



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 TUFÎȘ	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

Exercises in Literary Geography: where hajduks roam in the second half of the nineteenth century¹

Alexandra OLTEANU

Geocriticism and spatial criticism

Geocriticism and spatial criticism are two distinct but interconnected approaches used within the fields of literary and cultural studies. Both approaches focus on the relationship between literature, space and place, but they differ in their theoretical foundations and emphases. Geocriticism, as a term, was first coined by French geographer Bertrand Westphal in the early 2000^s (Westphal 2011, 14). Geocritics examine the ways in which literature engages with space, and they emphasize the geographical imagination and the representation of landscapes in literary texts. One of the foundational ideas of geocriticism is the notion of *topophilia*, introduced by geographer Yi-Fu Tuan; it refers to the love of place (Tuan 1990, 12). In Tuan's examination of topophilia, he elucidates the diverse methods through which individuals formulate a sense of sites and establish connections with their surroundings. This involves the cultivation of positive feelings and attitudes towards specific spaces. Tuan posits that these affective bonds play an essential role in shaping human perceptions of and engagement with the environment. Topophilia, in its full comprehensive scope, extends beyond the physical attributes of a location to encompass cultural, historical, and the symbolic connotations attached to it. Tuan underscores that these emotional affiliations contribute significantly to the multifaceted nature of human encounters with the environment. An appreciation of

¹ 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.

topophilia facilitates a profound exploration of the subjective and emotional dimensions inherent in human-landscape relationships, thereby offering insights into the intricate interplay between individuals and their surroundings:

The existence of a kinesthetic relationship between certain physical forms and human feelings is implied in the verbs we use to describe them: for example, mountain peaks and man-made spires “soar,” ocean waves as well as architectural domes “swell” [...] Moreover, architectural forms appear to influence our impression of size—of the way space expands and contracts to a degree that natural landforms rarely give. (Tuan 1990, 28)

The narrative of this work is infused with the concept of topophilia, which encapsulates the emotional connection between individuals and their surroundings. The exploration spans a diverse array of experiences, ranging from the profound allure of a cathedral to the vibrant hustle and colors of a market stall, the majestic presence of the Alps, the rhythmic embrace of waves on a beach, and the nuanced composition of planes within a landscape garden. Yi-Fu Tuan’s discerning observations extend to a myriad of worldly phenomena, from the massive to the microscopic that captivate his affectionate scrutiny (Tuan 1990, 78). Nevertheless, the scope of his inquiry transcends topophilia alone. It encompasses a broader spectrum of perceptions, attitudes, and values, intertwined with intricate depictions of landscapes and townscape from various time periods. The book delves into topics such as the distinctions between human sensory experiences and those of other species, the systems we use to classify the world around us, our views of the afterlife and the future, the ways in which landscape and scenery are defined, and the dynamic connection between symbols and realities. Themes related to contrasts between the primitive and the modern, between wilderness and paradise, and between rural and urban continuously emerge, submerge, and resurface, akin to playful seals in an archipelagic sea (Tuan 1990, 82).

Geocritics argue that literature not only reflects real geographical spaces but also constructs and influences our perceptions of these spaces. Prominent geocritics,

such as Edward Said in his work *Culture and Imperialism*, analyze the role of literature in shaping cultural perceptions of colonial spaces. Said's postcolonial geocriticism explores how literature contributes to the creation of geographical imaginaries that influence power dynamics and identity formation (Said 1994, 18). Dane Kennedy's article "Imperial History and Postcolonial Theory" plays a significant role in the ongoing discourse about the relationship between literary criticism and historical studies, specifically within the context of the interaction between "colonial discourse analysis" and imperial historiography. His balanced intervention thoughtfully analyzes the shortcomings found in the approaches of (post)Saidian literary critics regarding their engagement with historical evidence and historiographical methods. Additionally, Kennedy laments the widespread missed opportunities among imperial historians to fully embrace what is implied by the emergence of postcolonial cultural studies (Kennedy 1996, 352).

Other theorists like Franco Moretti focus on spatial patterns in literature, using mapping and distant reading techniques to uncover trends and relationships across different texts. Moretti's methodology enables him to discern overarching literary trends and patterns spanning temporal and spatial dimensions. Through this approach, Moretti endeavors to offer a macroscopic perspective on literary history, unveiling concealed structures that may elude traditional analytical methods. The ideological ramifications of Moretti's scholarly interventions manifest in the tension between macroscopic analysis and individual agency. Detractors contend that an exclusive focus on statistical trends runs the risk of disregarding the intricacies and complexities inherent in individual texts and authors. This threatens to homogenize literature, potentially neglecting the distinct cultural and historical contexts intrinsic to each work. Moreover, the reliance on quantitative data prompts questions about the potential risk of reducing literature to mere data points, thereby overlooking the subjective and qualitative facets of artistic expression. Comparing Moretti's macroscopic approach and Said's historiographical viewpoint reveals both converging and diverging features. In *Orientalism*, Said critiques Eurocentric perspectives within Western scholarship, advocating for a more nuanced understanding of cultural texts (Said 1994, 55). Said's emphasis on understanding

power dynamics, imperialism, and the construction of otherness introduces a layer of complexity to the discourse regarding literary spaces. Applying Said's critique of orientalism to Moretti's work should raise relevant questions about whether distant reading inadvertently perpetuates a form of occidentalism by reducing diverse literary traditions to universalized patterns. Scholars have engaged in debates surrounding whether Moretti's methodologies risk perpetuating the very hegemonic structures he aims to expose.

Hajduks and the space

As compared to geocriticism, spatial criticism—sometimes referred to as literary cartography—has roots in human geography and critical theory. The spatial turn in literary studies gained momentum in the late twentieth century, with theorists like Henri Lefebvre and Michel de Certeau emphasizing the social and cultural dimensions of space. Spatial critics argue that space is not merely a backdrop for literary events but actually a dynamic force that shapes and is shaped by human actions. Henri Lefebvre's *The Production of Space* is another work that examines the social construction of space and its impact on everyday life. Lefebvre introduces the idea of the *spatial triad* to understand the connection between perceived space, conceived space, and lived space (Lefebvre 1991, 22). Spatial critics draw on Lefebvre's concepts to argue how literature reflects and challenges societal spatial practices:

From the social standpoint, space has a dual “nature” and (in any given society) a dual general “existence.” On the one hand, one (i.e. each member of the society under consideration) relates oneself to space, situates oneself in space. One confronts both immediacy and an objectivity of one's own. One places oneself at the centre, designates oneself, measures oneself and uses oneself as a measure. (Lefebvre 1991, 175)

Lefebvre asserts that space transcends mere physicality, and he posits that it is elaborately molded by societal behaviors, ideologies, and power dynamics. He

introduces the conceptual distinctions of *conceived space* (representations of space), *perceived space* (experiential dimensions of space), and *lived space* (the quotidian, tangible reality). Lefebvre contends that generating space constitutes a social phenomenon that entails both tangible and symbolic. Furthermore, he critiques the oversimplified view of space as solely governed by economic or technical considerations. This theory has left an indelible mark on the fields of urban studies, geography, and cultural studies, challenging established paradigms surrounding space and furnishing a theoretical framework for scrutinizing the social construction of the environment. Michel de Certeau, in *The Practice of Everyday Life*, further contributes to spatial criticism by introducing the concepts of space strategies and space tactics (De Certeau 1984, 95). De Certeau underscores the inventive nature of routine activities, highlighting how individuals, through their ordinary practices, establish realms of agency within prevailing systems. Of particular significance is the chapter dedicated to the concept of *walking in the city*, wherein de Certeau explores how pedestrians, by charting their unique paths and narratives, actively defy the predetermined order imposed by urban planning:

Unlike Rome, New York has never learned the art of growing old by playing on all its pasts. Its present invents itself, from hour to hour, in the act of throwing away its previous accomplishments and challenging the future. A city composed of paroxysmal places in monumental reliefs. The spectator can read in it a universe that is constantly exploding. In it are inscribed the architectural figures of the *coincidatio oppositorum* formerly drawn in miniatures and mystical textures. On this stage of concrete, steel and glass, cut out between two oceans (the Atlantic and the American) by a frigid body of water, the tallest letters in the world compose a gigantic rhetoric of excess in both expenditure and production. (De Certeau 1984, 102)

Lefebvre and de Certeau present mutually reinforcing viewpoints on the interaction between space and everyday life. Lefebvre builds a macro-level theory, delving into the social production of space, while de Certeau creates a micro-level exploration, examining how individuals engage with and challenge spatial frameworks in their daily routines. The influence of both works is significant,

extending to cultural theory, urban studies, and broader understandings of the experiential dimensions of space. Spatial critics apply these ideas to literature, examining how fictional characters navigate and resist spatial constraints. Franco Moretti's spatial criticism entails examining how literature portrays space and how these representations contribute to understandings of social and cultural phenomena. The *Atlas of the European Novel 1800–1900* explores the geographic settings of European novels to glean insights into the complex connections between literature and geography:

The young hero of *The Captain's Daughter*, for instance, who dreams of going to Petersburg, is promptly dispatched in the opposite direction, to the eastern periphery of the Czarist empire. In *Waverley*, Charles Stewart never completes his march towards London: he lands in the North-West of Scotland, raises the Standard of Rebellion in the middle of the Highlands, crosses the Highland line, reaches Edinburgh, crosses the Anglo-Scottish border, reaches Derby—and then stops. He stops, in other words, exactly where Austen's England begins (Pemberley, the northernmost locality “seen” in her novels, is also in Derbyshire). And that Scott's world should end exactly where Austen's begins, and Austen's end where Scott's begins. (Moretti 1999, 33)

Franco Moretti's examination of the spatial trajectories of characters in *The Captain's Daughter* and *Waverley* provides a perceptive analytical framework for investigating the dynamics of literary geography in British novels. In both instances, intentional narrative shifts redirect the characters away from their original aspirations, guiding them towards the peripheries or boundaries of their respective empires. The juxtaposition of the protagonist's aspiration to reach Petersburg in *The Captain's Daughter* with his journey to the eastern periphery of the Czarist empire not only signifies a divergence in the narrative but also implies a broader commentary on the sociopolitical milieu of the period (Moretti 1999, 32). Moretti emphasizes the strategic deployment of geographical settings as a tool for examining and commenting on power dynamics, emphasizing the peripheries as crucial sites for narrative exploration. Likewise, in *Waverley*, the decision to interrupt Charles Stewart's march towards London reflects a conscious selection of geographical scope.

Moretti posits that this purposeful interruption corresponds to the geographical boundaries of Austen's England, particularly spotlighting Pemberley in Derbyshire as the most northerly setting in her novels (Moretti 1999, 33).

The observation that Scott's fictional world concludes precisely where Austen's commences, and vice versa, prompts contemplation of the symbolic significance embedded in these geographical intersections. This observation raises questions about the thematic implications of the chosen locations, indicating a nuanced connection between distinct literary realms. Moretti's analysis acts as a catalyst, prompting readers to investigate the function of literary geography not solely as a setting but also as a deliberate narrative device that shapes the nuances of the fictional worlds meticulously crafted by these authors. The essential inquiry that instigated this investigation was whether Moretti's concepts pertaining to the functions of peripheral spaces and the thematic implications of locations could be used in case studies within Romanian literature. The HAIRO corpus serves as the primary source of material for such an examination (Patras *et al.* 2020).

In the genre of Romanian Hajduk novels, there exists a profound correlation between character and its geographic space. The Hajduk's ability to traverse various landscapes is grounded in the exploration of an extensive territory characterized by diverse topographies. Elements such as evading authority, executing remarkable escapes, and cultivating a personal connection with a nurturing natural environment are portrayed in both the textual narrative and the series of illustrations integrated into the novels.

The illustrations in *Figures 1* and *2* depict a profound correlation between the central character of the Hajduk novel and the spatial context in which the narrative unfolds. The geographic settings of the Hajduks' exploits are intentionally chosen, serving to facilitate both the mobility of the noble outlaws and the sensational nature of their actions. The "engravings"/ illustrations capture instances of remarkable escapes, wherein the courage of the Hajduk Codreanu is heightened by spatial considerations. The act of leaping from a lofty cliff and vaulting over a tall wall—executing this feat amid the intense scrutiny of adversaries—reflects integral elements of a narrative relevant to literary geography.

I analyze the Hajduk novel as a hybrid subgenre that combines elements of historical, social, and adventure novels. These literary features reveal profound shifts in historical and cultural perceptions of space. The Hajduk novel illustrates transformative alterations in how individuals across diverse historical eras and cultures have conceived the world imaginatively. Hajduks establish a natural, feudal society in opposition to official authority, choosing the protective refuge of nature, specifically the forest, as the territory for shielding themselves from adversaries who were pursuing them. The city and the palace emerge as danger zones, where conspiracies against the Hajduks and the oppressed populace they safeguard are devised by corrupt rulers. The intrigues within the Hajduk novel attest that spatiality transcends the fact of being a mere framework or backdrop for the Hajduk's rebellion; rather, the space assumes a central role in the narrative's evolution. A potential explanation for the absence of the "great city" in the Hajduk novel is provided by Dinu Pillat in a commentary on an article authored by Al. A. Philippide. Philippide establishes a correlation between the scarcity of the adventure subgenre in Romania and the absence of native metropolises capable of serving as settings for sequences of unforeseen and captivating events:

Al. A. Philippide, the only Romanian writer who questioned why the adventure novel was almost non-existent in our literary landscape, attributes this phenomenon to the absence of the "great city," a true "inextricable labyrinth" that "offers at every step unforeseen things, unexpected encounters, strange coincidences, infinite possibilities of disappearance, hard-to-discover hiding places, and numerous and varied characters". The idea is advanced that it is challenging to conceive of an adventure novel unfolding anywhere other than in the "human forest" of a sprawling metropolis. The more confined the setting (provincial town, neighborhood, or a single house), the more the adventure easily risks—according to the author's opinion—becoming internal, and the novel turning into an ordinary one.² (Pillat, 270) [*author's translation*]

² Original quote: "Al. A. Philippide, singurul scriitor român care s-a întrebat de ce romanul de aventuri nu a fost aproape deloc cultivat la noi, găsește cauza în lipsa."

Al. A. Philippide's perspective from "The Adventure Novel and Romanian Society" finds an advocate in Dinu Pillat, a connoisseur of sensational novel subgenres from the nineteenth century. This is because the inventiveness of Romanian writers seemed to be conditioned, on the one hand, by scenarios and strategies mediated through translations and, on the other hand, by their ability to adapt elements drawn from foreign cultural capital to the local specificities.

Fredric Jameson's assertion that "our daily life, our psychic experiences, our cultural languages are dominated by categories of space rather than by categories of time" encapsulates a core concept in his analysis of postmodernism and its cultural implications (Jameson 1990, 15). As postmodern scholars, emphasizing spatial considerations, we often engage in an analysis aimed at identifying spatial categories. Is it possible that the prevalence of spatial categories might be a distortion effected by the postmodern paradigm perspective? If this is the case, a focused examination of space-related concepts (applied to texts from any historical period) would, by logical extension, be considered a postmodern interpretation, aligning with Jameson's framework. This shift signals a broader cultural transformation wherein traditional linear narratives of time are supplanted by a focus on spatial configurations and relationships. The ascendancy of spatial categories implies a reorientation in our perception and navigation of the world. Everyday life, psychological experiences, and cultural expressions increasingly reflect the influence of spatial considerations, underscoring the importance of both physical and symbolic spaces in shaping our contemporary reality. This perspective aligns with the fragmentation and decentering characteristic of postmodernism, where an emphasis on space contributes to a sense of dislocation and disorientation. Moreover, in a society marked by rapid globalization, commodification, and the proliferation of images and signs, the spatial dimensions of culture become paramount. Jameson's observation underscores a fundamental shift in our cultural logic, indicating a departure from the primacy of temporal narratives to a terrain where spatial categories play a central role in shaping our experiences, identities, and cultural expressions.

While geocriticism and spatial criticism share a focus on the relationship between literature and space, they differ in their theoretical foundations and

methodological approaches. Geocriticism tends to emphasize the geographical imagination and the representation of landscapes, often using mapping techniques, while spatial criticism draws on broader theories of space from human geography and critical theory. Through the exploration of the contributions of influential theorists like Edward Said, Franco Moretti, Henri Lefebvre, and Michel de Certeau, literary studies can gain insights into the ways literature both influences and is influenced by the spatial dimensions of human experience. These methodologies enhance our understanding of the nuanced connection between textual and spatial elements in literary and cultural analysis, contributing to a deeper understanding of these complex dynamics. Geocriticism and spatial criticism concur on the notion that space not only generates the narrative elements but also is reciprocally reshaped by the narrative.

The “spatial turn” in literary studies

Both geocriticism and spatial criticism provide unique perspectives for examining the interconnection between space and literature. A relevant work in terms of understanding the transformative impact of spatial thinking across disciplines is Barney Warf and Santa Arias’s *The Spatial Turn: Interdisciplinary Perspectives*. This work stands as a cornerstone in the examination of spatial thinking across disciplines. Published in 2009, the book captures the zeitgeist of a paradigm shift towards spatial considerations in literary studies. Its interdisciplinary approach not only encompasses geography and sociology but also extends to the realms of literary studies. The authors elucidate the theoretical foundations of the spatial turn, stressing the need to rethink conventional understandings of space and place (Warf and Arias 2009). This interdisciplinary perspective aligns with geocriticism, encouraging the blending of methodologies as a way to enrich analyses of literary spaces by drawing on insights from geography, anthropology, and cultural studies.

The geographical imagination, a key concept within geocriticism, aligns closely with Warf and Arias’s call to reevaluate spatial paradigms. Other theorists, such as Franco Moretti, leverage mapping techniques to unravel spatial patterns in literature, providing a tangible link between the imaginative landscapes crafted by authors and

the geographical spaces they represent. Warf and Arias's emphasis on the multidisciplinary nature of the spatial turn resonates with the geocritical approach, which advocates for exploring literary spaces in comprehensive ways by integrating insights from various disciplines. Concurrently, spatial criticism, or literary cartography, draws inspiration from theorists like Lefebvre and de Certeau. Lefebvre's spatial triad and de Certeau's notions of strategies and tactics become invaluable tools for spatial critics exploring how literature reflects and challenges societal spatial practices (Lefebvre 1991; de Certeau 1984). By incorporating *The Spatial Turn* into the spatial criticism framework, research gains a more comprehensive understanding of how literature participates in the construction and contestation of spatial narratives.

Robert Tally's view of space as an active presence, which not only informs but also potentially determines various narrative elements, underscores the critical importance of spatial considerations in the realm of literary analysis. Tally's viewpoint aligns seamlessly with the theoretical framework of literary geography, accentuating the pivotal role that both the physical and symbolic dimensions of space play in shaping the profound substance of a narrative. The observation that space informs the narrative implies that the geographical and environmental context is fundamental to understanding the dynamics inherent in a story. Furthermore, Tally's claim that space may have significant impact on a narration underscores the likely that space will have a significant impact on the narration. This implies that geographic setting, cultural milieu, and physical surroundings may impose constraints or present opportunities that profoundly shape characters' actions and the development of the plot.

From raw data to a “Hajduk” literary space

The HAI-RO: *Hajduk Novels in Romania During the Long Nineteenth Century: Digital Edition and Corpus Analysis Assisted by Computational Tools* project emerged under the auspices of the “Brâncuși” Program for Integrated Actions, fostering a dynamic collaboration between the “Alexandru Ioan Cuza” University of

Iași and Université Paris 3 Sorbonne Nouvelle. Initiated by the project’s principal investigators, Roxana Patraș and Ioana Galleron, the collaboration began as an experimental venture focusing on a series of TEI-XML novels for ELTeC (the “European Literary Text Collection” developed within the COST Action 16204: *Distant Reading for European Literary History*). This endeavor led to the formation of a dedicated French-Romanian team comprising Camelia Grădinaru, Ioana Lionte, Alexandra Olteanu, Lucreția Pascariu, Chiara Mainardi, and Ofra Lévy. In essence, the project addressed the deficiencies in resources and tools specifically tailored for Romanian Computational Literary Studies. Its research trajectory centered on characterizing the distinctive features inherent to “the Hajduk literature” by adapting corpus linguistics instruments to the nuances of literary texts printed in non-standardized Romanian. This endeavor was complemented by an examination of the external features developed throughout the historical evolution of a literary genre commonly categorized as “national.” The primary outcome of the project is the creation of a Romanian literary corpus spanning from 1850 to 1950, meticulously aligned with the XML-TEI international standard of digital editing and enriched with semantic annotations. This corpus is accessible through an Open Access system, hosted on the Nakala platform (based in France), and documented on HAL-SHS.³ This collection of Hajduk novels provides a rich source for exploring a dynamic literary space abundant in representations related to the motivations and circumstances behind characters’ actions, sentimental embellishments of space, and the thematic patterns highlighted by scholars such as Tuan, Moretti, Warf, and Arias.

The HAI-RO collection constitutes a heritage corpus designed to delve into a literary tradition. The research initiative originated within genre theory, with a particular emphasis on examining the generic cohesion of the Hajduk subgenre. The primary aim was to scrutinize a phenomenon characterized by imitation and cultural

³ HAL-SHS is an open access archive platform operated by the Centre pour la Communication Scientifique Directe (CCSD), a French national institution dedicated to providing free access to scholarly publications and research outputs in the humanities and social sciences. It serves as a repository for research materials, including articles, conference papers, and datasets, allowing researchers to share and access valuable academic resources worldwide.

transmission. This investigation was carried out through an analysis of the Hajduk hero's mobility; this analysis involved annotating spatial elements and examining narrative structures in mystery novels.

In the initial phase of mapping the literary landscape of the Hajduk novel, the HAI-RO project sought to establish a potential method for spatial annotation. This led to the identification of the following spatial categories: 1. places (*his village, poor wretched country*); 2. zones and regions (*the outskirts of Bucharest, in front of the house*); 3. imagined or archetypal spaces (topos) (*the other world, the universe, far from any village*); 4. elements related to an architectural plan (*at the church door, in the kitchen*); 5. spaces defined by body-related (e. g. clothing) elements (*near their breasts, from his pocket, on the head*); 6. transportation/access routes (streets, bridges); 7. vehicles (*in a carriage pulled by four horses*). In an effort to standardize the diverse spatial elements, a decision to extract toponyms was made, as a measure that increased the objectivity of the analysis and mitigated potential annotative ambiguities (Galleron *et al.* 2021). The optimal method for extracting toponyms was the use of UDPipeline (Straka *et al.* 2016). Dariah Geobrowser was used to generate a map of the Hajduk literary space.⁴ UDPIPE (short for *UD Parsing Pipeline*) is a natural language processing tool developed by Straka *et al.* in 2016. Dariah Geobrowser is a mapping tool used in this study to visualize the spatial data.

The identification and extraction of proper nouns hold significance within the realm of natural language processing (NLP), particularly in tasks such as information retrieval, named entity recognition (NER),⁵ and sentiment analysis. The challenge presented by Hajduk novels, characterized by a diverse array of character names, locations, and named entities, underscores the relevance of using UDPIPE for this

⁴ I would like to express my heartfelt appreciation to my colleagues Ioana Lionte and Lucretia Pascariu for their dedicated efforts in collecting information about geolocations. Their commitment and diligence significantly contributed to the success of this research.

⁵ Named Entity Recognition (NER) is a natural language processing task that involves identifying and categorizing entities within a text into predefined categories such as person names, organization names, locations, dates, and numerical expressions. NER is commonly used in information extraction, text mining, and other NLP applications to identify and extract important information from unstructured text data.

task. Developed by the Universal Dependencies project, UDPipe is an open-source library designed to process raw text across various languages. Its capabilities, including tokenization, lemmatization, syntactic parsing, part-of-speech tagging, and named entity recognition (NER), empower users to perform comprehensive linguistic analysis and information extraction tasks. While not specifically designed to extract toponyms (proper names) from texts, UDPipe can be used for this task, which involves leveraging the tool's tokenization and part-of-speech tagging capabilities for nouns and then generating CoNLL-U files.

Using UD Pipeline to extract toponyms and construct a map of the literary space within Hajduk novels involved multiple steps. Firstly, a careful selection of the novels was made, and these texts were then converted to .txt format and subsequently uploaded to the UDPipe system (*Figure 3*). The next step was tokenization, which involved breaking down the raw text into segmented units, known as tokens. Pre-trained models were used for analyzing the text and breaking it down into meaningful components. Part-of-speech (POS) tagging was then carried out, with UDPipe (as shown in *Figure 4*) used to assign grammatical categories (such as nouns, verbs, or adjectives) to each token. This process plays a crucial role in identifying proper nouns, which frequently denote names of characters, locations, or historical events. By employing techniques such as part-of-speech tagging, tokenization, and named entity recognition, researchers can effectively pinpoint these elements within the text. Preliminary results were obtained from the CoNLL-U file, from which all proper nouns were subsequently extracted (*Figure 4*).

The ensuing step involved manually curating nouns from the list of proper nouns. Once proper nouns that did not meet the criteria for toponyms had been eliminated, insights into the dynamic nature of the Hajduk subgenre unexpectedly came to light. The compilation of toponyms consisted of 597 entries, covering various types of Romanian settlements (villages, cities) as well as geographical regions with European, Asian, and American characteristics. This step revealed unexpected insights into the dynamic nature of the Hajduk novel.

The final list of toponyms was uploaded onto the Dariah Geobrowser, and any missing geographic coordinates were supplemented to ensure the precise location of

toponyms (*Figure 6*). The process of manually inputting geographical coordinates proves crucial to ensuring the quality of the data displayed on the prospective map, because this process contributes to eliminating ambiguities in those instances where toponyms are common to multiple geographical spaces, and to enhancing precision when it comes to locating both the sites where the Hajduks' adventures unfold and the spaces to which these sites refer.

The processes described above generated a map that illustrates the spatial features of Hajduk novels. UDPipe's adaptability is effective in extracting proper nouns from these novels, where efficient tokenization is critical to isolating individual words and facilitating analyses. UDPipe is well-suited to the context-rich textual content of a Hajduk corpus. Proper noun extraction, which is an essential preprocessing step in various NLP applications, contributes to the understanding of cultural and historical narratives intertwined with the literary space.

The Hajduks' literary space

The map of the Hajduks' literary space is revealing not only in geographical terms but also thematically and historically. The most common toponyms are those located within the Romanian space, while those scattered throughout the world appear less frequently, despite being potentially meaningful in a regional novel subgenre (*Figure 7*). The frequency of toponyms that mark specific geographic areas delineates genuine Hajduk hubs clustered around the cities of Bucharest, Tecuci, Bacău, Pașcani, and aligning smoothly along the sub-Carpathian arch (*Figure 8*). Considering the infrequent travels of the Hajduks beyond the boundaries of the principalities, international toponyms do not function as active spaces for the development of intrigue; instead, they play a passive role within the characters' discourses, serving as venues for exploring ideas and for supporting the characters' impulse to reflect. The authors might have incorporated various international toponyms to evoke a broader cultural or geopolitical context, providing a sense of interconnectedness with the wider world. The inclusion of global toponyms could be a deliberate artistic choice,

allowing the author to create a more expansive and imaginative literary universe, transcending strict geographical boundaries.

The international toponyms might carry symbolic or allegorical significance, representing themes, ideas, or conflicts beyond the immediate physical locations, adding depth to the narrative. In *Figure 8*, which illustrates the distribution of Romanian toponyms, it is evident that assertions regarding the existence of Hajduk camps in mountainous regions are not confirmed. On the contrary, this hypothesis should be taken with a pinch of salt. Hajduk novels suggest that the mobility of rural outlaws is attributed not to rugged terrain but to the accessibility of commercial routes and postal roads. Certainly, the figure of the Hajduk would not have instilled such fear if he had not possessed the authority to regulate the movement of people and goods throughout the principalities of Wallachia and Moldavia.

Juxtaposing *Figure 8*, which illustrates the literary realm of the Hajduks, with *Figure 9*, depicting the postal roads in Wallachia and Moldavia, reveals the Hajduks' familiarity with the pathways along which people and goods traveled within the principalities (*Figure 9*). The significant flow of traffic along these routes enabled Hajduks to use effective disguises, cross the territory inconspicuously, and gain quick access to information about potential targets and valuable assets. The rearrangement of Hajduk hubs in this revealing literary space reflects the political shifts of the latter half of the nineteenth century. The Hajduk novel underscores deep-seated connections with historical events, emphasizing the transfer of political power from Iași to Bucharest. Symbolically, Iași yields its authority, giving way to Bucharest.

Conclusions

The theoretical insights brought by geocriticism and spatial criticism methodologies, in conjunction with digital tools, provide a deep understanding of the Hajduks' literary space. The geographical choices of the authors are revealed through the strategic placement of toponyms on the map, and the extraction of toponyms was facilitated by the use of UDPipeline. This digital tool enabled the systematic extraction and analysis of toponyms from the text, contributing to the comprehensive

mapping of the Hajduks' literary space. The map of the Hajduks' literary space challenges assumptions about the rural outlaw camps in mountainous regions. The alignment of the Hajduks' literary territory with commercial routes and postal roads raises questions about the contrasting and conflicting dynamics between, on the one hand, subversive clandestine conduits represented by Hajduk roads and, on the other hand, the formal authorized networks embodied by postal routes. This alignment encourages reflection about the connection between rebellion and governance, as well as about the dichotomy between secrecy and transparency, elucidating the manifold roles that pathways play overtime. One may contemplate the intersections or disparities between Hajduk roads, often symbolizing resistance and opposition, and the meticulously controlled and regulated postal routes. Examining potential overlaps in specific regions prompts an exploration of what the coexistence or divergence of these pathways implies for the circulation of people, ideas, and information. This line of inquiry seeks to unearth the historical ramifications of such interactions, enriching our understanding of the complex historical milieu in which those pathways operated. The toponyms identified in the text shed light on the correlation between geographic spaces and characters in Hajduk novels.

The groups of toponyms are notably concentrated around Hajduk hubs such as Bucharest, Tecuci, Bacău, and Pașcani. This reveals the symbiotic relationship between the movement of goods and people along postal road maps in Wallachia and Moldavia and the tactics of the Hajduks to control not only their own mobility but also that of others. The intersection of the literary realm of the Hajduks with the maps detailing the circulation of goods and people encapsulates the characters' ability to exert authority over information flow, navigate territories inconspicuously, and execute rapid maneuvers. This exchange of political and symbolic power from Iași to Bucharest underscores the deep-seated connections between the Hajduk novel and the historical events it portrays.

A subsequent research project that examines Hajduk novels, while expanding the spatial perspective to incorporate verb analysis, has the potential to reveal interesting connections between various physical configurations such as architecture, landscapes, roads, and geographical features, alongside human emotions and

narrative depictions, and narrative depictions. Drawing on Yi-Fu Tuan's perceptive insights into the kinesthetic interplay between emotions and physical surroundings, the exploration of the verbs used to portray Hajduk landscapes offers a rich avenue for investigation. Through meticulous analysis of the dynamic language used in these novels, we can observe the implicit ways in which emotions and perceptions associated with the depicted spaces are conveyed. For instance, verbs characterizing mountain peaks as "soaring" or ocean waves as "swelling" not only mirror the physical attributes of the landscapes but also evoke a sense of grandeur or power. Furthermore, this study may consider the impact that architectural forms have on perceptions of space—as suggested by Tuan—by analyzing verbs that delineate spatial expansion and contraction within Hajduk narratives. Such an inquiry promises to enrich our understanding of the interaction between language, spatial representation, and emotional resonance within the realm of Hajduk novels.

The Hajduk novels exhibit a unique conceptualization of spatial dynamics, distinguishing between active and passive spaces. Active spaces go beyond their traditional role as mere settings, influencing the narrative trajectory. These dynamic arenas serve almost as theatrical stages where Hajduk protagonists navigate, resist, and orchestrate their actions and challenges. In contrast, passive spaces dispersed globally play a crucial role in serving a distinct imaginative function within the Hajduk discourse. They predominantly act as contemplative venues that facilitate exploration of broader cultural or geopolitical contexts. The inclusion of global toponyms in the novels can be interpreted as an artistic decision that fosters the creation of an expansive and imaginative literary realm, enriching the narrative with broader cultural or geopolitical contexts. This dichotomy between active and passive spaces not only enhances the complexity of the Hajduk narrative but also encourages readers to delve into the symbolic, thematic, and geopolitical dimensions integrated into the spatial configurations of these novels. The Hajduk literary space is an intriguing construct, providing readers with a distinctive perspective from which to examine the intersection of literature, space, and history.

Annex 1: Figures

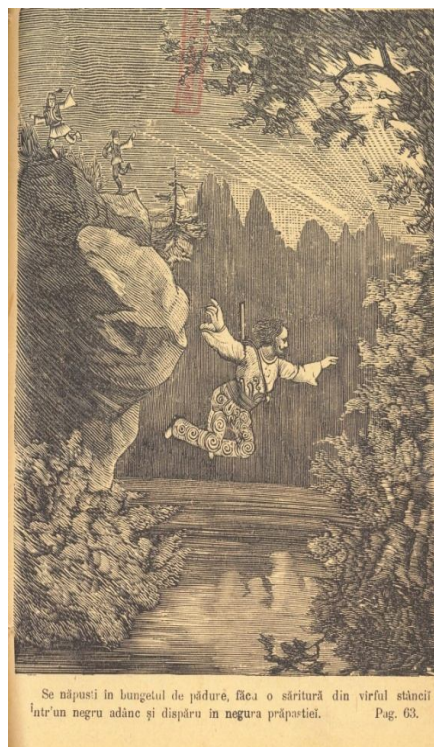


Figure 1: Sample1- illustrated page from a Hajduk novel



Figure 2: Sample2- illustrated page from a Hajduk novel

Service

The service is freely available for testing. Respect the CC BY-NC-SA licence of the models – **explicit written permission of the authors is required for any commercial exploitation of the system**. If you use the service, you agree that data obtained by us during such use can be used for further improvements of the systems at UFAL. All comments and reactions are welcome.

Model: ☐ UD 2.10 (description) ☒ UD 2.6 (description) ☐ Evalatin20 (description)

 romanian-rtt-ud-2.6-200830

Actions: ☒ Tag and Lemmatize ☒ Parse

▼ Advanced Options

Input Text

Output Text

Figure 3: Loading a Hajduk novel as an input file

Save Output File									
Id	Form	Lemma	UPosTag	XPosTag	Feats	Head	DepRel	Deps	Misc
# generator = UDPipe 2, https://lindat.mff.cuni.cz/services/udpipe									
# udpipe_model = romanian-rtt-ud-2.10-220711									
# udpipe_model_licence = CC BY-NC-SA									
# newdoc									
# newpar									
# sent_id = 1									
# text = Moldova în veacul al 17-lea Istoria ce vom povesti în această broşură, se petrece în a doua jumătate a veacului al şeppte-spre-decelea, cam pe la anul 1666, Moldovei în timpul acela, pentru ca faptele ce vom povesti să nu pară nici stranii nici imposibile.									
1	Moldova	Moldova	PROPN	Np	—	15	nsubj	—	TokenRange=0:7
2	în	în	ADP	Spsa	AdpType=Prep Case=Acc	3	case	—	TokenRange=8:10
3	veacul	veac	NOUN	Ncmsry	Case=Acc,Nom Definite=Def Gender=Masc Number=Sing	15	obl	—	TokenRange=11:17
4	al	al	DET	Tsms	Gender=Masc Number=Sing Poss=Yes PronType=Prs	5	det	—	TokenRange=18:20
5	17-lea	17-lea	NUM	Mc-p-d	Gender=Masc Number=Sing NumForm=Vword NumType=Ord	3	nummod	—	SpacesAfter=Vn TokenRange=21:27
6	Istoria	istorie	NOUN	Ncfsry	Case=Acc,Nom Definite=Def Gender=Fem Number=Sing	15	nsubj	—	TokenRange=29:36
7	ce	ce	PRON	Pw3--r	Case=Acc,Nom Person=3	9	obj	—	TokenRange=37:39

Figure 4: Results of POS tagging as an output file

Name	Address	Longitude	Latitude	GettyID	TimeStamp	TimeSpan begin	TimeSpan end
Absinia	Absinia	38.6521203	10.2116702				
Ada Kaleh	Ada Kaleh	28.650158	44.1993027				
Adjud	Adjud	27.1667	46.1167	1050000			
Adunați	Adunați	25.5993255	45.183573				
Africa	Africa	20.590230209764215	9.076845219425246				
Afumați	Afumați	26.2494303	44.524683				
Alina	Alina	23.7167	46.3167	1050002			
Albania	Albania	19.9999619	41.000028	2042117			
Alexandria	Alexandria	29.894378	31.199004	2002157			
Amaradia	Amaradia	23.3957565	45.0619549				
America	America	47.25	33.7333	2002460			
Anglia	Anglia	1.8542312130120822	54.77803614719509	7006106			
Ardeal	Ardeal	24.3740295	46.5971623				
Argeș	Argeș	24.8333	45	7002693			
Armășești	Armășești	24.0122869	45.0171702				
Asina	Asina	30.85937472728644	50.514493608956975				
Asina	Asina	44.0093	36.19099	7030345			
Atena	Atena	15.55	40.45	1043327			
Athos	Athos	24.3272583	40.1584206				
Austrelitz	Austrelitz	16.86746798780363	49.18110938160163				
Austria	Austria	13.84496310968734	47.533634513010865				
Babadag	Babadag	28.7833	44.8833	1050019			
Babel	Babel	25.4715813	45.4065778				
Băbeni	Băbeni	24.22675861915425	44.9700332				
Bablonia	Bablonia	26.9667	46.55	7004053			
Bacău	Bacău	44.4116729807046	33.56604907591689				
Bagdad	Bagdad	27.733333	47.133333	8711876			
Bahlui	Bahlui	14.25	41.3	7057200			
Baia	Baia de Aramă	22.8	45	1050023			
Băilești	Băilești	26.1111466	47.4802379				
Bulboacă	Bulboacă	27.7796997	46.5500131				

Figure 5: The final list of 597 toponyms

TGN search can be based on column Address or on column GettyID in your datasheet. Choose which one to use:

Address

The coordinates for the first place found are automatically added to the sheet if no existing entries are available. If there are multiple results, you may have to adjust the places manually!

- sets all places in the sheet with the highlighted address to the coordinates of the chosen place. This will overwrite existing entries!
- shows the chosen place in the Map selection and allows you to correct and refine coordinates or set unrecognized places.

Abisnia

No results, please try map selection. Map Set

Ada Kaleh

No results, please try map selection. Map Set

Adjud

Adjud: Vrancea | Romania | Europe | Wor Map Set

Afumați

No results, please try map selection. Map Set

Adunați

No results, please try map selection. Map Set

Alina

No results, please try map selection. Map Set

Map search

Search the Getty Thesaurus of Geographical Names, OpenStreetMap, or GeoNames to fill your map selection.

Atumati

Q TGN Q OSM Q GeoNames

Figure 6: The manual addition of geographical coordinates using Geobrowser



Figure 7: The world seen through Hajduks' words

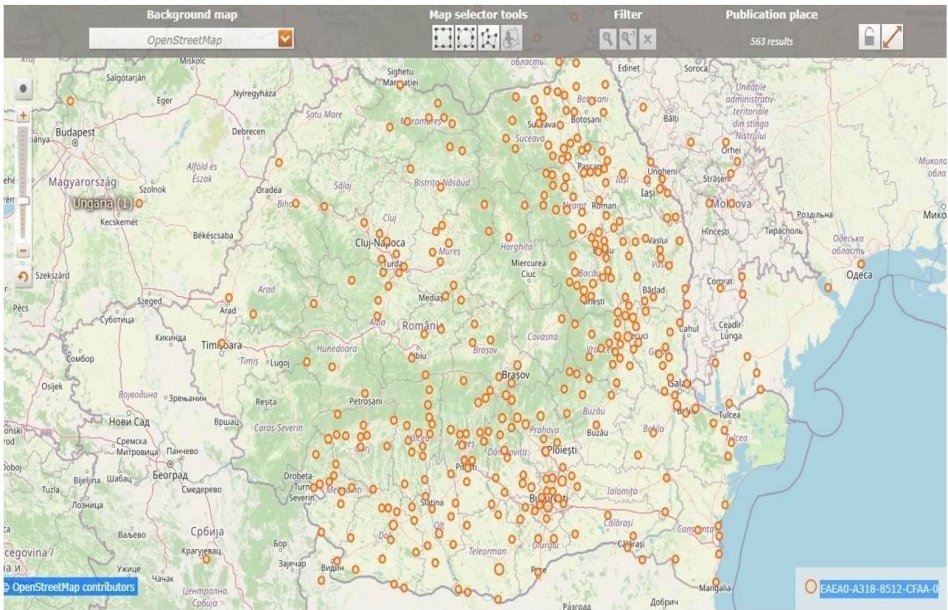


Figure 8: The crystallization of Hajduk hubs

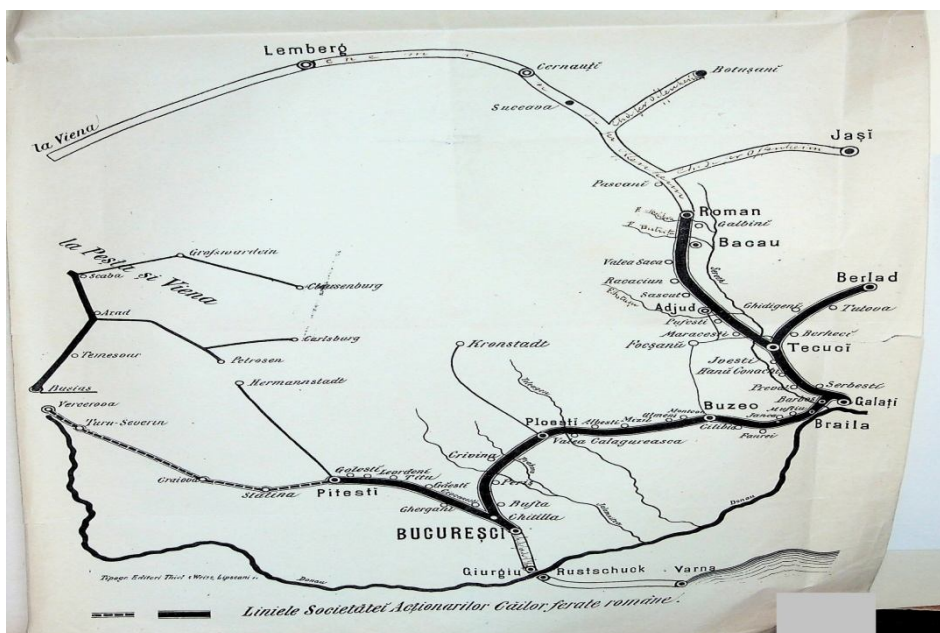


Figure 9: The map of postal roads in Wallachia

References

- De Certeau, Michel. 1984. *The Practice of Everyday Life*. Berkeley: University of California Press.
- Galleron, Ioana, Roxana Patraș, Camelia Grădinaru, and Frédérique Mélanie-Becquet. 2021. “À la recherche des haïdouks. Annoter les entités spatiales dans des romans roumains du XIXe siècle/Looking for Hajduks. Space Entities Annotation in Nineteenth-Century Romanian Novels.” *Humanités numériques* 3 (3). <https://doi.org/10.4000/revuehn.1399>.
- Jameson, Fredric. 1990. *Postmodernism, or the Cultural Logic of Late Capitalism*. Durham: Duke University Press.
- Kennedy, Dane. 1996. “Imperial History and Postcolonial Theory.” *The Journal of Imperial and Commonwealth History* 24 (3): 345–363.
- Lefebvre, Henri. 1991. *The Production of Space*. New Jersey: Wiley-Blackwell.
- Moretti, Franco. 1999. *Atlas of the European Novel 1800–1900*. London: Verso.
- Patraș, Roxana, Ioana Galleron, Lucreția Pascariu, Alexandra Olteanu, Ioana Lionte, and Camelia Grădinaru. 2020. *HAIRO: Corpus de Romane Haiducești (1850–1950)*. <https://www.nakala.fr/page/collection/11280/51a881d1>.
- Patraș, Roxana, Lucreția Pascariu, Alexandra Olteanu. 2023. “Patrimoniu și Inovație. HAIRO/ELTeC/Pop-Lite.” Paper presented at the SmartDiaspora conference, Timișoara.
- Said, Edward. 1994. *Culture and Imperialism*. New York: Vintage.
- Straka, Milan, Jan Hajič, and Jana Straková. 2016. “UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing.” In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia, May 2016.
- Tally, Robert T. (ed). 2014. *Literary Geographies: Spatiality, Representation and Narrative*. New York: Palgrave MacMillan.
- Tuan, Yi-Fu. 1990. *Topophilia: A Study of Environmental Perception, Attitudes, and Values*. New York: Columbia University Press.
- Warf, Barney, and Santa Arias. 2009. *The Spatial Turn: Interdisciplinary Perspectives*. London: Routledge.
- Westphal, Bertrand. 2011. *Geocriticism: Real and Fictional Spaces*. Translated by Robert T. Tally. New York: Palgrave MacMillan.