

ROMÂNIA
MINISTERUL EDUCAȚIEI NAȚIONALE
UNIVERSITATEA AGORA DIN MUNICIPIUL ORADEA



Doctor Honoris Causa



PROFESOR UNIVERSITAR DOCTOR

DAN TUFIȘ

Membru titular al Academiei Române
Director al Institutului de Cercetări în Inteligență Artificială
al Academiei Române

ORADEA
18 OCTOMBRIE 2019



UNIVERSITATEA AGORA
DIN MUNICIPIUL ORADEA

Căutare...

[Romana](#) [English](#)

[Contact](#) [Hartă site](#)

[Universitate](#) [Academic](#) [Studenti](#) [Licență](#) [Masterat](#) [Conferințe](#) [Cercetare](#)
[Calitate](#) [ICCCC2020](#) [Înscriere](#)



Mesajul rectorului

"Acces la succes!" (Moto UAO)"Echipa de management a Universității Agora din Oradea

Colaborare internațională: UAO-VGTU (2)

Universitatea Agora din Oradea are multe

Alte știri

Colaborare internațională: UAO-VGTU (1)



Universitatea Agora din Oradea,

Piata Tineretului nr. 8, 410526 Oradea, jud. Bihor,

Tel: +40 259 427 398, +40 259 472 513, Fax: +40 259 434 925, secretariat@univagora.ro,
rector@univagora.ro, www.univagora.ro

DOCTOR HONORIS CAUSA

AL UNIVERSITĂȚII AGORA DIN MUNICIPIUL ORADEA



**ACADEMICIAN
DAN TUFIȘ**

ORADEA, 18 OCTOMBRIE, 2019

Propunerea rectorului către Senatul Universității Agora

Academicianul Dan Tufiş este o personalitate științifică proeminentă în domeniul Inteligenței artificiale din România, cu o certă recunoaștere internațională, având 4 cărți de autor, 25 volume coordonate, 15 cursuri editate, 71 de capitole în cărți, 141 de articole în volumele unor conferințe, 64 de articole publicate în jurnale de specialitate, cu peste 3.000 de citări în Google Scholar.

Google Scholar



Dan Tufiş

FOLLOW

Professor of Artificial Intelligence, Institute for Artificial Intelligence, Romanian Academy

Verified email at racai.ro

Human Language Technolo... Machine Translation Semantic Web

TITLE	CITED BY	YEAR
The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages R Steinberger, B Pouliquen, A Widiger, C Ignat, T Erjavec, D Tufiş, ... arXiv preprint cs/0609058	551	2006
The Romanian Wordnet D Tufiş, E Barbu, VB Mititelu, R Ion, L Bozianu Romanian Journal on Information Science and Technology 7 (2-3), 105-122	370 *	2004
Sense discrimination with parallel corpora N Ide, T Erjavec, D Tufiş Proceedings of the ACL-02 workshop on Word sense disambiguation: recent ...	144	2002
Multext-east: Parallel and comparable corpora and lexicons for six central and eastern european languages	130	1998

Colaborarea domnului Dan Tufiş cu Universitatea Agora din Oradea (UAO) a început încă din anul 2006, cu ocazia primei ediții a *International Conference on Computers Communications and Control (ICCCC2006)* – organizată de UAO, unde a prezentat o interesantă comunicare invitată și a prezidat secțiunea de Inteligență artificială.

Imagini de la ICCC2006



J. Fodor – K. Hirota – D. Tufiş



K. Hirota—D. Tufiş –I. Dziţac...

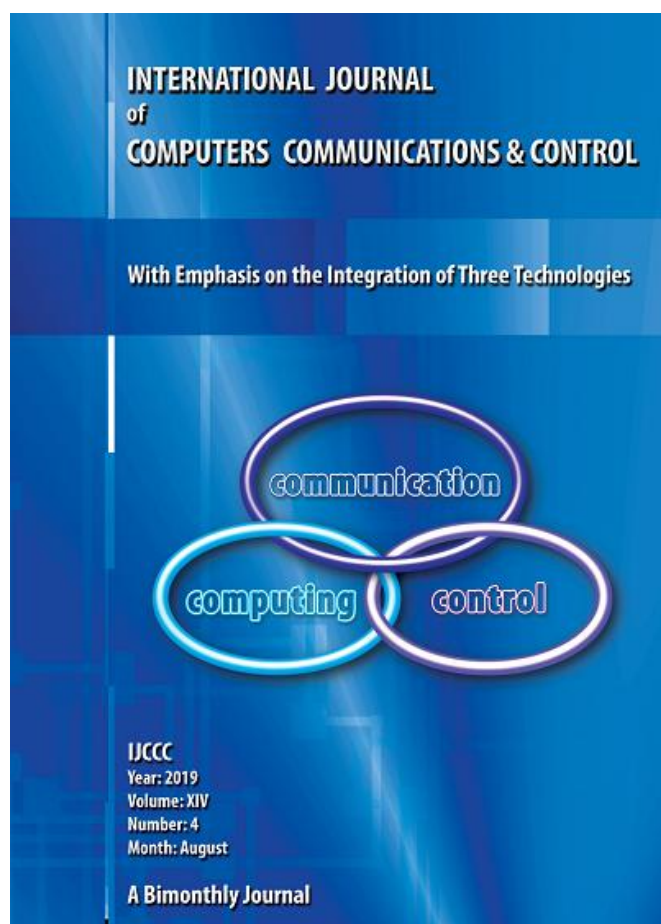


I. Dziţac — D. Tufiş—M.-J. Manolescu



Comunicare plenară la ICCCC2006: Dan Tufiş

Tot în anul 2006, domnul Dan Tufiş a acceptat să facă parte din colectivul editorial al *International Journal of Computers Communications & Control* (IJCCC) – revistă fondată la UAO în 2006 de către I. Dziţac, F.G. Filip şi M.J. Manolescu, care a început să fie indexată ISI Web of Science începând chiar cu ultimul număr din anul înfiinţării, ajungând în JCR2018 la un factor de impact de 1,585.



EDITORIAL TEAM

Dan TUFIS

Dan TUFIS
<http://www.racai.ro/~tufis/>
 Research Institute for Artificial Intelligence, Romanian Academy, Romania

Member of the Romanian Academy, Professor, Ph. D., Senior Scientist grade I.

- 1992 Ph.D. in Computer Science, *Polytechnic Institute of Bucharest*, Romania
- 1991 Post graduate studies in Computational Linguistics, *Linguistic Institute, University of California at Santa Cruz*, USA
- 1979 B.Sc. and M.Sc. in Computer Science, *Department of Computer Science, Polytechnic Institute of Bucharest*, Romania



Web of Science

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL

Impact Factor

1.585 **1.361**

2018 5 year

JCR ® Category	Rank in Category	Quartile in Category
AUTOMATION & CONTROL SYSTEMS	38 of 62	Q3
COMPUTER SCIENCE, INFORMATION SYSTEMS	108 of 155	Q3

Data from the 2018 edition of *Journal Citation Reports*

Publisher

CCC PUBL-AGORA UNIV, PIATA TINERETULUI 8, ORADEA, JUD, BIHOR, 410526, ROMANIA

ISSN: 1841-9836

eISSN: 1841-9844

Colaborarea noastră s-a consolidat în 2008, la ICCCC2008, cu ocazia prezenţei la Universitatea Agora a părintelui Logicii fuzzy, regretatul profesor Lotfi A. Zadeh (1921-2017). Cu acest prilej am organizat în colaborare un workshop exploratoriu în domeniul calculului în limbaj natural, la sfârşitul căruia am publicat în Editura Academiei Române un volum comun (editori: L.A. Zadeh, D. Tufiş, F.G. Filip, I. Dziţac).

Imagini de la ICCCC2008



**D. Tufiş-I. Dzitac – L.A. Zadeh-
M.-J. Manolescu – F.G. Filip**



**A. Manolescu—D. Tufiş – F.G. Filip
la ICCCC2008**



D. Tufiş—L.A. Zadeh



D. Tufiş—L.A. Zadeh – F.G. Filip



Având în vedere reputația științifică internațională, precum și contribuțiile benefice aduse imaginii și prestigiului instituției noastre, propun Senatului universitar acordarea titlului de Doctor Honoris Causa al Universității Agora din Oradea domnului prof.univ.dr. Dan Tufiş, membru titular al Academiei Române.

De asemenea, în acest scop, supun votului Senatului următoarea

COMISIE DE LAUDATIO

1. Acad. Florin Gheorghe Filip, Academia Română – Președinte.
2. Prof.univ.dr.ing. Mișu-Jan Manolescu, Universitatea Agora– Membru.
3. Prof.univ.dr. Adriana Manolescu, Universitatea Agora– Membru.
4. Prof.univ.dr. Ioana Moisil, Professor Honoris Causa al Universității Agora – Membru.
5. Dr. Bogdana Stanojevic – Cercetător I, Institutul de Matematică al Academiei Sârbe de Științe și Arte – Membru.

RECTOR,
Prof. univ. dr. habil. Ioan Dziţac



LAUDATIO

***adresat domnului profesor universitar doctor Dan Tufiş,
membru titular al Academiei Române,
cu ocazia decernării titlului de Doctor Honoris Causa
al Universităţii Agora din municipiul Oradea***



Preambul

Dan Tufiş s-a născut la 5 februarie 1954 în Bucureşti. A urmat liceul nr. 43 din Bucureşti (numit între anii 1977 şi 1990 Liceul de Matematica-Fizica nr. 3, iar acum Liceul Teoretic "A.I. Cuza"), fiind olimpic naţional la fizică în anii 1969-1971.

În anii 1974-1979 a urmat cursurile Facultăţii de Automatică şi Calculatoare a Universităţii „Politehnica” Bucureşti cu repartitie guvernamentală la Institutul Naţional de Cercetare în Informatică, Bucureşti. În 1991 a absolvit studii de specializare (masterat) în lingvistică computaţională la Institutul Lingvistic, Universitatea Santa Cruz, California, iar titlul de doctor în Calculatoare l-a obţinut în 1992 de la Facultatea de Automatică şi Calculatoare a Universităţii „Politehnica” Bucureşti, pentru teza „Mediu de dezvoltare a sistemelor de dialog în limbaj natural», sub coordonarea Prof. Mircea Petrescu. Obţine titlul de Cercetător Ştiinţific gradul I în anul 1992 la Institutul Naţional de Cercetare în Informatică, Bucureşti, iar în 2003 Universitatea A.I. Cuza, Iaşi îi conferă titlul de Profesor universitar. În anul 1997 este ales Membru Corespondent al Academiei Române, la Secţia de Ştiinţa şi Tehnologia Informaţiei, iar în anul 2011 este ales Membru Titular al Academiei Române. Din 2004 este conducător de doctorat (8 studenţi au finalizat cu brio tezele de doctorat, alţi doi urmand să sustină public tezele anul acesta). Între 2005-2010, reprezintă Secţia de Ştiinţa şi Tehnologia Informaţiei în Prezidiul Academiei Române, în prezent făcând parte din Biroul Secţiei. Din 1998 conduce Centrul de Cercetări Avansate, devenit în 2002 Institutul de Cercetări în Inteligenţă Artificială "Mihai Drăgănescu" al Academiei Române, din Bucureşti. În anul 2010, Universitatea "A.I. Cuza" din Iaşi îi acordă distincţia de "Profesor de Onoare".

În anul 2016 este decorat de Preşedintele ţării cu Ordinul "Steaua României" în grad de Cavaler.

Cercetările sale au fost recunoscute şi recompensate cu premiile Academiei Române (în 1989 şi 1994) cu numeroase diplome de merit şi distincţii. A fost nominalizat pentru premiul internaţional "Principe de Asturias" (în 2004).

Domnul Academician Dan Tufiş a publicat peste 300 de lucrări (4 volume de autor, 25 de volume coordonate, 71 capitole de carte, 141 de lucrări în volume ale conferinţelor naţionale şi internaţionale, 64 de lucrări în reviste de specialitate) şi a susţinut 82 de conferinţe invitate. Cf. Google Scholar este citat de peste 3100 ori având un h-index de 25. Face parte din colectivele de redacţie a 8 reviste ştiinţifice (International Journal on Language Resources and Evaluation, Springer Verlag, USA (from 2002), The Computer Science Journal of Moldova, Republica Moldova (from 2005), Proceedings of the Romanian Academy, Series A (din 1999)- topical editor, Romanian Journal on Science and Technology

of Information, Academia Română (from 1997), International Journal on Information and Control, ICI, România (from 1992), International Journal of Computers, Communication & Control, Universitatea „Agora” din Oradea (from 2006), Romanian Journal of Human – Computer Interaction, ACM SIGCHI Romania (from 2008), NOESIS – Travaux du comite roumain d’histoire et de philosophie de sciences, Academia Română (from 2007) –director (from 2015)).

A făcut parte din peste 150 de comitete științifice la conferințe internaționale și din 60 de comisii naționale și internaționale de acordare a titlului de doctor (Universitatea „A. I. Cuza” din Iași, Universitatea “Politehnica” din București, Universitatea București, Universitatea din Craiova, Academia Română, Universitatea “Luis Pasteur” din Strasbourg, Franța, Universitatea de Științe și Tehnologii din Lille, Franța, Institutul Politehnic din Grenoble, Franța, Universitatea Tehnică din Strasbourg, Franța, Universitatea din Pisa, Italia, Institutul de Tehnologie din Bombay, Institutul de Matematică și Informatică al Academiei Moldovei din Chișinău, Universitatea din Alicante, Spania).

Lucrările și/sau biografia prof. Dan Tufiş sunt incluse în numeroase enciclopedii:

- The second (revised) Handbook of Computational Linguistics, Oxford University Press, 2018
- Handbook of Linguistic Annotation, Springer, 2017
- Who’s Who in the World, 23rd Edition, Marquis Who’s Who[®], 2006-2012
- Who’s Who in Science and Engineering, 9th Edition, Marquis Who’s Who[®], 2006-2012
- Enciclopedia Personalităților din România, Hübners Who is Who, 2006-2012
- Enciclopedia of Language and Linguistics, second edition, Elsevier, 2005
- Membrii Academiei Române, Editura Academiei, 2004, 2010
- Who’s Who in Terminology, Unione Latine, 2003
- Who’s Who in Romania, Pegasus Press, 2002

Cercetările sale au fost sau au fost finanțate pe plan intern de: Institutul de Cercetări în Informatică, Comisia Națională pentru Informatică, Ministerul Cercetării și Tehnologiei, Academia Română, Ministerul Cercetării și Tehnologiei.

Iar pe plan extern de: Academia de Științe din Slovacia, Biroul Internațional de Schimburi Științifice (IREX), Comisia Europeană, Fundația Națională de Știință din SUA, Academia Regală din Anglia, Consiliul Național al Cercetării Științifice din Franța, Asociația Francofoniei (AUPELF-UREF), XEROX-Research Europe, Consiliul Federal al Cercetării din Elveția, Asociația Mondială de Lingvistică Computațională, NATO.

Anii de început

Activitatea științifică a profesorului Dan Tufiş este dedicată domeniului tehnologiilor limbajului natural, o ramură a științei și tehnologiei informației. Domeniul este unul preponderent tehnologic și aplicativ cu un înalt grad de creativitate și inovare. A început să studieze problematica sistemelor inteligente de prelucrare a limbajului natural în 1981, în 1982 propunând prima temă de cercetare din programul național, dedicată prelucrării limbajului natural prin metode ale inteligenței artificiale. La început a fost preocupat cu precădere de aspectele logice ale comunicării prin intermediul limbajului natural, dezvoltând metode și tehnici noi de reprezentare și prelucrare a cunoștințelor lingvistice. La scurt timp realizează primul sistem de dialog în limba română, SDLR, valorificat ulterior prin intermediul Bibliotecii Naționale de Programe.

În colaborare cu Centrul de Calcul al Universității „A.I. Cuza” din Iași, finalizează în 1983, un nou sistem de întrebare/răspuns în limbaj natural, independent de domeniul de aplicație și de limba de interogare. Programul, numit IURES, a constituit nu numai o premieră națională dar în multe privințe includea soluții inedite pe plan mondial. Astfel, schema de reprezentare a cunoștințelor realiza o

îmbinare a metodelor de reprezentare declarative cu cele procedurale. Se extindea conceptul de gramatică semantică prin introducerea (și implementarea) noțiunii de operator logico-lingvistic. IURES includea metode originale de navigare într-o rețea semantică cu moștenire multiplă. El a constituit obiectul a peste 30 de articole, comunicări științifice, rapoarte de cercetare și tehnice, care s-au bucurat de o deosebită apreciere, fiind citat de numeroși cercetători din țară și străinătate. Sistemul IURES a fost omologat internațional în 1988 și a constituit primul produs românesc de inteligență artificială exportat. Sistemul IURES este referit în enciclopedia de lingvistică computațională și în *Survey of the Current Status Research and Future Trends in Machine Translation and Natural Language Processing*, realizat în 1992 de *Japan Electronic Industry Development Association*.

În paralel cu activitatea în domeniul prelucrării limbajului natural, cercetătorul Dan Tufiș realizează, în anii '80, un mediu de programare funcțională, numit TC-LISP, care s-a impus în țară ca produs standard de programarea LISP pe minicalculatoare. Pentru mulți ani, toate realizările semnificative în domeniul inteligenței artificiale în România, până la apariția pe scară largă a calculatoarelor personale, s-au implementat în TC-LISP, limbaj care prezenta o serie de concepte de programare inedite în programarea LISP: spații virtuale multiple, aritmetică «chirurgicală», utilizarea controlată de utilizator a memoriei virtuale, programare paralelă etc.

Morfologia paradigmatică

În aceeași perioadă (1987-1989) realizează un sistem original de gestiune a dicționarilor de dimensiuni mari destinate sistemelor de prelucrare a limbajului natural. Cercetările în domeniul morfologiei și lexicologiei computaționale s-au concretizat în plan teoretic cu un model computațional original, morfologia paradigmatică. Dintre lucrările în care domnul Dan Tufiș a descris modelul morfologiei paradigmatică, *It Would Be Much Easier if WENT Were GOED*¹, prezentată la Conferința Europeană de Lingvistică Computațională în 1989, a fost cotate cea mai valoroasă contribuție, alături de comunicarea reputatului specialist american Ronald Kaplan de la Institutul de Cercetări Stanford. La aceeași conferință, independent de cercetările domnului Tufiș, dr. Jo Calder de la Universitatea din Edinburgh a propus un model similar numit tot “morfologie paradigmatică”. În momentul de față morfologia paradigmatică, alături de morfologia derivativă pe 2 niveluri, este considerată una dintre cele două modele morfologice unanim practicate (sub diferite variante) în tehnologia limbajului². Teoria morfologiei paradigmatică, a stat la baza implementării unui sistem de învățare automată a morfologiei limbilor naturale, numit PARADIGM, cercetări care au fost răsplătite cu premiul „Traian Vuia” al Academiei Române pe anul 1989.

Între anii 1993 și 1995, în colaborare cu Centrul de Studii Semantice și Cognitive din Geneva, dr. Dan Tufiș a dezvoltat un sistem integrat de prelucrări lingvistice numit Mac-ELU, considerat ca fiind un sistem de generația a 3-a (cea mai evoluată la nivelul anului 1993). Pe baza acestui sistem, colectivul condus de dr. Tufiș a lucrat la realizarea primului dicționar computațional românesc (bazat pe unificare) de mare acoperire lexicală. Dicționarul conținea peste 40.000 de intrări în formă leasă, pe baza cărora și a morfologiei paradigmatică a limbii române, puteau fi recunoscute și generate peste 1.000.000 de forme flexionate. Această cercetare a fost recompensată cu Premiul ”Tudor Tănăsescu” al Academiei Române în anul 1994.

Standarde lexicale și lexicografice, dezambiguizare morfo-lexicală, corpusuri mono- și multi-lingve

Între anii 1995-2000 a coordonat activitățile de cercetare în alte proiecte europene dintre care amintim: MULTEXT-EAST, TELRI (*Trans European Language Resources Infrastructure*), ELSNET (*Excelency in Language and Speech NETwork*) și CONCEDE (*Consortium for Central European Dictionary*

¹ D.Tufiș. “It Would Be Much Easier If WENT Were GOED”, in *Proceedings of the 4th European Conference of the Association for Computational Linguistics*, Manchester, 1989.

² Richard Sproat, “Morphology and Computation” MIT Press, 1992.

Encoding). Le menționăm, dintre multele în care a participat dr. Dan Tufiş în cariera sa, pentru că acestea au printre primele având ca obiectiv alinierea metodologică şi tehnologică la standardele şi recomandările internaţionale în domeniul ingineriei limbajului şi sinergizarea europeană a activităţilor naţionale în acest domeniu. Cercetările legate de standardizarea descrierilor morfo-lexicale au debutat la începutul anilor '90, coordonate de EAGLES (*Expert Advisory Group on Language Engineering Standards*), una dintre cele mai influente organizaţii profesionale europene în prelucrarea automată a limbajului. Cooptat în acest grup în 1994, dr. Tufiş a elaborat specificaţiile pentru codificarea dicţionarelor morfo-lexicale pentru limba română (1996), singura limbă est-europeană inclusă (la vremea respectivă) în standardele EAGLES.

Exploatănd descrierea morfologiei paradigmatică în termenii de atribut valoare, ca şi adecvarea ei la tehnicile de învăţare automată, dr. Tufiş dezvoltă în anul 1997 un model de proiectare automată, independent de limbă, a adnotatoarelor morfo-lexicale (generatoare de programe de dezambiguizare automată, la nivel morfo-lexical, a cuvintelor din texte arbitrare). Modelul dezambiguizării morfo-lexicale ierarhizate (cunoscut astăzi sub numele de *tiered-tagging*) este prezentat în 1999 şi implementat prima dată în adnotatorul morfo-lexical Q-Tag, reprezintă contribuţii larg citate în comunitatea internaţională. Unul din motivele aprecierii de care se bucură aceste contribuţii, pe lângă performanţele superioare altor abordări, este faptul că ele sunt independente de limbă şi sunt conforme unor standarde şi recomandări internaţionale asupra codificării morfo-lexicale.

Standardul EAGLES a fost extins (inclusiv pentru limba română) în cadrul proiectului european MULTEXT-EAST (finalizat în 1998), în care dr. Dan Tufiş a coordonat echipa din România. În cadrul acestui proiect s-a realizat nu numai extensia specificaţiilor de codificare, pe baza noilor recomandări ale ISO, şi TEI-P3 (*Text Encoding Initiative*) dar şi implementarea primului lexicon morfo-lexical (conţinând peste 400.000 de intrări la vremea respectivă) şi a primului corpus de limbă română (cu aproape 500.000 de cuvinte), cu o codificare aliniată la practica internaţională. Aceste resurse lingvistice, unice la vremea respectivă în limba română, atât prin cantitate dar mai ales prin calitate (prelucrările statistice au fost validate şi corectate manual) au constituit nucleul dezvoltărilor ulterioare ale corpusurilor şi lexicoanelor de limbă română existente actualmente în institutul pe care îl conduce. În prezent, lexiconul conţine peste 1.200.000 de intrări iar corpusurile construite de-a lungul anilor includ peste 2.000.000.000 de articole lexicale (cuvinte şi semne de punctuaţie) adnotate.

Colectarea şi prelucrarea corpusurilor de mari dimensiuni a fost o preocupare constantă a cercetărilor coordonate de prof. Dan Tufiş, una dintre cele mai importante realizări în această direcţie fiind contribuţia esenţială la cel mai important corpus paralel multilingual folosit în cercetările de traducere automată: JRC-Acquis. Corpusul, conţinând documente în 21 de limbi, care a fost finalizat de o echipă internaţională, coordonată de dr. Ralf Steinberger de la JRC-Ispra, şi distribuit comunităţii ştiinţifice în 2006 a inclus documente în limbile română şi bulgară ce au fost prelucrate şi documentate de colectivul coordonat de Prof. Dan Tufiş.

Între anii 2008-2010 a coordonat echipa românească în proiectele CLARIN şi FlaReNet proiecte care au pus bazele dezvoltării de anvergură a resurselor lingvistice pentru limba română în format standardizat şi interoperabil. Proiectele ACCURAT (2010-2012) şi MetaNet4U (2010-2013) au fost cele care au condus la internaţionalizarea şi valorificarea în context multilingual a cercetărilor proprii. Proiectul ACCURAT a fost determinant în lansarea cercetărilor privind extragerea de date de antrenament pentru sisteme de traducere automată din texte comparabile, nu din texte paralele cum se obișnuiește. Rezultatele acestui proiect au constituit volumul co-editat de Prof. Dan Tufiş şi apărut la Springer în 2019 "Using Comparable Corpora for Under-Resourced Areas of Machine Translation", în *series Theory and Applications of Natural Language Processing*, Springer, 2019, 978-3-319-99003-3. Proiectul MetaNet4U a pus bazele platformei de diseminare a resurselor lingvistice pentru limba română şi a pregătit lansarea unor proiecte de anvergură: CoRoLa (2014-2019), DruKoLa (2015-2019), ELRC (2015-2019), Presidency Translation Kit (2018-2020), ReTeRom (2018-2020), ROBIN (2018-2020).

Între anii 2014-2017 Prof. Dan Tufiş a coordonat crearea celui mai mare corpus de limbă română contemporană (scrisă şi vorbită), CoRoLa, în parteneriat cu Institutul de Informatică Teoretică din Iaşi al Academiei Române, în cadrul programului prioritar al Academiei Române. Proiectul, care a avut un ecou excepţional în comunitatea academică şi nu numai, a fost prelungit pentru încă doi ani. Corpusul CoRoLa, pe lângă dimensiunea impresionantă (peste 1.2 miliarde de articole lexicale), prelucrările şi meta-datele aferente, prezintă o caracteristică rar întâlnită la corpusurile de structură şi dimensiuni similare: textele incluse în corpus sunt obţinute pe baza unor acorduri de colaborare cu proprietarii drepturilor de proprietate intelectuală.

Lexicografia computaţională şi ontologii lexicale

În perioada 1997-1999 în cadrul proiectului european CONCEDE (*CONsortium for Central and Eastern Dictionaries Encoding*) dr. Dan Tufiş a fost unul din realizatorii schemei XML generice³ de codificare standardizată a dicţionarilor explicative. Schema de codificare, cunoscută sub numele CONCEDE, a fost folosită pentru implementarea unor dicţionare explicative pentru mai multe limbi (bulgară, cehă, engleză, estoniană, maghiară, slovenă şi desigur română). Un rezultat foarte semnificativ al activităţii în domeniul lexicografiei computaţionale a fost realizarea unui compilator pentru dicţionare în format tipografic (de exemplu, Word) ce analizează textul respectiv şi generează codul XML conform cu descrierea CONCEDE. Compilatorul, numit **DIC** se bazează pe gramatica convenţiilor tipografice specifice şcolii româneşti de lexicografie, fiind parametrizabil atât în raport cu convenţiile tipografice cât şi cu schema XML a codificării ţintă. Cu ajutorul acestui compilator, în anul 2000 a fost finalizată implementarea conformă cu schema CONCEDE a întregului dicţionar explicativ al limbii române (DEX, ediţia 1996). Această implementare profesională a DEX-ului permite regăsirea de informaţii lexicografice după o mulţime de criterii (categorie gramaticală, sufixe gramaticale sau lexicale, etimologie, variante, grupuri de litere conţinute în cuvântul temă, definiţii etc.). Aceste cercetări şi rezultatele obţinute au facilitat lansarea în anul 2001 a proiectului de ontologie lexicală pentru limba română, proiect de un deosebit impact pentru comunitatea ştiinţifică interesată de prelucrarea automată a limbii române.

Este vorba de proiectul european BalkaNet⁴, în care grupul de cercetare al profesorului Tufiş şi cel de la Facultatea de Informatică a Universităţii Alexandru Ioan Cuza din Iaşi au fost responsabile de implementarea componentei de limbă română a Wordnet-ului – tezaur lexical, dezvoltat iniţial pentru limba engleză, sub coordonarea profesorului George Miller de la Universitatea Princeton⁵. Global Wordnet Association⁶ a indexat 79 de proiecte de dezvoltare de wordnet-uri în peste 50 de limbi şi multe din ele urmăresc principiile şi metodele proiectului BalkaNet. Prin amploarea mondială a mişcării „wordnet”, prin volumul de resurse umane şi financiare implicate, acest domeniu al lexicografiei computaţionale îşi găseşte greu vreun rival în orice alt domeniu al informaticii.

De la finalizarea proiectului BalkaNet, dezvoltarea wordnet-ului românesc a continuat sub directă supervizare a prof. Tufiş, astfel că în prezent ontologia lexicală pentru limba română este printre cele mai mari din lume (conţine peste 59.348 de clase de echivalenţă sinonimică, peste 85.200 de sensuri ale celor peste 53.000 de cuvinte distincte şi mai mult de 138.500 de relaţii semantice şi lexicale, neluând în considerare relaţia de sinonimie care leagă literalii în synset-uri), poate cea mai completă sub aspect lexicologic (de exemplu puţine wordnet-uri ale altor limbi conţin definiţii, sau dacă le conţin, de multe ori sunt în limba engleză, importate direct din wordnet-ul original, dezvoltat la Princeton).

³ O schemă XML de codificare este o specificaţie formală a unui limbaj de adnotare textuală (un limbaj de programare) pentru care fiecare element de adnotare are un context de utilizare şi o semantică riguros descrise.

⁴ Proiectul a avut ca obiectiv crearea colecţiilor lexicale de tip wordnet pentru 5 limbi din zona balcanică: bulgară, greacă, română, sârbă, turcă, aliniate la nivel de concept cu Wordnet-ul englezesc.

⁵ www.wordnet.com

⁶ www.globalwordnet.org/

Achiziția automată de cunoștințe lexicale; alinierea textelor paralele la nivel de propoziție și cuvânt

Cercetările prof. Dr. Dan Tufiș în domeniul achiziției automate a cunoștințelor lexicale datează din anii 1997-1998, când domnia sa a dezvoltat un prim model inductiv, ce se baza pe echivalențele de traducere implicite existente între două texte reprezentând traduceri reciproce (bitext). În anii 2000-2002 au apărut și primele rezultate competitive care au dovedit că modelele statistice de identificare a co-ocurențelor cros-linguale constituie o metodă robustă de aliniere lexicală a cuvintelor unui bitext, de extragere automată a dicționarelor bilingve și, mai departe, de construire a modelelor de traducere (coloana vertebrală a unui sistem de traducere statistică). Problema corectitudinii alinierii lexicale a fost identificată ca una din ștrangulările tehnologice ale progresului în traducerea automată, astfel încât în anul 2003, în cadrul Conferinței Asociației Nord-Americane de Lingvistică Computațională (NAACL-2003) de la Edmonton, a fost organizată o competiție pe această problemă, pentru mai multe perechi de limbi, printre care și engleză-română.

Prof. Tufiș își mobilizează echipa și, în mai puțin de 2 săptămâni, reușesc să adapteze sistemul lor de aliniere, TREQ-AL, la cerințele organizatorilor, în două variante diferite. Sistemele ICIA⁷ ocupă primele două locuri, devansând competitori celebri, cum ar fi: XEROX Research Europe (XRCE-locurile 3, 5, 6, 7), Universitatea din Alberta, Canada (Proalign-locul 4), Universitatea din Montreal, Canada (Ralign-locul 8), Universitatea Carnegie Melon, Institutul de Tehnologii ale Limbii, USA (BiBr – locurile 9, 10, 11), Universitatea din Minnesota, Duluth, USA (UMD – locurile 12, 13), MIT Research Corporation, USA (Fourday – locul 14).

Doi ani mai târziu, prof. Tufiș și echipa sa propun un al doilea sistem, numit COWAL, care combinând mai multe soluții de aliniere obținute independent, se dovedește a fi din nou cel mai performant la următoarea competiție, organizată de data aceasta de către Asociația de Lingvistică Computațională (ACL), în iunie 2005 la Ann Arbor, Michigan. În competiția din SUA au fost înscrise în concurs 37 de sisteme de la universități și companii celebre (ISI-Universitatea din California, Universitatea din Maryland, Microsoft Research, Carnegie Mellon etc).

Din anul 2006, limba română este prezentă și în competițiile europene **CLEF**, care testează o gamă largă de problematici de inginerie lingvistică. Prof. Tufiș participă împreună cu doctoranzii săi la competițiile CLEF2006, ACL-SEMEVAL2007, CLEF 2007, CLEF2008, CLEF 2009 și CLEF 2010 (pe care de altfel, cu excepția competițiilor din 2008 și 2010, le câștigă, la concurență cu unele dintre cele mai repute centre de cercetare sau companii de software din lume). În anul 2006 colectivul prof. Dan Tufiș elaborează un sistem de întrebare-răspuns în univers de discurs deschis (web) cross-lingual (întrebarea în limba română, răspunsul căutat în documente de limbă engleză). În anul 2009, pentru prima oară, rezultatele competiției pentru sisteme de întrebare-răspuns în limbaj natural (**CLEF-ResPubliQA**) au putut fi comparate interlingual, întrucât întrebările de test (500) au fost aceleași în 7 limbi (bulgară, engleză, franceză, germană, italiană, română și spaniolă) răspunsurile trebuind a fi căutate în corpusul paralel al legislației europene „Acquis Communautaire”, disponibil în 22 din limbile oficiale ale Uniunii Europene. Sistemul realizat în colectivul coordonat de prof. Tufiș a câștigat din nou detașat, cu cel mai ridicat scor pe toate limbile, devansând toate celelalte 43 de sisteme competitori.

Dezambiguizarea semantică automată

Dezambiguizarea semantică (WSD – *Word Sense Disambiguation*) este o altă problemă cheie în traducerea automată. În ipoteza identificării din context a sensului cuvântului polisemantic din limba sursă, relațiile de echivalență semantică interlinguale codificate de wordnet-urile multilingve de tip BalkaNet permit identificarea exactă a cuvântului potrivit de traducere în limba țintă. Cu cât distincțiile între sensuri sunt mai fine, cu atât este mai dificilă însă rezolvarea problemei WSD. Metodologia

⁷ Institutul de Cercetări în Inteligență Artificială al Academiei Române, cunoscut în comunitatea internațională sub sigla RACAI (*Romanian Academy Centre for Artificial Intelligence*).

dezvoltată de prof. Tufiş pentru rezolvarea problemei WSD în texte paralele este printre cele mai avansate din lume. Sistemul dezvoltat⁸ a fost principalul mijloc de validare a corectitudinii semantice a wordnet-urilor dezvoltate în cadrul proiectului BalkaNet, fiind utilizat pentru toate cele 6 limbi ale proiectului. Mai multe lucrări publicate şi conferinţe invitate au tratat pe larg diferite aspecte conceptuale, algoritmice sau noi dezvoltări în dezambiguizarea automată a sensurilor cuvintelor în texte paralele.

WEB Semantic şi servicii web de prelucrare a limbajului natural

Una dintre direcţiile cele mai pregnante ale cercetării actuale în tehnologiile limbajului se încadrează domeniului web-ului semantic. Cercetările în această direcţie a prof. Dan Tufiş s-au focalizat pe problematica serviciilor web şi a aplicaţiilor complexe bazate pe prelucrări distribuite geografic. Începând din anul 2006, el coordonează activitatea de implementare a unei platforme de calcul pentru web-ul semantic, ce asigură servicii web lingvistice pentru limba română şi limba engleză. De curând, în colaborare cu Universitatea Marc Bloch din Strasbourg, serviciile ICIA asigură şi prelucrarea limbii franceze. Această platformă permite accesul de la distanţă la majoritatea instrumentelor şi resurselor dezvoltate de ICIA în ultimii 15 ani.

Anul acesta, pe situl Institutului de Cercetări pentru Inteligenţă Artificială a fost lansat Portalul Românesc de Tehnologii Lingvistice (relate.racai.ro) ce oferă acces liber la majoritatea instrumentelor şi resurselor lingvistice pentru limba română dezvoltate în proiectele de cercetare din ultimii ani (DruKoLa, CoRoLa, ReTeRom, Robin, Presidency Translation Kit, ELRC, Marcell).

Traducerea automată în şi din limba română

Majoritatea rezultatelor obţinute după anul 2000 în domeniul lingvisticii corpusului au permis lansarea unor cercetări sistematice privind traducerea automată din şi în limba română, începând cu anul 2003. Între anii 2005 şi 2010 au fost elaborate trei teze de doctorat în acest domeniu sub îndrumarea prof. Dan Tufiş, finalizate cu prototipuri funcţionale de sisteme de traducere din limba engleză în limba română şi invers, clădite pe modele, metode, algoritmi şi resurse lingvistice create sub coordonarea dumnealui în perioada anilor 1995-2008. Sistemele, antrenate pe resursele multilinguale, demonstrează un mare grad de generalitate şi reprezintă premise solide pentru realizarea unui sistem profesional de traducere automată, uşor adaptabil la orice pereche de limbi.

În anul 2018, în colaborare cu TILDE din Letonia, pornind de la sistemul eTRANSLATION dezvoltat de Directoratul pentru Traduceri al Comisiei Europene, o versiune îmbunătăţită cu pre- şi post-procesări neuronale a fost intens antrenată pe date paralele (RO-EN) şi date monolingve (RO) însumând peste un milion şi jumătate de propoziţii. Datele de antrenare sistemului de traducere au fost colectate din diferite domenii publice stocate şi prelucrate cu instrumentele implementate de echipa condusă de prof. Dan Tufiş. Sistemul de traducere, antrenat este disponibil în prezent, pe portalul RELATE (relate.racai.ro).

Implicarea în cercetarea naţională şi europeană

În cursul anilor, profesorul Dan Tufiş a coordonat ori a participat în 41 de proiecte de cercetare internaţionale, iar după anul 1997, cel al primirii în Academie, a fost responsabilul a 21 teme anuale de cercetare, înscrise în planul de cercetare al Academiei Române. A fost implicat direct în configurarea multor programe naţionale. De exemplu, a fost directorul Programului Naţional INFOSOC «Strategii şi soluţii pentru Societatea Informaţională – Societatea Cunoaşterii în România», între 2001-2002, o continuare a programului fundamental al Academiei Române dedicat problemelor strategice ale prelucrării automate a limbii române. A coordonat colectivul de elaborare al subprogramului “Tehnologia Limbajului” din strategia şi planul naţional de cercetare al MCT “Societatea Informaţională” (2005). Din aprilie 2009 face parte din Comisia de monitorizare a proiectelor de

⁸ WSDTool.

cercetare și diseminare a rezultatelor, organism al Consiliului Național al Cercetării Științifice Universitare.

Din 1994 este expert UNESCO în domeniile inteligenței artificiale, lingvisticii computaționale și al programării funcționale (LISP). În 2001 a fost ales în comisia guvernamentală de experți UNESCO responsabili de elaborarea proiectului de recomandări privind “Promovarea multilingvismului și a accesului universal în spațiul informațional”.

În perioada ianuarie 1997-ianuarie 1999 a fost membru al *Advisory Board* al Asociației Europene de Lingvistică Computațională, cea mai importantă asociație profesională în domeniul Prelucrării Limbajului Natural.

În anul 2001 a înființat Comisia de Informatizare pentru Limba Română în subordinea Secției de Știință și Tehnologia Informației, al cărui președinte este de la înființare. Această comisie constituie un organism consultativ și un forum pentru discutarea priorităților și problematicilor specifice prelucrării automate a limbii române. Ca organism executiv, mult mai larg, a fost înființat în același an Consorțiul Național pentru Informatizarea Limbii Române (CONSILR), cu scopul de a disemina contribuțiile publice (resurse și instrumente de prelucrare a limbii române) dar și ca forum de discuții cu toți partenerii interesați. Conferința CONSILR a ajuns în anul 2019 la cea de a 14-a ediție.

A participat adesea, ca expert al CE, la pregătirea planurilor de finanțare a cercetării europene în tehnologia limbajului și, desigur, în repetate rânduri, la evaluarea proiectelor propuse spre finanțare.

Este membru al Language Resource Board (National Contact Point) organism de conducere al proiectului ELRC (European Language Resource Coordination) și al European Language Grid (reprezentant național).

Contribuții la formarea unei școli românești de inginerie lingvistică

Profesorul Dan Tufiș este creatorul școlii românești de inginerie lingvistică, cu largă recunoaștere internațională. Un impresionant număr de tineri colaboratori din colectivele pe care le-a condus și-au obținut doctorate și lucrează actualmente în importante universități sau institute de cercetare din lume. Numeroși studenți ai cursurilor de master ale Universității București și „A.I. Cuza” din Iași sau la Școlile de Vară Eurolan au obținut titlul de doctor sau sunt doctoranzi la universități de prestigiu.

Directorul și academicianul Dan Tufiș a reușit să creeze în institutul pe care l-a condus timp de peste 20 ani un climat de cercetare incitant, să formeze și să păstreze în jurul său un colectiv, mereu reînnoit, pe care l-a format într-un spirit de competitivitate, care a stimulat înalta performanță. Credem că prof. Dr. Dan Tufiș este un exemplu demn de urmat asupra modului în care un conducător de colectiv reușește să implice tinerii cercetători în toate etapele activității de cercetare, de la aprofundarea și dezvoltarea propriilor idei, la elaborarea de lucrări științifice, de la analiza și evaluarea unor articole științifice, până la preluarea responsabilității unor componente importante în proiecte naționale și internaționale de cercetare, tinerii merituos recunoscându-i aceste calități și faptul că întotdeauna au fost promovați deschis, pe criterii exclusiv profesionale.

**Acad. Florin Ghorghe Filip,
Președintele Comisiei de Laudatio**



HOTĂRÂRE DE SENAT din 12.09.2019

Senatul Universității Agora din municipiul Oradea hotărăște, cu unanimitate de voturi:

Articol unic. Se acordă titlul și diploma "Doctor Honoris Causa al Universității Agora din municipiul Oradea" domnului profesor universitar doctor Dan TUFÎȘ, membru titular al Academiei Române, director al Institutului de Cercetări în Inteligență Artificială al Academiei Române. Ceremonia festivă de Laudatio și decernare a diplomei va avea loc în 18.10.2019.

Senatul Universității Agora:

1. Prof. univ. dr. Adriana MANOLESCU
2. Prof univ. dr. Elena - Ana IANCU
3. Conf.univ.dr. Ioan STAN
4. Conf.univ. dr. Casian-Valentin BUTACI
5. Lect.univ.dr. Radu Gheorghe FLORIAN
6. Stud. Patricia Carmen POLEAC
7. Stud. Loredana Gabriela MOCANU

Președinte Senat
Prof. univ. dr. Adriana Manolescu

