

**Prof. em. dr. D. Cristea – membru corespondent al Academiei Române
Secția de Știința și Tehnologia Informației**



Născut la 16 decembrie 1951, în Moinești, jud. Bacău.

1. Studii, diplome

- Școala primară și gimnazială în Moinești; primii 2 ani de liceu (1966 - 1968) la Liceul Teoretic din Moinești; ultimii 2 ani de liceu (1968 - 1970) la Liceul „I.L.Caragiale” din Ploiești.
- Diploma de absolvent în specialitatea Calculatoare, la Facultatea de Automatică și Calculatoare, Universitatea Politehnică București, în 1975.
- Diploma de absolvent al Facultății de Matematică, la Universitatea “Alexandru Ioan Cuza” (UAIC) din Iași, în 1981.
- Diploma de doctor inginer în specialitatea Calculatoare din partea Universității Politehnice București, Facultatea de Automatică și Calculatoare, în 1994.

2. Cariera profesională

- 1975-1976 – inginer stagiar la I.I.S. Țesătura Iași
- 1976-1989 – inginer de sistem la Centrul de Calcul al UAIC
- 1989-1995 – lector la Fac. de Matematică, secția Informatică a UAIC (din 1992 – devenită Facultatea de Informatică din Iași – FII)
- 1995-2002 – conferențiar la UAIC-FII
- 2000-2001 – director al firmei germano-române Langware, Iași
- 2000-2003 – decan al UAIC-FII
- 2000-2012 – membru în Senatul UAIC
- 2001-prezent – conducător de doctorat în Informatică la UAIC-FII
- iunie 2002-2008 – cercetător principal II (cu jumătate de normă), la Institutul de Informatică Teoretică din cadrul Filialei Iași a Academiei Române (ARFI-IIT)
- sept. 2002-2020 – profesor la UAIC-FII
- 2004-2011, oct. 2012 - oct. 2016 – președintele Comisiei de Informatică a CNATDCU
- 2011-2012, iul. 2020 - prezent – membru al Comisiei de Informatică a CNATDCU
- 2004-2012 – prodecan al UAIC-FII

- 2004-2012 – Director executiv al Școlii Doctorale din UAIC-FII
- 2008-2016 – Director al Departamentului de Cercetare din UAIC-FII
- 2008–oct. 2020 și apoi din ianuarie 2021 – cercetător principal I (cu jumătate de normă), la ARFI-IIT
- mai 2013 – membru în panelul de experți UEFISCDI pe domeniul Sisteme Inteligente
- 2017-2020 – Director al Școlii Doctorale din UAIC-FII
- 2018-2020 – Director interimar al CSUD-UAIC
- din 2019 – evaluator ARACIS în domeniul Informatică – școli doctorale
- 1 oct. 2020 – ieșirea la pensie de la UAIC

3. Stagii de cercetare, cursuri și conferințe invitate

Am efectuat stagii de cercetare în mai multe universități și institute de cercetare, printre care:

- La Universitatea Edinburgh și la LIMSI, Universitatea Paris-Sud, Orsay (1993) – unde am dezvoltat sistemele MICH & MULTI-MORPH (implementate în Lisp), modele de morfologie prin clasificare. Sistemul Multi-Morph era capabil să facă generare morfologică în mai multe limbi: spaniolă, română, rusă și franceză.
- În Laboratorul de Lingvistică Computațională al Universității din Veneția (1994, 1995) – unde am dezvoltat, împreună cu 3 colaboratori studenți, sistemul PROSODICS, un pachet de prelucrare a semnalului vocal.
- În IRST, *Istituto di Ricerca Scientifica e Tecnologica* din Trento, și la *University of Sains Malaysia*, Penang (1996, 2004), unde am pus bazele și apoi am rafinat cercetarea asupra structurii discursului, finalizată cu elaborarea Teoriei Nervurilor.

Am prezentat prelegeri și tutoriale invitate, la mai multe universități din țară și străinătate, ori în cadrul unor manifestări științifice; o parte din ele sunt listate mai jos:

- Tutoriale în cadrul Școlilor Internaționale EUROLAN (1993, 1995, 1997 etc.);
- Tutorial în cadrul *International Summer School in Computational Linguistics*, Tzikov Chark, Bulgaria (1992).
- *Université de Paris-Sud*, LIMSI-Orsay, France (1993).
- *Max Planck Institute for Psycholinguistics*, Nijmegen, Germany (1993).
- *Summer school on Contemporary Topics on Natural Language Processing*, Tuzlata, Bulgaria (1994).
- Seminar la Universitatea Ca Foscari din Veneția, Laboratorul de Lingvistică Computațională (1995).
- *University of Sains Malaysia*, Penang (1996 - o serie de 3 conferințe).
- *Loria-CNRS*, Nancy, France (1997).
- *University of Toronto*, Canada (1998).
- Universitatea din București (1999).
- *TALN-2000: 7e conférence annuelle sur le Traitement Automatique des Langues Naturelles*, Lausanne, France (2000).
- *University of Wolverhampton, School of Humanities, Languages and Social Sciences*, UK (2000).
- *University of Sheffield, School of Humanities, Languages and Social Sciences*, UK (2000).
- În Consorțiul pentru Tehnologia Limbii Române – ConsILR, întâlnirea de înființare a Consorțiului, Academie de Științe, Chișinău, Republica Moldova (2000).
- *University of Hamburg*, Germany (2002).

- *The International Conference Recent Developments in Natural Language Processing*, Borovets, Bulgaria (2003).
- *Istituto Trentino di Cultura*, Trento, Italy (2004).
- La ICON, Hyderabad, India (2004).
- *The LT4eL general project meeting*, Prague (2006).
- *National University of Singapore, School of Computing*, Singapore (2006, o serie de 2 conferințe).
- *Université de Marseille*, France (2007).
- La *The Spatial Cognition, Spatial Language and the Balkan Spatial Lexicon meeting*, Brussels (2009).
- În *LT Days*, Luxembourg (2009).
- La *Cicling 2010*, Iași (2010).
- În cadrul *The Transition Symposium from Cicling to Promise*, Iași (2010).
- În *The NLP Seminar, University of Sussex*, UK (2011).
- La *International Workshop on Intelligent Information Systems*, Chișinău (2011).
- La *The Crosslingual Language Technology Conference*, Univ. Hamburg, Germany (2012).
- Universitatea din Ternopil, Ucraina (2014, o serie de 2 conferințe).
- La *EneL meeting*, Bled, Slovenia (2014).
- Universitatea de Vest Timișoara (2022).

4. Modele teoretice și dezvoltări tehnologice

Preocupările predominante în activitatea mea de cercetare au fost legate de domeniul Prelucrării Limbajului Natural (PLN), cu un interes special spre dezvoltarea de modele de interpretare a limbajului, dezvoltarea de resurse lingvistice și de instrumente de prelucrare a limbajului, cu precădere cel scris, testarea modelelor, integrarea instrumentelor în aplicații și aprofundarea modelelor aflate în dezvoltare. La dezvoltarea modelelor, de cele mai multe ori am lucrat cu studenți de la toate nivelurile universitare, cu scopul dublu, de a-i pregăti ca cercetători și pentru sofisticarea și verificarea în practică a modelelor, pe datele de limbă pregătite, în general, tot de ei. Am convingerea fermă că un cercetător care este și cadru didactic nu are voie să ignore niciunul dintre aceste aspecte, ori altfel contribuția sa este imperfectă. Pe de altă parte, cred că actul predării trebuie permanent împrăștiat din cunoașterea generată de propria cercetare.

Descriu mai jos o seamă de modele și sisteme la a căror dezvoltare am avut contribuția esențială sau egală cu a colaboratorilor.

Modele și sisteme simbolice de întrebare-răspuns (colaboratori: Dan Tufiș, Alex Moruz, Adrian Iftene):

- **IURES** (1985: *Înțeleg Ușor Românește Eliminând Sintaxa*) – un sistem care comunică cu un interlocutor uman pe baza cunoștințelor reprezentate într-o rețea semantică. Noțiunea de rețea semantică a fost extinsă prin introducerea conceptelor sensibile la context, modelate ca noduri ale rețelei ce au atașate proceduri activabile automat la traversarea nodurilor de către procesul de căutare. Au fost de asemenea introduși operatorii logico-lingvistici și au fost dezvoltate metode originale de navigare în rețele semantice cu moștenire multiplă. Sistemul avea implementate rudimente de tratare a elipselor și de rezolvare a referințelor anaforice, iar răspunsul era generat tot în limbaj natural prin fraze construite ad-hoc.

- **QUERNAL I și QUERNAL II** – două sisteme de întrebare-răspuns care realizau comunicarea cu informațiile depozitate în baze de date, realizate între 1985 și 1989.

Întrebările erau prelucrate pe mai multe niveluri (lexical, sintactic, semantic) până la transformarea lor în expresii ale limbajului QRL (*Query Representation Language*), de asemenea propus și descris la nivel sintactic și semantic. Odată forma QRL a întrebării generată, ea era transformată de un traducător într-o exprimare a unui limbaj de interogare a bazelor de date (DTR, pe vremea aceea, ulterior SQL). Lansarea aplicației de gestionare a bazei de date asupra întrebării primite astfel în intrare, genera apoi răspunsul.

- Participări în concursuri de sisteme de întrebare-răspuns, la QA@CLEF, AVE, RTE, QA4MRE între anii 2006 și 2009, cu câștigarea de 10 ori a locului 1, o dată a locului 2 și de 4 ori la locului 3 în diverse exerciții.

Modele de reprezentare și procesare a discursului (colaboratori: Nancy Ide, Laurent Romary, Daniel Marcu, Valentin Tablan, Ionuț Pistol, Gabriela E.Dima, Oana Postolache, Eugen Ignat, Daniel Anechitei)

- **Teoria Nervurilor** (1998-2015). Cunoscută în literatură drept *Veins Theory* (VT), teoria a fost elaborată într-o seamă de lucrări, începând cu anul 1998. Ideea că structura discursului influențează referențialitatea era cunoscută într-o seamă de teorii mai vechi, precum *Attentional State Theory* (Grosz and Sidner). VT duce mai departe acest concept, propunând bazele unui calcul al unui domeniu de accesibilitate referențială, în care nuclearitatea nodurilor din arborele de structură are importanța majoră, în timp ce tipul relațiilor retorice sunt ignorate. Descifrarea domeniilor de accesibilitate induce însă și concluzii relative la coerența discursului. Tranzițiile teoriei *Centering* (Grosz, Joshi, Weinstein), limitate acolo doar la segmentele de discurs, pot fi acum extinse la nivelul întregului discurs. Ideea este că, într-un text coerent, segmentele discursului se întrepătrund, ceea ce face posibilă câte o „citire” a discursului din perspectiva fiecărei unități de discurs (în esență, egală cu o propoziție). VT a fost validată experimental pe corpusuri adnotate manual la structuri retorice, regularitățile pe care le pune în evidență fiind relevante statistic. Pe baza teoriei a fost construit un parser de discurs.

- modele și sisteme de **rezoluția anaforelor** (2001-20090): RARE (*Robust Anaphora Resolution System*) a fost sistemul realizat pe baza modelului teoretic propus.

- modele și sisteme de **rezumare automată** (2003-2013): modelul de rezumare propus extrage din textul original secvențe textuale care se combină într-un text mai scurt, dar coerent și care exprimă esența; la baza modelului au stat domeniile de accesibilitate ale VT; sistemul a fost implementat și exploatat în proiectul internațional ATLAS pentru realizarea de rezumatoare pentru 6 limbi.

Modele de configurare de aplicații și compoziții de servicii web (colaboratori: Ionuț Pistol, Paul Diac)

Ingenieria software, inclusiv cea dedicată domeniului PLN, utilizează de mulți ani module interschimbabile, interconectate în arhitecturi de procesare complexe, în care fiecare modul este descris printr-o intrare, o ieșire și accesul la anumite resurse, ceea ce face ca pentru designerul de sistem, funcționarea reală a unui modul să poate fi ascunsă (într-o cutie neagră). Începând din 2004 am lucrat la dezvoltarea și apoi rafinarea unui model de descriere ierarhică a schemelor de adnotare (seturi de etichete, atributele lor și restricții), ca noduri ale unui graf orientat aciclic. Ulterior modelul a fost augmentat cu procese, acestea fiind așezate pe arcele grafului. Abordarea face posibilă mixarea mai multor adnotări asupra fișierelor, simplificarea adnotărilor și calculul automat al traseelor de prelucrare (*workflows and services*). Două teze de doctorat au dezvoltat acest model: sistemul ALPE (Ionuț Pistol, 2014) și o descriere de algoritmi de compunere de servicii web, însoțiți de studii de complexitate (Paul Diac, 2020).

Realizări de lexicografie computațională (colaboratori principali: Nicolae Curteanu, Alex Moruz, Marius Răschip, Cristian Pădurariu, Petre Rebeja, Gabriela Haja, Marius Popescu):

- **eDTLR** (Dicționarul Tezaur al Limbi Române în format electronic): modele de descifrare a structurii dicționarului în vederea transformării acestora în resurse electronice interogabile multicriteriale (a se vedea în secțiunea Proiecte);

- **DeLORo** (*Deep Learning for Old Romanian*) – modele de interpretare a textelor vechi românești în caractere chirilice (a se vedea în secțiunea Proiecte).

Modele de descifrare a conținutului textelor (colaboratori principali: Ionuț Pistol, Daniela Gîfu, Andreea Macovei, Șerban Boghiu)

- **MappingBooks** – un model de interacționare a cititorului cu texte de geografie didactică sau ghiduri de călătorie, prin recunoașterea entităților geografice, a relațiilor spațiale menționate în text și apelarea la tehnici de realitate virtuală; modelul a fost urmat de crearea unei aplicații cu un înalt impact în educație (a se vedea în secțiunea Proiecte).

- **Modelul Câmpurilor Temporale** (*Time Yards Model*) – un model de reconstrucție a traseelor temporale ale personajelor care apar într-un roman de ficțiune. Segmentele temporale sunt întinderi de text contigue în care apar aceleași personaje, timpul este staționar sau în creștere liniară, iar istoria este depășită de același autor. Identificarea lor poate fi făcută cu modele neuronale, iar când ordinea relativă a acestora este refăcută, traseele românești ale oricărui personaj pot fi identificate, indiferent de idiosincraziile în redare ale autorului.

5. Participarea în proiecte finanțate din surse europene sau naționale

O listă completă a proiectelor finanțate în care am participat poate fi găsită la <http://profs.info.uaic.ro/~dcristea/research.html>. În cele ce urmează rezum doar câteva dintre cele în care am fost coordonator sau director de proiect.

- **BALKANET** – rețeaua balcanică de wordneturi (2001-2004): proiect în care s-a realizat o rețea de wordnet-uri ale limbilor din Balcani: bulgara, greaca, româna, sârba, turca, aliniate wordnet-ului englezesc, folosit drept coloană vertebrală pentru toate limbile din proiect. Ulterior, Ro-WN a continuat să fie dezvoltat la AR-ICIA.

- **LT4eL** – tehnologii lingvistice pentru eLearning (2005 – 2008): proiectul a deschis căi noi de abordare a problematicii eLearning, pentru recuperarea automată a conținutului documentelor textuale, prin angajarea de resurse și tehnologii lingvistice. Limbile abordate în cadrul proiectului au fost: engleza, germana, polona, româna, ceha și bulgara. Cunoștințe de natură semantică au fost integrate sub formă de ontologii.

- **RoLTech** – *Romanian Language Technology: Resources, Tools And Interfaces* (2006 – 2008): proiectul a urmărit dezvoltarea unei platforme web pentru depozitarea de resurse și tehnologii ale limbii române: dicționare, tezaure, corpusuri, modele statistice și simbolice, instrumente de prelucrare a limbii române, precum și o listă de documente, trimiteri la resurse și instrumente externe, proiecte semnificative și evenimente științifice dedicate domeniului PLN, titluri de cărți, colecții de lucrări privind limba română, evenimente științifice semnificative etc. Au fost dezvoltate 3 tipuri de aplicații: o interfață de însușire a morfologiei românești; o interfață care exploatează variațiile morfologice ale cuvintelor, pentru maximizarea rezultatelor motoarelor de căutare, însoțită de un corector ortografic; un serviciu de găsire a ocurențelor în colecții mari de texte, de ajutor pentru lexicografi în activitatea lor de creare de dicționare și lexicoane.

- **eDTLR** - Dicționarul Tezaur al Limbii Române în format electronic (2007 - 2010): proiectul, realizat în colaborare de 7 institute din țară, a dus la obținerea primei variante

electronice a marelui dicționar al limbii române, elaborat de Academia Română pe parcursul unui secol. Modelul de parsare a formatului tipărit, elaborat în cadrul proiectului, precum și structura de bază de date a dicționarului este folosit astăzi la elaborarea noii versiuni a Dicționarului, care se dezvoltă în cele 3 institute de lingvistică din țară integral electronic.

- **CLARIN** *Common Language Research Infrastructure* (2008 - 2010): unul dintre cele mai vizibile proiecte europene din domeniul PLN în secolul nostru. Obiectivele lui au fost să creeze o infrastructură de cercetare europeană care să permită integrarea și interoperabilitatea resurselor lingvistice și a tehnologiilor domeniului PLN. Resursele și serviciile, oferite prin intermediul unei rețele de centre specializate, sunt bazate pe tehnologii ale web-ului semantic, capabile să facă față deosebirilor de format, structură ori terminologie. Infrastructura creată în acei ani este deosebit de activă și deschisă, astfel încât noi resurse și servicii pot fi adăugate cu ușurință comunității CLARIN. Platformele interactive inteligente create de CLARIN facilitează cercetătorilor din științele sociale și umaniste accesul la resursele lingvistice și instrumentele de prelucrare a limbajului.

- **ALEAR** *Artificial Language Evolution on Autonomous Robots* (2008 – 2010): experimente atent controlate au arătat că agenți inteligenți își pot auto-organiza sisteme de comunicare cu caracteristici similare celor caracteristice limbilor umane. Colectivul condus de mine s-a ocupat de propunerea de modele care să explice dezvoltarea spontană, în comunități de agenți inteligenți, a unor elemente ale discursului, în esență, apariția referențelor pronominali și coerența discursului. Modelele noastre au dovedit că pronumele poate fi inventat spontan în urma unei serii de interacțiuni între agenți, însă nu același lucru se întâmplă în cazul coerenței discursului, unde abilitatea de a produce discursuri coerente nu reprezintă o adaptare căpătată în urma unui șir de experimente de comunicare între agenți, ci reprezintă manifestarea incorporării unui aparat cognitiv al memoriei imediate (asemănător stivei) și a unui principiu lingvistic minimalist, care favorizează economia în exprimare.

- **ATLAS** *Applied Technology for Language-Aided Content Management Systems* (2010 – 2013): proiectul s-a ocupat de crearea unor tehnologii lingvistice care să ajute persoanele care administrează și publică pe web conținut multilingv. Cu componentele dezvoltate în proiect, i-Librarian, i-Publisher și EUDocLib, utilizatorii pot crea cu ușurință, gestiona și publica conținut multilingv, precum și avea un acces ușor și intuitiv la o vastă colecție de documente ale legislației UE. Platforma ATLAS clasifică, rezumă, adnotează și traduce documente, indiferent de format și de limbă (tehnologiile dezvoltate în proiect au fost pentru limbile bulgară, engleză, germană, greacă, poloneză și română).

- **METANET4U** Resurse și instrumente de PLN (2010 – 2012): deși mult asemănător proiectului CLARIN, METANET4U a fost mai pragmatic în abordare, în cadrul lui fiind elaborată rețeaua de portale META-SHARE, pe care s-au stocat resursele și instrumentele elaborate în consorțiu.

- **DRuKoLa** *Comparing German and Romanian languages through corpus linguistics technologies* (2013-2017): interfața de interogare a corpusurilor KorAP, dezvoltată la Institut für Deutsche Sprache, Mannheim, a servit ca infrastructură tehnologică pentru implementarea Corpusului Limbii Române Contemporane (CoRoLa), apoi această interfață a fost folosită de lingviști pentru a compara structura lingvistică și sintactică a limbilor germană și română.

- **CoRoLa** Corpusul limbii române contemporane (2013-2017): construirea unui corpus incluzând aprox. 1 mld. cuvinte românești, provenit din publicații care acoperă perioada de după cel de-al Doilea Război Mondial, precum și aprox. 300 de ore de înregistrări vocale, toate culese din 17 domenii și mai multe stiluri literare. Textele, provenite din romane, articole de presă, bloguri, scrieri științifice, piese de teatru etc., au fost curățate, segmentate la borne de propoziție și cuvinte, iar cuvintele au fost adnotate automat cu informații morfosintactice. Accesul în căutare multicriterială este permis prin 3 interfețe. Corpusul poate fi folosit de persoanele interesate să învețe limba română din exemple, la

clasă – în scop educațional, dar și în cercetări lingvistice ori pentru crearea de modele computaționale ale limbii române. Toate datele de natură textuală și înregistrările vocale au fost cedate de deținătorii cu drepturi de proprietate intelectuală asupra lor, pe baza unor protocoale semnate.

- **MappingBooks** *Let's jump in the Book!* (2014-2017): Un proiect UEFISCDI care a dezvoltat un nou tip de produs electronic, adresat elevilor care studiază geografia și turiștilor. Tehnologia apelează la tehnici PLN, de cartografie web, la tehnici de realitate mixtă și de inteligență ambientală. Toponimele menționate în carte și alte referințe de interes pentru cititor devin active și pot ghida cititorul pe teren, sau sunt complementate cu informații obținute din internet.

- **ReTeRom** Resurse și tehnologii pentru dezvoltarea interfețelor om-mașină în limba română (2017-2020): un proiect complex, în care am fost responsabil de proiectul component COBILIRO (Corpus bimodal pentru limba română adnotat pe mai multe niveluri), care a vizat crearea unui tezaur cu resurse audio și textuale paralele, adnotate la diferite niveluri acustice și lingvistice, care să devină cea mai importantă referință pentru acest tip de resursă pentru limba română (COBILIRO), precum și a unei platforme care să găzduiască tezaurul și să permită accesul facil în căutare, editare și descărcare.

- **DeLORO** - *Deep Learning for Old Romanian* (2019-2022) – modele de interpretare a textelor vechi românești în caractere chirilice. Proiectul intenționează să dezvolte o tehnologie capabilă să descifreze documente vechi românești redactate în alfabetul chirilic, tipărite ori în scriere de mână fără ligamente, și să le translitereze în alfabet latin.

6. Publicații, conferințe

Lista de publicații include peste 250 de titluri. Ea poate fi consultată la adresa: <https://profs.info.uaic.ro/~dcristea/publications.html>. Între acestea, se regăsesc:

- 5 cărți de autor, în Springer, Editura Tehnică, Editura Academiei Române, Editura UAIC, dintre care 2 ca unic autor;

- 16 cărți editate, dintre care 12 în seria „Resurse lingvistice și instrumente pentru prelucrarea limbii române”, numită apoi *Linguistic Resources And Tools For Processing Natural Language*, în Editura Universității "Al.I.Cuza", Iași, ISSN 1843-911X;

- 38 de capitole de cărți publicate în editurile: North-Holland, Springer, Benjamin Publishing Books, INSTICC Press (Portugalia), Editura Universității "Al.I.Cuza", Editura Academiei Române, Bulgarian Academy of Sciences, Trinitas – Iași, Editura Expert – București;

- 25 de articole publicate în reviste de specialitate;

- 118 lucrări publicate în proceedings-uri de conferințe, dintre care: 69 internaționale și 45 naționale;

- 44 de eseuri și alte publicații despre știință, cele mai multe apărute în internet, dintre care 33 în limba engleză. Majoritatea eseurilor au apărut în CLARIN Newsletter, al cărui *editor-in-chief* am fost între anii 2008 și 2011;

- 7 cursuri publicate pe internet.

Mai jos dau o selecție din aceste publicații.

Cărți de autor sau editate în colaborare:

- cele 13 Proceedings ale conferinței *ConsILR – International Conference on Linguistic Resources and Tools for Natural Language Processing*, edițiile publicate: 2006,

2007, 2008, 2010, 2011-2012, 2013, 2014, 2015, 2016, 2018, 2019, 2020 și 2021 (ultima în curs de publicare), la Editura Universității „Alexandru Ioan Cuza” din Iași.

- 5 ediții BringITon! Catalogue, în anii: 2010, 2012, 2014, 2016, 2018, la Editura Universității „Alexandru Ioan Cuza” din Iași.

- Diana Trandabă, Elena Irimia, Verginica Barbu Mititelu, Dan Cristea, Dan Tufiș (2012) *The Romanian Language in the Digital Age / Limba română în era digitală*, in White Paper Series, Eds. Georg Rehm and Hans Uszkoreit, Berlin, Springer, ISBN 978-3-642-30702-7, 87 p.

- Dan Cristea (2002). *Formalisme și instrumente de descriere și prelucrare ale limbajului natural*, Editura Universității "Al.I.Cuza", Iași, 260 pag.

- Dan Cristea (2002). *Programarea bazată pe reguli*. Editura Academiei, București, 216 pag.

- Cristian Giumale, Dan Preoteșcu, Luca Șerbănați, Dan Tufiș, Gheorghe Tecuci, Dan Cristea (1987). *LISP*, Technical Publishing House, Bucharest, vol. 2.

O selecție de capitole de cărți:

- D. Cristea, C. Pădurariu, P. Rebeja, M. Onofrei (2020). *From Scan to Text. Methodology, Solutions and Perspectives of Deciphering Old Cyrillic Romanian Documents into the Latin Script*. In: *Knowledge, Language, Models, Volume in Honour of Prof. Galia Angelova on Her 65th Birthday*. Milena Slavcheva, editor. INCOMA Ltd. Shoumen, BULGARIA. ISBN 978-954-452-062-5, pp. 38-56.

- Cristea D. (2019). *La confluența dintre științele umaniste și informatică, conferința „Perspectives in Humanities and Social Sciences: Hinting at Interdisciplinarity”*, ediția a V-a, „Mapping Digital Futures”, Universitatea „Alexandru Ioan Cuza”, Iași pe 23-24 mai 2018, în Anca-Diana Bibiri, Camelia Grădinaru, Emanuel Grosu, Andreea Mironescu, Roxana Patraș (eds.) *Exploring the Digital Turn*, Editura Universității “Alexandru Ioan Cuza” din Iași, ISBN: 978-606-714-504-5.

- Dana Bădulescu, Dan Cristea (2018). *Reading Books Differently*, in *KnygoTyra 70*, ISSN 0204 – 2061, (online) ISSN 2345 – 0053, Vilnius University, Book Science.

- Dan Cristea, Daniela Gîfu, Mihaela Colhon, Paul Diac, Anca-Diana Bibiri, Cătălina Mărănduc, and Liviu-Andrei Scutelnicu (2015). *Quo Vadis: A Corpus of Entities and Relations*. In N.Gala, R.Rapp and G.B.Enguix (eds.): *Language Production, Cognition, and the Lexicon*, Springer International Publishing Switzerland.

- Daniel Anechitei, Dan Cristea, Ioannidis Dimosthenis, Eugen Ignat, Diman Karagiozov, Svetla Koeva, Mateusz Kopeć, Cristina Vertan (2013). *Summarizing Short Texts Through a Discourse-Centered Approach in a Multilingual Context*. In Neustein, A., Markowitz, J.A. (eds.), *Where Humans Meet Machines: Innovative Solutions to Knotty Natural Language Problems*. Springer Verlag, Heidelberg/New York (copyright with Springer).

- Dan Cristea and Ionuț Cristian Pistol (2012). *Multilingual Linguistic Workflows*, in Cristina Vertan and Walther v. Hahn (Eds.) *Multilingual Processing in Eastern and Southern EU Languages. Low-resourced Technologies and Translation*, Cambridge Scholars Publishing, UK.

- Dan Cristea, Adrian Iftene (2011). *Grounding Coherence Properties of Discourse*. In *ALEAR Final Report*, vol. II. *Embodied Cognitive Semantics*, Berlin, April.

- Dan Cristea, Adrian Iftene (2010). *Discourse Coherence – a Built-in Cognitive Mechanism?* In D. Tufiș and C. Forăscu (eds.) *Multilinguality and interoperability in*

language peocessing with emphasis on Romanian, Editura Academiei Române, București, ISBN 978-973-27-1972-5, pages 362 – 379.

- Lothar Lemnitzer, Cristina Vertan, Alex Killing, Kiril Ivanov Simov, Diane Evans, Dan Cristea, Paola Monachesi (2007). Improving the Search for Learning Objects with Keywords and Ontologies. In *Creating New Learning Experiences on a Global Scale, EC-TEL 2007*, Lecture Notes in Computer Science, vol. 4753/2007, ISBN 978-3-540-75194-6, pages 202-216.

- Dan Cristea (2002). Referențialitate și cursivitate în relație cu structura de discurs. In Dan Tufiș și Florin Gheorghe Filip (coords.): *Limba Română în Societatea Informațională - Societatea Cunoașterii*, Ed. Expert, București, pages 269-302.

O selecție de articole publicate în reviste:

- D. Bădulescu, D. Cristea (2020). Reading with the Body and the Bodies of Books, in *Philologica Jassyensia*, an XVI, nr. 1 (31), 2020, p. 13–22.

- Kupietz, M., Diewald, N., Trawiński, B., Cosma, R., Cristea, D., Tufiș, D., Váradi, T. and Wöllstein, A. (2020). Recent developments in the European Reference Corpus EuReCo. In Sylviane Granger & Marie-Aude Lefer (eds) *Translating and Comparing Languages: Corpus-based Insights. Corpora and Language in Use Proceedings 6*, Louvain-la-Neuve: Presses universitaires de Louvain.

- Cristea, D., Diewald, N., Haja, G., Mărănduc, C., Mititelu, V.B., Onofrei, M. (2019). How to Find a Shining Needle in the Haystack. Querying CoRoLa: Solutions and Perspectives. In *Revue Roumaine de Linguistique/Romanian Review of Linguistics*, vol. 64, nr.3, 2019, Publishing House of the Romanian Academy, ISSN: 0035-3957.

- Cristea D. (2019). La confluența dintre științele umaniste și informatică, conferința „Perspectives in Humanities and Social Sciences: Hinting at Interdisciplinarity”, ediția a V-a, „Mapping Digital Futures”, Universitatea „Alexandru Ioan Cuza”, Iași pe 23-24 mai 2018, în Anca-Diana Bibiri, Camelia Grădinaru, Emanuel Grosu, Andreea Mironescu, Roxana Patraș (eds.) *Exploring the Digital Turn*, Editura Universității “Alexandru Ioan Cuza” din Iași, ISBN: 978-606-714-504-5.

- Pădurariu C., Cristea, D. (2019). Solution for scanned documents segmentation and letter recognition, in *Proceedings of the 14th edition of the International Conference on Linguistic Resources and Tools for Natural Language Processing – ConsILR-2019*, Ed. Universității “Alexandru Ioan Cuza” din Iași, ISSB: 1843-911X.

- Cristea, D., Pădurariu, C., Boghiu, Ș., Gîfu, D., Onofrei, M., Trandabăț, D., Pistol, I.C., Bibiri, A.-D., Scutelnicu, A. (2019). The CoBiLiRo project: Building and Distributing a Bimodal Corpus for Romanian Language, in *Proceedings of the 14th edition of the International Conference on Linguistic Resources and Tools for Natural Language Processing – ConsILR-2019*, Ed. Universității “Alexandru Ioan Cuza” din Iași, ISSB: 1843-911X.

- Rebeja, P., Cristea, D. (2019). Tracking the Influence of the Two Historical Unions on the Evolution of the Romanian Language, in *Proceedings of the 14th edition of the International Conference on Linguistic Resources and Tools for Natural Language Processing – ConsILR-2019*, Ed. Universității “Alexandru Ioan Cuza” din Iași, ISSB: 1843-911X.

- Dan Cristea, Daniela Gîfu, Alex Moruz, Mihaela Onofrei, Laura Pistol, Andrei Scutelnicu, Cecilia Bolea (2017). An Insight into the Corpus of Contemporary Romanian. In *Memoirs of the Scientific Sections / Memoriile Secțiilor Științifice*, Series IV, Tome XL, Publishing House of the Romanian Academy, pp. 67-84, ISSN 1224-1407, ISSN (online) 2343-7049.

- Dan Cristea, Andreea Macovei (2017). Why Time Resembles Rail Yards? A Way to Look at Time in Text Books, in *Romanian Journal of Information Science and Technology*, Vol. 19, 1-2, 2016, pp. 31-43.

- Dan Cristea (2016). Natural Language Processing versus Logic. Pros and cons on the dispute whether logic is useful in the computational interpretation of language, in *Computer Science Journal of Moldova*, vol.24, no.3(72), pp. 339-416.

O selecție de articole apărute în proceedings de conferințe:

- D. Cristea, I. Pistol, Ș. Boghiu, A. Bibiri, D. Gîfu, A. Scutelnicu, M. Onofrei, D. Trandabăț, G. Bugeag (2020). CoBiLiRo: a Research Platform for Bimodal Corpora. In: *Proceedings of the 1st International Workshop on Language Technology Platforms (IWLTP 2020)*, European Language Resources Association (ELRA), George Rehm et al. (eds.), pp. 22-27.

- A. Scutelnicu, C. Mărănduc, D. Cristea (2020). Adding a Syntactic Annotation Level to the Corpus of Contemporary Romanian Language, *Proceedings of the 8th Workshop on Challenges in the Management of Large Corpora*, Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020 c European Language Resources Association (ELRA).

- Kupietz, M., Diewald, N., Trawiński, B., Cosma, R., Cristea, D., Tufiș, D., Váradi, T. and Wöllstein, A. (2020). Recent developments in the European Reference Corpus EuReCo. In Sylviane Granger & Marie-Aude Lefer (eds): *Translating and Comparing Languages: Corpus-based Insights. Corpora and Language in Use Proceedings 6*, Louvain-la-Neuve: Presses universitaires de Louvain.

- Rebeja P., Cristea D. (2020). A dual-encoding system for dialect classification. In *Proceedings of VarDial 2020, Seventh Workshop on NLP for Similar Languages, Varieties and Dialects*, Coling 2020, December 13, 2020.

- Alex Moruz, Andrei Scutelnicu, Dan Cristea (2018). in V.Păiș, D.Gîfu, D.Trandabăț, D.Cristea, D.Tufiș (eds.) *Proceedings of the 13th International Conference on Linguistic Resources and Tools for Processing Romanian Language*, noiembrie 22-23, Editura Universității “A.I.Cuza” din Iași, pp. 125-132, ISSN 1843-911X.

- Dan Cristea, Daniela Gîfu, Svetlana Cojocaru, Alexandru Colesnicov, Ludmila Malahov, Mihai Popescu, Mihaela Onofrei, Daniela Cecilia Bolea (2018). How To Find Out Whether The Romanian Language Was Influenced By The Two Historical Unions?, in V.Păiș, D.Gîfu, D.Trandabăț, D.Cristea, D.Tufiș (eds.) *Proceedings of the 13th International Conference on Linguistic Resources and Tools for Processing Romanian Language*, noiembrie 22-23, Editura Universității “A.I.Cuza” din Iași, pp. 77-88, ISSN 1843-911X.

- Dan Tufis, Dan Cristea (2018). A Bird's-eye View of Language Processing Projects at the Romanian Academy, in *Proceedings of Language Resources and Evaluation Conference (LREC)*, 7-12 May, Miyazaki, Japan.

- Daniela Gîfu, Dan Cristea, Svetlana Cojocaru (2018). Language Independent Entity Linking in Fiction, at the *International Conference on Mathematics, Informatics and Information Technologies dedicated to the illustrious scientist Valentin Belousov*, MITI 2018, April 19-21, 2018, Bălți, Republic of Moldova.

- Dan Cristea, Ionut Pistol, and Daniela Gîfu (2018). Connecting the Content of Books to the Web and the Real World, *Proceedings of the 19th International Conference on Computational Linguistics and Intelligent Text Processing*, CICLing 2018, 18-24 March 2018, Hanoi, Vietnam.

- Mihaela Colhon, Dan Cristea (2017). Improving an NP-Chunker by Exploiting the Inner NP Dependency Structure. In: *Proceedings of the International Conference on Human-Computer Interaction (RoCHI-2017)*, Craiova 11-12 Sept., pp. 163-168.
- Marc Kupietz, Andreas Witt, Piotr Bański, Dan Tufiş, Dan Cristea and Tamás Váradi (2017). EuReCo - Joining Forces for a European Reference Corpus as a sustainable base for cross-linguistic research. In: Bański, Piotr/Kupietz, Marc/Lüngen, Harald/Rayson, Paul/Biber, Hanno/Breiteneder, Evelyn/Clematide, Simon/Mariani, John/stevenson, Mark/Sick, Theresa (eds.): *Proceedings of the Workshop on Challenges in the Management of Large Corpora and Big Data and Natural Language Processing (CMLC-5+BigNLP) 2017*, including the papers from the Web-as-Corpus (WAC-XI) guest section. Birmingham, 24 July 2017, Mannheim: Institut für Deutsche Sprache, 2017. Pp. 15-19.
- Dan Cristea, Eveline Wandl-Vogt, Mihaela Onofrei, Andrei Scutelnicu (2017). Towards a technology for dictionary intermediated dynamic alignment of multilingual corpora. A vision and some concrete steps, *COST ENeL Meeting*, 24-25 February 2017, Budapest.
- Dan Cristea (2017). The Time Yards Model: Rethinking the way to look at texts. *Proceedings of SYNASC-2016*, p. 15-22.
- Verginica Barbu Mititelu, Dan Cristea, Ruxandra Cosma (2017). Corpus of Contemporary Romanian. Architecture, Annotation Levels and Analysis Tools. *Actes of the 16th International Conference of the Department of Linguistics: Romanian Linguistics, Romance Linguistics*, Bucharest, Nov. 25-26 2016.

Am scris în colaborare cu colegii mei peste 150 de rapoarte de cercetare la UAIC-FII și ARFI-IIT.

Conform Google Scholar, am fost citat de 1743 de ori, am h-index = 19 și i10-index = 42 (ian. 2022).

Am avut aproximativ 150 de luări de cuvânt în conferințe, o parte din acestea putând fi consultate în secțiunea „A selection of tutorials and conferences” din pagina mea de publicații, la care se adaugă zeci de luări de cuvânt (de câte 10-15 min) în întâlniri de proiect naționale și internaționale.

7. Eseuri și publicații literare

- Dan Petru Cristea (2013). Noul colonialism al online-ului. În *Echidistanțe – Revistă de cultură și științe*, Institutul European, martie, număr coordonat de Cristina Florescu.
- Dan Petru Cristea (2012). Oameni și dicționare. În *Echidistanțe – Revistă de cultură și științe*, Institutul European, iulie, număr coordonat de Cristina Florescu.
- Dan Petru Cristea (2011). La ce e bună cartea online. În *Caiete Critice*, martie, #3, Fundația Națională pentru Știință și Artă.
- M. Tadic, D. Cristea (2011). (eds.) *CLARIN Newsletter*, no. 13, March.
- M. Tadic, D. Cristea (2011). Editors' Foreword. În *CLARIN Newsletter*, no. 13, March.
- D. Cristea, Gabriela Haja (2011). Thesaurus Dictionary of Romanian Language in Electronic Form. În *CLARIN Newsletter*, no. 13, March.
- Dan Petru Cristea (2011) Ajutând limbile aflate în pericol. În *Caiete Critice*, februarie, Fundația Națională pentru Știință și Artă.
- Dan Petru Cristea (2011). Evoluția limbajului. Poate fi ea descifrată de roboți? În *Caiete Critice*, ianuarie, Fundația Națională pentru Știință și Artă.
- D. Cristea (2011). Oameni și dicționare. În *Echidistanțe – Revistă de cultură și științe*, Institutul European, Serie nouă nr. 10 (58), ISSN 1221-8987.

- M. Tadic, D. Cristea (2010). (eds.) CLARIN Newsletter, no. 11-12, December.
- M. Tadic, D. Cristea (2010). Editors' Foreword. In CLARIN Newsletter, no. 11-12, December.
- Dan Petru Cristea (2010). O dizolvare care nu funcționează în online. In Caiete Critice, #12, decembrie, Fundația Națională pentru Știință și Artă.
- Dan Petru Cristea (2010). Istoria individ cu individ. In Caiete Critice, #11, noiembrie, Fundația Națională pentru Știință și Artă.
- Dan Petru Cristea (2011). Cyberspațiul. In Caiete Critice, #7, Fundația Națională pentru Știință și Artă.
- Dan Petru Cristea (2010). Coaliții europene între umaniști și informaticieni. In Caiete Critice, octombrie, Fundația Națională pentru Știință și Artă.
- M. Tadic, D. Cristea (2010). (Eds.) CLARIN Newsletter, no. 9-10, March-June.
- M. Tadic, D. Cristea (2010). Editors' Foreword. In CLARIN Newsletter, no. 9-10, March-June.
- Dan Petru Cristea (2010). *Scaune de pluș. Roman care a primit premiul de debut al Editurii Cartea Românească pentru anul 2009*, Polirom, mai 2009.
- M. Tadic, D. Cristea (Eds.) (2009). CLARIN Newsletter, no. 8, December.
- M. Tadic, D. Cristea (2009). Editors' Foreword. In CLARIN Newsletter, no. 8, December.
- M. Tadic, D. Cristea (2009). (Eds.) CLARIN Newsletter, no. 7, September.
- M. Tadic, D. Cristea (2009). Editors' Foreword. In CLARIN Newsletter, no. 7, September.
- M. Tadic, D. Cristea (2009). (Eds.) CLARIN Newsletter, no. 6, June.
- M. Tadic, D. Cristea (2009). Editors' Foreword. In CLARIN Newsletter, no. 6, June.
- F. Binder, D. Cristea (2009). Training and dissemination in CLARIN – the potential to bridge gaps. In CLARIN Newsletter, no. 6, June.
- M. Tadic, D. Cristea (2009). (Eds.) CLARIN Newsletter, no. 5, March.
- M. Tadic, D. Cristea (2009). Editors' Foreword, In CLARIN Newsletter, no. 5, March.
- M. Tadic, D. Cristea (2008). (Eds.) CLARIN Newsletter, no. 4, December.
- M. Tadic, D. Cristea (2008). Editors' Foreword. In CLARIN Newsletter, no. 4, December.
- D. Cristea (2008). Romanian language resources and tools. In CLARIN Newsletter, no. 4, December.
- M. Tadic, D. Cristea (2008). (Eds.) CLARIN Newsletter, no. 3, October.
- M. Tadic, D. Cristea (2008). Editors' Foreword. In CLARIN Newsletter, no. 3, October.
- M. Tadic, D. Cristea (2008). (Eds.) CLARIN Newsletter, no. 2, July.
- M. Tadic, D. Cristea (2008). Editors' Foreword. In CLARIN Newsletter, no. 2, July.
- M. Tadic, D. Cristea (2008). (Eds.) CLARIN Newsletter, no.1, May.
- M. Tadic, D. Cristea (2008). Editors' Foreword. In CLARIN Newsletter, no.1, May.
- D. Cristea (2008). Kick-off meeting in Nijmegen. In CLARIN Newsletter, no.1, May.
- D. Cristea (2008). WP6: Dissemination and advisory services. In CLARIN Newsletter, no.1, May.
- D. Cristea, Hamish Cunningham, Nancy Ide, Rada Mihalcea, Dan Tufiş (2005). (Eds). EuroLan'05 - The Multilingual Web: Resources, Technologies and Prospects July 25-August 5 (CD publication).
- D. Cristea (2003). O dizolvare demodată. In Consemnări, Academia Română – Filiala Iași, no.1.

- D. Cristea (2002). Să ne jucăm cu cuvintele. Proiecte de prelucrare a limbajului natural. In Revista de Informatică, no. 1, Iași.

8. Recunoașteri

După cunoștința mea, sunt (sau am fost) membru al următoarelor academii, agenții, organizații, asociații și comisii:

- Academia Română (membru corespondent din 2015),
- Academia de Științe Tehnice din România, Secția de Știința și Tehnologia Informației (corespondent din 2002, titularizat în 2015),
- Asociația Română de Inteligență Artificială (membru fondator),
- Comisia CNATDCU pe domeniul Informatică (din 2011, președinte al Comisiei între 2004 - 2011 și 2012 - 2016),
- ARACIS – Agenția Română pentru Asigurarea Calității în Învățământul Superior, pentru domeniul școlilor doctorale;
- Consorțiul *Informatica Univesitaria*;
- Asociația Română de Lingvistică Computațională (membru fondator),
- Asociația Pedagogilor Creștini (membru fondator),
- Comisia de Informatizare pentru Limba Română din cadrul Academiei Române (neînregistrată) – în calitate de vicepreședinte pentru secțiunea Prelucrarea Limbajului Natural,
- *Association de Recherches Cognitives (ARC)* – France,
- *Association of Computer Machinery (ACM)* – USA,
- *SIGDial – Special Interest Group in Dialogue and Discourse* – USA,
- *Balkan Asociation of Fuzzy Systems* – Romania.

Am fost membru al panelului de evaluatori pentru programele de cercetare CHIST-ERA finanțate de CE.

Am activat ca membru al panelului de experți Cercetare-Dezvoltare-Inovare al UEFISCDI, pe domeniul Sisteme inteligente.

În decembrie 2003 am primit premiul “Grigore Moisil” al Academiei Române, Secția de Știința și Tehnologia Informației, pentru o lucrare publicată în colaborare în 2001 în revista *Information Science and Technology*, Editura Academiei Române.

În decembrie 2006 am primit premiul cercetării, acordat de CNCSIS.

În 2008 am primit premiul pentru activitatea de cercetare, acordat de UAIC.

În perioada 2000 – 2021 am făcut parte din aprox. 195 de comitete de program și științifice ale unor manifestări internaționale și naționale, ori ca recenzor pentru jurnale.

9. Conducere de doctorat, membru în comisii de doctorat și abilitare

Am finalizat în calitate de îndrumător 12 teze de doctorat, în cadrul Școlii Doctorale din UAIC-FII, dintre care 4 cu îndrumare în cotutelă. Actualmente am 6 studenți doctoranzi în îndrumare, încadrați în ȘD UAIC-FII, și o studentă în cotutelă cu ȘD UAIC-Litere.

Am fost membru în comisii de doctorat sau abilitare la următoarele universități și institute de cercetare din România: UAIC Iași – facultățile de Informatică, Litere și Filosofie, Academia Română –Institutul de Cercetări în Inteligență Artificială „M.Drăgănescu”, Universitatea „Babeș-Bolyai” din Cluj-Napoca – Facultatea de Matematică-Informatică, Universitatea Politehnica București – Facultatea de Automatică și Calculatoare, Universitatea

Tehnică „Gheorghe Asachi” Iași – Facultatea de Electronică și Telecomunicații, Universitatea Dunării de Jos din Galați – Facultatea de Matematică și Informatică, Universitatea „Ștefan cel Mare” din Suceava – Facultatea de Automatică și Calculatoare, Universitatea din București – Facultatea de Matematică-Informatică.

Am fost membru în comisii de doctorat sau abilitare la următoarele institute de cercetare și universități din străinătate: Universitatea din Neuchâtel (Elveția), Universitatea din Geneva (Elveția), Max-Planck Institute Nijmegen (Olanda), Universitatea din Strasbourg (Franța), Universitatea Lille I (Franța), Universitatea din Trento (Italia), Universitatea din Madrid (consultant), Institutul de Matematică – Informatică al Academiei de Științe a Moldovei.

10. Fotografii istorice



În ședința de deschidere a Școlii de Vară EUROLAN-2001, Iași (împreună cu: Romand Romary, Dumitru Oprea, Nancy Ide și Dan Tufiș)



În fața afișelor Școlii Doctorale a Facultății de Informatică de la UAIC și a Școlii internaționale EUROLAN, ediția 2019



Fotografie recentă cu o parte din colegii de la UAIC-FII, ARFI-IIT și ARFI-Philippide (în fotografie apar majoritar membri ai grupului comun de cercetare în limbaj natural: Corina Forăscu, Adrian Iftene, Anca Bibiri, Cecilia Bolea, Gabriela Haja, Isabelle Tamba, Diana Trandabăț, Paula Crucianu, Daniela Gîfu, Mihaela Onofrei, Andrei Scutelnicu)