



Științele limbajului

**Roxana Patraș
Loredana Cuzmici**
(editors)

**Tools, Methods, and Solutions
for the Exploration of Romanian Corpora**

INSTITUTUL EUROPEAN

Roxana Patraș, Loredana Cuzmici (editors), *Tools, Methods, and Solutions for the Exploration of Romanian Corpora*

© 2024 Institutul European Iași, pentru prezenta ediție

INSTITUTUL EUROPEAN

Iași, Str. Bălușescu nr. 2, et. I, O P 6, CP 1309

euroedit@hotmail.com.; www.euroinst.ro

Descrierea CIP a Bibliotecii Naționale a României

Tools, methods, and solutions for the exploration of Romanian corpora /

ed.: Roxana Patraș, Loredana Cuzmici. - Iași : Institutul European, 2024

Conține bibliografie

ISBN 978-606-24-0395-9

I. Patraș, Roxana (ed.)

II. Cuzmici, Loredana (ed.)

811.135.1

All rights reserved. No part of this work covered by the copyrights hereon may be reproduced or copied in any form or by any means – graphic, electronic, or mechanical, including photocopying, recording, taping – without written permission of publisher.

PRINTED IN ROMANIA

Roxana Patraş, Loredana Cuzmici
(editors)

TOOLS, METHODS, AND SOLUTIONS FOR THE EXPLORATION OF ROMANIAN CORPORA

INSTITUTUL EUROPEAN
2024

This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS–UEFISCDI, project number PN-III-P1-1.1-TE-2019-0127, within PNCDI III.

The revision and editing of this collection was performed by BolderWords translation & revision and was supported by the Altissia Chair in Digital Cultures and Ethics held by Chris Tanasescu at Université catholique de Louvain.

Table of Contents

INTRODUCTORY NOTES	7
Tactics: your DH is not my DH Roxana PATRAȘ	9
CREATING TOOLS FOR ROMANIAN CORPORA	23
RELATE: a modern processing platform for Romanian language Vasile PĂIȘ, Radu ION, Andrei-Marius AVRAM, Maria MITROFAN, Dan TUFIȘ	25
Tools and Resources for Digitizing Historical Romanian Documents Victoria BOBICEV, Tudor BUMBU, Ludmila MALAHOV, Alexandru COLESNICOV, Svetlana COJOCARU, Liudmila BURTSEVA, Cătălina MĂRÂNDUC	53
Digitalization of Romanian Lexicography Elena Isabelle TAMBA	85
ADAPTING TOOLS FOR ROMANIAN CORPORA	97
Postdigital Creativity: modeling poetry translation with multiplexes, neural networks, and large language models Raluca TANASESCU, Chris TANASESCU	99
Challenges of Digitization and Digitalization for the Study of Writing: from archival survey to process analysis Georgeta CÎȘLARU	125
Towards a Digital Corpus-Based Method for Assessing Language Level in EFL Student Writing: a case study on Romanian undergraduate literary analyses Alexandru ORAVIȚAN, Mădălina CHITEZ	139
DIGITIZING ROMANIAN HERITAGE	151
A Computer-Assisted Analysis of Eminescu's <i>Doină</i> Ioana GALLERON	153
Mapping Eastern European Political Discourse and Inequalities through Public Monuments: a digital cartography crowdsourcing project Voica PUȘCAȘIU	175
The Title of the Feuilleton Novel: a model of semantic annotation Lucreția PASCARIU	193
Exercises in Literary Geography: where hajduks roam in the second half of the nineteenth century Alexandra OLTEANU	227

<i>DIGITUS DEI EST HIC!</i>	251
Proxy Structures in Computational Analyses of Text Corpora	253
Laura PRICOP	
The Many Faces of Virtual: charting the family resemblance network for conceptual understanding	267
Matei Alexandru STOENESCU	
CARTOGRAPHIES OF THE ROMANIAN NOVEL: THEN AND NOW	297
The Recontextualization of a Genre	299
Loredana CUZMICI	
Notes on the Contributors	321

Tactics: your DH is not my DH

Roxana PATRAŞ

The present book should have been titled *Tactics*. Inspired by recent contributions to the state of the field (Rieder and Röhle 2012, Malazita 2021), we use Michel de Certeau's term in order to enhance the challenging situation that researchers of *less(er)-resourced languages* or *under-resourced languages* tend to be placed in, chiefly when their contributions are weighed up against the *big-tent* metaphor that has been the most impactful of all DH brands (Terras 2011, Svensson 2016). According to de Certeau, *tactics* is defined as a cluster of everyday practices devised by people or, if extrapolation is accepted, by any *realia* endowed with the ability to mobilize small-scale, lesser resources in order to move and evolve in a vast territory. Tactical navigation would thus refer not only to geographic workarounds—the shortcuts that walkers take in order to adapt to large urban environments—but also to a special navigation through the knowledge landscape in general. In a context marked by head-whirling concepts such as turbo-capitalism, datafication, infoxication, algorithmic civilization, virtual reality, big data, digital turn, technocratic culture, instrumentalist episteme, anglobalism, and so forth, it is crucial and paramount to raise awareness on the tactics that researchers of *less(er)-resourced languages*, “the DH neighbouring tribes” (Bode and Longley Arthur 2014) apply in order to enhance particularities or, as one of the European Language Resource Association [ELRA] reports states, to bring to the fore such tactics that are used simply to help them survive the digital world (Report of Special Interest Group on Under-resourced Languages [SIGUL], 2019, 2020).

The very same attributes *less(er)-resourced* and *under-resourced* also call for a deeper examination, as this might unveil and thus reveal, for instance, the origins of some particular terminology that, more often than not, have been embraced uncritically, like any other stereotype. Between 1982 and 2010, the European Bureau

for Lesser-Used Languages [EBLUL] put in circulation various reports regarding languages (mostly European) that, from a demographic and geographic perspective, were considered *lesser-used* and thus endangered. EBLUL and the European Charter for Regional or Minority Languages [ECRML], that ensued the bureau’s activity, were meant to represent and speak on behalf of Europe’s 50 million speakers using *regional* and *minority idioms* (*non-territorial languages* included), yet without providing clear criteria or definitions for the question under debate (Bernoville 2020). Since the classification highly depended on national states, the list of languages “protected under the Charter” reflected various policies and approaches to minority issues as well as an interwoven, very complex landscape in which many of the European idioms were tagged as both majority and minority languages. Romanian and its dialects, for instance, appeared as a minority language, and it was thus subjected to future safeguarding manoeuvres, in Bosnia and Herzegovina, Croatia, Hungary, Serbia and Ukraine. Already entangled by various historical factors, the picture of ethnic minorities’ language(s) would have looked even more opaquely “thick” if ECRML had also considered the minorities and diasporas developed as a consequence of massive labor migration waves.

The release of ECRML (1992) to a certain degree coincided with the establishment of the European Language Resources Association [ELRA] in 1995 as well as with the first boost of language technologies [LT] (Păiș and Tufiș 2022). This association aimed to create a catalogue of language resources (*see* ELRA catalogue at <https://catalogue.elra.info/en-us/>), in addition to supporting special interest groups like *Workgroup on Less-Resourced Languages* (2016) and *Special Interest Group on Under-resourced Languages* (2017). Apparently, the terminological emphasis changed from *less(er)-used* to *lesser/ under-resourced*, that is, from debatable minority considerations to more objective and seemingly universal *technicities*. However, the new understanding of language resourcefulness as a series of scalable technicist affordances—from *use(d)* to *resource(d)*—was not apt to either sort out the majority-minority conundrum or address the pejorative keys implied in prefixes such as “less” or “under”. As a consequence, in the program of the last SIGUL conference (2024), one might have come across a rich assortment of *less(er)-resourced languages*, many of them non-European: Turkish, Portuguese, Cicipu,

Effutu, Mocho', Upper Napo Kichwa, Toratán, Ulwa, Uzbek, Icelandic, Bahasa Sug, Chavacano, Filipino, Iranun, Maguindanaon, Meranaw, Sinama, Teduray, and Yakan, Occitan, Irish, Luxembourgish, *Romanian*, Bavarian, Ligurian, Croatian, Serbian, Bosnian, Montenegrin, Danish, Alsatian, Poitevin-Santonguais, Nepali, Armenian, Arbëresh, Persian (Farsi), Filipino, Cebuano, Kapampangan, Hiligaynon, Ilokano, Bicolano, Waray, and Tausug, Guarani Mbya, Somali, Basque, te reo Māori, 'ōlelo Hawai'i (Maite, Sakti, and Soria 2024).

In further support of the reasonable arguments summarized above and contrary to positive discrimination trends that have recently emerged in Romanian academia, the present volume proves that Romanian is anything but an *under-resourced language*. In fact, recent contributions point out that two of the top challenges that Romanian DH projects have to grapple with are firstly, the infrastructure, and secondly, the digitization policies (Schöch *et al.* 2021, Chitez, Rogobete, and Foitoş 2019, Patraş 2022). While (*digital*) *infrastructure* has been perceived until very recently as too general, static and nonhuman (McGrail, Nieves, and Senier 2021)—because it might designate a very diverse range of things, from services, hubs, and resources to epistemic scaffolding, unconscious institutional designs, and even “infrastructural imaginaries” (Parks 2015 *qtd. in* Lach and Pressman 2021)—the problem of digitization policies also revolves around national strategies that, going back to Michel de Certeau's reasoning, tend to manifest in a hegemonic and sluggish manner (1984). The tension between the hegemony of a research system, scarred across by the ploughs of Communist ideology and by the idleness of post-Communist transition, on the one hand, and the various tactics devised by independent DH researchers, ad-lib DH hubs, and spontaneous DH initiatives, on the other, might well enable us to contextualize the comparison trap described in an excellent report on the current status of Romanian language technologies:

Considering the neighboring EU member states in the region, the number of Romanian corpora is the lowest... Romanian lexical and conceptual resources represent approximately 10% of the available English resources, while grammar and language models represent 20%. Considering neighboring EU member states, Romanian resources of this kind are near the region's average... In spite of the

reduced number of available language resources [that is, *corpora*], applications for different natural language processing tasks were constructed for Romanian language. This includes lemmatization, part-of-speech tagging, dependency parsing, named entity recognition, syllabification, automatic speech recognition, text-to-speech, automatic translation, punctuation restoration, terminology annotation, text classification. Following the ELE project, over 100 tools and services for Romanian language were identified. This number represents 15% of the available English tools. Looking at language tools from neighboring EU member states, we find a similar number for Bulgarian and slightly higher for Hungarian tools (Păiș and Tufiș 2022, 7–8).

In a nutshell, attributes such as *under* and *less(er)* should, in all reality, be applied with a pinch of salt, chiefly when a zoomed-in picture of technology support across (European) languages shows that everything depends on the terms and grounds of comparison. For instance, among the 11,500 metadata records on tools/services listed in the European Language Grid (ELG) Catalogue, only English, French, German and Spanish fall within good-LT-support and moderate-LT-support groups, while the majority of other analyzed languages are “fragmentarily” supported. From the viewpoint of *use* (usage in terms of demographics)-*resource* (resourcefulness) balance, which involves, as we have already mentioned, a series of scalable technicist affordances, *Figure 1: Overall state of technology support for selected European languages* (Păiș and Tufiș 2022, 14) is even more informative because it enhances, within the fragmentary supported languages, 5 sub-groups, among which Finnish, Swedish and Dutch stand out as technically backed-up despite reduced demographics: a. Finnish, Italian, Portuguese and Dutch; b. Swedish and Polish; c. Catalan, Greek, Hungarian, Romanian, Estonian, Danish, Norwegian, Lithuanian, Czech, Latvian; d. Bulgarian, Slovak, Slovenian, Croatian, Basque, Welsh, Galician, Icelandic; e. Serbian, Maltese, Irish, Luxembourgish, and Bosnian.

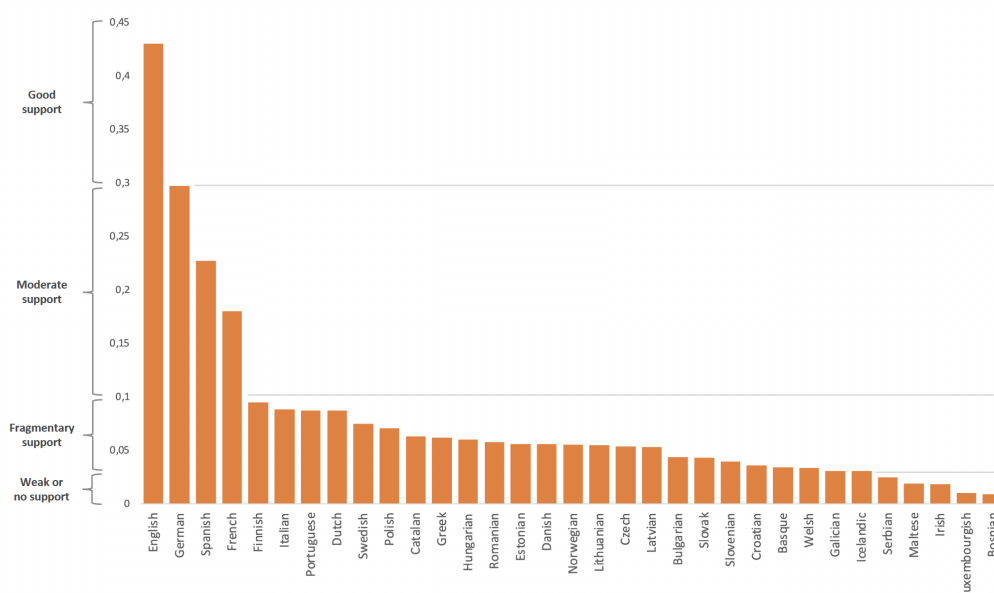


Figure 1: Overall state of technology support for selected European languages (2022)

Figure 1: Reproduced from *European Language Equality (ELE). Report on the Romanian Language, 2022*. Image credit Vasile Păiș and Dan Tuș

Despite lucid epistemic calibration on ground-truth data (e. g. ELG catalogue), this kind of insights might also carve out a logic of triangulation among “alternative peripheries,” which involves contrasting and negotiating the self-images of *peripheral languages* against three other cultural aspects: “the coerced and desired centres of their ‘dual dependency’ and the alternative periphery” (Ștefănescu 2022, 71). By surpassing the conventional postcolonialist critiques and drawing from recent input on the analogy postcolonialism-postcommunism (Chari and Verdery 2009), a transcolonial view on alternative peripheries and on alternative peripheral languages might also elucidate the majority-minority conundrum, detailed above, as well as the terms that have marked the approach, and consequently the interest, on languages usually researched and supported by the *big-tent* DH community (Weingart and Eichmann-Kalwara 2017). Indeed, if the emergence of DH as a research field has also built upon universalist and even utopian promises such as the Cold War confidence in apolitical technical expertise (Rieder and Röhle 2012, McCarty 2014, Risam 2018, Malazita 2021), then one must take into account the residual spectrality of research traditions too. The so-called *zombie concepts* (Braunstein and Warren 2021) favour all kinds of matches at the DH peripheries, for

instance the “elegiac” pairing between the computational impulse and the commemorative impulse (see also Birkerts 1994), the quixotic mélange in the formula “minimal computing” (Risam and Gil 2022) or the impetuously assertive disjunction “Your DH is not my DH” (Ray Murray 2015).

In an article entitled “Where is the ML in DH? And Where is the DH in ML? The Relationship between Modern Languages and Digital Humanities, and an Argument for a Critical DHML”, Thea Pitman and Claire Taylor avouch that “it is time for all the various and different versions of DH to compare and contrast their situatedness globally, hopefully eschewing the construction of any sense of centre/periphery or ‘one true DH’ as the exercise unfolds” (2017). This means acknowledging the linguistically-heterogeneous landscape of global DH and endorsing multilingualism—at least as a desideratum for alternative methods, tools and concepts—in some of the most esteemed DH events and publications (see, for instance, the special issue of “Cultural Analytics” devoted to the Arab DH). However, the openness to multilingualism and diversity does not even start to grease the giant, long-established (and some would say outmoded) wheels of global DH, which are still deeply entrenched, as postcolonial critics from Tara McPherson (2012) to Roopika Risam (2018) prove, by the colonial and neocolonial influences that have molded the digital cultural record so far, or, in short, by the systemic processes of canon replication.

Assisted by data from the formerly cited reports, we may add (while hoping to conduct proper research on this matter sometime in the near future) that canon replication secretly communicates with LT replication and dissemination in the DH peripheries. Beyond all talks about digital infrastructure and policies, that is, about *strategies*, technological scalarity is applicable only to groups of good- and moderately-supported languages and partially accounts for the uneven developments in terms of text processing (e.g. part-of-speech tagging, syntactic parsing), information extraction and retrieval (e. g. search and information mining), translation technologies (e.g. machine translation, computer-aided translation), natural language generation (e.g. text summarisation, simplification), speech processing (e.g. speech synthesis, speech recognition), image/video processing (e.g. facial expression recognition), human-computer interaction (e.g. tools for conversational systems).

However, as already revealed, the fetishization of technology and the compulsive focus on a digital preservation and tool-making paradigm do not allow for the problematization of digital disparities (Fernández L’Hoeste and Rodrigues 2020), for the nuanced approach on the locus of enunciation about DH concepts (Gallego Cuiñas 2024) as well as about situated DH tools, methods and solutions, which are the focal points of the present volume.

In an attempt to address (and hopefully redress) technology fetishization—as well as a whole spectrum of emotions, from awe and disapproval to submission and optimism—that hollow out at the DH peripheries and in the emergent DH communities, the book’s first two sections [*Creating tools for Romanian Corpora* and *Adapting Tools for Romanian Corpora*] follow a seamless and logical sequence all the way through to the last two sections [*Digitizing Romanian Heritage* and *Digitus Dei est Hic!*] by bridging *tools*, on the one hand, and *ethos*, on the other. When not straightforwardly protectionist, the European legaleeze on *less(er)-resourced languages* suggests that research stock should chiefly be invested in the creation of dedicated tools for endangered idioms as well as for a series of patrimonial elements enumerated in the UNESCO convention of 2003. However, as Vasile Păiș *et al.*, Victoria Bobicev *et al.* and Elena Isabelle Tamba attest, the integration of firm and already-tested solutions represents a cost-efficient and a human-empowering tactics. Addressing the fragmentation in the LT landscape and in Romanian academia, these contributions fully prove that instead of competing with one another, research projects and platforms (including NLP features such as annotation, transcription or digitization) should be designed to be interoperable and to interact with each other in order to create eco-systemical synergies. Bringing to the fore significant contributions to open access facilities, both citizen science and private initiatives are also touched upon from a critical angle, by reaffirming the FAIR principles, as well as by pleading for scholarly validation for all freely available LT resources (e.g. applications such as <https://dexonline.ro> and datasets created for various purposes).

The second section assembles contributions that share the perspective on corpora and archives as living, thus complex, realities—imaginable as *multiplexes* or as *multilayer networks*—that should thus be managed and modelled by keeping pace

with the ever-changing digital knowledge infrastructure (see also Bode 2018). Surmounting divides and embracing the concept of post-digitality, Raluca Tanasescu and Chris Tanasescu aptly point out that, while tools and code lines might belong to a largely-frequented set, it is the methods and approaches that could bring to the fore the wildly creative part of DH research. For instance, the relationship between the operation of translation and computer code reveals its fascinating versatility because it is rooted in the fact that, at its core, translating human-readable code into machine-executable instructions is a form of interpretation. Even if not targeted as such, the matters of translability and of foreign language proficiency are also at the core of Alexandru Oravițan and Mădălina Chitez's contribution, which describes a new method for assessing language levels in EFL student writing by using ROGER as a benchmark parallel corpus. Reflecting on the changes that the process of writing undergoes in the digital environment, Georgeta Cîșlaru's chapter *Challenges of Digitization and Digitalization for the Study of Writing: from archival survey to process analysis* introduces, through means of keylog data, a very ingenious approach to the textual genetics, as well as a new take on the long-established profiling of content creators (writers, for instance) as either *discoverers* or *executors*:

Compared to handwritten or typewritten data, texts produced by means of digital tools are formally and qualitatively different, and offer different research perspectives. Drafts become evanescent and not archival, and revisions leave no visible physical trace.

Section three *Digitizing Romanian Heritage* includes four use-cases, illustrative for the manner in which heritage texts and patrimonial corpora should be explored and promoted. Ioana Galleron's chapter draws from a larger initiative that aims at the prototypization of the first digital edition in Romania. The substantive results on the versions' melodic evolution and thematic patterns, both of them strikingly submerging or, on the contrary, raising up to the final version of the poem *Doină*, support the claim that the process of collation-comparison between the official scholarly edition and the collection of facsimiled manuscripts will be the corner stone of *Digital Mihai Eminescu*, if such project will ever catch the eye of

research strategists in Romania. While Ioana Galleron abstracts a poem's convolutions to a map, the other contributions also rely on different kinds of mapping: mapping the secret lives and conflicting ideologies of public monuments in Cluj-Napoca (Voica Pușcașiu), mapping the patterns in titling feuilleton novels (Lucreția Pascariu), and mapping the intersections of fictional spaces and real places ventured through by one of the most iconic Romanian heroes, the hajduk (Alexandra Olteanu). Occasioned and supported by various DH projects implemented at the Digital Humanities Laboratory (Institute of Interdisciplinary Research, "Alexandru Ioan Cuza" University of Iași), these chapters are also wrapped up in the *epilogue* signed by Loredana Cuzmici who, in line with Hirsch (2012), contributes a detailed analysis on the status of digital literary corpora and on their possible uses for pedagogical purposes.

Under the facetiously ominous title *Digitus Dei est Hic!*, the last section juxtaposes two philosophical approaches to the stunning complexity of concepts such as digitality, computability, and virtuality. While their methods seem rather dissimilar—Matei Alexandru Stoenescu explores amplification and Laura Pricop explores abstraction—these contributions provide further proof of a falsely-oppositional paradigm that has recently been discussed by Katherine Bode in "What's the Matter with Computational Literary Studies" (2023). In fact, the reader of this collection will perceive how the same rhizomatic concepts and practices could be addressed in different ways: by compiling an expansive corpus of networked semantic traits in order to empirically test Wittgenstein's concept of "family resemblances" or by identifying, in recent computational literary studies, the nodal importance of arguments by proxy in harnessing the conceptual complexities of literature and in steering away from instrumentalist shaming.

Following the notion of tactics, the chapters in this volume also critically inquire if digital humanities constitute "an opportunistic, strategic, and instrumental version of academic imperialism" (L'Hoeste and Rodrigues 2020, 14) or whether they have the potential to become a counterhegemonic practice against various conceits on lesser-resourced languages and regions (Thompson Klein 2015, Galleron and Idmhand 2020). We are grateful to our friends Chris Tanasescu, Raluca Tănăsescu, Ioana Galleron, Georgeta Cîșlaru, and Mădălina Chitez, whose mature

situatedness in both the global DH and Romanian DH communities has opened a new path for reflection on diasporic modes and alternative peripheries. We also extend our gratitude to the other colleagues who are active members of DH hubs around Romania and whose gentle feedback, in all circumstances, has made our own tactics much sleeker and more entertaining.

References

- Bernoville, Gabrielle. "Europe's Forgotten Words. The Case of the European Regional and Minority Languages." In *La Regionist. A Regions of Europe Society Journal*, Vol. 2, January–February, 2020, 3–8.
- Birkerts, Sven. *The Gutenberg Elegies: The Fate of Reading in an Electronic Age*. London: Faber and Faber, 1994.
- Bode, Katherine. "What's the Matter with Computational Literary Studies." *Critical Inquiry* 49, no. 4 (2023): 507–529.
- Bode, Katherine. *A World of Fiction Digital Collections and the Future of Literary History*. Ann Arbor: University of Michigan Press, 2018.
- Bode, Katherine and Paul Longley Arthur. "Collecting ourselves." In *Advancing Digital Humanities. Research, Methods, Theories*, edited by Paul Longley Arthur and Katherine Bode, 1–12. London: Palgrave Macmillan, 2015.
- Braunstein, Laura R., and Michelle R. Warren. "Zombies in the Library Stacks." In *People, Practice, Power. Digital Humanities outside the Center*, edited by Anne McGrail, Angel David Nieves, and Siobhan Senior, 70–83. Minneapolis–London: University of Minnesota Press, 2021.
- Chari, Sharad, and Katherine Verdery. "Thinking between the Posts: Postcolonialism, Postsocialism, and Ethnography after the Cold War." *Comparative Studies in Society and History* 51, no. 1 (2009): 6–34.
- Chitez, Mădălina, Roxana Rogobete, and Alexandru Foitoș. "Digital Humanities as an Incentive for Digitalisation Strategies in Eastern European HEIs: A Case Study of Romania." In *European Higher Education Area: Challenges for a New Decade*, edited by A. Curaj, Ligia Deca, Remus Pricopie, 545–565. https://doi.org/10.1007/978-3-030-56316-5_34.
- De Certeau, Michel. *The Practice of Everyday Life*. Berkley, Los Angeles, California: University of California Press, 1984.
- ELRA Catalogue. <https://catalogue.elra.info/en-us/>. Accessed in May 2024.
- European Charter for Regional or Minority Languages. <https://rm.coe.int/1680695175>. Accessed in May 2024.
- Fernández L'Hoeste, Héctor, and Juan Carlos Rodríguez. "Introduction: In and Out of Digital Humanities; Nations, Networks, and Practices in Latinx America." In *Digital Humanities in Latin America. Reframing Media, Technology, and Culture in Latin/o America*, edited by Héctor Fernández L'Hoeste and Juan Carlos Rodríguez, 1–38. Gainesville: University of Florida Press, 2020.
- Gallego Cuiñas, Ana. "Literature Seen Through Big Data and Artificial Intelligence: Key Concepts and Critical Challenges." In *Humanities and Big Data in Ibero-America*, 25–42. Berlin: De Gruyter, 2024.

- Galleron, Ioana, and Fatiha Idmhand. "Why Go from Texts to Data, or The Digital Humanities as A Critique of the Humanities." *Word and Text. A Journal of Literary Studies and Linguistics* 10 (2020): 53–69.
- Hirsch, Brett D. (ed.) *Digital Humanities Pedagogy. Practices, Principles and Politics*. Cambridge: Open Book Publishers. <https://www.openbookpublishers.com/books/10.11647/obp.0024>.
- Lach, Pamela R., and Jessica Pressman. "Digital Infrastructures: People, Place, and Passion—a Case Study of San Diego State University." In *People, Practice, Power. Digital Humanities outside the Center*, edited by Anne McGrail, Angel David Nieves, and Siobhan Senier, 189–202. Minneapolis–London: University of Minnesota Press, 2021.
- Malazita, James. "Epistemic Infrastructure, the Instrumental Turn, and the Digital Humanities." In *People, Practice, Power. Digital Humanities outside the Center*, edited by Anne McGrail, Angel David Nieves, and Siobhan Senier, 3–24. Minneapolis-London: University of Minnesota Press, 2021.
- McCarty, Willard. "Getting There from Here: Remembering the Future or Digital Humanities." *Literary and Linguistic Computing* 29, no. 3 (2014): 283–306.
- McGrail, Anne, Angel David Nieves, and Siobhan Senier. "Introduction." In *People, Practice, Power. Digital Humanities Outside the Center*, edited by Anne McGrail, Angel David Nieves, and Siobhan Senier, VII–XXII. Minneapolis-London: University of Minnesota Press, 2021.
- McPherson, Tara. "Why Are the Digital Humanities So White? or, Thinking the Histories of Race and Computation." In *Debates in the Digital Humanities*, edited by M. K. Gold, 139–60. Minneapolis: University of Minnesota Press, 2012.
- Melero, Maite, Sakriani Sakti, and Claudia Soria, eds. "Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages @ LREC-COLING 2024." 2024. <https://aclanthology.org/2024.sigul-1.pdf>.
- Păiș, Vasile, and Dan Tufiș. *European Language Equality (ELE). Report on the Romanian Language*. 2022. https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D1_29__Language_Report_Romanian_.pdf. Accessed in June 2024.
- Patraș, Roxana. "Digital Romania: A Novelty Effect?" *Hermeneia* 29 (2022): 127–148.
- Pitman, Thea, and Claire Taylor. "Where is the ML in DH? And Where is the DH in ML? The Relationship between Modern Languages and Digital Humanities, and an Argument for a Critical DHML." *DHQ* 11, no. 1 (2017).
- Ray Murray, Padmini. "Locating the Digital Humanities in India: Internationalisation, Globalisation, and Localisation." Plenary presentation at Digital Diversity 2015, University of Alberta, Edmonton, Canada, May 9, 2015.
- Report of Special Interest Group on Under-resourced Languages [SIGUL]. 2019–2020. https://www.elra.info/media/filer_public/2020/11/03/2019-20.pdf.
- Rieder, Bernhard, and Theo Röhle. "Digital Methods: Five Challenges." In *Understanding Digital Humanities*, edited by David M. Berry, 67–85. London: Palgrave Macmillan, 2012.
- Risam, Roopika, and Alex Gil. "Introduction: The Questions of Minimal Computing." *DHQ: Digital Humanities Quarterly* 16, no. 2 (2022).

- Risam, Roopika. *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Evanston: Northwestern University Press, 2018.
- Schöch, Christof, *et al.* "Creating the European Literary Text Collection (ELTeC): Challenges and Perspectives." *Modern Languages Open* (2021). DOI: <http://doi.org/10.3828/mlo.v0i0.364>.
- Svensson, Patrik. *Big Digital Humanities. Imagining a Meeting Place for the Humanities and the Digital*. Ann Arbor, Mich: University of Michigan Press, 2016.
- Ștefănescu, Bogdan. "The Complicated Selves of Transcolonialism: The Triangulation of Identities in the Alternative Peripheries of Global Post/Colonialism". *Metacritic Journal for Comparative Studies and Theory* 8.1 (2022). DOI: <https://doi.org/10.24193/mjcst.2022.13.04>.
- Terras, Melissa. "Peering inside the Big Tent: Digital Humanities and the Crisis of Inclusion." Melissa Terras' Blog, July 26, 2011. peering-inside-big-tent-digital.html.
- Thompson Klein, Julie. *Interdisciplining Digital Humanities: Boundary Work in an Emergent Field*. Ann Arbor: University of Michigan Press, 2015.
- Weingart, S. B., & Eichmann-Kalwara, N. "What's Under the Big Tent?: A Study of ADHO Conference Abstracts." *Digital Studies/le Champ Numérique* 7, no. 1 (2017): 6. DOI: <http://doi.org/10.16995/dscn.284>.