specifice de limbaj în texte scurte, s-a concentrat pe dezvoltarea unor *modele multidimensionale de complexitate textuală*. Acestea integrează metrice de evaluare a performanței pentru modelele de clasificare și/sau de suprafață, indici specifici la nivel lexical, semantic, pragmatic și structural al textului. Complementar, detecția formelor specifice de limbaj în texte scurte se axează pe evaluarea figurilor de stil (sarcasm, umor, ironie etc.) și pe analiza sentimentelor, având la bază două modele computaționale: un model centrat pe analiza opiniilor, derivat din analiza de sentimente, și un

PROCESARE AVANSATĂ A TEXTELOR: DE LA DIACRONIE LA DETECȚIA SEMNALELOR LINGVISTICE ÎN TEXTELE SCURTE

model multidimensional derivat din voce, text și imagine, dezvoltat în colaborare cu specialiști în procesarea imaginilor și a vocii.

Abordarea propusă integrează tehnici avansate de prelucrare a limbajului natural și se concentrează pe furnizarea de estimări atât calitative, cât și cantitative ale procesului de evaluare. Teza prezintă diverse cazuri de utilizare, argumentând că cele trei platforme, la a căror dezvoltare am contribuit (v. corpusul [RODICA](#) (*ROmanian DIachronic Corpus with Annotations*)), au fost supuse unor validări intense, demonstrând capacitatea lor de a reproduce rezultatele experților umani. Pornind de la analizele efectuate, am reușit să extindem perspectiva colaborării în ceea ce privește realizarea unei reprezentări multidimensionale a textului, în special pentru textele din rețelele de socializare. Astfel, unul dintre cele mai importante obiective ale modelelor noastre este de a spori înțelegerea ca detector de forme specifice de limbaj în texte scurte prin furnizarea de *output*-uri specifice și interpretabile.

3 Avantaje și perspective viitoare

Unul dintre principalele avantaje ale prezentei cercetări este *modularitatea* arhitecturii, care permite adăugarea constantă de noi componente pentru procesarea textului, imaginii și vocii, îmbunătățind continuu calitatea generală a sistemului. Un alt atu major este *portabilitatea*, având în vedere că soluțiile bazate pe agenți inteligenți sunt accesibile pe o multitudine de dispozitive IoT și smartphone-uri, oricând și oriunde. Remarc deschiderea companiilor IT din Iași, Constanța, Cluj-Napoca etc. de a se implica activ în astfel de proiecte, fiind interesate de dezvoltarea competențelor studenților și cercetătorilor noștri, în vederea oferirii de oportunități de stagiu în laboratoarele lor de cercetare. Am convingerea că soluțiile automate vor fi folosite din ce în ce mai intens în cadrul proceselor de dezvoltare de software. Prioritățile mele constau în menținerea și extinderea rețelei noastre de colaborări, atât cu mediul academic, cât și cu mediul privat, dezvoltând totodată în continuare arhitectura noastră modulară. Mai mult decât atât, sper și îmi doresc ca parcursul actual de dezvoltare a carierei, precum și perspectivele de cercetare curente să creeze premisele pentru noi proiecte, colaborări și publicații de impact.