



Universitatea
Transilvania
din Braşov

INTERDISCIPLINARY DOCTORAL SCHOOL

Faculty of Letters

Claudius-Marian TEODORESCU

Modelling and developing a digital lexicon for the Romanian language

SUMMARY

Scientific supervisor

Prof. dr. Mihaela GHEORGHE

BRAŞOV, 2025

Introduction

Is it possible to elaborate „complete reference system, consisting of a diachronic and, as much as possible, exhaustive presentation of the Romanian language, since the beginning of the Romanian writings until today, based upon phonetic, morphologic, and lexical material that is provided by the texts” (Costinescu 1981: 91)?

A first endeavour for elaboration of such „complete reference system” was initiated at the Linguistic Institute of Bucharest, under the coordination of Ion Dănilă and Radu Michăescu, and lasted for the period 1979 - 1996. A part of this system was developed at the "Sextil Pușcariu" Institute of Linguistics and Literary History from Cluj-Napoca. It was an immense effort, both theoretical, and practical, which abundantly show the necessity of such a system for the Romanian language; this effort was sustained by a big enthusiasm regarding the possibility of elaborating and using such resources, with the help of computers. The teams that were involved in the project were interdisciplinary ones, including linguists and IT specialists, and the coordination was provided by linguists.

After 30 years since the above-mentioned initiative, we can and we have to make an analysis of what was accomplished in Romanian philologic and linguistic research meanwhile, in this regard. The specialized literature indicates that many valuable digital resources and instruments were created during this period, which could be united in a system, in order to serve the assembly of goals correlated with the „complete reference system”. One can also found a huge progress of the possibilities offered by the computers, of which we can mentioned that some would have been hard to think of 30 years ago.

The field of information science borrowed, for the elaboration of the computer programming languages, many of the concepts of the field of syntactic analysis of the natural languages. In this way, it was a huge transfer of knowledge from linguistics to information science.

Considering what was mentioned above, armed with the methodologies and technologies provided by the information science, one can assume that the information science can nowadays offer to the Romanian philology and linguistics the instruments needed for accomplishing the desiderata that are specific to the mentioned initiative, by using an interdisciplinary approach, in order to better use the computers for developing digital philologic and linguistic resources.

Research objectives

The general objective of the current work is modelling and developing a general digital lexicon for the Romanian language. Based upon this general objective, I have defined two main objectives, as follows:

- modelling and developing a general digital bibliography for the Romanian language;
- modelling and developing a general digital corpus for the Romanian language.

Research methodology

The current work aims to review the Romanian research projects that are relevant as to the elaboration of a digital bibliography, a digital corpus, and a digital lexical, respectively.

For this, the work includes the needed steps for elaboration and publication of a digital resource, by using standardised solutions that were tested by the author. These steps are presented in a chapter for each of the three topics: digital bibliography, digital corpus, and digital lexicon.

For each of these three chapters, the work includes standardized modalities for data and metadata encoding, data models, various correlations, as well as the results that can be obtained.

All the research approaches are done from a philologic perspective. Briefly, the difference between the philologic perspective and the computational perspective, is as follows: the computational perspective considers that a 80% or 90% success rate of a process is satisfactory; the philologic perspective constantly implies a 100% success rate, namely to have a text that is totally and correctly transcribed, or structural, morphological, syntactical, or semantical annotated, etc.

Another facet of the approach presented in this work consists of recovering, as it was possible, of the “classical” methods and concepts for every of the three fields that are investigated (bibliography, corpus, lexicon), their transposition in the new associated fields (digital bibliography, digital corpus, digital lexicon), as well as adding of new methods and concepts that are specific for working with computers. This synthesis was done very rigorously, in order to preserve the “classical” knowledge thesaurus.

The structure of the work

The second chapter contains the current state of the art in Romania, for the fields: digital bibliographies, digital corpora, and digital lexicons, respectively. In order to analyze this state of the art, I have chosen eight sets of metadata about every research project, with each set dealt with in a separate section, so that to provide a comprehensive and consistent presentation and analysis.

The third chapter presents a methodology for elaboration and publication of philological and linguistic digital resources, starting with their legal status and continuing with data management models, data and metadata models, stages of elaboration, linking with other digital resources, storage, publication, up to examples of software products for digital resources. All these aspects were necessary to be dealt with, because the resources this work proposes (bibliography, corpus, lexicon) are digital. All the elements that are presented in this chapter were applied during the author’s scientific activity for elaboration and publication of digital resources. The chapter concludes with a section dedicated to the impact of research results, and with a simple methodology for elaboration of a *National Citation Index*.

With the fourth chapter, my intention was to review the main modalities of annotation the works referred to into a bibliography, with examples of metadata sets for seven digital manuscript catalogues. I also exposed the idea, which is not new at all, but it is presented in a new light, of a *General Bibliography for the Romanian Language*, which has to be developed in two directions: a

bibliographic one, for which I have presented modalities and examples of semantic annotation designated for obtaining a semantic graph of all the concepts treated withing the works referred to in a bibliography; and a *scientometric* one, for which I have presented the data model used for a similar initiative. Also, I have exposed few examples, from Romania or from abroad, of thematic bibliographies that are similar to *General Bibliography for the Romanian Language*.

The fifth chapter comprises the brief presentation of a methodology for obtaining a digital corpus, followed by the detailed presentation of 16 digital resources, for elaboration of which I have used this very methodology. The series of practical applications of the concept of *digital corpus* is continued with the presentation of a special corpus, the *General Corpus for the Romanian Language*, along with a short argumentation regarding the necessity of such a corpus.

The sixth chapter includes a proposition about a methodology for modelling and developing a digital lexicon. By lexicon, one can understand the component part of the grammar „that contains lexical entries which are associated with morphological, syntactical, and respectively, semantical information” (Atkins 1994: 183). As an expected continuation of the propositions for a *General Bibliography for the Romanian Language*, and a *General Corpus for the Romanian Language*, respectively, this chapter contains some arguments in favour of the proposition for elaboration of a *General Lexicon for the Romanian Language*, ideas about the input data, the lexicon data model, the description of the possibility of linking the lexicon with other similar digital resources, elaborated for other languages, as well as other ways for using it.

The seventh and, respectively, eighth chapter contain detailed explanations about the concepts that are new for the Romanian philology and linguistics, which are mentioned in this work.

Conclusions, contributions, and further research directions

The contributions brought by this work have both a theoretical, given the proposed interdisciplinary approach, and a practical nature. Most of the theoretical contributions are accompanied by methodologies for applying them. One can find below the main contributions:

- A methodology, which was tested by the author, for elaboration and free publication in Internet of digital resources.
- A methodology for elaboration of a *National Index of Citations* for the field of language sciences.
- Argumentations and examples for using free methods and resources in the field of scientometry, applied for digital bibliographies, for the elaboration of the *National Index of Citations*, mentioned above.
- Methodology for elaboration and publication in Internet, for anyone to consult, of a *General Bibliography for the Romanian Language*, and of a concept graph for the language sciences, for the Romanian language.
- Methodology for elaboration and publication in Internet, for anyone to consult, of a *General Corpus for the Romanian Language*.

- Methodology for elaboration and publication in Internet, for anyone to consult, of an ontology for Romanian text editions.
- Methodology for elaboration and publication in Internet, for anyone to consult, of a *General Lexicon for the Romanian Language*.
- Brief methodology for elaboration and publication in Internet, for anyone to consult, of an ontology for Romanian language's grammar.
- A first version of a typology for textual indexes.
- A detailed description of the components of a textual index.
- An interdisciplinary perspective regarding the digital textual indexes.
- A codification for digital textual indexes.
- Software products: static indexing engine; static search engine, to be used in a web browser.

Out of the above-mentioned contributions, I consider that the following activities can constitute research directions for a first stage of elaboration of the digital lexicon for the Romanian language:

- Elaboration and publication in Internet, for anyone to consult, of all the above-mentioned ontologies.
- Further developing and refining the mentioned software products.
- Elaboration of a *National Index of Citations* for the field of language sciences.

Bibliografie

- Abalkina, Anna. 2021. „Detecting a network of hijacked journals by its archive”. În *Scientometrics*, 126, 7123–7148. <https://doi.org/10.1007/s11192-021-04056-0>, accesat la 16.11.2024.
- Abalkina, Anna. 2021. „Publication and collaboration anomalies in academic papers originating from a paper mill: evidence from Russia”. În *arXiv*, 2112.13322, preprint. <https://arxiv.org/abs/2112.13322>, accesat la 16.11.2024.
- Abalkina, Anna. 2023. „Challenges posed by hijacked journals in Scopus”. În *Journal of the Association for Information Science and Technology*, 75(4):395-422. <https://doi.org/10.1002/asi.24855>, accesat la 16.11.2024.
- Academia Română. 1913-1949. *Dicționarul limbii române* [DA], ediție digitală anastatică în *CLRE. Corpus lexicografic românesc electronic*, 2021-2022, <https://dlr1.solirom.ro/>, accesat la 10.09.2023.
- Adams, Michael (ed.). 2010. *“Cunning passages, contrived corridors”: Unexpected Essays in the History of Lexicography*. Milano: Polimetrica. ISBN: 978-88-7699-208-7. https://ht.ac.uk/downloads/Hidden_Passages_-_HTOED_Material.pdf, accesat la 17.06.2022.
- AFLR. 2017. Analiza fonetică a limbii române: studiu și aplicații informatice. <http://softwinresearch.ro/index.php/ro/proiecte/aflr>, accesat la 29.07.2022.
- Ahmadi, Sina, Mathieu Constant, Karën Fort, Bruno Guillaume, John P. McCrae. 2021. „Convertir le Trésor de la Langue Française en Ontolex-Lemon : un zeste de données liées”. În *Journées LIFT 2021 - Linguistique informatique, formelle et de terrain*, decembrie 2021, Grenoble, France. <https://inria.hal.science/hal-03463294>, accesat la 05.11.2023.
- American Society for Indexing. 2015. *Best Practices for Indexing*, American Society for Indexing (ASI), 76 p. <https://asindexing.org/wp-content/uploads/2014/04/ASI-BestPracticesIndexing2015-linked-updated-copyright-v3-02-26-16.pdf>, accesat la 01.03.2024.
- Avram, Mioara. 2002. *Ortografie pentru toți: 30 de dificultăți, ediția a doua*, Litera Internațional: București - Chișinău, 286 p. ISBN 9975-904-75-0.
- B. T. S. Atkins, A. Zampolli (editori). 1994. *Computational Approaches to the Lexicon*, Oxford University Press: Oxford, New York, 479 p. ISBN 0-19-823979-3.
- Badia, Antonio, Daniel Lemire. 2011. „A Call to Arms: Revisiting Database Design”. În *Computing Research Repository*, SIGMOD Record 40 (3), 61-69. DOI 10.48550/arXiv.1105.6001. <https://arxiv.org/abs/1105.6001>, accesat la 26.10.2023.
- Baker, Collin F., Charles J. Fillmore, John B. Lowe. 1998. „The Berkeley FrameNet Project”. În *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, volum 1, 86-90. Association for Computational Linguistics. <https://aclanthology.org/P98-1013.pdf>, accesat la 02.07.2022.
- BalkaNET. 2004. BALKANET: A Multilingual Semantic Network for Balkan Languages. <https://dblab.upatras.gr/balkanet/>, accesat la 03.08.2022.

- Barborică, Elena, Mirela Teodorescu, Liviu Onu. 1972. *Introducere în filologia română. Orientări în tehnica cercetării științifice a limbii române*, Editura didactică și pedagogică: București, 144 p.
- Barbu, Ana-Maria, Emil Ionescu, Verginica Barbu Mititelu. 2006. „Romanian Valence Dictionary in XML Format”. În *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genova, Italia, mai 2006, p. 2467-2470. https://lrec-conf.org/proceedings/lrec2006/pdf/295_pdf.pdf, accesat la 07.08.2022.
- Barbu, Ana-Maria, Emil Ionescu. 2008. Proiectarea unui dicționar de valențe verbale destinat prelucrării limbajului natural. *Studii și cercetări lingvistice*, LIX, 2008, nr. 2, București, p. 295-306.
- Barbu, Ana-Maria. 2008. „Romanian Lexical Data Bases: Inflected and Syllabic Forms Dictionaries”. În *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, mai 2008, Marrakech, Maroc, p. 1937-1941. http://lrec-conf.org/proceedings/lrec2008/pdf/495_paper.pdf, accesat la 30.07.2022.
- Barbu, Ana-Maria. 2018. „Dictionary of Verbal Contexts for the Romanian Language”. În *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, p. 411-422. Ljubljana: Ljubljana University Press, Facultatea de Arte. <https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202018/118-4-2912-1-10-20180820.pdf>, accesat la 07.08.2022.
- Barbu, Ana-Maria, Verginica Barbu Mititelu, Cătălin Mititelu. 2022. „Aligning the Romanian Reference Treebank and the Valence Lexicon of Romanian Verbs”. În *Proceedings of the Thirteenth International Conference on Language Resources and Evaluation (LREC'22)*, 20-25 iunie 2022, Marsilia, Franța.
- Barbu, Violeta. 1988. Probleme și orientări în editarea literaturii medievale. *Limba Română*, an XXXVII, nr. 2, martie-aprilie, p. 143-153.
- Barbu Mititelu, Verginica, Dan Cristea, Ruxandra Cosma. 2016. *Corpus of Contemporary Romanian. Annotation Levels. CoRoLa and DRuKoLa*. Manuscris, <https://profs.info.uaic.ro/~dcristea/Talks/CoRoLa&DRuKoLa&LLOD.pptx.pdf>, accesat la 25.08.2022.
- Barbu Mititelu, Verginica, Dan Tufiș, Elena Irimia. 2018. „The Reference Corpus of the Contemporary Romanian Language (CoRoLa)”. În *Proceedings of LREC 2018*, European Language Resources Association (ELRA), Japonia, mai 2018, p. 1178-1185. <https://aclanthology.org/L18-1189.pdf>, accesat la 25.08.2022.
- Barend Mons, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino da Silva Santos, Mark D. Wilkinson, 2017. *Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud*. „Information Services & Use”, 2017, vol. 37, no. 1, p. 49-56. DOI: 10.3233/ISU-170824. <https://content.iospress.com/articles/information-services-and-use/isu824>, accesat la data de 01.11.2022.
- BDD. 2014. Baza de date „Diacronia”. <https://diacronia.ro/ro/indexing>, accesat la 26.08.2022.

- Becker, Devin, Evan Williamson, Olivia Wikle. 2020. „CollectionBuilder-CONTENTdm: Developing a Static Web ‘Skin’ for CONTENTdm-based Digital Collections”. În *code4lib*, nr. 49, 10.08.2020. <https://journal.code4lib.org/articles/15326>, accesat la 26.10.2024.
- Beckett, David, Tim Berners-Lee, Eric Prud'hommeaux, Gavin Carothers. 2014. *RDF 1.1 Turtle. Terse RDF Triple Language*. World Wide Web Consortium (W3C), <https://w3.org/TR/turtle/>, accesat la 10.07.2023.
- Bejinariu, Silviu-Ioan, Ramona Luca, Vasile Apopei, Florin Iftene. 2022. *Editarea textelor dialectale folosind transcrierea fonetică specifică limbii române*, Anuar de lingvistică și istorie literară, tom LXII, 2022 - In memoriam Stelian Dumistrăcel, <http://alil.academiaromana-is.ro/archives/3721>, accesat la 22.06.2023.
- BIBNAT. 2009. Biblioteca Națională a României. <https://bibnat.ro>, accesat la 07.09.2022.
- Bidu-Vrânceanu, Angela. 2000. *Lexic comun, lexic specializat*. București: Editura Universității din București, p. 1975, ISBN 978-973-575-486-0.
- Bidu-Vrânceanu, Angela, Cristina Călărășu, Liliana Ionescu-Ruxăndoiu, Mihaela Mancaș, Gabriela Pană Dindelegan. 2005. *Dicționar de științe ale limbii*, ediția a doua, București, Nemira & Co., ISBN 973-569-762-9, 606 p.
- BME. 2010. *Bibliografia Mihai Eminescu*. <https://biblacad.ro/bme.html>, accesat la 03.09.2022.
- Bocșa, Minerva. 1974. „Technique: Letters with Variable Values and the Mechanical Inflection of Rumanian Words”. În *American Journal of Computational Linguistics*, decembrie 1974, p. 38-52. <https://aclanthology.org/J74-3003.pdf>, accesat la 13.08.2022.
- Bocșa, Minerva. 1974. „Un algoritm pentru generarea cuvintelor limbii române”. *Revista de analiză numerică și teoria aproximației*, vol. 3, fascicula 1, p. 35-51. <https://ictp.acad.ro/ranta-ro-72-74/journal/article/view/40/40>, accesat la 13.08.2022.
- Bodoc, Alice. 2021. „Eterogenitatea formantului *cum* în limba română veche”. În *Philologica Jassyensia*, an XVII, nr. 1 (33), 2021, p. 43–56.
- Boian, Elena, Constantin Ciubotaru, Svetlana Cojocar, Alexandru Colesnicov, Ludmila Malahov, Mircea Petic. 2011. „Creation and Development of the Romanian Lexical Resources”. În *Proceedings of Recent Advances in Natural Language Processing*, RANLP 2011, 12-14 septembrie 2011, Hissar, Bulgaria, p. 678-685.
- Boldea, Iulian. 2019. *Vatra-dialog cu Alexandru Gafton: „Rostul științei este aflarea și înțelegerea realității întocmai cum este ea”*. *Vatra*, nr. 10-11/2019, p. 20-23. <https://revistavatra.org/2020/02/02/vatra-dialog-cu-alexandru-gafton-rostul-stiintei-este-aflarea-si-intelegerea-realitatii-intocmai-cum-este-ea/>, accesat la 28.08.2022.
- Brewer, Charlotte. 2013. „OED Online Re-launched: Distinguishing Old Scholarship from New”. *Dictionaries: Journal of the Dictionary Society of North America*, vol. 34: 101–126. doi:10.1353/dic.2013.0002. https://oed.hertford.ox.ac.uk/wp-content/uploads/2019/07/Brewer_2013a.pdf, accesat la 02.06.2022.
- Britannica. 1999. <https://britannica.com/topic/generative-grammar>, accesat la data de 20.08.2023.

- BNRP. 2010. Bibliografia Națională Retrospectivă a Periodicelor (1790-1952). <https://biblacad.ro/bnr/ppr.php>, accesat la 05.09.2022.
- Brut, Mihaela, Dumitru Irimia, Oana Panait. 2006. „Generare de concordanțe pentru dicționarul limbajului poetic eminescian”. În *Lucrările atelierului Resurse lingvistice și instrumente pentru prelucrarea limbii române*, Iași, p. 113-118.
- Busuioc, Monica, Ana-Maria Barbu, Victor Celac, Florin Vasilescu. 2008. *Baza de texte - instrument esențial pentru informatizarea activității lexicografice*, manuscris.
- Busuioc, Monica, Dan Caragea. 2018. ROMTEXT – a fundamental instrument for the new edition of the Dictionary of the Romanian Language. RRL, LXIII, 4, p. 409–414.
- Busuioc, Monica, Dan Caragea. 2019. *ROMTEXT: Flux de obținere și tratare a textelor*. SCL, LXX, nr. 1, București, p. 118–123.
- Bursuc, Alina Mihaela. 2013. „Pentru un dicționar de concordanțe ale numelor de rudenie”. În Ofelia Ichim; Luminița Botoșineanu, Daniela Butnaru, Marius-Radu Clim, Florin-Teodor Olariu, Elena Tamba (ed.), *Cultură și identitate românească - Tendințe actuale și reflectarea lor în diaspora*, 59-66. Editura Universității „Alexandru Ioan Cuza”: Iași.
- Buturugă, Maria, Cornelia Lepădatu. 2011. „Bibliografia retrospectivă a cărții românești. Bibliografie on-line”. Comunicare în cadrul Colocviului Internațional „*Biblioteca științifică în societatea informațională - realizări și perspective*”, cu ocazia *Târgului Internațional de Carte* organizat de Biblioteca Științifică Centrală „A. Lupan” a Academiei de Științe a Moldovei, 10-13 octombrie 2011, Chișinău. <https://biblacad.ro/ASM2011.pdf>, accesat la 02.09.2022.
- Byck, Jacques. 1957. *Sarcinile actuale ale filologiei române*, în *Limba română*, VI (1957), nr. 1, p. 11 ș. urm. <https://dspace.bcu-iasi.ro/handle/123456789/1638>, accesat la 23.07.2024.
- Cambose, Astrid. 2018. „Fonduri arhivistice inedite: răspunsurile la Chestionarul lingvistic B.P. Hasdeu”. În Ofelia Ichim, Luminița Botoșineanu, Daniela Butnaru, Marius-Radu Clim, Veronica Olariu (editori), *Clasic și modern în cercetarea filologică românească actuală*, Iași: Editura Universității „Alexandru Ioan Cuza”, p. 393-404. ISBN: 978-606-714-506-9. <https://diacronia.ro/en/indexing/details/V4921/pdf>, accesat la 22.08.2022.
- Capadisli, Sarven, Tim Berners-Lee, Ruben Verborgh, Kjetil Kjernsmo (ed.). 2020. *Solid Protocol*. W3C Solid Community Group. <https://solidproject.org/TR/protocol>, accesat la 01.08.2023.
- CATCR. 2016. *Catalogul on-line al cărților rare din Biblioteca Academiei Române*. <https://biblacad.ro/cercetare/2016/ManuscriseCarteRara/CartiRarePrezentare.pdf>, accesat la 03.09.2022.
- Cazacu-Davidescu, Lucia Mihaela. 2008. *Index tematico-cronologic*, în *Arheologia Moldovei*, vol. 31 (2008), p. 365-368. DOI: <https://doi.org/10.11588/amold.2008.0.30541>, accesat la 08.10.2023.
- CETRV. 2013. *Corpus electronic al textelor românești vechi (1521 – 1640)*. https://web.archive.org/web/20220120141242/http://media.lit.uaic.ro/?page_id=3914, accesat la 26.08.2022.

- Cherata, Sanda, Teodor Vușcan, Emma Tămăianu. 1994-1995. SILEX - Un sistem lexico-morfologic computerizat pentru analiza textelor românești. *Dacoromania serie nouă*, I: 201-212.
- Cherata, Sanda, Teodor Vușcan, Emma Tămăianu. 1996-1997. Silex - Funcțiile de lematizare și de generare a paradigmelor. *Dacoromania serie nouă*, II: 273-285.
- Cherata, Sanda. 1997. „CONCORD - Software System for Concordances of Romanian Poetical Texts”. În Dan Tufiș, Poul Andresen (ed.), *Recent Advances in Romanian Language Technology*, București: Editura Academiei Române, p. 55-62. <https://racai.ro/external/static/awde/cherata.html>, accesat la 31.08.2022.
- Cherata, Sanda, Manuela Mihăescu. 2008. Modele formale de reprezentare a informațiilor lexicale și terminologice în proiectul *SIASTRO*. *Dacoromania*, serie nouă, XIII, nr. 2, Cluj-Napoca, p. 151–169.
- Chicago. 2017. Manualul de stil Chicago. Ediția a 17-a, marcă înregistrată a Universității din Chicago. Chicago și Londra: Tipografia Universității din Chicago.
- CIDOC-CRM. 2024. *Definition of the CIDOC Conceptual Reference Model*. Bekiari, Chryssoula, George Bruseker, Erin Canning, Martin Doerr, Philippe Michon, Christian-Emil Ore, Stephen Stead, Athanasios Velios (editori), versiunea 7.1.3, februarie 2024. https://cidoc-crm.org/sites/default/files/Documents/cidoc_crm_version_7.1.3.html, accesat la 06.07.2024.
- Ciubotaru, Constantin, Valentina Demidova, Tudor Bumbu. 2019. „Generation of the Romanian Cyrillic lexicon for the period 1967 - 1989”. În *Proceedings of the Fifth Conference of Mathematical Society of Moldova IMCS-55*, 28 septembrie - 1 octombrie 2019, Chișinău, Republica Moldova, p. 309-316. https://ibn.idsi.md/sites/default/files/imag_file/309-316_1.pdf, accesat la 13.07.2022.
- Ciubotaru, Constantin. 2021. „Backtracking algorithm for lexicon generation”. *Computer Science Journal Of Moldova*, vol. 29, nr. 1 (85): 135-152. [http://math.md/files/csjm/v29-n1/v29-n1-\(pp135-152\).pdf](http://math.md/files/csjm/v29-n1/v29-n1-(pp135-152).pdf), accesat la 13.07.2022.
- Cojocaru, Svetlana. 1996. „Lexicon român: instrumentar, implementare, utilizare”. În Dan Tufiș (editor), *Limba și tehnologie*, București: Editura Academiei Române, 37-39.
- Condurachi, Em., Corneliu Bălăcescu, George Buzdugan, Octavian Iliescu, Mihai Ioniță, Octavian Luchian, Bucur Mitrea, Constantin Preda (redactori). 1974. *Buletinul Societății numismatice române*, anii XLII - LXVI (1948 - 1972), nr. 96 - 120, p. 143. <https://dspace.bcucluj.ro/handle/123456789/13842>, accesat la 08.10.2023.
- CoNLL-U. 2014. <https://universaldependencies.org/introduction.html>, accesat la data de 19.08.2023.
- Constantin, Alexandru, Silvio Peroni, Steve Pettifer, David Shotton, Fabio Vitali. 2016. The Document Components Ontology (DoCO). *Semantic Web*, vol. 7, nr. 2, pag. 167-181. <https://content.iospress.com/articles/semantic-web/sw177>, accesat la 03.10.2024.
- CoRoLa. 2014. *Corpus of Contemporary Romanian Language (CoRoLa)*. <https://corola.racai.ro/>, accesat la 25.08.2022.
- Costinescu, Mariana. 1981. Codicele Voronețean. Ediție critică, studiu filologic și studiu lingvistic de Mariana Costinescu. București: Minerva.

Cristea, Dan, Gabriela Haja, Alex Moruz, Marius Răschip, Mădălin Pătrașcu. 2011. „Statistici parțiale la încheierea proiectului eDTLR–Dicționarul tezaur al limbii române în format electronic”. În R. Zafiu, C. Ușurelu, H. Bogdan Oprea (editori), *Actele celui de-al 10-lea Colocviu al Catedrei de limba română*, vol. I (Gramatică și fonologie, lexic, semantică, terminologie, istoria limbii române, dialectologie și filologie), București: Editura Universității din București, p. 213-224. ISBN: 978-606-16-0046-5.

https://profs.info.uaic.ro/~dcristea/papers/Cristea_Haja_Moruz_Raschip_Patrascu.pdf, accesat la 24.08.2022.

CRVIH. 2014. Cartea românească veche în Imperiul Habsburgic (1691-1830). Recuperarea unei identități culturale. <http://diam.uab.ro/index.php?s=10&p=28>, accesat la 06.09.2022.

Curteanu, Neculai, G. Holban, S. Negulescu, C. Păvăloi, I. Păvăloi, A. Todirașcu. 1996. „Bază de cunoștințe lingvistice pentru limba română”. În Dan Tufiș (editor), *Limba și tehnologie*, București: Editura Academiei Române, 101-108.

Cyganiak, Richard, David Wood, Markus Lanthaler. 2014. *RDF 1.1 Concepts and Abstract Syntax*. <https://w3.org/TR/rdf11-concepts/>, accesat la 10.07.2023.

Dănăilă, Ion, Radu Michăescu. 1980. Banca de date fono-morfo-semantică a limbii române (BANDASEM). *Limba română*, vol. XXIX, nr. 5: 471-476.

Dănăilă, Ion. 1993. Pentru un inventar general al lexicului limbii române. *Limba română*, vol. XLII, nr. 2: 61-68.

Dănăilă, Ion. 1996. „Primul proiect românesc de dicționar lingvistic computerizat. Dicționarul confruntativ de sinonime, de analogii și de asociații al limbii române (DCSAAs)”. În Dan Tufiș (editor), *Limba și tehnologie*, București: Editura Academiei Române, 177-179.

Di Giambattista, Chiara, Ivan Heibi, Silvio Peroni, David Shotton. 2022. „OpenCitations: an Open e-Infrastructure to Foster Maximum Reuse of Citation Data”. În *International Journal of Digital Curation*, 2022, vol. 17, nr. 1, 5 p. 2-5. <http://dx.doi.org/10.2218/ijdc.v17i1.818>, accesat la 30.11.2024.

Di Iorio, Angelo, Silvio Peroni, Fabio Vitali. 2009. „Towards markup support for full GODDAGs and beyond: the EARMARK approach”. În *Balisage: The Markup Conference 2009*, 11-14 august 2009. <https://balisage.net/Proceedings/vol3/html/Peroni01/BalisageVol3-Peroni01.html>, accesat la 15.10.2024.

Diaconescu Ștefan. 2006. „Crearea resurselor lingvistice cu ajutorul unui limbaj specializat”. În *Workshop on Linguistic resources and tools for Romanian language processing*, Iași.

Diaconescu Ștefan, Ionuț Dumitrașcu. 2007. „Complex Natural Language Processing System Architecture”. În Corneliu Burileanu, Horia-Nicolai Teodorescu (ed.), *Advances in Spoken Language Technology*, 228-240. București: Editura Academiei Române.

Diaconescu Ștefan, Cristi Ingineru, Felicia Codîrlașu, Monica Rizea, Oana Bulibașa. 2009. „General System for Normal and Phonetic Inflection”, în *Proceedings of the 5-th Conference on Speech Technology and Human-Computer Dialogue, 2009 SpeD '09*, 98-107. București: Editura Academiei Române.

- Diaz, Chris. 2018. „Using Static Site Generators for Scholarly Publications and Open Educational Resources”. În *code4lib*, nr. 42, 08.11.2018. <https://journal.code4lib.org/articles/13861>, accesat la 26.10.2024.
- Digiteca Arcanum. 2022. *Bază de date cu periodice digitalizate*. <https://adt.arcanum.com/ro/>, accesat la 08.02.2023.
- DLR¹. 1913 - 2005. *Dicționarul limbii române*. <https://dlr1.solirom.ro/>, accesat la 24.07.2023.
- DOI Foundation. 2024. *DOI Handbook*. <https://doi.org/10.1000/182>, accesat la 13.02.2024.
- Dragomirescu, Adina. 2020. Poziția subiectului în interogative directe parțiale și relative în limba română veche. *Studii și cercetări lingvistice*, LXXI, iulie - decembrie 2020, 189-198.
- ELTeC. 2017. European Literary Text Collection. <https://distant-reading.net/>, accesat la 23.08.2022.
- ELTeC-rom. 2017. Romanian Novel Collection (ELTeC-rom). <https://github.com/COST-ELTeC/ELTeC-rom>, accesat la 23.08.2022.
- Farrar, Scott, William D. Lewis. 2007. „The GOLD Community of Practice: An Infrastructure for Linguistic Data on the Web”. În *Language Resources and Evaluation*, vol. 41, nr. 1 (feb., 2007), pp. 45-60. <https://doi.org/10.1007/s10579-007-9016-x>, accesat la 02.04.2025.
- Goddard, Cliff, Anna Wierzbicka. 2016. *Words and Meanings. Lexical Semantics Across Domains, Languages and Cultures*, Oxford University Press: Londra, 316 p. ISBN: 978-0-19-878355-8.
- Grimm Jacob, Wilhelm Grimm et al. 1838 - 1961. *Dicționarul limbii germane în formă digitală*, <http://dwb.uni-trier.de/de/>, accesat la 03.04.2024.
- HaiRO. 2019. HaiRO. <https://proiectulbrancusihairo.wordpress.com>, accesat la 22.08.2022.
- Hanks, Patrick. 2006. „The Organization of the Lexicon: Semantic Types and Lexical Sets”, în Elisa Corino, Carla Marengo, Cristina Onesti (ed.), *Proceedings of the 12th EURALEX International Congress*, Torino, Italia, 6-9 septembrie 2006, pp. 1165-1168, Edizioni dell'Orso, ISBN 88-7694-918-6. <https://euralex.org/publications/the-organization-of-the-lexicon-semantic-types-and-lexical-sets/>, accesat la 19.03.2025.
- Harmelen, Frank van, Vladimir Lifschitz, Bruce Porter, editori. 2007. *Handbook of Knowledge Representation*, Elsevier Science, 1034 p., ISBN: 9780080557021.
- Harris, Steve, Andy Seaborne. 2013. *SPARQL 1.1 Query Language. W3C Recommendation 21 March 2013*, <https://w3.org/TR/sparql11-query/>, accesat la 16.09.2023.
- Harzing, Anne-Wil. 2019. „Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science?”. În *Scientometrics* (2019) 120:341–349. <https://doi.org/10.1007/s11192-019-03114-y>, accesat la 16.11.2024.
- Hasan, Ismael, Javier Parapar, Roi Blanco. 2008. „Segmentation of Legislative Documents Using a Domain-Specific Lexicon”. În *19th International Workshop on Database and Expert Systems Applications*, Torino, Italia, 2008, pp. 665-669. DOI: <https://doi.org/10.1109/DEXA.2008.45>.
- Haspelmath, Martin. 2020. „The structural uniqueness of languages and the value of comparison for language description”. În *Asian Languages and Linguistics*, 1:2 (2020), pp. 346–366. ISSN

2665-9336, E-ISSN 2665-9344, <https://doi.org/10.1075/alal.20032.has>, accesat la 02.04.2025.

Haspelmath, Martin. 2021. „General linguistics must be based on universals (or non-conventional aspects of language)”. În *Theoretical Linguistics*, 2021, 47(1–2): pp. 1–31. <https://doi.org/10.1515/tl-2021-2002>, accesat la 02.04.2025.

Hendricks, Ginny, Dominika Tkaczyk, Jennifer Lin, Patricia Feeney. 2020. Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies* 2020; 1 (1): 414–427. doi: https://doi.org/10.1162/qss_a_00022, accesat la 02.04.2025.

Hess, D. J. 1997. *Science Studies: An advanced introduction*. New York: New York University Press.

Holban, Corina, Felicia Codîrlașu, Andrei Mineă, Ștefan Diaconescu. 2014. „Sense-tagging of Romanian glosses”, în Mihaela Colhon, Adrian Iftene, Verginica Barbu Mititelu, Dan Cristea, Dan Tufiș (ed.), *The 10th International Conference on Linguistic Resources and Technologies for Romanian Language (ConsILR 2014)*, Craiova, septembrie 2014. ISSN: 1843-911X. DOI: 10.13140/2.1.2948.2246.

HTML. 2011. *HTML. Living Standard*, Web Hypertext Application Technology Working Group, <https://html.spec.whatwg.org/>, accesat la 27.08.2023.

HTOED. 2022. Historical Thesaurus of the Oxford English Dictionary. <https://ht.ac.uk>, accesat la 17.06.2022.

Huet, Gérard. 2019. „Sanskrit lexicography, past and future”. În *Research on the Language and Script in Buddhist Sutras*, ed. Pr Li Wei. Hangzhou: Hangzhou Buddhist Academy. <http://gallium.inria.fr/~huet/PUBLIC/Hangzhou.pdf>, accesat la 09.06.2022.

ILB. 2019. *Institutul de Lingvistică „Iorgu Iordan - Al. Rosetti” din București. Plan de activitate. 2019.* https://acad.ro/bdar/rapInt2019/01lit/plan/2_InstLingvIordanRosetti_Plan.pdf, accesat la 24.08.2022.

Ion Radu, Elena Irimia, Dan Ștefănescu, Dan Tufiș. 2012. „ROMBAC: The Romanian Balanced Annotated Corpus”. În *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, mai 2012, Istonbul, Turcia, p. 339-344. https://lrec-conf.org/proceedings/lrec2012/pdf/218_Paper.pdf, accesat la 29.08.2022.

INTELLIT ICI. 2018. *Prezervarea și valorificarea patrimoniului literar românesc folosind soluții digitale inteligente pentru extragerea și sistematizarea de cunoștințe (INTELLIT)*. Proiect de cercetare desfășurat în cadrul Institutului de Istorie și Teorie Literară „G. Călinescu”, în perioada 2018-2020. <https://intellit.ici.ro/prezentare/informatii-proiect/>, accesat la 28.06.2023.

INTELLIT IITL. 2018. *Prezervarea și valorificarea patrimoniului literar românesc folosind soluții digitale inteligente pentru extragerea și sistematizarea de cunoștințe (INTELLIT)*. Proiect de cercetare desfășurat în cadrul Institutului de Istorie și Teorie Literară „G. Călinescu”, în perioada 2018-2020. <https://inst-calinescu.ro/proiecte-de-cercetare/>, accesat la 28.06.2023.

Jakubíček, Miloš. 2010. *Enhancing Czech Parsing with Complex Valency Frames*, disertație, Universitatea Masaryk, Facultatea de informatică, Brno, Cehia. <https://is.muni.cz/th/amdcn/dp.pdf>, accesat la 04.04.2025.

- Joshi, S. D., J. A. F. Roodbergen. 1991. *The Aṣṭādhyāyī of Pāṇini: With Translation and Explanatory Notes*, vol. I, New Delhi: Sahitya Akademi. ISBN: 81-260-0751-6.
- Kalibatiene, Diana, Olegas Vasilecas. 2011. *Survey on Ontology Languages*, în Grabis, J., Kirikova, M. (editori) *Perspectives in Business Informatics Research. BIR 2011. Lecture Notes in Business Information Processing*, vol 90, Springer: Berlin - Heidelberg. https://doi.org/10.1007/978-3-642-24511-4_10, accesat la 04.04.2025.
- Kant, Immanuel. 1999. *Metafizica moravurilor*, traducere de Rodica Croitoru, ediția a doua, revizuită, București: Antaios, 352 p.
- Kiparsky, Paul. 1991. „Economy and the construction of the Sivasutras”, în M. M. Deshpande și S. Bhate (ed.), *Paninian Studies*. Ann Arbor, Michigan: University of Michigan Press. <https://web.stanford.edu/~kiparsky/Papers/siva-t.pdf>, accesat la 07.06.2022.
- Klimek, Bettina, John P. McCrae, Christian Lehmann, Christian Chiarcos, Sebastian Hellmann. 2017. OnLiT: An Ontology for Linguistic Terminology. *Proceedings of Language, Data, and Knowledge - First International Conference, LDK 2017*, volume 10318 of *Lecture Notes in Computer Science*, page 42--57. Springer Verlag,
- Koss, Mike. 2017. *A Directed Acyclic Word Graph implementation in TypeScript/JavaScript*. <https://github.com/mckoss/dawg>, accesat la data de 24.08.2023.
- Kuebler, Sandra, Heike Zinsmeister. 2015. *Corpus Linguistics and Linguistically Annotated Corpora*. Bloomsbury: Bloomsbury Academic. 320 p. ISBN: 9781441164476.
- Kukich, Karen. 1992. „Techniques for Automatically Correcting Words in Text”. În *ACM Computing Surveys*, vol. 24, nr. 4, decembrie 1992. <https://dl.acm.org/doi/pdf/10.1145/146370.146380>, accesat la 15.10.2023.
- Le Robert. 2020. *Le Robert Dico en ligne*, Éditions Le Robert. <https://dictionnaire.lerobert.com/>, accesat la 08.11.2023.
- Mareș, Alexandru. 2009. „Școala filologică din Institutul de lingvistică Iorgu Iordan - Al. Rosetti”. În *Limba română*, LVIII, nr. 1. București, pag. 65 - 121. <https://diacronia.ro/ro/indexing/details/A587/pdf/>, accesat la 16.12.2023.
- Mărănduc, Cătălina, Cenel Augusto Perez, Victoria Bobicev, Cătălin Mititelu, Florinel Hociung, Valentin Roșca, Roman Untilov, Petru Rebeja. 2015. https://github.com/UniversalDependencies/UD_Romanian-Nonstandard, accesat la data de 19.08.2023.
- Meier, Wolfgang. 2022. „Generating a high voltage site by going static”. În *e-editiones*, 06.12.2022. <https://e-editiones.org/posts/community-event-going-static/>, accesat la data de 26.10.2024.
- Meirovici, Betina. 1970. *Adevărul literar și artistic*. Vol. III.1: (1920 - 1939). Index pe materii, Iași: BCU „Mihai Eminescu”, 782 p.
- Mincă, Andrei, Ștefan Diaconescu. 2013. „An Approach to Reduce Part of Speech Ambiguity Using Semantically Annotated Lexicon Definitions”. În *Proceedings of the 2nd International Conference on*

Management Science and Industrial Engineering, noiembrie 2013. ISBN: 978-90786-77-87-1. ISSN: 1951-6851. DOI: 10.2991/msie-13.2013.94.

Mischler, Ælfwine. 2018. Book Indexes — Part 1: Basic Vocabulary. *Mischler Editorial, indexing, copyediting, proofreading*.
<https://americaneditor.wordpress.com/2018/05/21/book-indexes-part-1-basic-vocabulary/>, accesat la 28.02.2024.

Mititelu, Verginica Barbu, Ștefan Daniel Dumitrescu, Dan Tușiș. 2014. „News about the Romanian Wordnet”. În *Proceedings of the Seventh Global Wordnet Conference*, Tartu, Estonia, ianuarie 2014, p. 268–275. <https://aclanthology.org/W14-0137.pdf>, accesat la 06.08.2022.

MLD. 2009. *Monumenta linguae Dacoromanorum*.
<http://consilr.info.uaic.ro/~mld/monumenta/istoric.html>, accesat la 31.08.2022.

Munteanu, Eugen. 2021. *Monumenta linguae Dacoromanorum. Scurt istoric al unui proiect filologic major (I). Ziarul de Iași*, 13.02.2021.
<https://ziaruldeiasi.ro/stiri/monumenta-linguae-dacoromanorum-scurt-istoric-al-unui-proiect-filologic-major-i--279058.html>, accesat la 01.09.2022.

Newson, Kaitlin. 2017. „Tools and Workflows for Collaborating on Static Website Projects”. În *code{4}lib*, nr. 38, 18.10.2017. <https://journal.code4lib.org/articles/12779>, accesat la data de 26.10.2024.

Olmsted, Hugh. 1994. „Modelling the genealogy of Maksim Grek's collection types: the «Plectogram» as visual aid in reconstruction”. În Flier, Michael S., Daniel Rowland, *Medieval Russian Culture*, volumul al doilea, Los Angeles, Londra: Univ. of California Press, p. 107-133.

Opăriuc, Loredana. 2009. „Dicționarul limbajului poetic eminescian”. În Viorica S. Constantinescu, Cornelia Viziteu, Lucia Cifor, *Studii eminescologice*, nr. 11, Cluj-Napoca: Casa de editură Atlas-Clusium, p. 220-229. ISSN 1454-9115. <http://bcu-iasi.ro/docs/studii-eminescologice-nr11.pdf>, accesat la 31.08.2022.

Osborne, Timothy, Kim Gerdes. 2019. „The status of function words in dependency grammar: A critique of Universal Dependencies (UD)”. În *Glossa: a journal of general linguistics*, 2019, vol. 4. <https://doi.org/10.5334/gjgl.537>, accesat la 15.02.2025.

PALIROM. 2010. Pachet de aplicații lingvistice pentru limba română.
<http://softwinresearch.ro/index.php/ro/proiecte/palirom>, accesat la 28.07.2022.

Pamfil, Viorica. 1968. „Observatii cu privire la frecventa cuvintelor în limba secolului al XVI-lea”. În *Studii și cercetări lingvistice*, an 19, nr. 3, p. 297-302.
<https://dspace.bcu-iasi.ro/handle/123456789/3391>, accesat la 08.04.2025.

Până Dindelegan, Gabriela. 1972. „Reflecții pe marginea „teoriei cazurilor” a lui Charles J. Fillmore”. În *Studii și cercetări lingvistice*, an. 23, nr. 1, p. 49-57.
<https://dspace.bcu-iasi.ro/handle/123456789/4705>, accesat la 08.03.2025.

- Pană Dindelegan, Gabriela. 1992. *Sintaxă și semantică. Clase de cuvinte și forme gramaticale cu dublă natură (adjectivul, adverbul, prepoziția, forme verbale nepersonale)*. București: Universitatea din București, Facultatea de litere, 142 p.
- Pană Dindelegan, Gabriela, Rodica Zafiu, Adina Dragomirescu, Alexandru Nicolae, Adnana Boioc Apintei. 2023. *Dicționar de termeni gramaticali și concepte lingvistice conexe*. București: Univers Enciclopedic Gold.
- Peev Luciana, Lidia Bibolar, Jodal Endre. 1997. „A Formalization Model of the Romanian Morphology”. În Dan Tufiş, Poul Andresen (ed.), *Recent Advances in Romanian Language Technology*, București: Editura Academiei Române, p. 72-77.
- Peev, Luciana, Felicia Șerban. 2006. „Linguistic premises for designing a text-to-speech synthesizer”. În *Revista română de lingvistică*, LI, 3-4, p. 441-456, București. <https://lingv.ro/RRL%203-4%202006%20Peev,%20Serban.pdf>, accesat la 22.07.2022.
- Peev, Luciana, Felicia Șerban. 2008. „Metode de analiză lingvistică a textelor în limba română pentru extragerea terminologică. Instrumente și resurse”. În *Actele Seminarului Internațional „Instrumente pentru asistarea traducerii”*, București, 28-29 februarie 2008. <https://unilat.org/Library/Handlers/File.ashx?id=c5da92d6-ef18-4e7d-b853-0769a17a4f0c>, accesat la 13.08.2022.
- Petic, Mircea. 2014. „Tendințe actuale în procesarea limbajului natural pentru limba română”. În Mircea Petic, *Tradiție și inovare în cercetarea științifică*, ediția a 3-a, Materialele Colloquia Professorum, 12 oct. 2012, Bălți: Universitatea de Stat „Alecu Russo”, p. 245-249.
- Pragglejaz Group. 2007. „MIP: A Method for Identifying Metaphorically Used Words in Discourse”. În *Metaphor and Symbol*, 22:1, 1-39, DOI: 10.1080/10926480709336752, accesat la 04.01.2024.
- Pustejovsky, James. 1991. *The generative lexicon*. Waltham, MA: Association for Computational Linguistics.
- Pustejovsky, James. 1996a. *The generative lexicon*. Cambridge, MA, London, England: The MIT Press.
- Pustejovsky, James, Branimir Boguraev (editori). 1996b. *Corpus Processing for Lexical Acquisition*. Cambridge, MA: The MIT Press.
- Pustejovsky, James. 1998. „The Semantics of Lexical Underspecification”. În *Folia Linguistica* XXXII/3-4, pp. 323-347. <https://doi.org/10.1515/flin-2017-1004>, accesat la 23.03.2025.
- ReTeRom. 2022. Resurse și tehnologii pentru dezvoltarea interfețelor om-mașină în limba română. <https://racai.ro/p/reterom>, accesat la 30.07.2022.
- RoLingva. 2009. Colectiv RoLingva - SC Software ITC Cluj SA, Cluj-Napoca. <https://web.archive.org/web/20090227061255/http://rolingva.ro/rolingva.php>, accesat la 12.08.2022.
- roTenTen. 2016. *roTenTen: Corpus of the Romanian Web*. <https://sketchengine.eu/roenten-romanian-corpus/>, accesat la 26.08.2022.
- RoWordNet. 2013. Romanian WordNet. <http://ws.racai.ro:9191/narratives/batch1/RoWordNet.pdf>, accesat la 06.08.2022.

- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker, Jan Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*. https://framenet.icsi.berkeley.edu/fndrupal/the_book, accesat la 27.06.2022.
- Ruppenhofer, Josef, Hans C. Boas, Collin F. Baker. 2017. „FramNet”. În Pedro A. Fuertes-Olivera (ed.), *The Routledge Handbook of Lexicography*, p. 383-398, Londra: Routledge. ISBN pentru eBook: 9781315104942. DOI: <https://doi.org/10.4324/9781315104942>. http://sites.la.utexas.edu/hcb/files/2017/12/RuppenhoferBoasBaker_2017_Proofs.pdf, accesat la 03.07.2022.
- Russo, Demostene. 1910. *Studii și critice. O carte asupra Învățăturilor lui Pseudo-Neagoe. Răspuns unui critic. Cărți de bună-cuviință. Un catalog de mss grecești*, București, Inst. de Arte Grafice „CAROL GOEBL” S-r I. St. Rasidescu, Strada Doamnei, 16. https://bcub.ro/lib2life/Studii%20si%20critice_Russo%20Demostene_Bucuresti_1910.pdf, accesat la 10.09.2023.
- Russo, Demostene. 1912. *Critica textelor și tehnica edițiilor*, în Buletinul Comisiei istorice a României 1 (1912), București, Atelierele grafice Socec, p. 1-100.
- Schmidt, Desmond Allan. 2010. „The inadequacy of embedded markup for cultural heritage texts”. În *Lit. Linguistic Comput.* 25(3): 337-356. <https://eprints.qut.edu.au/38436/1/c38436.pdf>, accesat la 15.10.2024.
- SenDIS. 2017. Dezambiguizarea sensurilor cuvintelor folosind rețele lexicon. <http://softwinresearch.ro/index.php/ro/proiecte/sendis>, accesat la 28.07.2022.
- Shieber, Stuart, Hans Uszkoreit, Fernando Pereira, Jane Robinson și Mabry Tyson. 1983. „The Formalism and Implementation of PATR-II”. În Barbara J. Grosz and Mark E. Stickel (ed.), *Research on Interactive Acquisition and Use of Knowledge (Final report of SRI Project 1894)*. Menlo Park, California: SRI International. <https://dash.harvard.edu/bitstream/handle/1/23492376/formalism%20and%20implementation%20of%20patr-ii%20-%20shieber%20-%201983.pdf>, accesat la 14.07.2022.
- SIMBNR. 2011. *Sistem integrat on-line pentru managementul bibliografiei naționale retrospective*. <https://biblacad.ro/seminarii/SIMBNR.pdf>, accesat la 02.09.2022.
- Simionescu, Radu. 2011. „Hybrid POS Tagger”. În *Proceedings of the Workshop “Language Resources and Tools with Industrial Applications”*, Cluj-Napoca, 30-31 august 2011, p. 21-28. Iași: Editura Universității „Alexandru Ioan Cuza” din Iași.
- Sgarbas, Kyriakos N., Nikos D. Fakotakis, George N. Kokkinakis. 2000. *A Straightforward Approach to Morphological Analysis and Synthesis*. Proceedings COMLEX 2000, Workshop on Computational Lexicography and Multimedia Dictionaries, Kato Achaia, Greece, September 22-23, 2000, p. 31-34.
- Sohyeong, Kim, Taejin Choi, Aeran Yoon, Sung-Soo Seol. 2013. „Korea Citation Index and Its Macro Bibliometrics”. În *Asian Journal of Innovation and Policy*, 2.2: p. 194-211. <http://dx.doi.org/10.7545/ajip.2013.2.2.194>, accesat la 29.11.2024.

Sperberg-McQueen, C. M., Claus Huitfeldt. 2004. „GODDAG: A Data Structure for Overlapping Hierarchies”, în *DDEP-PODDP 2000*, P. King și E.V. Munson (ed.), Lecture Notes in Computer Science 2023, Berlin: Springer, p. 139-160. <https://cmsmcq.com/2000/poddp2000.html>, accesat la 18.01.2024.

Stan, Aurelia. 1961. *Indice de materii*, în *Cercetări de lingvistică*, an.6, nr. 2 , p. 479-482.

Stan, Camelia. 2003. *Gramatica numelor de acțiune din limba română*, București, Editura Universității din București, ISBN 973-575801-6, <https://biblioteca-digitala.ro/?pub=8123-gramatica-numelor-de-actiune-din-limba-romana>, accesat la 26.03.2025.

Stanciu, Măriuca. 2014. „BYZANTION – un proiect de valorificare a colecției de manuscrise a Bibliotecii Academiei ... și mai mult decât atât”. În Florin G. Filip, Doina Banciu, Mircea Regneală, Măriuca Stanciu (coordonatori), *Colecții de patrimoniu în era digitală*, București: Editura Academiei Române, p. 18-27. https://biblacad.ro/workshop/VolumWorkshopBAR_vf.pdf, accesat la 25.08.2022.

Sterian, Florin. 2017. *Bibliografia românească de lingvistică (BRL, 59, 2016). Lucrări de lingvistică apărute în țara noastră în cursul anului 2016. Limba română*, LXVI (2), p. 123-269.

Șerban, Felicia, Luciana Peev, Lidia Bibolar, Dana Bucerzan. 1996. „Baza de date a limbii române. Fonetica și fonologie”. În Dan Tufiș (editor), *Limba și tehnologie*, București: Editura Academiei Române, p. 157-160.

Șerban Felicia, Dana Bucerzan, Luciana Peev, Lidia Bibolar. 1997. „Database of the Romanian Language Phonetics and Phonology”. În Dan Tufiș, Poul Andresen (ed.), *Recent Advances in Romanian Language Technology*, București: Editura Academiei Române, p. 188-194. <https://racai.ro/external/static/awde/serban.html>, accesat la 08.08.2022.

Șerbănescu, Andra. 1989. Sesiunea științifică anuală a Facultății de Filologie din București (24-26 noiembrie 1988). *Limba română*, an. XXXVIII, nr. 2 martie-aprilie, p. 161-163. <https://dspace.bcu-iasi.ro/handle/123456789/7188>, accesat la 24.08.2024.

Șerbănescu, Andra. 1991. Structuri retorice și organizare sintactică în *Manuscrisul de la Ieud. Limba română*, an. XL, nr. 1-2, p. 85-90.

Șerbănescu, Andra. 1991. Despre statutul lui *uite* vs *uită-te*. *Studii și cercetări lingvistice*, XLII, nr. 3-4, p. 151-163, București.

Șerbănescu, Andra. 1994. Pentru un dicționar sintactic al verbelor românești. *Studii și cercetări lingvistice*, XLV, nr. 3-4, p. 133-150, București.

Șuteu, Flora, Laura Vasiliu. 1983. Indicele revistei „Limba română” pe anul 1983. *Limba română*, an. XXXII, nr. 6, p. 553-569. <https://dspace.bcu-iasi.ro/handle/123456789/6696>, accesat la 19.08.2024.

Tămăianu-Morita, Emma, Sanda Cherata, Cornel Vilcu. 2006-2007. Analiza sintagmatică a textelor românești prin mijloace informatice: proiectul *SIASTRO. Dacoromania*, serie nouă, XI–XII, Cluj-Napoca, p. 77-87.

- Tănăsescu, Florin-Teodor. 2009. „Remus Răduleț și terminologia tehnică”. În *Studii și comunicări*, vol. II, Cluj-Napoca: Mega Print, p. 97-120. https://studii.crist.ro/doc/2009/2009_06.pdf, accesat la 11.08.2022.
- TenTen. 2016. *TenTen Corpus Family*. <https://sketchengine.eu/documentation/tenten-corpora/>, accesat la 26.08.2022.
- The Indexer. 2011. *Indexing societies around the world: a brief survey*, în The Indexer. The International Journal of Indexing, vol. 29, nr. 4, decembrie 2011. <https://theindexer.org/wp-content/uploads/2020/07/indexing-societies.pdf>, accesat la 13.09.2023.
- Timotin, Alexandru, Florin Teodor Tănăsescu. 1997. „Structures for a Thesaurus of Technical Terminology”. În Dan Tufiş, Poul Andresen (ed.), *Recent Advances in Romanian Language Technology*, Bucureşti: Editura Academiei Române, p. 105-120.
- Tittel, Sabine Tittel, Christian Chiarcos. 2018. „Historical lexicography of Old French and linked open data: transforming the resources of the Dictionnaire étymologique de l'ancien français with OntoLex-Lemon”. În *Proceedings of the 6th Workshop on Linked Data in Linguistics: Towards Linguistic Data Science*, 12 mai 2018, Miyazaki, Japonia. ISBN: 979-10-95546-19-1. <https://opus.bibliothek.uni-augsburg.de/opus4/files/104091/104091.pdf>, accesat la 04.11.2023.
- Toma, Ştefan-Adrian, Adriana Stan, Mihai-Lica Pura, Traian Barsan. 2017. „MaRePhoR - An Open Access Machine-Readable Phonetic Dictionary for Romanian”. În *Proceedings of the 9th Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, Bucureşti. http://adrianastan.com/papers/2017_SPED_Marephor.pdf, accesat la 29.07.2022.
- Tufiş, Dan, Lidia Diaconu, Călin Diaconu, Ana Maria Barbu. 1996. „Dicţionar al limbii române destinat traducerii automate”. În Dan Tufiş (editor), *Limbaj şi tehnologie*, Bucureşti: Editura Academiei Române, p. 93-100.
- Tufiş, Dan, Ana-Maria Barbu. 1997. „A Reversible and Reusable Morpho-Lexical Description of Romanian”. În Dan Tufiş, Poul Andresen (ed.), *Recent Advances in Romanian Language Technology*, Bucureşti: Editura Academiei Române, p. 83-93.
- Tufiş, Dan, Dan Cristea. 2002. „Methodological issues in building the Romanian Wordnet and consistency checks in Balkanet”. În *Proceedings of The Third International Conference on Language Resources and Evaluation*, Las Palmas, Insulele Canare, Spania, 29 mai - 31 mai 2002. <https://racai.ro/media/Tufis-Cristea-LREC2002.pdf>, accesat la 03.08.2022.
- Tufiş, Dan, Eduard Barbu, Verginica Barbu Mititelu, Radu Ion, Luigi Bozianu. 2004. „The Romanian Wordnet”. În *Romanian Journal of Information Science and Technology*, vol. 7, nr. 1-2, 2004, p. 107-124. http://dblab.upatras.gr/balkanet/journal/12_Romanian1RJIST-RACAI1corect.pdf, accesat la 03.08.2022.
- Tufiş, Dan, Verginica Barbu Mititelu, Alexandru Ceaşu, Luigi Bozianu, Cătălin Mihăilă, Margareta Manu Magda. 2006. „Noi dezvoltări ale WordNET-ului românesc”. În *Lucrările atelierului Resurse lingvistice şi instrumente pentru prelucrarea limbii române*, Iaşi, 3 noiembrie 2006, p. 17-21.

- Tuفیș, Dan. 2008. „Ro-WordNet: ontologie lexicală pentru limba română”. *Academica*, XVIII (208–209): 30–34, februarie-martie 2008. ISSN 1220–5737.
- Tuفیș, Dan, Maria Mitrofan, Vasile Păiș, Radu Ion, Andrei Coman. 2020. „Collection and Annotation of the Romanian Legal Corpus”. În *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, Marsilia, Franța, 11–16 mai 2020, p. 2773–2777. <http://lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.337.pdf>, accesat la 31.07.2022.
- UaicPosTagger. 2016. UaicPosTagger. <https://github.com/radsimu/UaicNlpToolkit>, accesat la 31.07.2022.
- Unicode Consortium. 2022. *The Unicode Standard. Version 15.0*. ISBN: ISBN 978-1-936213-32-0. <https://unicode.org/versions/Unicode15.0.0/UnicodeStandard-15.0.pdf>, accesat la 21.05.2023.
- van Eck, Nees Jan, Ludo Waltman, Vincent Lariviere, Cassidy Sugimoto. 2018. „Crossref as a new source of citation data: A comparison with Web of Science and Scopus”. În *CWTS. Meaningful metrics*, 17 ianuarie 2018. <https://cwts.nl/blog?article=n-r2s234>, accesat la 16.11.2024.
- Vintilă-Rădulescu, Ioana. 2002. „Resurse lingvistice pentru limba română elaborate la Institutul de Lingvistică Iorgu Iordan”. În Dan Tuفیș, Florin Gh. Filip (coordonatori) *Limba Română în Societatea Informațională - Societatea Cunoașterii*, 21-33. București: Expert.
- Visser, Martijn, Nees Jan van Eck, Ludo Waltman. 2021. „Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic”. În *Quantitative Science Studies*, vol. 2, fascicula 1, 2021. https://doi.org/10.1162/qss_a_00112, accesat la 16.11.2024.
- Vîlcu, Cornel. 2008. Preliminarii teoretice la analiza gramaticală în proiectul *SIASTRO*: nivelul sintagmatic. *Dacoromania*, serie nouă, XIII, nr. 2, Cluj-Napoca, p. 117–126.
- Vossen, Piek T. J. M. 2013. „93. WordNet: Principles, developments and applications”. În Rufus Hjalmar Gouws, Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (editori), *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*, p. 1307-1320, Berlin, Boston: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110238136.1307>.
- Wilkinson et al. 2016. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hoof R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. *The FAIR Guiding Principles for scientific data management and stewardship*. „Sci Data”, 8, 2016, 3. DOI: 10.1038/sdata.2016.18. <https://ncbi.nlm.nih.gov/pmc/articles/PMC4792175/>, accesat la data de 01.11.2022.
- WordNet. 2011. WordNet. A lexical database for English. <https://wordnet.princeton.edu>, accesat la 03.07.2022.

